STOCHASTIC MODELLING OF LONG-TERM PERSISTENCE IN STREAMFLOW SEQUENCES

by

P. Enda O'Connell

Thesis submitted for the degree of
Doctor of Philosophy in the University of London

June 1974

To

Jane Rosemary

ABSTRACT

A simple stochastic model is proposed which is capable of reproducing the observed short-term and long-term behaviour of annual streamflow. The model, a mixed autoregressive moving average or ARIMA (1,0,1) process, is specifically developed for generating synthetic flows for use in the design and operation of water resource systems.

The application of stochastic streamflow simulation in the planning and management of water resources is outlined initially, and some of the associated problems are discussed. In a survey of existing models for generating synthetic flows, particular attention is paid to discrete-time fractional Gaussian noise (dfGn) which has been found in many cases to adequately describe observed long-term persistence in streamflow. The general complexity associated with the generation of approximations to dfGn on a digital computer underlines the necessity for a simple alternative model for generating synthetic flows with the desired short-term and long-term properties.

The ARIMA (1,0,1) process is found to have the necessary properties, and its elegant simplicity, as opposed to the complexity of available approximations to fGn, facilitates the quantification of its asymptotic and small sample properties through analysis and Monte Carlo sampling experiments, respectively. The model may be considered to be a reasonable approximation to dfGn, and is formulated so as to maintain the desired statistical resemblance between historic and synthetic flows. The impact of long-term persistence on reservoir design is illustrated using a simple design procedure.

Some difficulties are encountered in obtaining solutions to the matrix equations which specify the multisite ARIMA (1,0,1) model. The conditions for an acceptable solution are identified and numerical techniques

are employed to solve the equations in the general case. A modified multi-site ARIMA (1,0,1) process is developed which allows an analytical solution to the matrix equations to be obtained. The modified process should prove adequate for the majority of applications.

Parameter estimation for the ARIMA (1,0,1) process is explored using moment and maximum likelihood techniques applied to small samples generated by the process. In a considerable proportion of cases, solutions are unobtainable, and those obtained are found to be highly biased and variable for both estimation procedures.

Finally, the physical bases for alternative stochastic models of streamflow are examined, and the ARIMA (1,0,1) process is found to have a reasonable basis.

ACKNOWLEDGEMENTS

The writer wishes to express his gratitude to both Professor T.O'Donnell, who encouraged the writer's earlier efforts in this field, and Professor J. R. D. Francis for acting as supervisors at various stages of this work. The writer is particularly indebted to Dr. J. R. Wallis whose helfpul criticism and encouragement has contributed considerably. Dr. N. C. Matalas has also contributed significantly through helpful comments and suggestions for which the writer is grateful. Discussions with the following have also proved helpful: Professor D. R. Cox; Mr. P. M. Johnston; Professor J. E. Nash; Mr. J. R. Slack and Dr. G. Weiss.

The unselfishness of the writer's colleagues, Professor J. R. D. Francis, Dr. M. J. Hall, Professor T. O'Donnell and Miss E. M. Shaw is gratefully acknowledged for facilitating the leave of absence granted to pursue his work. Finally, the unfailing patience and care of Misses J. R. Stephens and E. H. Bridgman in typing this thesis are very much appreciated.

Computer facilities provided by the IBM Research Centre are gratefully acknowledged.

TABLE OF CONTENTS

SYMBOLS AND ABBREVIATIONS

Note: As far as possible, a continuous time variable has been denoted, for example, as $X(t)$ while a discrete time variable has been denoted as X_t. An exception occurs in Chapter 4, where the symbol $X(t)$ is used to denote a discrete time variable to simplify the notation for multisite processes.
To accommodate the reader, an abbreviated list of the most frequently used symbols has been included on a pull-out page at the back of the thesis.

∇	Backward difference operator
$\hat{\cdot}$	Denotes 'estimate of'
$\tilde{\cdot}$	In Chapter 1, denotes 'estimate of'; elsewhere, denotes 'approximate value of'
$\forall$	Denotes 'for all'
$\{\cdot\}$	Denotes 'conditional expectation of '
A	Abbreviation for Anderson
$\underline{A}$	(m x m) matrix of coefficients for multisite generating process
$\underline{A}'$	(m x m) matrix of coefficients for multisite log-normal generating process
AR	Abbreviation for 'autoregressive'
ARIMA	Abbreviation for 'autoregressive integrated moving average'
ARMA	Abbreviation for 'autoregressive moving average'
a	Lower bound of a log-normally distributed random variable
a'	Constant characterizing behaviour of population range $R_p(T)$ of Brownian motion $B(t)$
a_h	Constant characterizing behaviour of population range $R_p(T)$ of fractional Brownian motion $B_h(t)$
a_i	In Chapter 4, lower bound of a log-normally distributed random variable at site i; in Chapter 5, harmonic coefficient for frequency i
a_t	Random variable denoting moving average component of ARMA (p,q) process
a'_i	Infiltration ratio for period i
a_{ii}	Element of matrix $\underline{A}$

a.c.	Abbreviation for autocorrelation
B	In Chapter 2, parameter of a fast fractional noise generator; elsewhere, backward shift operator
BL	Abbreviation for 'Broken Line'
$B(t)$	Random variable denoting Brownian motion at time t
$B_h(t)$	Random variable denoting fractional Brownian motion at time t
$B'_h(t,\delta)$	'Derivative' of fractional Brownian motion
$\underline{B}$	(m x m) matrix of coefficients in multisite generating process
$\underline{B}'$	(m x m) matrix of coefficients in multisite log-normal generating process
$\underline{B}^*$	(m x m) lower triangular matrix of coefficients in multisite generating process
b'	Constant characterizing behaviour of the adjusted range R(T) for Brownian motion
b_h	Constant characterizing behaviour of the adjusted range R(T) for fractional Brownian motion
b_i	Harmonic coefficient for frequency f_i
b_j	Regression coefficient relating flow in month (j+1) with flow in month j
b_t	Random variable denoting autoregressive component of the ARMA (p,q) process
b_{ik}, b_{jk}	Elements of matrix $\underline{B}$
C_a	Minimum reservoir storage to meet demands for sequence generated by ARIMA (1,0,1) process
C_m	Minimum reservoir storage to meet demands for sequence generated by lag-one Markov process
C_p	Minimum reservoir storage to meet demands
C(j)	Reservoir storage at end of period j
$C(f_j)$	Cumulative discrete periodogram at frequency f_j
$C^{(hf)}(1)$	Error in approximating $C_h(1)$ with $D_h(1)$
$C_h(k)$	Autocovariance at lag k for discrete time fractional Gaussian noise
$C_h^a(k)$	Approximation to $C_h(k)$

$C_h(\tau,\delta)$	Autocovariance function at lag τ for the 'derivative' of fractional Brownian motion $B_h'(t,\delta)$
C.L.	Abbreviation for 'confidence limits'
CP	Abbreviation for 'cumulative periodogram'
$\underline{C}$	(m x m) matrix of coefficients in multisite generating process
$\underline{C}'$	(m x m) matrix of coefficients in multisite log-normal generating process
c	Distance in time separating values of ε_j in simple Broken Line process
c_i	Value of c for i^{th} simple Broken Line process
c_{ik}, c_{jk}	Elements of matrix $\underline{C}$
D	Differential operator
D^β	Fractional differential operator with β non-integral
D_k	Cumulative sum of the first k values of a random variable in a sequence of length n
D^*_k	Cumulative departure from sample mean
D_n	Cumulative sum of a sequence of n values of a random variable
$D_h(k)$	Approximation to $C_h(k)$
$\underline{D}$	Matrix of second derivatives of log-likelihood function
d	Order of differencing for ARIMA (p,d,q) process
d(j)	Demand on reservoir in period j
d_{ij}	Element of matrix $\underline{D}$
dfGn	Abbreviation for 'discrete time fractional Gaussian noise'
$E(\cdot)$	Expectation operator
$E[\cdot]_n$	Denotes expected value of quantity within brackets in samples of size n
e,exp	The exponential
e(t)	Random variable denoting a simple Broken Line process
F	Forward shift operator
F_i	Baseflow contribution to flow in period i from within period precipitation

$F_1(t/h,M)$	Random variable at time t denoting Type I approximation to discrete time fGn with parameters h and M
$F_2(t/h,M)$	Random variable at time t denoting Type II approximation to discrete time fGn with parameters h and M
$f(n,\rho_1,\varphi)$	Small sample correction factor for variance of ARIMA (1,0,1) process
$f(z;\underline{\xi})$	Probability density function
$f(\underline{\varphi},\underline{\theta})$	Function of parameter sets $\underline{\varphi}$ and $\underline{\theta}$ in likelihood function for ARIMA (p,o,q) process
fGn	Abbreviation for 'fractional Gaussian noise'
G_i	In Chapter 3, inverse of i^{th} root of polynomial $\varphi(B)$; in Chapter 5, baseflow contribution to flow in period i from over-period storage
$g(n,\rho_1)$	Small sample correction factor for standard deviation of lag-one Markov process
$g(n,\rho_1,\varphi)$	Small sample correction factor for standard deviation of ARIMA (1,0,1) process
H	Estimate of h
$\bar{H}$	Mean of 10 values of H
H_j	Inverse of j^{th} root of polynomial $\theta(B)$
h	Hurst coefficient
h_i	Hurst coefficient at site i
$I(f_i)$	Periodogram ordinate at frequency f_i
$I_{[jc,(j+1)c]}^{(t)}$	Indicator function for Broken Line process
$I(\underline{\alpha})$	Information matrix for set of parameters $\underline{\alpha}$
$\underline{I}$	Identity matrix
K	Estimate of h
K_i	Estimate of h at site i
K_α	Factor determining width of confidence limits for cumulative periodogram test
$K_h(s)$	Kernel for discrete-time fractional Gaussian noise
$K_h(s,\delta)$	Kernel for the derivative of fractional Brownian motion $B_h'(t,\delta)$

$K_1(t-\frac{\xi}{10} \mid h,M)$	Kernel for Type I approximation to dfGn
k_i	i^{th} constant in partial fraction expansion of $\varphi^{-1}(B)$
$L(Z_1,Z_2,..Z_n;\underline{\xi})$	Likelihood function
l_1	Loss associated with Type I error
l_2	Loss associated with Type II error
$l(\underline{\xi})$	Log-likelihood function for set of parameters $\underline{\xi}$
$l(\underline{\alpha}, \sigma_\epsilon)$	Log-likelihood function for ARIMA (p,o,q) process
$l^*(\underline{\varphi},\underline{\theta}, \sigma_\epsilon)$	Conditional log-likelihood function for ARIMA (p,o,q) process
M	Memory parameter in Type I and II approximations to dfGn
M_i	Total streamflow in period i
M_n	Maximum of D_1, D_2,...,D_n
M(t)	Moment generating function for real values of t
$M(t_1,t_2)$	Moment generating function for real values of t_1 and t_2
MA	Abbreviation for moving average
$\underline{M}_o$	(m x m) lag-zero cross-correlation matrix with elements $\rho_{ij}(0)$
$\underline{M}_1$	(m x m) lag-one cross-correlation matrix with elements $\rho_{ij}(1)$
$\underline{M}_2$	(m x m) lag-two cross-correlation matrix with elements $\rho_{ij}(2)$
$\underline{M}_o'$	(m x m) lag-zero cross-correlation matrix with elements $\rho_{ij}'(0)$
$\underline{M}_1'$	(m x m) lag-one cross-correlation matrix with elements $\rho_{ij}'(1)$
$\underline{M}_2'$	(m x m) lag-two cross-correlation matrix with elements $\rho_{ij}'(2)$
m	In Chapter 4, number of sites; in Chapter 5, number of backward or forward forecasts used in computing $S(\varphi,\theta)$
m_j	j^{th} constant in partial fraction expansion of $\theta^{-1}(B)$
m_n	minimum of D_1, D_2,..., D_n
m.s.e.	Abbreviation for 'mean square error'

NIP	Abbreviation for normal independent process
n	Sample duration in discrete time
n_m	Value of n at which break to h = 0.5 law occurs in pox diagram
n_s	Sub-sample length in discrete time
n_o	Minimum value of n_s used in calculating H
P_i	Effective precipitation in period i
P_n	Maximum of cumulative departures from sample mean
P_r	r^{th} peak in sequent peak algorithm
P(f)	Cumulative continuous periodogram ordinate
p	Order of autoregressive process
p(f)	Continuous power spectrum
Q	Parameter in fast fractional noise approximation
Q_n	Minimum of cumulative departures from sample mean
q	Order of moving average process
q_i	Monthly flow at time point i
R, R_n	Adjusted range or range of cumulative departures from sample mean in discrete time
R_p	Range of cumulative departures from population mean in discrete time
R_t	Tree ring width for year t
R(T)	Adjusted range in continuous time
$R_p(T)$	Population range in continuous time
R/S	Rescaled range in discrete time
R_{ij}	Element of matrix $\underline{R}$
$\underline{R}$	Matrix of second derivatives of sum of squares function $S(\underline{\alpha},\underline{W})$ with respect to elements of $\underline{\alpha}$
r'_i	Proportion of direct runoff for period i
r.h.s.	Abbreviation for 'right-hand-side'
S	Sample standard deviation
$S_{\bar{x}}$	Estimate of standard deviation of sub-sample means

$S(\underline{\alpha})$	Sum of squares function for set of parameters $\underline{\alpha} = (\underline{\varphi},\underline{\theta})$
$S(\underline{\alpha},\underline{W})$	Sum of squares function for set of parameters $\underline{\alpha}$ and set of observations $\underline{W}$
$S(\underline{\varphi},\underline{\theta})$	Sum of squares function for ARIMA (p,o,q) process
$S(\varphi,\theta)$	Sum of squares function for ARIMA (1,0,1) process
$S^*(\underline{\varphi},\underline{\theta})$	Conditional sum of squares function for ARIMA (p,o,q) process
$\underline{S}$	(m x m) matrix of coefficients for multisite ARIMA (1,0,1) process
s	Backward shift operator
st. dev.	Denotes 'standard deviation of'
T	Sample duration in continuous time
T_r	r^{th} trough in sequent peak algorithm
T_t	Trend component in tree ring width for year t
T'_i	Effective time of occurence of a storm equivalent to all storms recurring in period i
$\underline{T}$	(m x m) matrix of coefficients in multisite ARIMA (1,0,1) process
th_1	Parameter in fast fractional noise approximation
$\underline{U}$	(m x m) matrix of coefficients in multisite ARIMA (1,0,1) process
$\underline{U}_j$	Value of matrix $\underline{U}$ after j^{th} iteration
V_h	Variance of unit increments of fractional Brownian motion
$V(\underline{\alpha})$	Variance-covariance matrix of set of parameters $\underline{\alpha}$
$V(\varphi,\theta)$	Large-sample variance-covariance matrix of parameters of ARIMA (1,0,1) process
Var	Denotes 'variance of'
$\underline{V}$	(m x m) matrix of coefficients for multisite ARIMA (1,0,1) process
$\underline{V}_j$	Value of matrix $\underline{V}$ after j^{th} iteration
W_m	Weight attached to m^{th} lag-one Markov process in fast fractional noise approximation
W_t	Random variable at time t with zero expectation

$\underline{W}$	In Chapter 4, (m x m) matrix of coefficients for multisite generating process; in Chapter 5, set of values $[W_1, W_2, \ldots, W_n]$
$\underline{W}^*$	Set of values $[W_{t-1}, W_{t-2}, \ldots, W_{t-p}]$
$w(j)$	Waste-water in period j
X_t	Discrete-time random variable at time t
$X_t^{(hf)}$	High frequency component of fast fractional noise approximation
$X_t^{(lf)}$	Low frequency component of fast fractional noise approximation
$X_t^{(m)}$	General Broken Line process
$X_t(\omega)$	Discrete-time random variable at time t, with ω denoting an element in the sample space
$X(t,\omega)$	Continuous-time random variable at time t, with ω denoting an element in the sample space
$X_i(t)$	Discrete-time random variable at time t for site i.
$\bar{X}$	Mean of a sequence $X_1, X_2, \ldots, X_n$
$\underline{x}(t)$	(m x 1) matrix of standardized discrete time random variables at time t
$x(j)$	Inflow to reservoir in period j
Y_t	Normally distributed discrete time random variable at time t
$Y_i(t)$	Normally distributed discrete time random variable at time t for site i
$y_t(\rho_1^{(m)})$	Component lag-one Markov process of fast fractional noise approximation
$\underline{y}(t)$	(m x 1) matrix of standardized discrete time normally distributed random variables at time t
Z_t	Discrete-time random variable at time t
Z_α	Standard normal deviate for probability level α
z_i	Standardized monthly deviate for time point i
α	Level of significance
α'	Level of development
α_i	Element of set $[\alpha]$

$\underline{\alpha}$	Set of parameters
$[\alpha]$	Set of parameters
β_{ij}	Element of matrix $\underline{\beta}$
β^*_{ij}	Element of matrix $\underline{\beta}^*$
$\underline{\beta}$	(m x m) lower triangular matrix
$\underline{\beta}^*$	(m x m) lower triangular matrix
$\Gamma[\cdot]$	Denotes Gamma Function
γ	Roughness parameter of Broken Line process
γ_i	Coefficient of skewness of random variable $X_i(t)$
γ_k	Autocovariance for lag k
γ_x	Skewness of random variable X_t
γ_η	Skewness of random variable η_t
Δt	Increment of time
$\Delta B_h(t)$	Increment of fractional Brownian motion
$\underline{\Delta}$	(m x m) matrix such that $\underline{\Delta}\underline{\Delta}^T = \underline{I}$
δ	Small positive quantity
δ_t	Normally and independently distributed random variable at time t with zero mean and unit variance
ε	Small positive time increment
ε_t	Normally and independently distributed random variable at time t with zero mean and unit variance
$\bar{\varepsilon}$	Mean of sequence $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$
$\underline{\varepsilon}$	Set of values $[\varepsilon_1, \varepsilon_2, \ldots \varepsilon_n]$
$\underline{\varepsilon}^*$	Set of values $[\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots, \varepsilon_{t-q}]$
$\underline{\varepsilon}(t)$	(m x 1) matrix of normally and independently distributed random variables with zero mean and unit variance
η_t	Independent random variable at time t with zero mean, unit variance and skewness γ_η
θ	Parameter of first order moving average
θ_q	q^{th} moving average parameter

$\theta(B)$	Polynomial in backward shift operator B with coefficients θ_j
$\theta(F)$	Polynomial in forward shift operator F with coefficients θ_j
$\hat{\theta}(1)$	Maximum likelihood estimate of θ
$\hat{\theta}(m)$	Moment estimate of θ
$\underline{\theta}$	Set of parameters $[\theta_1, \theta_2, \ldots, \theta_q]$
λ_a	Average minimum reservoir storage for sample size n for ARIMA (1,0,1) process
λ_m	Average minimum storage for sample size n for lag-one Markov process
μ_i	Population mean of random variable $X_i(t)$
μ_j	Population mean flow for month j
μ_x	Population mean of random variable X_t
μ_y	Population mean of random variable Y_t
$\underline{\mu}$	(m x 1) matrix with elements μ_i
$\underline{\mu}'$	(m x 1) matrix with elements μ'_i
ξ_i	Element of set $\underline{\xi}$
$\underline{\xi}$	Set of parameters
π	Constant equal to 3.1416...
π_j	Parameter of infinite autoregression
$\pi(B)$	Polynomial in backward shift operator B with coefficients π_j
ρ, ρ_1	Lag-one autocorrelation coefficient
ρ_j	Correlation coefficient between flow in month (j+1) and flow in month j
ρ_k	Lag-k autocorrelation coefficient
ρ_x	Lag-one autocorrelation coefficient of random variable X_t
ρ_y	Lag-one autocorrelation coefficient of random variable Y_t
$\rho_1^{(m)}$	Lag-one autocorrelation coefficient for m^{th} lag-one Markov process in fast fractional noise approximation
$\rho_1^{(hf)}$	Lag-one autocorrelation coefficient of high frequency component in fast fractional noise approximation

$\hat{\rho}_1^*$	Estimate of ρ_1 from synthetic sequence
$\hat{\rho}_1^{(c)}$	Corrected estimate of ρ_1
ρ_{xy}	Cross-correlation coefficient between random variables X_t and Y_t
$\rho(\tau)$	Autocorrelation function at lag τ in continuous time
$\rho_x(\tau)$	Autocorrelation function at lag τ for general Broken Line process
$\rho''(0)$	Second derivative of autocorrelation function at origin
$\rho_i(k)$	Lag-k autocorrelation coefficient of random variable $X_i(t)$
$\rho_i'(k)$	Lag-k autocorrelation coefficient of random variable $Y_i(t)$
$\rho_{ij}(k)$	Lag-k cross correlation coefficient between random variables $X_i(t)$ and $X_j(t)$
$\rho_{ij}'(k)$	Lag-k cross correlation coefficient between random variables $Y_i^{(t)}$ and $Y_j(t)$
σ	Population standard deviation
σ_a	Population standard deviation of ARIMA (1,0,1) process
σ_i	In Chapter 2, population standard deviation for i^{th} simple Broken Line process; in Chapter 4, standard deviation of random variable $X_i(t)$
σ'_i	Standard deviation of random variable $Y_i(t)$
σ_j	Population standard deviation of flows in month j
σ_m	Population standard deviation of lag-one Markov process
σ_x	Population standard deviation of the random variable X_t
σ_y	Population standard deviation of the random variable Y_t
$\sigma_{\bar{x}}$	Population standard deviation of sub-sample means
σ_ϵ	Standard deviation of the random variable ε_t
φ	Parameter of first order autoregressive process
φ_i	Autoregressive parameter at site i
φ_p	p^{th} autoregressive parameter
$\varphi(B)$	Polynomial in backward shift operator B with coefficients φ_j
$\varphi(F)$	Polynomial in forward shift operator F with coefficients φ_j

$\hat{\varphi}^{(1)}$	Maximum likelihood estimate of φ
$\hat{\varphi}^{(m)}$	Moment estimate of φ
$\underline{\varphi}$	Set of parameters $[\varphi_1, \varphi_2, \ldots, \varphi_p]$
$\chi^2(K-p-q)$	Value of chi-square for (K-p-q) degrees of freedom
ψ_j	Parameter of infinite moving average
$\psi(B)$	Polynomial in backward shift operator B with coefficiants ψ_j
Ω	Sample space for random variable.

INTRODUCTION

"On a multitude of rivers of the earth, long series of dry and wet years are observed that are practically incredible in successions of accidental values".

Kritsky and Menkel (1970)

The empirical observation embodied in the above quotation has inspired numerous investigations in the past to determine laws, either deterministic or probabilistic, which would quantify the persistence in annual streamflow. Among such investigations, the work of Hurst (1951, 1956) stands out. He made the remarkable empirical discovery that a host of geophysical time series, including streamflow, obeyed one universal, probabilistic law, specified by one parameter, $0 < h < 1$, which governed the duration and intensity of periods of above and below average flow. Even more remarkable was the fact that the overall average value of h found by Hurst, 0.73, was in strong disagreement with the value of h of 0.5 predicted by all the then available probabilistic theories.

The result continued to baffle and intrigue statisticians and engineers for a number of years, until Mandelbrot (1965) and Mandelbrot and Van Ness (1968) evolved the necessary probability theory to explain Hurst's findings, and proposed discrete-time fractional Gaussian noise (dfGn) as a new model of annual streamflow. The theory is based on the existence of long-term persistence, measured by values of h in the range $0.5 < h < 1$, suggesting that events in the distant past still influence present events and that long periods of high or low streamflow can be extremely long indeed. In contrast, short-term persistence, characterized by a value of h equal to 0.5, refers to the belief that present events are not influenced by past events or are influenced only by very recent past events.

The application of Mandelbrot's theory to streamflow was strikingly illustrated by Mandelbrot and Wallis (1968, 1969a,b,c,d,e), who proposed algorithms for generating approximations to dfGn on a digital computer. Further approximations to dfGn, namely, fast and filtered dfGn, have since been developed by Mandelbrot (1971a) and Matalas and Wallis (1971b),respectively.

However, there are difficulties associated with the practical application of dfGn in the design of water resource systems. Only one generating process, filtered dfGn, has been suitably documented for generating synthetic streamflows, but the process is rather cumbersome and expensive to use. Fast dfGn, as implied by its name, aims to achieve economy in computation, but the generating process itself is still rather complicated, and has not as yet been sufficiently documented for practical application. In addition, Matalas and Wallis (1974) have pointed out that the small sample properties as well as the population properties of a generating process are needed before the correct resemblance can be maintained between historic and synthetic flows.

A requirement thus exists for a simple generating process with the necessary long-term behaviour which can easily be employed in practical design situations. This thesis is concerned with the development of the ARIMA (1,0,1) process, already proposed by O'Connell (1971) and shown to provide an alternative to the fGn model of Hurst's time series. As a result of its simplicity, the small sample properties of the process can be obtained either through theoretical analysis or Monte Carlo sampling experiments. The new process is particularly suitable for the investigation of the practical consequences of long-term persistence in the design of water resource systems.

Chapter 1 deals with the fundamentals of stochastic streamflow simulation, and considers its potential in the design of water resource systems.

Associated problems are discussed under the headings of model choice, distribution choice and parameter estimation, and decision theory is suggested as a possible means of resolving some of the problems.

Chapter 2 traces the historical development of synthetic hydrology, and particular attention is paid to studies attempting to account for Hurst's findings. In this context, the failure of models possessing the properties of the increments of Brownian motion is noted. In contrast, fractional Brownian motion provides the correct basis for a model, and an attempt is made to give a logical integrated treatment of the properties of the increments of fractional Brownian motion (discrete-time fractional Gaussian noise) from mathematical theory through to computational algorithms. The main difficulties associated with the practical application of dfGn, namely, the formulation of generating processes and the estimation of h, the Hurst coefficient, are discussed. Some of the very recent contributions to synthetic hydrology are also reviewed.

The properties of a simple stochastic model of long-term persistence, the ARIMA (1,0,1) process, are explored in chapter 3, and its potential as an adequate and useful alternative to dfGn for simulation purposes is exposed. Some of the background material presented initially, relies on the exposition of Box and Jenkins (1970), and is included for completeness. Using the lag-one Markov or ARIMA (1,0,0) process and dfGn as foils, and extensive simulation experiments as evidence, the ARIMA (1,0,1) process is shown to possess both the better attributes of the alternative models, namely, elegant simplicity and the ability to model short-term and long-term persistence adequately. The small sample properties of estimates of the variance, lag-one autocorrelation and h are established and are used in formulating the model to maintain the required resemblance between historic and synthetic flows. The preservation of skewness is also considered

as is the case of a log-normal ARIMA (1,0,1) process. Some simple simulation experiments are reported which illustrate the effects of long-term persistence and small sample biases on reservoir design.

Multisite ARIMA (1,0,1) processes with the required temporal and spatial correlation structures are developed in chapter 4. Existing multisite models are first reviewed and the absence of a manageable multivariate process for generating flows with the required long-term and short-term properties is evident. Some difficulties are encountered in finding solutions to the matrix equations which specify the multisite ARIMA (1,0,1) process. The conditions for an acceptable solution are established, and a simple, iterative, numerical procedure is used to solve the matrix equations. The properties of available analytical solutions are examined in detail and one particular solution is found to be acceptable if lagged cross-correlations of order greater than zero are not of interest. The latter formulation is found to have a convenient interpretation in terms of long-term persistence. The generation of multivariate log-normal sequences is also considered in addition to the standard case where attention is confined only to first and second order moments.

The estimation of the parameters of the ARIMA (1,0,1) process using the techniques proposed by Box and Jenkins (1970) forms the subject matter of chapter 5. While these techniques may prove adequate for the longer time series frequently available in business and industry, their potential is shown to be somewhat limited when applied to shorter sequences. Some of the background material on parameter estimation using the Box and Jenkins approach is presented initially. The power of some of their goodness-of-fit tests in detecting long-term persistence is explored, and found to be rather low. Moment and maximum likelihood estimation techniques are applied to short to moderate length sequences generated by an ARIMA (1,0,1)

process, and, in a large proportion of cases, neither technique provides estimates lying within the parameter space of the process. When acceptable estimates are available, they are found to be characterized by excessive bias and variability for both techniques ; only when the sample size exceeds considerably the length of historic annual streamflow sequences do the better properties of maximum likelihood techniques tend to emerge. The application of the ARIMA (1,0,1) process to some records of annual streamflow is finally illustrated.

In chapter 6, the main conclusions of the thesis are presented. In this context, the physical bases of fGn and the ARIMA (1,0,1) process are compared, as no clear recommendation for the use of either model can be made on statistical grounds. Some recent work suggests that the ARIMA (1,0,1) model has a reasonable physical basis, while the physical basis of dfGn has not been explored. In addition, the ARIMA (1,0,1) process has been formulated so as to achieve statistical resemblance between synthetic and historic sequences of equal length, which gives it an important advantage over approximations to dfGn. However, for any particular design situation, the model choice dilemma remains, and may be best approached using decision theory. With design periods of more than about 50 years becoming unrealistically long in present day circumstances, work should perhaps concentrate on determining the importance of long-term persistence in such cases.

Chapter 1

THE RATIONALE AND FRAMEWORK OF SYNTHETIC HYDROLOGY

Advances in stochastic modelling and simulation techniques in hydrology in recent years have led to the development of improved methods of water resource system planning. Stochastic techniques of flow synthesis or generation are frequently referred to collectively as operational or synthetic hydrology.

In section (1.1) streamflow is viewed as a stochastic process, while section (1.2) outlines the basic assumptions underlying synthetic hydrology, and some of the difficulties associated with model choice, distribution choice and parameter estimation. Section (1.3) outlines briefly the potential of synthetic hydrology in the design and operation of water resource systems. Finally, in section (1.4) a decision theory framework is considered which may help to overcome some of the shortcomings of synthetic hydrology.

1.1 Streamflow as a stochastic process

Hydrological processes i.e. rainfall, streamflow, tend to behave in a rather complex fashion; the developing science of hydrology is primarily concerned with quantifying the temporal and spatial behaviour of such processes. As in other fields, such as fluid mechanics, the study of complex natural processes has produced two distinct approaches. Those who have allied themselves to a deterministic approach have sought to explain the behaviour of streamflow and its natural component processes in terms of known physical laws. On the other hand, the apparently random nature of some hydrological processes have led the advocates of probability theory to suggest that a probabilistic mechanism underlies the evolution of such phenomena. The latter

approach is frequently conjectured, in particular by deterministes, to be an admission of a state of ignorance: however, the argument for the application of either approach must surely revolve not around philosophical standpoints in hydrology but around how the design and management of water resource systems may best be approached. Probability theory affords a means of formulating an objective methodology for assessing the inherent risks and uncertainties associated with hydrological design in the face of inadequate data. Thus, the field of stochastic hydrology has emerged partly in response to the need for such a methodology, and partly in response to the need for a better understanding of the structure of hydrological processes.

Within the field of stochastic hydrology, hydrological processes are hypothesised as stochastic processes; a definition of a stochastic process is, perhaps, pertinent. Within the literature on stochastic processes, a universal definition barely emerges. Some authors reserve the term stochastic process for random phenomena which change with time (Prabhu, 1965; Jenkins and Watts, 1968); however, a more acceptable and general concept is obtained if change with respect to reference frames other than time is allowed (Bartlett, 1966; Cox and Miller, 1966). Hence, a stochastic process may formally be defined as an indexed or ordered set of random variables $[X(s,\omega) : s \in S, \omega \in \Omega]$ where s is an element of the index set S, and Ω is the sample space for the random variable $X(s,\omega)$ for each value of s.

In dealing with hydrological processes, the reference frame will generally be the time axis, although reference frames in space sometimes arise. In the former situation, the stochastic process will have as its index set T, the set of time points; further consideration can, without loss of generality, be confined to stochastic processes with time as a reference frame. For a continuous process, the index

set T' may be specified as $T' = [t : -\infty < t < \infty]$ while for a stochastic process in discrete time the index set may be specified as $T' = [0, \pm 1 \pm 2...]$ or $T' = [0,1,2...]$. Hence the continuous stochastic process for streamflow may be denoted by $[X(t,\omega) : t \in T', \omega \in \Omega]$, and a single realization of the process, an observed streamflow sequence may be denoted by $[X(t,\omega_o) : t \in T', \omega_o \in \Omega]$. In this notation, $X(t,\omega)$ denotes the random variable associated with time t, and $X(t,\omega_o)$ denotes a variate value, for example, observed streamflow, of $X(t,\omega)$. The indexing of the random variables implies that particular importance is attached to their order; ordering is necessary as interdependence will generally exist between the random variables $X(t,\omega)$ and $X(t+\delta,\omega)$ where δ represents a shift in the time origin.

As already noted, a hydrological process may be hypothesised as a stochastic process; interdependence between the random variables at times t and $t+\delta$ may be attributed to the action of atmospheric and catchment storage. In this context, the stochastic process represents a mathematical or stochastic model of the hydrological process. In almost all cases, only a single realization of a stochastic process is available, and that for only a finite set of time points $[t_1,...t_n]$. The available sample is generally termed a time series and represents a sample of size n from the n^{th} order multivariate process $[X_t(\omega) : t \in T', \omega \in \Omega]$. The principal aim of time series analysis is to infer the probability law of the stochastic process from a single realization of the process.

For purposes of application, a stochastic model must be specified in terms of a set of parameters, which specify in turn the probability law of the stochastic process; in this form the model is sometimes referred to as a generating process. The values of the parameters will

be unknown, and statistical estimates of these parameters must be abstracted from the observed realization of the process.

