STOCHASTIC MODELLING FOR FORECASTING MAHI RIVER INFLOWS

Ekkgh unh & vUrokg dk iokuekuu gr ÁlÒkO; Áfr:i.k

BY

NEERAJ CHHAJED

Thesis

Master of Engineering

In

Agricultural Engineering

(Soil and Water Conservation Engineering)

2004

Department of Soil and Water Engineering

College of Technology and Engineering

Maharana Pratap University of Agriculture and Technology, Udaipur-313001

MAHARANA PRATAP UNIVERSITY OF AGRICULTURE AND TECHNOLOGY, UDAIPUR

COLLEGE OF TECHNOLOGY AND ENGINEERING, UDAIPUR

Dated: 03 – 09 – 2004

CERTIFICATE - I

This is to certify that **Mr. Neeraj Chhajed** has successfully completed the comprehensive examination held on 29th June, 2004 as required under the regulations for Master of Engineering in Soil and Water Conservation Engineering.

(Dr. RAJ VIR SINGH)

HEAD

DEPARTMENT OF SWE

COLLEGE OF TECHNOLOGY AND ENGINEERING, UDAIPUR

MAHARANA PRATAP UNIVERSITY OF AGRICULTURE AND TECHNOLOGY, UDAIPUR

COLLEGE OF TECHNOLOGY AND ENGINEERING, UDAIPUR

Dated: 03 – 09 – 2004

CERTIFICATE - II

This is to certify that this thesis entitled **“Stochastic Modelling for Forecasting Mahi River Inflows”** submitted for the degree of **Master of Engineering** in agriculture in the subject of **Soil and Water Conservation Engineering**, embodies bonafide research work carried out by **Mr. Neeraj Chhajed** under my guidance and supervision and that no part of this thesis has been submitted for any other degree. The assistance and help received during the course of investigation have been fully acknowledged. The draft of the thesis was also approved by the advisory committee on 13th July, 2004.

(Dr. RAJ VIR SINGH)
HEAD
SWE DEPARTMENT

(Dr. ARUN SHARMA)
MAJOR ADVISOR

(Dr. A. N. MATHUR)

DEAN

CTAE, UDAIPUR

MAHARANA PRATAP UNIVERSITY OF AGRICULTURE AND TECHNOLOGY, UDAIPUR
COLLEGE OF TECHNOLOGY & ENGINEERING, UDAIPUR

DATED: 03 – 09 – 2004

CERTIFICATE – III

THIS IS TO CERTIFY THAT THIS THESIS ENTITLED **"STOCHASTIC MODELLING FOR FORECASTING MAHI RIVER INFLOWS "** SUBMITTED BY **MR. NEERAJ CHHAJED** TO MAHARANA PRATAP UNIVERSITY OF AGRICULTURE & TECHNOLOGY, UDAIPUR, IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF **MASTER OF ENGINEERING** IN AGRICULTURAL ENGINEERING IN THE SUBJECT OF **SOIL AND WATER CONSERVATION ENGINEERING**, WAS AFTER RECOMMENDATION BY THE EXTERNAL EXAMINER AND DEFENDED BY THE CANDIDATE BEFORE THE FOLLOWING MEMBERS OF THE EXAMINATION COMMITTEE. THE PERFORMANCE OF THE CANDIDATE IN THE ORAL EXAMINATION ON HIS THESIS HAS BEEN FOUND SATISFACTORY; WE THEREFORE, RECOMMEND THAT THE THESIS BE APPROVED.

(Dr. ARUN SHARMA)
MAJOR ADVISOR

(Er. P. K. SINGH)
ADVISOR

(Prof. R. K. AERON)
ADVISOR

(Dr. S. M. MATHUR)
DRI NOMINEE

(Dr. RAJ VIR SINGH)
HEAD
DEPTT. OF SWE

(Dr. A.N. MATHUR)
DEAN
CTAE, UDAIPUR

APPROVED

(DR. L.L. SOMANI)
DIRECTOR RESIDENT INSTRUCTION
MPUAT, UDAIPUR

MAHARANA PRATAP UNIVERSITY OF AGRICULTURE AND TECHNOLOGY, UDAIPUR

COLLEGE OF TECHNOLOGY AND ENGINEERING, UDAIPUR

DATED: 03 – 09 – 2002

CERTIFICATE - IV

THIS IS TO CERTIFY THAT **MR. NEERAJ CHHAJED** STUDENT OF **MASTER OF ENGINEERING** IN AGRICULTURE IN THE SUBJECT OF **SOIL AND WATER CONSERVATION ENGINEERING**, DEPARTMENT OF SOIL AND WATER ENGINEERING HAS MADE ALL CORRECTIONS/MODIFICATIONS IN THE THESIS ENTITLED **"STOCHASTIC MODELLING FOR FORECASTING MAHI RIVER INFLOWS"** WHICH WERE SUGGESTED BY THE EXTERNAL EXAMINER AND THE ADVISORY COMMITTEE IN THE ORAL EXAMINATION HELD ON 02 – 09 – 2004. THE FINAL COPIES OF THE THESIS DULY BOUND AND CORRECTED WERE SUBMITTED ON 03 – 09 – 2004.

(Dr. ARUN SHARMA)
MAJOR ADVISOR

(Dr. RAJ VIR SINGH)

HEAD

DEPARTMENT OF SOIL AND WATER ENGINEERING,
CTAE, UDAIPUR

(Dr. A. N. MATHUR)

DEAN

CTAE, UDAIPUR

ABSTRACT

A mathematical model of monthly river inflow is a useful tool for generation of synthetic data and for a single month or multiple months ahead forecasting. The synthetic data sequences are needed in the simulation studied for the design and operation of water storage, conveyance and control structures. Monthly forecasts are needed for the operation of reservoir, agricultural planning and watershed management practices. In view of this present study was undertaken to develop and validate appropriate stochastic model for Mahi river inflows. Monthly inflow data of Mahi river for a period of 76 years (1928-2003) were collected. Turning point test, Kenall's correlation test and regression test performed on annual data confirmed that the inflow series is random and trend free. Fourier analysis was performed to obtain the stochastic series by removing the periodic component from the monthly inflow series. Stochastic series was standardized and normalized to make the inflow series normally distributed. ACF, PACF, IACF and IPACF were analysed to identify the class and order of stochastic models to represent the Mahi river inflows. By the use of identification stage, twelve ARMA models in which AR term ranges from 1 to 6 and MA term varies to 1 and 2, were identified for investigation.

The first 74 years data were used for estimation of the identified models parameters. The parameters of only six models, namely, ARMA(1,1), ARMA(2,1), ARMA(3,1), ARMA(1,2), ARMA(2,2), ARMA(4,2) were found to be significant. The other six models have large standard error than value of their parameters. All the six models pass the validation tests.

Regeneration of series was made by all the six models for the period of 1928 to 2001 and basic statistical characteristics were compared with those of the actual inflow series. It was found that the statistical characteristics of the actual series were distorted in the generated series. Minimum mean square error (MMSE) criterion was used for selection of best model. It revealed that ARMA (3,1) model was proved to be most appropriate model. One-time step ahead monthly forecast were made by the selected model ARMA (3,1) for the period of 2002-2003. Values of correlation coefficient between observed and forecasted monthly inflow series for the two year period (2002-2003) was found to be 0.912 showing adequacy of developed model.

vu{ki.k

unh vUrokg dk ekfld xf.krh; ifr:i vkdMk d df=e mRiknu rFkk ,d ekg ;k vf/kd ekg vkx d iokueku d fy, mi;kxh lk/ku gA df=e vkdMk dh J.kh ty lxg.k] iog.k rFkk fu;=.k ljpuk dh vk;ktuk rFkk xfrfof/k;k d fy; vko';d gA ekfld iokueku dh vko';drk tyk'k; icU/ku] df"k vk;ktuk rFkk ty xg.k {k= ic/ku xfrfof/k;k d fy; vko';d gA vr% iLrr v/;;u ekgh unh&vUrokg d mi;Dr ilHkkO; ifr:i d fodkl ,o ekU;dj.k d mnn'; d lnHk e fd;k x;kA ekgh unh d ekfld vUrokg vkdM 76 o"kh; dky ¼1928&2003½ d fy, ,d= fd; x;A okf"kd vUrokg J.kh ij oru&fcUn leUoh{kk] dUMky lglcU/k leUoh{kk ,o lekJ;.k leUoh{kk d fu"iknu] vUrokg J.kh d lelHkkfod ,o idfr jfgr gku dh if"V djr gA Qkfj;j fo'y"k.k ekfld vUrokg J.kh l vkofrr ?kVd gVkdj ilHkkO; J.kh ikIr dju d fy; fd;k x;kA ilHkkO; J.kh d iekfidj.k rFkk ileku;r% forj.k }kjk vUrokg J.kh dk lfpr forj.k fd;k x;kA ekgh unh vUrokg dk fu:fir dju oky ilHkkO; ifr:ik dk ox ,o Øe fu/kkfjr dju d fy; Lolglc/k Qyu] vkf'kd Lolglc/k Qyu] ifryke Lolglc/k Qyu] ifryke vkf'kd Lolglc/k Qyu dk fo'y"k.k fd;k x;kA fu/kkj.k iØe d mi;kx l ckjg LolekJ;h xfreku ek/; ox d ifr:i vUo"k.k d fy; fu/kkfjr fd; x; ftue LolekJ;h in 1 l 6 rFkk xfreku ek/; in 1 l 2 rd xfr djrk gA

fu/kkfjr ifr:ik d ikpyk dk vkdyu dju d fy, iFke 74 o"kh; vkdMk dk i;kx fd;k x;kA ;g ik;k x;k fd doy N% ifr:ik vFkkr LolekJ;h xfreku ek/; ¼1]1½] LolekJ;h xfreku ek/; ¼2]1½] LolekJ;h xfreku ek/; ¼3]1½] LolekJ;h xfreku ek/; ¼1]2½] LolekJ;h xfreku ek/; ¼2]2½ rFkk LolekJ;h xfreku ek/; ¼4]2½ d ikpy gh lkFkd gA 'k"k N% ifr:ik dh ekud =fV;k mud ikpyk d eku l vf/kd gA lHkh N% ifr:i ekU;dj.k ifj{k.kk e [kj mrjA

lHkh N% ifr:ik d }kjk J.kh dk iu% mRiknu 1928 l 2001 d dky d fy; fd;k x;k ,o vk/kkjh lkf[;dh; vfHky{k.kk dh ryuk okf"kd J.kh d vUrokg J.kh d x.kkdk l dh xbA ;g ik;k x;k fd okfLrfod J.kh d lkf[;dh; x.kkd iu% mRikfnr J.kh e fodr gk x; gA loJ"B ifr:i d puko gr y?kre ek/; ox =fV fl)kUr dk mi;kx fd;k x;kA mld

}kjk LolekJ;h xfreku ek/; ¼3]1½ ifr:i lokf/kd mi;Dr ifr:i fl) gvkA pu g, ifr:i LolekJ;h xfreku ek/; ¼3]1½ d }kjk ,d dky pj.k vkx d okf"kd iokueku 2002&2003 d dky d fy; yxk; x;A nk o"k ¼2002&2003½ dh okLrfod rFkk iokuekfur J.kh d lglc/k x.kkd dk eku 0-912 ik;k x;k tk fodflr ifr:i dh mi;Drk crkrk gA

1. INTRODUCTION

Water is the key input for crop production and the future development of agriculture depends upon its efficient utilization and management. Planning of water resources projects based on scientific data in respect of present water supply and forecasting its future behavior is one of the important factors of water management. A mathematical model of monthly river inflow is a useful tool for generation of synthetic data and for a single or multiple months ahead forecasting. The synthetic data sequences are needed in the simulation studies for the design and operation of water storage, conveyance and control structures. Monthly forecasts are needed for the operation of reservoirs, agricultural planning and watershed management practices. Before being able to generate synthetic sequences and forecast future values, models have to be obtained which adequately describe the past data. These models should ideally preserve all statistical characteristics of the observed data. River inflow data of sufficiently long duration are essential for decisions in project design, planning and operation. In the absence of a long term data, sequences of river inflow are required to be generated for long duration using the statistical characteristics of the available short term data.

The river inflow phenomenon is periodic- stochastic in nature. The periodicities are attributed to the astronomical cycles and consequently to the periodicity in the energy supply from the sun over various areas of the earth's surface and the further interactions and responses of various earths' environments. The stochastic component appears due to various random processes in the air over oceans and at the continental surface and in the relevant geophysical environments. A mathematical model representing a stochastic process is called "stochastic model". It has a certain mathematical form or structure and a set of parameters. Differences always exist between the true and estimated models and between the true and estimated model parameters. These differences represent modelling uncertainties. One way of decreasing such uncertainties is by selecting the model which best represents the physical reality of the system.

The modelling of river inflow processes has followed two approaches; the deterministic or physical simulation of the hydrologic system and the statistical or stochastic simulation of the system. In the deterministic approach, the hydrologic system is described and represented by theoretical and/or empirical physical relationships. In the stochastic approach, however a type of model is assumed, aimed to represent the most relevant statistical characteristics of the historic series. Within this approach, the most widely used models have been the autoregressive models. Subsequently, other deterministic and stochastic models have been reported in the literature. The forecasts are either based on physical understanding of the process or on statistical analysis of the process popularly termed as stochastic approach. Recently a combination of physical and stochastic approach is gaining more popularity in the field of hydrologic forecasting. In the present study the stochastic approach is adopted. Major emphasis is laid on the time series approach.

A stochastic model is the mathematical abstraction of an empirical process and is governed by time dependent probabilistic laws. The word stochastic (which was apparently suggested 300 years ago by Jacob Bernoulli of Switzerland) means, according to its Greek origin, to contemplate or to conjecture. Roughly speaking it can be regarded as synonymous with chance, random or probabilistic but more precisely the interdependence of random variables should be account for.

River inflow simulation based on historic inflows, or the generation of synthetic hydrologic data by means of stochastic models has obtained the name of operational hydrology. The purpose of this generation is to use the derived records as inputs to subsequent hydrologic analysis, and to apply simulation in water resource systems design and management.

In most of the river basins, except in under developed region, 20 to 50 years of precipitation and runoff records are available. For annual time series such periods may not be sufficient to determine the structure of time series and the probability distribution parameters with accuracy. In comparison with annual data, monthly series are twelve times longer and thus have about 250 to 600 items which can be considered as sufficient for the determination of both deterministic and stochastic components of the series. At the same time, the use of monthly values does not yield as much detail in the analysis of the two types of components as would the daily values, therefore, it does not involve so much computation (Jochanan, 1971).

Hydrologic modelling is a procedure in which one or more of the phases of the hydrologic cycle are represented by a simplified system (Mc Cuen, 1973). River inflow at a particular point in a river represents a time dependent non linear response to rainfall. Designs that are sensitive to rapid variation in river inflow, e.g. flood protection schemes, can be based on an analysis of rainfall and the response pattern of the catchments, with base inflow assessed separately. Alternatively, projects such as storage reservoirs which are less sensitive to short term variation in river inflow can be designed using monthly average inflows. However, many problems of design and control fall between these two conditions. The design of a regulating reservoir, for instance, has to be based on control rules for release of water, which have to take into account the varying travel time in the river. A daily record of the river inflows, which can be used for simulating the behavior of the reservoir using alternative operating policies, is therefore essential.

The problem of short term river inflow forecasting is of considerable practical importance. Dam construction, reservoir operation, flood control and waste water disposal are a few situations where future knowledge of river inflows will improve management of water resources.

Stochastic linear models are fitted to hydrological data for two main reasons; to enable forecasts of the data one or more time periods ahead and to enable the generation of sequences of synthetic data. These techniques are of considerable importance to the design and operation of water resource systems. Short sequences of data lead to uncertainties in the estimation of model parameters and to doubts about the appropriateness of particular time series models. A premium is placed on models that are economical in terms of the number of parameters required. One such family of models is multiplicative seasonal autoregressive integrated moving average (ARIMA) models that have been described by Box and Jenkins (1976). The ARIMA process is a powerful time series modeling and forecasting technique which possesses flexibility for the inclusion of many time series characteristics. The ARIMA methodology has gained enormous popularity in many areas and research practice confirms its power and flexibility. However, because of its power and flexibility, ARIMA is a complex technique; it is not easy to use, it requires a great deal of experience, and although it produces satisfactory results, those results depend on the researcher's level of expertise.

Autoregressive (AR) and autoregressive integrated moving average (ARIMA) models have an important place in the stochastic modelling of hydrologic data. Autoregressive type models have been used in many of such studies especially to represent river inflow sequences. More recent studies have concentrated on the long term (low frequency) features of inflow sequences.

The recent advances in the fields of operations research and computer technology have had an enormous impact on synthetic hydrology. More and more hydrologists and engineers are using synthetic sequences in the design, operation and management of water resource systems. Even though the determination of an optimum solution by linear or dynamic programming for a given river inflow sequence is deterministic, the stochastic nature of the river inflow enters the system through synthetic river inflow sequences.

Stochastic processes deal with continuous or discrete state and time parameters. The analysis of time series is done to understand the mechanism that generate the data and to produce likely future sequence if required. These are attempted by making inferences regarding the underlying laws of the stochastic process from one or more sequences of recorded observations and then to postulate a model that fits the data, which again used for estimation purposes. At first it is necessary to identify and analyze the different components of time series and then generate the future sequence (Kottegoda, 1980).

In the arid and semi-arid state of Rajasthan the rivers play an important role in agricultural planning. A typical characteristic of rivers of Rajasthan is that they are monsoonal, having sufficient water flowing during monsoon season, and then slowly become dry. The wells are dug on the banks of these rivers which help in irrigation. The river inflow simulation for southern Rajasthan has not been made so far. Therefore an attempt of modelling the Mahi river inflow is made in the present study so that management of water resources of this river may be done in more systematic way.

Mahi Irrigation Project, Banswara is a newly constructed project. Such prediction model may be useful in generating the sequences of river inflow for longer durations using the characteristics of the available short term data for operation of water Storage structure, reservoir, and agricultural planning and watershed management practices. In view of above description the present study is undertaken with the following objectives:

I. Identification of different stochastic models best representing the Mahi river inflows.
II. Estimation of parameters of different stochastic models and their validation.
III. Forecasting the Mahi river inflows on monthly basis by different stochastic models.
IV. Comparison of different stochastic models and selection of the most appropriate one.

2. REVIEW OF LITERATURE

2.1 General

The objective of present study is to formulate stochastic models for forecasting river inflows. A mathematical model representing a stochastic process is called "stochastic model". Yevjevich (1972) described the stochastic process as the mathematical abstraction of an empirical process, with its development governed by the probability laws. Nearly all the time oriented processes can be characterised as stochastic processes or as a combination of deterministic and stochastic processes. A time series represents a set of observations that measure the variation in time of some dimension of a phenomenon such as precipitation, wind speed, river inflow etc. The time series can be considered in continuous or discrete form. Most practical applications in hydrology consider the discrete form primarily as it is easier to handle the discrete time series on a digital computer. Time series can be classified as stationary time series or nonstationary time series. If the expected value of statistical parameters does not change with time, the time series is said to be stationary, otherwise nonstationary.

A simple time series model could be represented by a single probability distribution function f(x: θ) with parameters $\theta = (\theta_1, \theta_2\ldots)$ valid for all positions t=1, 2… and without any dependence between x_1, x_2… For instance, if x is normal with mean μ and variance σ^2, the time series model can be conveniently written as;

$$x_t = \mu + \sigma\,\varepsilon_t \qquad \text{... (2.1)}$$

where ε_t is also normal with mean zero and variance one and ε_1, ε_2, ….. are independent. In equation (2.1) the model has the parameters μ and σ and since they are constants (do not vary with time) the model is stationary. The structure of the model is simple since the variable x_t is a function only of the independent variable ε_t and so x_t is also independent.

A time series model with dependence structure can be formed as;

$$\varepsilon_t = \phi\,\varepsilon_{t-1} + \xi_t \qquad \text{... (2.2)}$$

Where ξ_t is an independent series with mean zero and variance $(1-\phi^2)$, ε_t is the dependent series, and ϕ is the parameter of the model. In Eq. (2.2) ε_t is a dependent series because in addition to being a function of ξ_t, it is a function of the same variable ε at time t-1. If ε_t in Eq. (2.1) would be represented by the dependent model of Eq. (2.2) then x_t would also become dependent model. In this case the parameters of the model x_t would be μ, σ and ϕ. Since the parameters of the above models are constants, the models are stationary representing stationary time series or stationary stochastic processes. Non-stationary models would result if such parameters would vary with time.

In view of river inflow being a stochastic process, and the study being on stochastic modelling of river inflow, a brief review of various aspects of stochastic modelling of river inflows has been presented under the following broad categories as per the notations and description used by Box and Jenkins (1976) and Hipel and McLeod (1994).

(i) Linear stationary models
(ii) Linear nonstationary models
(iii) Seasonal models
(iv) Application of various models for forecasting river inflows

2.2 Linear Stationary Models

2.2.1 The general linear process

A general linear stochastic model supposes a time series to be generated by a linear aggregation of random shocks. A stochastic process can be represented as the output from a linear filter whose input is a white noise $a_{t,}$ that is

$$Z_t = a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \ldots\ldots = a_t + \sum_{j=1}^{\infty} \psi_j a_{t-j} \qquad \text{... (2.3)}$$

where $Z_t = Z_{t-\mu}$ is the deviation of the process from some origin, or from its mean μ if the process is stationary. The general linear process of Eq. (2.3) represents Z_t as a weighted sum of present and past values of the "white noise" process a_t. The white noise process a_t is regarded as a series of shocks which drive the system and consists of a sequence of uncorrelated random variables with mean zero and constant variance, that is

$$E[a_t] = 0 \qquad Var[a_t] = \sigma^2_a$$

Since the random variables a_t are uncorrelated, their autocovariance function is

$$\gamma_k = E\,[\,a_t\,.\,a_{t+k}] = \begin{cases} \sigma_a^2 & k = 0 \\ 0 & k \neq 0 \end{cases} \qquad \text{... (2.4)}$$

Thus, the autocorrelation function of white noise has the particularly simple form

$$\rho_k = \begin{cases} 1 & k = 0 \\ 0 & k \neq 0 \end{cases} \qquad \text{... (2.5)}$$

The model represented by Eq. (2.3) implies that, under suitable conditions $\tilde{Z}_t$ is a weighted sum of past values of the $\tilde{Z}$'s plus an added shock a_t, that is

$$\begin{aligned} \tilde{Z}_t &= \pi_1\,\tilde{Z}_{t-1} + \pi_2\,\tilde{Z}_{t-2} + \ldots + a_t \\ &= \sum_{j=1}^{\infty} \pi_j \tilde{Z}_{i-j} + a_t \end{aligned} \qquad \text{... (2.6)}$$

The Eq. (2.6) may be thought of as one where the current deviation $\tilde{Z}_t$ from the level μ, is "regressed" on past deviations $\tilde{Z}_{t-1}$, $\tilde{Z}_{t-2}$ of the process.

2.2.2 Autoregressive model of order p: AR (p)

The general linear processes of Eq (2.3) and Eq. (2.6) contain an infinite number of parameters ψ_1 and π_1 and hence are not of practical use. Considering parsimony only first p of the weights are considered non zero and the model is represented as

$$\tilde{Z}_t = \phi_1 \tilde{Z}_{t-1} + \phi_2 \tilde{Z}_{t-2} + \ldots + \phi_p \tilde{Z}_{t-p} + a_t \qquad \text{... (2.7)}$$

where the symbols ϕ_1, ϕ_2, … ϕ_p represent the finite set of weight parameters. The process defined by Eq. (2.7) is called an autoregressive process of order p or AR (p) process. In particular, the autoregressive processes of first order (p=1) and of second order (p=2)

$$\tilde{Z}_t = \phi_1 \tilde{Z}_{t-1} + a_t$$

and

$$\tilde{Z}_t = \phi_1 \tilde{Z}_{t-1} + \phi_2 \tilde{Z}_{t-2} + a_t$$

are of considerable practical importance (Box and Jenkins 1976). The autoregressive model of Eq. (2.7) can be written in the equivalent form

$$(1 - \phi_1 B - \phi_2 B^2 - \ldots \phi_p B^P)\ \tilde{Z}_t = a_t$$

or $$\phi(B)\ \tilde{Z}_t = a_t \qquad \text{... (2.8)}$$

where B is a backward shift operator defined by $B\tilde{Z}_t = \tilde{Z}_{t-1}$ and $B^m \tilde{Z}_t = \tilde{Z}_{t-m}$. Eq. (2.8) implies $\tilde{Z}_t = 1/\phi(B)\ a_t = \phi^{-1}(B)\ a_t$

Hence, the autoregressive process can be thought of as the output $\tilde{Z}_t$ from a linear filter with transfer function ϕ^{-1} (B) when the input is white noise a_t. Autoregressive (AR) models have been extensively used in hydrology and water resources since the 1960's for modelling annual and periodic hydrologic time series.

The order for the AR (p) model with constant parameters to be stationary the set of parameters $\phi_1, \phi_2 \ldots \phi_p$ must satisfy the stationarity conditions. These conditions are satisfied if the roots of the characteristic equation (Yevjevich, 1972)

$$\mu^p - \phi_1 \mu^{p-1} - \phi_2 \mu^{p-2} - \ldots - \phi_p = 0 \qquad \text{... (2.9)}$$

lie inside the unit circle. That is, we must have $|\mu_j| < 1$, i=1…, p, where μ_j are the roots of the solution of Eq. (2.9).

