COMPARISON BETWEEN TWO SIMPLE LINEAR CATCHMENT MODELS FOR THE YALA RIVER BASIN

BY

JAMES MAGEZI-AKIIKI BYABAKAMA

A THESIS SUBMITTED IN PART FULFILMENT FOR THE DEGREE OF MASTER OF SCIENCE IN METEOROLOGY OF THE UNIVERSITY OF NAIROBI.

DECLARATION

This thesis is my original work and has not been presented for a degree in any other University.

JMagezi-akiiki 18/06/87

JAMES MAGEZI-AKIIKI BYABAKAMA

This thesis has been submitted for examination with our approval as University Supervisors

LJOgallo

DR. L.J. OGALLO

18/6/87

MR. F.M. MUTUA

Department of Meteorology,
University of Nairobi,
P.O. Box 30197,
NAIROBI,
Kenya.

CONTENTS

CHAPTER 3

LIST OF FIGURES

Figure Page

Figures

Figures | Pages

Figures Pages

Figures

Figure

LIST OF TABLES

Table Page

ABSTRACT:

In this study, two simple linear catchment models were used to model the rainfall-runoff characteristics over Yala catchment.

The first model was based on the response function. In this model a linear stochastic process for the rainfall-runoff system which yields the Wiener-Hopf integral equation was assumed. The method of maximum likelihood was then used to compute optimum response functions. The average response function together with the computed areal rainfall were finally used to estimate streamflow during the verification period.

In the second method, a multivariate autoregressive model was used in which linear transfer functions of different orders were fitted to the daily areal rainfall and streamflow data. The optimum autoregressive model was again used to estimate discharge during the verification period.

The results from the two approaches were lastly compared statistically.

Prior to fitting the models, the statistical characteristics of the computed daily areal rainfall and daily discharge series were investigated through time series analysis. The characteristics of the auto-correlation, cross-correlation, spectral and cross-spectral functions were examined.

Yala river basin is located in drainage area one in Western Kenya. The study was carried out in this drainage basin using daily rainfall and discharge records for nine years (1968 to 1976). The first seven years were used for calibration of the models, while the last two years were used for verification of the skill of the fitted models. The calender year was divided into two seasons namely the long and short rain seasons. The models were fitted for both seasons.

Results from spectral analysis, indicated periods of 2 - 4 days, 5 - 7 days and 8 - 10 days in both rainfall and discharge records. The results from the cross-correlation and coherence analyses showed significant correlations between time lags of 2 - 6 days with peaks at a lag of 3 days during the first rain season and a lag of 4 days during the second rain season.

The response functions had peaks at a constant time lag of 3 days during the first rain season and 4 days during the second rain season.

The results from the fitted autoregressive models indicated that the autoregressive models with four rainfall and two discharge lag terms fitted the discharge data best.

It was noted that although both models could not adequately estimate the extreme discharge values, the response function model was better in real time

forecasting. This model however, was poor in giving the actual estimates of the observed discharge. On the other hand, the autoregressive models gave good estimates of the magnitudes of the discharges. The peak discharge values were, however, shifted in some cases. It was therefore suggested that the response function models which gave good estimates of the time of occurrence of the discharge peaks would be the best models for the Yala river basin for real time forecasting.

CHAPTER ONE

1.0. INTRODUCTION

In most developing countries, river gauging stations have been recently established as compared to rainfall networks. Rainfall records are therefore generally longer than runoff records. River gauging stations are also fewer, more expensive to instal and maintain and are occasionally located in remote places. For these reasons, discharge records usually have many discontinuities. It is therefore necessary to simulate discharge data using some hydrologic techniques. Many methods have been used to simulate discharge data. In this study, the hydrological modelling technique has been used to simulate discharge data from rainfall records.

Hydrological models are becoming increasingly important tools for planning and management of water resources especially in forecasting and synthetic problems including all branches of the National Economy that are either directly or indirectly related to water utilization. They are specially important in planning and operating hydroelectric power plants and domestic, industrial and irrigation water supply systems as well as in certain requirements of Navigation and timber rafting.

A different kind of forecast is required for each purpose. Navigation requires forecasts of the water level in shipping lanes, mainly in the vicinity of shoals and landing stages. Hydroelectric power plants, industrial and irrigation water supply systems call for forecasts of the mean monthly, seasonal and annual flows as a basis for regulation of reservoirs and planning discharge. Timber rafting on the other hand requires forecasts of the mean daily stages during the passage of floods when great masses of timber are floated down small rivers in the forest regions.

In addition to their use in current planning of water resources utilization, hydrological forecasts are widely employed to predict and warn against dangerous hydrological phenomena such as floods. Overbank flooding which is a natural feature, results in great losses,for example, loss of life and property, injury, inconvenience and other indirect losses caused by floods.

Embankments can be constructed to control the floods from spreading over the floodplains and cause damage to property. However, considering the high costs involved in the construction of such flood control structures, an alternative is to set up some flood forecasting procedure.

Forecasts of flood events are of great value as they can generally reduce the damages caused by floods especially for densely populated industrial and residential regions which are located in floodplains. There are two ways in which this beneficial effect may be achieved.

Firstly and most obviously, warning of a catastrophic flood event enables people to evacuate a danger area. If sufficient lead-time is provided, vulnerable possessions may be removed from the danger zone and preparations can be made such as sandbagging to minimize property damage.

The second method of achieving benefits through forecasting is to use flood frequency analysis in the design of structures within the floodplain and for floodplain zoning. As examples, a knowledge of magnitude-frequency relationships should be used in the design of dams, highway bridges, railway bridges, culverts and water supply systems. Floodplain zonning ensures that housing and industrial developments are not located in high-risk zones, or at least if they are located in such areas, there can be no justification for compensation in the inevitable event of flood damage. High risk zones along the banks of rivers susceptible to floods should otherwise be used for activities compatible with this risk, such as parkland, recreation areas, some types of crop production or grazing.

A scientifically sound hydrological forecast can only be produced on the basis of a correct understanding of the physical nature of the interaction between the hydrosphere and the atmosphere. The basic laws governing seasonal fluctuations in river and reservour regimes can be determined from an analysis of observational data. The advent of hydrological forecasting thus became possible only after a considerable amount of observational material had been accumulated and the laws governing the movement of water over a basin area had been derived therefrom.

A hydrological model which is used in forecasting is a simplified representation of a complex system. Every model may have a few specific purposes, such as forecasting and control, and the model need only have just enough significant detail to satisfy these purposes. Thus the basic premise in model building is that complicated systems (all real systems are usually complicated) do not always need complicated models. For instance, the complete phenomenological description of river flow processes is very complicated, but still one can get relatively simple models of riverflow processes which yield adequate performance in forecasting and control. Models with a degree of complexity beyond a certain level often perform poorly in comparison with some simpler models. For example, including factors like infiltration or evaporation as input parameters without knowing the details of the factors which affect them such as the type of soil, type of vegetation, and topography, could lead to erroneous results from the model. Often if a model for a given process involves a large number of parameters, it is a good indication that we have to consider an entirely different family of models for that given process. Thus it is advisable to fit relatively simple models to the given data and to increase the complexity of the model only if the simpler model is not satisfactory.

There are several ways in which most hydrological models developed to date can be classified. They may be classified into two broad classes namely conceptual and empirical models (Clarke, 1973). Empirical models do not attempt to describe the internal structure and the response of the system being modelled. The system is treated as a "black box". In this case a simple relationship which matches the input and output of the system is sought. Conceptual models on the other hand involve complex systems of equations based on the physical laws and their theoretical concepts governing hydrologic processes.

Both empirical and conceptual models may further be classified as either deterministic or stochastic. In deterministic models, the processes are assumed to have no probabilistic distributions. The processes are then regarded as being free from random variations. One of the principal disadvantages of the deterministic models of systems is the absence of an effective method of comparison of various possible models that can be constructed by using the same empirical data, such that the comparison is relevant to the ultimate purpose of the models. For instance, a criterion such as least squares (between the output of the model and the observed input) can be used to compare the models. But the result of the comparison may not always be very relevant for forecasting or control, as systems that satisfy the least squares criterion may yield poor predictions. Furthermore, the utility of the model is intimately related to the "Generalization Capability" of the model. A model can be said to have generalization capability if the model constructed by using n-observations of the system does not differ very much from a model constructed by using a subsequent set of n-observations. Stochastic models on the other hand regard hydrologic processes to be characterised by statistical properties. The construction of models in this probabilistic framework allows us to compare different models and to give a precise meaning to words such as "significant" and "negligible" which is lacking in models constructed in a deterministic context. Since the class of stochastic models encompasses the class of deterministic models, searching for the best stochastic model does not entail any loss of generality.

Hydrological models can on the other hand be divided into a slightly more detailed classification as:

(i) Regression models

(ii) Conceptual models

(iii) Watershed models

Regression models seek to relate the input rainfall to the output discharge by statistical correlations. Sometimes other variables may be included to improve the correlation.

Some conceptual models are based on the unit hydrograph theory. The essence of the unit hydrograph theory is the assumption of linearlity in the relation between rainfall and discharge. Having made this assumption, it remains to establish the relation between the characteristics of the catchment and the response in the discharge to a pre-determined input rainfall; that is, the "indicial response". The most frequently used indicial response is the "instantaneous unit hydrograph" defined as the unit hydrograph of runoff caused by a unit volume of rainfall input, generated instantaneously and uniformly over the whole catchment. This relationship between the input rainfall and output discharge for a linear system is the convolution equation:

$$g(t) = \int_{-\infty}^{t} h(t-\tau)\ i(\tau)\ d\tau \quad \text{---------------} \quad (1)$$

in which g(t) is the output at time t after the beginning of the input, $h(t-\tau)$ denotes the unit hydrograph ordinate at time $(t-\tau)$ from the origin of the unit hydrograph, $i(\tau)$ is rainfall input at time τ from the beginning of the input. Most conceptual models involve a basic approach of determining the indicial runoff response function h(t). Although equation (1) may vary in form, Nash (1959), Dooge (1960) and O'Donell (1973) have suggested methods for solving the convolution equation analytically. An analytical solution of equation (1) is of the form:

$$h = \frac{s}{k\Gamma n}\ e^{-t/k}\ (t/k)^{n-1} \quad \text{------------} \quad (2)$$

in which

h = impulse response function which is a function of time.

s = total storage of the basin

k = linear storage parameter

Γ = gamma function

n = hypothetical number of linear reservoirs or channels.

Watershed models usually divide the catchment into numerous subcatchments. The water balance concept is applied to estimate the rainfall excess which appears as direct runoff by routing the flows from each subcatchment to the outlet of the basin. The use of watershed models requires very accurate original data.

Faced with this range of modelling techniques, the choice of a suitable method for modelling is at first sight not easy. Inevitably, there are two major factors affecting the choice, namely the information required from the method and the amount and quality of data available. The amount and quality of data from a drainage basin, both for geometry of the channel and flood plain and the previous discharges and water levels, is another significant factor in the choice of the modelling methods. If the data is of poor quality, then nothing is gained from using a very sophisticated and complicated technique which in most cases yields very un-realistic results. In such a case then, the use of the simplest model is preferable. On the other hand, if the data amount is limiting, then a model such as the conceptual model which uses the least amount of data for it's calibration would be a preferable choice.

In this study two different simple linear models have been used to study the rainfall-runoff relationships in the Yala river basin of Kenya. The results from the two approaches have then been compared.

Before the estimation of streamflow from the daily rainfall records, it was found necessary to first investigate the statistical characteristics of the computed daily areal rainfall and observed daily discharge in the Yala river basin through time series analysis. The auto-correlation, cross-correlation, spectral and cross-spectral characteristics of both the streamflow and areal rainfall series have been studied. In the auto-correlation and spectral analyses, periodic fluctuations in the two series were examined while in the cross-correlation and cross-spectral analyses, the cross-correlations, co-spectrum, quadrature, coherence and phase angles were computed in an attempt to examine relationships between the rainfall and streamflow series.

Some of the relevant hydrologic modelling approaches which have been used in previous studies are briefly reviewed in the following section.

1.1. LITERATURE REVIEW

In the past recent years, many efforts have been directed towards the understanding of the rainfall-runoff phenomenon, not only in the East African region, but also in many other parts of the world. The following is a review of some of the relevant research studies.

Since the introduction of the unit hydrograph by Sherman (1932), linear systems analysis has played an important role in relating input-output components in the rainfall - runoff modelling. The main assumptions in the unit hydrograph theory are linearlity and time invariance of the parameters. These assumptions imply that:

(i) The effect of a particular input component will be the same irrespective of when it occurs and irrespective of the state of the input or output at the time of occurance. (time invariance) and

(ii) The effect is proportional to the cause. The relationship or the operation of transforming the input into the output is implied uniquely by any corresponding input - output pair and in principle at least, this relationship can be abstracted and used to find the output for any arbitrary input or the input corresponding to any given output (linearlity and stability).

Although it is now realised that the assumption of linearlity is not strictly correct for natural catchment systems, this method is still widely used because it is simple and the results obtained are sufficiently accurate. However in the case of high-flows, rainfall - run-off systems are likely to manifest considerable non-linearlities.

In these cases of high flows, a non-linear approach to the rainfall - run-off modelling has been found to give more realistic results (Hino, 1970)

Since it is difficult to take into account all details involved in the rainfall - runoff mechanism, it is necessary to accept certain simplifying assumptions on the basis of which the prediction model for the output can be constructed. The input and output components may be considered in a lumped fashion. That is the rainfall input is represented by the average areal rainfall as measured at various rainfall gauging stations and streamflow output is measured at the outlet of the catchment.

Many studies have been carried out in both linear and non-linear rainfall - runoff modelling. Below are some of the examples.

Amorocho (1963) extended the convolution linear equation to include non-linear terms in his equation. A time variable linear system is represented by the following equation:

$$f(t) = \int_{-\infty}^{\infty} h(t-\tau)\ x(\tau)\ d\tau \quad \text{-------------} \quad (3)$$

where $h(t-\tau)$ is the impulse response function of the system. This can be made to conform with the Volterra condition of $h(t-\tau) = 0$ for $t \leq \tau$.

Amorocho (1963) considered a general case of a non-linear, time variable system which he expressed by the summation of the first term identical to equation (3) and a series of similar terms of progressively high dimensions. The non-linear response of the catchments under rainfall input was represented by means of functional series. This involves consideration of higher powers of the elements of the inflow and leads to a more accurate prediction of the forms and magnitudes of flood events than is possible by the linear approximation.

Eagleson et al (1966) determined a basin's unit impulse response from coincident records of storm rainfall and runoff by using the "black box" analysis. They selected individual storms and obtained their corresponding direct runoff after separating base flow from the flood hydrographs. In the analysis, no assumptions were made regarding the internal structure of the system other than those of linearlity and stationarity. They considered a physical system (initially relaxed) for which the input and output time functions are exactly related by the convolution integral. The integral equation expresses the system output as a function of the system input and the system unit impulse response. The basic problem was to solve the convolution equation for a meaningful response function when the input and output are related by a system that is not truly linear. For a meaningful solution of this type they introduced "maximum likelihood prediction" (optimum linear prediction). The optimum impulse response functions which are used in the prediction of streamflow are determined from the auto-covariances of rainfall and cross-covariances of rainfall and streamflow.

Hino (1970) used the linear theory of a lumped hydrologic system to a long range rainfall runoff system. He evaluated rainfall-runoff response functions using daily rainfall and daily streamflow records. He obtained good prediction of streamflow for the case of light and moderate rainfall, but for very heavy rainfall, the prediction of streamflow was not accurate. To get good prediction for the case of heavy rainfall, he modified the linear prediction procedure taking into account the non-linearity involved in the case of heavy rainfall-runoff system. The method used was to map by a non-linear transform both rainfall and runoff to a space where the transformed variables of rainfall and runoff become a nearly linear system. By computing the coherence function of rainfall and runoff, the coherence values approached unity thereby confirming the linearlity of the system in the transformed domain. Then he applied the theory of linear prediction in the transformed space and results obtained were close to the observed streamflow.

Tojiro and Ikebuchi (1971) studied long range runoff response based on the Information Theory. First statistical properties of daily rainfall and streamflow series were studied through correlation analysis. After time invariant linearization of the runoff system according to the physical mechanisms in the runoff, the unit impulse response function was derived from the Wiener-Hopf theory (equation). Next, statistical unit hydrographs that took into account both the variations of the water content in the sub-surface stratum during rain seasons and also the daily snowmelt water input to the system in the snowmelt season were evaluated. They found that the linear system method was effective for analysis and synthesis of long-range runoff system responses and satisfactory agreement was obtained between the observed and the predicted discharge, making it possible to predict the daily river discharge and to supplement the lack of data of river flows.

Delleur and Rao (1971) used the Fourier, Laplace and z-transforms to determine the Kernel function (impulse response) of catchments from discrete rainfall and runoff values. They found that the fourier transform gives a good understanding and interpretation of the mathematical properties of the impulse response function, but that the z-transform which is the discrete equivalent of the laplace transform, is easier to program and requires less computer time.

Blackie (1972) developed a simple conceptual model using the data from the East African Agricultural and Forest Research Organisation (EAAFRO) catchments. Previous analysis of the data had been based on the water balance approach. The model treats rainfall and groundwater storage as the two types of inputs while the output consists of evaporation, ground water storage and surface runoff. The outputs are then routed to obtain streamflow. After some developments on his model, Blackie used a nine-parameter model with a unit time interval of one day and obtained reasonable representation of the streamflow from the Kimakia catchments. The results of this model were indicative but not conclusive regarding the hydrologic changes that resulted from changes in land use.

Diskin et al (1972) studied the role of lag in a quasi-linear analysis of the surface runoff system. They developed a quasi-linear approach in modelling rainfall - runoff system. In this approach the linear convolution integral is assumed to represent the rainfall - runoff relationship during any one storm, but the Kernel function is taken to be different for various storms. The changes in the shape of the Kernel function may be represented in terms of the variations in the value of the lag defined as the time difference between centroids of input and output. Using the lag as a basis for the dimensionless plotting of the Kernel function leads to a curve that is nearly constant for all storm events in a given catchment. If the shape of the dimensionless curve is known as well as the factors affecting the value of the lag for any particular storm, it is possible to obtain a good fit to observed floods and to predict future floods for the catchment concerned.

Kovacs and Morth (1974) demonstrated that they could obtain a useful representation of the flood hydrographs of the Tana river at Garissa (Kenya) with an expression for the increase or decrease, δD, in terms of the rainfall recorded on the one-degree square bounded by the equator and 1°S and longitudes 37°E and 38°E of the following form:

$$\delta D = 0.6\ (R-0.1) - 0.09D \qquad (4)$$

where:

D = Discharge in cubic feet per second multiplied by a factor of 10^4

R = Daily rainfall in inches.

For the purpose of water resource planning, this model is too simple and would require the introduction of other parameters to vary the equation in a non-linear fashion.

Tavares (1975) regressed daily river discharge on daily rainfall and previous runoff values. His results showed that the regression of runoff on rainfall at lag zero and the runoff at one day lag had smaller residual variance than regression of runoff on rainfall only.

Todini and Bouillot (1975) presented an on-line rainfall runoff model of the Sieve river in Italy. To estimate the model parameters, they used instrumental variable maximum likelihood estimator. The data used were two hourly interval concurrent rainfall and runoff time series. A transfer function with autoregressive component of order one and moving average component of order one was identified and fitted. Using the theory of Kalman filters, the model was shown to have good prediction performance.

Whitehead and Young (1975) fitted a transfer-function model to daily rainfall and runoff data from Bedford Ouse river in eastern London. The rainfall series was found to be significantly correlated up to lag of three days. Therefore, the rainfall series was prewhitened using an auto-regressive model of order three. The prewhitened rainfall series was then cross correlated with similarly filtered residual flow series. A rainfall runoff model of second order autoregressive and second order moving average processes was identified. Instrumental variable maximum likelihood method was used to estimate the model parameters after which a water quality model was incorporated.

The Sacramento model (TAMS 1977) has been employed by the UNDP/WMO Hydrometeorological project on some of the rivers that flow into Lake Victoria. The model input is the average daily rainfall over the catchment and the model requires characteristic values for at least 16 parameters. It is very difficult to evaluate the performance of this model partly because of the amount of work involved in calibrating the sixteen parameters and also because only preliminary results after the calibration are so far available. Furthermore the results obtained from the preliminary analysis were not very reliable since very few raingauges were considered in each catchment. No provision was also made to acommodate the variation of the catchment sizes which varied from a few hundred to several thousand square kilometres.

Clark and Brutsaer (1978) have worked on a non-liner analysis of the relationship between rainfall and runoff for extreme floods. A practical method was developed to obtain optimal Kernels or response functions of the two-term truncated volterra series, describing rainfall - runoff system. The Kernels were determined on the basis of the three historical events and were applied to predict the unrelated event of a tropical storm, 'Agnes'. They found out that a non-linear model could predict the flood hydrograph from an extreme and complex storm more closely than the linear model. They also showed that the linear model resulted in an underestimate of the peak flow in the case of a very large and catastrophic flood.

Bennett (1979) applied Box-Jenkins techiniques using data from the East African Agricultural and Forest Research Organization (EAAFRO) catchments. The data were in the form of monthly time series for the years 1958 to 1968 from Lagan and Sambret basins. Some of his results showed that a transfer-function with autoregressive component of order two and a moving average of also order two fitted the monthly rainfall - runoff series. He also carried out cross-spectral analysis on the Lagan catchments monthly rainfall - runoff data. But this analysis yielded results which were hard to interpret.

Mutua (1979) applied the Wiener-Hopf theory of optimum dynamic systems for the determination of stable runoff impulse response functions of the Nzoia River drainage basin from coincident records of input rainfall and output discharge for a data observation interval of two hours in the upper sub-catchment and of eight hours in the whole catchment. With the few continuous and reliable records available the model was able to simulate hydrographs which were close to the actual (observed) hydrographs. Other simulated hydrographs showed departures from the observed hydrographs.

Kato (1982) assumed a linear stochastic process for the rainfall runoff system which yields Wiener-Hopf type of integral equation which relates the auto-covariance function of rainfall and streamflow through the catchment response function. He derived the optmum response functions using the Wiener-Hopf equation for the three sub-catchments of the Ruvu basin of Tanzania. Then using the average of the computed response functions and computed areal rainfall, he estimated stream flows during two rain seasons. The study showed that there was a good agreement between actual and estimated stream flow for the smallest and medium sized sub-catchments. For the large sub-catchment, there was a significant difference between actual and estimated streamflow in the case of very heavy rainfall.

Mutulu (1984) developed a multivariate Autoregressive model in which he fitted linear transfer function models to different sets of concurrent daily areal rainfall and streamflow series. He used daily rainfall and streamflow records for the Yala and Sondu river basins of Kenya and considered two wet seasons extending from April to June and August to October over the years 1972 - 1978 in fitting the model during the calibration stage. The models which were found to be statistically adequate were used to predict streamflow records for the periods April - June, 1979 and August - October 1979 for the Yala river basin and April to June 1979 and August to October 1980 for the Sondu river basin. The study showed that the models were able to reproduce the main features of the observed hydrographs. However, persistent departures of the forecasts from the observed streamflow values were evident especially with some high flow events.

The objectives of this study are discussed in the following section.

1.2. OBJECTIVES OF THE STUDY

Many methods have been used to study the relationship between rainfall and runoff. Some of these have been discussed in the last section.

In this study, the following two simple linear models were fitted for the Yala river catchment.

(i) In the first approach, a linear stochastic process for the rainfall-runoff system which yields the Wiener-Hopf type of integral equation was assumed. This equation relates the auto-covariance function of rainfall and discharge through the catchments response function. The method of maximum likelihood (optimum linear representation was used to compute optimum response functions from the Wiener-Hopf equation for each pair of time series in the calibration period. The average response functions were then evaluated and used together with the computed areal rainfall to estimate streamflow in the verification period.

(ii) In the second approach, a multivariate auto-regressive model in which linear transfer function models are fitted to different sets of concurrent daily areal rainfall and discharge series was used. A similar study was carried out for the Yala catchment by Mutulu (1984) but for a different period of years. So the transfer function models from Mutulu's study were recalibrated for the long term means and the ones found to be statistically adequate were used for prediction.

A period of nine years was used in this study extending from 1968 to 1976 with the first seven years used in the calibration of the model.

Lastly the results from the two approaches were compared.

Due to the complexity of the rainfall - runoff processes, simple rainfall - runoff models are usually preferred. Most of these simple models require the following assumptions:

(i) that the effect of a particular input component will be the same irrespective of when it occurs and irrespective of the state of the input or output at the time of occurrance (time invariance)

(ii) that the effect is propotional to the cause. The relationship or the operation of transforming the input into the output, is implied uniquely by any corresponding input - output pair and in principle at least, this relationship can be obstructed and used to find the output for any arbitrary input or the input corresponding to any given output (linearlity and stability),

(iii) that the input and output components are considered in a lumped fashion. That is the rainfall input is represented by the average areal rainfall as measured at certain locations in the catchment and the streamflow output is measured at the outlet of the catchment.

The fact that the process of runoff generation on a catchment is non-linear and time variant can scarcely be doubted. The fact that the runoff resulting from an individual rainfall event is often a residual after certain storages and deficiencies in the catchment have been made up implies a departure from linearlity and the difference between dry season and wet season responses to the same rainfall event implies time variance. The discharge at any given instant is dependent not just on the recent history of rainfall on the catchment but also on the evaporation, particularly the amount since the previous rainfall. Neglect of this highly seasonal factor in any analysis of rainfall and discharge records may imply that the relationship obtained would itself exhibit seasonal variation or even that no consistent relationship could be identified. This difficulty is further compounded by the fact that in most such analyses, the relationship is sought between the recorded discharge and an index of areal mean rainfall obtained by averaging the records of a number of raingauges. The non-representativeness of the recorded rainfall and the distribution of rainfall within the catchment are usually neglected.

Therefore,in practice, in analysing systems which are not exactly linear and time invariant, or where the data are subject to errors, problems may arise both in identifying the operation or in computing an input corresponding to a given output function of time.

However, linear systems have a remarkable property that given any input-output record, the operation may be obtained analytically without further assumptions.

If the data are without errors, the expression of the operation will similarly be error free, but even if the data are subject to random errors and the system only approximately linear, an estimate of the operation can still be found without further assumptions. The reliability of such an estimate depends on the extent of the sample data, on the errors of those data and on the extent and nature of the departure of the actual operation from the linear assumption. The method of analysis best suited to a particular case depends on the nature of the data, mainly on whether continuous or discrete, periodic or non-periodic and whether capable of expression by analytic functions of time.

In order to moderate the effects of these shortcomings, the following points were considered in this study:

(i) Two wet seasons were used. It is expected that during the wet seasons, the soil is saturated most of the time and therefore most of the rainfall received will contribute to runoff.

(ii) The region of study was picked from the rainy part of Western Kenya. In this region there are very few if any dry days during the wet seasons. This minimizes the effects caused by evaporation.

These two points give support to the assumptions of linearlity and time invariance.

It can further be noted that:

(i) The Yala river basin has nearly homogeneous soils and vegetation except near the mouth of the river towards Lake Victoria where it is swampy.

(ii) The basin does not have very high mountains.

(iii) The Yala River basin is small and it is situated in the region of high and uniformly distributed rainfall.

These three points give support to the method of lumping the input and output components. Before studying the rainfall-runoff mechanisms for the Yala river basin, time series analysis was carried out to give further evidence of the significance of the above assumptions.

1.3. CLIMATOLOGY AND OTHER GENERAL CHARACTERISTICS OF THE STUDY REGION

The Yala river catchment (area 1F) is located in the Lake Victoria drainage area which is drainage area one in Kenya. Fig.1 shows the location of the Yala river catchment on the map of Kenya.

The rainfall in the Lake Victoria drainage area is mainly convective in form of thunderstorms and showers from cumulus clouds and occurs mainly in the afternoons (Tomset, 1975 and Asnani and Kinuthia, 1979). This Lake Victoria drainage area which is characterised by humid climate has no real dry seasons. However, the seasonal rainfall patterns exhibit a bimodal distribution associated with the intertropical convergence zone which migrates

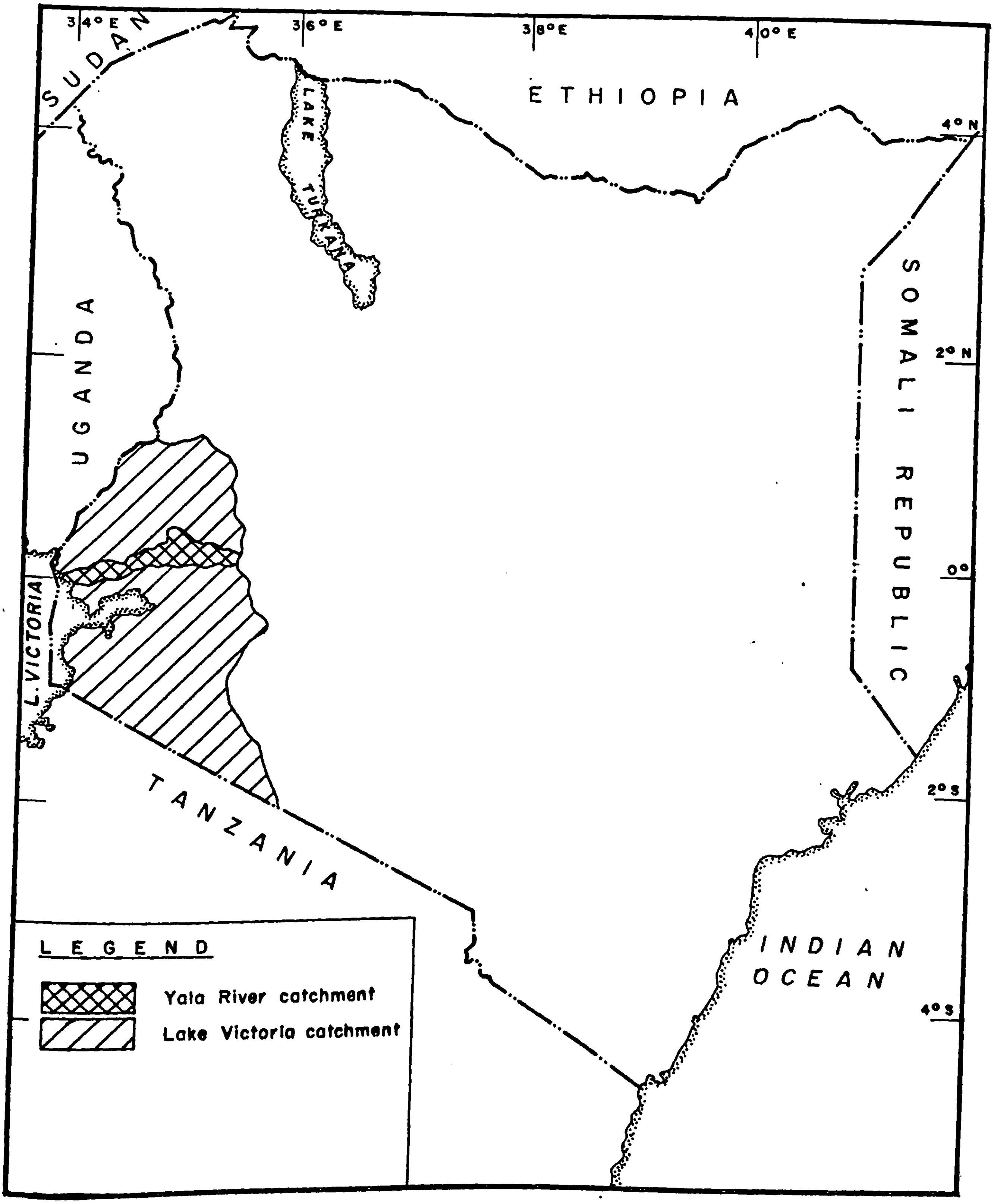


Fig. 1 MAP OF KENYA SHOWING THE LOCATION OF YALA RIVER BASIN

seasonally across the region with the overhead sun. The two rain seasons are centred around March - May (long rains) and September - November (short rains). Substantial rainfall is also received during the months of July/August due to moisture influx from the Atlantic ocean and the Congo/Zaire basin. The Yala river drains an average area of 2,388 km² at the Yala gauging station 1FG1... The basin area extends from Elgeyo escarpment and Nandi hills to lake Victoria. The basin axis extends roughly 170 km East to West. The catchment is characterised by poorly drained planosols in the upper reaches of adverse topography and small holdings in the middle and lower reaches. Poor swampy soils dominate the river mouth.

The mean annual rainfall over the Yala basin is around 1500 mm and the potential annual evaporation is of the order of 1200 mm. A typical mean monthly rainfall distribution for a station in Yala basin is shown in Figure 2 and Figure 3 shows the mean annual rainfall of the basin.

Major drainage and irrigation efforts have been under way since 1967. Power development opportunities are limited due to inadequate reservoir storage. However, several damsites have been identified (TAMS 1980). The catchment has a relatively high incidence of flooding during heavy rains.

1.4. DATA REQUIREMENTS

Rainfall records showed that 37 stations in the Yala river catchment have standard raingauges which report

KAIMOSI TEA ESTATE LTD (30 YEARS MEAN)

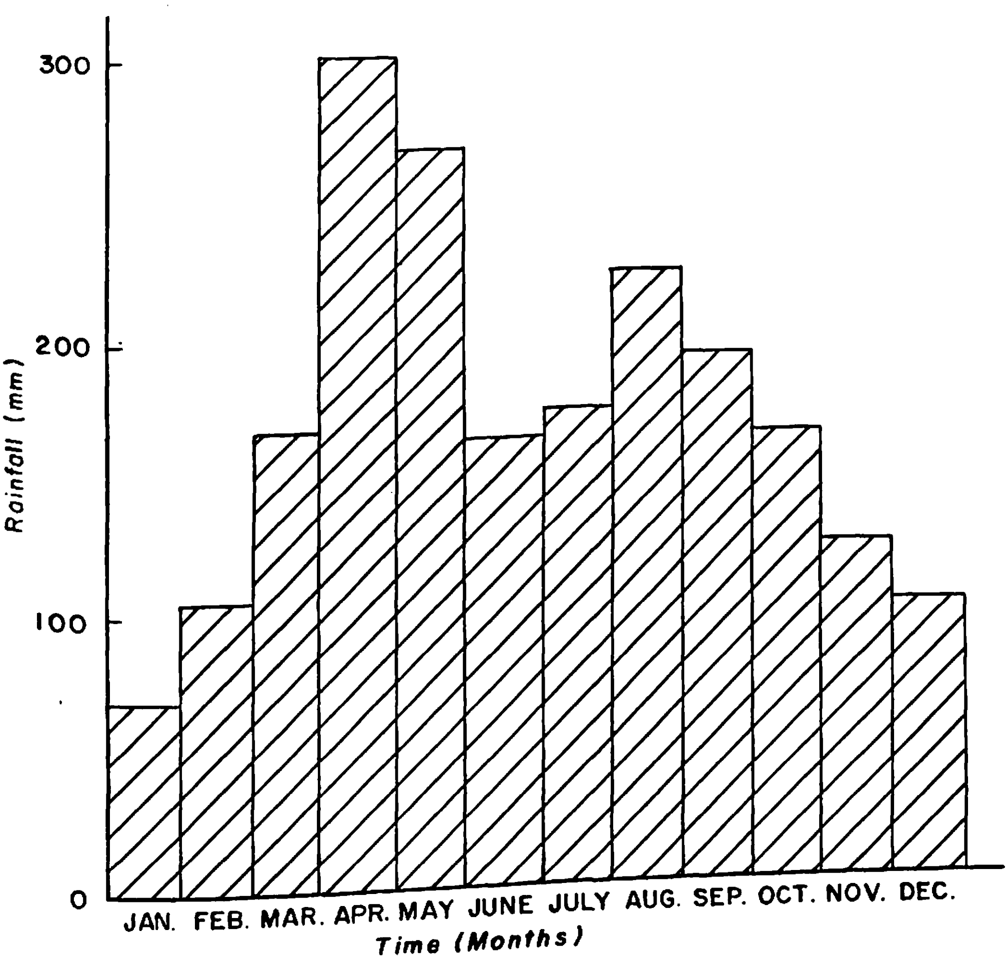


Fig. 2 MONTHLY RAINFALL DISTRIBUTION FOR A TYPICAL STATION IN YALA BASIN

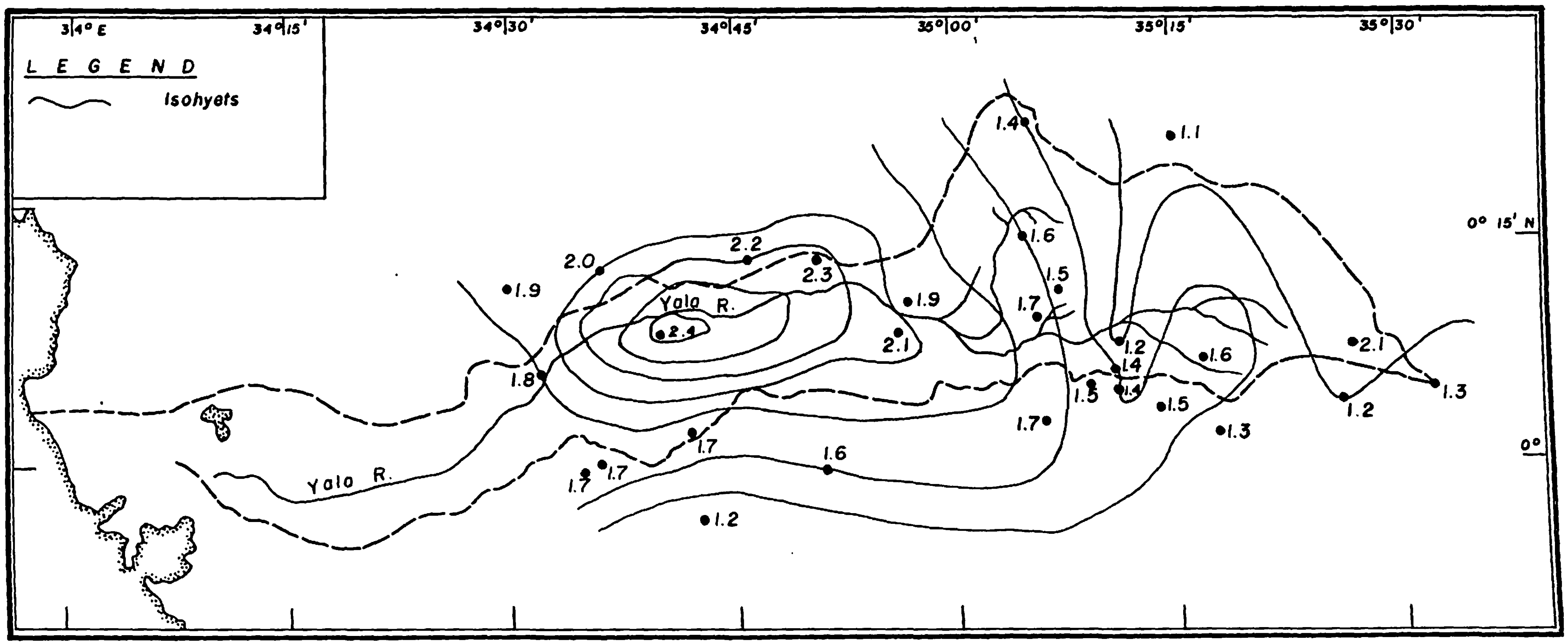


Fig. 3 MEAN ANNUAL RAINFALL (x 100^3 mm) OF YALA BASIN

daily rainfall values. However many stations had missing records and as a result, only 14 stations within the Yala basin were finally selected for this study. 15 other stations around the Yala river basin boundary were selected and used in the isopercentile estimation of areal rainfall.

This study was restricted to two relatively wet seasons. The first wet season was taken to run form April to June (91 days) while the second season extended from August to October (92 days). These two seasons April - June and August - October will subsequently be referred to as seasons 'one' and 'two' respectively.

1.4.1 RAINFALL DATA

Daily rainfall values for the 14 stations in the Yala river basin were obtained for the years 1968 to 1976 both years inclusive from the Kenya Meteorological Department. Annual total rainfall values were also obtained for the above 14 stations and for the 15 other stations around the Yala river basin boundary for use in the isopercentile areal rainfall estimation. Table 1 lists the 29 selected rainfall stations with their characteristics. The spatial distribution of these stations is given in Figure 4.

1.4.2 DISCHARGE DATA

The corresponding discharge data for the years 1968 to 1976 were obtained from the Ministry of Water Development

TABLE 1: LIST OF RAINFALL STATIONS USED IN ANALYSIS FOR YALA BASIN

STATION NUMBER AS IN FIG. 4	STATION REG. NUMBER	STATION NAME	STATION LOCATION LATITUDE	STATION LOCATION LONGITUDE	YEAR OPENED
1.	8934031	ST. MARY'S SCHOOL YALA	0^o 06'N	34^o 32'E	1935
2	8934103	VIHIGA AGRIC. OFFICE	0^o 02'N	34^o 43'E	1950
3	8934041	YALA MWIHILA SEC. SCHOOL	0^o 11'N	34^o 37'E	1940
4	8934028	KAKAMEGA FOREST STATION	0^o 14'N	34^o 52'E	1935
5	8934072	KAIMOSI TEA ESTATE LTD.	0^o 09'N	34^o 56'E	1947
6	8934078	KAIMOSI FARMER'S TRAINING CENTRE	0^o 13'N	34^o 57'E	1958
7	8935127	KAPSABET KABIYET HEALTH CENTRE	0^o 24'N	35^o 05'E	1954
8	8935062	UNIVERSITY OF EASTERN AFRICA	0^o 15'N	35^o 05'E	1938
9	8935018	KAPSABET DISTRICT OFFICE	0^o 12'N	35^o 06'E	1904
10	8935112	NANDI FOREST STATION	0^o 12'N	35^o 04'E	1950
11	8935134	KESSUP FOREST RESERVE	0^o 39'N	35^o 31'E	1955
12	8935152	NANDI HILLS AGRIC. OFFICE	0^o 07'N	35^o 11'E	1962
13	8935120	KIBABRET ESTATE LTD.	0^o 08'N	35^o 17'E	1952
14	8935080	NABKOI FOREST STATION	0^o 08'N	35^o 28'E	1947
15	9034011	MASENO VETERINARY STATION	0^o 00'S	34^o 35'E	1934
16	9034078	MASENO GOVT. INSTITUTE	0^o 00'S	34^o 36'E	1959
17	9034060	KISUMU NEW PRISON	0^o 04'S	34^o 43'E	1951
18	9034069	KIBOS KISUMU WATER SUPPLY	0^o 00'S	34^o 49'E	1955
19	8935033	NANDI HILLS, SAVANI ESTATE	0^o 03'N	35^o 06'E	1929
20	8935161	NANDI HILLS, KIBPARI TEA ESTATE	0^o 05'N	35^o 09'E	1958
21	8935095	KAPSABET NANDI TEA FACTORY	0^o 06'N	35^o 11'E	1947
22	8935071	NANDI, SIRET TEA CO. LTD.	0^o 04'N	35^o 14'E	1944
23	8935001	KABNGENDUI, KIBET FARM	0^o 02'N	35^o 18'E	1914
24	8935148	KIPKURERE FOREST STATION	0^o 05'N	35^o 25'E	1959
25	8935137	TIMBOROA FOREST STATION	0^o 04'N	35^o 32'E	1954
26	8935074	ELDORET, GARA FALLS ESTATE	0^o 22'N	35^o 15'E	1946
27	8934104	MUKUMU GIRLS SECONDARY SCHOOL	0^o 13'N	34^o 46'E	1960
28	8934002	BUKURA F.TR. CENTRE	0^o 13'N	34^o 37'E	1924
29	8934040	BUTERE HEALTH CENTRE	0^o 13'N	34^o 30'E	1939

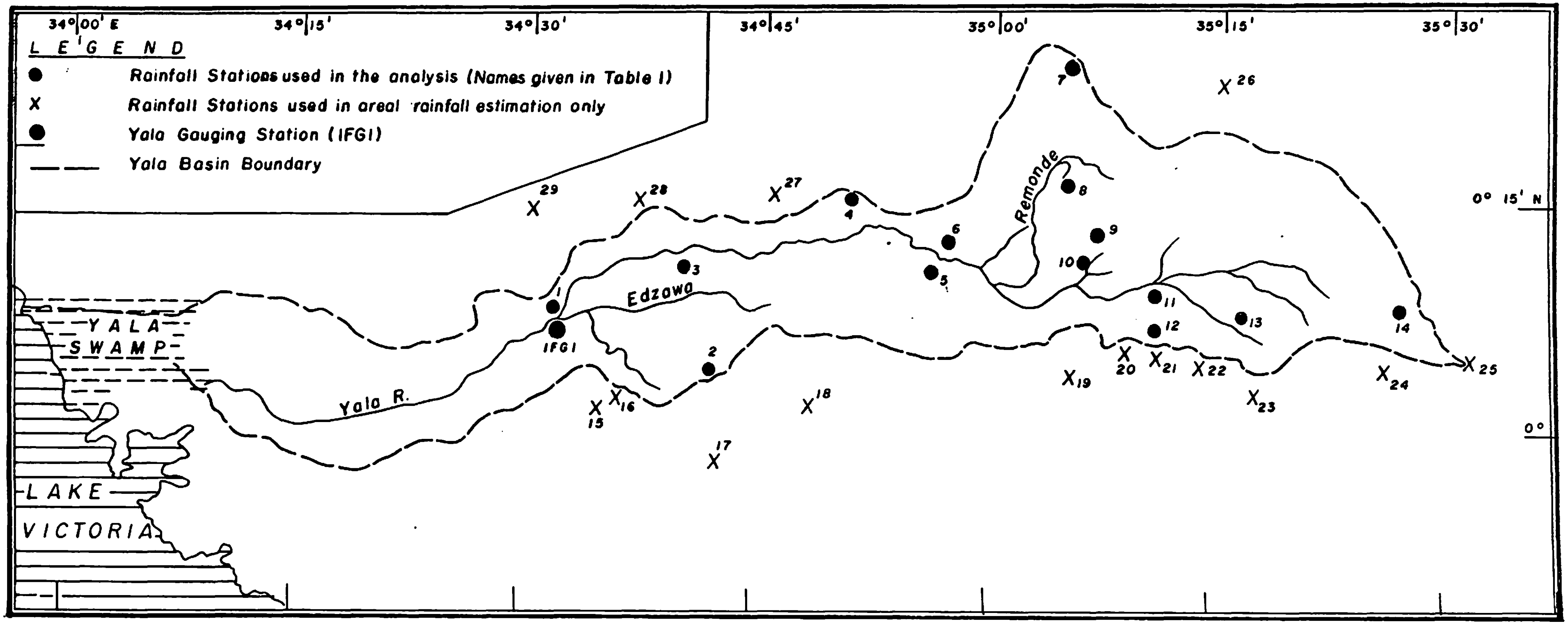


Fig. 4 THE YALA RIVER BASIN

Headquarters for both seasons one and two. These were observations made at the Yala river gauging station 1FG1. This station has an automatic recorder and has a record of about 30 years. A check on the rating curves of the Yala river gauging station 1FG1 indicated that the station had stable controls for the period of study. Records at random were checked by using the rating curves to ascertain their accuracy. The location of this gauging station is given in Figure 4.