Stochastic modelling may be undertaken solely with the purpose of providing insight into the structure of hydrological processes. However, combined with simulation techniques on digital computers, the stochastic modelling of streamflow has realized its potential more fully, and has emerged as a powerful tool for use in water resource system planning. Today, stochastic simulation techniques form an integral part of the methodology of water resource system design because they afford the planner a means of making realistic projections of future flows within the system and of assessing the risks involved in failing to meet the demands to be placed on the system. In order to give the technique of synthesising likely future flow sequences a proper identity, the proponents of this approach have coined the equivalent titles of operational and synthetic hydrology.

Synthetic hydrology helps to overcome the inadequacies and uncertainties associated with the use of historic streamflow sequences only in the design of water resource systems. However, because the history of flows for a particular stream provides the only available information on the future behaviour of the stream, any projections of future flows must essentially be based on the historic sequence. Synthetic hydrology satisfies this criterion; it is important to note, however, that while very many synthetic flows may be generated, new information is not created. Synthetic hydrology is merely a sophisticated technique for ~~building~~ utilising the information contained in historic records. The assumptions and the problems associated with the generation of synthetic sequences have come to be more fully understood within the last few years and there have been a number of further innovations in the field.

1.2 Synthetic Hydrology

1.2.1 Assumptions

The successful application of synthetic hydrology in the design of water resource systems inevitably depends on a proper understanding by the planner of the underlying assumptions. To generate synthetic sequences certain assumptions about the statistical behaviour of the generating process of streamflow are necessary. A convenient assumption is that annual streamflow is a stationary stochastic process, i.e., the process is in a state of statistical equilibrium and its statistical characteristics are independent of absolute time. In essence, this assumption suggests that the variability of a historic record reflects the variability to be expected in a future flow sequence. Practically all the statistical literature dealing with stochastic processes pertains to stationary stochastic processes and a theory for non-stationary processes hardly exists. However, observed streamflow may not represent natural streamflow due to some artificial influences acting on the streamflow regime, causing the statistical characteristics of the flows to change with time. Such influences may however be difficult to quantify and there may be no apparent reason for transient low frequency effects. Provided there are stationary stochastic processes which can adequately characterize observed annual streamflow behaviour, there seems to be no valid reason for invoking non-stationarity, in the presence of which nothing useful can be said about the future behaviour of the process. If monthly or daily streamflows are considered, a quantifiable source of non-stationarity arises, which can be coped with through considering that monthly flows, or daily flows within a month constitute a stationary stochastic process.

Given the generating mechanism of streamflow, an ensemble of

future flow sequences within a water resource system could be generated. Each synthetic sequence within the ensemble would, according to probability theory, have an equal chance of being realized over the design life of the system. However, the generating process of streamflow, is, unfortunately, unknown. The available historic sequence represents only one realization from the ensemble of flow sequences necessary to characterize a stochastic process. In this situation, the generating process must be approximated using information conveyed by a historic sequence. This approach invokes the assumption of ergodicity, which allows averages over the ensemble, which is not available, to be replaced by time averages over the historic sequence. The ensemble of future flow sequences, each spanning the economic life of the water resource system, may then be generated, and used to aid decision making in the planning of water resource systems.

1.2.2 Model Choice

As already noted, the generating process of streamflow is unknown, and a model must be chosen to approximate the underlying generating mechanism. In this situation, the planner must use all the quantifiable characteristics of a historic sequence to guide his choice. The generating process of streamflow may be specified by a set of population parameters $[\alpha] = (\alpha_1, \alpha_2, \ldots, \alpha_j)$, the values of which are unknown. Estimates of $\alpha_i \; \forall i$ may be obtained from historic sequences. However, all the salient properties of historic sequences may not be captured by the set $[\alpha]$. The system design will, in general, be insensitive to certain members of the set $[\alpha]$ which may be reduced accordingly. A generating process is now required which will generate synthetic sequences resembling the historic sequences in terms of the set $[\alpha]$. At this stage a model may not be available with the required capabilities and a compromise may have to be adopted. The net result is that, until

recently, the designer has probably adopted a simple, widely used and will documented generating process and hoped that he has produced realistic design responses. The lag-one Markov process has filled precisely such a role in synthetic hydrology. For this process, the set $[\alpha]$ usually comprises means, variances, skewnesses and lag-one serial correlation coefficients of individual historic sequences and lag-zero cross-correlations between sequences. However, deficiencies in the Markovian generating process have recently called its widespread use into question; in this context, members of the set $[\alpha]$ pertaining to persistence warrant special discussion.

Persistence is the tendency for high flows to be followed by high flows and low flows to be followed by low flows, and is present to a greater or lesser extent in most hydrological time series. The degree of observed persistence will obviously depend on the time interval of the observations. However, persistence is measured by the autocorrelation coefficient, ρ_k which is a measure of the degree of linear dependence existing between observations separated by k time units. In the past, synthetic hydrology has relied almost entirely on ρ_1, the first autocorrelation coefficient, to measure persistence in hydrologic sequences. Since persistence undoubtedly plays a major role in determining the sizes of storage reservoirs within a water resource system, adequate measures of persistence must be included in the set $[\alpha]$. The lag-one autocorrelation coefficient, however, apparently measures only high frequency behaviour in a time series and consequently cannot capture any long term persistence, which, if it exists, can be the critical factor in determining a reservoir design. The lag-one Markov process is not the designer's panacea in this situation; for this model, the only measure of persistence included in the set $[\alpha]$

is the lag-one autocorrelation coefficient, which specifies the autocorrelation function of the process completely. Consequently, if long-term persistence exists in historic sequences, synthetic sequences generated by a lag-one Markov process will not reproduce such effects, and may well lead to underdesign of reservoirs, especially if high levels of development are being considered. (Wallis and Matalas, 1972). With the increasing demands on water resources throughout the world, high levels of development are to be expected, and consequently the correct modelling of long-term persistence constitutes an important issue.

The most important innovation in synthetic hydrology in recent years has been the introduction of a new model, called fractional Gaussian noise (fGn)(Mandelbrot and Wallis, 1968, 1969a,b,c). Approximations to fGn enable long-term effects to be reproduced in synthetic sequences. With these approximations, an extra parameter, h, the Hurst coefficient is added to the set $[\alpha]$, where $0.5 < h < 1$. h is a direct measure of long-term persistence and traces its origin in hydrology to the studies of Hurst (1951, 1956). Hurst's extensive studies, based on some 800 time series pertaining to streamflow, rainfall, tree rings, mud varves and temperature yielded an average estimate of 0.73 for h, which suggests that designers must consider the matter of persistence and alternatives to simple Markovian processes for which $h = 0.5$. However, the simpler of the original approximations to fGn proposed by Mandelbrot and Wallis (1969a) called Type 2, possessed some deficiencies, resulting in undesirable low and high frequency properties. However, adequate approximations, called filtered fGn's have now been successfully developed for generating synthetic sequences by Matalas and Wallis (1971b).

However, there have been deterrents associated with using fGn. As a generating process, it is more complicated than the lag-one Markov

process and demands a considerable investment in computer time. A fast fGn generator has been proposed (Mandelbrot, 1971a) but its adaptation for synthetic hydrology has not as yet been attempted. It is perhaps pertinent to pose the question, is the mathematical exactness of fGn, in preserving values of h explicitly, warranted? Perhaps simple approximations, computationally easier and incorporating alternate measures of persistence would yield design results not differing significantly from those provided by fGn. Contenders for such a role are low order ARIMA processes (Box and Jenkins, 1970), as one in this class, the ARIMA (1,0,1) process, has been shown by O'Connell (1971) to incorporate a measure of long-term persistence, and to provide an adequate model of Hurst's time series. From the standpoint of synthetic hydrology, low order ARIMA processes are simple to generate, and may be viewed as approximations to fGn.

More recently, a new model, called the Broken Line (BL) process, has been proposed for use in synthetic hydrology, (Rodriguez et al , 1972 ; Mejia et al , 1972 and Garcia et al , 1972), and a number of advantages are claimed for it over existing generating processes. However, the advantages quoted have been disputed (Mandelbrot, 1972). An interesting concept has been proposed by Garcia et al (1972) who suggested that synthetic sequences should contain observed historic sequences. However, this approach is not a unique advantage of BL methodology, and some evidence is required that it possesses distinct advantages over conventional approaches.

While the modelling of daily streamflow is not of primary concern here, an important innovation, the shot noise model, has recently emerged in this field (Weiss, 1973). Foremost among the advantages of the shot noise model are (i) the ability to model recession effects adequately and (ii) the ability to generate synthetic flows which are

always positive without invoking transformations.

A more detailed review of the various models available for generating synthetic flows is delayed until chapter 2, with particular emphasis laid on the modelling of long-term persistence.

1.2.3 Choice of Distribution

In the preceding section, attention revolved around choosing a model which would ensure that synthetic sequences resembled historic sequences in terms of the set $[\alpha]$. Included in this set were the first three moments of the distribution function of the historic flows. However, this distribution function is again, unknown. This does not constitute any problem if the designer is merely interested in preserving the first three moments of historic sequences, as this may be achieved without invoking the assumption of any particular distribution. Consideration of the lag-one Markov process,

$$X_t = \rho X_{t-1} + \sqrt{1 - \rho^2}\, \varepsilon_t \qquad (1.1)$$

where ε_t is an independently distributed random variable, illustrates this point. For this process, relations may be established between the first three moments of the X_t and those of the ε_t. Thus, given the moments of X_t, the moments of ε_t may be found. In generating the process, ε_t may be sampled from any arbitrary distribution, provided that distribution has the required first three moments. The choice of distribution for the ε_t will probably be dictated by available computational algorithms for generating ε_t. A similar strategy may also be employed with other generating processes if only the first three moments of the observed flows are to be preserved.

However, if,in addition, the designer is interested in a specific distribution for the X_t, problems may arise. In general, the difficulty

arises in specifying the distribution of the ε_t, given that of the X_t. The case where X_t is normal presents no difficulty as the sum of normal variates with a given mean and standard deviation is normal, and X_t may be written as an infinite summation of the ε_t. The case where X_t is distributed as log-normal or as gamma requires a different approach as the sum of two log-normal variates is not log-normal and the sum of two gamma variates is not, generally, gamma. In the case of the log-normal, normal variates, Y_t may be generated where $Y_t = \log X_t$ and these may then be transformed to X_t values. Relations have been established which enable the moments of the Y_t to be defined, given those of the X_t (Aitchison and Brown, 1957). Matalas and Wallis (1974) have shown that the Markov process cannot be used to generate flows X_t distributed as gamma, if ε_t is distributed as gamma. The distribution of ε_t yielding a gamma distribution for X_t is unknown.

The pertinent issue here is whether or not the choice of distribution is going to influence the system design. Proof that streamflow conforms to any particular distribution is extremely difficult to establish using statistical tests, given the sample sizes usually available in hydrology. Consequently, some research is required on the impact of distribution choice for streamflow on water resource system design.

1.2.4 Parameter Estimation

Hydrology is usually afflicted with a paucity of data and a designer usually has to accept large standard errors and biases associated with the statistical estimates he obtains. Estimates of members of the set $[\alpha]$, denoted by $\hat{\alpha}_i \; \forall \; i$, measured within historic sequences of size n suffer from such afflictions, and this situation necessitates formal definitions of the statistical resemblance to be

maintained between historic and synthetic sequences. Matalas and Wallis (1974) have provided definitions along the following lines.

From a synthetic sequence of length n where n is the length of the historic sequence, estimates of α_i $\forall$ i, denoted by $\tilde{\alpha}_i$ are obtained. The synthetic sequence is usually said to resemble the historic sequence if $\tilde{\alpha}_i \rightarrow \hat{\alpha}_i$ $\forall$ i as $n \rightarrow \infty$. This type of resemblance will be referred to as Type A. Resemblance between historic and synthetic sequences of size n will now be considered. An ensemble of N synthetic sequences of size n is generated. For each synthetic sequence of length n, parameter estimates $\tilde{\alpha}_i$ $\forall$ i are obtained. The N values of $\tilde{\alpha}_i$ $\forall$ i are averaged over the ensemble for each value of i, yielding average values, α_i^*. If $\alpha_i^* \rightarrow \alpha_i$ $\forall$ i as $N \rightarrow \infty$, then each synthetic sequence of length n is said to resemble the historic sequence. This latter resemblance, referred to as Type B, would appear to be what the designer requires, although the difference between these two definitions of statistical resemblance has generally been unappreciated by hydrologists. Nevertheless, estimator properties and loss functions may determine which type of resemblance is the best to maintain.

In generating synthetic sequences, $\hat{\alpha}_i$, a parameter estimate obtained from the historic sequence of length n, assumes the role of a population value, and if $\hat{\alpha}_i$ is an unbiased estimate of α_i, then the two definitions of statistical resemblance are equivalent. However, biased statistical estimates have been encountered in synthetic hydrology (Wallis and Matalas, 1971). In the presence of persistence, estimates of the variance and the lag-one autocorrelation in small samples will be biased downwards, regardless of what generating mechanism is postulated. In these cases resemblance of Type A is obtained but not of Type B for which case $\alpha_i^* \rightarrow \beta_i$ as $m \rightarrow \infty$ where $\beta_i < \alpha_i$ $\forall$ i.

Unfortunately, a further source of bias, the magnitude of which will be unknown, must be tolerated in adopting synthetic hydrology for system design. The estimates of the parameters $\alpha_i \; \forall \; i$ from a historic sequence are quite unlikely to be equal to their respective population values. Corresponding parameter estimates derived from synthetic sequences will be distributed about $\hat{\alpha}_i \; \forall \; i$ rather than about the true population values. Thus the system design must of necessity, be under or over-designed with respect to the design which would result if the population of flows were available. The parameter estimates $\hat{\alpha}_i \; \forall \; i$ are said to be operationally biased with respect to the system's design; this bias is quite distinct from statistical bias in the parameter estimates themselves. Even though estimates of the members of the set $[\alpha]$ may be statistically unbiased, operational bias will still arise, due purely to sampling variability in estimates of the set of parameters $[\alpha]$. Operational bias is one of the limitations of synthetic hydrology; however, alternative design methods, inevitably based on the historic record, must also suffer to a greater or lesser extent from sampling variability in the historic record.

Operational bias may be minimized, but never eliminated. Efficient estimation techniques, such as the method of maximum liklihood, should be used whenever possible. Regionalization techniques have been suggested (Benson and Matalas, 1967) as a means of utilizing the spatial relations among the members of the set $[\alpha]$ to augment the lengths of the historic sequences, and thereby lessen the operational bias whose magnitude varies inversely with sequence length. However, a great deal more research is needed before the effects of various regionalization techniques upon the set $[\alpha]$ are known or fully understood.

A further nuisance, again deriving its origin in statistical

estimation, may be encountered in generating synthetic sequences. The mathematical structure of the model may be inconsistent with the parameter estimates derived from historic sequences. (Matalas and Wallis, 1971a ; Slack, 1973). However, more appropriate estimation techniques may be employed (Crosby and Maddock, 1970 ; Valencia and Schaake, 1972) to circumvent this problem.

1.3 Water Resource System Design and Operation

In the past twenty years there has been a tremendous increase in the world's population attended by a consequent threat of a shortage of a number of natural resources. Water resources are not an exception, and the demand for more accurate evaluation of water resource system design has become more acute in the face of increasing demands for water products and services. The design problem has in the past fallen largely within the domain of engineering hydrology; however, in recent years economic theory has come to play a major role in system planning. The interacting roles of engineering hydrology and economics can best be studied through the methodology of systems analysis or operations research, which has now become widely accepted in the planning and management of water resources. The increasing use of techniques of optimization and simulation, underlying systems analysis, in the field of water resources, has been closely linked with the development of high speed large memory digital computers. Techniques of simulation and optimization, which afforded a study of the interactions between engineering hydrology and economics, were perhaps first applied on a large scale by the Harvard Water Program (Maas et al., 1962).

The general simulation procedure proposed by the Harvard group consisted essentially of the following steps:

1. Define the probability distribution and temporal and spatial interdependency assumptions. Essentially this step represents the specification of a generating process in time and space.
2. Abstract statistical estimates of the parameters of the generating process from historic sequences of natural streamflow. This step involves the approximation of the generating process.
3. Use the approximated generating process in conjunction with a randomization routine to produce several equally likely sequences of synthetic streamflows of the required length.

The alternative assumptions which can be made about interdependency assumptions and probability distributions have been discussed in sections (1.2.2) and (1.2.3), respectively. The length of each synthetic streamflow sequence depends on the operational horizon, and will generally be of the order of 50 - 100 years.

An excellent review of the conjoint use of simulation and optimization techniques in water resource system planning and management has been given by Roefs (1968), who characterizes the planning process as the following sequence of decisions which the planner must take:

1. the specific project to be built
2. the time at which the project should be built
3. the size to which the project should be built
4. the target output which should be set
5. the operation rules for the project
6. the "real-time" operational control decisions

The first three decisions may be characterized as planning decisions. The fourth decision may be regarded as a risk allocation decision, while the fifth decision is the operation plan. The sixth decision may be viewed as a "real-time" management decision, which is

the adjustment of decisions within a finer time scale to fit the operation rules. (Roefs, 1968). However, the planning process is complicated by the fact that the planning decisions are interdependent, and are also dependent on the fourth and fifth decisions. In other words the project size cannot be decided without considering the definition of an optimal target for that project and the definition of an optimal set of operating rules for that project and target.

In the situation of an existing project with existing targets, the problem reduces to the manageable task of finding operating rules. A straight simulation approach may be adopted, or simulation may be combined with a deterministic optimization technique. The former approach is described in block diagram form in figure (1.1) and may be summarised as follows (Roefs, 1968):

1. simulate several sequences of streamflows
2. define a set of operational decisions
3. test the defined decision set against each of the simulated sequences
4. repeat steps 2 and 3 several times
5. compare the benefits and choose the best decision set of those run.

However, if the project is large, the number of defineable decision sets may be such that the simulation scheme will not reach optimality within any reasonable time.

Incorporation of a deterministic optimization routine into the simulation scheme results in a more attractive approach for defining operating rules which is given in block diagram form in figure (1.2) as the following sequence of steps (Roefs, 1968):

1. simulate a sequence of streamflows
2. find the optimal operation policy (set of releases) for that sequence
3. save the results

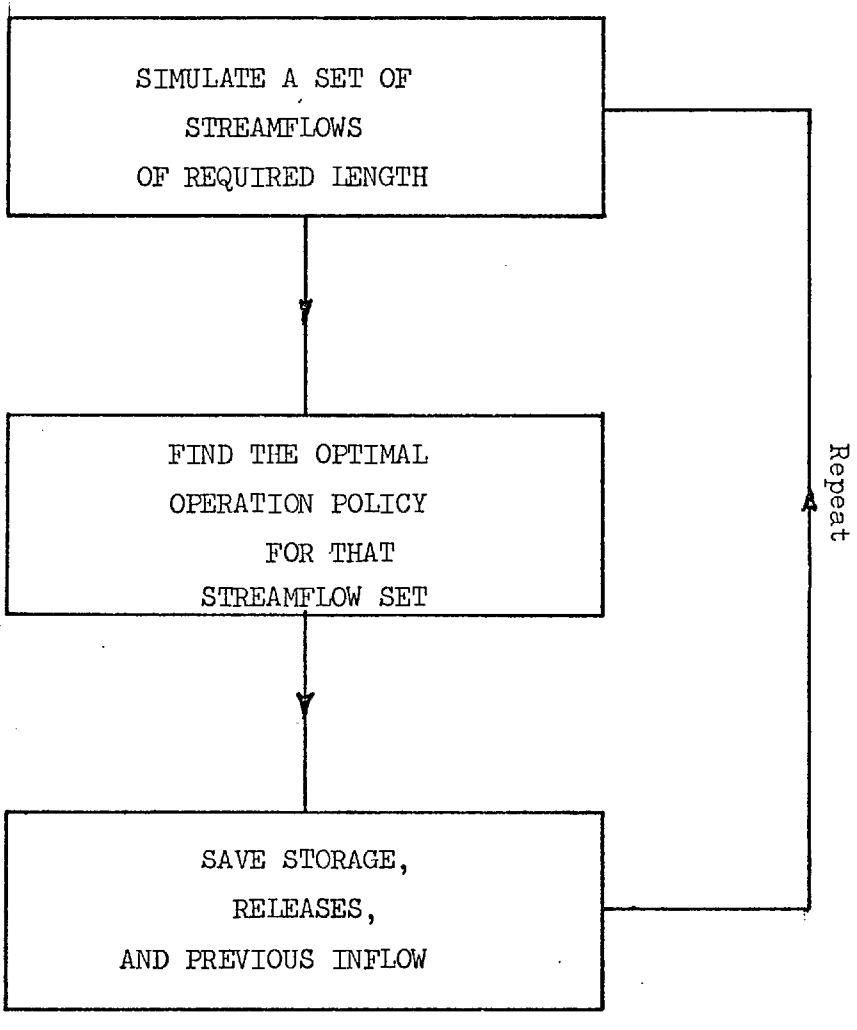


Figure (1.1) (after Roefs, 1968)

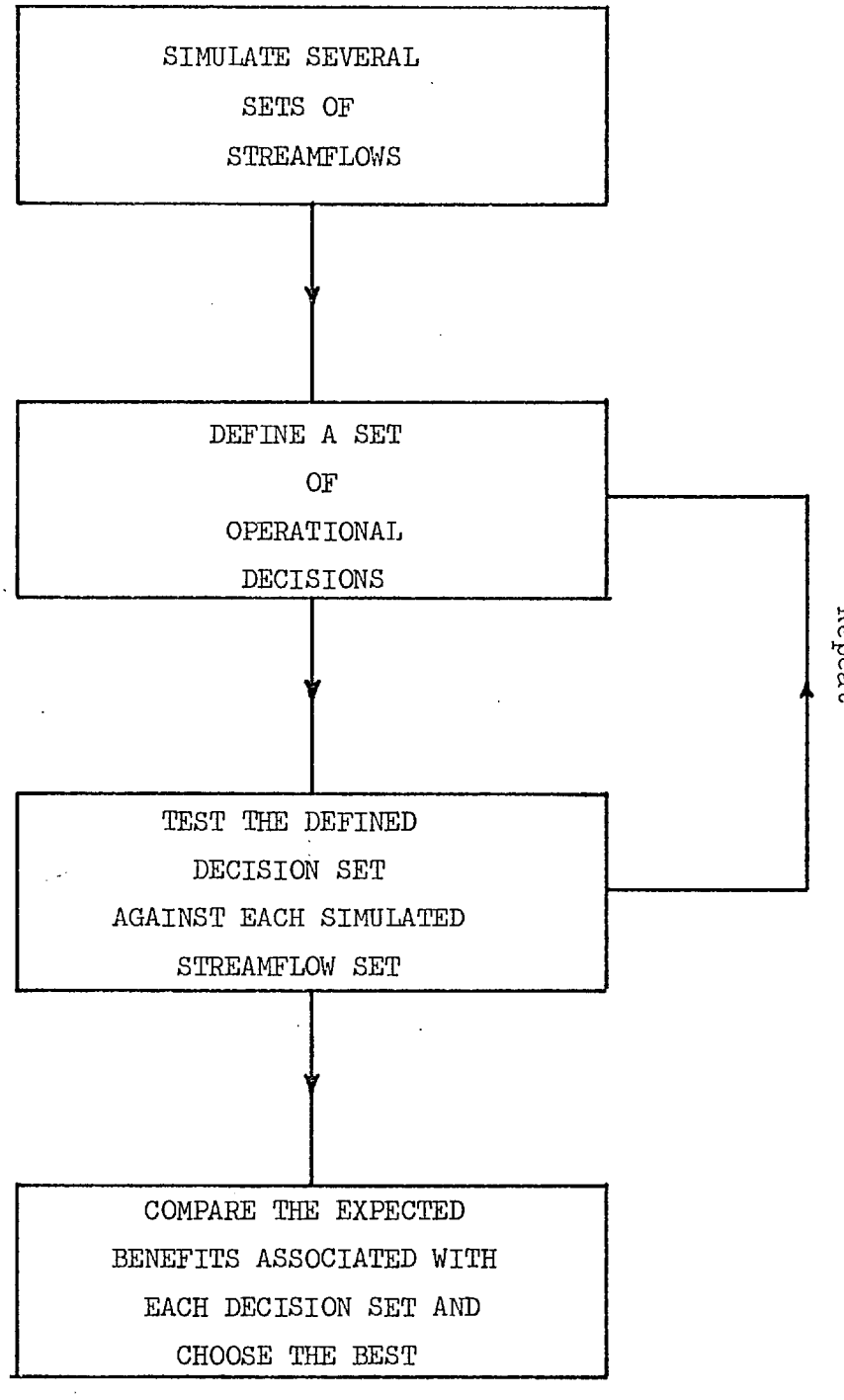


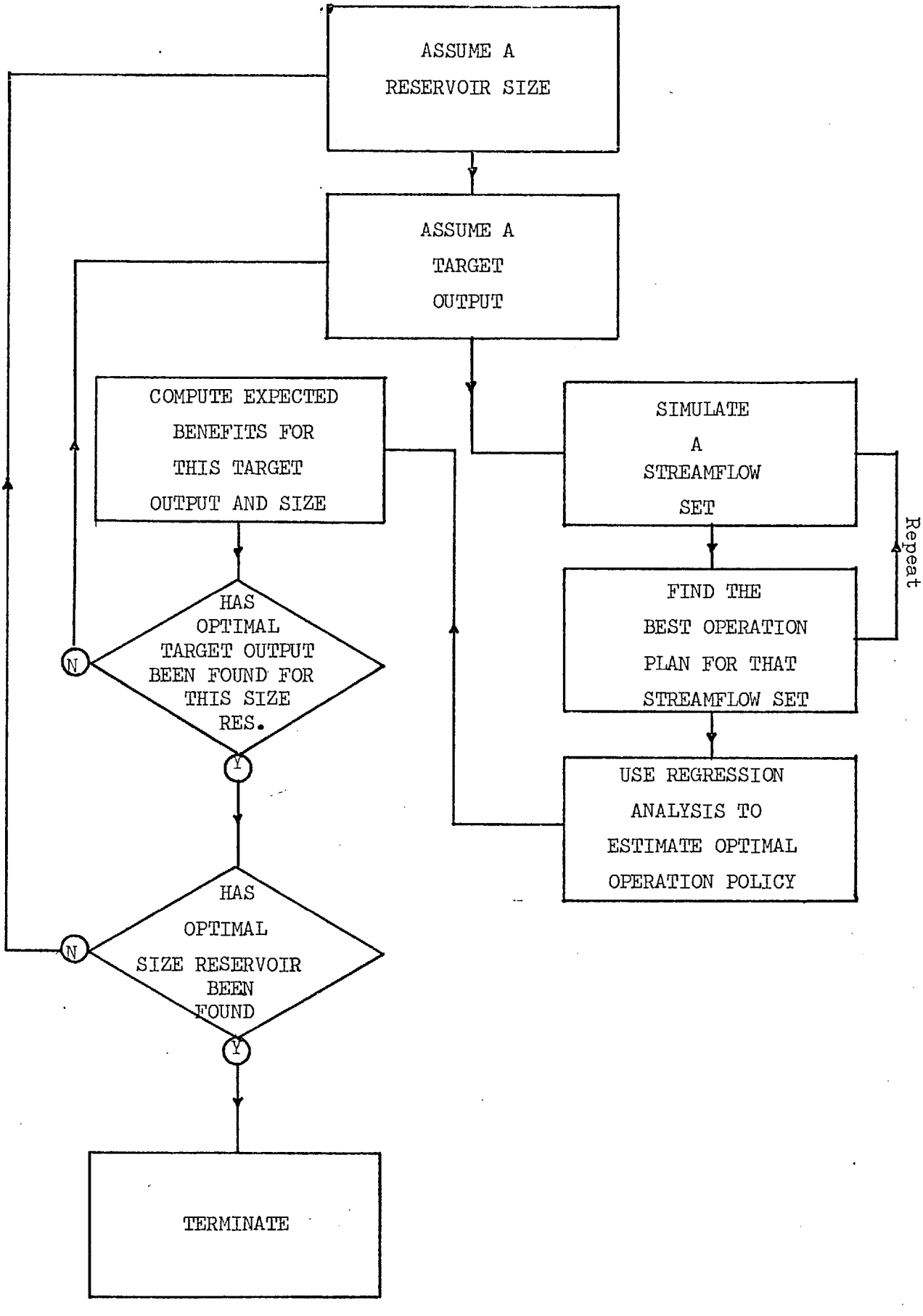
Figure (1.2) (after Roefs, 1968)

4. repeat steps 1,2 and 3 several times
5. perform a regression analysis in which release is the dependent variable and storage and inflow are the independent variables

Suitable contenders for the role of optimization routine are linear and dynamic programming. It is worthwhile noting that given a streamflow sequence the problem of determining optimal operating rules using linear or dynamic programming is a deterministic one; the stochastic nature of streamflow is recognised through simulating several equally likely sequences.

In the design situation where a water resource system does not already exist, the overall optimal design and operation of the system is sought. In general, the problem will be too formidable to permit a solution, although simulation affords a promising course of action which may be viable in certain cases. If, for example, in the case of a single reservoir system, the reservoir site and timing of construction happened to be known, then the reservoir size and target output could be defined through an iterative process suggested by Roefs (1968), which is given in block diagram form in figure (1.3) as the following series of steps:

1. assume a reservoir size;
2. assume a target output.
3. For this reservoir size and target yield:
   a. simulate a streamflow sequence;
   b. find the best operation plan for that streamflow set
   c. repeat steps a and b several times;
   d. use a regression analysis to determine the optimum operation plan
4. estimate benefits for the assumed target output and size
5. repeat steps 2 - 4 until the optimal target output for a given



Figure(1.3) (after Roefs, 1968)

size has been found;

6. estimate benefits for the reservoir at this size

7. repeat steps 1 - 6 until the optimal reservoir size is found.

Essentially, the previously defined optimization-simulation technique for defining operating rules is embedded within the overall iterative process.

In a more complex design situation, the number of alternative designs which can be defined will generally be enormous, and the design problem and problem of finding operating rules have generally been tackled separately. Nevertheless, in investigating alternative designs, the straight simulation approach has been widely used, mainly because the number of feasible designs can be reduced considerably through the skill and experience of the planner. However, it will generally be impossible to prove formally that the design is optimal.

All the foregoing simulating schemes require a set of streamflow sequences, each having the same length as the operational horizon. The size of the set will depend on the level of statistical confidence required in the results, and on the inherent variability of the problem, which in turn is a function of the streamflow regime, the level of development and the shape of benefit or penalty functions. Large projects in arid regions with irregular benefit functions would be expected to exhibit a high inherent variability.

A further technique for finding operating rules is that of stochastic optimization, which does not require simulation. The stochastic nature of streamflow is taken account of through probability and interdependency matrices which are incorporated into a single step optimization algorithm. However, the scope of this approach is limited on the grounds of computational feasibility, even for a

single reservoir case; to a much lesser extent, the scope of dynamic programming is also limited on similar grounds.

Recent advances in water resource system planning suggest that synthetic hydrology will continue to play an important role. Multi-regional planning models are being formulated which utilize mixed integer programming for selecting and scheduling for construction a minimum cost configuration of water resources projects in Puerto Rico. (Moody, 1973). Uncertainties enter into the model from a number of sources, one of which is the stochastic nature of streamflow, and which may be handled through synthetic hydrology. Synthetic streamflow sequences are also being used in the design of water resource systems for England and Wales (Sexton and Jamieson, 1973).

1.4 Decision Theory

In the preceding sections, the main advantages and limitations of synthetic hydrology have been discussed. Some of the limitations arise from uncertainties resulting from lack of information, particularly with respect to model choice and distribution choice. Decision theory may help in the future to provide solutions to the designer's dilemma in these situations.

Consider the issue of distribution choice. Sufficient information may not be contained in the historic sequence to distinguish between proposed alternative distributions. Fiering and Jackson (1971) advocate the following use of economic considerations and decision theory as a solution to this dilemma. Suppose the planner has narrowed the distributions under consideration to two. Designs using synthetic streamflows based on each distribution are then evolved. An optimal design corresponding to each distribution is identified via some economic objective function. The next step is to evaluate the consequences, in terms of economic loss, of accepting the design corresponding to

the first distribution, if the second were the true distribution. Conversely, the economic loss involved in assuming the second distribution when the first was, in fact, the true one must be evaluated. The optimal design will suggest the appropriate distribution for the flows. A similar strategy may be applied to the problem of model choice, and an attempt has been made to implement such a strategy by O'Connell and Wallis (1973). Indeed decision theory might also be applied to isolate the members of the set $[\alpha]$ to which the design is insensitive. For example, what is the economic loss associated with the non-preservation of kurtosis in generated flows when a design has been evolved using flows preserving kurtosis?

1.5. Summary

Within the past decade, synthetic hydrology has become more sophisticated as a design procedure. New models have been introduced which enable the preservation of a broader spectrum of characteristics of historic sequences. In the future, the interaction of decision theory and synthetic hydrology may result in the emergence of a more powerful design methodology.

