Training Course

On

Predictions in Ungauged Basins

[July 26 – 30, 2010]

CHAPTER-11

STOCHASTIC MODELLING AND STREAM FLOW FORECASTING

By

Sanjay Kumar,
Scientist-C

Organized under Hydrology Project - II

National Institute of Hydrology
Roorkee – 247667 (Uttrakhand)

CHAPTER-11

STOCHASTIC MODELLING AND STREAM FLOW FORECASTING

Dr. Sanjay Kumar,
Scientist-C
National Institute of Hydrology, Roorkee-247667

11.0 INTRODUCTION

The hydrological forecast involves the estimation of the future state of the hydrological variables in space and time (WMO/UNESCO, 1991). Many times it is in real time and for a particular location. River flow forecasting is one of the most important hydrological forecast besides other forecasts such as those for avalanches, soil moisture and groundwater levels, together with forecasts for navigable rivers and lakes. In addition there are also the more specialized forecasts (WMO 1994) for water quality, including the travel time of slugs of pollutants, for dam-breaks, and like phenomena. Hydrological forecasts are valuable for various purposes such as river regulation, hydropower production, irrigation scheduling, water supply and also for managing and mitigating the flood disasters.

A hydrological forecast needs to be accurate and timely to be effective. These two features are dependent on a number of factors; primarily the availability of hydrological and meteorological data, the speed of receipt of these data at the forecast centre, the effectiveness of the forecasting method employed and the time taken to disseminate the forecast to users. In addition the knowledge of the characteristics of a particular river basin is also important to a forecast.

Most currently used methods of flow forecasting are based on one or more types of hydrological model. Such a model attempts to describe the quantitative relations between the precipitation, the other inputs and the outputs, primarily the flow. At one end of the spectrum (WMO 1994) these models are simple empirical 'black boxes', which do not try to mimic the physical processes involved. At the other end are those conceptual models, which aim to recreate the physical processes and the concepts about the behavior of a basin in a series of mathematical expressions. In the region between there is a great variety of models; (deterministic, stochastic, lumped, distributed, event driven, continuous), which can all be the basis of a flood forecasting system.

The time series models provide an alternative approach as compared to deterministic or conceptual approach to hydrological forecasting. These models require long times series for their best operation and are considered as 'black box' type. Identification of the components of the structure of the time series is important to these methods; namely the trend, the periodicity and the random component. A large number of time series models have been developed which use these and other methods of time series analysis (Salas, 1992).

This lecture covers the principles of time series analysis and various steps involved in the identification of time series models.

11.1 TIME SERIES ANALYSIS IN HYDROLOGICAL FORECASTING

A time series is a collection of observations generated sequentially through time. The special features of a time series are that the data are ordered with respect to time, and that successive observations are usually expected to be dependent. The methodology adopted to study the history of movement in time of the variables, e.g. hourly discharge, daily rainfall etc., is called Time Series Analysis (TSA). Prerequisites for application are observed data sequences. The analysis proceeds with the various components of a time series. In the following sections a general overview of TSA, the terminology used in TSA, and brief description of Box-Jenkins (B-J) modelling approach for univariate series has been provided.

In contrast to other areas of statistics, the characteristic feature of TSA is that the observations occur in the temporal order. The implication is that we shall, among other things, be interested in the relationship between the values from one time to the next; that is in the serial correlations along the series, when we come to consider several series, it becomes necessary to consider not only correlations between series, but also the serial correlations within each series. Furthermore, the extent to which one series leads, or lags, another becomes an important part of statistical modelling.

Thus, the primary objective underlying in studying time series is to understand the mechanism that generates the data and also, but not necessarily, to produce likely future sequences or to forecast events over a short period of time. These are attempted by making inferences regarding the underlying laws of the stochastic process from one or more sequences of recorded observations and then by postulating model that fits the data, which are again used for estimation purposes. The objectives of time series analysis are as follows:

(i) To obtain a concise description of the features of a particular time series process.

(ii) To construct a model to explain the time series behaviour in terms of other variables and to relate the observations to some structural rules of behaviour.

(iii) Based on the results of (i) and (ii), to use the analysis to forecast the behaviour of the series in the future based upon the knowledge of the past.

(iv) To control the process generating the series by examining what might happen when some model parameters are altered, or by establishing policies that intervene only when the process deviates from a target by more than a prescribed amount.

Definitions of the terms used in the TSA have been presented in the next section in alphabetical order. Mathematical formulae and physical interpretation have been given wherever necessary.

11.2 Definition of Terms

Autocorrelation function (ACF): It is an important guide to the properties of a time series and measures the correlation between observations at different distances (lags) apart. These coefficients help in identifying the probability model that generated the data. It is defined as

$$\rho_k = \frac{E[(x_t - \mu)(x_{t+k} - \mu)]}{E[(x_t - \mu)^2]}$$

The set of values ρ_k and the plot of ρ_k against k = 1, 2, ..., are known the autocorrelation function (ACF) or correllgram. In practice, however, the autocorrlation coefficients are usually calculated by computing the series of autocovariance coefficients, $\{c_k\}$, from the observed time series. Thus, the sample estimate of γ_k is

$$c_k = \frac{1}{n}\sum_{t=1}^{n-k}(x_t - \bar{x})(x_{t+k} - \bar{x})$$

and the estimated or sample ACF (SACF) is

$$r_k = \frac{c_k}{c_o}.$$

The ACF has following three important properties.

Property 1: The ACF is an even function. That is

$$\rho_k = \rho_{-k}$$

Property 2: $|\rho_k| \leq 1$

Property 3: Lack of uniqueness.

