IDENTIFICATION, ESTIMATION AND VALIDATION OF SOME RIVER CATCHMENT MODELS WITH APPLICATION

By

M. HALIM SALEM, H.T. DORRAH AND M.A. ABO EL-ELA

CAIRO

October 1981

جامعة القاهرة - معهد ماساتشوستس للتكنولوجيا
برنامج التخطيط التكنولوجي
CAIRO UNIVERSITY - MASSACHUSETTS INSTITUTE OF TECHNOLOGY
TECHNOLOGICAL PLANNING PROGRAM

PREFACE

This report is directed towards the identification, estimation and validation of some physical data based river catchment models. Two general classes of models, with a variety of mathematical formulations and estimation methodologies, are presented. The first class is the linear stochastic difference equation models, while the second is the transfer function models selected using the minimum mean-square error criterion.

A case study of the Waki River catchment located near Lake Albert has been examined to demonstrate the applicability of the above models. Using the input precipitation over this catchment and the corresponding measured output discharge, it has become possible to digitally simulate the two proposed models and to scrutinize the main statistical characteristics of their output data sequence. The validity of the residual sequences generated by different structures of these models for the prespecified estimation conditions has also been investigated.

The salient features of the two best fitted linear stochastic difference equation model and noisy transfer function model have then been discussed in a comparative pattern in order to achieve a better representation for the Waki River catchment. As a general view, it is concluded that the application of linear stochastic difference equation models is pragmatic both for estimation and prediction of the given catchment output discharge.

ACKNOWLEDGEMENTS

The authors would like to express their appreciation to Professor Rafael Bras and Professor Peter S. Eagleson, of the Ralph M. Parsons Laboratory of Water Resources and Hydrodynamics of the M.I.T., for commenting on the original manuscript. Special thanks are extended to Professor M.F. Sakr of Cairo University, Dr. Hassan Ibrahim, Director of the Water Resources Development Institute, Ministry of Irrigation, and to Eng. A. Afifi of the Nile Control Department, Cairo, for all the help they provided.

This study was completely sponsored by the Cairo University/ M.I.T. Technological Planning Program which is funded by a contract between the Agency for International Development, United States Department of State and the M.I.T. Technology Adaptation Program. The views and opinions expressed in this report, however, are those of the authors and do not necessarily reflect those of the sponsors.

Many thanks are extended to the staff members of the Cairo University Program Office for their unlimited services and typing the report.

TABLE OF CONTENTS

PAGE

PAGE

PAGE

LIST OF SYMBOLS

$[y(\cdot)]$	A sequence of observed output data, measured for a given river catchment (in mm/day).
$\bar{y}$	Mean value of the observed output data sequence $[y(\cdot)]$.
σ_y	Standard deviation of the observed output data sequence $[y(\cdot)]$.
γ_y	Skewness coefficient of the observed output data sequence $[y(\cdot)]$.
$[x(\cdot)]$	A sequence of observed input data, measured for a given river catchment (in mm/day).
$\bar{x}$	Mean value of the observed input data sequence $[x(\cdot)]$.
σ_x	Standard deviation of the observed input data sequence $[x(\cdot)]$.
γ_x	Skewness coefficient of the observed input data sequence $[x(\cdot)]$.
N	Number of observations for either the input or the output data sequences.
$[\tilde{y}(\cdot)]$	A sequence of normalized output data.
$[\hat{\tilde{y}}(\cdot)]$	A sequence of estimated normalized output data.

$[\hat{y}(\cdot)]$	A sequence of estimated output data.
$[x_d(\cdot)]$	A sequence of dalayed input data.
$[\tilde{x}_d(\cdot)]$	A sequence of normalized delayed input data.
k	Lag at which either the input or the output data sequence is observed (day).
$[\hat{y}(k\|k-1)]$	An estimate for the output sequence at lag k given k-1 past observations.
P	Order of the autoregressive model.
q	Order of the moving average model.
ζ	An observation set for the output data sequence.
$\underline{U}$	Impulse response vector.
$\underline{U}_{LS}$	Unconstrained estimate for the impulse response vector $\underline{U}$.
K_o	Kernel length.
$P(\cdot)$	Probability of an event $(\cdot)$.
$E(\cdot)$	Expected value of an event $(\cdot)$.
L_i	Maximum Likelihood function.
J_i	One-step ahead prediction index.

$\forall$	For all values.
π	A constant equals to 3.14159
W_1	Frequency of variation of the normalized output data (radians/day).
τ	Time delay factor.
$\in$	Subset notation sign.
ε	Noise vector for the transfer function model.
$[w(\cdot)]$	A sequence of zero-mean Gaussian distribution random variable with unknown variance.
ψ_i	The ith weighting parameter for the linear filter subclass of stochastic model.
$[y_D(\cdot)]$	A sequence of deviated output data from its mean value $\bar{y}$.
t	Time interval in days.
ϕ	The ith weighting parameter for the autoregressive subclass of stochastic model.
θ_i	The ith weighting parameter for the moving average subclass of stochastic model.
E_i, H_i	The ith weighting parameters for the continuous transfer function model.

R, S	Orders of the weighting parameters E's and H's respectively.
δ_i, ω_i	The ith weighting parameters for the discrete transfer function model.
r, s	Orders of the weighting parameters δ and ω respectively.
R_K	Covariance at lag K.
$\underline{\underline{H}}$	Nxk_o matrix composed of normalized and delayed input data to the noisy transfer function model.
J	Performance index.
$\underline{\underline{V}}$	Symmetric positive definite NXN matrix.
$\underline{\underline{G}}$	NxK_o matrix.
M	Number of observation sets.
$\underline{i}$	Unitary vector of dimension Mxl.
$\underline{0}$	Null vector of dimension $k_o$xl.
$\underline{\underline{I}}$	Identity matrix of dimension NxN.
θ_c	A simplified quadratic performance index.
$\eta(\cdot)$	Residual sequence generated via the noisy transfer function model.

$\bar{\eta}$	Mean value of the sequence $\eta(\cdot)$.
$\hat{\sigma}_\eta$	Estimated variance of the sequence $\eta(\cdot)$.
θ	A biasing limit for the test of zero mean.
D_1	A t-distributed variable.
η_o	Threshold for the test of zero mean value of residuals.
$\bar{R}(k)$	Value of the observed correlogram at lag k.
$R(k)$	Value of theoritical correlogram at lag k.
$R^j(k)$	An estimate of the jth observed correlogram at lag k.
$R^M(k)$	An estimate of the actual observed correlogram at lag k.
α_j	The jth parameter for the linear stochastic difference equation model.
ϕ_j	The jth weighting function for the linear stochastic difference equation model.
$\underline{a}$	Estimated parameter vector for the linear stochastic difference equation model.
$\underset{=}{S}$	Covariance matrix.

$[\hat{w}(\cdot)]$	Estimated residual sequence obtained during the tuning stage of the linear stochastic difference equation model.
$[\bar{w}(\cdot)]$	Residual sequence of estimation stage for the linear stochastic difference equation model.
n	Number of autoregressive terms.
m	Number of corrective error terms.
n_3	Number of sinusoidal terms.
n_1	An integer value equal to $n + m + n_3$.
C_i	The <u>ith</u> class of linear stochastic difference equation model.
$\hat{\phi}_i$	An estimate for the conditional maximum likelihood function.
ν_i	Field of conditional maximum likelihood functions.
$F(\bar{w})$	Continuous cumulative distribution function of $\bar{w}$.
$[w_{(i)}]$	Order statistics of $[w(i)]$.
Z	Test statistic for the test of normality.
$L(Z)$	Limiting cumulative function of $D_N \sqrt{N}$.
H_o	Null hypothesis.
d_c	Threshold value for the test of normality.

Γ_{n_2}	$n_2 x n_2$ matrix composed of autocorrelation coefficients of residuals at different lags.
$\beta(w)$	Test statistic for serial independence.
β_1	Threshold value for the test of serial independence.
$e(k)$	Error of prediction al lag k.
$R_{yx}(k)$	Cross-correlation coefficient of y and x with lag k.
M_1	The most acceptable linear stochastic difference equation model.
E_o	Mean value of residual sequence $[\bar{w}(\cdot)]$.
E_1	Absolute mean value of residual sequence $[\bar{w}(\cdot)]$.
E_2	Mean square value of residual sequence $[\bar{w}(\cdot)]$.
M_4'	The most successful noisy-transfer function model.

CHAPTER I
INTRODUCTION

CHAPTER I

INTRODUCTION

1.1 ART OF MODELING

The word "model" is used in many situations to describe the physical system at hand. Consequently, there is a strong difference of opinion as to the appropriate use of the model. It may suggest a photographic replication of the system under study which reflects all its ramifications so that the model may adequately represent that system.

Usually, complicated physical systems, such as river catchments, do not need an inextricable mathematical model to describe it. Thus, it is advisable to select a relatively simple model to a given system and increase the complexity of that model only if the simplest one is not satisfactory.

Briefly, the class selection methods furnish only the best class among a list of chosen classes. There is no guarantee that the best fitting model from the best class given by the class selection methods is the most appropriate one, i.e., it may not pass the validation tests. Thus, we should consider all the possible classes relevant for the physical system under consideration.

Practically, the best fitting model is that model which passes all the validation tests and have a relatively small number of parameters among the various prespecified classes.

1.2 OBJECTIVE OF STUDY AND SCOPE OF THE WORK

This research work is directed to the identification, estimation, and validation of some stochastic models suitable for river catchments.

Two families of models are discussed in some details. The first family is the linear stochastic difference equation models, while the second is the transfer function models selected using the minimum mean-square error criterion. The choice of the adequate model from either two families, for a given river catchment, is treated in the following steps :

i) Estimation of the parameters in a model using the given physical observations. This is usually known as the tuning step of the model.

ii) Choice of the appropriate structure by means of some class selection techniques.

iii) Verification of the validity of the selected structure by means of "goodness of fit" test and by a direct comparison of the various statistical characteristics of both the observed and estimated output data sequences.

Once the appropriate structure is selected, its one-step ahead prediction capability is checked by the straight forward comparison of the predicted and observed output data sequences within some prespecified levels of classification.

The following is a brief outline of the main parts of this report :

Chapter II discusses pertinent details of the model building problem as well as some alternative structures of models.

Chapter III presents an important model structure which is commonly used for river catchments. The possibility of using either the generalized least-square or constrained estimator to evaluate the unknown parameters of that noisy-transfer function model is also scrutinized. The validity of the proposed model is then examined in order to achieve a better estimatability conditions.

In Chapter IV, a family of univariate linear stochastic difference equation models is suggested for representing the given physical data sequence. Moreover, some methods are given for estimating the unknown parameters of these models. The nature of model validation is also discussed by using some goodness of fit tests.

In Chapter V, the Waki river catchment is selected as a case study to demonstrate the applicability of the above models. A complete description of this catchment is given from both the geological, meteorlogical and hydrological view points.

Chapter VI investigates the availability of using either the noisy-transfer function model or the univariate linear stochastic difference equation model, with different concepts for each, to represent Waki river catchment. The forecasting capability of the two successful models, each developed from a prespecified family, is also tested for the given catchment.

Chapter VII presents a summary of the report as well as its main findings.

CHAPTER II

CHOICE OF AN APPROPRIATE MODEL

CHAPTER II

CHOICE OF AN APPROPRIATE MODEL

2.1 INTRODUCTION

The choice of an appropriate model for a given physical data such as river catchments is necessarily iterative, i.e. it is a process of evaluation and adaptation. Usually, when the physical mechanism of a phenomenon is completly understood, it may be possible to write down a mathematical expression which depicts it exactely, thus we obtain an ideal mathematical model. Although, insufficient information may be available initially to write an adequate mechanistic model. Nevertheless, an adaptive strategy can sometimes lead to such a model. On the other hand, the rather complete knowledge or large experimental resources needed to produce a mechanistic model are not available and we must then resort to a stochastic model tuned by observed physical data [Box and Hunter (1965)].

2.2 ITERATIVE APPROACH TO MODEL BUILDING

In fitting dynamic models, a theoretical analysis can sometimes tell us not only the appropriate form of the model but also can furnish good estimates of the numerical values of its parameters. The various stages of the iterative approach are:

i) From the interaction of theory and practice, a useful class of models, for the purpose at hand, is considered.

ii) Because this class is too extensive to be conveniently fitted directly to the physical data, rough methods for identifying subclass of these models are sought. Such methods of model identification employ data and knowledge of the system to suggest an appropriate parsimonious subclass of models which may be utilized to yield rough preliminary estimates of the model's parameters.

iii) The rough estimates obtained during the identification stage can now be used as commencing values in more refined iterative methods for estimating these parameters.

iv) Diagnostic checks are applied with the object of uncovering possible lack of fit. If a permissible lack of fit is indicated, the model is ready to use, but if any inadequacy is found, the iterative cycle of identification, estimation and diagnostic checking is re-iterated until a suitable mathematical representation is attained.

2.3 GENERAL CLASSES OF PHYSICAL DATA BASED MODELS

2.3.1 Deterministic Models

It is sometimes possible to derive an empirical model, based on physical laws, which permits the calcualtion of some time-dependent quantities, almost exactly, at any instant of time. If exact calculations are attainable, such a model is entirely deterministic.

2.3.2 Stochastic Models

In diverse cases, we have to consider a time-dependent phenomenon comprising many unknown factors and can not render the application of a deterministic model possible. Thus, it may be easier to derive a model which can be used to calculate the probability of a future value lying between two specified limits. Such a class of models is called a stochastic model which is introduced to achieve an optimal forecasting and control tasks for the physical processes. The main subclasses of these stochastic models are:

2.3.2a The Linear Filter Subclass

Usually, a physical system in which successive values are highly dependent can be usefully regarded as generated from a series of independent random variable $w(t)$ by what is called a linear filter [Yule (1927)]. The linear filtering operation simply assumes a weighted sum of previous observation, so that

$$y(t) = \bar{y} + w(t) + \psi_1 w(t-1) + \psi_2 w(t-2) + \ldots \tag{2.1}$$

where the weights ψ_1, ψ_2, ..., may be finite or infinite and the parameter $\bar{y}$ is the mean value of the process $y(\cdot)$.

2.3.2b The Autoregressive Subclass

In this subclass, the current values are expressed as a finite linear aggregate of the previous values and a random $w(t)$. Let us denote the deviation of the process $y(\cdot)$ from its mean value $\bar{y}$ at equally spaced time intervals $t, t-1, \ldots, t-p$, by $y_D(t), y_D(t-1), \ldots, y_D(t-p)$ respectively. This gives

$$y_D(t) = \phi_1\, y_D(t-1) + \ldots + \phi_p\, y_D(t-p) + w(t) \tag{2.2}$$

which is called an autoregressive (AR) model of order p.

2.3.2c Moving Average Subclass

In this subclass, it is considered that the deviation of the system output from its mean value be linearly dependent on a finite number of previous random variables. That is

$$y_D(t) = w(t) - \theta_1\, w(t-1) - \ldots - \theta_q\, w(t-q) \tag{2.3}$$

which is referred to as the moving average (MA) model of order q.

2.3.2 d Mixed Autoregressive Moving Average Subclass

To achieve greater flexibility in fitting mathematical models, it is advantageous to include both autoregressive and moving-average terms to the model. This will lead to the mixed autoregressive moving-average (ARMA) model. The notation ARMA (p,q), represents an ARMA model with p consecutive AR terms $y_D(t), \ldots, y_D(t-p)$ and another q consecutive MA terms $w(t), \ldots, w(t-q)$. This model is expressed mathematically as

$$y_D(t) = \phi_1 y_D(t-1) + \ldots + \phi_p y_D(t-p) + w(t) - \theta_1 w(t-1) - \ldots - \theta_q w(t-q) \cdot \quad (2.4)$$

2.3.3 The Transfer Function Models

In these models, the deviation of the input $[x(\cdot)]$ and the output $[y(\cdot)]$ from their appropriate mean values are related by a linear differential equation of the form

$$(1 + E_1 D + \ldots + E_R D^R)\, y_D(t) = (H_0 + H_1 D + \ldots + H_S D^S)\, x_D(t-\tau), \qquad (2.5)$$

where D is the differential operator, the E's and H's are unknown parameters and τ is a time delay factor.

In a similar way, for discrete data systems, we can represent the transfer function between the quantities x_D and y_D each measured at equispaced time intervals, by the corresponding difference equation

$$(1 - \delta_1 B - \ldots - \delta_r B^r)\, y_D(k) = (\omega_o - \omega_1 B - \ldots - \omega_s B^s)\, x_D(k-b) \qquad (2.6)$$

or simply

$$y_D(k) = V(B)\, x_D(k), \qquad (2.7)$$

where V(B) designates the transfer function of the given physical system.

The problem of estimating the transfer function V(B) is, however, practically complicated due to the presence of some undefined noises. Therefore, we adjust the ideal transfer function model (2.7) to be in the form

$$y_D(k) = V(B)\, x_D(k) + w(k), \tag{2.8}$$

where w(·) is a zero-mean Gaussian distribution random variable whose variance is to be determined from the tuning process employing the physical data.

2.4 CLASS SELECTION OF MODELS

In selecting an appropraite class of models among a number of possible candidates, we need a suitable criterion which may be specified according to the goal of model building. Sometimes, many common criteria such as mean-square error may not lead to a better model selection. Hence, we shall work with a more sensitive criterion such as the likelihood or one-step ahead prediction approaches.

2.5 VALIDATION OF THE SELECTED MODELS

Once the appropriate class of models is selected, we must investigate how will that class represents the given physical data sequence, this is referred to as validation test of the model.

The first approach for validation testing is to check the validity of the assumptions behind the model. But to confirm the validity of the model, we have to directly compare the principle characteristics of the model output such as correlogram, power spectrum and histogram with these of the physical system. We accept the model if the discrepancy between the two sets of actual and simulated data characteristics is within one or two standard deviation limits of the actual data characteristics, which is inversely proportional to

$\sqrt{N}$, N being the number of observations. This acceptance criterion represents the most common used second approach for validation testing. Other validation tests will be considered later in more details.

2.6 SOME FEATURES OF STOCHASTIC MODELS

2.6.1 Stationarity

A stochastic model is said to be strictly stationary if its properties are unaffected by a change of time origin, i.e. if the joint probability distribution associated with m-observations, made at any set of times t_1, t_2, ..., t_m, is the same as that associated with other m-observations made at $t_1 + k$, $t_2 + k, \ldots, t_m + k$, where k is an arbitrary time shift operator [Papoulis (1965)].

Moreover, a stochastic model can be regarded as weakly stationary representation if the mean and covariance of its output series [y(·)] exist and satisfy

$$E\,[y(t)] = E\,[y(t+k)] \tag{2.9}$$

as well as

$$E\left\{\Big[y(t) - E\,[y(t)]\Big]\Big[y(t+k) - E[y(t+k)]\Big]\right\} = R_k \tag{2.10}$$

where E [(·)] is the expected value of a sequence (·) and R_k is the covariance at lag k [kashyap and Rao (1976)].

Most of the physical processes are stationary for finite period of time but there is, of course, no sudden transition from stationary to non-stationary behaviour.

In doubtful cases, there may be an advantage in employing the non-stationary models rather than the stationary alternative. It is advisable to select the nonstationary models for those systems whose mathematical representation requires some periodic and/or time-dependent terms. On the other hand, the stationarity of a given stochastic model may ensure its convergence to a stable estimates of the unknown parameters involved by that model [Box and Jenkins (1970)].

2.6.2 Invertibility

A stochastic model is said to be invertable if the added noise sequence can be recovered, with probability one or in the mean-square sense, from a semi-infinite history of input and output data sequences. The concept of invertibility forms the basis of parameter estimation and prediction in systems with moving average terms, but it is automatically achieved by the other systems.

Definitely, the invertable stochastic models are relevant for keeping the main statistical characteristics of the added noise sequence [kashyap and Rao (1976)].

CHAPTER III

ANALYSIS OF THE NOISY_TRANSFER FUNCTION MODEL

CHAPTER III

ANALYSIS OF THE NOISY-TRANSFER FUNCTION MODEL

3.1 INTRODUCTION

In this chapter, some numerical methods are described for identifying, fitting and checking the noisy-transfer function model when simultaneous pairs of observations of the input and output data are available at a discrete time intervals.

3.2 IDENTIFICATION OF THE NOISY-TRANSFER FUNCTION MODEL

Alternatively, the noisy-transfer function model of (2.8) can be written in the following matrix form [Natale and Todini (1976)]

$$\underline{y} = \underline{\underline{H}}\,\underline{U} + \underline{\varepsilon} \tag{3.1}$$

where:

i) $\underline{y}$ is Nx1 vector designating the noramlized deviation of the output sequence from its mean value and can be written as

$$\underline{y} = \begin{bmatrix} y(1) \\ y(2) \\ \cdot \\ \cdot \\ \cdot \\ y(N) \end{bmatrix}. \tag{3.2}$$

ii) $\underline{\underline{H}}$ is $N x k_o$ matrix denoting the delayed normalized deviation of the input data sequence from its mean value which is related to the model output sequence at any time interval, and may be expressed as

$$\underline{\underline{H}} = \begin{bmatrix} \tilde{x}_d(1) & 0 & \cdot & \cdot & \cdot & \\ \tilde{x}_d(2) & \tilde{x}_d(1) & \cdot & \cdot & \cdot & \\ \tilde{x}_d(3) & \tilde{x}_d(2) & \cdot & \cdot & \cdot & \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ \tilde{x}_d(N) & \tilde{x}_d(N-1) & \cdot & \cdot & \cdot & \tilde{x}_d(N-k_o+1) \end{bmatrix} \tag{3.3}$$

where k_o is the kernel length.

iii) $\underline{U}$ is $k_o x1$ vector comprising the parameters of the impulse response vector, and is written as

$$\underline{U} = \begin{bmatrix} U(1) \\ U(2) \\ \cdot \\ \cdot \\ \cdot \\ U(k_o) \end{bmatrix} . \tag{3.4}$$

iv) $\underline{\varepsilon}$ is Nx1 vector denoting the input noise to the model at equispaced time intervals, and is given by

$$\underline{\varepsilon} = \begin{bmatrix} \varepsilon(1) \\ \varepsilon(2) \\ \cdot \\ \cdot \\ \cdot \\ \varepsilon(N) \end{bmatrix} . \tag{3.5}$$

3.2.1 Least-Square Estimation of the Impulse Response Vector

Usually, the least-Square (LS) estimator can be invoked if the statistical characteristics of the noise vector $\underline{\varepsilon}$ are unknown, which is the most general case. In fact, by definition, the LS estimator is that estimator which minimizes the quadratic performance index

$$J = \frac{1}{2}\,\underline{\varepsilon}^T\,\underline{\underline{V}}^{-1}\,\underline{\varepsilon} \tag{3.6}$$

where $\underline{\underline{V}}$ is a symmetric positive definite matrix.

The performance index J can be written in the form of the impulse response vector $\underline{U}$ as follows

$$J = \frac{1}{2}\,(\underline{y} - \underline{\underline{H}}\underline{U})^T\,\underline{\underline{V}}^{-1}(\underline{y} - \underline{\underline{H}}\underline{U}). \tag{3.7}$$

The necessary condition for the existence of an extreme value is that

$$\left.\frac{\partial J}{\partial \underline{U}}\right|_{\underline{U} = \hat{\underline{U}}_{LS}} = 0 \tag{3.8}$$

which yields

$$\hat{\underline{U}}_{LS} = (\underline{\underline{H}}^T\,\underline{\underline{V}}^{-1}\,\underline{\underline{H}})^{-1}\;\underline{\underline{H}}^T\,\underline{\underline{V}}^{-1}\,\underline{y}\,, \tag{3.9}$$

where $\hat{\underline{U}}_{LS}$ is the least-square estimate of the impulse response vector $\underline{U}$. On the other hand, the sufficient condition for the existence of a minimum is then satisfied by

$$\frac{\partial^2 J}{\partial\,\underline{U}^2} \geq 0. \tag{3.10}$$

This is attained only if the matrix $(\underline{\underline{H}}^T\underline{\underline{V}}^{-1}\underline{\underline{H}})$ is positive definite.

4.2.2 The Constrained Estimation of the Impulse Response Vector

An improvement in the accuracy of the estimated impulse response vector can be produced by considering some priori additional information, which can reduce the field of the choice of $\underline{U}$ [Natale and Todini (1976)]. A natural way of obtaining this reduction is to impose a set of constraints that must be satisfied by the true and estimated values of the impulse vector $\underline{U}$.

In many hydrological systems, which are mathematically balanced, it is possible to impose upon the impulse response vector $\underline{U}$ a set of linear constraints, namely $\underline{\underline{G}}\underline{U} = \underline{i}$, which expresses the continuity equation. But, for those physical systems which can be described by a positive autocorrelation and cross-correlation coefficients, it is more convenient to assume

$$\underline{U} \geq \underline{0}, \tag{3.11}$$

which represents an inequality constraint that must be satisfied by the estimated response vector $\hat{\underline{U}}$. Sometimes, we have to consider both the equality and inequality constraints based on some mathematical and physical consideration [Natale and Todini (1976)].

For instance, the solution of the constrained estimation problem can be found by searching for the minimum of

$$J_c = \frac{1}{2}(\underline{y} - \underline{\underline{H}}\underline{U})^T \underline{\underline{V}}^{-1}(\underline{y} - \underline{\underline{H}}\underline{U}) \tag{3.12}$$

which reduces to

$$\theta_c = \frac{1}{2}\ \underline{U}^T \underline{\underline{H}}^T \underline{\underline{V}}^{-1} \underline{\underline{H}}\underline{U} - \underline{U}^T \underline{\underline{H}}^T \underline{\underline{V}}^{-1} \underline{y} \tag{3.13}$$

subject to

$$\underline{\underline{G}}\,\underline{U} = \underline{i} \text{ and/or } \underline{U} \geq \underline{0}, \tag{3.14}$$

where:

i) $\underline{y}$ is Nx1 vector representing the system output, at equispaced time intervals, substracted from the estimated mean value of the noise sequence $[\varepsilon(\cdot)]$.

ii) $\underline{\underline{H}}$ is an $N \times k_o$ matrix composed of the delayed system input sequence $[x_d(\cdot)]$, and may be written as

$$\underline{\underline{H}} = \begin{bmatrix} x_d(1) & 0 & \cdot & \cdot & \cdot & 0 \\ x_d(2) & x_d(1) & \cdot & \cdot & \cdot & 0 \\ x_d(3) & x_d(2) & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ x_d(N) & x_d(N-1) & \cdot & \cdot & \cdot & x_d(N-k_o+1) \end{bmatrix} . \tag{3.15}$$

iii) $\underline{\underline{G}}$ is $M \times k_o$ matrix containing the continuity coefficients for an M input vectors.

iv) $\underline{i}$ is an Mx1 unitary vector and $\underline{0}$ is $k_o \times 1$ null vector. Thus

$$\underline{i} = \begin{bmatrix} 1 \\ 1 \\ \cdot \\ \cdot \\ \cdot \\ 1 \end{bmatrix} \quad \text{and} \quad \underline{0} = \begin{bmatrix} 0 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{bmatrix} . \tag{3.16}$$

v) $\underline{\underline{V}}$ is an NxN covariance matrix of the noise sequence.

3.3 ASSUMPTIONS ABOUT THE COVARIANCE MATRIX

As stated previously, either the constrained or unconstrained estimates of the impulse response vector $\underline{U}$ need apriori evaluation of the noise covariance matrix $\underline{\underline{V}}$. Unfortunately, it is not possible to resolve the nature of the noise vector by looking at the residual sequence, thus it is assumed to be a white noise so that the covariance matrix becomes

$$\underline{\underline{V}} = \begin{bmatrix} \sigma^2_{11} & 0 & \cdot & \cdot & \cdot & \\ & \sigma^2_{22} & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \ddots & & \cdot \\ & \cdot & & & & \cdot \\ & \cdot & & & & \sigma^2_{NN} \end{bmatrix} . \tag{3.17}$$

Practically, to set up the noise covariance matrix we consider that, [Natale and Todini (1976)],

$$\underline{\underline{V}} = \sigma^2 \, \underline{\underline{I}} \tag{3.18}$$

where $\underline{\underline{I}}$ is an NxN identity matrix and σ is the standard deviation of the noise sequence.

Finally, the previous constrained optimization problem could be solved using the quadratic programing technique as the performance index θ_c is a concave function [Wilson (1963)].

3.4 ESTIMATION OF THE MODEL OUTPUT DATA SEQUENCE

Let

$$\left.\begin{aligned} \bar{y} &= \frac{1}{N} \sum_{i=1}^{N} y(i) \\ \sigma_y &= \frac{1}{N} \left[\sum_{i=1}^{N} [y(i) - \bar{y}]^2 \right]^{\frac{1}{2}} \end{aligned}\right\} \tag{3.19}$$

be respectively the mean and variance of the observed output data sequence $[y(\cdot)]$, which can be used together with the unconstrained impulse response vector estimated before to generate the current estimates of the model output sequence as follows

$$\left.\begin{aligned} \hat{y}(k) &= \bar{y} + \sigma_y \sum_{i=1}^{k_u} \hat{U}_{LS}(i)\, \tilde{x}_d(k-i+1) \\ &\forall k = 1, 2, \ldots, N. \end{aligned}\right\} \tag{3.20}$$

This can be reduced, in the case of constrained estimator, to

$$\hat{y}(k) = \sum_{i=1}^{k_u} \hat{U}(i)\, x_d(k-i+1). \tag{3.21}$$

Alternatively, the residual sequence $[n(\cdot)]$ for either the constrained or unconstrained estimates of the model output data sequence $[y(\cdot)]$ may be given by

$$\left.\begin{aligned} n(k) &= y(k) - \hat{y}(k) \\ &\forall k = 1, 2, \ldots, N. \end{aligned}\right\} \tag{3.22}$$

3.5 VALIDATION TESTS USING RESIDUALS OF ESTIMATION

Usually, some validation tests are applied to check the adequacy of the generated residual sequence for the priori estimation conditions, such as

$$\bar{\eta} = 0 \tag{3.23}$$

which is called the zero-mean test [kashyap and Rao (1976)].

3.5.1 Test of Zero Mean

On the basis of residuals [$\eta(\cdot)$], we have to choose one of the following assumptions:

$$\left.\begin{aligned} S_o &: \eta(k) = w(k) \quad \text{, or} \\ S_1 &: \eta(k) = \theta + w(k) \\ &\forall\, k = 1, 2, \ldots, N \text{ ,} \end{aligned}\right\} \tag{3.24}$$

where $w(\cdot)$ is a sequence of zero mean random variable with distribution $N(0,\rho)$, and θ is a biasing limit. Let

$$\left.\begin{aligned} \bar{\eta} &= \frac{1}{N} \sum_{i=1}^{N} \eta(i) \text{ , and} \\ \hat{\rho} &= \frac{1}{N-1} \sum_{i=1}^{N} [\eta(i) - \bar{\eta}]^2 \end{aligned}\right\} \tag{3.25}$$

be the mean and variance of the residual sequence respectively. Define

$$D_1 = (N / \hat{\rho})^{\frac{1}{2}}\, \bar{\eta} \tag{3.26}$$

where D_1 is t-distributed variable with N-1 degrees of freedom independent of $\hat{\rho}$ [Kashyap and Rao (1976)]. Hence, we can employ the following decision rule

$$\text{IF } |D_1| \begin{cases} < \eta_0 & \text{Accept } S_o \\ \geq \eta_0 & \text{Reject } S_o, \end{cases} \tag{3.27}$$

such that the threshold η_0 could be chosen from the table of t-distribution with the corresponding degree of freedom and required significant level. For large values of physically based observation, one may consider

$\eta_0 = 1.64$ at 95% significant level , and

$\eta_0 = 1.28$ at 90% significant level.

3.5.2 Correlogram of Residuals with Two Standard Deviation Limits

Anderson (1971) showed that, the autocorrelation coefficients of a sequence of zero-mean white noise are, approximately, normally distributed with zero mean and variance 1/N.

Let

$$R(k) = \frac{1}{(N-k)\,\hat{\rho}} \sum_{j=1}^{N} \eta(j)\,\eta(j-k) \tag{3.28}$$

be the theoretical correlogram of the residual sequence $[\eta(\cdot)]$. Thus, for a zero-mean white noise, the coefficients R(k) at any lag k, k being greater than zero, should be:

a) Small in comparison to unity.
b) Lie between the range $\pm\, 2/\sqrt{N}$ with probability of nearly 0.95.

2

3.6 VALIDATION TESTS BASED ON COMPARISON OF THE VARIOUS CHARACTERISTICS OF OBSERVED AND ESTIMATED DATA

In these tests, we will directly compare the theoretical characteristics of the observed and estimated output sequences. Of course, we can compare only few characteristics such as correlograms and power spectrums [Kashyap and Rao (1976)].

3.6.1 Comparison of Correlograms

Let

$$\bar{R}(k) = \frac{1}{(N-k)\,\sigma_y^2} \left[\sum_{j=1}^{N} [y(j) - \bar{y}]\,[y(j-k) - \bar{y}] \right], \tag{3.29}$$

$$R(k) = \lim_{N \to \infty} \bar{R}(k), \text{ and}$$

$$\sigma(k) = \Big[E\,[R(k) - \bar{R}(k)] \Big]^{\frac{1}{2}}$$

where $\bar{y}$, σ_y^2 denotes respectively the mean and variance of the output sequence $[y(\cdot)]$.

The graph of $\bar{R}(k)$ versus k, for fixed N, is called the observed correlogram whereas R(k) versus k is called the theoretical correlogram of the same output sequence $[y(\cdot)]$. The degree of fit between these two correlograms can be quantitatively expressed in a manner consistent with the available observation size. Let

$$R^M(k) = \frac{1}{M} \sum_{j=1}^{M} R^j(k), \text{ and}$$

$$\sigma^M(k) = \frac{1}{M} \sum_{j=1}^{M} \Big[[R^j(k) - R^M(k)]^2 \Big]^{\frac{1}{2}}, \tag{3.30}$$

where:

i) M is a reasonable number of independent observation sequence for the model output which can be generated by the appropriate simulation of the model.

ii) $R^j(k)$ is an estimate of the jth observed correlogram at lag k.

iii) $R^M(k)$ indicates an estimate of the actual observed correlogram at lag k.

iv) $\sigma^M(k)$ is an estimate of $\sigma(k)$.

Practically, the observed correlogram can be regarded as being a good fit to the theoretical correlogram of the model if the following relationship is satisfied

$$R^M(k) - 2\,\sigma^M(k) \leq R(k) \leq R^M(k) + 2\,\sigma^M(k) \tag{3.31}$$

and hence the model can be considered as adequate in representing the actual physical system.

3.6.2 Comparison of Power Spectrum

Similarly, the qualitative decision rules may be used to test the resemblance between the observed and theoretical power spectrums. The theoretical and observed power spectrums may be evaluated as shown in Appendix A.

CHAPTER IV
ANALYSIS OF SOME STOCHASTIC LINEAR MODELS

CHAPTER IV

ANALYSIS OF SOME STOCHASTIC LINEAR MODELS

4.1 INTRODUCTION

In this chapter, we consider the structure of stochastic linear models described by a finite univariate difference equation. This class of models has a variety of terms such as autoregressive terms, moving average terms and deterministic trend function.

4.2 DESCRIPTION OF THE PROPOSED MODEL

It is convenient, though not necessary, to assume that, [Kashyap and Rao (1972)], the stochastic process $[y(\cdot)]$ obeys the following stationary stochastic difference equation:

$$y(k) = \sum_{j=1}^{n} \alpha_j \phi_j [k-1, y(k-1), \ldots, y(k-n), U(k-1), \ldots, U(k-n_3)] + \sum_{j=1}^{m} \alpha_{n+j} w(k-j) + w(k) \tag{4.1}$$

where $w(\cdot)$ is the disturbance sequence whose statistical characteristics are unknown except for

$$E\left[w(k) \phi_j [k-1, y(k-1), \ldots, y(k-n)]\right] = 0, \quad j=1, 2, \ldots, n \tag{4.2}$$

$$E[w(k) w(k-j)] = 0, \quad j=1, 2, \ldots, m \tag{4.3}$$

where $E(\cdot)$ indicates the expected value of $(\cdot)$.

Usually, the deterministic trend sequence [U(·)] is introduced to reflect the variation of data from its mean value during an interval M of time. This sequence is expressed as

$$U(i) = \alpha_o + \sum_{j=1}^{n_3} [\alpha_{n+m+2j-1} \cos W_j i + \alpha_{n+m+2j} \sin W_j i] \tag{4.4}$$

where the frequency of variation W_j is defined as

$$W_j = 2\pi j/M , \quad j = 1, 2, \ldots, N. \tag{4.5}$$

Alternatively, when the sequence [y(·)] is strictly positive, we could assign the following multiplicative form of the difference equation

$$y(k) = \prod_{j=1}^{n} [\phi_j (k-1, y(k-1), \ldots, y(k-n), U(k-1), \ldots, U(k-n_3)]^{\alpha_j}$$

$$\prod_{j=1}^{m} w(k) [w(k-j)]^{\alpha_{n+j}} \tag{4.6}$$

where the parameters n, n_3 and m in both (4.1) and (4.6) are chosen to achieve, in the mean square sense, a better prediction ability. Moreover, the function $\phi_j(\cdot)$ can be expressed as

$$\phi_j(k) = [y(k), y(k-1), \ldots, y(k-n+1), 1, \cos W_1 k,$$

$$\sin W_1 k, \ldots, \cos W_{n_3} k, \sin W_{n_3} k] \tag{4.7}$$

where W_j is the frequency of variation defined at the jth time interval.

4.3 ESTIMATION OF THE PARAMETER VECTOR

We shall present a heuristic development of the recursive algorithm for computing the vector $\underline{\alpha}$. Alternatively, (3.1) may be written as

$$y(k) = \underline{\alpha}^T \underline{Z}\ (k-1) + w(k) \tag{4.8}$$

where

$$\underline{\alpha}^T = [\alpha_o, \alpha_1, \ldots, \alpha_{n+m+2n_3}] \tag{4.9}$$

and

$$\underline{Z}^T\ (k-1) = [\phi_1\ (k-1), \ldots, \phi_n\ (k-n), w(k-1), \ldots, w(k-m)]. \tag{4.10}$$

Let $\underline{a}(i)$ be an estimate for the N-dimension vector $\underline{\alpha}$ computed by using the following recursive algorithm [Kashyap and Rao (1972)]

$$\left.\begin{aligned} \underline{a}(i+1) &= \underline{a}(i) + \underline{\underline{S}}(i+1)\ \underline{Z}(i)\ [y(i+1) - \underline{a}^T(i)\ \underline{Z}(i)] \\ \underline{\underline{S}}(i+1) &= \underline{\underline{S}}(i) - \underline{\underline{S}}(i)\ \underline{Z}(i)\ \underline{Z}^T(i)\ \underline{\underline{S}}(i)\ /\ [1+\underline{Z}^T(i)\ \underline{\underline{S}}^T(i)\ \underline{Z}(i)] \\ \hat{w}(i+1) &= y(i+1) - \underline{a}^T\ (i+1)\ \underline{Z}(i), \qquad i = 1, 2, \ldots, N-1 \end{aligned}\right\} \tag{4.11}$$

where $[w(\cdot)]$ is an estimate for the residual sequence $w(\cdot)$ whose final estimates may be given by

$$\bar{w}(k) = y(k) - \underline{a}^T{}_F\ \underline{Z}(k-1), \qquad k = 1, 2, \ldots, N, \tag{4.12}$$

where $\underline{a}_F$ denotes the final estimate of the parameter vector $\underline{\alpha}$.

Practically, the above algorithm should be initialized before it is operated in the recursive model (4.11). Therefore, either one of the following procedures may be invoked:

4.3.1 The First Procedure

Let the available data be designated by [y(j)], where j=1, 2, ..., N. Thus, the algorithm commences as follows [Kashyap and Rao (1972)]

$$\left.\begin{aligned} &\underline{a}(0) = \underline{0}\ , \quad \underline{\underline{S}}(0) = \underline{\underline{I}} \\ &y(j) = 0\ , \quad j = -1, -2, \ldots, -n \\ &w(\kappa) = 0\ , \quad \kappa = -1, -2, \ldots, -m \end{aligned}\right\} \tag{4.13}$$

4.3.2 The Second Procedure

Let the available data be denoted by [y(j)], such that j=-p, -p+1, ..., where p is an integer greater than or equal to 2n. Hence, the procedure for initialization is [Kashyap and Rao (1972)]

$$\underline{\underline{S}}(0) = \sum_{j=-(p-n_1)}^{o} [\underline{Z}(j-1)\ \underline{Z}^T(j-1)]^{-1}$$

and (4.14)

$$\underline{a}(0) = \underline{\underline{S}}(0)\ [\sum_{j=-(p-n_1)}^{o} \underline{Z}(j-1)\ y(j)]$$

where n_1 is an integer given by

$$n_1 = n + n_3 + m. \tag{4.15}$$

On the other hand, the values w(1), w(2), ..., are all generated from a Gaussian random number generator with zero-mean and variance equal to the sample variance of [y(0), y(-1), ..., y(-p)].

The first procedure is easier to implement, while the second procedure leads to a better prediction for small values of k.

Obviously, the parameters of the multiplicative structure (4.6) may be identified by a same manner as the additive structure (4.1) but with a natural logarithmic transformation technique [Kashyap and Rao (1972)].

4.4 CLASS SELECTION OF UNIVARIATE STOCHASTIC MODELS DESCRIBED BY A LINEAR DIFFERENCE EQUATION

One of the popular methods for comparing some proposed classes of the univariate stochastic models which are depicted by a linear difference equation is the method of hypothesis testing. Even though, the theory of that method is elegant [Kashyap and Rao (1976)], as it involves arbitrary quantities such as significant levels. Furthermore, it has limited applicability in the sense that it can handle, essentially, two classes of models at a time. Hence, two other approaches may be involved to select an appropriate class of models among q-proposed classes.

4.4.1 The Likelihood Approach

The decision rule can be expressed as follows:

i) For every proposed class C_i, i = 0, 1, ..., q-1, find the conditional maximum likelihood estimate $\hat{\phi}_i$ of $\overset{o}{\phi}_i$ given that $\overset{o}{\phi}_i \epsilon \nu_i$ using the given observation [ζ = y(j), j = 1, 2, ..., N]. Then compute the corresponding value of likelihood function L_i as follows

$$L_i = \ln p\,(\zeta\,,\,\hat{\phi}_i) - n_i,\ \hat{\phi}_i = (\underline{a}_{F_i},\,\hat{\rho}_i) \tag{4.16}$$

where $p(\cdot\,,\,\cdot)$ denotes the conditional probability and n_i is the dimension of the vector $\phi_i^o = [\underline{a}_i,\,\rho_i^o]$.

ii) Choose the class which yield the maximum value of L_i among $[L_i,\ i = 0,\ 1,\ \ldots,\ q-1]$. Specifically, for the simplified model (4.8), the mathematical expressions for $\hat{\rho}_i$ and L_i are given be Kashyap and Rao (1976) as follows

$$\hat{\rho}_i = \frac{1}{N-m_1} \sum_{k=m_1+1}^{N} [y(k) - \underline{a}_{F_i}^T\ \underline{Z}(k-1)]^2 \tag{4.17}$$

and

$$L_i \simeq \frac{N}{2} \ln \hat{\rho}_i - n_i \tag{4.18}$$

where m_1 is the number of terms involved by C_o.

4.4 2 The Prediction Approach

This method allows the comparison of a number of different classes of models C_i, $i = 0,\ 1,\ \ldots,\ q-1$, simultaneously, where $C_i = [S_i,\ \nu_i,\ \Omega_i]$, provided that they do not have average terms [Kashyap and Rao (1976)]. Thus, consider the indices

$$J_i = \frac{1}{N-1} \sum_{k=2}^{N} [y(k) - \hat{y}_i(k|k-1)]^2 \tag{4.19}$$

where $i = 0,\ 1,\ \ldots,\ q-1$.

Practically, if there was only one class C_{i_o} such that the index J_{i_o} is the smallest among the set $[J_i, i = 0, 1, \ldots, q-1]$, we select that class. Alternatively, if more than one class can yield same minimum value of J_i, the given data will be assigned to one of these classes according to other subsidary measure such as minimal complexity.

4.4.3 Discussion of the Various Class Selection Methods

Among all the above presented methods, the likelihood approach is very versatile, theoretically sound and furnishes, in practice, reasonable results. It can simultaneously handle a number of classes, including those having moving average terms or log-tranformed terms.

One of the most distinguished merits of the likelihood approach is that, it does not involve the use of arbitrary quantities such as significant levels. One shortcoming of the likelihood approach for the determination of the order of AR models is, however, that the determined order is often higher than is necessary for passing the validation tests.

The hypothesis testing approach is more ambitious, since there is an attempt to obtain a decision rule with certain prespecified probability of error. But, in practice, it can handle only two classes at a time and even these two classes must be made up of generalized AR models.

The prediction approach is valid for systems possessing moving average terms. It is instructive to analyze the difference between the estimates of the mean-square prediction error obtained during the design of the predictor and that obtained during its testing. The difference between the two mean-square errors is examined to determine whether they are due to sampling variations only or to the poor quality of model.

On the other hand, the recursive prediction approach is especially useful with systems in which some of the parameters may vary with time. Alternatively, the prediction approach is apt to yield models that may not pass the validation tests [Kashyap and Rao (1976)].

4.5 VALIDATION OF THE FITTED MODEL

Practically, no model form ever represents completely the physical process. It follows that, given sufficient physical data, statistical tests can discredit models which could, nevertheless, be entirely adequate for the purpose at hand. Clearly, the validation tests must be such that they place the model in jeopardy, i.e. they must be sensitive to discrepancies which are likely to happen. However, if validation tests, which have been thoughtfully devised, are applied to a model fitted by a reasonable large number of data and fail to show serious discrepancies, then we shall rightly feel more comfortable about using that model.