Chapter 2

THE DEVELOPMENT OF SYNTHETIC HYDROLOGY

Since the pioneering work of Thomas and Fiering (1962), there has been an explosive growth in the volume of literature pertaining to the use of simulation techniques within the field of stochastic hydrology. A considerable portion of the literature relates to the adoption, for simulating synthetic streamflows, of models which already existed within the literature on stochastic processes. Considerable emphasis has been laid on the ability of the models to preserve important properties of the historic flows within sequences of synthetic flows. On the other hand, stochastic hydrology has yielded some new models specifically developed because existing documented models were inadequate. Indeed, hydrologists all too frequently assume that models applicable within other fields are readily applicable to hydrological processes without giving due consideration to the physical context of the observed data.

In this review, only literature pertaining to the development and application of stochastic models primarily for the simulation of synthetic streamflow sequences will be considered. Some historical pre-computer attempts at generating synthetic streamflows are described in section (2.1) while section (2.2) traces early developments in generating synthetic flows on a digital computer. The range of cumulative departures, Hurst's law, the Hurst phenomenon and various theoretical studies of the range are discussed in section (2.3), while section (2.4) describes early attempts at studying the behaviour of the range through simulation. In section (2.5) fractional Gaussian noise is reviewed, under the headings of (1) theoretical properties (2) approximations proposed and simulation experiments carried out by

Mandelbrot and Wallis (1969a,b,c,e) and (3) estimation of h, the Hurst coefficient. Further approximations to fractional Gaussian noise, namely, fast fractional Gaussian noise, the ARIMA (1,0,1) process and the Broken Line process are discussed in section (2.6) while some further recent developments in synthetic hydrology are outlined in section (2.7).

2.1 Historical Developments

The problem of determining the reservoir capacity necessary to meet a given demand was studied almost a hundred years ago by Rippl (1883), and this same problem has been largely responsible for the advent and subsequent development of techniques for generating synthetic streamflows. Rippl's approach relies entirely on a mass curve analysis of the historic record, and while the approach may provide useful information on the behaviour of low yield reservoirs, the method is open to serious criticism when over-year storage is involved. Fiering (1967) cites some of the deficiencies of the Rippl method as follows:

(1) the analysis is based solely on the historical record even though the same flow sequence is unlikely to recur in the future

(2) the mass diagram provides no information on the risk of water shortages in periods of low flow; and

(3) the length of the historical record and the economic life of the structure under design are likely to differ; since the storage capacity indicated by a Rippl diagram tends to increase with length of record, the estimated capacity may not be compatible with the active life of the scheme.

The deficiencies in Rippl's approach were recognised from an early

stage. Hazen (1914) attempted to increase the lengths of historic records by compiling a 300-year synthetic record through combining the data from 14 streams; each record of annual flows was standardized by dividing each value by the mean annual flow of the particular river; the records were then combined in order of increasing coefficient of variation. The synthetic record was then used in compiling design charts giving reservoir capacities for specific coefficients of variation of streamflow and projected level of development which attempted to overcome deficiencies (1) and (2) in the mass curve procedure. However, the synthetic record still only represented a limited range of conditions, and much more important, ignored and destroyed the autocorrelation structure of the flows.

Sudler (1927), again grappling with the storage problem, wrote 50 annual runoff values on to cards, shuffled and dealt the cards 20 times to give a 1000 year record. While the sequence of occurrence of the flows is altered, the autocorrelation structure was again destroyed, and all 20 sequences had the same moments.

Barnes (1954) was the first to use proper random sampling in simulating artificial streamflow sequences. He used a table of random numbers to sample from a normal distribution with the same mean and variance as the historical data. However, the approach again ignored the autocorrelation structure of the flows.

2.2 Early Developments in Synthetic Hydrology

Thomas and Fiering (1962) were apparently the first to combine the use of an electronic digital computer, the sampling of random numbers on the computer, and a model which took account of the dependence and seasonality in streamflow, in order to generate artificial sequences

of streamflows, called synthetic streamflows, for use in the design of water resource systems. The model, based on monthly streamflow, takes account of seasonality in the monthly means, standard deviations and lag-one autocorrelation coefficients through treating the flows for each month as a separate population. Thus, three parameters are required to specify the model for each month, which may be written as

$$q_{i+1} = \mu_{j+1} + b_j(q_i - \mu_j) + \sigma_{j+1}(1 - \rho_j^2)^{\frac{1}{2}} \varepsilon_{i+1} \qquad (2.1)$$

where q_{i+1} and q_i are the flows during the $(i+1)^{th}$ and $(i)^{th}$ month, respectively; μ_j and μ_{j+1} are the mean monthly flows for month j and j+1 where j = 1,2,...,12. The term b_j is a regression coefficient derived from regressing the flows of month (j+1) on those of month j, and $\varepsilon_{i+1}\sigma_{j+1}(1-\rho_j^2)^{\frac{1}{2}}$ is the error term in the regression. If the flows are normal or have been transformed to assure normality, then ε_{i+1} is a standard normal deviate with zero mean and unit variance. Alternatively, the model may be written as

$$\frac{q_{i+1} - \mu_{j+1}}{\sigma_{j+1}} = \rho_j \frac{q_i - \mu_j}{\sigma_j} + (1 - \rho_j^2)^{\frac{1}{2}} \varepsilon_{i+1} \qquad (2.2)$$

or in terms of standardized deviates

$$z_{i+1} = \rho_j z_i + (1 - \rho_j^2)^{\frac{1}{2}} \varepsilon_{i+1} \qquad (2.3)$$

which emphasises the underlying Markovian nature of the model.

In applying the models, estimates of μ_j, σ_j, and ρ_j, j = 1,2,...,12, are derived from the historic flows, and are then used in generating synthetic flows together with ε_{i+1} values which are sampled from a normal distribution with zero mean and unit variance, thus ensuring the stochastic nature of the flows.

Thomas and Fiering (1962) also proposed a similar form to equation (2.1) for modelling the spatial correlations existing between concurrent

streamflow records; this model, together with other multisite generating mechanisms, is discussed in chapter 4.

In an important contribution, Matalas (1967) gave the methodology for generating synthetic streamflow a sound mathematical basis, and discussed some important issues such as operational bias and the importance of synthetic flows conforming to a particular distribution. The lag-one Markov process, specified with mean μ_x, variance σ_x^2 and lag-one autocorrelation ρ_x as

$$(X_t - \mu_x) = \rho_x(X_{t-1} - \mu_x) + (1 - \rho_x^2)^{\frac{1}{2}} \sigma_x \varepsilon_t \qquad (2.4)$$

was considered as the basic generating process for synthetic flows. Matalas illustrated how resemblance between historic and synthetic flows in terms of skewness could be achieved in a number of ways, illustrating the approach in all cases with the lag-one Markov process. One approach consisted of replacing the random component ε_t in equation (2.4) with a random component η_t defined as

$$\eta_t = \frac{2}{\gamma_\eta}\left(1 + \gamma_\eta \frac{\varepsilon_t}{6} - \frac{\gamma_\eta^2}{36}\right)^3 - \frac{2}{\gamma_\eta} \qquad (2.5)$$

where the skewness of η_t, denoted by γ_η, is related to the skewness of the process X_t, γ_x, by

$$\gamma_x = \frac{(1 - \rho_x^2)^{\frac{3}{2}}}{(1 - \rho_x^3)} \gamma_\eta \qquad (2.6)$$

Assuming that ε_t is a normal independent process with zero mean and unit variance, denoted by NIP(0,1), then X_t will be approximately distributed as gamma with zero mean, unit variance and skewness γ_x. From an estimate of γ_x, denoted as $\hat{\gamma}_x$, an estimate of γ_η is defined via equation (2.6) whence the component η_t may be generated using equation (2.5).

Equation (2.4) with ε_t replaced by η_t then yields flows which are approximately distributed as gamma. Matalas (1967) also proposed a technique for generating flows which are exactly distributed as gamma, but only for particular values of γ_x.

Synthetic flows which conform to a three-parameter log-normal distribution, and which resemble historic flows in terms of $\hat{\mu}_x$, $\hat{\sigma}_x$, $\hat{\gamma}_x$ and $\hat{\rho}_x$ may be generated as follows (Matalas, 1967). If a is assumed to be the lower bound of a variate X_t, where $(X_t - a)$ is log-normally ditributed, then $Y_t = \ln(X_t - a)$, where ln denotes $\log_e$, is normally distributed. The mean μ_x, variance σ_x^2 and skewness γ_x are related to the lower bound a and to the mean μ_y and variance σ_y^2 of Y_t by

$$\mu_x = a + e^{[\sigma_y^2/2 + \mu_y]} \tag{2.7}$$

$$\sigma_x^2 = e^{[2(\sigma_y^2 + \mu_y)]} - e^{[\sigma_y^2 + 2\mu_y]} \tag{2.8}$$

$$\gamma_x = \frac{e^{3\sigma_y^2} - 3e^{\sigma_y^2} + 2}{(e^{\sigma_y^2} - 1)^{3/2}} \tag{2.9}$$

(Aitcheson and Brown, 1957). If the flows Y_t are generated by a lag-one Markov process with mean μ_y, variance σ_y^2 and lag-one autocorrelation ρ_y, then ρ_x is given as

$$\rho_x = \frac{e^{\sigma_y^2 \rho_y} - 1}{e^{\sigma_y^2} - 1} \tag{2.10}$$

which may easily be inverted to yield ρ_y as a function of σ_y and ρ_x; however, the flows in X space will derive from a non-linear process, and thus will not constitute a lag-one Markov process.

The procedure for generating synthetic flows such that they will resemble historic flows in terms of $\hat{\mu}_x$, $\hat{\sigma}_x$, $\hat{\gamma}_x$ and $\hat{\rho}_x$ is as follows.

Estimates of μ_y, σ_y, ρ_y and a are derived from the estimates $\hat{\mu}_x$, $\hat{\sigma}_x$, $\hat{\gamma}_x$ and $\hat{\rho}_x$ through solving equations (2.7 - 2.10). Using a lag-one Markov process formulated in terms of $\hat{\mu}_y$, $\hat{\sigma}_y$ and $\hat{\rho}_y$, synthetic flows are generated in Y-space, whence exponentiation and the addition of the constant a yields the flows X_t.

An alternative approach would be to transform the flows into log-space first before estimating μ_y, σ_y and ρ_y in the usual fashion from a sequence of Y-values. However, synthetic flows will then resemble historic flows in terms of those estimates $\hat{\mu}^*_y$, $\hat{\sigma}^*_y$ and $\hat{\rho}^*_y$ but not in terms of $\hat{\mu}_x$, $\hat{\sigma}_x$ and $\hat{\rho}_x$, which would appear to be the required resemblance from a physical standpoint. However, the choice of an estimation procedure must ultimately depend on the loss which may result from basing a decision on the incorrect value of a parameter rather than on its true value. If the question of which estimation procedure should be used cannot be resolved analytically i.e. the loss function cannot be formulated directly in terms of that parameter, then in a design situation, investigation of either approach through simulation may be the only answer. Matalas (1967) discussed the issue of distribution choice for the flows, and suggested that the impact of alternative distributions on the design situation be investigated in a similar fashion. In this context, a decision theoretic framework for assisting in resolving this issue has been proposed by Fiering and Jackson (1971) and has been discussed in section (1.4).

Matalas (1967) also discussed problems relating to parameter estimation, and noted that the magnitude of the operational bias discussed in section (1.2.4) depends on the standard error of estimate of the parameter. Regionalization (Benson and Matalas, 1967) affords a means of reducing the standard error of a particular parameter through providing a regionalized estimate with smaller variance. However, the

relationships existing between streamflow parameters and physiographic and meteorological parameters of a river basin have neither been adequately quantified or fully understood to date, so that the worth of this suggested method of reducing operational bias is largely unknown. Maximum likelihood techniques have also been suggested because they provide large sample estimates which are asymptotically unbiased and have smaller variance than all other estimates. However, it does not necessarily follow that these properties hold within small samples, and indeed little is known of the small sample properties of maximum likelihood estimates. Further, the derivation of maximum likelihood estimates requires an assumption about the underlying probability distribution of the flows, and different assumptions will provide different numerical values for the parameters. Maximum likelihood estimates will generally be more difficult to derive than moment estimates, as a system of non-linear equations will frequently have to be solved numerically.

Matalas (1967) also presented a multivariate lag-one Markov process for generating synthetic flows resembling historic flows in terms of spatial as well as temporal correlations, but the details of this and other multivariate approaches will be presented in chapter 4.

While low order moments provide a convenient means of structuring a generating process, the process may not perform satisfactorily from an operational viewpoint, i.e. long synthetic sequences may not on average reflect more extreme events than short historical sequences. In the context of storage design, events might be the duration and intensity of drought periods, quantities for which an explicit resemblance between historic and synthetic sequences is generally very difficult to define. However, one such quantity which is closely related to storage design and which has been studied within the framework

of mathematical statistics is the range of cumulative departures from the sample mean, first introduced into hydrology by Hurst (1951, 1956). Some of the proponents of synthetic streamflow generation (Matalas and Huzzen, 1967; Fiering, 1967) attempted to examine the operational performance of the lag-one Markov process through comparing the behaviour of the range of cumulative departures for historic and synthetic streamflow sequences. A definition of the range of cumulative departures now follows, together with a brief review of the work done on the range up to the time when investigators sought to reconcile the behaviour of the range for observed historical sequences with the expected behaviour of the range for the then current models of streamflow.

2.3 The Range of Cumulative Departures

In order to help in the physical interpretation of the range, the following formulation is adopted. Let $X_1, X_2, \ldots, X_n$ denote a sequence of annual inflows into a reservoir over n years. Let $\bar{X}$ denote the sample mean and define

$$D_n = \sum_{i=1}^{n} X_i$$

where D_n represents the total amount of water flowing into the reservoir in the n years. To maintain a constant outflow equal to the average inflow would require the removal of $\frac{1}{n}\sum_{i=1}^{n} X_i$ each year. With such a rate of removal, $\frac{k}{n}\sum_{i=1}^{n} X_i$ is the amount removed over the first k years. Hence

$$D_k^* = \sum_{i=1}^{k} X_i - \frac{k}{n}\sum_{i=1}^{n} X_i$$

$$= D_k - kD_n/n$$

represents the excess or deficiency relative to the amount removed up to the k^{th} year. Defining

$$P_n = \max D_k^* \qquad 1 \leq k \leq n$$

and $$Q_n = \min D_k^* \qquad 1 \leq k \leq n$$

as the largest excess and greatest deficiency respectively over the steady outflow during the n years, the quantity

$$R_n = P_n - Q_n$$

is known as the range of cumulative departures from the sample mean. The "ideal" reservoir represented by this concept becomes extremely large over a long time span and accordingly is never economically justifiable. However, the variation of R_n with n provides an estimate of the yield that might be maintained from any given capacity of reservoir. R_n is sometimes referred to as the adjusted range, and, for notational simplicity, will be referred to as R.

2.3.1 Hurst's Law and the Hurst Phenomenon

Interest in the statistic R was stimulated by Hurst's studies (1951, 1956) of long-term storage requirements on the River Nile. For approximately 900 annual time series comprising streamflow and precipitation records, stream and lake levels, tree rings, mud varves, atmospheric pressure and sunspots, Hurst found R to vary with n as

$$\frac{R}{S} \sim n^h \tag{2.11}$$

where S denotes the sample standard deviation of the time series of length n, and h is a constant. The quantity R/S will be referred to as the rescaled range so as to distinguish it from the adjusted range, R. The coefficient h was estimated through the following relationship

$$\frac{R}{S} = \left(\frac{n}{2}\right)^K \tag{2.12}$$

where K denotes the resulting estimate of the population coefficient, h.

Thus K was defined for each time series as

$$K = \frac{\text{Log } R - \text{Log } S}{\text{Log } n - \text{Log } 2} \qquad (2.13)$$

and ranged from 0.46 to 0.96 with a mean of 0.729 and a standard deviation of 0.092 over all phenomena. The records employed by Hurst varied in length from 30 to 2000 years.

Hurst (1951) employed some simple coin tossing experiments to arrive at a theoretical value for the expected or mean value of R for a normal independent process, which he gave as

$$E[R] = \left[\frac{n\pi}{2}\right]^{0.5} \sigma \qquad (2.14)$$

or $$\frac{E[R]}{\sigma} = 1.25\, n^{0.5}$$

Independently, Feller (1951), using the theory of Brownian motion derived equation (2.14) which is an asymptotic result, but without invoking the assumption of normality for the underlying process. The form of equation (2.12) which was used by Hurst for estimating h, was apparently suggested by equation (2.14) wherein the term (n/2) appears.

The disagreement between the average value of K, 0.73, derived by Hurst and the value of the exponent in equation (2.14), 0.5, puzzled many statisticians and engineers alike at the time, and a number of suggestions were put forward as to the origin of the discrepancy. Hurst (1951) himself conjectured that non-randomness as evidenced by the tendency in natural time series for high values to follow high values and low values to follow low values might be a possible explanation. Langbein (1956), in a discussion on Hurst's (1956) paper, reinforced Hurst's work with the observation that the variance of the sample mean for Hurst's data tended to exhibit greater variation than if the data were random, and could be approximated by the relationship

$$S_{\bar{x}} = \frac{S}{n^{0.28}} \qquad (2.15)$$

where $S_{\bar{x}}$ denotes the standard deviation of subsample means and S is the standard deviation of the sample. As the mean value of the exponent K found by Hurst was approximately 0.72, equation (2.15) suggested that

$$\frac{S_{\bar{x}}}{S} = n^{h-1} \tag{2.16}$$

Langbein also deduced that skewness was an unlikely explanation, as was serial correlation of a Markovian nature. He suggested that the persistence in Hurst's data was of a more complex nature.

Feller (1951) suggested that the discrepancy might be accounted for by autocorrelation of a Markovian nature in the time series, but Barnard (1956) discredited this suggestion, pointing out that no simple set of correlations could account for Hurst's result. Skewness was at that time considered fleetingly as a possible reason, but Feller's result and the fact that a considerable number of Hurst's time series followed the Gaussian distribution tended to rule out this possibility. Hence, no acceptable explanation was forthcoming.

Subsequent to the theoretical work of Feller (1951), a number of further theoretical results were derived for the adjusted range R and related statistics. Anis and Lloyd (1953) derived some results for finite n, but defined the range as follows. For the sequence $X_1, X_2, \ldots, X_n$, the sequence of partial sums $D_1, D_2, \ldots, D_n$ was defined as

$$\begin{aligned} D_1 &= X_1 \\ D_2 &= X_1 + X_2 \\ &\vdots \\ D_n &= X_1 + X_2 + \ldots X_n \end{aligned}$$

From the partial sums the quantities

$$M_n = \max (0, D_1, D_2, \ldots, D_n)$$

$$m_n = \min (0, D_1, D_2, \ldots, D_n)$$

are defined whence a definition of the range, denoted as R_p emerges as

$$R_p = M_n - m_n \tag{2.17}$$

R_p is sometimes referred to as the population range, as its definition implies knowledge of the population mean of the process.

Feller (1951) had observed that the sampling properties of R_p were inferior to those of R, and in any case, R rather than R_p was the definition used by Hurst. However, Anis and Lloyd showed that the mean value of R_p for finite n and independent variates with a common distribution was given as

$$E[R_p]_n = \sqrt{\frac{2}{\pi}} \sum_{r=1}^{n-1} r^{-\frac{1}{2}} \tag{2.18}$$

Anis (1955, 1956) derived the higher moments of M_n for small n. Solari and Anis (1957) reverted again to working with R and derived the mean and variance of R for finite n for the case where the underlying process is normal. For the case of a common mean and unit variance for the underlying process and $n \geq 2$, they showed that

$$E[R]_n = \sqrt{\frac{n}{2\pi}} \sum_{s=1}^{n-1} s^{-\frac{1}{2}} (n-s)^{-\frac{1}{2}} \tag{2.19}$$

which has the asymptotic value $(\frac{n\pi}{2})^{\frac{1}{2}}$ derived by Hurst (1951) and Feller (1951). Feller's comment on the greater sampling stability of R over that of R_p was also borne out by Solari and Anis' result for the variance of P_n when compared with the results derived by Anis (1955).

However, while the work of Anis and Lloyd (1953), Anis (1955, 1956) and Solari and Anis (1957) was of some interest, it could not explain the discrepancy between the average value of 0.73 for the exponent K in equation (2.12) obtained from the extensive data analysis of Hurst. The assumption of h = 0.5 was inherent in all the analyses, and, in any

case, Hurst had worked with the statistic R/S rather than R or R_p, but the importance of this point was not to be realised until much later. Nevertheless, the problem still continued to niggle statisticians such as Moran (1959) who wrote reinforcing Barnard's (1956) earlier contention, ... "the exponent 0.73 could not occur unless the serial dependence were of a very peculiar kind because with all the plausible models of serial dependence, the series of partial sums is always approximated by a Bachelier-Weiner process when the time scale or economic horizon is sufficiently large."

Moran (1964), working again with the quantity R_p, deduced that for moderate n, Hurst's result could, after all, be explained by skewness, and used highly skewed distributions with very large second moments about the mean to buttress his argument. He noted that such distributions were unlikely to provide a good fit to Hurst's data, but suggested that distributions with large but finite second moments (such as a Cauchy distribution with a truncated density) would still provide the required behaviour in R_p. His analysis, however, did not permit any conclusions to be drawn about the behaviour of R/S.

In a lucid summary of the work that had been done up to that time on the range, Lloyd (1967) noted that, ... "We are then in one of those situations, so salutary for theoreticians in which empirical discoveries stubbornly refuse to accord with theory. All of the researches described above lead to the conclusion that in the long run E[R] should increase like $n^{0.5}$, whereas Hurst's extraordinarily well documented empirical law shows an increase like n^K where K is about 0.7. We are forced to the conclusion that either the theorists interpretation of their own work is inadequate or their theories are falsely based: possible both conclusions apply."

Lloyd coined the term "Hurst phenomenon" to describe the discrepancy between Hurst's average value of 0.73 for K and the expected value of 0.5 for h deduced from theory. He examined closely the behaviour of equation (2.19) for small to moderate n and concluded that the asymptotic value of 0.5 was approached too quickly to explain Hurst's findings, but conceeded that for an underlying non-normal distribution there was a possibility that convergence to 0.5 might be extremely slow. This was in accord with Moran's (1964) explanation but both Lloyd and Moran had for some strange reason confined all considerations to the behaviour of R and R_p while Hurst had worked only with the rescaled range R/S. Hence, any explanation based on R or R_p would have to presume that the behaviour of R/S closely resembled that of R and R_p.

2.4 Synthetic Hydrology and the Range

Up to the time of Moran's (1964) contribution, studies of the range had been confined to the domains of data analysis and mathematical statistics. However, simulation had emerged as a new tool whereby the Hurst phenomenon could be investigated more exhaustively. Matalas and Huzzen (1967) and Fiering (1967) were the first to investigate some properties of the rescaled range within the framework of synthetic hydrology, which was at that time in the early stages of development. Matalas and Huzzen carried out some sampling experiments on the rescaled range R/S using the lag-one Markov process given in equation (2.4). Employing both normal and log-normal distributions for the process, they generated sequences of various lengths with lag-one autocorrelation in the range $0 < \rho_1 < 1$, and estimated the Hurst coefficient h using the estimator proposed by Hurst as given by equation (2.13). The results of their simulations for the case of an underlying normal distribution

for the process are given in table (2.1).

n \ ρ_1	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
5	.59	.60	.61	.62	.63	.63	.64	.65	.65	.66
10	.66	.67	.69	.70	.73	.74	.76	.77	.79	.80
25	.64	.67	.70	.72	.74	.76	.79	.81	.84	.86
50	.63	.66	.68	.71	.73	.75	.77	.80	.84	.87
100	.61	.65	.66	.69	.71	.73	.75	.79	.82	.87
500	.59	.61	.62	.64	.66	.68	.70	.73	.77	.82
1000	.58	.60	.61	.63	.65	.67	.69	.71	.75	.80

Table (2.1): Values of $\widetilde{E}[K]_n$ from 10000 realizations from a lag-one Markov Process. (After Matalas and Huzzen, 1967)

For sample size n in the range $25 \leq n \leq 100$ and $0.1 < \rho_1 < 0.4$, the estimated expected value of K, $\widetilde{E}[K]_n$ is about 0.7, which the authors noted to be roughly compatible with results derived from some annual streamflow sequences assembled by Yevjevich (1963). However, the authors indicated that their results offered no substantial proof that the lag-one Markov process could be used to model annual streamflow. The variation of $\widetilde{E}[K]_n$ with n in table (2.1) suggests that the process cannot adequately model Hurst's law as given in equation (2.11), as a plot of log (R/S) against log (n) would not be linear. It is also evident from the results corresponding to $\rho = 0.0$ that Hurst's method of estimating h results in a positive bias.

The authors conducted the same set of experiments reported in table (2.1) for cases where the underlying distribution was log-normal with skewness coefficient γ_x in the range $0.2 \leq \gamma_x \leq 2$. They observed that the effect of skewness on the values of $\widetilde{E}[K]_n$ was negligible and concluded that skewness offered no explanation of the Hurst phenomenon.

Fiering (1967) carried out an extensive analysis of the behaviour of R/S for single lag and multi-lag autoregressive processes, as part of a general approach to reservoir design. Attempts to reconcile results derived from these models with the findings of Hurst, and the

behaviour of R/S for some U.S. streams proved futile. It was found that a 20-lag model was necessary to ensure that Hurst's law held for values of n up to about 60. As in the case of the Matalas and Huzzen simulations, a relationship giving R/S as an invariant power of n was not realized, in contradiction to Hurst's findings. Numerical problems prevented an extension of Fiering's approach to models with a higher number of lags than 20.

Yevjevich (1967) performed some simulations involving the population range R_p, and extended Anis and Lloyd's (1953) result, equation (2.18), to embrace models exhibiting linear dependence, such as autoregressive and moving average schemes. He then verified some theoretical results using extensive simulation experiments. However, no attempt was made to relate this work to that of Hurst: indeed because of the definition of the range employed, this would have been very difficult.

2.5 Fractional Gaussian Noise

In a significant contribution to stochastic hydrology, Mandelbrot and Wallis (1968, 1969a,b,c,d,e) proposed a rigourous theory which adequately accounted for the Hurst phenomenon, and also proposed computer orientated algorithms for generating observations for which R/S followed Hurst's law, equation (2.11), with h preselected to lie anywhere in the range $0 < h < 1$ with the exception of $h = 0.5$.

The explanation of the Hurst phenomenon finds its roots in the concept of long-term persistence in the presence of which the interdependence between values of a process at points in time far distant from each other is small but non-negligible. A stochastic process which possessed the intensity of interdependence required to provide a precise model for the geophysical records examined by Hurst was first proposed by Mandelbrot (1965), and a mathematical foundation was laid by Mandelbrot

and Van Ness (1968), and Mandelbrot and Wallis (1969c). Some interesting theoretical properties of the process will now be presented.

2.5.1 Some Theoretical Considerations

(a) Brownian Motion

Brownian motion, B(t), sometimes referred to as a Bachelier or Wiener process, forms a convenient starting point. The function B(t) is continuous, and its most significant property is that for every $\varepsilon > 0$ the increments $B(t+\varepsilon) - B(t)$ (defined for t as a multiple of ε) are Gaussian and independent with zero mean and variance equal to ε. Hence, Brownian motion provides one framework for the definition of white noise. Brownian motion is also a "self-similar" process in that if the time scale is changed in the ration T where $T > 0$, the function $[B(t) - B(0)]$ and $T^{-0.5}[B(tT) - B(0)]$ are generated by the same probabilistic mechanism. (Mandelbrot and Wallis, 1969c). By letting $t = 1$, it may be deduced that the

$$\text{st.dev.}\,[B(t+T) - B(t)] = T^{0.5} \tag{2.20}$$

for every T. A further consequence of self-similarity is that the population range $R_p(T)$ and adjusted range R(T) of the continuous function B(t) obey the following laws asymptotically

$$E[R_p(T)] = a' T^{0.5} \tag{2.21}$$

$$E[R(T)] = b' T^{0.5} \tag{2.22}$$

where a' and b' are constants. Feller (1951) derived equations (2.21) - (2.22) and evaluated the constants a' and b'. Feller also claimed that the population range and adjusted range of the function B(t) would be virtually identical to the corresponding respective quantities for the function B(t) sampled at discrete time intervals $t = 1,2,3,\ldots n$; hence the form of equation (2.14). Equations (2.20) - (2.22) are sometimes referred to as "$T^{0.5}$ laws", and their failure to account for the Hurst phenomenon has previously been noted.

(b) Fractional Brownian Motion

In order to provide a mathematical basis for a suitable model for geophysical phenomena, Mandelbrot and Van Ness (1968) introduced fractional Brownian motion, the increments of which constitute a stationary Gaussian process, referred to as fractional Gaussian noise. Fractional Brownian motion (fBm), $B_h(t)$, may be defined from ordinary Brownian motion B(t) by forming the integral

$$B_h(t) - B_h(0) = \frac{1}{\sqrt{h+0.5}} \left[\int_{-\infty}^{0} \left[(t-u)^{h-0.5} - (-u)^{h-0.5} \right] dB(u) + \int_{0}^{t} (t-u)^{h-0.5} \, dB(u) \right]$$

$$0 < h < 1 \qquad (2.23)$$

where dB(u) is an infinitesimal increment of ordinary Brownian motion.

In order to maintain consistency with the notation of Mandelbrot and Van Ness (1968) and Mandelbrot and Wallis (1969c), the symbol H employed by those authors is equivalent to the symbol h employed here. An alternative definition of fBm is given by

$$B_h(t_2) - B_h(t_1) = \frac{1}{\sqrt{h+0.5}} \left[\int_{-\infty}^{t_2} (t_2 - u)^{h-0.5} dB(u) - \int_{-\infty}^{t_1} (t_1 - u)^{h-0.5} dB(u) \right] \qquad (2.24)$$

which is useful in illustrating that fBm is derived by weighting past values of white noise by $(t-u)^{h-0.5}$. However, both integrals in equation (2.24) are divergent, even though their difference is convergent, so that equation (2.24) is mathematically incorrect. However, by putting $t_2 = t$ and $t_1 = 0$ and rewriting the first integral, equation (2.23) is

obtained, and the first integral on the rhs of equation (2.23) is now convergent.

Fractional Brownian motion is a continuous process with self-similar increments such that if the time scale is changed in the ratio T, where $T > 0$, the functions $[B_h(t) - B_h(0)]$ and $T^{-h}[B(tT) - B_h(0)]$ are generated by the same probabilistic mechanism (Mandelbrot and Wallis, 1969c). By letting $t = 1$ and equating the variances of the two functions it may be shown (Appendix 2.1) that

$$\mathrm{Var}[B_h(t + T) - B_h(t)] = T^{2h} V_h \qquad 0 < h < 1 \qquad (2.25)$$

where V_h is a constant defined as the variance of unit increments of fBm. For $h = 0.5$, V_h reduces to unity, $B_h(t)$ becomes ordinary Brownian motion and equation (2.20) is obtained.

A further consequence of self-similarity is that the population range $R_p(T)$ and adjusted range $R(T)$ of the continuous function $B_h(t)$ obey the following laws asymptotically

$$E[R_p(T)] = a_h T^h \qquad 0 < h < 1 \qquad (2.26)$$

$$E[R(T)] = b_h T^h \qquad (2.27)$$

where a_h and b_h are constants. Under the assumption that equation (2.27) holds for the discrete case, and that $E[R] \sim n^h$ implies $E[R/S] \sim n^h$ asymptotically, then the increments of fBm form the necessary basis for modelling Hurst's law.

(c) Continuous time Fractional Gaussian Noise

In order to obtain a continuous time representation of the increments of fractional Brownian motion, it would appear logical to obtain the derivative of $B_h(t)$ to yield $B_h'(t)$ as stationary continuous time fractional Gaussian noise. However, the local behaviour of $B_h(t)$ (and $B(t)$) is so

erratic that $B_h(t)$ does not have a derivative, as

$$\lim_{\delta \longrightarrow 0} \left[\frac{B_h(t+\delta) - B_h(t)}{\delta} \right] = \infty \tag{2.28}$$

and, consequently, $B_h'(t)$ can exist only if some smoothing procedure is applied beforehand to $B_h(t)$. A smoothed process $B_h(t,\delta)$ may be defined (Mandelbrot and Van Ness, 1968) as

$$B_h(t,\delta) = \frac{1}{\delta}\int_t^{t+\delta} B_h(v)dv \qquad (\delta > 0) \tag{2.29}$$

whereupon the derivative of $B_h(t,\delta)$ is derived as

$$B_h'(t,\delta) = \frac{1}{\delta}\left[B_h(t+\delta) - B_h(t)\right] \tag{2.30}$$

where $B_h'(t,\delta)$ is stationary continuous time fractional Gaussian noise which is not differentiable.

The autocovariance function of $B_h'(t,\delta)$ has been given by Mandelbrot and Van Ness (1968) as

$$C_h(\tau,\delta) = \tfrac{1}{2}V_h\,\delta^{2h-2}\left[\left[\frac{|\tau|}{\delta}+1\right]^{2h} - 2\left|\frac{\tau}{\delta}\right|^{2h} + \left|\left|\frac{\tau}{\delta}\right| - 1\right|^{2h}\right] \tag{2.31}$$

where τ is the lag. For $\tau = 0$ the process has a finite variance given by

$$C_h(0,\delta) = V_h\,\delta^{2h-2} \tag{2.32}$$

For $0.5 < h < 1$, $C_h(\tau,\delta)$ is positive and finite for all τ such that

$$\int_0^{\infty} C_h(\tau,\delta) = \infty \tag{2.33}$$

while, for $0 < h < 0.5$

$$\int_0^{\infty} C_h(\tau,\delta) = 0 \tag{2.34}$$

The distinctly different forms of $C_h(\tau,\delta)$ for $0 < h < 0.5$ and $0.5 < h < 1$ are illustrated in figure (2.1).