The methods of analysis will be discussed in the next chapter.

CHAPTER TWO

2.0. METHODOLOGY

In this chapter the methods of analysis used in the study will be discussed.

Before subjecting the records to time series analysis, the daily rainfall series were lumped through areal averaging over the whole catchment. With the computed areal rainfall series as inputs to the catchment systems, rainfall - runoff mechanisms were examined for the catchment.

2.1. ESTIMATION OF AREAL RAINFALL

There are various techniques which can be used in the computation of areal rainfall over a catchment. They include:

(a) Thiessen polygon method

(b) Chidley and keys method

(c) Unweighted arithmetic mean method

(d) Isohyetal method

(e) Isopercentile method

The choice of the method depends largely on the characteristics of the catchment concerned. Some previous studies have however indicated that the isopercentile method gives relatively better results than the other methods (Basalirwa, 1980). Another advantage with the isopercentile method is that it reduces the need for a consistently reporting network (Chow, 1964). It is also the method recommended when a catchment has closed isohyetes due to the influence of diverse physiographic features which is the case with the Yala river catchment. The isopercentile method which is an extension of the isohyetal method was chosen for this study.

The two methods are reviewed briefly in the following sections.

2.1.1. ISOHYETAL METHOD

In this method, rainfall values at various locations (rainfall gauging stations) within and around the catchment and for appropriate durations are plotted on a suitable map. Isohyets are then drawn at equal intervals and the areas between consecutive pairs of isohyets are estimated with the help of a planimeter. These areas between consecutive pairs of isohyets are then divided by the total catchment area to form weights. These weights are then multiplied by the corresponding mean rainfall between a pair of isohyets to form products. The sum of these products is the estimated mean areal rainfall for the catchment.

2.1.2. THE ISOPERCENTILE METHOD

This method requires that a minimum of twenty years of annual rainfall for each station in the catchment be available. Unfortunately very few stations have continuous data for such a long period. In such cases the method is applied to shorter records. The mean annual rainfall for each station is computed and using the isohyetal method, the average annual general areal rainfall for the catchment is estimated. The catchment areal rainfall is then given by:

$$\overline{X}_i = \frac{X_a}{N} \sum_{n=1}^{N} \frac{X_{i,n}}{X_n} \quad \text{-------------------- (5)}$$

where:

$\overline{X}_i$ = areal rainfall average on a given day i.

X_a = Average annual general areal rainfall, normally determined over a period of thirty years using the isohyetal method.

N = Total number of stations

$X_{i,n}$ = Rainfall total of the n^{th} station and on a given day i.

X_n = Mean annual rainfall for a particular station n taken over the same period as that used in determining X_a.

2.2. TIME SERIES ANALYSIS

Before showing the application of the linear theory to the prediction of stream flow, the statistical characteristics of the computed areal rainfall and discharge series were investigated both in time and frequency domains. The functions which have been studied are auto-correlation, cross-correlation, spectrum, cross-spectrum, coherence and phase angle.

The auto-correlation and spectral density functions were used to investigate stationarity in the daily areal rainfall and streamflow series, to aid in building prewhitening filters for the areal rainfall series, to detect cycles in both series and to help in the rainfall-runoff model diagonostic checking. The cross-correlation function was used to determine the lag of maximum response between stream flow and rainfall and to lay down the initial hypothesis as to the orders of rainfall - runoff models. Lastly the cross-spectral density estimates, coherence and phase angle were used to confirm the significance of the cycles observed in the spectral estimates of both rainfall and streamflow, to confirm the assumption of linearlity and to give an idea as to how the rainfall and streamflow series lead or lag behind each other.

It is necessary that the rainfall and streamflow series be stationary if the functions mentioned above have to give meaningful results. A stochastic process is first order stationary if its mean is constant with time. The variance should also be constant with time. It is stationary in the covariance if the covariance depends only on the time lag and not the position of the time. A stochastic process is second order stationary when it is stationary in both the mean and the covariance.

If all higher moments are independentof time, the stochastic process is stationary in the strict sense and it is simply referred to as stationary.

If the time series is not stationary the auto-correla and cross-correlation functions do not approach zero at infinit (Wiener, 1949) and so no meaningful information can be got from the functions. A simple technique commonly used to achieve stationarity is differencing and this technique was adopted in this study.

2.2.1. ESTIMATION OF AUTO-CORRELATION FUNCTION

The investigation of the sequencial properties of a series by auto-correlation analysis is already a classical technique. The use of auto-correlations in searching for periodicities is quite old (Craig 1916, Yule 1927 and Walker 19 It is used to determine the linear dependence among the successi values of a time series that is usually characterised by a constant time interval of observation. Though there may be some disadvantages in using this technique in comparison with other techniques the auto-correlation technique is useful in hydrology, provided it is applied appropriately. There is nothi unusual in using the linear correlation of the lagged values of a series although the non-linear correlation may also be used. However, the simplicity of the linear approach, the already developed estimation techniques and distributions of the estimate parameters for this approach, the lack of physical justification for non-linear correlation and the limitations of sample sizes give sufficient support for the use in the hydrology of the linear auto-correlation (Yevjevich, 1972).

Let $x_1, x_2, x_3, \ldots\ldots\ldots, x_N$, be a discrete time series of N points which is stationary with mean zero, then the auto-correlation function estimate is given by:

$$R_x(j) = \frac{\sum_{t=j+1}^{T} x(t) \cdot x(t-j)}{\sum x_t^2} \qquad \text{---------} \quad (6)$$

$$j = 0, 1, 2, \ldots\ldots\ldots\ldots m$$

where j is the lag number and m is the maximum lag number chosen which does not normaly exceed N/4 (Chatfield, 1975). In this study, the sampling interval is unity (one day).

The autocorrelation function has the following general properties:

(i) $R_x(o) = 1$ (follows from the definition)

(ii) $|R_x(j)| \leq 1$ for all j

(iii) $R_x(-j) = R_x(j)$ for all j when x(j) is a real valued function.

Thus, an *autocorrelation* function is symmetric about the origin and attains it's maximum value of unity at the origin.

The autocorrelation measures the serial dependence between observations j time units apart. Such a correlation coefficient might be used to test the null hypothesis of independence against the alternative that there is dependence between observations j time units apart. In most situations however, if there is dependence between observations j time units apart, it would be assumed that there is dependence between observations that are

1, 2, ... and j-1 time units apart. The j^{th} order autocorrelation would customarily be used to test the null hypothesis of the j^{th}-order dependence (Anderson, 1971).

If x_1, x_2, ... x_N are independent and identically distributed as random variables with arbitrary mean, it can be shown that:

$$E\left[R_x(j)\right] = -\frac{1}{\sqrt{N}} \qquad (7)$$

and

$$Var\left[R_x(j)\right] = \frac{1}{N} \qquad (8)$$

N = Number of observations

The 95% confidence limits are then approximated by:

$$CL_{95} = -\frac{1}{N} \pm \frac{2}{\sqrt{N}} \qquad (9)$$

When N is large, the above equation is further approximated to:

$$CL_{95} = \pm \frac{2}{\sqrt{N}} \qquad (10)$$

All values of $R_x(j)$ which fall outside these limits are regarded as significantly different from zero at the 5% level.

The values of the autocorrelation at lag of one day, $[R_x(1)]$ provides some information on the characteristics of the time series. Large values of $R_x(1)$ may suggest red noise or spectral peaks at long periods while small values indicate independent series (white noise) i.e. series with only random variations.

The graphical plot of the auto-correlation function $R_x(j)$ against the time lag j is known as the correlogram. The nature of the correlogram can conjecture about the generating process of the time series. For a periodic process, the correlogram is periodic. It decays exponentially for a first order Markov process while it is in form of damped harmonics for mixed processes. In practice the number of observations are generally small and the correlograms may fail to damp out as theoretically expected because the observed auto-correlations are subject to inflation due to sampling errors (Granger and Hatanaka, 1964). The rate at which the auto-correlation decays may then be interpreted as a measure of the memory of the process. It should be noted however that there are processes for which the auto-correlation does not decay to zero as the time lag tends to infinity but instead behaves in a sinusoidal manner. This type of correlogram corresponds to a process which has a sinusoidal memory (Priestley, 1981).

The auto-correlation functions were computed for both the daily areal rainfall and streamflow series and correlograms plotted.

In cases where the series is not purely periodic or when it has more than one cycle the correlogram may fail to show the cycles clearly. Thus in most cases the method of spectral analysis is frequently applied.

2.2.2. ESTIMATION OF SPECTRAL DENSITY FUNCTIONS

Spectral analysis may be used to detect cycles or oscillations in the time series at certain frequencies. The function used in this case is the spectral density function, $D_x(f)$. The cycles appear as peaks in the plot of $D_x(f)$ against frequency (f) and these peaks correspond to frequencies which account for a large percentage of the total variance. The significant cycle periods in the series would then be given by the inverses of the above frequencies.

Spectral density function is the fourier transformation of the autocorrelation function. The autocorrelation function (R_x) and the spectral density function (D_x) are theoretically related to each other in the following form:

$$D_x(f) = \int_{-\infty}^{\infty} R_x(\tau)\, e^{ik\tau}\, d\tau \qquad \text{------------------} \quad (11)$$

where

$K = 2\pi f$

f = frequency

τ = time lag

Since the spectral density function and the autocorrelation function are even functions of their respective arguments, the above equation may be expressed as a one-sided cosine transform in the following form:

$$D_x(f) = \int_0^\infty Rx(\tau) \cos k\tau d\tau \quad \text{--------------- (12)}$$

Expressed in the discrete form, the above equation becomes:

$$D_x(f) = 2h\left[R_x(0) + \sum_{m=1}^{n-1} R_x(mh) \cos\left(\frac{\pi m f}{f_c}\right) + R_x(nh) \cos\left(\frac{\pi n f}{f_c}\right)\right] \quad \text{---------------- (13)}$$

where:

$m = 1, 2, \ldots\ldots\ldots n$

$h = 1$

$f_c = \frac{1}{2}$ (the cutoff frequency)

In order to get a consistent estimate of the spectral density functions ($D_x(f)$), the random variables which are usually superimposed on the cycle peaks in the plot of $D_x(f)$ against f must be removed. This is achieved by the use of windows (weights). In this study the Parzen window was used and it is given by:

$$\left.\begin{aligned} P_m &= 1 - 6\left(\frac{m}{n}\right)^2 + 6\left(\frac{m}{n}\right)^3, \; m=0, 1, \ldots n/2 \\ &= 2\left[1 - \left(\frac{m}{n}\right)\right]^3, \; m = n/2 + 1, \ldots n \\ &= 0, \quad m > n \end{aligned}\right] \quad \text{---- (14)}$$

Where n is the truncation point.

The final smoothed spectral density estimates are given by:

$$D'x\left(\frac{pf_c}{n}\right) = 2h\left[P_o\ R_x\ (0) + \sum_{m=0}^{n-1} P_m R_x(mh)\ \mathrm{Cos}\left(\frac{\pi mP}{n}\right)\right] \text{----} \quad (15)$$

For normalised records, the area beneath the autospectral function is theoretically equal to unity.

In this study, the smoothed spectral estimates were computed and the graphs of D'(f) against 'f' were plotted to examine the cyclical components in the areal rainfall and discharge time series. The statistical significance of the peaks can be tested using the white-noise or red-noise hypothesis depending on the process of the time series concerned. In this study both white and red noise hypotheses were assumed and the chi-square test was used in testing the null hypothesis.

2.2.3. ESTIMATION OF THE CROSS CORRELATION FUNCTION

If $[x(i)]$ and $[y(i)]$ are discrete time series of N points which are stationary with mean zero, then the cross-correlation function estimate for the two normalized series $[x(i)]$ and $[y(i)]$ is given by:

$$R_{xy}\ (j) = \frac{1}{N-j} \sum_{i=1}^{N-j} x(i).\ \ y(i+j) \text{ --------------- } (16)$$

$j = 0, 1, 2, \ldots\ldots, m$

and

$$R_{yx}(j) = \frac{1}{N-j} \sum_{i=1}^{N-j} y(i).\ x(i+j) \text{ ----------------- } (17)$$

$j = 0, 1, 2, \ldots\ldots, m$

The cross-correlation function $R_{xy}(j)$ can be used to describe the intensity of relationship between two time series at various time lags j. If the series $[x(i)]$ and $[y(i)]$ are un-correlated,

$$E\left[R_{xy}(j)\right] = 0 \quad \text{------------------------} \quad (18)$$

and

$$Var\left[R_{xy}(j)\right] = 1/N \quad \text{------------------------} \quad (19)$$

The 95% confidence limits can then be approximated by:

$$CL_{95} = -1/N \pm \frac{2}{\sqrt{N}} \quad \text{------------------------} \quad (20)$$

A plot of the cross-correlation function $[R_{xy}(j)]$ against time lag (j) gives the nature of the relationship between the two series and a peak in this plot at any lag j may indicate that one series is related to the other significantly when delayed by time j units. The cross correlation function is also required in order to subject the two time series to cross-spectral analysis.

In this study, the cross-correlation functions were computed for each pair of daily areal rainfall and streamflow series and the cross-correlograms were plotted.

2.2.4. <u>CROSS-SPECTRAL ANALYSIS</u>

Cross-spectral analysis may be used for examining the relationship between two time series. The theory of spectral analysis for a single time series can be extended to multivariate time series. In this particular case of bivariate time series, the fourier transformation is carried out on the cross-correlation function. The theoretical relationship between the cross-correlation function (R_{xy}) and the cross-spectral density function (D_{xy}) may be given in the form:

$$D_{xy}(f) = 2 \sum_{k=-\infty}^{\infty} R_{xy}(\tau) e^{-ik\tau} \quad \text{----------------} \quad (21)$$

Equation (21) may be split into two parts:

$$D_{xy}(f) = C_{xy}(f) - iQ_{xy}(f) \quad \text{-----------------} \quad (22)$$

where the real part $C_{xy}(f)$ is called the coincident spectral density function (cospectrum) and the imaginary part $Q_{xy}(f)$ is the quadrature spectral density function.

The co-spectrum represents the in-phase covariance and the quadrature represents the out of phase covariance between x(i) and y(i) at any given frequency.

The phase function (phase angle) $\theta_{xy}(f)$ which describes how the two time series lead or lag each other at any given frequency f may be expressed by:

$$\theta_{xy}(f) = \tan^{-1} \frac{Q_{xy}(f)}{C_{xy}(f)} \quad \text{-----------------} \quad (23)$$

The coherence function which is the measure of the degree of relationship (correlation) between x(i) and y(i) over a range of frequencies is given by:

$$COH^2_{xy}\ (f) \;=\; \frac{D^2_{xy}\ (f)}{D_x(f)\ D_y(f)} \qquad \text{------}(24)$$

$$= \frac{C_{xy}{}^2\ (f)\ +\ Q^2_{xy}\ (f)}{D_x(f)\,.\ D_y(f)}$$

This should theoretically satisfy:

$-1 \leq Coh^2\ xy \leq 1$ for all frequencies.

In order to evaluate the discrete approximations of C_{xy} (f) and $Q_{xy}(f)$, it is necessary to first evaluate the two intermediate functions A_{xy} (mh) and B_{xy} (mh) which are given as:

$$A_{xy}\ (mh) = \tfrac{1}{2}\left[\ R_{xy}(mh)\ +\ R_{yx}(mh)\ \right] \qquad \text{------------}(25)$$

and

$$B_{xy}(mh) = \tfrac{1}{2}\left[R_{xy}(mh)\ -\ R_{yx}(mh)\right] \qquad \text{-------------}(26)$$

Then

$$C_{xy}(f) = 2h\left[\ A_{xy}(0)\ +\ 2\sum_{m=1}^{n-1} A_{xy}(mh)\ \mathrm{Cos}\left(\frac{\pi\ mf}{f_c}\right)\right.$$

$$\left.+\ A_{xy}(nh)\ \mathrm{Cos}\ \left(\frac{\pi\, nf}{f_c}\right)\right] \qquad \text{--------------}(27)$$

and

$$Q_{xy}(f) = 2h\left[2\sum_{m=1}^{n-1} B_{xy}(mh)\ \mathrm{Sin}\ \left(\frac{\pi mf}{f_c}\right)\right.$$

$$\left.+\ B_{xy}\ (nh)\ \mathrm{Sin}\ \left(\frac{\pi\, nf}{f_c}\right)\right] \qquad \text{---------}(28)$$

As in the case of auto-spectral analysis, Parzen window (P_m) as described in Section 2.2.2. was used in smoothing the above cross spectral estimates.

In this section all the spectral functions described above were estimated and the corresponding graphs were plotted.

After the study of the statistical properties of the daily areal rainfall and streamflow time series, the rainfall-runoff mechanisms were examined.

2.3. RAINFALL RUNOFF MODELS

Due to the complexity of the various physical phenomena involved in the natural catchments and also due to scarcity of data, rainfall runoff systems are often approximated by a linear system. This means that the internal details of the physical mechanisms are not examined, but that the catchment is treated as a lumped 'black box' system converting input into output. The main input into the system is rainfall and the main output is streamflow.

Hydrological systems are in general time lag systems. This means that the output of these systems at each instant of time depends on the input during some time before that instant.

Since it is difficult to take into account all details involved in the rainfall-runoff mechanism, it is necessary to accept certain simplifying assumptions on the basis of which the prediction model for the output can be constructed. The input and output components may be considered in a lumped fashion. That is rainfall input is represented by the average areal rainfall as measured at certain locations (rainfall gauging stations) and streamflow output is measured at the outlet of the catchment. Then the process of rainfall - runoff may be considered as a linear system as in the case of the Wiener-Hopf approach.

2.3.1. WIENER-HOPF EQUATIONS

In the Wiener-Hopf theory, it is assumed that the input into the system is stationary, that is, it's statistical properties do not vary with time. So the linear input-output model, based on this Wiener-Hopf approach, represents the simplest possible assumption of a causal relationship between an input x(t) and an output y(t) both expressed as functions of time. It implies that:

(i) the effect of a particular input component will be the same irrespective of when it occurs and irrespective of the state of the input or output at the time of occurance (time invariance).

and

(ii) the effect is propotional to the cause. The relationship or the operation of transforming the input into the output is implied uniquely by any corresponding input - output pair and in principle at least, this relationship can be obstructed and used to find the output for any arbitrary input or the input corresponding to any given output. (linearlity and stability)

Linear systems have a remarkable property that given any input-output record, the operation may be obtained analytically without any further assumptions. If the data are without errors, the expression of the operation will similarly be error free. Even if the data are subject to random errors and the system only approximately linear, an estimate of the operation can still be found without further assumptions. The reliability of the estimate depends on the extent of the sample data on the errors of those data and on the extent and nature of the deperture of the actual operation from the linear assumption. The method of analysis best suited to a particular case depends on the nature of the data, mainly on whether continuous or discrete, periodic or non-periodic and whether capable of expression by analytic functions of time (Nash and Foley, 1982).

The operation of the system on a continuous input function can be expressed as a linear differential equation in which case the property of time invariance will be reflected in the constancy of the coefficients. For a system in which the cause preceeds the effect and in which no input or output occurs prior to time t=0, (physical realizability), the relationship of the input-output impulse response is expressed by the convolution integral equation as:

$$\left.\begin{aligned} y(t) &= \int_0^t x(t)\ h(t-\tau)\ d\tau \\ &= \int_0^t x(t-\tau)\ h(t)\ d\tau \end{aligned}\right] \qquad \text{-------------} \quad (29)$$

where:

$y(t)$ = output

$x(t)$ = input

$h(t)$ = response function.

i.e. The value of the output $y(t)$, is given as a weighted linear sum over the entire history of the input. When the input function is expressed as a series of pulses or mean values over successive short intervals T, the response to a unit pulse of duration T may be a more convenient expression of the operation than the impulse response. The input-output relationship expressed in terms of the pulse response (discrete form) is:

$$y_n = \sum_{r=0}^{m} x(n-r)\ h(r) \qquad \text{---------------------} \quad (30a)$$

or

$$\begin{aligned} y_n &= x(n)\ h(0) + x(n-1)\ h(1) + \ldots\ldots \\ &\quad + x(n-m)\ h(m) \qquad \text{---------------------} \quad (30b) \end{aligned}$$

where it is assumed that the effect of a disturbance in x lasts only through m intervals of duration T. In the presence of data errors or imperfection in the linearlity of the system described by equations (30) above, an error term must be added to give:

$$y(n) = x(n)\ h(0) + x(n-1)\ h(1) + \ldots\ldots$$

$$+ x(n-m)\ h(m) + e_n \qquad \text{---------------} \quad (31)$$

This equation expresses the dependence of y on x as a linear regression in which the successive terms of 'y' constitute the values of the dependent variable and the corresponding values of 'x' and (n-1) previous values are the 'n' independent variables.

In a true linear system, the unit response function obtained through solution of equation (30) is a unique characterization of the system, which is independent of the particular x(t) and y(t) used in it's determination.

The basic problem is to solve equation (30) for a meaningful h(t) when y(t) and x(t) are related to a system which is not truly linear. The solution may be obtained by the maximum likelihood (optimum linear representation) method.

2.3.1.1. MAXIMUM LIKELIHOOD METHOD

Let

y(i) = observed output

$\hat{y}(i)$ = predicted output

$x(i)$ = observed input

following the development of Wiener (1949), it is required that the predicted output be an exact solution of the convolution equation. i.e.

$$\hat{y}(i) = \sum_{t=1}^{\infty} h(t)\, x(i - t + 1)\, \Delta t \qquad \text{------------ (32)}$$

In this study the time interval Δt is equal to one. The difference between the observed output $y(i)$ and the predicted output $\hat{y}(i)$ is given by:

$$\begin{aligned} d(i) &= y(i) - \hat{y}(i) \\ &= y(i) - \sum_{t=1}^{\infty} h(t)\, x(i-t+1) \end{aligned} \qquad \text{----------- (33)}$$

The integral square error, e, is given by the sum of the squares of these differences.

$$\sum_{t=1}^{\infty} d^2 (i) = e = \sum_{i=1}^{\infty} \left[y(i) - \sum_{t=1}^{\infty} h(t)\, x(i-t+1) \right]^2 \text{---(34)}$$

A stable linear system should give the minimum value of e. This implies that the differential of e must be equal to zero.

i.e.

$$\frac{\delta e}{\delta h(j)} = 0,\ j = 1,2,\ \ldots\ldots,\ \infty \qquad \text{------------- (35)}$$

and this gives:

$$\sum_{i=1}^{\infty} y(i)\, x(i-j+1) = \sum_{i=1}^{\infty} x(i-j+1) \cdot \sum_{t=1}^{\infty} h(t)\, x(i-t+1) \text{ --- (36)}$$

$$\text{for } j \geqslant 1$$

Let i-j+1 = t on the left hand side of equation (36) and also note that $x(\delta)$ and $y(\delta)$ are zero for $\delta \leq 0$.

Then equation 36 becomes

$$\sum_{t=1}^{\infty} x(t)\ y(t+j-1) = \sum_{t=1}^{\infty} h(t) \sum_{i-j+1}^{\infty} x(i-j+1)\ x(i-t+1) \quad \text{---} \quad (3$$

if i-j+1 = v on the right hand side of equation (37), we get:

$$\sum_{t=1}^{\infty} x(t)\ y(t+j-1) = \sum_{t=1}^{\infty} h(t) \sum_{v=1}^{\infty} x(v)\ x(v+j-t) \quad \text{------} \quad (38)$$

The discrete time auto-covariance function is given by definition as:

$$C_{xx}(j) = \sum_{v=1}^{\infty} x(v)\ x(v+j)\ \Delta v, \text{ all integer } j, \quad \text{----------} \quad (39)$$

and the cross-covariance function is defined as:

$$C_{xy}(\delta) = \sum_{t=1}^{\infty} x(t)\ y(t+\delta)\ \Delta t, \text{ all integer } \delta, \quad \text{---------} \quad (40)$$

In this study $\Delta v = \Delta t = 1$, so equation (38) becomes the discrete time form of the Wiener-Hopf equation:

$$C_{xy}(j-1) = \sum_{t=1}^{\infty} h_{opt}(t)\ C_{xx}(j-t) \quad \text{----------------} \quad (41)$$

The solution of equation (41) leads to minimum integral square error. Thus the unit impulse response determined through solving equation (41) is optimum and hence stable and it will subsequently be referred to as $[h_{opt}(t)]$.

Equation (41) is both a necessary and sufficient condition for defining the optimum linear system H_{opt} (t) as long as C_{xx} (δ) is neither zero nor constant.

When the theory of minimum mean square error linear prediction is extended to the case of a finite observation period T, equation (41) becomes:

$$C_{xy}\ (j-1) = \sum_{t=1}^{T} h_{opt}(t)\ C_{xx}\ (j-t) \quad \text{-------------} \quad (42)$$

In the matrix form, equation (42) becomes:

$$[\ C_{xy}] = [C_{xx}]\ [h_{opt}] \quad \text{-------------------} \quad (43)$$

where

$$[C_{xy}] = \begin{bmatrix} C_{xy}\ (0) \\ C_{xy}\ (1) \\ C_{xy}\ (2) \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ C_{xy}\ (k) \end{bmatrix}, \quad [h_{opt}] = \begin{bmatrix} h_{opt}\ (0) \\ h_{opt}\ (1) \\ h_{opt}\ (2) \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ h_{opt}\ (m) \end{bmatrix}. \quad \text{-----} \quad (44)$$

$$C_{xx} = \begin{bmatrix} C_{xx}(0) & C_{xx}(1) & C_{xx}(2) & \cdots\cdots & C_{xx}(j) \\ C_{xx}(1) & C_{xx}(0) & C_{xx}(1) & & C_{xx}(j-1) \\ C_{xx}(2) & C_{xx}(1) & C_{xx}(0) & & C_{xx}(j-2) \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ C_{xx}(j) & C_{xx}(j-1) & C_{xx}(j-2) & & C_{xx}(0) \end{bmatrix} \quad \text{--- (45)}$$

The symbols j,k, and m stand for the numbers of lags of the auto-covariance function, the cross-covariance function and the period of the response function respectively. In equation (45) the property of the auto-covariance function is used. It is always an even function with a maximum at time lag zero.

Since the auto-covariance matrix is generally a square matrix and the lags of the auto-convariance and cross covariance are equal, equation (43) becomes:

$$[h_{opt}] = [C_{xx}]^{-1} [C_{xy}] \quad \cdots\cdots\cdots\cdots\cdots\cdots \quad (46)$$

In this study daily areal rainfall time series were generated using the daily records of two three month seasons. Season one extending from April to June (91 days) and season two extending from August to October

(92 days) for the Yala river catchment. The daily areal rainfall was computed using the isopercentile method.

For each pair of time series (rainfall and discharge), the auto-covariances (C_{xx}) and the cross-covariances (C_{xy}) were computed for various time lags. More than one maximum time lags were used in the computation of the auto and cross covariances but the largest was 55 days. It was assumed that the rainfall after 55 days would no longer affect streamflow.

The computed values of the auto and cross convariances were then arranged to form the matrix equations of equation (45). The system of the simultaneous equations was then solved to obtain $h_{opt}(t)$ by using the computer package F4 ACSL (ICL1900). Then for each season the average response function was evaluated.

Finally streamflow was estimated for each season and for two years using the computed areal rainfall for the catchment and the average response function with the help of the following equation:

$$\hat{y}(i) = \sum_{t=0}^{m} h(i-t)\, x(t) \qquad \text{------------------} \quad (47)$$

$$i = 0, 1, 2, \ldots m$$

where

$\hat{y}(i)$ = estimated discharge

h(i) = response function

x(t) = computed areal rainfall

2.3.2. MULTIVARIATE AUTOREGRESSIVE MODEL APPROACH

In this approach which is stochastic in nature, a three stage cycle is often followed namely, the identification of the rainfall-runoff relationship. At this stage the initial hypothesis as to the order of the model is given. The second stage is the estimation of the un-known model parameters of the identified model. Finally the adequacy of the model is tested through diagonostic checking. If the identified model is not statistically adequate, the cycle is repeated until an adequate one is obtained. Once a model is found to be statistically adequate, it can then be adopted for forecasting.

A similar study has already been carried out by Mutulu (1984) for Yala catchment and for the same seasons as used in this study but for a different period of years.

For rainfall-runoff model identification, he used the prewhitening method described by Box and Jenkins (1970). In this method he used a prewhitening filter for the rainfall series to transform both the rainfall and discharge time series. From the analysis of the auto-correlation, partial auto-correlation and the spectral density functions of the rainfall series $[x_t]$ he adopted the first order auto-regressive (Markov) process as representative of the series.

This model is of the form:

$$x(t) = R(1)\ x(t-1) + \alpha(t) \quad \text{--------------} \quad (48)$$

where:

$x(t)$ = computed areal rainfall on day t

$\alpha(t)$ = random disturbance

$R(1)$ = lag-one autoregressive coefficient (auto-correlation) and it is taken as the maximum likelihood parameter of the model (Yule, 1927 and Walker, 1931)

From the above model, the corresponding prewhitening filter for the rainfall is:

$$\alpha(t) = x(t) - R(1)\ x(t-1) \quad \text{--------------} \quad (49)$$

similarly for discharge $[y(t)]$, the corresponding transformation filter is:

$$\beta(t) = y(t) - R(1)\ y(t-1) \quad \text{--------} \quad (50)$$

Note that R(1) varies from one period of study to another. So using the filter of equations (49) and (50). He transformed both the rainfall and stream flow series. He then subjected the transformed series to auto-correlation and spectral analyses and the results showed that the rainfall series had been converted into white noise (independent series) and the discharge series was also transformed. These results justified the use of the above filters.

Mutulu (1984) further found in his study that the second or third order distributed lag effects from rainfall were present in the generating models of the Yala river basin system and that second order auto-regressive dynamics were operative in the Yala river basin. He therefore introduced three tentative transfer function models of different orders namely:

(i) TF(2, 2.): $y_t = a_1 y_{t-1} + a_2 y_{t-2} + b_o x_t + b_1 x_{t-1} + \delta_t$ ------- (51)

(ii) TF(3, 2): $y_t = a_1 y_{t-1} + a_2 y_{t-2} + b_o x_t + b_1 x_{t-1} + b_2 x_{t-2} + \gamma_t$ --- (52)

(iii) TF(4, 2.): $y_t = a_1 y_{t-1} + a_2 y_{t-2} + b_o x_t + b_1 x_{t-1} + b_2 x_{t-2} + b_3 x_{t-3} + \eta_t$ -------- (53)

where

x = Rainfall

y = discharge

δ_t, γ_t and η_t are random disturbances assumed to be normally distributed and with zero mean. The notation

TF(q, p) is used above and subsequently to denote a transfer function with distributed lag component of order 'q' and autoregressive component of order p.

In this study, the above models were recalibrated for the long term means. The parameters were determined by the maximum likelihood method. The significance of the parameters was tested using the students t-test against the null hypothesis with (N-S) degrees of freedom. S is the number of parameters and N is the total number of observations.

The mean models fitted above were lastly used to forecast discharge during the verification period.

The skill of the models fitted in the above two methods in forecasting streamflow was tested using the multiple correlation coefficient (R^2) and the Smirnov-Kolmogorov tests.

The various tests which can be used to test for goodness of fit are briefly reviewed in the following section.

2.4. TESTING OF GOODNESS OF FIT

Various statistical tests are available to test the goodness of fit. In general the procedure used in applying these tests is:

(a) State the null-hypothesis (H_o) of goodness of fit.

(b) Choose a statistical test.

(c) Select a significance level.

(d) Assume the sampling distribution of the statistical test under H_o.

(e) On the basis of b, c, and d, define the region of rejection.

(f) Compute the values of the statistical test. If the value is in the region of rejection, reject H_o otherwise accept it.

There are two types of statistical tests, namely, parametric and non-parametric. Parametric tests are associated with certain assumptions about the sample being tested. These assumptions are:

(a) The observations must be independent.

(b) The observations must be drawn from normally distributed populations.

(c) These populations must have the same variance.

Generally, non-parametric tests need only assumption (a) and in addition, non-parametric tests do not require as high a level of data measurement as do parametric tests.

The most commonly used tests are:

(a) Multiple correlation coefficient (R^2)

(b) The students t-test.

(c) The chi-square test.

(d) Durbin-watson test

(e) The Portmanteau lack of fit test.

(f) The Smirnov-Kolmogorov test.

2.4.1. THE MULTIPLE CORRELATION COEFFICIENT TEST (R)

The multiple correlation coefficient (R) is the measure of linear association of many variables, including one dependent and the others mutually independent. It is expressed by the ratio of the standard deviation of the estimated values to the standard deviation of the observed values of the dependent variable. This multiple correlation coefficient R, measures the combined influence of all independent variables in terms of the difference between the observed and the estimated values of the dependent variable. It may be expressed as:

$$R = \frac{S_e}{S} = \sqrt{1 - \frac{P^2}{S^2}} \qquad \text{.........} \quad (54)$$

$$= \sqrt{\frac{b_2 \Sigma\Delta x_1 \Delta x_2 + b_3 \Sigma\Delta x_1 \Delta x_3 + \ldots \Delta b_m \Sigma\Delta x_1 \Delta x_m}{\Sigma(\Delta x_1)^2}}$$

where:

S_e = standard deviation of estimated values

S = Standard deviation of observed values

P = Standard deviation of residuals.

The multiple correlation coefficient is a dimensionless measure of linear association while the standard deviation is a measure having the dimension of the variable x_1. Because 'S' is always the same regardless of the number of variables, a decrease of the standard deviation of residuals (P) will cause an increase of the multiple correlation coefficient .

The multiple correlation coefficient can also be defined as the correlation coefficient of linear association of two variables, one being the observed values and the other being the estimated values of the dependent variable based on all the other independent variables.

The square of the multiple correlation coefficient (R^2) is referred to as the coefficient of multiple determination and ($1 - R^2$) is the coefficient of multiple non-determination.

The coefficient of multiple determination (R^2) indicates the percentage of the variance of the dependent variable (S^2) that has been accounted for mathematically by the multiple linear correlation of the variable x_1 on $x_2, \ldots\ldots, x_m$.

For use as a goodness of fit test, the coefficient of multiple determination may be expressed as:

$$R^2 = \frac{[A]^T\ [y]^T\ [y] - N(\bar{y})^2}{[y]^T\ [y] - N(\bar{y})^2} \qquad \text{------------------ (55)}$$

where

$$[y] = \begin{bmatrix} y_o \\ y_1 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ y_N \end{bmatrix} \qquad \text{---------------- (56)}$$

$$[V] = \begin{bmatrix} -y_1 & -y_2 \cdots\cdots & -y_{-p} & x_o & x_1 & \cdots\cdots & x_{-q} \\ -y_o & & & & & & \\ \cdot & & & & & & \\ \cdot & & & & & & \\ \cdot & & & & & & \\ \cdot & & & & & & \\ \cdot & & & & & & \\ -y_{N-1} & -y_{N-2} \cdots & -y_{N-p} & x_N & \cdots\cdots & & x_{m-q} \end{bmatrix} \quad \text{----- (57}$$

$$[A] = \begin{bmatrix} a_1 \\ a_2 \\ \cdot \\ \cdot \\ \cdot \\ a_p \\ b_o \\ b_1 \\ \cdot \\ \cdot \\ \cdot \\ b_q \end{bmatrix} \quad \text{---------------- (58)}$$

and

$\bar{y}$ = mean of y - series.

The test of significance is given under the Null hypothesis.

$$H_o: A_1 = \cdots\cdots = A_2 = 0 \qquad \text{-------------- (59)}$$

This assumes that the overall model has no effect on the system output. The 'F' statistic with (s-1, N-S) degrees of freedom can be expressed as:

$$F = \frac{R^2/(S-1)}{(1-R^2)/(N-S)} = \frac{\text{Regression mean square}}{\text{Residual mean square}} \quad \text{------} \quad (60)$$

2.4.2. THE STUDENTS T-TEST

The student's t-test is given by the following statistic:

$$t = \frac{|\bar{x}_1 - \bar{x}_2|}{\delta\sqrt{1/n_1+1/n_2}} \quad \text{----------------} \quad (61)$$

where

$\bar{x}_1$ and $\bar{x}_2$ are the arithmetic means of the observed and predicted data.

n_1 and n_2 are the sample sizes of observed and predicted data and δ is the unknown population standard deviation estimated from the sample variances S_1^2 and S_2^2 as

$$\delta = \sqrt{\frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2}} \quad \text{---------------} \quad (62)$$

If t is greater than the tabulated value of the students t - distribution at a given significance level α and $n_1 + n_2 - 2$ degrees of freedom then the null-hypothesis of goodness of fit is rejected.

2.4.3. THE CHI-SQUARE TEST

Let a discrete variable have k mutually exclusive ... a continuous variable be divided into k class intervals also mutually exclusive events. Let f_i, i = 1, 2, ... k, represent the relative frequencies of these random events or class intervals for a sample of size N, and let the discrete or continuous distribution functions have the probability mass of random events or the probability of class intevals as p_i, i = 1, 2, .. k. For the observed sample size N, the samples absolute frequencies are $N_i = Nf_i$ in which f_i are the relative frequencies and the expected values of absolute frequencies are NP_i. The chi-square parameter can be represented by the following equation:

$$X^2 = \sum_{i=1}^{k} \frac{(N_i - NP_i)^2}{NP_i} \quad \text{--------------------} \quad (63)$$

The above parameter will have chi-square distribution with the number of degrees of freedom (NDF) equal to K-1, for a sufficiently large N.

The chi-square test prescribes the critical value X_o^2 for a given significance level α so that for $x^2 < x_o^2$ the Null hypothesis of a good fit is accepted, otherwise for $x^2 \geqslant x_o^2$ it is rejected.

In general, the number of class intervals should be $k \geqslant 5$ and the expected values of absolute frequencies should be $NP_i \geqslant 5$ for the distribution with the number of degrees of freedom equal to k-1 to well approximate the distribution of the parameter given in the above equation. If $k < 5$, the expected absolute frequency should be $NP_i > 5$.

2.4.4. THE DURBIN-WATSON TEST

The Durbin-Watson statistic which can be used to test goodness of fit may be defined as follows:

$$D = \sum_{i=1}^{n-1} (\Gamma_{i+1}-\Gamma_i)^2 / \sum_{i=1}^{n} (\Gamma_i - \bar{\Gamma})^2 \quad \text{------------} \quad (64)$$

where $\Gamma_1, \ldots\ldots \Gamma_n$ are the n residuals (i.e. difference between observed and estimated values) and $\bar{\Gamma}$ is the mean of the residuals.

If 'D' is equal or greater than the theoretical value for a given significance level α. The null hypothesis for goodness of fit is rejected.

2.4.5. THE PORTMANTEAU LACK OF FIT TEST

The portmanteau statistic Q which is used to test whether the residuals of the model form independent series may be given as:

$$Q = N \sum_{k-1}^{L} \Gamma_k^2 \quad (e_t) \quad \text{---------------------} \quad (65)$$

where $\Gamma_k(e_t)$ is the autocorrelation of the residual series (e_t) and L the maximum lag considered. Q is approximately distributed as chi-squared with L-S degrees of freedom (x^2 (L-S)). At a given significance level the residual series (e_t) forms an independent series if Q is less than the theoretical value of x^2 (L-S). Hence the model is adequate.

2.4.6. THE SMIRNOV-KOLMOGOROV TEST

If two subsamples are from the same population then the cumulative probability distributions of both subsamples should be fairly close to each other. If the two cumulative distributions are too far apart at any point, then the subsamples may be from different populations. Thus a large enough deviation between the two cumulative distributions is evidence for rejecting H_o.

Let the sample data x_1, x_2, ... x_n be in ascending or descending order x_1', x_2', ... x_n' and the plotting positions $P(x_i') = m/(n+1)$ be determined for these values and the emperical distribution plotted. Also let the true population value for $P(x_i')$ be $F(x)$ and their maximum difference be defined as:

$$D = \max \left| F(x) - P(x_i') \right| \qquad \text{------------------} \quad (66)$$

This statistic 'D' has a sampling distribution. Using the Smirnov-Kolmogrov statistic in the goodness of fit test, cumulative probability distributions for both the observed and computed data are computed and the largest absolute difference between the cumulative distribution functions of the observed and computed values is determined from equation (66) over the entire range of data.

If D exceeds the critical value for a given number of sample points and a given significance level in the Lilliefor's table, the Null hypothesis of goodness of fit is rejected. (YEVJEVICH, 1972)

In this study, the multiple correlation coefficient and the Smirnov-Kolmogorov tests were used to test the goodness of fit of the estimated discharge at the 5% significance level.

Finally the results from the two approaches were compared by computing and comparing their corresponding standard errors.

2.5. COMPUTATION OF STANDARD ERRORS

A measure of variability of the resulting event magnitudes is the standard error of estimate. The standard error of estimate can be defined as:

$$S = \left[\left[\sum_{i=1}^{N} (x_i - x_i')^2 \right] / N \right]^{\frac{1}{2}} \quad \text{.........} \quad (67)$$

where:

x_i = observed value for day 'i'.

x_i' = estimated value for day 'i'

N = number of points in the series

S = standard error of estimate.

In this study, the values of the standard errors of estimate 's' were computed for both methods used in the prediction. The value of these standard errors of estimate were compared and the method which gave the smallest standard error was taken as the best.

The results obtained from the study are discussed in the next chapter.

CHAPTER THREE

3.0. RESULTS AND DISCUSSION

In this chapter the results obtained from the various analyses will be discussed under separate sections. The subsequent time series analyses have been carried out on standardised data. The data were standardised in order to have zero mean and unit variance.