4.5.1 Test of Normality

The goodness of fit between the histogram of residuals and the fitted normal distribution may be visually judged by the first Kolmogrov-Smirnov test as follows:

Given a sample of N-independent and identically distributed set of residuals $\bar{w}(1)$, $\bar{w}(2)$, ..., $\bar{w}(N)$, with continuous cumulative distribution function $F(\bar{w})$, the first Kolmogrov-Smirnov test calculates the difference, in absolute value, between the usual normal distribution function $F_N(\bar{w})$ and the theoretical cumulative distribution function $F(\bar{w})$. For this purpose:

i) The order statistics $[\bar{w}_{(i)}]$ are determined by sorting the set $[\bar{w}(i)]$ into an ascending order.

ii) The measured cumulative distribution function is expressed as follows:

$$F(w) = \begin{cases} 0 & \text{for } \bar{w} < \bar{w}(1) \\ k/N & \text{for } \bar{w}(k) \leq \bar{w} < \bar{w}(k+1)\ , \quad k = 1, 2, \ldots, N-1 \\ 1 & \text{for } \bar{w}(N) \leq \bar{w}. \end{cases} \tag{4.20}$$

iii) The maximum deviation D_N, in absolute value, between the measured and theoretical distribution can be written as

$$D_N = \underset{1 \leq \bar{w}_{(i)} \leq N}{\text{Max}} \left| F_N(\bar{w}) - F(\bar{w}) \right| . \tag{4.21}$$

Since $F_N(\bar{w})$ and $F(\bar{w})$ are nondescending functions, the result is

$$D_N = \underset{1 \leq k \leq N}{\text{Max}} \left| F_N(\bar{w}_{(k)}) - F(\bar{w}_{(k)}) \right| . \tag{4.22}$$

Define

$$L(Z) = \underset{N \to \infty}{\text{lime}} \quad p[D_N \sqrt{N} < Z] \tag{4.23}$$

where D_N is a random variable, $p(\cdot)$ denotes the probability of an event $(\cdot)$ and $L(Z)$ is the limiting cumulative function of $D_N \sqrt{N}$.

iv) The probability that Z being greater than or equal to the computed value of $D_N \sqrt{N}$ can be written as

$$p(Z) = 1 - L(Z) \tag{4.24}$$

where

$$L(Z) = \begin{cases} 0 & \text{for } Z \leq 0 \\ 1-2 \sum_{k=1}^{\infty} (-1)^{k-1} \exp(-2k^2 Z^2) & \text{for } Z > 0. \end{cases} \tag{4.25}$$

When Z is very small, the series (4.25) converges slowly, but, using Jacobi's Theta functions $\theta_2\,(u,t)$ and $\theta_4\,(u,t)$, defined by

$$\left.\begin{aligned}\theta_2\,(u,t) &= 2 \sum_{k=o}^{\infty} \exp\,[i\,\pi\,(k+1/2)^2\,t]\,\cos\,[(2k+1)\,u]\\ \theta_4\,(u,t) &= 1- \sum_{k=o}^{\infty} (-1)^{k-1}\,\exp\,(i\pi k^2 t)\,\cos(2k\,u)\end{aligned}\right\} \tag{4.26}$$

and invoking the Jacobi imaginary transformation

$$\theta_4\,(0,t) = (-it)^{-\frac{1}{2}}\;\theta_2\,(0,\,-1/t), \tag{4.27}$$

it follows that

$$\begin{aligned}L(Z) &= \theta_4\,(0,\,2iZ^2\,/\,\pi)\\ &= \frac{\sqrt{2\pi}}{Z}\sum_{k=1}^{\infty} \exp\,[-\,(2k-1)^2\,\pi^2\,/\,8Z^2]\end{aligned} \tag{4.28}$$

which converges quickly when Z is small, see Wittaker and Watson This gives

$$L(Z) = \begin{cases} 0 & \\ \dfrac{\sqrt{2\pi}}{Z}\displaystyle\sum_{k=1}^{3} \exp -(2k-1)^2\,\pi^2\,/\,8Z^2 + E_1\,(Z) & \text{for } 0.27 \leq Z < 1.0\\ 1-2\displaystyle\sum_{k=1}^{4}\,(-1)^{k-1}\,\exp\,(-2k^2Z^2) + E_2(Z) & \text{for } 1.0 \leq Z < 3.1\\ 1 & \text{for } 3.1 \leq Z < \infty \end{cases}$$

where

$$\begin{aligned}E_1(Z) &\leq 6(10^{-15}) \quad \text{when } Z < 1\text{ , and}\\ E_2(Z) &< 10^{-20} \quad \text{when } Z \geq 1\text{ .}\end{aligned} \tag{4.29}$$

Decision Rule:

For the value of D_N given in (4.22), define a null hypothesis H_o which assumes that both the measured and theoretical distributions are identical, then the decision rule for accepting or rejecting H_o is expressed as

$$\text{If } D_N\sqrt{N} \begin{cases} \leq d_c & \rightarrow \quad \text{Accept } H_o \\ > d_c & \rightarrow \quad \text{Reject } H_o \end{cases} \tag{4.30}$$

where the threshold d_c is chosen as
d_c = 1.36 at 95% significant level , and
d_c = 1.22 at 90% significant level.

4.5.2 Test of Serial Independence

We will determine whether the residual sequence given in (4.12), is serially correlated [Whittle (1951 and 1952)].
Let

$$C_i = [S_i,\ \nu_i,\ \Omega_i]\ ,\ i = 0,\ 1$$

$$\begin{aligned} S_o &: \bar{w}(k) = w(k) \\ S_1 &: \bar{w}(k) = \sum_{j=1}^{n_2} \theta_j\, w(k-j) + w(k) \end{aligned} \tag{4.31}$$

where $w(\cdot)$ is an independent Gaussian random variable with zero mean and variance $\hat{\rho}$, $\hat{\rho}\varepsilon\Omega$

$\theta = [\theta_1\ \theta_2 \ldots \theta_{n_2}]$, and $\nu_1 = [\theta : \theta \neq 0]$ with $\nu_o = [0]$.

Let $\hat{\rho}_o$, $\hat{\rho}_1$ be the residual variances of the best fitting models for the given data in the two classes C_o and C_1 respectively, and introduce $\hat{R}_k$ as the physically measured covariance at lag k, so that

$$\hat{R}_k = \frac{1}{N-k} \sum_{i=k+1}^{N} \bar{w}(i)\, \bar{w}(i-k). \tag{4.32}$$

Then, we have

$$\hat{\rho}_o = \hat{R}_o, \quad \hat{\rho}_1 = \det \Gamma_{n_2} / \det \Gamma_{n_2-1} \tag{4.33}$$

where Γ_{n_2} is $n_2 \times n_2$ matrix and

$$(\Gamma_{n_2})_{i,j} = \hat{R}_{|i-j|} \;;\; i, j = 1, 2, \ldots, n_2 \tag{4.34}$$

The test statistics is given by

$$\beta(\bar{w}) = \left(\frac{N}{n_2} - 1\right)\left(\frac{\hat{\rho}_o}{\hat{\rho}_1} - 1\right) \tag{4.35}$$

which is approximately follows an F-distribution with two degrees of freedom n_2 and $N-n_2$ for large value of N provided that $[\bar{w}(\cdot)]$ obeys C_o.

Decision Rule:

For the value $\beta(\bar{w})$ defined before, we can accept either C_o or C_1 according to the following decision rule

$$\beta(\bar{w}) \begin{cases} \leq \beta_1 & \rightarrow \quad \text{Accept } C_o \\ > \beta_1 & \rightarrow \quad \text{Accept } C_1 \end{cases} \tag{4.36}$$

where β_1 is chosen by the corresponding significant level and n_2 is considered as 0.1 N or 0.05 N.

4.6 DATA NORMALIZATION

In order to remove the periodicities of a given data sequence $[y(\cdot)]$, two types of normalization can be performed [Kashyap and Rao (1976)]. These are

$$\tilde{y}(k) = [y(k) - \tilde{y}] / \sigma_y \tag{4.37}$$

or

$$y(k) = \log_{10} y(k) \tag{4.38}$$

where $\tilde{y}$, σ_y are the sample mean and standard deviation of the given data sequence $[y(\cdot)]$ respectively.

Usually, the data given by the normalized models can reproduce the mean and variance with a very satisfied significance, but the prediction errors with the normalized data may be larger than the original models, see Kashyap and Rao (1976).

Clearly, the transformation given by (4.37) may be satisfactory for those models of additive structures, while the otner transformation (4.38) may be suitable for the multiplicative structures.

4.7 RECURSIVE PREDICTION OF THE OUTPUT DATA

Let $\hat{\tilde{y}}(k+1|k)$ be an estimate of the natural one-step ahead prediction $\tilde{y}(k+1)$, then

$$\hat{\tilde{y}}(k+1|k) = \underline{a}_F^T \, \underline{Z}(k) + w(k) \tag{4.39}$$

where

$$\underline{Z}^T(k) = [\phi_1(k), \ldots, \phi_n(k), w(k-1), \ldots, w(k-m)] \tag{4.40}$$

and the vector $\underline{a}_F$ is the final estimate of the parameter vector $\underline{a}$. The noise sequence $[w(\cdot)]$ is generated from a Gaussian random number generator with zero mean and variance similar to that of the residual sequence $[\bar{w}(\cdot)]$. The prediction error is defined as

$$e(k+1) = y(k+1) - \hat{y}(k+1 \mid k) \tag{4.41}$$

where

$$\hat{y}(k+1|k) = \sigma_y \hat{\tilde{y}}(k+1|k) + \bar{y} \tag{4.42}$$

for the additive structure, and

$$\hat{y}(k+1|k) = 10^{\hat{\tilde{y}}(k+1 \mid k)} \tag{4.43}$$

for the multiplicative structure.

It is important to distinguish between $\bar{w}(k)$, which is only a residual, and $e(k)$ which is a difference between the predicted and actual quantities. The convergence properties of the algorithm (4.11) can be attained by considering the $\phi_j(\cdot)$, $j = 1, 2, \ldots, n$, as linearly independent events whose cumulative mean square vlaue, $\sum_{j=1}^{k} \phi_j(k)/k$, is bounded for all values of k, see Kashyap and Rao (1972).

CHAPTER V
DESCRIPTION OF THE CASE STUDY
(WAKI RIVER CATCHMENT)

CHAPTER V

DESCRIPTION OF THE CASE STUDY (WAKI RIVER CATCHMENT)

5.1 INTRODUCTION

The case study represents a hydrologic system whose input and output daily records are as illustrated in Tables (5.1) and (5.2) respectively. These data denote the daily precipitation over the Waki River catchment, located near lake Albert, and the corresponding daily discharge. This catchment lies between longitudes 31° 18′ and 31° 39′ E, and latitudes 1° 40′ and 1° 28′ N. The catchment is drained by two main streams, Waki and Siba, see WMO (1972).

5.2 TOPOGRAPHY OF WAKI CATCHMENT

The topography of the Waki catchment is shown in Fig. (5.1). It can be observed that the catchment is steep at its southern part but its steepness drops gradually when moving towards the Waki-II hydrological station. The maximum and minimum elevations are about 1402 m and 991 m respectively, while the average surface area of the catchment is 475 Km^2, [WMO (1972)].

5.3 SOIL OF WAKI CATCHMENT

The soil types found in the catchment are as shown in Fig. (5.2). The percentages of area covered by each soil type are:

i) Shallow dark brown or black sandy loams 3.5%

ii) Reddish and reddish brown gritty clay loams 39.7%

Table (5.1)

LIST OF PRECIPITATION OVER RIVER WAKI CATCHMENT (IN MM/DAY)

DAY	APR.	MAY	JUNE	JULY	AUG.	SEP.	OCT.	NOV.
1	4.90	14.30	2.90	21.50	2.70	0.00	0.90	15.10
2	3.30	1.10	2.60	0.00	11.20	3.40	22.90	10.80
3	6.40	4.30	0.00	1.10	8.10	6.00	4.20	0.20
4	4.00	9.60	0.00	0.40	7.20	7.90	8.10	5.30
5	4.30	1.30	2.50	5.20	0.50	7.80	2.40	4.30
6	9.30	4.00	2.30	2.70	2.70	2.10	2.90	1.40
7	4.10	0.40	14.40	6.20	0.00	13.70	0.00	0.00
8	1.20	7.80	4.10	0.10	11.90	10.90	5.90	6.50
9	1.50	6.60	2.00	0.00	6.60	10.40	1.50	5.10
10	6.10	0.80	0.00	0.00	0.80	14.90	0.00	0.00
11	10.10	1.70	0.00	0.00	5.70	4.50	2.10	0.00
12	2.80	3.70	0.60	0.00	18.10	2.60	25.40	0.00
13	1.40	5.00	0.00	0.60	2.10	9.80	11.60	2.80
14	4.70	2.70	7.60	0.30	0.00	4.90	7.80	10.50
15	21.00	0.00	15.20	0.00	7.50	1.10	18.40	0.90
16	11.30	0.00	1.40	0.00	1.80	1.30	0.20	19.90
17	10.10	4.90	0.70	3.20	7.40	11.70	6.70	2.70
18	2.20	11.20	0.00	0.00	1.60	7.60	4.20	0.00
19	2.70	2.70	3.80	9.90	3.10	0.40	8.50	3.60
20	20.20	0.70	2.80	9.80	0.70	0.00	9.50	0.00
21	0.80	13.00	7.60	6.50	5.70	0.90	9.20	6.00
22	18.60	0.90	0.00	13.20	9.40	4.10	5.00	6.90
23	19.00	4.60	0.00	9.70	11.10	6.60	9.20	5.70
24	0.60	3.40	1.10	12.20	24.50	2.80	0.30	0.80
25	0.50	0.00	0.70	0.00	1.30	0.00	0.20	0.00
26	10.10	8.70	0.60	0.00	2.00	1.70	0.00	1.00
27	0.00	4.00	0.30	0.00	3.90	0.00	8.20	0.00
28	0.00	12.10	4.20	0.00	0.00	0.00	10.90	0.00
29	0.00	7.10	0.10	5.30	0.10	3.60	0.10	0.00
30	11.20	0.70	14.30	0.00	2.30	0.40	4.70	0.00
31		0.00		0.60	1.30		0.80	0.00

YEAR 1970.

Table (5.1) Cont'd.

DAY	APR.	MAY	JUNE	JULY	AUG.	SEP.	OCT.	NOV.
1	4.60	1.30	0.00	2.60	14.60	13.40	4.10	0.00
2	11.60	1.10	1.00	0.50	0.00	6.70	0.10	0.00
3	0.50	1.60	7.70	0.80	0.00	2.70	0.50	0.00
4	6.70	0.00	15.90	3.60	6.40	1.40	8.50	2.80
5	0.40	5.90	22.70	13.20	0.00	6.20	12.20	33.50
6	2.00	0.00	1.90	11.50	17.40	2.00	10.60	0.00
7	5.50	26.80	0.20	5.50	0.10	3.50	1.20	2.70
8	5.40	4.00	0.00	0.40	1.00	4.50	0.00	1.60
9	0.90	4.80	0.00	7.90	4.00	0.00	4.50	0.00
10	1.50	0.20	0.00	0.00	0.10	0.00	0.00	1.80
11	13.00	0.00	0.00	0.00	2.30	9.90	0.00	9.50
12	29.10	1.90	0.90	0.00	0.40	0.60	0.30	15.30
13	9.50	1.50	0.00	6.20	12.70	0.00	6.70	0.50
14	1.00	0.00	2.10	1.60	0.60	0.50	0.30	1.70
15	0.00	2.10	15.10	1.50	4.50	0.00	0.00	0.00
16	5.20	17.40	4.90	4.40	1.10	0.00	0.00	0.00
17	2.20	0.00	0.00	1.60	0.00	3.40	10.30	0.00
18	2.20	0.00	0.00	0.00	1.80	0.00	7.40	0.00
19	7.80	24.50	0.00	4.00	0.10	3.30	6.00	1.40
20	23.10	4.90	0.70	14.00	0.40	0.00	2.40	0.00
21	0.00	13.20	2.70	0.60	3.00	3.00	3.10	0.00
22	4.20	2.40	3.50	5.50	7.60	0.00	5.10	0.00
23	23.30	0.50	0.00	10.20	1.90	2.30	2.90	0.00
24	0.20	0.40	0.00	1.90	14.00	1.70	7.60	0.20
25	3.30	1.30	4.90	14.80	9.10	5.20	1.20	0.30
26	13.40	4.40	0.20	3.30	0.00	1.50	8.10	2.50
27	0.00	0.00	0.90	3.90	3.30	0.60	13.20	3.30
28	0.00	5.60	0.70	9.60	3.70	41.10	14.50	28.90
29	3.50	0.00	0.80	1.40	10.40	0.00	0.00	2.00
30	6.10	0.20	7.00	0.00	0.60	0.00	3.70	0.00
31		2.10		1.80	0.60		0.40	

YEAR 1971

Table (5.2)

LIST OF DISCHARGE FROM RIVER WAKI CATCHMENT (IN MM/DAY).

DAY	APR.	MAY	JUNE	JULY	AUG.	SEP.	OCT.	NOV.
1	0.4800	0.8000	0.7300	0.4900	0.4700	0.6600	0.6100	0.9400
2	0.5000	0.9100	0.7000	0.6100	0.4800	0.6400	0.6200	1.0500
3	0.5100	0.8400	0.6700	0.5800	0.5300	0.6400	0.8000	1.1300
4	0.5300	0.8200	0.6200	0.5500	0.5700	0.6700	0.8100	1.0400
5	0.5500	0.8700	0.5900	0.5200	0.6000	0.7000	0.8700	1.0400
6	0.5500	0.8100	0.5700	0.5300	0.5900	0.7400	0.8500	1.0300
7	0.5900	0.7800	0.5600	0.5200	0.5800	0.7200	0.8400	0.9800
8	0.6000	0.7300	0.6300	0.5300	0.5600	0.8100	0.8000	0.9300
9	0.5800	0.7600	0.6200	0.5100	0.6200	0.8900	0.8400	0.9600
10	0.5700	0.7600	0.6000	0.4800	0.6400	0.9400	0.8300	0.9600
11	0.6000	0.7200	0.5600	0.4600	0.6200	1.0300	0.7900	0.9100
12	0.6500	0.7000	0.5300	0.4400	0.6300	0.9900	0.7900	0.8500
13	0.6400	0.7000	0.5100	0.4300	0.7500	0.9300	1.0400	0.8100
14	0.6200	0.7000	0.4800	0.4200	0.7200	0.9700	1.1200	0.7900
15	0.6300	0.6900	0.5100	0.4100	0.6800	0.9500	1.1400	0.8700
16	0.7600	0.6600	0.5900	0.4000	0.7000	0.9000	1.3000	0.8400
17	0.8200	0.6300	0.5800	0.3900	0.6700	0.8400	1.1600	1.0200
18	0.8600	0.6400	0.5500	0.4000	0.7000	0.9000	1.1800	0.9800
19	0.8200	0.7100	0.5200	0.3900	0.6800	0.9200	1.1500	0.9200
20	0.7800	0.6900	0.5200	0.4200	0.6700	0.8500	1.1800	0.9100
21	0.9400	0.6700	0.5200	0.4600	0.6400	0.7900	1.2200	0.8500
22	0.8800	0.7400	0.5500	0.4800	0.6600	0.7500	1.2300	0.8700
23	1.0100	0.7100	0.5200	0.5400	0.7100	0.7500	1.1900	0.9100
24	1.1400	0.7100	0.5000	0.5800	0.7700	0.7600	1.2200	0.9200
25	1.0200	0.7100	0.4800	0.6300	0.9700	0.7500	1.1100	0.8800
26	0.9300	0.6700	0.4700	0.5900	0.9000	0.7100	1.0200	0.8400
27	0.9600	0.7100	0.4600	0.5500	0.8500	0.6900	0.9700	0.8100
28	0.8700	0.7100	0.4400	0.5200	0.8200	0.6600	1.0100	0.7800
29	0.7900	0.7900	0.4500	0.5000	0.7600	0.6300	1.0800	0.7500
30	0.7400	0.8100	0.4300	0.5100	0.7100	0.6300	1.0000	0.7300
31		0.7600		0.4900	0.6900		0.9900	

YEAR 1970

Table (5.2) Cont'd.

DAY	APR.	MAY	JUNE	JULY	AUG.	SEP.	OCT.	NOV.
1	0.4100	0.5300	0.5500	0.3700	0.4700	0.5000	0.6200	0.6900
2	0.4200	0.5200	0.5100	0.3700	0.5400	0.5700	0.6200	0.6400
3	0.4600	0.5100	0.4900	0.3600	0.5100	0.5900	0.5900	0.6100
4	0.4200	0.5100	0.5100	0.3500	0.4800	0.5900	0.5800	0.5800
5	0.4400	0.4900	0.5900	0.3500	0.5000	0.5600	0.6200	0.5700
6	0.4100	0.5100	0.7300	0.4000	0.4900	0.5800	0.6900	0.8400
7	0.4000	0.4900	0.6900	0.4400	0.5800	0.5700	0.7400	0.7900
8	0.4100	0.6700	0.6300	0.4500	0.5400	0.5700	0.7000	0.7600
9	0.4200	0.6700	0.5700	0.4300	0.5200	0.5700	0.6600	0.7200
10	0.4000	0.6600	0.5300	0.4600	0.5300	0.5400	0.6600	0.6700
11	0.3900	0.6200	0.4900	0.4300	0.5100	0.5100	0.6200	0.6500
12	0.4400	0.5900	0.4600	0.4100	0.5000	0.5500	0.5900	0.6900
13	0.6000	0.5800	0.4300	0.4800	0.4800	0.5300	0.5600	0.7900
14	0.6200	0.5600	0.4100	0.4900	0.5400	0.5100	0.5800	0.7400
15	0.5900	0.5400	0.4000	0.4700	0.5200	0.4900	0.5600	0.7200
16	0.5400	0.5300	0.4700	0.4500	0.5300	0.4700	0.5400	0.6700
17	0.5400	0.6400	0.4700	0.4500	0.5100	0.4500	0.5200	0.6300
18	0.5100	0.6100	0.4400	0.4300	0.4800	0.4600	0.5700	0.5900
19	0.4900	0.5800	0.4200	0.4100	0.4700	0.4500	0.6000	0.5700
20	0.5100	0.7500	0.4000	0.4100	0.4500	0.4500	0.6300	0.5500
21	0.6300	0.7300	0.3900	0.4700	0.4400	0.4400	0.6200	0.5300
22	0.5800	0.8000	0.3800	0.4500	0.4400	0.4400	0.6200	0.5100
23	0.5600	0.7600	0.3800	0.4600	0.4700	0.4300	0.6200	0.4900
24	0.7100	0.7100	0.3700	0.4900	0.4500	0.4300	0.6100	0.4700
25	0.6300	0.6700	0.3500	0.4700	0.5200	0.4300	0.6400	0.4600
26	0.6000	0.6400	0.3600	0.5400	0.5500	0.4500	0.6200	0.4500
27	0.6500	0.6300	0.3600	0.5200	0.5200	0.4400	0.6500	0.4600
28	0.6000	0.6000	0.3500	0.5200	0.5200	0.4300	0.7300	0.4700
29	0.5500	0.6100	0.3500	0.5400	0.5200	0.7100	0.8200	0.6600
30	0.5300	0.5800	0.3500	0.5200	0.5600	0.6700	0.7600	0.6400
31		0.5500		0.4800	0.5300		0.7400	

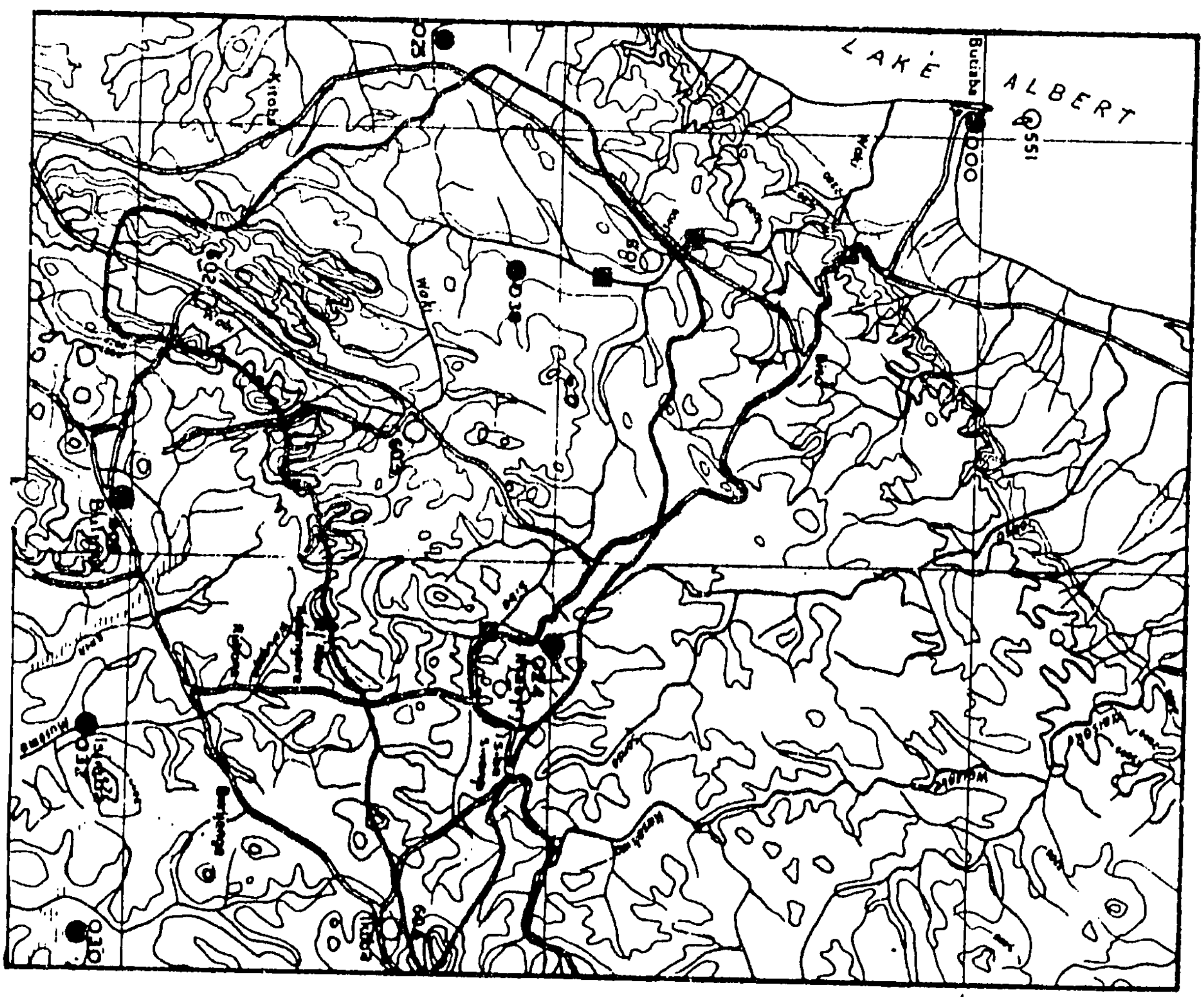
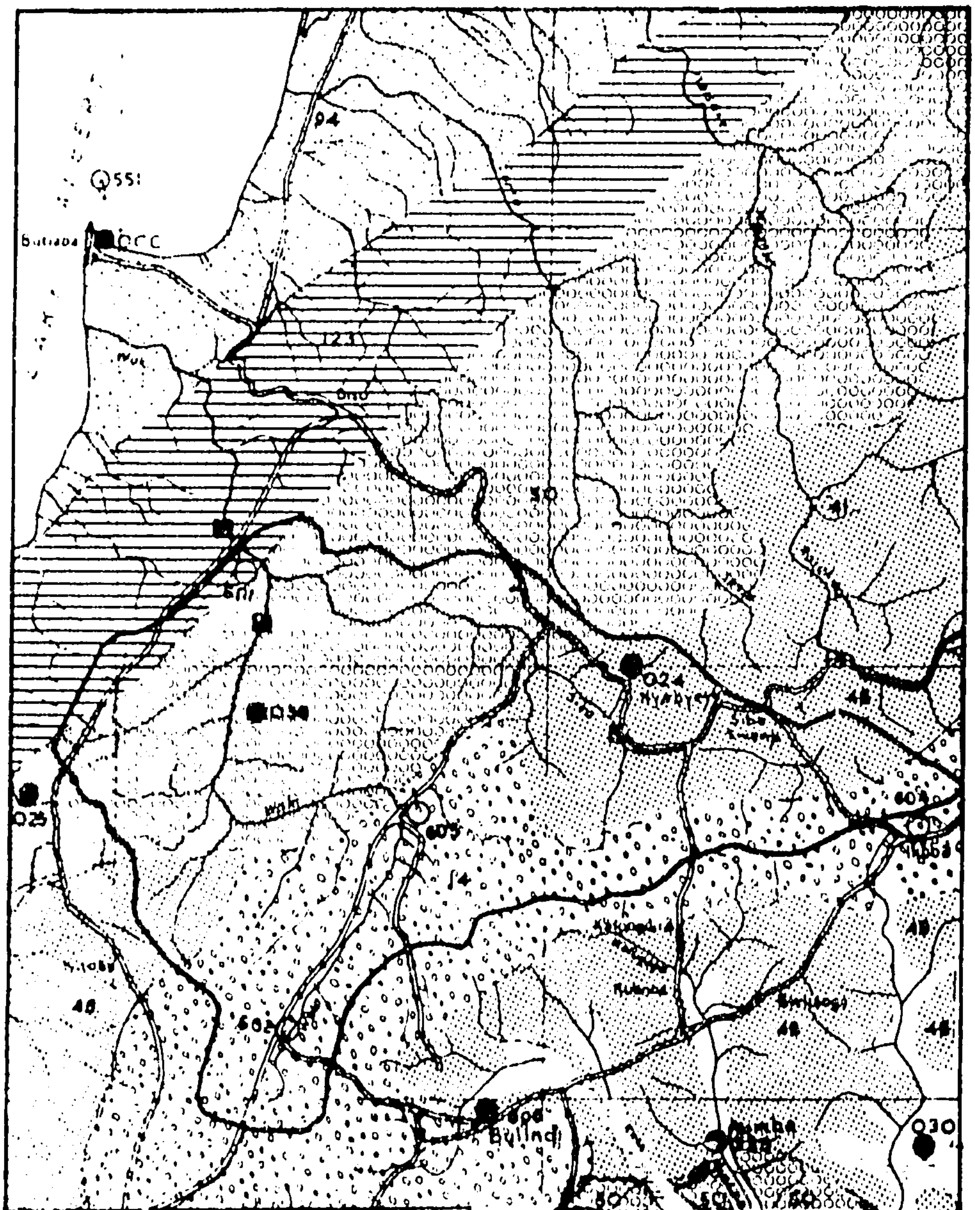


Fig.(5.1) WAKI II CATCHMENT_RELIEF.



SOILS

	Series	Soil Type	Material
94	WEIGA COMPLEX	Black clays and sands	Recent Lake and river Alluvium
123	BUGANGARI	Shallow dark brown or black sandy loams often very stony	Granites gneisses, schists amphibolites
50	RUKIRI COMPLEX	Reddish and reddish brown gritty clay loams	B.C. granites and gneisses and schists
41	BURULI CATENA	Reddish brown sandy loams and loams on laterite	B.C. gneisses and granites
14	KITONYA CATENA	Dark red clay loams occasionally lateritised.	K.A. phyllites
48	HOIMA CATENA	Yellowish red clay loams occasionally shallow over phyllites	Bunyoro series tillites and phyllites

Fig. (5.2) WAKI II CATCHMENT- SOILS.

iii) Dark redy clay loams occasionally lateritized 35.2%

iv) Yellowish red clay loams occasionally shallow over phyllites 21.6%

5.4 GEOLOGY OF WAKI CATCHMENT

The geological structure of the catchment is illustrated in Fig. (5.3). The percentages of areas for the two types of rock formation in the catchment are:

i) Undifferentiated gneisses including elements of P(B) and, in the north, granulite facies rocks 36.9%

ii) Bunyoro series and Kyoga series: shales arkoses and quartzites with tillites, like rocks in the Bunyoro series 63.1%

5.5 VEGETATION OF WAKI CATCHMENT

The vegetation types in the Waki catchment are given in Fig. (5.4). The percentages of area covered with the different types of vegetation are:

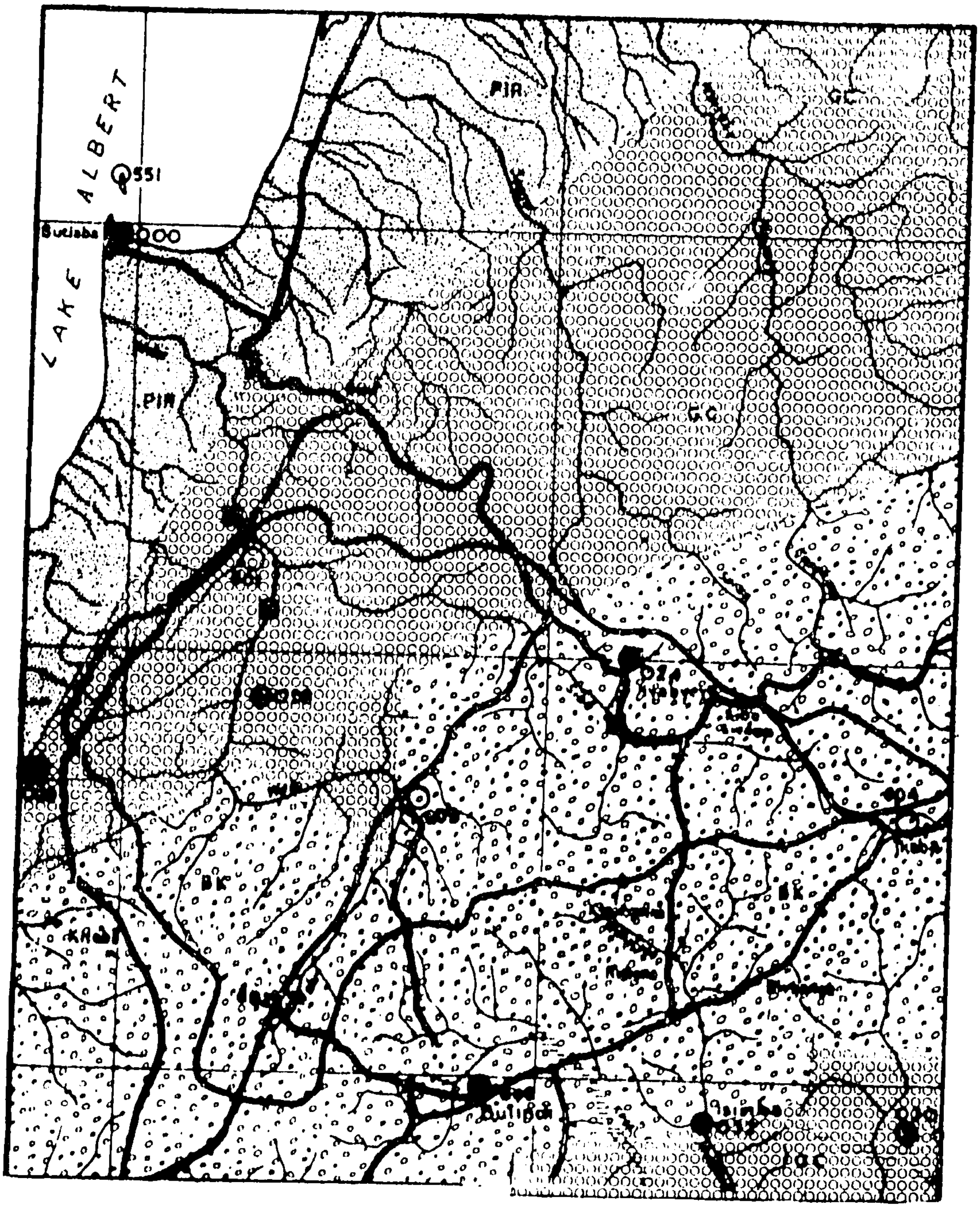
i) Dry combretum savannah 13.8%

ii) Moist combretum savannah 10.8%

iii) Medium altitude moist semi-deciduous forests 26.6%

iv) Forest / savannah mosaics 47.7%

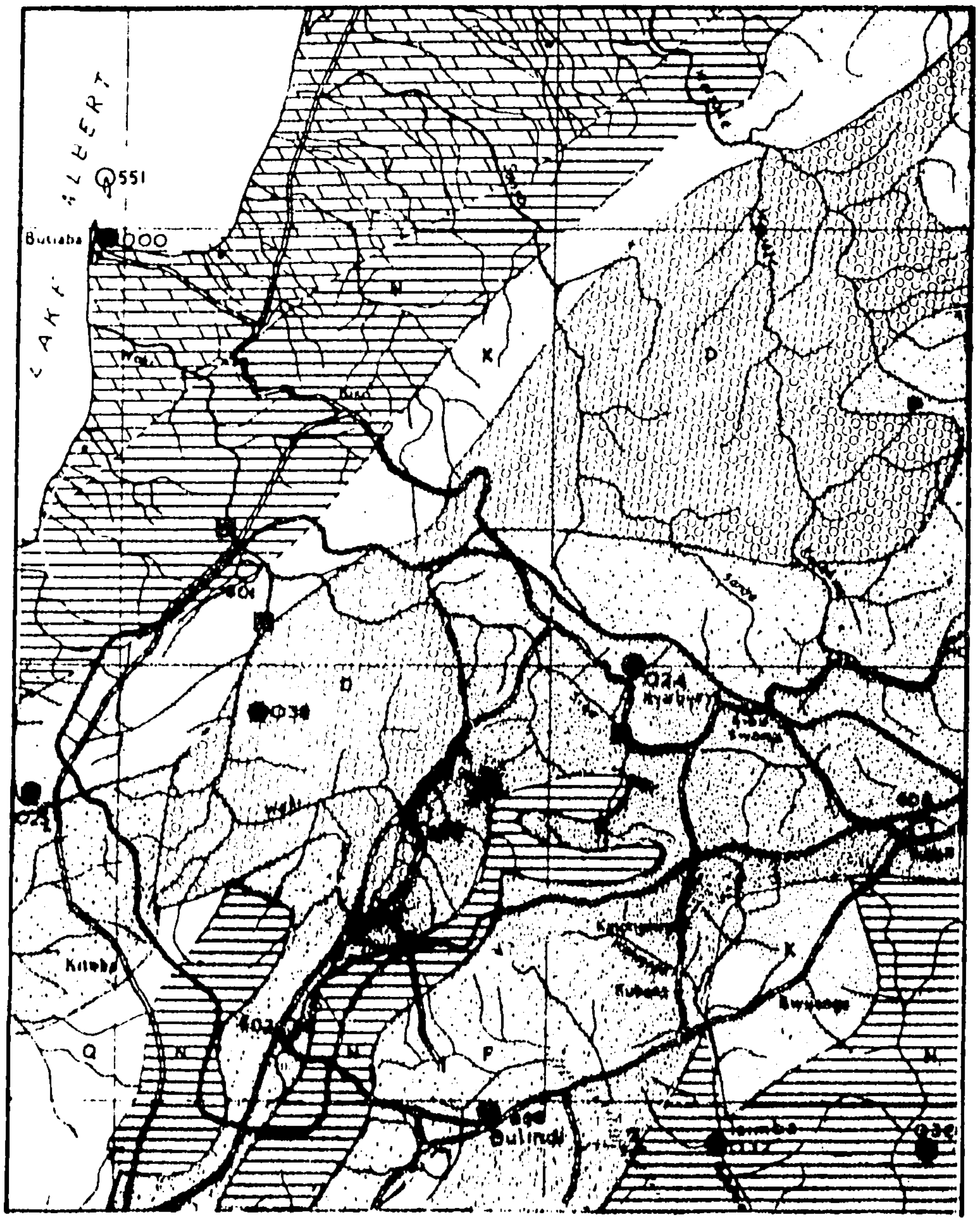
v) Grass savannah 1.1%



GEOLOGY

PIR	Rift Valley Sediments
G-C	Undifferentiated gneisses including elements of P(B) and, in the north, granulite facies rocks
B-K	Bunyoro series and Kyoga series; shales arkoses and quartzites with tillites-like rocks in the Bunyoro series.

Fig.(5.3) WAKI II CATCHMENT_GEOLOGY.



VEGETATION

V	Dry thickets
N	Dry combretum savannas
K	Moist combretum savannas
D	Medium altitude moist semi-deciduous forests
P	Forest/savannah mosaics
Q	Grass savannahs

Fig.(5.4) WAKI II CATCHMENT_VEGETATION.

5.6 AREA VERSUS ELEVATION FOR WAKI CATCHMENT

Areas of the Waki catchment between contours of 200 feet intervals are given in Table (5.3). Using the relationship between area and elevation shown in Fig. (5.5), it can be seen that an area of 365 Km^2 lies between 3250 and 3640 feet with change in elevation of 490 feet, while the remaining area of 110 Km^2 lies between 3640 and 4600 feet with change in elevation of 960 feet. Weighting the elevation of the two areas, the mean elevation of the catchment comes to 3601 feet approximately, see WMO (1972).

5.7 CLIMATE OF WAKI REGION

There are two climatological stations near the catchment. Station Masindi is located to the east, and station Butiaba lies to the north-west. Statistics of climatic elements of temperature, humidity, rainfall and wind speed for these two stations are given in Tables (5.4) and (5.5) respectively.

5.8 OBSERVATIONAL NETWORKS OVER WAKI REGION

5.8.1 Meteorological Stations

The meteorological stations existing within and around the Waki catchment are shown in Fig. (5.6). The particulars of these stations are illustrated in Table (5.6). It can be observed that there is a dense network of rain gauges in Siba sub-catchment and one self-recording rain gauge in the whole of Waki-II catchment. Most of these stations started its operation in 1970, [WMO (1972)].

5.8.2 Hydrological Stations

Waki-I, Waki-II and Siba are the main hydrological stations found within the Waki catchment. The first station lies on Waki tributary upstream and

Table (5.3)

WAKI II

AREA VS ELEVATION FOR RIVER WAKI II CATCHMENT

Elevation range	Area in Sq. Kms	Cumulative area Sq. Kms.
3250' - 3400'	51.7	51.7
3400' - 3600'	213.4	265.1
3600' - 3800'	121.2	386.3
3800' - 4000'	38.7	425.0
4000' - 4200'	26.0	451.0
4200' - 4400'	18.5	469.5
4400' - 4600'	5.2	474.7

Table (5.4)

CLIMATOLOGICAL STATISTICS FOR SATATION MASINDI

Lat. 01°41'N Long. 31°43'E Alt. 1146 meters

Month	Temperature (1931-1954)				Relative Humidity 1200 GMT. %	Rainfall (1907-1962)			Average wind speed (1938-1962)	
	Average					Monthly Total				
	$\frac{\text{Max.} + \text{Min.}}{2}$ C°	Mean Range C°	Mean Max. C°	Mean Min. C°		Average (mm)	Highest (mm)	Lowest (mm)	0600 GMT. Knots	1200 GMT. Knots
January	23.8	14.2	30.9	16.7	41	29	103	0	4	10
February	24.1	14.1	31.2	17.1	43	55	183	0	4	9
March	24.0	12.8	30.4	17.6	49	103	227	12	4	9
April	23.3	11.5	29.1	17.6	59	157	287	61	4	7
May	22.9	10.7	28.2	17.5	64	148	292	40	4	7
June	22.3	11.2	27.9	16.7	64	99	242	31	3	7
July	21.6	10.6	26.9	16.3	63	111	242	40	3	7
August	21.5	10.7	26.9	16.2	65	141	275	46	4	7
September	21.9	11.5	27.7	16.2	63	143	233	61	4	7
October	22.5	11.7	28.4	16.7	60	144	277	41	4	8
November	22.9	12.2	29.0	16.8	53	118	340	3	4	8
December	22.9	12.9	29.3	16.4	51	44	105	0	4	8
Year	22.8	12.0	28.8	16.8	56	1292	1717	1009	4	8

Table (5.5)

CLIMATOLOGICAL STATISTICS FOR STATION BUTIABA

Lat. 01°50'N Long. 31°20'E Alt. 621 meters

Month	Temperature (1931-1954) Average				Relative Humidity 1200 GMT. %	Rainfall (1904-1962) Monthly Total			Average wind speed (1938-1954)	
	$\frac{\text{Max.} + \text{Min.}}{2}$ C°	Mean Range C°	Mean Max. C°	Mean Min. C°		Average (mm)	Highest (mm)	Lowest (mm)	0600 GMT. Knots	1200 GMT. Knots
January	26.1	7.9	30.1	22.2	66	14	55	0	4	7
February	26.5	7.5	30.2	22.7	67	31	179	0	5	7
March	26.5	7.2	30.1	22.9	68	56	162	13	3	7
April	25.9	7.3	29.6	22.3	70	101	205	24	3	6
May	25.7	7.2	29.3	22.1	70	96	234	8	3	6
June	25.3	7.3	29.0	21.7	69	55	191	4	4	6
July	24.8	7.0	28.3	21.3	70	68	269	5	5	6
August	24.5	6.5	27.8	21.3	70	86	169	22	5	6
September	25.1	7.4	28.8	21.4	70	75	125	10	5	6
October	25.5	7.3	29.1	21.8	71	84	184	14	4	6
November	25.6	7.4	29.3	21.9	69	72	280	3	4	7
December	25.7	7.8	29.6	21.8	67	27	105	0	4	6
Year	25.6	7.4	29.3	21.9	69	165	1263	400	4	6

Table (5.6)
EXISTING METEOROLOGICAL STATIONS
AT WAKI - II CATCHMENT

Sr. No.	Name	Regis-tered No.	Type	Latitude	Longitude	Altitude (Feet)	Date of Start
1.	Waki	8831150	Rainfall	1°43'N	31°22'E	3250	5.7.68
2.	Karongo	8831062	Rainfall	1°41'N	31°30'E	3550	6.9.70
3.	Nyantonzi	8831065	Rainfall	1°39'N	31°29'E	3600	5.9.70
4.	Bubwa	8831149	Rainfall	1°37'N	31°27'E	3500	4.7.68
5.	Kisabagwa	8831048	Rainfall	1°32'N	31°24'E	3900	3.7.68
6.	Siba	8831038	Rainfall	1°39'N	31°23'E	3400	1968
7.	Nyabyeya	8831024	Hydromet St.	1°40'N	31°32'E	3900	
8.	Bwinamira	8831056	Rainfall	1°38'N	31°32'E	3550	18.4.70
9.	Budongo	8831057	Rainfall	1°39'N	31°34'E	3650	15.4.70
10.	Nyankwanzi	8831060	Rainfall	1°37'N	31°34'E	3650	17.4.70
11.	Kitonozi	8831064	Rainfall	1°38'N	31°39'E	3850	4.9.70
12.	Kyabagenyi	8831063	Rainfall	1°38'N	31°35'E	3550	9.9.70
13.	Kikobwa	8831066	Rainfall	1°38'N	31°38'E	3750	2.9.70
14.	Kimanya	8831068	Rainfall	1°35'N	31°31'E	3700	4.9.70
15.	Kaangoire	8831059	Rainfall	1°35'N	31°33'E	3700	16.4.70
16.	Bulyango	8831067	Rainfall	1°38'N	31°33'E	3600	10.9.70
17.	Kabango	8831058	Rainfall	1°39'N	31°35'E	3650	14.4.70

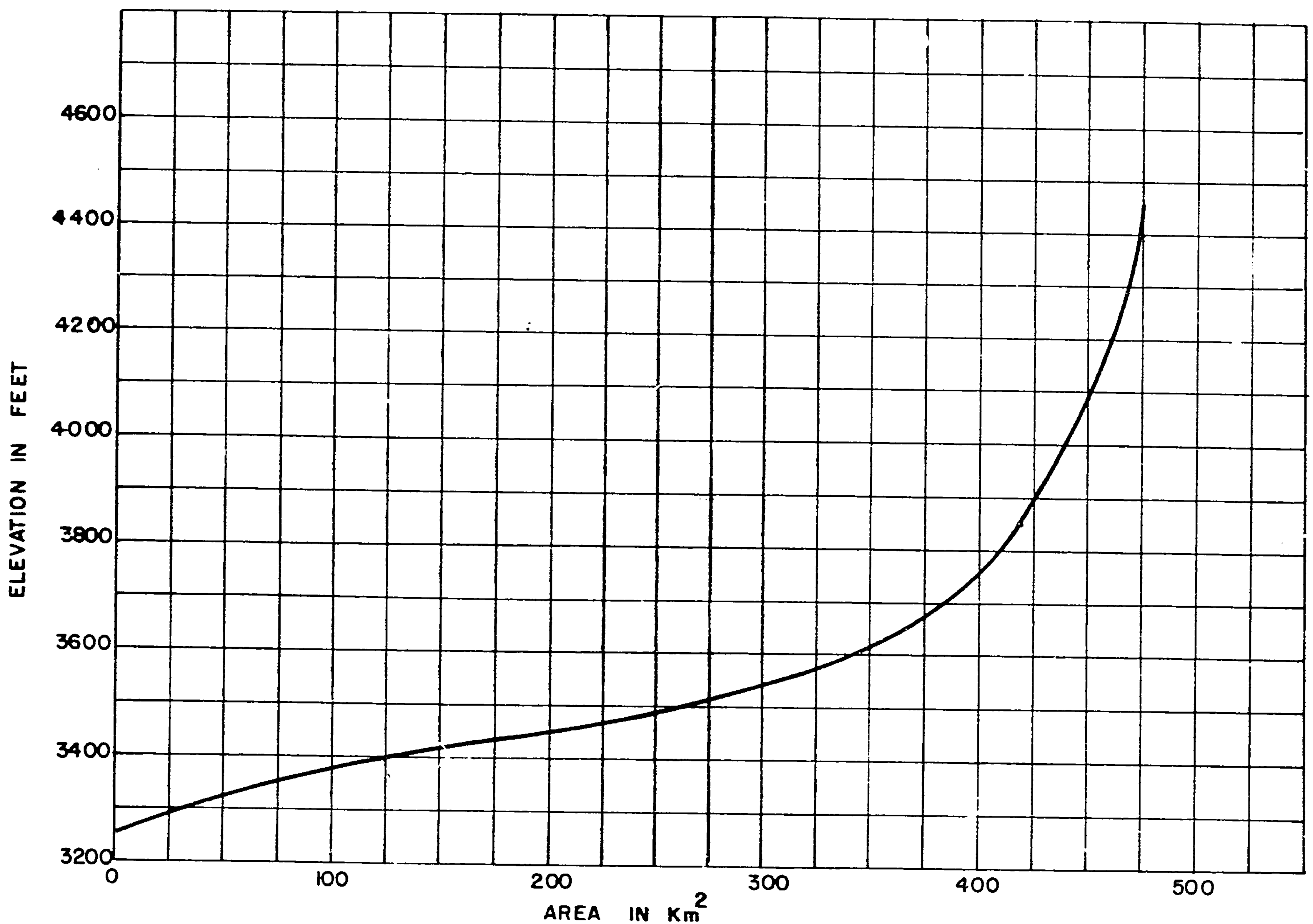


Fig. (5.5) WAKI II AREA_ELEVATION CURVE.