To use the autocorrelations for identifying models, it is necessary to know when ρ_k is effectively 0, or is significantly different from 0. For this purpose the standard error of the sample autocorrelations is required. For lags k greater than some value q beyond which the theoretical ACF may be deemed to have died out, Bartlett (1946, 1966) has shown that an approximate estimate of the autocorrelation is given by:

$$\text{var}(r_k) = \frac{1}{n}\left[1 + 2\sum_{i=1}^{q}\rho_i^2\right], \qquad k > q$$

However, in practice population autocorrelations are replaced with the estimated autocorrelations r_k to obtain an estimate of the approximate variance of r_k as

$$\text{var}(r_k) = \frac{1}{n}\left[1 + 2\sum_{i=1}^{q}r_i^2\right], \qquad k > q$$

The square root of above equation is called the *large sample standard error*, SE (r_k) of r_k. For purely random series, it can be shown (Kendall et al., 1983) that

$$E(r_k) \approx -\frac{1}{n}$$

$$\text{var}(r_k) \approx \frac{1}{n}$$

and that r_k is asymptotically normally distributed under weak conditions. Thus, having plotted the correllogram, approximate 95% confidence limits at $-1/n \pm 2/\sqrt{n}$ can be plotted, which often approximated as $\pm 2/\sqrt{n}$. Observed values of r_k which fall outside these limits are significantly different from zero at the 5% level.

Cross-correlation function (CCF): The cross-correlation function between two time series x_t and y_t is given by

$$CCF(k) = \frac{Cov(y_t, x_{t+k})}{\sqrt{\text{var}(y_t) \cdot \text{var}(x_2)}}$$

and is a measure of the relationship is affected by the within series correlations, appropriate transformations must be performed in both series before the CCF is estimated.

Partial autocorrelation function (PACF): The partial autocorrelation function at lag k is defined as the correlation between time series terms k lags apart, after the correlation due to intermediate terms has been removed.

The lag k partial autocorrelation is the partial regression coefficient ϕ_{kk} in the kth order autoregression.

$$x_t = \phi_{k1} x_{t-1} + \phi_{k2} x_{t-2} + \ldots + \phi_{kk} x_{t-k} + a_t$$

and it measures the additional correlation between x_t and x_{t-k} after adjustments have been made for the intermediate variables x_{t-1}, x_{t-2}, …, x_{t-k+1}. The ϕ_{kk} is obtained from Yule-Walker equations (Mills, 1991). However, the sample PACF (SPACF) is calculated by fitting AR models of increasing order: the estimate of the last coefficient in each model is the sample partial correlation. Box and Jenkins (1976) present a recursive method for calculating SPACF, which is attributed to Durbin (1960).

Qvenouille (1949) showed that the variance of the estimate of the partial autocorrelations, $\hat{\phi}_{kk}$ is approximately equal to

$$\text{var}(\hat{\phi}_{kk}) = \frac{1}{\sqrt{n}}$$

where, n equals the number of observations after suitable differencing, and p represents the first partial autocorrelations that are assumed to be non-zero.

The approximate standard errors are equal to

$$SE(\hat{\phi}_{kk}) = \frac{1}{\sqrt{n}}$$

Thus, the values of SPACF which lie within the confidence interval of $\pm 2/\sqrt{n}$ are said to significantly not different from 0 at 95% significance level.

***Prewhitening*:** Prewhitening refers to the process of removing the dominant autocorrelation structures from the dependent and the independent series. This is done to avoid spurious correlation that might exist between the two series due to high "within" autocorrelation. It amounts to first obtaining the appropriate univariate models for each series involved, and then, at the second stage, cross correlating the (residual) white noise series. The process is known as "prewhitening" since it produces a series closer to "white-noise".

Stationarity: Most of the probability theory of time series is concerned with *stationary* time series, which is, broadly speaking, a series which contains no systematic change in mean (no trend), variance and periodic variations; and for this reason TSA requires to turn a non-stationary series into a stationary one so as to use this theory.

A time series is *weakly, second-order or covariance stationary* if

(i) $\mu(t) = \mu$ *and* $\gamma(t) = \gamma_o$ *for all t*

(ii) $\gamma(t, t+k) = \gamma_k$ *for all t and k*

Condition (ii) implies that two observations, k time periods apart, have the same covariance no matter where they occur in the series. Further, conditions (i) and (ii) imply that we can define *autocorrelation* between x_t and x_{t+k} as

$$\rho_k = corr(x_t, x_{t+k}) = \frac{Cov(x_t, x_{t+k})}{[\text{var}(x_t)\text{var}(x_{t+k})]^{1/2}} = \frac{\lambda_k}{\lambda_o}$$

Since these conditions apply only to the first- and second-order moments of the process, this is known as *second-order stationarity.*

A time series is said to be strongly, strictly, stationary if the joint density functions depend only upon the relative locations of the observations. That is if $f\{x(t_1), x(t_2), \ldots, x(t_n)\}$ denotes the joint density of the observations at times t_1, t_2, …, t_n, we require that if $f\{x(t_1+k), x(t_2+k), \ldots, x(t_n+k)\} = f\{x(t_1), x(t_2), \ldots, x(t_n)\}$ for all choices of the $\{t_i\}$. In particular, this means that

$f\{x(t)\} = f\{x(t+k)\}$

$f\{x(t_1), x(t_2)\} = f\{x(t_1+k), x(t_2+k)\}$ *and so on.*

Usually for TSA the *less restrictive* definition of stationarity is considered.

Stationarity and Invertibility Conditions: Stationarity and invertibility conditions impose restrictions on the parameters of the AR and MA processes respectively and ensure variance stability and existence of unique MA process for a given ACF. These two conditions provide a diagnostic tool to check stationarity of the fitted model. If the model fails to fulfill these two conditions, it implies that some transformation, such as additional differencing, is required to induce stationarity.

White Noise: White noise is a sequence of random shocks drawn from a fixed distribution, with zero mean and constant variance. For a white noise (or random series) $\rho_k = 0$, $k > 0$ and is approximately N (0,1/n), so that, if a time series is random, 19 out of 20 of the values of r_k can be expected to lie between $\pm 2/\sqrt{n}$, at 95% significance level.

11.3 Time Series Notation

Difference Operator

If $\{x_t\}$ represents the raw data or the series obtained after making a transformation to stabilise the variance, the regular and the seasonal difference of first-order is then defined respectively as

$$\nabla x_t = x_t - x_{t-1} \qquad \textit{and} \qquad \nabla_s x_t = x_t - x_{t-s}$$

where, s is the span of the seasonal cycle.

In general, combinations of seasonal and consecutive differencing are written as

$$\nabla_s^D \nabla^d x_t,$$

where D represents the order of seasonal difference operator, s the span, and d the order of the consecutive difference operator.