3.1. RESULTS FROM TIME SERIES ANALYSIS

Previous studies showed that the rainfall in Yala river basin is corrupted by a first order (Markov) process, (Mutulu, 1984). So, before carrying out time series analysis both the areal rainfall and streamflow series were transformed using the following filter in order to obtain a random series (Box and Jenkins, 1968)

$$\alpha_t = x_t - \Gamma_1 x_{t-1} \qquad (68)$$

where:

α_t = transformed value for day t

x_t = observed value for day t

Γ_1 = value of the autocorrelation function at lag one.

Note that the value of Γ_1 was changing from one season to another.

3.1.1. AUTO-CORRELATION ANALYSIS

Examples of the results from correlogram analysis are given in figures 5,6,7 and 8 and in appendix A.

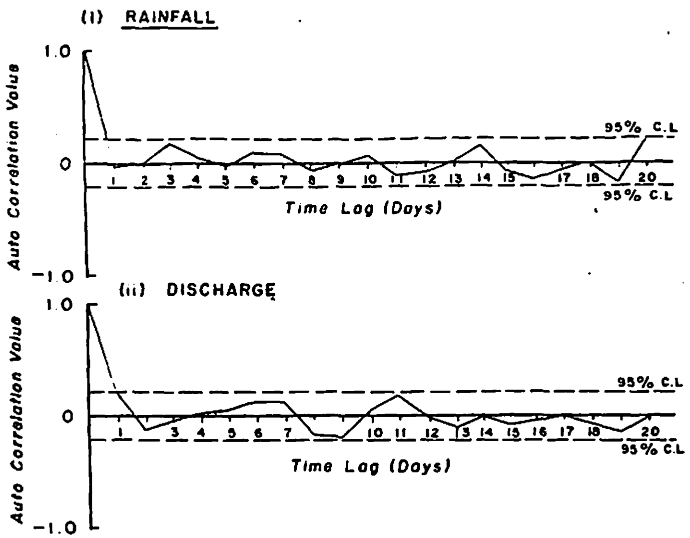


Fig. 5 CORRELOGRAMS FOR SEASON ONE YEAR TWO (1969)

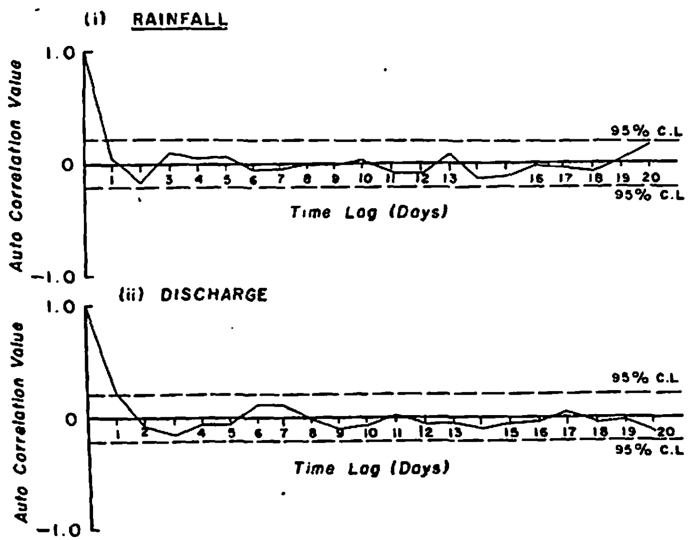


Fig. 6 CORRELOGRAMS FOR SEASON ONE YEAR SEVEN (1974)

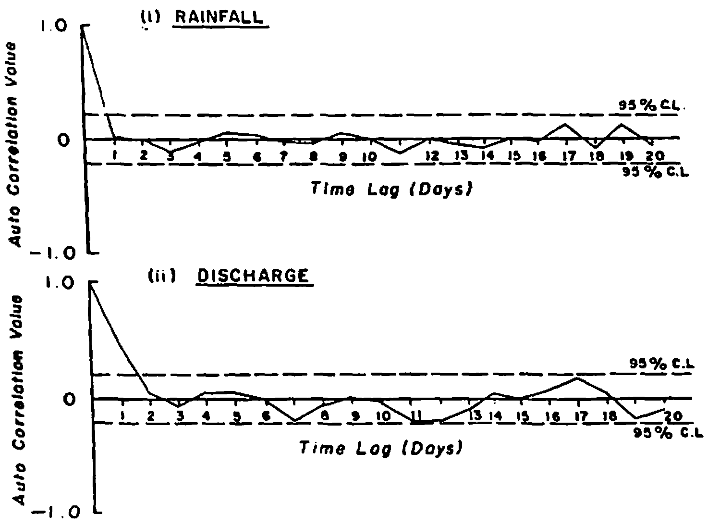


Fig 7 CORRELOGRAMS FOR SEASON TWO YEAR THREE (1970)

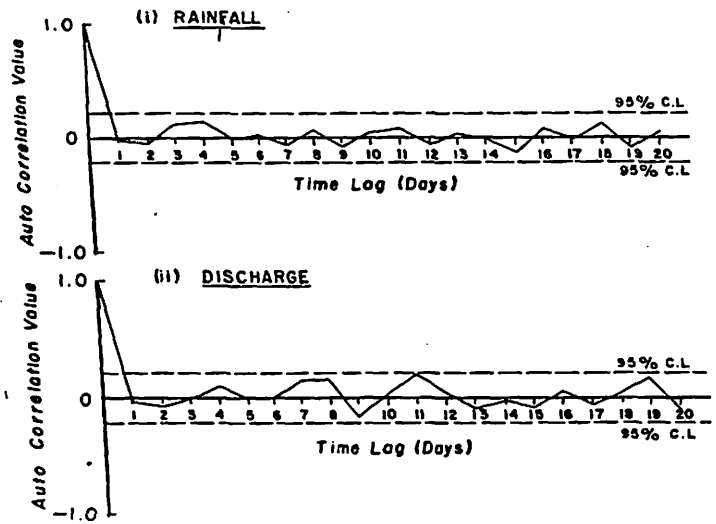


Fig. 8 CORRELOGRAMS FOR SEASON TWO YEAR FOUR (1971)

The statistical significance of the autocorrelation values (R_x) observed in this study were tested as indicated in equation (10). It was observed that at the 5% significance level, (R_x) was significantly different from zero when it was outside the confidence limits of $-0.21 \leq R_x \leq 0.21$.

The following general features were also observed:

(i) The null hypothesis of no significanct correlation at the 5% significance level for the lag one auto-correlation values could be accepted for all the years and in both seasons in the case of the daily areal rainfall series. For the daily discharge series, it could be accepted in only some cases and rejected in others.

This could have resulted from the fact that the transformation filter was used to prewhiten only the daily areal rainfall series.

(ii) At the 5% significance level, all auto-correlation values were not significantly different from zero for both the daily rainfall and discharge records after the first time lag.

(iii) There were no purely harmonic fluctuations in the correlograms. Many correlograms generally failed to damp out to zero at large time lags. This may indicate that the fluctuations in the daily rainfall and discharge series could not show purely periodic

3.1.2. PATTERNS OF THE SPECTRAL AND CO-SPECTRUM ESTIMATES

Some examples of the results obtained from spectral analysis are given in figures 9, 10, 11, and 12 and Appendix B. In general the following features were observed.

(i) The cycles observed in the daily areal rainfall and streamflow series in season one had dominant periods centered around 2 - 4 days and 5 - 7 days.

(ii) The cycles observed in season two had periods ranging between 2 - 4.3 days 5 - 6 days and 8 - 10 days.

(iii) In both seasons, the maximum peaks were centered between 3 - 5 days.

Since the periods observed in the rainfall series were also reflected in the streamflow series, it may be concluded that the Yala river basin has little damping effect on the rainfall. This means that the catchment responds directly to rainfall by rise of flow in the river. This may also indicate that the rainfall - runoff system for the Yala river basin was fairly linear and time invariant during the two seasons considered.

The results from the co-spectrum analysis are illustrated in figures 13 - 16 and Appendix C. In general the fluctuations observed in the cospectrum estimates are similar to those observed in the spectra of rainfall and streamflow. This confirms the

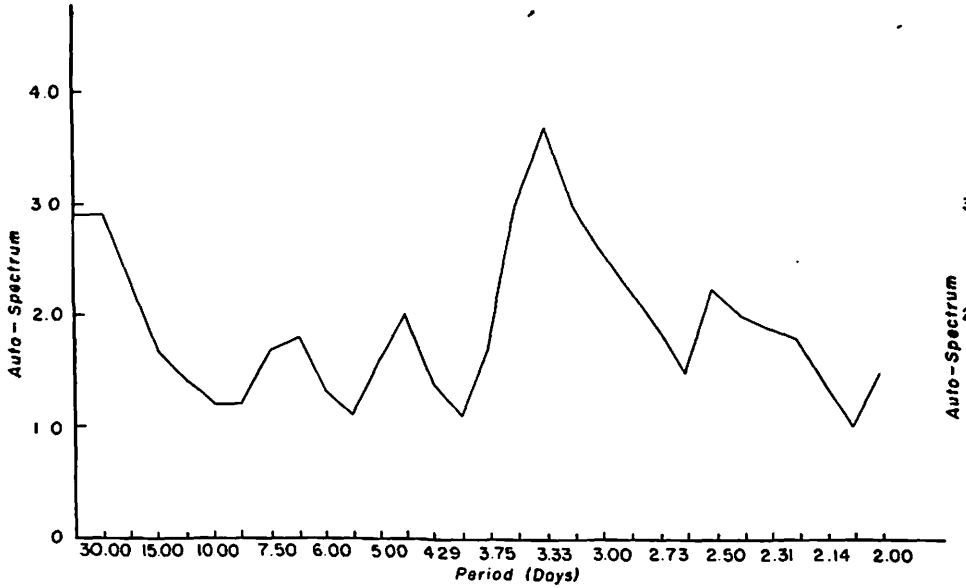


Fig. 9 RAINFALL AUTO-SPECTRUM FOR SEASON ONE IN YEAR TWO (1969)

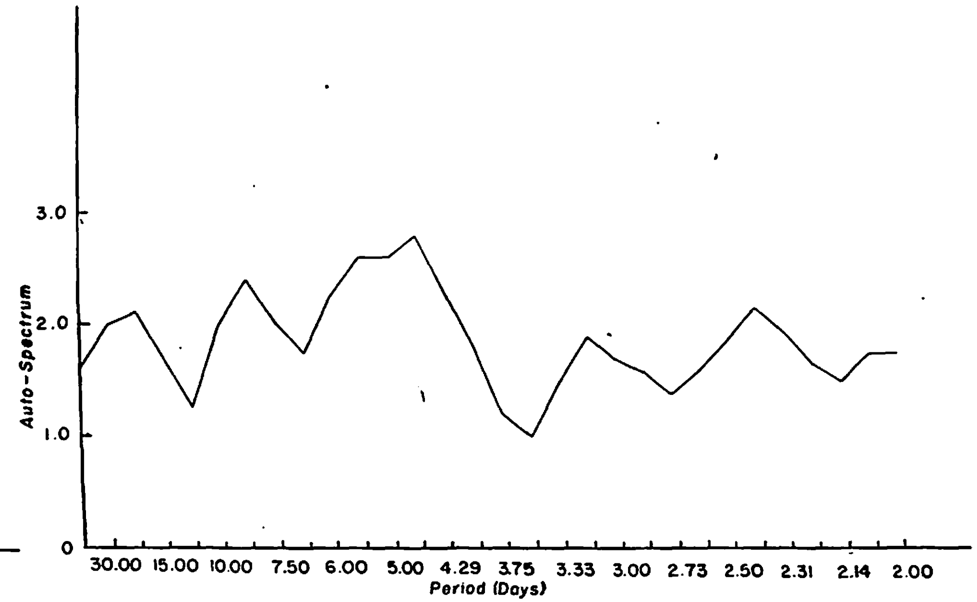


Fig. 10 RAINFALL AUTO-SPECTRUM FOR SEASON TWO IN YEAR THREE (1970)

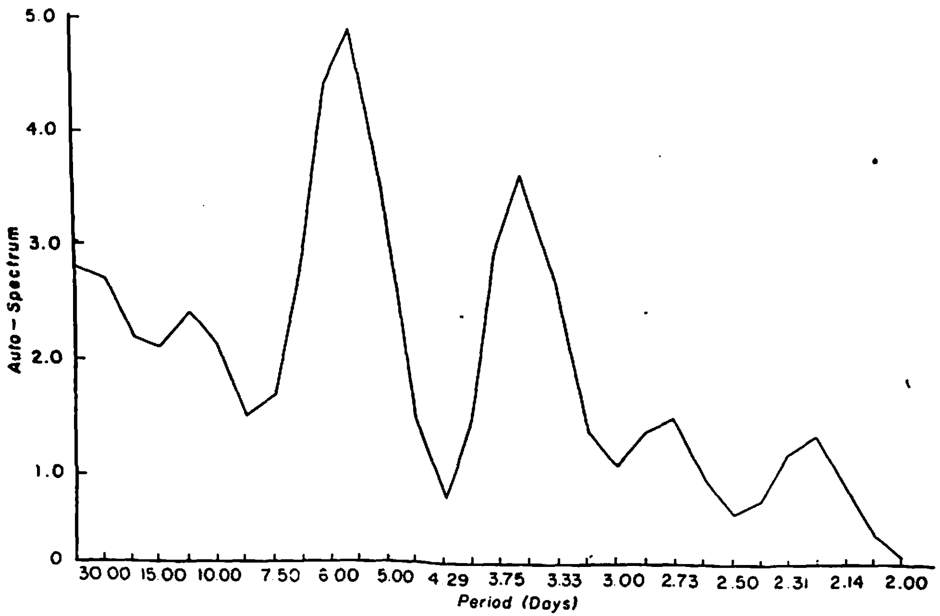


Fig. 11 DISCHARGE AUTO-SPECTRUM FOR SEASON ONE IN YEAR TWO (1969)

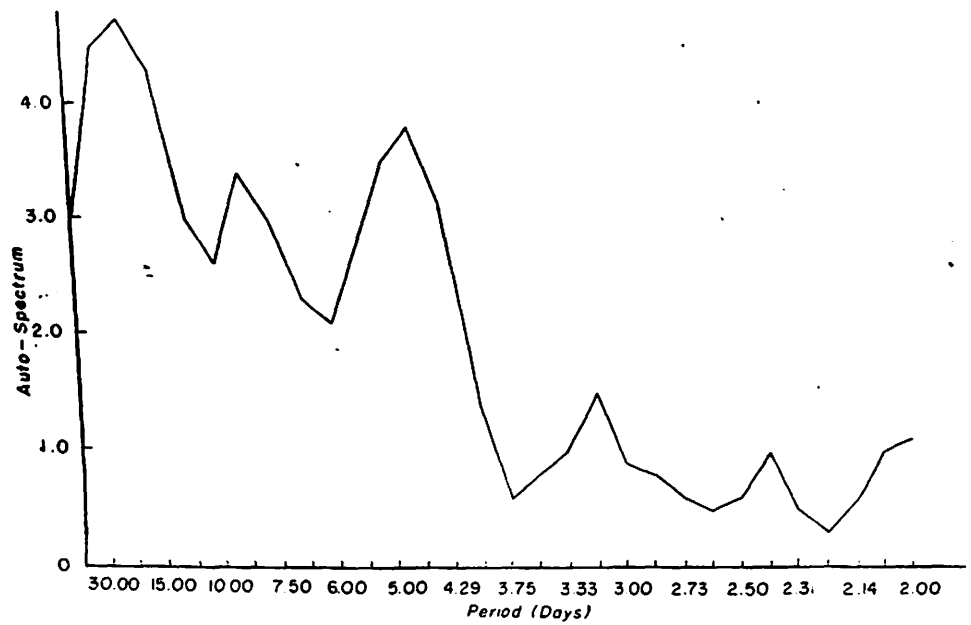


Fig. 12 DISCHARGE AUTO-SPECTRUM FOR SEASON TWO YEAR THREE (1970)

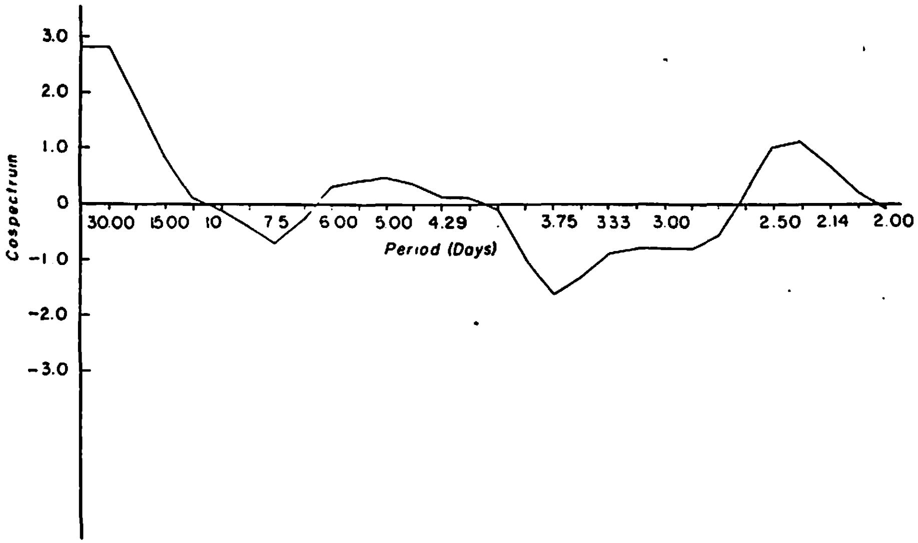


Fig. 13 COSPECTRUM FOR SEASON ONE IN YEAR TWO (1969)

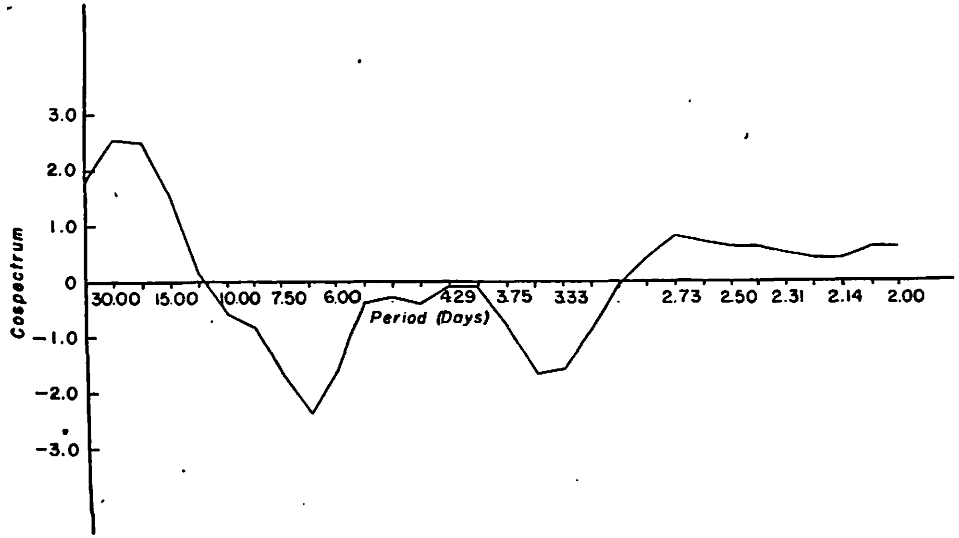


Fig. 14 COSPECTRUM FOR SEASON ONE IN YEAR THREE (1970)

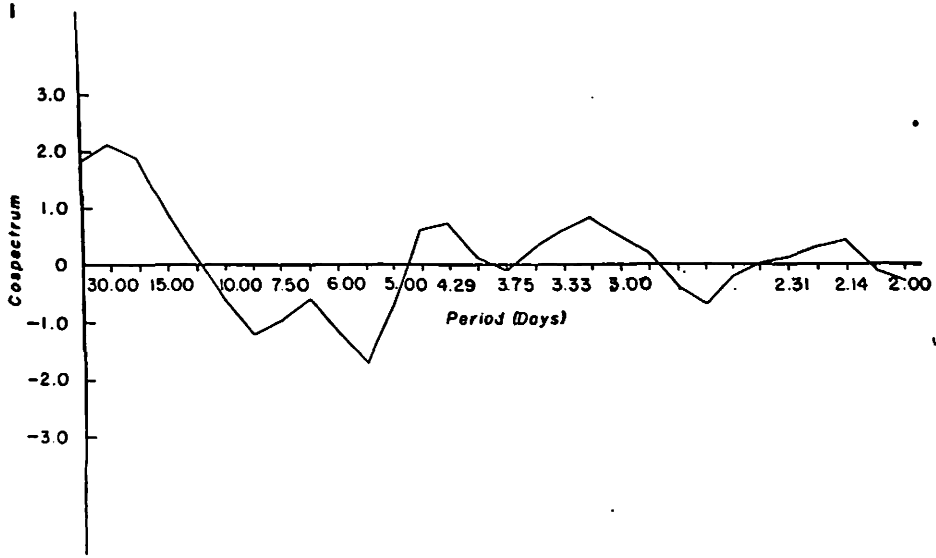


Fig. 15 COSPECTRUM FOR SEASON TWO IN YEAR THREE (1970)

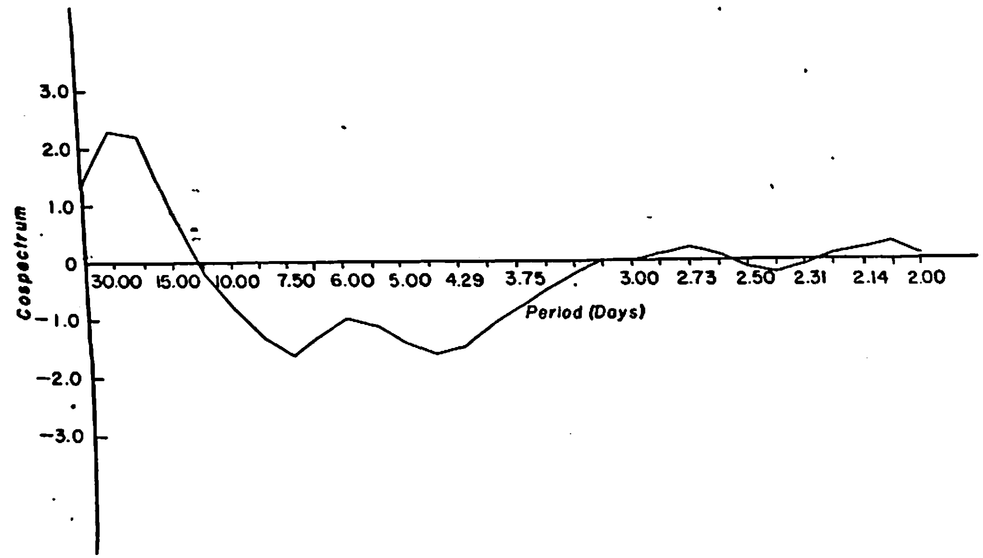


Fig. 16 COSPECTRUM FOR SEASON TWO IN YEAR FIVE (1972)

physical reality of some of the cycles observed under spectral analysis.

3.1.3. CROSS-CORRELATION AND COHERENCE FUNCTIONS

Figures 17 and 18 and Tables 2 show the results from cross-correlation analysis.

The results obtained for the previous sections are also reflected in the cross-correlation and coherence patterns.

TABLE 2a: RESULTS OF SEASON ONE CROSS-CORRELATION ANALYSIS

	SEASON ONE	
YEAR	LAG OF MAX. CR.	VALUE OF MAX. CR
1 (1968)	3 DAYS	0.62
2 (1969)	3 DAYS	0.45
3 (1970)	2 DAYS	0.72
4 (1971)	3 DAYS	0.55
5 (1972)	3 DAYS	0.68
6 (1973)	3 DAYS	0.58
7 (1974)	3 DAYS	0.75

NOTE: CR = CROSS-CORRELATION

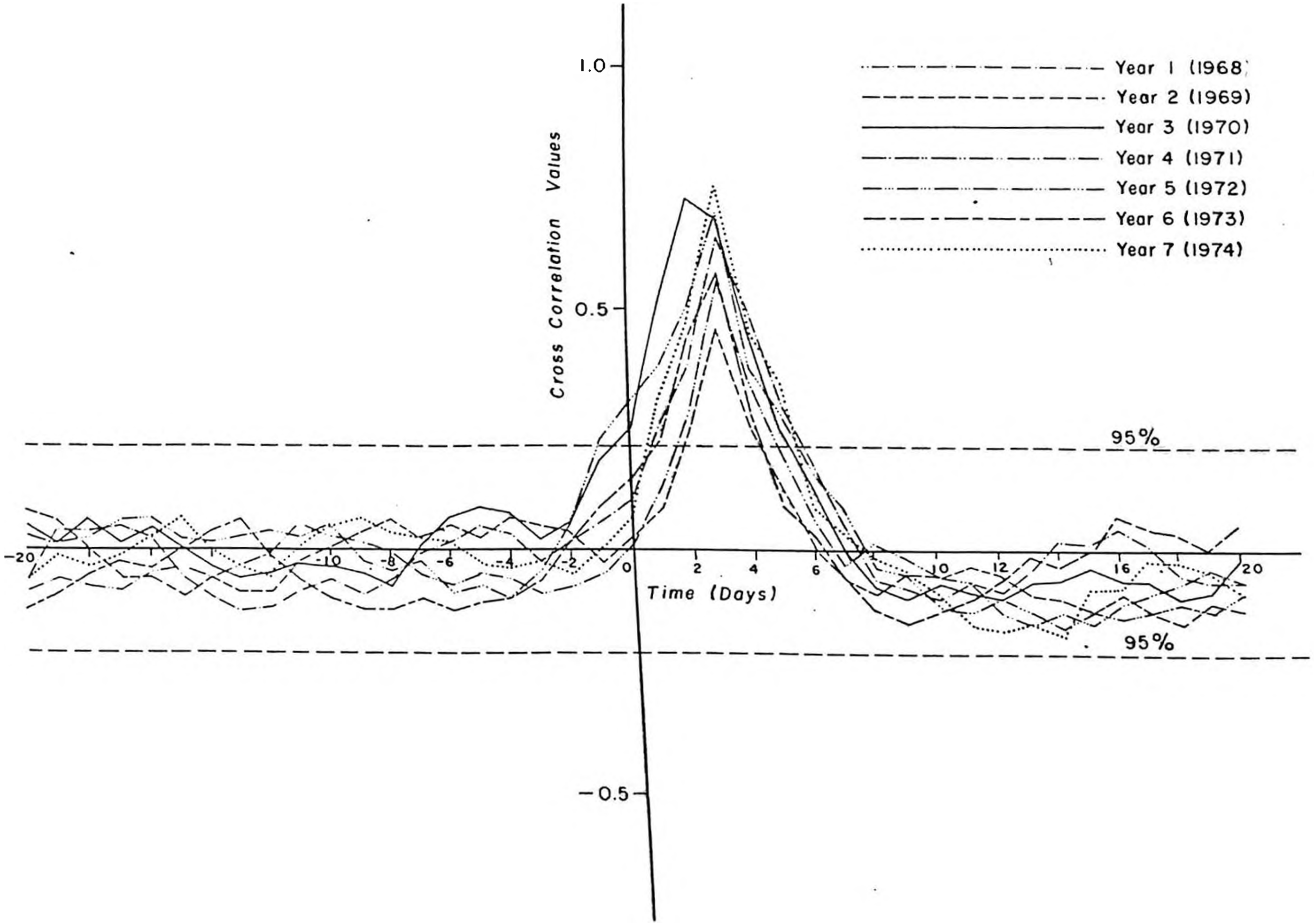


Fig. 17 SEASON ONE CROSS CORRELOGRAMS

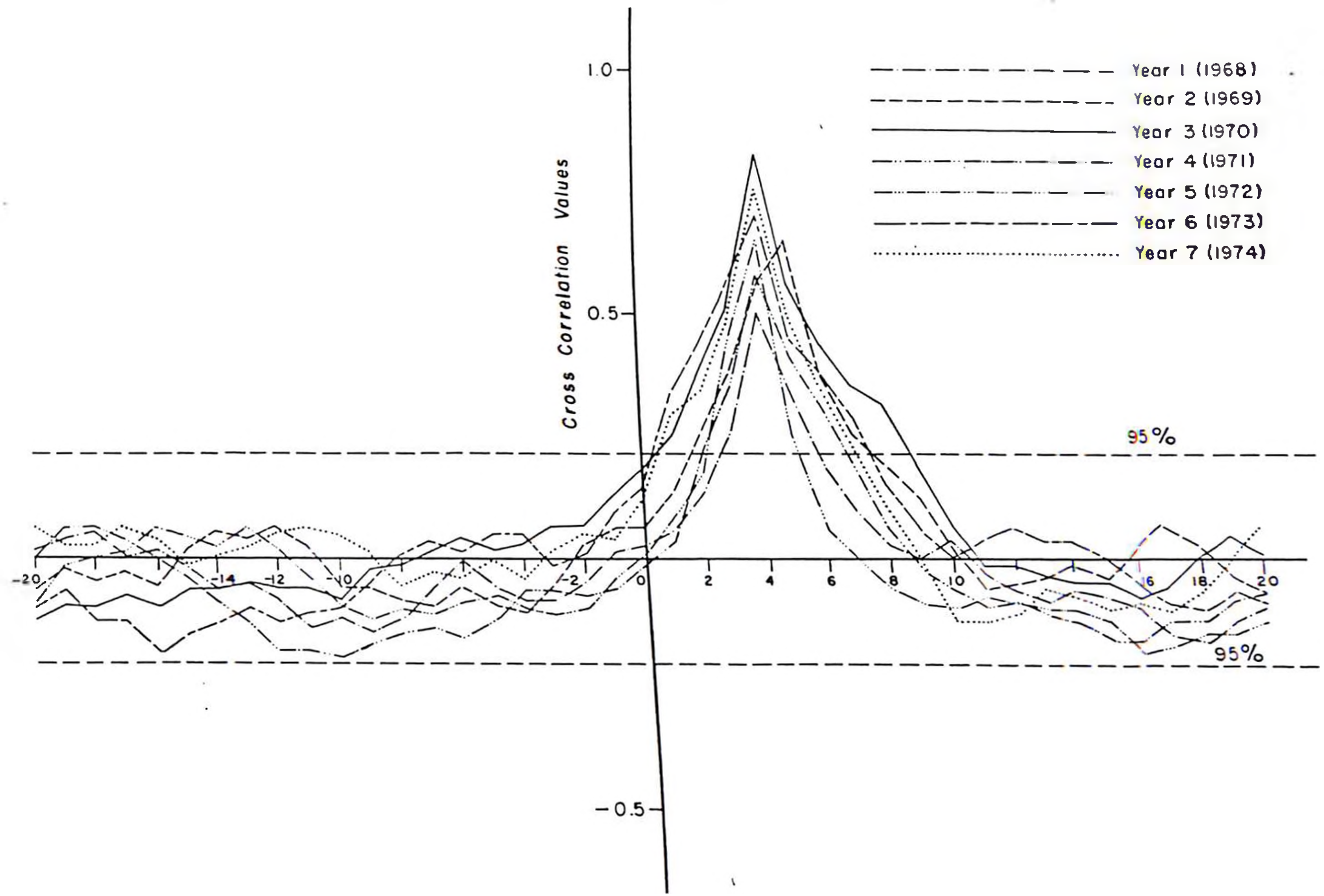


Fig. 18 SEASON TWO CROSS CORRELOGRAMS

TABLE 2b: RESULTS OF SEASON TWO CROSS-CORRELATION ANALYSIS

SEASON TWO		
YEAR	LAG OF MAX. CR	VALUE OF MAX. CR
1 (1968)	4 DAYS	0.50
2 (1969)	5 DAYS	0.52
3 (1970)	4 DAYS	0.83
4 (1971)	4 DAYS	0.58
5 (1972)	4 DAYS	0.65
6 (1973)	4 DAYS	0.70
7 (1974)	4 DAYS	0.75

NOTE CR = CROSS-CORRELATION

It can further be observed from these figures that:

(i) Rainfall and streamflow were significantly correlated between the time lags of two to four days during season one and between three to five days in season two.

(ii) During season one, the rainfall - streamflow cross-correlations were generally maximum at a lag of three days with peak lag correlation values ranging between 0.45 to 0.75.

(iii) For season two, the cross-correlation values were generally maximum at a lag of four days with peak lag correlation values ranging between 0.50 to 0.83.

Since the time lags for maximum cross-correlation values were constant in both seasons, we may assume that the rainfall - runoff system for the Yala river catchment may be approximated by a linear system.

Figures 19 to 22 and Appendix D show the results obtained from the coherence analysis.

Highest coherence generally corresponded to the periods observed under spectral analysis. This indicates highest relationship between rainfall and streamflow at those period ranges.

The linearity of the system with statistically stationary input may, also be inferred from the coherence function values. If the system is truely linear, the values of the coherence approach unity.

The values of the coherence function obtained in this study were very high (close to one) as can be seen from the Figures 19 to 22. It can therefore be said that it is reasonable enough to assume a linear approach in the estimation of streamflow in this study.

3.1.4. QUADRATURE SPECTRUM AND PHASE ANGLE

The results obtained from the quadrature spectrum analysis are illustrated in Figures 23 to 26 and Appendix E.

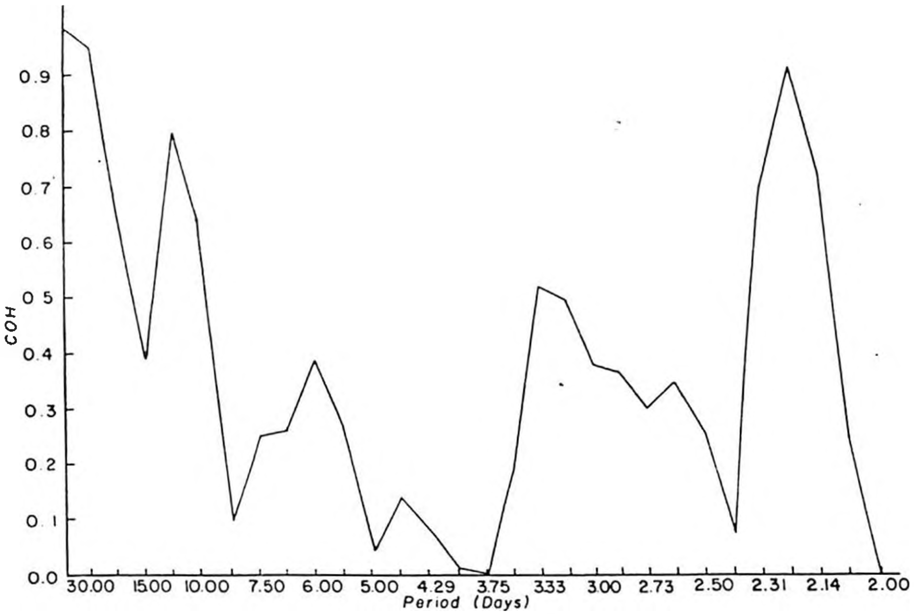


Fig. 19 COHERENCE FUNCTION FOR SEASON ONE IN YEAR TWO (1969)

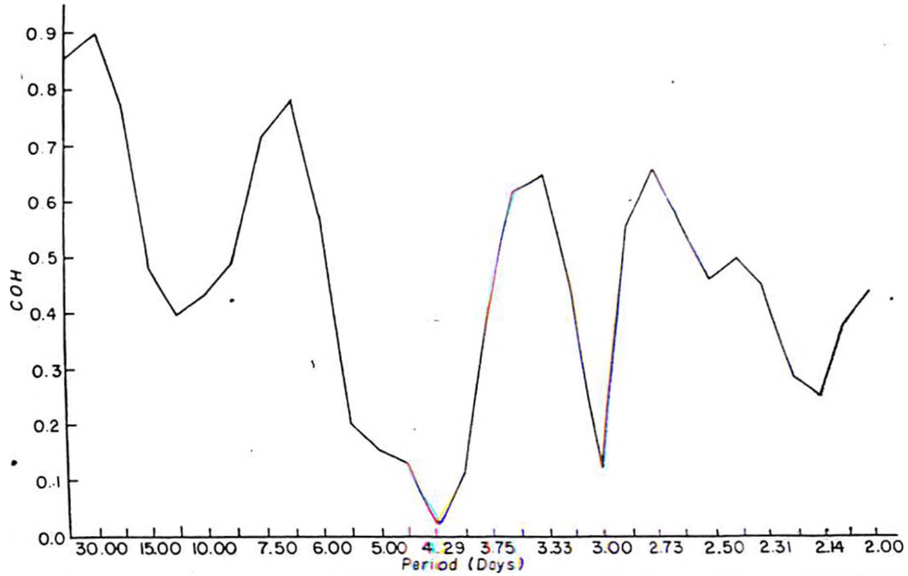


Fig.20 COHERENCE FUNCTION FOR SEASON ONE IN YEAR THREE (1970)

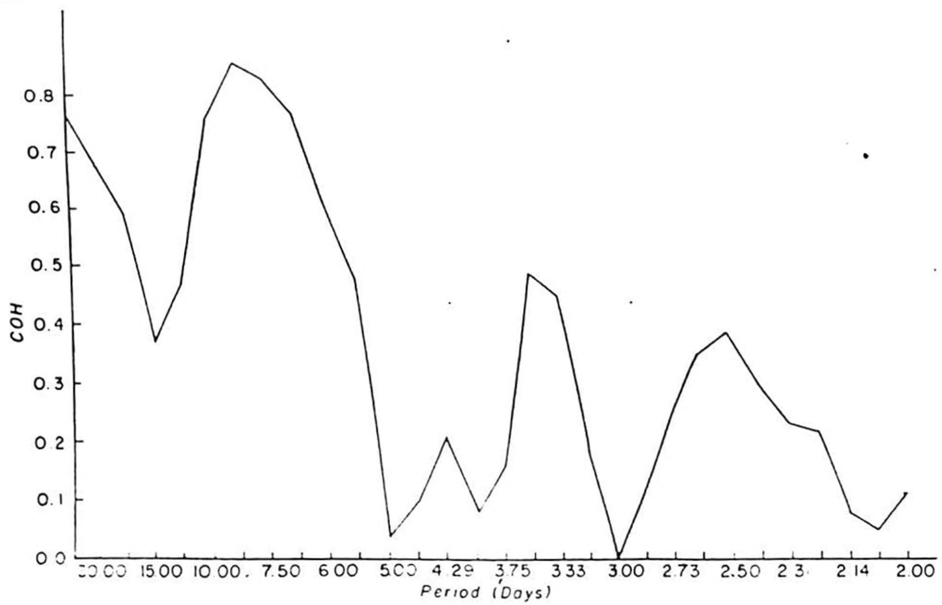


Fig 21 COHERENCE FUNCTION FOR SEASON TWO IN YEAR ONE (1968)

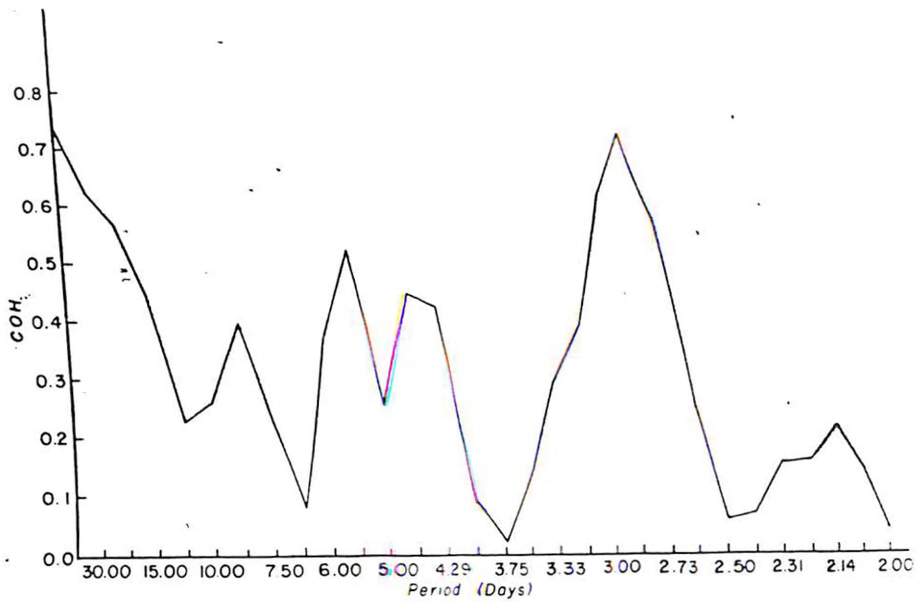


Fig 22 COHERENCE FUNCTION FOR SEASON TWO IN YEAR THREE (1970)

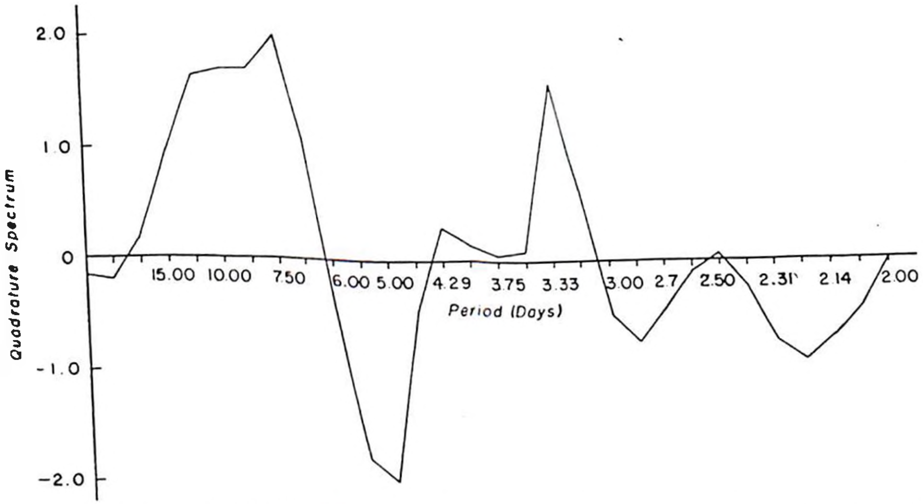


Fig. 23 QUADRATURE SPECTRUM IFOR SEASON ONE IN YEAR TWO (1969)

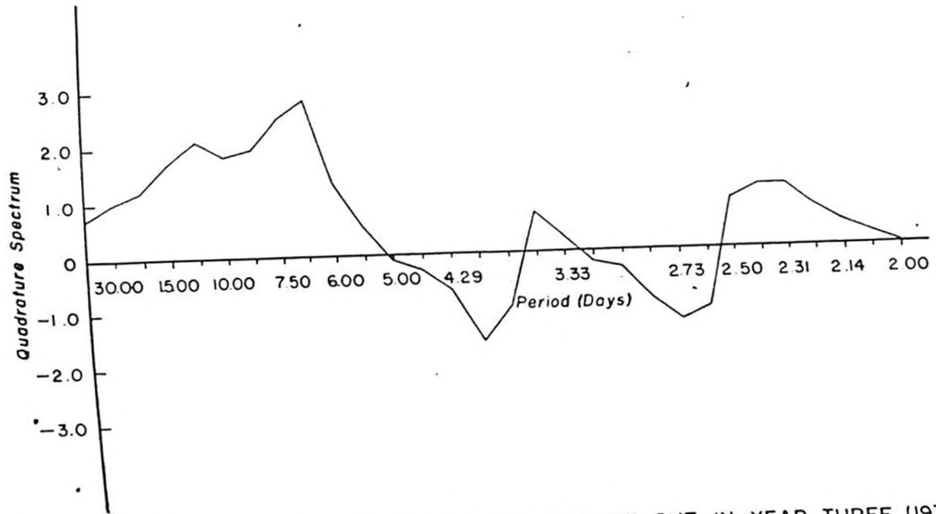


Fig. 24 QUADRATURE SPECTRUM FOR SEASON ONE IN YEAR THREE (1970)

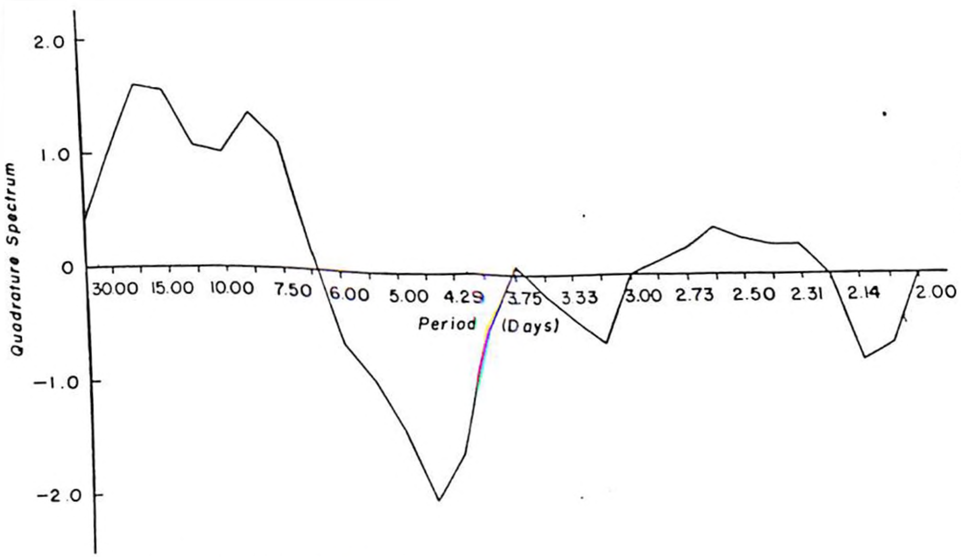


Fig. 25 QUADRATURE SPECTRUM FOR SEASON TWO IN YEAR THREE (1970)

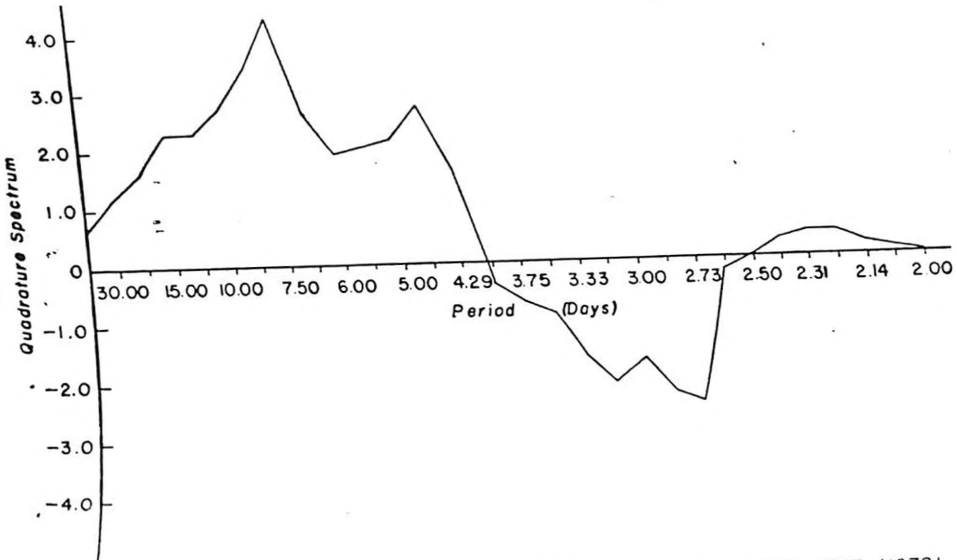


Fig. 26 QUADRATURE SPECTRUM FOR SEASON TWO IN YEAR FIVE (1972)

The shape of the quadrature spectrum is opposite that of the cospectrum. This indicates that the two series are 90° out of phase.