2.2.2.1 Autocorrelation function (ACF)

The covariance between Z_t and a value Z_{t+k} which is k time lags removed from Z_t is theoretically defined in terms of the autocovariance γ_k at lag k given by

$$\gamma_k = \mathrm{Cov}\,[Z_t . Z_{t+k}] = E\,[(Z_t - \mu)(Z_{t+k} - \mu)] \qquad \text{... (2.10)}$$

when k = 0, the autocovariance is the variance and consequently $\gamma_o = \sigma_z^2$. A more convenient normalized quantity to deal with than γ_k is the theoretical autocorrelation coefficient which is defined at lag k as

$$\rho_k = \frac{\gamma_k}{\gamma_0} \qquad \text{... (2.11)}$$

The autocorrelation coefficient, ρ_k is dimensionless and, therefore, independent of the scale of measurement. The possible values of ρ_k range from -1 to 1, where ρ_k has a magnitude of unity at lag zero. The autocorrelation coefficient ρ_k is also called the theoretical autocorrelation function (ACF) or serial correlation coefficient (Jenkins and Watts, 1968). The sample ACF is useful for identifying what type of time series model to fit to a given time series. To determine which values of the estimated ACF are significantly different from zero, confidence limits are included on the graph. For this the variance of the sample ACF after lag q is given by

$$\text{Var}\,[\gamma_k] = \frac{1}{N}\left(1 + 2\sum_{j=1}^{q}\rho_j^{\,2}\right) \qquad \text{for } k > q \qquad \text{... (2.12)}$$

When a normal process is uncorrelated and $\rho_k = 0$ for $k > 0$, the variance of γ_k for $k > 0$ is approximately 1/N from Eq. (2.10). The square root of the estimated variance is standard deviation commonly referred to as standard error (SE). To obtain the 95 percent confidence interval (or equivalently the 5% significance interval), at a given lag, 1.96 times the SE above and below the axis is plotted.

In order to study the properties of the theoretical ACF for a stationary AR (p) process, Eq.. (2.7) is firstly multiplied by $(Z_{t-k} - \mu)$ to obtain

$$(Z_{t-k} - \mu)\,(Z_t - \mu) = \phi_1(Z_{t-k} - \mu)\,(Z_{t-1} - \mu) + \phi_2\,(Z_{t-k} - \mu)\,(Z_{t-2} - \mu) + \ldots + \phi_p\,(Z_{t-k} - \mu)\,(Z_{t-p} - \mu) + (Zt_{-k} - \mu)\,a_t \qquad \text{... (2.13)}$$

By taking expected values of Eq. (2.13), the difference equation for the autocovariance function of the AR (p) process is

$$\gamma_k = \phi_1\gamma_{k-1} + \phi_2\gamma_{k-2} + \ldots + \phi_p\,\gamma_{k-p}, \quad k > 0 \qquad \text{... (2.14)}$$

The term $E\,[(Zt_{-k} - \mu)\,a_t]$ is zero for k>0 because Z_{t-k} is only a function of the disturbances upto time t-k and a, is uncorrelated with these shocks. To determine an expression for the theoretical ACF for the AR (p) process, Eq. (2.14) is divided by γ_0 to obtain ρ

$$\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2} + \ldots + \phi_p\,\rho_{k-p}, \quad k > 0$$

This equation can be equivalently written as

$$(1 - \phi_1 B - \phi_2 B^2 - \ldots . \phi_p B^P)\ \rho_k = \phi(B)\ \rho_k = 0 \qquad , k > 0 \qquad \text{... (2.15)}$$

where B operates on k instead of t. The general solution of the difference Eq. (2.15) is (Box and Jenkins, 1976)

$$\rho_k = A_1 G_1^K + A_2 G_2^K + \ldots . + A_P G_P^K \qquad \text{... (2.16)}$$

where, G_I^{-1} ,G_2^{-1} ...G_p^{-1} are distinct roots of the characteristic Eq. $\phi(B)=0$ and the A_K's are constants. If a root G_1^{-1} is real then $|G_I^{-1}| > 1$ due to the stationarity conditions. Hence $|G_j| < 1$ and $A_j\ G_j^k$ in Eq. (2.16) forms a damped exponential which geometrically decays to zero as k increases.

2.2.2.2 Yule-Walker equations

The AR (p) parameters, ϕ_p are expressed in terms of theoretical ACF. The resulting set of linear equations are called the Yule-Walker equations and are given by

$$\begin{array}{lllllll}
\rho_1 & = & \phi_1 & + & \phi_2\rho_1 & + \ldots . + \phi_p\ \rho_{k-1} \\
\rho_2 & = & \phi_1\rho_1 & + & \phi_2 & + \ldots . + \phi_p\ \rho_{k-2} \\
\ldots & & \ldots & & \ldots & \ldots \\
\ldots & & \ldots & & \ldots & \ldots \\
\rho_p & = & \phi_1\rho_{p-1} & + & \phi_2\rho_{p-2} & + \ldots . + \phi_p
\end{array} \qquad \text{... (2.17)}$$

The solution of Eq. (2.17) for the parameters in terms of ϕ the autocorrelations may be written

$$\phi \quad = \quad P_p^{-1}\ \rho_p \qquad \text{... (2.18)}$$

Where

$$\phi = \begin{bmatrix} \phi_1 \\ \phi_2 \\ .. \\ .. \\ .. \\ \phi_p \end{bmatrix} \qquad \rho_p = \begin{bmatrix} \rho_1 \\ \rho_2 \\ .. \\ .. \\ .. \\ \rho_p \end{bmatrix} \qquad P_p = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \ldots & \rho_{p-1} \\ \rho_1 & 1 & \rho_1 & \ldots & \rho_{p-2} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \rho_{p-1} & \rho_{p-2} & \rho_{p-3} & \ldots & 1 \end{bmatrix}$$

2.2.2.3 The partial autocorrelation function (PACF)

If the ACF of an AR process attenuates and does not truncate at a specified lag, it is advantageous to define a function which does cut off for an AR process. Such a device is useful to employ in conjunction with the sample ACF and other tools for identifying the type of model to fit to a given data set.

Let ϕ_{kj} be the j[th] coefficient in a stationary AR process of order k so that ϕ_{kk} is the last coefficient. The Yule-Walker [Eq. (2.17)] can then be equivalently written as-

$$\begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-2} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \phi_{k1} \\ \phi_{k2} \\ .. \\ .. \\ .. \\ \phi_{kk} \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ .. \\ .. \\ .. \\ \rho_k \end{bmatrix} \quad \text{... (2.19)}$$

The coefficient ϕ_{kk} is a function of the lag k and is called the theoretical partial autocorrelation function (PACF). Because of the definition of theoretical PACF, it is equal to zero after lag p for an AR (p) process. The possible values of ϕ_{kk} range from -1 to 1.

When plotting ϕ_{kk} against lag k, approximate confidence limits must be given in order to decipher values of the estimate PACF which are significantly different from zero. If the process is AR (p), the sample PACF is not significantly different from zero after lag p. Based upon the hypothesis that the process is AR (p), the estimated values of the PACF at lags greater than p are approximately normally independently distributed with a SE given by (Barndorff-Nielsen and Schou, 1973).

$$\text{SERIES } [\phi_{kk}] = \frac{1}{\sqrt{N}}$$

where, N is the length of the time series. Solving Eq. (2.19) for k=1, 2, 3 … successively, The following PACF are obtained

$$\phi_{11} = \rho_1$$

$$\phi_{22} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}}$$

$$= \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} \qquad \text{... (2.20)}$$

and

$$\phi_{33} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_1 \\ \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_1 & \rho_3 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}}$$

In general, for ϕ_{kk} the determinant in the numerator has the same elements as that in the denominator, but with the last column replaced by ρ_k.

2.2.3 Moving average model of order q: MA (q)

The moving average process of order q is denoted by MA (q) and is written as

$$\tilde{Z}_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \ldots - \theta_q a_{t-q} \qquad \text{... (2.21)}$$

Equation (2.21) is a special case of Eq. (2.3) when only the first q of the ψ weights are non zero and the symbols used are $-\theta_1$, $-\theta_2$, …., $-\theta_q$ for the finite set of weight parameters. Eq. (2.21) can also be written in the equivalent form

$$\tilde{Z}_t = (1-\theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q)\, a_t$$

or

$$\tilde{Z}_t = \theta(B)\, a_t \qquad \text{... (2.22)}$$

Hence the moving average process can be thought of as the output Z_t from a linear filter with transfer function θ (B), when the input is white noise a_t. MA (1) process in Eq. (2.22) can be equivalently written as an infinite AR process by invoking binomial theorem.

$$\tilde{Z}_t = (1-\theta_1 B)\, a_t$$

or $a_t = (1-\theta_1 B)^{-1}\ \tilde{Z}_t$

$$= (1 + \theta_1 B + \theta_1^2 B^2 + \theta_1^3 B^3 + \ldots\ldots)\ \tilde{Z}_t \qquad \text{... (2.23)}$$

The infinite series (Eq. 2.23) converges for $|\theta_1| < 1$, consequently, the stationary MA (1) process can only be meaningfully expressed as an infinite AR process if a restriction is placed upon MA parameter. This restriction is referred to as the invertibility condition and is independent of the stationarity requirements of a process. The characteristic equation, for a MA (q) process is

$$\theta(B) = (1-\theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q) = 0 \qquad \text{... (2.24)}$$

For a MA (q) process to be invertible, the roots of the characteristic Eq. (2.24) must lie outside the unit circle.

The theoretical ACF for MA (q) process is found to be

$$\rho_k = \frac{-\theta_k + \theta_1\theta_{k+1} + \ldots. + \theta_{q-k}\theta_q}{1 + \theta_1^2 + \ldots.. + \theta_q^2} \qquad k = 1, 2, \ldots, q \qquad \text{... (2.25)}$$

$$= 0$$

It is shown in Eq. (2.23) that MA (1) process can be equivalently written as an infinite AR process. In general, any finite invertible MA process can be expressed as an infinite AR process. Because the PACF is theoretically defined to be zero after lag p for a finite AR (p) process, the PACF of a MA (q) process must therefore attenuate at increasing lags or equivalently an infinite AR process.

2.2.3.1 Duality between autoregressive and moving average processes

The aspects of duality between autoregressive and finite moving average processes are as follows (Box and Jenkins, 1976)

(i) In a stationary autoregressive process of order p, a_t can be represented as a finite weighted sum of previous Z's, or Z_t as an infinite weighted sum

$$\tilde{Z}_t = \phi^{-1}(B)\ a_t$$

of previous a's. Also, in an invertible moving average process of order q, Z_t can be represented as finite weighted sum of previous a's or a_t as an infinite weighted sum

$$\theta^{-1}(B)\ \tilde{Z}_t \ = \ a_t$$

of previous $\tilde{Z}$'s.

(ii) The finite MA process has an autocorrelation function which is zero beyond a certain point, but since it is equivalent to an infinite AR process, its partial autocorrelation function is infinite in extent and is dominated conversely, the AR process has a partial autocorrelation function which is zero beyond a certain point, but its autocorrelation function is infinite in extent and consists of damped exponentials and/or damped sine waves.

(iii) For an autoregressive process of finite order p, the parameters are not required to satisfy any conditions to ensure invertibility. However, for stationarity, the roots of ϕ (B) =0 must lie outside the unit circle. Conversely, the parameters of the MA process are not required to satisfy any conditions to ensure stationarity. However, for invertibility of the MA process, the roots of θ (B) = 0 must lie outside the unit circle.

2.2.4 Mixed autoregressive -moving average model: ARMA (p, q)

A finite moving average process

$$\tilde{Z}_t = a_t - \theta_1 a_{t-1} = \theta (1- \theta_1 B) a_t \qquad | a_t | < 1$$

can be written as an infinite autoregressive process.

$$\tilde{Z}_t = -\theta_1 \ \tilde{Z}_{t-1} \ -\theta_2 \tilde{Z}_{t-2} \ - \dots + a_t$$

Hence, if the process were really MA (1) a non parsimonious representation in terms of an autoregressive model is obtained. Conversely, an AR (1) could not be parsimoniously represented using a moving average process. In practice, to obtain a parsimonious parameterization, it will sometimes be necessary to include both autoregressive and moving average terms in the model. Thus,

$$\tilde{Z}_t = \phi_1 \tilde{Z}_{t-1} + \phi_2 \tilde{Z}_{t-2} + \ldots. + \phi_p \tilde{Z}_{t-p} + a_t - \ldots - \theta_q a_{t-q}$$

or $$\phi(B)\, \tilde{Z}_t = \theta\,(B)\, a_t \qquad \text{... (2.26)}$$

is called the mixed autoregressive-moving average model of order (p, q) or ARMA (p, q). The ARMA (1, 1) model can be written as

$$\tilde{Z}_t - \phi_1 \tilde{Z}_{t-1} = a_t - \theta_1\, a_{t-1}$$

Since Eq. (2.26) may be written

$$\begin{aligned} \tilde{Z}_t &= \phi^{-1}(B)\ \ \theta\,(B)\, a_t \\ &= \frac{\theta(B)}{\phi(B)}\, a_t \\ &= \frac{1-\theta_1(B)-\ldots.-\theta_q B^q}{1-\phi_1(B)-\ldots.-\phi_p B^p}\, a_t \end{aligned}$$

The mixed autoregressive-moving average process can be thought of as the output Z_t from a linear filter, whose transfer function is the ratio of the polynomials θ(B) and ϕ(B), when the input is white noise a_t. The ARMA (p, q) process, will define a stationary process, provided that the characteristic equation ϕ(B) = 0 has all its roots lying outside the unit circle. Similarly, the roots of θ(B) =0 must lie outside the unit circle if the process is to be invertible.

2.2.4.1 Autocorrelation function (ACF)

The theoretical ACF for an ARMA (p, q) process is derived in a fashion which is similar to that used for an AR process in Section 2.2.2.1 by multiplying both sides of Eq. (2.26) by Z_{t-k} and taking expectations.

$$\gamma_k = \phi_1\,\gamma_{k-1} + \phi_2\,\gamma_{k-2} + \ldots\ldots. + \phi_p\,\gamma_{k-p} + \gamma_{za}(k) - \theta_1\,\gamma_{za}(k-1) - \theta_2\,\gamma_{za}(k-2) + \ldots..+\theta_q\,\gamma_{za}(k-q) \qquad \text{... (2.27)}$$

where $\gamma_k = E\,[Z_{t-k}, Z_t]$ is the theoretical autocovariance function and $\gamma_{za}(k) = E\,[Z_{t-k}\,, a_t]$ is the cross covariance function between Z_{t-k} and a_t. Since Z_{t-k} is dependent only upon the shocks which have occurred upto time t-k, it follows that

$$\gamma_{za}(k) = 0 \qquad , \; k > 0$$

$$\gamma_{za}(k) \neq 0 \qquad , \; k \leq 0 \qquad \text{... (2.28)}$$

Because of the γ_{za} (k) terms in Eq, (2.27), it is necessary to derive other relationships before it is possible to solve for the autocovariance. This can be affected by multiplying Eq. (2.26) by a_{t-k} and taking expectations to get

$$\gamma_{za}(-k) - \phi_1 \gamma_{za}(-k+1) - \phi_2 \gamma_{za}(-k+2) - \ldots\ldots - \phi_p \gamma_{za}(-k+p)$$

$$= -[\theta_k]\, \sigma_a^2 \qquad . \qquad \text{… (2.29)}$$

where,

$$[\theta_k] = \begin{cases} \theta_k & , k = 1,2,\ldots.q \\ -1 & , k = 0 \\ 0 & , k = otherwise \end{cases}$$

Equations (2.27) and (2.29) can be employed to solve for the theoretical autocovariance function for an ARMA (p, q) process. For k > q, Eq. (2.27) reduces to

$$\gamma_k - \phi_1 \gamma_{k-1} - \phi_2 \gamma_{k-2} - \ldots\ldots. - \phi_p \gamma_{k-p} = 0$$

or $$\phi(B)\, \gamma_k = 0 \qquad \text{... (2.30)}$$

If k > r = max (p, q), Eq. (2.30) may be used to calculate the γ_k directly from the previous values. For k = 0, 1, 2… r, Eq. (2.29) can be used to solve for the cross covariances, γ_{za} (k), and then substituting the γ_{za} (k) into Eq.(2.27). The resulting equations can be solved to determine the theoretical auto covariance function for any ARMA (p, q) process where the values of the parameters are known. The theoretical ACF can then be determined by simply dividing by the variance.

Thus, for the ARMA (p, q) process, there will be q autocorrelations $\rho_q, \rho_{q-1}, \ldots\ldots, \rho_1$ whose values depend directly, through Eq. (2.27) on the choice of the q moving average parameters θ, as well as on the p autoregressive parameters ϕ. Also, the ρ values $\rho_q, \rho_{q-1} \ldots.. \rho_{q-p+1}$ provide the necessary starting values for the difference equation $\phi(B)\, \rho_k = 0$, where $k \geq q + 1$ which then entirely determines the autocorrelations of higher lags. If q-p < 0, the whole autocorrelation function ρ_j, for j = 0, 1, 2, ….. will consist of a mixture of damped exponentials and/or damped sine waves, whose nature is dictated by the polynomial ϕ(B) and the starting values. If, however, q-p $\geq$ 0 there will be q-p+1 initial values $\rho_0, \rho_l \ldots\ldots \rho_{q-p,}$ which do not follow this general pattern. These facts are useful in identifying mixed series.

2.2.4.2 Partial autocorrelation function, (PACF)

As a result of the MA operator, the ARMA (p, q) process in Eq. (2.26) can be written as an infinite AR process given by

$$a_t = \theta^{-1} (B) \phi (B) \tilde{Z}_t$$

where θ^{-1} (B)is an infinite series in B. Hence, the partial autocorrelation function of a mixed process is infinite in extent. It behaves eventually like, the partial autocorrelation function of a pure moving average process, being dominated by a mixture of damped exponentials and/or damped sine waves, depending on the order of the moving average and the values of the parameters it contains.

2.3 Linear Nonstationary Models

The ARMA models described in Section 2.2.4 can be fitted to stationary hydrologic series, such as the annual series. For nonstationary series such as monthly, weekly and daily series, the nonstationarity can be removed by the periodic standardardization. This procedure leads to useful models for the synthetic generation and forecasting of hydrologic series. However, the number of parameters required is generally large. For example, an ARMA (1, 1) model applied to monthly series requires 27 parameters (12 monthly means, 12 monthly standard deviations, ϕ_1, θ_1 and σ_a^2). However, a nonstationary time series can be transformed into a stationary series by alternate ways which leads to models requiring fewer parameters. If the series does not have a fixed mean but its successive changes or differences are stationary, then ARMA models can be extended to nonstationary series by working with their differences. It is possible to take the first, second, or in general, the d^{th} difference, which leads to simple nonperiodic ARIMA (p, d, q) models (also known as nonseasonal ARIMA models). It is also possible to take periodic or seasonal differences at lag s such as the 12th difference of monthly series, which leads to periodic ARIMA $(P, D, Q)_s$ models (also known as seasonal ARIMA models). The combination of nonperiodic and periodic ARIMA models leads to the multiplicative ARIMA (p, d, q) x $(P, D, Q)_s$ model which consists of a seasonal ARMA (P,Q) fitted to the D^{th} seasonal difference of the data coupled with an ARMA (p, q) model fitted to the d^{th} difference of the residuals of the former model.

2.3.1 Autoregressive integrated moving average model: ARIMA (p, d, q)

In the previous sections an important class of stochastic models, known as stationary models, which assume that the process remains in equilibrium about a constant mean level, have been discussed. When the process has no natural mean, it is called a nonstationary process and such series do not vary about a fixed mean. Such series may exhibit homogeneous behaviour of a kind. In particular, although the general level about which fluctuations are occurring may be different at different times, the broad behaviour of the series, when differences in level are allowed for, may be similar. A general model, which can represent homogeneous nonstationary behaviour, is of the form.

$$\phi\ (B)\ (1-B)^d\ \tilde{Z}_t = \theta\ (B)\ a_t \qquad \ldots (2.31)$$

where ϕ (B) is a stationary autoregressive operator. Introducing backward difference operator ∇ which can be written in terms of B, such that

$$\nabla\ Z_t = Z_t - Z_{t-1} = (1-B)\ Z_t$$

and since $\nabla^d Z_t = \nabla^d Z_t$, for $d \geq 1$ model (2.31) can be written as

$$\phi\ (B)\nabla^d Z_t = \theta\ (B)\ a_t \qquad \ldots (2.32)$$

Equivalently, the process is defined by the two equations

$$\phi\ (B)\ W_t = \theta\ (B)\ a_t \qquad \ldots (2.33)$$

and $\quad W_t = \nabla^d Z_t \qquad \ldots (2.34)$

The process defined by equations (2.33) and (2.34) provides a powerful model for describing stationary and nonstationary time series and is called an autoregressive integrated moving average (ARIMA) process of order (p, d, q). The process is defined by

$$W_t = \phi_1 W_{t-1} + \ldots .. + \phi_p W_{t-p} + a_t - \theta_1 a_{t-1} - \ldots -\theta_q a_{t-q} \qquad \ldots (2.35)$$

with $W_t = \nabla^d Z_t$. If W_t is replaced by Z_t-μ, when d=0, the model (2.35) includes the stationary ARMA model (2.26), as a special case, and also the pure autoregressive model (2.7) and the pure moving average model (2.21). The model (2.32) is equivalent to representing the process Z_t as the output from a linear filter (unless d=0 this is an unstable linear filter), whose input is white noise a_t. Alternatively, it can be regarded as a device for transforming the highly dependent and possibly nonstationary process Z_t to a sequence of uncorrelated random variables a_t; that is for transforming the process to white noise.

If in ARIMA model (2.32), the autoregressive operator $\phi(B)$ is of order p, the d^{th} difference is taken, and the moving average operator θ (B) is of order q, then it is called an ARIMA model of order (p, d, q) or simply an ARIMA (p, d, q) process.

2.4 Seasonal Models

Seasonal autoregressive integrated moving average (SARIMA) are useful for modelling seasonal time series in which the mean other statistics for a given season are not stationary across the year. The basic ARIMA model in its seasonal form is described as (Box and Jenkins, 1976, Hipel et al., 1977), a straight forward extension of the nonseasonal ARMA and ARIMA models described earlier.

Let Z_l, $Z_{2...}$, Z_n represent a sequence of seasonal observations. If for example there were n years of data for which each year contains s seasons this would mean that total number of data η would be equal

$$\nabla_S^D Z_t^\lambda = (1-B^S)^D Z_t^\lambda \qquad \text{...(2.38)}$$

For purposes of explanation, a time series consisting of monthly observations is considered. To model correlation among, say, June observations in the differenced series, one may wish to introduce appropriate model parameters. More specifically, to accomplish this task of linking June observations together the model of the following form is used.

$$\Phi(B^S)\ \nabla_S^D Z_t^\lambda = \Theta(B^S)\ \alpha_t \qquad \text{... (2.39)}$$

where Φ (B^S) and Θ (B^S) are the seasonal autoregressive (AR) and seasonal moving average (MA) operators, respectively, and α_t is a residual series which may contain nonseasonal correlation. Both the AR and MA operators are defined in order to describe relationships within the same season. In particular, the seasonal AR operator is defined as

$$\Phi(B^S) = 1 - \Phi_1 B^S - \Phi_2 B^{2S} - \ldots\ldots - \Phi_P (B^{PS})$$

where Φ_1 is the i^{th} AR parameter and P is the order of the AR operator. Because the power of each differencing operator is always an integer multiple of s, only the observations within each season are related to one another when using this operator. Hence, for the case of June observations in a monthly series, only the June observations are connected together using $\Phi\ (B^S)$. To describe the relationship of the residuals, $\alpha_{t,}$ within a given season, the seasonal MA operator is defined using

$$\Theta\ (B^S) = 1 - \Theta_1 B^S - \Theta_2 B^{2S} - \ldots. - \Theta_Q B^{QS}$$

where Θ_i is the i^{th} MA parameter and Q is the order of the MA operator. Since the exponents of B in $\Theta\ (B^S)$ are always integer multiples of s, the residuals in the same season are linked with one another when using $\Theta\ (B^S)$.

Theoretically a separate model as in Eq. (2.39) could be defined for each season of the year. However, to keep the model as parsimonious as possible, it is assumed that Eq. (2.39) can be used for all the seasons. Therefore, the assumption is made that the correlation within all of the seasons is the same. For the case of monthly data this means that the relationship among all of the June observations is exactly the same as each of the other months.

The error components or residuals, α_1, may contain nonseasonal nonstationarity which can be removed by using the nonseasonal differencing operator defined in Section 4.2.1 as

$$\nabla^d \alpha_t = (1\text{-}B)^d \alpha_t \qquad \ldots (2.40)$$

where d is the order of the nonseasonal differencing operator which is selected just large enough to remove all of the nonseasonal nonstationarity. The sequence produced using Eq. (2.40) is theoretically a stationary nonseasonal series. The nonseasonal correlation can then be captured by writing the ARMA model in Eq. (2.26) or Eq. (2.33) as

$$\phi\ (B) \nabla^d \alpha_t = \theta\ (B)\ a_t \qquad \ldots (2.41)$$

where $\phi\ (B)$ is the nonseasonal AR operator of order p defined as

$$\phi\ (B) = 1\text{-}\ \phi_1 B - \phi_2 B^2 - \ldots.. - \phi_p B^p$$

and $\phi\ (B)$ is the non seasonal MA operator of order q written as

$\theta (B) = 1 - \theta_1 B - \theta_2 B^2 - ..... - \theta_q B^q$

The a_t's are the innovations which are normally identically independently distributed (IID) with a mean of zero and variance σ_a^2.