Backward Shift Operator

It is defined as

$$Bx_t = x_{t-1} \qquad \textit{and} \qquad B^s x_t = x_{t-s}$$

In general, $B^k x_t = x_{t-1}$. Therefore, the relationship between ∇ and B can be expressed as

$$\nabla = 1 - B$$

Autoregressive Operator Polynomial

The regular and seasonal AR operator polynomial order p and P are defined respectively as

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

$$\Theta_P(B) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \ldots - \Phi_P B^{Ps}$$

Moving Average Operator Polynomial

The regular and seasonal MA operators of order q and Q are defined respectively as

$\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q$ and

$\Theta_Q(B) = 1 - \Theta_1 B^s - \Theta_2 B^{2s} - \ldots - \Theta_Q B^{Qs}$

General Multiplicative Seasonal Model

The broad class of Box and Jenkins (B-J) models called the multiplicative ARIMA models can be expressed in the form of the operators defined above as

$$\phi_p(B)\Phi_p(B^s)\nabla_s^D\nabla^d x_t = \theta_q(B)\Theta(B^s)a_t$$

General Transfer Function Model

Similarly the general class of TFN models can also be expressed in terms of the above defined operators as

$$x_t = \frac{\omega(B)\Omega(B^s)}{\delta(B)\Delta(B^s)} z_{t-b} + \frac{\theta(B)\Theta(B^s)}{\phi(B)\Phi(B^s)} a_t$$

with

$$x_t = \nabla_s^{D'}\nabla^{d'} Y_t$$

$$z_t = \nabla_s^{D}\nabla^{d} X_t$$

11.4 Box-Jenkins Models

Box-Jenkins time series models can be classified according to the number of variables included in the model. A time series model consisting of just one variable is called univariate time series model. A univariate time series will use only current and past values on one variable. For example if we forecast the water table next month or two months from now using a univariate model, we could use only the current and past water table data. The assumption in the formation of such model is that the factors, which influence the water table, have not changed or are not expected to change sufficiently. A time series model, which uses other variables to describe the behavior of the desired series, is called a multiple time series model. The model expressing the dynamic relationship between these variables is called a transfer function model. The usefulness of transfer function can be better appreciated if we consider an example. Suppose that beside water table fluctuation data, we also have rainfall and groundwater withdrawl data for the corresponding period. Then, by constructing a transfer function model, we could exploit the dynamic relationship between water table fluctuations, rainfall and groundwater withdrawl. For instance, if we discover that changes in rainfall and groundwater withdrawl this month (or any other time scale) will trigger a response in groundwater table two months from now, we may then be in a much better position than if we had just used a univariate model to predict future water table conditions.

Finally, a special form of transfer function model is called intervention model. The special characteristic of such a model is not the number of variables in the model, but one of the explanatory variables captures the effect of an intervention, a policy change, or a new law.

In the following sections only univariate Box-Jenkins time series models are described. The models involve non-seasonal and seasonal univariate models. For more detail on transfer function and intervention models one may refer to Chatfield (1996), Harvey (1993), Kendall and Ord (1990), Kendall et al. (1983), Box et al. (1994), Mills (1991), Vandaele (1983), etc. Chatfield (1996) has given a very good and frugal review of some relevant books on time series.

11.4.1 Autoregressive (AR) models

A time series is said to be governed by a first-order autoregressive process if the current value of the time series, x_t , can be expressed as linear function of the previous value of the series and a random shock a_t. If we denote the previous value of the series by x_{t-1} , this process then can be written as

$$x_t = \phi_1 x_{t-1} + a_t \qquad (1)$$

where ϕ_1 is the autoregressive parameter which describes the effect of a unit change in x_{t-1} on x_t and which needs to be estimated. The equation (1) is called an autoregressive process of order 1, AR(1). The order of the process corresponds to the number of parameter that need to be estimated. An autoregressive process of order p, AR(p), contains p number of parameters and can be expressed as

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \ldots\ldots\ldots\ldots + x_{t-p} + a_t \qquad (2)$$

The random shocks a_t also known as errors or white noise series, are assumed to be normally and independently distributed with mean zero, constant variance σ_a^2 and independent of x_{t-1} i.e

$$\begin{aligned} E(a_t) &= 0 \\ E(a_t a_s) &= \sigma_a^2 \quad \text{if} \quad t = s \\ E(a_t a_s) &= 0 \quad \text{if} \quad t \neq s \\ E(a_t, x_{t-1}) &= 0 \end{aligned} \qquad (3)$$

Fig. 1 shows a schematic representation of an auto regressive process.

x_{t-1}

x_t a_t

Fig.1 Schematic representation of an Autoregressive Operator

The stationarity condition introduces restrictions on the parameter values (ϕ's) of autoregressive process. It can be shown that for a stationary first order autoregressive process, the absolute value of parameter ϕ_1 must be less than 1 i.e $|\phi_1| < 1$. There are certain properties based on autocovariance and autocorrelation, which can be used to distinguish an AR process from any other process. It can be shown that autocovariance of an AR(1) process at lag k is

$$\lambda_k = \phi_1^k \lambda_0 \tag{4}$$

Similarly it can be shown that autocorrelation of an AR(1) process at lag k is

$$\rho_k = \frac{\lambda_k}{\lambda_0} = \phi_1^k \tag{5}$$

Another distinguish feature of the autoregressive process is its long memory i.e current observation x_t is influenced by shocks a_t that occurred in distant past. For an AR(1), stationary process ($|\phi_1| < 1$), the effect of shocks will gradually dissipate, whereas this not the case if the process is nonstationary. Stationarity conditions, as well as autocovarience, autocorrelation and memory function characteristics can be derived for an autoregressive process of any order however, they are not as simple as for an AR(1) process but complicated function of ϕ_i's parameter.