Results from the phase function analysis are given in Figures 27 - 30 and Appendix F.

The outputs of the phase function indicate that phase leads dominate the catchments responses. This means that the rainfall lag terms are more dominant. There is an indication that rainfall terms up to 4 days lags have a maximum effect on the Yala catchment response. This observation agrees with the results obtained under cross - correlation and spectral analysis.

3.2. RESPONSE FUNCTIONS

The results of the response functions are illustrated in Figures 31 to 34. Theoretically, all the values of the response functions are supposed to be positive and the area under the curve should be unity. For a system which is exactly linear and time invariant, the response functions for all the years are supposed to be the same.

It should be noted, however, that in deriving response function ordinates by least squares, it is often found that a condition of numerical instability results in the estimates of the ordinates of the recession being very badly defined, sometimes taking negative values and sometimes oscillating wildly.

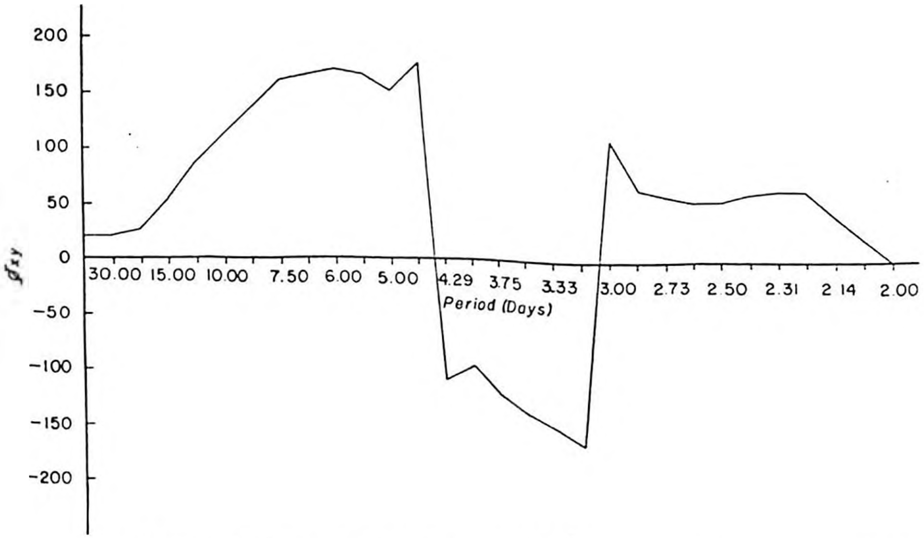


Fig. 27 PHASE FUNCTION FOR SEASON ONE IN YEAR THREE (1970)

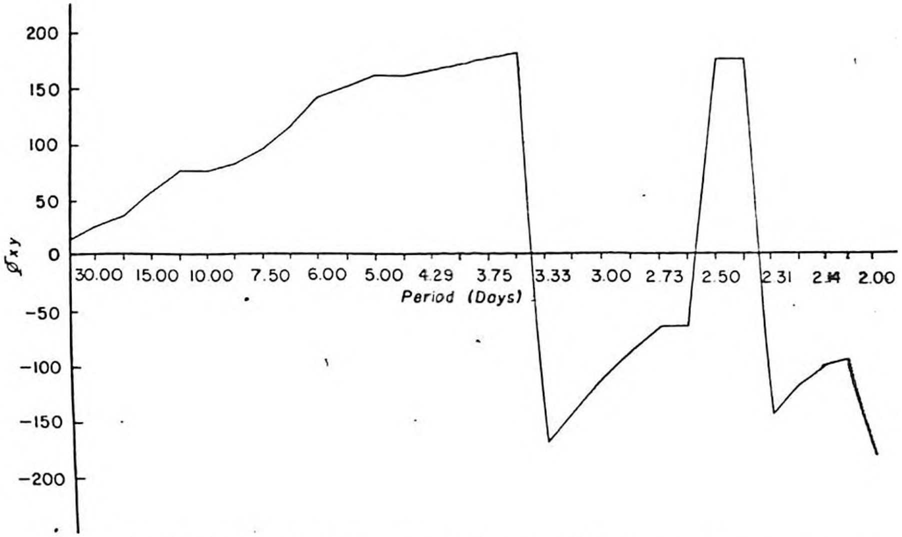


Fig. 28 PHASE FUNCTION FOR SEASON ONE IN YEAR SEVEN (1974)

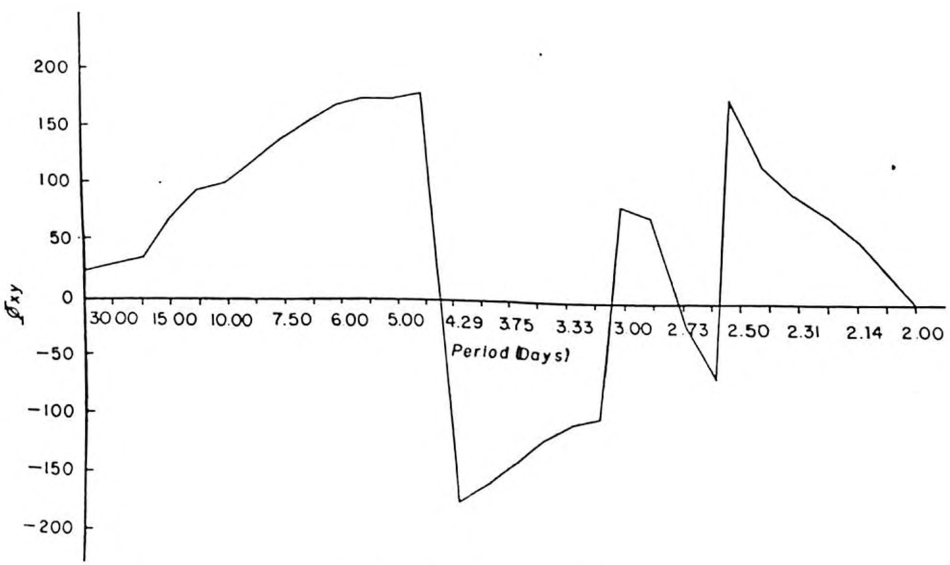


Fig. 29 PHASE FUNCTION FOR SEASON TWO IN YEAR FIVE (1972)

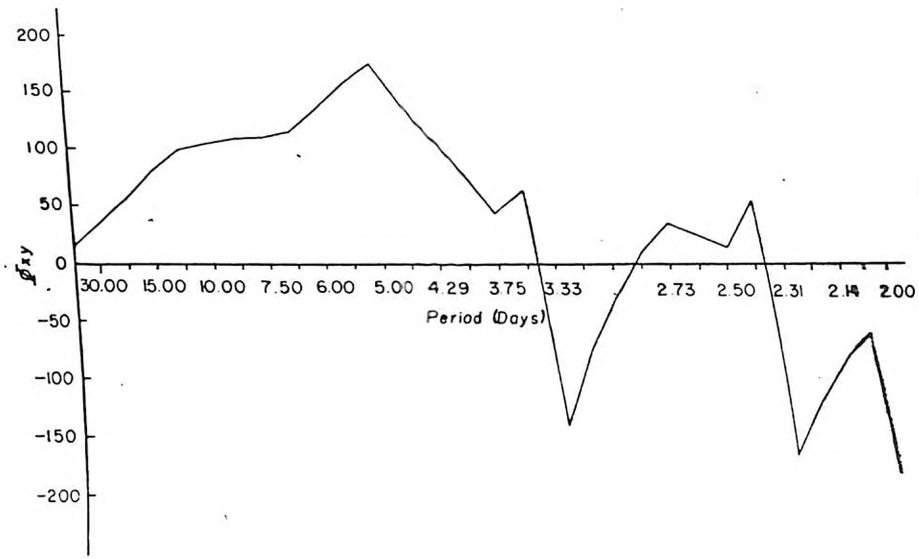


Fig. 30 PHASE FUNCTION FOR SEASON TWO IN YEAR SIX (1973)

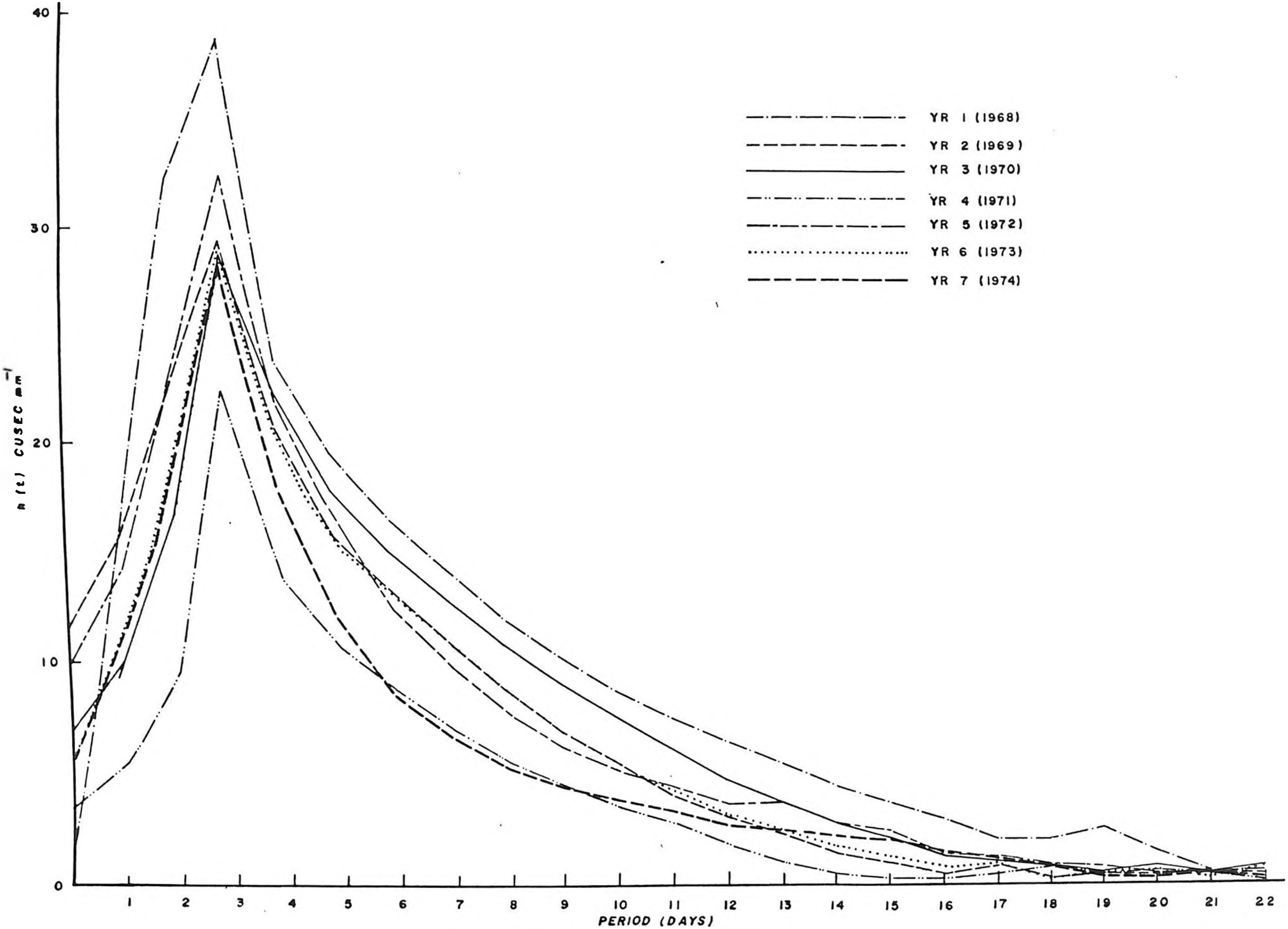


Fig. 31 SEASON ONE — RESPONSE FUNCTIONS

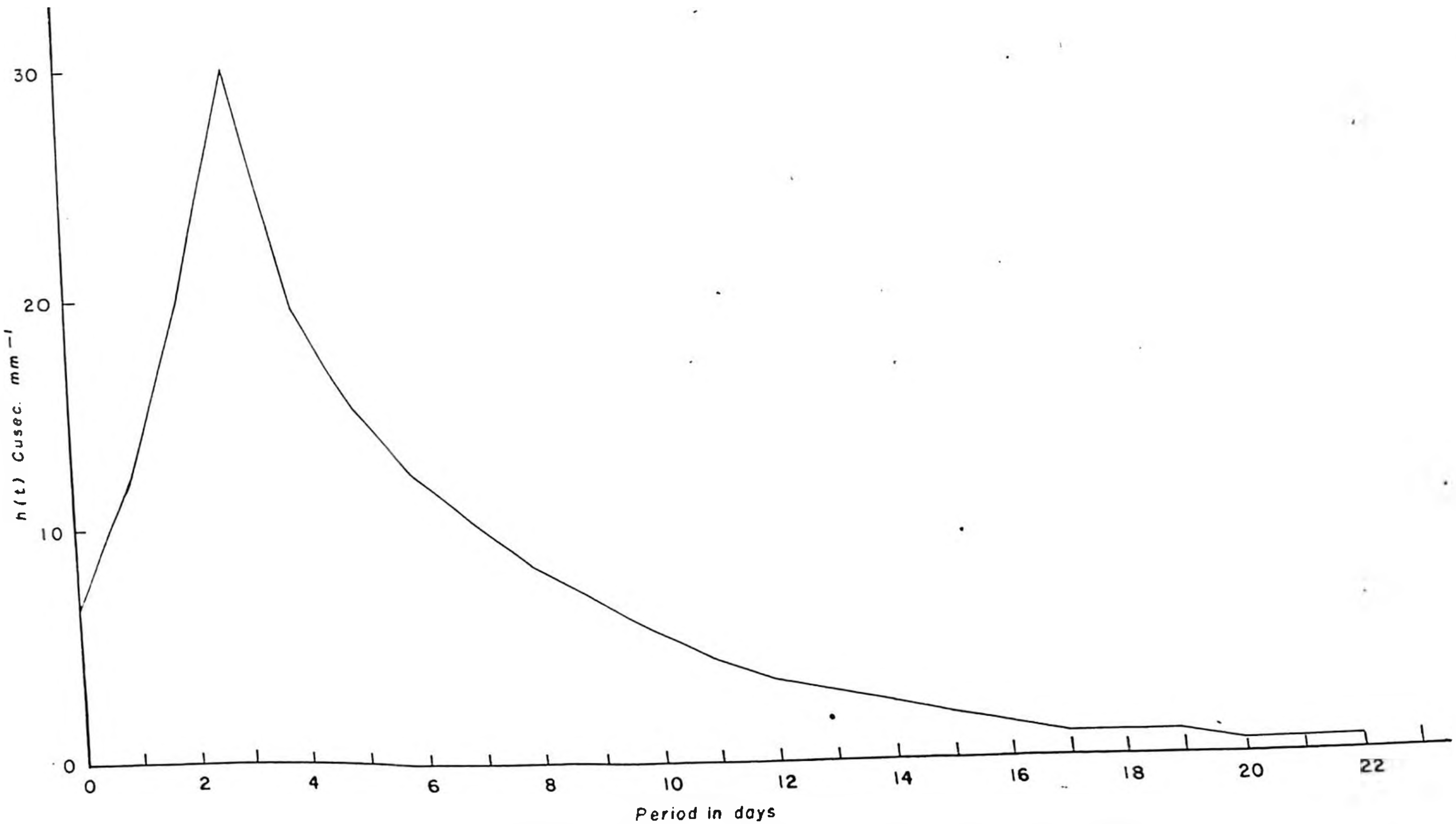


Fig. 32 SEASON ONE — AVERAGE RESPONSE FUNCTION

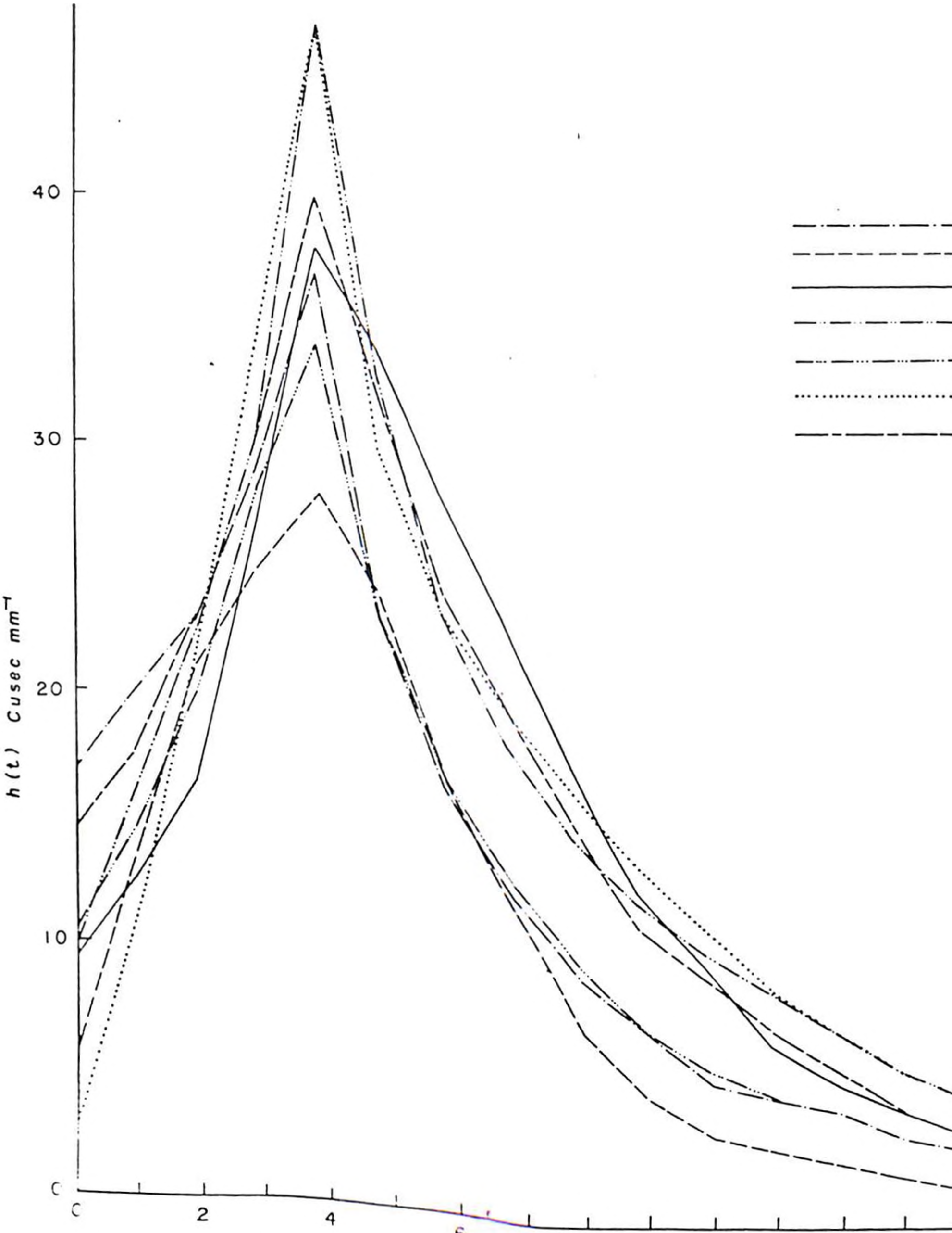
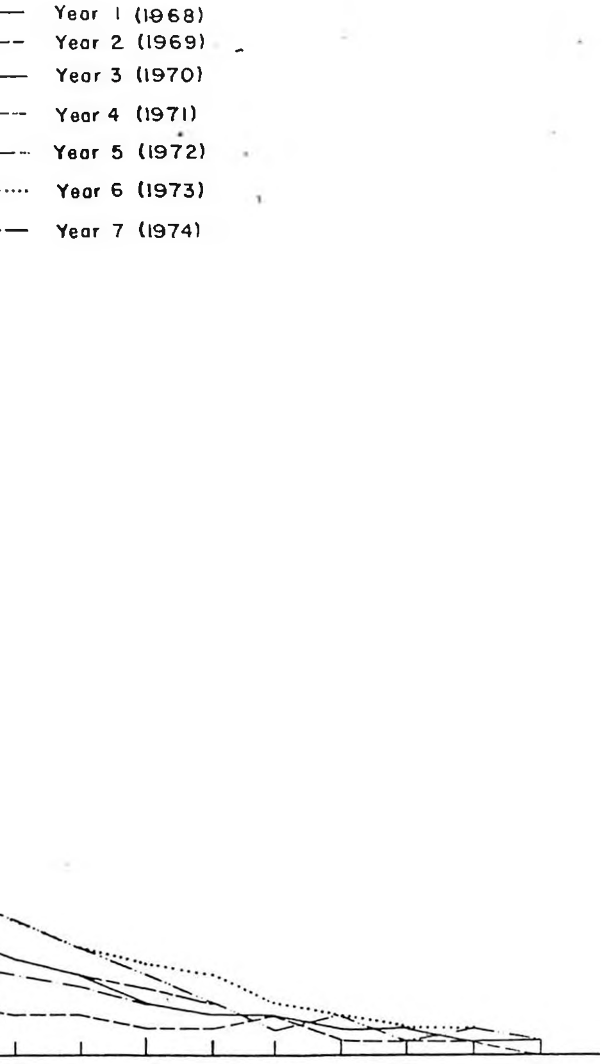


Fig. 33 SEASON TWO — RESPONSE FUNCTIONS


Year 1 (1968)
Year 2 (1969)
Year 3 (1970)
Year 4 (1971)
Year 5 (1972)
Year 6 (1973)
Year 7 (1974)
14
16
18
20
22

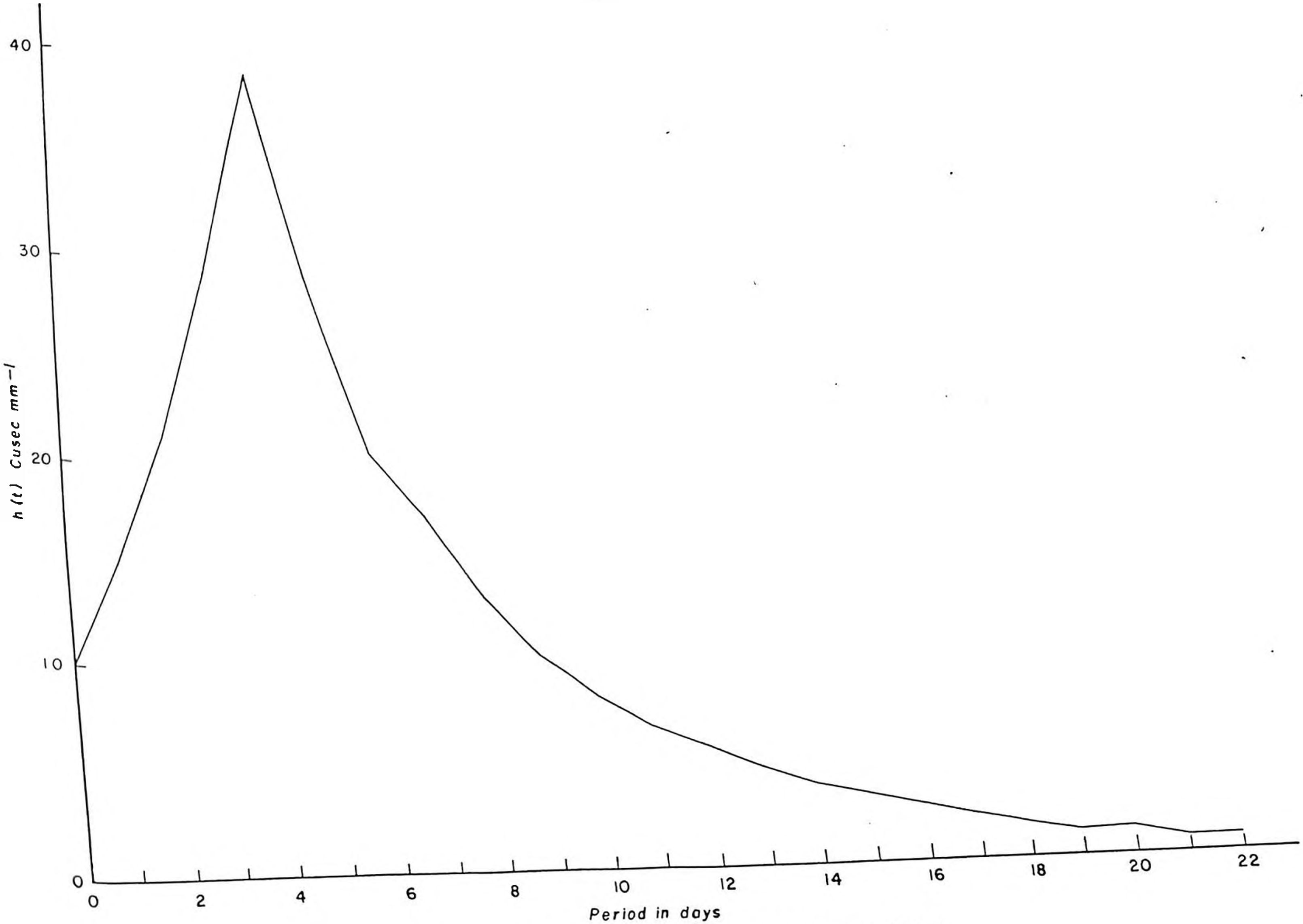


Fig. 34 SEASON TWO MEAN OF RESPONSE FUNCTION

This effect may be more pronounced the greater the degree of *linear dependence* between the independent variables. In this case the input values of Rainfall x(t) with lags t = 0, 1, 2, ... n. Thus the effect is dependent on the degree of serial correlation which is exhibited by the series of the input values. The longer the input and output series, and the more random the input, the less marked is this troublesome effect.

A comparison of the ordinates of the response functions with their standard errors may indicate the accuracy with which the ordinates have been determined and this comparison together with the maximum lag of all positive values of the ordinates can help to indicate the base time of the pulse response or the time of termination of the effect of a disturbance in the input.

Generally, the following points were observed from the results

(i) In season one there was a peak in the response functions at a lag of 3 days. This agrees well with the results obtained from the previous analysis.

(ii) In season two, the peaks appear at a lag of four days which also agrees well with the previous results.

The constancy of the results observed from the response functions also gives some indications on the linearity of the rainfall - runoff systems.

(iii) The patterns of the response functions across the various time lags indicate that the river flow was never zero during the study period.

(iv) The response functions further show some differences in the shapes from one year to another. This may indicate that the rainfall - runoff system for the Yala river basin is not exactly linear and time invariant. This may also be a reflection of the high interannual variations of rainfall over this region.

The avarage response fuctions are shown in Figures 32 and 34. These were used in the prediction of streamflow during the verification period. They were obtained by averaging the response function ordinates of all the seven years in the calibration period.

This averaging technique has a limitation in that it tends to smoothen out the extreme values. This may lead to the over-estimation or under-estimation of some streamflows.

3.3. TESTS OF GOODNESS OF FIT FOR THE RESPONSE FUNCTION MODEL

The two methods which were used to test the goodness of fit of the models in this study were the multiple correlation coefficient and Smirnov-Kolmogorov tests.

The results of the multiple correlation coefficient test (R) for the response function model are summarised in Tables 3 and 4.

TABLE 3: RESULTS FOR YEAR ONE, COEFFICIENT OF MULTIPLE DETERMINATION (R^2), VARIANCE RATIO (F) AND RESIDUAL STANDARD DEVIATION (δe)

	R^2	F	δe
SEASON ONE	0.81	131.6	0.42
SEASON TWO	0.75	85.8	0.50

TABLE 4: RESULTS FOR YEAR TWO, COEFFICIENT OF MULTIPLE DETERMINATION (R^2), VARIANCE RATIO (F) AND RESIDUAL STANDARD DEVIATION (δe)

	R^2	F	δe
SEASON ONE	0.79	91.2	0.48
SEASON TWO	0.73	78.6	0.51

In general, the following points could be observed from these results:

(i) The rainfall-discharge response function model explained over 73% of the total discharge variance in all cases.

(ii) In both years, the model explained higher variance in the first season than the second season. This may have resulted from the fact that the streamflow values were lower in season one than season two for both years and this led to their accurate estimation by the linear model. It can further be observed that the differences in the response functions are smaller in season one than season two. This may imply that the assumption of time - invariance may have been more accurate for season one than season two. It is also well known that the rainfall characteristics during the long rain season (season one) are more reliable and greater than those observed during the short rain season (season two).

The results from the Smirnov-Kolmogorov test also showed that the response function model fitted the data well during the two seasons used.

3.4. RESULTS FROM THE AUTOREGRESSIVE MODELS

The results of the significance of the model parametes and the goodness of fit tests for the autoregressive models will be discussed in this section.

3.4.1. COMPUTED MODEL PARAMETERS FOR THE AUTOREGRESSIVE MODELS

Tables 5 - 7 summarize some of the results observed with the auto-regressive models. The following points were generally noted:

TABLE 5: Yala T.F. (2,2) mean model Parameters, coefficient of multiple determination (R^2), variance ratio (F) and residual standard deviation (δe)

SEASON	a(1)	a(2)	b(0)	b(1)	YEAR	R^2	F	δe
ONE	0.97	-0.09	-0.12	0.06	1	0.82	131.6	0.47
	s	n	s	n	2	0.79	91.2	0.48
TWO	1.13	-0.15	0.01	0.12	1	0.75	85.8	0.50
	s	n	n	s	2	0.73	78.6	0.51

TABLE 6 : Yala T.F. (3,2) mean model parameters, coefficient of multiple determination (R^2), variance ratio (F) and residual standard deviation (δe)

SEASON	a(1)	a(2)	b(0)	b(1)	b(2)	YEAR	R^2	F	δe
ONE	1.00	-0.10	-0.05	0.01	0.17	1	0.86	125.7	0.38
	s	n	n	n	s	2	0.81	111.7	0.34
TWO	0.94	-0.01	-0.01	0.07	0.12	1	0.76	66 .7	0.49
	s	n	n	s	s	2	0.77	70.2	0.41

TABLE 7: Yala T.F. (4,2) mean model parameters, coefficient of multiple determination (R^2), variance ratio (F) and residual standard deviation (δe)

SEASON	a(1)	a(2)	b(0)	b(1)	b(2)	b(3)	YEAR	R^2	F	δe
ONE	1.00	-0.11	-0.03	-0.06	0.17	0.14	1	0.87	111.2	0.36
	s	n	n	s	s	s	2	0.84	86.5	0.39
TWO	1.01	-0.01	0.01	0.02	0.11	0.12	1	0.83	81.6	0.40
	s	n	n	n	s	s	2	0.80	76.2	0.43

(i) The coefficient of discharge at a lag of one day (a(1)), was always the largest in all cases.

(ii) The coefficient of discharge at a lag of two days, (a(2)), was much smaller than a(1) and not significant in all cases.

(iii) The coefficient of rainfall input at lag zero, b(O) was not significant in all cases.

(iv) The coefficient of rainfall input at a lag of one day (b(1)), was positive in all cases but significant in only some cases.

(v) The coefficients of rainfall input at lags of two and three days, b(2) and b(3), respectively were positive and significant in all cases.

From these results it may be concluded that the daily discharge was significantly related to discharge at a lag of one day in all cases. It was also significantly related to rainfall input at lags of two and three days in all cases.

These results agree well with those observed under time series analysis.

3.4.2 TESTS OF GOODNESS OF FIT FOR THE AUTOREGRESSIVE MODELS

The results of the multiple correlation coefficient test are summarised in Tables 5 - 7. These results showed that:

(i) The autoregressive model T.F. (2,2) explained 82% of the total runoff variance in season one of the first year while 75% was explained in season two. The corresponding proportions of the variance explained in the second year were 79% and 73% respectively.

(ii) The autoregressive model T.F. (3, 2) explained 86% of the total runoff variance in season one of the first year while 76% was explained in season two. *The corresponding proportions of the variance* explained in the second year were 81% and 77% respectively.

(iii) The autoregressive model T.F. (4, 2) explained 87% of the total runoff variance in season one of the first year while 83% was explained in season two. The corresponding proportions of the variance explained in the second year were 84% and 80% respectively.

(iv) The total variance explained by the models generally increased with the orders of the models. Similar characteristics were indicated from some previous results of time series analyses which included the patterns of the auto-correlation, cross-correlation and coherence functions.

(v) In all cases, the models explained a higher percentage of the total runoff variance in season one than season two. The differences in the model parameters obtained in season one were relatively smaller than those of season two. This indicates that the rainfall - runoff system for the Yala river basin was relatively more time invariant in season one than in season two, during the period of study. The stability of the rainfall characteristics during the first season has also been highlighted in some earlier parts of the text.

The Smirnov-Kolmogorov test was also used to test the goodness of fit of the discharge estimated by the autoregressive models. The results showed that all the models had good fits at the 5% significance level.

It can be noted that the models used in this study explained between 73% and 87% of the total variance in the discharge. The difference of 13% - 27% can be attributed to:

(i) Lack of representativeness of the data which is caused by taking the rainfall input and discharge output in a lumped fashion. Rainfall input is taken as an areal mean rainfall obtained by averaging the records of a number of raingauges. During the computation of the areal mean rainfall, the non-representativeness of the recorded rainfall and the distribution of the rainfall within the catchment are neglected. This is further amplified if the catchment characteristics like soil type, vegetation and orography are not homogeneous.

(ii) Inadequacy of the linear assumption which may be caused by the fact that the process of transforming rainfall into runoff is not truely linear. Other factors which affect runoff like catchment storage and evaporation were not considered.

(iii) Inadequacy of the assumption of time invariance. The fact that the response of a catchment to the same rainfall event varies from one season to another can scarcely be doubted. In defining the wet seasons used in this study, it was assumed that during these seasons, the soils were saturated such that most of the rainfall contributed directly to runoff. Thus the assumption of time invariance. But this may not be practically true since the effects of other factors like evaporation etc., may not be negligible. Due to the variability of the rainfall within the seasons used, variations would also be expected in the response functions. These factors may introduce some time variance in the rainfall - discharge system and this might have introduced the large variations observed in some of the results.

3.5. COMPARISON BETWEEN THE MODELS

In order to compare the results obtained with the various models, the standard errors of estimate as given in Equation (67) were computed for all the models. Tables 8 and 9 give the computed standard errors for the various models. It can be observed from these tables that:

TABLE 8: RESULTS OF THE STANDARD ERRORS OF ESTIMATE FOR SEASON ONE (APRIL - JUNE)

TYPE OF MODEL	STANDARD ERROR OF ESTIMATE	
	YEAR ONE (1975)	YEAR TWO 1976
T.F. 4,2	6.9×10^1	3.9×10^1
T.F. 3,2	1.8×10^2	2.1×10^2
T.F. 2,2	2.6×10^2	7.7×10^1
RESPONSE FUNCTION	3.7×10^2	4.2×10^2

TABLE 9 RESULTS OF THE STANDARD ERRORS OF ESTIMATE FOR SEASON TWO (AUGUST - OCTOBER)

TYPE OF MODEL	STANDARD ERROR OF ESTIMATE	
	YEAR ONE (1975)	YEAR TWO (1976)
T.F. 4,2	1.7×10^2	1.9×10^1
T.F. 3,2	2.3×10^2	4.7×10^1
RESPONSE FUNCTION	2.8×10^2	1.6×10^2
T.F. 2,2	5.7×10^2	2.5×10^2

(i) The T.F. (4, 2) model had the smallest standard error of estimate in all cases.

(ii) T.F. (4,2) and T.F. (3,2) models performed better than the response function model in all cases.

(iii) For the case of the autoregressive models, the standard of performance increased with the orders of the models.

(iv) The over all performance was better in season one than in season two.

It should however be noted that the performances of most of the models were generally good as can be seen from the percentages of the total discharge variance that was accounted for by the various models and the magnitudes of the standard errors of estimate.

3.6. VERIFICATION OF THE SKILLS OF THE MODELS USING PREDICTED DISCHARGE

In order to verify the results obtained from the various models, data from another period which was not used to estimate the models parameters are often required. In this study the years 1975 - 1976 were used to verify the performances of the various models. Some of the results are shown on Figures 35 to 50. It can be observed from these graphs that:

(i) The performances of all the models were reasonably good.

(ii) The predicted hydrographs exhibited the major features of the actual hydrographs.