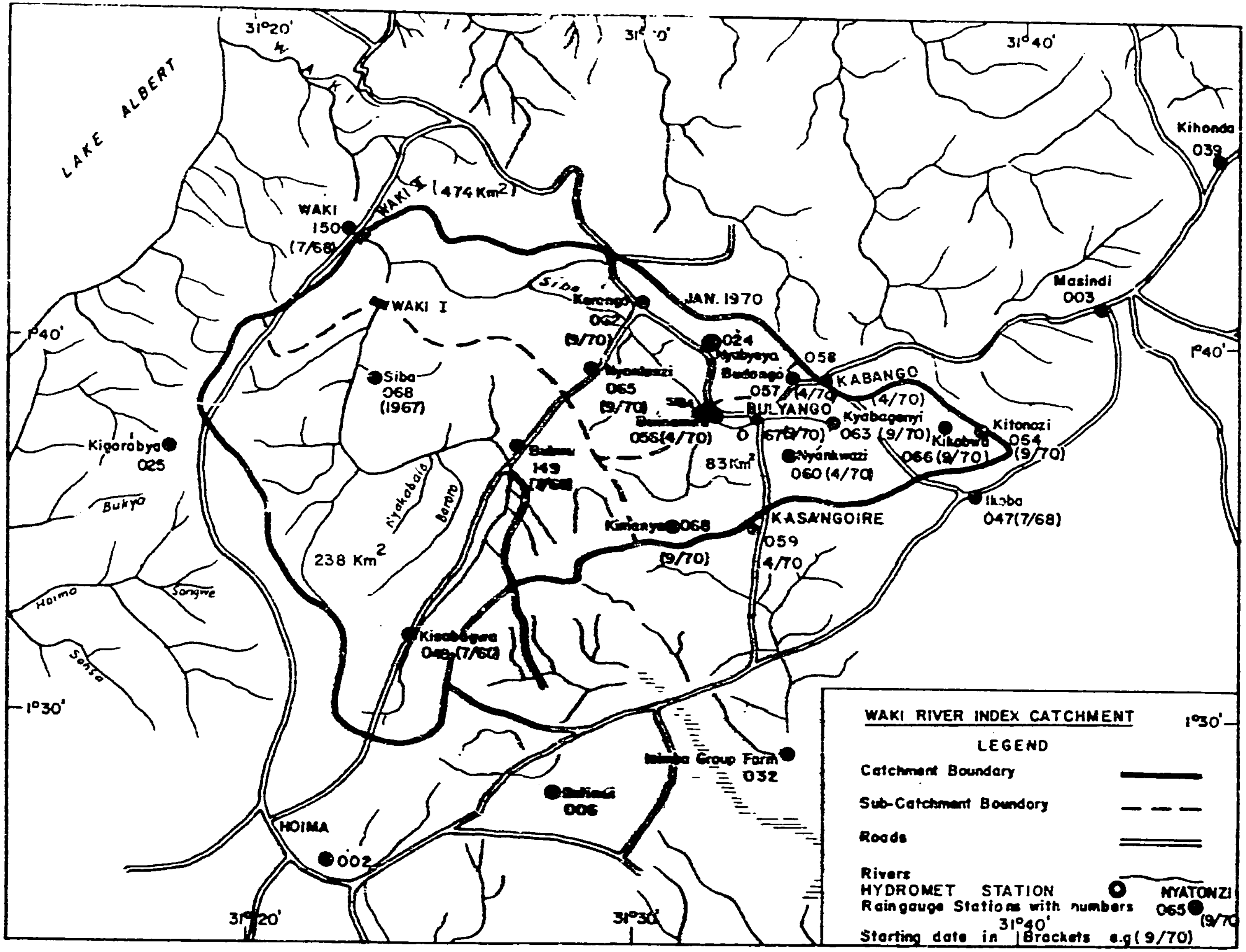

LAKE ALBERT
31°20'
31°40'
1°40'
1°30'
Kihonda
039
Masindi
003
WAKI
150
(7/68)
(474 Km2)
WAKI I
JAN. 1970
Siba
068
(1967)
Kigorobya
025
Bukya
Hoima
Songwe
Sonso
238 Km2
058
KABANGO
(4/70)
BULYANGO
Kyabagenyi
063 (9/70)
Kitonozi
054
(9/70)
83 Km2
Nyankwazi
060 (4/70)
Ikoba
047(7/68)
KASANGOIRE
059
4/70
(9/70)
HOIMA
002
032
006
31°30'
WAKI RIVER INDEX CATCHMENT
LEGEND
Catchment Boundary
Sub-Catchment Boundary
Roads
Rivers
HYDROMET STATION
Raingauge Stations with numbers
Starting date in Brackets e.g (9/70)
NYATONZI
065

Fig.(5.6)

near forestry station while the others are located on the main Waki River and Siba River respectively.

5.9 HYDROLOGICAL ANALYSIS OF DATA

5.9.1 Daily, Monthly and Annual Runoff

For the three hydrological stations of Waki catchment, runoff is evaluated after applying shift corrections to the observed gauges according to the following equation

$$G_o = \frac{1}{8}\left(G_{2.1} + 3G_{1.2} + 3G_{2.2} + G_{1.3}\right) \qquad (5.1)$$

where

G_o : Mean daily gauge.

$G_{1.2}$: First reading of the day under consideration.

$G_{1.3}$: First reading of the next day.

$G_{2.1}$: Second reading of the previous day.

$G_{2.2}$: Second reading of the day under consideration.

The percentages of monthly to annual runoffs are illustrated in Table (5.7). The average values of these percentages range from 4.8 to 12.4 which means that the variations of monthly runoff are not high.

5.9.2 Rainfall - Runoff Relationship

Runoff coefficient for some months of the period of observed data are shown in Table (5.8). Obviously there is a high influence of the storage capacity of the catchment on the hydrological regime since runoff coefficients higher than unity have been obtained in some months. The percentage of annual runoff to annual rainfall ranges from 11 to 15 which is very low.

Table (5.7)

Waki - II Runoff Coefficient

Year & Month	Rain-fall (mm)	Runoff (mm)	Runoff coefficient %	Year & Month	Rain-fall (mm)	Runoff (mm)	Runoff coefficient %
1967				1970			
Nov.	165.2	24.1	15	Jan.	49.3	14.0	28
Dec.	12.2	14.6	120	Feb.	20.5	8.5	42
1968				Mar.	153.5	12.6	8
Jan.	15.9	6.3	40	Apr.	221.1	24.8	11
Feb.	43.8	6.5	15	May	153.6	20.5	13
Mar.	59.6	8.4	14	June	80.1	13.3	17
Apr.	183.6	9.2	5	July	120.9	13.8	11
May	180.4	21.2	12	Aug.	176.0	21.0	12
June	89.0	12.7	14	Sep.	140.3	22.9	16
July	58.5	8.4	14	Oct.	210.6	32.9	16
Aug.	166.6	14.8	9	Nov.	95.0	22.9	24
Sep.	147.7	13.5	9	Dec.	10.5	14.3	136
Oct.	147.5	13.6	9	Annual	1431.0	221.6	15
Nov.	125.5	12.4	10	1969			
Dec.	103.1	18.4	18	Jan.	119.2	12.1	10
Annual	1321.2	145.6	11	Feb.	93.1	11.0	12
				Mar.	124.0	12.5	10
				Apr.	104.9	7.9	8
				May	216.6	21.1	110
				June	88.5	13.2	115
				July	74.8	10.8	14
				Aug.	91.9	9.8	11
				Sep.	118.7	11.7	10
				Oct.	177.8	15.8	9
				Nov.	164.9	19.7	12
				Dec.	88.1	32.5	37
				Annual	1462.5	178.1	12

Table (5.8)

WAKI II

Total Runoff Recession Data

q_o = Initial discharge (C.M.S)

q_t = Discharge (C.M.S) after 12 hours

Period of hydrograph	q_o	q_t	Period of hydrograph	q_o	q_t
2 – 10 December 1967	5.60	5.43	11 – 19 December 1969	6.75	6.30
	5.43	5.22		6.30	5.60
	5.22	4.97		5.60	5.19
	4.97	4.65		5.19	4.77
	4.65	4.30		4.77	4.57
	4.30	4.07		4.57	4.30
	4.07	3.83		4.30	4.09
	3.83	3.64		4.09	3.97
	3.64	3.38		3.97	3.90
	3.38	3.20		3.90	3.77
	3.20	3.00		3.77	3.64
	3.00	2.90		3.64	3.50
	2.90	2.75		3.50	3.37
	2.75	2.65		3.37	3.25
	2.65	2.50		3.25	3.20
	2.50	2.40	25 – 28 April 1970		
2 – 6 May 1968	7.37	6.26		10.25	8.00
	6.25	5.44		8.00	6.70
	5.44	4.90		6.70	5.85
	4.90	4.37		5.85	4.95
	3.37	3.65		4.95	4.28
	3.65	2.97	26 – 29 August 1970		
	2.97	2.55		7.00	5.75
	2.55	2.25		5.75	4.90
25 – 28 November 1969				4.90	4.17
	6.45	5.40		4.17	3.75
	5.40	4.55		3.75	3.40
	4.55	4.09			
	4.09	3.75			
	3.75	3.30			

5.9.2.1 Relationship based on monthly values

For this relationship effective rainfall is used in order to introduce the effect of soil moisture on runoff. The effective rainfall has been calculated from two months of observed data using weighting factors of 0.9 and 0.1, 0.8 and 0.2, 0.7 and 0.3 and so on. Using the rank test, the effective rainfall computed with weighting coefficients of 0.7 and 0.3 is found to be the best. The coefficient of correlation of monthly runoff and monthly effective percipitation is found to be 0.63.

It was found that rainfall - runoff relationship based on monthly data could not be improved further with all months put together. Perhaps a better relationship could be obtained if each month was taken separately.

5.9.2.2 Relationship based on ten-day values

In the view of short time data available for Waki-II catchment, rainfall-runoff relationship was attempted on the basis of ten-day values. Ten-day rainfall, ten-day mean discharge and Antecedent Precipitation Index (API) were used in multiple correlation technique for each month of observed data. After several trials with various API values, it was found that API calculated by the following equation furnishes the best relationship [WMO (1972)]

$$API = o.8P_1 + 0.4P_2 + 0.1P_3 \qquad (5.2)$$

where P_1, P_2 and P_3 are rainfalls of first, second and third ten-day periods.

The coefficient of correlation computed from these relationships came to 0.92 which is quite satisfactory.

5.9.3 Ground Flow Recession Curve

From the observed hydrographs, two hydrographs where the falling limb had reached the ground flow, are selected and plotted on semi-logarithmic paper as illustrated in Figs. (5.7) and (5.8). The ground flow recession is exponentially decayed according to

$$q_t = q_o K^t \tag{5.3}$$

where

q_o : Initial discharge.

q_t : Discharge at time t.

K : Recession constant.

The straight line portion at the end of the falling limb of the two hydrographs gives part of ground flow hydrograph. The value of recession constant K at time t equals to 24 hours is found to be 0.98 in both cases.

5.9.4 Total Runoff Recession Curve

In the separation of compound hydrographs, information of total runoff recession can sometimes be useful. Therefore, a number of observed hydrographs with different peaks are selected and for each hydrograph, values of discharge at intervals of 12 hours are read out starting after the inflection point on the falling limb. A plot of q_o vs q_t was done together for the data of these hydrographs given in Table (5.8) as shown in Fig. (5.9). There is a considerable scattering in the plotted point because the falling limbs of these hydrographs are generally distorted by rain falling over the Waki catchment even after the hydrograph peak has been reached. Therefore, the falling limb of the total runoff hydrograph does not represent simple depletion of the channel storage but is mixed with more surface runoff coming into the streams.

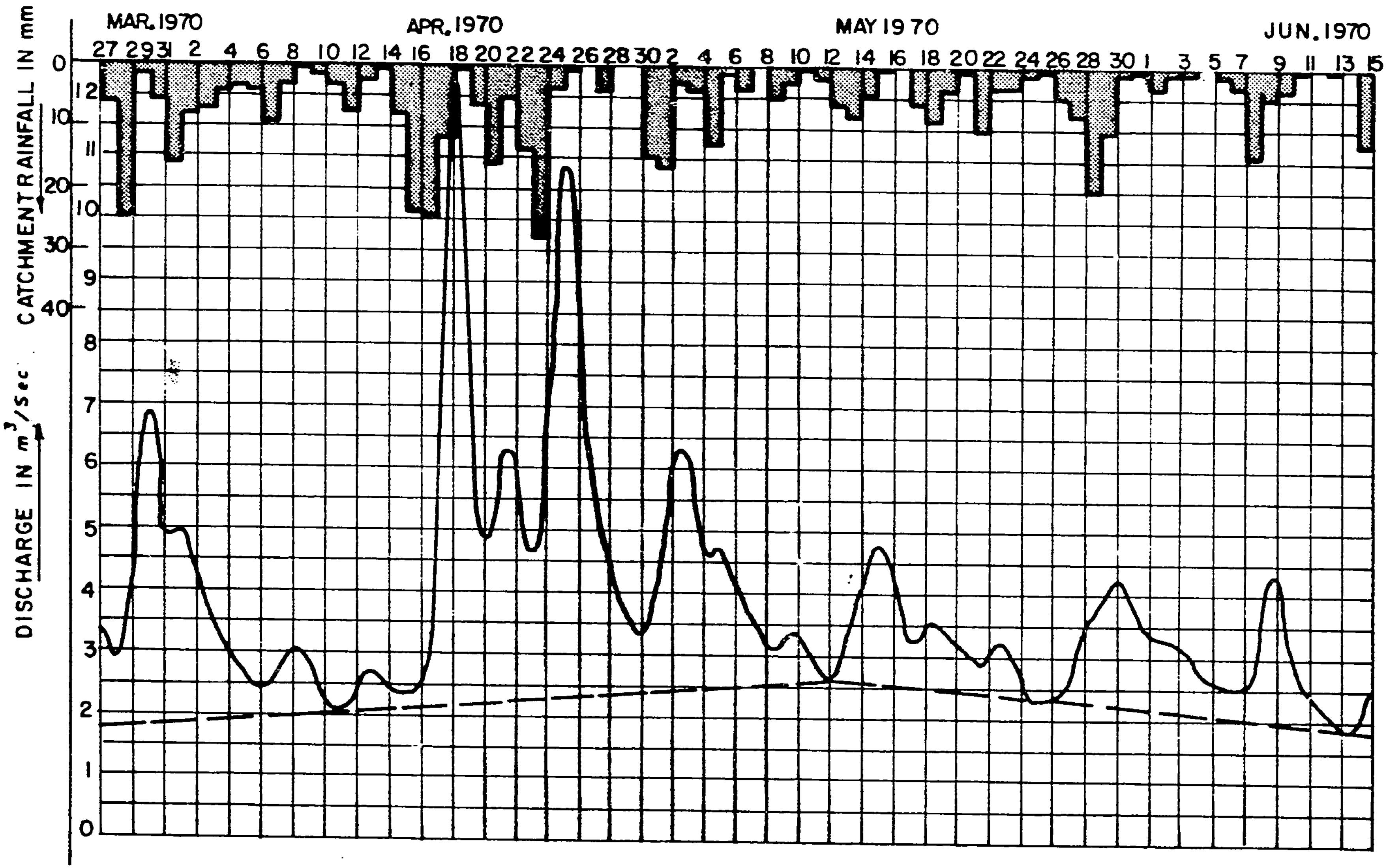


Fig.(5.7) WAKI II OBSERVED HYDROGRAPH AND RAINFALL.

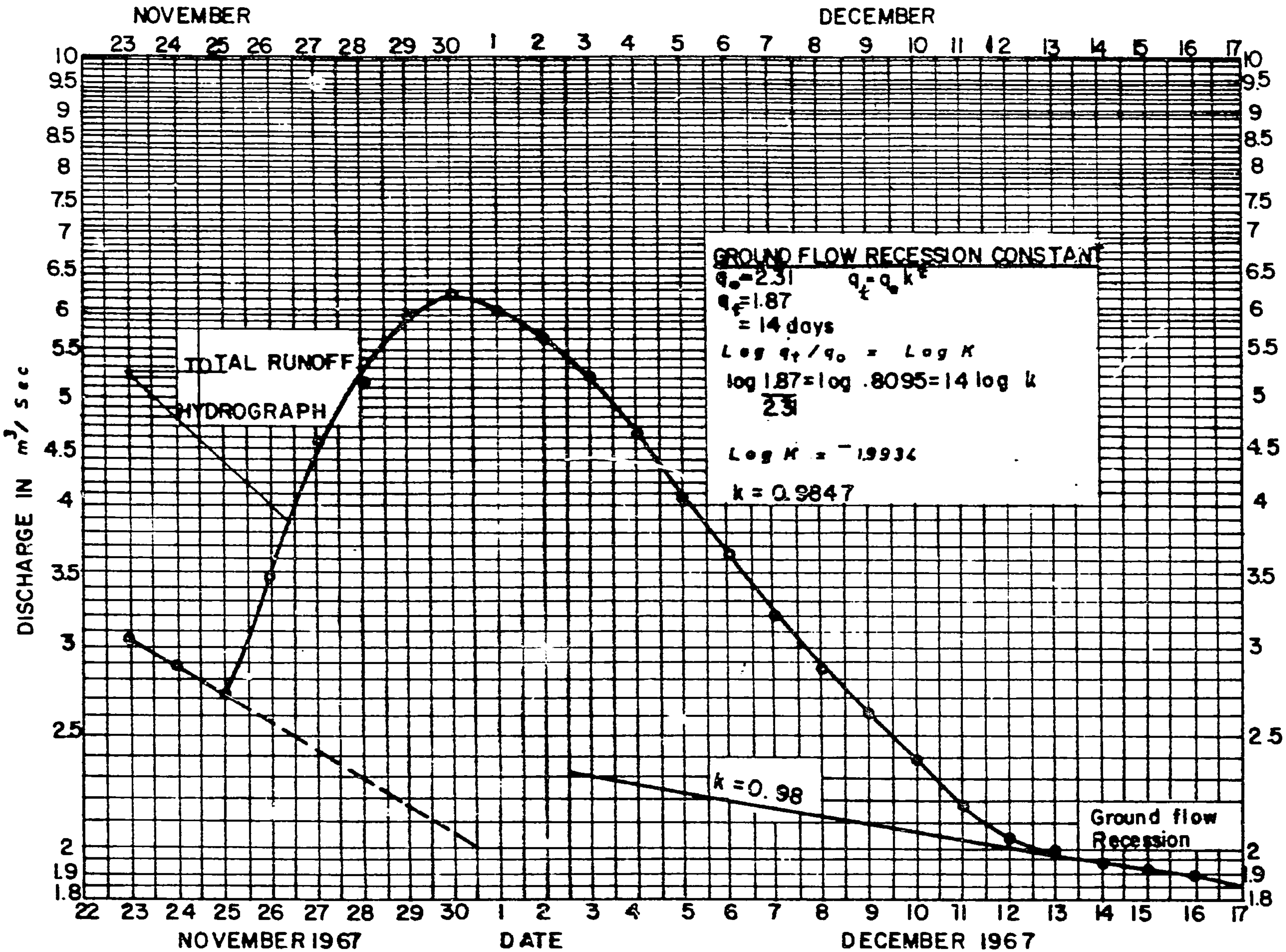


Fig.(5.8) WAKI II COMPOUND HYDROGRAPH (WITHOUT DISTINCT SEPARATE PEAKS) 1967.

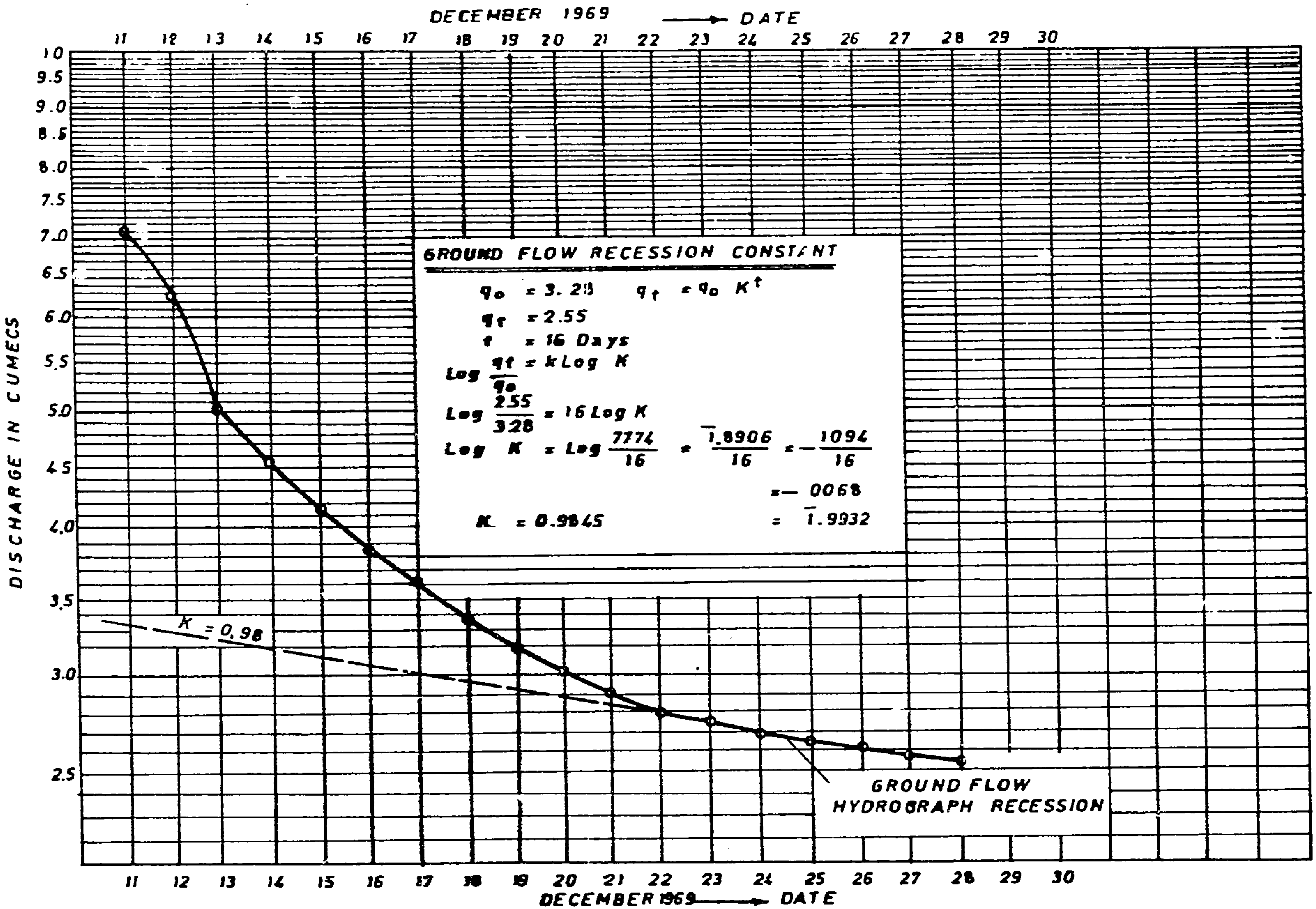


Fig.(5.9) WAKI II _ FALLING LIMB OF HYDROGRAPH.

5.10 ANALYSIS OF TYPICAL HYDROGRAPHS OF WAKI CATCHMENT

The pattern of rainfall of Waki catchment is such that the falling limbs of the hydrographs reach base flow after a long period and the hydrographs are mostly compound. During the rainy season it nearly rains every day and a real break is unusual. For separating the base flow from direct runoff, a simpler procedure is applied. The base flow hydrograph is fixed by joining the lowest points reached by the daily discharge hydrograph when rainfall stopped for some days or was very small. The base flow hydrograph is shown in Fig. (5.10).

As mentioned earlier, it is impossible to find a simple hydrograph, therefore the compound hydrograph observed from 16th to 30th April for 1970 was selected to analyse the unit-hydrograph. As shown in Fig. (5.11), the selected hydrograph has three peaks. Each of these peaks has been produced by three separable rain spells. This hydrograph is therfore composed of three simple hydrographs. The first hydrograph is then used for the determination of the unit hydrograph and its final configuration is shown in Fig. (5.11).

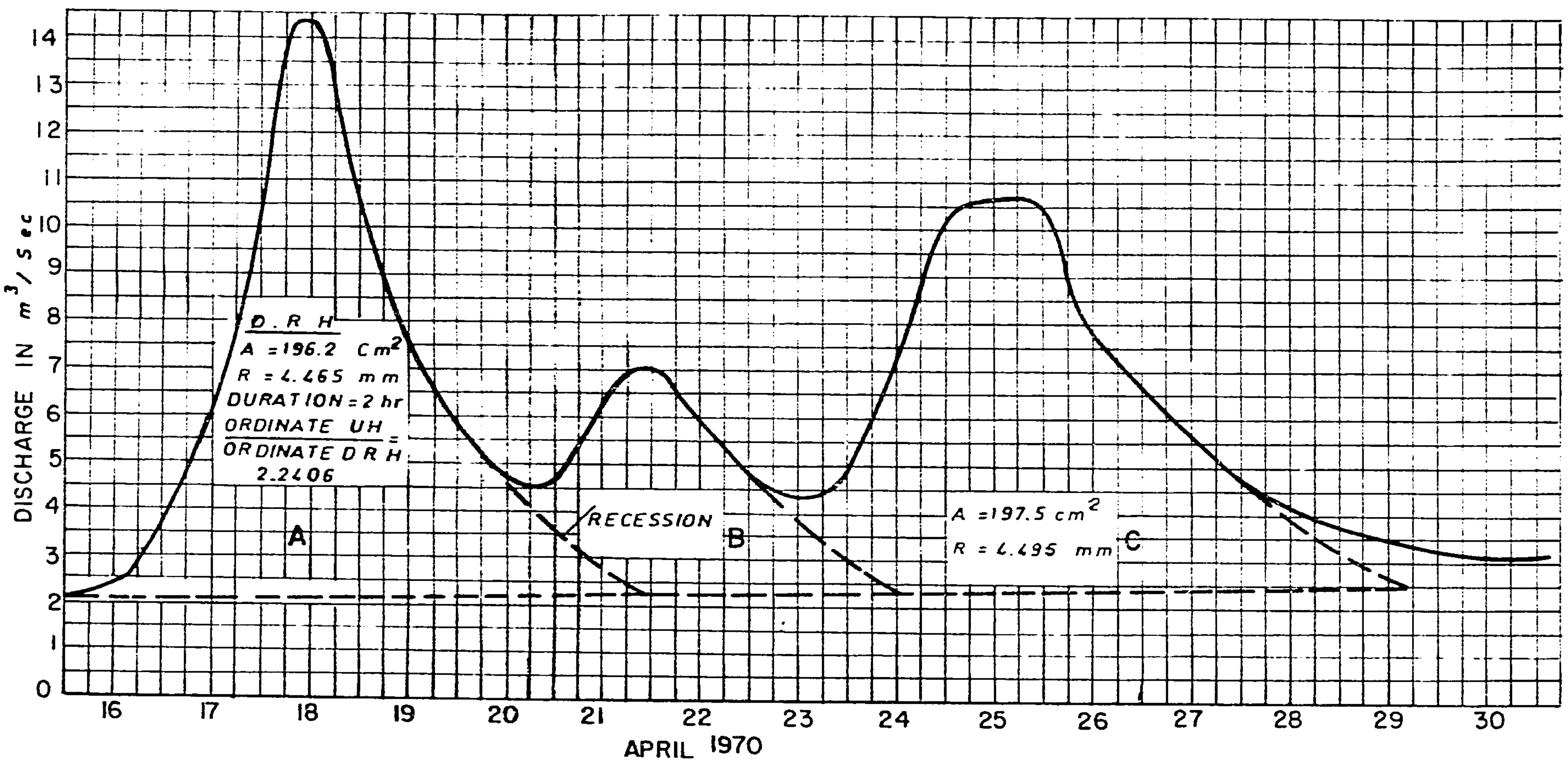


Fig (5.10) WAKI II HYDROGRAPH ANALYSIS.

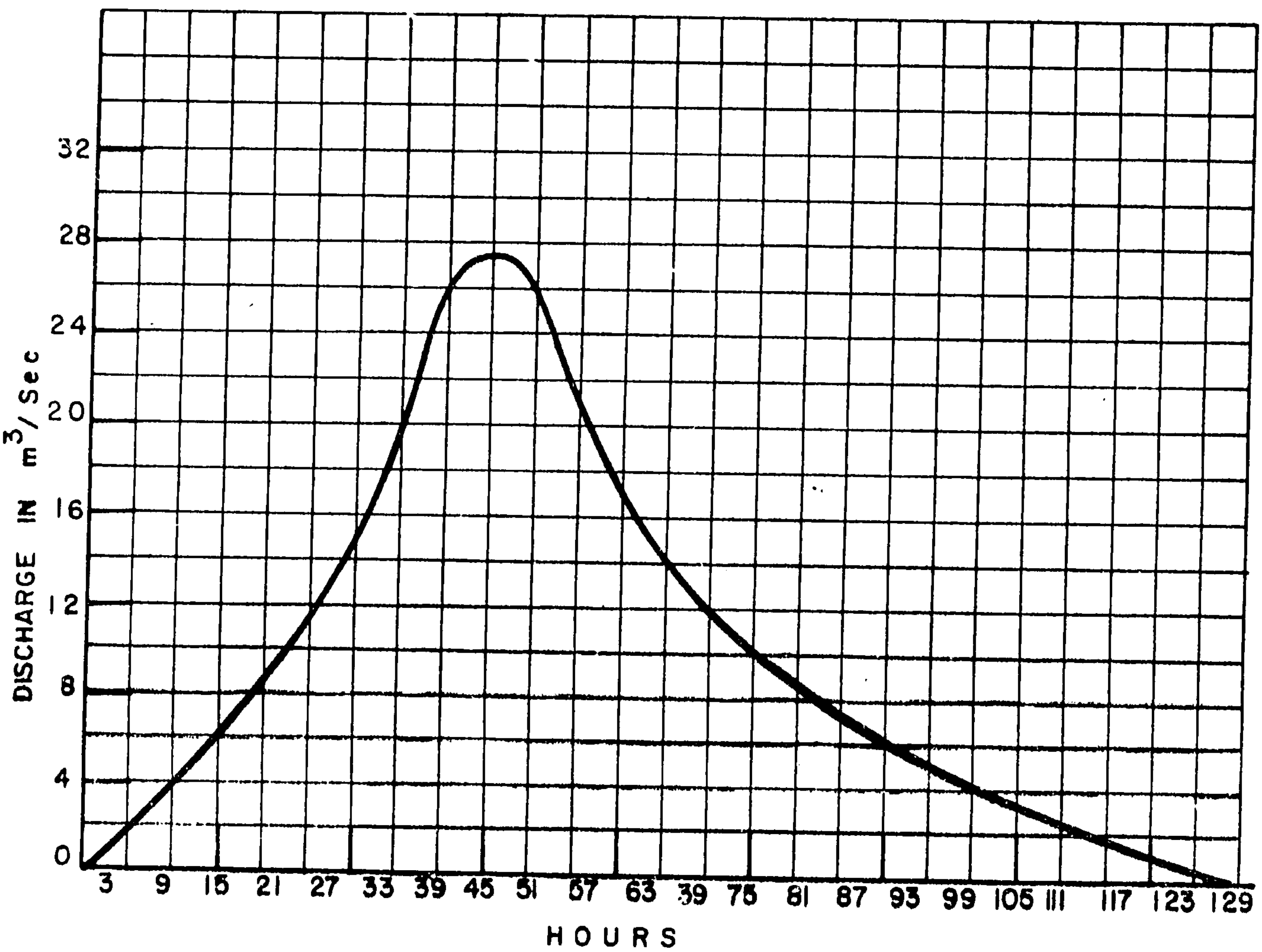


Fig.(5.11) UNIT HYDROGRAPH OF WAKI II RIVER
(DURATION 1/2 HOUR).

CHATER VI
APPLICATION OF THE MODEL BULLDING TECHNIQUES TO WAKI RIVER CATCHMENT

CAHPTER VI

APPLICATION OF THE MODEL BUILDING TECHNIQUES TO WAKI RIVER CATCHMENT

6.1 INTRODUCTION

The construction of mathematical models from observed time series is practiced in a variety of disciplines, including engineering, ecology and applied statistics with specific objectives. For example, Kashyap and Rao have suggested the stochastic difference equation models to represent some hydrological systems.

In application, a plausible classes of models can be obtained by the inspection of the given time series and examination of their characteristics. Consequently, the availability of using either the noisy-transfer function model or the linear stochastic difference equation model for an appropriate simulation of the case study previously presented in Chapter V will be studied in some details.

6.2 SOME FEATURES OF THE CASE STUDY

The data used for this study is selected in the rainy season of Waki catchment which includes eight successive months, starting with April, to avoid data non-stationarity. Therefore, the data length for both the input sequence $[x(\cdot)]$ and output sequence $[y(\cdot)]$ illustrated in Tables (5.1) and (5.2) respectively includes 488 points [WMO (1972)].

6.2.1 Statistical Characteristics of the Observed Data

Consider $\bar{y}$, σ_y and γ_y as the observed mean, standard deviation and skewness coefficient of the measured output data $[y(\cdot)]$, whereas the same notations for the input data $[x(\cdot)]$ are $\bar{x}$, σ_x and γ_x respectively. The variations of these notations with the sample size for both the input rainfall and output discharge are elucidated in Figs. (6.1) and (6.2) respectively.

The cross-correlation coefficient of the output discharge $[y(\cdot)]$ and the input rainfall $[x(\cdot)]$ at different time lags k have been calculated using the formula

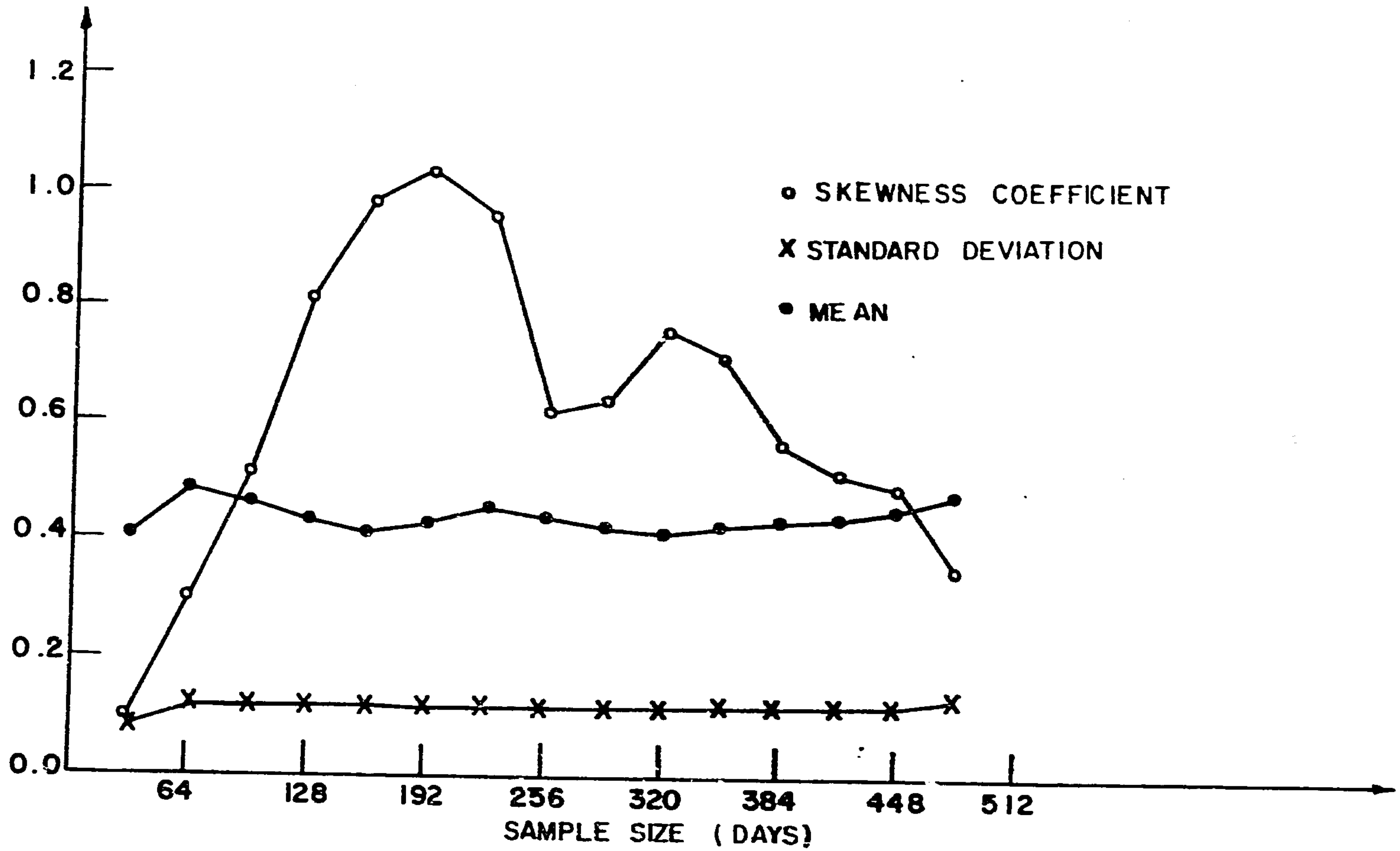
$$R_{yx}(k) = \frac{1}{\sqrt{\sigma_y \sigma_x}\,(N-k+1)} \sum_{i=1}^{N-k+1} [y(i) - \bar{y}]\,[x(k+i-1) - \bar{x}]. \tag{6.1}$$

This yields the results shown in Fig. (6.3), where the maximum value has been located at a time lag equals three days. In practice, this value of time lag represents a very suitable estimate for the time delay factor τ.

Consider the correlograms of measured rainfall and output discharge shown in Figs. (6.4) and (6.5). The first correlogram reflects considerable fluctuations compared with that of the output discharge which shows a little variability. Consequently, the smoothed raw estimates of the power spectrum evaluated for the output discharge reveals a small damping response as delineated in Fig. (6.6). Finally, the probability of both the measured input rainfall and output discharge are shown in Figs. (6.7) and (6.8) respectively.

6.3 APPLICATION OF THE NOISY-TRANSFER FUNCTION MODEL

The basic premise of this study is the appropriate selection of an estimation methodology which yields an adequate results for the case study. . Therefore, we shall consider different structures of the noisy-transfer fun-



Fig(6.1) VARIATION OF THE MEAN, STANDARD DEVIATION AND SKEWNESS COEFFICIENT OF THE MEASURED DISCHARGE DATA WITH THE SAMPLE SIZE.

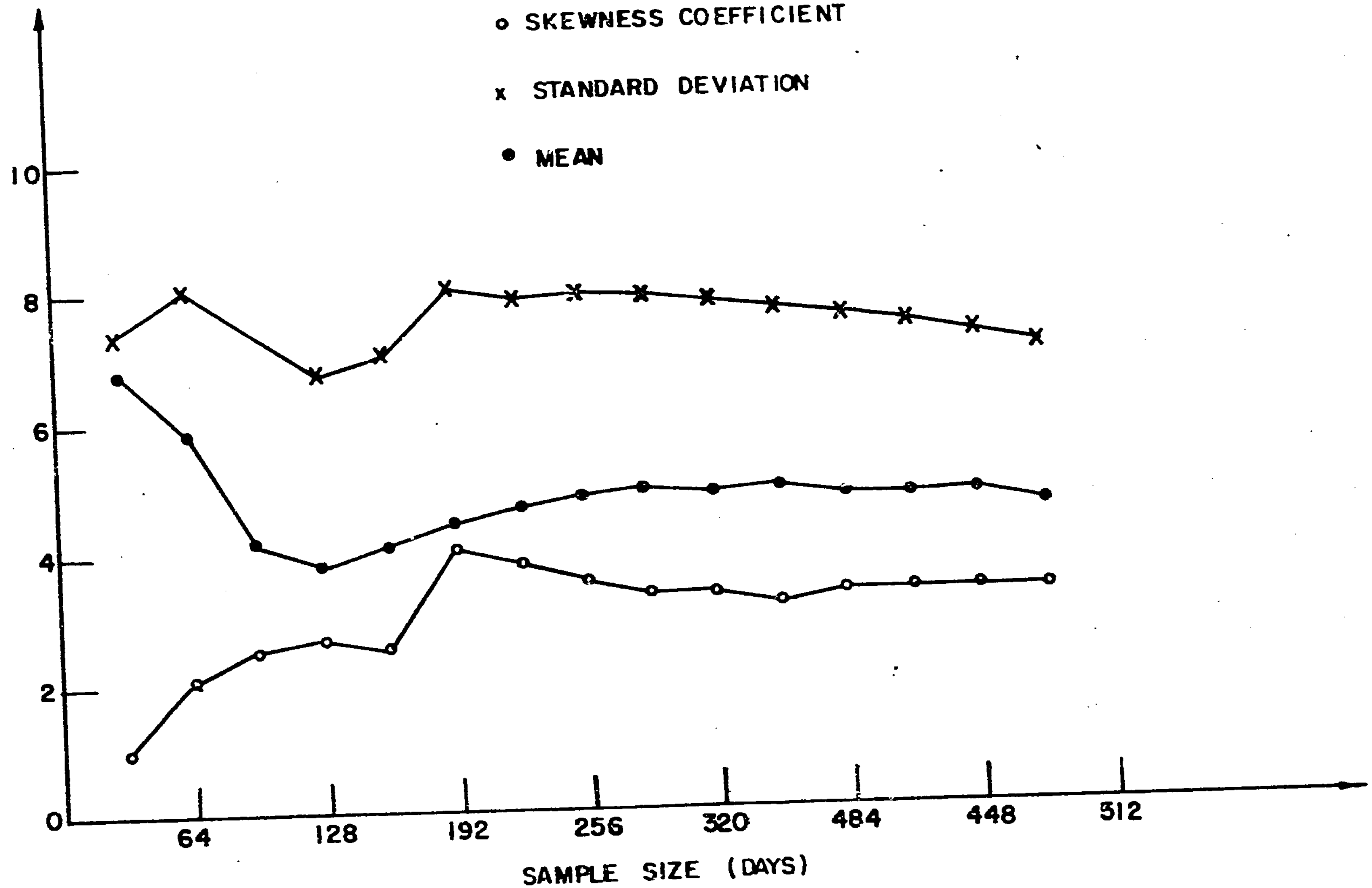


Fig.(6.2) VARIATION OF THE MEAN, STANDARD DEVIATION AND SKEWNESS COEFFICIENT OF THE MEASURED RAINFALL WITH THE SAMPLE SIZE.