11.4.2 Moving Average (MA) model

The autoregressive model of first order AR(1) in equation (1) would be extended to include past errors to see if they can improve on the time series representation of the data i.e. the AR(1) model in equation (1) could be modified as

$$x_t = \phi_1 x_{t-1} + a_t - \theta_1 a_{t-1} \quad (6)$$

Where a_{t-1} represent the error at period t-1, and θ_1 is called the moving average parameter, which describe the effect of the past error on x_t. A Moving Average (MA) model is obtained from equation (6) by omitting the lagged variable x_{t-1}. Mathematically, we express a first order moving average MA(1) model as

$$x_t = a_t - \theta_1 a_{t-1} \quad (7)$$

Where θ_1 is the moving average parameter. The equation (7) expresses the current value of the series x_t as a linear function of the current and previous errors or shocks, a_t and a_{t-1}. The equation (7) is called moving average process of order 1, MA(1). The order of the process corresponds to the number of parameter that needs to be estimated. A moving average process of order q, MA(q), contains q number of parameters and can be expressed as

$$x_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \ldots\ldots\ldots - \theta_q a_{t-q} \quad (8)$$

The random shocks, a's, in a moving average process are assumed to be normally and independently distributed with mean zero and constant variance σ_a^2 i.e they satisfy equation (3). Fig. 2 shows a schematic representation of a moving average process.

x_t a_t, a_{t-1}

Fig. 2: Schematic representation of an Moving Average (MA) operator

The stationarity condition introduces no restrictions on the parameter values (θ's) of moving average process. However for a 1 order moving average process, MA(1), the parameter θ_1 must satisfy $|\theta_1| < 1$. This condition is called the invertibility condition of an MA(1) process. There are properties based on autocovariance and autocorrelation, which can be used to distinguish an MA process from any other process. It can be shown that autocovariance of an MA(1) process at lag 0, 1 and $k \geq 2$ can be expressed as

$\lambda_0 = (1+\theta_1^2)\sigma_a^2$

$\lambda_1 = -\theta_1\sigma_1^2$, and

$\lambda_k = 0 \quad \text{for } k \geq 2$

Therefore, the autocorrelations can be defined as

$$\rho_1 = -\frac{\theta_1}{(1+\theta_1^2)}, \text{ and}$$

$\rho_k = 0 \quad \text{for } k \geq 2$

It can be noticed that whereas the autocorrelations for AR(1) process die out gradually, they die out abruptly for an MA(1) process. It follows directly from equation (7) that the memory of an MA(1) process lasts only for one period i.e. the system will feel the full impact of a_t at time t, but its effect at time t+1 will be proportional to θ_1.

Invertability condition, as well as autocovarience, autocorrelation and memory function characteristics can be derived for a moving average process of any order however, they are not as simple as for an MA(1) process but complicated function of θ_i's parameter. A parsimonious representation of model is possible using a moving average process wherein the number of parameters to be estimated can be drastically reduced. Vandaele (1983), demonstrate the equivalence between an MA(1) process and an AR(∞) process (an autoregressive model with an infinite number of parameter). It is important to represent a time series process with as few parameters as possible as the parameters have to be estimated with a finite number of data points.

11.4.3 Mixed Autoregressive Moving Average (ARMA) models

A process with an autoregressive and a moving average term is called (mixed) autoregressive moving average process and is represented by equation (6). This may reproduced as in equation (9)

$$x_t = \phi_1 x_{t-1} + a_t - \theta_1 a_{t-1} \qquad (9)$$

The autoregressive moving average models are denoted as ARMA(p q) and represented as in Fig.3. The p refers to the number of autoregressive parameters, and the q to the number of moving average parameters. Model as in equation (9) is an ARMA(1 1) process. The higher order ARMA(p, q) can be expressed as

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \ldots\ldots + \phi_p x_{t-p} + a_t - \theta_1 a_{t-1} - \ldots\ldots\ldots - \theta_q a_{t-q} \qquad (10)$$

x_{t-1}

x_t a_t $a_t\, a_{t-1}$

Fig. 3: Schematic representation of an (ARMA) operator

The stationarity condition introduces restrictions on the parameter values (ϕ 's & θ 's) of ARMA(p, q) process. However for a ARMA(1, 1) process the parameter ϕ_1 and θ_1 must satisfy $|\phi_1| < 1$ and $|\theta_1| < 1$ respectively. The autocovariance and autocorrelation, of an ARMA(1, 1) process at lag 0, 1 and for $k \geq 2$ can be expressed as

$$\lambda_0 = \frac{1+\theta_1^2 - 2\theta_1\phi_1}{1-\phi_1^2}\sigma_a^2$$

$$\lambda_1 = \frac{(1-\phi_1\theta_1)(\phi_1-\theta_1)}{1-\phi_1^2}\sigma_1^2, \text{ and}$$

$$\lambda_k = \phi_1\, \lambda_{k-1} \quad \text{for } k \geq 2$$

Therefore, the autocorrelations can be defined as

$$\rho_1 = \frac{(1-\phi_1\theta_1)(\phi_1-\theta_1)}{(1+\theta_1^2 - 2\theta_1\phi_1)}, \text{ and}$$

$$\rho_k = \phi_1\, \rho_{k-1} \quad \text{for } k \geq 2$$

It can be noticed that the autocorrelations for ARMA(1, 1) process die out gradually as for a stationary AR(1) process. It is therefore difficult to distinguish between an ARMA(1, 1) process and an AR(1) process based on these properties. However the first autocorrelation of ARMA(1, 1) process is influenced by both the AR and the MA parameters.

11.4.4 Autoregressive Integrated Moving Average (ARIMA) model

Any of the models discussed so far cannot be used if the series is not stationary. Transformation of the data will be necessary to induce stationarity. It is shown that for an AR(1) model to be stationary, ϕ_1 must be less than 1. A special model is obtained if $\phi_1 = 1$. In this case, the AR(1) process in equation (1) can be written as

$$x_t = x_{t-1} + a_t$$

or as $$x_t - x_{t-1} = a_t \qquad (11)$$

The equation (11) is known as a random walk model, a model in which changes are brought about by a white noise series. The memory function of the random walk model demonstrates that a random walk process is a stochastic process wherein successive random shocks accumulate over time. This means that the effect of a shock incurred at time t will never die out which is a general characteristic of a nonstationary series. The equation (11) suggest a simple way to model the random walk. The first difference of the series, $x_t - x_{t-1}$ constitutes a stationary series whose observations are random shocks a_1 , a_2 , ... a_t . In other words the differencing transformed the random walk into a white noise process. Thus the non stationary series often transformed into stationary series by differencing. Suppose that the first differences of a series are stationary. Then by defining the difference between consecutive values of x_t as

$$w_t = x_t - x_{t-1} \qquad (12)$$

we could replace x_t and x_{t-1} in the mixed autoregressive moving average model, the ARMA(1,1) model with w_t and w_{t-1} to obtain

$$w_t = \phi_1 w_{t-1} + a_t - \theta_1 a_{t-1} \qquad (13)$$

Modeling of x_t expressed by equation (12) and (13) is called an autoregressive integrated moving average (ARIMA) model. This model is labeled an ARIMA(1 1 1) model. In general, autoregressive integrated moving average models are denoted as ARIMA(p d q). The number inside the parentheses refer to the order of the autoregressive process, the degree of differencing required to induce the stationarity, and the order of the moving average process respectively. It can be shown that the actual realizations, x_t can be written as an infinite sum of past and present differences transformation i.e.