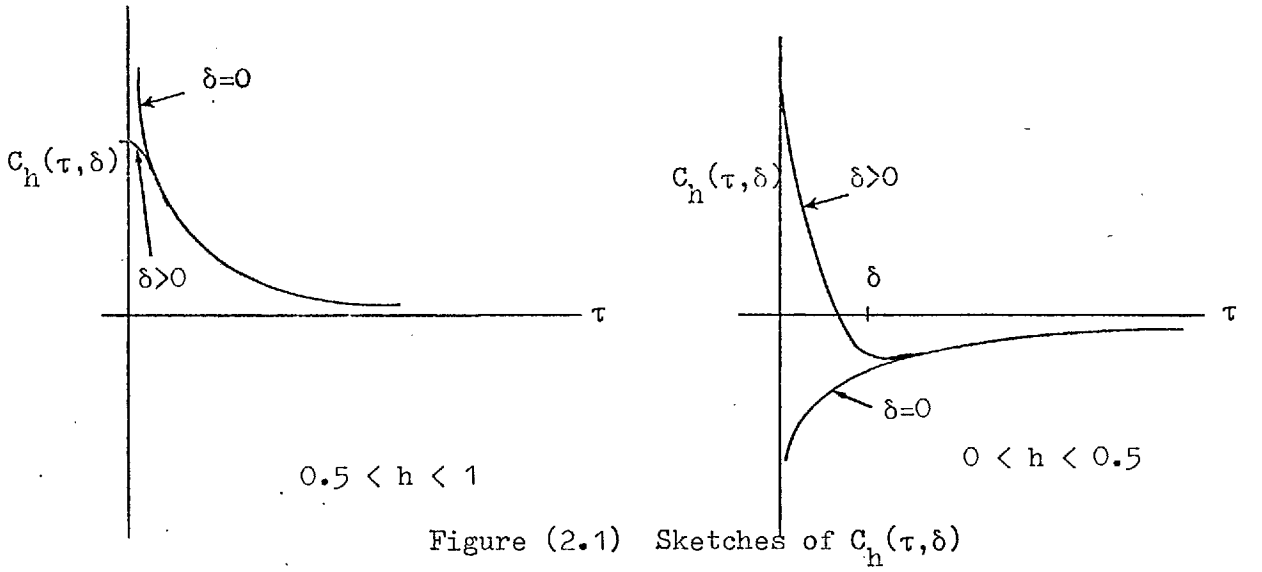


Figure (2.1) Sketches of $C_h(\tau,\delta)$

If $\tau \gg \delta$

$$C_h(\tau,\delta) \sim V_h\, h(2h-1)\, |\tau|^{2h-2} \qquad (2.35)$$

and $C_h(\tau,\delta)$ tends to zero as $|\tau| \to \infty$ which means that the process $B_h'(\tau,\delta)$ is ergodic. For $0.5 \lesssim h < 1$, $B_h'(t,\delta)$ exhibits strong positive long-run dependence, with the intensity of this dependence increasing as $h \to 1$. The function $C_h(\tau,\delta)$ approaches zero so slowly as to result in the divergence of the integral of that function. Hence, while the dependence for large lags will be small the cumulative effect of this dependence is strongly felt in the behaviour of R(T) as shown by equation (2.27), which is an asymptotic result.

In order to express $B_h'(t,\delta)$ in a form suitable for extrapolation, the integral given in equation (2.23) may be re-written for time $(t+\delta)$ to yield the function $[B_h(t+\delta) - B_h(0)]$. By forming the difference

$$[B_h(t+\delta) - B_h(0)] - [B_h(t) - B_h(0)] = [B_h(t+\delta) - B_h(t)]$$

and substituting for $[B_h(t+\delta) - B_h(t)]$ into equation (2.36), $B_h'(t,\delta)$ is defined as follows :

$$B_h'(t,\delta) = \frac{1}{\delta\sqrt{h+0.5}} \int_{-\infty}^{t} [(t+\delta-u)^{h-0.5} - (t-u)^{h-0.5}]\, dB(u) + \int_{t}^{t+\delta} (t+\delta-u)^{h-0.5}\, dB(u) \qquad (2.36)$$

By substituting $(s+\delta)$ for u, $B_h'(t,\delta)$ may be written as

$$B_h'(t,\delta) = \frac{1}{\delta\sqrt{h+0.5}} \left[\int_{-\infty}^{t-\delta} [(t-s)^{h-0.5} - (t-s-\delta)^{h-0.5}]\,dB(s+\delta) + \int_{t-\delta}^{t} (t-s)^{h-0.5}\,dB(s+\delta)\right] \quad (2.37)$$

$$= \frac{1}{\delta\sqrt{h+0.5}} \int_{-\infty}^{t} K_h(t-s,\delta)\,dB(s+\delta) \quad (2.38)$$

which corresponds to the form given by Mandelbrot and Van Ness (1968).

The kernel $K_h(t-s,\delta)$ may be defined as

$$K_h(s,\delta) = [\delta\sqrt{h+0.5}]^{-1} s^{h-0.5} \qquad (s \leqslant \delta)$$

$$= [\delta\sqrt{h+0.5}]^{-1} [s^{h-0.5} - (s-\delta)^{h-0.5}] \quad (s > \delta) \quad (2.39)$$

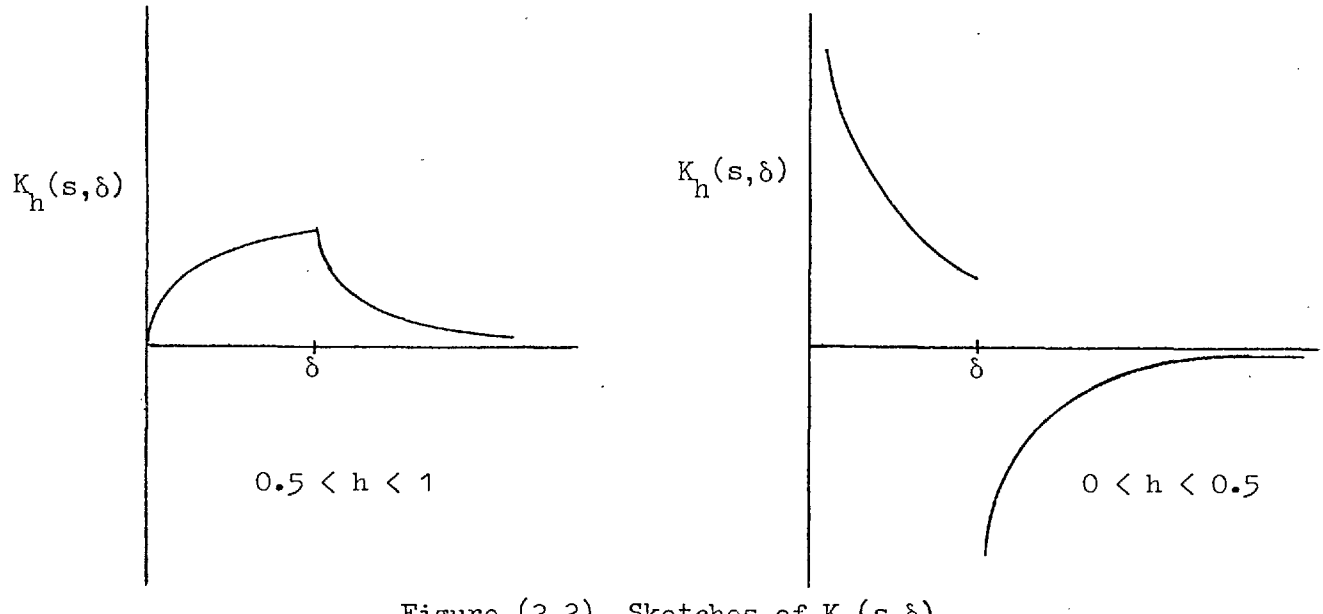


Figure (2.2) Sketches of $K_h(s,\delta)$

The kernal $K_h(s,\delta)$ which is sketched in figure (2.2) assumes distinctly different shapes in the ranges $0 < h < 0.5$, where $\int_0^\infty K_h(s,\delta)\,ds = 0$, and $0.5 < h < 1$, where $\int_0^\infty K_h(s,\delta)\,ds = \infty$. In the case of $h = 0.5$, $K_h(s,\delta) = 0$.

In an examination of the system which transforms white noise into fractional noise, Clarke (1971) has found the governing differential equation to be of the form

$$D^{\beta}y(t) = a(t) \qquad (2.40)$$

where a(t) and y(t) denote the input and output to the system, respectively, D denotes the differential operator $\frac{d}{dt}$ and β is a fractional power. However, some care is required in interpreting equation (2.40) as exemplified by Clarke, who showed that y(t) may be written as

$$y(t) = \frac{1}{\Gamma\beta} \int_{-\infty}^{t} (t - s)^{\beta-1} a(s)\, ds \qquad (2.41)$$

For $\beta = h+0.5$, equation (2.41) is identical to the Mandelbrot and Van Ness (1968) representation of Fractional Brownian motion. A value of $\beta = h-0.5$ would correspond to the differential of fractional Brownian motion which does not exist. Clarke circumvents this problem by noting the equivalence of the discrete time form of equation (2.41) with $\beta = h-0.5$ to the discrete type 2 approximation to fractional Gaussian noise proposed by Mandelbrot and Wallis (1969a). Clarke also points out that while the representation of y(t) given in equation (2.41) is non-linear in the parameter β, the system governed by equation (2.40) is linear in the systems theory sense as D^{β} is a linear operator.

(d) Discrete-time Fractional Gaussian Noise

By defining $B_h'(t,\delta)$ for integer values of t only and choosing $\delta = 1$, discrete time fractional Gaussian noise (dfGn) is obtained. Mandelbrot and Wallis (1969c) have defined the sequence of increments of $B_h(t)$ thus obtained as

$$\Delta B_h(t) = [B_h(t) - B_h(t-1)] \qquad (2.42)$$

while equation (2.30) suggests

$$\Delta B_h(t) = [B_h(t+1) - B_h(t)] \tag{2.43}$$

which is given by Mandelbrot and Wallis (1968). However, equation (2.43) is perhaps conceptually difficult to justify ; nevertheless, the statistical properties of either definition are identical. Thus $X_t = \Delta B_h(t)$ may now be considered as a model of Hurst's geophysical time series. In order that X_t be an acceptable model, the relationships given by equations (2.25) and (2.27) must be shown to hold for the discrete case.

Defining

$$D_n = \sum_{t=1}^{n} X_t$$

$$= \sum_{t=1}^{n} [B_h(t) - B_h(t-1)]$$

$$= B_h(n) - B_h(0)$$

whereupon, from equation (2.25)

$$\text{Var}[B_h(t+n) - B_h(t)] = \text{Var}[B_h(n) - B_h(0)]$$

$$= n^{2h} V_h \tag{2.44}$$

where t and n assume integer values only. Consequently, equation (2.25) holds for the discrete case as noted by Mandelbrot and Wallis (1969c). From equation (2.44)

$$\text{Var}[B_h(t+1) - B_h(t)] = \text{Var } X_t$$

$$= V_h = \sigma^2 \tag{2.45}$$

Thus, it may be deduced that

$$\text{Var } (\bar{X}) = \sigma^2 n^{2h-2}$$

where $\bar{X} = D_n/n$ or

$$\frac{\sigma_{\bar{x}}}{\sigma} = n^{h-1} \tag{2.46}$$

which is exactly equivalent to equation (2.16) deduced by Langbein (1956) for Hurst's data, and referred to as Langbein's corollary of Hurst's law by Mandelbrot and Wallis (1968). However, the validity of this relationship for small n is unknown.

The autocovariance function of $\Delta B_h(t)$ may be derived from equation (2.31) as

$$C_h(k) = \tfrac{1}{2}V_h \left[|k+1|^{2h} - 2k^{2h} + |k-1|^{2h} \right] \qquad (2.47)$$

which is defined for integer k, and which for large k may be approximated as

$$C_h(k) \sim \left[V_h\, h(2h-1)\right] k^{2h-2} \qquad (2.48)$$

An important property of $C_h(k)$ is that it satisfies

$$\sum_{k=0}^{\infty} C_h(k) = \infty \qquad (2.49)$$

in contrast with low order autoregressive and moving average type models which are characterized by the property that $\sum_{k=0}^{\infty} C_h(k)$ is finite. Hence, the fractional noise model with $h > 0.5$ suggests that the effect of past events on present behaviour dies out extremely slowly as distinct from models for which $\sum_{k=0} C_h(k)$ is finite when the effect of the past dies out extremely quickly. Thus, long-term persistence is synonymous with $0.5 < h < 1$ and short-term persistence is synonymous with $0 < h \leqslant 0.5$

The behaviour of the range and adjusted range for dfGn will follow closely the form of equations (2.26) and (2.27), respectively, for large n and hence

$$E[R_p] = a_h\, n^h \qquad 0 < h < 1 \qquad (2.50)$$

$$E[R] = b_h\, n^h \qquad 0 < h < 1 \qquad (2.51)$$

However, for small n, Mandelbrot and Wallis (1969c) have shown that the value of R_p will be less than the corresponding quantity in continuous

time, $R_p(T)$, with the difference disappearing with increasing n. However, simulation experiments were necessary to examine the behaviour of R for small n, as well as the behaviour of R/S for both small and large n.

2.5.2 Generating Processes and Simulation Experiments

A generating process for dfGn may be formulated by writing equation (2.38) with $\delta = 1$ to yield

$$X_t = B_h(t+1) - B_h(t)$$

$$= \frac{1}{\Gamma(h+0.5)} \int_{-\infty}^{t} K_h(t-s)\, dB(s+1) \qquad (2.52)$$

where the form of $K_h(t-s)$ may be deduced from equation (2.39). Equation (2.52) cannot be evaluated exactly on a computer, and its accurate numerical approximation involves 3 approximations (Mandelbrot and Wallis, 1969c).

(i) The span $-\infty < s < t$ must be replaced by a finite span $t - M < s < t$, which introduces a low frequency error term. The parameter M then defines the memory of the process, and represents the time span over which equations (2.46),(2.50) and (2.51) are approximated.

(ii) A discrete grid ε must be selected for the variable of integration s, and the infinitesimal $dB(s+1)$ must be replaced by a finite difference $B(s+1+\varepsilon) - B(s+1)$, which may be sampled as a Gaussian random variable with zero mean and variance ε.

(iii) The terms in the resulting summation must be evaluated to a finite number of decimals.

Mandelbrot and Wallis (1969c) proposed a type I approximation with the grid ε arbitrarily selected as $\varepsilon = 0.1$ and defined the resulting approximation to dfGn as

$$F_1(t \mid h,M) = \sum_{g=10(t-M)}^{10t-1} K_1\left(t - \frac{g}{10} \,\middle|\, h,M\right) \varepsilon_g \qquad (2.53)$$

$$0 < h < 1$$

where the kernal $K_1(t - \frac{g}{10})$ is the kernal $K(t-s)$ evaluated at the points $(t - \frac{g}{10}) = \frac{1}{10}, \frac{2}{10}, \ldots, (M - \frac{1}{10})$, M, and ε_g are independent Gaussian random variables with zero mean and variance $\frac{1}{10}$.

Obviously the calculation of equation (2.53) is cumbersome, as 10M independent Gaussian random variables must be summed to generate each value of the type I approximation to dfGn. In an attempt to circumvent this problem, Mandelbrot and Wallis (1969c) selected a coarser grid of $\varepsilon = 1$ for the variable of integration and, noting that $K_1(s|h,\infty) \sim (h-0.5)s^{h-1.5}$ for large s, defined a type II approximation as

$$F_2(t|h,M) = (h-0.5) \sum_{i=t-M}^{t-1} (t-i)^{h-1.5} \varepsilon_i \qquad (2.54)$$

$$0.5 < h < 1$$

which involves a summation of M independent Gaussian random variables to generate each value. In defining the type I and II approximations to dfGn, Mandelbrot and Wallis have omitted the scaling constant $(\sqrt{h+0.5})^{-1}$ present in equation (2.52).

Extensive computer experiments using the type I and type II approximations to dfGn were carried out by Mandelbrot and Wallis (1969a,b,e) to justify their use as generating processes, and to obtain further results not readily attainable through theoretical analyses. Using a digital graph plotter, Mandelbrot and Wallis (1969a) have illustrated that approximations to dfGn with $h > 0.5$ are characterized by a wealth of low frequency effects, with apparent periodic movements frequently perceptible which, without knowledge of the nature of the generating process, might lead to the belief that "hidden periodicities" were present. Low frequency movements tend to increase with h, with short-term movements becoming less pronounced. The low frequency nature of dfGn with $h > 0.5$ is exemplified by the sample variance S^2 which satisfies (Mandelbrot and Wallis, 1969a)

$$E[S^2] = \sigma^2 - \sigma^2 n^{2h-2} \qquad (2.55)$$

which means that, for any given value of n, the sample variance underestimates the population variance by an increasing amount as h increases from 0.5 to 1. Equation (2.55) has as a consequence that the variability of the sample mean increases with increasing h, as may be deduced from equation (2.46). The difficulties associated with using a past sample average as a predictor of a future sample average of dfGn have also been studied extensively by Mandelbrot and Wallis (1969a).

The statistical properties of the rescaled range R/S for approximations to dfGn were extensively investigated by Mandelbrot and Wallis (1968b,e). By generating sequences of 9000 values using a type I generating process with a preselected value of h, Mandelbrot and Wallis noted that, apart from a short initial transient for $n < 20$, the slope of plots of log(R/S) against log n was linear with slope h up to a value of $n = M^*$. M^* is related to M, the span of the moving average in the generating process, or the memory, and M^*/M is at least 1. Only computer storage limitations prevent making M infinite in accordance with theory. Hence, Mandelbrot and Wallis (1969b) claimed that

$$\frac{R}{S} \sim n^h \qquad 0 < h < 1$$

for dfGn which is asymptotically valid and is the exact form of Hurst's law as given by equation (2.11). They also claimed that the distribution of $(R/S)\, n^{-h}$ was independent of n on the basis of the observed plots of log (R/S) against log n which seems strong in the light of the large skewness of R/S observed from further simulation experiments. They also noted that the type II approximation was inferior to the type I, except for values of h of the order of 0.9 and higher when the lack of high frequencies in the type II approximation passes unnoticed.

The potential of the rescaled range as a measure of long-term persistence in time series was also established by Mandelbrot and Wallis (1969b,d).

In comparing the relative merits of the rescaled range R/S and spectral analysis in detecting long-term persistence, they found R/S to be much more sensitive than spectral analysis in detecting long-term persistence, particularly when "log-spectral" plots are used. However, when very strong periodic cycles are present, spectral analysis would be superior. Thus, the frequent failure of spectral analysis to verify the presence of hidden periodicities in long geophysical records might be attributed to the fact that the low frequency movements observed are a characteristic of the random nature of the generating process.

In assessing the robustness of the statistic R/S, Mandelbrot and Wallis (1969e) established that R/S is distribution free for the case of an independent random process, such that Hurst's law with $h = 0.5$ is obeyed even for small n, no matter what the underlying distribution is. Thus, no explanation of the Hurst phenomenon could be based on skewness. However, both R and R_p are distribution dependent, and, in the presence of skewness, may exhibit long transients suggesting $h > 0.5$, as shown by Mandelbrot and Wallis (1969e). Moreover, R is also much more variable than R/S when the underlying distribution is skewed. Hence an erroneous explanation of the Hurst phenomenon such as that proposed by Moran (1964) could be based on skewness if attention is confined only to the statistics R or R_p. In the case where the underlying distribution is Gaussian, the scaling of R by S is unimportant, and it is sufficient to work with R. Some non-linear transformations were also applied to sequences of approximate dfGn, showing that Hurst's law is invariant with respect to moderate non-linearity, but in the presence of extreme non-linearity, this invariance breaks down, and the slope of the plot of log (R/S) against log n is depressed. However, if a transformation is applied to the variate ε_g in equation (2.53) before applying the type I kernal to generate values of

$F_1(t|h,M)$, R/S was found to be invariant under such a transformation.

Shortcomings in the procedure used by Hurst (1951, 1956) for the estimation of h (equation (2.13)) have been pointed out by Mandelbrot and Wallis (1968, 1969d). The method uses only the value of R/S for the total sample length to estimate h, and assumes that the line defining the estimate of h always passes through the point of abscissa log 2 and ordinate log 1 = 0. The effect of this estimation procedure is to accentuate the effect of the initial transient when $h < 0.72$, yielding estimates of h which are too high, and to underestimate h when $h > 0.72$. The variability of estimates of h about the true value is also underestimated.

A more general estimation procedure was proposed by Mandelbrot and Wallis (1969b,d) whereby R/S was computed for various subsample lengths of the record, and the origin of the plot was allowed to vary freely. The new method was then used to estimate h for a variety of geophysical records, resulting in a further reinforcement of Hurst's original conclusion that $h > 0.5$. The longest records analysed showed no tendency for estimates of h to tend to 0.5 as the record length increased. Thus the span of statistical dependence would appear to be as long as the longest geophysical records available.

2.5.3 Estimation of h

The fact that dfGn tends to be dominated by low frequencies (i.e. the spectrum has a large concentration of variance at low frequencies) means that small sample statistical properties will differ considerably from their corresponding population quantities. The properties of estimators of h are of primary importance, and have been investigated by Wallis and Matalas (1970) for independent processes, lag-one Markov processes and an approximation to dfGn.

Two estimation procedures were considered, one being the original procedure used by Hurst which provides an estimate denoted by K, and the other being an adoption of the method proposed by Mandelbrot and Wallis (1969b) which provides an estimate denoted by H. The latter method consists of the following steps : -

(i) a record of length n is divided into N subsamples of length n_s where $3 < n_s < n$. The subsamples may or may not overlap, and the selection of the subsample lengths n_s is made such that a uniform spacing of n_s is achieved on a logarithmic scale

(ii) the rescaled range R/S is computed for each subsample of length n_s

(iii) A least squares line is fitted to the mean R/S for each subsample size in the range $n_0 < n_s < n$, where n_0 is chosen on the grounds of an initial non-linear transient in the plot of log(R/S) against log n, as shown in figure (2.3). The slope of the fitted least squares line yields the estimate H.

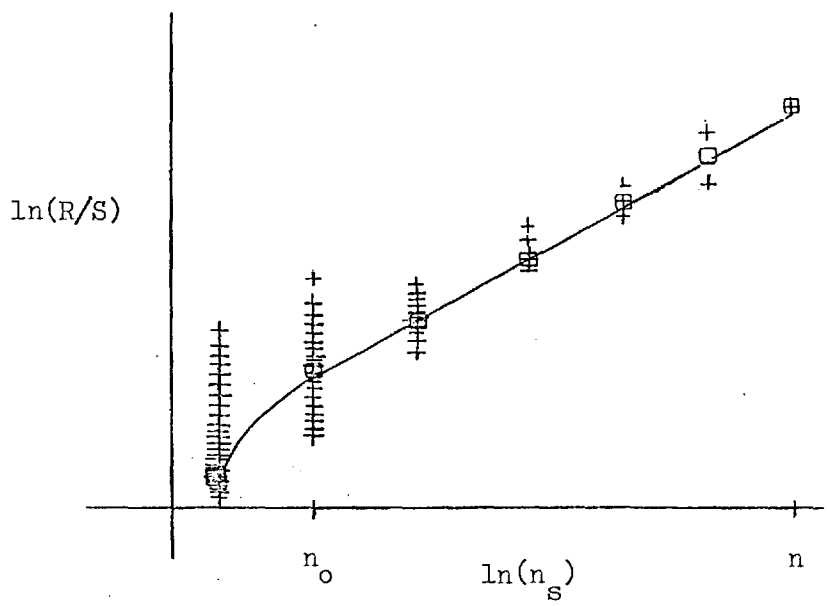


Figure (2.3). Schematic plot of ln(R/S) against $\ln(n_s)$. Each value of R/S is denoted by a cross and the mean values of R/S for a given value of n_s are denoted by little squares.

For a Gaussian independent process, Matalas and Wallis (1970) found that both K and H are biased estimators, with K displaying greater bias than H ; however, K was found to have smaller variance than H. The bias in both estimators decreases slowly with n, the sample size ; for example, the mean value of K for n = 1000 was found to be 0.58. In calculating H, Wallis and Matalas (1970) generally used a value of $n_0 = 10$; a higher value of n_0 would yield a less biased but more variable estimate. If only non-overlapping subsamples were used, a more biased and variable estimate resulted. The selection of n_0, as well as the selection of a procedure for subdivision of a sample, is obviously subjective.

Similar results were found for the case of a log-normal independent process although estimates were generally slightly less variable than for the Gaussian case. Sampling from a normal and log-normal lag-one Markov process showed that K and H have similar properties in the presence of Markovian autocorrelation and emphasised further the long-run deficiencies of the Markov process. Limited sampling experiments were carried out using a type I approximation to dfGn with values of h in the range $0.5 < h < 1$. Again, K was found to be more biased than H ; for h less than about 0.7, both H and K are biased upwards, while for $h > 0.7$, H and K are biased downwards. Hence, H and K are approximately unbiased in the neighbourhood of 0.7. The sampling experiments were not sufficiently extensive to permit conclusions to be drawn about their relative variability, although H is probably a more variable estimator than K for $0.5 < h < 1.0$. Estimates of H and K on a monthly basis for 25 streams in the Potomac estuary with 30-40 years of record tended to suggest that for the streams examined, $h > 0.5$, and that H is more variable than K. Both H and K were found to vary seasonally ; however, the excessive temporal and spatial variability of H and K did not allow a regionalization of h.

2.6 Further Approximations to dfGn

2.6.1 Fast fractional Gaussian Noise

In response to criticism of the initial type I and type II approximations, Mandelbrot (1971a)proposed a fast fractional noise generator for values of h in the range $0.5 < h < 1$. The type I approximation, while following the theoretical form of the kernel of dfGn closely, proves expensive and cumbersome to compute while the type II approximation is deficient in high frequencies and the low frequency approximation is satisfactory only when h is close to 1.

Rather than approximating the kernel $K_h(s)$ given by equations (2.39) and (2.52), Mandelbrot worked with the autocovariance function of dfGn which is given by equation (2.47) and is an equivalent specification of dfGn.

In approximating $C_h(k)$ Mandelbrot first noted that $C_h(k)$ is the second finite difference of the function $\frac{k^{2h}}{2}$ which for large k may be approximated by its second derivative which is $h(2h-1)k^{2h-2}$ as given by equation (2.48). As $C_h(k)$ defines the autocovariance function for continuous time as well as discrete time fractional Gaussian noise, defined as $\Delta B_h(t) = [B_h(t) - B_h(t-1)]$ for all t and integer values of t, respectively, the function k^{2h-2} may be expressed in terms of its Laplace transform as follows

$$k^{2h-2} = \frac{\int_0^\infty e^{-ku}\, u^{1-2h}\, du}{\Gamma(-2h+2)}$$

whereupon

$$h(2h-1)k^{2h-2} = \frac{\int_0^\infty e^{-ku}\, u^{1-2h}\, du}{2\,|\Gamma(-2h)|}$$

$$= \int_0^\infty e^{-ku}\, W(u)\, du \qquad (2.56)$$

where $W(u) = u^{1-2h}/2\Gamma(-2h)$. The integral in equation (2.56) is very

conveniently the autocovariance of a first order Gauss-Markov process for which the lag-one autocorrelation is e^{-u} and the variance is infinitesimal and equal to W(u) du. The integral in equation (2.56) thus forms the basis for the approximation, which for practical purposes must be evaluated between finite limits on a discrete grid which means that approximate dfGn is derived as the summation of a finite number of suitably weighted and independent discrete time first order Gauss-Markov processes. Mandelbrot establishes an upper limit on the integral in equation (2.56) by neglecting high frequency effects, which are handled separately afterwards, and a lower limit by neglecting low frequency effects which will not be noticed within a finite sample, to yield the approximation

$$C_h^a(k) = \frac{\int_{1/TQ}^{-\log (th_1)} e^{-ku}\, u^{1-2h}\, du}{2\left|\sqrt{(-2h)}\right|} \qquad (2.57)$$

where th_1 is a high frequency threshold, T is the sample duration and Q controls the low frequency threshold. Following a change of variable from u to B^{-v} with $B > 1$, which results in a greater emphasis on low frequencies, the resulting integral is then evaluated over a discrete grid to give the final approximation to $C_h(k)$ as

$$D_h(k) = \sum_{m=1}^{L} \frac{B^{1-h} - B^{-1+h}}{4(h-1)\left|\sqrt{-2h}\right|} B^{-2(1-h)m}\, e^{-kB^{-m}} \qquad (2.58)$$

with L defined as the nearest larger integer to $[\log (Qn)/\log B]$ where n is the sample size. The generating process is defined by Mandelbrot as

$$X_t^{(1f)} = \sum_{m=1}^{L} W_m\, y_t(\rho_1^{(m)}) \qquad (2.59)$$

where

$$\rho_1^{(m)} = e^{-B^{-m}} \qquad (2.60)$$

$$W_m^2 = \frac{B^{1-h} - B^{-1+h}}{4(h-1)\left|\sqrt{-2h}\right|} B^{-2(1-h)m} \qquad (2.61)$$

and

$$y_t(\rho_1^{(m)}) = \rho_1^{(m)}\, y_{t-1}(\rho_1^{(m)}) + \sqrt{1-\rho_1^{(m)}}\,\varepsilon_{t,m} \qquad (2.62)$$

The factors B and Q control the quality of the approximation, which improves as $B \to 1$ and $Q \to \infty$, while the number of variables $y_t(\rho_1^{(m)})$ to be weighted and summed is proportional to log n rather than 10n and n as for the type I and II approximations, respectively.

As some low and high frequencies have been neglected the approximation represented by equation (2.59) will have a variance less than unity which is given as

$$D_h(0) = \sum_{m=1}^{L} W_m^2 \qquad (2.63)$$

and the residual variance is then

$$1 - D_h(0) = 1 - \frac{B^{-(1-h)}\, h(2h-1)}{\sqrt{3-2h}} \qquad (2.64)$$

It is perhaps worth pointing out that equation (2.64) is incorrectly printed in Mandelbrot's paper.

Mandelbrot (1971a) suggests using white Gaussian noise with a variance of $(1 - D_h(0))$ to model the residual variance. However, a high frequency error, represented by the autocovariance

$$C^{(hf)}(1) = C_h(1) - D_h(1) \qquad (2.65)$$

still remains. Rather than using white Gaussian noise, a lag-one Gauss-Markov process may be used with a variance of $(1 - D_h(0))$ and a lag-one autocorrelation given by Mandelbrot as

$$\rho_1^{(hf)} = 2^{2h-1} - 1 + \sum_{m=1}^{L} W_m(1-\rho_1^{(m)}) - \frac{B^{-(1-h)}\, h(2h-1)}{\sqrt{3-2h}} \qquad (2.66)$$

An attempt to reproduce this result has been made by deriving $D_h(1)$ from equation (2.58) and defining the lag-one autocorrelation of the Gauss-Markov

process as

$$\rho_1^{(hf)} = \frac{C(1,h) - D_h(1)}{1 - D_h(1)} \tag{2.67}$$

$$= \frac{2^{2h-1} - 1 - \sum_{m=1}^{L} W_m^2 \rho_1^{(m)}}{1 - \sum_{m=1}^{L} W_m^2}$$

$$= \frac{2^{2h-1} - 1 - \left[\sum_{m=1}^{L} W_m^2 \rho_1^{(m)}\right]}{1 - \frac{B^{-(1-h)}\, h(2h-1)}{\sqrt{3-2h}}} \tag{2.68}$$

which does not appear to correspond with Mandelbrot's result. If $\sum_{m=1}^{L} W_m^2 \rho_1^{(m)}$ in equation (2.68) is written as $\sum_{m=1}^{L}\left[W_m^2 - W_m^2(1-\rho_1^{(m)})\right]$ then the r.h.s. of equation (2.66) with W_m written as W_m^2 is equal to the numerator of equation (2.68)

The high frequency term is generated as

$$X_t^{(hf)} = \rho_1^{(hf)} X_{t-1}^{(hf)} + \sqrt{1 - \rho_1^{(hf)^2}}\, \varepsilon_t \tag{2.69}$$

and finally approximate dfGn with zero mean and unit variance is given by

$$X_t = X_t^{(lf)} + \sqrt{1 - D_h^2(0)}\, X_t^{(hf)} \tag{2.70}$$

Alternatively, the high frequency process may be used to fit observed high frequency characteristics of a historic sequence. However, details of how this might be achieved have not been given by Mandelbrot, and the necessary documentation for generating fast dfGn with prespecified statistical properties has not been given nor has the non-Gaussian case been treated. For example, if X_t in equation (2.70) is to preserve a variance σ^2 estimated from a historic sequence, how should this variance be divided between the low frequency and high frequency processes? One

method would be to replace $C_h(1)$ in equation (2.65) by $\hat{\rho}_1$, the estimated lag-one autocorrelation, thereby defining $\rho_1^{(hf)}$, whence dfGn with zero mean and unit variance may be generated and scaled using σ. However, the problem of simultaneously modelling short run and long run effects needs to be more fully explored while the selection of B and Q needs more documentation. The quality of approximation should perhaps be viewed in the light of the sensitivity of Hurst's law to various values of B and Q. Further documentation of fast dfGn is required before it can be freely used in water resource system simulations.