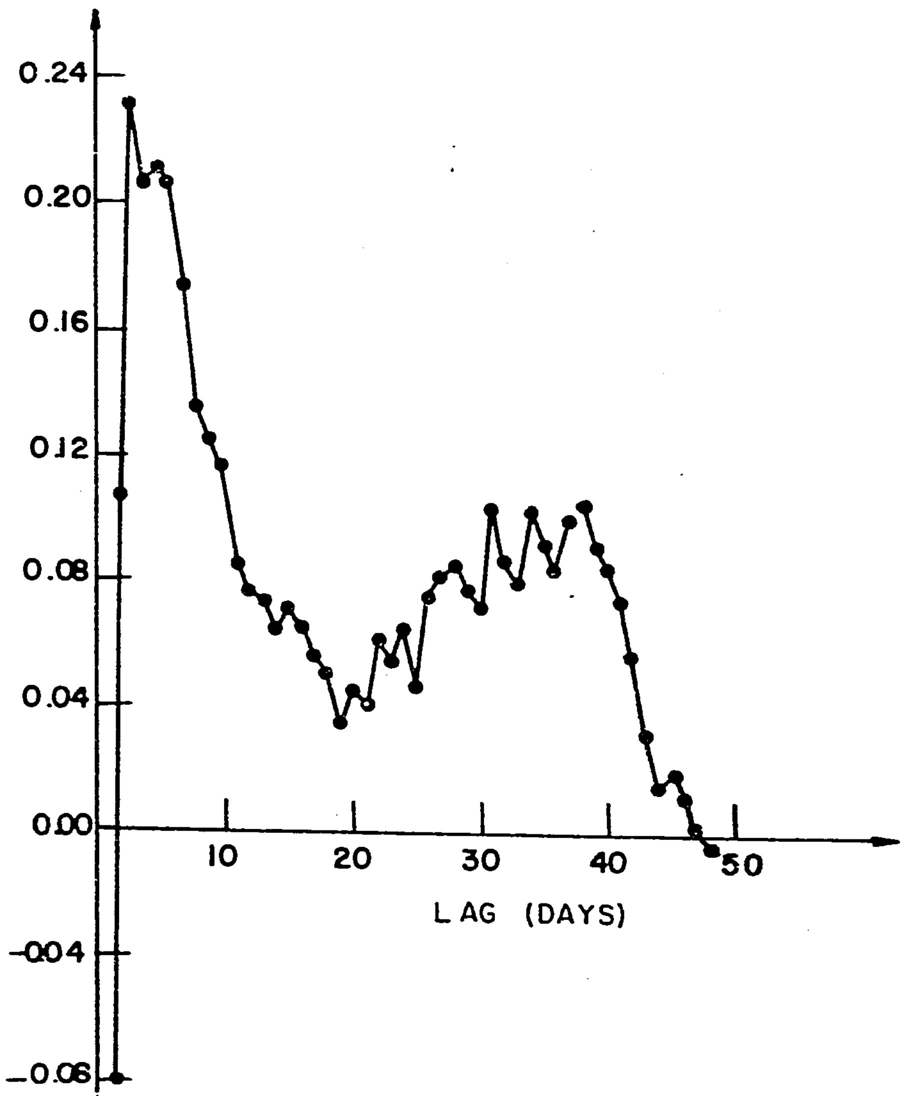
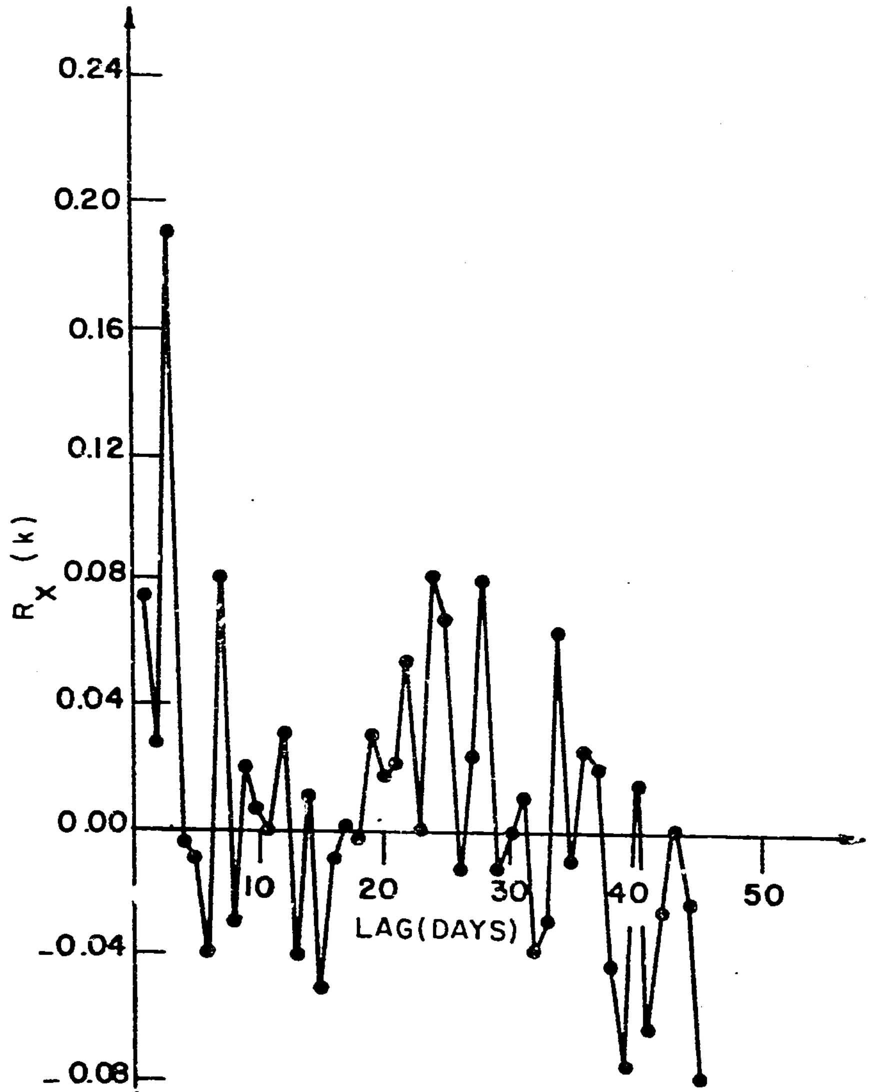


Fig.(6.3) CROSS-CORRELATION COEFFICIENT AT DIFFERENT LAGS.



Fig, (6.4) CORRELOGRAM OF THE MEASURED RAINFALL DATA.

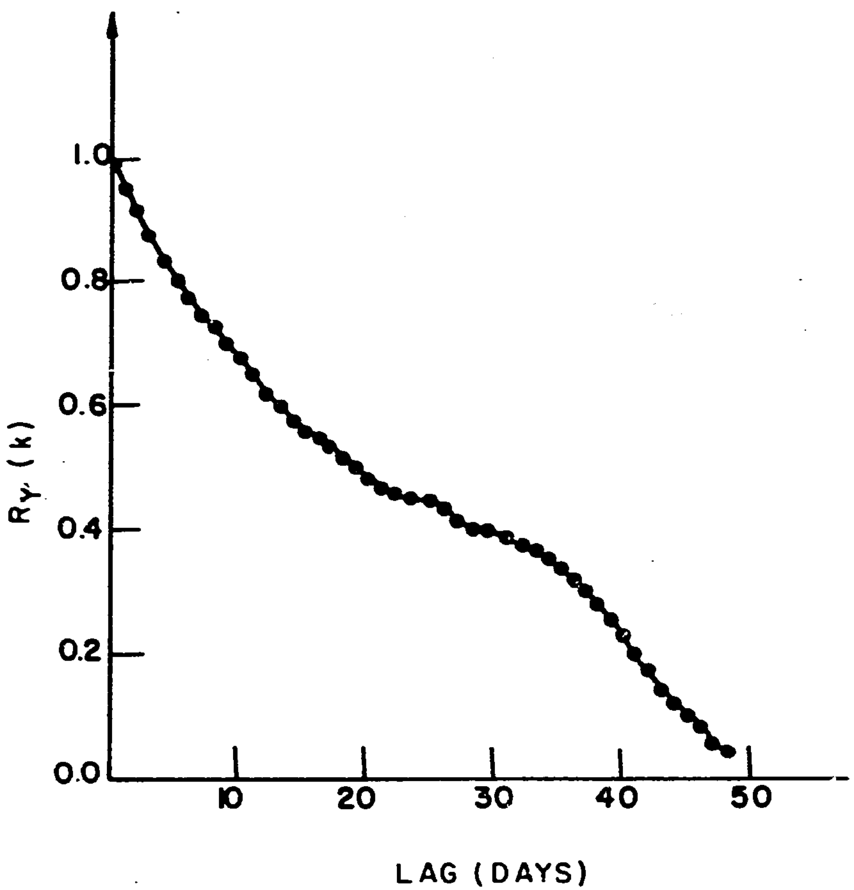


Fig.(6.5)CORRELOGRAM OF THE MEASURED DISCHARGE DATA.

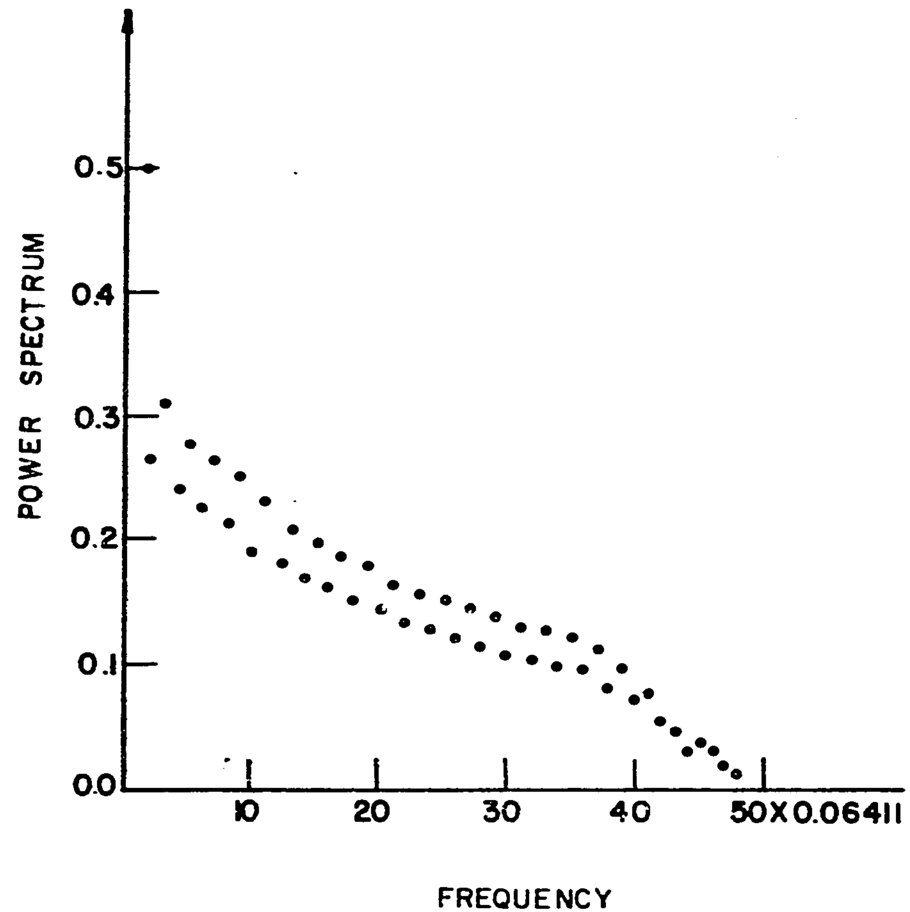


Fig.(6.6)POWER SPECTRUM OF THE MEASURED RAINFALL DATA.

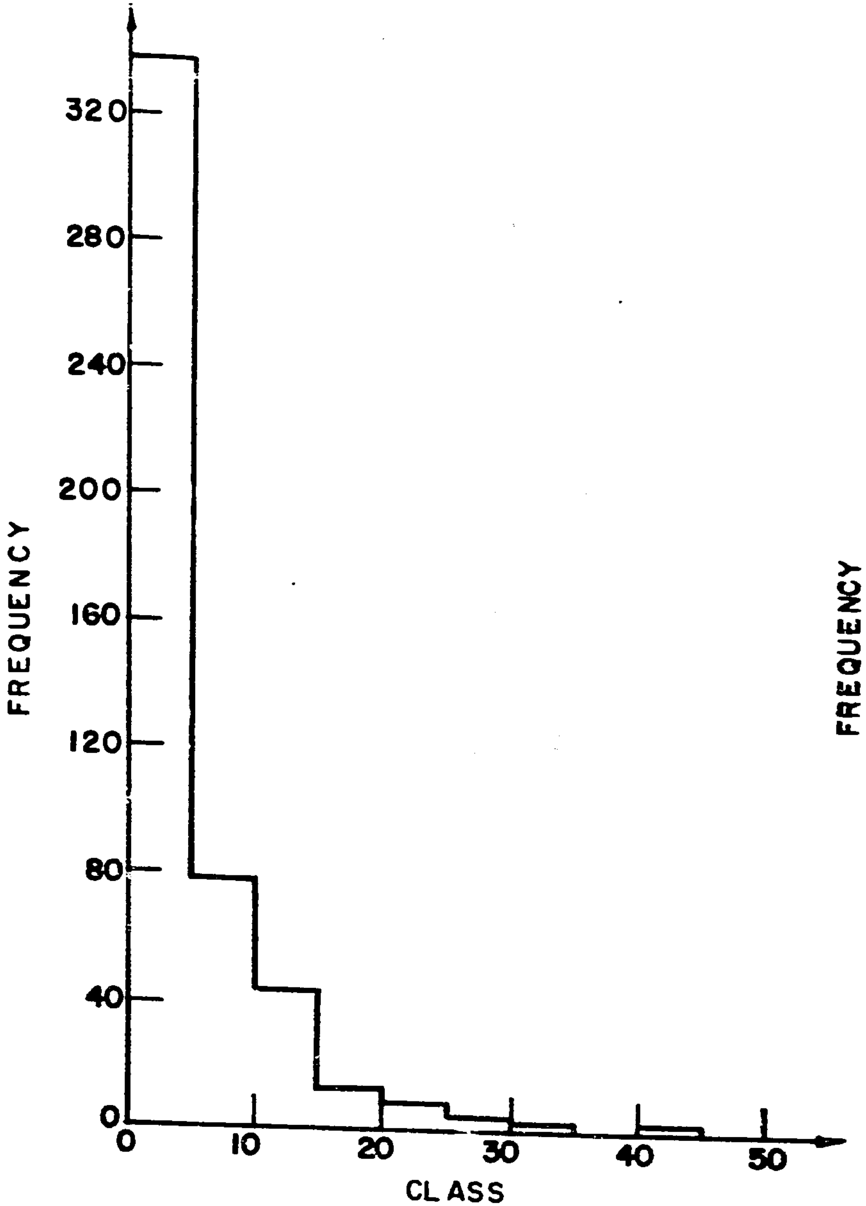


Fig.(6.7) HISTOGRAM OF THE MEASURED RAINFALL DATA.

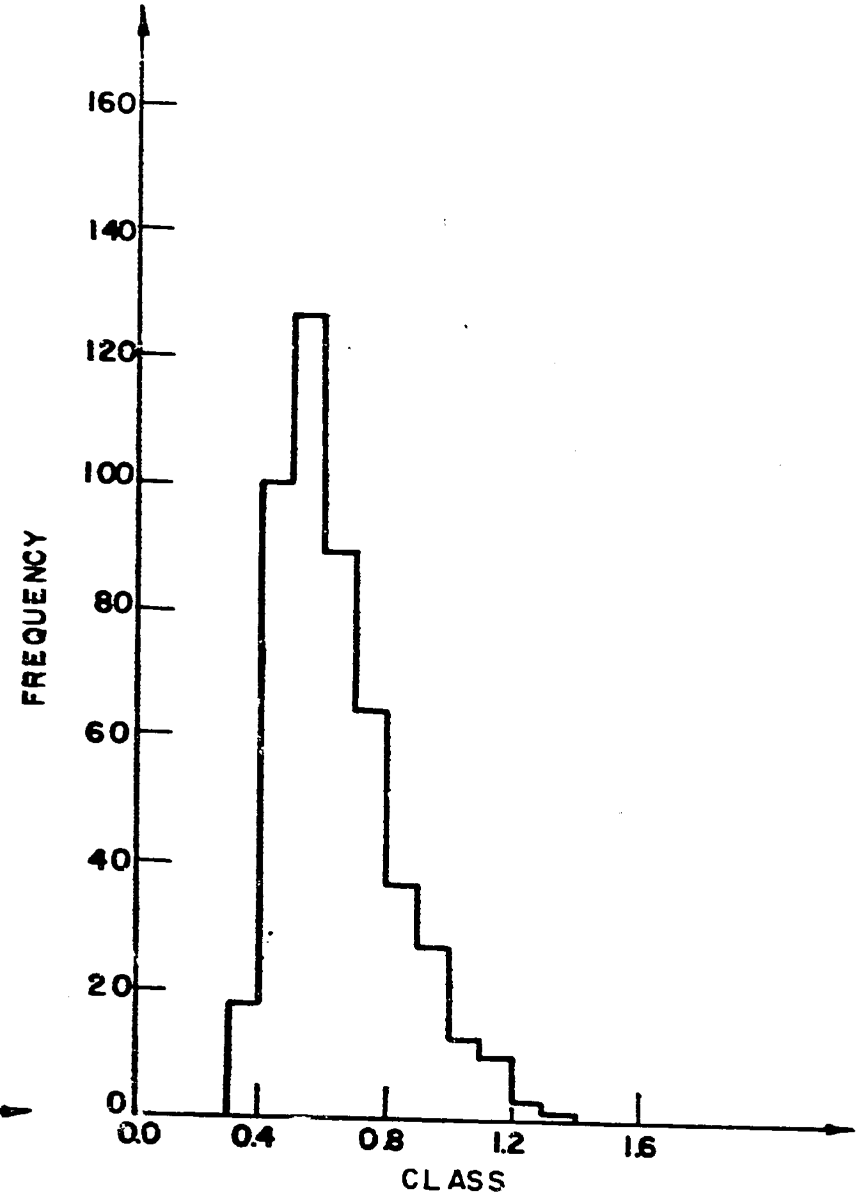


Fig.(6.8) HISTOGRAM OF THE MEASURED DISCHARGE DATA.

ction model on the basis of causality principle. Systematicaly, these structures can be described as follows:

i) The normalized values of the measured input rainfall sequence [x(·)] are mathematically delayed as

$$\ddot{x}_d(k) = \begin{cases} x(k-\tau) & \text{for } k \geq \tau + 1 \\ 0 & \text{for } k < \tau + 1 \end{cases} \tag{6.2}$$

to achieve a better coincidence with the similar values of the output discharge sequence [y(·)]. Practically, the kernel length k_o can be chosen, such that

$$\hat{U}(k_o-1) > \hat{U}(k_o), \tag{6.3}$$

and

$$\sum_{i=1}^{k_o} \hat{U}(i) = \sqrt{\sigma_y / \sigma_x} \ . \tag{6.4}$$

Then, the unconstrained numerical solution may be invoked, together with (6.3) and (6.4), to obtain the values of the impulse response vector $\underline{U}$ since the matrix ($\underline{\underline{H}}^T\ \underline{\underline{V}}^{-1}\ \underline{\underline{H}}$) appears to be ill-conditioned in most of the usual cases [Abadie (1970)].

The evaluated impulse response vector $\hat{\underline{U}}$ together with (6.2) are invoked to estimate the output of the first model M_1', as follows

$$\left.\begin{aligned} \hat{y}(k) &= \sigma_y \left[\sum_{i=1}^{k_o} \hat{U}(i)\ \tilde{x}_d(k-i+1) \right] + \bar{y} \\ \forall k &= 1, 2, \ldots, N. \end{aligned}\right\} \tag{6.5}$$

ii) Further, it is alleged that the autoregressive models have to be preferred since they can achieve much better estimatability conditions for those systems whose complete mathematical description is not available. Thus, the normalized discharge is used to generate the vector $\underline{\tilde{y}}$, as follows

$$\underline{\tilde{y}} = \begin{bmatrix} \tilde{y}(2) \\ \tilde{y}(3) \\ \cdot \\ \cdot \\ \cdot \\ \tilde{y}(N) \end{bmatrix} \tag{6.6}$$

as well as the following matrix

$$\underline{H} = \begin{bmatrix} \tilde{y}(1) & 0 & \cdot & \cdot & \cdot & 0 \\ \tilde{y}(2) & \tilde{y}(1) & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & \cdot & & \cdot \\ \cdot & \cdot & & & \cdot & \cdot \\ \tilde{y}(N-1) & \tilde{y}(N-2) & \cdot & \cdot & \cdot & \tilde{y}(N-k_o) \end{bmatrix}. \tag{6.7}$$

Obviously, the necessary and sufficient condition for estimating the kernel length k_o is

$$\sum_{i=1}^{k_o} \hat{U}(i) = 1. \tag{6.8}$$

The unconstrained numerical solution, together with (6.8), are used to evaluate the impulse response vector $\underline{U}$. Thus, the current estimates of the output data generated from the model M_3 may be expressed as follows

$$\hat{y}(k) = \begin{cases} \sigma_y[\sum_{i=1}^{k_o} \hat{U}(i)\, \tilde{y}(k-i)] + \bar{y}, & k=2,\ 3,\ \ldots,\ N \\ \sigma_y\, \tilde{y}(k) + \bar{y} & ,\ \text{for } k=1. \end{cases} \tag{6.9}$$

Consequently, the one-step ahead predicted values of the output discharge $[y(\cdot)]$ may be defined as

$$\hat{y}(k+1) = \sum_{i=1}^{k_o} \hat{U}(i)\, \hat{y}(k-i+1), \qquad k=1, 2, \ldots, N. \tag{6.10}$$

iii) As mentioned earlier, the constrained approach may lead to a considerable improvement in the accuracy of estimated output data. Thus, it is advisable to consider the numerical solution of the optimization problem (3.13) together with the two constraints of (3.14).

Specifically, the incomplete mathematical balance of the system under study strengthen the hypothesis of inequality constraint alone. Thus, the optimization problem reduces to

$$\text{Min } \theta_c = \frac{1}{2} \underline{U}^T \underline{H}^T \underline{V}^{-1} \underline{\underline{HU}} - \underline{U}^T \underline{H}^T \underline{V}^{-1} \underline{Y}$$

$$\text{subject to } \underline{U} \geq 0, \tag{6.11}$$

where the kernel length k_o may be evaluated using (6.3), (6.4) together with (6.11).

The impulse response vector $\underline{U}$ that minimizes the previous optimization problem is then invoked to transfer the delayed input data of the model M_5 into its output part according to (3.21).

iv) Unfortunately, the three impulse response vectors obtained before demonstrated an oscillatory pattern due to the irrepresentability of the observed input and/or output data [Blanke et al. (1970)]. Thus, it is relevant to point out that, these oscillatory vectors may be mathematically smoothed using the Hamming window algorithm discussed in Appendix A. Consequently, we can obtain another three models M_2', M_4' and M_6'.

Practically, all necessary estimates can be evaluated using the computer program listed in Appendix B. Let

$$\eta(k) = y(k) - \hat{y}(k)$$
$$\forall k = 1, 2, \ldots, N \qquad (6.12)$$

be the residuals of estimation at lag k.

The numerical values of the impulse response vectors for the previous models as well as the mean and variance of each residual sequence $[\eta(\cdot)]$ are summarized in Table (6.1), where

$$\bar{\eta} = \frac{1}{N} \sum_{i=1}^{N} \eta(i), \qquad (6.13)$$

and

$$\sigma_\eta = \frac{1}{N-1} \sum_{i=1}^{N} [\eta(i) - \bar{\eta}]^2. \qquad (6.14)$$

6.4 VALIDATION TESTS OF THE NOISY-TRANSFER FUNCTION MODEL

Kashyap and Rao (1976) have suggested that the appropriate class of models can be obtained by investigating its validation for the prespecified estimation conditions. Thus, we shall use the validation tests discussed in Chapter III to select an adequate model among the six noisy-transfer function models presented before.

6.4.1 Test of the Goodness of Fit

Usually, the goodness of fit between the two histograms of observed and estimated discharges may be checked by using the second Kolmogrov-Smirnov test given in Appendix C. Consequently, the statistical responses of the six models

Table (6.1) SUMMARY OF THE NOISY_TRANSFER FUNCTION MODELS.

MODEL	DELAY FACTOR	THE IMPULSE RESPONSE												$\bar{\eta}$	σ_η
		DURATION	U (1)	U (2)	U (3)	U (4)	U (5)	U (6)	U (7)	U (8)	U (9)	U (10)	U (11)		
M_1'	3	10	-1.0189	0.2175	-0.1230	0.1279	-0.1702	-0.024	-0.1282	0.1273	-0.0028	0.0463	-0.0201	0.00111	0.2[illegible]199
M_2'	3	5	-0.4522	-0.1455	0.0123	-0.0039	-0.0704	-0.0971	—	—	—	—	—	0.00072	0.23355
M_3'	0	5	-1.0000	2.0551	-0.2485	0.2260	-0.3162	0.2030	—	—	—	—	—	0.00018	0.13818
M_4'	0	5	0.4050	0.8218	0.3901	-0.0078	-0.0720	-0.0357	—	—	—	—	—	-0.00045	0.15504
M_5'	3	10	0.0000	0.0160	0.0000	0.0118	0.0000	0.0208	0.0000	0.0188	0.0000	0.0163	0.0071	-0.24178	0.21834
M_6'	3	10	0.0073	0.0086	0.0054	0.0064	0.0075	0.0112	0.0031	0.0102	0.0091	0.0104	0.0113	-0.21652	0.18276

$\bar{\eta}$: MEAN OF THE RESIDUALS.

σ_η : STANDARD DEVIATION OF RESIDUALS.

are illustrated in Table (6.2).

As a general view, the test statistics of the model M_4' are acceptable on both the 0.95 and 0.90 significant levels, while the other model M_3' may be accepted on the second level only.

6.4.2 Test of Zero-Mean Value of Residuals

Obviously, the estimators of the output data sequence may be unbiased for those models whose residual sequence has a zero-mean value. Thus, the results shown in Table (6.3) insure the validity of the unconstrained models M_1', M_2', M_3' and M_4' for the zero-mean value and consequently the unbiasing condition.

6.4.3 Validation Tests Based on the Comparison of Various Characteristics of Observed and Estimated Discharges

For an appropriate reduction for the field of choice, we may consider only the two successful models M_3' and M_4'. Specifically, the correlograms, power spectrums, histograms, and the normalized cumulative histograms of these two models compared with the corresponding characteristics of the observed output data are illustrated in Figs. (6.9) to (6.14).

These results indicate that :

i) The standard deviation $\sigma^M(k)$ governed by (3.30) is found to be 0.24 which represents a very convenient qualitative decision limit for both models.

ii) The correlogram, power spectrum and histogram of the generated data using M_4' are quite similar to those of the observed output data. Thus the qualitative validation test strengthen the hypothesis of choice M_4'.

Table(6.2) RESULTS OF THE SECOND KOLMOGROV_SMIRNOV TEST

MODEL	TEST STATISTIC	LAG 5	10	15	20	25	30	35	40	45	50
M_1'	Z	1.5814	1.7889	1.4606	0.7906	1.2728	1.4201	1.7928	1.9007	1.7919	1.5000
	$\epsilon_1 = 0.05$	R	R	R	A	A	R	R	R	R	R
	$\epsilon_2 = 0.10$	R	R	R	A	R	R	R	R	R	R
M_2'	Z	1.5811	2.0125	1.6432	0.9437	1.4142	1.9365	2.3905	2.6833	2.6352	2.4000
	$\epsilon_1 = 0.05$	R	R	R	A	R	R	R	R	R	R
	$\epsilon = 0.10$	R	R	R	A	R	R	R	R	R	R
M_3'	Z	1.2649	1.3416	1.2780	1.1067	0.9899	0.9036	0.8366	1.2298	1.1595	1.1000
	$\epsilon_1 = 0.05$	A	A	A	A	A	A	A	A	A	A
	$\epsilon_2 = 0.10$	R	R	R	A	A	A	A	R	A	A
M_4'	Z	0.3162	0.8944	0.9128	0.9486	0.8485	0.7746	0.9562	0.8944	0.6325	0.6000
	$\epsilon_1 = 0.05$	A	A	A	A	A	A	A	A	A	A
	$\epsilon_2 = 0.10$	A	A	A	A	A	A	A	A	A	A
M_5'	Z	1.5811	2.2361	2.7386	3.0042	2.8284	2.5819	2.5099	2.6833	2.8461	3.2000
	$\epsilon_1 = 0.05$	R	R	R	R	R	R	R	R	R	R
	$\epsilon_2 = 0.10$	R	R	R	R	R	R	R	R	R	R
M_6'	Z	1.5811	2.3361	2.7386	2.6979	2.4042	2.1947	2.7348	2.9236	2.4244	2.8000
	$\epsilon_1 = 0.05$	R	R	R	R	R	R	R	R	R	R
	$\epsilon_2 = 0.10$	R	R	R	R	R	R	R	R	R	R

A : ACCEPT THE NULL HYPOTHESIS Ho.

A : REJECT THE NULL HYPOTHESIS Ho.

Table (6.3) RESULTS OF THE ZERO MEAN TEST.

TEST STATISTIC	MODELS					
	M_1'	M_2'	M_3'	M_4'	M_5'	M_6'
D_1	0.04534	0.3300	0.02524	0.01069	11.43045	11.18838
$\epsilon_1 = 0.05$	A	A	A	A	R	R
$\epsilon = 0.10$	A	A	A	A	R	R

A : ACCEPT So.

R : REJECT So.

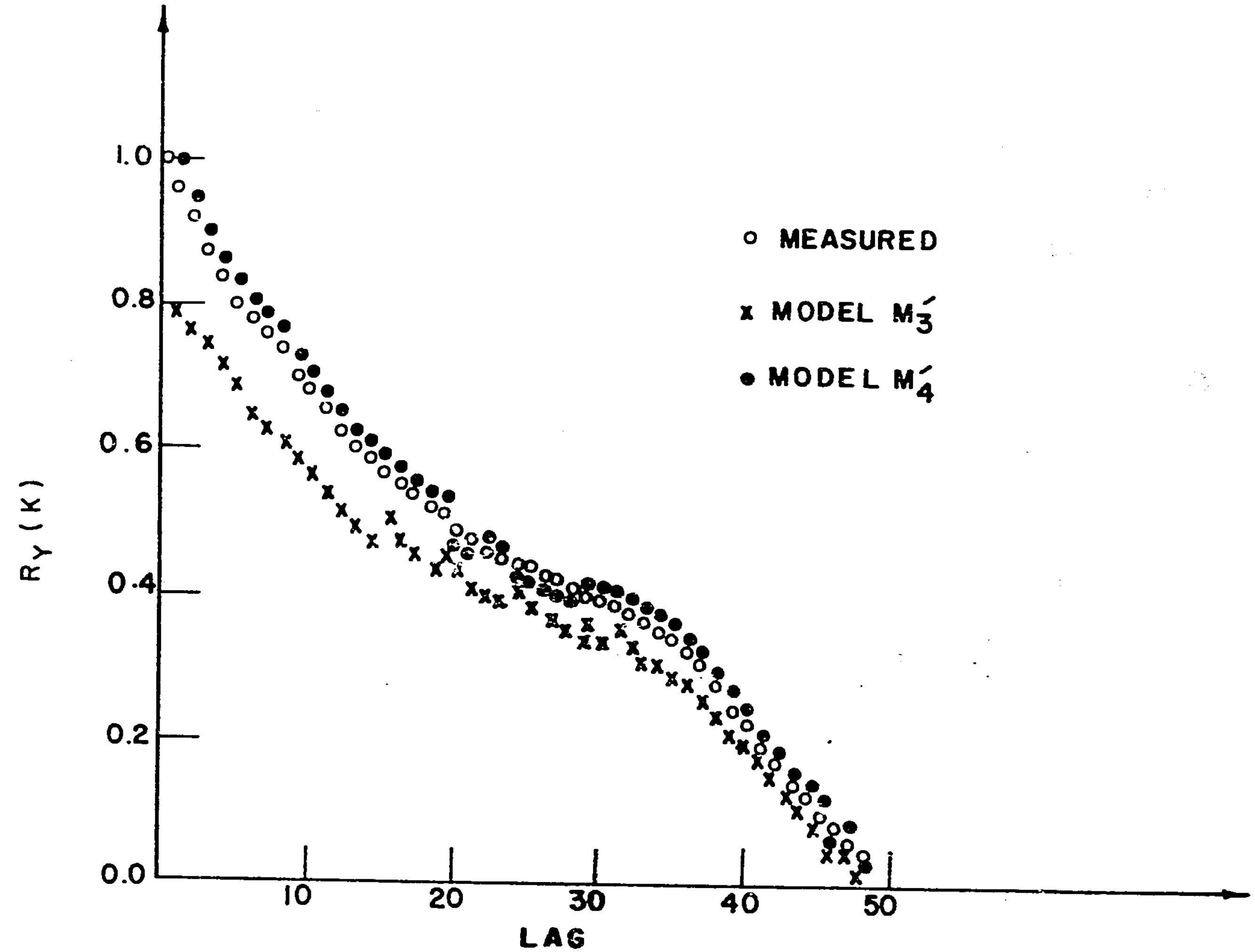


Fig.(6.9) CORRELOGRAMS OF THE MEASURED AND ESTIMATED DISCHARGE DATA.

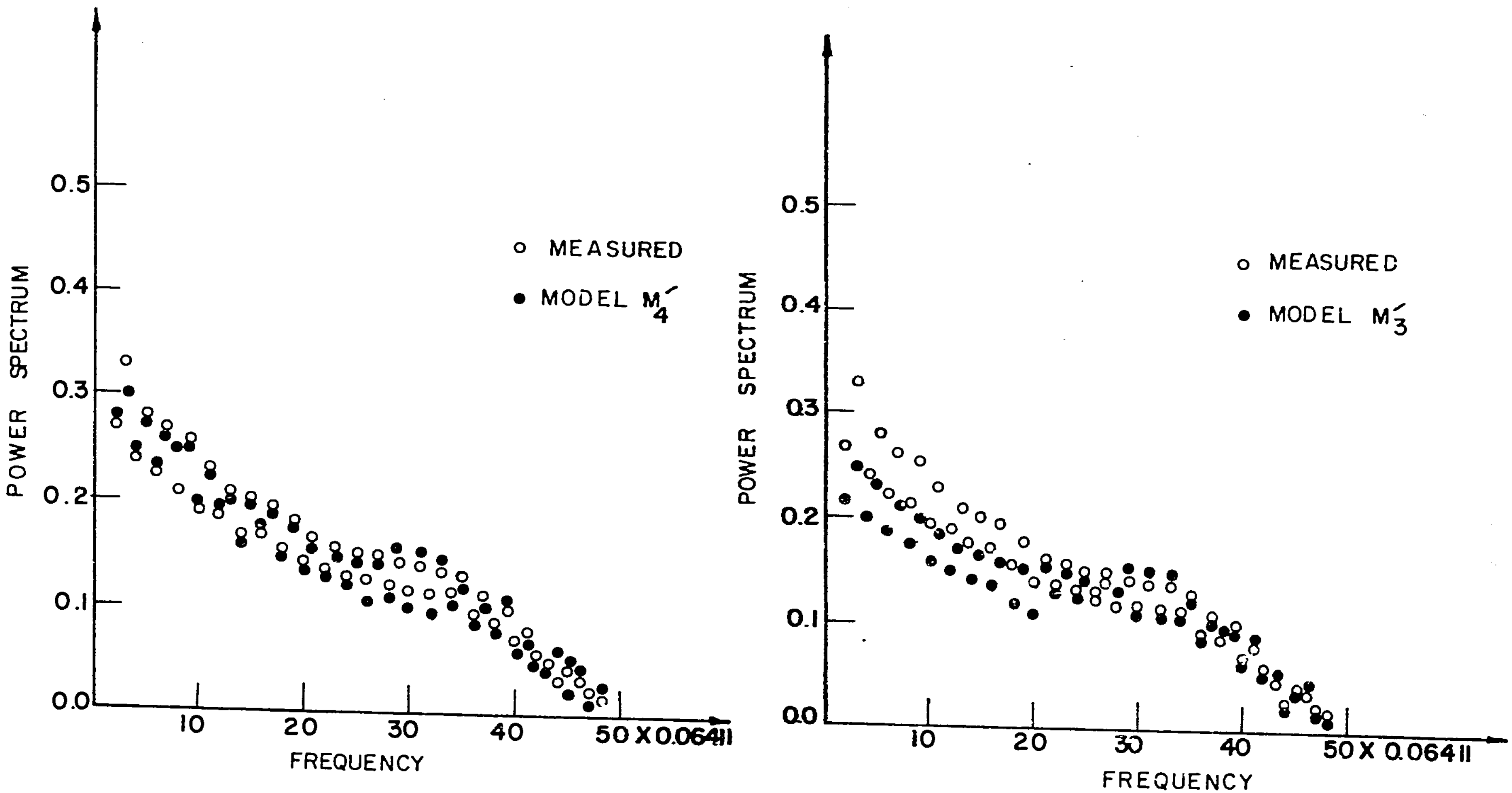


Fig. (6.10) POWER SPECTRUMS OF THE MEASURED AND ESTIMATED DISCHARGE DATA.

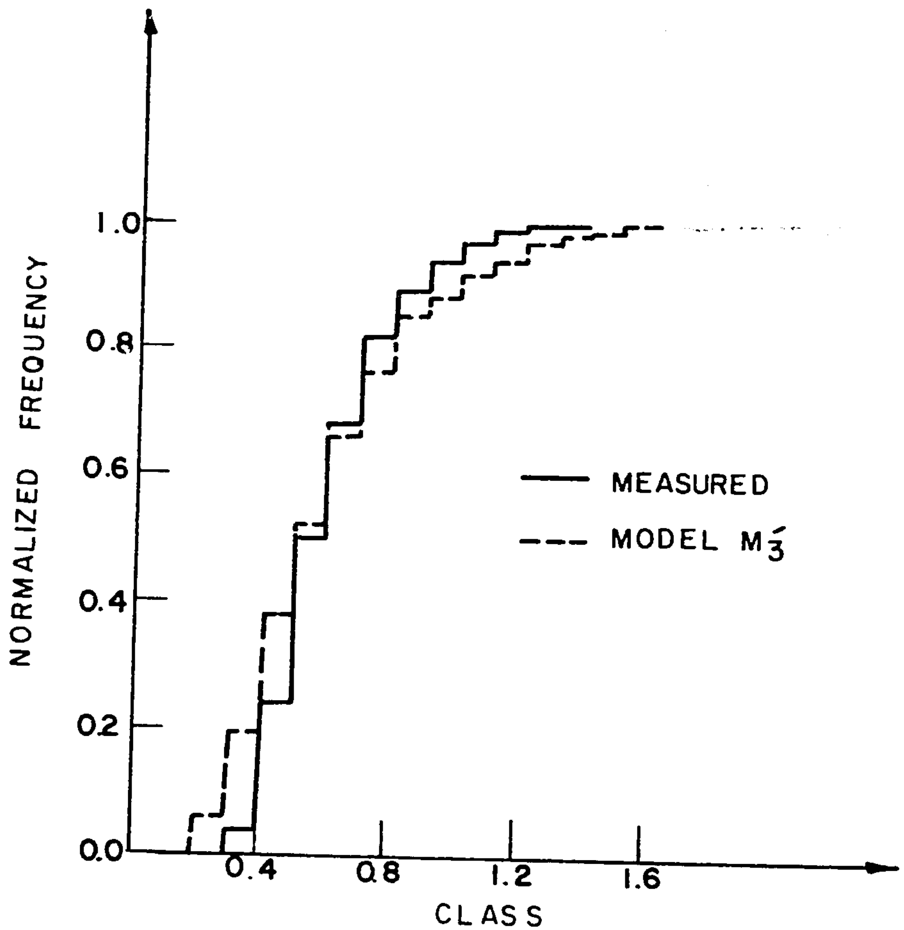


Fig. (6.11) NORMALIZED CUMULATIVE HISTOGRAMS OF MEASURED AND ESTIMATED DISCHARGE DATA.

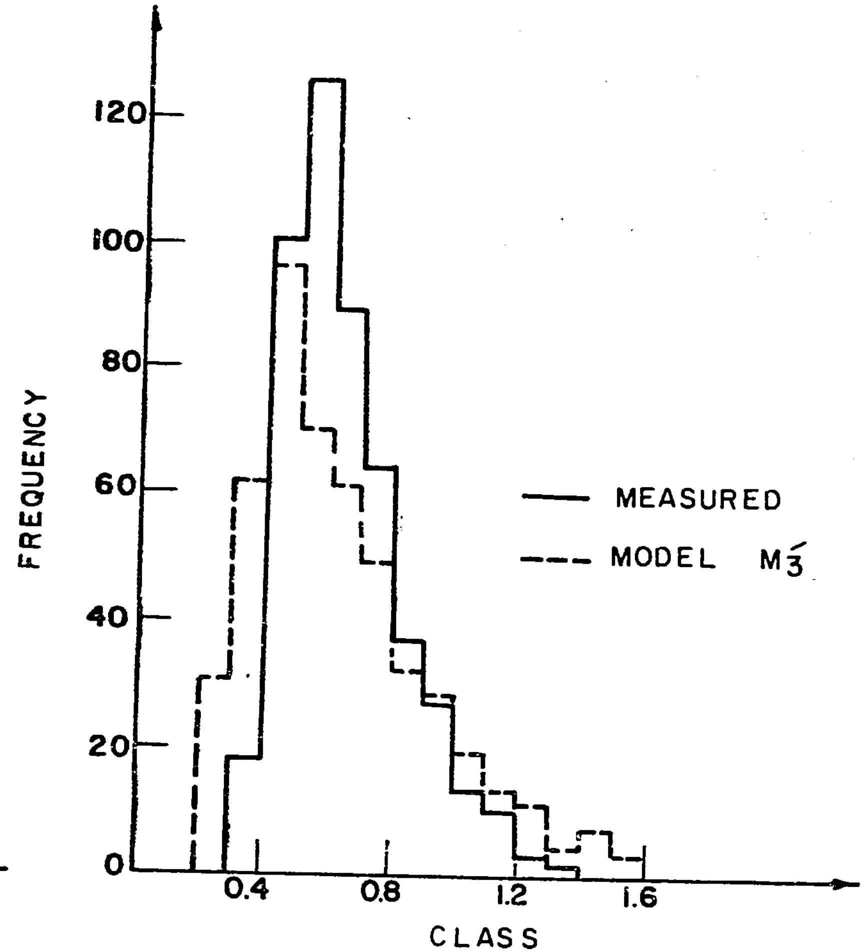


Fig (6.12) HISTOGRAMS OF MEASURED AND ESTIMATED DISCHARGE DATA.

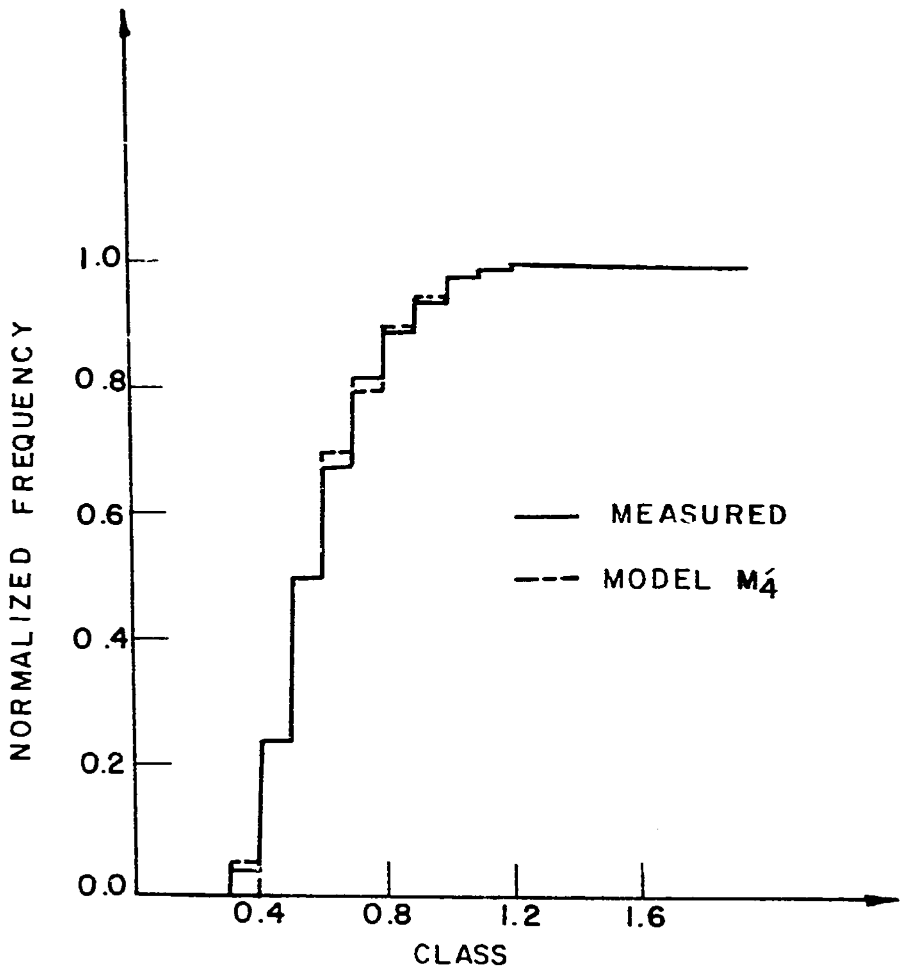


Fig.(6.13) NORMALIZED CUMULATIVE HISTOGRAMS OF THE MEASURED AND ESTIMATED DISCHARGE DATA.

