Chapter One

Introduction

1.1 Study Objectives

The design and planning of projects in water resource management have long been recognized as an important task in human society. This is supported by the fact that man have been mitigating floods and droughts for thousands of years.

The prediction of characteristics of future water supply and planning is often unreliable due to the complex interactions of natural and anthropological elements. In order to simplify the above interactions, mathematical modelling of chronologically recorded series, which are believed to have definite structural patterns, is playing a considerably important role in these projects.

Most hydrologic time series are subjected to two main factors, random and non-random. Non-random factors appear as seasonality-periodic fluctuations that recur within the same ranges each year. These seasonal movements normally dominate the behavior of the observations and therefore hinder us in interpreting and reacting to the important non-seasonal movements in the series.

When making decisions in the planning and operation of a water resource system, it is essential that the available series should provide as much hydrologic information as possible. According to Yevjevich (1972), the extraction of seasonality and underlying trend from the original data will not affect the basic structure of the series and ruin the information contained in the data. Therefore seasonal movements as well as trend components are normally removed from the original series so that the resulting series, the stochastic component, will have a simpler pattern to study.

Our first objective in this study is to estimate and remove the seasonal and trend components of hydrologic time series with the hope that the residual or the stochastic component will turn out to be a stationary random process.

In Chapter Two, we shall discuss the common seasonal adjustment techniques employed in the study of hydrologic data. The classical method for removing seasonal effects from monthly hydrologic data is to standardize the data, i.e., to deduct the corresponding monthly mean from the original observations of each calendar month and then divide them by the monthly standard deviation. Another commonly used method is harmonic analysis, where the total variance of each of the periodic parameters of the series is divided among six harmonics. Significant harmonics are then used to approximate the periodic parameters. Normally, the periodic parameters are the monthly means and standard deviations of the data.

In the last few decades, several relatively sophisticated seasonal adjustment methods have been developed and used in economic and business studies. Some of the more established methods use the Census X-11 procedure, the X-11-*ARIMA* procedure, the *ARIMA*-model-based approach, the State-space model etc. Some of these techniques are surveyed in the second chapter and the applicability of these techniques in hydrologic study is discussed.

Seasonally adjusted river flows are often found to have high skewness and long-term correlation structures. Over the years, the standard procedure for removing skewness from these series is by performing instantaneous transformations. Gaussian models are then used to describe the statistical behavior of the transformed series.

Unfortunately, as commented by Sim (1987b), these Gaussian-type models are inadequate for monthly flows which are highly skewed. He pointed out that in order to transform the original process $\{X_t\}$ into a Gaussian process, there must exist an instantaneous transformation $f(.)$ such that $Y_t = f(X_t)$, and the finite dimensional distributions of Y_t's are multivariate normal. However in practice, it is unlikely that every subset of the Y_t's is multivariate normal. In addition, Weiss (1975) has proved that

(a) a process which is not time-reversible cannot be a Gaussian process and

(b) a transformed process is time-reversible if and only if its original series is time-reversible.

As a consequence, a process which is time-irreversible can nowhere be transformed to a Gaussian process.

In view of the above-mentioned difficulties in fitting Gaussian models to highly skewed series and the increasing interest for non-Gaussian models, particularly the Gamma models, in a variety of fields such as hydrology and meteorology (Lewis, et. al., 1986), the study of non-Gaussian modelling is now of great importance.

Our second objective in this study is to construct probabilistic models for the stochastic component of hydrologic time series, to analyze its properties and subsequently use it in conjunction with the trend and seasonal components for purposes such as simulation of hydrologic data and prediction.

In Chapter Three, we shall review some well established Gaussian autoregressive-moving average processes. A new Gamma-like first-order mixed autoregressive-moving average *ARMA*(1, 1) process is suggested as an alternative to the mixed Gamma first-order autoregressive-moving average *MGARMA*(1, 1) process of Sim (1987a). A detailed discussion of the process is given in Section 3.4.2. We shall fit this model to the stochastic component of the monthly flows of Perak river, and we shall compare the performance of this model with that of the *MGARMA*(1, 1) model.

For effective planning, development, construction and operation of a water resource system, it is crucial to obtain an optimum solution of all the variables that are involved in the project. However, neither the classical empirical methods that have been used extensively in hydrology and water resource projects nor the analytical methods solve the problems with sufficient accuracy. Moreover, the recorded hydro-meteorological data are too short to be used for long-term planning. Consequently, generating new samples of time series may be a more realistic and more accurate method to overcome the limitations just stated.