Because of the form of Eq. (2.41), the correlation among seasons is the same no matter what season is being dealt with. Hence, when entertaining monthly data, the correlation between, say, the June and May observations is defined to be the same as that between any other adjacent months such as October and September.

To define the overall seasonal model, Eq. (2.41) and Eq. (2.39) are combined. This is accomplished by solving for α_t in Eq. (2.41) and substituting this result into Eq (2.39) to obtain the SARIMA (seasonal autoregressive integrated moving average) model.

$$\phi (B)\, \Phi (B^s)\, \nabla^d \nabla_s^D Z_t^{\lambda}. = \theta (B)\, \Theta (B^S)\, a_t \qquad \text{... (2.42)}$$

Because the operators in Eq. (2.42) are multiplied together rather than summed, this model is often called a multiplicative SARIMA model.

When fitting the SARIMA model to a given time series of length η, if necessary, the data are first transformed using the Box-Cox transformation in Eq. (2.36). Following this, the data can be differenced both seasonally and nonseasonally. It does not matter which differencing operation is carried out first. One then obtains the stationary series given by

$$W_t = \nabla^d \nabla_s^D Z_t^{\lambda} \qquad \text{... (2.43)}$$

where the length of the W_t, series is $\eta' = \eta\text{-d-sD}$. The seasonal and nonseasonal correlation in the W_t sequence is modelled by using the seasonal and nonseasonal AR and MA operators respectively. Hence, W_t is modelled by employing

$$\phi (B)\, \Phi (B^s)\, W_t = \theta (B)\, \Theta (B^S)\, a_t \qquad \text{... (2.44)}$$

In some applications, W_t may be stationary seasonal series which is not obtained by differencing the original series. The model in Eq. (2.44) is called seasonal ARMA or SARMA model of the W_t series.

2.4.1 Notation of SARIMA models

An economic notation for summarizing the structure of the SARIMA model in Eq. (2.42) is $(p, d, q) \times (P, D, Q)_S$. The first set of brackets contains the orders of the nonseasonal operators while the orders of the seasonal operators are listed inside the second set of brackets. More specifically, p, d, and q stand for the orders of the nonseasonal AR, differencing and MA operators, respectively. In the second set of brackets, P, D and Q give the orders of the seasonal AR, differencing and MA operators, respectively. The subscript s appearing to the right of the second set of brackets points out the number of seasons per year.

For the case of monthly data for which s=12, a specific example of a SARIMA model is $(2, l, 1) \times (1, 1, 2)_{12}$. By utilizing Eq. (2.42) this model is written using a finite difference equation as

$$(1-\phi_1 B - \phi_2 B)\,(1-\Phi B^{12})\,(1-B)\,(1-B_{12})\,Z_t^{\lambda} = (1-\theta B)\,(1-\Theta B^{12} - \Theta B^{24})\,a_t$$

If the data are stationary, nonseasonal or seasonal differencing is not required. A stationary model is indicated as $(p, 0, q) \times (P, 0, Q)_S$. Because this model is stationary, sometimes it is referred to as a SARMA (i.e. seasonal autoregressive-moving average) $(p, q) \times (P, Q)_S$ model.

The summary notation for a pure MA model is $(0, d, q) \times (0, D, Q)_s$. When a model contains no MA parameter, the SARIMA model is written as $(p, d, 0) \times (P, D, 0)_s$.

When a model is purely nonseasonal, the notation of nonseasonal models, should be used, Hence, a stationary nonseasonal ARMA model is simply indicated by ARMA (p, q) instead of SARMA $(p, q) \times (0, 0)_1$. Likewise, a nonstationary nonseasonal ARIMA model is denoted as ARIMA (p, d, q) rather than the more cumbersome notation given by SARIMA $(p, d, q) \times (0, 0, 0)_1$.

2.5 Application of Various Stochastic Models

Sharma Arun Kumar (1998) reviewed extensively about the application of stochastic models. Here only some important and latest reviews are presented.

Bidwell and Griffitsch (1994) developed a model for flood forecasting for basins with highly nonlinear inflow response to rainfall with a low data input and resources requirement, suited to telemetered rainfall and river stage data, was developed. Rainfall-stage and upstream stage-downstream stage relationships were identified as autoregressive moving average (ARMA} models with time-varying parameters. Parameter values were estimated in adaptive real time mode by incorporating the ARMA models into the Kalman filter algorithm.

Bender and Simonovle (1994) used seasonal autoregressive integrated moving average (SARIMA) model for forecasting monthly water supply. Models were developed and applied to three types of river-basin data within a sensitivity analysis of flow scenarios. Ranking of model performance for possible system scenarios suggested a set of rules to govern the choice of a single model to produce the best available forecast. SARIMA models appeared to be more flexible for natural inflows with low upstream storage capacity and high variability. Deseasonalized ARMA models may be better suited to natural inflow systems that have a large storage capacity, lower variability and greater response lags to precipitation events.

Srinivasan and Thandaveswara (1995) fitted lower order periodic autoregressive/ autoregressive moving average (PAR/PARMA) models to the monsoon-dependent river inflows of southern India, measured at Chunchanakatte (Cauvery river), Akkihebbal (Hemavathi river) and Unduwadi (Lakshmanathirtha river). Power transformation with periodic exponents (PTPE) and Wilson-Hilferty transformation (WHT) were found to be the most suited ones. In all the three cases, the more commonly used log and square-root transformations were rejected by the normality test of residuals. The periodic skewness was in general reproduced better by WHT, while periodic mean and periodic variance were better reproduced by PTPE. Even the negative skewness coefficients were reproduced by WHT. However, any of the periodic models could not reproduce the periodic skewness or preserve the periodic correlations of low flows with high skewness and low correlations. It was concluded that the most recent periodic models (PAR/PARMA) do not seem to perform to expectation in modelling a few cases of the highly variable monsoon dependent river flows.

Samani et. al. (1995) applied time series techniques to Ghara- Aghaj flow records, in order to generate forecast values of the mean monthly river inflows. The autoregressive models of order one and two (AR1, AR2) moving average model of order one and ARMA (1, 1) model were fitted to the stationary series, where the AR2 model yields results which are statistically compatible with the past records.

Gorantiwar et. al. (1995) applied autoregressive {AR) models upto fourth order to the historical annual river inflows and logarithmic transformation of the historical annual river inflows of Barkar river (Maithon dam site) to generate synthetic annual river inflows that will resemble historical sequence in terms of statistical properties such as mean, standard deviation, skewness coefficient and lag-one serial correlation coefficient. Mean, standard deviation and lag one serial correlation coefficient were preserved in generated river inflows of all AR models applied to original river inflows as well as logarithmic transformation of original streamflows as statistical properties of historical sequence fell in the 95 per cent confidence limit of statistical properties of generated sequence. However, skewness coefficient was preserved in the generated river inflows of all AR models applied to logarithmically transformated river inflows only. Diagnostic checking of residual series showed that AR model of fourth order applied to logarithmically transformated river inflows fit the annual river inflow data well in preservation of all the properties.

Narulkar (1995) used time series modelling with basic ARIMA model in its seasonal form and used a multivariate structure for four reservoirs located in hydro meteorologically homogenous areas of MRP command area, Madhya Pradesh. The data were analyzed in single site perspective as well as multisite perspective. In multisite case a simple seasonal time series model, the seasonal AR (1) model was attempted and the result of forecast with all the models were compared to evolve an appropriate model for forecasting of inflows to MAP system.

The seasonal characteristics of precipitation and temperature impart periodic behaviour on the resulting components of the hydrologic cycle. The nonstationarity produced by these seasonal characteristics could be taken care by passing through a linear-time invariant filter or a deseasonalisation technique as far as the correlation structure is not periodic. But, the monthly or weekly hydrologic time series usually display a periodic correlation structure. This necessitates a recourse to the periodic AR or ARMA models, which allows the period i.e. variation of the model parameters apart from satisfying the basic assumption of second order stationarity. A striking feature of the periodic AR or ARMA models is that the parameters of each season can be estimated independently of the parameters of the other seasons (Srinivasan, 1995).

Raman et. al. (1995) used five regression models namely, runoff coefficient model, single linear regression, monthly linear regression, monthly linear regression with stochastic description for residuals and a double regressed model for extending the streamflow data of Kudhiraigar basin. It was found that the monthly linear regression model with stochastic description for the residuals was the most suitable for extending the flow recorded among the various competing models. This model performed very well by having minimum error criteria in comparison to the other models considered.

Srinivasan (1995) applied periodic Gaussian univariate models with different transformations to the southwest monsoon based flows of the Cauvery river in southern India. The models considered were PAR (1), PAR (2), PARMA (1, 1) and Thomas Fiering 3-parameter model (only log transformation). The transformations attempted were Natural logarithm, square root, square root (log), Wilson-Hilferty and Log (Wilson-Hilferty). The diagnostics as well as the verification of monthly historical flow showed that the periodic model of ARMA (1, 1) type with Wilson-Hilferty transformation performed the best in terms of the overall reproduction of basic statistics.

Several attempts have been made in the past to model hydrological processes such as monthly river inflows in dry regions. One of the crucial problem in modelling this type of process is the handling of zero inflows. Chebaane and Salas (1995) presented a stochastic model which enables the reproduction of the percentage of zero inflows in each month, the monthly mean and variance, and the month-to-month correlation of the intermittent inflows. The model considers the intermittent monthly inflow process as a product of a periodic binary discrete process and a periodic continuous process. Both the discrete and the continuous processes are periodic first-order autoregressive. Parameter estimation has been developed based on the method of moments, method of transition probability and method of maximum likelihood.

Fernandez and Pizarro (1996) used linear transfer function models with precipitation series as input to estimate statistical properties such as monthly means and variances of resulting runoff series. They suggested empirical relationships based on data from watershed in the mountainous zone of Central Chile to estimate parameters of low order transfer function models and some of their properties.

Sharma et. al. (1997) used kernel estimates of the joint conditional probability density functions to generate synthetic river inflow sequences. River inflow is assumed to be a Markov process with time dependence characterised by a multivariate probability density function. Kernel methods were used to estimate this multivariate density function Simulation process by sequentially resampling from the conditional density function derived from the Kernel estimate of the underlying multivariate probability density function. This is a nonparametric method for the synthesis of river inflow that is data-driven and avoids prior assumptions as to the form of dependence (e.g. linear or nonlinear) and the form of the probability density functions (e.g. Gaussian). They showed using synthetic examples with known underlying models that the nonparametric method presented in more flexible than the conventional models used in stochastic hydrology and is capable of reproducing both linear and nonlinear dependence. The effectiveness of this model is illustrated through its application to simulation of monthly river inflow from the Beaver River in Utah.

Montan.ari and Rosso (1997) stated that since Hurst (1951) detected the presence of long-term persistence in hydrologic data, new estimation methods and long-memory models have been developed. The lack of flexibility in representing the combined effect of short and long memory has been the major limitation of stochastic models used to analyze hydrologic time series. They considered a fractionally differenced auto regressive integrated moving average (FARIMA) model. In contrast to using traditional ARIMA models, this approach allows the modelling of both short and long term persistence in a time series. A framework for identification and estimation is presented. The data do not have to be Gaussian. The resulting model, which replicates the sample probability density of the data, can be used for the generation of long synthetic series. They presented an application to the monthly and daily inflows of Lake Maggiore, Italy.

Benalaya et. al. (1998) collected rainfall data from six weather stations of northern Tunisia with the objective of identifying the temporal rainfall trends. The rainfall time series were described in terms of mean, standard deviation, coefficient of variation, coefficient of skewness and statistical normality tests. The series were stabilized and used to produce estimates of monthly rainfall.

Sharma Arun Kumar (1998) evaluated various stochastic models for forecasting Jakham river monthly inflows in southern Rajasthan. Autocorrelation function, Partial autocorrelation function, inverse autocorrelation function, inverse Partial autocorrelation functions were analysed to identify the class and order of stochastic models to represent Jakham river inflows. The parameters of the identified models were estimated by conditional least square method. Validation of the identified models suggested that only ARMA (3, 0), ARMA (2, 1) and ARMA (3, 1) IIAR constrained models passed the tests. The minimum mean square error criterion was used to select the best model. It revealed that ARMA (3, 0) model was proved to be the best model.

Singh (1998) studied the persistence structure causing interannual variability in monsoon rainfall by stochastic modelling of monsoon rainfall data at 50 different stations across India. For this, correlograms and partial correlograms were developed from 41 years of data at each station. Due to the resulting non-existence of persistence structure in the data, statistical modelling of these data was carried out instead of quantitative estimation of monsoon rainfall at specified recurrence intervals. To overcome the difficulties in choice of a statistical distribution, the method of power transformation was used. The results of the models were verified with the observed rainfall at all stations. Maps were developed which can be used for quantitative estimation of monsoon rainfall at ungauged locations.

Reddy and Devendra Kumar (1999) developed a time series model for average monthly rainfall and applied the same on Bino watershed of Ramganga river. They found that the autoregressive model of order one fitted best to the dependent stochastic component.

Bhakar, S.R. (2000) developed stochastic model for weekly evaporation using 20-year data. Validation of the developed model was done by comparison of the estimated values with measured values. Stochastic model was found to predict evaporation very accurately. Stochastic models were also developed for estimation of daily Wheat evapotranspiration and Green gram evapotranspiration was found to predict the daily crop evapotranspiration very accurately.

Montanari et. al. (2000) introduced a seasonal fractional integrated moving average (ARIMA) model, with both short and long term persistent periodic components. The estimation of the parameter was carried out by applying the Whittle's approximation to the Gaussian maximum likelihood function, which yielded asymptotically consistent estimates. The method was applied to the monthly flows of the Nile river at Aswan and the results were compared with ones obtained by applying heuristic procedures.

Raghuwanshi and Wallender (2000) applied time-domain methodology to forecast daily ET_0 values for Davis, California. Stochastic process of daily crop ET was characterized by both the first order autoregressive AR (1) and autoregressive moving average ARMA (1, 1) models. The ARMA (1, 1) model with the least square method of estimation preserved variance and kurtosis better than the AR (1) model. For both the parameter estimation methods, ARMA (1, 1) model performed better than the AR (1) model. ARMA (1, 1) model with the least squares estimation method was the best model for forecasting ET_0 and can be used to forecast ET_0 one day ahead for Davis CIMIS station.

Koutosoyiannis (2001) proposed a methodology for coupling stochastic models of hydrologic processes applying to different time scales so that the time series generated by different models be consistent. Given two multivariate time series generated by two separate stochastic models of the same hydrologic process, each applying to different time scale. A transformation was developed that appropriately modified the time series of the lower-level time scale so that this series becomes consistent with the time series of higher level and an appropriate correlation between the two time series was established.

Chen and Rao (2002) explained that hydrological monthly series are stationary; a segmentation algorithm is applied so that non-stationary series are identified and partioned into stationary segments. Four sets of hydrologic data were analysed, including monthly stream flow, temperature, precipitation and Plamer's drought severity index series. The first order differenced standardized monthly series were also analysed in addition to the standardized monthly series. The results indicated that more series of monthly streamflow and Plamer's drought severity index are identified as non-stationary than stationary, while more series of monthly temperature and precipitation are stationary. In general, standardized hydrologic monthly series, either differenced or not, are non-stationary.

Jha, V. (2002) developed stochastic models for the estimation of weekly soil moisture content for various depths using 19 years data (1980-1998). The developed models were validated using latest three-year data (1999-2001). The developed stochastic models were found to predict the soil moisture accurately at different depths.

Lefevre, M. (2002) modeled the logarithm of the flow of a river in Quebec, Canada. Using the model, forecasts of the river flow were made for up to 7 days ahead. The forecasts were found to be much more accurate for 1 day ahead than those produced by a deterministic model called PREVIS. The stochastic model also outperforms PREVIS for 2 days ahead forecasts and is comparable to PREVIS for 3 days ahead forecasts.

Pandey, P.K. (2002) developed stochastic model for estimation of daily black gram evapotranspiration and Kumar, A. (2003) developed stochastic model for estimation of daily Okra evapotranspiration. The developed stochastic models were found to predict daily evapotranspiration very accurately.

Phoon et. al. (2002) presented a practical inverse approach for forecasting non-linear hydrological time series. The approach involved calibrating delay time, embedding dimension and number of nearest neighbors simultaneously using a single definite criteria, namely, optimal prediction accuracy and verifying that the optimal parameters had wider applicability outside the scope of calibration and demonstrating that chaotic behaviour was presented when optimal parameters were used in conjunction with existing system characterization tools. The proposed approach was shown to be better than the standard approach for a theoretical chaotic time series (Mackey-Glass) and two real runoff time series

Patil, R.M. (2003) made stochastic modeling for water deficit by using 24 years (1976-1999) data. The time series was found trend free. The generated series was compared with observed water deficit series. Developed autoregressive model for Rahuri was validated by predicting two years ahead and compared with observed water deficit series. The results indicate a high degree of model fitness to observed data series.

Verma, A. (2004) developed stochastic model on monthly rainfall of Kota, Rajasthan. The historical data series were normalized by square root transformation. Fourier analysis was used for determination of periodic components and number of significant harmonics was determined by standardized to remove the periodic components. The parameters of AR models were estimated by the general recursive formula proposed by Kottegoda (1980) whereas the parameters of ARMA models were computed by the use of Yule-Walker equations. Box-Pierce Portmanteau lack of fit test and Akaike Information Criterion were used to select the models. The performance of the models in generation and forecasting values of monthly rainfall data were evaluated quantitatively and qualitatively by comparison of historical and selected model correlograms and goodness of fit tests such as mean forecast error, mean absolute error, root mean square error and integral square error. The low values of errors suggest the applicability of AR (1) and ARMA (1, 1) models for real time forecasting of monthly rainfall data of Kota, Rajasthan.

A few of the significant works in literature dealing with the development and/or application of periodic Gaussian models to river inflow time series are: Crolay and Rao (1976), Tao and Delleur (1976), McLeod and Hipel (1978), Hirsch (1979), Salas et. al. (1980), Salas et al. (1982), Vecchia et. al. (1983), Vecchia (1985).

The literature available on the application of stochastic models for forecasting river inflows is enormous. In the present study a brief review on hydrologic time series with particular emphasis on river inflow time series modelling has been presented. It may be mentioned that not much effort has been put by hydrologists in identifying a most suitable model. Probably, more attention is required in this regard. It may further be mentioned that although extensive work has been done on monthly flow modelling, the use of such models for modelling monthly flows for Indian rivers may require some modification. This is because many Indian rivers have nearly zero flows during non-monsoon season. Case studies regarding application of time series models to Indian rivers will be highly informative and good contribution to hydrology literature as very little has been done in the application side.

3. MATERIALS AND METHODS

3.1 General

Time series modelling for either data generation or forecasting of hydrologic variables is an important step in the planning and operational analysis of water resource systems. The stochastic simulation of river inflows is a technique with potential for aiding in the analysis of water resource systems. Stochastic simulation is based on the premise that a river inflows record is only a single realization in time. The statistical properties of the recorded river inflows give sample estimates of the possible future long-term river inflows. Stochastic models of river inflows generation are becoming more accepted as a tool in hydrologic planning and design. A stochastic model is employed to preserve certain estimates of statistical properties of a measured record and to use those estimates to generate many equally likely sequences. The equally likely sequence can be used in place of the historical record, and a spectrum of equally likely optimal design may be developed. For many hydrologic studies and particularly in the planning and operation of water resource systems, the need for stochastic simulation has been recognized for quite some time. Before being able to generate synthetic sequences and/or forecast future values, models have to be found which describe past data adequately. Ideally these models should preserve all properties of the, observed data, in practice, however, this can not be achieved.

Autoregressive (AR) and autoregressive integrated moving average (ARIMA) models have an important place in the stochastic modelling of hydrologic data. Generation of river inflows by such models has been developed for annual, monthly and daily time series, of which the monthly ones offer the most opportunity for development and use. The reason for this is that monthly series were found to be most appropriate for analysis (Roesuer and Yevdjevich, 1966). Monthly values show the basic structure of precipitation and runoff series with both deterministic and stochastic components which are necessary for establishing the models to generate operational hydrology.

The present study was undertaken to develop the stochastic model of Mahi river inflows in southern Rajasthan. The purpose of this chapter is to briefly describe the procedure adopted for development and analysis of the stochastic model for Mahi river inflows to achieve the objectives described earlier.

3.2 Description of the Study Area

Mahi river basin is located in south-eastern Rajasthan, between latitudes $23^0 04'$ and $24^0 35'$ N and longitudes $73^0 18'$ and $74^0 52'$ E. It lies south of Banas basin, its eastern edge borders Chambal basin in Madhya Pradesh, and its western edge borders Sabarmati basin. Mahi river originates in the Mahi Kanta hills in the Vindhyachal range, in the western part of Madhya Pradesh, and enters Rajasthan in Banswara District, near Chandangarh. It leaves the State at Salakari village. On an average the river is about 100 - 130 m wide and it flows mostly through rocky terrain. Its banks are steep, though not very high. Mahi river basin extends over parts of Banswara, Chittorgarh, Dungarpur and Udaipur Districts. The total catchment area of the basin is 16,985 Sq. km according to the 1:250,000 scale topographical maps published by the Survey of India. The Catchment area upto the Mahi Bajaj Sagar Dam site is 6,149 Sq. km. orographically, the basin is marked by hilly terrain belonging to the Aravali chain. Ground elevations in the southern hilly part range from +465 to +1046 m above MSL approximately, while the alluvial plain elevations range from +208 to +262 m approximately. The mean annual rainfall over the Mahi basin is around 700 mm, of which about 94 per cent falls during the four monsoon months (June-September).

The main tributaries of the Mahi river are the Anas, Hiran, Eru and Chap rivers, in Banswara District. Of these, only the Anas river is perennial. The Jakam and Gomti rivers are the next most important downstream tributaries of the Mahi river, originating from Chittorgarh and Udaipur Districts, respectively. In Dungarpur District, the last lap of the Mahi river in Rajasthan, the main tributary is the Som river. Another tributary, the Moran, a seasonal river, also flows through this District. There are three major projects (Mahi Bajaj Sagar, Jakham and Jaisamand), two medium and 220 minor irrigation projects in the Mahi river basin, as well as some small irrigation systems (covering less than 20 ha.) constructed and operated by Panchayat Samities. There are no tanks in Mahi Bajaj Sagar project catchment. At present there is no planning of developments and upstream reservations on the project. The salient features of Mahi Bajaj Sagar project are given in Table 3.1.

3.3 Data Collection

Monthly inflow data for 76 years i.e. from 1928 to 2003 for Mahi river at Mahi Bajaj Sagar Dam site has been collected from the office of the Chief Engineer, the Mahi Bajaj Sagar

Project, Banswara. Monthly inflow series for 76 years i.e. 1928 to 2003 was used for stochastic model building. Total annual inflow has also been calculated from the collected data.

Table 3.1 Salient features of Mahi Bajaj Sagar Project, Banswara

S.No.	Particulars	Value
1	Location	Near village Bor Khera about 16 Km North-East of Banswara town.
2	Catchment area	6149.00 Sq. km. Rajasthan 1515 Sq. km. Madhya Pradesh 4634 Sq. km.
3	Reservoir data	
	a) Top of dam	EL. 284.50 m.
	b) Maximum water level	EL. 281.50m.
	c) Minimum draw down level	EL. 259.00 m.
	d) Water spread area	142.90 Sq. km. Rajasthan 135.04 Sq. km. Madhya Pradesh 7.86 Sq. km.
	e) Gross capacity at FRL. 280.75 m	20648 lac. cum
	f) Live storage capacity	17180 lac. cum
4	Unit-1	
	Dam and appurtenant works	
	A) Earthen dam	
	i) Length of left flank	2320.25 m.
	ii) Length of right flank	353.75 m.
	iii) Road width at top	8.00 m.
	iv) Top of road	EL. 284.50 m.
	B) Masonry dam	
	Non-over flow section	
	i) Total length	135.00 m.
	C) Gated spillway	
	i) Length	300.00 m.
		Total length of dam 3109 m.
	ii) Crest EL	268.50 m.
	iii) Crest Gates	16 Nos. Radial Gates
		Size: 15 x 13.75
	iv) Routed flood discharge	23270 cumecs
	v) Peak flood discharge	25590 cumecs
5	Unit – II (Canal system)	
	A) Left main canal FSL	231.11 m
	i) Discharge at head	62.53 cumecs
	ii) Length	36.12 km
	iii) CCA	40570 ha
	B) Right main canal FSL	231.53 m
	i) Discharge at head	30 cumecs
	ii) Length	71.72 km
	iii) CCA	35910 ha
	C) Bhungra canal FSL	257.50 m

	i) Discharge at head	3.197 cumecs
	ii) Length	39.80 km
	iii) CCA	3490 ha
6	Unit –III (Common civil works)	
	A- Surface power house No. 1	
	i) Installed capacity	2 x 25 MW
	ii) Type of turbine	Francis
	iii) Design head adopted	40 m
	B- Balancing reservoir no. 1 (Kagdi pick up weir)	
	i) FSL	236.00
	ii) Total length of dam	952 m
	iii) Gross balancing capacity	4.24 m cum
	iv) Nos. of gates	5 nos. Radial gates size : 6m x 4.40m
	v) Catchment area	35.22 Sq. km
	C- Balancing reservoir no. 2	
	i) FSL	220.50
	ii) Length of penstock	370.0 m
	iii) Penstock diameter	5.3 m
	iv) Balancing capacity	22.4 lac cumec
	v) Length of dam	2956 m
	D- Power house no. 2	
	i) Installed capacity	2 x 45 MW
	ii) Average power potential	20 MW
	iii) Design head	83.81 m
7	Benefit Cost ratio	1.51 : 1
8	Cost per hectare of CCA	Rs. 75,100

3.4 Mathematical Procedure of Analysis

The mathematical procedure adopted for formulation of a model has been discussed in the following subsections.