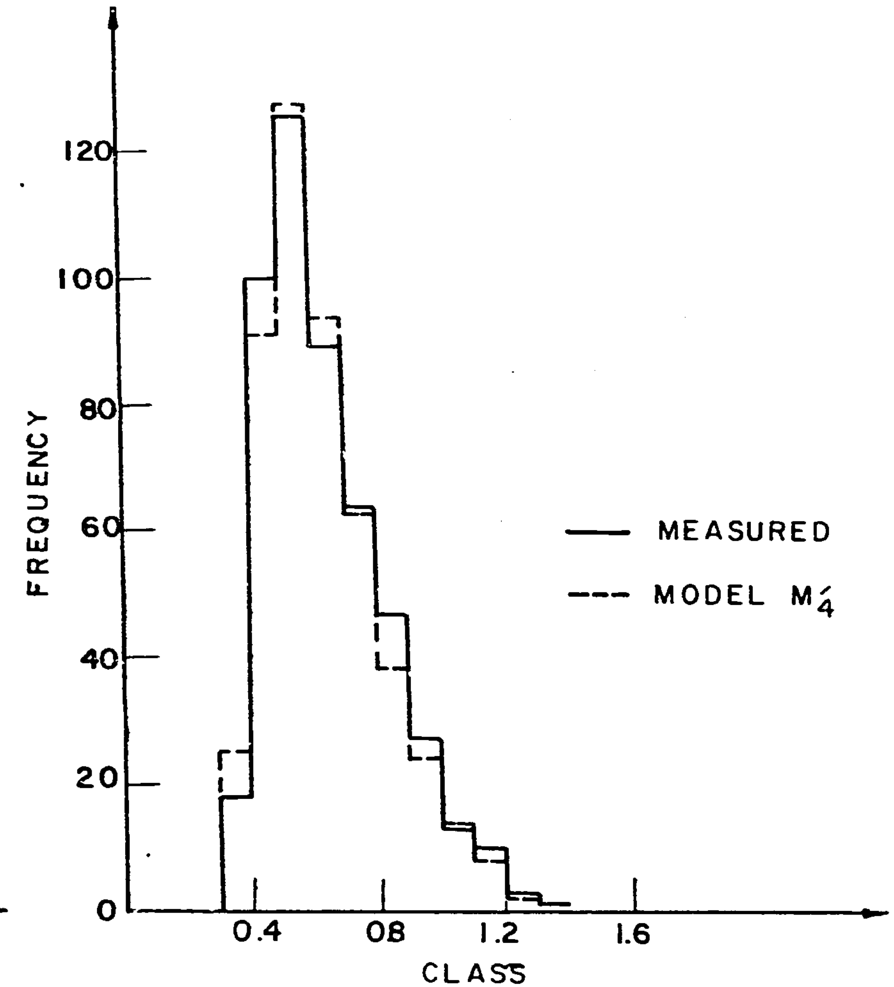


Fig.(6.14) HISTOGRAMS OF MEASURED AND ESTIMATED DISCHARGE DATA.

6.5 THE QUALITATIVE CHARACTERISTICS OF RESIDUALS

The ten-days mean-values of residuals obtained by using the two successful models M_3' and M_4' are delineated in Fig. (6.15).

The correlograms of daily residuals, evaluated via the two models M_3' and M_4', are illustrated in Figs. (6.16) and (6.17) respectively. It can be observed that the coefficients R(k) of the first residual sequence are more acceptable than those of the second sequence. since they lie within the specified standard deviation limit.

The smoothed raw estimates of the power spectrum for both daily residual sequences are shown in Figs. (6.18) and (6.19), which demonstrate a considerable variability but with a negligable magnitudes w.r.t. $S(w_o)$.

Finally, the histogram of residuals generated by the most successful model M_4' and its normalized cumulative values are shown in Figs. (6.20) and (6.21). These histograms coincide with the normal distribution N(-0.00045, 0.16), see Clark (1969).

6.6 APPLICATION OF THE LINEAR STOCHASTIC DIFFERENCE EQUATION MODEL

In this section, the linear stochastic difference equation model is applied to the physical system under study. The multiplicative and additive structures are utilized with the following assumptions:

i) The proposed model has only autoregressive terms of a variable order n.

ii) In addition to these n-autoregressive terms, another mth order term representing the residuals may be fedback to the output part of the model inorder to achieve a corrective pattern.

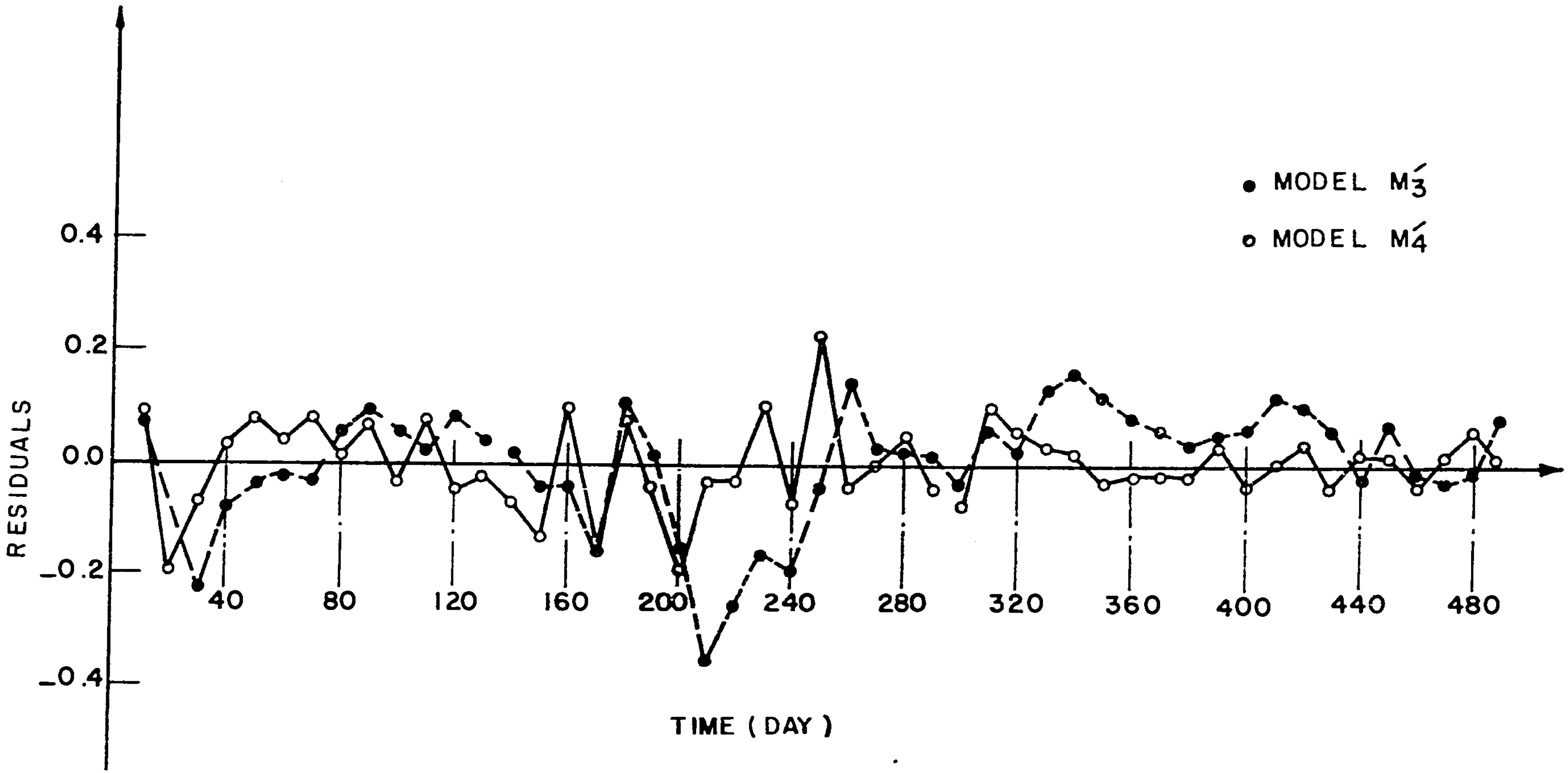


Fig.(6.15) VARIATION OF THE TEN-DAYS MEAN OF RESIDUALS FOR BOTH MODELS M_3' AND M_4' WITH TIME.

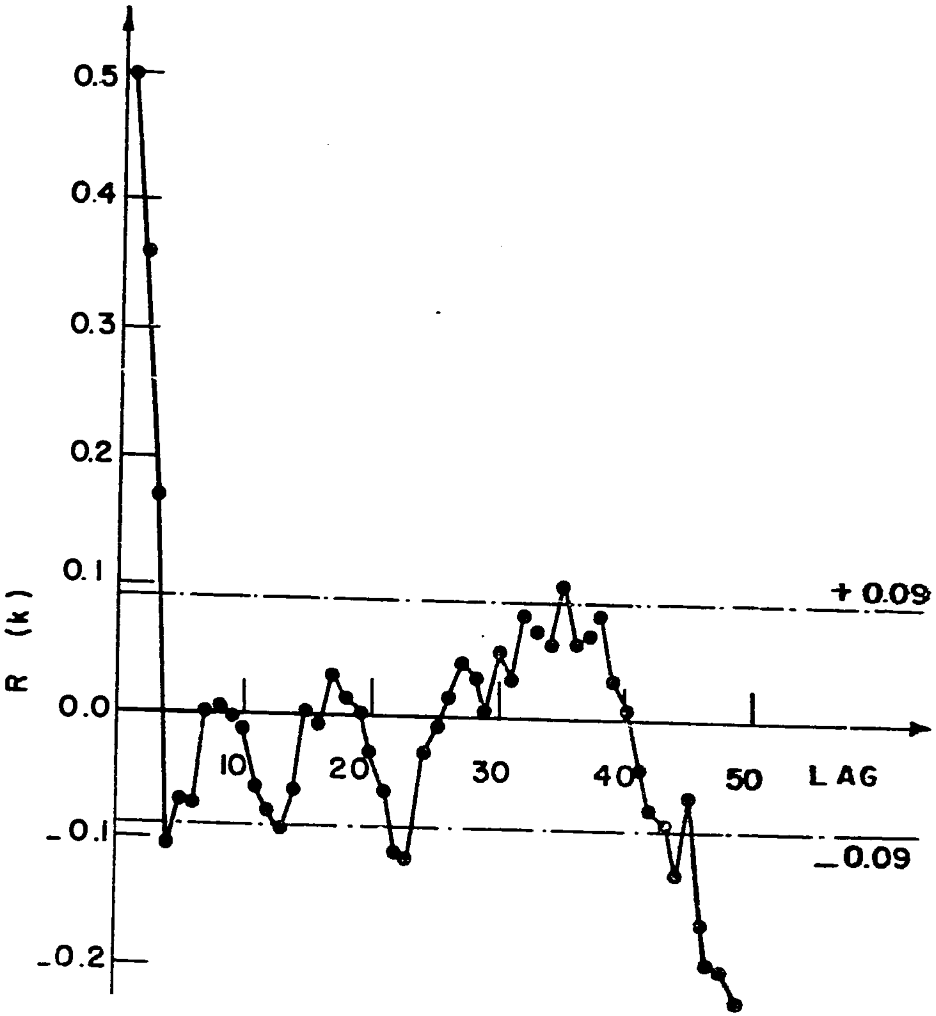
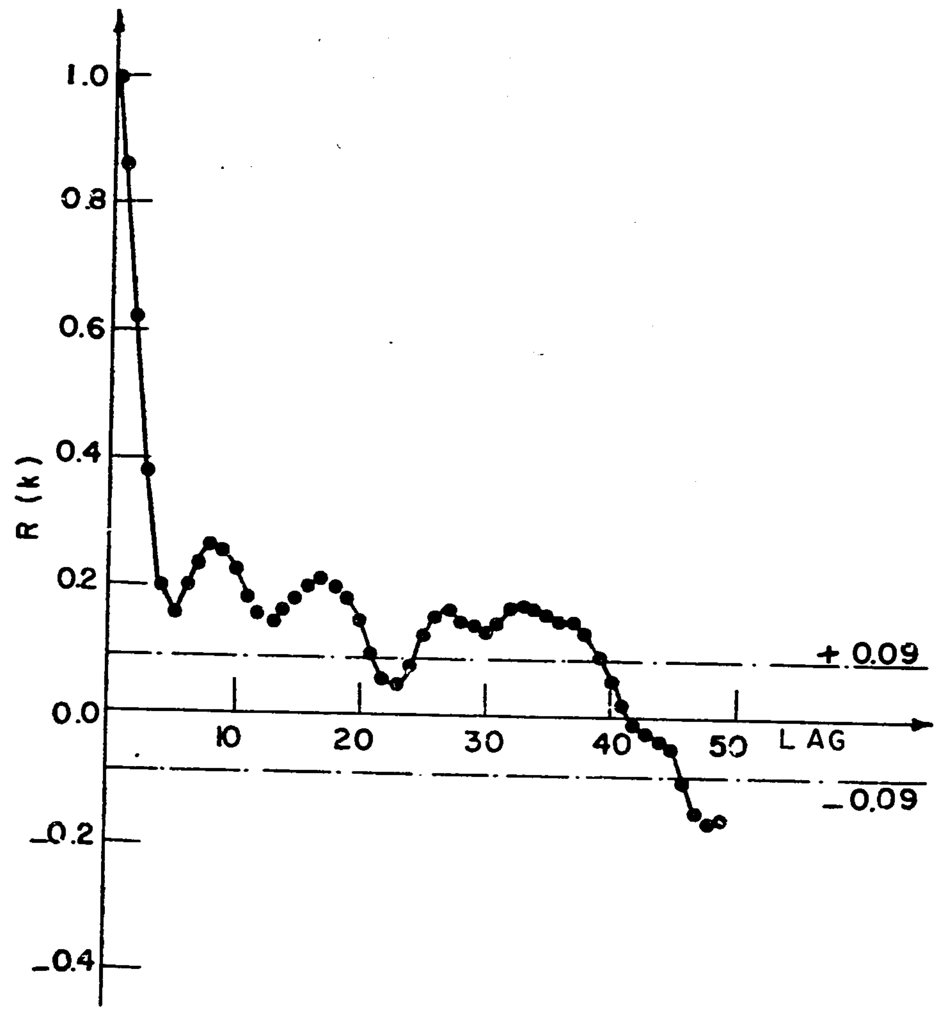
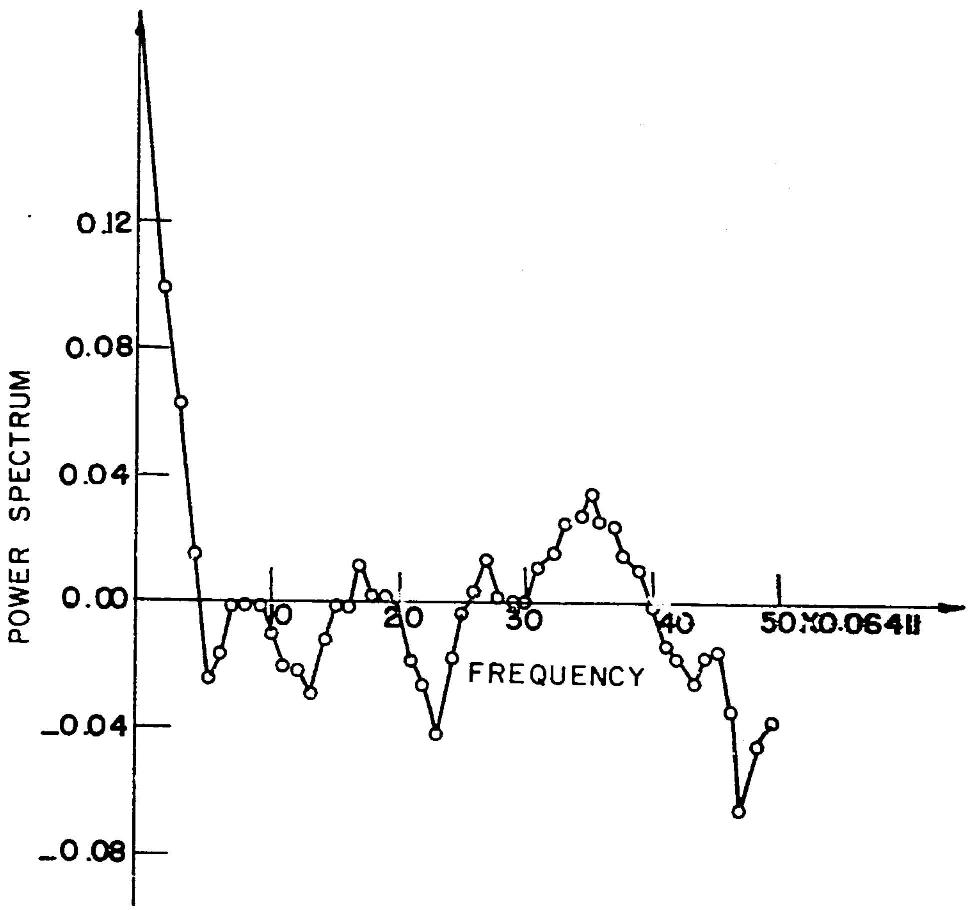


Fig.(6.16) CORRELOGRAM OF RESIDUALS WITH TWO STANDARD DEVIATION LIMITS FOR M_4'.



Fig(6.17) CORRELOGRAM OF RESIDUALS WITH TWO STANDARD DEVIATION LIMITS FOR M_3'.



Fig(6.18) POWER SPECTRUM OF RESIDUALS (MODEL M_4').

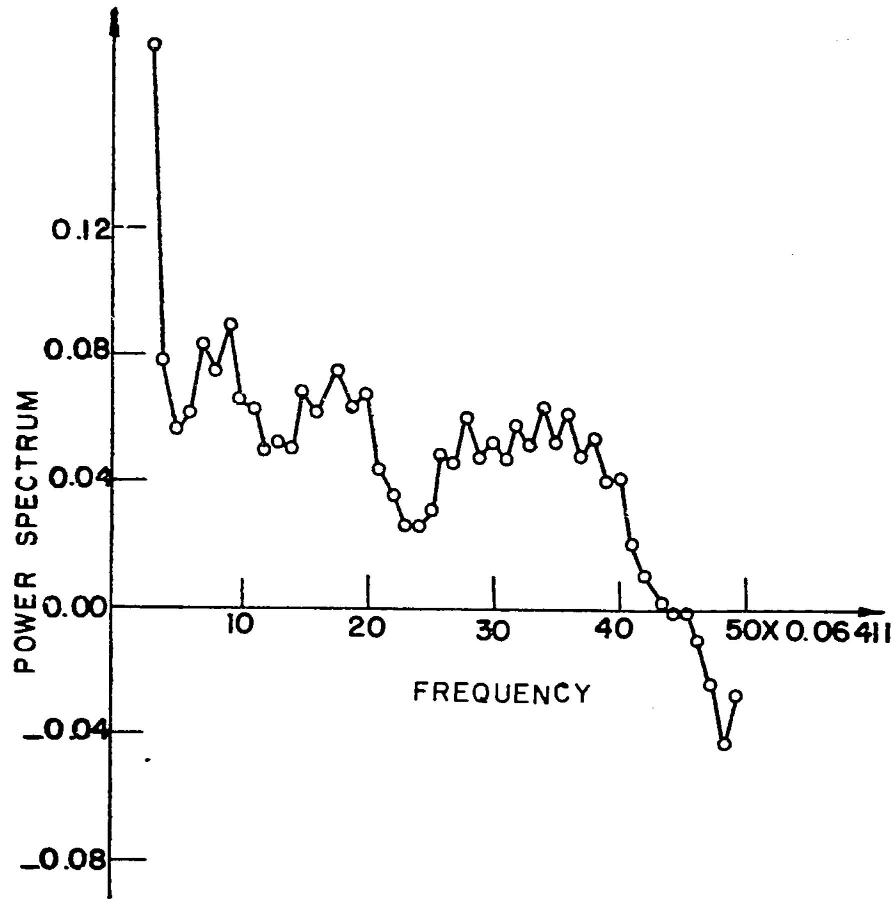


Fig.(6.19) POWER SPECTRUM OF RESIDUALS (MODEL M_3').

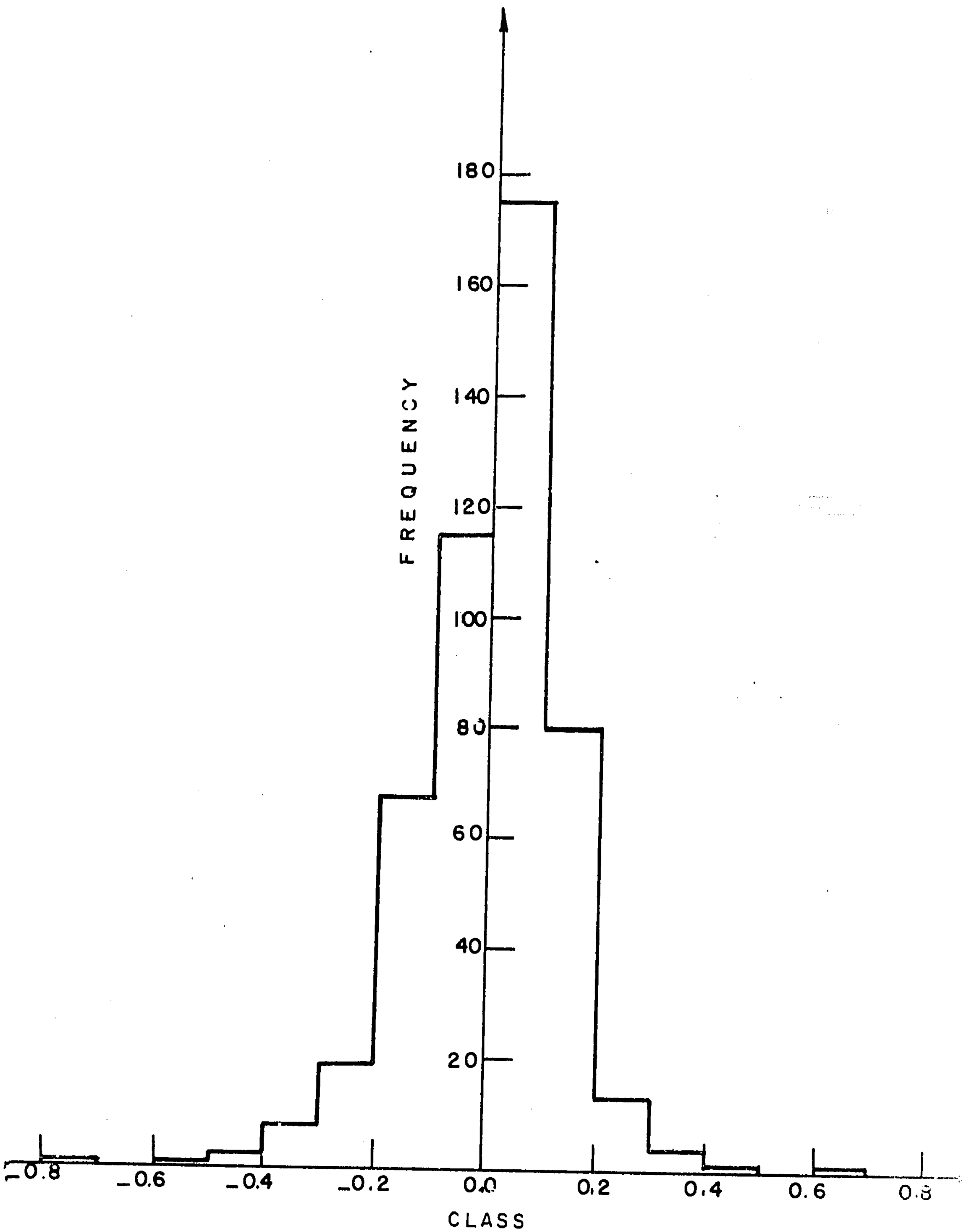


Fig. (6.20) HISTOGRAM OF RESIDUALS (MODEL M4́).

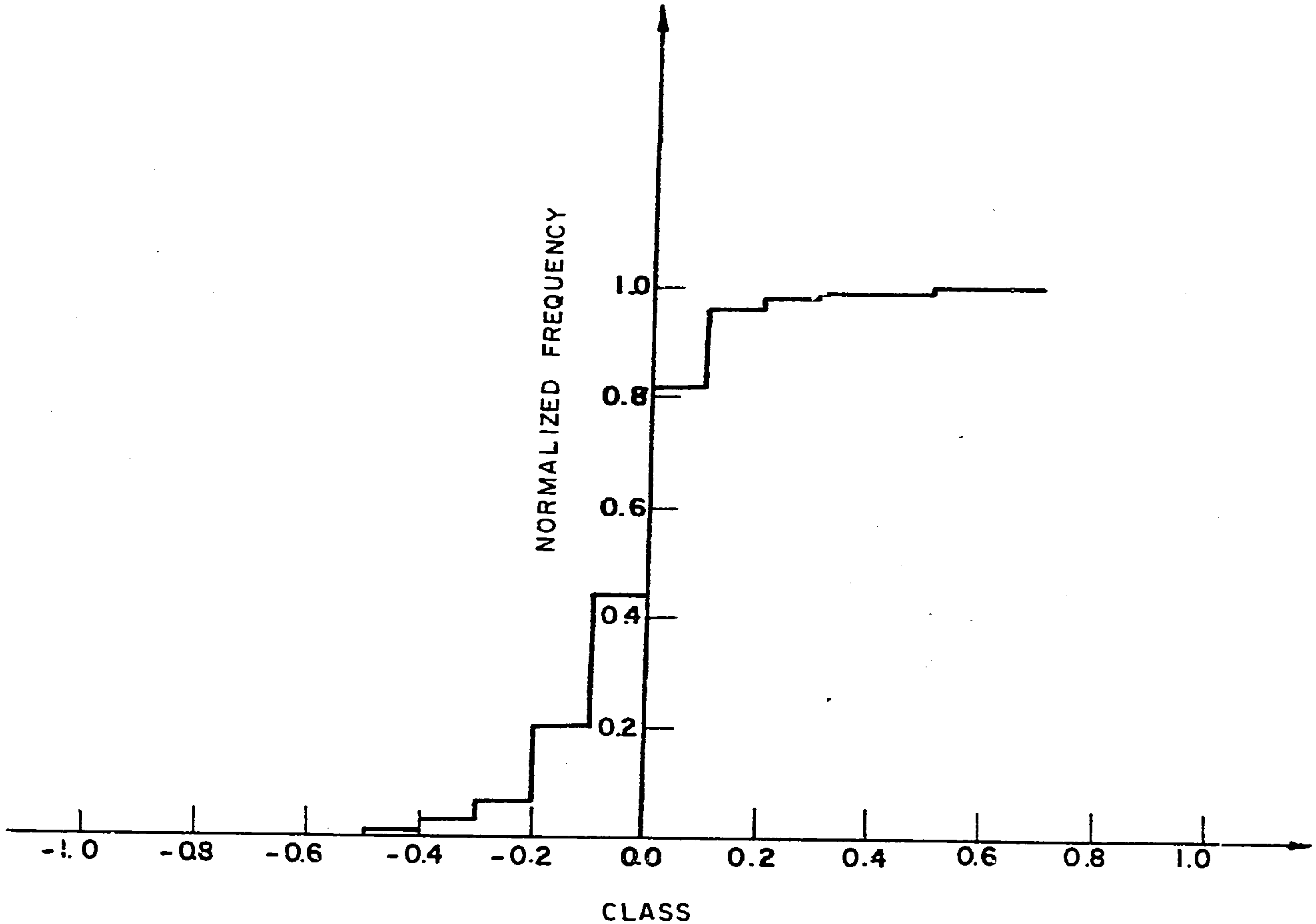


Fig.(6.21) NORMALIZED CUMULATIVE HISTOGRAM OF RESIDUALS
(MODEL M_4').

iii) Another sinusoidal term of frequency $2\pi j/244$, $j=1, 2, \ldots, N$, is added to the nth order autoregressive model to trace the daily oscillations of the data.

The number of autoregressive, residuals and sinusoidal terms for both the additive and multiplicative models are illustrated in Table (6.4).

6.7 ESTIMATION OF THE PARAMETER VECTOR

Using the recursive algorithm (4.11) together with first procedure of initialization, the parameter vector $\underline{a}(i)$, $i=1, 2, \ldots, N$, is identified. The final values of the estimated parameter vector $\underline{a}$ as well as the mean, absolute mean and mean-square values of residuals for both the additive and multiplicative sturctures are shown in Tables (6.5) and (6.6), where

$$E_o = \frac{1}{N} \sum_{i=1}^{N} \bar{w}(i),$$

$$E_1 = \frac{1}{N} \sum_{i=1}^{N} \left|\bar{w}(i)\right|$$

and

$$E_2 = \frac{1}{N} \sum_{i=1}^{N} [\bar{w}(i)]^2 \tag{6.15}$$

indicate respectively the mean, absolute mean and mean-square values of the residual sequence $[w(\cdot)]$.

6.8 CLASS SELECTION OF THE LINEAR STOCHASTIC DIFFERENCE EQUATION MODEL

Among the different classes of the linear stochastic difference equation model illustrated in Table (6.4), the most acceptable model can be obtained

Table (6.4) LIST OF PARAMETERS FOR THE ADDITIVE AND MULTIPLICATIVE MODELS.

PARAMETER	MODELS							
	M_1 & M_9	M_2 & M_{10}	M_3 & M_{11}	M_4 & M_{12}	M_5 & M_{13}	M_6 & M_{14}	M_7 & M_{15}	M_8 & M_{16}
n	2	3	4	5	2	3	2	3
m	—	—	—	—	2	2	—	—
n_3	—	—	—	—	—	—	1	1

n : NUMBER OF AUTOREGRESSIVE TERMS.

n_3 NUMBER OF SINUSOIDAL TERMS.

m : NUMBER OF ERROR TERMS.

M_i ; i = 1 — 8 : ADDITIVE MODELS.

M_i ; i = 9 — 16 : MULTIPLICATIVE MODELS.

Table(6.5) SUMMARY OF THE RESULTS OF THE ADDITVE MODELS

a_0	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_9	a_{10}	E_0	E_1	E_2
1	$\tilde{Y}(K-1)$	$\tilde{Y}(K-2)$	$\tilde{Y}(K-3)$	$\tilde{Y}(K-4)$	$\tilde{Y}(K-5)$	$\sin W_1 K$	$\cos W_1 K$	$\bar{W}(K-1)$	$\bar{W}(K-2)$			
0.000024	0.989151	0.034769	—	—	—	—	—	—	—	0.000013	0.006702	0.000072
0.000048	0.990626	0.074076	0.039635	—	—	—	—	—	—	0.006876	0.036015	0.001222
0.000038	0.992343	0.077380	0.083911	0.04464	—	—	—	—	—	0.000049	0.007058	0.000081
0.000170	0.996483	0.08546	0.091488	0.138120	0.094106	—	—	—	—	0.000086	0.007720	0.000105
0.000012	0.978184	-0.041485	—	—	—	-0.043227	0.022563	—	—	0.000140	0.015308	0.000186
-0.000047	0.979631	-0.073253	0.032415	—	—	-0.042782	0.023852	—	—	0.000212	0.015996	0.000212
0.002097	0.579018	0.360508	—	—	—	—	—	0.449080	-0.002664	0.000169	0.022560	0.000276
0.001883	0.549516	0.235870	0.151411	—	—	—	—	0.477697	0.148088	0.000285	0.023390	0.000309

E_o : MEAN VALUE OF THE RESIDUALS.

E_1 : ABSOLUTE MEAN VALUE OF THE RESIDUALS.

E_2 : MEAN SQUARE VALUE OF THE RESIDUALS.

W_1 : THE MAIN FREQUNCY OF THE OBSERVED OUTPT DATA.

Table (6.6) SUMMARY OF THE RESULTS OF THE MULTIPLICATIVE MODELS.

a_0	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9			
1	$\tilde{Y}_2(K-1)$	$\tilde{Y}_2(K-2)$	$\tilde{Y}_2(K-3)$	$\tilde{Y}_2(K-4)$	$\tilde{Y}_2(K-5)$	SIN W_1K	COS W_1K	$\bar{w}(K-1)$	$\bar{w}(K-2)$	E_0	E_1	E_2
-0.025724	0.579371	0.301895	—	—	—	—	—	—	—	0.006826	0.029192	0.001146
-0.022711	0.538083	0.223713	0.133374	—	—	—	—	—	—	0.013548	0.059435	0.002382
-0.021777	0.532248	0.212368	0.049375	0.049375	—	—	—	—	—	0.006878	0.046378	0.001467
-0.021103	0.530324	0.209142	0.098336	0.028958	0.038258	—	—	—	—	0.006711	0.024044	0.001212
-0.03433	0.558548	0.282152	—	—	—	-0.011022	0.008888	—	—	0.013527	0.070874	0.002625
-0.030421	0.524503	0.214345	0.120754	—	—	-0.010336	0.007069	—	—	0.013244	0.048891	0.002432
-0.038117	0.497887	0.320971	—	—	—	—	—	0.227203	0.046531	0.019652	0.090558	0.003420
-0.035376	0.433579	0.217850	0.181635	—	—	—	—	0.254027	0.071917	0.019138	0.068282	0.003199

E_0 : MEAN VALUE OF THE RESIDUALS.

E_1 : ABSOLUTE MEAN VALUE OF THE RESIDUALS.

E_2 : MEAN SQUARE VALUE OF THE RESIDUALS.

W_1 : THE MAIN FREQUNCY OF THE OBSERVED OUTPUT DATD.

by using the class selection procedure depicted previously in Chapter IV.

6.8.1 The Likelihood Approach

According to (4.16) the likelihood function L_i, i=0, 1, ..., 15, is evaluated for each proposed model. It is found that, the additive model M_1, furnishes the largest value of the likelihood function L_i. Consequently, the given data may be assigned to that successful model.

6.8.2 The Prediction Approach

Using the estimated parameter vector $\underline{a}$ together with the noise sequence $[w(\cdot)]$ generated via a Gaussian random variable generator whose mean and variance are quite similar to those of the residual sequence $[\bar{w}(\cdot)]$, the one-step ahead prediction of the output discharge can be obtained. The quality of predicted values may be checked by using (4.19). Finally, the values of L_i and J_i, i=0,1 , ..., 15, for both the additive and multiplicative models together with their corresponding rank are illustrated in Table (6.7).

6.9 VALIDATION TESTS OF THE LINEAR STOCHASTIC DIFFERENCE EQUATION MODEL

It is convenient to test the validity of the proposed models illustrated in Tables (6.5) and (6.6) for the utility condition (4.3), together with the normality of the generated residual sequence $[\bar{w}(\cdot)]$.

6.9.1 Test of Serial Independence

Using the residual sequence for both the additive and multiplicative models in addition to the computer program listed in Appendix D, the test statistics $\beta(\bar{w})$ are computed according to (4.35). The decision of acceptance

Table(6.7) RESULTS OF THE LIKELIHOOD APPROACH AND PREDICTION APPROACH FOR THE CLASS SELECTION OF THE ADDITIVE AND MULTIPLICATIVE MODELS.

	ADDITIVE MODELS								MULTIPLICATIVE MODLS							
	M_1	M_2	M_3	M_4	M_5	M_6	M_7	M_8	M_9	M_{10}	M_{11}	M_{12}	M_{13}	M_{14}	M_{15}	M_{16}
L_i	232648	1633.57	2294.74	2230.42	2082.90	2059.98	1995.61	1968.05	1651.24	1472.22	1589.99	163609	1448.52	1466.66	1384.96	1400.27
$J_i \times 10^5$	7.170	121.448	8.024	10.392	18.448	21.026	27.430	30.647	114.130	236.74	145.50	119.96	260.89	241.21	339.89	317.93
RANK	1	10	2	3	4	5	6	7	8	12	11	9	14	13	16	15

$$L_i = -(N/2)\ln \hat{P}_i - n_i$$

$$J_i = \sum_{k=2}^{N} \left[y(k) - \hat{y}(k/k-1) \right] / N-1$$

or rejection the class C_o may be made by comparing the values of $\beta(\bar{w})$, at different lags, with those of the F-distribution function having n_2 and $N-n_2$ degrees of freedom, where n_2 is the corresponding lag. The response for both the additive and multiplicative models to that test is illustrated in Tables (6.8) and (6.9) respectively.

Briefly, acceptance of C_o insures the serial independency of the specified residual sequence $[\bar{w}(\cdot)]$.

6.9.2 Test of Normality

As discussed before, the histogram of estimated residual sequence $[\bar{w}(\cdot)]$ can be compared with the standard normal distribution curve, having the same mean and variance, by employing the first Kolomgrov-Smirnov test. The test statistics as well as the decision of acceptance or rejection the null hypothesis H_o for both the additive and multiplicative models are elucidated in Tables (6.10) and (6.11). On the other hand, the probability of acceptance of the null hypothesis H_o for the most successful model M_1 is illustrated in Table (6.12).

Finally, the variations in coefficients of the two successful models M_1 and M_3 with sample size are demonstrated in Figs. (6.22) and (6.23) respectively. It can be observed that, these coefficients exhibit significant changes with the variation of sample size.

6.10 COMPARISON OF THE TWO BEST FITTED MODELS M_4' AND M_1

Using the two output data sequences generated by the best fitted noisy-transfer function model M_4' and the successful linear stochastic difference equation model M_1, the major features of these two models can be summarized as follows

Table (6.8) RESULTS OF THE SERIAL CORRELATION TEST FOR THE ADDITIVE MODELS.

MODEL	TEST STATISTIC	LAG 5	10	15	20	25	30	35	40	45	50
M_1	$\eta(\bar{w})$	0.5630	0.5272	0.5054	0.3961	0.3186	0.2767	0.2227	0.2056	0.1772	0.1612
	$\epsilon_1 = 0.05$	A	A	A	A	A	A	A	A	A	A
	$\epsilon_2 = 0.10$	A	A	A	A	A	A	A	A	A	A
M_2	$\eta(\bar{w})$	1.7622	1.6514	1.7221	1.6614	1.4324	1.3733	1.4141	1.4001	1.3910	1.4120
	$\epsilon_1 = 0.05$	R	R	R	R	R	R	R	R	R	R
	$\epsilon_2 = 0.10$	R	R	R	R	R	R	R	R	R	R
M_3	$\eta(\bar{w})$	0.4468	0.4253	0.4013	0.3961	0.3512	0.3213	0.2015	0.1732	0.1701	0.1651
	$\epsilon_1 = 0.05$	A	A	A	A	A	A	A	A	A	A
	$\epsilon_2 = 0.10$	A	A	A	A	A	A	A	A	A	A
M_4	$\eta(\bar{w})$	10.3210	5.2752	4.5010	4.2010	3.9221	3.5268	3.1519	2.528	2.3187	2.1525
	$\epsilon_1 = 0.05$	R	R	R	R	R	R	R	R	R	R
	$\epsilon_2 = 0.10$	R	R	R	R	R	R	R	R	R	R
M_5	$\eta(\bar{w})$	15.9621	12.7151	10.5140	9.8155	8.6631	7.4340	6.1416	5.5501	5.9030	6.5220
	$\epsilon_1 = 0.05$	R	R	R	R	R	R	R	R	R	R
	$\epsilon_2 = 0.10$	R	R	R	R	R	R	R	R	R	R
M_6	$\eta(\bar{w})$	5.6713	5.7314	6.9132	6.5143	6.3152	6.0152	5.4220	5.1107	4.3143	4.9212
	$\epsilon_1 = 0.05$	R	R	R	R	R	R	R	R	R	R
	$\epsilon_2 = 0.10$	R	R	R	R	R	R	R	R	R	R
M_7	$\eta(\bar{w})$	6514 3	4.9182	3.7170	3.2517	2.8157	2.0103	1.8132	1.5152	1.4130	1.3730
	$\epsilon_1 = 0.05$	R	R	R	R	R	R	R	R	R	R
	$\epsilon_2 = 0.10$	R	R	R	R	R	R	R	R	R	R
M_8	$\eta(\bar{w})$	8.8173	9.1320	10.3107	9.7541	9.2512	8.7373	7.7125	7.5412	6.5962	6.2130
	$\epsilon_1 = 0.05$	R	R	R	R	R	R	R	R	R	R
	$\epsilon_2 = 0.10$	R	R	R	R	R	R	R	R	R	R

A : ACCEPT Co .

A : REJECT Co .

Table (6.9) RESULTS OF THE SERIAL CORRELATION TEST OF THE MULTIPLICATIVE MODELS.

MODEL	TEST STATISTIC	LAG 5	10	15	20	25	30	35	40	45	50
M_9	$\eta(\bar{w})$ $\epsilon_1 = 0.05$ $\epsilon_2 = 0.10$	15.1086 R R	7.5086 R R	4.9641 R R	3.8013 R R	3.0891 R R	2.6171 R R	2.1356 R R	1.9523 R R	1.6291 R R	1.4892 R
M_{10}	$\eta(\bar{w})$ $\epsilon_1 = 0.05$ $\epsilon_2 = 0.10$	3.4409 R R	3.2643 R R	3.0776 R R	2.8788 R R	2.6653 R R	2.4331 R R	2.2251 R R	1.9626 R R	1.5732 R R	1.1126 R R
M_{11}	$\eta(\bar{w})$ $\epsilon_1 = 0.05$ $\epsilon_2 = 0.10$	4.4682 R R	4.2534 R R	3.4952 R R	3.1258 R R	3.1415 R R	2.8871 R R	2.5950 R R	2.1696 R R	2.0606 R R	1.3606 R R
M_{12}	$\eta(\bar{w})$ $\epsilon_1 = 0.05$ $\epsilon_2 = 0.10$	3.4839 R R	3.3051 R R	3.1161 R R	2.9145 R R	2.6983 R R	2.4634 R R	2.2111 R R	1.9351 R R	1.5863 R R	1.1246 R R
M_{13}	$\eta(\bar{w})$ $\epsilon_1 = 0.05$ $\epsilon_2 = 0.10$	3.4782 R R	3.2997 R R	3.1110 R R	2.9100 R R	2.6942 R R	2.4594 R R	2.1898 R R	1.9309 R R	1.5783 R R	1.3214 R R
M_{14}	$\eta(\bar{w})$ $\epsilon_1 = 0.05$ $\epsilon_2 = 0.10$	3.2077 R R	3.0282 R R	2.8374 R R	2.[illegible]331 R R	2.4317 R R	2.2197 R R	2.1117 R R	1.8940 R R	1.5604 R R	1.1116 R R
M_{15}	$\eta(\bar{w})$ $\epsilon_1 = 0.05$ $\epsilon_2 = 0.10$	6.5447 R R	5.5842 R R	5.4047 R R	4.8661 R R	3.5360 R R	3.2844 R R	3.0520 R R	2.7917 R R	2.2201 R R	1.8410 R R
M_{16}	$\eta(\bar{w})$ $\epsilon_1 = 0.05$ $\epsilon_2 = 0.10$	6.7230 R R	5.6378 R R	5.4251 R R	4.8726 R R	3.5382 R R	3.2861 R R	3.0536 R R	2.7970 R R	2.2207 R R	1.6905 R R

A : ACCEPT Co.

R : REJECT Co.

Table (6.10) RESULTS OF THE FIRST KOLMOGROV_SMIRNOV TEST OF THE RESIDUAL NORMALITY FOR THE ADDITIVE MODELS.

MODEL	TEST STATISTIC	LAG									
		5	10	15	20	25	30	35	40	45	50
M_1	Z	0.3162	0.4472	0.3651	0.3162	0.2828	0.2582	0.3586	0.3354	0.3163	0.3000
	$\epsilon_1 = 0.05$	A	A	A	A	A	A	A	A	A	A
	$\epsilon_2 = 0.10$	A	A	A	A	A	A	A	A	A	A
M_2	Z	1.6324	1.9487	1.6708	1.5477	1.4743	1.4243	1.3873	1.5976	1.5590	1.5000
	$\epsilon_1 = 0.05$	R	R	R	R	R	R	R	R	R	R
	$\epsilon_2 = 0.10$	R	R	R	R	R	R	R	R	R	R
M_3	Z	0.3162	0.4472	0.3651	0.3162	0.2828	0.3873	0.4781	0.4472	0.4216	0.4000
	$\epsilon_1 = 0.05$	A	A	A	A	A	A	A	A	A	A
	$\epsilon_2 = 0.10$	A	A	A	A	A	A	A	A	A	A
M_4	Z	1.1160	0.5783	0.9309	0.9198	0.4595	1.6942	1.9100	2.2202	2.2997	2.4000
	$\epsilon_1 = 0.05$	A	A	A	A	A	R	R	R	R	R
	$\epsilon_2 = 0.10$	A	A	A	A	A	R	R	R	R	R
M_5	Z	0.6324	0.4472	0.3651	0.3612	1.2843	1.8190	1.8551	1.5410	1.6228	1.5000
	$\epsilon_1 = 0.05$	A	A	A	A	R	R	R	R	R	R
	$\epsilon_2 = 0.10$	A	A	A	A	R	R	R	R	R	R
M_6	Z	1.1160	1.5783	1.9309	2.1998	2.4594	2.6942	2.9100	3.1110	3.2997	3.4782
	$\epsilon_1 = 0.05$	A	R	R	R	R	R	R	R	R	R
	$\epsilon_2 = 0.10$	A	R	R	R	R	R	R	R	R	R
M_7	Z	1.1125	1.5733	1.9269	2.2251	2.4331	2.6653	2.8788	3.0776	3.2643	3.4409
	$\epsilon_1 = 0.05$	A	R	R	R	R	R	R	R	R	R
	$\epsilon_2 = 0.10$	A	R	R	R	R	R	R	R	R	R
M_8	Z	1.6325	1.4472	1.3651	1.3162	1.2728	1.5351	1.4270	1.4001	1.3977	1.3000
	$\epsilon_1 = 0.05$	R	R	R	R	R	R	R	R	R	R
	$\epsilon_2 = 0.10$	R	R	R	R	R	R	R	R	R	R

Z : STATISTIC OF THE FIRST KOLMOGROV_SMIRNOV TEST.

A : ACCEPT THE HYPOTHESIS OF NORMALITY.

R : REJECT THE HYPOTHESIS OF NORMALITY.

Table (6.11) RESULTS OF THE FIRST KOLMOGROV- SMIRNOV TEST OF THE RESIDUAL NORMALITY FOR THE MULTIPLICATIVE MODELS.