$$x_t = w_t + w_{t-1} + w_{t-2} + \ldots\ldots \qquad (14)$$

The general structure of an ARIMA process can be summerised as in Fig. 4

Fig 4: Schematic representation of an ARIMA operator

The moving average operator transforms the white noise into an intermediate series e_t. For a first-order moving average filter, e_t would be represented as

$$e_t = a_t - \theta_1 a_{t-1} \quad (15)$$

The autoregressive operator transfer e_t into another intermediate series w_t. Again for a first order autoregressive filter, w_t would be expressed as

$$w_t = \phi_1 w_{t-1} + e_t \quad (16)$$

And finally the integration operator transforms w_t to x_t. The above described models are also called ARIMA models. An AR(p) model can always be denoted as an ARIMA(p,0,0) model. An MA(q) model can be represented as an ARIMA(0,0,q) and an ARMA(p,q) model as as ARIMA(p,0,q).

11.5 Seasonal Box-Jenkins Models

Seasonal patterns are often encountered in time series observed at quarterly or monthly intervals. For example with monthly observations, where s (seasonality)=12, it is expected that x_t depend on term such as x_{t-12} or perhaps x_{t-24} as well as terms such as x_{t-1}, x_{t-2}. It is therefore logical, to try to exploit the correlation between the same quarter or months in successive years, as well as the correlation between successive quarters or months, to model such time series. Box-Jenkins (1970) extended and generalized the ARIMA models to analyze seasonal variations. The following sections describe seasonal Box-Jenkins models.

11.5.1 Seasonal Autoregressive (SAR) models

A time series (observed, say, at quarterly intervals) is governed by a first-order seasonal autoregressive process if the current value of the series x_t can be expressed as a linear function of the value of the series attained one year ago, y_{t-s} (s=4), and a random shock, a_t i.e.

$$x_t = \Phi_1 x_{t-s} - a_t \quad (17)$$

Backshift operator can be used to express equation (17) as

$$(1 - \Phi_1 B^s)x_t = a_t \quad (18)$$

Where Φ_1 is the seasonal autoregressive parameter. Equation (18) could be denoted as an SAR(1) model i.e. seasonal autoregressive model of order 1. In general a seasonal autoregressive model of order P can be written as

$$\Phi(B^S)w_t = a_t \tag{19}$$

Where

$$\Phi(B^S) = 1 - \Phi_1 B^S - \Phi_2 B^{2S} - \ldots\ldots - \Phi_P B^{PS} \tag{20}$$

$$w_t = \nabla_S^D \nabla^d x_t \tag{21}$$

The seasonal and consecutive difference operators (∇_S^D , ∇^d) used to induce stationarity in the series x_t. The autocorrelation function of a SAR model is similar to the regular AR model, except that the values of the autocorrelations appear at multiples of the span i.e. the autocorrelations of SAR(1) model have non-zero values at lags that are multiples of the span. That is,

$\rho_{sk} = \Phi_1^k$ where s=spn, $k > 0$

11.5.2 Seasonal Moving Average (SMA) models

A time series is said to be governed by a first-order seasonal moving average process if the current value of the series x_t can be represented by a current shock, a_t and a shock occurring exactly s observation earlier, a_{t-s}, where s equals the span of the seasonal model. Such a model is written as

$$x_t = a_t - \Theta_1 a_{t-s} \tag{22}$$

using backshift operator

$$x_t = (1 - \Theta_1 B^S) a_t$$

Where Θ_1 is called seasonal moving average parameter. The equation (22) represents a first order seasonal moving average, SMA(1) model. A SMA model of order Q can be expressed as

$$w_t = \Theta(B^S) a_t \tag{23}$$

where

$$\Theta(B^S) = 1 - \Theta_1 B^S - \Theta_2 B^{2S} - \ldots\ldots - \Theta_Q B^{QS} \text{ and} \tag{24}$$

$$w_t = \nabla_S^D \nabla^d x_t \tag{25}$$

The seasonal and consecutive difference operators (∇_S^D , ∇^d) used to induce stationarity in the series x_t. The autocorrelation function of a SMA model is similar to the regular MA model, except that the values of the autocorrelations appear at multiples of the span i.e. the autocorrelations of SMA(1) model have only one non-zero values, which will occur at the lag corresponding to span. That is,

$$\rho_S = -\frac{\Theta_1}{(1+\Theta_1^2)} \quad \text{where s=span}$$

11.5.3 Mixed Seasonal (SARMA) models

A process which, is a combination of seasonal autoregressive and seasonal moving average processes is called a mixed seasonal process and can be expressed as

$$\Phi(B^S)w_t = \Theta(B^S)a_t \qquad (26)$$

Where $\Phi(B^S)w_t$ and $\Theta(B^S)a_t$ are expressed in equation (21) and (24) respectively. Similarly the difference operators have been expressed in equation (25). The autocorrelations of a seasonal mixed process have nonzero autocorrelations at lags which are multiple of the span.

11.5.4 General multiplicative seasonal models

If all the models explained so far can be combined into one model then this broad class of models is called multiplicative ARIMA models, and can be expressed as

$$\phi(B)\Phi(B^S)w_t = \theta(B)\Theta(B^S)a_t \qquad (27)$$

where

$\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p$

$\Phi_p(B) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \ldots - \Phi_p B^{Ps}$

$\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q$ and

$w_t = \nabla_S^D \nabla^d x_t$

The equation (4.21) can also be summarized as

ARIMA(p, d, q) x (P, D, Q)

These models when analyzed properly can yield good fits to time series data and generate accurate forecasts.