2.6.2. The ARIMA (1,0,1) Process

In response to the need for a simple, fast, generating process which combines desirable short run and long run effects, O'Connell (1971) proposed a mixed moving average autoregressive process or in the terminology of Box and Jenkins (1970), an ARIMA (1,0,1) process expressed as

$$X_t = \varphi X_{t-1} + \varepsilon_t - \theta\varepsilon_{t-1} \tag{2.71}$$

which has the following autocorrelation function

$$\left.\begin{aligned} \rho_1 &= \frac{(\varphi-\theta)(1-\varphi\theta)}{(1+\theta^2-2\varphi\theta)} \\ \rho_k &= \varphi\rho_{k-1} \end{aligned}\right\} \tag{2.72}$$

Obviously $\sum_{k=0}^{\infty} \rho_k$ converges ; and therefore the ARIMA (1,0,1) process may be said to lie "in the Brownian domain of attraction" for which h = 0.5. However, for suitably chosen values of φ and θ, the process maintains good agreement with Hurst's law with h > 0.5 for values of n considerably in excess of those necessary to constitute an explanation of the Hurst phenomenon. Short run and long run effects may be modelled simultaneously through matching estimates of ρ_1 and h from historic records of length n with their corresponding expectations. As with the type I and II approximations,

Hurst's law is obeyed over a finite span which can be as large as n = 10000 for the ARIMA (1,0,1) process ; ultimately, however h → 0.5. In the case of the type I and II approximations, the parameter h is incorporated in the generating mechanism while for the ARIMA (1,0,1) process the value of h must be defined from plots of log (R/S) against log n up to very large values of n as demonstrated by O'Connell (1971). The ARIMA (1,0,1) process can thus be viewed as an approximation to dfGn, or as an explanation of the Hurst phenomenon in its own right ; only geophysical records longer than those currently available would rule out the latter possibility.

As it is the purpose of this thesis to explain and develop more fully the properties of the ARIMA (1,0,1) process, it will not be expanded on further here.

2.6.3. Filtered Fractional Gaussian Noise

A further approximation to dfGn has been developed by Wallis and Matalas (1971) and Matalas and Wallis (1971b). Using the original type II approximation given by equation (2.54) as a starting point, they derived the autocorrelation function of the process as

$$\rho_k = \frac{\sum_{i=0}^{M-1-k} (M-i)(M-i-k)^{h-1.5}}{\sum_{i=0}^{M-1} (M-i)^{2h-3}} \qquad (2.73)$$

and noted that the process yielded values of ρ_1 too high for modelling annual streamflow and also much in excess of values of ρ_1 for dfGn itself, resulting in undesirable high frequency properties. A filtered type II approximation to dfGn was proposed by Matalas and Wallis (1971b) as

$$X_t = (h - 0.5) \sum_{i=pt-M}^{pt-1} (pt - i)^{h-1.5} \varepsilon_i \qquad (2.74)$$

where $p \geq 1$ is an integer. Thus, values X_t, X_{t+1}, X_{t+2},... are generated by weighting and summing the sequences of independent random variables

$[\varepsilon_{p-M}, \varepsilon_{p-M+1},\ldots,\varepsilon_{p-2}, \varepsilon_{p-1}]$, $[\varepsilon_{2p-M}, \varepsilon_{2p-M+1},\ldots,\varepsilon_{2p-2}, \varepsilon_{2p-1}]$, $[\varepsilon_{3p-M}, \varepsilon_{3p-M+1},\ldots,\varepsilon_{3p-2}, \varepsilon_{3p-1}]$,..., respectively. It follows that the process X_t given by equation (2.74) is equivalent to sampling values at times pt from the type II approximation given by equation (2.54). The mean, variance, and skewness of the filtered process are the same as for the type II approximation itself, and have been derived by Matalas and Wallis (1971b)as

$$\mu_x = (h-0.5) \sum_{i=0}^{M-1} (M-i)^{h-1.5} \mu_\varepsilon \tag{2.75}$$

$$\sigma_x^2 = (h-0.5 \sum_{i=0}^{M-1} (M-i)^{2h-3} \sigma_\varepsilon^2 \tag{2.76}$$

$$\gamma_x = \frac{\sum_{i=0}^{M-1} (M-i)^{3h-4.5}}{\left[\sum_{i=0}^{M-1} (M-i)^{2h-3}\right]^{1.5}} \tag{2.77}$$

However, the autocorrelation function of the filtered process is given as ρ_{kp}, k = 0,1,2,..., where ρ_k is given by equation (2.73). Increasing values of p yield successively decreasing values of ρ_1 which allows the desirable flexibility of being able to preserve estimates of h and ρ_1, simultaneously, although the fact that p is an integer means that ρ_1 does not vary continuously with p. Computation time for filtered fractional noise is only fractionally in excess of computation time for the type II approximation.

Using filtered fractional noise with prespecified values of ρ_1 and h, Wallis and Matalas (1971) conducted further sampling experiments which showed that H and K, estimators of h, are biased for small samples of the filtered type II approximation, with the bias for any given value of h decreasing with increasing p and sample size n. The problem of detecting

long-term persistence in small samples was extensively investigated using simulation experiments, with the finding that available tests for independence, such as the Anderson test (1942) and a variety of runs tests, lacked power in the presence of persistence, both short-term and long-term, with the power decreasing as the intensity of persistence increased. The lack of power in the case of the Anderson test of $\hat{\rho}_1$ may be largely ascribed to the fact that estimators of ρ_1 are biased towards zero in the presence of persistence, with the bias increasing with increasing persistence. In any case, the Anderson test is only suitable for detecting high frequency non-randomness, and no formal test of significance exists for a measure of long-term persistence. Wallis and Matalas (1971) also observed that oft-noted fluctuations in sample correlograms for large lags may be consistent with underlying long-term persistence rather than being ascribed to chance.

2.6.4. The Broken Line Process

In a recent series of papers, Rodriguez, Mejia and Dawdy (1972) and Mejia, Rodriguez and Dawdy (1972) have presented a new model, the Broken Line (BL) process, which is claimed to have distinct advantages over Markovian and fractional noise models. The BL process, originally devised by Ditlevsen (1971) to check by simulation some results in first passage theory, is composed of a sum of a series of simple BL processes, each of which is given as

$$e(t - lc) = \sum_{j=0}^{\infty} \left[\varepsilon_j + \frac{(\varepsilon_{j+1} - \varepsilon_j)(t - jc)}{c}\right] I_{[jc,(j+1)c]}(t) \tag{2.78}$$

where $\varepsilon_0, \varepsilon_1, \varepsilon_2, \ldots, \varepsilon_j, \ldots$ are a series of identically and independently distributed Gaussian random variables with zero mean and variance σ^2, l is a uniformly distributed random variable over the interval $(0,1)$, c is

the distance in time separating values of ε_j, and $I_{[jc,(j+1)c]}(t)$ is an indicator function which satisfies

$$I_{[jc,(j+1)c]}(t) = 1 \qquad jc < t < (j+1)c$$
$$= 0 \qquad \text{otherwise} \tag{2.79}$$

A schematic representation of $e(t)$ is given in figure (2.4). The process $e(t)$ is a continuous time process with a mean, variance and autocorrelation function as follows

$$E[e(t)] = 0 \tag{2.80}$$

$$\text{Var}[e(t)] = E[e(t)^2] = \frac{2}{3}\sigma^2 \tag{2.81}$$

$$\rho(\tau) = \begin{cases} 1 - \frac{3}{4}(\tau/c)^2[2 - (\tau/c)] & 0 \leqslant \tau \leqslant c \\ \frac{1}{4}[2 - (\tau/c)]^3 & c < \tau \leqslant 2c \\ 0 & \tau > 2c \end{cases} \tag{2.82}$$

The general BL process is then defined as

$$X^{(N)}(t) = \sum_{i=1}^{N} e_i(t) \tag{2.83}$$

where $e_i(t)$ is a simple BL process with parameters c_i and σ_i. The variance and autocorrelation function of $X^{(N)}(t)$ are given as

$$\text{Var}[X^{(N)}(t)] = \frac{2}{3}\sum_{i=1}^{N} \sigma_i^2 \tag{2.84}$$

$$\rho_x(\tau) = \frac{\sum_{i=1}^{N} \sigma_i^2 \rho_i(\tau)}{\sum_{i=1}^{N} \sigma_i^2} \tag{2.85}$$

where $\rho_i(\tau)$ denotes the autocorrelation function of $e_i(t)$. The parameters c_i and σ_i are required to satisfy restrictions which are given by Mejia

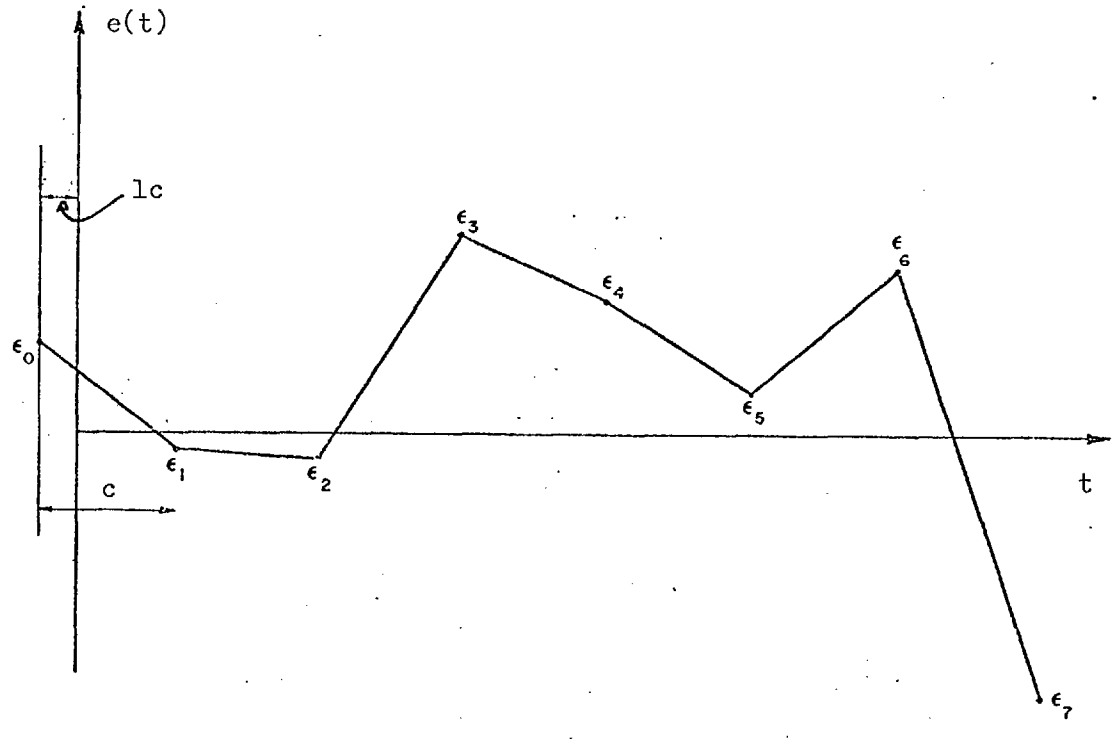


Figure (2.4) A schematic representation of a simple Broken Line process.

et al (1972) as

$$c_i = c_1 \gamma^{i-1} \qquad \gamma > 1$$

$$\sigma_i = \sigma_1 \left(\frac{c_1}{c_i}\right)^{0.5} = \sigma_1 (1/\gamma^{i-1})^{0.5} \tag{2.86}$$

Other restrictions may be specified, depending on the parameters to be preserved. The memory of the process, or lag τ at which the autocorrelation function becomes zero is $2C_N$ and is controlled by 1 and N. In general, the high frequency properties are governed by the "shorter" lines and the "longer" lines control the low frequency properties. Mejia et al (1972) have shown how BL may be derived as an approximation to dfGn.

In proposing the BL process as a model of geophysical time series, Rodriguez et al (1972) and Mejia et al (1972) have claimed distinct advantages for the model over Markovian and fractional noise models. The deficiencies of Markovian models quoted are largely those previously enunciated by Mandelbrot and Wallis (1968), while the main criticism levelled at fGn is that it does not possess a derivative, and, that, consequently, the second derivative of the autocorrelation function at the origin, denoted by $\rho''(0)$, does not exist for fGn. The existence of $\rho''(0)$ which measures the shape of $\rho(\tau)$ at the origin, means that the process must be four times differentiable, which implies that the high frequency properties of the process must satisfy certain "smoothness" criteria. For the BL process $\rho''(0)$ exists, and hence crossing properties, such as extreme events, run lengths and run sums, may be defined for the continuous time BL Gaussian process. However, the slight non Gaussian character of BL has been pointed out by Mandelbrot (1972) and consequently the applicability of crossing theory to continuous BL is somewhat dubious.

In the context of modelling geophysical time series, which are sampled as discrete values at equi-spaced time points, or averages over

equal time intervals, the special advantages claimed for BL over fGn disappear. At the present, only the statistical properties of the continuous BL process are known, and these cannot be correctly matched with corresponding quantities measured from discrete or averaged data. Furthermore, the crossing properties of a discrete or averaged BL process are unknown. However, assuming that $\rho''(0)$ may be estimated from discrete time series, (which for the case of a monthly or annual time unit would appear to be extremely difficult) what are the grounds for preserving such an estimate? For daily flows, Weiss (1973) has shown that models with a Gaussian basis have limited application. Are simulations using approximations to dfGn (BL may be considered as an approximation to dfGn (Mandelbrot, 1972)) which preserve and do not preserve estimates of $\rho''(0)$ liable to lead to very different results? Consideration of the original motivation behind BL and fGn permits some tentative conclusions to be drawn.

Discrete time fGn was developed specifically as a model of Hurst's annual time series, where frequencies higher than $f = \frac{1}{2\Delta t}$ where $\Delta t = 1$ year are missed out in the data sampling procedure, and are therefore irrelevant as far as subsequent simulation modelling is concerned. Indeed successive smoothing of fractional Brownian motion as suggested by Mandelbrot and Van Ness (1968) would yield a version of fGn for which $\rho''(0)$ existed but the long-run properties of which would be identical to those of the present version of fGn. As fGn was developed primarily with long-run properties in mind, the non existence of $\rho''(0)$ for fGn is unimportant as pointed out by Mandelbrot (1972). Indeed the non existence of $\rho''(0)$ is not particular to fGn ; continuous time white Gaussian noise itself does not possess a derivative but this does not hamper its use as a basis for discrete time models. While the motivation for dfGn is soundly based on the empirical law of Hurst, the motivation for the BL process as a model

of discrete time hydrological time series is somewhat obscure. The BL process has been shown to have potential in modelling turbulence phenomena (Nordin, McQuivey and Mejia, 1972) where the measurement of very high frequencies allows a very close approximation to the underlying continuous process, and $\rho''(0)$ becomes measurable and may be modelled. In fact, $\rho''(0)$ has a specific interpretation in turbulence theory as a measure of the microscale of turbulence, first defined by G.I. Taylor (1921).

In the context of modelling long-term persistence in annual geophysical time series, $\rho''(0)$ must be viewed in the following light. As its very existence derives from high frequency properties in the first place, it is difficult to view it as an effective measure of low frequency effects. Certainly, mathematical relationships may exist between $\rho''(0)$ and run sums and run lengths for the continuous process, but do estimates of $\rho''(0)$ represent an effective measure of such quantities? For example a mathematical relationship exists between ρ_1 and h for fGn given by equation (2.47), but estimates of ρ_1 could not sensibly be expected to effectively measure low frequency effects represented by h. As the very definition of h itself is based on the intensity and duration of periods of above (or below) average flow, it is highly unlikely that an alternative measure of droughts would be largely at variance with the run sums and run lengths implicitly and neatly preserved by h. If an alternative and better measure is to be found and accepted, then it must be shown to be readily measurable, and follow some well defined pattern of relationship such as Hurst's law within a large data base. The methods proposed by Mejia et at (1972) for measuring $\rho''(0)$ within discrete time series are based on the sample spectrum and correlogram, both of which require lengthy computation and both of which are notorious for their sampling instability. No attempt has been made to show that the run sums and run lengths preserved through

an estimate of $\rho''(0)$ are consistently at variance with those preserved by a measured h. As far as high frequency short term effects are concerned there does not appear to be any strong evidence to suggest that such effects cannot be modelled through preserving ρ_1, the lag-one autocorrelation coefficient.

A further issue relating to the methodology of BL requires clarification. Rodriguez et al (1972) argue that Mandelbrot and Wallis (1968, 1969a,b,c,d) did not study the Noah and Joseph effects as originally defined by them, but rather that they merely studied the Hurst phenomenon. This contention appears inaccurate, as the occurence of long periods of above (or below) average flow (the Joseph effect) is directly defined by Hurst's law which is the basis for dfGn. The Noah effect was shown to be both unnecessary and insufficient in accounting for the Joseph effect, and may be handled separately through skewed marginal distributions as shown by Mandelbrot and Wallis (1969e). The main use made of $\rho''(0)$ by Mejia et al (1972) has been to define the Noah effect, based on the assumption of a Gaussian marginal distribution, through a roughness parameter γ which is directly related to $\rho''(0)$. However, as streamflow is bounded at or near zero, the Noah effect, if present, manifests itself through positive skewness, and can hardly be in accord with a Gaussian assumption.

Perhaps the most important and useful concept to emerge from BL methodology is that synthetic sequences should preserve a "memory" of the historic sequence i.e. equally likely projections over an economic time horizon should all be conditional on a common past. (Garcia, Mejia and Dawdy, 1972). However, some rather arbitrary procedures have been invoked by Garcia et al in applying this concept, which is not unique to BL methodology. A careful evaluation of the merits of this concept should prove extremely advantageous to users of synthetic hydrology.

2.7 Further Developments in Synthetic Hydrology

In recent years, a number of new models have emerged for generating synthetic streamflows, and methodologies have been developed for ensuring that resemblance is maintained between historic and synthetic sequences in terms of statistical parameters which are thought to exert an important influence on system design. However, the question of how this statistical resemblance should be maintained is one of considerable importance particularly when small sample expectations of parameters differ greatly from respective population expectations. Such small sample biases are particularly severe for estimates of the variance and lag-one autocorrelation when low frequency effects are strong, as illustrated by Wallis and Matalas (1971). Thus, two types of statistical resemblance may be obtained, as defined by Matalas and Wallis (1974) and given in section (1. 2.4). Resemblance in terms of statistics measured in both synthetic and historic sequences of size n would appear to be the most logical goal to achieve, but knowledge of small sample properties of a process is necessary so that appropriate bias corrections may be administered. Wallis and O'Connell (1972) have considered small sample bias corrections for the lag-one Markov process, and noted that estimates of ρ_1 defined as

$$\left[\hat{\rho}_1\right]_n = \frac{\sum_{t=1}^{n-1} (X_{t+1} - \bar{X})(X_t - \bar{X})}{\sum_{t=1}^{n} (X_t - \bar{X})^2} \tag{2.87}$$

satisfy (Kendall, 1954)

$$E\left[\hat{\rho}_1\right]_n = \rho - \frac{1}{n}(1 + 4\rho) \tag{2.88}$$

For n = 20 and $\rho_1 = 0.4$, $E\left[\hat{\rho}_1\right]_n = 0.27$, so that the bias is quite severe in small samples, even for a lag-one Markov process. If $E\left[\hat{\rho}_1\right]_n$ is replaced by its sample estimate and equation (2.88) is rearranged, an unbiased estimate of ρ_1 is obtained as

$$\left[\hat{\rho}_1^{(c)}\right]_n = \frac{n}{n-4}\left[\hat{\rho}_1\right]_n + \frac{1}{n-4} \tag{2.89}$$

which, if used in the lag-one Markov generating mechanism will ensure that estimates of ρ_1 derived from synthetic sequences of size n, denoted by $[\hat{\rho}^*_1]_n$, will satisfy

$$E[\hat{\rho}^*_1]_n = [\hat{\rho}_1]_n \qquad (2.90)$$

as required. However, the variance of $[\hat{\rho}_1^{(c)}]_n$ is larger than that of $[\hat{\rho}_1]_n$ by a factor of $(n/n-4)^2$, so a trade-off between bias and variance is implied. Nevertheless, application of the bias correction has been shown to reduce type II errors considerably when applying tests of significance to estimates of ρ_1 (Wallis and O'Connell, 1972). In a design situation knowledge of a loss function can offer some guidance as to which estimator should be used, and perhaps resolve the issue as to which type of resemblance should be maintained between historic and synthetic sequences.

Estimates of the sample variance from a lag-one Markov process are also biased and satisfy (Matalas, 1966)

$$E(S^2) = \sigma^2 \left[1 - \frac{2}{n(n-1)} \left[\frac{n\rho_1(1-\rho_1) - \rho_1(1-\rho_1^n)}{(1-\rho_1)^2}\right]\right] \qquad (2.91)$$

An unbiased estimate of σ^2 may be defined from equation (2.91), provided ρ_1 is known. In practice however, only an estimate of ρ_1 will be available, and complications arise which are considered in some detail by O'Connell and Wallis (1973).

In the case of approximations to dfGn, small sample biases in estimates of ρ_1, σ^2 and **h** have been shown to be acute ; however the complexity of the available approximations precludes the analytical derivation of small sample properties. Some results have been derived through Monte Carlo simulation (Wallis and Matalas, 1971), and techniques for applying bias corrections using analytical and Monte Carlo results have been considered by Matalas and Wallis (1974). The problem of biased estimates and the

application of bias corrections has been discussed by Slack (1972).

With the availability of a number of models for generating annual streamflow, the problem of model choice presents itself, and constitutes an important issue, as some simulations by Wallis and Matalas (1972) have shown that approximations to dfGn yield reservoir sizes considerably in excess of those yielded by the lag-one Markov process when levels of development greater than 0.80 are considered. Historic sequences of the length usually available on an annual time scale cannot be relied upon to provide reliable guidance as to the presence or absence of long-term persistence. Using the distribution of R/S, the rescaled range, derived through Monte Carlo simulations, Wallis and O'Connell (1973) have attempted to separate sequences generated by a long memory ARIMA (1,0,1) process from sequences generated by a short memory lag-one Markov process. For sample sizes $n < 100$ reliable separation was not possible.

The possibility of using Bayesian decision theory to assist in model choice and parameter estimation problems has already been considered in section (1. 4), and an attempt to implement such an approach has already been made by O'Connell and Wallis (1973). A generating process is postulated as being that of the real world, and an optimal design evolved on this basis using the sequent peak algorithm. Assumptions are then made concerning the identity of the real world, and the expected regrets accruing from each assumption may be evaluated using an assumed loss function. The procedure is repeated for each postulated generating process for the real world, with the assumed generating process yielding the minimum overall regrets representing the appropriate choice. Considerable difficulties have however been encountered with the experimental design, and ultimate guidance on model choice would appear to be conditional upon the design process.

2.8 Summary

The foregoing survey of models for generating synthetic flows shows that while the theoretical properties of fGn are entirely desirable, the application of fGn in the design of water resource systems is still a rather complex undertaking. The Broken Line process, while simple in concept, has not been developed sufficiently for practical application. The necessity is seen to exist for a generating process with the mathematical and computational simplicity of the lag-one Markov process but the long-term properties of which are close to those of fGn.

Chapter 3

A SIMPLE STOCHASTIC MODEL OF LONG-TERM PERSISTENCE

Statisticians have been directing their attentions to the analysis of time series for over half a century. The realization that a Gaussian independent process did not adequately describe the irregular behaviour of a number of naturally occurring phenomena started a search for stochastic models which incorporated measures of dependence. In his historic contribution, Yule (1927) introduced the idea of representing such irregular phenomena as a linear aggregation of random shocks. Many of the models being widely used in time series analysis today embody this fundamental concept.

With the advent of electronic computers, time series analysis entered a new era. Much of the theory previously formulated could now be applied and new areas of application emerged such as communications theory, econometrics, meteorology, and more recently, hydrology. As previously noted, a number of these models have been specifically adapted for use in synthetic hydrology. The present chapter is largely devoted to examining the properties of one of a class of linear stochastic models, called ARIMA models. The model, the ARIMA (1,0,1) process, is found to provide an adequate model of Hurst's law, and is accordingly developed for use in synthetic hydrology.

Section (3.1) contains a brief appraisal of linear stationary models, and relies largely on the treatment of Box and Jenkins (1970). Specifically, the AR(p), MA(q) and ARMA(p,q) processes are presented under the headings of (a) stationarity and invertibility conditions, and (b) autocorrelation function. The general ARIMA(p,d,q) process which subsumes the AR(p), MA(q) and ARMA(p,q) processes is formulated.

Within section (3.2) the deficiencies of the ARIMA (1,0,0) process in modelling Hurst's law are noted, while in section (3.3) the closely related ARIMA (1,0,1) process is examined in considerable detail and found to be an adequate model of Hurst's time series. Extensive use is made of sample function plots and "pox diagrams", and some small sample properties are derived through Monte Carlo simulation, which are then used in formulating the process for generating synthetic sequences. Some simple reservoir design experiments are reported in section (3.4) which illustrate the effects of long-term persistence and small sample biases on reservoir design. A brief reference to some experiments with higher order ARIMA models is made in section (3.5).

3.1. Linear Stationary Models

The following simple mathematical operators warrent definition as they facilitate a more concise presentation. Let X_t represent a stochastic process in discrete time. Define a backward shift operator B, (Box and Jenkins, 1970), as

$$B\,X_t = X_{t-1}$$

whence $$B^m X_t = X_{t-m}$$

Define a backward difference operator ∇ as follows:

$$\nabla X_t = X_t - X_{t-1} = (1-B)X_t$$

The operator ∇ has for its inverse the summation operator s given by:

$$\nabla^{-1}X_t = sX_t = \sum_{i=0}^{\infty} X_{t-i}$$

$$= X_t + X_{t-1} + X_{t-2} + \ldots$$

$$= (1 + B + B^2 + \ldots)X_t$$

$$= (1 - B)^{-1} X_t$$

3.1.1 Equivalent Forms for the General Linear Process

Let $\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots$ be a series of uncorrelated random variables, termed white noise, with zero mean, variance σ_ε^2, and autocorrelation function

$$\rho_k = \begin{cases} 1 & k = 0 \\ 0 & k > 0 \end{cases}$$

The process X_t, defined as

$$\begin{aligned} X_t &= \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \cdots \\ &= \sum_{j=1}^{\infty} \psi_j \varepsilon_{t-j} + \varepsilon_t \end{aligned} \tag{3.1}$$

is known as a general linear process, and constitutes a weighted summation of present and past values of the white noise process ε_t. The process is linear in the parameters $\psi_j \; \forall \; j$. Equation (3.1) suggests that X_t can be written alternatively as

$$\begin{aligned} X_t &= \pi_1 X_{t-1} + \pi_2 X_{t-2} + \cdots + \varepsilon_t \\ &= \sum_{j=1}^{\infty} \pi_j X_{t-j} + \varepsilon_t \end{aligned} \tag{3.2}$$

where X_t is "regressed" on previous values of the process. As equations (3.1) and (3.2) are equivalent representations, relations may be obtained between the ψ weights and the π weights. Using the backward shift operator, equation (3.1) may be written as

$$X_t = (1 + \sum_{j=1}^{\infty} \psi_j B^j) \varepsilon_t$$

or

$$X_t = \psi(B) \varepsilon_t \tag{3.3}$$

where

$$\begin{aligned} \psi(B) &= 1 + \sum_{j=1}^{\infty} \psi_j B^j \\ &= \sum_{j=0}^{\infty} \psi_j B^j \end{aligned}$$

with $\psi_0 = 1$. B is treated as a dummy variable whose j^{th} power is

the coefficient of ψ_j. Similarly, equation (3.2) may be written as

$$(1 - \sum_{j=1}^{\infty} \pi_j B^j) X_t = \varepsilon_t$$

or $$\pi(B) X_t = \varepsilon_t \qquad (3.4)$$

where $$\pi(B) = 1 - \sum_{j=1}^{\infty} \pi_j B^j$$

The application of $\psi(B)$ to both sides of equation (3.4) yields

$$\psi(B)\pi(B) X_t = \psi(B)\varepsilon_t = X_t$$

$\therefore$ $$\psi(B)\pi(B) = 1$$

or $$\pi(B) = \psi^{-1}(B)$$

Consequently the weights $\pi_j \; \forall \; j$ may be defined from the weights $\psi_j \; \forall \; j$ and vice versa (Box and Jenkins, 1970). To illustrate the duality between equation (3.1), termed an infinite moving average, and equation (3.2), termed an infinite autoregression, consider the case where

$$\psi_1 = -\theta, \; \psi_j = 0 \quad \text{for } j > 1$$

Therefore, $$X_t = \varepsilon_t - \theta \varepsilon_{t-1} \qquad (3.5)$$

$$= (1 - B\theta)\varepsilon_t$$

and $$(1 - B\theta)^{-1} X_t = \varepsilon_t$$

Expanding $(1 - B\theta)^{-1}$ yields

$$(1 + \theta B + \theta^2 B^2 + \theta^3 B^3 + \ldots) X_t = \varepsilon_t$$

which is equivalent to

$$X_t = -\theta X_{t-1} - \theta^2 X_{t-2} - \theta^3 X_{t-3} - \ldots + \varepsilon_t \qquad (3.6)$$

so that equation (3.5), which constitutes a first order moving average, may equivalently be expressed as the infinite autoregression of equation (3.6).

The autocovariance function of the general linear process of

equation (3.1) may be expressed as (Box and Jenkins, 1970)

$$\gamma_k = \sigma_\epsilon^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+k} \tag{3.7}$$

For k = 0, equation (3.7) reduces to

$$\gamma_0 = \sigma_\epsilon^2 \sum_{j=0}^{\infty} \psi_j^2 \tag{3.8}$$

which is the variance of the process X_t. For the process given by equation (3.1) to have a finite variance, the weights ψ_j must decrease rapidly enough to ensure that

$$\sum_{j=0}^{\infty} \psi_j^2 \tag{3.9}$$

converges.

3.1.2 Stationarity and Invertibility Conditions for a Linear Process

The convergence of the series in equation (3.9) ensures that the process has a finite variance. In addition, the autocovariances and autocorrelations of the process must satisfy a set of conditions to ensure stationarity. These conditions can be embodied in the condition that the series $\psi(B)$ converges for $B \leq 1$. For example, consider equation (3.1) with

$$\pi_1 = \varphi_1, \ \pi_j = 0 \quad \text{for } j > 1$$

Therefore $X_t = \varphi_1 X_{t-1} + \varepsilon_t$ (3.10)

$$(1-B\varphi_1)X_t = \varepsilon_t$$

whence $$X_t = (1-B\varphi_1)^{-1} \varepsilon_t$$

Consequently, for $B \leq 1$, the series

$$1 + B\varphi_1 + B^2\varphi_1^2 + \ldots$$

diverges for $\varphi_1 \geq 1$. Consequently, stationarity requires $\varphi_1 < 1$.

A restriction must be enforced on the π weights in equation

(3.2) to ensure that a condition known as invertibility be satisfied for this process. This condition for a moving average corresponds to the stationarity condition for an autoregression but the conditions are essentially independent. Consideration of the particular case of equation (3.5) serves to illustrate this point. The series

$$1 + \theta B + \theta^2 B^2 + \ldots$$

diverges for $B \leqslant 1$ and $|\theta| \geq 1$. However, equation (3.6) represents a stationary process for all values of θ. The condition that the series $\pi(B)$ converges for $B \leq 1$ is known as invertibility.

3.1.3 The AR(p) Process

The representations in equations (3.1) and (3.2) are of little practical use insofar as estimating the infinite series of parameters ψ_j and π_j $\forall$ j is concerned. Finite data series may only allow reliable estimates of ψ_j for $j \leq q$ and of π_j for $j \leq p$ where q and p are generally not greater than 2. Consequently only the properties of such processes are of practical interest.

The special case of equation (3.2) where $\pi_j = 0$ $\forall$ $j > p$ is termed an autoregressive process of order p: or more succinctly, an AR(p) process. Equation (3.2) may now be written as

$$X_t = \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \ldots + \varphi_p X_{t-p} + \varepsilon_t \qquad (3.11)$$

where the symbols $\varphi_1, \ldots \varphi_p$ are now used to denote the finite set of parameters. Using the backward shift operator B the process may be written as

$$(1 - \varphi_1 B - \varphi_2 B^2 - \ldots - \varphi_p B^p) X_t = \varepsilon_t$$

or $$\varphi(B) X_t = \varepsilon_t.$$

Consequently, $$X_t = \varphi^{-1}(B)\, \varepsilon_t \qquad (3.12)$$

represents the output from a linear filter with transfer function $\varphi^{-1}(B)$.

(a) Stationarity and Invertibility

The finite set of parameters $\varphi_1,\ldots,\varphi_p$ must satisfy a set of conditions for stationarity. Factoring the polynomial $\varphi(B)$ yields

$$\varphi(B) = (1 - G_1B)(1 - G_2B) \ldots (1 - G_pB)$$

Thus, in equation (3.12), $\varphi^{-1}(B)$ may be expanded in partial fractions to yield

$$X_t = \varphi^{-1}(B)\varepsilon_t = \sum_{i=1}^{p} \frac{k_i}{(1 - G_iB)} \varepsilon_t$$

Stationarity requires that $\varphi^{-1}(B)$ be a convergent series for $|B| \leq 1$; as a result, $|G_i| < 1$ for $i = 1,2,\ldots p$. Equivalently, the roots of $\varphi(B) = 0$, referred to as the zeroes of the polynomial $\varphi(B)$, must lie outside the unit circle. The equation $\varphi(B) = 0$ is known as the characteristic equation for the AR(p) process.