MODEL	TEST STATISTIC	LAG									
		5	10	15	20	25	30	35	40	45	50
M_9	Z $\epsilon_1 = 0.05$ $\epsilon_2 = 0.10$	1.6324 R R	1.6709 R R	1.5477 R R	1.4246 R R	0.5163 A A	0.7171 A A	0.6708 A A	0.6315 A A	0.6000 A A	0.9912 A A
M_{10}	Z $\epsilon_1 = 0.05$ $\epsilon_2 = 0.10$	0.9486 A A	0.6718 A A	0.5471 A A	1.4745 R R	1.3205 A R	1.4242 R R	1.3872 R R	1.5976 R R	1.5590 R R	0.5000 A A
M_{11}	Z $\epsilon_1 = 0.05$ $\epsilon_2 = 0.10$	0.9571 A A	0.7602 A A	0.5421 A A	0.7721 A A	0.8312 A A	0.8872 A A	1.1271 A A	1.4213 R R	1.5271 R R	1.5000 R R
M_{12}	Z $\epsilon_1 = 0.05$ $\epsilon_2 = 0.10$	0.5572 A A	0.7814 A A	0.9452 A A	1.1243 A A	1.1571 A A	1.1742 A A	1.2017 A A	1.2571 A R	1.3740 R R	1.5000 R R
M_{13}	Z $\epsilon_1 = 0.05$ $\epsilon_2 = 0.10$	0.7324 A A	0.7211 A A	0.6814 A A	0.7415 A A	0.8999 A A	1.1215 A A	1.5432 R R	1.6780 R R	1.7641 R R	1.8000 R R
M_{14}	Z $\epsilon_1 = 0.05$ $\epsilon_2 = 0.10$	0.7324 A A	0.7211 A A	0.6915 A A	0.7621 A A	0.9120 A A	1.1714 A A	1.7450 R R	1.7785 R R	1.8417 R R	2.1071 R R
M_{15}	Z $\epsilon_1 = 0.05$ $\epsilon_2 = 0.10$	1.3162 A A	1.6708 R R	1.5477 R R	1.4743 R R	1.4242 R R	1.3872 R R	1.3685 R R	1.3951 R R	1.4216 R R	1.6000 R R
M_{16}	Z $\epsilon_1 = 0.05$ $\epsilon_2 = 0.10$	1.3162 A A	1.3407 A A	1.5386 R R	1.4372 R R	1.4245 R R	1.4010 R R	1.3850 R R	1.3871 R R	1.4320 R R	1.7000 R R

Z : STATISTIC OF THE FRIST KOLMOGROV-SMIRNOV TEST.

A : ACCEPT THE HYPOTHESIS OF NORMALITY.

R : REJECT THE HYPOTHESIS OF NORMALITY.

Table (6.12) RESULTS OF THE FIRST KOLMOGROV_SMIRNOV TEST FOR THE ADDITIVE MODEL M_1.

TEST STATISTICS	LAG (DAYS)									
	5	10	15	20	25	30	35	40	45	50
Z	0.316228	0.447214	0.365148	0.316228	0.282843	0.258199	0.358569	0.335410	0.316228	0.300000
$\epsilon_1 = 0.05$	A	A	A	A	A	A	A	A	A	A
$\epsilon_2 = 0.10$	A	A	A	A	A	A	A	A	A	A
PROB.	0.99965	0.988261	0.99942	0.999965	0.999998	1.00000	0.999524	0.999871	0.999965	0.999991

A : ACCEPT THE MULL HYPOTHESIS Ho.

PROB: PROBABILITY OF ACCEPTANCE OF THE NULL HYPOTHESIS Ho.

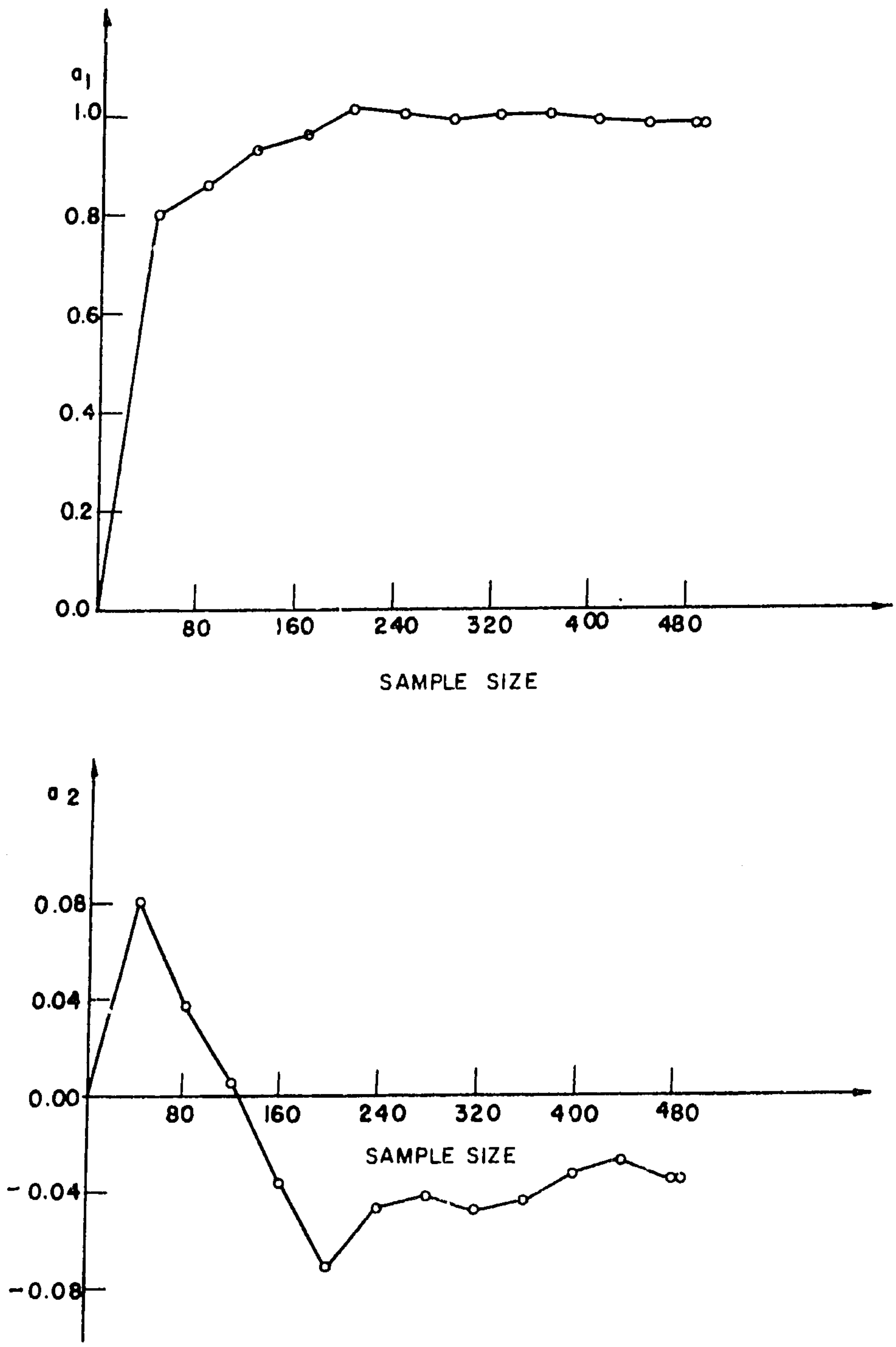


Fig. (6.22) VARIATION OF THE COEFFICIENTS OF THE MODEL M_1 WITH THE SAMPLE SIZE IN DAYS.

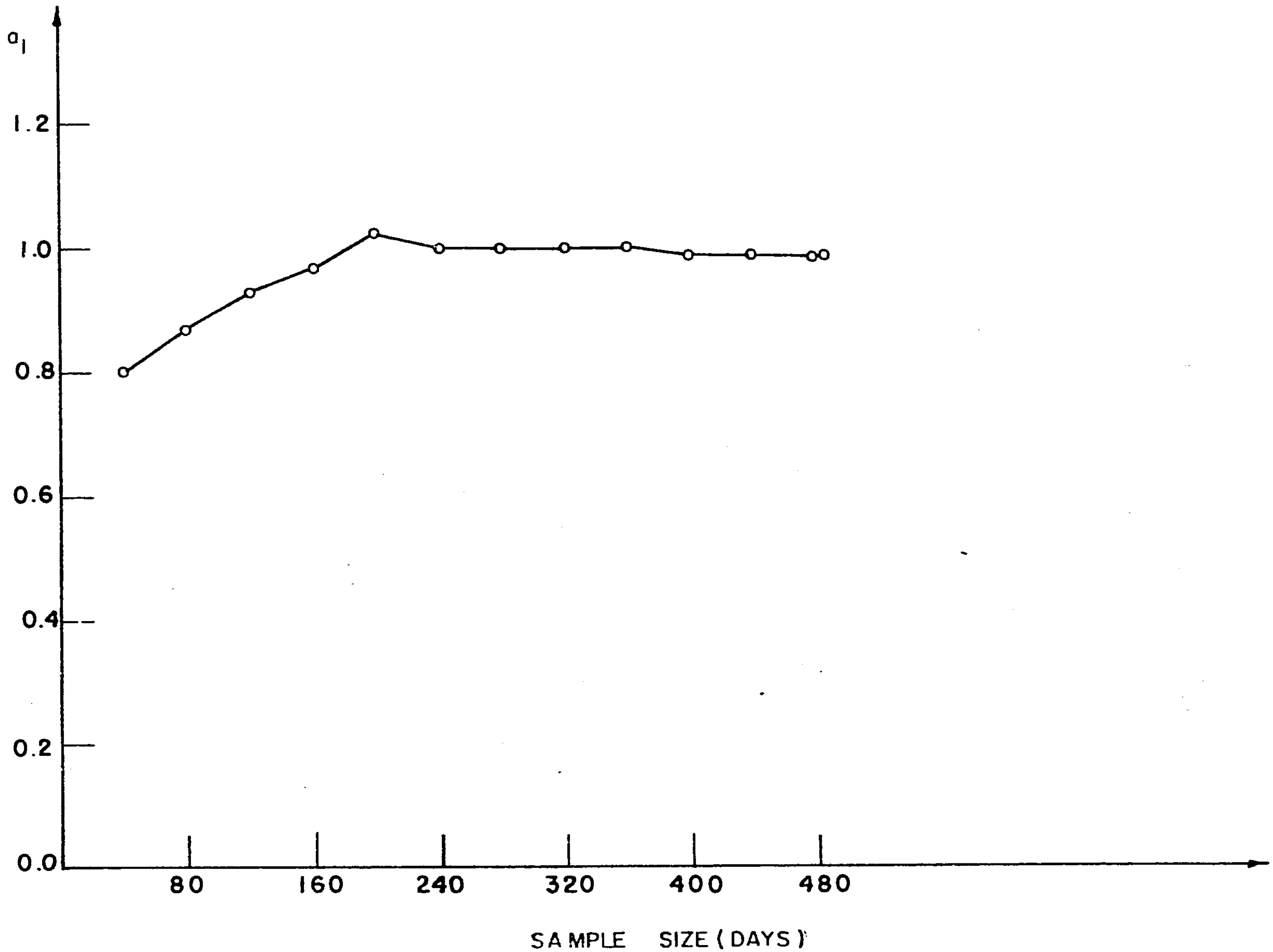


Fig.(6.23) VARIATION OF THE COEFFICIENTS OF THE MODEL M_3 WITH THE SAMPLE SIZE.

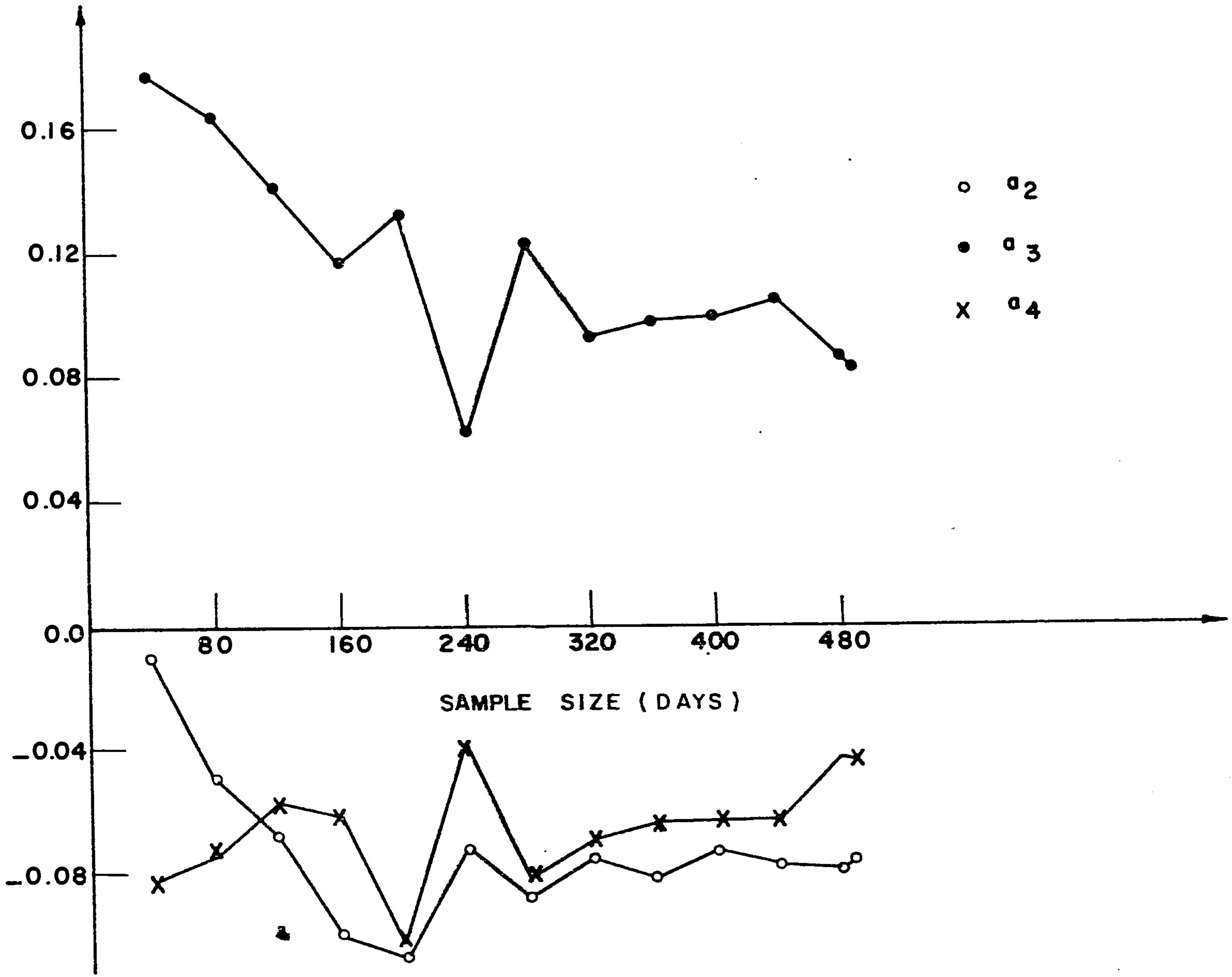


Fig. (6.23) CONT.'D

i) Simulation Capability:

The discrepancy between the statistical characteristics of the observed and generated sequences discriminates its simulation capability. Thus, some statistical characteristics such as correlograms, power spectrums, histograms and cumulative histograms of the two output sequences generated by M_4' and M_1 are compated with those of the measured output discharge. The results of that comparative procedure are illustrated in Tables (6.13) to (6.15), which confirm the ability of model M_1 to generate an adequate output sequence.

ii) Estimatability:

Some bisidary estimation conditions play an active part in the model selection techniques. The estimatability of a given model may insure its ability to generate an accurate estimates of parameters as well as appropriate statistics of residuals. Consequently, the significance of estimated parameters for the two successful models M_4' and M_1 may be tested as suggested by Clark (1969). The numerous mathematical operations needed to evaluate the impulse response vector $\underline{U}$ lead to a marginal significance of its coefficients, whereas the parameters of M_1 estimated by the recursive algorithm (4.11) show a small variability and better level of significance. On the other hand, the discrepancy between the histogram of residuals and the normal distribution curve, with similar mean and variance, is more acceptable for M_1 rather than M_4'. Furthermore, the histogram of residuals as well as its cumulative values for the successful model M_1 are shown in Figs. (6.24) and (6.25) respectively.

iii) Forecasting:

According to the general classification of monthly output data illustrated in Fig. (6.26), the forecasting ability of the two successful models M_4' and M_1 can be quantatively compared via Fig. (6.27). Clearly, the one-step ahead prediction capability of the model M_1 is much better compared with that of M_4'.

Fig. (6.13) COMPARISON OF THE CORRELOGRAMS OF THE MEASURED AND GENERATED DISCHARGE DATA FOR THE TWO SUCCESSFUL MODELS M_4' AND M_1.

TYPE OF DATA	LAG (DAYS)									
	5	10	15	20	25	30	35	40	45	50
MEASURED	0.806272	0.677308	0.577811	0.495149	0.444279	0.401704	0.342013	0.227529	0.112896	0.053637
GENERATED BY M_4'	0.819805	0.687749	0.583387	0.501245	0.450134	0.407119	0.346501	0.230257	0.109087	0.058214
GENERATED BY M_1	0.804605	0.675017	0.574804	0.492689	0.442457	0.399962	0.340875	0.226302	0.111361	0.052620

Table (6.14) COMPARISON OF POWER SPECTRUMS FOR THE MEASURED AND GENERATED DISCHARGE DATA FOR THE SUCCESSFUL MODELS M'_4 AND M_1.

TYPE OF DATA	FREQUENCY (RADIANS / DAY)									
	$\pi/10$	$\pi/5$	$3\pi/10$	$2\pi/5$	$\pi/2$	$3\pi/4$	$7\pi/10$	$4\pi/5$	$9\pi/10$	π
MEASURED	0.502841	0.008400	0.360824	0.005531	0.278187	0.00361	0.215088	0.001280	0.071462	0.017034
GENERATED BY M'_4	0.52904	0.00000	0.371395	0.00000	0.286564	0.00000	0.220589	0.00000	0.069447	0.00000
GENERATED BY M_1	0.501801	0.008372	0.358947	0.005503	0.277045	0.003595	0.214372	0.001273	0.070491	0.016711

Table (6.15) COMPARISON OF THE MEASURED AND GENERATED DISCHARGE HISTOGRAMS FOR THE TWO SUCCESSFUL MODELS M_4' AND M_1

TYPE OF DATA	CLASS INTERVAL										
	0.3-04	0.4-0.5	0.5-0.6	0.6-0.7	0.7-0.8	0.8-0.9	0.9-1.0	1.0-1.1	1.1-1.2	1.2-1.3	1.3-1.4
OBSERVED	18	100	126	89	64	37	27	13	10	3	1
GENERATED By M_4'	25	91	128	94	63	38	24	14	8	2	1
GENERATED By M_1	14	104	127	96	60	39	26	9	11	2	-

COMPARISON OF THE MEASURED AND GENERATED DISCHARGE CUMULATIVE HISTOGRAMS FOR THE TWO SUCCESSUF MODELS M_4'&M_1

TYPE OF DATA	CLASS INTERVAL										
	0.3-0.4	0.4-0.5	0.5-0.6	0.6-0.7	0.7-0.8	0.8-0.9	0.9-1.0	1.0-1.1	1.1-1.2	1.2-1.3	1.3-∞
OBSERVED	18	118	244	333	397	434	461	474	484	487	488
GENERATED By M_4'	25	116	244	338	401	439	463	477	485	487	488
GENERATED By M_1	14	118	245	341	401	440	466	475	486	488	488

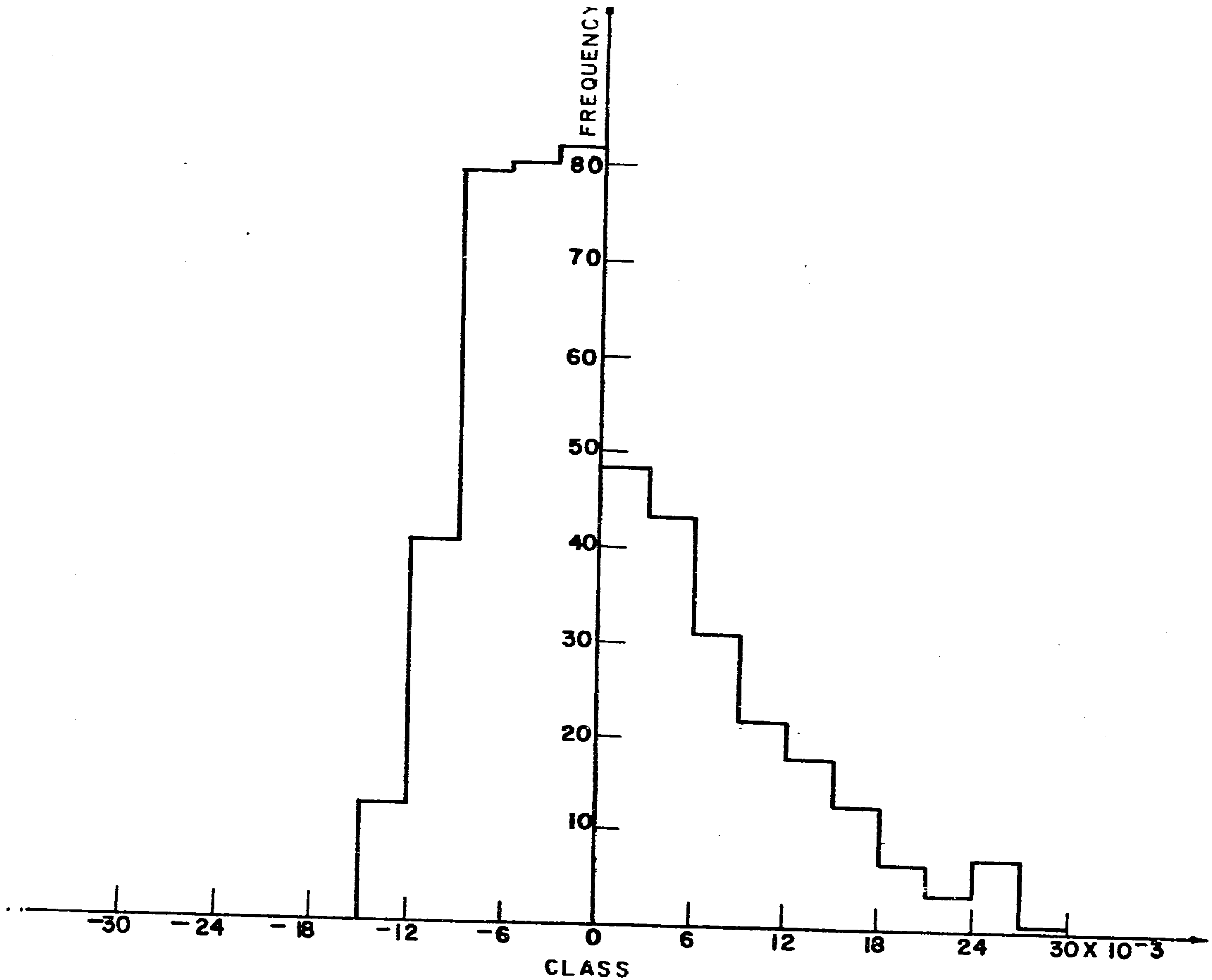


Fig. (6.24) HISTOGRAM OF THE RESIDUALS FOR THE MODEL M_1.

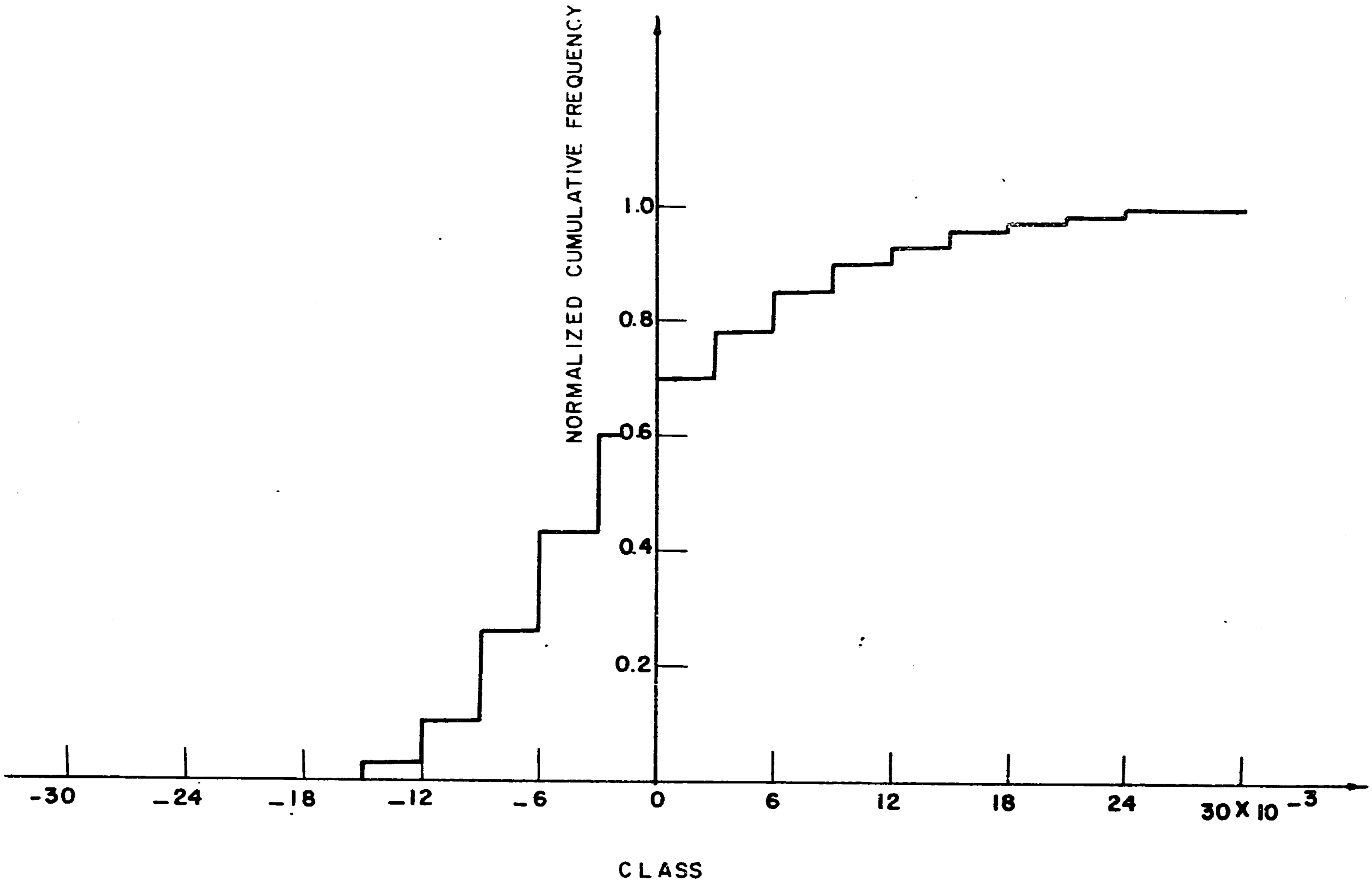


Fig.(6.25) NORMALIZED CUMULATIVE HISTOGRAM OF THE MODEL M_1 RESIDUALS.

TYPE OF LEVEL	LOWER LIMIT	UPPER LIMIT
I VERY LOW DISCHARGE	0	9.8
II LOW DISCHARGE	9.8	14.6
III MEDIUM LOW DISCHAR.	14.6	17.0
IV MEDIUM DISCHARGE	17.0	21.8
V MEDUIM HIGH DISCH.	21.8	24.2
VI HIGH DISCHARGE	24.2	29.0
VII VERY HIGH DISCHAR.	29.0	∞

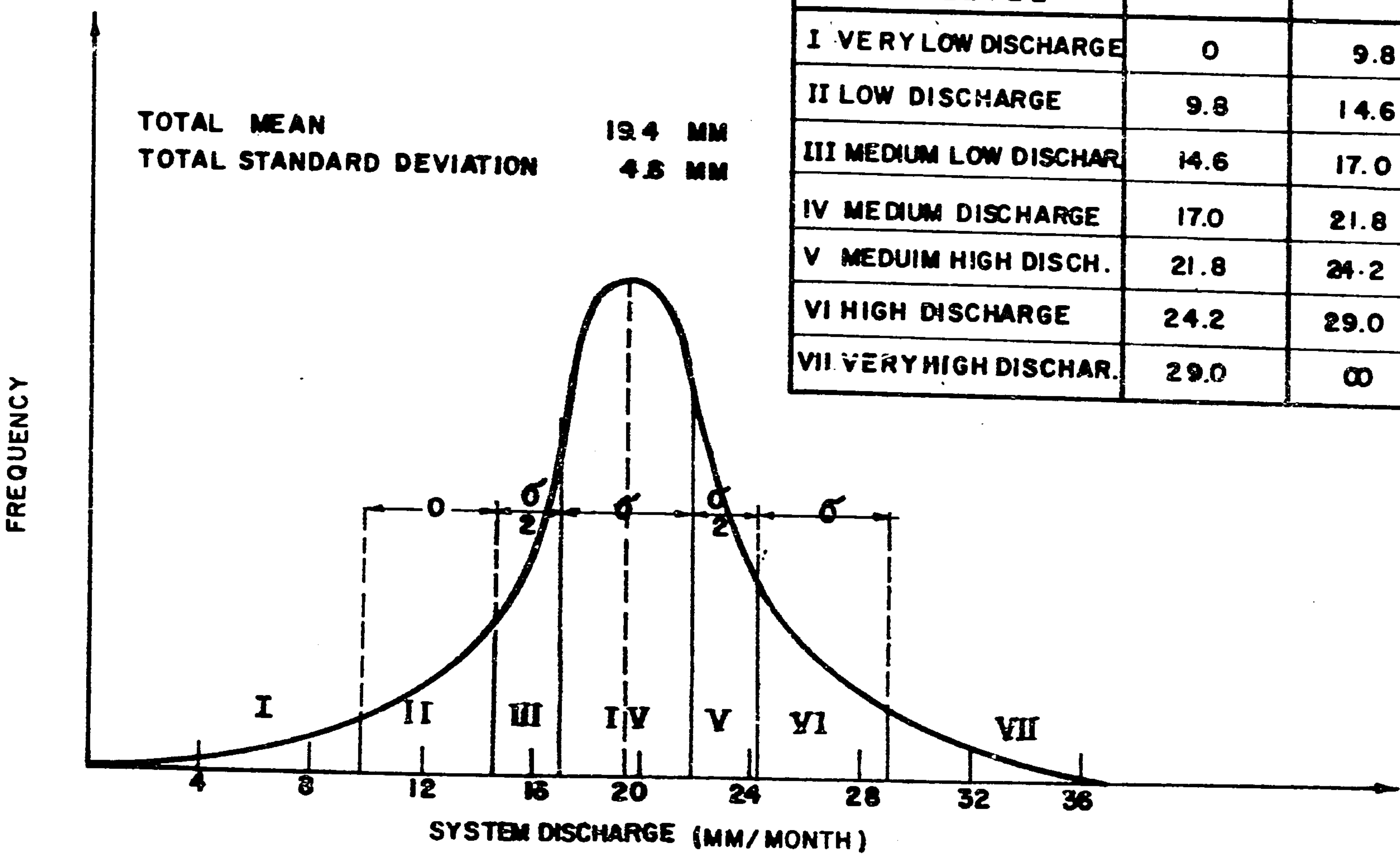


Fig. (6.26) CLASSIFICATION OF THE SYSTEM DISCHARGE.

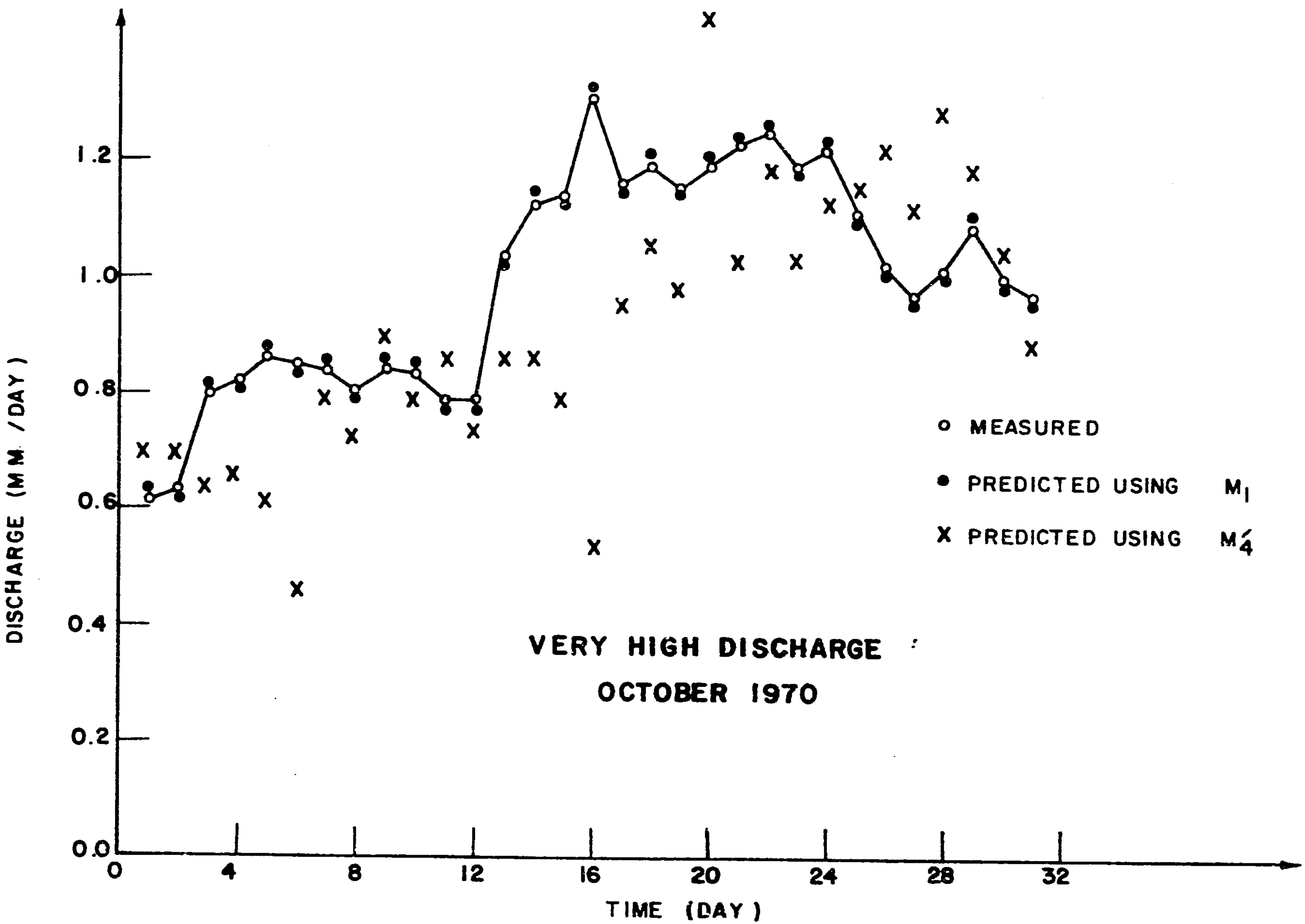


Fig. (6.27) A PLOT OF MEASURED AND PREDICTED DISCHARGE DATA FOR M_1 AND M_4' (TOTAL DISCHARGE IS 30.8 MM/MONTH).

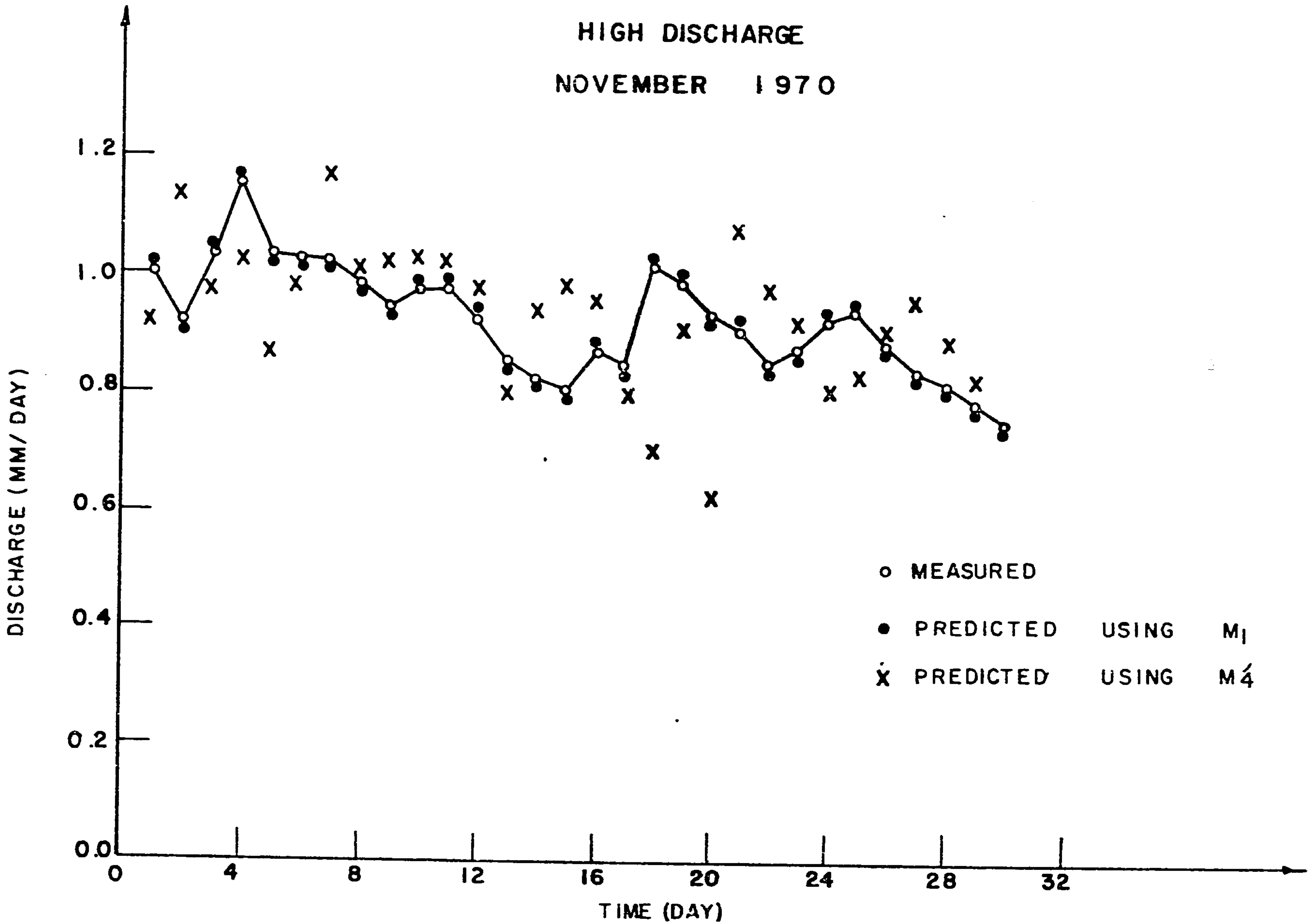


Fig. (6.27) CONT.'D (TOTAL DISCHARGE IS 27.3 MM/MONTH).

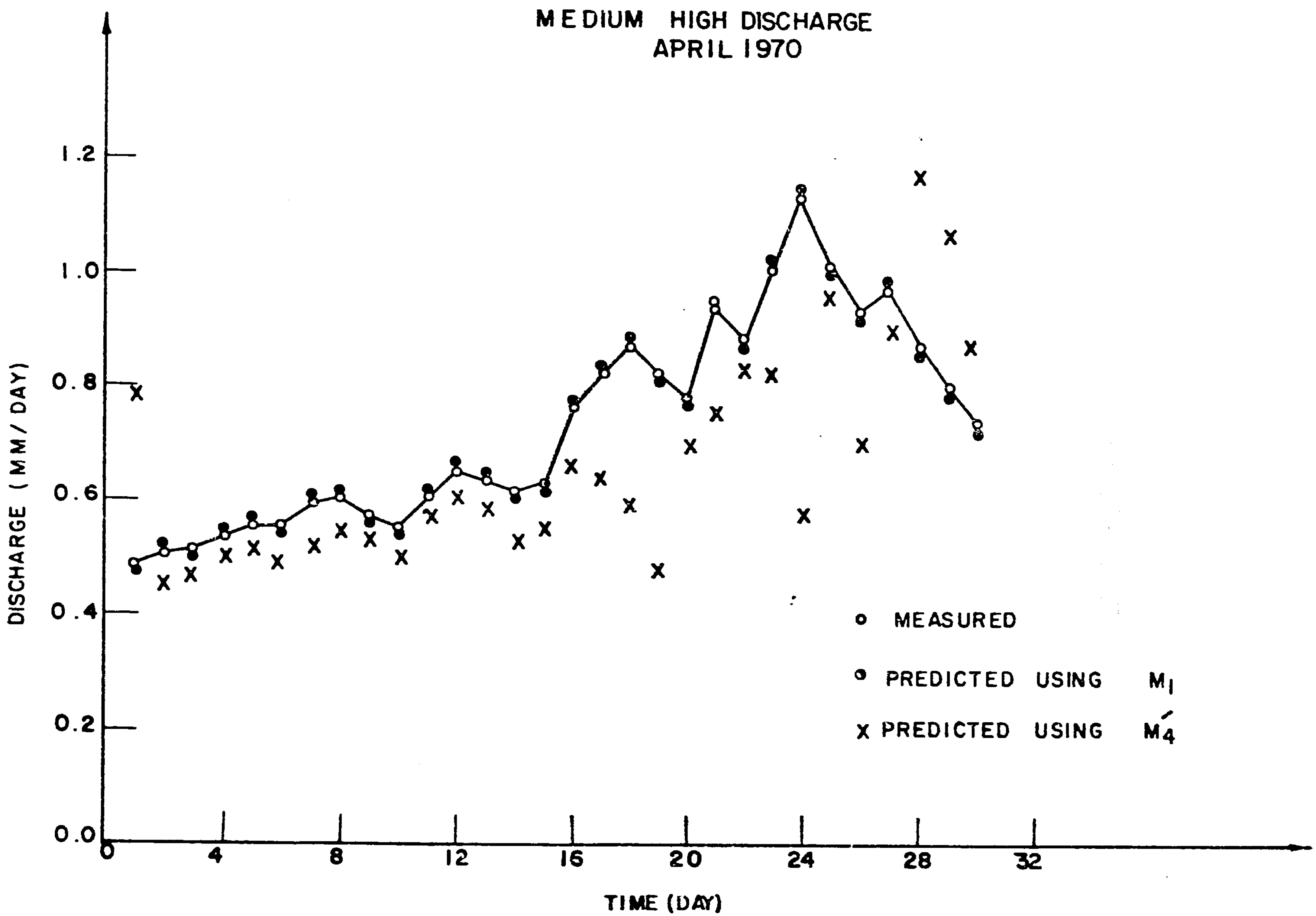


Fig.(6.27) CONT.'D (TOTAL DISCHARGE IS 21.9 MM/MONTH).

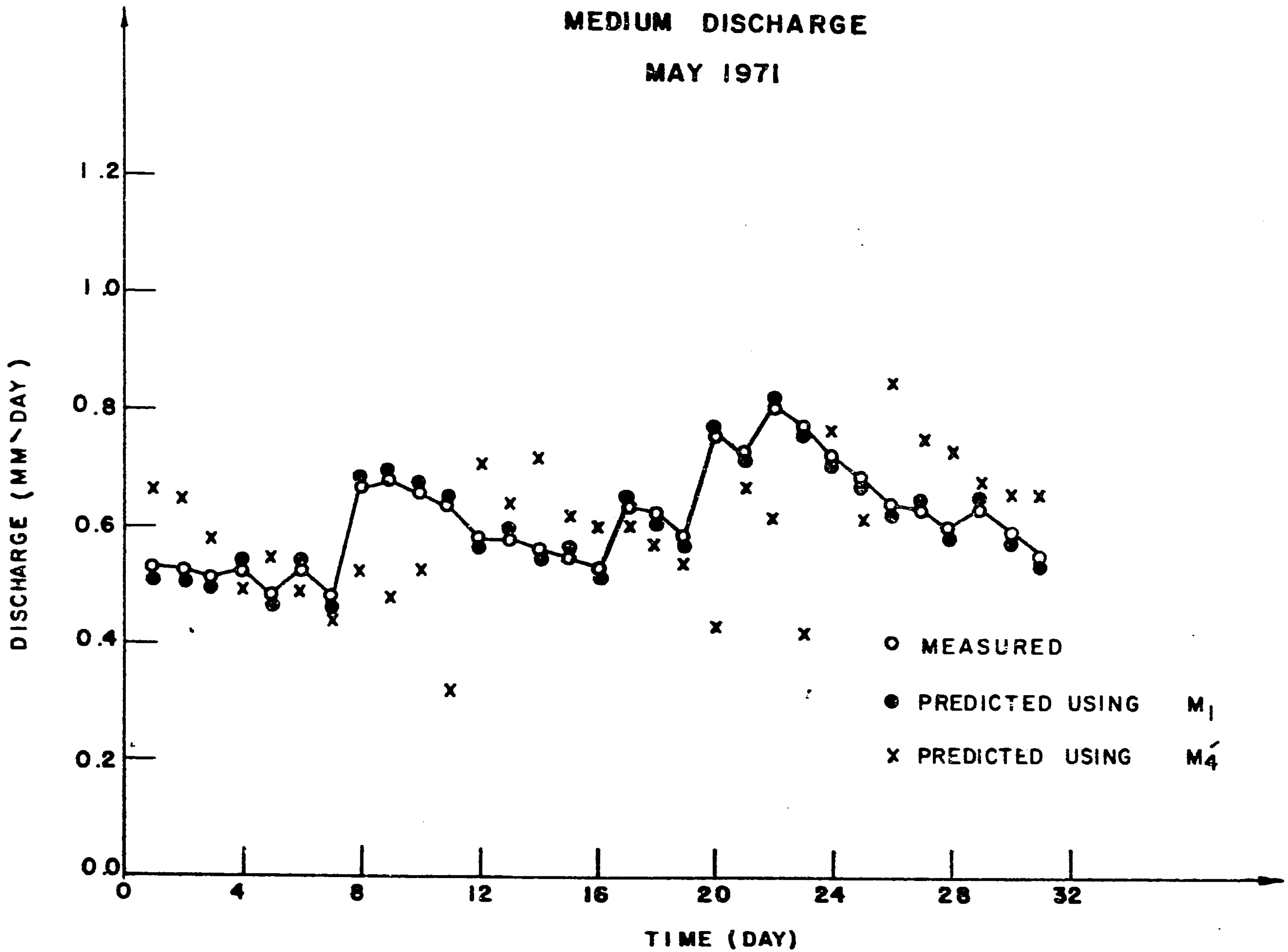


Fig.(6.27) CONT.'D (TOTAL DISCHARGE IS 18.6 MM/MONTH).