3.4.1 Basic statistical characteristics of time series

Statistical analysis of data is the first step for its mathematical modelling. The most common statistical characteristics are defined below:

The basic statistical characteristic of a time series X_t=1, ... N is the **sample mean** given by

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \qquad \ldots (3.1)$$

where,

x_i =data sequence
$\bar{x}$ =arithmetic mean

N=number of data point

It measures the central tendency of x_t or determines where the series is located as a whole.

The second important statistical characteristic of a time series in the **sample variance** s^2 given by

$$s^2 = \frac{1}{N-1}\sum_{t=1}^{N}(X_t - \bar{X})^2 \qquad \ldots (3.2)$$

The estimate of Equation (3.2) is the most generally used in statistical hydrology. The square root of s^2 is called the **standard deviation**. Related to the mean and standard deviation is the **coefficient of variation** σ/μ and $s/\bar{x}$ for the population and sample, respectively. Coefficient of variation (C_V) is calculated as

$$C_V = \frac{s}{\bar{X}} \times 100 \qquad \ldots (3.3)$$

If the estimated value of coefficient of variation is significantly different than zero, the significant variability of the time series with respect to time is concluded and the series can be assumed to contain nondeterministic component which can be modelled on stochastic theory. Just as the mean measures the location of a time series X_t, the standard deviation measures the dispersion or the spread of the series around the mean $\bar{X}$. A small value of s means that the values $X_1, X_2, ..., X_n$ do not defer much from $\bar{X}$, while a large value of s generally means that the X's have a large spread around $\bar{X}$.

The sample **skewness coefficient** of a time series may be determined by

$$C_s = \frac{\sum (X_t - \bar{X})^3}{N(N-1)(N-2)s^3} \qquad \ldots (3.4)$$

with s obtained from Equation (3.2). The coefficient of skewness measures the asymmetry of time series.

Kurtosis characterized the relative peakness or flatness of a distribution compared with the normal distribution. Positive kurtosis indicates a relatively peaked distribution, whereas negative kurtosis indicates a relatively flat distribution. Kurtosis can be expressed as:

$$Kurtosis = \left\{ \frac{N(N+1)}{(N-1)(N-2)(N-3)} \sum_{i=1}^{N} \left(\frac{x_i - \overline{X}}{s} \right)^4 \right\} - \frac{3(N-1)^2}{(N-2)(N-3)} \qquad \text{... (3.5)}$$

The autocovariance function measures the degree of linear autodependence (self-dependence) of a time series. The autocovariance C_k between X_t and X_{t+k} may be determined by

$$C_k = \frac{1}{N} \sum_{t=1}^{n-k} (X_t - \overline{X})(X_{t+k} - \overline{X}) \quad , 0 \leq k < N \qquad \text{… (3.6)}$$

where C_k is usually called the lag-k autocovariance, k represents the time lag (or distance) between the correlated pairs (X_t, X_{t+k}). $\overline{X}$ is the sample mean of Eq. (3.1) and N is the sample size. For the particular case that k=0, C_0 becomes the variance s^2 of Eq. (3.2). The sample autocovariance C_k of Eq. (3.6) is a biased estimate of the population autocovariance function r_k. An unbiased estimate can be obtained by using (N-k), instead of N in the denominator of Eq. (3.6). In either case such estimators are referred as the open series estimators. These estimators have only N-k terms in the cross-products of Eq. (3.6).

A dimensionless measure of linear dependence is obtained by dividing C_k of Eq. (3.6) by C_0 such operation gives

$$r_k = \frac{C_k}{C_0} = \frac{\sum_{t=1}^{N-K} (X_t - \overline{X})(X_{t+k} - \overline{X})}{\sum_{t=1}^{N} (X_t - \overline{X})^2} \qquad \text{… (3.7)}$$

where r_k is called the (lag k) **autocorrelation coefficient**, the **serial correlation coefficient** or the **autocorrelation function (ACF)**.The plot of r_k versus k is generally called the **correlogram.** The sample autocorrelation coefficient r_k is an estimate of the population coefficient ρ_k. The most currently used simple measure of time dependence is the first serial correlation coefficient of the population.

For strictly random sequences, the value of r_1 necessarily differs from zero only by sampling variation; for sequences showing strong persistence, its value is close to one. Negative values of the lag-one serial correlation are possible hydrological sequences; they imply that large values in the sequence tend to be followed by small ones, and vice-versa (Clarke, 1973).

3.4.2 Complex characteristics of periodic time series

The statistical characteristics such as the mean, variance, skewness coefficient and serial correlation of periodic hydrologic time series can be determined by Eq. (3.1) through Eq. (3.7). However, such equations would only give the statistical characteristics as a whole and they will not show the effect of the annual cycle (except for the case of r_k). In order to take into account such effect the characteristics must be determined for each time interval within the year.

Considering the periodic time series $X_{v,t}$ where v denotes the year and t denotes the time interval within the year, the sample mean for the time interval t is determined by

$$\overline{X}_t = \frac{1}{N}\sum_{v=1}^{N} X_{v,t} \quad , t = 1 \ldots w \qquad \ldots (3.8)$$

where, N is the number of years of record and w is the number of time intervals in the year, 12 for the monthly series. The sample mean $\overline{X}_t$ is an estimate of the population mean μ_t.

The sample variance for time t is given by

$$S_t^2 = \frac{1}{(N-1)}\sum_{v=1}^{N}(X_{v,t} - \overline{X}_t)^2 \qquad \ldots (3.9)$$

It is an estimate of the population variance σ^2_t. Similarly, the sample skewness coefficient for time t is

$$C_{k\text{-}t} = \frac{\sum_{v=1}^{N}(X_{v,t} - \overline{X}_t)^3}{N(N-1)(N-2)s_t^3} \qquad \ldots (3.10)$$

where X_t is given by Eq. (3.8) and s_t is obtained from Eq. (3.9). The skewness coefficient $C_{k\text{-}t}$ of Eq. (3.10) is an estimate of the population skewness coefficient r_t.

The correlation structure of the time series $X_{V,t}$ may be determined for each time interval t by

$$r_{k,t} = \frac{\frac{1}{N}\sum_{v=1}^{N}(X_{v,t} - \overline{X}_t)(X_{v,t-k} - \overline{X}_{t-k})}{s_t . s_{t-k}} \quad \ldots (3.11)$$

where $r_{k,t}$ is the sample lag-k correlation coefficient which is an estimate of the population correlation coefficient $\rho_{k,t}$. When t-k < 1 in Eq. (3.11), N is replaced by N-1, $X_{v,t-k}$ is replaced by $X_{V-1,w+t-k}$, and X_{t-k} is replaced by X_{w+t-k}. The estimates X_t, S^2_t, C_{k-t} and $r_{k,t}$ are generally called the periodic or seasonal statistical characteristics of the series $X_{v,t}$.

3.4.3 Tests for randomness and trend

A time series can be decomposed into a deterministic component, which could be formulated in a manner that allow exact prediction of its value, and a stochastic component, which is always present in the data and cannot strictly be accounted for as it is made by random effect. The time series X (t) can be represented by a decomposition model of the additive type as follow:

$$X(t) = T(t) + P(t) + S(t) \quad \ldots (3.12)$$

where,

T (t) = trend component, t = 1, 2, 3…… N

P (t) = periodic component

S (t) = stochastic component, including dependent and independent parts

A river inflow is treated as a random or stochastic process. The Justification is that river inflow is a function of precipitation and other processes like physiographical characteristics of the river basin as well as of the temporal and spatial distribution of the contributing rainfall. Catchment retention, diurnally and seasonally varying evaporation losses, transpiration from plants and infiltration which depends upon soil characteristics, vegetation and antecedent rainfall are important constituents in any rainfall equation. River inflows are also supplemented to varying extents by ground water that fluctuates sporadically, conveyance properties of water courses as affected by natural erosion and siltation, changes in land use and vegetation, water resource developments and other form of interventions. All these factors in combination make river inflow a highly complex process within the hydrologic cycle.

Also, there is some repetitiveness in climatic patterns, thus river inflow and other hydrological sequences are characterized by variability and oscillatory behavior. The concept of random process can not completely explain all these. However, with the current state of knowledge, this seems to be the most feasible solution.

For detecting the trend, hypotheses of no trend was made and following statistical tests, as suggested by Kottegoda (1980), has performed for randomness of Mahi river inflow sequences.

i. Turning point test
ii. Kendall's rank correlation test
iii. Regression test for linear trend.

These tests are described in the following paragraphs.

3.4.3.1 Turning point test

In an observed sequence X_t, i = 1, 2, 3……… N, a turning point, p, occur at time t = i, if X_i is, either greater than X_{i-1} and X_{i+1} or less than the two adjacent values. The expected number of p in a random series is

$$E\ (p) = \frac{2(N-2)}{3} \qquad \ldots (3.13)$$

Variance (p) can be shown as

$$Var\ (p) = \frac{(16N-29)}{90} \qquad \ldots (3.14)$$

Consequently, Z can be expressed as a standard measure,

$$Z_{cal} = \frac{p - E(p)}{\{Var(p)\}^{1/2}} \qquad \ldots (3.15)$$

which is treated approximately as standard normal deviate. This was compared with its table value, at 5% level of significance to test the trend. If the calculated value of Z is within the limit, then hypothesis of no trend is accepted.

The turning point test also indicates the randomness or non- randomness of time series. If there are too few or too many turning points then series indicates non- randomness otherwise the series is random.

3.4.3.2 Kendall's rank correlation test

If the series is thought to have a trend component, Kendall's rank correlation test can be used to test the significance. This measures the 'disarray' in the data. It is particularly effective if the underlying trend is of a linear type. This test, which is also referred to as the τ test, is based on the proportionate number of subsequent observations which exceed a particular value. For a sequence X_l, X_2,…, X_N, the standard procedure is to determine the number of times, say p in all pairs of observations (X_j, X_i; j > i) that X_j is greater than X_i. The ordered (i, j) subsets are (i = 1, j = 2, 3, 4 … N), (i = 2, j = 3, 4, 5,….. , N), …….., (i = N-1, i = N). The test is based on the static

$$\tau = \frac{4p}{N(N-1)} - 1 \qquad \dots (3.16)$$

$$Var(\tau) = \frac{2(2N+5)}{9N(N-1)} \qquad \dots (3.17)$$

$$Z = \frac{\tau}{(Var(\tau))^{1/2}} \qquad \dots (3.18)$$

The standard measure Z was again compared with its table value at 5 per cent level of significance and hypothesis of no trend is tested. If the calculated value of Z is within its table value, then it can be concluded that the trend is not present in the data, and the sequence is random.

3.4.3.3 Regression test for linear trend

If Kendall's rank correlation test shows a trend, then this type of test can be used if it is thought that the trend is approximately linear. Standard methods of linear regression are used for the purpose. A linear model of the following type is fitted to the sequence of data.

$$X = X_0 + \alpha . t + \xi_t \qquad \text{\`} \qquad \dots (3.19)$$

The hypothesis to be tested in this case is $\alpha \neq 0$. The first step is to estimate α and its variance which is denoted by $S_{\hat{\alpha}}$. The static t = $\hat{\alpha} / S_{\hat{\alpha}}$ is then tested by using student's t test. It is assumed that the residuals ξ_t are stationary, sequentially independent and normally distributed.

If trend is present, it may be removed by differencing, regression, square root transform, log transform or any other methods. After removing the trend a trend free series can be obtained.

3.4.4 Periodic component

The periodic component (as given P (t) in Eq. 3.14) concerns an oscillating movement, which is repetitive over a fixed interval of time (Kottegoda 1980). The existence of P (t) was identified by the time plot of the time series. The oscillating shape of the time series verifies the presence of P (t), with the seasonal period P, at the multiple of which peak of estimation can be made by Fourier analysis followed by the tests for significant harmonics. The time series X (t) is expressed in the Fourier form as under:

$$P(t) = A_O + \sum_{K=1}^{\infty}\left[A_K Cos\left(\frac{2Kt\pi}{p}\right) + B_K Sin\left(\frac{2Kt\pi}{p}\right)\right] \quad \ldots (3.20)$$

where,

$$\mathrm{A}_0 = \frac{1}{\mathrm{N}}\sum_{\mathrm{t}=1}^{\mathrm{N}} \mathrm{x}(\mathrm{t}) \quad \ldots (3.21)$$

$$A_k = \frac{2}{N}\sum_{t=1}^{N} x(t) Cos\left(\frac{2Kt\pi}{p}\right) \quad \ldots (3.22)$$

and

$$B_K = \frac{2}{N}\sum_{t=1}^{N} x(t) Sin\left(\frac{2Kt\pi}{p}\right) \quad \ldots (3.23)$$

where,

K =number of significant harmonics

p =base period

N =number of observation points

A_K and B_K = Fourier coefficients,

When p is even, then

$$A_M = \frac{1}{N}\sum_{i=1}^{N} X_t Cos\left(\frac{2\pi Kt}{P}\right) \qquad \dots (3.24)$$

$$B_M = 0 \qquad \dots (3.25)$$

These coefficients are obtained by a least square fit of the data to the K^{th} harmonics components, and then a least square approximation can be given by the finite series

$$P(t) = A_O + \sum_{K=1}^{M}\left[A_K Cos\left(\frac{2Kt\pi}{p}\right) + B_K Sin\left(\frac{2Kt\pi}{p}\right)\right] \qquad \dots (3.26)$$

where,

M is the number of significant harmonics (maximum, p/2). For later use, it was more convenient to use the alternate form for P (t) given as under:

$$P(t) = A_O + \sum_{t=1}^{M} D_K Cos\left(\frac{2Kt\pi}{p} - \theta_K\right) \qquad \dots (3.27)$$

where,

$$D_K = \sqrt{A_K^2 + B_K^2} \qquad \dots (3.28)$$

and

$$\theta_K = Arc\tan\left(\frac{A_K}{B_K}\right) \qquad \dots (3.29)$$

In Equation (3.26) if $M \to \infty$, P (t) $\to$X (t) than X (t) can be represented satisfactory by Equation (3.20) only. However, it may not be practical or desirable to allow the condition $M \to \infty$. Thus the appropriate approach would be the selection of M which contains only those harmonics which are significantly contributing towards X (t). With this as the objective, following tests were conducted to select an appropriate value of M:

(i) Analysis of variance
(ii) Fourier decomposition of mean square.

(iii) Cumulative periodigram

In general, periodicities can be represented by one or two harmonics in monthly series and by four to six harmonics in weekly series. In such case the other harmonics are treated as noise and are passed on to stochastic component.

3.4.4.1 Test of analysis of variance

If the points τ = 1, 2, 3…, p indicate the time span of periodicity, then periodic function of the periodic mean, m_τ, estimated from the observed time series X_1, X_2, X_3, …, X_n are given by:

$$m_\tau = \frac{1}{n}\sum_{i=1}^{n} X_{\tau+p(i-1)} \quad \text{… (3.30)}$$

where n = N/p is the number of years of data. To estimate the number and coefficients of the significant harmonics, the coefficient α_K and β_K were determined from:

$$\alpha_K = \frac{2}{p}\sum_{r=1}^{p} m_\tau Sin\left(\frac{2Kt\pi}{p}\right) \quad \text{, for k = 1, 2, …., p/2 -1} \quad \text{… (3.31)}$$

$$\alpha_{p/2} = 0 \quad \text{… (3.32)}$$

$$\beta_K = \frac{2}{p}\sum_{r=1}^{p} m_\tau Cos\left(\frac{2Kt\pi}{p}\right) \quad \text{, for k = 1, 2, …., p/2 -1} \quad \text{… (3.33)}$$

and

$$\beta_{p/2} = \frac{1}{p}\sum_{r=1}^{p} m_\tau (-1)^r \quad \text{… (3.34)}$$

These coefficients were tested through the analysis of variance for half the base period in order to obtain the F-ratios. In this analysis, null hypothesis was that the variance explained by a harmonic K, which was $(N/2)(\alpha_K^2 + \beta_K^{2})$, where N is the total sample size, is zero. F ratio was found out by mean squared values divided by unexplained variance. If this values of F ratio is less than its table value at 1% level of significance, the corresponding harmonics was selected otherwise test is to be repeated for higher harmonics until the obtained value of F ratio was less than its table value.

3.4.4.2 Fourier decomposition of mean square

The contribution of the individual harmonics towards the mean square was calculated and the number of harmonics, which were dominantly contributing to mean square, was selected as the significant harmonics.

3.4.4.3 Cumulative periodigram test

A graphical method was employed for selecting the significant harmonics in Fourier series fit of a periodic estimate. The mean squared deviation, MSD, of periodic estimate around the mean of the periodic estimate was determined:

$$\text{MSD(u)} = \frac{1}{P}\sum_{\tau=1}^{p}(u_\tau - \bar{u})^2 \qquad \dots (3.35)$$

Where,

MSD(u) = Mean squared deviation

u_τ = Periodic estimate

$\bar{u}$ = Mean of periodic estimate

$$= \frac{1}{P}\sum_{\tau=1}^{p} u_\tau$$

Mean square deviation MSD (j) of each harmonics j was calculated by the following expression:

$$\text{MSD(j)} = \frac{1}{2}(A_j^2 + B_j^2) \qquad j = 1, 2, \dots, p \qquad \dots (3.36)$$

The cumulative periodigram, P_t was determined by the following equation:

$$P_t = \frac{\sum_{j=1}^{i} MSD(j)}{MSD(u)} \qquad i = 1,2, \dots, p \qquad \dots (3.37)$$

A graph was drawn between P_t and the number of harmonics for selecting the significant harmonics. The significant harmonics were selected up to the fast increase in P_t and the rest of harmonics were rejected. The periodic component was then removed from the time series using the harmonic constants. The remaining component was the stochastic component, which was used for time series modelling.

3.5 Time Series Modelling

For reservoir operation, forecasts of inflows are considered to be an important factor. The forecasts are either based on physical understanding of the process or on statistical analysis of the process popularly termed as the stochastic approach. In the present study the stochastic approach is adopted. Major emphasis is laid on the time series approach. Time series analysis belongs to major statistical techniques used in extraction of information on hydrologic and water resources random variables from the observed data. Once the required information is extracted in the form of statistical parameters, the forecasting of the variables with certain degree of confidence is possible. However, the analysis to extract the information with accuracy is quite a difficult task.

Box and Jenkins (1976) have systematically discussed the time series models. Most of the recent advances in time series analysis are based on the basic work of Box and Jenkins. The time series models of the type are termed as Box Jenkins models. The general model discussed by them and as used in the present study is called as **auto regressive integrated moving average (ARIMA)** model. Major application of time series models in hydrology are in the stochastic generation of sequences or forecasting future events of a hydrological behavior.

Box and Jenkins (1976) and Hipel et. al. (1977) has formalized the modelling process and has described the process to be iterative and composed of 3 major steps,

i. Identification of the model
ii. Determination of the parameters of selected model
iii. Validation of the selected model.

For each step stated above the procedures have been developed and much of the research efforts in stochastic analysis have been put on the betterment of the modelling process to cover the problems occurring in the fields of wide variety.

In the present study the procedure adopted for the time series modelling with univariate models is confined to the descriptions presented by Box and Jenkins (1976), Hipel et al. (1977), Mcleod et al. (1977) and Mujumdar and Kumar (1990). The basic ARIMA model used in the present study in its seasonal form is given by Box and Jenkins, (1976), and Hipel et al. (1977).

Let Z_1, Z_2 … Zn be a discrete time series measured at approximately equal time intervals. An ARIMA model is given as

$$\phi (B)\ \Phi (B^{s})\ (W (t)-\mu) = \theta (B)\ \Theta (B^{S})\ a_t \qquad \text{… (3.38)}$$

where,

t = discrete time step equal to one month.

s = seasonal length equal to 12 month values in a year.

B = backward shift operator defined by B (w (t)) = w (t-1) and B^s (w (t)) = w (t-s).

μ = mean level of the process to be modeled.

a_t = normally independently distributed white noise residual with mean 0 and variance σ^2_a.

ϕ (B) = $1- \phi_1 B - \phi_2 B^2 - ….. - \phi_p B^p$ nonseasonal autoregressive (AR) operator or polynomial of order p and the ϕ_i, i=1,2, …. p is the nonseasonal AR parameters.

$(1-B)^d$ = ∇^d non seasonal differencing operator of order d to produce non seasonal stationarity of the d^{th} differences, usually d=0,1 and 2.

$\Phi (B^s)$ = $\Phi (B^S) = 1 - \Phi_1 B^S - \Phi_2 B^{2S} - …… - \Phi_P (B^{PS})$, seasonal AR operator of order P and Φ_i, i=1,2, ….. P are the seasonal AR parameters.

$(1-B^S)^D$ = ∇s^D seasonal differencing operator of order D to produce seasonal stationarity of the Dth differenced data, usually D=0, 1 and 2 and S corresponds to the seasonal recurrence (usually S=12,24).

w(t) = $\nabla^d \nabla s^D Z_t^{\lambda}$ the stationary series formed by differencing, series (n=N-d-sD) is the number of terms in the w(t) series.

θ (B) = $1-\theta_1 B - \theta_2 B^2 - \ldots\ldots - \theta_q B^q$ non seasonal moving average (MA) operator or polynomial of order q and θ_i, i= 1,2, …. q are the non seasonal MA parameters.

$\Theta (B^S)$ = $1 - \Theta_1 B^S - \Theta_2 B^{2S} - \ldots . - \Theta_Q B^{QS}$ seasonal MA operator of order Q and Θ_i, i=1, 2,….. Q are the seasonal MA parameters.

Z_t^λ = transformed series whose distribution is assumed to be approximately normal. The transformation of original non normal series is carried out to render the data to be normal by Box-Cox transformation.

The model is identified by symbol ARIMA (p, d, q) $(P, D, Q)_s$. Equation (3.38) can be generalized in many ways by considering or omitting the different terms in the models. The resulting models range from simple autoregressive (AR) model to an ARIMA model. And the selection of the terms depend on the inherent properties of the given time series. The ARIMA model is termed as nonstationary model or the seasonal model since it takes into account the seasonal nonstationarity of the data. In the model explained by the Eq. (3.38) if the seasonal terms are absent and the differencing term d is set equal to zero then the model is called as stationary autoregressive moving average (ARMA) model [identified by (p, q)]. Further if the model has only the autoregressive term then the model is called as the AR model or it is called as MA model if only the moving average terms are present in the model. The equation can also model the data which have non contagious parameters that is intermediate parameters are forced to take zero value.

3.5.1 Standardization and normalization of time series variables

The first step in time series modelling is to standardize the time series. The standardization of the series is done by

$$Y_{v,t} = \frac{X_{v,t} - \overline{X}_t}{\sigma_t} \qquad \ldots (3.39)$$

in which $y_{v,t}$ is stationary stochastic component in the mean and variance, $X_{v,t}$ denote the monthly river inflow in the month t of the year v, $\overline{X}_t$ and σ_t denote the monthly means and standard deviations with ranging from 1 to 12.

The standardized series is transformed to ensure normalcy of data sequence and residuals. Many transformations are available in the literature and Box- Cox transformation is most commonly employed. The transformation is also known as power transformation and is given as

$$Z_t^{\lambda} = \frac{[(Y_t + Const)^{\lambda} - 1]}{\lambda} \qquad \lambda \neq 0$$

$$Z_t^{\lambda} = \ln(Y_t + Const) \qquad \lambda = 0 \qquad \ldots (3.40)$$

where, Const is a constant term added to standardized data to make the complex data set to be positive, λ is another constant whose value is varied to get desired transformation. The values of Const and λ are varied till the coefficient of skewness of the transformed Z series becomes near zero. Here after, the analysis is done of the standardized transformed normal time series Z_t of monthly river inflows.

3.6 Identification of Models

The first and the foremost important step in the modelling of a time series is the identification of the tentative model type to be fitted to the data set. It is also to be understood that identification is necessarily inexact. It is the stage at which graphical methods are particularly useful and judgment must be exercised. However, it should be clearly understood that preliminary identification commits only to tentatively entertaining a class of models which will later be efficiently fitted and checked. An engineer is usually confronted with selecting the most suitable model from a large set of possible models for fitting to a given time series. In the present study the procedure stated by Hipel et. al. (1977) is adopted. The requirements of identification procedure are

a) Plot of the original series
b) Plot of standardized series
c) Autocorrelation function (ACF) analysis
d) Partial autocorrelation function (PACF) analysis
e) Inverse autocorrelation function (IACF) analysis
I) Inverse partial autocorrelation function (IPACF) analysis

These are described as under:

a. A visual inspection of a graph of the given observations against time can often reveal autocorrelation, seasonality, nonstationarity, trends and extreme values etc. Thus it can reveal both obvious and also less apparent statistical characteristics of the data.

b. Plot of standardized series can reveal about the stationarity of the series after it has been standardized as using Eq.(3.39).

c. By utilizing Eq. (2.11) the **sample autocorrelation function (ACF)** of a time series can be calculated and then plotted against lag k upto a maximum lag of approximately N/4 where N is the length of the series. If theoretical ACF is assumed to be zero after lag q, Eq. (2.12) is used to calculate the confidence limits. When the sample ACF of a stationary series is plotted along with its appropriate confidence limits upto a lag of about N/4, the following general rules may be invoked to help to determine the orders of p and q in AR; MA, or ARMA model.