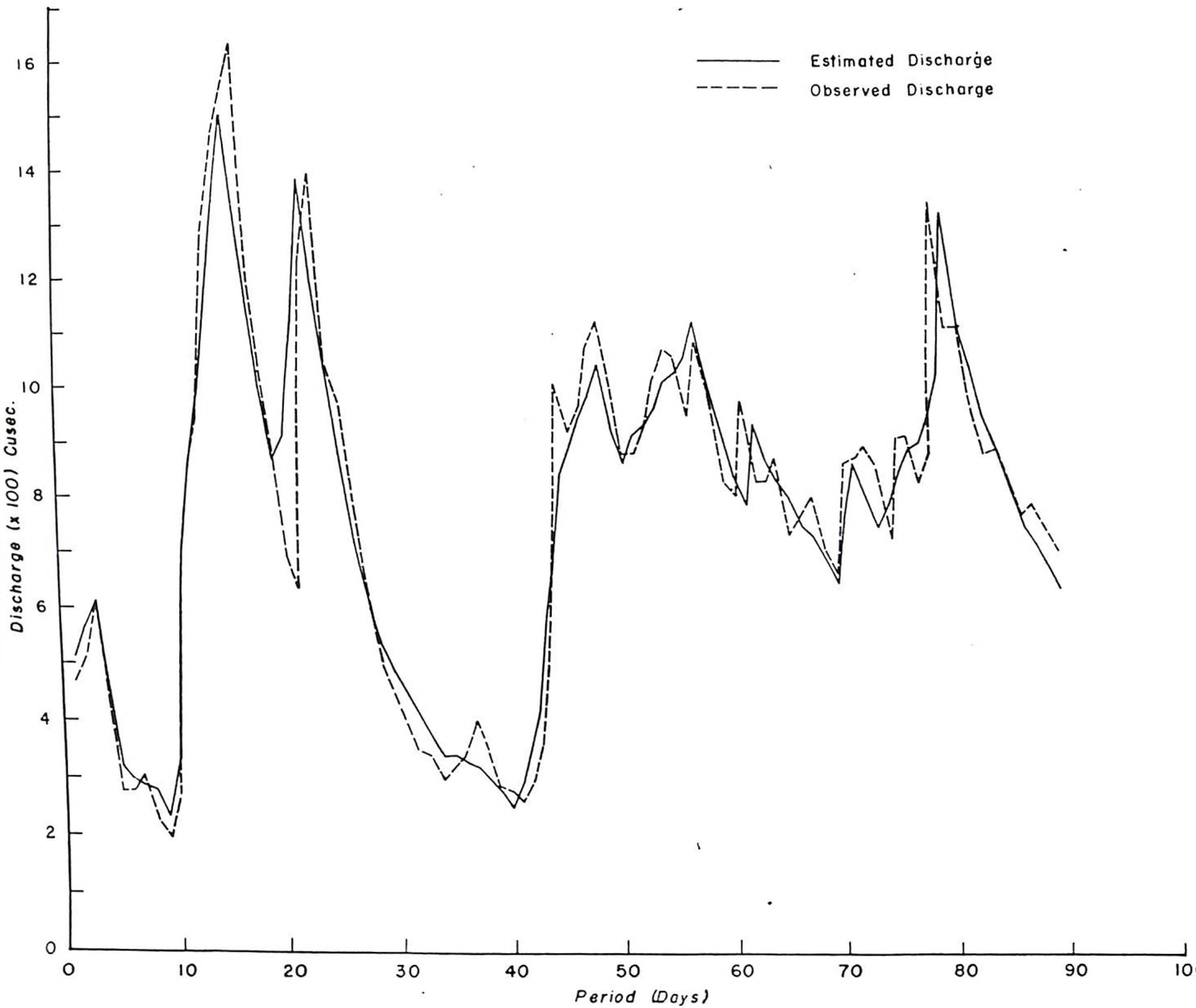


Fig. 35 SEASON ONE YEAR ONE (1975) — PREDICTION BY THE RESPONSE FUNCTION MODEL

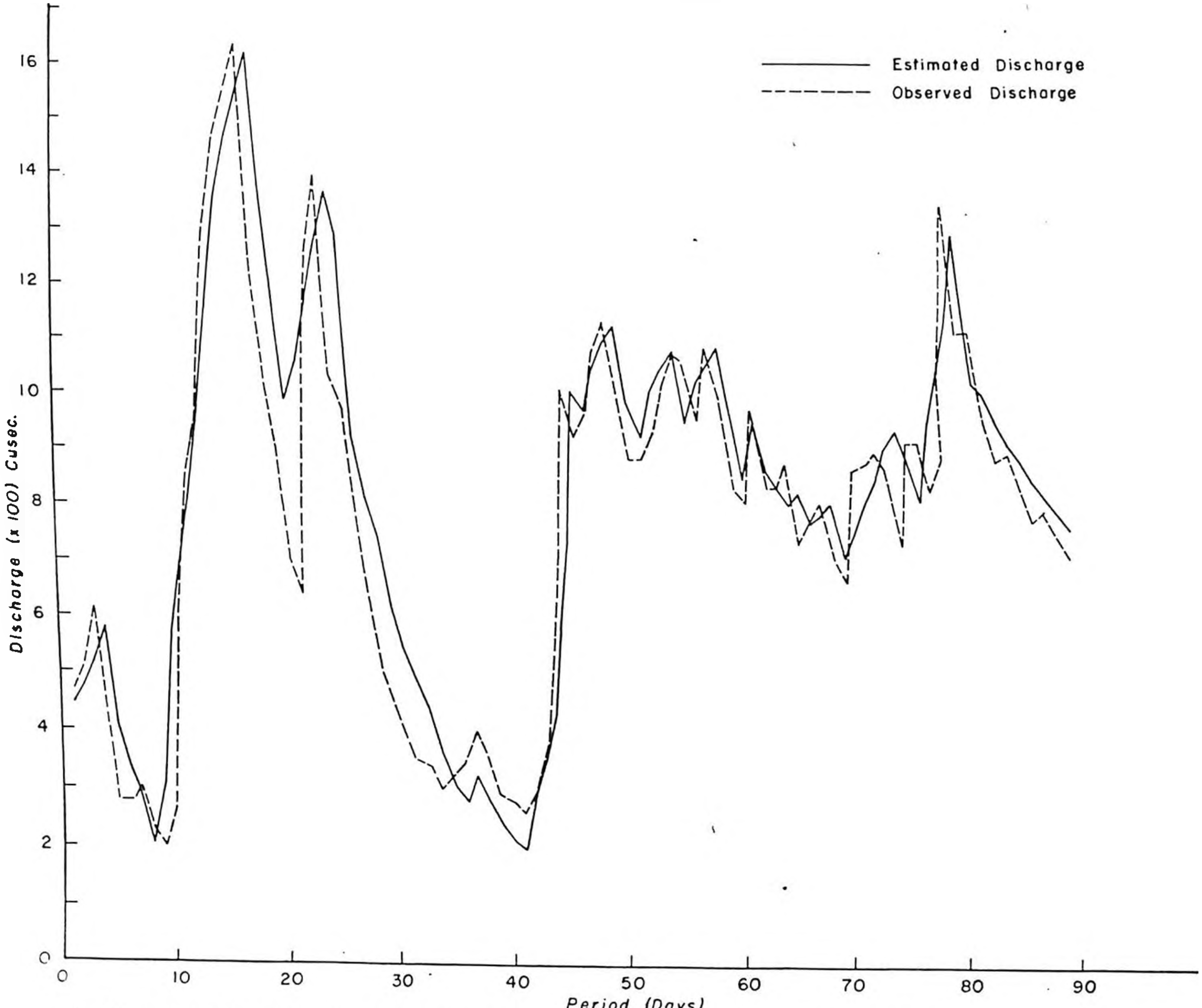


Fig. 36 SEASON ONE YEAR ONE (1975) — PREDICTION BY T.F. 2.2 MODEL

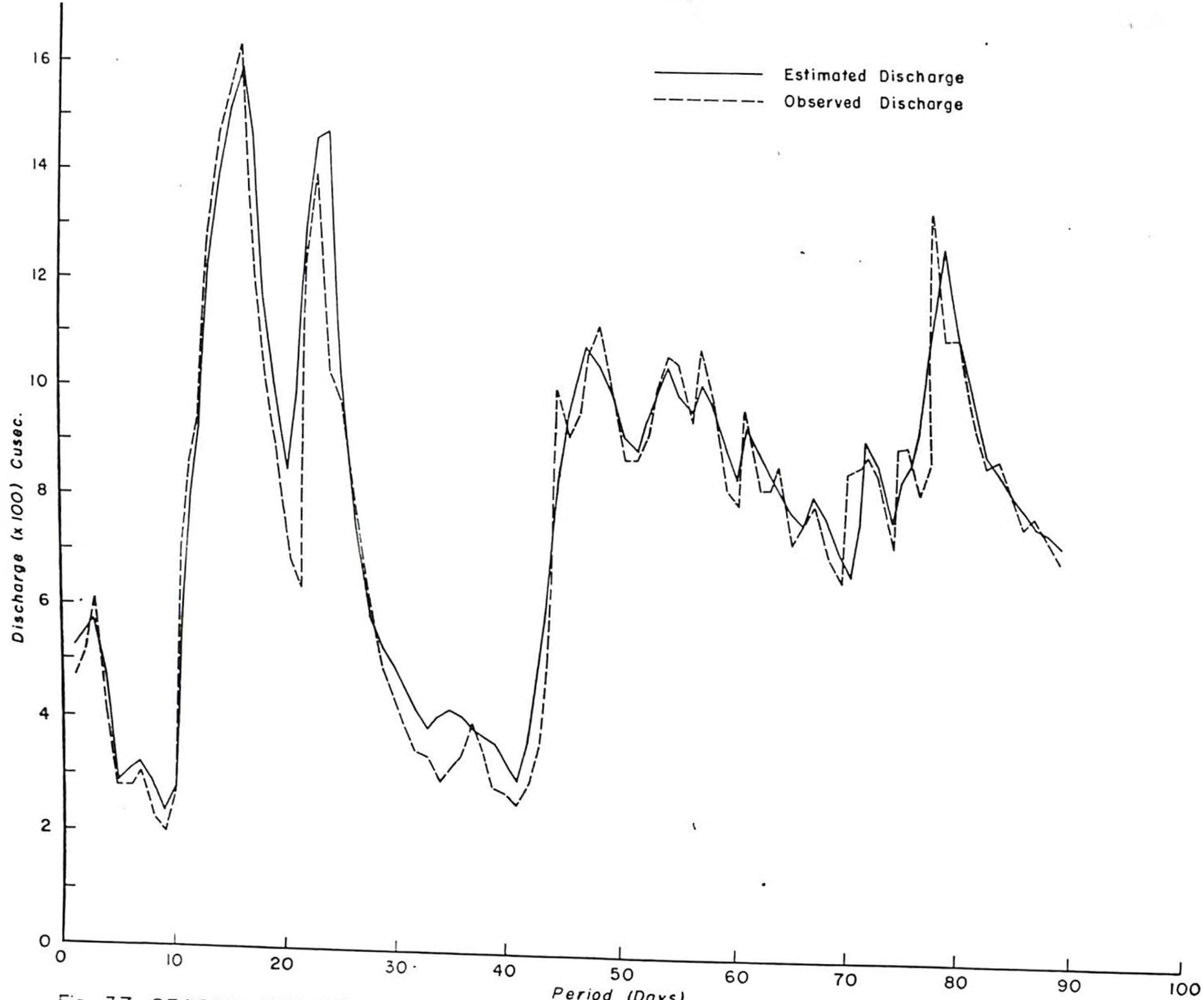


Fig. 37 SEASON ONE YEAR ONE (1975) — PREDICTION BY THE T. F. 3.2 MODEL

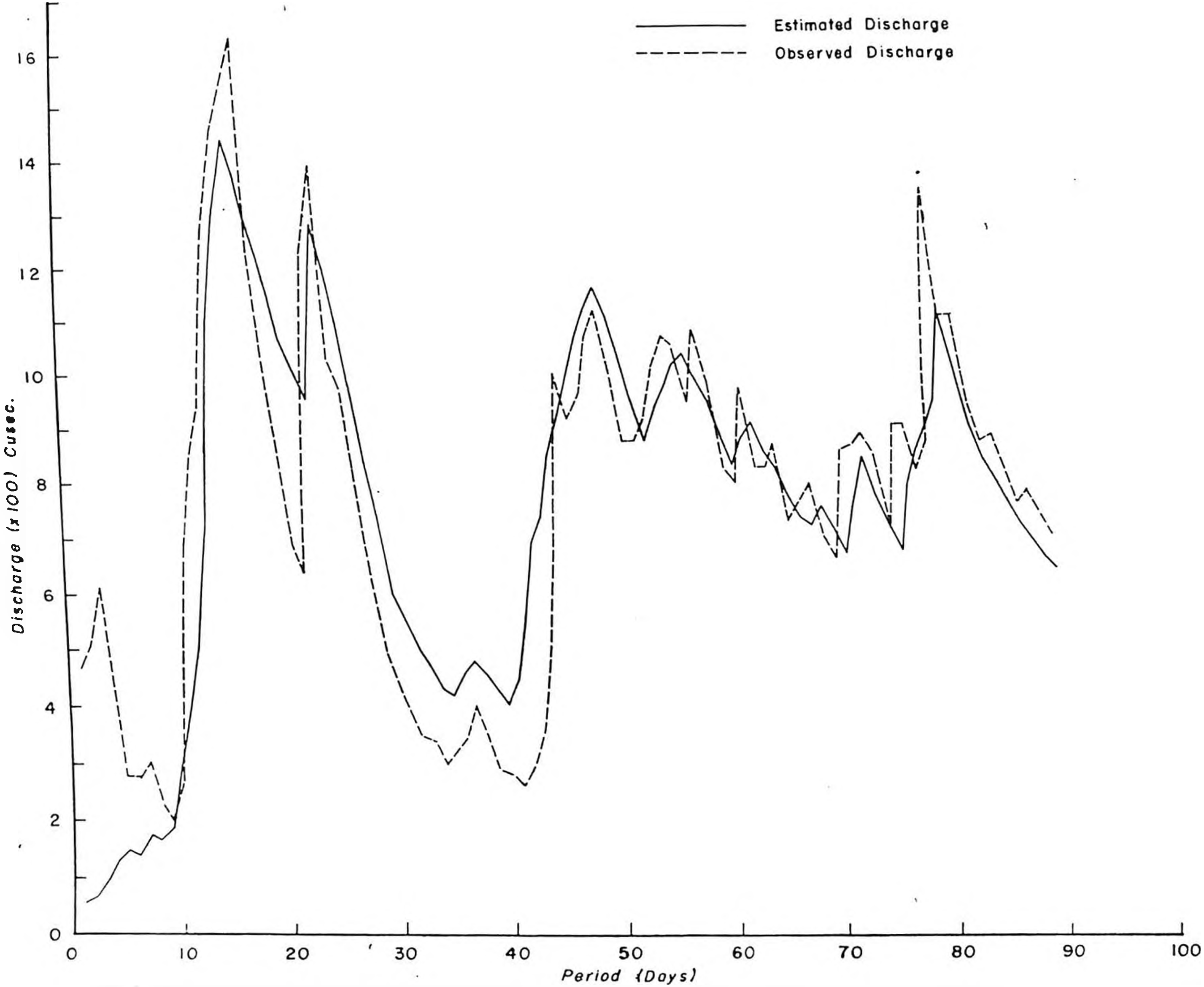


Fig. 35 SEASON ONE YEAR ONE (1975) — PREDICTION BY THE T.F. 4.2 MODEL

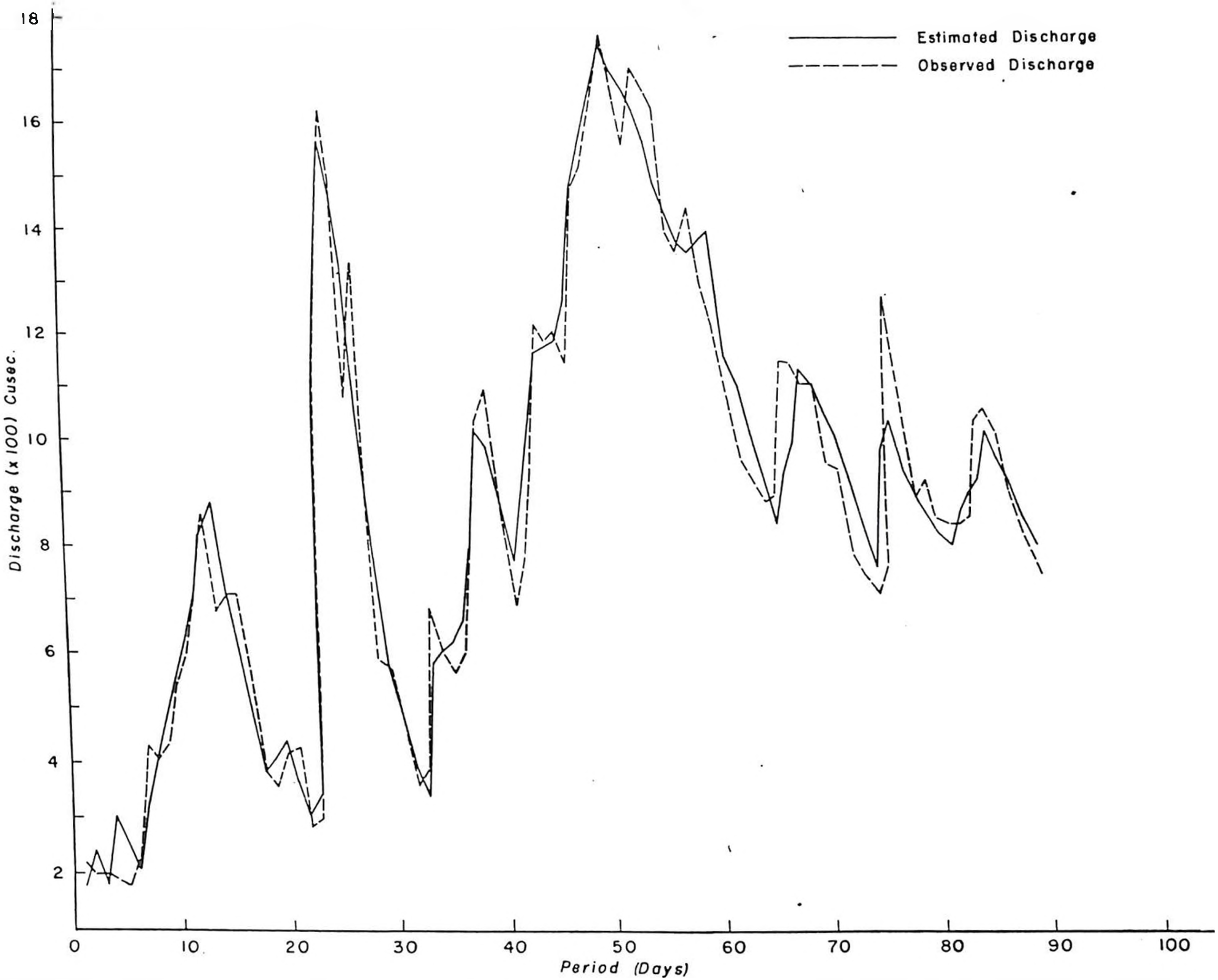


Fig. 39 SEASON ONE YEAR TWO (1976) — PREDICTION BY THE RESPONSE FUNCTION MODEL

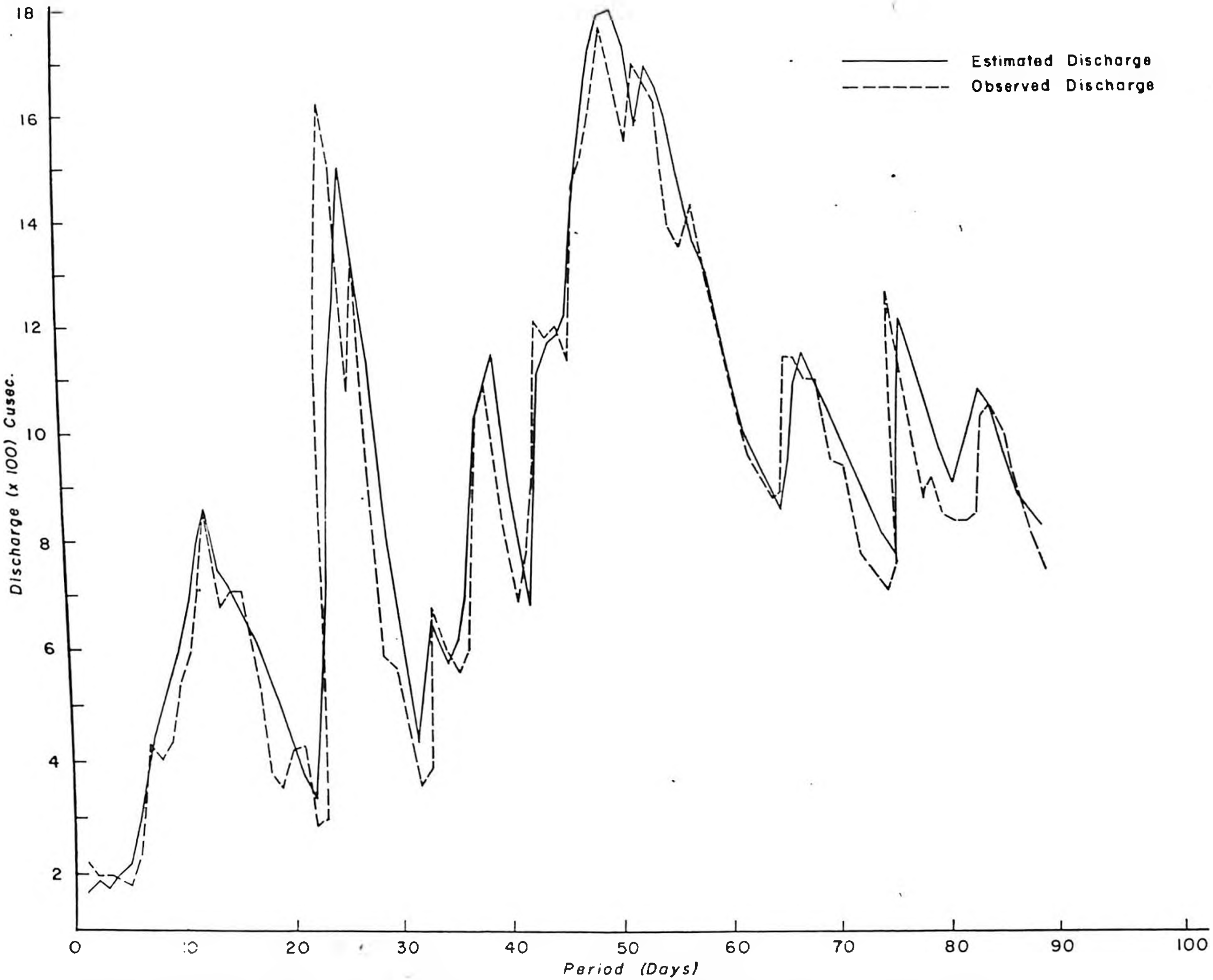


Fig. 40 SEASON ONE YEAR TWO (1976) – PREDICTION BY THE T.F. 2.2 MODEL

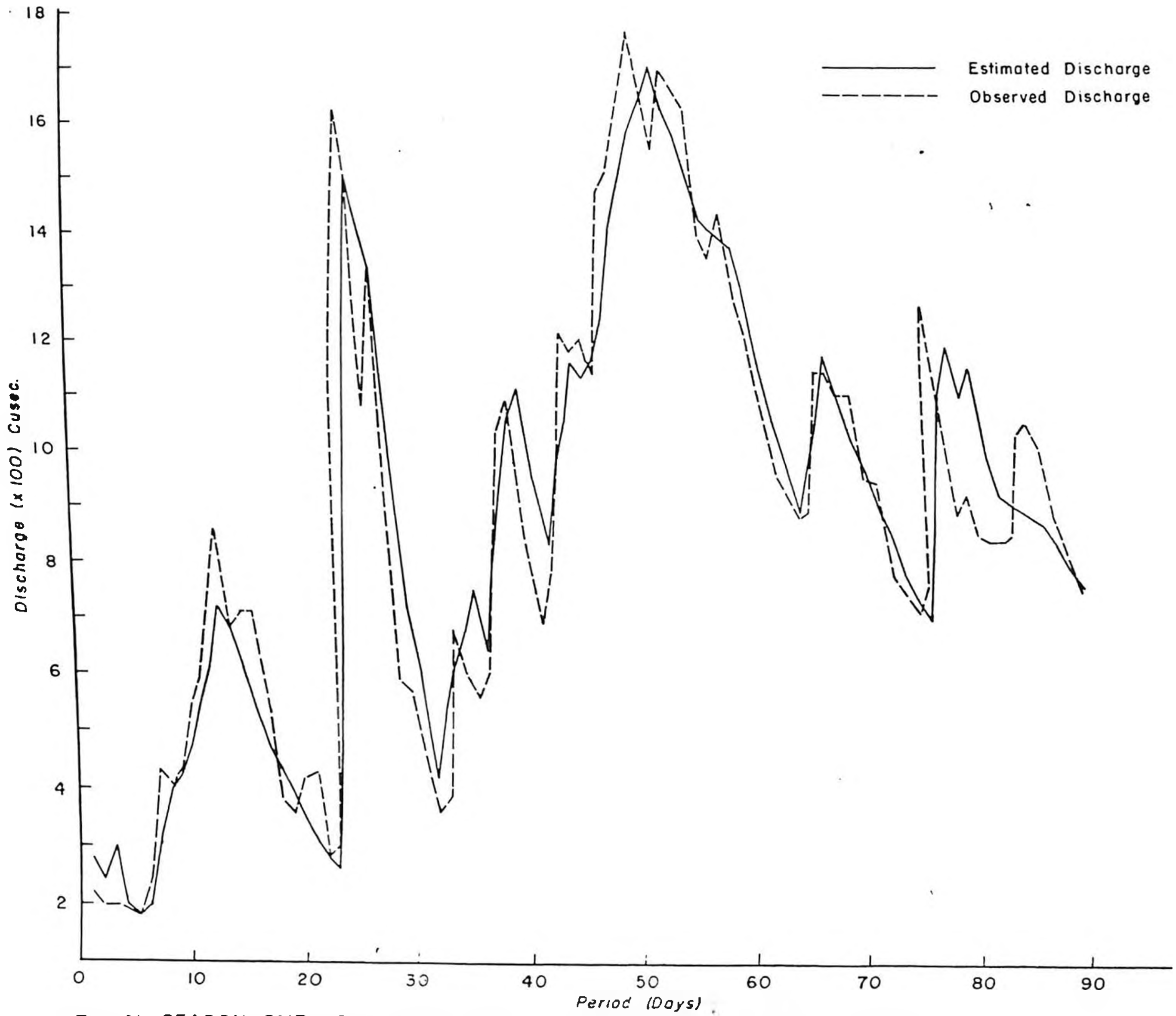


Fig 41 SEASON ONE YEAR TWO (1976) — PREDICTION BY T.F 3.2 MODEL

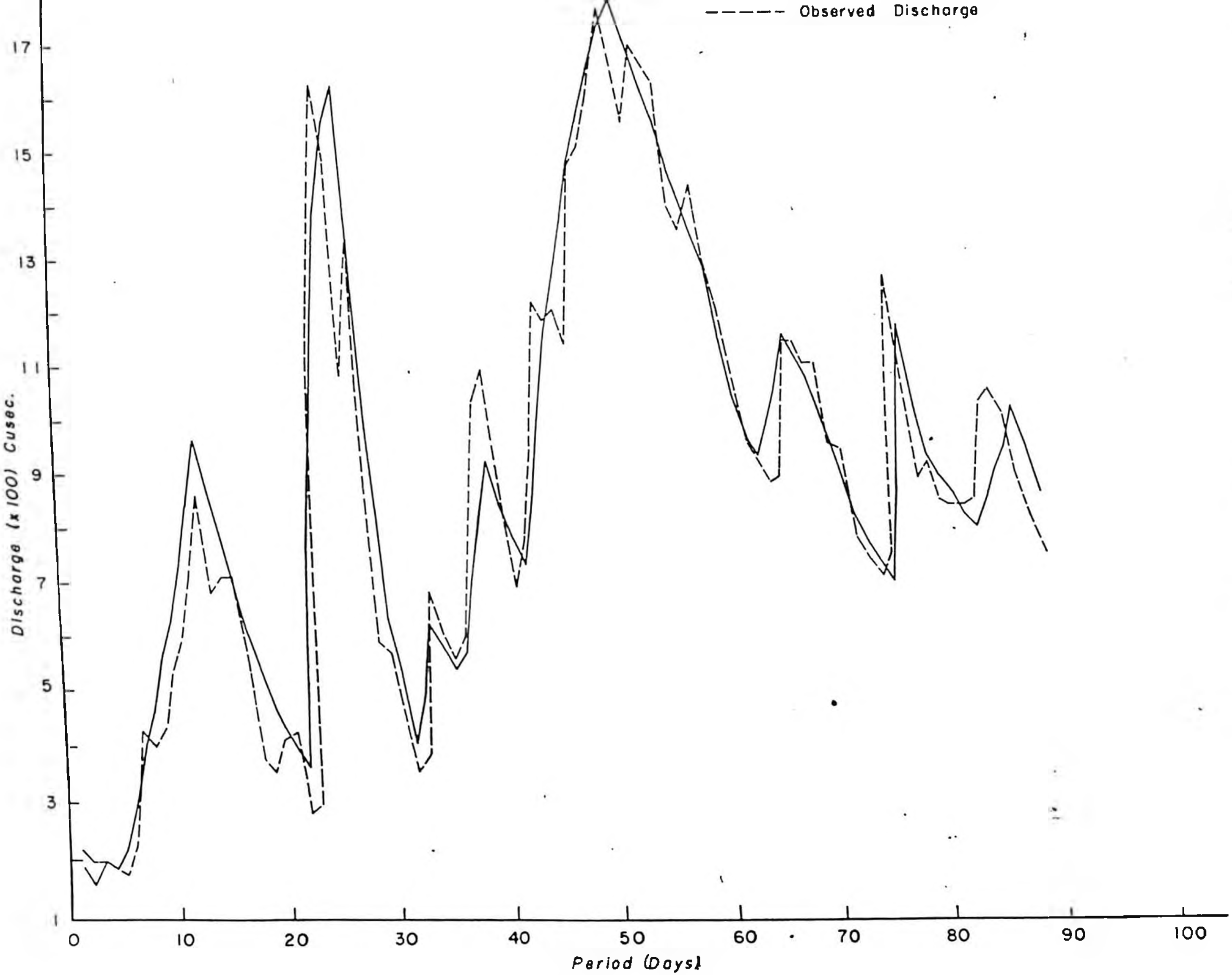


Fig. 42 SEASON ONE YEAR TWO (1976) – PREDICTION BY T.F. 4.2 MODEL

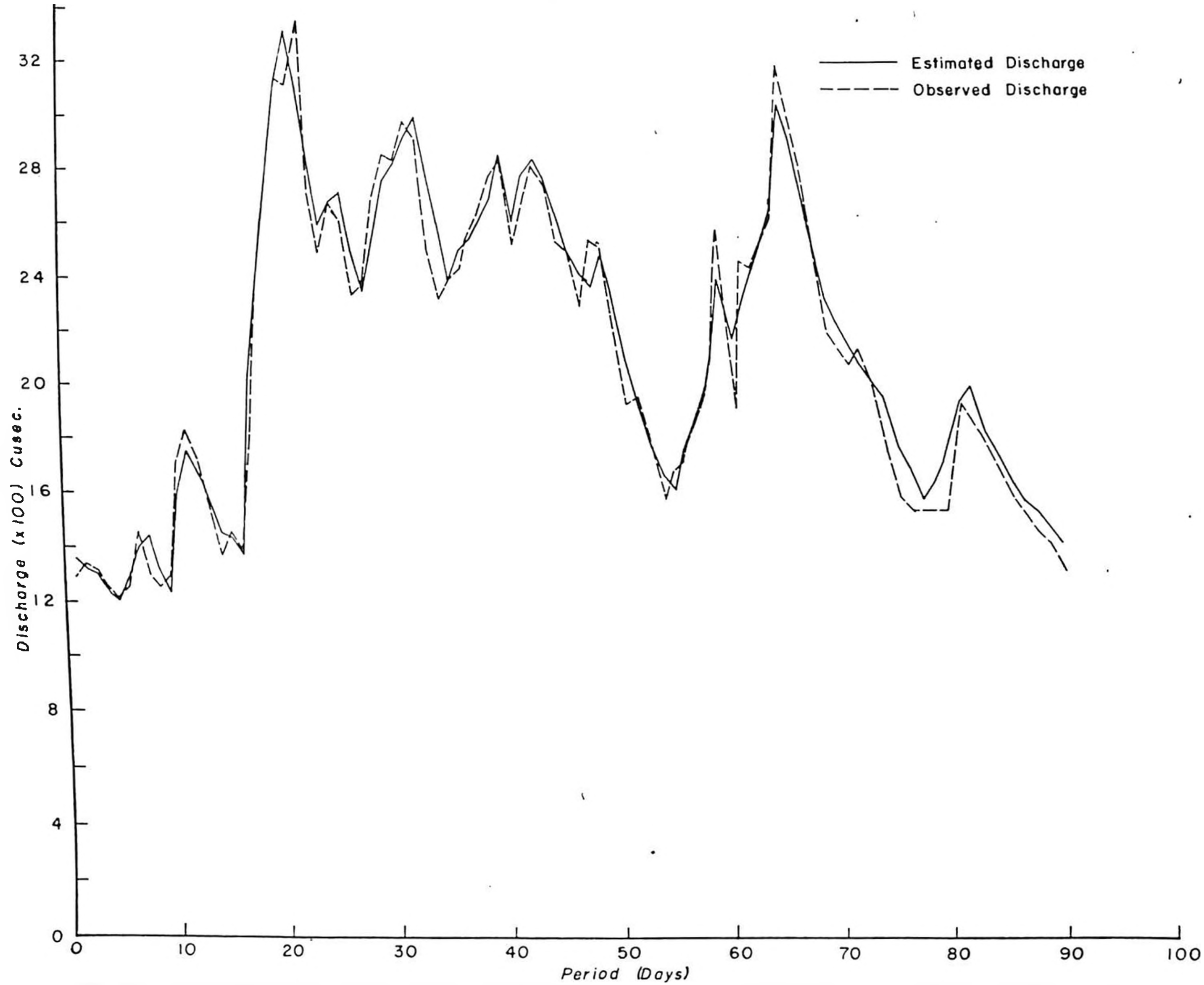


Fig. 43 SEASON TWO YEAR ONE (1975) — PREDICTION BY THE RESPONSE FUNCTION MODEL

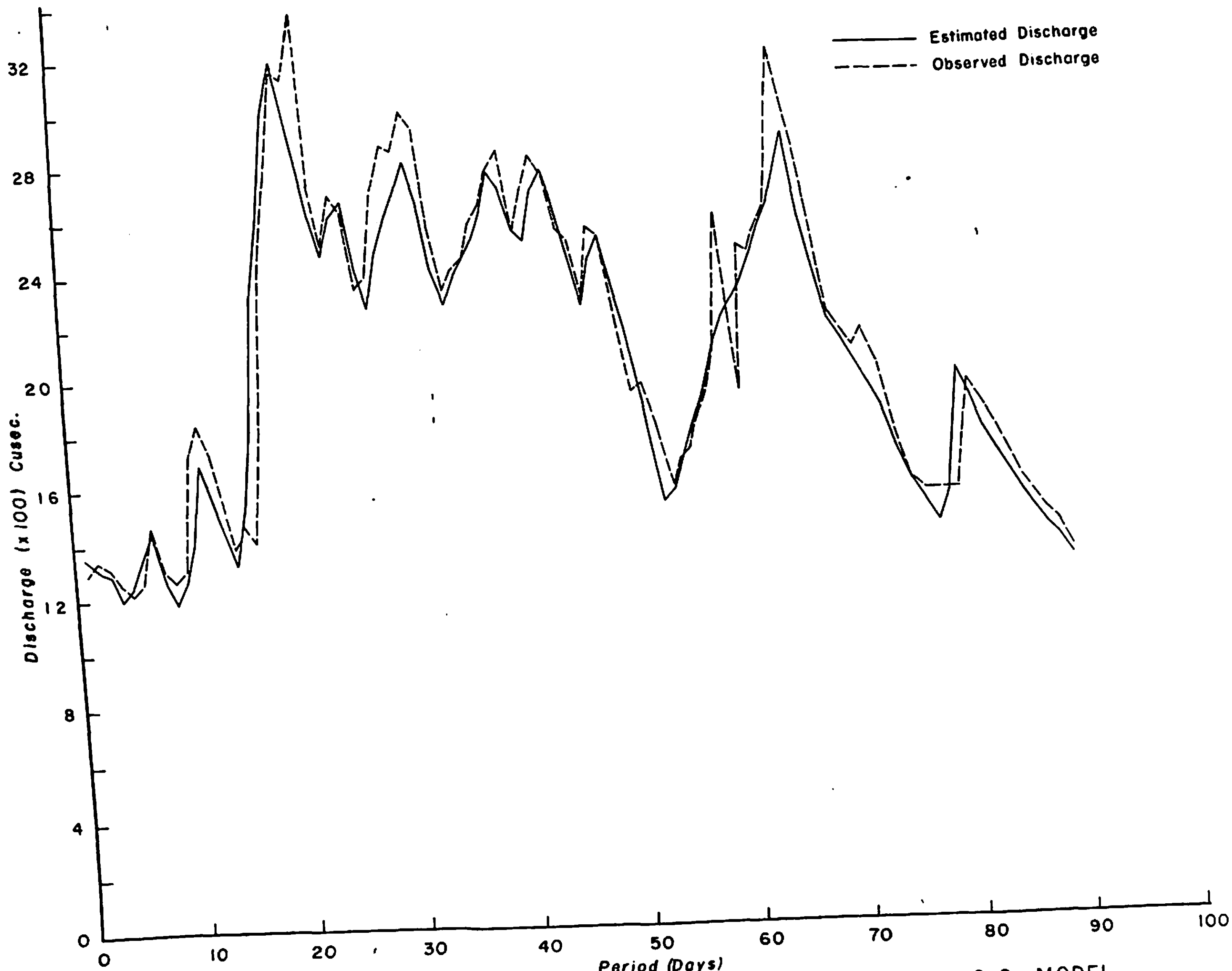


Fig. 44 SEASON TWO YEAR ONE (1975) — PREDICTION BY THE T.F. 2.2. MODEL

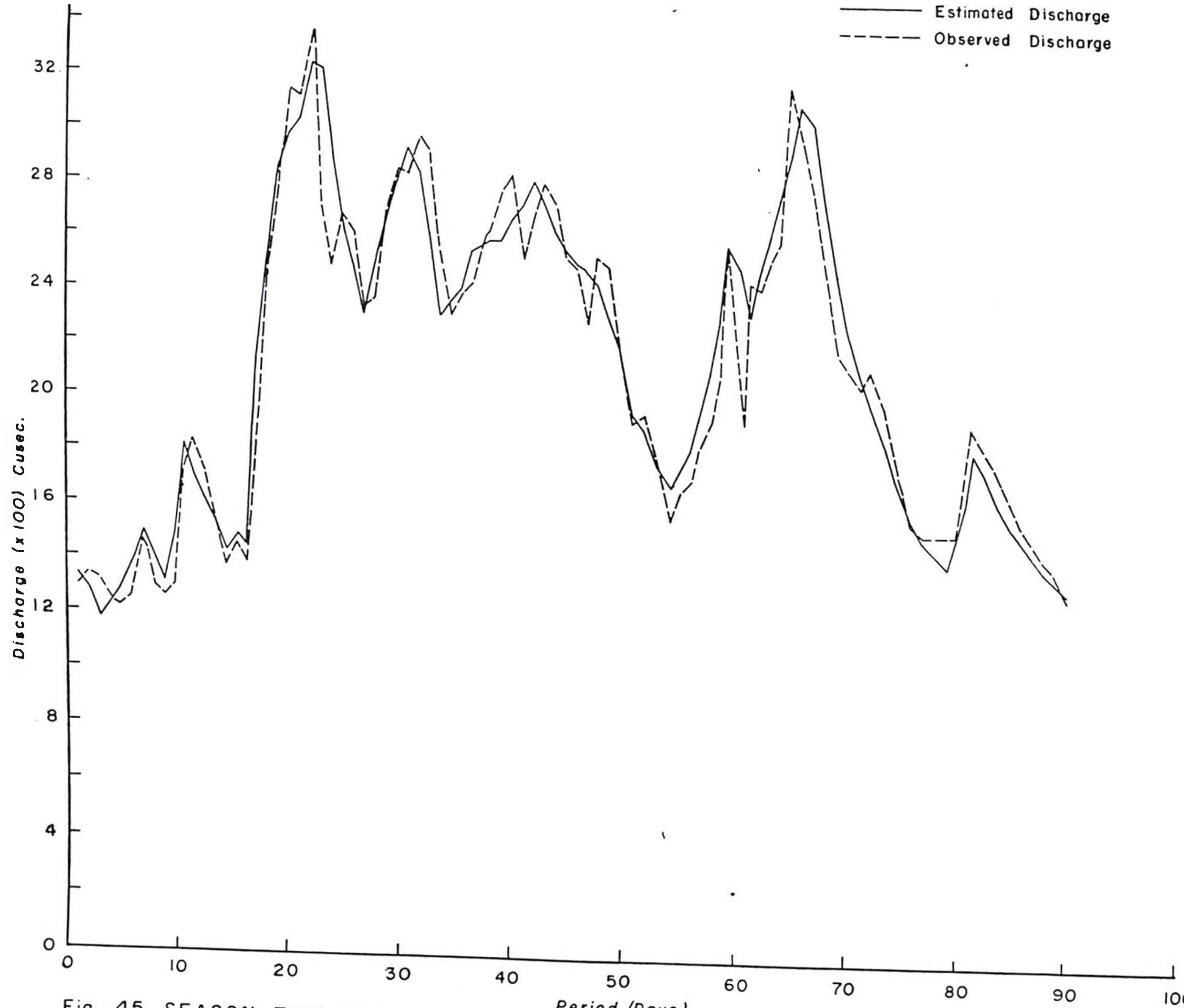


Fig. 45 SEASON TWO YEAR ONE (1975) — PREDICTION BY THE T.F. 3.2 MODEL

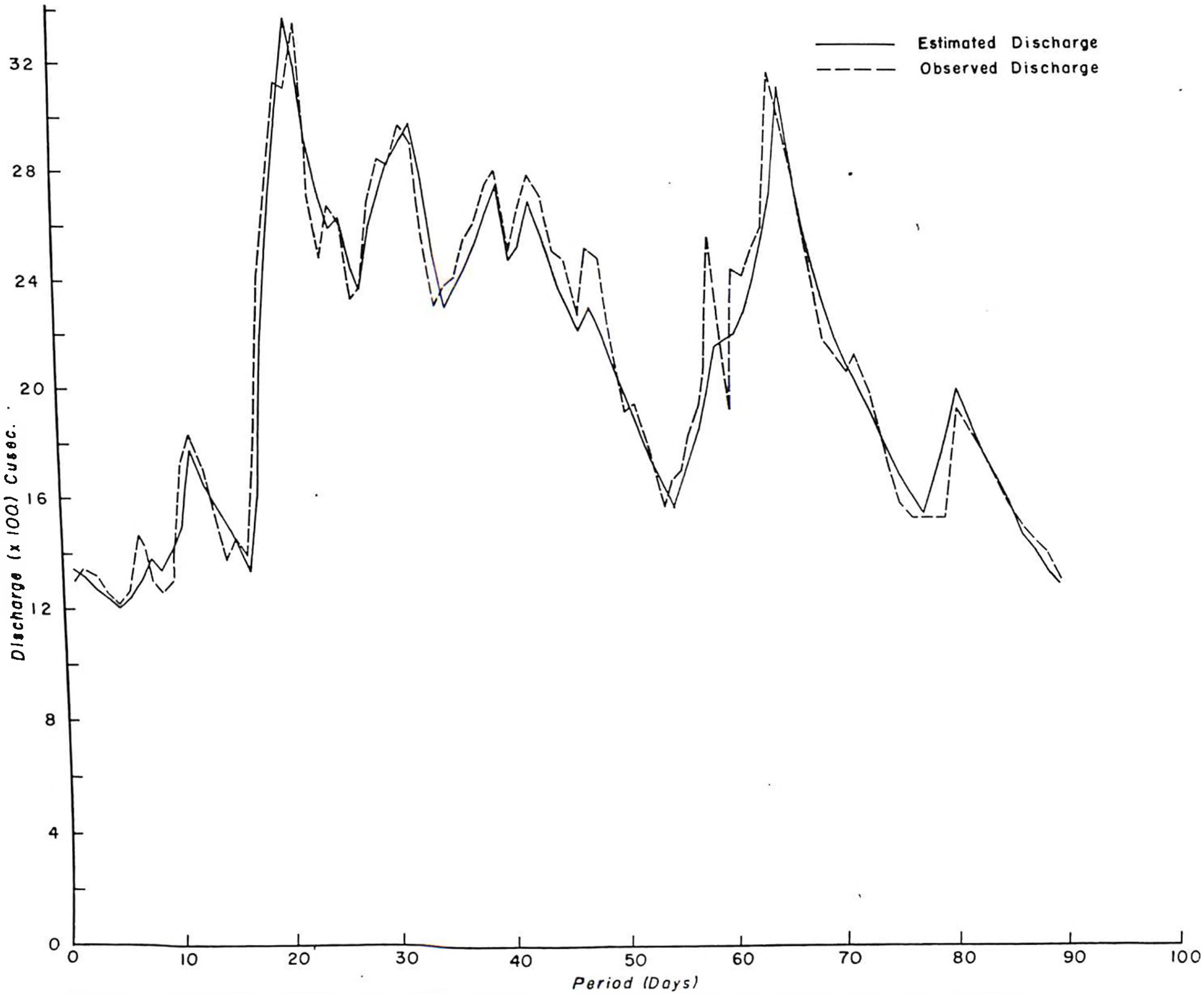


Fig. 46 SEASON TWO YEAR ONE (1975) — PREDICTION BY THE T.F. 4.2 MODEL

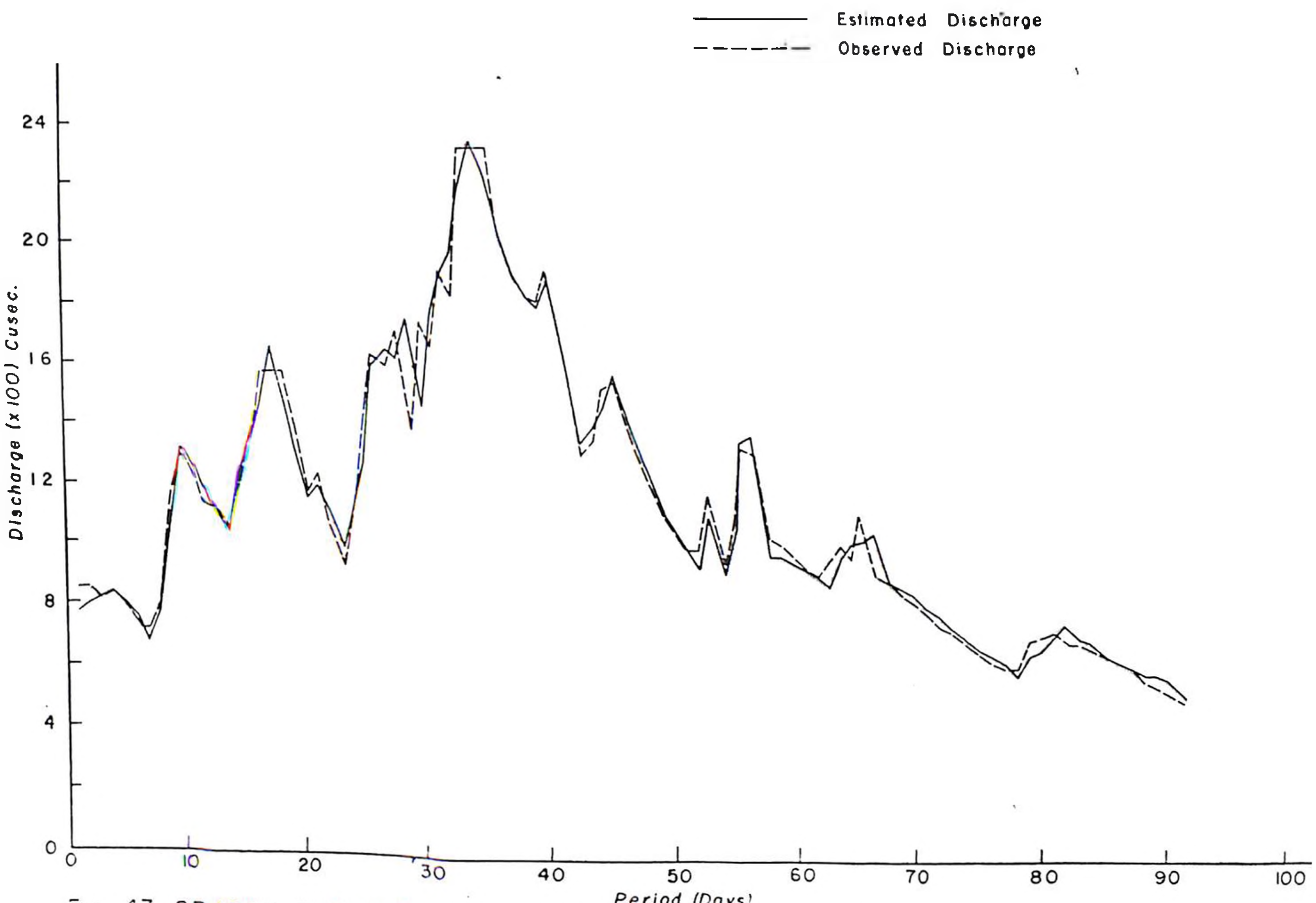


Fig. 47 SEASON TWO YEAR TWO (1976) — PREDICTION BY THE RESPONSE FUNCTION MODEL

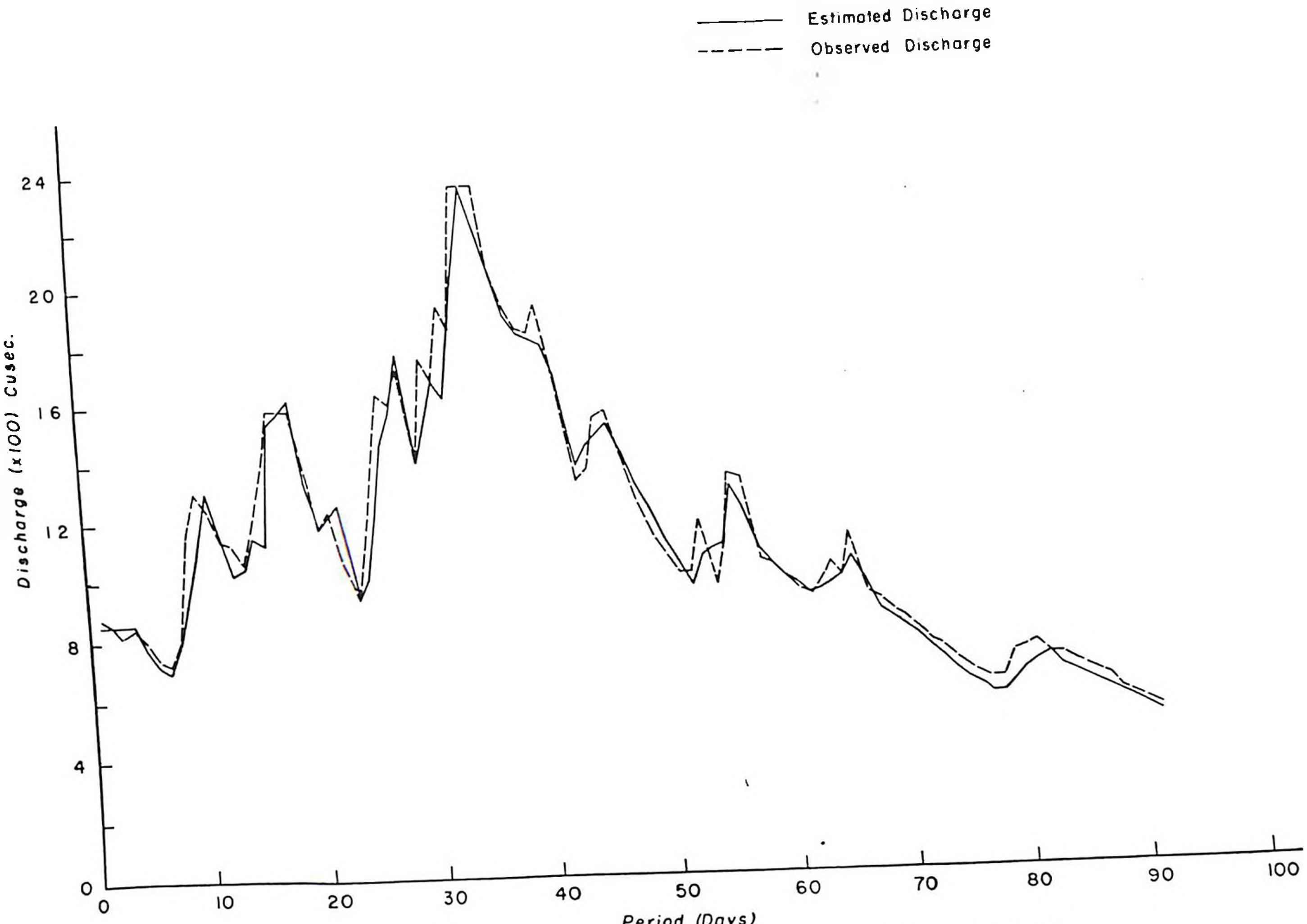


Fig. 48 SEASON TWO YEAR TWO (1976)—PREDICTION BY THE T.F. 2.2 MODEL

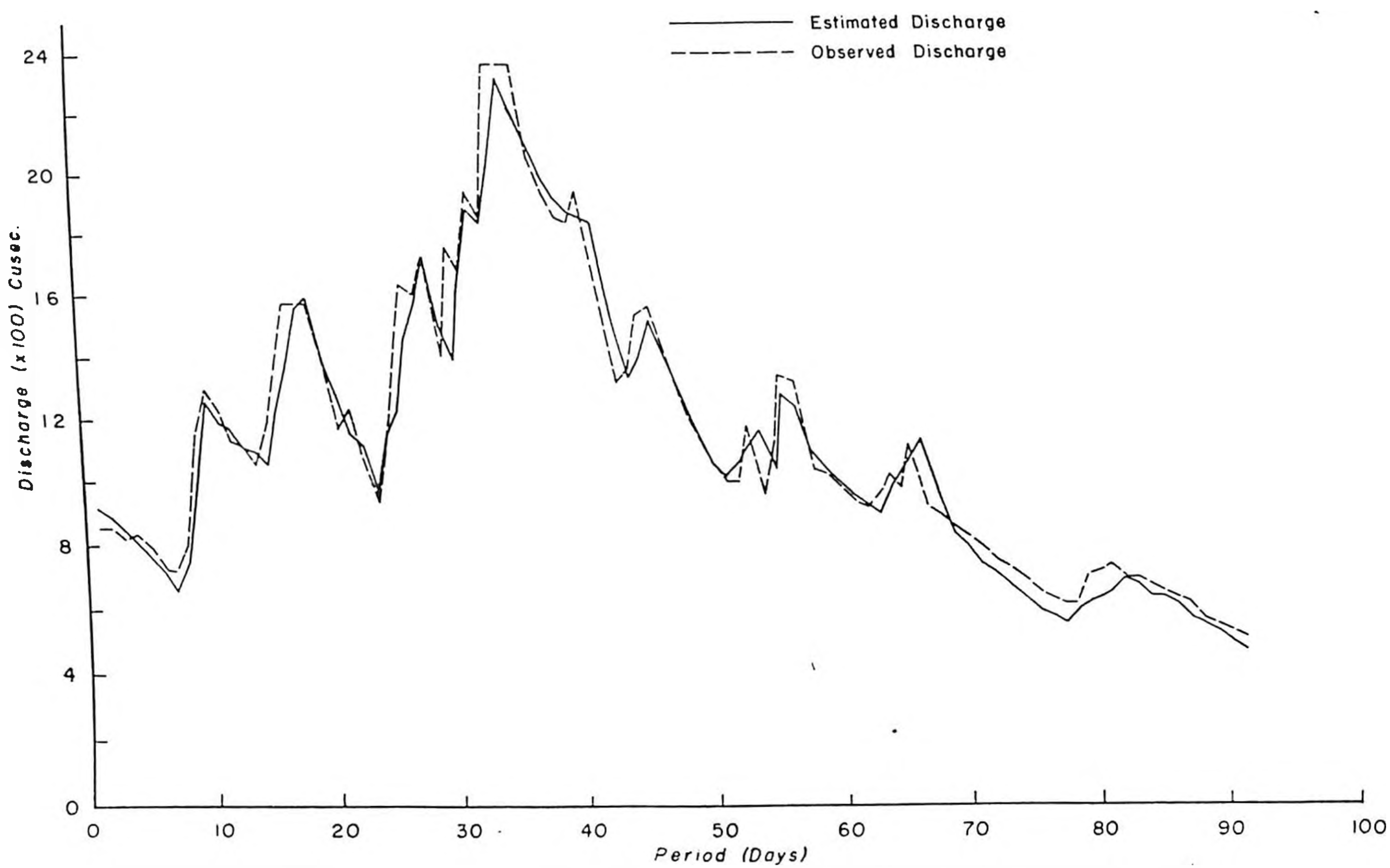


Fig 49 SEASON TWO YEAR TWO (1976) — PREDICTION BY THE T.F. 3.2 MODEL

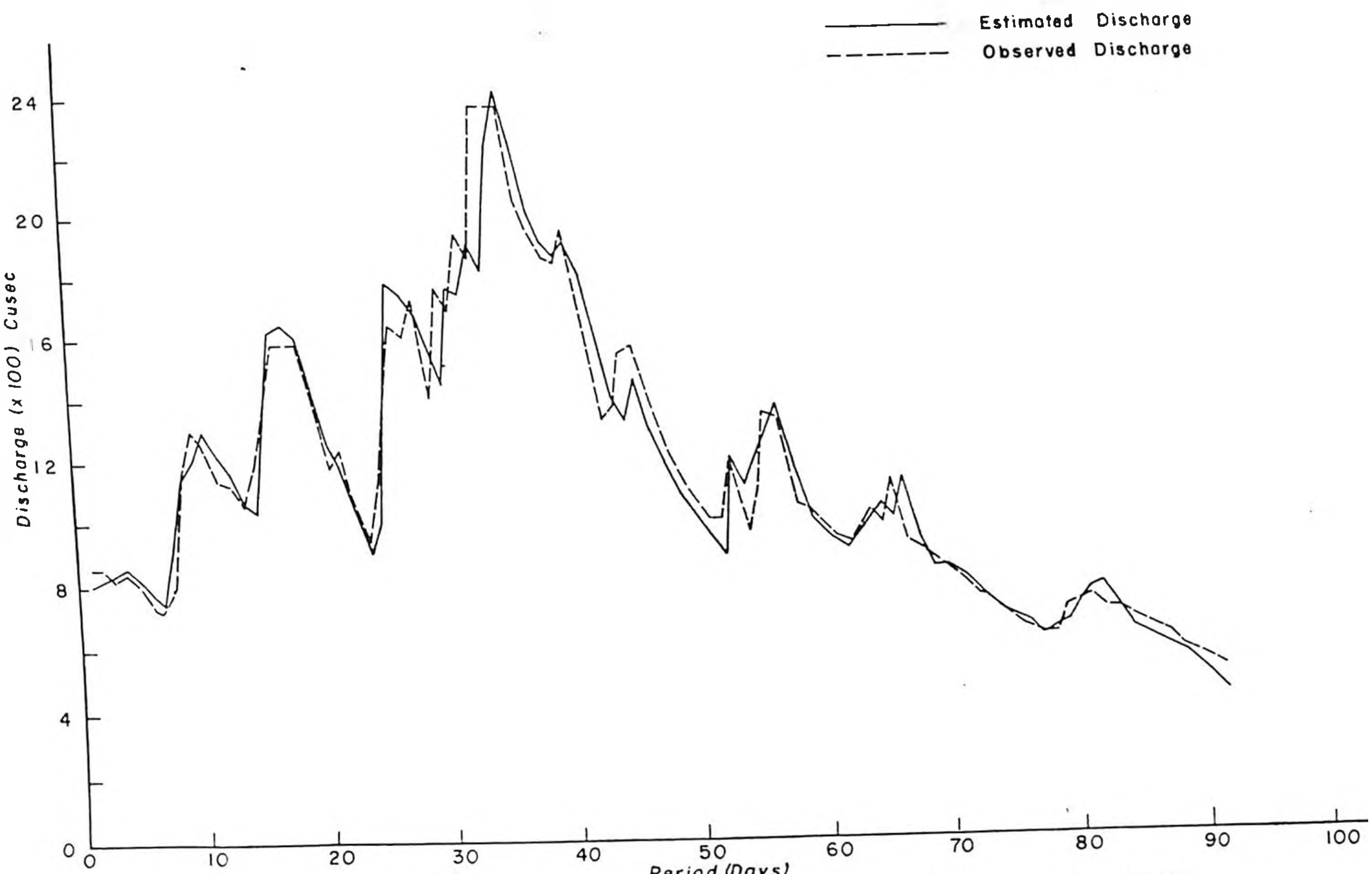


Fig 50 SEASON TWO YEAR TWO (1976) – PREDICTION BY THE T.F. 4.2 MODEL

(v) The autoregressive models, however, displaced flood peaks more than the response function model. This type of model has a disadvantage in that it would fail to accurately predict the actual time of occurance of a flood peak. Thus a warning of an expected flood based on such a model will give erroneous "lead time". This could make it very difficult for the authorities concerned to make precise decisions as to the actual time of evacuation of the people and property from the danger zone in case of an expected flood as it may occur earlier than expected and therefore take them by surprise. Such a mistake can lead to a loss of lives and property.

(vi) The response function models were more accurate in prediction of low-flows than the autoregressive models. A model which overpredicts low flows has a disadvantage in that a storage structure constructed on the basis of such a model could turn out to be too big as the actual flows will be lower than the predicted ones. This could lead to less water than was planned for, to flow downstream of the storage. In extreme cases the flow may completely stop downstream. This could paralyse all the downstream activities which are utilizing water from the river. On the other hand a model which underpredicts low flows may lead to overstorage which can even result into overbank flooding in the neighbourhood of the reservoir.

This can result into losses due to interuption of activities in the flooded plain and loss of water due to increased infiltration and evaporation.

Generally, the prediction of floods more accurately than lowflows would be better because the effects of floods are very disastrous and take place within a very short time. The accurate prediction of the time of occurance of a flood may even be more important than the accurate prediction of the flood magnitudes, more so if the errors in the estimation of the flood magnitudes are small. On the other hand, effects of inaccurate prediction of low-flows are not as serious as those of floods, they take place over a long period of time and they can be detected before it is too late. The other advantage of low flows is that if an inaccuracy in prediction is detected, precautions can be taken to moderate the effects of such an error. For example, in case of an overprediction, water use during the dry season can be regulated and thus reduce the effects caused by the overprediction.

It should be noted that the auto-regressive models would be the best in the estimation of the magnitudes of the peak discharges. When the standard errors are considered the T.F. (4, 2) model seemed to be the best auto-regressive model as it had the smallest standard error. The auto-regressive models however were poor in forecasting the actual time of occurance of the extreme flows.

Combining all these factors one may suggest that the response function models which give good estimates of the onsets of extreme flow would be preferable over the Yala basin.

3.7. ACHIEVEMENTS AND RECOMMENDATIONS

This study has provided an apportunity for two linear models to be fitted and compared for the Yala river basin rainfall - discharge system with a view of finding out which of the two is the best. The following recommendations have been suggested based on the results obtained in this study:

(i) Real time forecasting:

A model which accurately predicts the actual time of occurance of flood peaks and low flows would be recommended for real time forecasting which is useful in flood warning services. The results explained above indicate that the response function model can do better in real time forecasting than the auto-regressive model. Although the results showed that the response function model was under-predicting flood peaks more than the auto-regressive model, the effect of this drawback can be moderated by applying an updating procedure during the verification period. One example of an updating procedure is to fit an auto-regressive model on the residual errors in the calibration period of the form:

$$\acute{e}_i = a_1 e_{i-1} + a_2 e_{1-2} + \dots + a_n e_{i-n} \quad \text{----------} \quad (69)$$

where

$\hat{e}_i$ = updated residual error for date i

e_i = actual residual error for date i.

After the order of the above auto-regressive equation is determined and the constants $a_1 \dots a_n$ computed, the correction can then be applied to the forecasts during the verification period.

The response function model has an extra advantage in that it requires only rainfall values as inputs.

(ii) <u>Synthesis of data</u>

In synthesis of data for planning economic utilization of water discharges in a river, a model which accurately predicts the flood magnitudes and low flows would be more appropriate. The results from this study indicate that the auto-regressive model would do better than the response function model in this respect. One short-coming of the auto-regressive model is that in order to predict discharge on a given day, both discharge and rainfall data must be available on the previous days. Inspection of the daily records from the Yala river catchment showed that discharge data is very often not continuous as compared to rainfall. However this limitation may be over-come by using the models to estimate the previous missing runoff data required for forecasting. It should be noted however that this method of filling in the missing data should be used with caution as the estimated data will not be as error free as the observed data. The performance of the auto-regressive models in the prediction of the actual time of occurance of extreme flows can be improved by including a filter which can adjust the times of occurance as predicted by the model.

It should also be noted that the methods used to compute areal rainfall averages may not be accurate. Other methods based on empirical orthogonal factors etc., could be employed (Johnson 1980, Ogallo 1986).

CHAPTER FOUR

SUMMARY, CONCLUSIONS AND SUGGESTIONS FOR FURTHER WORK

4.1. SUMMARY AND CONCLUSIONS

Two simple linear models have been fitted to different sets of concurrent daily areal rainfall and discharge series from the Yala river basin. Two wet seasons extending from April to June and August to October over the years 1968 to 1974 have been considered during the calibration of the models.

Before fitting the models, it was necessary to investigate the statistical properties of the daily areal rainfall and discharge series for the different seasons through time series analysis. The functions which have been computed comprise of the autocorrelation, cross-correlation, spectrum and cross-spectrum. The results have indicated that the daily areal rainfall and daily discharge series have almost similar statistical properties, indicating a close similarity between rainfall and discharge series. The dominant periodicities existing in both the areal rainfall and discharge series had dominant peaks centred within 2 - 10 days.