Since the series

$$\varphi(B) = 1 - \varphi_1B - \varphi_2B^2 - \ldots - \varphi_pB^p$$

is finite, no restrictions are imposed on the parameters φ_i, $i = 1,\ldots p$, to ensure invertibility.

(b) Autocorrelation Function

The autocorrelation function of the AR(p) process is a function of the set of parameters φ_i, $i = 1,2,\ldots, p$, and satisfies the following recurrence relationship:

$$\rho_k = \varphi_1\rho_{k-1} + \varphi_2\rho_{k-2} + \ldots + \varphi_p\rho_{k-p} \qquad k > 0 \quad (3.13)$$

A solution of the difference equation (3.13) shows that in general, ρ_k consists of a mixture of damped exponentials and damped sine waves.

The parameters φ_i, $i = 1,2,\ldots p$ are related to the autocorrelations ρ_i, $i = 1,2,\ldots p$ through a set of linear equations as follows:

$$\begin{aligned}
\rho_1 &= \varphi_1 + \varphi_2\rho_1 + \ldots + \varphi_p\rho_{p-1} \\
\rho_2 &= \varphi_1\rho_1 + \varphi_2 + \ldots + \varphi_p\rho_{p-2} \\
&\vdots \qquad\qquad \vdots \\
\rho_p &= \varphi_1\rho_{p-1} + \varphi_2\rho_{p-2} + \ldots + \varphi_p
\end{aligned} \tag{3.14}$$

which may be solved for the parameters φ_i, $i = 1,2,\ldots p$.

3.1.4. The MA(q) Process

The special case of equation (3.1) where $\phi_j = 0 \;\forall\; j > q$ is termed a moving average process of order q, or, more succinctly, a MA(q) process. Equation (3.1) may now be written as

$$X_t = \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \ldots - \theta_q\varepsilon_{t-q} \tag{3.15}$$

where the symbols $-\theta_1,\ldots, -\theta_q$ are now used to denote the finite set of parameters. Again using the shift operator B the process may be written as

$$X_t = (1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q)\,\varepsilon_t$$

or $$X_t = \theta(B)\,\varepsilon_t$$

Consequently X_t represents the output from a linear filter with transfer function $\theta(B)$.

(a) Stationarity and Invertibility

The series

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

is finite and consequently no restrictions are required on the parameters of the MA(q) process to ensure stationarity. However, the set of parameters θ_i, $i = 1,2,\ldots,q$ must satisfy a set of conditions to ensure invertibility. The polynomial $\theta(B)$ may be factored as follows

$$\theta(B) = (1 - H_1 B)(1 - H_2 B) \ldots (1 - H_q B)$$

$$= \prod_{j=1}^{q} (1 - H_j B)$$

Thus, in equation (3.16), $\theta^{-1}(B)$ may be expanded in partial fractions to yield

$$\varepsilon_t = \sum_{j=1}^{q} \frac{m_j}{(1 - H_j B)} X_t$$

Stationarity requires that $\theta^{-1}(B)$ form a convergent series for $|B| \leq 1$: as a result, $|H_j| < 1$ for $j = 1,2,\ldots,q$. This is equivalent to the specification that the roots of $\theta(B) = 0$, referred to as the zeroes of the polynomial $\theta(B)$, must lie outside the unit circle. The equation $\theta(B) = 0$ is known as the characteristic equation for the process.

(b) Autocorrelation Function

The autocorrelation function of the MA(q) process is given as

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

$$= 0 \qquad k > q \qquad (3.16)$$

and is a non linear function of the parameters θ_i, $i = 1,2,\ldots q$. Apart from the case where $q = 1$, equation (3.16) with $k = 1,\ldots,q$, must be solved iteratively for the parameters θ_j, $j = 1,2,\ldots,q$.

3.1.5 The ARMA (p,q) Process

Within section (3.1.1) it was noted that an infinite number of autoregressive parameters is required to achieve an equivalent representation of an MA(1) process. Similarly, an infinite number of moving average parameters is required to characterize an AR(1) process. Consequently, if a process were a mixture of autoregressive and moving average terms, then a minimal parameterization of such a process could not be obtained through either an AR(p) or an MA(q) process. Hence, a process consisting of a mixture of moving average and autoregressive terms must be considered. The process

$$X_t = \varphi_1 X_{t-1} + \ldots + \varphi_p X_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \ldots - \theta_q \varepsilon_{t-q}$$

which may be expressed as (3.17)

$$(1 - \varphi_1 B - \varphi_2 B^2 - \ldots - \varphi_p B^p) X_t = (1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q) \varepsilon_t$$

or $$\varphi(B) X_t = \theta(B) \varepsilon_t \quad (3.18)$$

is termed a mixed moving average autoregressive process of order (p,q) which may be abbreviated to ARMA(p,q). Dividing through in equation (3.18) by $\varphi(B)$ yields

$$X_t = \frac{\theta(B)}{\varphi(B)} \varepsilon_t$$

suggesting that the ARMA (p,q) process may be considered as the output from a linear filter with transfer function $\frac{\theta(B)}{\varphi(B)}$ when the input is white noise ε_t. There are two equivalent conceptions of the process:

(a) an AR(p) process

$$\varphi(B) X_t = a_t$$

where a_t follows an MA(q) process $a_t = \theta(B) \varepsilon_t$

(b) an MA(q) process

$$X_t = \theta(B) b_t$$

where b_t follows an AR(p) process $\varphi(B) b_t = \varepsilon_t$

Consequently the ARMA (p,q) process may also be thought of as the output from two linear filters in series with transfer functions $\varphi^{-1}(B)$ and $\theta(B)$ when the input is white noise ε_t. The order of the filters is immaterial.

(a) Stationarity and Invertibility

The existence of moving average terms on the right of equation (3.17) does not affect the argument established in section (3.1.4) for stationarity of a process, provided that the roots of the characteristic equation

$$\varphi(B) = 0$$

lie outside the unit circle. Similarly, the roots of

$$\theta(B) = 0$$

must lie outside the unit circle to ensure invertibility of the process. Thus, the stationarity and invertibility conditions are essentially independent for an ARMA (p,q) process.

(b) Autocorrelation Function

The autocorrelation function for an ARMA (p,q) process is defined as

$$\rho_k = \varphi_1\rho_{k-1} + \varphi_2\rho_{k-2} + \ldots + \varphi_p\rho_{k-p} \qquad k \geq q+1 \qquad (3.19)$$

or

$$\varphi(B)\rho_k = 0$$

Consequently, for an ARMA (p,q) process, the values of the q autocorrelations $\rho_q, \rho_{q-1},\ldots,\rho_1$ will depend on the parameters θ_j, $j = 1,2,\ldots,q$ and φ_i, $i = 1,2,\ldots,p$. Further, the p autocorrelations $\rho_q, \rho_{q-1},\ldots,\rho_{q-p+1}$ provide the necessary starting values for the difference equation (3.19), which then determines the behaviour of the autocorrelation function entirely for higher lags. If $(q-p) < 0$, the entire autocorrelation function ρ_k for $k = 0,1,2\ldots$ will consist of a mixture of damped exponentials and/or damped sine waves, whose nature is determined by the polynomial $\varphi(B)$ and the starting values. If, however, $(q-p) \geq 0$ there will be $(q-p)$ initial values $\rho_1, \rho_2,\ldots,\rho_{q-p}$ which do not follow this general pattern.

Thus, in general, the parameters φ_i, $i = 1,2,\ldots,p$ may be obtained through solving a set of linear equations obtained from equation (3.19) with $k = q+1, q+2,\ldots,q+p$. The parameters θ_i may then be obtained through solving a set of q non linear equations in terms of $\rho_1, \rho_2,\ldots,\rho_q$, where $\rho_1, \rho_2,\ldots,\rho_q$ are functions of the parameters φ_i and θ_i.

3.1.6 The ARIMA (p,d,q) Process

The stationary ARMA (p,q) process may be operated on to obtain a non-stationary process which still involves only (p + q) parameters. This may be achieved by applying the simple summation operator

$$sX_t = X_t + X_{t-1} + \cdots = \sum_{h=-\infty}^{t} X_h$$

successively to the X_t values which follow an ARMA (p,q) process. Reapplication of the operator s yields

$$s^2X_t = sX_t + sX_{t-1} + \cdots = \sum_{i=-\infty}^{t} \sum_{h=-\infty}^{i} X_h$$

The operator s may be applied d times to the process X_t; this in turn implies that if the difference operator

$$\nabla = s^{-1}$$

is applied to the process

$$Z_t = s^d X_t$$

d times, then the stationary process

$$\varphi(B)X_t = \theta(B)\,\varepsilon_t$$

is obtained. Consequently, the process Z_t may be defined as

$$\varphi(B)\,\nabla^d Z_t = \theta(B)\,\varepsilon_t \qquad (3.20)$$

Equation (3.20) is termed an autoregressive integrated moving average process, or more succinctly, an ARIMA (p,d,q) process. The general ARIMA (p,d,q) process subsumes AR(p) models, MA(q) models, ARMA(p,q) models (d = 0), and the integrated forms of all three, and is represented in block diagram form in figure (3.1) as a succession of filtering operations on white noise. The mathematical-statistical properties of the non-

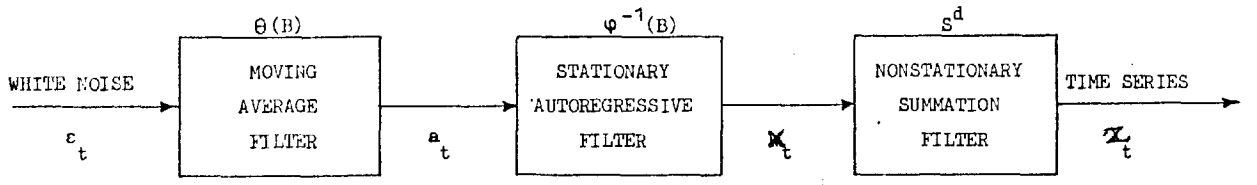


Figure (3.1). Block diagram of ARIMA (p,d,q) process.

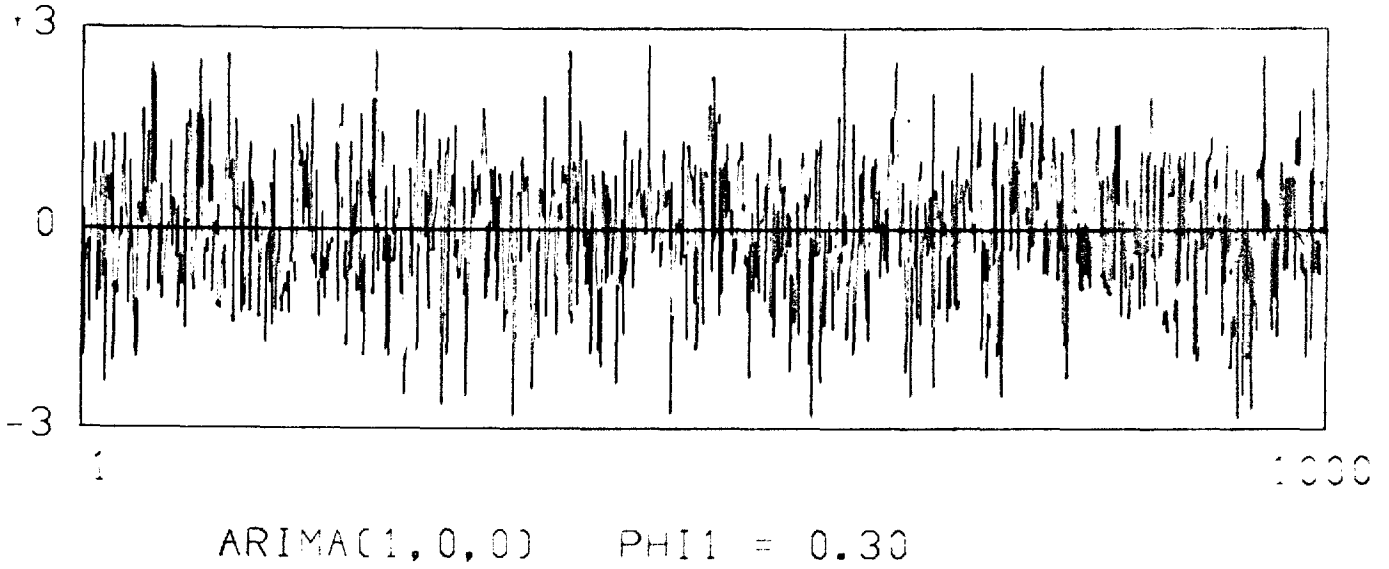


Figure (3.2) Plot of 1000 points from an ARIMA (1,0,0) or lag-one Markov process with $\varphi = \rho_1 = 0.3$. The sample has been standardized to have zero mean and unit variance.

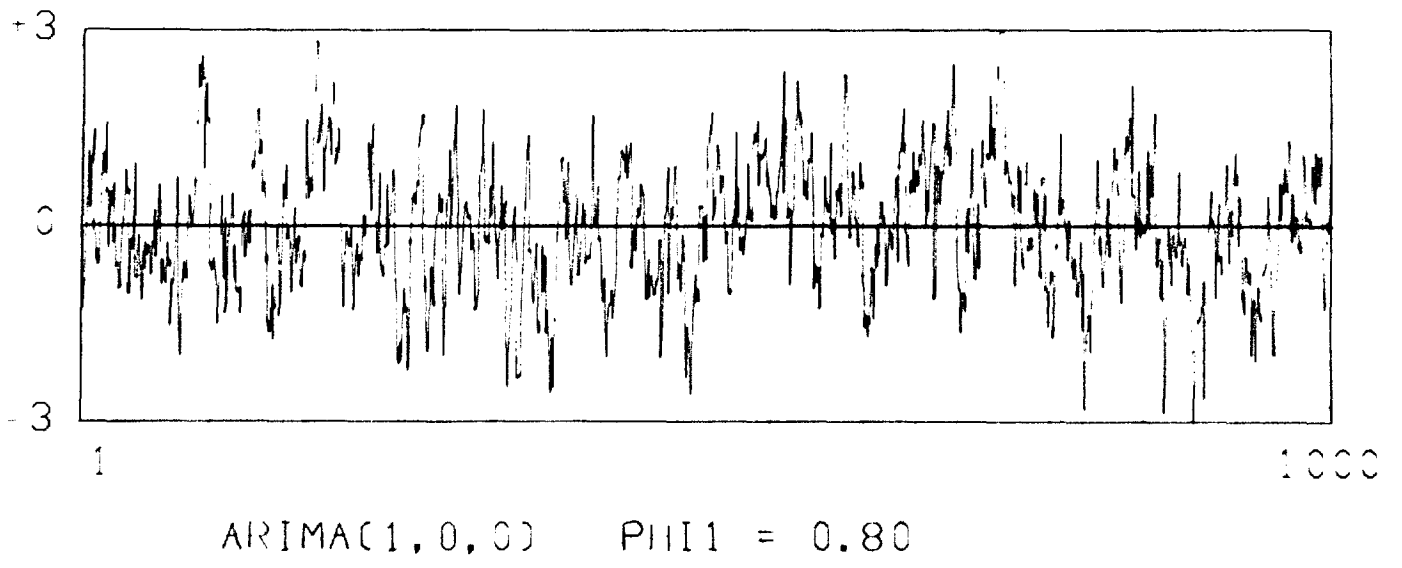


Figure (3.3) Plot of 1000 points from an ARIMA (1,0,0) or lag-one Markov process with $\varphi = \rho_1 = 0.8$. The sample has been standardized to have zero mean and unit variance.

stationary ARIMA (p,d,q) process with d > 0 are unmanageable as these properties are a function of time. While a non-stationary ARIMA process may be used for forecasting applications, such a process cannot be used to generate synthetic sequences (Watts, 1972).

While there may be physical grounds for claiming that annual streamflow should not be considered stationary, especially over long time spans, only stationary models can be used for generating synthetic sequences; in fact stationary processes can often model what appear to be non-stationary effects. This stems from the fact that there is no sudden transition from stationary to non-stationary behaviour; for example, a stationary ARIMA (1,0,1) process merges into a non-stationary ARIMA (0,1,1) across a boundary defined by the stationarity condition for the former process. As a result, finite realizations from either process with parameter values near the boundary may be statistically inseparable. Further, stationary ARIMA (p,d,q) processes (d = 0) with parameter values near the boundary frequently possess autocorrelation functions which damp out rather slowly, suggesting that such processes might be useful in modelling long-term persistence. The simplest model in the ARIMA class which possesses such an autocorrelation function is the ARIMA (1,0,1) process. Closely related to the ARIMA (1,0,1) process is the ARIMA (1,0,0) or lag-one Markov process, which has been unsuccessful in accounting for Hurst's findings.

The ARMA notation is sufficient to describe stationary mixed models; however, as non-stationary ARIMA models will be referred to occasionally, the more general ARIMA notation will be adhered to.

3.2 The ARIMA (1,0,0) Process

The first order autoregressive or AR(1) process is defined as

$$X_t = \varphi_1 X_{t-1} + \varepsilon_t$$

or $$(1 - B\varphi_1)X_t = \varepsilon_t \qquad (3.21)$$

Thus,

$$\varphi(B)X_t = \varepsilon_t$$

Equation (3.21) is frequently referred to as a lag-one Markov process.

3.2.1 Parameter Space

From the conditions established for stationarity in section (3.1.4) the expansion

$$\varphi^{-1}(B) = 1 + \varphi_1 B + \varphi_1^2 B^2 + \ldots$$

must form a convergent series for $B \leq 1$ to ensure stationarity. Consequently, the corresponding parameter space is

$$-1 < \varphi_1 < 1$$

As already noted, no conditions are imposed on AR(p) processes to ensure invertibility.

3.2.2 Autocorrelation Function

The autocorrelation function of the ARIMA (1,0,0) process satisfies the difference equation (3.13) with p = 1, resulting in

$$\rho_k = \varphi_1 \rho_{k-1} \qquad k > 0$$

which reduces to

$$\rho_1 = \varphi_1$$

$$\rho_k = \varphi_1^k \qquad k \geq 0 \qquad (3.22)$$

Thus, the autocorrelation function decays exponentially and when $\varphi_1 < 0$, oscillates in sign. Consequently φ_1 defines completely the shape of the autocorrelation function for higher lags, which greatly restricts the range of shapes that the function may assume. As a result only persistence of a special kind may be modelled adequately with this process.

3.2.3 Sample Functions

Sample function plots of 1000 values standardized to have zero mean and unit variance for $\varphi_1 = 0.3$ and $\varphi_1 = 0.8$ are presented in figures (3.2) and (3.3) respectively. The high frequency behaviour in figure (3.2) is not unlike that encountered within annual streamflow sequences where an average value of $\rho_1 = \varphi_1$ is about 0.3. However, the distinct absence of low frequency behaviour in figure (3.2) is noticeable, and while figure (3.3) portrays these effects, the "local" behaviour of the series is now too smooth to adequately represent annual streamflow.

3.2.4 Behaviour of h

As previously noted, a linear stationary stochastic process which has $R/S \sim n^h$ with h a constant in the range $0.5 < h < 1$, possesses a unique autocovariance function i.e. that of dfGn. Accurate approximations to dfGn such as the type 1 proposed by Mandelbrot and Wallis (1969a) ensure that Hurst's law with $0.5 < h < 1$ is closely followed within a finite transient which can be made arbitrarily long. Other generating processes which seek to model Hurst's law must be viewed in the light of how well the autocorrelation function of dfGn is approximated or equivalently, how well Hurst's law is modelled. In addition, some attention must be paid to ensure that the high frequency properties of observed streamflow are reproduced.

The lag-one Markov process is traditionally fitted in the high frequency domain using an estimate of ρ_1, the lag-one autocorrelation coefficient which specifies the autocorrelation function uniquely. For values of ρ_1 representative of annual streamflow, R/S grows faster than $n^{0.5}$ for small n and is a curvilinear function of n, but quickly assumes proportionality to $n^{0.5}$ thereafter. Even within the short

transient, Hurst's law is not closely followed; in fact, if this transient is described functionally by Hurst's law, h is then represented as varying with n.

The Hurst estimator of h, K, was used by Matalas and Huzzen (1967) to define the expected value of K as a function of n and ρ_1 for the lag-one Markov process using Monte Carlo simulation, and their results have already been reproduced in table (2.1). For values of ρ_1 which are representative of annual streamflow, $\tilde{E}[K]_n$ decreases monotonically towards 0.5 with increasing n, thus reflecting the short memory nature of the process. The value of $\tilde{E}[K]_n$ for any particular n cannot be increased without increasing ρ_1 which in most situations will result in the distortion of the high frequency characteristics of observed streamflow.

The inability of the lag-one Markov process to model observed low frequency and high frequency properties of a record may be emphasised through reference to the observed annual flows of the Colorado River at Lees Ferry, where for n = 61 years, ρ_1 = 0.22 and K = 0.82. Wallis and Matalas (1971) have pointed out that it is extremely unlikely that such estimates of K and ρ_1 could be provided by a realization from a Markovian generating mechanism. In general, either ρ_1 or K may be preserved in synthetic sequences but not both.

3.2.5 Formulation for Synthetic Hydrology

The lag-one Markov process has been extensively documented for synthetic hydrology by Matalas (1967) and Fiering and Jackson (1971), and a brief review of this documentation has been presented in section (2.2).

3.3 The ARIMA (1,0,1) Process

The ARIMA (1,0,1) process incorporates one autoregressive term and one moving average term and may be defined as

$$X_t - \varphi_1 X_{t-1} = \varepsilon_t - \theta_1 \varepsilon_{t-1} \qquad (3.23)$$

$$(1 - B\varphi_1)X_t = (1 - B\theta_1)\varepsilon_t$$

$$\varphi(B)X_t = \theta(B)\varepsilon_t$$

Conceptually, the process may be formulated in either of two ways:

(a) as an ARIMA (1,0,0) process applied to the process a_t as:

$$X_t = \varphi_1 X_{t-1} + a_t$$

where

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

(b) as an ARIMA (0,0,1) process applied to the process b_t as

$$X_t = b_t - \theta_1 b_{t-1}$$

where

$$b_t = \varphi_1 b_{t-1} + \varepsilon_t$$

Thus, the mixed ARIMA (1,0,1) process may embody features of both the ARIMA (1,0,0) process and the ARIMA (0,0,1) process, thereby allowing the combination of low frequency properties of the former process with high frequency properties of the latter process.

3.3.1 Parameter Space

The stationarity condition for the ARIMA (1,0,1) process is identical to that for the ARIMA (1,0,0) process: i.e.,

$$-1 < \varphi_1 < +1$$

while the invertibility condition is as for the ARIMA (0,0,1) process: i.e.,

$$-1 < \theta_1 < +1$$

The parameter space is represented as a square as shown in figure (3.4).

For convenience the subscripts on φ_1 and θ_1 will now be dropped.

3.3.2 Autocorrelation Function

The autocorrelation function for the ARIMA (1,0,1) process will, for lag-one, be determined by φ and θ jointly, as

$$\rho_1 = \frac{(\varphi - \theta)(1 - \varphi\theta)}{1 + \theta^2 - 2\varphi\theta} \qquad (3.24)$$

whereas φ determines the behaviour of ρ_k for $k \geq 2$ as

$$\rho_k = \varphi\rho_{k-1} \qquad k \geq 2 \qquad (3.25)$$

Depending on where φ and θ fall within the parameter space shown in figure (3.4) the autocorrelation (a.c.) function displays different behavioural patterns and may conveniently be examined within the subregions 1-6. The process offers considerable additional flexibility over the ARIMA (1,0,0) process with respect to the shape of the a.c. function. In general, the absolute value of the a.c. decays exponentially from ρ_1 onwards. For fixed φ, θ defines ρ_1 and consequently determines the high frequency behaviour of the process. The sign of ρ_1 is determined by the sign of $(\varphi-\theta)$ and dictates from which side of zero the exponential decay takes place. The line $\varphi = \theta$ represents the special case where the process degenerates to white noise.

Subregion 1:

Herein, while φ is positive the sign of $(\varphi-\theta)$ is negative and consequently, so is the a.c. function for all lags. The process is consequently rich in high frequencies: however, high values of φ allow a slow decay in the a.c. function thereby allowing a large "effective memory".

Subregion 2:

Within this area, φ is negative and thus defines an oscillatory geometric decay in ρ_k for $k > 1$. ρ_1 is always negative as dictated

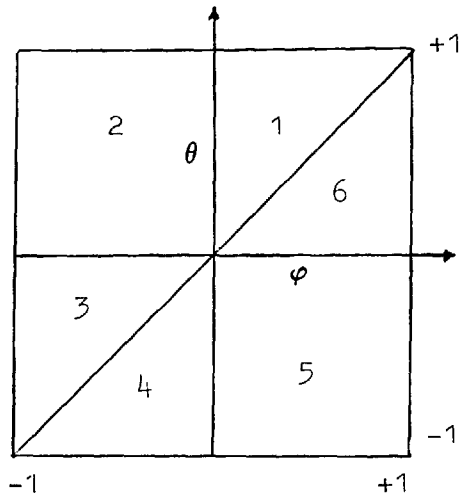


Figure (3.4) Parameter space for the stationary invertible ARIMA (1,0,1) process.

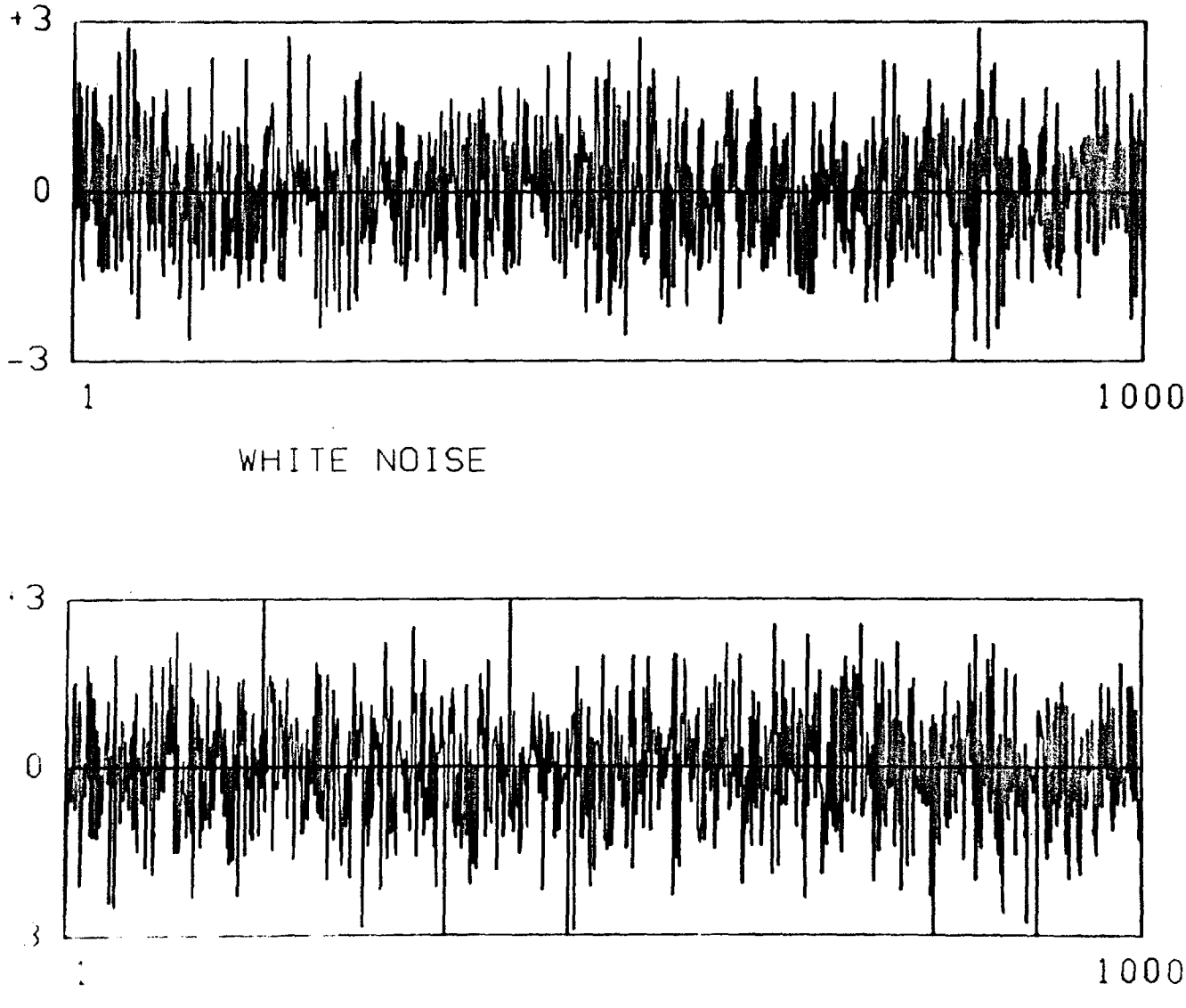


Figure (3.5) Plots of 1000 points of discrete white noise. Each sample of 1000 points has been standardized to have zero mean and unit variance.

by the sign of $(\varphi-\theta)$. Approaching the boundary $\varphi = -1$ the a.c. function decays slowly from the value of ρ_1 which is jointly defined by φ and θ.

Subregion 3:

Here, φ and θ are both negative where $\varphi < \theta$. As a result the process is highly oscillatory and thus rich in high frequencies; on the boundary $\varphi = -1$ the a.c. function fails to damp out.

Subregion 4:

$\varphi > \theta$ and as a result ρ_1 is positive. However, the sign of φ dictates an oscillatory decay for the a.c. function and the resulting process is rich in high frequencies.

Subregion 5:

The shape of the a.c. function here is not unlike that of the ARIMA (1,0,0) process; high values of φ result in high values of ρ_1, which is positive.

Subregion 6:

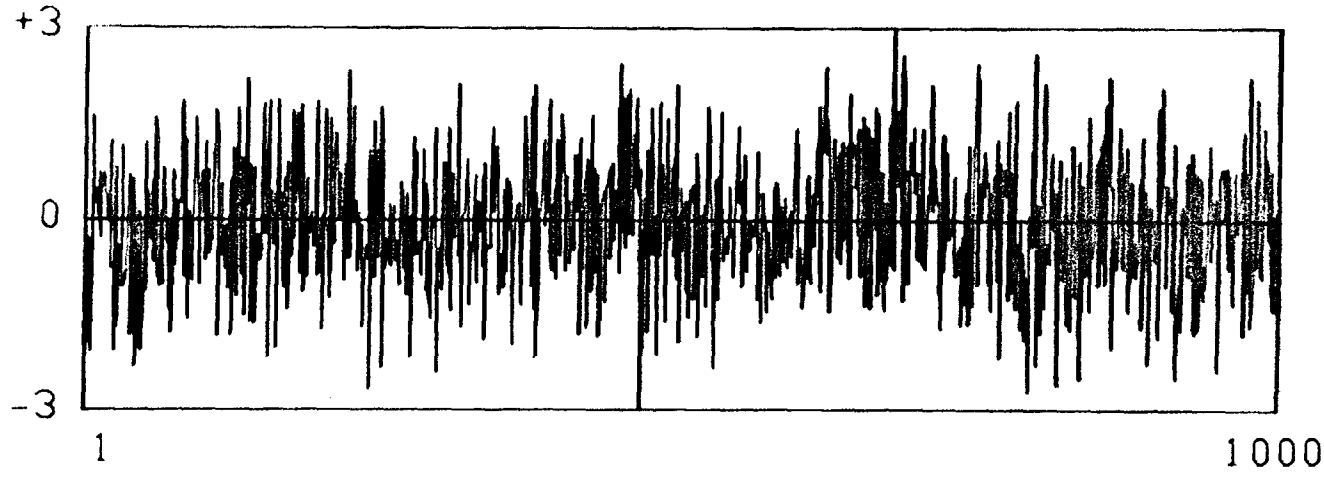
This subregion is of particular interest from the viewpoint of a plausible model of annual streamflow. For all values of φ and θ, ρ_1 is positive and may assume any value in the range $0 < \rho_1 < \varphi$. High values of φ and θ combine to yield low values of ρ_1. For φ fixed, decreasing θ results in increasing ρ_1. For θ fixed, increasing φ results in the autocorrelation function dying out more slowly and hence in a longer "effective memory". From the standpoint of modelling low and high frequency effects a desirable blend may be defined here. For example, $\varphi = 0.90$ and $\theta = 0.80$ combine to yield $\rho_1 = 0.14$. Long run or low frequency effects result from the high value of φ while ρ_1 preserves high frequency effects.