   (i) If the series can be modelled by a white noise model, then ACF is not significantly different from zero after lag zero.

   (ii) For a pure MA model, ACF cuts off and is not significantly different from zero after lag q.

   (iii) When ACF damps out and does not appear to truncate, this suggests that AR terms are needed to model the time series.

Failure of the estimated autocorrelation function to die out rapidly suggests that the underlying stochastic process is nonstationary in Z_t but possibly as stationary in ∇Z_t or in some higher difference.

d. The theoretical definition for the **partial autocorrelation function (PACF)** is given by the Yule-Walker equations in Eq. (2.19).

The partial autocorrelation function ϕ_k (k) in an AR process of order k is a measure of the linear association between ρ_j and ρ_{j-k} for $j \leq k$. It is the k^{th} auto regressive coefficient and $\phi_k(k)$ for k=1 ,2, … , is the partial autocorrelation function.

The difference equation for an AR (k) model is

$$\rho_j = \phi_1(k)\,\rho_{j-1} + \phi_2(k)\rho_{j-2} + \ldots. + \phi_k\,(k)\rho_{j-k} \qquad j=1, \ldots ,k \qquad \ldots (3.41)$$

where $\phi_j(k)$ is the j^{th} autoregressive coefficient of the AR (k) model. The partial autocorrelation is given by the last coefficient $\phi_k(k)$, k= 1,2,…. Eq. (3.41) constitutes the set of linear equations.

$$
\begin{array}{llllllllll}
\phi_1(k) & + & \phi_2(k)\,\rho_1 & + \ldots + & \phi_k(k)\,\rho_{k-1} & = & \rho_1 \\
\phi_1(k)\,\rho_1 & + & \phi_2(k)\,\rho_0 & + \ldots + & \phi_k\,\rho_{k-2} & = & \rho_2 \\
\phi_1(k)\,\rho_2 & + & \phi_2(k)\,\rho_1 & + \ldots + & \phi_k\,\rho_{k-3} & = & \rho_3 \\
\ldots & & \ldots & \ldots & \ldots & & \ldots \\
\ldots & & \ldots & \ldots & \ldots & & \ldots \\
\phi_1(k)\,\rho_{k-1} & + & \phi_2(k)\,\rho_{p-2} & + \ldots + & \phi_k(k)\rho_0 & = & \rho_k
\end{array} \quad \ldots (3.42)
$$

which may be written as,

$$
\begin{bmatrix}
1 & \rho_1 & \rho_2 & \ldots & \rho_{k-1} \\
\rho_1 & 1 & \rho_1 & \ldots & \rho_{k-2} \\
\rho_2 & \rho_1 & 1 & \ldots & \rho_{k-3} \\
\ldots & \ldots & \ldots & \ldots & \ldots \\
\ldots & \ldots & \ldots & \ldots & \ldots \\
\rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \ldots & 1
\end{bmatrix}
\begin{bmatrix}
\phi_1(k) \\ \phi_2(k) \\ \phi_3(k) \\ .. \\ .. \\ \phi_k(k)
\end{bmatrix}
=
\begin{bmatrix}
\rho_1 \\ \rho_2 \\ \rho_3 \\ .. \\ .. \\ \rho_k
\end{bmatrix}
$$

… (3.43)

or

$$P_k \ \phi_k = \rho_k$$

$$\phi_k = P_k^{-1}\,\rho_k\,,\ k = 1,2,\ \ldots \qquad \ldots (3.44)$$

Thus the partial autocorrelation function ϕ_k (k) is determined by successively applying Eq, (3.44).

On the hypothesis that the process is AR (p) the estimated $\phi_k(k)$ for $k > p$ is asympotically normal with mean zero and variance 1/N.

$$\text{Var}\,[\phi_k(k)] = 1/N$$

Thus, the standard error (S.E.) of the estimated partial autocorrelation $\phi_k(k)$ is

$$\text{S.E.}(\phi_k(k)) = \sigma[\phi_k(k)] = 1/\sqrt{N} \qquad \ldots (3.45)$$

Hence 1- α probability limits for zero partial autocorrelation may be determined by

$$[-U_{1-\alpha/2} / \sqrt{N}, + U_{1-\alpha/2} / \sqrt{N}] \qquad \ldots (3.46)$$

where $U_{1-\alpha/2}$ is the 1 - α /2 quantile of the standard normal distribution, N is the sample size and α is the probability limit. The limits of Eq.(3.45.) are used to give some guide as to whether theoretical partial autocorrelations are particularly zero beyond a particular lag. Thus 95 percent confidence limits are calculated and plotted 1.96 times the S.E. for $\phi_{k,k}$ above and below the horizontal axis.

When used in conjuction with an identification aid such as a plot of the sample ACF, the estimated PACF is useful for determining the number of AR and MA parameters. The following general characteristics of the PACF may be of assistance in model identification.

i. When the series is white noise, the estimated values of the PACF are not significantly different from zero for all lags.
ii. For a pure AR model, the sample PACF truncates and is not significantly different zero from zero after lag p.
iii. If the sample PACF attenuates and does not appear to cut off, this may indicate that MA parameters are needed in the model.

(e) **The Inverse autocorrelation function** (IACF) of a time series is defined as the ACF associated with the reciprocal of the spectral density function of the series. The theoretical IACF, r_{ik} can also be specified in an alternative equivalent fashion within the time domain. When considering the ARIMA (p, d, q) Process, the theoretical IACF is defined to be the ACF of the (q, d, p) process. The estimate r_{ik} for the theoretical IACF at lag k can be obtained from

$$r_{ik} = [-\hat{\phi}_k + \sum_{i=1}^{r-k} \hat{\phi}_i \hat{\phi}_{i+k}][1 + \sum_{i=1}^{r} \hat{\phi}_i^{\,2}]^{-1} \qquad \ldots (3.47)$$

To utilize the IACF for model identification calculate and plot r_{ik} versus lag k, where r_{ik} can go from -1 to + 1. From knowledge of the distribution of r_{ik}, confidence limits can be drawn on the graph of the sample IACF. The variance of r_{ik} after lag p is given by

$$\text{Var}(r_{ik}) = \frac{1}{N}[1 + 2\sum_{j=1}^{p} r_{ij}^{\ 2}] \quad , k > p \qquad \ldots (3.48)$$

When using the sample IACF for model identification to ascertain the order p and q, the following rules may be utilized.

i. If the series can be modelled by a white noise model r_{ik} is not significantly different from zero after lag zero.

ii. For a pure AR model, r_{ik} truncates and is not significantly different from zero after lag p. In practice it has been found that the IACF is useful for identifying AR models where some of the AR parameters should be constrained to zero. at the same lags at which the AR parameters are zero, the sample IACF often possesses values that are not significantly different from zero,

iii. When r_{ik} attenuates and does not appear to cutoff, this indicates that MA terms are needed to model the time series.

The sample IACF along with other identification graphs are very helpful when employed together for identifying ARMA models. The sample ACF and IACF are recommended for model identification rather than the sample ACF and PACF.

(f) Hipel et al., (1977) provided the original definition of the **Inverse partial autocorrelation function** (IPACF) as the PACF of an ARMA (q, p) process. To define mathematically the theoretical lPACF, consider the inverse Yule-Walker equations given by

$$\begin{bmatrix} 1 & r_{i1} & r_{i2} & \ldots & r_{i,k-1} \\ r_{i1} & 1 & r_{i1} & \ldots & r_{i,k-2} \\ \ldots & \ldots & \ldots & & \ldots \\ \ldots & \ldots & \ldots & & \ldots \\ r_{i,k-1} & r_{i,k-2} & r_{i,k-3} & \ldots & 1 \end{bmatrix} \begin{bmatrix} \theta_{k1} \\ \theta_{k2} \\ \ldots \\ \ldots \\ \theta_{kk} \end{bmatrix} = \begin{bmatrix} r_{i1} \\ r_{i2} \\ \ldots \\ \ldots \\ r_{ik} \end{bmatrix} \ldots (3.49)$$

where r_{ik} is the theoretical IACF at laq k and θ_{kj} is the j^{th} coefficient in a MA process of order k such that θ_{kk} is the last coefficient.

The coefficient θ_{kk} is called the theoretical IPACF. To obtain an estimate of θ_{kk} solve the inverse Yule-Walker equation for θ_{kk}.

For model identification θ_{kk} is plotted against lag k for the same number of lags as were chosen for the sample IACF. The values of the sample IPACF can range from -1 to +1. When employing the plot of the IPACF for model identification, the following properties can be kept in mind

i. When the time series is white noise, the sample IPACF is not significantly different from zero after lag zero.
ii. For a pure MA model, θ_{kk} cuts off and is not significantly different from zero after lag q.
iii. When the sample IPACF damps out and does not appear to truncate, this suggests that AR terms are needed in order to suitably model the series.

The inherent characteristics of the sample IPACF are similar to those of the sample ACF. Even though this pair of functions possesses the same general properties for identifying an ARMA model to fit to a series, the two functions are defined differently. In a given situation, for instance, one identification function may more clearly reveal characteristics of the data than the other. Consequently both IPACF and ACF are recommended for application to the series under consideration. Likewise, common relationships also exist between PACF and IACF, and both of these functions should also be used in the application. The general attributes of all these useful identification functions are summarized in Table 3.2.

Table 3.2 Properties of Four Identification methods

Identification method	Type of model					
	Non seasonal			Seasonal		
	AR(p)	MA(q)	ARMA(p,q)	AR(p)	MA(q)	ARMA(p,q)
ACF	Attenuates or Tail off	Truncates after lag p	Attenuates or Tail off	Attenuates at every seasonal lag	Truncates at Q+SQ	Attenuates
PACF	Truncates after lag p	Attenuates or Tail off	Attenuates or Tail off	Truncates at P+SP	Attenuates at every seasonal lag	Attenuates
IACF	Truncates after lag p	Attenuates or Tail off	Attenuates or Tail off	Truncates at P+SP	Attenuates at every seasonal lag	Attenuates
IPACF	Attenuates or Tail off	Truncates after lag p	Attenuates or Tail off	Attenuates at every seasonal lag	Truncates at Q+SQ	Attenuates

3.7 Estimation of Parameters

After the identification of the model, the parameters of the selected models were estimated. The parameters of the identified models have been estimated by the statistical analysis of the data series. The most popular of the approaches of the parameter estimation are:

i. The method of moments
ii. The method of least-squares
iii. The method of maximum likelihood

The last method is most exact method of parameter estimation, since it utilizes full information from the data set. Box and Jenkins (1976) have discussed the methods in detail. The recent research in the methods is also summarized by Hipel and McLeod (1994). The method used in the present study is the conditional least square estimation method as explained by Box and Jenkins (1976). The parameters are, in general, estimated in two steps: a preliminary or initial estimate and exact estimate.

3.7.1 Initial estimates

The Yule-Walker equation given in section 2.2.2.2 provides simple initial estimates for the parameters of the autoregressive processes. For the MA process the equivalent of the Yule-Walker equation is given as

$$\gamma_k = \mathrm{Cov}\,[Z_t Z_{t-k}] = \sigma_\varepsilon^2 \sum_{j=0}^{q-k} \theta_j \theta_{j+k} \qquad , k \le q$$

$$= 0 \qquad , k > q \qquad \ldots (3.50)$$

For k=0, the variance is:

$$\mathrm{Var}[Z_t] = \gamma_0 = \sigma_\varepsilon^2 \sum_{j=0}^{q} \theta_j^{\,2} \qquad \ldots (3.51)$$

Eq. (3.50) and (3.51) may be solved iteratively for the θ parameters; however, their statistical efficiency is less than that of the Yule-Walker equations for the autoregressive models. Eq. (3.51) and (3.50), are rewritten in the form

$$\hat{\sigma}_{\varepsilon}^{2} = \frac{C_0}{1+\hat{\theta}_1^{2}+....+\hat{\theta}_q^{2}} \qquad \ldots (3.52)$$

and $$\hat{\theta}_j = -(\frac{C_j}{\hat{\sigma}^2} - \hat{\theta}_1 \hat{\theta}_{j+1} - \hat{\theta}_2 \hat{\theta}_{j+2} - \ldots\ldots - \hat{\theta}_{q-j} \hat{\theta}_q) \qquad \ldots (3.53)$$

where C_0 and C_j are the estimators of the variance and autocovariance respectively, and the ^ indicates estimates.

The unknown $\hat{\theta}$'s are assumed to be zero in the first iteration, and improved values of $\hat{\sigma}_{\varepsilon}^{2}$ and $\hat{\theta}_j$ are obtained successively. For the MA (1) process these equations become

$$\hat{\sigma}_{\varepsilon}^{2} = \frac{C_0}{1+\hat{\theta}_1^{2}}$$

$$\hat{\theta}_j = (\frac{C_1}{\hat{\sigma}^2})$$

For the ARMA process the p autoregressive parameters $\hat{\phi}_1$, $\hat{\phi}_2$... $\hat{\phi}_p$ are estimated first. The autocovariance are independent of the MA parameters. A new series is now constructed which is the difference between the original series and the one formed by the AR model constructed with the parameters $\hat{\phi}_1$, $\hat{\phi}_2$... $\hat{\phi}_p$, namely.

$$Z'_t = Z_t - \hat{\phi}_1 Z_{t-1} - \ldots - \hat{\phi}_p Z_{t-p} \qquad \ldots (3.54)$$

This series presumably contains only the MA portion of the process. Its Autocovariance estimates C'_0, C'_1... C'_q are calculated and the parameters θ_1, θ_2, ..., θ_q are estimated by means of Eq. (3.52) and (3.53) applied iteratively.

3.7.2 Exact estimates

Having obtained preliminary estimates of parameters of the tentative model, efficient estimates of the parameters are needed that take into account all the information contained in the data. The maximum likelihood estimates satisfy this requirement. Box and Jenkins, (1976) showed that the maximum likelihood estimates are essentially the same as the least squares estimates if the ε's given by

$$\varepsilon_t = Z_t - \phi_1 Z_{t-1} - \phi_2 Z_{t-2} - \dots \phi_p Z_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} \qquad \dots (3.55)$$

are normally distributed. Therefore the study was concerned with the evaluation of the sum of the squares of the residuals

$$S(\phi,\theta) = \sum_{t=1}^{N} (\varepsilon_t)^2 \qquad \dots (3.56)$$

The sum of the square of the residuals is understood to depend on the parameters ϕ and θ, the Z_t series and the starting values of the ε's. Therefore the set of parameters ϕ and θ are found which minimizes the sum-of-squares function. The variance of the residuals is then estimated by

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{N} S(\hat{\phi},\hat{\theta}) \qquad \dots (3.57)$$

A steepest descent algorithm is used to obtain the maximum likelihood estimate of the parameters and the residuals. For example, for the MA (1) model

$$Z_t = \varepsilon_t - \theta_1 \varepsilon_{t-1}$$

The residuals are given by

$$\varepsilon_t = Z_t + \theta_1 \varepsilon_{t-1}$$

and the sum of the squares of the residuals is thus,

$$S(\theta) = \sum_{t=1}^{N} (\varepsilon_t)^2 = \sum_{t=1}^{N} (Z_t + \theta_1 \varepsilon_{t-1})^2$$

To calculate the sum a starting value of ε_0 may be taken as zero, which is its mean value.

Thus

$\varepsilon_1 = Z_1$

$\varepsilon_2 = Z_2 + \theta_1 \varepsilon_1 \quad = \quad Z_2 + \theta_1 Z_1$

$\varepsilon_3 = Z_3 + \theta_1 \varepsilon_2 \quad = \quad Z_3 + \theta_1 Z_2 + \theta_1^2 Z_1$

etc.....

The residuals are seen to be nonlinear functions of the parameters. Since the MA model was constrained to be invertible, namely $| \theta_1 | < 1$, the above series converges and depends on the past observation and on θ_1. By the use of maximum likelihood estimate the exact estimation of parameter is done.

3.8 Validation of Models

The assumption used in building the stochastic models of Mahi river were:

i. The residual series a_t has zero mean.
ii. The residual series is independent to each other.

For Validation purposes following tests are carried out to examine whether the assumption used in building the model are in fact valid for the model selected:

1) ACF of residual series
2) Significance of residual mean
3) Box- Pierce Portemanteau lack of fit test

All the validation tests are carried out on the residual series only. Residual series is constructed by following equation

$$a_{(t)} = Z_{(t)} - \sum_{j=1}^{p} \phi_j \cdot Z_{(t-1)} + \sum_{j=1}^{q} \theta_j \cdot a_{(t-1)} \qquad \dots (3.58)$$

3.8.1 ACF of residual series

The autocorrelation function of residuals are calculated with their confidence limits. Autocorrelogram of residuals, a_t of the models are constructed for judging the mutual dependency. If the autocorrelogram of residuals, a_t is within the corresponding limits the residual obtained from the model are mutually independent as they are not significantly different from zero.

3.8.2 Significance of residual mean

The purpose of this test is to examine the validity of the assumption that the series {w (t)} has zero mean. For this purpose a statistic, $\eta(w)$, is defined as:

$$\eta\,(w) \quad = \quad N^{1/2}\ \overline{w}\ /\ \hat{\rho}^{\,1/2} \qquad \ldots (3.59)$$

where $\overline{w}$ is the estimate of the residual mean; and

$\hat{\rho}$ is the estimate of the residual variance.

The statistic, η (w), is approximately distribution as t (α, N-1), where α is the significance level at which the test is being carried out. If the value of η (w) $\leq$ t (α, N-1), then the mean of the residual series is not significantly different from zero and hence the series passes the test.

3.8.3 The Box- Pierce Portemanteau lack of fit test

Consider that a time series x_t of size N is represented by an ARIMA (p, d, q) model. Assume that after d differences the ARMA (p, q) series Z_t, t=1, …., N-d is obtained and assume further that in such models ε_t is the residual series. The statistics for the Portemanteau lack of fit test is

$$Q = (N-d)\sum_{k=1}^{L} {r_k}^2(\varepsilon) \qquad \ldots (3.60)$$

where $r_k(\varepsilon)$ is the correlogram of the residuals ε_t and L is the maximum lag considered. The static Q is approximately Chi-square distribution with L-p-q degree of freedom. The value of L is taken to be order of 10 – 30 percent of the sample size N. The test was carried out for different values of L. The adequacy of the ARMA model for x_t or of the ARMA model for Z_t may be checked by comparing χ^2 (L-p-q) of a given significance level. If Q < χ^2 (L-p-q), ε_t is an independent series and so that models are inadequate.

3.9 Regeneration of Inflows

Having fitted the models to actual data, they are used to regeneration of the observed time series. It is kept in mind that estimation errors in the parameters are small and they will not seriously affect the regeneration. In the present study minimum mean square error regeneration were done directly from the following equation

$$Z_{t+L} = \phi_1 Z_{t+L-1} + \phi_2 Z_{t+L-2} + ... \phi_p Z_{t+L-p} + a_{t+L} - \theta_1 a_{t+L-1} - \theta_2 a_{t+L-2} - ..... - \theta_q a_{t+L-q} \quad ... (3.61)$$

By taking the conditional expectations, designated here by a square bracket,

$$Z_t(L) = [Z_{t+L}] = \phi_1 [Z_{t+L-1}] + \phi_2 [Z_{t+L-2}] + ... \phi_p [Z_{t+L-p}] + a_{t+L} - \theta_1 a_{t+L-1} - \theta_2 a_{t+L-2} - ..... - \theta_q a_{t+L-q} \quad ... (3.62)$$

The forecasting function is then obtained by noting that the present and past values (with a subscript equal to or less than t) have occurred and no longer random. Thus, for past values the conditional expectation is the value itself, i.e. $[Z_{t-j}] = Z_{t-j}$, j = 0, 1, 2 … For the future values the conditional expectation is the forecast, i.e. $[Z_{t+j}] = Z_{t+j}$, j = 0, 1, 2 … and likewise for the past values of the random terms $[a_{t-j}] = a_{t-j}$ and the conditional expectation of their future values is zero, i.e. $[a_{t+j}] = 0$, for j = 1, 2…

To calculate the conditional expectations which occur in the Eq. (3.62) we note that if j is a nonnegative integer, then.

$$[Z_{t-j}] = E_t [Z_{t-j}] = Z_{t-j}, \qquad j = 0, 1, 2 ...$$

$$[E_{t+j}] = E_t [Z_{t+j}] = Z_t^j, \qquad j = 1, 2, 3 ...$$

$$[a_{t-j}] = E_t [a_{t-j}] = a_{t-j} = Z_{t-j} - Z_{t-j-1} \qquad j = 0, l, 2 ...$$

$$[a_{t+j}] = E_t [a_{t+j}] = 0 \qquad j = 1, 2, 3 ...$$

Therefore to obtain the regenerated Z_t (L), the terms on the right hand side of Eq. (3.62) are treated according to the following rules

The Z_{t-j} (j = 0, 1, 2…) which have already happened at origin t, are left unchanged.
The Z_{t+j} (j = 1, 2…), which have not yet happened, are replaced by their regenerated Z_t^λ at origin t.
The a_{t-j} (j =0, 1, 2…) which have happened, are available from

$$Z_{t-j} - \hat{Z}_{t-j-1}$$

The a_{t+j} (j =0, 1, 2…) which have not yet happened, are replaced by zeroes.

The generation process is started off initially by setting unknown a_t equal to their unconditional expected values of zero. It is noted that the regeneration were obtained on transformed standardized normal inflow series of Mahi river. Therefore the actual regeneration is obtained by applying inverse Box-Cox transformation to the generated values obtained. The values now obtained are in standardized form and they are converted back in original units of inflow in TMC by multiplying by standard deviation of that month and adding the mean flow of that month and then add the periodic component. The actual and generated monthly Mahi river inflows were plotted together to compare the generated values with the actual observations.

3.10.1 Evaluation of regeneration performance

The generated data were used to assess the regeneration performance of the models. Results of the stochastic models were evaluated to assess the regeneration performance. The models were evaluated with regard to several errors such as mean forecast error, mean absolute error, root mean square error and integral square error.

3.10.1.1 Mean forecast error

Raghuwanshi et al. (2000) used mean forecast error to evaluate the performance of the time series models of daily evapotranspiration. The mean forecast error (MFE) was computed for monthly runoff series by using the following equation:

$$\mathrm{MFE} = \frac{\sum_{i=1}^{N} X_c(t) - \sum_{i=1}^{N} X_0(t)}{N} \qquad \ldots (3.63)$$

Where,

X_c (t)	=	computed runoff values
X_0 (t)	=	observed runoff values
N	=	number of observations

3.10.1.2 Mean absolute error

Raghuvanshi et al. (2000) evaluated the performance of autoregressive and autoregressive moving average models by mean absolute error. It was computed by using the following equation:

$$\text{MAE} = \frac{\sum_{i=1}^{N} \left| X_c(t) - X_0(t) \right|}{N} \qquad \ldots (3.64)$$

3.10.1.3 Root mean square error

The root mean square was used for evaluation of the performance of the models. The root mean square error was computed by following equation:

$$\text{RMSE} = \left[\frac{\sum_{i=1}^{N} \left[X_c(t) - X_0(t) \right]^2}{N} \right]^{1/2} \qquad \ldots (3.65)$$

3.10.1.4 Integral square error

Singh et al. (1991) used the integral square error (ISE) as a measure of goodness of fit of a time series model for air temperature. The integral square error was computed using the equation:

$$\text{ISE} = \frac{\sqrt{\sum_{i=1}^{N} \left[X_c(t) - X_0(t) \right]^2}}{\sum_{i=1}^{n} X_0(t)} \qquad \ldots (3.66)$$

3.11 Comparison of Different Models

Having generated the inflows by all the selected models fitted to the Mahi river inflow data, the basic statistical characteristics such as mean, standard deviation, variance, coefficient of variation, coefficient of skewness, kurtosis and lag one serial correlation coefficients are determined for all the generated series and they are compared with the basic statistical characteristics of the actual series.

3.12 Selection of most appropriate model

For selection of the most appropriate model for Mahi river inflow forecasting, first the inherent process of model building is considered which involve identification, estimation and validation of different models. A popular decision rule for selection of appropriate model in the time series literature is the Akaike Information Criteria (AIC). However, investigations, both theoretical as well as numerical, have indicated flaws in the AIC rule (Mujumdar, 1990). Firstly, the AIC has no optimal property, i.e. it does not minimize the average value of any criterion function. Secondly, the AIC rule is not consistent, i.e. the probability that the decision rule will choose a wrong model does not go to zero even when the number of observations tends to infinity.

Rao et. al. (1982) has given a rule that is consistent. They use the criterion of **Minimum mean square error (MMSE).** This criterion is used to select the most appropriate model in the present study. Firstly the mean square error (MSE) is estimated for each model, which is given as

$$\text{MSE} = \frac{\sum_{i=1}^{N} e(i)^2}{N} \qquad \ldots (3.67)$$

where e(i) is the error at time t given as e(i) = Z_t - $\hat{Z}_t$, in which Z_t and $\hat{Z}_t$ are the original and generated inflow in TMC at time t. Then select the model that results in the least value of the MSE. That model is the selected as the most appropriate model.