In the first approach optimum response functions have been computed for the actual daily areal rainfall and discharge series. In the second approach three different orders namely, T.F (2,2), T.F. (3,2), and T.F. (4.2) have been fitted to the long term means of daily areal rainfall and daily discharge series. Statistical tests were carried out to check the validity of the transfer function models. To find out the accuracy of the response functions and the transfer functions, different rain seasons have been selected and used for the estimation of discharge. The estimated discharge was compared to the actual discharge during those seasons. The rain seasons used in testing the skill of the models were chosen in such a way that they were different from those used during the calibration stage of the models.

The results have shown that the variations in the response functions for the different seasons are small and the values of the coherence function are also relatively high. This may indicate that the rainfall - discharge system for the Yala river basin behaves in an almost linear fashion.

The forecasts compared well to the actual observations for both models. However persistent departures of the forecasts from the actual hydrographs were obtained for some events of heavy rainfall.

The Smirnov-Kolmogorov and multiple correlation coefficient tests have been used to test the goodness of fit of the discharge estimated by the models. The results showed that all the models have good fits at the 5% significance level.

The results of the computed standard error of estimate have shown that the autoregressive models with four rainfall and two discharge lag terms (T.F. (4,2)) had the smallest standard error in all cases.

4.2. SUGGESTIONS FOR FURTHER WORK

The study was based on a relatively short period, therefore the performance of the models may not be always accurate. Thus models similar to these but based on longer durations, say 30 years or more should be constructed.

The method of using the two sets of three calender months each to define the two wet seasons may not have defined the two seasons correctly. So another study can be carried out in which the wet seasons are better defined like considering the daily rainfall totals and the number of dry days between any given two wet days as some of the conditions for defining the seasons.

Lastly the model forecasts may be improved by applying an updating procedure during the verification period. An example of an updating procedure is to fit an autoregressive model on the residual errors in the calibration period of the form:

$$\hat{e}_i = a_1 e_{i-1} + a_2 e_{i-2} + \dots + a_n e_{i-n} \quad \text{-------- 70}$$

where:

$\hat{e}_i$ = updated residual error for date i

e_i = actual residual error for date i

After the order of the auto-regressive Equation (70) is determined and the constants a_1 .... a_n computed, the correction can then be applied to the forecasts during the verification period. Inclusion of a filter which adjusts the times of occurance of the flood peaks and low flows can further improve the performance of the models.

The methods used to compute areal rainfall averages may not be very accurate. Other methods based on empirical orthogonal factors etc. could be employed in another study.

ACKNOWLEDGEMENTS

I am greatly indebted to Dr. L.J. Ogallo and Mr. F.M. Mutua both of the University of Nairobi for their guidance and encouragement as University Supervisors during the preparation of this work.

I owe special thanks to the German Academic Exchange Service (DAAD) for providing to me the financial support in the course of my studies at the University of Nairobi.

I would also wish to express my gratitude to the Ministry of Water Development and the Kenya Meteorological Department for making the data used in this study accessible.

Mrs. Kamau deserves special thanks for her patience while typing the manuscript. Thanks are also extended to Mr. E. Okumu for drafting the diagrams in this work.

Finally, I am indebted to my wife Beatrice and daughters Damaris and Catherine for their patience and continued moral support in the course of the study.

REFERENCES

Amorocho, J., (1963): Measures of the linearlity of Hydrologic systems. J. Geophy. Res., Vol.68, No.8, 2237-2249

Anderson, T.W., (1971): The statistical analysis of time series. John Wiley and Sons, Inc.

Asnani, G.C. and J.H. Kinuthia, (1979): Diurnal variation of precipitation in East Africa. E.A. Met. Dept. memo No.8

Basalirwa, C.P.K., (1980): Estimation of areal rainfall in some catchments of the Upper Tana river. M.Sc Thesis University of Nairobi, Kenya.

Bennett, R.J., (1979): Spatial time series; Analysis, Forecasting and Control. Pion Limited.

Blackie, J.R., (1972): The application of conceptual models to two East African Catchments. M.Sc Thesis Imperial College of Science and Technology

Box, G.E.P. and G.M. Jenkins, (1968): Some recent advances in forecasting and control. J. Roy. Stat. Soc., C16, 91-109

Box, G.E.P. and G.M. Jenkins, (1970): Time series analysis, forecasting and control. Holden-Day. Inc.

Chatfield, C., (1975): The analysis of time series.

Chow, Y.T., (1964): Handbook of applied hydrology. Ed. Y.T. Chow, McGraw Hill N.Y.

Clark, C.L. and W. Brutsaer, (1978): A non-linear analysis of the relationship between Rainfall and Runoff for extreme floods. Water Res. Res., Vol.14, No.1, 75-83.

Clarke, R.J., (1973): Mathematical models in hydrology. Irrigation and Drainage paper NO.19, FAO, Rome.

Craig, J.I., (1916): The periodogram and method of correlation. Brit. Ass. adv. Sc. Rept., 86

Delleur and Rao, (1971): Linear Systems analysis in hydrology - The transform approach, the kernel oscillations and the effect of Noise. Systems Approach to hydrology, 116-128.

Diskin, M.H. and A. Boneh, (1972): properties of Kernels for time invariant, initially relaxed second order surface runoff systems. J. Hydrology Vol.17, 115-141.

Dooge, J.C.I., (1960): A general theory of the unit hydrograph. J. Geophy. Res., Vaol.64, No.2, 241-256.

Eagleson, P.S., Mejia, R. and F. March, (1966): Computation of optimum realizable unit hydrographs. Water Res. Vol. 2, NO.4, 755-764

Granger, C.W.J. and M. Hatanaka, (1964): Spectral analysis for economic time series. Princeton University Press.

Hino, M., (1970) : Runoff forecasts by linear predictive filter. J. Hydrology Div., ASCE, Vol.96, No.HY3, 681-702.

Johnson, D.M., (1980): An index of Arizona summer rainfall development through Eigen Vector Analysis. J. App. met., Vol.19, 849-856

Kato, P., (1982): Rainfall Runoff relationships in the Ruvu Basin of Tanzania. M.Sc. Thesis, University of Nairobi, Kenya

Kovacs, G. and H.T. Morth, (1974): Tech. memo. NO.20. East African Meteorological Department, Nairobi.

Mutua, F.M. (1979): Rainfall-Runoff model for the Nzoia Drainage Basin. M.Sc. Thesis, University of Nairobi, Kenya

Mutulu, P.M. (1984): Stochastic modelling of rainfall-runoff data for Sondu and Yala River Basins. M.Sc. Thesis University of Nairobi, Kenya

Nash J.E., (1959): Systematic determination of unit hydrograph parameters. J. Geo. Res., Vol.64, No.1, 111-114.

Nash, J.E. and J.J. Foley, (1982): Linear models of rainfall - Runoff Systems. Water Resources PUblications.

O'Donell, T., (1966): Methods of Computation in hydrograph analysis and synthesis App. Sci. Res., 65-103

Ogallo, L.J. (1986): Stochastic modelling of regional annual rainfall anomalities in East Africa. J. App.Stat. Vol. 13, No.1., 49-56.

Priestley, M.B., (1981): Spectral analysis and time series, Volume 1. Academic press.

Sherman, L.K., (1932): Streamflow from rainfall by the unit hydrograph method. Eng. News-Record, Vol.108 501-505

TAMS, (1977): TAMS Rainfall - Runoff model National master water plan, stage 1 working paper No.4. Ministry of water development, Kenya.

TAMS, (1980): TAMS - National Master water plan, stage 2, Vol. 2 Ministry of water development, Kenya.

Tavares, L.V., (1975): Watershed runoff simulation; A comparative study of main methods. In Computer simulation of water Resources systems Ed. G.C. Vansteekiste (NHPC).

Todini, E. and D. Bouillot, (1975): A rainfall-runoff kalman filter model. In system simulation in water resources. Ed. G.C. Vansteekiste (NHPC).

Tojiro, I. and S. Ikebuchi, (1971): A study of long range runoff system response based on information theory. Systems approach to Hydrology., 80-101

Tomset, J.E., (1975): The diurnal variation of precipitation in East Africa. E.A. Met. Dept. Tech. memo. No.25

Walker, C.J., (1931): The periodicities in series of related terms. Proc. Roy. Soc., A131, 518

Whitehead, P and P. Young, (1975): A dynamic stochastic model for water quality in part of the Bedford-ouse river system. In Computer simulation of water resources systems. Ed. g.c. vansteekiste, (NHPC).

Wiener, N., (1949): _Extrapolation, interpolation and smoothing of stationary time series_. John Wiley, New York.

Yevjevich, V. (1972): _Stochastic processes in Hydrology_. Water resources publications.

Yevjevich, V. (1972): _Probability and Statistics in Hydrology._ Water resources publications.

Yule, G.V., (1927): On a method of investigating periodicities in disturbed series with special reference to wolfer's sunspot numbers. _Trans. Roy. Soc._, A, 226-267