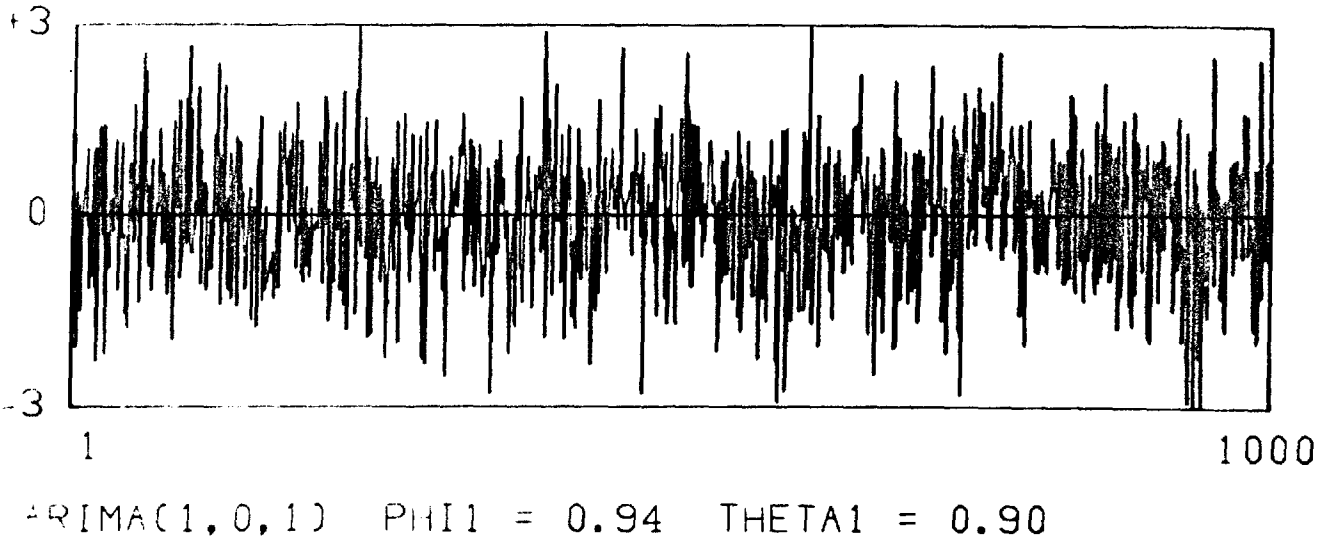
3.3.3 Sample Functions

In order to illustrate the nature of the ARIMA (1,0,1) process graphically, a value of $\varphi = 0.94$ has been chosen and successively decreasing values of θ combined with φ to generate sample function plots of 1000 values standardized to have zero mean and unit variance. The case of $\varphi = 0.94$ and $\theta = 0.94$ in figure (3.5) corresponds to white noise and is included as a starting point. Neither low frequencies or high frequencies preponderate, the distribution of variance being independent of frequency. However, as θ is decreased, ρ_1 increases, and high frequencies gradually give way to low frequencies. Graphically, figures (3.6-3.9) illustrate the gradual emergence of low frequency behaviour as θ decreases, and the mixing of low frequency and high frequency behaviour which is generally necessary to model annual streamflow. As $\theta \to 0$ however, $\rho_1 \to \varphi$ and the process degenerates into a "locally smooth" ARIMA (1,0,0) process. As φ decreases the low frequencies give way to "medium" frequencies; examples of this behaviour are given in figures (3.10) and (3.11). The range of values of ρ_1 which may be modelled decreases as φ decreases; φ, in fact, represents an upper limit on ρ_1 in the case where $\theta = 0$. In general, the sample functions of the ARIMA (1,0,1) process for $\varphi > \theta > 0$ are very similar to those for filtered dfGn approximations with comparable values of ρ_1.

3.3.4 Behaviour of h

The ARIMA (1,0,1) process is within the "Brownian domain of attraction", and theoretically, h equals 0.5 for this process. However, finite memory approximations to dfGn also fall within the Brownian domain. Nevertheless, the filtered approximation to dfGn developed by Matalas and Wallis (1971b) has the parameter h explicitly incorporated, thus ensuring that $R/S \sim n^h$ with $0.5 < h < 1$ within a finite transient.





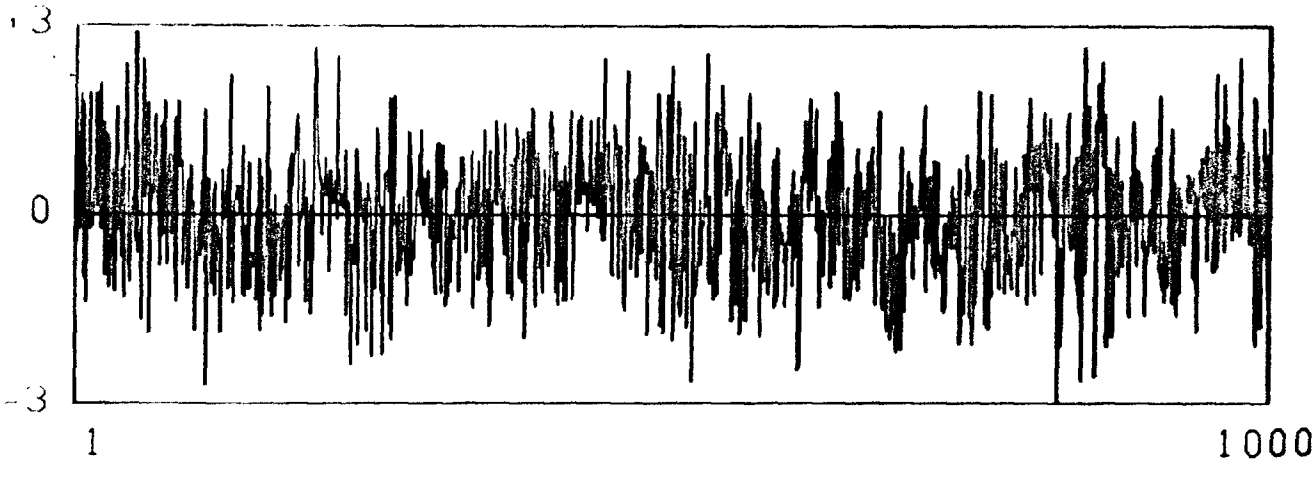
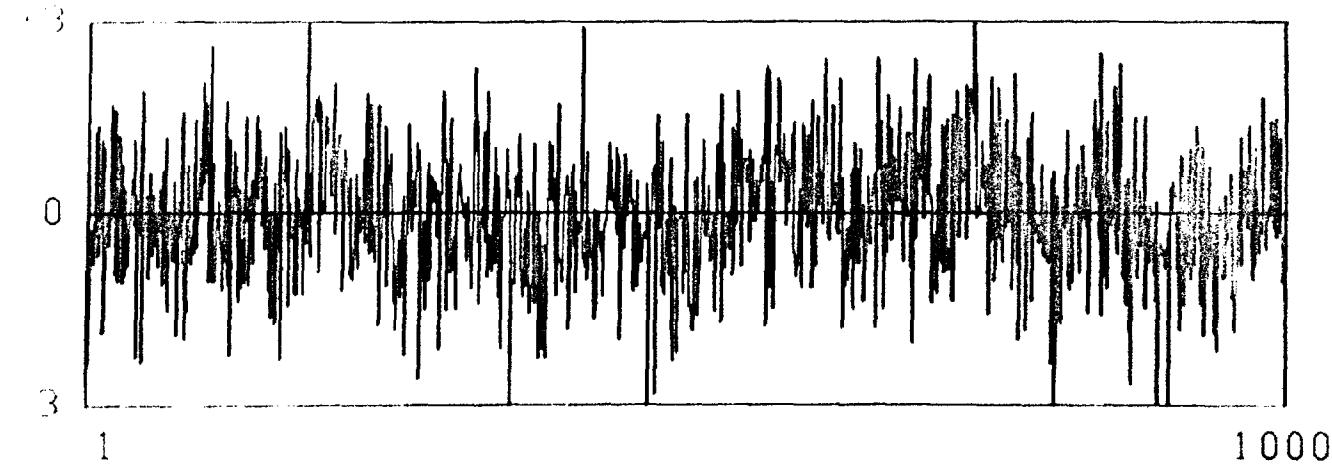
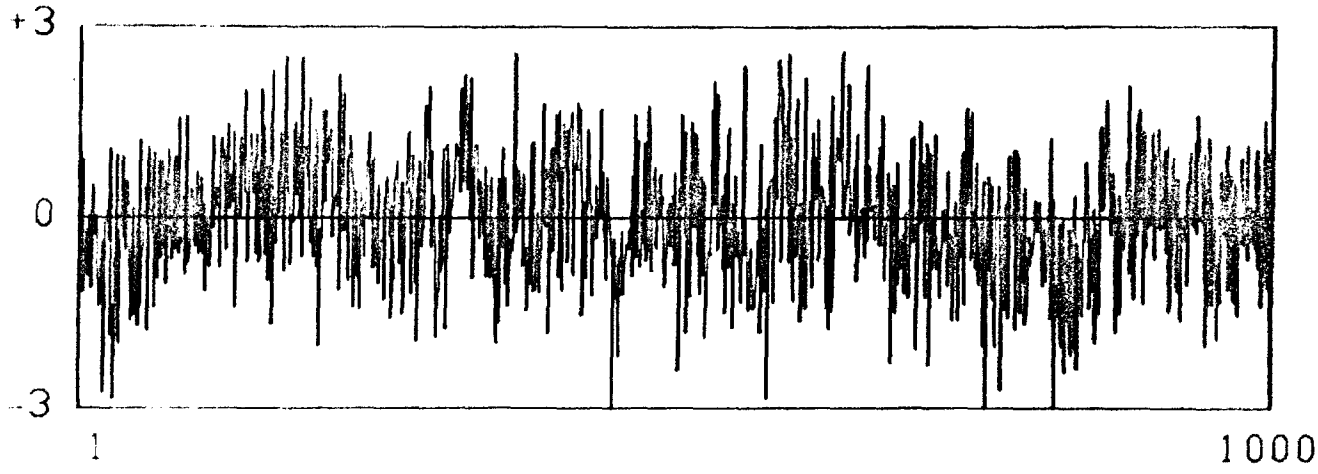


Figure (3.6) Plots of 1000 points of an ARIMA (1,0,1) process with $\varphi = 0.94$ and $\theta = 0.92, 0.90, 0.88$. Each 1000 points has been standardized to have zero mean and unit variance.





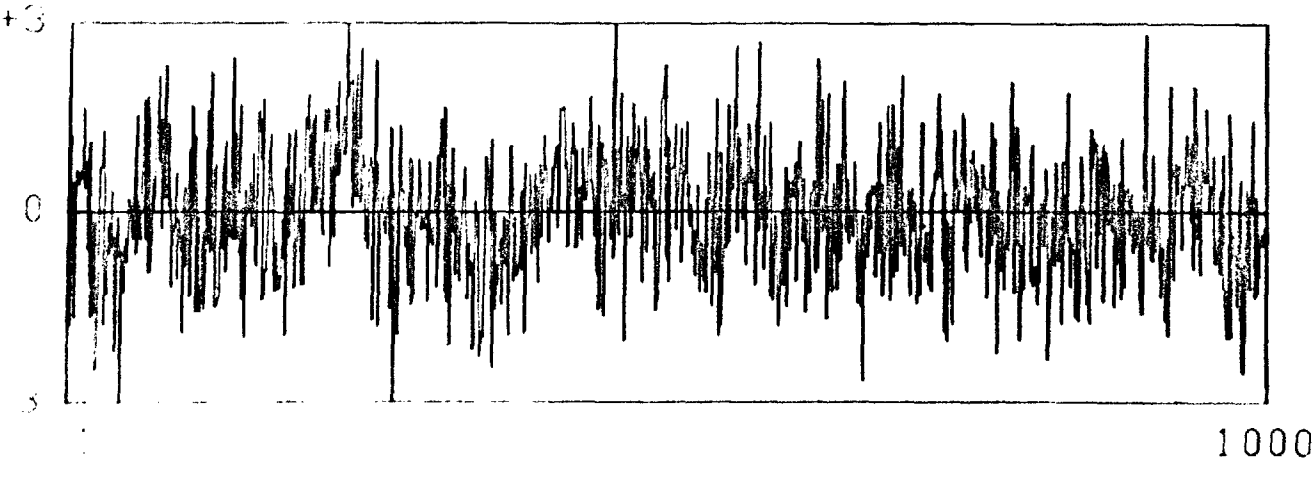
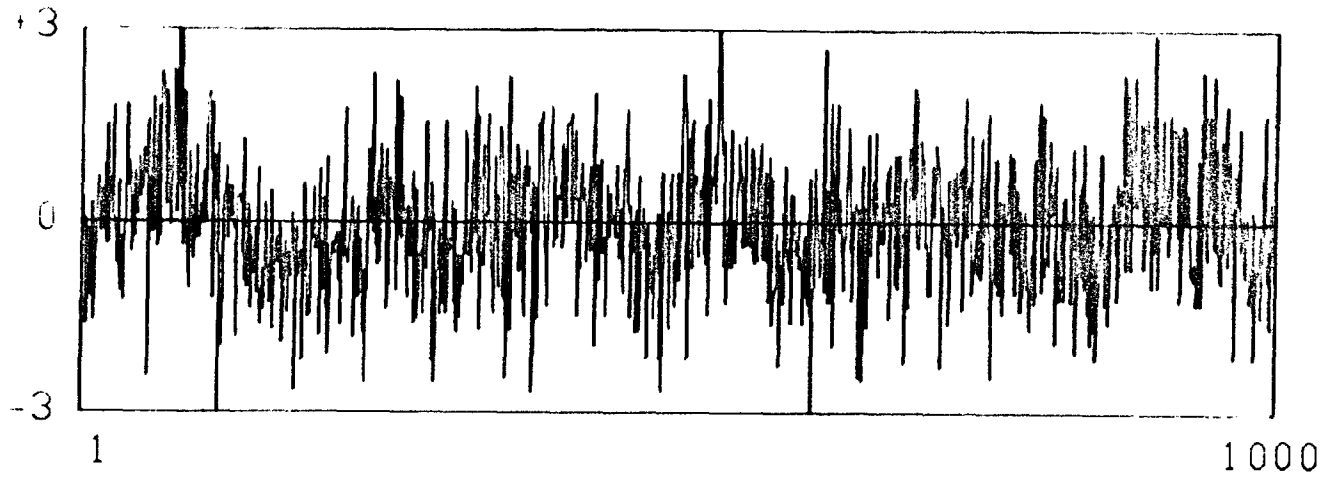
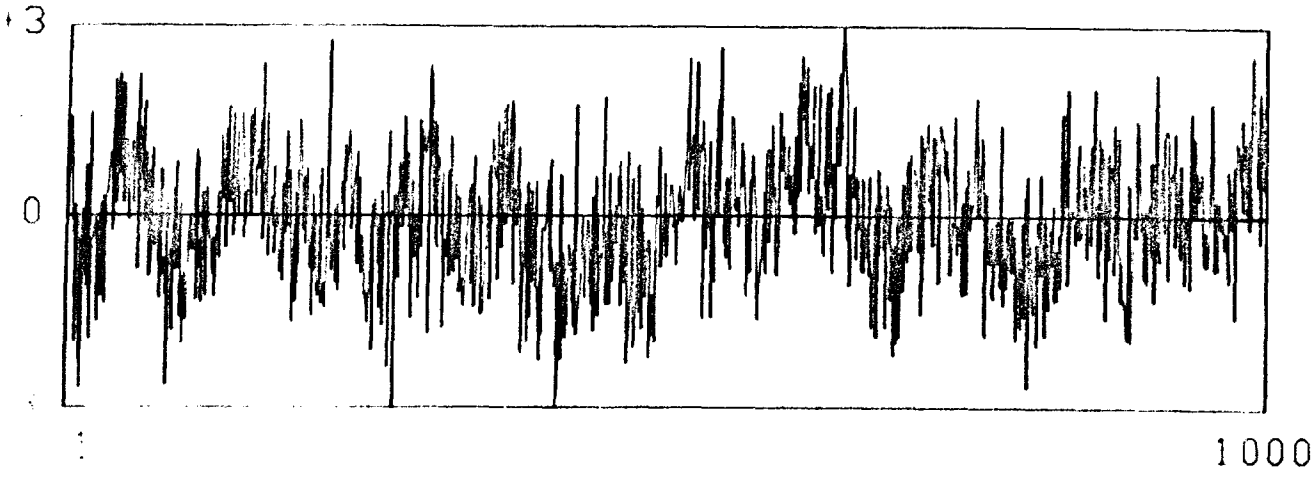


Figure (3.7) Plots of 1000 points from an ARIMA (1,0,1) process with $\varphi = 0.94$ and $\theta = 0.86$, 0.84, 0.82. Each sample of 1000 points has been standardized to have zero mean and unit variance.





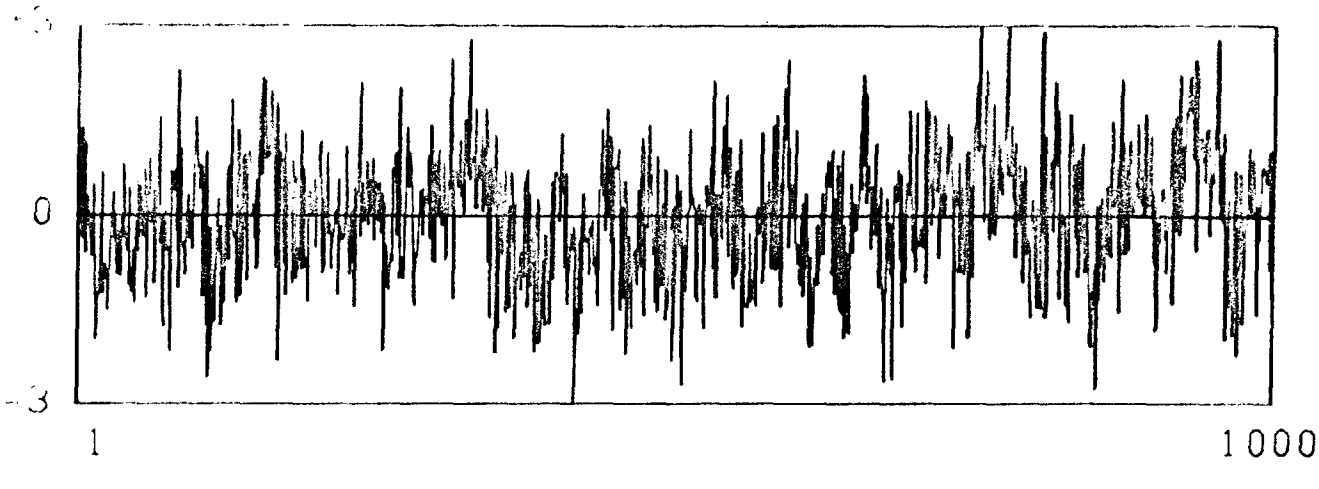
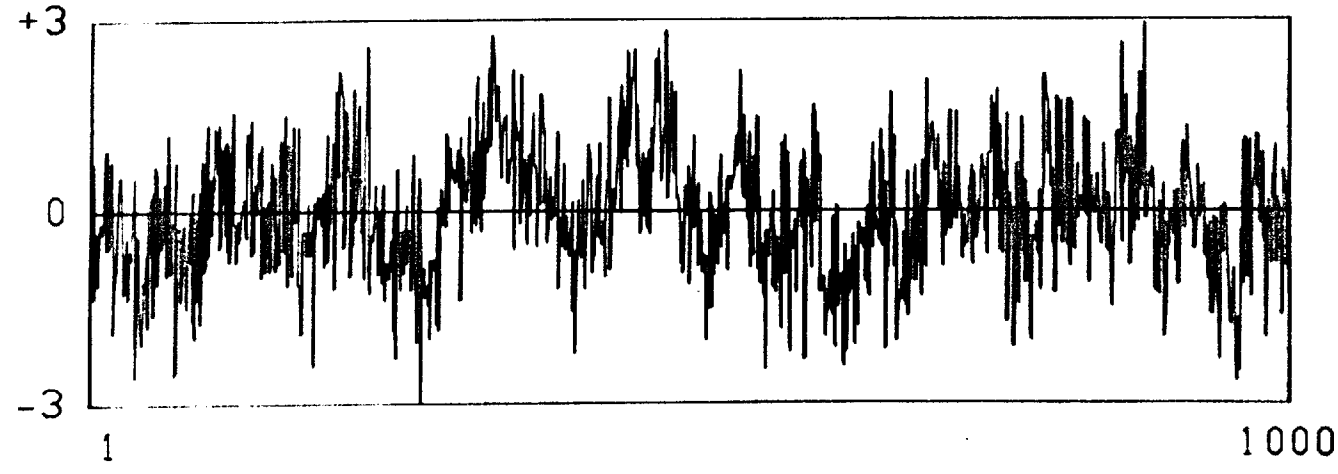
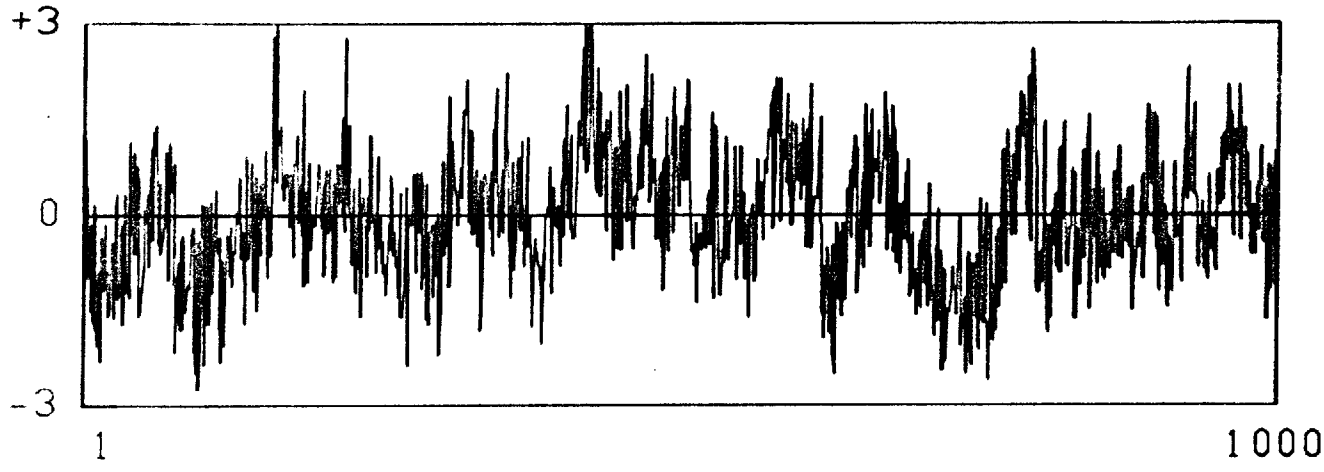


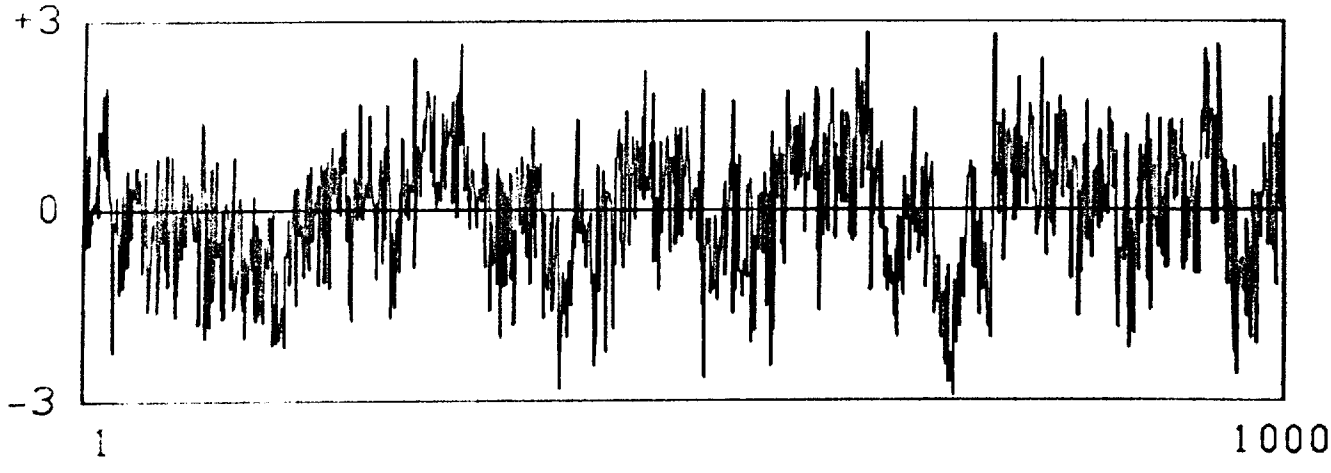
Figure (3.8) Plots of 1000 points of an ARIMA (1,0,1) process with $\varphi = 0.94$ and $\theta = 0.80, 0.78, 0.76$. Each sample of 1000 points has been standardized to have zero mean and unit variance.



ARIMA(1,0,1) PHI1 = 0.94 THETA1 = 0.74



ARIMA(1,0,1) PHI1 = 0.94 THETA1 = 0.72



ARIMA(1,0,1) PHI1 = 0.94 THETA1 = 0.70

Figure (3.9) Plots of 1000 points of an ARIMA (1,0,1) process with $\varphi = 0.94$ and $\theta = 0.74, 0.72, 0.70$. Each sample of 1000 points has been standardized to have zero mean and unit variance.

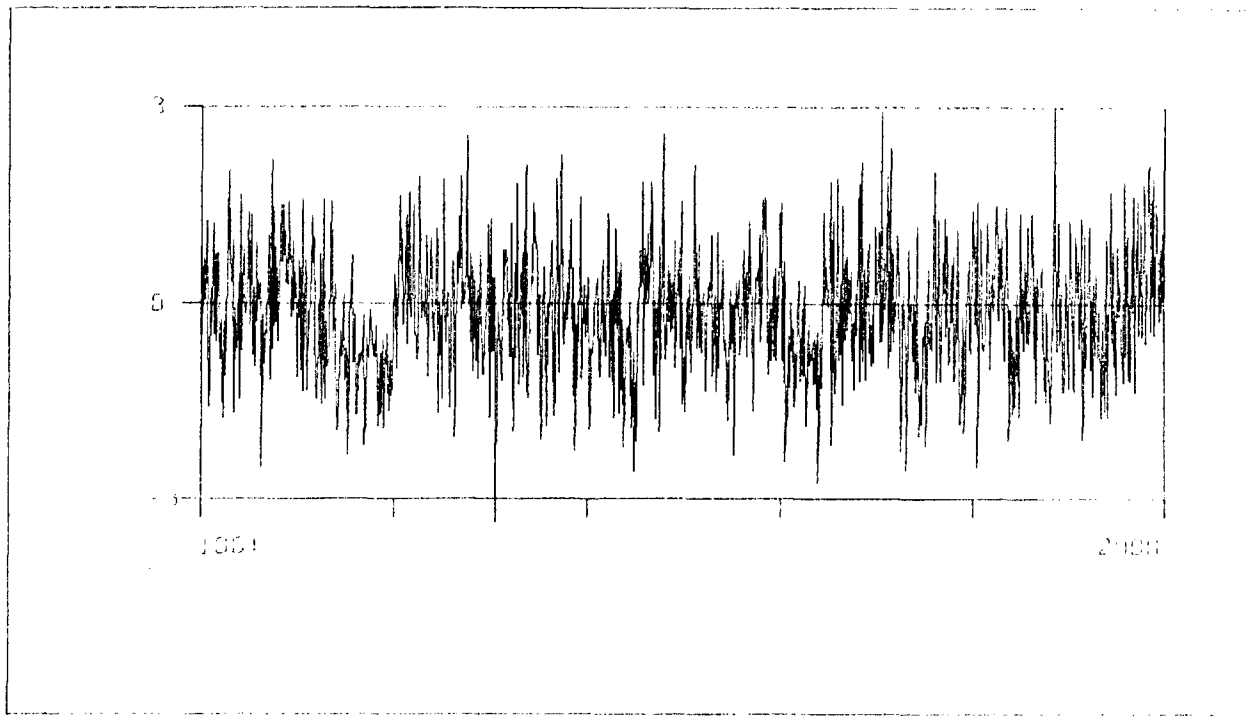


Figure (3.10) Plot of 1000 points from an ARIMA (1,0,1) process with $\varphi = 0.85$ and $\theta = 0.70$. The sample has been standardized to have zero mean and unit variance.

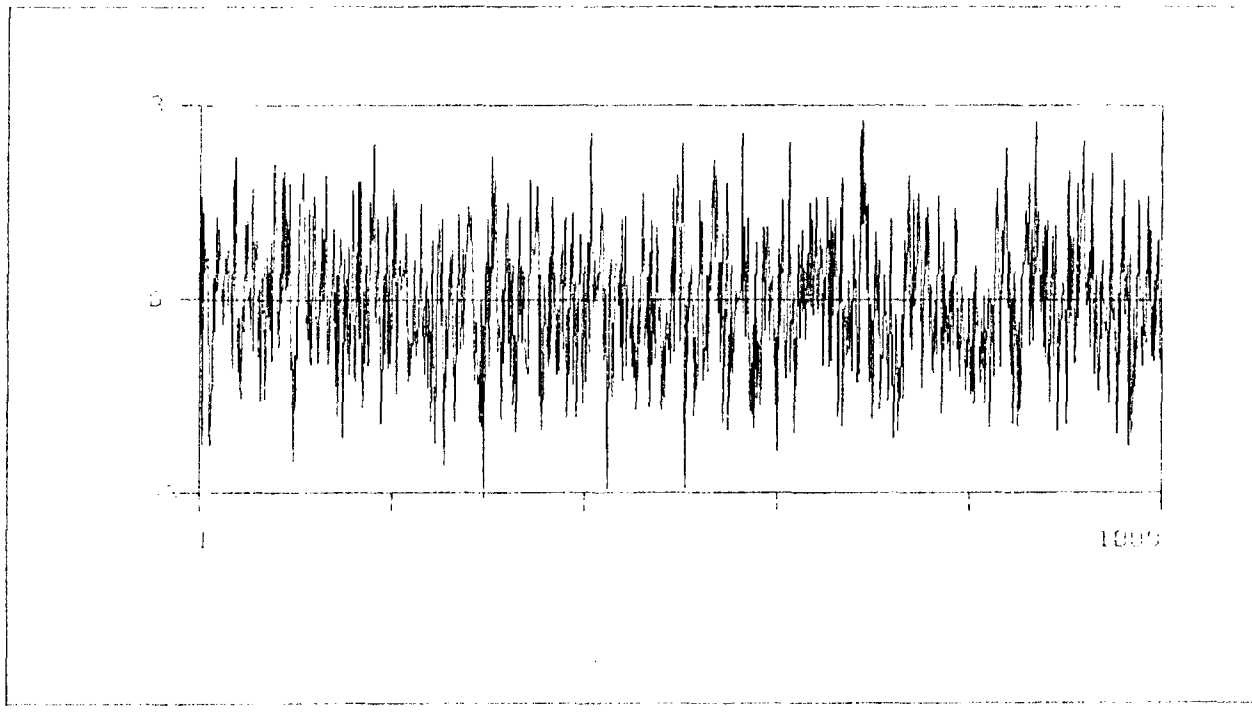


Figure (3.11) Plot of 1000 points from an ARIMA (1,0,1) process with $\varphi = 0.70$ and $\theta = 0.50$. The sample has been standardized to have zero mean and unit variance.

Provided that the sample size n is less than the memory of the process, sample estimates of h will have h as their population value. Thus, any process which approximates Hurst's law adequately within a finite span may also be considered to have an underlying (although explicitly unknown) population value of h where again $0.5 < h < 1$. In the subsequent analysis of the behaviour of h for the ARIMA (1,0,1) process, h is viewed in this latter sense.