3.12 Forecasting of Inflow

The most appropriate model is used for forecasting of monthly inflow values, for the year 2002 and year 2003. Forecasting performance of the model has evaluated by graphical comparison of historical and forecasted series.

For the analysis some computer program were developed in Fortran-77 language, which are given in Appendices. The help of SPSS software has also taken for the present research works.

4. RESULTS AND DISCUSSION

The objective of the present study was to formulate an appropriate mathematical model, based on stochastic concept, which may be used to describe the time dependence structure of monthly inflow series of Mahi river of Southern Rajasthan. The inflow data for 76 years i.e. from 1928 to 2003 for Mahi river at Mahi Bajaj Sagar Dam site has been collected from the office of the Chief Engineer, Mahi Bajaj Sagar Project, Banswara (Appendix A-1).There is no obstruction to inflow at Dam site and hence data are Virgin and as such can be used for model building. River Mahi river is a monsoonal river, as the inflow of Mahi river during seven months of dry season (from November to May) is less than 5 per cent of annual inflow. The total inflow is occurring only during wet season (from June to October). It was assumed in this study that flow during **dry** season is **Zero** and hence only five month series is used for analysis and model building. The annual inflow has been tested for its randomness. Fourier analysis has done to detect periodic component. After removal of periodic component stochastic models has been fitted to the monthly inflow series of Mahi river and regeneration of data was done for first 74 years data series. From best selected model forecasting has been done for remaining two years i.e. 75th and 76th year. In this chapter the results of the study are discussed under the following heads:

4.1 Statistical Characteristics of Monthly Inflow Series of Mahi River.

Statistical analysis of data is the first step for its mathematical modelling. Sample mean($\bar{x}$), sample standard deviation(s), variance (s^2), coefficient of variation (C.V.), sample skewness coefficient (C_s), kurtosis (C_k) and lag-one serial correlation coefficient (r_1) are the statistical characteristics used for this purpose. Table 4.1 depicts these characteristics of Mahi river monthly inflow series from 1928 to 2003 (76 years).

Mean

Mean monthly Mahi river inflow varies from 110.15 thousand M cum. (TMC) in the month of June to 1028.02 TMC in the month of August. Figure 4.1 shows that the maximum inflow occurs during July, August, and September. The mean annual inflow is 2724.84 TMC.

Standard deviation

Table 4.1 shows that sample standard deviation, s, varies from 97.58 TMC in the month of October to 760.42 TMC in the month of September. Accordingly, the variance s^2 is lowest as 9522.11 in the month of October to highest 578236.80 in the month of September. Whereas the mean monthly inflow is highest as 1028.02 TMC in the month of August followed by mean monthly inflow of 761.69 TMC in the month of September, whereas the standard deviation is higher (760.42 TMC) in the month of September as compared to 746.76 TMC in the month of August. As evident from Figure 4.1 the range (mean ± standard deviation) of Mahi river inflow is highest for the month of September.

Table 4.1 Statistical characteristics of monthly inflow series of Mahi river

Month	Mean $\overline{X}$ TMC	Standard deviation s TMC	Variance s^2	Coefficient of Variation CV	Coefficient of Skewness C_s	Kurtosis	Lag one serial correlation coefficient r_1
June	110.15	141.06	19897.88	1.28	2.35	7.26	0.300
July	707.67	549.78	302261.66	0.78	0.67	-0.55	0.014
August	1028.02	746.76	557656.31	0.73	1.21	1.15	0.235
September	761.69	760.42	578236.80	1.00	1.53	2.80	0.052
October	117.30	97.58	9522.11	0.83	1.92	7.56	-0.093
Seasonal	2724.84	1527.86	2334347	0.56	0.95	1.23	0.155

Coefficient of variation

The **coefficient of variation** C.V. ranges from 0.73 to 1.28, which signifies the importance of time variability of the monthly inflow series and along with its exact prediction. This justifies the treatment of the Mahi river inflow series as a stochastic or a random series. Coefficient of variation is highest as 1.28 in month of June as compared to lowest in the month of August. Although the mean monthly inflow is higher (1028.02 TMC) in the month of August as compared to 761.69 TMC in the month of September, the standard deviation (dispersion or the spread of the series around the mean) and coefficient of variation is very high (760.42 TMC and 1.00 respectively) in the month of September as compared to the standard deviation of 746.76 TMC and coefficient of variation of 0.73 in the month of August. For the mean annual inflow of 2724.84 TMC of Mahi river the annual standard deviation is 1527.86 TMC and annual coefficient of variation is 0.56.

Coefficient of skewness

The monthly and annual data is highly skewed to the right as all values of **Coefficient of skewness** are positive and are very high. The highest coefficient of Skewness (2.35) is observed in the month of June and the next lower value (1.92) in the month of October. This again indicates that the river Mahi is a monsoonal river. Coefficient of Skewness of annual series is 0.95.

Kurtosis

The **Kurtosis** of monthly inflow is negative (-0.55) or platykurtic distribution for the month of July and highly positive (7.56) or Leptokurtic for the month of October and 7.26 for the month of June.

Serial correlation coefficient

The **lag-one serial correlation coefficient**, r_1 is negative for the months of October (-0.093) and positive for the remaining months (varies from 0.014 in the month of July to 0.300 in the month of June.)

4.2 Annual Data; Tests for Randomness and Trend

The first step in data analysis is to construct a sequence or segment of a time series from the known observations, called a trace or a time plot. This trace gives a visual indication of the significance and the relative strength of the inherent behavioral patterns. Although, simulation of annual river inflow is not a common practice, however, to generate monthly river inflow volumes, the year-to-year structure between annual inflow volumes should also be preserved. It is advisable to use only annual data for the analysis of **trend** (Kottegoda, 1980). In this way the periodic component of the river inflow is suppressed, and it will not matter if the stochastic component of the inflow has a highly non-normal distribution. The trend in a hydrological time series, if it appears, is in effect part of a low-frequency oscillatory movement induced by climatic factors or through changes in land use and catchments characteristics. The annual inflow data of Mahi river inflow from 1928 to 2003 are shown in Figure 4.2 and also given in appendix A-1. By inspection of Figure 4.2 there seems to be some type of irregular oscillation which occurs once every 3, 4 years, as seen from the peaks and troughs in Figure 4.2. Therefore, a 3-year and 5-year centered moving average $[(X_{t-1} + X_t + X_{t+1})/3$ and $(X_{t-2} + X_{t-1} + X_t + X_{t+1} + X_{t+2})/5]$ has taken and plotted on the same Figure. Although, the range between maximum and minimum values of annual inflow is reduced and the number of oscillations is also much less, however, no rising or falling trend of annual inflow is visible from the Figure 4.2. Hence, it is doubtful whether any suspected systematic effect is significant or not.

For this reason the following statistical tests were performed on annual inflow data for randomness and detecting the trend.

4.2.1 Turning point test for randomness of the series

The annual values X_i of Mahi river inflow series was used for the test. There were 47 turning points in the sequence. Based on this value of p, the test statistic (Z_{cal}) was estimated as given in section 3.4.3.1. Calculation for turning point and test statistic are given in Appendix B-1 and B-2. The estimated value of test statistics was within the acceptable range at 1 per cent level of significance. Hence the (null) hypothesis of no dependence in the series is not rejected at 1 per cent level of significance.

4.2.2 Kendall's rank correlation test

The annual inflow data of Mahi river, for Kendall's rank correlation test, started with 3495.44 as the first base number, and the succeeding number was examined, and a score of +1 is made each time the base number is exceeded. Kendall's point was obtained for this base number. This is repeated 74 times for each of the other base numbers. Test statistic (Z_{cal}) was estimated according to section 3.4.3.2. The results are shown in Appendix B-3 and B-4. The estimated value of test statistics was within the acceptable range at 1 per cent level of significance, which again shows the absence of trend component in the data.

4.2.3 Regression test for linear trend

The regression test for linear trend shows that at 5 per cent level of significance α is not significant. Hence, the null hypothesis $\alpha = 0$ is not rejected and the series is trend free. Calculations are shown separately in Appendix B-5.

From the above analysis of annual data by turning point test, Kendall's rank correlation test and regression test it is confirmed that the annual inflow series of Mahi river is a random series and trend component in the inflow series is absent and the observed series maybe treated as random and trend free series.

4.3 Periodic Component

Time series consists of mainly three components i.e. trend component, periodic component and stochastic component. As trend component is absent in the time series, stochastic component can be obtained by removal of periodic component. Periodic time series is usually not stationary which can be expanded into a Fourier series representation using equation (3.26). For representing the periodic component of monthly inflow series, the number of harmonics significantly contributing to periodicity was identified through the following three different test approaches.

4.3.1 Analysis of variance

In practice periodicities can be represented by one or two harmonics in monthly series and by four to six harmonics in daily series (Kottegoda, 1980). The actual number to be fitted can be found through an analysis of variance. Using the periodic means estimated from observed monthly inflow series, parameters α and β were determined using equations (3.31) to (3.34) for half the base period of series.

Table 4.2 shows the estimate of the parameter along with amplitude and explained variance for monthly inflow series.

Table 4.2 Fourier series coefficients for monthly Mahi river inflow series (1928 - 2003)

Order	Alpha	Beta	Amplitude	Explained variance	Cumulative explained variance
1	-323.18	-407.00	519.70	79.59	79.59
2	-31.32	-20.67	37.53	0.42	80.00
3	0.00	-23.46	23.46	0.16	80.16

The analysis of variance is given in Table 4.3. α and β parameters were evaluated to obtain the F ratio. The harmonics for which F ratio were greater than at least table value of F at 0.05 level of significance has been considered as significant harmonics. Only first harmonic was found to contribute significantly to periodicity.

Table 4.3 Analysis of variance of monthly Mahi river inflow series (1928 - 2003)

<table>
<tr><th rowspan="2">S. No.</th><th rowspan="2">Harmonic</th><th rowspan="2">Variation</th><th rowspan="2">Degree of freedom</th><th rowspan="2">Sum of Squares</th><th rowspan="2">Mean Squares</th><th rowspan="2">F_{cal}</th><th colspan="2">F_{tab}</th></tr>
<tr><th>0.01</th><th>0.05</th></tr>
<tr><td>1</td><td>2,3</td><td rowspan="2">α,β</td><td>3</td><td>372116.11</td><td>124038.7</td><td rowspan="2">0.28</td><td rowspan="2">3.78</td><td rowspan="2">2.60</td></tr>
<tr><td>2</td><td>Residual</td><td>370</td><td>161280552</td><td>435893.4</td></tr>
<tr><td>3</td><td>1</td><td rowspan="2">α,β</td><td>2</td><td>51317023.3</td><td>25658512</td><td rowspan="2">87.67</td><td rowspan="2">4.60</td><td rowspan="2">2.99</td></tr>
<tr><td>4</td><td>Residual</td><td>377</td><td>110335645</td><td>292667.5</td></tr>
<tr><td></td><td>Total</td><td></td><td>379</td><td>161652668</td><td>426524.2</td></tr>
</table>

4.3.2 Fourier decomposition of mean square

The numbers of significant harmonics that represent the periodic component were estimated by evaluating the Fourier series coefficient A_k and B_k. The Fourier series coefficient A_k and B_k along with amplitude and phase angle (θ) for the corresponding harmonic were computed using Eq. (3.22), (3.23), (3.24), (3.25), (3.28), and (3.29). The contribution of the individual harmonic towards the mean square has been shown under the explained variance and those harmonics dominantly contributing to mean square are selected as the significant harmonics. The results are shown in Table 4.4 monthly inflow series. The results from Table 4.4 indicated that first harmonic contributed 31.66 per cent of total variance caused by periodic components while only 0.22 per cent contributed by the rest harmonics.

Table 4.4 Fourier decomposition of periodic component in monthly Mahi river inflow series (1928 - 2003)

Order	A_k	B_k	Amplitude	Theta	Explained variance	Cumulative explained variance
1	-406.99	-323.18	519.70	0.67	31.66	31.66
2	-20.67	-31.32	37.53	0.99	0.17	31.83
3	-20.67	0.00	20.67	0.00	0.05	31.88

4.3.3 Cumulative periodigram test

In this test mean square deviation values and Pi values were calculated from Eq. (3.35) to (3.37) and shown in Table 4.5 for monthly Mahi river inflow series. A cumulative periodigram of Pi values versus order i were constructed and shown in Figure 4.3. From the cumulative periodigram it can be observed that first harmonic appeared to be the part of fast increase and after that periodigram remains almost constant which may be treated as non-significant.

Table 4.5 Cumulative periodigram of monthly Mahi river inflow series (1928 - 2003)

Order	MSD(u)	MSD(j)	Cumulative MSD (j)	Pi
1	135748.85	135044.83	135044.83	1.00
2	135748.85	704.07	135748.89	1.00
3	135748.85	213.67	135962.56	1.00

These three criteria were found to be consistent, so first harmonic is treated as significantly contributing towards periodicity and remaining are as white noise. The Fourier series coefficient, A_k and B_k, were substituted in the Eq. (3.26) and deterministic periodic

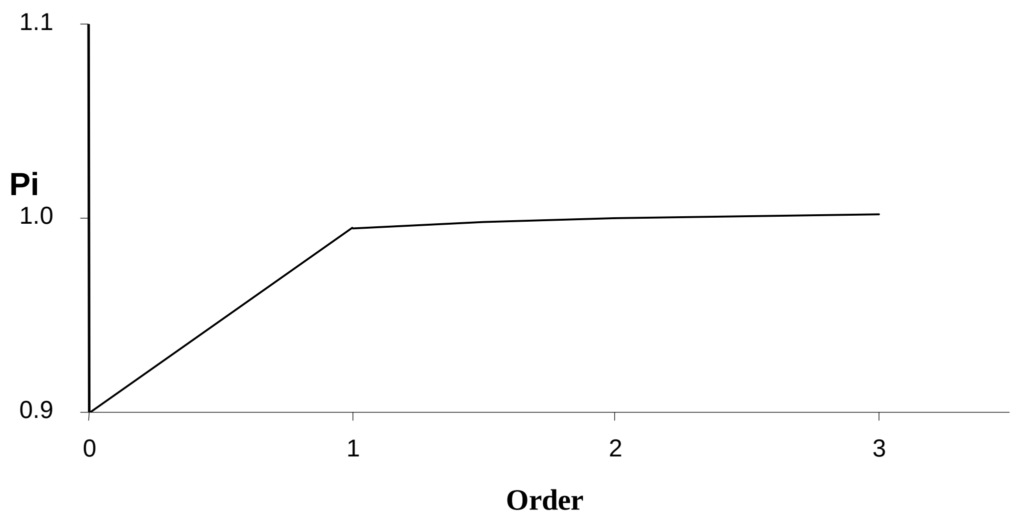

Figure 4.3 Cumulative periodigram of observed monthly Mahi river inflow series (1928-2003)

component, P_t have been computed for all values of t, where t is total time period, which is 380 for monthly inflow series. After determining the periodic component it was then removed by deducting from the observed time series. The remaining series is a stochastic component part, which is required to be fitted by an autoregressive model of suitable order.

4.4 Modelling of Mahi River Inflows

The overall process of model construction of Mahi river inflow consists of identification of models, estimation of parameters of selected models and validation of the selected models. Most probability theory and statistical techniques applied to hydrologic time series analysis are developed assuming that the variables are normally distributed. It is necessary to transform the stochastic series of Mahi river to normal before carrying out the actual process of model building. The AR, MA, and ARMA models can be applied to stationary nonseasonal time series. ARIMA models can be fitted to nonstationary nonseasonal time series where the annual values of the statistical characteristics such as mean and variance is changing over years, for example, the average annual cost of hydroelectric power and annual consumption of water of an expanding metropolis constitute the time series which increase in magnitude over time, hence ARIMA model can be fitted to such series. Seasonality is removed first by standardizing the series as follows:

4.4.1 Standardization and normalization of time series variables

Sharma (1978) and Salas et. al., (1980) have discussed the method to stationarise the series by using Eq. (3.39). The results of standardized series are shown in Table 4.6. The time plot of standardized series of monthly Mahi river inflows is shown in Appendix C-2. Appendix C-2 shows no seasonality effect in the monthly observation.

After the series has been standardized by subtracting the mean inflows in each season from the seasonal data and dividing it by the standard deviation of the seasonal inflows, the next step is normalizing of data. The probability theory and statistical techniques applied to river inflow time series analysis are developed assuming that the variables are normally distributed (Gaussian). For normalization Box-Cox transformation given in Eq. (3.40) was used with the procedure stated in the following paragraph.

The constant in Eq. (3.40) is chosen just large enough to cause all entries in Z_t^{λ} to be positive. Although we may have a separate Box- Cox transformation for each season of the year, in order to reduce the number of parameters, it is assumed that the same λ and constant are used for each season. The purpose of the Box-Cox transformation is to rectify anomalies such as heteroscedasticity and non normality in the residuals of the ARMA model fitted to the deseasonalized time series.

For transformation of the data, an iterative procedure was adopted. The initial value of λ was taken to be 0.25 and the Eq. (3.40) was applied to compute the data transformation. After the transformation, the skewness coefficient of the data was checked for an approximate limit on the skewness coefficient. That is given by

$$-0.02 \leq Cs \leq 0.02$$

If the computed coefficient of skewness was within the specified, range the transformation was accepted. Otherwise the value of λ was reduced by 0.01 at each iteration and the transformation process is repeated.

Table 4.6 Standardized monthly inflow series for Mahi river

-0.78	1.86	0.49	-0.67	-0.01
0.33	0.26	-0.51	-0.94	-0.34
0.33	1.30	-0.48	0.05	0.36
-0.78	-0.97	0.83	-0.24	0.06
-0.78	0.30	-0.88	0.26	-0.05
1.07	0.11	0.39	1.30	0.88
-0.33	-1.04	0.37	0.25	0.05
-0.34	-0.64	-1.05	0.29	-0.37
0.38	-1.12	-1.22	0.21	-0.55
1.54	1.87	-0.95	0.55	0.60
2.31	0.65	-0.88	-0.94	-0.16
-0.78	-1.12	-0.25	-0.13	-0.39
0.00	-0.53	0.15	-0.64	-0.22
-0.78	-0.23	1.81	-0.87	0.32
-0.40	0.75	0.32	-0.13	0.37
-0.20	0.79	-0.73	-0.35	-0.09
0.07	1.72	1.88	-0.74	1.04
1.23	1.12	-0.22	0.66	0.70
2.42	1.11	0.54	-0.27	0.71
-0.78	-1.01	-0.30	0.47	-0.15
0.93	-0.03	-0.39	-0.49	-0.15
-0.78	-0.02	-0.81	0.98	0.13
-0.78	1.85	-0.54	2.11	1.20
-0.46	-0.59	-1.04	-0.94	-0.83
-0.05	2.18	-0.07	-0.94	0.26
0.01	-1.14	-0.53	-0.63	-0.63
-0.32	0.16	-0.73	3.97	1.40
0.97	-1.12	-0.04	0.96	0.24
-0.19	-0.10	-0.90	-0.56	-0.48
-0.42	-1.12	-0.30	-0.87	-0.67
-0.78	-0.36	-0.34	1.32	0.34
-0.25	0.78	0.23	1.03	0.81
-0.14	-1.12	-0.54	-0.93	-0.77
-0.52	-0.12	-0.70	2.70	0.83
-0.78	0.11	-0.47	1.17	0.37
-0.78	-0.40	0.08	-0.27	-0.12
0.12	-0.14	-0.35	-0.44	-0.21
-0.78	-1.07	-1.05	-0.87	-1.12
-0.48	-1.04	-0.72	-0.58	-0.59

0.35	-0.83	-1.17	0.48	-0.55
-0.78	0.86	0.45	-0.83	-0.61
-0.56	-0.37	0.21	0.56	-0.34
2.63	-0.47	-0.85	0.36	1.24
0.69	0.77	0.02	0.57	0.48
-0.32	-0.32	-0.29	-0.86	-0.88
-0.53	1.10	0.87	2.61	1.42
-0.55	-0.42	-0.23	-0.92	1.34
-0.06	-0.84	-0.21	0.04	-0.32
1.49	0.03	2.45	0.35	0.66
4.88	2.19	2.17	1.70	2.15
1.76	2.17	3.20	0.86	-1.20
0.20	-1.11	0.25	-0.98	-1.20
0.60	-0.37	0.72	-0.36	-1.20
-0.22	0.25	1.51	-0.53	-1.20
-0.55	-0.30	0.26	-0.73	-0.89
-0.26	-0.24	-0.22	-0.37	0.60
-0.51	-1.01	2.52	-0.73	-1.01
-0.78	-1.28	-0.93	-1.00	2.08
-0.53	2.00	-0.04	-1.00	-1.20
-0.37	-1.29	0.53	-0.81	-1.20
-0.69	-0.84	-0.04	-0.81	5.06
0.20	-0.72	-0.36	-0.53	-0.93
-0.64	-0.03	1.13	0.54	0.27
-0.44	1.05	1.05	-0.72	-0.39
-0.42	-1.10	-1.04	-0.22	0.17
-0.28	1.29	-0.17	-0.63	0.26
1.59	0.45	2.36	1.56	0.53
-0.69	0.45	-0.91	0.01	-0.33
-0.50	1.36	0.87	0.73	-0.84
-0.64	0.16	1.45	-0.54	-0.52
-0.66	-0.45	-0.47	0.99	0.72
-0.68	-0.99	-1.29	-0.86	-0.39
-0.78	-1.00	-1.20	-0.98	-1.20
0.34	-0.85	-1.00	-0.94	-1.16
-0.78	-1.28	-1.14	-0.75	-1.20
0.26	0.09	-0.56	-0.10	-0.97

By iterative procedure the value of λ was found to -0.23 with value of constant equal to 2.0. The value of coefficient of skewness with these values of λ and constant was obtained as 0.017, which is quite close to zero and this transformed series has become normal. The results of transformed standardized normal monthly inflow series of Mahi river are shown in Table 4.7. These values are used for stochastic model parameter estimation.

4.5 Identification of Models

When modelling a given data set, a large number of models are often available for consideration. The purpose of the identification stage is to ascertain the subset of models that appear to hold more promise for adequately modelling the time series. For the case of nonseasonal ARIMA models it is necessary to determine the order of differencing if homogenous nonstationarity is present, to ascertain the approximate number of AR and MA parameters that are required. When the observations are stationary, differencing is, of course, not required and one must only decide upon the ARMA model parameters that are needed for adequately describing the time series that may be transformed using a Box-Cox transformation.

By employing the simple graphical identification tools described in Section 3.6, the number of models which are worthwhile entertaining is reduced to just a few models. In application, whether to select an ARMA model or an ARIMA model is readily evident from the identification studies.

4.5.1 Time plot of original series

In constructing an appropriate model for Mahi river inflow series, the procedure for selection of the appropriate type of model among AR, MA, ARMA, ARIMA and seasonal ARIMA models and the orders for the selected model is as followed. The observed **virgin data** are listed in Appendix A-1. The time plot of the original monthly inflow series is shown in Appendix C-1. The plot shows that the monthly inflow of Mahi river in TMC from 1928 to 2003 are seasonal. For this reason the seasonal ARIMA or seasonal ARMA models may be required to describe the monthly series. However, the series is apparently stationary and does not have a trend when considered annual cycles.