11.6 Model Building Strategy for the Box-Jenkins Models

The Box-Jenkins iterative approach for constructing linear time series models consists of four steps. These are as follows:

(i) Identification of the preliminary specification of the model

(ii) Estimation of the parameters of the model
(iii) Diagnostic checking of the model
(iv) Forecasting future realizations

In the identification stage, one selects the order of consecutive and seasonal differencing required to make the series stationary, as well as specifies the order of the regular and seasonal autoregressive and moving average polynomial necessary to adequately represent the time series model. After a tentative model has been identified, the parameters of that model are estimated. Various diagnostic checks are then applied to determine whether or not the model adequately represents the data. A new model must be identified if any inadequacies are detected and the cycle of identification, estimation, and diagnostic checking repeated. Finally, the model, which passes all the checks, is used to generate forecasts. It should be noted that models, which adequately depict the behaviour of the data at time, may generate forecasts, which are not at all acceptable. Whenever this occurs, one may have to go back to the identification stage and restart the process.

11.6.1 Identification

The preliminary stage in the model identification is to plot the time series data to detect the incorrect data, keypunch errors and gross outliers etc. and to check whether any transformation is required to make the series stationary. Variance stabilization and mean stabilization are two types of transformations used for imparting stationarity to the time series.

There is no adequate formal test to check the variance non-stationarity (Granger and Newbold, 1977). However, one practical way to test for variance non-stationarity in cases where the standard deviation or the variance seems to be proportional to the local mean of the series is to divide the series into intervals of equal "length", and find the mean and corresponding standard deviation of each segment. If the local mean is found to be proportional to the local standard deviation, the variance stabilising transformation is the *logarithmic* one. If the local mean is found to be proportional to the local variance, the variance stabilising transformation is the *square root transformation* (Mills, 1991, for detailed discussion). Other transformations could also be evaluated to induce stationarity. A general class of such transformations[1], called power transformations, can be specified. However, in practice logarithmic and square root transformations are generally preferred and

[1] Box and Cox (1964) suggested the following transformation:

$$y_t = \begin{cases} \left(x_t^{\lambda} - 1\right)/\lambda & \lambda \neq 0 \\ \log_e x_t & \lambda = 0 \end{cases}$$

where λ is the transformation parameter.

used. For mean stabilization the time series plot and the correlogram help in assessing as to whether the series contains any trend or seasonality. The series is made stationary by proper differencing – regular differencing for trend removal and seasonal differencing to remove seasonality.

Once the series is made stationary, the next step is to identify a general class of tentative model using sample ACF and PACF. Table 1 shows the properties of ACF and PACF for different B-J seasonal and non-seasonal models that help in initial model identification.

Table 1 (a) Properties of the ACF and PACF Non-seasonal B-J Models

Model	ACF	PACF
AR (p)	Tails off. Exponential and/or sine wave decay, may contain damped oscillations	Cuts off after lag p (p spikes)
MA (q)	Cuts off after lag q (q spikes)	Tails off. Dominated by linear combination of damped exponentials and/or sine waves, may contain damped oscillations.
ARMA (p,q)	Tails off after q-p lags. Exponential and/ or sine wave decay after q-p lags	Tails off after p-q lags. Dominated by exponential and/or sine waves after p-q lags.

Table 1 (b) Properties of the ACF and PACF Seasonal B-J Models

Model	ACF	PACF
AR (p), Seasonal AR (P)	Tails off.	Cuts off after lag p + sP.
MA (q), Seasonal MA (Q)	Cuts off after lag q + sQ	Tails off.
Mixed Models	Tails off after (q + sQ) – (p + sP) lags. Exponential and/ or sine wave decay after (q + sQ) – (p + sP) lags.	Tails off after (p + sP) - (q + sQ) lags. Exponential and/ or sine wave decay after (p + sP) - (q + sQ) lags.

11.6.2 Estimation

The next step after identifying a particular model from the general class of multiplicative models is to estimate the vectors of parameters,

$\underline{\phi} = [\phi_1, \phi_2, \ldots, \phi_p]'$, $\underline{\Phi} = [\Phi_1, \Phi_2, \ldots, \Phi_P]'$, $\underline{\theta} = [\theta_1, \theta_2, \ldots, \theta_q]'$ and $\underline{\Theta} = [\Theta_1, \Theta_2, \ldots, \Theta_Q]'$

The parameter estimation can be done using various available algorithm and optimization techniques by minimizing the sum of squared residuals, SSR. That is one chooses $\hat{\underline{\phi}}, \hat{\underline{\Phi}}, \hat{\underline{\theta}}$ *and* $\hat{\underline{\Theta}}$ as estimators of $\underline{\phi}, \underline{\Phi}, \underline{\theta}$ *and* $\underline{\Theta}$ respectively so that SSR,

$$S\left(\hat{\underline{\phi}}, \hat{\underline{\Phi}}, \hat{\underline{\theta}}, \hat{\underline{\Theta}}\right) = \sum_{t=1}^{n} \hat{a}_t^2$$

is a minimum.

11.6.3 Diagnostic Checking

Three groups of tests or diagnostic checks are performed to evaluate the model adequacy. If inadequacies are detected then more likely that a different specification of the model is required to fit the data. The three groups of diagnostic checks are described below.

Stationarity Analysis

The fitted model must be checked for the stationarity and invertibility conditions imposed on the equations of the AR and MA polynomials of order p and q respectively. expressed by replacing B with z and regarding it as a complex variable and not as an operator, i.e.

$$\left.\begin{aligned} \phi(z) &= 1 - \phi_1 z - \phi_2 z^2 - \ldots - \phi_p z^p = 0 \quad \text{and} \\ \theta(z) &= 1 - \theta_1 z - \theta_2 z^2 - \ldots - \theta_q z^q = 0 \end{aligned}\right\} \qquad (28)$$

The stationarity conditions for the AR (p) follow from the requirements that the roots of the auxiliary Equation (28) should be greater than one in absolute value. If any root is ≤1 in absolute value, the process is non-stationary and needs to be differenced or transformed to make it stationary. Similarly, the conditions for invertibility are that the roots of Equation (28) should be greater than one in absolute value. When the reports are not available explicitly, a partial check is available by computing the values of above equations at x = ±1. The condition can be satisfied only if

$$|\phi_p| < 1, |\theta_q| < 1, \phi(1) > 0, \phi(-1) > 0, \theta(1) > 0 \text{ and } \theta(-1) > 0 \qquad (29)$$

The conditions in Equation (29) are necessary but *not* sufficient.