The first prerequisite for reproducing properties of observed time series in the simulated samples is the proper generation of samples of the stationary stochastic component. In Chapter Four, we shall generate a synthetic series from the proposed Gamma *ARMA*(1, 1) model of Section 3.4.2 and investigate whether the statistical properties, i.e., the mean, variance, skewness and serial coefficients of the generated data resemble those of the raw data. From our investigations, we find that statistical properties of the generated series closely resemble those of the original data.

1.2 Study Aids

In this study, the software package "STATGRAPHICS" is used to model our observed river flows to Gaussian *ARIMA* models. In Chapter Three, the software package "MATHEMATICA", implemented in the Apple Macintosh IICX, is used to draw three dimensional graphs of bivariate distributions, conditional expectations and variances of the non-Gaussian processes discussed in this study. The HP 7475 plotter is used to plot graphs generated by "STATGRAPHICS".

1.3 The Perak River and Data Collection

The hydrologic data used in this study are the monthly stream flows of Perak river in Malaysia. The catchment area of Perak river is 7769 km^2 and most of the areas are covered by virgin jungles. Three-fifths of the soil of the catchment is derived from granite. The maximum length and breadth of the catchment are 145 km and 90 km respectively. About 85% of the catchment is steep mountainous country rising to heights of 7000 feet, while the remainder lies along the flood plains of the river and its tributaries.

This historical series consists of 264 monthly observations (January 1948 to December 1969). The discharge measurements were taken between 32.45 m to 37.31 m above mean sea level and were measured in cubic feet per second per square mile. These data are taken from the stream flows records published by the Drainage and Irrigation Division, Ministry of Agriculture and Fisheries of Malaysia. The monthly data of Perak river are given in Table 1.1.

Table 1.1 *Monthly Mean Discharge (Cubic Feet Per Second Per square Mile) Of The Perak River*

Year	Jan.	Feb.	Mar.	Apr.	May.	Jun.	Jul.	Aug.	Sep.	Oct.	Nov.	Dec.
1948	6.40	3.56	2.45	2.95	4.26	2.25	2.29	2.42	2.01	3.06	3.14	1.76
1949	1.26	1.81	1.22	1.83	2.75	1.88	2.13	1.98	3.31	5.30	4.53	6.82
1950	4.55	4.13	2.09	3.37	4.04	2.50	1.23	1.90	2.32	2.12	3.06	2.18
1951	6.03	3.25	1.93	1.98	2.08	1.69	1.92	7.11	3.40	4.19	8.00	6.14
1952	5.20	2.97	4.10	4.02	3.91	2.11	1.87	1.25	1.54	2.72	4.68	3.10
1953	2.19	1.72	1.33	2.30	2.45	2.62	2.26	1.24	2.69	5.02	5.06	3.39
1954	5.05	1.78	1.33	1.65	1.97	1.58	1.69	1.66	1.24	4.86	3.12	5.15
1955	3.96	3.06	2.00	2.22	3.21	1.90	1.77	2.29	2.63	4.37	5.53	4.11
1956	3.40	2.65	3.62	2.83	2.79	3.76	2.27	2.10	2.18	5.83	6.97	4.72
1957	3.62	2.20	2.14	2.89	3.96	2.93	2.09	1.81	2.16	3.06	5.06	6.40
1958	3.76	3.21	1.96	1.78	2.84	1.84	1.56	1.67	1.62	2.95	3.83	2.22
1959	1.61	1.46	1.48	1.58	2.46	2.15	2.05	2.16	2.37	3.38	4.84	4.06
1960	2.92	1.79	1.68	1.94	1.91	1.02	1.43	0.73	1.00	1.10	2.40	2.64
1961	2.98	1.34	1.22	1.69	1.74	1.09	1.20	1.03	1.46	1.96	2.42	1.94
1962	2.18	0.98	1.25	1.23	2.24	1.16	1.18	1.36	1.46	2.73	2.24	2.90
1963	1.94	0.92	1.05	1.01	1.08	1.32	1.16	1.27	1.08	3.41	4.46	3.07
1964	1.15	1.07	1.04	1.04	1.54	1.36	2.81	1.76	2.66	2.04	3.87	2.24
1965	1.17	1.12	1.09	1.42	2.19	0.98	1.31	1.77	2.53	5.98	6.99	7.37
1966	4.84	2.68	2.33	2.29	2.93	2.57	1.74	2.20	2.62	8.74	7.46	7.74
1967	10.2	5.09	4.27	3.36	4.57	2.53	2.07	1.76	2.03	3.19	7.08	5.71
1968	2.12	1.19	0.96	1.20	1.68	1.83	1.54	1.16	1.15	2.37	1.55	1.95
1969	2.33	1.07	0.98	1.09	2.55	1.82	1.21	2.01	1.40	4.65	4.74	6.95