4.5.2 Time plot of standardized series

The time plot of standardized monthly inflow series of Mahi river is shown in Appendix C-2. It is evident that the seasonality effect and month to month nonstationarity effects are completely

Table 4.7 Standardized normal monthly inflow series of Mahi river

0.19	1.16	0.82	0.28	0.63
0.77	0.74	0.38	0.06	0.48
0.77	1.04	0.40	0.66	0.78
0.19	0.03	0.92	0.53	0.67
0.19	0.76	0.12	0.74	0.62
0.99	0.69	0.79	1.04	0.94
0.48	-0.04	0.78	0.74	0.66
0.48	0.30	-0.05	0.75	0.46
0.78	-0.13	-0.26	0.72	0.36
1.10	1.16	0.05	0.84	0.86
1.24	0.87	0.11	0.06	0.57
0.19	-0.13	0.53	0.58	0.45
0.64	0.37	0.70	0.30	0.54
0.19	0.54	1.15	0.12	0.76
0.44	0.90	0.76	0.58	0.78
0.55	0.91	0.23	0.47	0.60
0.67	1.13	1.17	0.23	0.98
1.03	1.00	0.54	0.88	0.89
1.26	1.00	0.84	0.51	0.89
0.19	-0.01	0.50	0.82	0.57
0.95	0.63	0.45	0.40	0.57
0.19	0.63	0.17	0.97	0.70
0.19	1.16	0.36	1.21	1.02
0.41	0.33	-0.04	0.06	0.15
0.62	1.22	0.61	0.06	0.74
0.64	-0.15	0.37	0.30	0.30
0.49	0.71	0.23	1.47	1.07
0.96	-0.13	0.62	0.96	0.73
0.56	0.60	0.09	0.35	0.40
0.44	-0.13	0.50	0.12	0.28
0.19	0.47	0.48	1.05	0.77
0.53	0.91	0.73	0.98	0.92
0.58	-0.13	0.36	0.06	0.20
0.37	0.59	0.25	1.30	0.92
0.19	0.68	0.41	1.01	0.78
0.19	0.45	0.67	0.51	0.59
0.69	0.58	0.47	0.42	0.55

0.19	-0.07	-0.05	0.12	-0.13
0.40	-0.04	0.24	0.33	0.33
0.77	0.16	-0.19	0.82	0.36
0.19	0.93	0.81	0.15	0.32
0.35	0.46	0.72	0.85	0.48
1.29	0.40	0.14	0.78	1.03
0.89	0.91	0.65	0.85	0.82
0.49	0.49	0.51	0.13	0.11
0.37	1.00	0.94	1.29	1.07
0.36	0.44	0.54	0.08	1.05
0.62	0.15	0.55	0.66	0.49
1.09	0.65	1.26	0.78	0.88
1.56	1.22	1.22	1.13	1.21
1.14	1.22	1.37	0.93	-0.23
0.72	-0.12	0.74	0.02	-0.23
0.86	0.46	0.89	0.47	-0.23
0.54	0.74	1.09	0.37	-0.23
0.36	0.50	0.75	0.23	0.11
0.52	0.53	0.54	0.46	0.86
0.38	-0.01	1.27	0.23	-0.01
0.19	-0.34	0.07	0.00	1.20
0.37	1.19	0.62	0.00	-0.23
0.46	-0.35	0.83	0.17	-0.23
0.26	0.14	0.62	0.17	1.57
0.72	0.24	0.47	0.37	0.07
0.30	0.63	1.00	0.84	0.75
0.42	0.98	0.98	0.24	0.45
0.43	-0.11	-0.04	0.54	0.71
0.51	1.04	0.56	0.30	0.74
1.11	0.81	1.25	1.10	0.83
0.26	0.81	0.08	0.65	0.48
0.38	1.06	0.94	0.90	0.15
0.30	0.71	1.08	0.37	0.37
0.28	0.42	0.40	0.97	0.89
0.27	0.01	-0.35	0.13	0.45
0.19	0.00	-0.22	0.02	-0.23
0.77	0.14	0.00	0.06	-0.17
0.19	-0.35	-0.16	0.22	-0.23
0.74	0.68	0.35	0.60	0.03

removed and the series exhibits a random series even on monthly basis. The standardized time plot of Appendix C-2 does not show any trend or nonstationarity in the monthly inflow series and there is no need of applying seasonal ARIMA or seasonal ARMA model to the inflow series of Mahi river if transformed data are used. Since, the standardized series does not show any seasonality or nonstationarity, only nonseasonal models were used.

4.5.3 Autocorrelation function (ACF) analysis

The standardized normal series of Table 4.7 was used for calculating the sample autocorrelation function (ACF) of the series at different lags. The ACF and their 95% confidence limits were computed up to a lag of 95 and are shown in Appendix D-1. The values of ACF of Mahi river against lag have been plotted in Figure 4.4. As can be seen in Figure 4.4 the estimated ACF of Mahi river inflow has significant non zero values at lower lags (upto 6 lag) and does not truncate at specified lag. Because the theoretical ACF of an AR process behaves in this fashion, this indicates that perhaps some type of model which contains an AR component should fit to the Mahi river inflows.

4.5.4 Partial autocorrelation function (PACF) analysis

Because the ACF of a process attenuates and does not truncate at a specified lag it is advantageous to consider partial autocorrelation function (PACF) which does cut off for an AR process.

The approximate 95% confidence limits for PACF have been taken as $\pm\ 1.96/\sqrt{N} = \pm 1.96/\sqrt{380} = \pm\ 0.092$. The estimates of PACF have been computed upto a lag of N/4 i.e. upto a lag of 95, and are shown in Appendix D-2. The values of PACF of Mahi river against lag have been plotted along with 95% confidence limits in Figure 4.5. It can be seen that there are rather large value for the estimated PACF at lag 1 and 2. Since the PACF does not truncate and attenuates at lag 22, 46, 63, 66, 68, 71, 81, 82, 87 and 90, it indicates that MA parameters are needed in the model, and it will be appropriate to fit some type of ARMA (p, q) model to the data of Mahi river.

4.5.5 Inverse autocorrelation function (IACF) analysis

The IACF for Mahi river has been computed upto a lag of 95 and IACF along with 95% confidence limits are shown in Appendix D-3. The IACF values against lag have been plotted in Figure 4.6. The IACF plot is a useful tool in identifying the type of ARMA model to be fitted to the series. It can be seen in Figure 4.6 the estimated IACF of Mahi river inflow has significant non zero values at lower lags and follow a damped exponential curve. It indicates the strong presence of moving average parameter.

4.5.6 Inverse partial autocorrelation function (IPACF) analysis

The IPACF of an ARMA (p, q) process can be defined as PACF of an ARMA (q, p) process. The IPACF for Mahi river have been computed upto lag 95 and sample IPACF along with 95% confidence limits are shown in Appendix D-4. The IPACF values against lag plotted for Mahi river in Figure 4.7. The IPACF are significantly different from zero at lower lag (1, 2, 3, 4, 5, 8, 9, etc.) and decay closely to zero after 24 lag. This further confirms the presence of AR terms of lower lags.

4.5.7 Application of identification tools

After examining the time plot of original time series for Mahi river in Appendix C-1, it was seen that the time series of Mahi river inflow exhibits a seasonal pattern. The standardized series for Mahi river is plotted in Appendix C-2 which shows that the series has become stationary and seasonal effects from the original series have also been removed and now only AR, MA or ARMA model may be fitted to the Mahi river inflows.

For further identifying the order of AR and MA terms the ACF, PACF IACF and IPACF are plotted in Figure 4.4 to Figure 4.7. The estimated ACF of Mahi river inflow has significant non zero values at lower lags (upto 6 lag) and does not truncate at specified lag. This indicates the presence of AR term in the model. The estimated PACF has large value at lag one and two. This fact shows presence of moving average term upto lag 2. It has also been suggested by Mujumdar (1990) that AR parameters of upto order 6 and MA parameters of upto 2 would in general, serve the purpose. The sample IACF and IPACF curves strongly showed the need of both autoregressive and moving average terms in the model (i.e. ARMA model) but does not clearly reveal the order of model that could be selected as candidate model. This fact clearly indicates the need of both AR and MA term in the models, so pure AR and MA models were discarded from the study. Therefore, only the following models which represent Mahi river inflow are investigated: ARMA(1,1), ARMA(2,1), ..., ARMA(6,1), ARMA(1,2), ARMA(2,2) ….and ARMA(6,2).

Description of identified models of Mahi river inflows

The models selected for Mahi river inflow are of ARMA family which may be written as

$$Z_{(t)} = \sum_{j=1}^{p} \phi_j Z_{(t-1)} - \sum_{j=1}^{q} \theta_j a_{(t-1)} + \dots + a_{(t)} \qquad \text{... (4.1)}$$

where,

$Z_{(t)}$	is the series being modeled.
P	is the number of AR parameters
ϕ_j	is the j^{th} AR parameters
q	is the number of MA parameters
θ_j	is the j^{th} MA parameter
[$a_{(t)}$, t=1, 2,..]	is the residual series.

The important assumption involved in these models is that residual series [$a_{(t)}$] has zero mean with terms which are independent to each other. Subsequent analysis is made for above selected models and results are discussed in the following sections.

4.6 Estimation of Parameters

Identification methods used previously are rough procedures applied to a data set to indicate the kind of representational model which is worth further investigation. We got some idea of the p and q needed in general ARMA model and now we have to obtain initial guesses for the parameters. The tentative model so obtained will provide a starting point for the application of the more formal and exact estimation of parameters.

A time series contains only partial information about the phenomenon under study. Therefore, the true or population values of the parameters of a model fitted to the series are not known. Thus, there is uncertainty about the estimation of the model parameters. This uncertainty for a specified parameter estimate is quantified by the term called standard error of the estimate.

4.6.1 Initial estimates for the parameters

The parameters of identified models which represent Mahi river inflow are estimated at two levels of increasing accuracy; an initial estimate of parameters and an exact non-linear estimation.

For making initial estimates of the parameters the data of Mahi river stochastic inflow series are first transformed using Box-Cox transformation in order to alleviate problems with non normality and changing variance. Standardized normal monthly inflow series of Mahi river is shown in Table 4.7. First 74 year data (n = 370) were used for making initial and final estimate of the parameters. The same 74 year data were used for validation of these parameters. Initial estimates and exact estimates of the parameters of Mahi river inflow models were obtained with the help of SPSS software package. The results of the computation of initial estimates of selected models are shown in Table 4.8.

Table 4.8 Initial estimates of parameters for identified models of Mahi river inflow

S. No.	Model	AR Parameters	MA parameters
1	ARMA(1,1)	$\phi_{1=}$ 0.94659	$\theta_{1=}$ -0.26971
2	ARMA(2,1)	$\phi_{1=}$ 1.08456 $\phi_{2=}$ -0.10299	$\theta_{1=}$ 0.16013
3	ARMA(3,1)	$\phi_{1=}$ 0.98880 $\phi_{2=}$ -0.05563 $\phi_{3=}$ 0.03230	$\theta_{1=}$ 0.19374
4	ARMA(4,1)	$\phi_{1=}$ 1.41087 $\phi_{2=}$ -0.30052 $\phi_{3=}$ -0.11185 $\phi_{4=}$ -0.13699	$\theta_{1=}$ 0.69182
5	ARMA(5,1)	$\phi_{1=}$ 1.07055 $\phi_{2=}$ -0.11019 $\phi_{3=}$ -0.00667 $\phi_{4=}$ -0.05626 $\phi_{5=}$ 0.09049	$\theta_{1=}$ 0.42612
6	ARMA(6,1)	$\phi_{1=}$ 0.70372 $\phi_{2=}$ 0.02316 $\phi_{3=}$ 0.05369 $\phi_{4=}$ -0.00793 $\phi_{5=}$ 0.11346 $\phi_{6=}$ 0.06816	$\theta_{1=}$ 0.11900
7	ARMA(1,2)	$\phi_{1=}$ 0.97576	$\theta_{1=}$ 0.05960 $\theta_{2=}$ -0.48047
8	ARMA(2,2)	$\phi_{1=}$ 0.67512 $\phi_{2=}$ 0.28458	$\theta_{1=}$ -0.07946 $\theta_{2=}$ -0.30708
9	ARMA(3,2)	$\phi_{1=}$ -0.38224 $\phi_{2=}$ 0.52582 $\phi_{3=}$ -0.04697	$\theta_{1=}$ -0.86767 $\theta_{2=}$ -0.56191

10	ARMA(4,2)	$\phi_{1=}$0.48192 $\phi_{2=}$0.73964 $\phi_{3=}$-0.18467 $\phi_{4=}$-0.11714	$\theta_{1=}$-0.02190 $\theta_{2=}$0.19810
11	ARMA(5,2)	$\phi_{1=}$1.45692 $\phi_{2=}$-0.78318 $\phi_{3=}$0.13668 $\phi_{4=}$-0.00291 $\phi_{5=}$0.15584	$\theta_{1=}$0.73936 $\theta_{2=}$-0.65260
12	ARMA(6,2)	$\phi_{1=}$0.16098 $\phi_{2=}$0.60419 $\phi_{3=}$-0.00612 $\phi_{4=}$-0.01155 $\phi_{5=}$0.08293 $\phi_{6=}$0.11728	$\theta_{1=}$-0.26303 $\theta_{2=}$0.04737

4.6.2 Exact estimation of parameters

Having obtained the initial estimates of parameters of the identified models to represent Mahi river inflows, these are used as input to estimate the exact value of parameters. The exact estimates of parameters take into account all the information contained in the data. After the identification and initial estimates of the models, the exact estimation of parameters is computed. The method used in the present study is the unconditional least square estimation method as explained by Box and Jenkins (1976) and described in Section 3.7.2. Table 4.10 shows the exact estimates of autoregressive and moving average parameters and their standard error values associated with the parameters. For a parameter to be significant, its absolute value must be larger than the standard error. In parameter estimation it is noted that in ARMA(4, 1), ARMA(5, 1), ARMA(6, 1), ARMA(5, 2) and ARMA(6, 2) the absolute value of some of autoregressive parameter ϕ is less than its standard error, while in ARMA(3, 2), ARMA(5, 2) and ARMA(6, 2) model the absolute value of moving average parameter θ, is less than its standard value. Therefore it can be argued that in these models parameters are significantly different from zero even at 1% significance level and therefore these models should be considered. Whereas in other models, the values of parameters were greater than their standard error.

Therefore it may be emphasized that only ARMA(1, 1), ARMA(2, 1), ARMA(3, 1), ARMA(1, 2), ARMA(2, 2) and ARMA{4, 2) models should be considered. The mathematical form of the above six ARMA models are given in Table (4.11).

Table 4.9 Exact estimates of parameters for identified models of Mahi river inflow

S. No.	Model	Variance of Residual	AR Parameters	Standard error of AR parameters	MA parameters	Standard error of MA parameters
1	ARMA(1,1)	0.14963365	$\phi_{1=}0.99361355$	0.00545077	$\theta_{1=}0.84568118$	0.03018629
2	ARMA(2,1)	0.14593257	$\phi_{1=}1.2042700$ $\phi_{2=}-0.2061261$	0.05593384 0.05507130	$\theta_{1=}0.9294811$	0.02424835
3	ARMA(3,1)	0.14533209	$\phi_{1=}1.2262880$ $\phi_{2=}-0.1499627$ $\phi_{3=}-0.0766103$	0.00686318 0.05428975 0.05373475	$\theta_{1=}0.9704864$	0.01663213
4	ARMA(4,1)*	0.14520651	$\phi_{1=}1.2246561$ $\phi_{2=}-0.1559043$ $\phi_{3=}-0.0299447$ $\phi_{4=}-0.0389555$	0.00143322 0.05368313 0.08248522 0.05358397	$\theta_{1=}0.9825172$	0.01520880
5	ARMA(5,1)*	.145897130	$\phi_{1=}1.2325351$ $\phi_{2=}-0.1648910$ $\phi_{3=}-0.0249249$ $\phi_{4=}-0.0877462$ $\phi_{5=}0.0448066$	0.03851909 0.07073628 0.08256637 0.07059271 0.03838120	$\theta_{1=}0.9772526$	0.01656580
6	ARMA(6,1) *	0.14503123	$\phi_{1=}1.2222396$ $\phi_{2=}-0.1395392$ $\phi_{3=}-0.0335092$ $\phi_{4=}-0.0815394$ $\phi_{5=}0.0958744$ $\phi_{6=}-0.0636384$	0.04431458 0.07491090 0.08288857 0.07407574 0.04426714 0.04027176	$\theta_{1=}0.9847857$	0.01699372
7	ARMA(1,2)	0.14652578	$\phi_{1=}0.99677559$	0.00349451	$\theta_{1=}0.74068169$ $\theta_{2=}0.15348236$	0.05191439 0.05206479
8	ARMA(2,2)	0.14543665	$\phi_{1=}1.4768093$ $\phi_{2=}-0.4772368$	0.01173329 0.01116057	$\theta_{1=}1.2133995$ $\theta_{2=}-0.2418099$	0.05159098 0.05189396
9	ARMA(3,2) *	0.14670653	$\phi_{1=}0.47937257$ $\phi_{2=}0.67793212$	1.5012355 1.7824806	$\theta_{1=}0.20652149$ $\theta_{2=}0.67596262$	1.5034142 1.3929274

			$\phi_3=-0.16032115$	0.2891514		
10	ARMA(4,2)	0.14529265	$\phi_1=0.55034552$ $\phi_2=0.69431204$ $\phi_3=-0.14818205$ $\phi_4=-0.09659738$	0.24830025 0.24823160 0.04937536 0.04934607	$\theta_1=0.30071459$ $\theta_2=0.67736496$	0.22718304 0.21571274
11	ARMA(5,2) *	0.14578869	$\phi_1=0.85993480$ $\phi_2=0.31329799$ $\phi_3=-0.10060837$ $\phi_4=-0.11216831$ $\phi_5=0.03943231$	0.40104002 0.45786373 0.06222641 0.06222293 0.06692465	$\theta_1=0.60636076$ $\theta_2=0.37372862$	0.39054761 0.38517485
12	ARMA(6,2) *	0.14616469	$\phi_1=0.41809608$ $\phi_2=0.85234325$ $\phi_3=-0.15711763$ $\phi_4=-0.10353397$ $\phi_5=-0.03940869$ $\phi_6=0.02949446$	0.34957431 0.37749541 0.07282482 0.06556263 0.05929551 0.04906482	$\theta_1=0.16869191$ $\theta_2=0.80718436$	0.32637125 0.31971130

* Model in which some of parameters have smaller value than corresponding standard error. These models were discarding from further study.

4.6.3 Model structure:

Model structure of time series constitutes the sum of trend, periodic component and stochastic components. As the observed inflow series was found to be trend free, the sub models of periodic and stochastic components are added together to form newly developed model structure of the inflow series. The mathematical structure of the additive model as describe by Eq. (3.14) can be represented and given in Table 4.10.

Table 4.10 Model structure of Mahi river inflow series

S. No.	Model	Periodic series	Stochastic series
1	**ARMA(1,1)**	$544.97 - 406.99 \cos\left(\frac{2.\pi.t}{5}\right) - 323.18 \sin\left(\frac{2.\pi.t}{5}\right)$	$Z_t = 0.99361355\ Z_{t-1} - 0.84568118\ a_{t-1} + a_t$
2	**ARMA(2,1)**		$Z_t = 1.2042700\ Z_{t-1} - 0.2061261\ Z_{t-2} - 0.9294811\ a_{t-1} + a_t$
3	**ARMA(3,1)**		$Z_t = 1.2262880 Z_{t-1} - 0.1499627\ Z_{t-2} - 0.0766103 Z_{t-3} - 0.9704864\ a_{t-1} + a_t$
4	**ARMA(1,2)**		$Z_t = 0.99677559\ Z_{t-1} - 0.74068169\ a_{t-1} - 0.15348236\ a_{t-2} + a_t$
5	**ARMA(2,2)**		$Z_t = 1.4768093\ Z_{t-1} - 0.4772368\ Z_{t-2} - 1.2133995\ a_{t-1} + 0.2418099\ a_{t-2} + a_t$
6	**ARMA(4,2)**		$Z_t = 0.55034552\ Z_{t-1} + 0.69431204\ Z_{t-2} - 0.14818205 Z_{t-3} - 0.09659738 Z_{t-4} - 0.30071459\ a_{t-1} - 0.67736496\ a_{t-2} + a_t$

The above models identified to represent Mahi river inflow was tested for validation and forecasting was made from these models.

4.7 Validation of Models

The models to represent Mahi river inflow have been identified and the parameters estimated, validation tests are then to be applied to the fitted models, for the assumptions used in building these stochastic models i.e. the residual series $a_{(t)}$ has zero mean and independent to each other is present in the residual series. The following tests are carried out to examine whether the assumption used in building the model are in fact valid for the model selected:

1) ACF of residual series
2) Significance of residual mean

3) Box- Pierce Portmanteau lack of fit test

All the validation tests are carried out on the residual series only. Residual series is constructed by Eq. (3.58).

4.7.1 ACF of residual series

The residuals $a_{(t)}$ were calculated for all the six models and autocorrelation function of the residual series $a_{(t)}$ were computed for diagnostic checking. Autocorrelograms of residuals, $a_{(t)}$ of all six models upto 74 lags were constructed for judging the mutual dependency and are shown in Figure 4.8 to Figure 4.13 respectively. It is evident from the autocorreograms that the residuals obtained from all the six models are mutually independent as they lie between the confidence limits and not significantly different from zero.

4.7.2 Significance of residual mean

The purpose of this test is to examine the validity of the assumption that the series $a_{(t)}$ has zero mean. The value of test statistics Q was computed by Equation (3.59) and is given in Table 4.11. The calculated values of test statistics were compared with the table value of t statistic at 5 per cent level of significance. The comparisons of test statistic with tabulated value of t statistic are given in Table 4.11. It is observed that the residual series passes the test in all cases. This must be true when models are fitted to the standardized series.

Table 4.11 Significance of residual mean test

Model	N	Residual mean	Variance of Residual	Statistic	Tabulated
ARMA(1,1)	370	0.0170	0.150	0.846	1.645
ARMA(2,1)	370	0.0082	0.146	0.411	1.645
ARMA(3,1)	370	-0.0023	0.145	-0.115	1.645
ARMA(1,2)	370	0.0104	0.147	0.523	1.645
ARMA(2,2)	370	0.0016	0.145	0.081	1.645
ARMA(4,2)	370	-0.0056	0.145	-0.282	1.645

4.7.3 Box-Pierce Portmanteau lack of fit test

Box Pierce Portmanteau lack of fit test was used for checking the adequacy of autoregressive models. The value of test statistics Q was computed by Equation (3.60) and is given in Table 4.12. The test is normally carried out L= 0.15N in Equation (3.60). However the test was carried out for different values of L. The calculated values of test statistics were compared with the table value of Chi-square at 5 per cent level of significance. The comparison of test statistics with tabulated value of Chi-square is given in Table 4.12. It can be seen from the table that the values of test statistics for all the six AR models are found to be less than tabulated value of Chi-square. Therefore, all the six AR models gave good fit and are acceptable.

Table 4.12 Box-Pierce Portmanteau lack of fit test

$Y^2_{0.95}$ (N) Model	**N=19 $Y^2_{tabulated}$ = 30.14**	**N=37 $Y^2_{tabulated}$ = 55.16**	**N=56 $Y^2_{tabulated}$ = 78.07**	**N=74 $Y^2_{tabulated}$ = 99.18**
ARMA(1,1)	25.82	40.99	63.44	82.68
ARMA(2,1)	16.10	31.66	61.59	79.23
ARMA(3,1)	15.88	30.60	61.14	78.13
ARMA(1,2)	17.00	32.72	61.70	79.58
ARMA(2,2)	15.23	30.16	59.96	77.07
ARMA(4,2)	14.93	28.26	56.36	72.56

4.8 Regeneration of Inflows

After passing the validation test all six models were used for regeneration of Mahi river inflow series as given in section 3.9. The generation process is started off initially by setting unknown a_t equal to their unconditional expected values of zero. The actual monthly Mahi river inflow series has been shown in Appendix A-1. The generated monthly Mahi river inflow series by different models have been shown in Appendix E-1 through Appendix E-6. The generated values start of time t =1 i.e., June 1928 and end at t = 370, i.e. October, 2001. Thus 370 values at lead time one have been computed.

The actual and generated monthly Mahi river inflows are plotted together to compare the generated values with the actual observations. This comparison is shown in Figure 4.14 through Figure 4.19 for different models fitted to Mahi river inflow series. It is observed that seasonal monsoonal pattern of inflow I series is maintained in generated values by all the six models.

It is noted that the regeneration were obtained on transformed standardized normal inflow series of Mahi river. Therefore the actual regeneration is obtained by applying inverse Box-Cox transformation to the generated values obtained. The values now obtained are in standardized form and they are converted back in original units of inflow in TMC by multiplying by standard deviation of that month and adding the mean flow of that month and then add the periodic component.

It may be noted that the stochastic models for forecasting river inflows have been used in United Kingdom, United States of America and Australia where the rivers have perennial flow by snow melting and effect of monsoon is not so much predominant on river inflow, whereas in India the rivers are monsoonal rivers and rainfall phenomenon is highly erratic, and it can not be predicted in terms of actual value, only it may be predicted that the coming month or year will have high or low rainfall as compared to the normal rainfall. The river inflow is a rainfall dependent phenomenon. Mahi river is a monsoonal river and the models used in this study express the deviations from the mean inflow and random element. It may be observed from Figure 4.14 to Figure 4.19 that the generated inflows by all the models follow the same pattern of high and low inflow as of actual inflow.

4.8.1 Evaluation of regeneration performance

The generated data were used to assess the regeneration performance of the models were made by tests of goodness of fit. Results of the stochastic models were evaluated to assess the regeneration performance. The models were evaluated with regard to several errors such as mean forecast error, mean absolute error, root mean square error and integral square error. The result is shown in Table 4.13.

Table 4.13 Performance evaluation of regeneration process

Model	MFE	MAE	RMSE	ISE
ARMA(1,1)	-104.45	332.79	760.55	521.88
ARMA(2,1)	-98.95	333.09	757.39	517.54
ARMA(3,1)	-92.20	335.24	753.87	512.74
ARMA(1,2)	-100.53	332.80	759.43	520.34
ARMA(2,2)	-94.85	334.38	754.70	513.88
ARMA(4,2)	-90.04	337.94	755.00	514.28

4.9 Comparison of Different Models

Having generated inflows by all the six models fitted to the Mahi river inflow data, the basic statistical characteristics such as mean, standard deviation, variance, coefficient of variation, coefficient of skewness, kurtosis and lag one serial correlation coefficients are determined for all the generated series and they are compared with the basic statistical characteristics of the actual series. The statistics for actual series were computed for all the 380 available data whereas the statistics for the generated series has been calculated for generated 370 data. The results are shown In Table 4.14

Mean

Table 4.14 shows the comparison of mean of actual and generated series of Mahi river inflow. It is observed that all the models have lower generated mean for all the months. Table 4.14 shows the percent deviation of mean monthly inflows in forecast by different models. All the model underestimates the inflow of June month by 26.66 percent to 31.62 percent. The ARMA (4, 2) model underestimates the mean by 12.68 percent in the month of August to 26.66 percent in the month of July. The mean inflow of months of August and September which are highly susceptible to forecast errors, ARMA (4, 2) underestimates as 12.68 percent and 18.52 percent respectively. The percentage of deviation from actual monthly mean is higher in all the other models as compared to ARMA (4, 2). Underestimate of mean monthly inflow varies from 13.53 percent to 31.62 percent by other models. Thus ARMA (4, 2) preserves the monthly means in a better way as compared to all other models.