APPENDIX A: EXTRA EXAMPLES OF CORRELOGRAMS

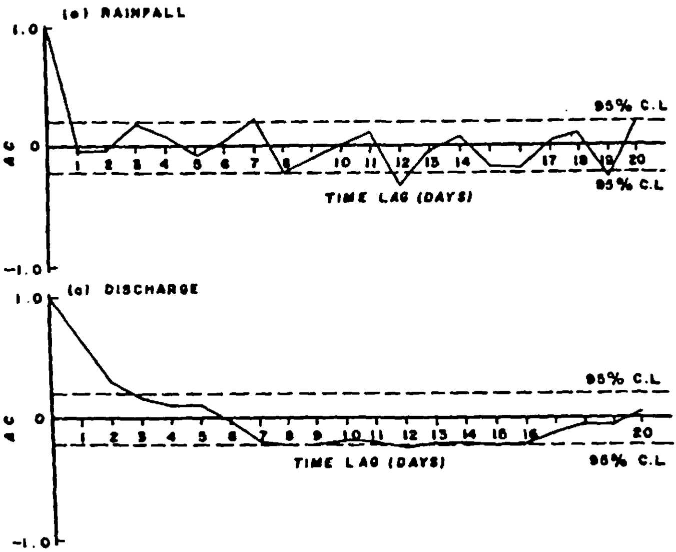


Fig 51 CORRELOGRAMS FOR SEASON ONE YEAR ONE 1968

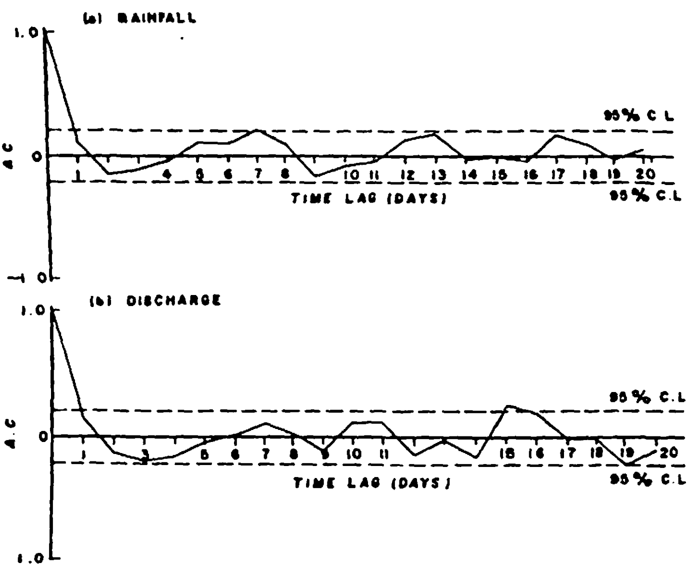


Fig 53 CORRELOGRAMS FOR SEASON ONE YEAR FOUR 1971

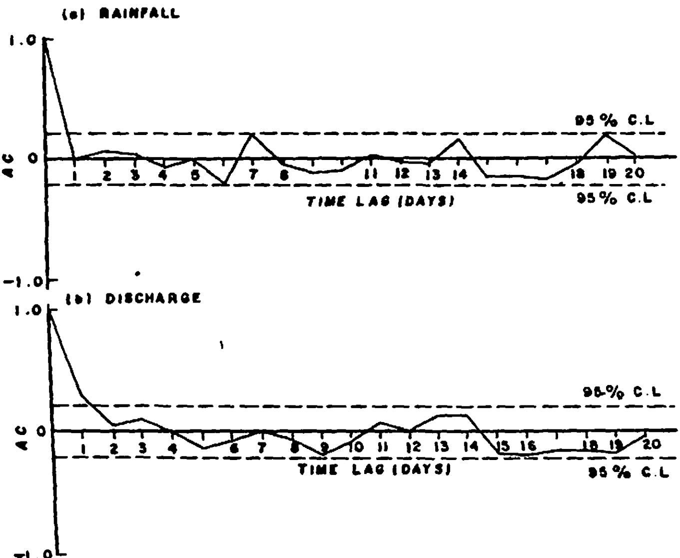


Fig. 52 CORRELOGRAMS FOR SEASON ONE YEAR THREE 1970

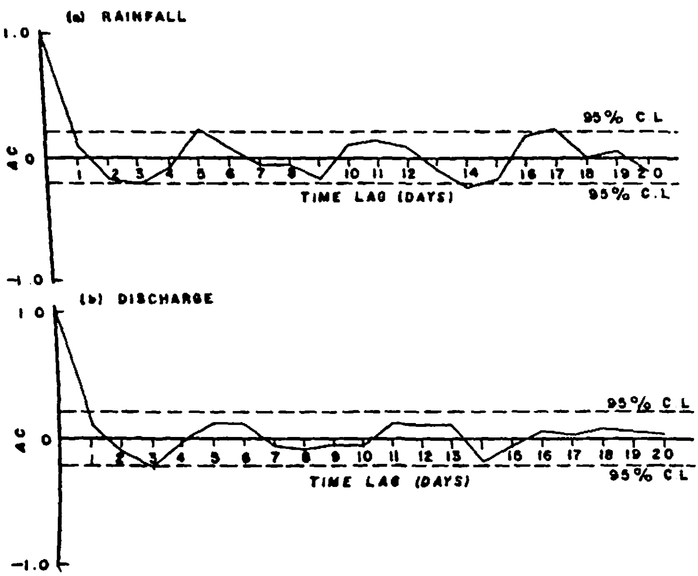


Fig. 54 CORRELOGRAMS FOR SEASON ONE YEAR SIX 1973

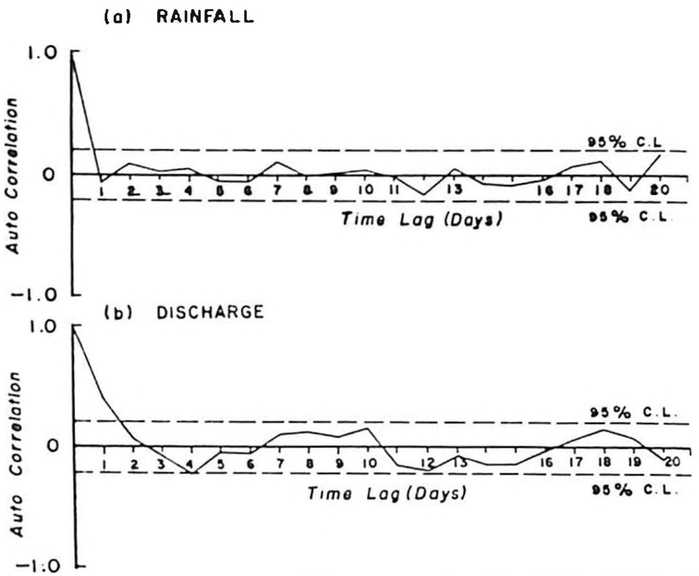


Fig. 55 CORRELOGRAMS FOR SEASON TWO YEAR ONE 1968

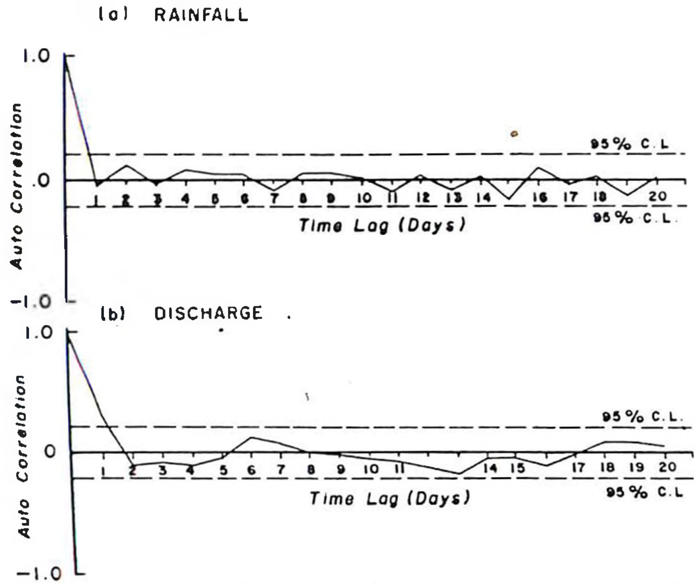


Fig. 56 CORRELOGRAMS FOR SEASON TWO YEAR TWO 1969

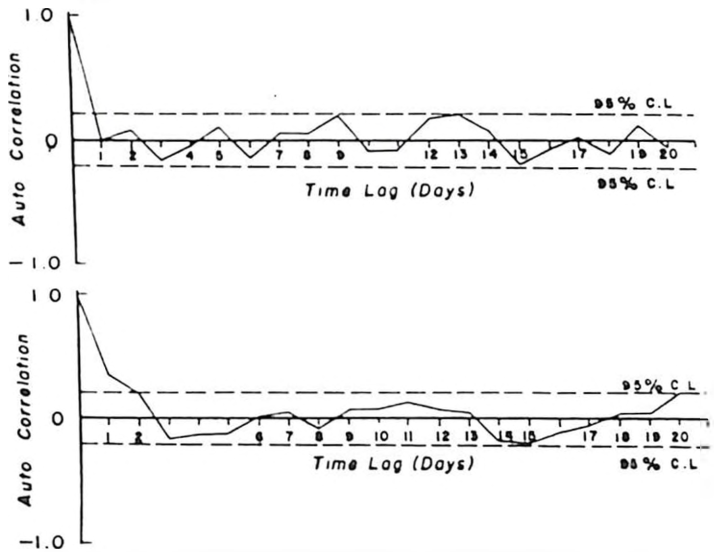


Fig 57 CORRELOGRAMS FOR SEASON TWO YEAR SIX 1973

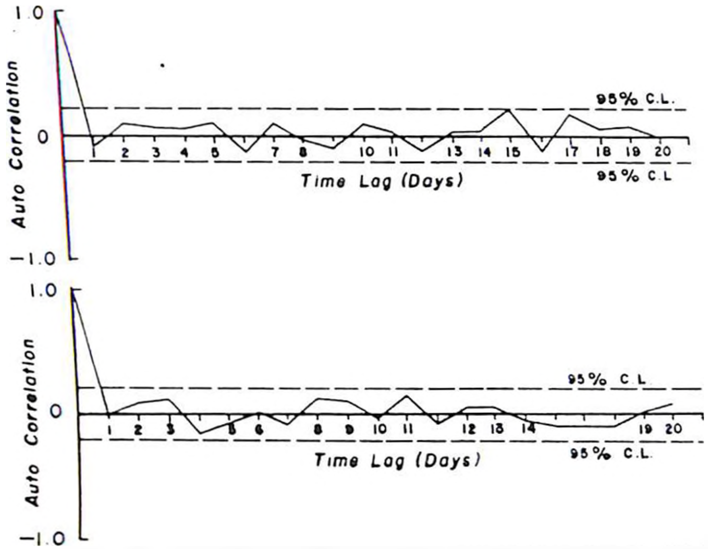


Fig. 58 CORRELOGRAMS FOR SEASON TWO YEAR SEVEN 1974

APPENDIX B: Extra examples of spectral functions

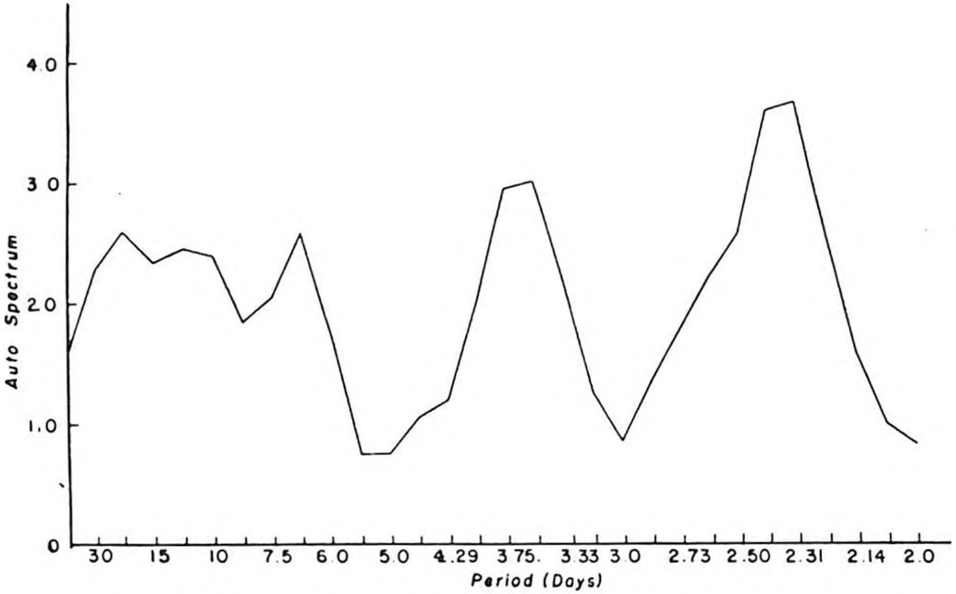


Fig 59 RAINFALL AUTO SPECTRUM FOR SEASON ONE IN YEAR THREE 1970

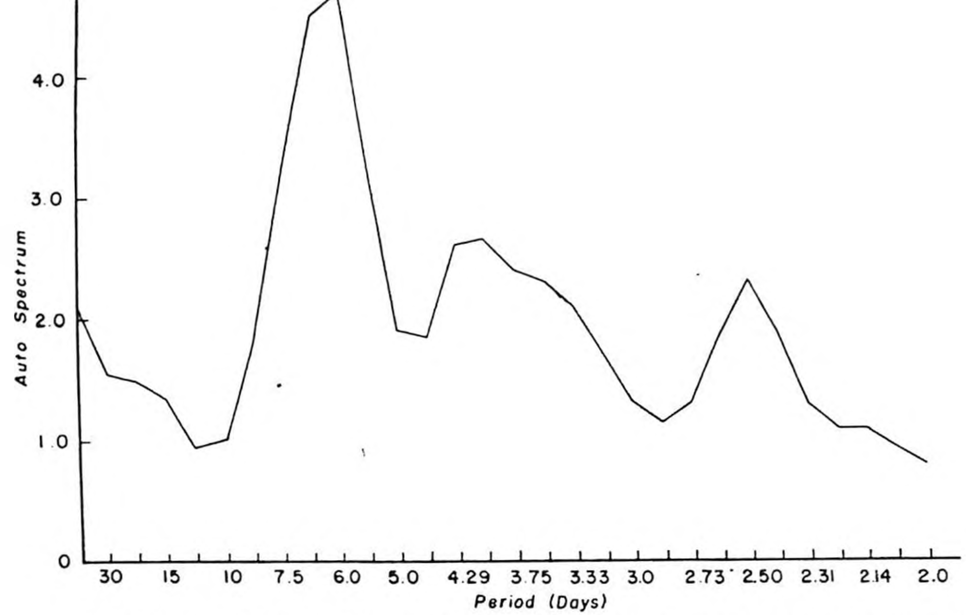


Fig. 60 RAINFALL AUTO SPECTRUM FOR SEASON ONE IN YEAR FOUR 1971

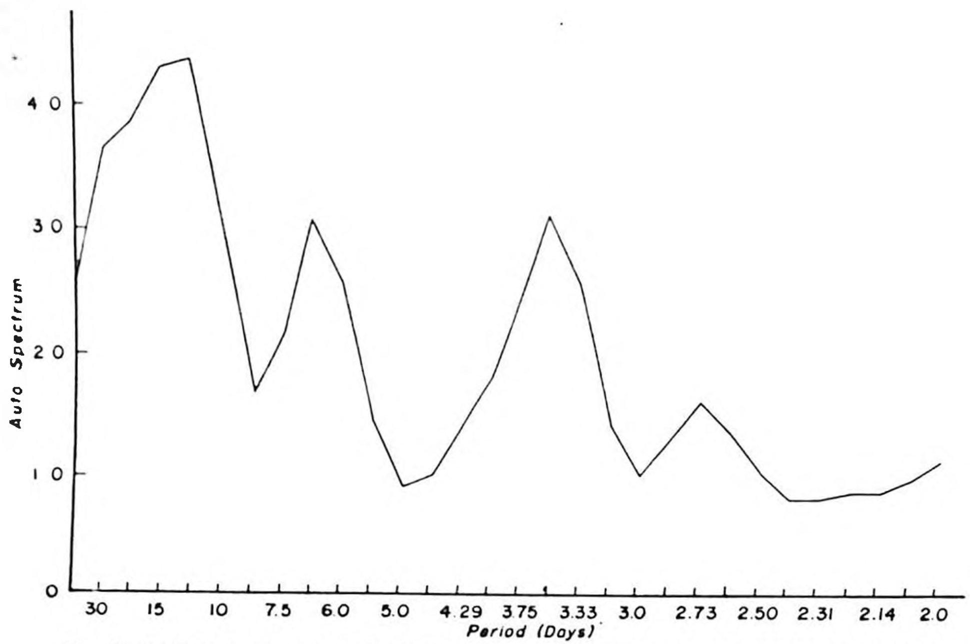


Fig. 61 DISCHARGE AUTO SPECTRUM FOR SEASON ONE IN YEAR THREE 1970

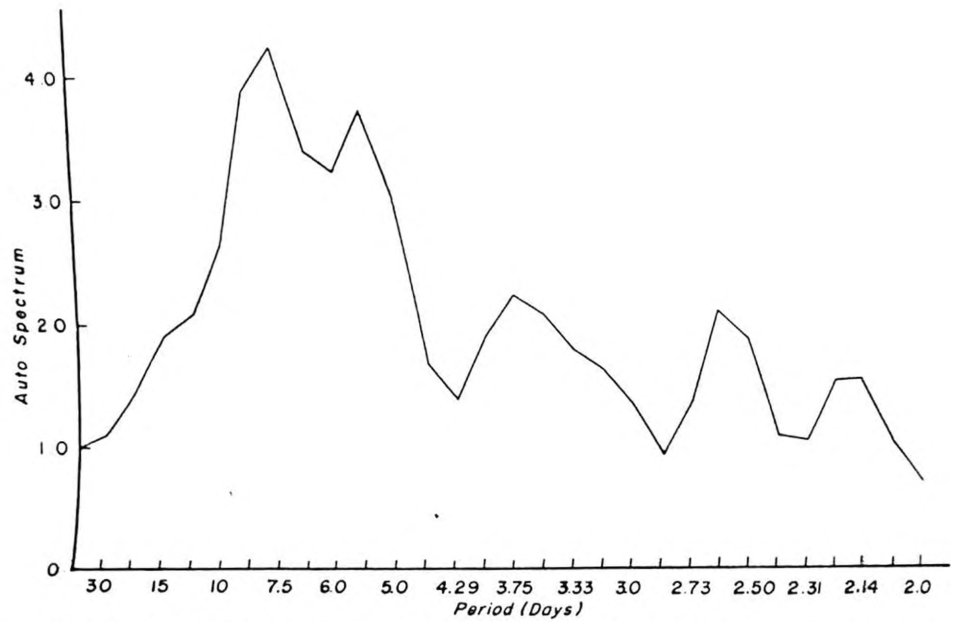


Fig.62 DISCHARGE AUTO SPECTRUM FOR SEASON ONE IN YEAR FOUR 1971

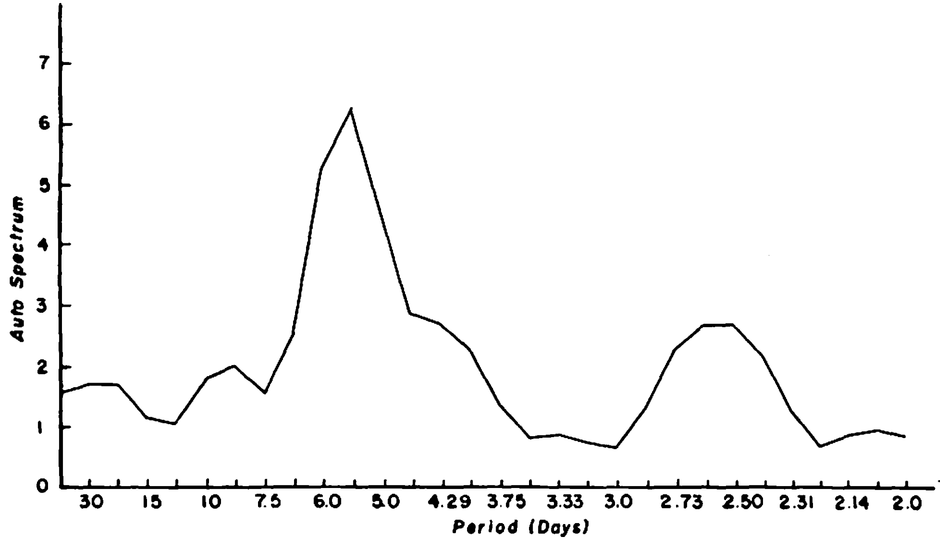


Fig. 63 RAINFALL AUTO SPECTRUM FOR SEASON ONE IN YEAR SIX 1973

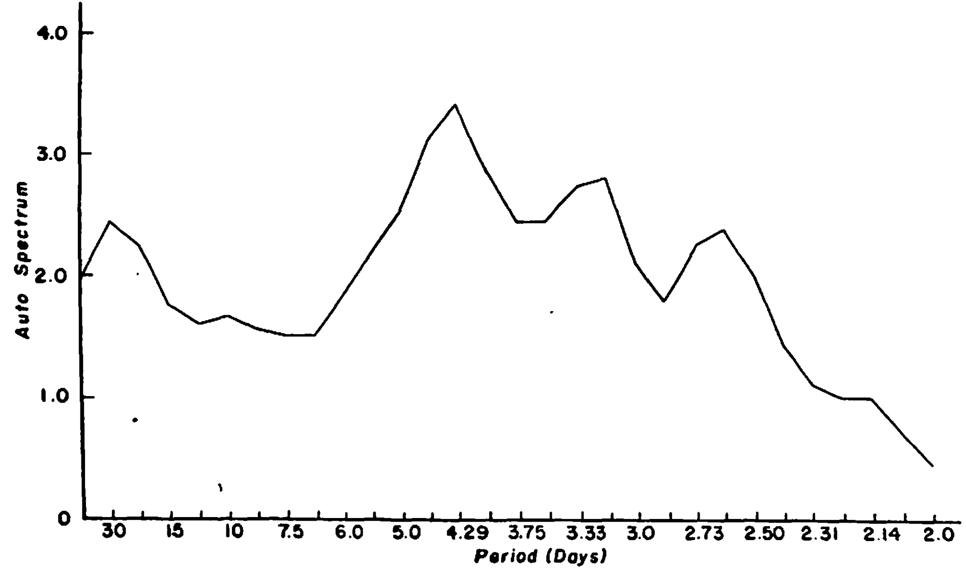


Fig. 64 RAINFALL AUTO SPECTRUM FOR SEASON ONE IN YEAR SEVEN 1974

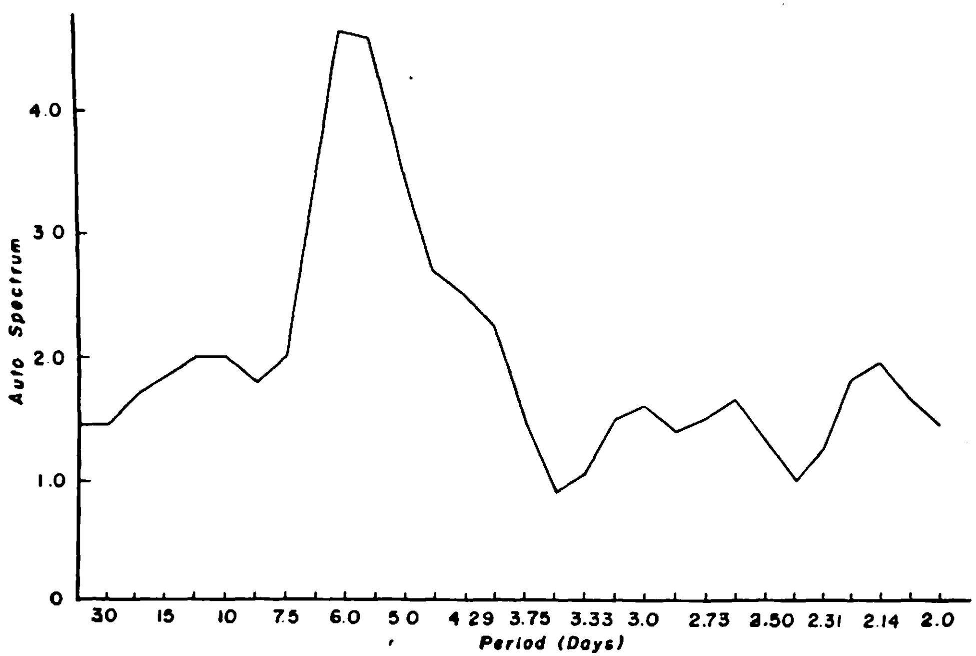


Fig. 65 DISCHARGE AUTO SPECTRUM FOR SEASON ONE IN YEAR SIX 1973

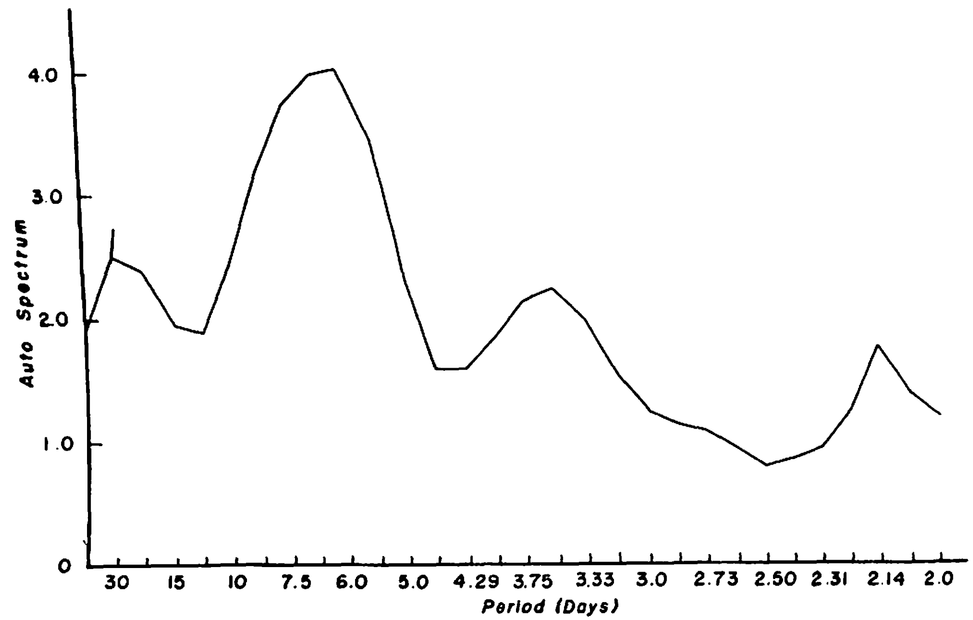


Fig. 66 DISCHARGE AUTO SPECTRUM FOR SEASON ONE IN YEAR SEVEN 1974

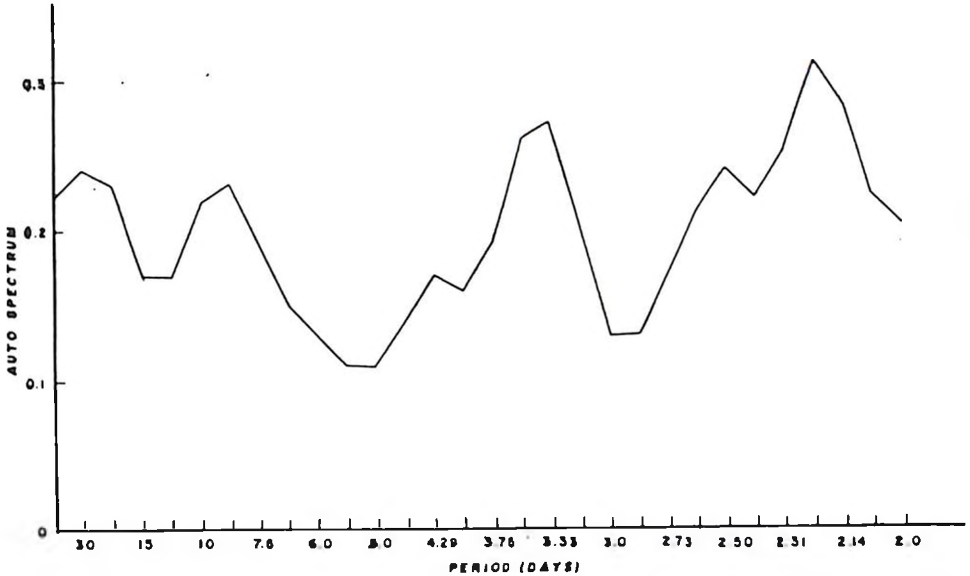


Fig 67 RAINFALL AUTO SPECTRUM FOR SEASON TWO IN YEAR ONE 1968

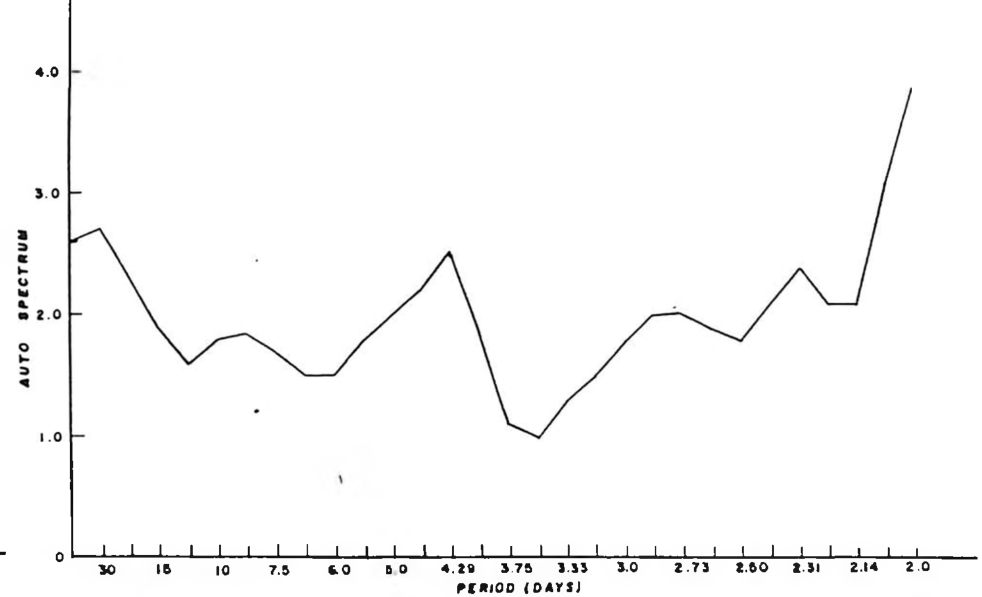


Fig. 68 RAINFALL AUTO SPECTRUM FOR SEASON TWO IN YEAR TWO 1969

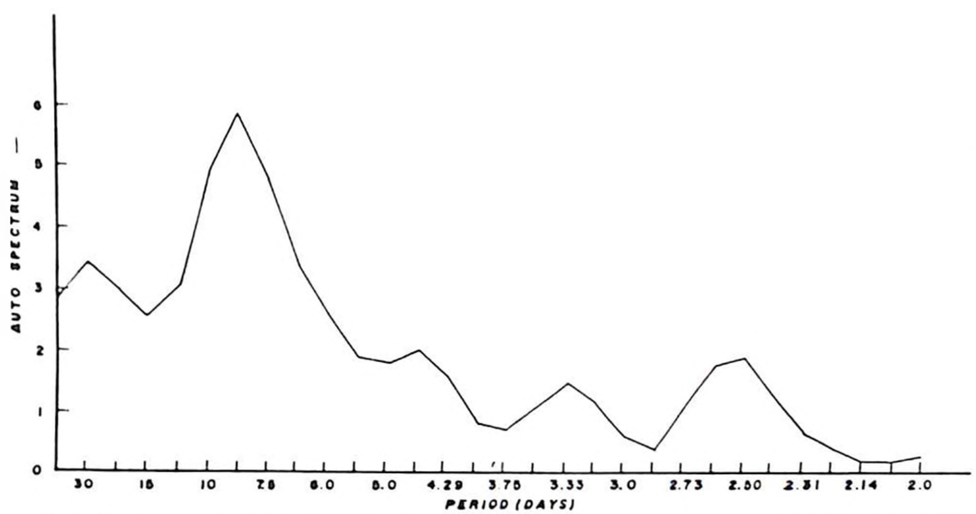


Fig.69 DISCHARGE AUTO SPECTRUM FOR SEASON TWO IN YEAR ONE. 1968

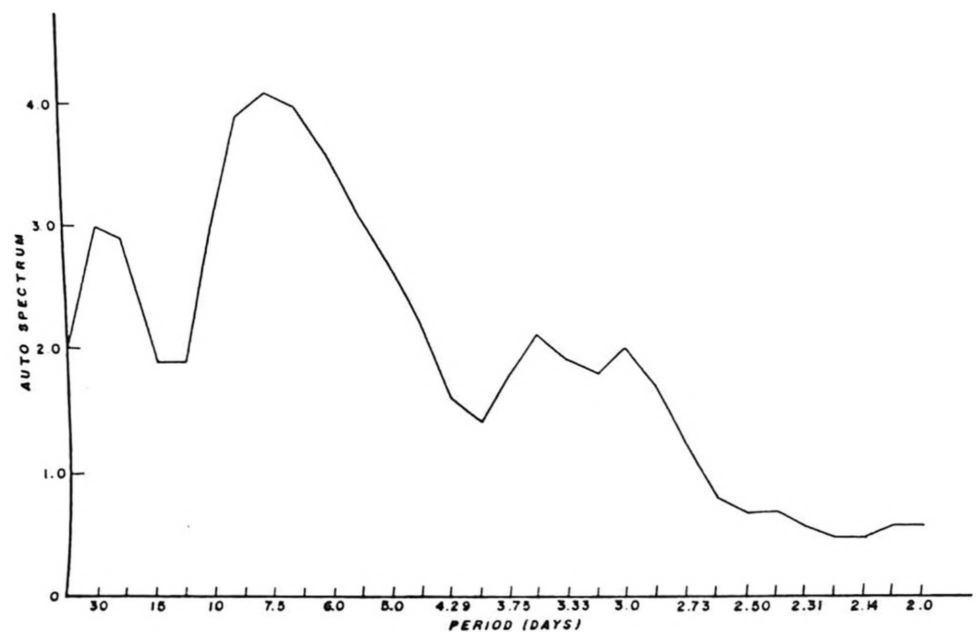


Fig. 70 DISCHARGE AUTO SPECTRUM FOR SEASON TWO IN YEAR TWO 1969

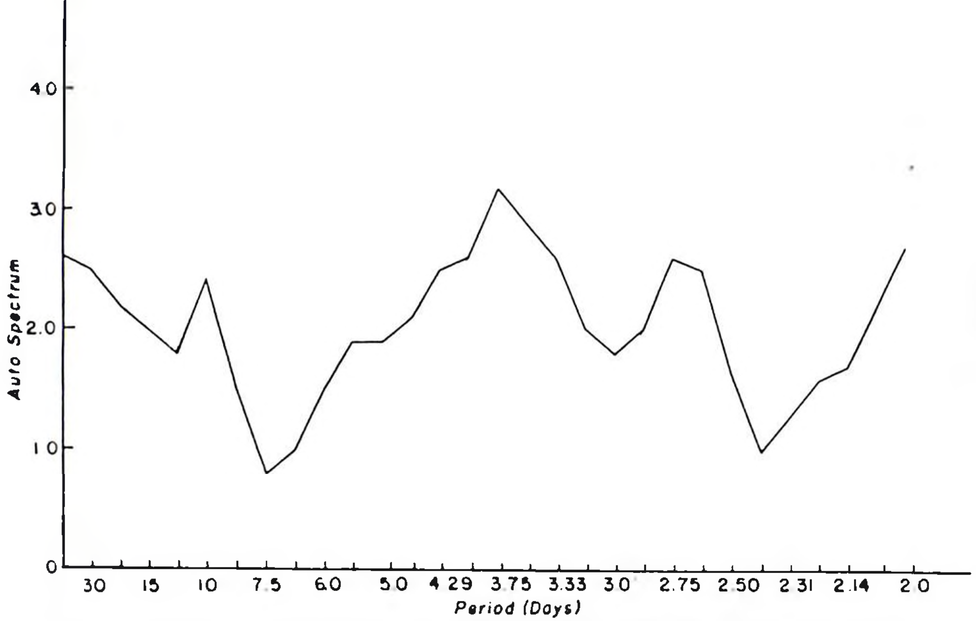


Fig 71 RAINFALL AUTO SPECTRUM FOR SEASON TWO IN YEAR FOUR 1971

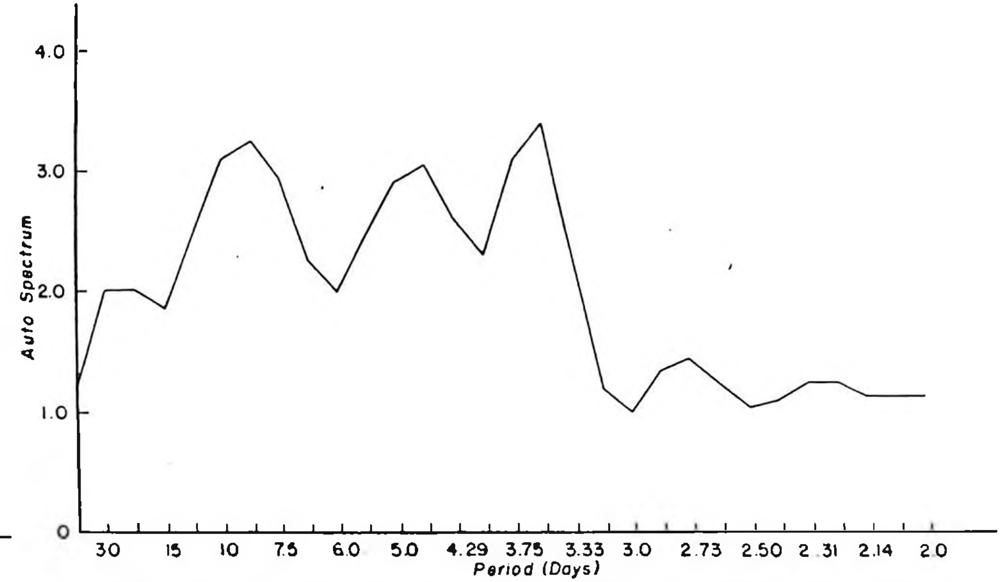


Fig. 72 RAINFALL AUTO SPECTRUM FOR SEASON TWO IN YEAR FIVE 1972

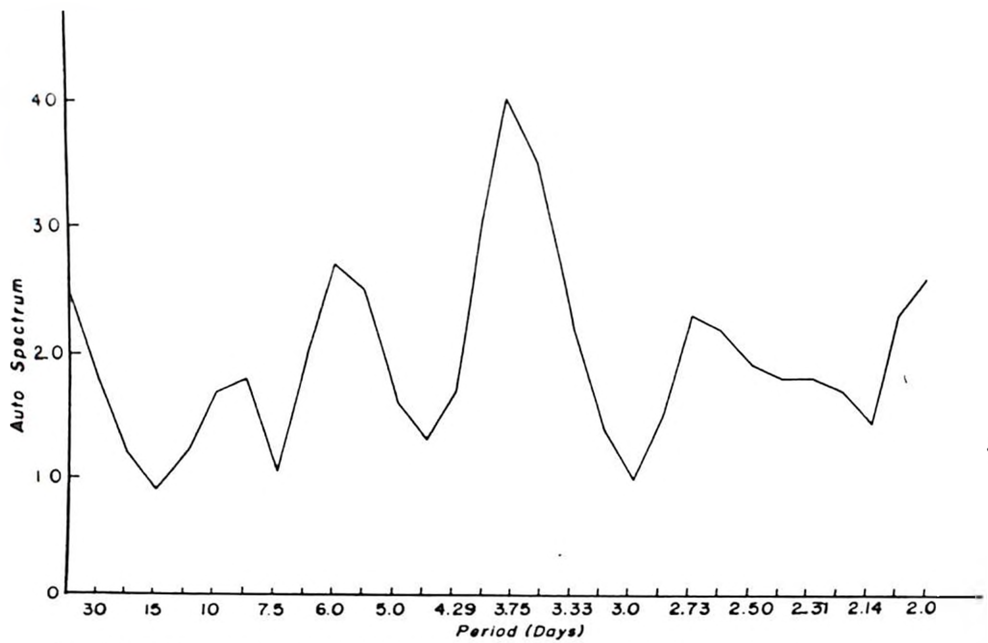


Fig. 73 DISCHARGE AUTO SPECTRUM FOR SEASON TWO IN YEAR FOUR 1971

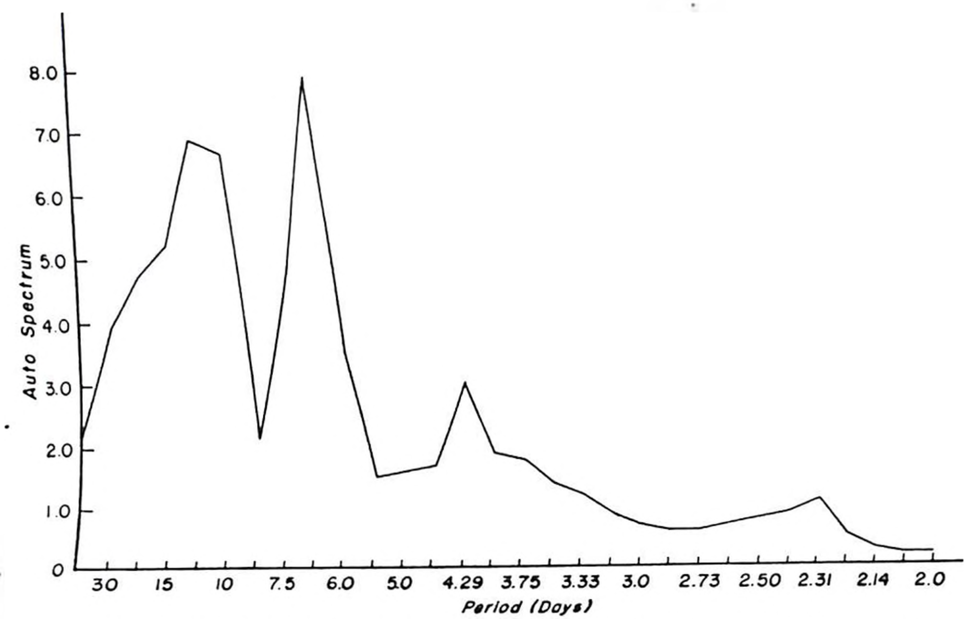


Fig. 74 DISCHARGE AUTO SPECTRUM FOR SEASON TWO IN YEAR FIVE 1972

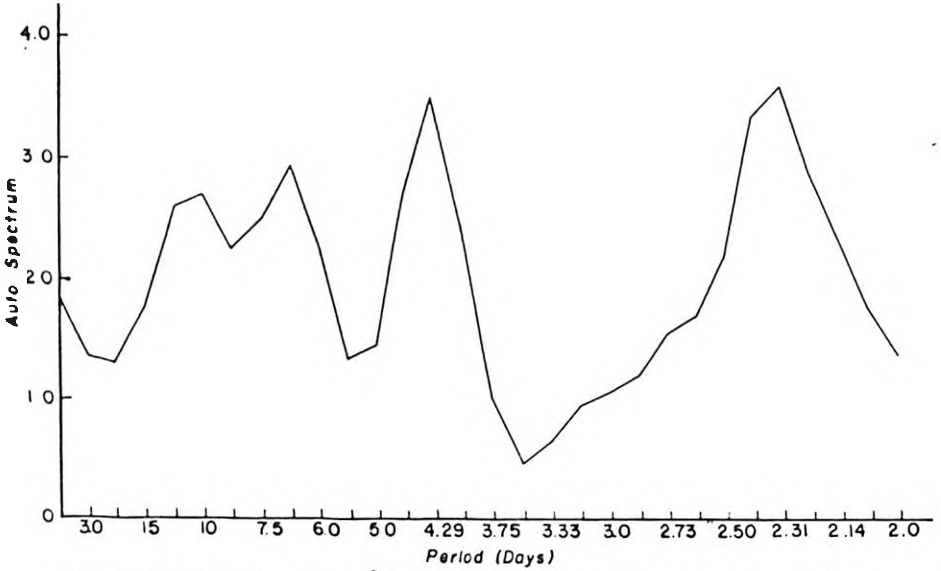


Fig. 75 RAINFALL AUTO SPECTRUM FOR SEASON TWO IN YEAR SIX 1973

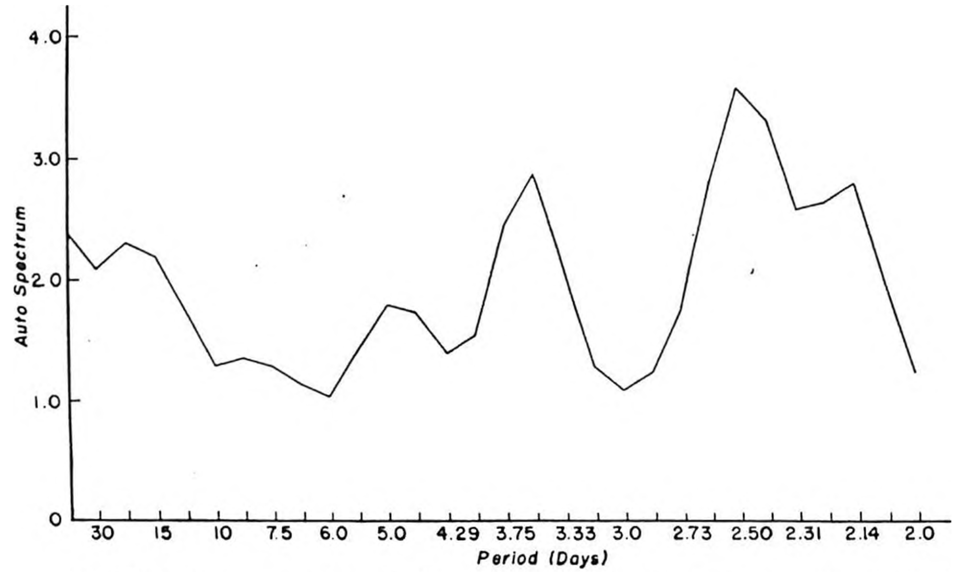


Fig. 76 RAINFALL AUTO SPECTRUM FOR SEASON TWO IN YEAR SEVEN 1974

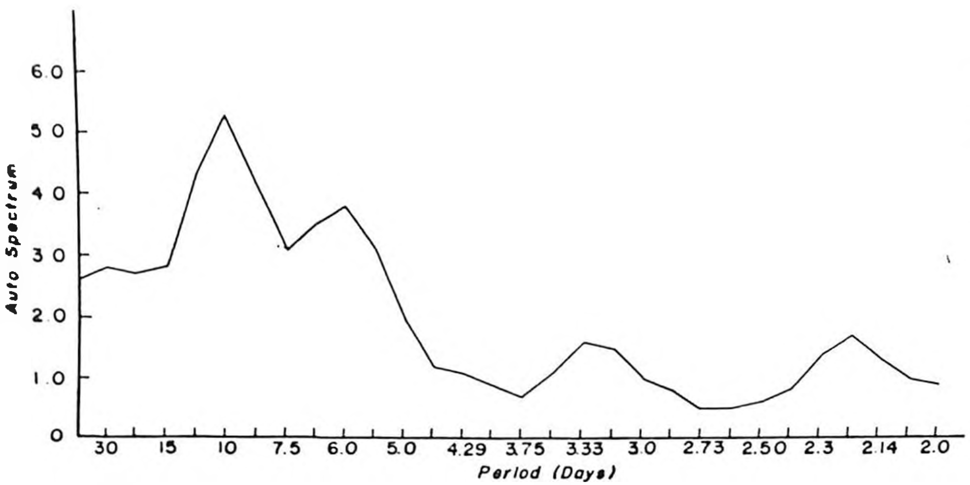


Fig. 77 DISCHARGE AUTO SPECTRUM FOR SEASON TWO IN YEAR SIX 1973

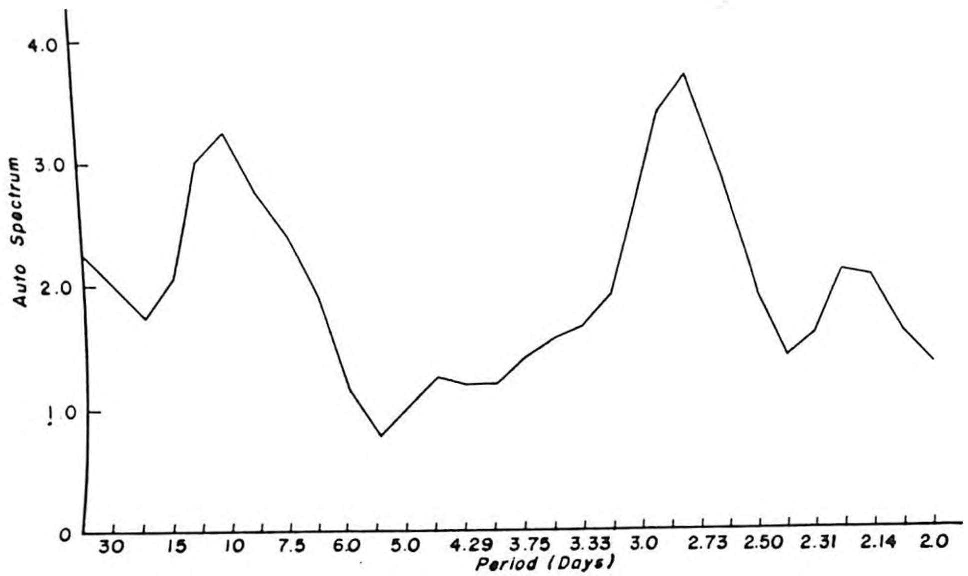


Fig. 78 DISCHARGE AUTO SPECTRUM FOR SEASON TWO IN YEAR SEVEN 1974

APPENDIX C: Extra examples of cospectrum

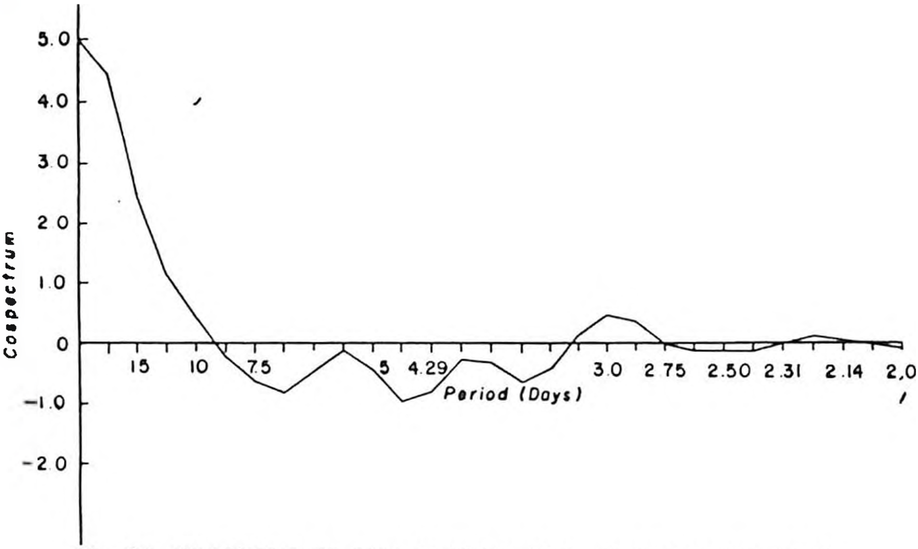


Fig. 79 COSPECTRUM FOR SEASON ONE IN YEAR ONE 1968

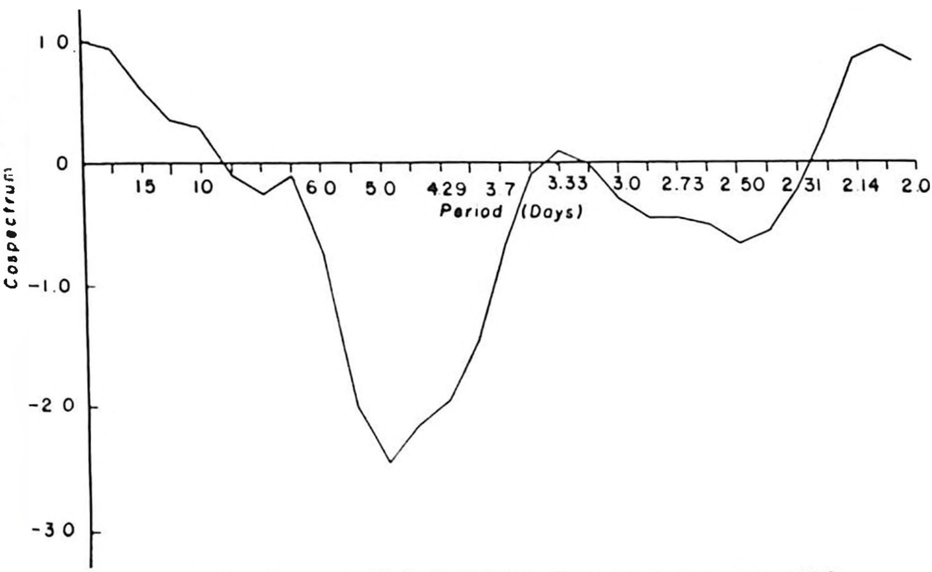


Fig. 81 COSPECTRUM FOR SEASON ONE IN YEAR SIX 1973