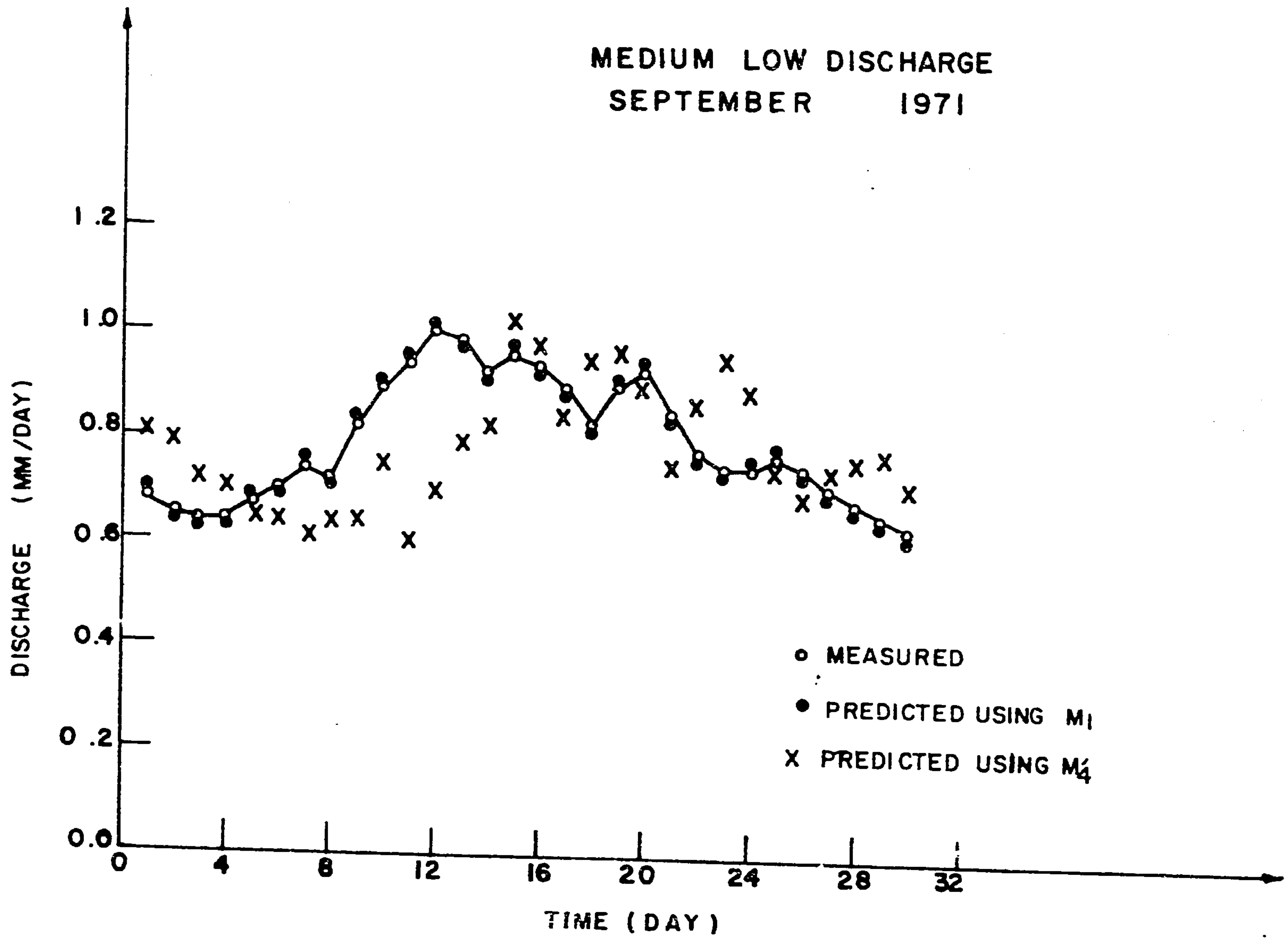


Fig.(6.27) CONT.'D (TOTAL DISCHARGE IS 15.4 MM/MONTH)

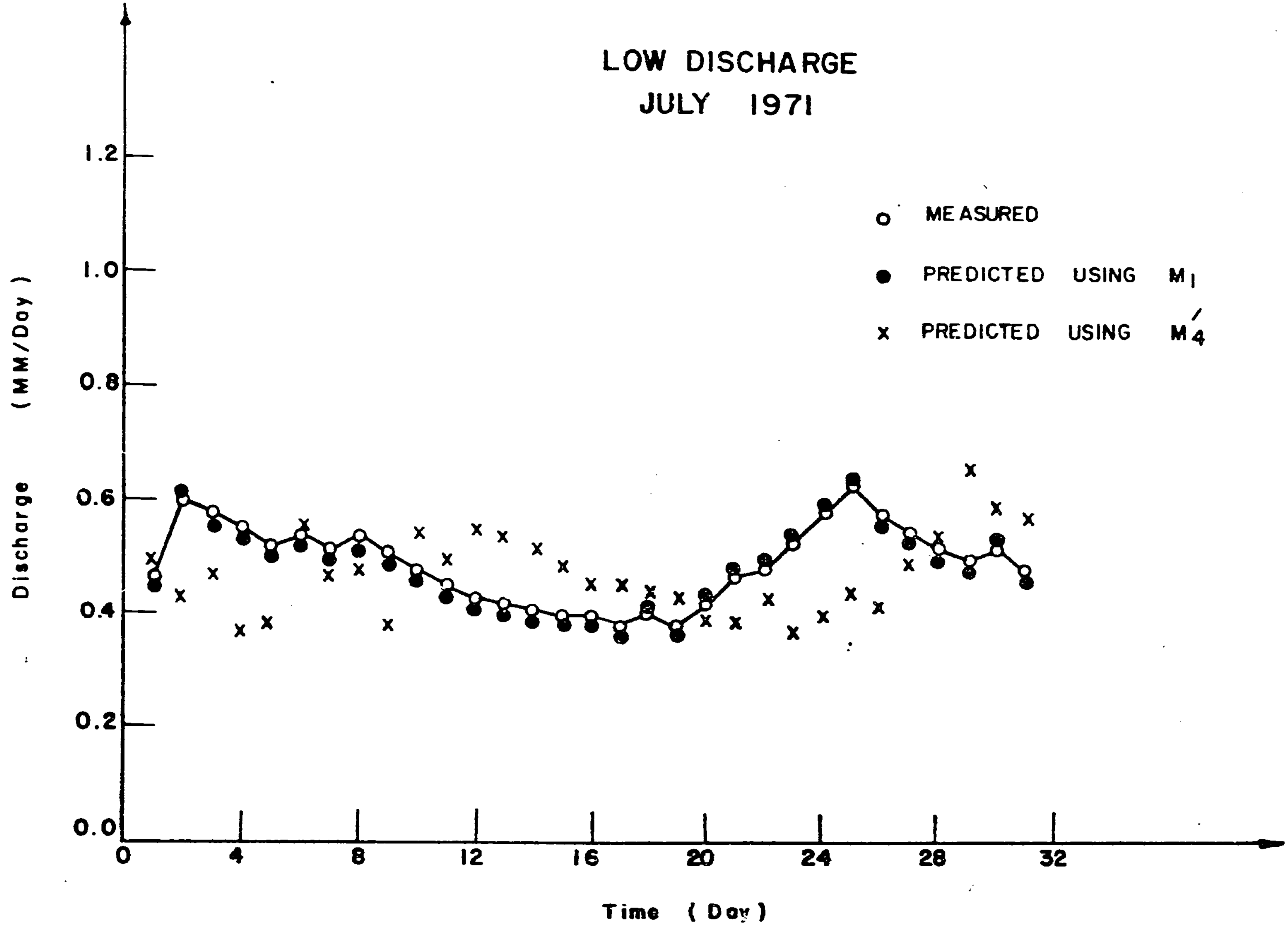


Fig.(6.27) CONT.D (TOTAL DISCHARGE IS 13.9 MM/MONTH).

CHAPTER VIII
CONCLUSIONS

CHAPTER VII

CONCLUSION

The major emphasis of this endeavour has been the identification, estimation and validation of noisy-transfer function and linear stochastic difference equation models appropriate for the representation of physical hydrological systems. A case study of the Waki River catchment, located near to lake Albert, has been selected. Using the input precipitation and the output discharge measured during the rainy season of that catchment, it has became possible to simulate the two proposed models on the digital computer together with the main statistical characteristics of their output data. Moreover, the validity of the residual sequences, generated by the different structures of these models, for the prespecified estimation conditions has also been investigated.

The important features of the two tuned noisy-transfer function and linear stochastic difference equation models have been quantitatively examined in a comparative pattern in order to achieve the best representation of the Waki catchment. As a general view, the performance of linear stochastic difference equation model is more favourable than that of the noisy-transfer function model.

The main findings of this work can now be summarized as follows:

i) The application of linear stochastic difference equation models is pragmatic for both prediction and estimation of the river catchment response.

ii) The linear stochastic difference equation models yield excellent prediction for the most given classifications, whereas the predictability of the noisy-transfer function models is restricted by using their autoregressive structure during the low level of output data.

iii) The multiplicative sturcture of the linear stochastic difference equation models has failed to attain the same accuracy obtained by the additive structure, this is mainly due to its inadequacy to the physical system at hand. Moreover, it is advisable to fit a relatively simple class of models and increase its complexity only if the simplest class proves to be unsatisfactory.

iv) The identification procedure of the linear difference equation model is equivalent to specifying the suitable number of autoregressive, corrective error and/or sinusoidal terms necessary for an adequate results. Alternatively, the basic premise in identifying the noisy-transfer function model is the evaluation of its appropriate kernel length.

v) It is advantageous to invoke the constrained estimators to evaluate the parameters of noisy-transfer function model adequate for some river catchment systems whose complete mathematical balance is available, together with the representability of their measured data. On the other hand, the recursive parameter estimation of the linear stochastic difference equation models is relevant for both the additive and multiplicative structures, provided that a proper data transformation procedure is manipulated.

vi) The validation of the two proposed families of models for the prespecified estimation conditions was checked both by examining their residuals and comparing the basic statistics of their generated output data such as mean, variance, correlogram, histograms and power spectrum with the others of observed sequence. It has been demonstrated that, the appropriate class of models should pass all validation tests at the required significance level in order to vendicate its adequacy for the system at hand.

The most fruitful area of future research would be the implementation of partitioned estimation technique together with the pre-whitening of the input data to the noisy-transfer function model. In addition, the sensitivity of linear stochastic difference equation model to the recursive manipulation of corrective error terms obtained via the Fourier analysis of residuals is suggested for further studies. Finally, it is recommended that the methodologies presented in this work be invoked to other physical systems in diverse areas of engineering and applied sciences, as well as to multi-input multi-output situations.

REFERENCES

REFERENCES

1) Anderson, T.W. (1970), "The Statistical Analysis of Time Series", John Wiley and Sons, New York.

2) Abadie, J. (1970), "Integer and Non-Linear Programing" North-Holland Publishing Co., Amsterdam.

3) Blank, D., Delleur, J.W. and Giorgini, A. (1971), "Oscillatory Kernel Functions in Linear Hydrologic Systems", Water Resources Research, Vol. 7, No. 5.

4) Box, G.E.P., and Hunter, W.G. (1965), "The Experimental Study of Physical Mechanisms", Technometrics, Vol. 7, No. 23.

5) Box, G.E.P., and Jenkins, G.M. (1970), "Time Series Analysis, Forecasting and Control", California, Holden Day, p. 553.

6) Clark, R.H. (1969), "College Statistics", Thomas Nelson and Sons, London, p. 267.

7) Diskin, M.H. and Boneh, A. (1973), "Determination of Optimal Kernels for Second-Order Stationary Surface Runoff Systems", Water Resources Research, Vol. 9, No. 2.

8) Dixon, W.J. (1970), "Biomedical Computer Programs", University of California Publications in Automatic Computations No. 2, University of California Press, Berkeley, California.

9) Dorrah, H.T. and Nichols, S.T. (1975), "Nonlinear Parameter Estimation via a Stochastic Steepest Descent Approach", Automatic Control Theory and Applications, Vol. 3, No. 2.

10) Dorn, W.S. (1963), "Nonlinear Programing-A survey", Management Science, Vol. 9, No. 2.

11) Feller, W. (1948), "On the Kolmogrov-Smirnov Limit Theorems for Empirical Distributions", Annals of Math. Stat., p. 177-189.

12) Gnedenko, B.V. (1962), "The Theory of Probability", Chelsea Publishing Co., New York, p. 384-401.

13) Hannon, E.J. (1960), "Time Series Analysis", Methuen, London.

14) Kamal Abo El-Hassan (1980), "Some Autoregressive Moving Average Models for the River Nile", M. Sc. Thesis, Cairo University, Cairo.

15) Kashyap, R.L. and Rao, A.R. (1973), "Real Time Recursive Prediction of River Flows", Automatica, Vol. 9, p. 175 -183.

16) Kashyap, R.L, and Rao, A.R. (1974), "Stochastic Modeling of River Flows", IEEE Trans. on Automatic Control, Vol. AC - 10, No. 6.

17) Kashyap, R.L, and Rao, A.R. (1976), "Dynamic Stochastic Models from Empirical Data", Academic Press, New York, p. 334.

18) Lilliefors, H.W. (1967), "On the Kolmogrov-Smirnov Test for Normality with Mean and Variance Unknown", J.A.S.A., p. 399-402.

19) Mass, N.J. and Senge, P.M. (1978), "Alternative Tests for the Selection of Models Variables", IEEE Trans. on System, Man and Cybernetics, Vol. SMC-8, No. 6.

20) Natale, L. and Todini, E. (1974), "A Constrained Parameter Estimation Technique for Linear Models in Hydrology", Proceedings of Workshop on Mathematical Models in Hydrology, IBM Piza Scientific Center, Piza, Italy.

21) Natale, L. and Todini, E. (1976), "A Stable Estimator for Linear Models", Water Resources Research, Vol. 12, No. 4.

22) Papoulis, A. (1965), "Probability, Random Variables and Stochastic Processes, McGraw-Hill Book Co., New York.

23) Sage, A.P. and Melsa, J.L. (1974), "Estimation Theory with Applications to Communications and Control", McGraw-Hill Book Co., New York.

24) Whittle, P. (1951), "Hypothesis Testing in Time Series Analysis", Almquist and Wiksell, Upsala.

25) Whittle, P. (1952), "Tests of Fit in Time Series Analysis", Biometrika, Vol. 39, p. 309-318.

26) Wilson, R.B. (1963), "A Simplified Algorithm for Concave Programming", Harvard University, Boston.

27) Wittaker, E.T. and Watson, G.N. (1952), "A Course of Modern Analysis", Cambridge University Press, Cambridge, England, p. 462-476.

28) WMO (World Meteorological Organization) (1972), Technical Reports, Geneva, Vol. I to IV.

29) Yule, G.U. (1927), "On a Method of Investigating Periodicities in Disturbed Series, with Spectial Reference to Wölfer's Sunspot number", Phil.Trans., p. 226-267.

APPENDIX A

ESTIMATION OF THE POWER SPECTRUM

APPENDIX A

ESTIMATION OF THE POWER SPECTRUM

The power spectral density function can be obtained by using the following formula [Kamal Abo El-Hassan (1980)]

$$PS(w_h) = \frac{2}{\pi} \sum_{k=0}^{M} E_k \, \gamma(k) \cos \frac{khx}{M} \qquad (A.1)$$

where w_h is the frequency in radians per unit time,

$$w_h = \frac{h\pi}{M} , \qquad h = 0, 1, \ldots, M \qquad (A.2)$$

and

$$E_k = \begin{cases} 1 & \text{for } 0 < k < M \\ \frac{1}{2} & \text{for } k = 0, \; M \end{cases} \qquad (A.3)$$

such that $\gamma(k)$ is the normalized autocorrelation function at lag k and M is an integer nearest to 0.1N or 0.05N, such that N denotes the number of observations.

These estimates are then smoothed employing the Hamming window alogrithm to obtain more refined values of the power spectrum, that is

$$\left.\begin{array}{ll} \text{for} & k = 0: S(w_o) = 0.54\ PS(w_o) + 0.46\ PS(w_1), \\ \text{for} & 0<k<M: S(w_k) = 0.23\ PS(w_{k-1}) + 0.54\ PS(w_k) + 0.23\ PS(w_{k+1}), \\ \text{and for} & \\ & k = M: S(w_M) = 0.54\ PS(w_M) + 0.46\ PS(w_{M-1}) . \end{array}\right\} \qquad \text{(A.4)}$$

The accuracy of computation was checked for the above procedure by evaluating

$$(\pi/M)\left[\frac{1}{2}\,[S\ (w_o) + S\ (w_M)] + \sum_{k=1}^{M-1} S\ (w_k)\right] \qquad \text{(A.5)}$$

which must be equal to $\gamma(0)$, see Dixon (1970).

APPENDIX B
LIST OF THE DIGITAL CMPUTER PROGRAM FOR THE NOISY-TRANSFER FUNCTION MODEL

```
PROGRAM(SOLT)
INPUT 1=CRO
OUTPUT 2=LPO/160
TRACE 0
END




MASTER RAIN
C     ************************************************************************
C     OBJECT:
C     THIS PROGRAM IS USED TO IDENTIFY THE RELATIONSHIP BETWEEN INPUT
C     RAINFALL AND OUTPUT DISCHARGE FOR WAKI CATCHEMENT.
C     THIS HYDROLIC SYSTEM CAN BE CONSIDERED AS LINER CONSTRAINED ONE
C     UNITS OF INPUT RAINFALL ARE MM,WHICH IS THE SAME AS DISCHARGE
C     DATA SOURCE : NILE BASIN SERIES FOR YEARS 1970,1971,1973,AND 1974
C     THE OBSERVED DATA IS RECORDED DAILY FOR THE GIVEN CATCHEMENT
C     ////////////////////////////////////////////////////////////////////////
C
      DIMENSION PORTAT(488),VINPUT(488),APIX(488),QH(488),B(130),A(78),
     *X(12),S1(12),U(12),XSTAR(12),R(12,37),IBB(12),IQ(12),IR(12),IB(37)
     *,JC(37),STAT(12),V(12)
      DIMENSION R1(49),R2(49),PS(49)
      DOUBLE PRECISION A,B,X,U,V,S,SQ,XSTAR,EPS
      INTEGER TI,TC,PP,QQ,TB,FL,TYPE
      COMMON /A1/S(2),SQ(2),ND
      COMMON /A2/MJ,MI,MD
      DATA TF,PP,QQ/10,'PP','QQ'/
      DATA ML,PHI/49,3.1415927/
C     ************************************************************************
C     READING FORMATS
 1000 FORMAT(7I2)
 1030 FORMAT(2F4.1,2F4.0)
 1040 FORMAT(8F0.0)
 1050 FORMAT(2I4)
C
C     MAIN PROGRAM OUTPUT FORMATS
 2000 FORMAT(1H1,//,10X,'THE C.L.S IS LINEAR AND TIME INVARIANT',
     */,10X,45(1H*))
 2010 FORMAT(1H1,//,10X,'THE C.L.S IS NONLINEAR AND TIME INVARIANT',
     */,10X,45(1H*))
 2020 FORMAT(1H1,//,10X,'THE C.L.S IS NONLINEAR AND TIME VARIANT',
     */,10X,45(1H*))
 2030 FORMAT(/,10X,'THE ORDINARY LEAST SQUARE IS USED TO ESTIMATE'
     *,' THE IMPULSE RESPONSE VECTOR',/,10X,80(1H*))
 2040 FORMAT(/,10X,'ONLY INEQUALITY CONSTRAINTS ARE USED TO ESTIMATE'
     *,' THE IMPULSE RESPONSE VECTOR',/,10X,81(1H*))
 2060 FORMAT(//,10X,'X-VALUES:',/,6(6X,F8.5),/)
C     ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
C
      READ(1,1050) ND,MD
      READ(1,1000) IP,IS,IU,IW,NT,LAG,TC
      READ(1,1040) (PORTAT(I),I=1,ND)
      NV,NC,KS,IFL=0
      TI,IAL=0
```

```
      KF=ND
      NS=IW-IU+1
      IF(NT.EQ.0) GO TO 20
      READ(1,1030) XM,AL,FHI,TR
      IF(AL.NE.0.0) IAL=1
   20 TI=TA+NT+1
      KI=TI
      IF(IU.EQ.0) GO TO 40
      DO 30 I=KI,ND
   30 VINPUT(I)=PORTAT(I)
      GO TO 60
   40 IF(IFL) 50,50,60
   50 READ(1,1040) (VINPUT(I),I=KI,ND)
   60 IF(LAG.EQ.0) GO TO 90
      JF=KF
   70 VINPUT(JF)=VINPUT(JF-LAG)
      IF(JF.EQ.(KI+LAG))GO TO 75
      JF=JF-1
      GO TO 70
   75 IF(TYPE.EQ.QQ) GO TO 90
      DO 80 I=KI,JF
   80 VINPUT(I)=0.0
   90 MV=MD-NC
      TB=TB+1
      NV=NV+TB
      IF(NV.LE.MV) GO TO 100
      GO TO 260
  100 FL=1
      DO 160 J=1,MD
      X(J)=0.0
      U(J)=0.0
      V(J)=0.0
      S1(J)=0.0
      IEB(J)=0
  160 IQ(J)=0
      IF(TC-1) 190,170,170
  170 DO 180 K=1,MD
  180 IQ(K)=1
  190 CALL INPUT1(VINPUT,PORTAT,TI)
  210 CALL MATR(VINPUT,PORTAT,TI,TB,FL,A,B,NV,NC)
      IZ=NV+NC
      JZ=3*IZ+1
      CALL ALGO(A,B,X,U,V,S1,IEB,IQ,R,XSTAR,IB,JC,IR,NV,NC,IZ,JZ,KA)
      SK=0.0
      DO 220 N=1,NV
  220 SK=SK+X(N)
      IF(KA.NE.0.OR.SK.EQ.0.0) GO TO 260
      WRITE(2,1060) (X(I),I=1,NV)
      CALL QHAT(VINPUT,PORTAT,QH,X,TB,TI,KS,NS)
      CALL ERROR(PORTAT,QH,STAT,TI,APIX,IW)
      WRITE(2,2000)
      GO TO 250
      IF(IAL) 230,230,240
  230 WRITE(2,2010)
      GO TO 250
  240 WRITE(2,2020)
  250 IF(TC.EQ.0) WRITE(2,2030)
      IF(TC.EQ.1) WRITE(2,2040)
      IF(TC.EQ.2) WRITE(2,2050)
      CALL WRITE2(STAT,PORTAT,QH,X,NV)
      CALL MISE(PORTAT,QH,APXX,ND)
  260 STOP
      END
```

```
      SUBROUTINE MISR(Y,YE,YR,ND)
C     PURPOSE:
C     IT TESTS THE RESIDUAL VECTOR YR.
C     ********************************************************************
C
      DIMENSION Y(488),YE(488),YR(488)
      DOUBLE PRECISION COR(50),GAMA(50,50),CORO
      EQUIVALENCE (COR(1),GAMA(1,1))
      DATA E0,E1,E2,U,IFCOD,S/4*0.0,1,1.0/
      DATA IC,IA,ML,ISIZE/2*0,50,5/
C     ********************************************************************
C     OUTPUT FORMATS:
 1000 FORMAT(10X,3HE0=,F10.6,10X,3HE1=,F10.6,10X,3HE2=,F10.6,/)
 1010 FORMAT(//,10X,'TESTING OF THE RESIDUALS:')
 1020 FORMAT(10X,'TEST:2',/,10X,2HZ=,F10.6,/,10X,5HPROB=,F10.6,/,10X,
     *'MAXIMUM DIFFERENCE DN=',F10.6,/)
 1030 FORMAT(10X,'SECTION:2',/,10X,'KOLMOGROV SMIRNOV TEST.',/,10X,
     *'TEST:1',/,10X,2HZ=,F10.6,/,10X,5HPROB=,F10.6)
 1040 FORMAT(10X,'SECTION:1',/,10X,'MEANS OF THE RESIDUAL.')
 1050 FORMAT(10X,'SECTION 3',/,10X,'THE F-TEST.',/,10X,'VALUE=',F10.6,
     *10X,'LAG=',I3,/,10X,60(1H*),/)

 1060 FORMAT(10X,'MLL=',I2,10X,'I=',I2,10X,'GAMA(MLL,I)=',D26.20)
C     ====================================================================
C
      ISUM=IA+IC+1
      WRITE(2,1010)
      NN=ND-ISUM
      DO 10 I=ISUM,ND
      E0=E0+YR(I)/NN
      E1=E1+ABS(YR(I)/NN)
      E2=E2+(YR(I)**2)/NN
   10 CONTINUE
      WRITE(2,1040)
      WRITE(2,1000) E0,E1,E2
      CALL AUTO(YE,ND,ML,IFCOD,COR,CORO)
      CALL AUTO(YR,NN,ML,ISUM,COR,CORO)
      DO 20 J=2,ML
      DO 20 I=J,ML
      GAMA(I,J)=GAMA(I-1,J-1)
      GAMA(J-1,I)=GAMA(I,J-1)
   20 CONTINUE
      CALL SINA(GAMA,ML,IREV)
      DO 30 I=ISIZE,ML,ISIZE
      MLL=I
      CALL KOLM1(YR,I,IER,IFCOD,U,S,PROB1,Z1)
      CALL KOLM2(Y,YE,I,I,Z2,PROB2,DN)
      WRITE(2,1030) Z1,PROB1
      WRITE(2,1020) Z2,PROB2,DN
      WRITE(2,1060) MLL,I,GAMA(I,I)
      FF=CORO/GAMA(I,I)
      FT=(FF-1.0)*((NN/I)-1)
      WRITE(2,1050) FT,I
   30 CONTINUE
      CALL KOLM1(YR,ND,IER,IFCOD,U,S,PROB1,Z1)
      CALL KOLM2(Y,YE,ND,ND,Z2,PROB2,DN)
      WRITE(2,1030) Z1,PROB1
      WRITE(2,1020) Z2,PROB2,DN
      CALL CCOR(Y,YR,ML,ND)
      RETURN
      END
```

```
      SUBROUTINE SMIRN(X,Y)
C     PURPOSE:
C     CALCULATES VALUES OF THE LIMITING DISTRIBUTION FUNCTION FOR
C     THE KOLMOGROV-SMIRNOV STATISTIC.
C     ******************************************************************
C
      DOUBLE PRECISION X,C1,C2,C4,C8,Y
      IF(X-0.27) 1,1,2
    1 Y=0.0
      GO TO 9
    2 IF(X-1.0) 3,6,6
    3 C1=EXP(-1.233701/X**2)
      C2=C1*C1
      C4=C2*C2
      C8=C4*C4
      IF(C8-1.0E-25) 4,5,5
    4 C8=0.0
    5 Y=(2.506628/X)*C1*(1.0+C8*(1.0+C8*C8))
      GO TO 9
    6 IF(X-3.1) 8,7,7
    7 Y=1.0
      GO TO 9
    8 C1=EXP(-2.0*X*X)
      C2=C1*C1
      C4=C2*C2
      C8=C4*C4
      Y=1.0-2.0*(C1-C4+C8*(C1-C8))
    9 RETURN
      END
```

```
      SUBROUTINE SINA(B,KC,IREV)
      DOUBLE PRECISION B(50,50),TEMP
      IREV=0
      DO 20 I=1,KC
      K=I
    9 IF(B(K,I)) 11,10,11
   10 K=K+1
      IF(K-KC) 9,9,51
   11 IF(I-K) 12,14,51
   12 DO 13 M=1,KC
      TEMP=B(I,M)
      B(I,M)=B(K,M)
   13 B(K,M)=TEMP
      IREV=IREV+1
   14 II=I+1
      IF(II.GT.KC) GO TO 51
      DO 17 M=II,KC
   18 IF(B(M,I)) 19,17,19
   19 TEMP=B(M,I)/B(I,I)
      DO 16 N=1,KC
   16 B(M,N)=B(M,N)-B(I,N)*TEMP
   17 CONTINUE
   20 CONTINUE
   51 RETURN
      END
```

```
      SUBROUTINE NDTR(X,P,D)
      AX=ABS(X)
      T=1.0/(1.0+0.2316419*AX)
      D=0.3989423*EXP(-X*X/2.0)
      P=1.0-D*T*((((1.330274*T-1.821256)*T+1.781478)*T-0.3565638)*
     *T+0.3193815)
      IF(X) 1,2,2
    1 P=1.0-P
    2 RETURN
      END


      SUBROUTINE OLS(LX,LY,M1,M2,IN1,IN2,W,Y,RO,A,B,NV,NC)
C     MATRIX CALCULATIONS
      DIMENSION W(488),Y(488),A(78),B(130)
      INTEGER UX,UY,UL
      DOUBLE PRECISION A,B,XX,XY,AX,S,SD,RO
      COMMON /A1/S(2),SD(2),ND
      UX=M1+LX-1
      UY=M2+LY-1
      LX1=LX+1
      LY1=LY+1
      IF(LX.NE.LY) GO TO 4
      DO 3 J=LY,UY
      JM=J-LY
      M=JM+1
      I=LX+J*(J-1)/2
      AX=0.0
      DO 1 K=M,ND
      XX=W(IN1+K)
      XY=Y(IN2+K-JM)
      AX=AX-XX*XY
    1 CONTINUE
      A(I)=AX/RO
      IF(JM.LE.0) GO TO 3
      DO 2 I1=LX1,J
      I=I1+J*(J-1)/2
      A(I)=A(I-J)
    2 CONTINUE
    3 CONTINUE
      RETURN
    4 DO 7 J=LY,UY
      JM=J-LY
      M=JM+1
      I=LX+J*(J-1)/2
      AX=0.0
      DO 5 K=M,ND
      XX=W(IN1+K)
      XY=Y(IN2+K-JM)
      AX=AX-XX*XY
    5 CONTINUE
      A(I)=AX/RO
      UL=MINO(LX+JM,UX)
      IF(LX1.GT.UL) GO TO 7
      DO 6 I1=LX1,UL
      I=I1+J*(J-1)/2
      A(I)=A(I-J)
    6 CONTINUE
    7 CONTINUE
```

```
      DO 10 I1=LX1,UX
      IM=I1-LX
      M=IM+1
      I=I1+LY*(LY-1)/2
      AX=0.0
      DO 8 K=M,ND
      XX=W(IN1+K-IM)
      XY=Y(IN2+K)
      AX=AX-XX*XY
    8 CONTINUE
      A(I)=AX/R0
      UL=MINO(LY+IM-1,UY)
      IF(LY1.GT.UL) GO TO 10
      DO 9 J=LY1,UL
      I=I1+J*(J-1)/2
      A(I)=A(I-J)
    9 CONTINUE
   10 CONTINUE
      RETURN
      END
```

```
      FUNCTION VALUE(X,I,J,K,A,B,NV)
      DIMENSION A(78),B(130),X(12)
      DOUBLE PRECISION A,B,X,AU,VX,TN
      DATA TN/0.D/
      IF(K.NE.0) GO TO 8
C     OPERATION ON THE OBJECTIVE FUNCTION (K=0)
      IF(I.NE.0) GO TO 7
      IF(J.NE.0) GO TO 4
C     CALCULATION OF THE VALUE OF THE OBJECTIVE FUN.
      VX=TN
      NT=NV*(NV+1)/2
      DO 3 JN=1,NV
      AU=0.0
      DO 1 IN=1,JN
      IEX=IN+JN*(JN-1)/2
      AU=AU+A(IEX)*X(IN)
    1 CONTINUE
      JJ=JN+1
      IF(JJ.GT.NV) GO TO 3
      DO 2 IN=JJ,NV
      IEX=JN+IN*(IN-1)/2
      AU=AU+A(IEX)*X(IN)
    2 CONTINUE
    3 VX=VX+(AU/2.-(NT+JN))*X(JN)
      VALUE=VX
      RETURN
C     CALCULATION OF THE FIRST DERIVATIVE OF THE OBJ.-FUN.
    4 NT=NV*(NV+1)/2
      VX=-A(NT+J)
      DO 5 IN=1,J
      IEX=IN+J*(J-1)/2
      VX=VX+A(IEX)*X(IN)
    5 CONTINUE
      JJ=J+1
      IF(JJ.GT.NV) GO TO 61
      DO 6 IN=JJ,NV
      IEX=J+IN*(IN-1)/2
```

```
      VX=VX+A(IEX)*X(IN)
    6 CONTINUE
   61 CONTINUE
      VALUE=VX
      RETURN
C     CALCULATION OF THE SECOND DERIVATIVE OF THE OBJECTIVE FUN.
    7 IEX=I+J*(J-1)/2
      IF(I.GT.J) IEX=J+I*(I-1)/2
      VALUE=A(IEX)
      RETURN
    8 IF(I.NE.0) GO TO 11
      IF(J.NE.0) GO TO 10
      NW=NV+1
      VX=B(K*NW)
      DO 9 IN=1,NV
      IEX=IN+NW*(K-1)
      VX=VX+B(IEX)*X(IN)
    9 CONTINUE
      VALUE=VX
      RETURN
C     CALCULATION OF THE FIRST DERIVATIVE OF THE K-CONSTRAINT
   10 IEX=J+(NV+1)*(K-1)
      VALUE=B(IEX)
      RETURN
C     CALCULATION OF THE SECOND DERIVATIVE OF THE K CONSTRAINT
   11 VALUE=0.0
      RETURN
      END
```

```
      SUBROUTINE ERROR(Q,QH,STAT,NING,RES,IW)
C     IT CALCULATE THE STATISTICS OF THE RESIDUALS
C     ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
C
      DIMENSION Q(488),QH(488),RES(488),STAT(12)
      DOUBLE PRECISION AMEAN,SD,XX,XY,XZ,PM,ENM,PS,PS1,TV,S,SS
      COMMON /A1/S(2),SS(2),ND
      DATA AMEAN,SD,PM,ENM,XZ,PS,PS1,QM,QHM/9*0.0/
      NN=NING+1
      DO 3 I=1,ND
      XX=QH(I)
      XY=Q(I)
      XX=XX-XY
      RES(I)=XX
      AMEAN=AMEAN+XX
      SD=SD+XX*XX
      IF(XY.LE.QM) GO TO 1
      KPH=I
      QM=XY
    1 IF(PM.LT.XX) PM=XX
      IF(ENM.GT.XX) ENM=XX
      IF(XX*XZ.LT.0) GO TO 2
      PS1=PS1+XX
      GO TO 3
    2 PS=PS+PS1*PS1
      PS1=XX
    3 XZ=XX
```

```
      NPH=SQRT(FLOAT(ND))/2.
      LU=MINO(ND,KPH+NPH)
      LL=MAX0(1,KPH-NPH)
      DO 4 I=LL,LU
      IF(QH(I).LE.QHM)GO TO 4
      QHM=QH(I)
      K=I
    4 CONTINUE
      IF(K.EQ.LL.OR.K.EQ.LU) LL=-2
      XX=DFLOAT(ND)
      TV=SS(NN)-S(NN)*S(NN)/XX
      STAT(1)=AMEAN/XX
      STAT(2)=DSQRT((SD-(AMEAN*AMEAN)/XX)/(XX-1.0))
      IF(IW-1) 7,5,7
    5 CALL TEST(RES,ND,ND,30)
    6 CALL OUTPT1(RES,3)
    7 STAT(3)=(TV-(SD-AMEAN*AMEAN/XX))/TV
      STAT(4)=PS/SD
      STAT(5)=PM
      STAT(6)=ENM
      STAT(7)=(QHM-QM)/QM*100.0
      STAT(8)=K-KPH
      STAT(9)=LL+1
      STAT(10)=LU-1
      STAT(11)=KPH
      STAT(12)=QM
      RETURN
      END


      SUBROUTINE ALGO(AT,BT,X,U,V,S1,IBB,IQ,R,XSTAR,IB,JC,IR,NV,NC,
     *IZ,JZ,KAPUT)
C     MATHEMATICAL PROGRAMING
C     ****************************************************************
C
      DIMENSION R(12,37),XSTAR(12),AT(78),BT(130),X(12)
     *,U(12),V(12),S1(12),IBB(12),IQ(12),IR(12),IB(37),JC(37)
      DOUBLE PRECISION X,U,XSTAR,EPS,AT,BT,V
      COMMON /A2/ML,MI,MD
C     INITIAL PARAMETER VALUS SELECTION
      ITM=0
      NKL=2
      EPS=1.0E-5
      ID=1
      IZE=0
      EPZ=1.0E-25
      OBJ=-1.0E+36
      KPO=3
      NN=NV+NC
      LA=(2*NN)+1
      LAN=LA+NN
      NVP=NV+1
      KF=-1
      IF(NN-MD) 1,1,997
C     INITIAL BASIS DESCRIPTION
    1 NQ=0
      K=NN
      DO 7 N=1,NN
      IF(IQ(N)) 3,2,3
```

```
    2 IB(K)=N
      J=NN+K
      IB(J)=NN+N
      K=K-1
      GO TO 6
    3 NQ=NQ+1
      J=NN+NQ
      IF(IBB(N)) 5,4,5
    4 IB(NQ)=NN+N
      IB(J)=N
      GO TO 6
    5 IB(NQ)=N
      IB(J)=NN+N
    6 J=LA+N
    7 IB(J)=J
      IB(LA)=LA
      IF(ITM) 997,930,8
C     CHECK CONSISTENCY OF INITIAL VALUES
    8 J=0
      DO 11 N=1,NN
      IF(IB(N)-NV) 11,11,9
    9 IF(IB(N)-NN-NV) 10,10,11
   10 J=J+1
   11 CONTINUE
      IF(J-NV) 12,12,997
C     APPROXIMATE THE SADDLE FUNCTION BY A QUADRATIC
   12 KQF=0
   13 KQF=KQF+1
      KL=0
C     ESTABLISH COLUMN LOCATIONS AND VARIABLE VALUES
   14 DO 15 J=1,LAN
   15 JC(J)=J
      DO 16 J=1,NV
   16 XSTAR(J)=X(J)
      DO 17 K=1,NC
      J=NV+K
   17 XSTAR(J)=U(K)
C     FILL THE TABLEAU
      DO 26 I=1,NN
      DO 18 J=NVP,LAN
   18 R(I,J)=0.0
      J=NN+I
      K=LA+I
      R(I,J)=1.0
      R(I,K)=1.0
      IF(I-NV) 19,19,25
   19 DO 22 J=1,I
      A=VALUE(X,I,J,IZE,AT,BT,NV)
      DO 21 K=1,NC
      IF(U(K)) 20,21,20
   20 A=A+U(K)*VALUE(X,I,J,K,AT,BT,NV)
   21 CONTINUF
      R(I,J)=A
   22 R(J,I)=A
      R(I,LA)=-VALUE(X,IZE,I,IZE,AT,BT,NV)
      K=NV
   23 K=K+1
```

```
      IF(K-NN) 24,24,26
   24 R(I,K)=VALUE(X,IZE,I,K-NV,AT,BT,NV)
      R(K,I)=-R(I,K)
      GO TO 23
   25 R(I,LA)=VALUE(X,IZE,IZE,I-NV,AT,BT,NV)
   26 IBB(I)=0
      DO 28 N=1,NN
      A=R(N,LA)
      DO 27 J=1,NV
   27 A=A+X(J)*R(N,J)
   28 R(N,LA)=A
C     INVERT THE MATRIX OF BASIC COLUMNS
      NP=0
   30 NP=NP+1
      IF(NP-NN) 31,31,39
   31 JP=IB(NP)
C     FIND MAXIMAL PIVOT
   32 A=0.0
      DO 35 I=1,NN
      IF(IBB(I)) 997,33,35
   33 AA=ABS(R(I,JP))
      IF(AA-A) 35,34,34
   34 A=AA
      IP=I
   35 CONTINUE
      IF(A-EPZ) 960,960,36
   36 IR(NP)=IP
      IBB(IP)=1
C     EXECUTE PIVOTING OPERATION
   37 KPI=1
   38 GO TO 900
C     OPTIMIZE THE QUADRATIC PROGRAM
   39 IF(NQ)997,72,40
C     CKECK FOR OBTIMALITY
   40 AP=0.0
      AD=0.0
      DO 46 N=1,NQ
      I=IR(N)
      AA=R(I,LA)
      IF(IB(N)-NV) 42,42,41
   41 IF(IB(N)-NN-NV) 44,44,42
   42 IF(AA-AP) 43,46,46
   43 AP=AA
      NFP=N
      GO TO 46
   44 IF(AA-AD) 45,46,46
   45 AD=AA
      NFD=N
   46 CONTINUE
C     CHECK PRIMAL FEASIBILITY
   47 IF(AP) 50,48,997
   48 IF(AD) 49,72,997
   49 NFP=NFD
   50 NPC=NN+NFP
      IRFP=IR(NFP)
   51 IBFP=IB(NFP)
C     LOCAL PIVOT ROW
   52 LP=IB(NPC)
      JP=JC(LP)
      IPN=NFP
```

```
  912 IF(ABS(R(IRFP,JP))-EPZ) 55,55,913
  913 CONTINUE
      AA=R(IRFP,LA)/R(IRFP,JP)
      IF(AA) 53,55,56
   53 IF(R(IRFP,JP)-EPZ) 55,55,54
C     PROBLEM NOT CONCAVE
   54 KE=5*KE
      WRITE(2,9003)
 9003 FORMAT(/10X,21H PROBLEM NOT CONCAVE /)
   55 AA=1.0E+36
      IPN=0
   56 DO 62 N=1,NQ
      I=IR(N)
      A=R(I,LA)
      IF(A) 62,57,59
   57 IF(R(I,JP)) 62,62,58
   58 R(I,LA)=EPZ+1.0E-25
      GO TO 52
   59 IF(R(I,JP)-EPZ) 62,62,60
   60 A=A/R(I,JP)
      IF(A-AA) 61,61,62
   61 AA=A
      IPN=N
   62 CONTINUE
      IF(IPN) 997,940,67
C     UNBOUNDED SOLUTION
   63 KE=7*KE
      IF(ITM) 997,997,64
   64 DO 65 K=1,NC
   65 U(K)=1.0+1.10*U(K)
      IB(NFP)=LP
      IB(NPC)=IBFP
   66 GO TO 98
   67 IP=IR(IPN)
      KPI=2
      GO TO 900
   68 KP=IB(IPN)
      JC(LP)=JC(KP)
      JC(KP)=JP
      IB(NFP)=LP
      IF(IPN-NFP) 69,70,69
   69 IPPN=NN+IPN
      IB(NPC)=IB(IPPN)
      IB(IPPN)=KP
      IB(IPN)=IBFP
      IR(NFP)=IP
      IR(IPN)=IRFP
      GO TO 52
   70 IB(NPC)=IBFP
   71 GO TO 40
   72 KVA=1
      IF(ITM) 997,73,920
   73 KVA=2
   74 GO TO 920
   75 JP=IB(1)
```

```
      JP=JC(JP)
      JPK=IB(2)
      JPK=JC(JPK)
      JPKK=IB(3)
      JPKK=JC(JPKK)
      GO TO (2083,2084),ID
 2084 WRITE(2,2085) JP,JPK,JPKK,KVA,KL,KQF,NQ
 2085 FORMAT(10X,'JP,...,NQ',/,10X,7(I5,5X))
 2083 KEY=0
   76 KL=KL+1
      IF(KL-KQF*KPO) 761,761,94
  761 CONTINUE
      Y,YY=0.0
C     CALCULATE THE R.H.S OF THE EQUATION
      DO 79 J=1,NV
      A=VALUE(X,IZE,J,IZE,AT,BT,NV)
      DO 78 K=1,NC
      IF(U(K)) 77,78,77
   77 A=A+U(K)*VALUE(X,IZE,J,K,AT,BT,NV)
   78 CONTINUE
   79 R(J,JP)=A+V(J)
   80 DO 81 K=1,NC
      J=NV+K
   81 R(J,JP)=-VALUE(X,IZE,IZE,K,AT,BT,NV)+S1(K)
C     CKECK FOR CONVERGENCE
      KP=0
      DEL=0.0
   83 DO 90 K=1,NN
      N=IR(K)
      A=0.0
   84 DO 85 I=1,NN
      J=LA+I
   85 A=A+R(I,JP)*R(N,J)
      R(N,JPK)=A
      IF(ABS(A)-(1.0E-25)) 87,87,851
  851 CONTINUE
      YY=YY+A*A
      Y=Y+A*R(N,JPKK)
      AA=ABS(A/R(N,LA))
      IF(AA-DEL) 87,87,86
   86 DEL=AA
   87 IF(K-NQ) 88,88,90
   88 IF(R(N,LA)-A+EPZ) 89,90,90
   89 KP=K
   90 R(N,LA)=R(N,LA)-A
      IF(KEY) 892,892,890
  890 IF(Y) 894,892,892
  892 YYY=YY
      KEY=KEY+1
      DO 893 N=1,NN
  893 R(N,JPKK)=R(N,JPK)
      GO TO 899
  894 KEY=0
      KL=KL-1
  895 TH=-Y/(YYY-Y)
  896 KP=0
      DO 898 N=1,NN
      A=R(N,LA)+R(N,JPK)+TH*R(N,JPKK)
      IF(A+EPZ) 897,897,898
  897 KP=N
  898 R(N,LA)=A
  899 CONTINUE
      IF(KP) 997,91,40
```

```
   91 GO TO (1091,2091),ID
 2091 WRITE(2,2092) A,AA,Y,YY,DEL,TH,KEY,KP
 2092 FORMAT(//,10X,'A,....,KP',/,10X,6(E15.3,5X),2(I5,5X))
 1091 GO TO (9001,9002),NKL
 9001 KVA=4
      GO TO 920
 9002 KVA=3
      GO TO 920
   92 IF(DEL-EPS) 94,94,93
   93 IF(KL+3-KQF*KPO) 76,76,98
   94 DO 95 J=1,NV
      IF(DABS(XSTAR(J)-X(J))-EPS*DABS(XSTAR(J))) 95,95,98
   95 CONTINUE
      DO 96 K=1,NC
      J=NV+K
      IF(DABS(XSTAR(J)-U(K))-EPS*DABS(XSTAR(J))) 96,96,98
   96 CONTINUE
      ITM=KQF
   97 KAPUT=-KE-1
      GO TO 950
   98 IF(KQF-ITM) 13,996,996
  900 A=R(IP,JP)
      IF(ABS(A)-EPZ) 901,901,906
  901 KE=3*KE
      IF(ITM) 997,997,902
  902 IF(IPN-NFP) 903,904,903
  903 IB(NFP)=LP
      IB(NPC)=IBFP
  904 DO 905 J=1,NV
  905 X(J)=1.0+1.10*X(J)
      GO TO 98
  906 DO 907 I=1,NN
  907 R(I,JP)=-R(I,JP)/A
      R(IP,JP)=1.0/A
      DO 911 K=NP,LAN
      J=IB(K)
      J=JC(J)
      IF(J-JP) 908,911,909
  908 AA=R(IP,J)
      IF(AA) 909,911,909
  909 DO 910 I=1,NN
  910 R(I,J)=R(I,J)+AA*R(I,JP)
      R(IP,J)=AA/A
  911 CONTINUE
      GO TO (30,68),KPI
C     IDENTIFICATION OF VARIABLE VALUES
  920 DO 921 J=1,NV
      X(J)=0.0
  921 V(J)=0.0
      DO 922 K=1,NC
      U(K)=0.0
  922 S1(K)=0.0
      DO 929 N=1,NN
      I=IR(N)
      J=IB(N)
      IF(J-NN) 923,923,926
```

```
  923 IPB(J)=1
      IF(J-NV) 924,924,925
  924 X(J)=R(I,LA)
      GO TO 929
  925 J=J-NV
      U(J)=R(I,LA)
      GO TO 929
  926 J=J-NN
      IBB(J)=1
      IF(J-NV) 927,927,928
  927 V(J)=R(I,LA)
      GO TO 929
  928 J=J-NV
      S1(J)=R(I,LA)
  929 CONTINUE
      GO TO (73,97,970,92,999),KVA
  930 LAN=LA
      GO TO 8
  940 DO 942 N=1,NQ
      J=NN+N
      J=IB(J)
      J=JC(J)
      A=R(IRFP,J)
      IF(A-AA) 941,942,942
  941 AA=A
      IPN=N
  942 CONTINUE
      IF(AA+EPZ) 67,63,63
C     STORE INVERSE OF BASIC MATRIX
  950 DO 953 N=1,NN
      II=IR(N)
      I=IB(N)
      IF(I-NN) 952,952,951
  951 I=I-NN
  952 DO 953 J=1,NN
      JJ=LA+J
  953 R(I,J)=R(II,JJ)
      RETURN
  960 IF(JP-NN) 961,961,962
  961 IB(NP)=JP+NN
      GO TO 31
  962 IB(NP)=JP-NN
      GO TO 31
C     CHECK THE OBJECTIVE VALUE OBJ
  970 AOBJ=OBJ
      OBJ=VALUE(X,IZE,IZE,IZE,AT,BT,NV)
      IF(ABS(AOBJ-OBJ)-EPS*0.1*ABS(AOBJ)) 94,94,92
C     ERROR EXIT
  996 KE=2*KE
  997 KVA=5
      GO TO 929
  999 KAPUT=KE
      RETURN
      END
```

```
      SUBROUTINE WRITE2(STAT,OUT,QH,V,NV)
      DIMENSION STAT(12),V(12),QH(488),OUT(488)
      DOUBLE PRECISION V
      WRITE(2,410)
      WRITE(2,411) STAT(1)
      WRITE(2,412) STAT(2)
      WRITE(2,413) STAT(3)
      WRITE(2,415) STAT(4)
      WRITE(2,416) STAT(5)
      WRITE(2,417) STAT(6)
      WRITE(2,418) STAT(7)
      WRITE(2,420) STAT(9)
      WRITE(2,421) STAT(10)
      WRITE(2,422) STAT(12)
      WRITE(2,300)
      WRITE(2,400) (V(I),I=1,NV)
      WRITE(2,423)
      CALL OUTPT1(OUT,1)
      CALL OUTPT1(QH,2)
      RETURN
  300 FORMAT(//,10X,'VALUES OF THE IMPULSE RESPONSE FUNCTION')
  400 FORMAT(6(4X,F8.4))
  410 FORMAT(//,10X,'STATISTICS OF THE RESIDUALS')
  411 FORMAT(/,10X,'MEAN OF THE RESIDUALS',14X,'=',F14.6)
  412 FORMAT(10X,'STANDARD DEVIATION OF RESIDUALS',4X,'=',F14.6)
  413 FORMAT(10X,'DETERMINATION COEFFICIENT',10X,'=',F14.6)
  415 FORMAT(10X,'COEFFICIENT OF PERSISTANCE',9X,'=',F14.6)
  416 FORMAT(10X,'MAXIMUM POSITIVE ERROR',13X,'=',F14.6)
  417 FORMAT(10X,'MAXIMUM NEGATIVE ERROR',13X,'=',F14.6)
  418 FORMAT(10X,'PERCENTAGE ERROR BETWEEN PEAKS',5X,'=',F14.6)
  420 FORMAT(10X,'INDEX OF LOWER LIMIT OF SEARCH',5X,'=',F14.6)
  421 FORMAT(10X,'INDEX OF UPPER LIMIT OF SEARCH',5X,'=',F14.6)
  422 FORMAT(10X,'MAXIMUM OBSERVED RUNOFF',12X,'=',F14.6)
  423 FORMAT(/,10X,'DAILY RECORDED AND ESTIMATED DISCHARGES IN MM FOR WA
     *KI CATCHMENT',/,10X,68(1H=),/)
      END


      SUBROUTINE TEST(A,N,N1,ML)
      DIMENSION A(488)
      DOUBLE PRECISION XN,SX,SXX,SDX,X1,X2,XX,QT
      DATA SX,SDX,QT/3*0.0/
      XN=DFLOAT(N)
      ML=ML+1
      DO 1 I=1,N
      XX=A(I)
      SX=SX+XX
      SDX=SDX+XX*XX
    1 CONTINUE
      SX=SX/XN
      SDX=DSQRT((SDX-SX*SX*XN)/(XN-1.D0))
      DO 3 J=1,ML
      SXX=0.0
```

```
      DO 2 I=J,N
      X1=A(I-J+1)-SX
      X2=A(I)-SX
      SXX=SXX+X1*X2
    2 CONTINUE
      XN1=XN-DFLOAT(J)
      SXX=SXX/(XN1*SDX*SDX)
      IF(J.GT.1) QT=QT+SXX*SXX
    3 CONTINUE
      QT=QT+DFLOAT(N1)
      SDX=SDX*SDX
      WRITE(2,100) SX,SDX
      WRITE(2,200) QT
      RETURN
  100 FORMAT(/,10X,'INNOVATION MEAN=',F10.6,/,10X,'INNOVATION VARIA'
     *,'NCE=',F10.6,/)
  200 FORMAT(10X,'Q-TEST=',F10.6)
      END
```

```
      SUBROUTINE OUTPT1(OUT,LL)
C     OBJECT:
C     IT WRITES THE OUTPUT RESULTS.
C     ***********************************************************************
C
      DIMENSION OUT(488)
  424 FORMAT(8(I8,F10.6))
 1000 FORMAT(/,5X,'RECORDED DISCHARGE:-')
 1010 FORMAT(/,5X,'ESTIMATED DISCHARGE:-')
 1020 FORMAT(/,5X,'THE RESIDUAL:-')
C     -----------------------------------------------------------------------
C
      GO TO (10,15,25),LL
   10 WRITE(2,1000)
      GO TO 30
   15 WRITE(2,1010)
      GO TO 30
   25 WRITE(2,1020)
   30 DO 20 I=1,61
      I1=I+61
      I2=I1+61
      I3=I2+61
      I4=I3+61
      I5=I4+61
      I6=I5+61
      I7=I6+61
      WRITE(2,424) I,OUT(I),I1,OUT(I1),I2,OUT(I2),I3,OUT(I3),I4,OUT(I4),
     *I5,OUT(I5),I6,OUT(I6),I7,OUT(I7)
   20 CONTINUE
      RETURN
      END
```

```
      SUBROUTINE MATR(VINP,OUT,NI,TEMP,FL,A,B,NV,NC)
C     AUTOCORRELATION AND CROSS CORRELATION PROGRAM.
C     ------------------------------------------------------------------
C
      DIMENSION VINP(488),OUT(488),A(78),B(130)
      INTEGER TEMP,FL
      DOUBLE PRECISION A,B,S,SD,RO
      COMMON /A1/S(2),SD(2),ND
      IN1,IN2=0
      IX=1
      L1=TEMP
      IY=IX
      L2=L1
      IF(FL.EQ.0) GO TO 1
      RO=SD(1)
      CALL OLS(IX,IY,L1,L2,IN1,IN2,VINP,VINP,RO,A,B,NV,NC)
    1 IY=IY+L2
      L2=IX
      RO=DSQRT(SD(1)*SD(2))
      CALL OLS(IX,IY,L1,L2,IN1,IN2,VINP,OUT,RO,A,B,NV,NC)
      RETURN
      END


      SUBROUTINE SMOOTH(V,NV,X)
C     IT SMOOTH THE OSCILATORY KERNAL FUNCTION ACCORDING TO HAMING
C     ALGORITHM.
C     ******************************************************************
C
      DIMENSION V(12),X(12)
      DOUBLE PRECISION X,V
      NVV=NV-1
      DO 2 I=1,NV
      IF(I.EQ.1) X(I)=0.54*V(I)+0.46*V(I+1)
      IF(I.GT.1.AND.I.LE.NVV) X(I)=0.23*V(I-1)+0.54*V(I)+0.23*V(I+1)
      IF(I.EQ.NV) X(I)=0.54*V(I)+0.46*V(I-1)
    2 CONTINUE
      RETURN
      END


      SUBROUTINE SMOS(V,NV)
C     IT SMOOTHS AND WRITES THE POWER SPECTRUM BY USING THE HAMING
C     WINDOW ALGORITHM.
C     //////////////////////////////////////////////////////////////////
      DIMENSION V(50),X(50)
      NVV=NV-1
      DO 10 I=1,NV
      IF(I.EQ.1) X(I)=0.54*V(I)+0.46*V(I+1)
      IF(I.GT.1.AND.I.LE.NVV) X(I)=0.23*V(I-1)+0.54*V(I)+0.23*V(I+1)
      IF(I.EQ.NV) X(I)=0.54*V(I)+0.46*V(I-1)
   10 CONTINUE
      WRITE(2,2060) (X(I),I=1,NV)
      RETURN
 2060 FORMAT(//,10X,'X-VALUES:',/,10(5X,F10.6),/)
      END
```

```
      SUBROUTINE INPUT1(VINP,OUT,N )
      DIMENSION VINP(488),OUT(488)
      DOUBLE PRECISION SD,SI,SDX,Y,S
      COMMON /A1/S(2),SD(2),ND
      DO 20 I=1,N
      SI=0.0
      SDX=0.0
      KA=(I-1)*ND
      DO 10 J=1,ND
      K=J+KA
      Y=VINP(K)
      SI=SI+Y
      SDX=SDX+Y*Y
   10 CONTINUE
      S(I)=SI
      SD(I)=SDX
   20 CONTINUE
      K=N+1
      SI=0.0
      SDX=0.0
      DO 30 I=1,ND
      Y=OUT(I)
      SI=SI+Y
   30 SDX=SDX+Y*Y
      S(K)=SI
      SD(K)=SDX
      RETURN
      END
```

```
      SUBROUTINE CONV(X,Y,Z,NX,NY,IS)
C     IT CALCULATES THE CONVOLUATION OF VECTOR Y WITH X
C     ***************************************************************
C
      DIMENSION X(12),Y(488),Z(488)
      DOUBLE PRECISION X,YY,ZZ
      JM=1
      IF(IS.LT.0) JM=2
      DO 3 J=JM,NY
      ZZ=0.0
      JX=J
      IF(IS.LT.0) JX=J-1
      IU=MINO(JX,NX)
      IF(IU-1) 3,1,1
    1 DO 2 I=1,IU
      IX=I-1
      IF(IS.LT.0) IX=I
      YY=Y(J-IX)
    2 ZZ=ZZ+X(I)*YY
      YY=Z(J)
      Z(J)=YY+ZZ
      IF(IS.LT.0) Y(J)=Z(J)
    3 CONTINUE
      RETURN
      END
```

```
      SUBROUTINE CCOR(X,Y,K,N)
C     CROSS CORRELATION COEFFICIENT PROGRAM
C     X,Y:INPUT ARRAYS N,N
C     K  :NO.OF CORRELATION COEFFICIENT REQUIRED
C
      DIMENSION X(488),Y(488)
   12 FORMAT(10X,'R',I2,'=',F6.4)
   20 FORMAT(//,10X,'CROSS CORR. COEF.')
      WRITE(2,20)
      DO 4 J=1,K
      JJ=J-1
      S=0.0
      S1=0.0
      S2=0.0
      S3=0.0
      S4=0.0
      L=N-JJ
      DO 2 I=1,L
      S=S+X(I)*Y(I+JJ)
      S1=S1+X(I)
      S2=S2+X(I)*X(I)
    2 CONTINUE
      I=JJ+1
      DO 3 M=I,N
      S3=S3+Y(M)
      S4=S4+Y(M)*Y(M)
    3 CONTINUE
      R=(S-S1*S3/L)/SQRT((S2-S1*S1/L)*(S4-S3*S3/L))
      WRITE(2,12) JJ,R
    4 CONTINUE
      RETURN
      END
```

```
      SUBROUTINE QHAT(P,Q1,Q2,X,TE,N,KS,NS)
      DIMENSION X(12),P(488),Q1(488),Q2(488)
      INTEGER TE
      DOUBLE PRECISION X,S,SQ,RO
      COMMON/A1/S(2),SQ(2),ND
      K=1
      IS=NS
      NN=N+1
      DO 10 I=1,ND
   10 Q2(I)=0.0
      K1=K+TE-1
      IF(KS.EQ.1) GO TO 25
      RO=DSQRT(SQ(NN)/SQ(N))
      DO 20 J=K,K1
   20 X(J)=X(J)*RO
   25 IF(IS.EQ.1) GO TO 30
      GO TO 40
   30 CALL CONV(X,P,Q2,TE,ND,IS)
   40 RETURN
      END
```

```
      SUBROUTINE KOLM2(X,Y,N,M,Z,PROB,DN)
C     TESTS THE DIFFERENCE BETWEEN TWO SAMPLE DISTRIBUTION FUNCTIONS
C     USING THE KOLMOGROV-SMIRNOV TEST.
C     X:(N*1)INPUT VECTOR.
C     Y:(M*1)INPUT VECTOR.
C     PROB:THE PROBABILITY OF THE STATISTIC BEING.GE.Z.
C     Z:OUTPUT VARIABLE CONTAINING THE GREATEST VALUE WITH RESPECT
C       TO THE SPECTRUM OF X AND Y.
C     ****************************************************************
C
      DIMENSION Y(488),X(488)
C     STORE X INTO ASCENDING ORDER .
      DO 5 I=2,N
      IF(X(I)-X(I-1))1,5,5
    1 TEMP=X(I)
      IM=I-1
      DO 3 J=1,IM
      L=I-J
      IF(TEMP-X(L)) 2,4,4
    2 X(L+1)=X(L)
    3 CONTINUE
      X(1)=TEMP
      GO TO 5
    4 X(L+1)=TEMP
    5 CONTINUE
C     SORT Y INTO ASCENDING ORDER .
      DO 10 I=2,M
      IF(Y(I)-Y(I-1)) 6,10,10
    6 TEMP=Y(I)
      IM=I-1
      DO 8 J=1,IM
      L=I-J
      IF(TEMP-Y(L))7,9,9
    7 Y(L+1)=Y(L)
    8 CONTINUE
      Y(1)=TEMP
      GO TO 10
    9 Y(L+1)=TEMP
   10 CONTINUE
C     CALCULATE DN=ABS(FN-GM) OVER THE SPECTRUM OF X AND Y .
      XN=FLOAT(N)
      XN1=1.0/XN
      YM=FLOAT(M)
      XM1=1.0/XM
      I,J,K,L=0
      DN=0.0
   11 IF(X(I+1)-Y(J+1))12,13,18
   12 K=1
      GO TO 14
   13 K=0
   14 I=I+1
      IF(I-N) 15,21,21
   15 IF(X(I+1)-X(I))14,14,16
   16 IF(K) 17,18,17
C     CALCULATE THE MAXIMUM DIFFERENCE,DN .
   17 DN=AMAX1(DN,ABS(FLOAT(I)*XN1-FLOAT(J)*XM1))
      IF(L) 22,11,22
   18 J=J+1
      IF(J-M) 19,20,20
```

```
   19 IF(Y(J+1)-Y(J)) 18,18,17
   20 L=1
      GO TO 17
   21 L=1
      GO TO 16
C     CALCULATE THE STATISTIC Z .
   22 Z=DN*SQRT((XN*XM)/(XN+XM))
C     CALCULATE THE PROBABILITY ASSOCIATED WITH Z .
      CALL SMIRN(Z,PROB)
      PROB=1.0-PROB
      RETURN
      END


      SUBROUTINE KOLM1(X,N,IER,IFCOD,U,S,PROB,Z)
C     TESTS THE DIFFERENCE BETWEEN THE EMPIRICAL AND THEORITICAL
C     DISTRIBUTIONS USING THE KOLMOGOROV SMIRNOV TEST.
C     X    :INPUT VECTOR OF N INDEPENDANT OBSERVATIONS.
C     PROB :THE PROBABILITY OF STATISTIC BEING .GE. TO Z.
C     IFCOD:CODE OF THE THEORITICAL DISTRIBUTION FUNCTION.
C     U,S  :STATISTICS OF VECTOR X ACCORDING TO IFCODE.
C     IER  :ERROR INDEX VALUE.
C     ****************************************************************
C
      DIMENSION X(488)
C     NON DECREASING ORDER OF X(I) .
      IER=0
      DO 5 I=2,N
      IF(X(I)-X(I-1))1,5,5
    1 TEMP=X(I)
      IM=I-1
      DO 3 J=1,IM
      L=I-J
      IF(TEMP-X(L)) 2,4,4
    2 X(L+1)=X(L)
    3 CONTINUE
      X(1)=TEMP
      GO TO 5
    4 X(L+1)=TEMP
    5 CONTINUE
C     COMPUTES MAXIMUM DEVIATION DN .
      NM1=N-1
      XN=N
      DN=0.0
      FS=0.0
      IL=1
    6 DO 7 I=IL,NM1
      J=I
      IF(X(J)-X(J+1)) 9,7,9
    7 CONTINUE
    8 J=N
    9 IL=J+1
      FI=FS
      FS=FLOAT(J)/XN
      IF(IFCOD-2) 10,13,17
```

```
10 IF(S) 11,11,12
11 IER=1
   GO TO 29
12 Z=(X(J)-U)/S
   CALL NDTR(Z,Y,D)
   GO TO 27
13 IF(S) 11,11,14
14 Z=(X(J)-U)/S+1.0
   IF(Z) 15,15,16
15 Y=0.0
   GO TO 27
16 Y=1.0-EXP(-Z)
   GO TO 27
17 IF(IFCOD-4) 18,20,26
18 IF(S) 19,11,19
19 Y=ATAN((X(J)-U)/S)*0.3183099+0.5
   GO TO 27
20 IF(S-U) 11,11,21
21 IF(X(J)-U) 22,22,23
22 Y=0.0
   GO TO 27
23 IF(X(J)-S) 25,25,24
24 Y=1.0
   GO TO 27
25 Y=(X(J)-U)/(S-U)
   GO TO 27
26 IER=1
   GO TO 29
27 EI=ABS(Y-FI)
   ES=ABS(Y-FS)
   DN1=AMAX1(ES,EI)
   DN=AMAX1(DN1,DN)
   IF(L-N) 6,8,28
28 Z=DN*SQRT(XN)
   CALL SMIRN(Z,PROB)
   PROB=1.0-PROB
29 RETURN
   END
```