Residual Analysis

If a model adequately depicts the ARIMA process governing the series, then the *residuals* of the model, defined as the difference between the observed and fitted values, should be *white noise*. The *whiteness* of the residuals is tested by time series plot and correlogram. The time plot will reveal any outliers and any obvious autocorrelation or cyclic

effects. The correlogram will enable a more close examination of autocorrelation effects. If the residuals are truly *white noise*, then their ACF should have no spikes and the correlogram is such that each autocorrelation coefficient is N (0, 1/n). Thus, the autocorrelations, r_k, which lie, say, outside the range $\pm 2/\sqrt{n}$ (i.e. outside the approximate 95%large sample confidence limits) are significantly different from zero.

Instead of looking at, r_ks one at a time, Box and Jenkins (1970) describe Portmanteau lack-of-fit test, (Box and Pierce, 1970) with modification by Ljung and Box (1978), which looks at the first k values of the correlogram at once, for analysing the residual autocorrelations. The Portmanteau test contrasts the null hypothesis:

Ho: $r_1 = r_2 = \ldots = r_k = 0$

Against the general alternative H1: not all $r_j = 0$. Based on the residual correlogram, the modified Box-Pierce or the Ljung – Box (L-B) Q statistic is stated as

$$Q \equiv Q(K) = n(n+2)\sum_{k=1}^{k} \frac{1}{n-k} r_k^2(\hat{a}) \qquad (30)$$

where, n denotes the length of the series after any differencing. If the fitted model is appropriate (i.e. if the residuals are *white noise*) Q is approximately distributed as Chi-square distribute variable with (K-p-q-P-Q) degrees of freedom, where p, q, P and Q are the numbers of AR, MA, SAR and SMA parameters respectively in the model. However, this and other similar tests have poor power properties (e.g. Davies and Newbold, 1979) and as suggested by Chatfield (1996) just 'looking' at few values of r_k, particularly at lags 1, 2 and first seasonal lag (if any), and examining if any are significantly different from zero using the crude limits of the series $\pm 2/\sqrt{n}$ is sufficient to test the *whiteness* of the residuals.

Metadiagnosis

This includes omitting or fitting extra parameters where the model is over specified or underspecified as described below.

11.6.4 Overspecified Model: Omitting Parameters

A useful check on the model adequacy is to evaluate whether the fitted model does not contain redundant parameters. The statistical significance of an estimated parameter can be evaluated by calculating the t-ratio, which is the ratio of the parameter estimate to the standard error. A parameter is significantly different from zero, if the t-ratio is equal to

greater than 2 in absolute value. An insignificant parameter is an indication that the model is overspecified and simplification of the model is possible.

11.6.5 Underspecified Model: Fitting Extra Parameters

In order to verify that the tentatively identified model contains the appropriate number of parameters to represent the data, a more elaborate model is fitted to the data. For instance, if the fitted model is ARMA (p, q), more elaborate models ARMA (p + 1, q) and ARMA (p, q + 1) are fitted to the data. The model is then tested to see whether the additional parameters improve the fit significantly. This is seen by examining residual variances; if the white noise variances reduced by 10% by fitting an overfit model, then the overfit model is appropriate.

11.6.6 Forecasting

Once a satisfactory model is found using above described procedure, then it can be used to forecast the future values of the observed time series. For example, if the observed time series is x_1 , x_2 ,....., x_n then the estimation of future values such as x_{n+k} where k is called lead time is called forecasting. The forecast of x_{n+k} made at time n for k steps ahead is denoted as $x(n, k)$. Broadly the forecasting methods may be classified into three groups:

(a) Subjective, (b) Univariate, (c) Multivariate

Subjective forecasts (Armstrong, 1985; Wright and Ayton, 1987) are made using judgement, intuition and any other relevant information. Univariate forecasts are based on a model fitted only to past values of a time series so that $x(n, k)$ depends only on x_n , x_{n-1} .....Methods of this type are also called naïve or projection methods. Multivariate forecast of a variable depends partly on values of one or more series. Methods of this type are also called causal methods. In practice, however, a forecasting procedure may involve a combination of the above approaches. In particular, hydrological forecasts involve combining statistical predictions with the subjective knowledge and insight of people involved in the field. It is important to realise that no single method is universally applicable and a analyst must choose a method which is most appropriate for a given set of conditions. Kendall (1993) gives a review of empirical comparisons of different time series forecasting methods and their suitability to different conditions.

Although point forecast are sometimes adequate, a prediction interval is often helpful to indicate the future uncertainty. The prediction interval can be calculated assuming that the fitted model holds true in the future or can be calculated on an empirical basis from the fitted errors (Chatfied, 1993a). If we assume that the error terms a_t are normally distributed, then the forecast distribution of $x(n, k)$ i.e. $f_{n,k}(x)$, will be distributed as a normal random variable with mean $x(n, k)$ and variance Var $[a_n(k)]$. With this error distribution it is straight forward

to make probability statement about future observations. Therefore the 95% large sample confidence interval for $x(n, k)$ is

$$x_n(k) \pm 1.96\ SE[a_n(k)] \tag{31}$$

Where SE denotes the standard error of the forecast error defined as the square root of variance in equation below

$$\text{Var}\,[a_n(k)] = E\{a_n^2(k)\} = \sigma_n^2 \sum_{j=0}^{k-1} \psi_j^2 \quad \text{with } \psi_0 = 1 \tag{32}$$

In actual calculation of the confidence limits, we replace the error learning coefficient with their estimates and σ_a^2 with its estimate.