Table 4.14 Comparison of actual and regenerated Statistical characteristics of Mahi river inflow series

Model	June	July	August	September	October	Seasonal
Mean, TMC						
Actual	110.15	707.67	1028.02	761.69	117.30	2724.84
ARMA(1,1)	78.48	592.04	875.28	597.52	97.93	2241.25
ARMA(2,1)	80.37	599.61	881.20	608.41	99.15	2268.74
ARMA(3,1)	82.71	607.70	893.78	617.78	100.50	2302.47
ARMA(1,2)	79.76	597.47	878.54	606.25	98.80	2260.82
ARMA(2,2)	81.78	604.49	888.97	614.02	99.95	2289.21
ARMA(4,2)	83.45	610.59	897.71	620.66	100.86	2313.28
Percent deviation of mean						
ARMA(1,1)	31.62	16.34	14.86	21.55	16.52	17.75
ARMA(2,1)	29.74	15.27	14.28	20.12	15.48	16.74
ARMA(3,1)	27.40	14.13	13.06	18.89	14.32	15.50

ARMA(1,2)	30.34	15.57	14.54	20.41	15.77	17.03
ARMA(2,2)	28.32	14.58	13.53	19.39	14.80	15.99
ARMA(4,2)	26.66	13.72	12.68	18.52	14.01	15.10
Standard deviation						
Actual	141.06	549.78	746.76	760.42	97.58	1527.86
ARMA(1,1)	48.13	182.64	249.03	271.98	33.82	748.40
ARMA(2,1)	42.07	163.19	236.27	252.01	29.53	621.94
ARMA(3,1)	39.33	138.58	211.60	222.08	25.33	521.15
ARMA(1,2)	42.69	168.56	237.43	255.67	30.30	649.43
ARMA(2,2)	40.16	146.46	219.72	231.88	26.94	556.09
ARMA(4,2)	37.60	134.43	200.27	208.25	24.34	477.06
Variance						
Actual	19897.9	302261.7	557656.3	578236.8	9522.1	2334346.9
ARMA(1,1)	2316.7	33356.6	62016.2	73975.7	1144.0	560099.9
ARMA(2,1)	1769.8	26632.5	55823.6	63507.2	871.8	386811.2
ARMA(3,1)	1546.9	19203.5	44774.9	49317.6	641.8	271600.8
ARMA(1,2)	1822.0	28413.5	56374.1	65366.4	918.0	421756.6
ARMA(2,2)	1613.1	21451.1	48276.1	53766.8	725.6	309239.7
ARMA(4,2)	1414.1	18071.5	40108.9	43369.5	592.4	227589.7
Coefficient of variation						
Actual	1.28	0.78	0.73	1.00	0.83	0.56
ARMA(1,1)	0.61	0.31	0.28	0.46	0.35	0.33
ARMA(2,1)	0.52	0.27	0.27	0.41	0.30	0.27
ARMA(3,1)	0.48	0.23	0.24	0.36	0.25	0.23
ARMA(1,2)	0.54	0.28	0.27	0.42	0.31	0.29
ARMA(2,2)	0.49	0.24	0.25	0.38	0.27	0.24
ARMA(4,2)	0.45	0.22	0.22	0.34	0.24	0.21
Coefficient of skewness						
Actual	2.35	0.67	1.21	1.53	1.92	0.95
ARMA(1,1)	0.87	1.23	1.53	1.81	1.67	1.71
ARMA(2,1)	0.19	1.27	0.72	1.23	0.79	1.49
ARMA(3,1)	-0.05	0.93	0.60	0.71	0.62	1.15
ARMA(1,2)	0.35	1.32	0.85	1.43	0.90	1.57
ARMA(2,2)	0.03	1.06	0.65	0.92	0.71	1.29
ARMA(4,2)	-0.12	0.72	0.50	0.55	0.51	0.95

Kurtosis						
Actual	7.26	-0.55	1.15	2.80	7.56	1.23
ARMA(1,1)	2.53	3.47	5.24	6.86	5.92	6.35
ARMA(2,1)	1.21	2.58	1.45	3.30	0.95	4.90
ARMA(3,1)	0.35	1.92	0.29	1.25	0.07	3.16
ARMA(1,2)	1.76	2.82	2.31	4.34	1.69	5.43
ARMA(2,2)	0.58	2.21	0.69	1.98	0.38	3.88
ARMA(4,2)	0.28	1.09	-0.02	0.40	-0.31	2.08
Lag-one serial correlation coefficient						
Actual	0.300	0.014	0.235	0.052	-0.093	0.155
ARMA(1,1)	0.477	0.565	0.609	0.606	0.591	0.652
ARMA(2,1)	0.285	0.587	0.407	0.550	0.448	0.650
ARMA(3,1)	0.120	0.392	0.242	0.393	0.294	0.459
ARMA(1,2)	0.367	0.633	0.483	0.595	0.509	0.693
ARMA(2,2)	0.157	0.451	0.283	0.445	0.330	0.514
ARMA(4,2)	0.059	0.236	0.114	0.280	0.146	0.324

Standard deviation

Table 4.14 shows the comparison of monthly standard deviations of actual and generated inflow series by different models. The monthly standard deviations by different models are generally low as compared to those of actual series. This difference is obvious because the generated series have been obtained from transformed standardized normal series of actual series. The ARMA (1,1) model preserves the standard deviation better as compared to all other models.

Variance

The monthly variances of actual and generated series are shown in Table 4.14. The variance is a better statistics to compare the models August and September months have the highest variance in actual series. The variance of month of August and September are better retained by ARMA (1,1) model. The annual variance is also better retained by ARMA (1,1) model.

Coefficient of variation

The monthly coefficients of variation of actual and generated series are shown in Table 4.14. It is observed that coefficients of variation are more than 1.0 in the month of June in the actual inflow series. The coefficient of variation is better retained by ARMA (1,1) model.

Coefficient of skewness

The coefficients of skewness for all the months for actual and generated series by different models are shown in Table 4.14. The Mahi river inflows are highly skewed in all the months. The skewness of actual series is highest at 2.35 in the month of June and lowest 0.67 in the month of July. Although the regeneration was made by transforming the actual series to normal having skewness coefficient as zero. However, when the regeneration was inverse transformed the skewness was observed in all the generated series. The inverse transformation does not entirely restore the 'hidden information' which was earlier caused by the transformation. For the months August September and October the skewness is better preserved by ARMA (1,1) model.

Kurtosis

The kurtosis of the actual and generated series is shown in Table 4.14. The kurtosis is as high as 7.26 and 7.56 in the month of June and October respectively for actual series. The kurtosis is distorted in all the months in generated series by all the models.

Lag-one serial correlation coefficient

The lag-one serial correlation coefficients for actual and generated series are shown in Table 4.14. In actual series for the month of September the correlation coefficient is negative. The generated series shows high coefficient by ARMA (1,1) model and low serial correlation coefficient by ARMA (4,2) models.

4.10 Selection of Most Appropriate Model

For selection of the most appropriate model for Mahi river inflow forecasting, first the inherent process of model building is considered which involve identification, estimation and validation of different models.

In identification stage, total number of 12 models i.e. ARMA(1,1), ARMA(2,1), ..., ARMA(6,1), ARMA(1,2), ARMA(2,2) ….and ARMA(6,2) were selected. In parameter estimation it was noted in Section 4.6.2 that in ARMA (4,1), ARMA (5,1), ARMA(6,1), ARMA(5,2) and ARMA(6,2) the absolute value of some of autoregressive parameter ϕ is less than its standard error and in ARMA (3,2), ARMA(5,2) and ARMA(6,2) model the absolute value of moving average parameter θ, is less than its standard value. For a parameter to be significant its absolute value must be larger than its standard error. Therefore at this stage itself, it is doubtful that (4, 1), ARMA (5,1), ARMA(6,1), ARMA (3,2), ARMA(5,2) and ARMA(6,2) models will forecast the Mahi river inflow adequately. Therefore six models namely ARMA(1,1), ARMA (2,1), ARMA(3,1), ARMA(1,2), ARMA(2,2) and ARMA{4,2) models are expected to provide better results.

Validation of all the six models was tested for significance of residual mean and their independence. It is seen in Section 4.7 that all the six models pass these tests.

After regeneration by all the six models, the basic statistical characteristics of actual and generated series were compared. All the six models have mixed results as far as basic statistical characteristics of actual and generated series are compared. Even periodic models PAR/PARMA considered by other authors are unable to retain the basic statistical characteristics of monsoon rivers in India.

Therefore to select the most appropriate model out of the six models considered the criterion of **minimum mean square error** of forecast was considered. The mean square error of forecast for 370 values was calculated as given below in Table 4.15.

Table 4.15 Mean square error in regeneration by different models

S.No.	Model	MSE in transformed domain	MSE in inflow domain
1	ARMA(1,1)	0.149889	576878
2	ARMA(2,1)	0.145681	572087
3	ARMA(3,1)	0.144971	**566781**
4	ARMA(1,2)	0.146269	575181
5	ARMA(2,2)	0.144944	568044
6	ARMA(4,2)	**0.144345**	568487

From the above comparison it is seen that that ARMA (4, 2) is giving the minimum mean square error of 0.144345 in transformed domain. But ARMA (3, 1) model is having minimum mean square error of 566781 in inflow domain. Considering all above discussion of selection of most appropriate model it is concluded that **ARMA (3, 1)** model is the most appropriate model out of the twelve identified models for forecasting Mahi river inflows. The model structure for representing the Mahi river inflow series are given below:

$$X_t = 544.97 - 406.99 \, Cos \left(\frac{2\pi.t}{5}\right) - 323.18 \, Sin \left(\frac{2\pi t}{5}\right) + 1.2262880Z_{t-1} - 0.1499627 \, Z_{t-2} - 0.0766103Z_{t-3} - 0.9704864 \, a_{t-1} + a_t \quad \text{... (4.2)}$$

4.11 Forecasting of Inflow

The most appropriate model, i.e., ARMA (3, 1) was used for forecasting of monthly inflow values, for the year 2002 and year 2003 (Table 4.16). Forecasting performance of the model was evaluated by graphical comparison of historical and forecasted series (Figure 4.20). Values of correlation coefficient between observed and generated monthly inflow series for the two year period (2002-2003) was found to be 0.912, which was tested at 5 % level of significance, which show adequacy of developed model.

Table 4.16 Forecasted Inflows with historical inflows for year 2002 and 2003.

S.No.	Historical Inflow in TMC	Forecasted inflow in TMC
June, 2002	0.00	0.00
July, 2002	1.77	85.58
Aug., 2002	174.32	123.58
Sep., 2002	192.41	40.00
Oct., 2002	0.00	8.29
June, 2003	146.30	22.00
July, 2003	759.23	307.58
Aug., 2003	607.37	460.91
Sep., 2003	686.24	499.18
Oct., 2003	22.33	26.49

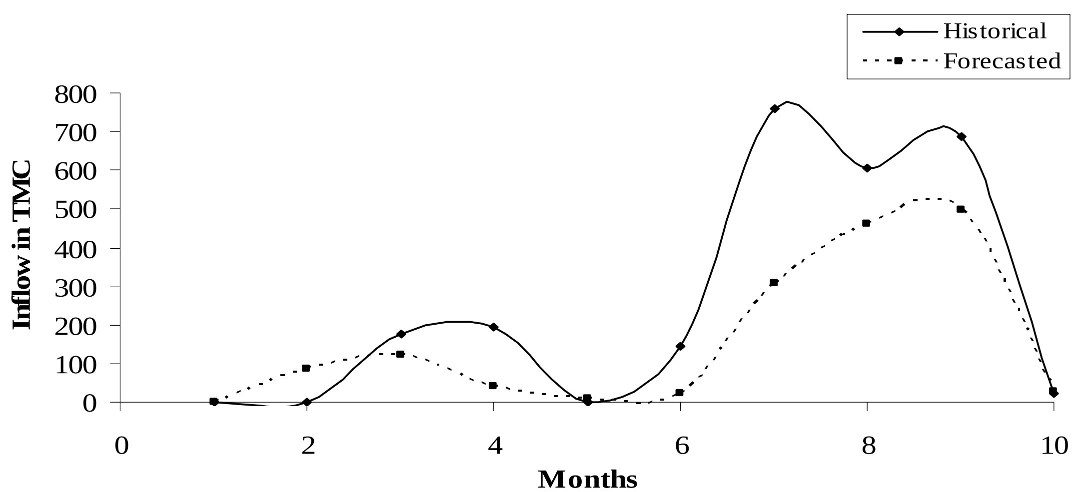

Fig. 4. 20 Comparision of historical and forecasted series

In order to further evaluate the performance of the model the mean forecast error, mean absolute error, root mean square error and Integral Square were computed using Equation (3.63) to Equation (3.66) respectively. The values of the error are given in Table (4.17).

Table 4.17 Performance of forecasted model:

	Mean Forecast Error	**Mean Absolute Error**	**Root Mean Square Error**	**Integral Square Error**
ARMA(3,1)	**-101.64**	**120.889**	**175.727**	**119.229**

The low values of error which is about 20 per cent of the mean monthly flow (544.97 TMC), suggest the adequacy use of the model for real time forecasting of monthly inflow series of Mahi river.

Considering the over-all process of model building and comparing the results obtained from forecasting by all the six models, the Mahi river inflow can be best forecasted by ARMA(3,1) model which may be used for water resource planning by making one-time step ahead forecast of monthly river inflow.

5. SUMMARY AND CONCLUSIONS

In recent years, stochastic modelling and time series analysis have gained tremendous importance in hydrological studies. This is because primary hydrological variables like rainfall; river inflows etc. are essentially stochastic in nature and are not amenable to solution by the classical statistical methods. The main objective of studying hydrologic time series is to understand the mechanism that generates the data so that the future sequence may be simulated or to forecast the future events over a short period of time (forecasting). These are attempted by making Inferences regarding the underlying laws of the stochastic process from the historical data and then by postulating a model that fits the data which can in turn be used for simulating/forecasting the future values. A mathematical model of monthly river inflow is useful for generation of the synthetic data and for single or multiple months ahead forecasting. The synthetic data sequences are needed in simulation studies for the design and operation of water storage, conveyance and control structures.

One or multiple months ahead forecasts are needed for the operation of reservoirs, agricultural planning and watershed management practices. Various mathematical models for generating the river inflows exist such as autoregressive (AR), moving average (MA), autoregressive moving average (ARMA) and autoregressive integrated moving average (ARIMA) models. Such models have an important place in the stochastic modelling of hydrologic data and are of value in handling the short-run problem of modelling the seasonal variability in a stochastic river inflow series.

The objective of this study was to formulate an appropriate mathematical model, based on stochastic concept, which may be used to describe the time dependence structure of monthly inflow series of Mahi river of Southern Rajasthan. The inflow data for 76 years i.e. from 1928 to 2003 for Mahi river at Mahi Bajaj Sagar Dam site has been collected from the office of the Chief Engineer, Mahi Bajaj Sagar Project, Banswara. River Mahi river is a monsoonal river. The total inflow is occurring only during wet season (from June to October). It is assumed in this study that flow during **dry** season is **zero** and hence only five month series was used for analysis and model building. The annual inflow series were tested for its randomness. Fourier analysis was done to detect periodic component. After removal of periodic component stochastic models were fitted to the monthly inflow series of Mahi river. Thereafter fitted models were validated and regeneration of data was done for first 74 years data series. Actual series and generated series were compared for selection of most appropriate model. From the most appropriate model forecasting was done for remaining two years i.e. 75th and 76th year. The results of the study are summarized as follows:

1) Mean monthly Mahi river inflow varies from 110.15 TMC in the month of June to 1028.02 TMC in the month of August. The maximum inflow occurs during July, August, and September. The sample standard deviation varies from 97.58 TMC in the month of October to 760.42 TMC in the month of September.
2) Mahi river is a monsoonal river having annual inflow series as a random series.
3) Periodic component can be represented by Fourier analysis. No. of significant harmonics was found to only one.
4) Monthly stochastic inflow series can be represented by linear stationary stochastic models after suitably standardizing and normalizing the series.
5) Out of twelve models identified representing the Mahi river inflows, only six has parameter value greater than their standard error, which are used for further studies.
6) All the six models passed the validation tests.
7) After passing the validation test the all the six models were used for regeneration of Mahi river inflow series. It is observed that seasonal monsoonal pattern of inflow series is maintained in generated values by all the six models.
8) The identified models underestimate the mean monthly inflow of Mahi river from 12.68 % to 31.62 %. Other statistical characteristics were not found to be restoring in regenerated series.
9) The MMSE criterion is used for selecting most appropriate model. The most appropriate model found was ARMA(3,1). The model structure of ARMA(3,1) model is

$$X_t = 544.97 - 406.99 \text{ Cos}\left(\frac{2\pi .t}{5}\right) - 323.18 \text{ Sin}\left(\frac{2\pi\, t}{5}\right) + 1.2262880Z_{t-1} - 0.1499627\, Z_{t-2} - 0.0766103Z_{t-3} - 0.9704864\, a_{t-1} + a_t$$

10) With best appropriate model i.e. ARMA(3,1) forecasting was done for the year 2002 and 2003.
11) Values of correlation coefficient between observed and forecasted monthly inflow series for the two year period (2002-2003) was found to be 0.912 showing adequacy of developed model.

REFERENCES

Barndorff-Neilsen, O. and G. Schou. 1973. On the parameterization of autoregressive models by partial auto correlations. Journal of Multivariate Analysis 3: 408-419.

Benalaya, A.; F. De-Troch, H. Sebei, H. Wheater and C. Kisby. 1998. Characterization and stochastic modelling of monthly rainfall time series. Hydrology in a changing environment. Vol I. Proceedings of the British. Hydrological Society International Conference, Exeter, UK, July 1998. 53: 61.

Bender, M. and Stohodan Simonovle. 1994. Time-series modelling for long-range stream flow forecasting. Journal of water Resources Planning and Management, ASCE, 120(6): 857-870.

Bhakar, S.R. 2000. Modelling of evaporation and evapotranspiration under climatic conditions of Udaipur. Pd.D. Thesis submitted to MPUAT, Udaipur.

Bidwell, V.J. and G.A. Griffitsch. 1994. Adaptive flood forecasting; an application to the Waimakariri river, J. of Hydrol., New Zealand, 32: 1-15.

Box, G.E.P. and G.M. Jen Kins. 1976. Time Series Analysis, forecasting and control. Holden-Day, San Francisco, Calif: 553 pp.

Chebaane Mohamed and J.D. Salas. 1995. Product periodic autoregressive process for modelling intermittent monthly streamflows, Water Resour. Res. 31: 1513-1518.

Chen, H.L. and A.A. Rao. 2002. Testing hydrologic time series for stationary. Journal Hydrologic Engineering, 7(2): 129-136.

Clarke, R.T. 1973. Mathematical Models in Hydrology, FAO Irrigation and Drainage Paper 19, U.N.Food and Agriculture Organisation, Rome. pp 282.

Fernadez, B. and Pizarro, G. 1996. Statistical estimation of runoff characteristics of watersheds in Central Chile. Hydrologic Science, 41: 735-749.

Gorantiwar S.D. and M. Majumdar and P.S. Pampattiwar. 1995. Application of autoregressive models of different orders to annual streamflows of Barkar river with their logarithmic transformation. Journal of Applied Hydrology, 8:33-39.

Hipel, K.W., A.I. Mcleod and W.C. Lennox. 1977. Advances in Box Jenkins modelling: 1. Model construction. Water Resour. Res. 13: 567-575.

Hipel, K.W., and A.I. Mcleod. 1994. Time series modelling of water resources and environmental systems. Development in Water Science, 45, Elsevier, New York.

Hirsch, R.M. 1979.Synthetic hydrology and water supply reliability. Water Resour. Res. 15: 1603-1615.

Jenkins, Gwilym M. and Donald G. Watts. 1968. Spectral analysis and its applications, Holdend-day Inc. San Fran cisco, Calofornia: 525.

Jha, V. 2002. Modelling of soil moisture. M.E. Thesis submitted to MPUAT, Udaipur.

Jochanan Bonne. 1971. Stochastic simulation of monthly streamflow by a multiple regression model utilizing precipitation data. J. Hydrol. 10: 285-310.

Kottegoda, N.T. 1980. Stochastic Water Resources Technology. The Macmillan Press Ltd., London: pp. 384.

Koutsoyiannis, D. 2001. Coupling stochastic models of different time scale. Water Resour. Res., 37(2): 379-391.

Kumar, A. 2003. Modelling of Okra evapotranspiration under climatic conditions of Udaipur. M.E. Thesis submitted to MPUAT, Udaipur.

Lefebvre, M. 2002. Using a lognormal diffusion process to forecast river flows. Water Resour. Res. 38(6):

McCuen Richard H. 1973. The role of sensitivity analysis in hydrologic modelling. J. Hydro. 18: 37-53.

Mcleod, A.T., K.W. Hipel and W.C. Lennox. 1977. Advances in Box – Jenkins modeling 2. Applications. Water Resour. Res., 13:557-586.

Montanari Alberto and Renzo Rosso. 1997. Fractionally differenced ARIMA models applied to hydrologic time series. Identification, estimation and simulation, Water Resour. Res., 33: 1035-1044.

Montanari, A.; R. Rosoo and M. S. Taqqu. 2000. A seasonal fractional ARIMA model applied to the Nile River monthly flows at Aswan. Water Resour. Res., 36(5): 1249-1259.

Mujumdar, P.P. and D. Nagesh Kumar. 1990. Stochastic models of streamflow; some case studies. Hydrological Sciences Journal, 35: 395- 410.

Narulkar, Sandeep M. 1995. Optimal real time operation of multireservoir systems for irrigation scheduling. Ph.D. Thesis submitted to I.I.T., Bombay.

Pandey, P.K. (2002). Modelling of Blackgram evapotranspiration under climatic conditions of Udaipur. M.E. Thesis submitted to MPUAT, Udaipur.

Patil, R.M. 2003. Stochastic modelling of water deficit for Rahuri region. M.E. thesis submitted to MPUAT, Udaipur.

Phoon, K.K.; M.N. Islam; C.Y. Liaw and S.Y.Liong. 2002. Practical inverse approach for forecasting non linear hydrological time series. Journal Hydrologic Engineering, 7(2): 116-128.

Quenouille, M.H. 1949. Approximate test of correlation in time series. Journal of Royal Statistical Society, Series B, 11: 68:84.

Raghuwanshi, N.S. and W.W. Wallender. 2000. Forecasting daily evapotranspiration for a grass reference crop. Agricultural Engineering Journal, 9(1): 1-16.

Raman, H.; S. Mohan and P. Padalinathan.1995. Models for extending stream flow data: a case study. Hydrologic Sci. Journal-des Science Hydrologiques, 40(3): 381-393.

Reddy, K.M. and D. Kumar, 1999. Time series analysis of monthly rainfall for Bino watershed of Ramganga river. Journal of Agriculture Engineering ISAE, 36(4): 19-29.

Salas, J.D. and J.T.B. Obeysekera. 1982. ARMA model identification of hydrologic time series. Water Resour. Res.. 18: 1011-1070.

Salas, J.D., J.W. Delleur, V.Yevjevich and W.L. Lane. 1980. Applied modeling of hydrologic time series. Water Resources Publications, Litteton, Colorado; p. 484.

Samani, N.,E. Raeissi and A.R. Soltani. 1995. Stochstic synthesis of draughts for reservoir storage design. Journal of Applied Hydrology, 8: 9-18.

Sharma, A.; R.V.Singh; S.M. Narulkar and P.K. Dashora. 2003. Identification of Appropriate stochastic model for forecasting Monsoonal river inflows. Journal of Agricultural Engineering 12(3-4): 191-208.

Sharma, Arun Kumar. 1998. Stochastic modelling for forecasting Jakham river inflows. Pd.D. Thesis submitted to MPUAT, Udaipur.

Sharma, Ashish, David, G. Tarboton and Upmanu Lall. 1997. Streamflow simulation; A nonparametric approach, Water Resour. Res., 33: p. 291-308.

Sharma, T.C. 1978. A stochastic model of monthly riverflow. Journal of Agriculture Engineering, ISAE. 15: 98-106.

Singh, C.V. 1998. Long-term estimation of monsoon rainfall using stochastic models. International Journal Climatology, 18(4): 1611-1624.

Srinivasan, K. 1995. Stochastic modelling of monsoon river flows. Journal of Applied Hydrology. 8: 51-57.

Srinivasan, K. and B.S. Thandaveswara. 1995. Stochastic modelling of monsoon dependent flows of south Indian rivers. J. Institution of Engineer (India), 75: 220-227.

Tao, P.C. and J.W. Delleur. 1976. Multistation, multiyear synthesis of hydrologic time series by disaggregation. Water Resour. Res. 12:1303-1312.

Vecchia, A.V. 1985. Periodic autoregressive moving average (PARMA) modelling with applications to water resources. Water Resour. Bull. 21:721-730.

Vecchia, A.V., J.T.B. Obeysekera, J.D. Salas and D.C. Boes. 1983. Aggregation and estimation for low- order periodic ARMA models. Water Resour. Res. 19: 1279-1306.

Verma, A. 2004. Stochastic modelling on monthly rainfall of Kota, Rajasthan. M.E. Thesis Submitted to GBPUAT, Pantnagar.

Yevjevich, V. M. 1972. Stochastic Processes in Hydrology. Water resources Publication, Fort Collins, Colorado, p.277.