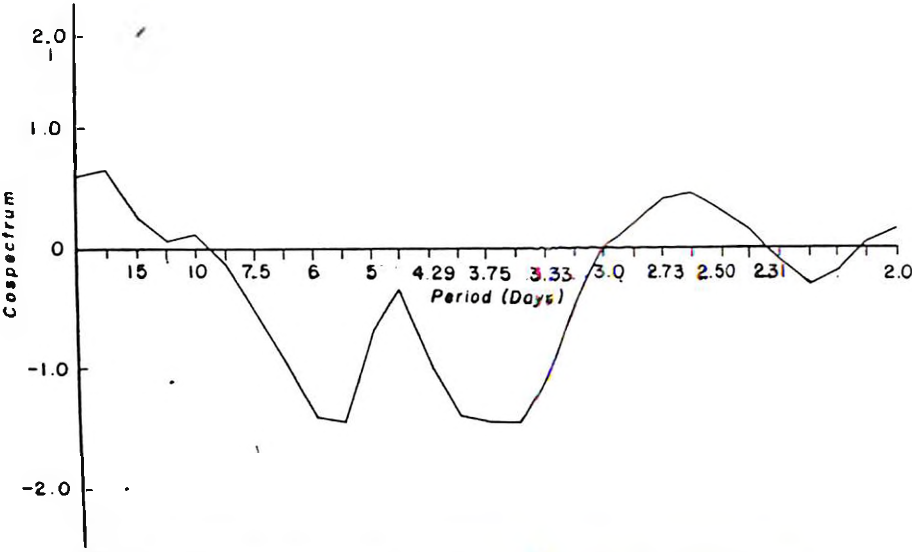


Fig. 80 COSPECTRUM FOR SEASON ONE IN YEAR FOUR 1971

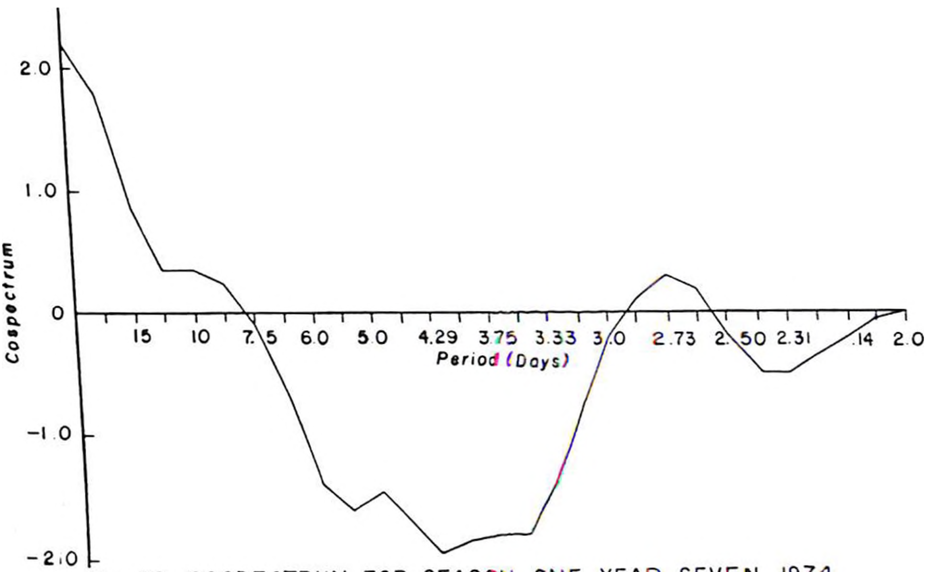


Fig. 82 COSPECTRUM FOR SEASON ONE YEAR SEVEN 1974

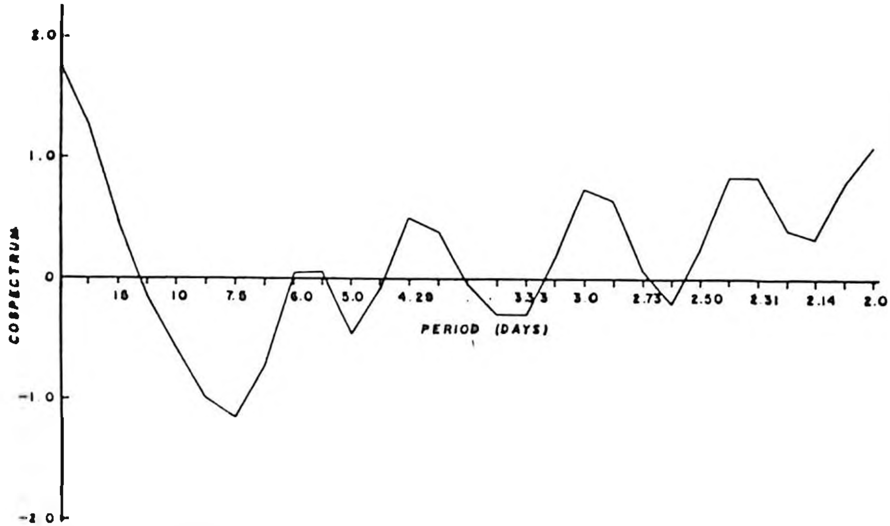


Fig. 83 COSPECTRUM FOR SEASON TWO IN YEAR TWO 1969

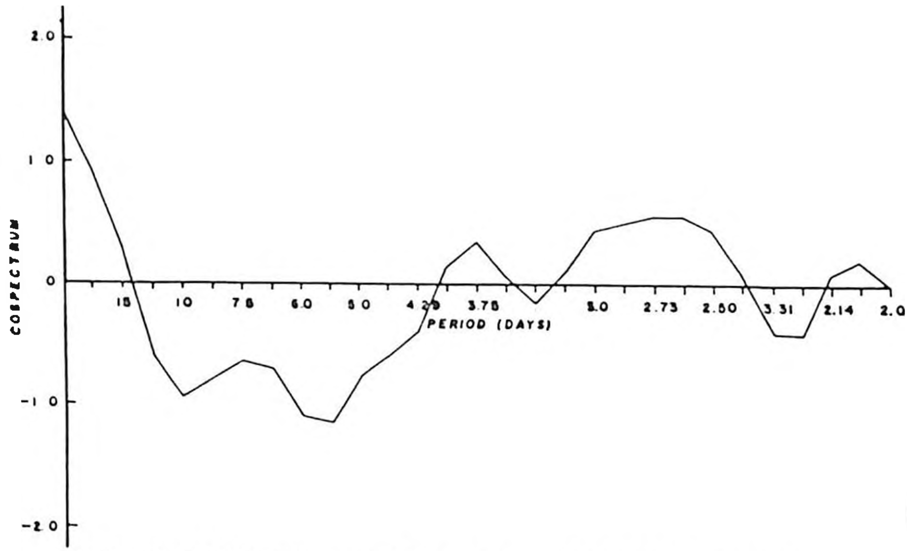


Fig 85 COSPECTRUM FOR SEASON TWO IN YEAR SIX 1973

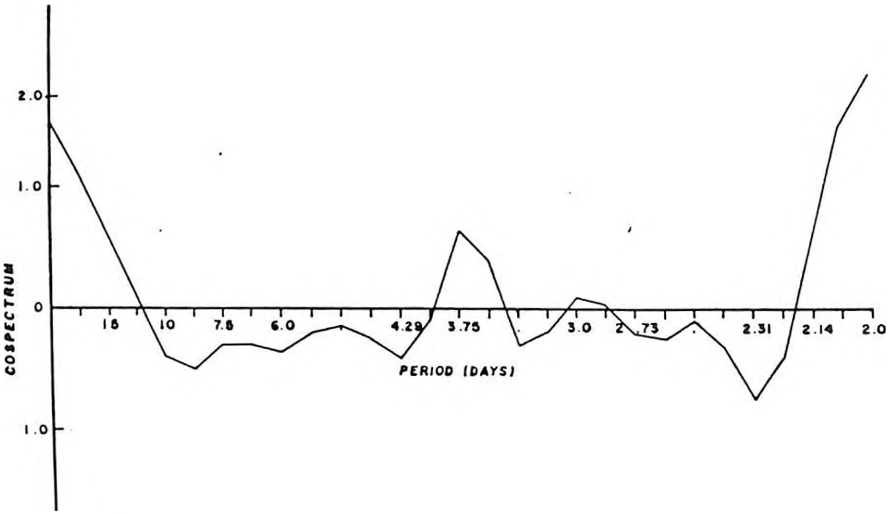


Fig. 84 COSPECTRUM FOR SEASON TWO IN YEAR FOUR 1971

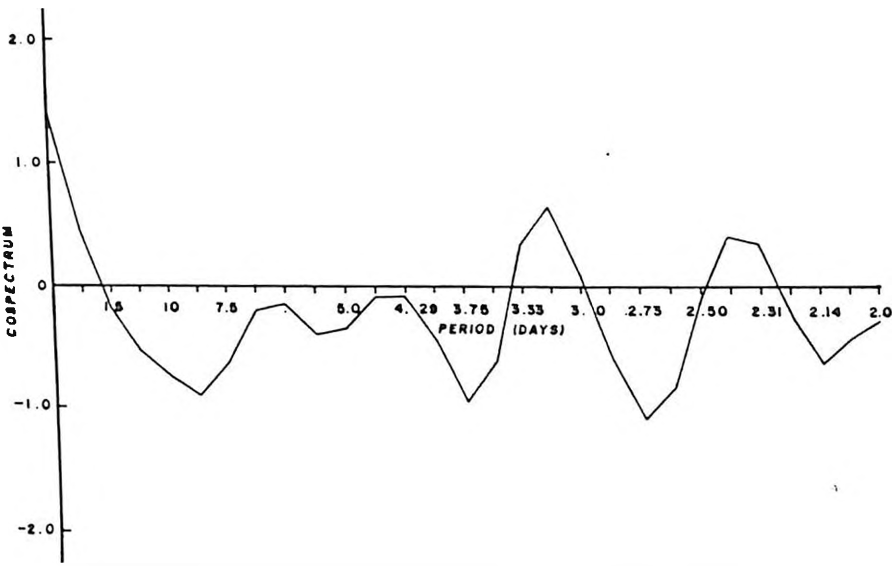


Fig. 86 COSPECTRUM FOR SEASON TWO IN YEAR SEVEN 1974

APPENDIX D: Extra examples of coherence function

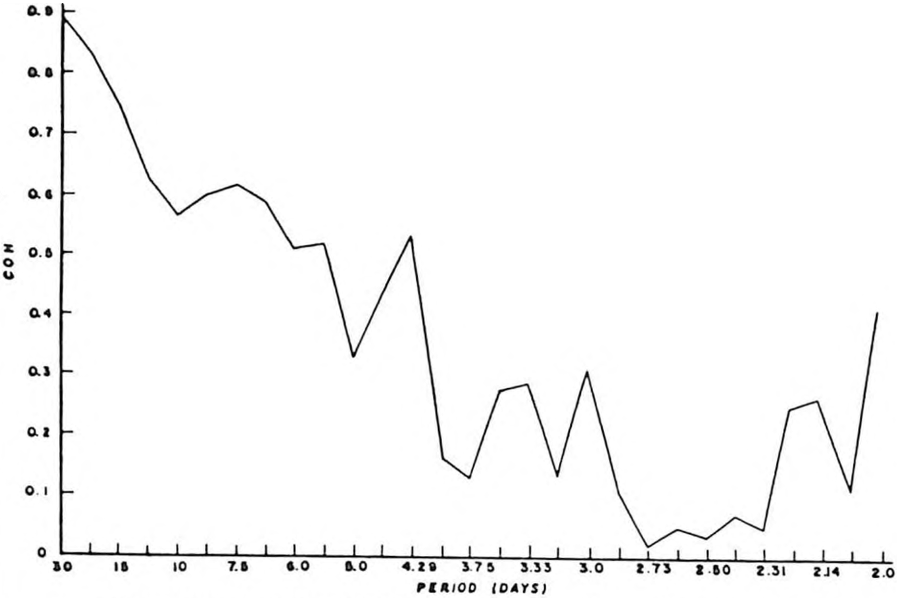


Fig. 87 COHERENCE FUNCTION FOR SEASON ONE IN YEAR ONE 1968

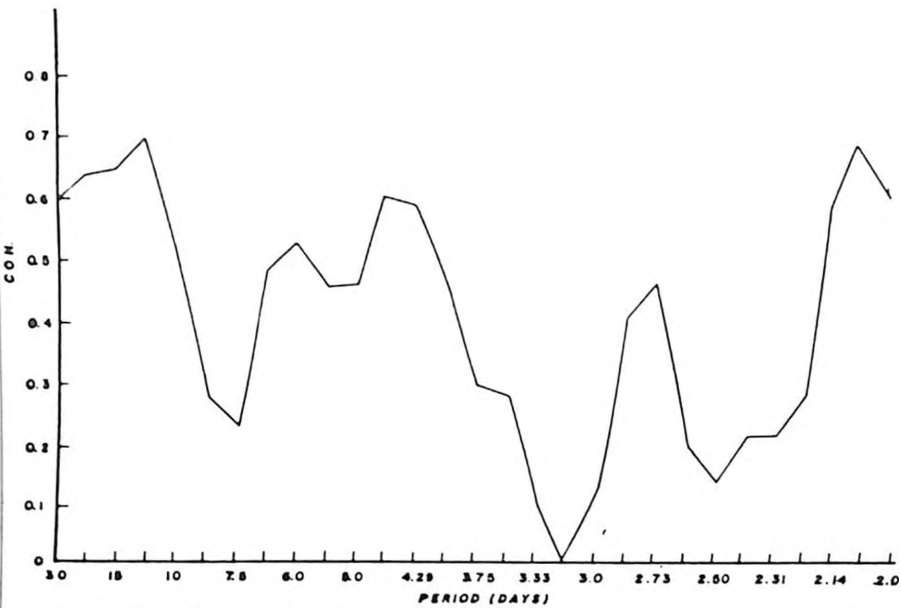


Fig. 89 COHERENCE FUNCTION FOR SEASON ONE IN YEAR SIX 1973

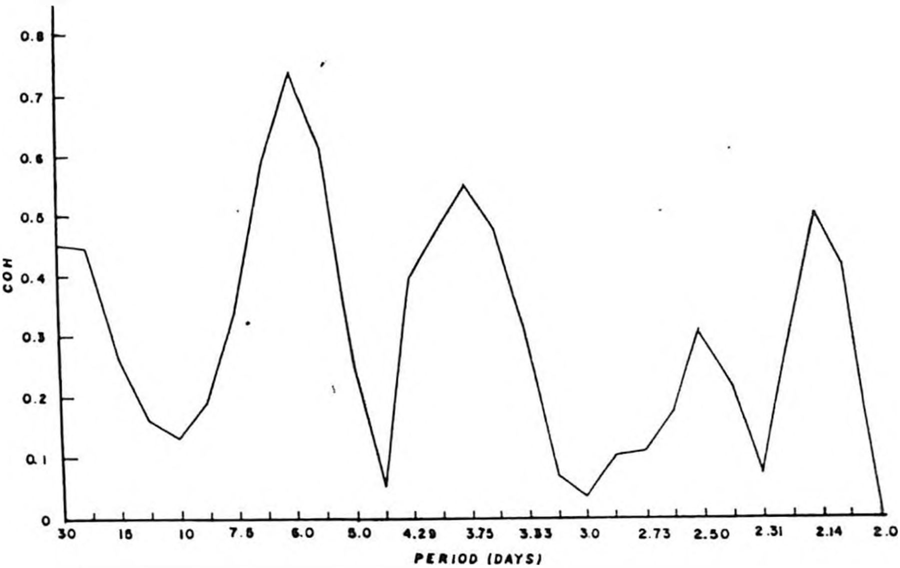


Fig. 88 COHERENCE FUNCTION FOR SEASON ONE IN YEAR FOUR 1971

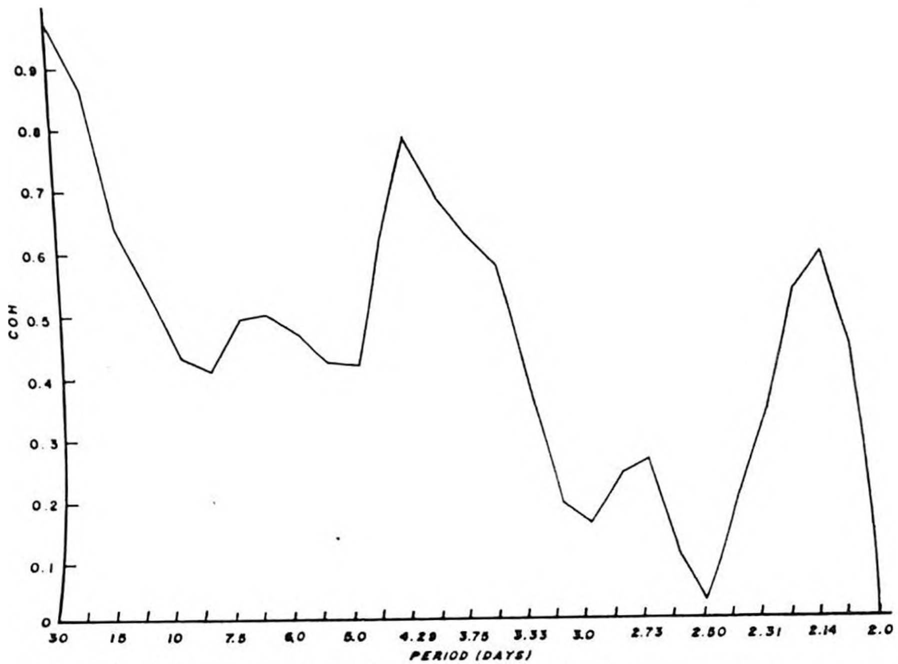


Fig. 90 COHERENCE FUNCTION FOR SEASON ONE IN YEAR SEVEN 1974

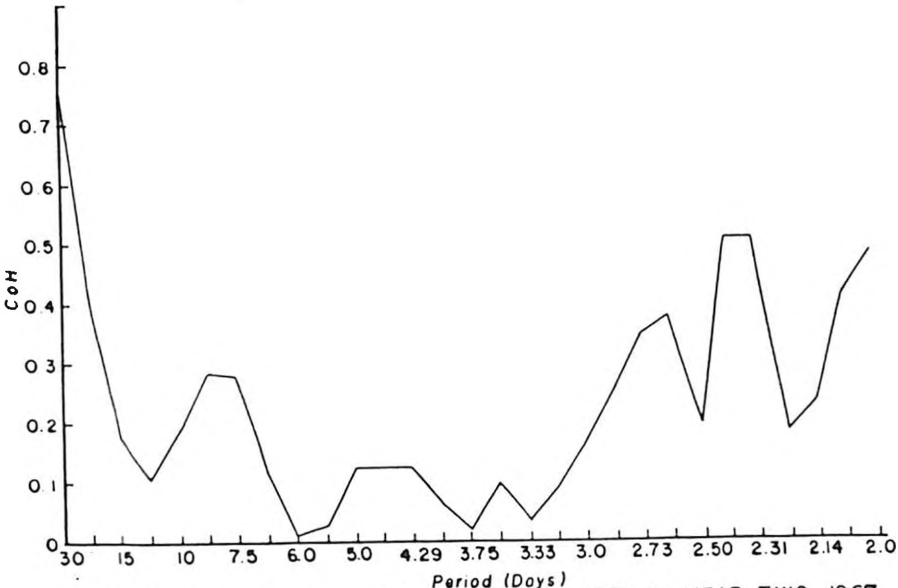


Fig. 91 COHERENCE FUNCTION FOR SEASON TWO IN YEAR TWO 1967

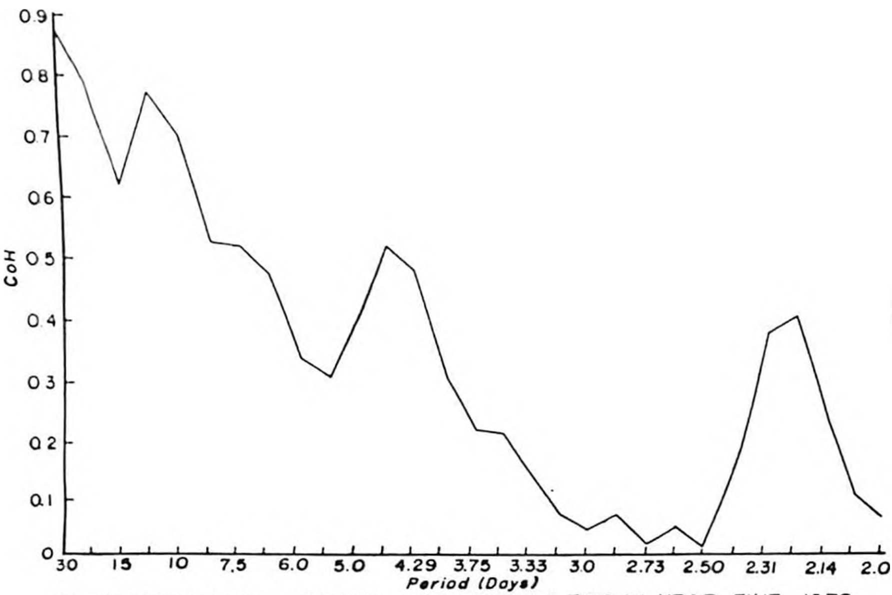


Fig. 93 COHERENCE FUNCTION FOR SEASON TWO IN YEAR FIVE 1972

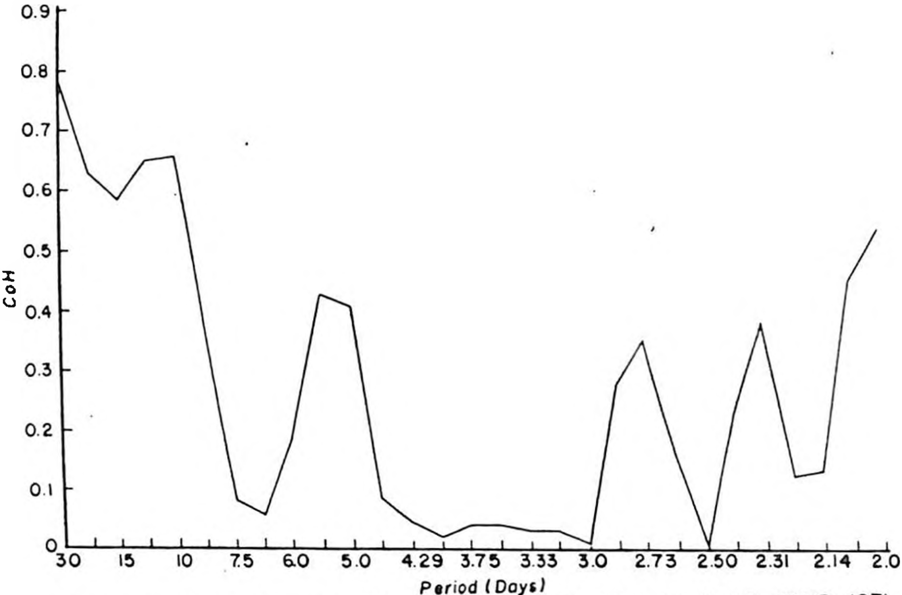


Fig. 92 COHERENCE FUNCTION FOR SEASON TWO IN YEAR FOUR 1971

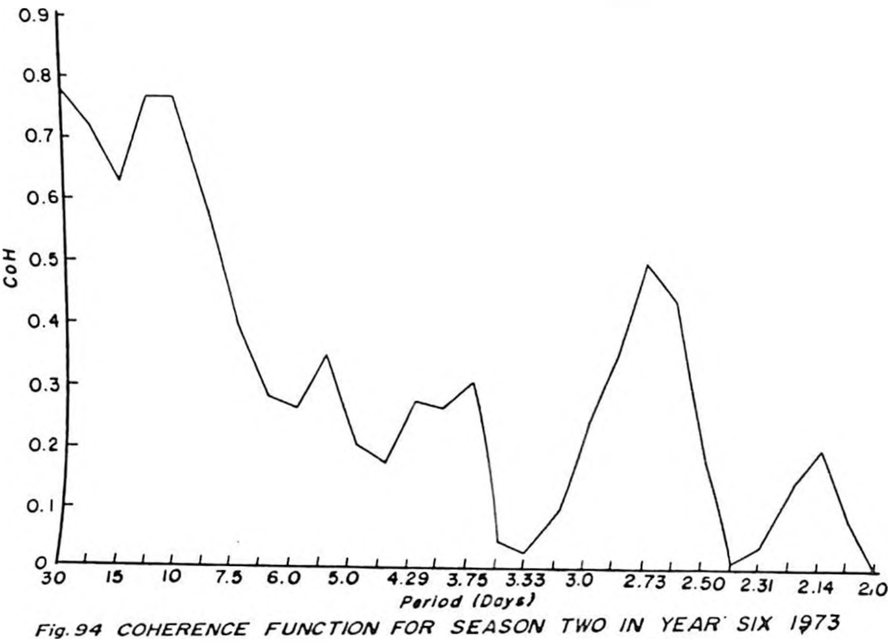


Fig. 94 COHERENCE FUNCTION FOR SEASON TWO IN YEAR SIX 1973

APPENDIX E: Extra examples of quadrature spectrum

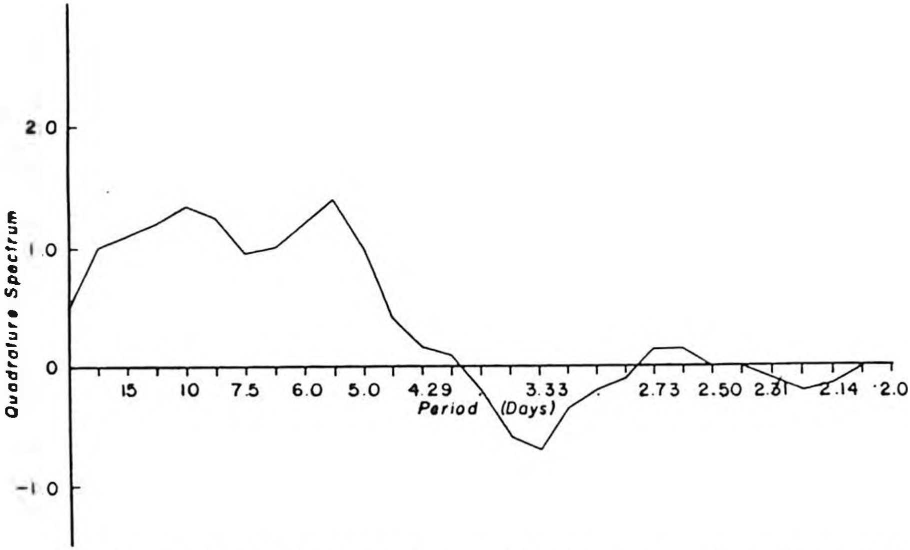


Fig. 95 QUADRATURE SPECTRUM FOR SEASON ONE IN YEAR ONE 1968

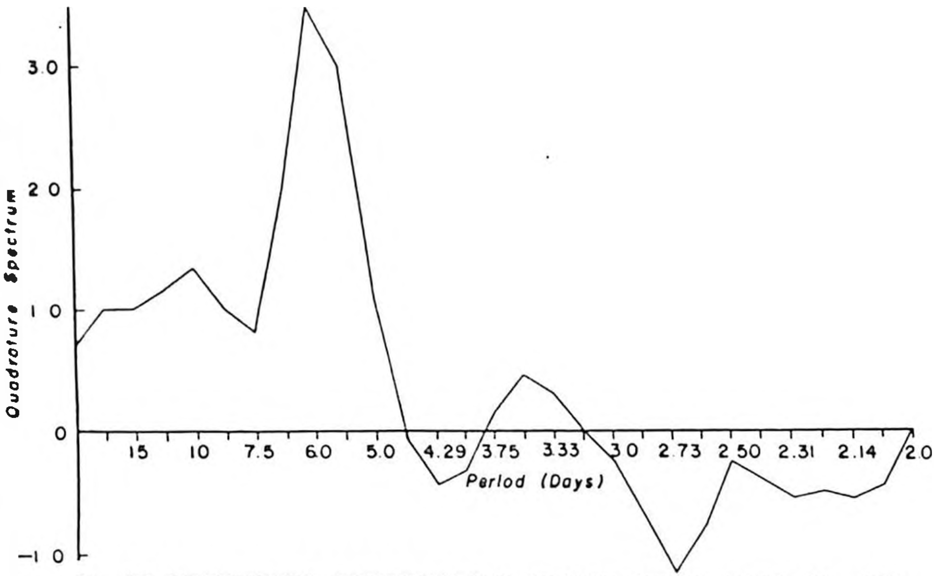


Fig. 97 QUADRATURE SPECTRUM FOR SEASON ONE IN YEAR SIX 1973

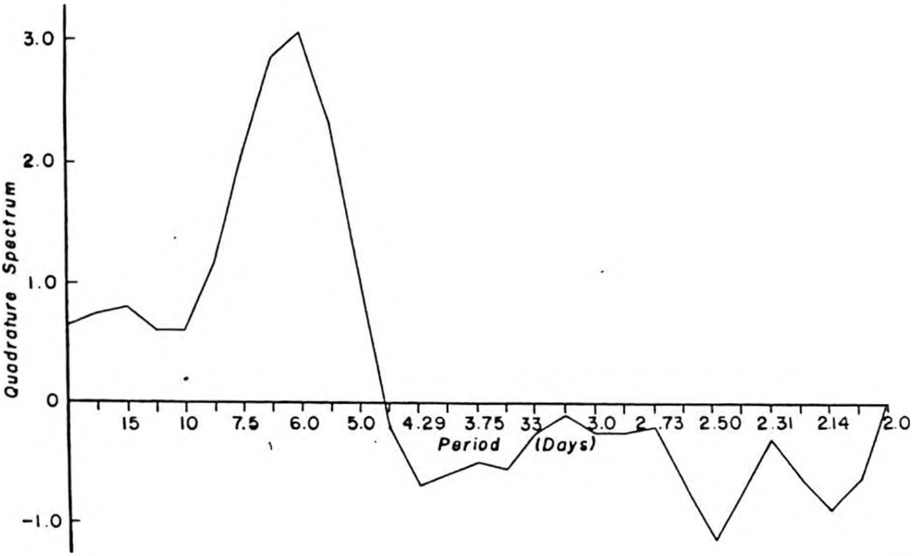


Fig 96 QUADRATURE SPECTRUM FOR SEASON ONE IN YEAR FOUR 1971

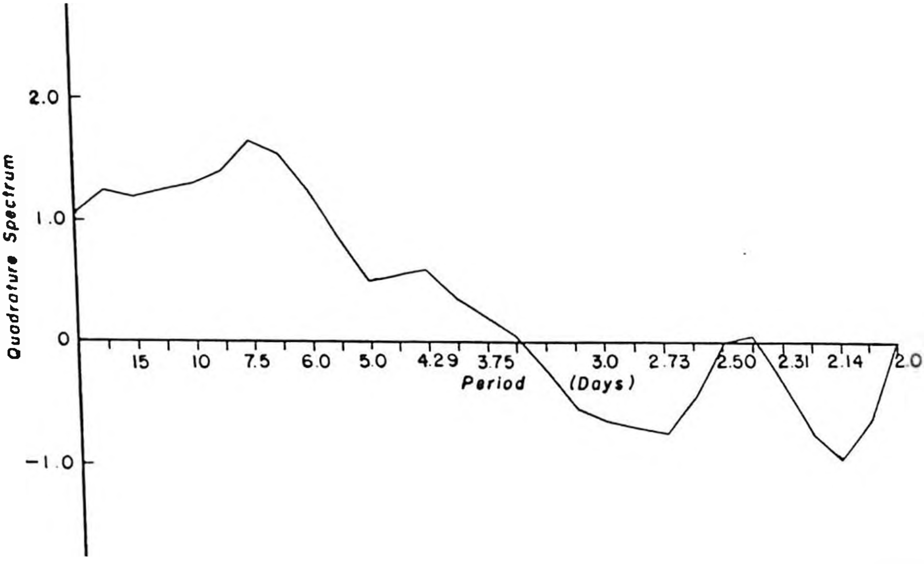


Fig 98 QUADRATURE SPECTRUM FOR SEASON ONE IN YEAR SEVEN 1974

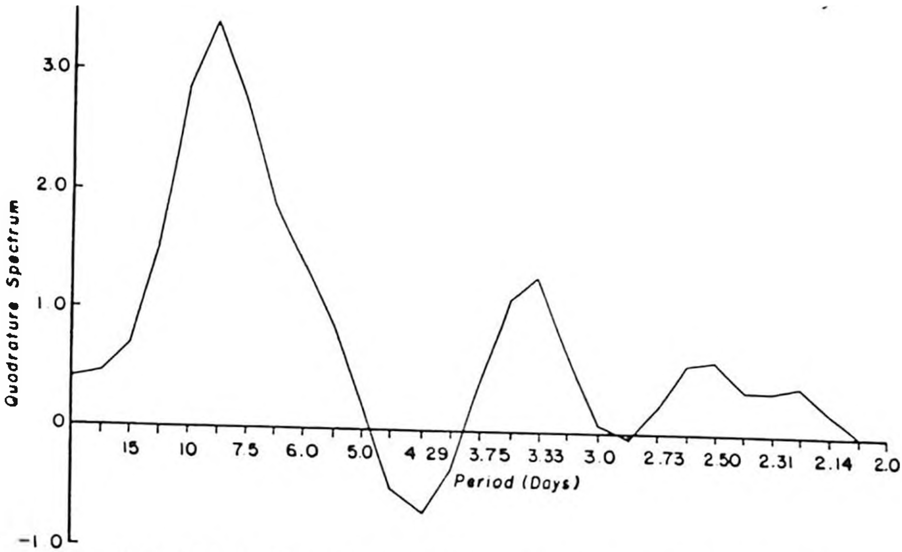


Fig. 99 QUADRATURE SPECTRUM FOR SEASON TWO IN YEAR ONE 1968

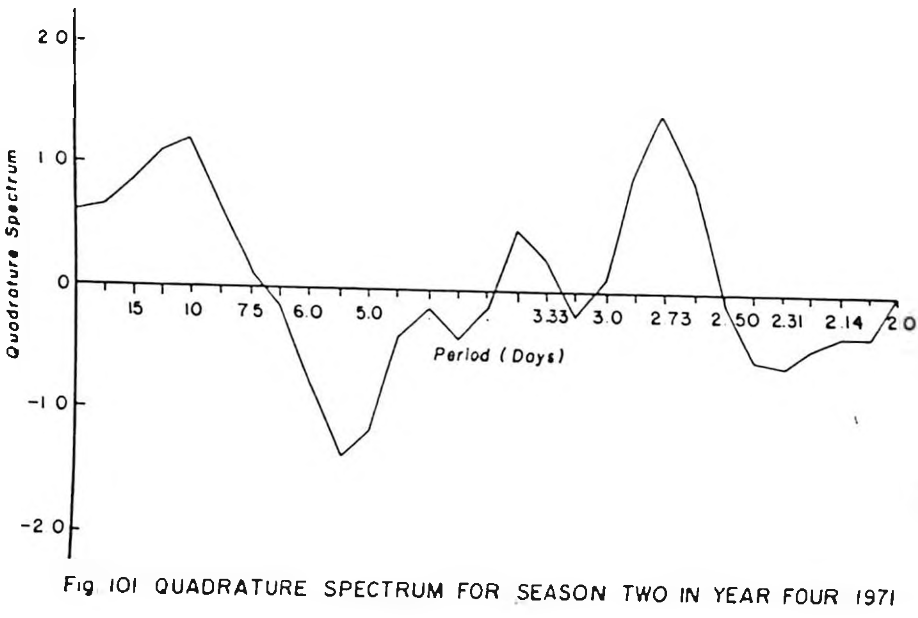


Fig 101 QUADRATURE SPECTRUM FOR SEASON TWO IN YEAR FOUR 1971

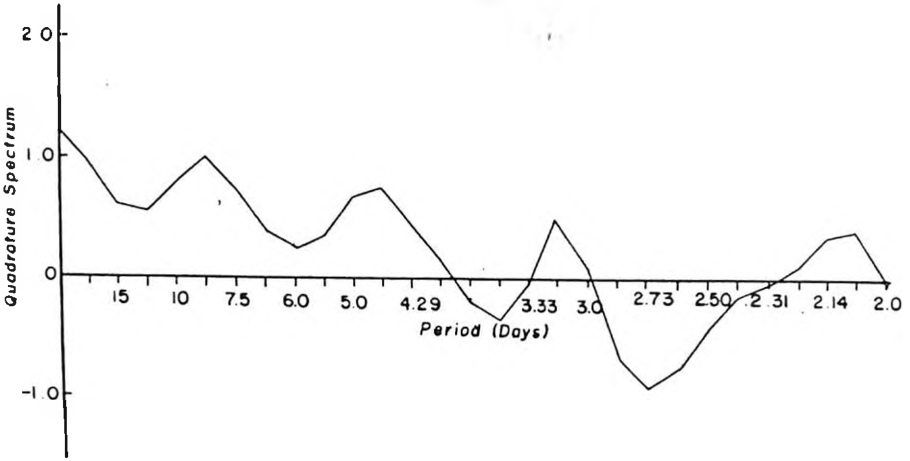


Fig. 100 QUADRATURE SPECTRUM FOR SEASON TWO IN YEAR TWO 1969

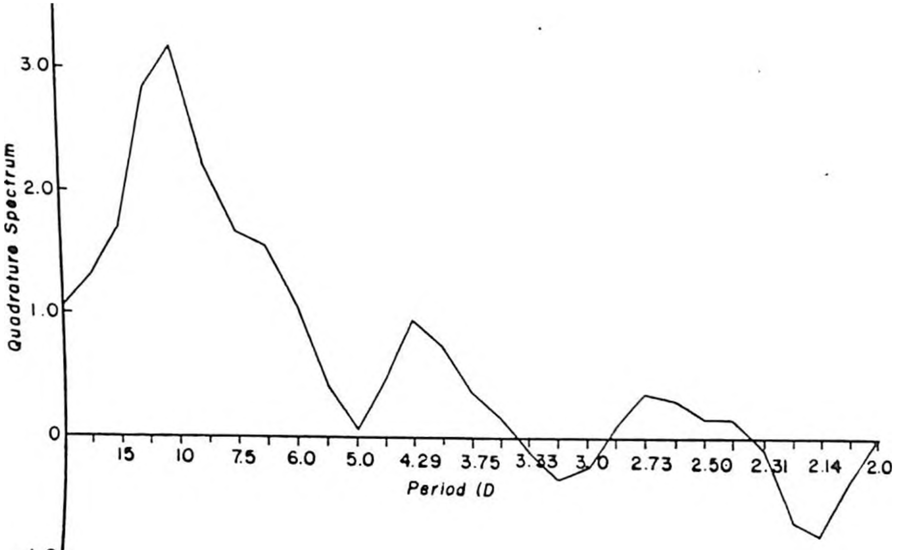


Fig. 102 QUADRATURE SPECTRUM FOR SEASON IN TWO IN YEAR SIX 1973

APPENDIX F: Extra examples of the phase function

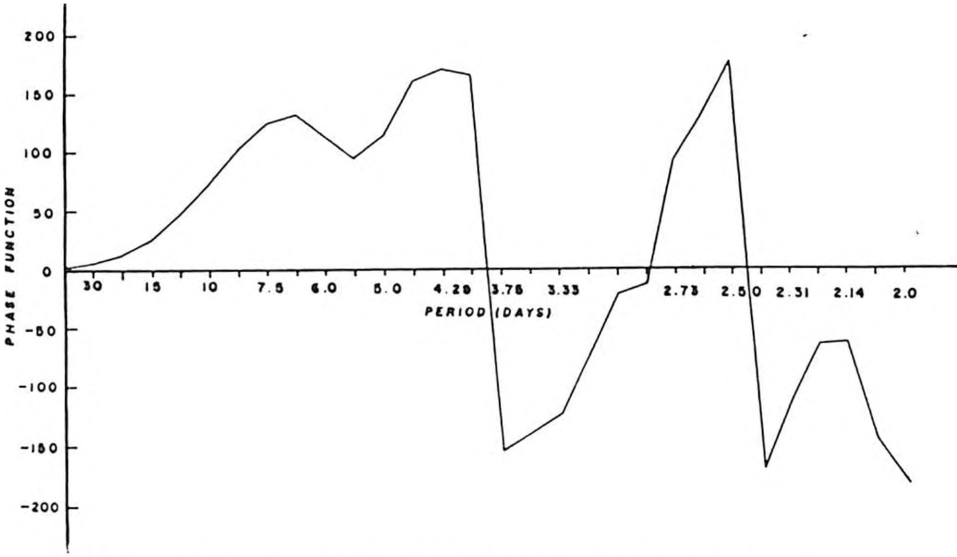


Fig 103 PHASE FUNCTION FOR SEASON ONE IN YEAR ONE 1968

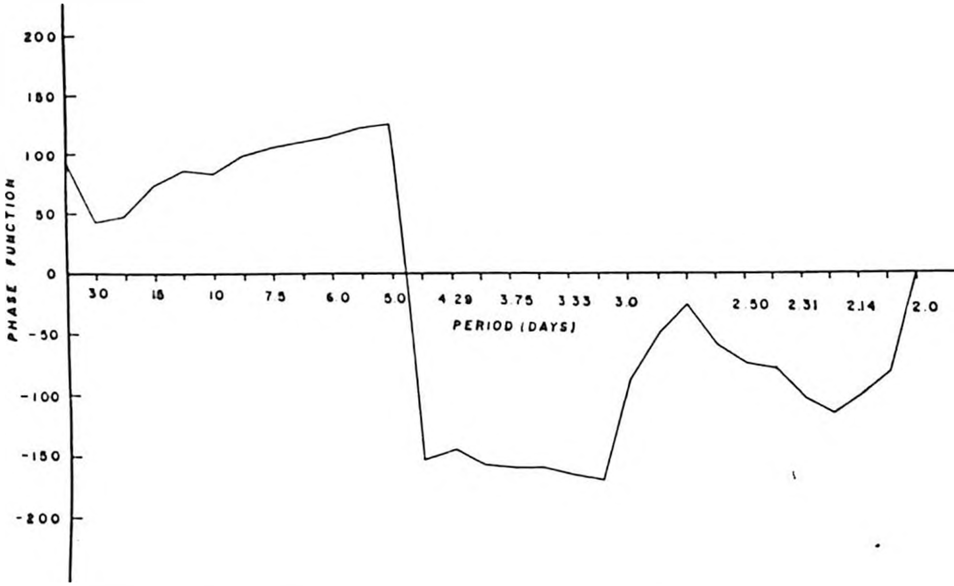


Fig. 105 PHASE FUNCTION FOR SEASON ONE IN YEAR FOUR 1971

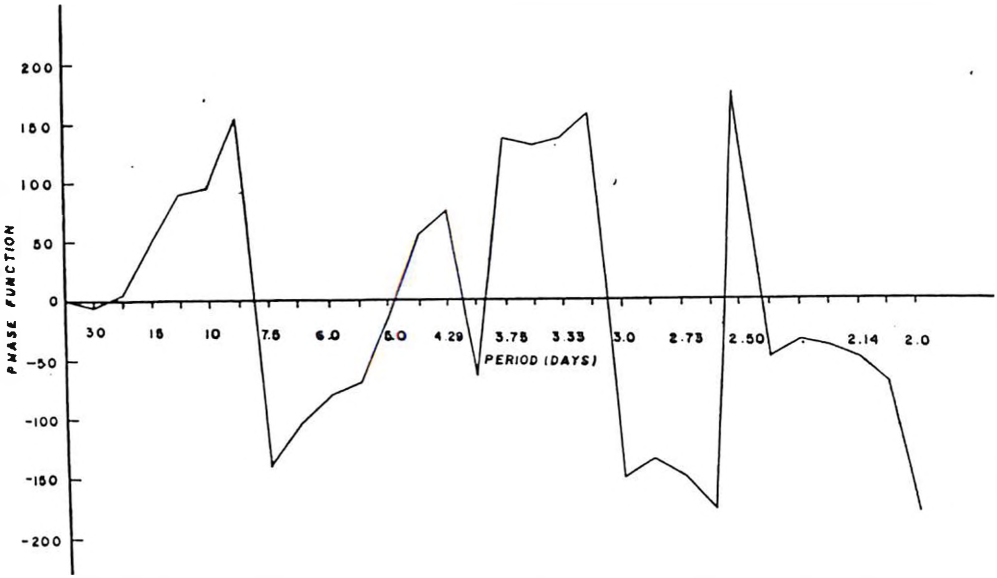


Fig. 104 PHASE FUNCTION FOR SEASON ONE IN YEAR TWO 1969

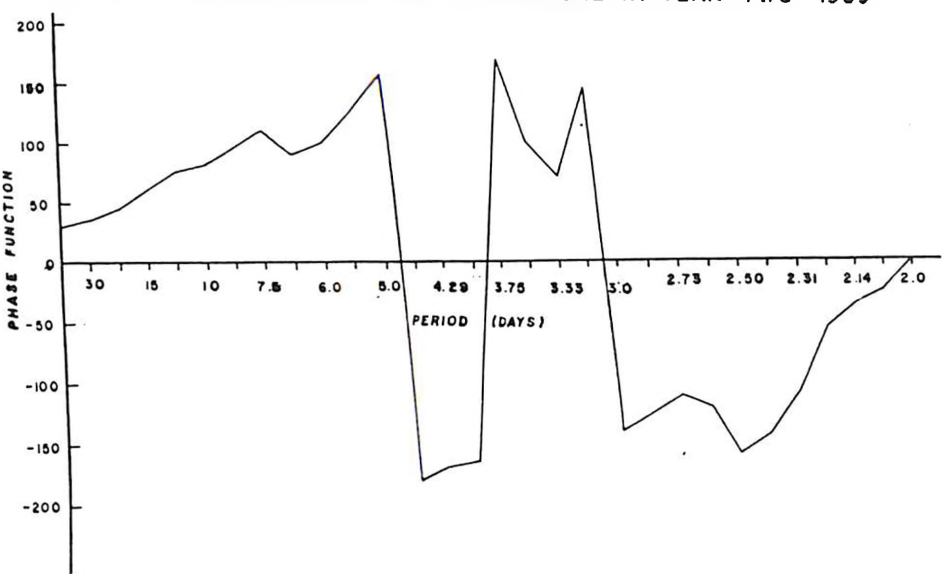


Fig. 106 PHASE FUNCTION FOR SEASON ONE IN YEAR SIX 1973

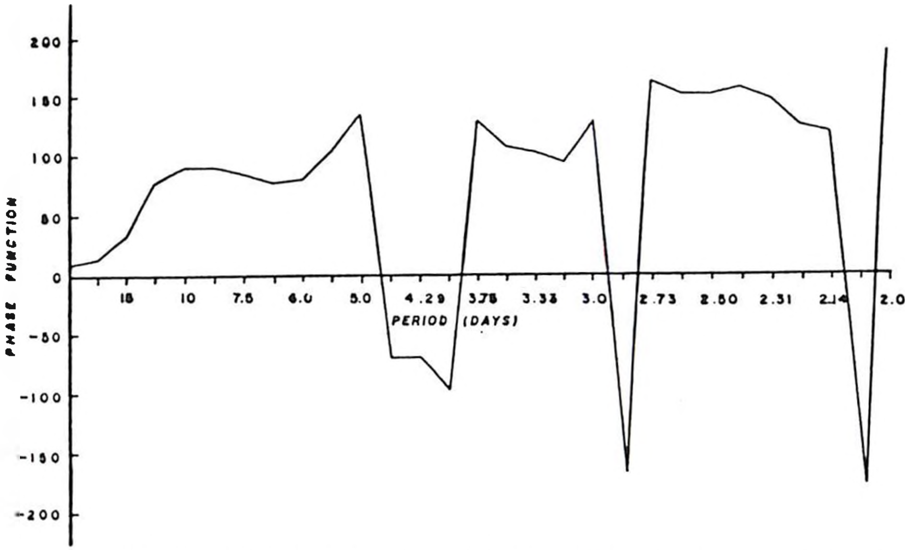


Fig. 107 PHASE FUNCTION FOR SEASON TWO IN YEAR ONE 1968

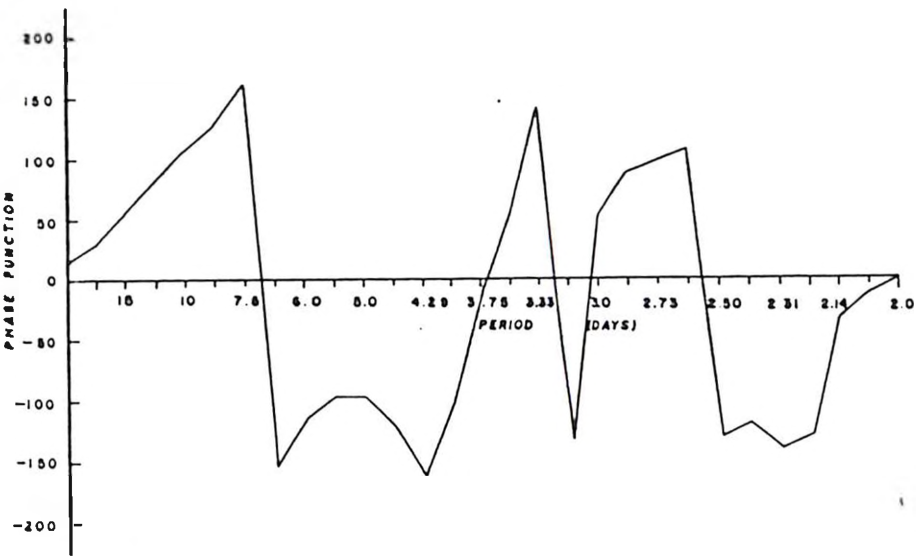


Fig. 109 PHASE FUNCTION FOR SEASON TWO IN YEAR FOUR 1971

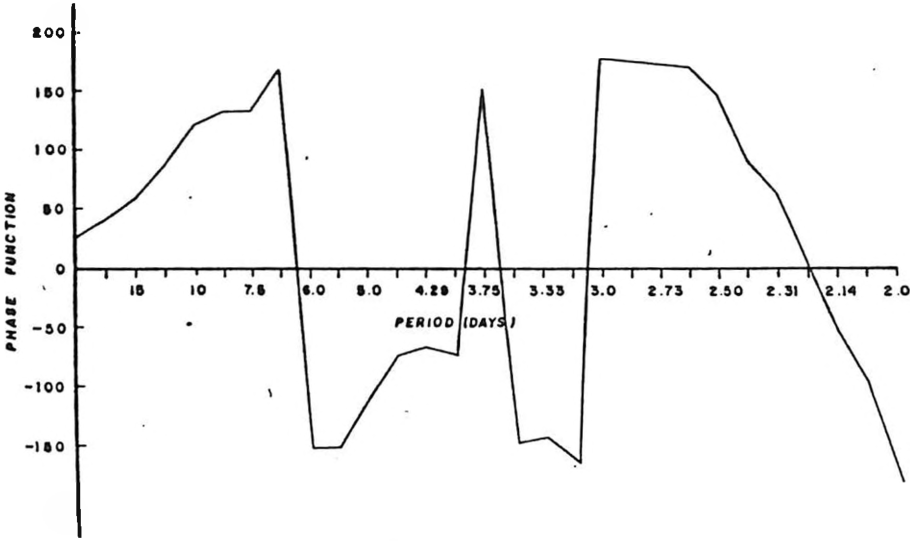


Fig 108 PHASE FUNCTION FOR SEASON TWO IN YEAR THREE 1970

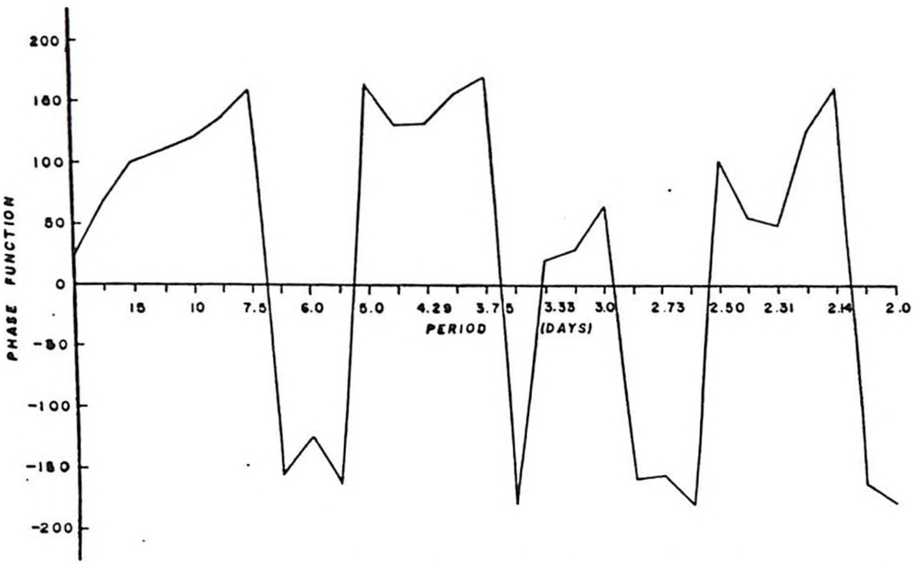


Fig. 110 PHASE FUNCTION FOR SEASON TWO IN YEAR SEVEN 1974