The models described in previous sections are assumed to be applied to discrete series that are transformed to second order stationary i.e the distribution of random component is normal. In practice this assumption would generally be sufficient in case of annual series. However modifications are necessary for coping with closely spaced data i.e hourly, daily, weekly and even for monthly data which are highly skewed (Kottegoda, 1980; Salas etal.,1980)). These closely spaced river flow data are auto-correlated and not normally distributed. No standard procedure is available which can cope with both non-normality and serial correlation (Swift and Janacek, 1991). In practice, however, one could fit a linear model using least squares for a long series and rely on asymptotic results. This method provides no information on the distribution of the errors. Therefore the prediction intervals of the forecast cannot be calculated. Swift and Janacek, (1991) describe a model for auto-correlated, non-normal data which allows (i) a general class of correlation structure and (ii) any marginal distribution, and for which maximum likelihood estimates (MLEs) can be found.

11.7 Algorithm for Adaptive Parameter Estimation

Due to many interrelated physiographic and meteorological factors, the parameters of models as described above for one hydrograph may not be an optimal fit to another hydrograph. This means, that every flood hydrograph may give different parameter estimates of the models described above. Thus the parameters estimated from one storm can only be taken as an initial estimate for another storm. The adaptive algorithms are therefore used for updating the parameters as more and more readings of the rainfall-runoff process become available.

In forecasting floods, it is desirable to obtain the parameters recursively as the flood develops. This real time surveillance of the hydrological system is desirable, especially when the parameters are not constant but time dependent. This necessitates the formation of adaptive filter models.

Adaptive filter models are used for online identification where the output from the model is continuously assessed against the actual system output. The residual errors between the actual and forecast output is used as a feedback information into the parameter adjustment mechanisms. Adaptive filter model updates the model parameters using the previous estimate of the model parameters and a function of the forecast error process. The updating algorithms are of the general form:

$$\theta(k+1) = \theta(k) + k(k+1)\, e(k+1) \qquad (33)$$

where

$\theta(k+1)$ = Updated parameter vectors at time (k+1)

$\theta(k)$ = Previous estimate

$k(k+1)$ = Weighting factor

$e(k+1)$ = Residual error between the observed and the forecast values.

Kumar (1980) discussed these adaptive algorithms in detail and used them for on line updating of model parameters.

11.8 REMARKS

The application of time series modelling techniques is explained to examine the historical sequences of hydrological time series for generating likely future sequences over a short period of time. These Black-Box techniques are particularly useful when there is a lack of complete understanding of hydrological processes, resulting in uncertainties in the magnitude and frequencies of future events. The Box-Jenkins time series models and various steps involved in their identification, estimation and diagnosis are described. The data generated through these models can be used for simulating water resources systems and various other purposes. Adaptive filter algorithm has also been explained for on-line updating of model parameters especially when the parameters are not constant but time dependent. These adaptive filters are important to update time series models in real time for short term forecasting of hydrological sequences.

REFERENCES

Armstrong, J. S., Long range forecasting, 2nd ed., New York: Wiley, 1985

Bartlett, M. S., On the theoretical specification and sampling properties of autocorrelated time series, J. of the Royal Stat. Soc., Series B, 8, 1(April), 27-41, 85-97, 1946

Bartlett, M. S., An introduction to stochastic process with special reference to methods and applications, 2nd ed., Cambridge, England: Cambridge University Press, 1966

Box, G. E. P., and G. M. Jenkins, G. C. Reinsel, Time series analysis, forecasting and control, 3rd ed., Englewood cliffs, N.J., Prentice Hall, 1994

Box, G. E. P., and G. M. Jenkins, Time Series Analysis; Forecasting and Control, first edition, Holden Day, San Francisco, Calif., 1970

Box, G. E. P., and G. M. Jenkins, Time Series Analysis; Forecasting and Control, revised edition, Holden Day, San Francisco, Calif., 1976

Box, G. E. P., and D. A. Pierce, Distribution of residual correlations in autoregressive integrated moving average time series models, Journal of the American Statistical Association, 65, 332 (December), 1509-1526, 1970

Chatfield, C., Calculating interval forecasts (with discussion), J. Bus. Econ. Stats., 11, 121-144, 1993(a)

Chatfield, C., The analysis of time series: An introduction, 6th ed., Chapman and Hall, U.K., 1996

Davies, N. and P. Newbold, Some power studies of a portmanteau test of time series model specification, Biometrika, 66, 1 (April), 153-155, 1979

Durbin, J., The fitting of time series models, Review of the international institute of statistics, 28, 3, 233-244, 1960

Harvey, A. C., Time series models, 2nd ed., Hemel Hempstead: Harvester Wheatsheaf, 1993

Hipel, K. W., and A. I. McLeod, Time Series Modelling of Water Resources and Environment Systems, vol 45, Dev. Water Sci., Elsevier Science, New York, 1994

Kendall, Sir Maurice and Ord, J. K., Time series, 3rd ed., Sevenoaks: Edward Arnold, 1990

Kendall, Sir Maurice; Stuart, A, and Ord, J. K., The advanced theory of statistics, vol 3, 4th ed., London, Griffin, 1983

Lettenmaier, D. P., and Wood, E. F., Hydrologic Forecasting, in : Handbook of Hydrology, (Maidment D. R. Ed)McGraw-Hill, 26-1-26-30.

Ljung, G. M, and G. E. P., Box , On a measure of lack of fit in time series models, Biometrika, 65, 2(August), 297-303, 1978

Mills, T. C., Time series techniques for economists, Cambridge: Cambridge University Press, 1991

Quenouille, M. H., Approximate tests of correlation in time series, Journal of The Royal Statistical Society, series B, 11, 1, 68-84, 1949

Salas, J. D., Analysis and modeling of hydrologic time series in : Handbook of Hydrology, (Maidment D. R. Ed)McGraw-Hill, 19-72.

Vandaele, W., Applied time series and Box-Jenkins models, Academic Press INC., 1983

WHO/UNESCO, Report on water resources assessment, World Meteorological Organization/UNESCO, 64.

WHO, Guide to Hydrological Practices, 5th Edition, WHO Geneva, 735.

Wright, G. and Ayton, P., Judgemental forecasting, Chichester: Wiley, 1985.