APPENDIX C

THE SECOND KOLMOGROV_SMIRNOV TEST

APPENDIX C

THE SECOND KOLMOGROV-SMIRNOV TEST

The goodness of fit between the two histograms of observed and generated sequences may be checked by using the second Kolmogrov-Smirnov test.

Let F and G be the cumulative distribution functions of the generated and observed sequences respectively, N_1 and N_2 be the length of these two sequences. Let H_o be the hypothesis that both cumulative distribution functions were obtained from the same population series. Then, the test statistics d can be expressed as

$$d = \sqrt{\frac{N_1 N_2}{N_1+N_2}} \cdot \max_{-\infty<\delta<\infty} \left| F_{N_1}(\delta) - G_{N_2}(\delta) \right| \tag{C.1}$$

Decision Rule

The decision rule for accepting or rejecting the null hypothesis H_o is given by

$$d \begin{cases} \leq d_c & \rightarrow \quad \text{Accept } H_o \\ > d_c & \rightarrow \quad \text{Reject } H_o \end{cases} \tag{C.2}$$

where the threshold d_c may be expressed as

$$d_c = \begin{cases} 1.36 \text{ at } 95\% \text{ significant level} \\ 1.22 \text{ at } 90\% \text{ significant level.} \end{cases} \tag{C.3}$$

APPENDIX D

LIST OF THE DIGITAL COMPUTER PROGRAM FOR THE LINEAR STOCHASTIC DIFFERENCE EQUATION MODEL

```
      PROGRAM(RAIN)
      INPUT 1=CRD
      OUTPUT 2=LPO/160
      TRACE 0
      END
```

```
      MASTER RAO
C     ******************************M*********************************
C     THIS PROGRAM IDENTIFY THE NECESSARY PARAMETERS FOR RAO AND KASHAP
C     DAILY DATA MODEL.THESE PARAMETERS ARE THEN USED FOR THE PREDICTION
C     OF DAILY STREAM FLOW Y AT ANY INSTANT I.
C     DESCRIPTION OF PARAMETERS:
C     Y(I) :A SEQUENCE OF DAILY INPUT DATA THE REQUIRED LENGTH IS ND.
C     YE(I):A SEQUENCE OF DAILY ESTIMATED OUTPUT DATA(STREAMFLOW).
C     YR(I):A SEQUENCE OF DAILY RESIDUAL.
C     A :VECTOR OF UNKNOWN PARAMETERS THE NECESSARY DIMENSION IS L.
C     Z :VECTOR CONTAINS CERTAIN FUNCTIONS OF Y(I) AND YR(I).
C     S :(L*L) MATRIX.
C     B :WORK VECTOR OF DIMENSION L.
C     I1:TRANSFORMATION PARAMETER.
C     I2:ANOTHER TRANSFORMATION PARAMETER.
C     I3:CONSTANT EQUAL TO 1
C     I4:CONSTANT EQUAL TO 2
C     I5:CONSTANT EQUAL TO 3
C
      COMMON /A2/Z(6),Y(976),YE(976),YR(976),A(6),S(6,6)
      DIMENSION B1(6),B2(6),XSTAR(6,6),VOUT(976)
      COMMON /C1/AMEAN,STDEV,ASK
      COMMON /A1/L,ND
      DATA ML,ISIZE/50,5/
C     ================================================================
C
C     READING FORMAT
 1000 FORMAT(3I4)
 1010 FORMAT(8F0.0)
 1020 FORMAT(10I2)
C
C     MAIN PROGRAM OUTPUT FORMATS:-
 2000 FORMAT(//,10X,'VALUES OF PARAMETER VECTOR A:-')
 2010 FORMAT(6(6X,F10.6))
 2020 FORMAT(//,10X,'THE ADDITIVE MODEL IS USED FOR PREDICTION THE DAILY
     * DATA.')
 2030 FORMAT(//,10X,'THE MULTIPLICATIVE MODEL IS USED FOR PREDICTION THE
     * DAILY DATA.')
 2040 FORMAT(10X,'ONLY THE PERIOD FROM APRIL TO NOVEMBER IS CONSIDERED.'
     *,/,10X,'YEARS OF OBSERVATION ARE 1970,1971,1973 AND 1974.',/,10X,
     *'NAME OF THE CATCHEMENT:WAKI RIVER CATCHEMENT.')
C     ////////////////////////////////////////////////////////////////
```

```
C     //////////////////////////////////////////////////////////////////
C
      READ(1,1000) I1,I2,I3,I4,ND,L,IER,IAUT,ISC,IP,LAG,NYEAR,I5
      READ(1,1010) (Y(I),I=1,ND)
      IF(I1-3) 10,20,20
   10 WRITE(2,2020)
      GO TO 30
   20 WRITE(2,2030)
   30 WRITE(2,2040)
C
      CALL EQUO
      CALL OUTPT1(Y,I3)
      CALL PARA(Y,ND,AMEAN,STDEV,ASK)
      CALL TRANS(I1,Y,ND,AMEAN,STDEV)
      CALL PRINT1(IAUT,IER,IP,ISC,LAG)
      CALL ZGEN(0,IER,IAUT,ISC,IP)
      WRITE(2,2000)
      IF(LAG.EQ.0) GO TO 70
      JF=ND
   40 Y(JF)=Y(JF-LAG)
      IF(JF.EQ.(LAG+1)) GO TO 50
      JF=JF-1
      GO TO 40
   50 DO 60 I=1,JF
      Y(I)=0.0
   60 CONTINUE
   70 DO 100 I=1,ND
      IF(IER.NE.0) YR(I)=Y(I)
      CALL MARS(A,Z,I3,SC,XSTAR)
      CALL VARC(I3,I4)
      CALL MULT(I4,B1)
      SCC=Y(I)-SC
      DO 80 J=1,L
      A(J)=A(J)+SCC*B1(J)
      IF(IER.NE.0) YR(I)=YR(I)-A(J)*Z(J)
   80 CONTINUE
      WRITE(2,2010) (A(K),K=1,L)
      CALL ZGEN(I,IER,IAUT,ISC,IP)
  100 CONTINUE
      DO 110 I=1,ND
      CALL ZGEN(I,IER,IAUT,ISC,IP)
      YE(I)=0.0
      DO 110 J=1,L
      YE(I)=YE(I)+A(J)*Z(J)
  110 CONTINUE
      CALL ERROR
      CALL OUTPT1(YR,I5)
      CALL PARA(YE,ND,AMEAN,STDEV,ASK)
      CALL TRANS(I2,YE,ND,AMEAN,STDEV)
      CALL OUTPT1(YE,I4)
      CALL TEST(IAUT,ISC,ML,ISIZE)
      STOP
      END
```

```
      SUBROUTINE TEST(IA,IC,ML,ISIZE)
C     PURPOSE:
C     IT TESTS THE RESIDUAL VECTOR YR.
C     ******************************************************************
C
      COMMON /A1/L,ND
      COMMON /A2/Z(6),Y(976),YE(976),YR(976),A(6),SS(6,6)
      DOUBLE PRECISION COR(50),GAMA(50,50),CORO
      EQUIVALENCE (COR(1),GAMA(1,1))
      DATA E0,E1,E2,U,IFCOD,S/4*0.0,1,1.0/
C     ******************************************************************
C     OUTPUT FORMATS:
 1000 FORMAT(10X,3HE0=,F10.6,10X,3HE1=,F10.6,10X,3HE2=,F10.6,/)
 1010 FORMAT(//,10X,'TESTING OF THE RESIDUALS:')
 1020 FORMAT(10X,'TEST:2',/,10X,2HZ=,F10.6,/,10X,5HPROB=,F10.6,/,10X,
     *'MAXIMUM DIFFERENCE DN=',F10.6,/)
 1030 FORMAT(10X,'SECTION:2',/,10X,'KOLMOGROV SMIRNOV TEST.',/,10X,
     *'TEST:1',/,10X,2HZ=,F10.6,/,10X,5HPROB=,F10.6)
 1040 FORMAT(10X,'SECTION:1',/,10X,'MEANS OF THE RESIDUAL.')
 1050 FORMAT(10X,'SECTION 3',/,10X,'THE F-TEST.',/,10X,'VALUE=',F10.6,
     *10X,'LAG=',I3,/,10X,60(1H*),/)
 1060 FORMAT(10X,'MLL=',I2,10X,'I=',I2,10X,'GAMA(MLL,I)=',D26.20)
C     ==================================================================
C
      ISUM=IA+IC+1
      WRITE(2,1010)
      NN=ND-ISUM
      DO 10 I=ISUM,ND
      E0=E0+YR(I)/NN
      E1=E1+ABS(YR(I)/NN)
      E2=E2+(YR(I)**2)/NN
   10 CONTINUE
      WRITE(2,1040)
      WRITE(2,1000) E0,E1,E2
      CALL AUTO(YE,ND,ML,IFCOD,COR,CORO)
      CALL AUTO(YR,NN,ML,ISUM,COR,CORO)
      DO 30 I=ISIZE,ML,ISIZE
      CALL KOLM2(Y,YE,I,I,Z2,PROB2,DN)
      WRITE(2,1020) Z2,PROB2,DN
   30 CONTINUE
      CALL KOLM2(Y,YE,ND,ND,Z2,PROB2,DN)
      WRITE(2,1020) Z2,PROB2,DN
      CALL CCOR(Y,YR,ML,ND)
      RETURN
      END


      SUBROUTINE ERROR
C     IT COMPUTES THE RESIDUALS VECTOR YR .
C     ******************************************************************
C
      COMMON /A1/L,ND
      COMMON /A2/Z(6),Y(976),YE(976),YR(976),A(6),S(6,6)
      DO 10 I=1,ND
      YR(I)=Y(I)-YE(I)
   10 CONTINUE
      RETURN
      END
```

```
      SUBROUTINE ZGEN(IGEN,IER,IAUT,ISC,IP)
C     IT GENERATES THE Z VECTOR FOR THE GIVEN INSTANT
C     ************************************************************************
C
      COMMON /A2/Z(6),Y(976),YE(976),YR(976),A(6),S(6,6)
      Z(1)=1.0
C     GENERATE THE IAUT ORDER AUTOREGRESIVE TERMS.
      DO 10 I=1,IAUT
      INDEX=IGEN-I+1
      IF(INDEX.LE.0) Z(I+1)=0.0
      IF(INDEX.GT.0) Z(I+1)=Y(INDEX)
   10 CONTINUE
C     GENERATE THE SECOND ORDER ERROR TERM IF ANY.
      IF(IER) 40,40,20
   20 IER1=IAUT+2
      IER2=IAUT+3
      DO 30 I=IER1,IER2
      JNDEX=IGEN-I+IER1
      IF(JNDEX.LE.0) Z(I)=0.0
      IF(JNDEX.GT.0) Z(I)=YR(JNDEX)
   30 CONTINUE
   40 IF(ISC) 60,60,50
C     GENERATE SIN AND COS TERMS IF ANY.
   50 ISC1=IAUT+2
      ISC2=IAUT+3
      Z(ISC1)=SIN(44.0*IGEN/1708.0)
      Z(ISC2)=COS(44.0*IGEN/1708.0)
   60 IF(IP) 90,90,70
C     GENERATE PERIODIC TERMS IF ANY.
   70 IP1=IAUT+2
      Z(IP1)=0.0
      DO 80 I=1,7
      II=I-4
      Z(IP1)=(Z(IP1)+Y(IGEN-244+II))/7.0
   80 CONTINUE
   90 RETURN
      END
```

```
      SUBROUTINE EQU0
C     IT INITIALIZE BOTH VECTOR A AND MATRIX S.
C     ************************************************************************
C
      COMMON /A1/L,ND
      COMMON /A2/Z(6),Y(976),YE(976),YR(976),A(6),S(6,6)
      DO 10 I=1,L
      A(I)=0.0
      DO 10 J= ,L
      IF(I.EQ.J) S(I,J)=1.0
      IF(I.NE.J) S(I,J)=0.0
   10 CONTINUE
      RETURN
      END
```

```
      SUBROUTINE MULT(M,B)
C     PURPOSE:
C     PERFORMS MATRIX AND VECTOR MULTIPLICATION.
C     A:INPUT VECTOR OF DIMENSION L .
C     X:INPUT MATRIX OF DIMENSION (LXL).
C     B:OUTPUT VECTOR OF DIMENSION L .
C     M:PERFORMANCE INDEX .
C     IF M=1:B=A*X .
C     IF M=2:B=X*A .
C     ************************************************************************
C
      COMMON/A2/Z(6),Y(976),YE(976),YR(976),A(6),X(6,6)
      COMMON /A1/L,ND
      DIMENSION B(6)
      GO TO (10,30),M
   10 DO 20 I=1,L
      B(I)=0.0
      DO 20 J=1,L
      B(I)=B(I)+Z(J)*X(J,I)
   20 CONTINUE
      GO TO 50
   30 DO 40 I=1,L
      B(I)=0.0
      DO 40 J=1,L
      B(I)=B(I)+X(I,J)*Z(J)
   40 CONTINUE
   50 RETURN
      END
```

```
      SUBROUTINE MARS(B1,B2,L,SCATB,ABT)
C     PURPOSE:
C     IT GIVES THE PRODUCT OF MULTIPLICATION OF A TRANSPOSED VECTOR B1
C     AND THE OTHER VECTOR B2 WHICH A SCALAR SCATB FOR L=1.
C     IT ALSO GIVES THE PRODUCT OF MUTIPLICATION OF VECTOR B1 AND A
C     TRANSPOSED VECTOR B2 WHICH A MATRIX ABT FOR L=2.
C     ************************************************************************
C
      DIMENSION B1(6),B2(6),ABT(6,6)
      COMMON /A1/LL,ND
      GO TO (10,30),L
   10 SCATB=0.0
      DO 20 I=1,LL
      SCATB=SCATB+B1(I)*B2(I)
   20 CONTINUE
      RETURN
C
   30 DO 40 I=1,LL
      DO 40 J=1,LL
      ABT(I,J)=B1(I)*B2(J)
   40 CONTINUE
      RETURN
      END
```

```
      SUBROUTINE AUTO(A,N,L,ISUM,R1,CO)
      DIMENSION A(976),R1(50),R2(50)
      DOUBLE PRECISION R1,CO,SUM,AVER
      PHI=22.0/7.0
      AVER=0.0
      IF(N-L) 50,50,60
   50 R1(1)=0.0
      GO TO 150
   60 WRITE(2,200)
  100 DO 110 I=ISUM,N
  110 AVER=AVER+A(I)
      FN=N
      AVER=AVER/FN
C     CALCULATE AUTOCOVARIENCES .
      DO 130 J=1,L
      NJ=N-J+1
      SUM=0.0
      DO 120 I=ISUM,NJ
      IJ=I+J-1
  120 SUM=SUM+(A(I)-AVER)*(A(IJ)-AVER)
      FNJ=NJ
      R1(J)=SUM/FNJ
      R2(J)=R1(J)/R1(1)
      K=J-1
      WRITE(2,300) K,R1(J),R2(J)
  130 CONTINUE
      CO=R1(1)
      CALL POWER(L,PHI,R2)
  150 RETURN
  200 FORMAT(//,10X,61(1H*),/,20X,'K',9X,'AUTO(K)',9X,'AUTO(K)/AUTO(0)',
     */,10X,61(1H*))
  300 FORMAT(20X,I2,2(7X,F10.6))
      END


      SUBROUTINE VARC(I3,I4)
C     PURPOSE:
C     THIS SUBROUTINE UPDATES THE S MATRIX.
C     A:VECTOR OF UNKNOWN PARAMETERS.
C     Z:VECTOR OF FUNCTIONS OF THE INPUT STREAMLOW.
C     S:UPDATED S MATRIX.
C     L:NUMBER OF UNKNOWN PARAMETERS.
C     ND:LENGTH OF INPUT DATA.
C     ******************************************************************
C
      DIMENSION B1(6),B2(6),XSTAR(6,6)
      COMMON /A1/L,ND
      COMMON /A2/Z(6),Y(976),YE(976),YR(976),A(6),S(6,6)
      CALL MULT(I4,B1)
      CALL MULT(I3,B2)
      CALL MARS(B1,B2,I4,SC1,XSTAR)
      CALL MULT(I3,B1)
      CALL MARS(B1,Z,I3,SC,XSTAR)
      DO 10 I=1,L
      DO 10 J=1,L
      S(I,J)=S(I,J)-XSTAR(I,J)/(1.0+SC)
   10 CONTINUE
      RETURN
      END
```

```
      SUBROUTINE OUTPT1(OUT,LL)
C     OBJECT:
C     IT WRITES THE OUTPUT RESULTS.
C     ************************************************************************
C
      DIMENSION OUT(976)
  424 FORMAT(8(I8,F10.6))
 1000 FORMAT(/,5X,'RECORDED DISCHARGE:-')
 1010 FORMAT(/,5X,'ESTIMATED DISCHARGE:-')
 1020 FORMAT(/,5X,'THE RESIDUAL:-')
C     ------------------------------------------------------------------------
C
      GO TO (10,15,25),LL
   10 WRITE(2,1000)
      GO TO 30
   15 WRITE(2,1010)
      GO TO 30
   25 WRITE(2,1020)
   30 DO 20 I=1,61
      I1=I+61
      I2=I1+61
      I3=I2+61
      I4=I3+61
      I5=I4+61
      I6=I5+61
      I7=I6+61
      WRITE(2,424) I,OUT(I),I1,OUT(I1),I2,OUT(I2),I3,OUT(I3),I4,OUT(I4),
     *I5,OUT(I5),I6,OUT(I6),I7,OUT(I7)
   20 CONTINUE
      RETURN
      END

      SUBROUTINE PRINT1(IA,IE,IP,IS,LA)
C     PURPOSE:
C     IT WRITES THE INPUTS.
C     ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
C
      COMMON /A1/L,ND
      COMMON /A2/Z(6),Y(976),YE(976),YR(976),A(6),S(6,6)
      COMMON /C1/AMEAN,STDEV,ASK
C
C     THE NECESSARY FORMATS:
 2000 FORMAT(10X,'NUMBER OF AUTOREGRESSIVE TERMS=',I2,/,10X,'NUMBER OF E
     *RROR TERMS=',I2,/,10X,'NUMBER OF PERIODIC TERMS=',I2,/,10X,'NUMBER
     * OF SIN AND COS TERMS=',I2,/,10X,'LAG=',I2,/,10X,'NO. OF DATA',
     *'=',I4,/)
 2010 FORMAT(10X,'PARAMETER SELECTION FOR RAO AND KASHYAP MODEL')
 2020 FORMAT(3X,'VALUES OF TRANSFORMED DISCHARGE')
 2030 FORMAT(8(3X,F7.4))
 2040 FORMAT(/,10X,'MEAN OF DISCHARGE=',F10.6,/,10X,'STANDARD DEVIATION
     *OF DISCHARGE=',F10.6,/,10X,'SKEWNESS COEFFICIENT OF DISCHARGE=',
     *F10.6,/)
C     ************************************************************************
C
      WRITE(2,2010)
      WRITE(2,2000) IA,IE,IP,IS,LA,ND
      WRITE(2,2040) AMEAN,STDEV,ASK
      WRITE(2,2020)
      WRITE(2,2030) (Y(I),I=1,ND)
      RETURN
      END
```

```
      SUBROUTINE KOLM1(X,N,IER,IFCOD,U,S,PROB,Z)
C     TESTS THE DIFFERENCE BETWEEN THE EMPIRICAL AND THEORETICAL
C     DISTRIBUTIONS USING THE KOLMOGOROV SMIRNOV TEST.
C     X      :INPUT VECTOR OF N INDEPENDANT OBSERVATIONS.
C     PROB  :THE PROBABILITY OF STATISTIC BEING .GE. TO Z.
C     IFCOD:CODE OF THE THEORITICAL DISTRIBUTION FUNCTION.
C     U,S   :STATISTICS OF VECTOR X ACCORDING TO IFCODE.
C     IER   :ERROR INDEX VALUE.
C     ******************************************************************
C
      DIMENSION X(976)
C     NON DECREASING ORDER OF X(I) .
      IER=0
      DO 5 I=2,N
      IF(X(I)-X(I-1))1,5,5
    1 TEMP=X(I)
      IM=I-1
      DO 3 J=1,IM
      L=I-J
      IF(TEMP-X(L)) 2,4,4
    2 X(L+1)=X(L)
    3 CONTINUE
      X(1)=TEMP
      GO TO 5
    4 X(L+1)=TEMP
    5 CONTINUE
C     COMPUTES MAXIMUM DEVIATION DN .
      NM1=N-1
      XN=N
      DN=0.0
      FS=0.0
      IL=1
    6 DO 7 I=IL,NM1
      J=I
      IF(X(J)-X(J+1)) 9,7,9
    7 CONTINUE
    8 J=N
    9 IL=J+1
      FI=FS
      FS=FLOAT(J)/XN
      IF(IFCOD-2) 10,13,17
   10 IF(S) 11,11,12
   11 IER=1
      GO TO 29
   12 Z=(X(J)-U)/S
      CALL NDTR(Z,Y,D)
      GO TO 27
   13 IF(S) 11,11,14
   14 Z=(X(J)-U)/S+1.0
      IF(Z) 15,15,16
   15 Y=0.0
      GO TO 27
   16 Y=1.0-EXP(-Z)
      GO TO 27
   17 IF(IFCOD-4) 18,20,26
   18 IF(S) 19,11,19
   19 Y=ATAN((X(J)-U)/S)*0.3183099+0.5
```

```
      GO TO 27
   20 IF(S-U) 11,11,21
   21 IF(X(J)-U) 22,22,23
   22 Y=0.0
      GO TO 27
   23 IF(X(J)-S) 25,25,24
   24 Y=1.0
      GO TO 27
   25 Y=(X(J)-U)/(S-U)
      GO TO 27
   26 IER=1
      GO TO 29
   27 EI=ABS(Y-FI)
      ES=ABS(Y-FS)
      DN1=AMAX1(ES,EI)
      DN=AMAX1(DN1,DN)
      IF(IL-N) 6,8,28
   28 Z=DN*SQRT(XN)
      CALL SMIRN(Z,PROB)
      PROB=1.0-PROB
   29 RETURN
      END


      SUBROUTINE KOLM2(X,Y,N,M,Z,PROB,DN)
C     TESTS THE DIFFERENCE BETWEEN TWO SAMPLE DISTRIBUTION FUNCTIONS
C     USING THE KOLMOGROV-SMIRNOV TEST.
C     X:(N*1)INPUT VECTOR.
C     Y:(M*1)INPUT VECTOR.
C     PROB:THE PROBABILITY OF THE STATISTIC BEING.GE.Z.
C     Z:OUTPUT VARIABLE CONTAINING THE GREATEST VALUE WITH RESPECT
C       TO THE SPECTRUM OF X AND Y.
C     ***************************************************************
C
      DIMENSION X(976),Y(976)
C     STORE X INTO ASCENDING ORDER .
      DO 5 I=2,N
      IF(X(I)-X(I-1))1,5,5
    1 TEMP=X(I)
      IM=I-1
      DO 3 J=1,IM
      L=I-J
      IF(TEMP-X(L)) 2,4,4
    2 X(L+1)=X(L)
    3 CONTINUE
      X(1)=TEMP
      GO TO 5
    4 X(L+1)=TEMP
    5 CONTINUE
C     SORT Y INTO ASCENDING ORDER
      DO 10 I=2,M
      IF(Y(I)-Y(I-1)) 6,10,10
    6 TEMP=Y(I)
      IM=I-1
      DO 8 J=1,IM
      L=I-J
      IF(TEMP-Y(L))7,9,9
    7 Y(L+1)=Y(L)
    8 CONTINUE
      Y(1)=TEMP
```

```
      GO TO 1
    9 Y(L+1)=TEMP
   10 CONTINUE
C     CALCULATE DN=ABS(FN-GM) OVER THE SPECTRUM OF X AND Y .
      XN=FLOAT(N)
      XN1=1.0/XN
      XM=FLOAT(M)
      XM1=1.0/XM
      I,J,K,L=0
      DN=0.0
   11 IF(X(I+1)-Y(J+1))12,13,18
   12 K=1
      GO TO 14
   13 K=0
   14 I=I+1
      IF(I-N) 15,21,21
   15 IF(X(I+1)-X(I))14,14,16
   16 IF(K) 17,18,17
C     CALCULATE THE MAXIMUM DIFFERENCE,DN .
   17 DN=AMAX1(DN,ABS(FLOAT(I)*XN1-FLOAT(J)*XM1))
      IF(L) 22,11,22
   18 J=J+1
      IF(J-M) 19,20,20
   19 IF(Y(J+1)-Y(J)) 18,18,17
   20 L=1
      GO TO 17
   21 L=1
      GO TO 16
C     CALCULATE THE STATISTIC Z .
   22 Z=DN*SQRT((XN*XM)/(XN+XM))
C     CALCULATE THE PROBABILITY ASSOCIATED WITH Z .
      CALL SMIRN(Z,PROB)
      PROB=1.0-PROB
      RETURN
      END


      SUBROUTINE PARA(T,N,AMEAN,AST,ASK)
C     IT COMPUTES MEAN,STANDARD DEVIATION AND SKEWNESS OF THE VECTOR T.
C     AMEAN:MEAN VALUE
C       AST:STANDARD DEVIATION
C       ASK:SKEWNESS COEFFICIENT
C     ******************************************************************
C
      DIMENSION T(976)
      AN=N
      SUM=0.0
      DO 10 I=1,N
   10 SUM=SUM+T(I)
      AMEAN=SUM/AN
      SUM=0.0
      SUM1=0.0
      DO 20 I=1,N
      SUM=SUM+((T(I)-AMEAN)**2)
      SUM1=SUM1+((T(I)-AMEAN)**3)
   20 CONTINUE
      AST=SQRT(SUM/AN)
      ASK=SUM1/(AN*AST**3)
      RETURN
      END
```

```
      SUBROUTINE POWER(ML,PHI,R2)
C     OBJECT:
C     IT CALCULATES AND WRITES THE POWER SPECTRUM PS.
C     ********************************************************************
      DIMENSION R2(49),PS(49)
 1000 FORMAT(10X,34(1H*),/,14X,'WH',15X,'PS(I)',/,10X,34(1H*))
 2000 FORMAT(12X,F10.6,8X,F10.6)
C
      WRITE(2,1000)
      DO 15 I=1,ML
      II=I-1
      WH=PHI*II/ML
      PS(I)=0.0
      IF(I.EQ.1.OR.I.EQ.ML) EK=0.5
      IF(I.NE.1.AND.I.NE.ML) EK=1.0
      DO 10 J=1,ML
      JJ=J-1
      PS(I)=PS(I)+(EK*R2(I)*COS(PHI*JJ*II/ML))
   10 CONTINUE
      PS(I)=2.0*PS(I)/PHI
      WRITE(2,2000)WH,PS(I)
   15 CONTINUE
      RETURN
      END


      SUBROUTINE SMOS(V,NV)
C     IT SMOOTHS AND WRITES THE POWER SPECTRUM BY USING THE HAMING
C     WINDOW ALGORITHM.
C     ////////////////////////////////////////////////////////////////////
      DIMENSION V(50),X(50)
      NVV=NV-1
      DO 10 I=1,NV
      IF(I.EQ.1) X(I)=0.54*V(I)+0.46*V(I+1)
      IF(I.GT.1.AND.I.LE.NVV) X(I)=0.23*V(I-1)+0.54*V(I)+0.23*V(I+1)
      IF(I.EQ.NV) X(I)=0.54*V(I)+0.46*V(I-1)
   10 CONTINUE
      WRITE(2,2060) (X(I),I=1,NV)
      RETURN
 2060 FORMAT(//,10X,'X-VALUES:',/,10(5X,F10.6),/)
      END
```