Agreement with Hurst's law

Following Mandelbrot and Wallis (1969b), a simulation approach was adopted to examine the ability of the ARIMA (1,0,1) process to model Hurst's law accurately for values of n comparable with the longest geophysical records available. A detailed account of the simulation experiments conducted for this latter purpose has been presented by O'Connell (1971); only a brief description will be furnished here. A range of values of φ and θ, $\varphi > \theta$, yielding pertinent values of ρ_1 was selected, samples of length 9000 were generated for each combination of φ and θ and "pox diagrams" were constructed after the fashion of Mandelbrot and Wallis (1969b). The least squares procedure used by Wallis and Matalas (1970) for estimating the slope of a "pox diagram" was adopted, with the resulting estimate of h denoted by H. Over 250 "pox diagrams" corresponding to a number of combinations of φ and θ were constructed, and good overall agreement with Hurst's law was observed up to moderate to large values of n; figures (3.12)-(3.16) illustrate the agreement. Some curvature is however evident in the "pox diagrams" when φ is greater than about 0.95. From the definition of the autocorrelation function of the process for $k \geq 2$ given by equation (3.25), φ obviously controls the "effective memory" of the process, which corresponds to the value of the lag, k, at which ρ_k becomes effectively zero. As φ decreases, the values of R/S in the

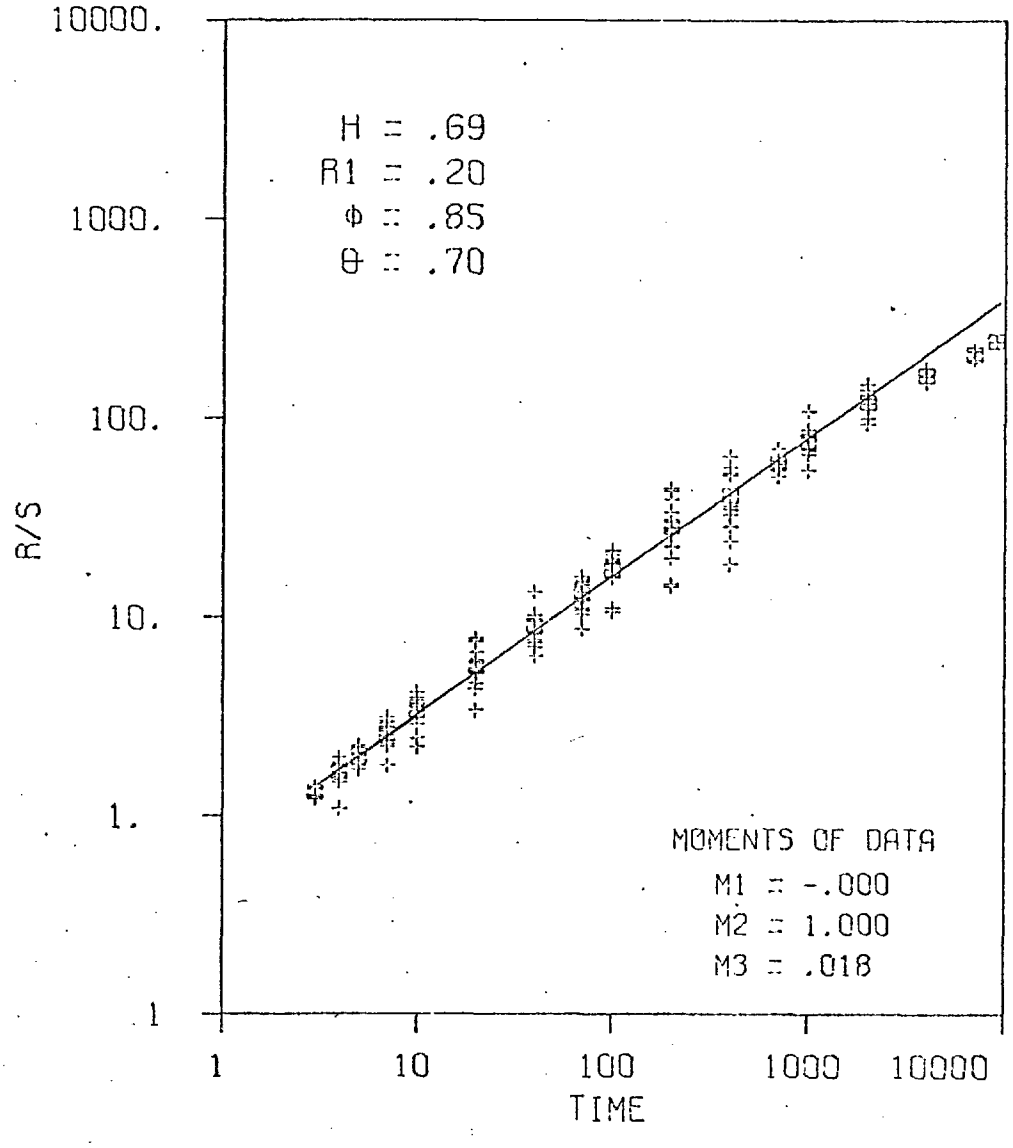


Figure (3.12) Pox diagram of $\ln(R_n/S_n)$ versus $\ln(n)$ (time) for a series of 9000 values from an ARIMA (1,0,1) process with $\varphi = 0.85$ and $\theta = 0.70$. The mean values of R_n/S_n have been plotted as little squares. Note that ρ_1 is designated as R1.

The selected values of subsample lengths were 3,4,5,7,10,20,40, 70,100,200,400,700,1000,2000,4000,7000,9000. For every value of $n_s < 500$, the subsample starting time was made equal to 1, 100, 200,..., 1400. For every value of $n_s > 500$, the subsample starting time was made equal to 100, 2000,..., 8000 or $(9000 - n_s + 1)$ whichever was the smaller (as used by Mandelbrot and Wallis 1969b).

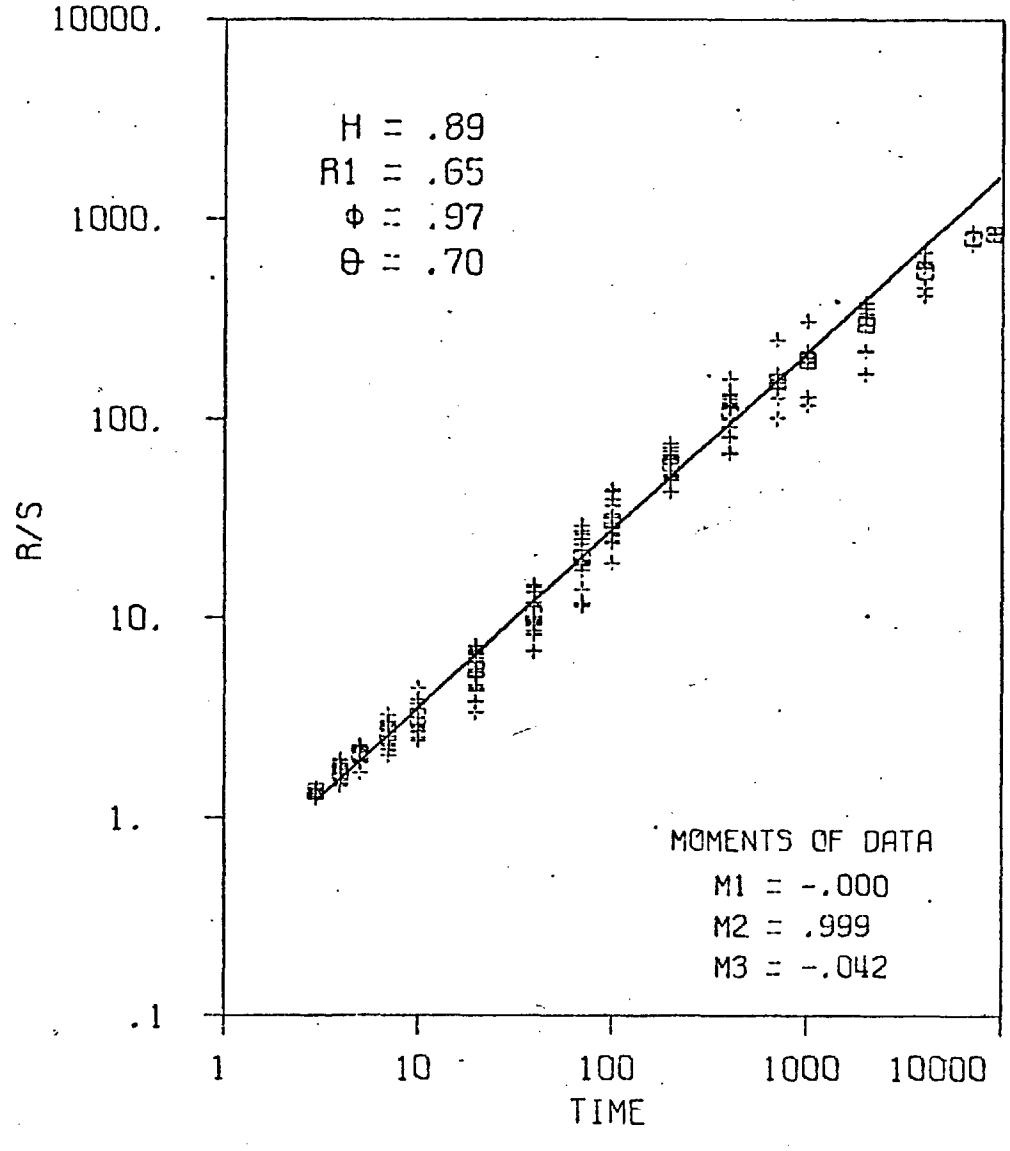


Figure (3.13) Pox diagram of $\ln(R_n/S_n$ versus ln(n) (time) for a series of 9000 values from an ARIMA (1,0,1) process with φ = 0.97 and θ = 0.70. Details are as for figure (3.12).

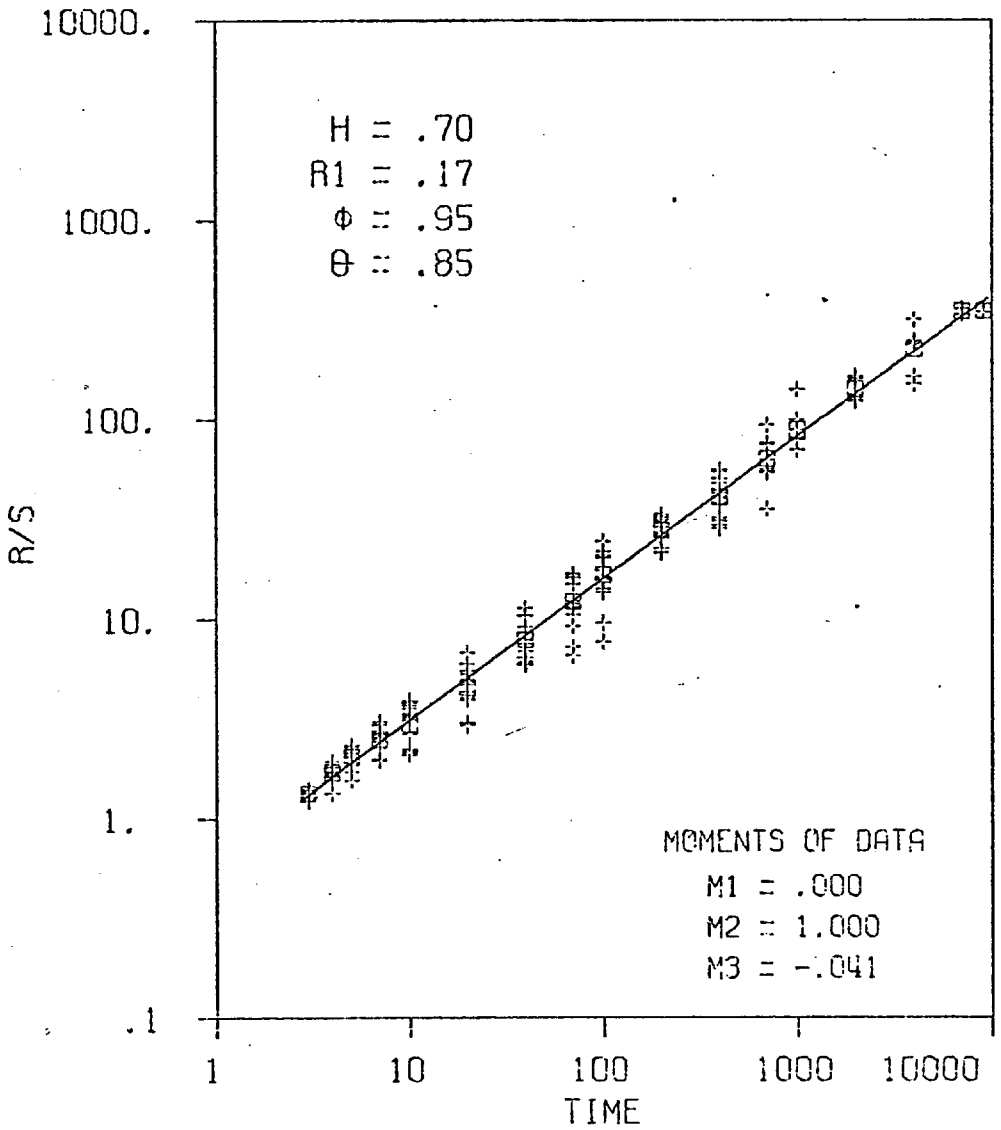


Figure (3.14) Pox diagram of $\ln(R_n/S_n)$ versus $\ln(n)$ (time) for a series of 9000 values from an ARIMA (1,0,1) process with $\varphi = 0.95$ and $\theta = 0.85$. Details are as for figure (3.12).

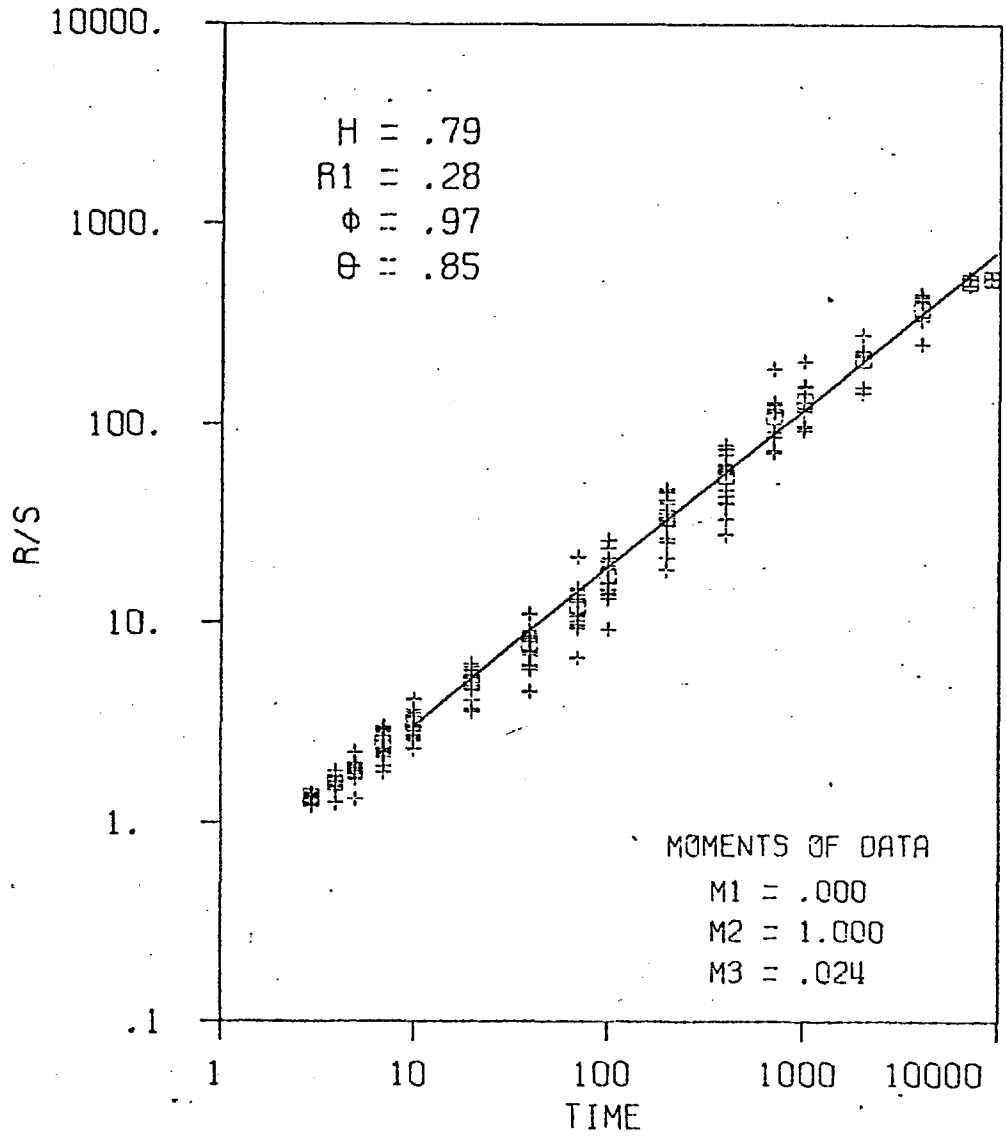


Figure (3.15) Pox diagram of $\ln(R_n/S_n)$ versus $\ln(n)$ (time) for a series of 9000 values from an ARIMA (1,0,1) process with $\varphi = 0.97$ and $\theta = 0.85$. Details are as for figure (3.12).

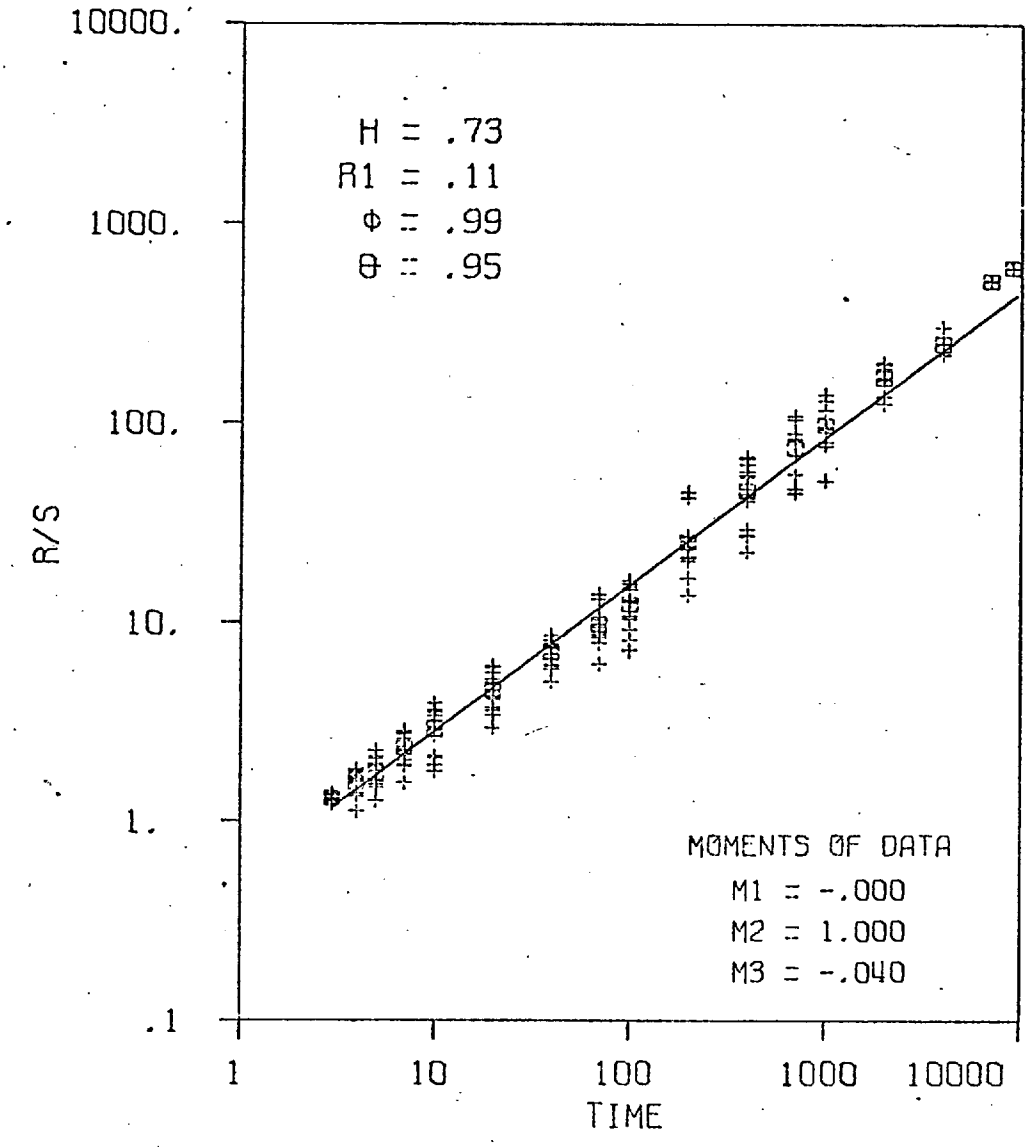


Figure (3.16) Pox diagram of $\ln(R_n/S_n)$ versus $\ln(n)$ (time) for a series of 9000 values from an ARIMA (1,0,1) process with $\varphi = 0.99$ and $\theta = 0.95$. Details are as for figure (3.12).

"pox diagrams" tend to "tail off" at smaller values of n. The value n at which a deviation from Hurst's law occurs is denoted by n_m and is intimately related to the "effective memory"; for smaller values of H, n_m, in general, tends to exceed the "effective memory"; however, for larger values of H, n_m approaches the "effective memory"; a similar result has been observed by Mandelbrot and Wallis (1969b) for the type 1 approximation to dfGn.

While it must be realized that h equals 0.5 ultimately for the ARIMA (1,0,1) process, figures (3.12) - (3.16) illustrate that for $n < n_m$, $H > 0.5$, where n_m may under certain conditions be as large as 10,000. Consequently, the ARIMA (1,0,1) process offers an explanation of Hurst's findings in terms of long-term persistence typified by values of φ approaching unity. The question as to whether or not h ultimately assumes a value of 0.5 in nature must await the availability of longer geophysical records.

The foregoing explanation of Hurst's findings has been in terms of H, which is an estimator of h. Mandelbrot and Wallis (1969d) have verified Hurst's findings using H rather than K which Hurst himself used as an estimator. However, even for very large samples Wallis and Matalas (1970) have shown that H is a biased estimator of h for a normal independent process, a Markov process and a type 1 approximation to dfGn. Based on the hypothesis that the generating mechanism in question is that of the real world, all estimates of h obtained from historic sequences must be treated as biased and consequently any explanation of Hurst's findings must be offered in terms of biased estimates of h.

In an attempt to define h for the ARIMA (1,0,1) process for $n \leq n_m$ an estimate, $\bar{H}$, was derived as the mean of ten values of H, each defined as the slope of a "pox diagram" for $n < n_m$, n_m=1000; such a

definition combats to some extent the large sampling variability exhibited by H, even for very large samples (Wallis and Matalas,1970). The results of these simulations are presented in table (3.1). There is evidence of some residual bias in $\bar{H}$ even for such large samples; for $\varphi = \theta$, $h = 0.5$ while $\bar{H} > 0.5$. For a particular choice of φ and θ, Wallis (1971) illustrated the nature of the variation of the estimated expected value of H with n; from figure (3.17), $\tilde{E}[H]_n$ obviously changes with n. For smaller values of n, the variation in $\tilde{E}[H]_n$ may be attributed to bias inherent in the estimator H; while for larger n a gradual approach to $h = 0.5$ is observed. This approach to $h = 0.5$ apparently commences at smaller values of n than individual "pox diagrams" suggest. A type 1 approximation to dfGn with $h \approx 0.8$ and $M = 500$ would produce a similar plot, as H is biased downwards when $h > 0.7$. However, if the memory parameter M is extended towards infinity then H tends towards its asymptotic value of h as $n \rightarrow M$. Where design horizons of the order of 50 - 100 years are under consideration, then it should be sufficient to use approximations to dfGn with memories of the order of a few hundred years for generating synthetic sequences. For this purpose the ARIMA (1,0,1) process would appear to be quite sufficient. Extensive inspection of "pox diagrams" suggests that the best agreement with Hurst's law is achieved with values of φ in the range 0.80 - 0.95, with the value of n_m generally being greater than 500. For any particular value of φ, the value of n_m decreases as θ decreases, and the approximation to dfGn becomes poorer.

3.3.5 Small Sample Properties of Estimates of σ^2, ρ_1 and h

For the purpose of generating synthetic sequences which reproduce observed long run effects, an estimate of h must be provided from an

φ		.99	.98	.97	.95	.90	.85	.80
θ = .95	ρ_1	.111	.051	.026	.000	-.038	-.076	-.094
	$\bar{H}$	.737	.676	.634	.576	.490	.439	.403
θ = .90	ρ_1	.350	.205	.139	.073	.000	-.042	-.076
	$\bar{H}$	.839	.781	.737	.673	.576	.515	.472
θ = .85	ρ_1	.562	.384	.287	.179	.061	.000	-.044
	$\bar{H}$	.892	.840	.800	.738	.639	.576	.529
θ = .80	ρ_1	.706	.540	.433	.300	.140	.057	.000
	$\bar{H}$	.920	.875	.839	.782	.687	.623	.576
θ = .75	ρ_1	.797	.659	.558	.418	.229	.126	.055
	$\bar{H}$	.937	.897	.864	.812	.723	.661	.614
θ = .70	ρ_1	.856	.745	.657	.523	.322	.203	.119
	$\bar{H}$	.947	.911	.881	.833	.750	.690	.645
θ = .65	ρ_1	.895	.807	.732	.612	.411	.282	.188
	$\bar{H}$	.953	.920	.892	.848	.770	.714	.670
θ = .60	ρ_1	.921	.851	.789	.684	.493	.360	.260
	$\bar{H}$	.956	.926	.900	.858	.785	.732	.691
θ = .55	ρ_1	.939	.883	.832	.742	.566	.435	.331
	$\bar{H}$	.959	.929	.904	.866	.797	.747	.707
θ = .50	ρ_1	.952	.907	.864	.788	.629	.503	.400
	$\bar{H}$	.960	.932	.908	.871	.806	.759	.721

Table (3.1) : Values of $\bar{H}$ and ρ_1 for a selection of values of φ and θ Each value of $\bar{H}$ is the mean of ten values of H, each derived as the slope of a least squares line fitted to the mean values of R_n/S_n for values of $n_s \leqslant 1000$ given in figure (3.12). The total sample size in all cases was 9000.

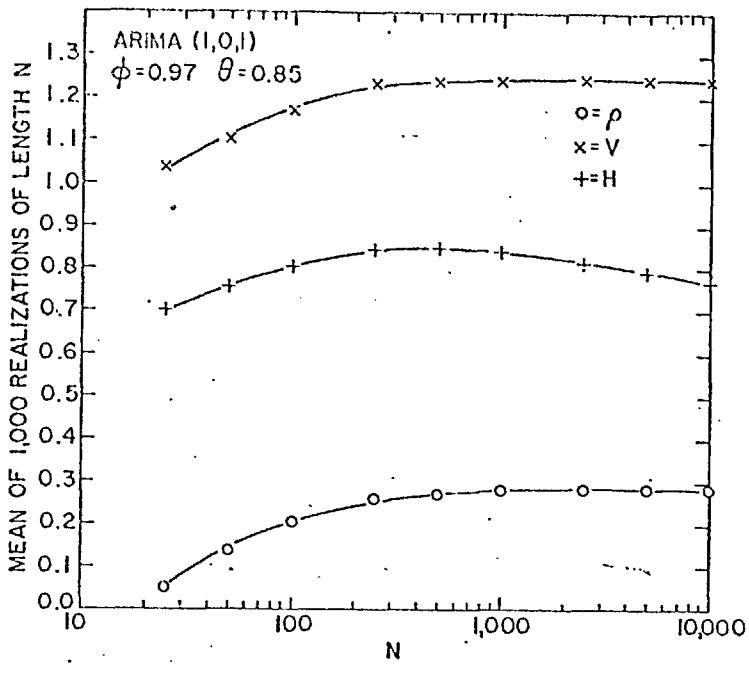


Figure (3.17) Values of $\tilde{E}[\hat{\rho}_1]_n$, $\tilde{E}[s^2]_n$ and $\tilde{E}[H]_n$ as a function of n for an ARIMA (1,0,1) process with $\varphi = 0.97$ and $\theta = 0.85$. Note that S^2 is designated as V, and n as N. Each point is based on 1000 realizations (after Wallis, 1971).

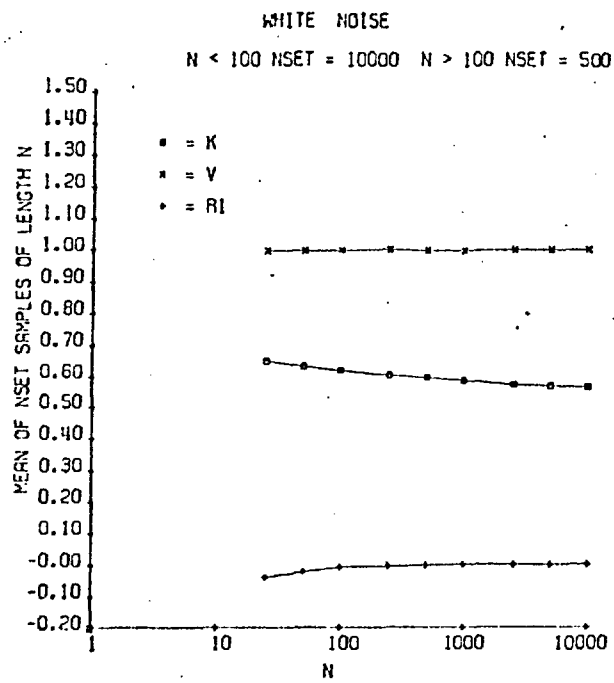


Figure (3.18) Values of $\tilde{E}[K]_n$, $\tilde{E}[s^2]_n$ and $\tilde{E}[\hat{\rho}_1]_n$ as a function of n for a white noise process. Note that S^2 is designated as V, ρ_1 as R1 and n as N. Each point is based on NSET realizations, where NSET = 10000 for $n \leq 100$ and NSET = 500 for $n > 100$.

historic sequence of length n. However, both K and H are known to be biased estimators and consequently, any historic record estimate of h requires a bias correction based on the assumption of a particular generating mechanism. In the case of approximations to dfGn, the analytic derivation of small sample bias corrections for estimates of h would appear to be highly unlikely. The biases in estimates of h may be quantified through Monte Carlo simulation only if h itself is known; however, simulation experiments with accurate approximations to dfGn involve a prohibitive investment in computer time. In addition, the application of bias corrections sometimes implies a loss of efficiency in an estimate.

The problem of bias corrections may be circumvented through defining $E[K]_n$ or $E[H]_n$, the expected value of K or H in samples of size n for a process. Preservation of an estimate of h within synthetic sequences then involves matching the observed estimate of h, either H or K, with $E[H]_n$ or $E[K]_n$ for the process. Invariably, Monte Carlo simulation will be necessary to define $E[K]_n$ or $E[H]_n$ for a process, which is feasible in the case of the ARIMA (1,0,1) process because of its simplicity. The flexibility of the ARIMA (1,0,1) process also allows the preservation of estimates of ρ_1 so that both low and high frequency properties of observed sequences may be modelled.

In defining the small sample properties of estimates of h for a process, a choice must be made between H and K. In general, H exhibits smaller bias but larger variance than K, so a trade-off between bias and variance is implied, and, in the absence of a loss function, no clear recommendation can be made. Mandelbrot and Wallis (1969d) have pointed out the deficiencies in K; however, H also suffers from a major deficiency in that no universal rule exists for defining the set of sub-series to be used, or for estimating the slope of the

'pox diagram'. However, K does not suffer from such a deficiency and is also quicker to compute. Accordingly, K was used in the simulation experiments to define $E[K]_n$ for the ARIMA (1,0,1) process.

The simulation experiments for defining $E[K]_n$ were extended to large values of n to help in assessing the quality of the approximation to dfGn afforded by the ARIMA (1,0,1) process for various values of φ and θ. Samples of size n = 25, 50, 100, 250, 500, 1000, 2500, 5000, 10000 were generated and estimates of the variance σ^2 and the lag-one autocorrelation ρ_1 in addition to K were derived from each sequence. The variance σ^2 was estimated as

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2 \tag{3.26}$$

but has since been shown to satisfy

$$E[S^2]_n = \sigma^2 \left[1 - \frac{2\rho_1}{n(n-1)} \left[\frac{n(1-\varphi) - (1-\varphi^n)}{(1-\varphi)^2}\right]\right] \tag{3.27}$$

where ρ_1 is defined by equation (3.24). The lag one autocorrelation ρ_1 was estimated as

$$\hat{\rho}_1 = \frac{\left[\sum_{i=1}^{n-1} X_i - \frac{1}{n-1}\sum_{i=1}^{n-1} X_i\right]\left[\sum_{i=1}^{n-1} X_{i+1} - \frac{1}{n-1}\sum_{i=1}^{n-1} X_{i+1}\right]}{\sqrt{\left[\sum_{i=1}^{n-1} (X_i - \frac{1}{n-1}\sum_{i=1}^{n-1} X_i)^2\right]\left[\sum_{i=1}^{n-1} (X_{i+1} - \frac{1}{n-1}\sum_{i=1}^{n-1} X_{i+1})^2\right]}} \tag{3.28}$$

Estimates of $E[K]_n$, $E[\hat{\rho}_1]_n$ and $E[S^2]$, denoted by $\tilde{E}[K]_n$, $\tilde{E}[\hat{\rho}_1]_n$ and $\tilde{E}[S^2]_n$, were then derived by averaging the respective statistics over the total number of realizations.

For $n \leqslant 100$, 10000 samples of size n were generated; stability criteria necessitated such a large number of realizations which emphasizes the tremendous variability of statistics derived from small samples in the presence of low frequency effects. Even with 10000 realizations,

slightly unstable estimates were sometimes noted. For n > 100, extremely stable estimates were not required, and 100 realizations were generated to provide an overall assessment of the quality of the approximation to dfGn.

The particular case of $\varphi = \theta$ or white noise was again used as a starting point, and $\tilde{E}[K]_n$, $\tilde{E}[\hat{\rho}_1]_n$ and $\tilde{E}[S^2]_n$ are presented in graphical form in figure (3.18). The bias in $\tilde{E}[K]_n$ is evident, while $\tilde{E}[\hat{\rho}_1]_n$ is an approximation of

$$E[\hat{\rho}_1]_n = -\frac{1}{(n-1)} \qquad (3.29)$$

which was derived by Anderson (1942). $\tilde{E}[S^2]_n$ is unbiased. The results of the experiments for φ=0.98 and a series of values of θ are presented in figures (3.19 - 3.21). Figure (3.19) corresponds to φ=0.98 and θ=0.94, which is quite close to white noise; nevertheless, comparison with figure (3.18) shows that a distinct difference is observed in the behaviour of K for both processes. For φ=0.98 and θ=0.94, ρ_1=0.075; hence, even with minimal lag-one autocorrelation, the effects of long-term persistence are decidedly evident. A lag-one Markov process with $\varphi=\rho_1$=0.075 would yield results indistinguishable from figure (3.18) because of the rapid decay of the autocorrelation function. It should be noted that in figure (3.19) the values of K do not tend to 0.5; this observation is in line with the conclusion in the preceding section that for smaller values of h, the value of n at which a tailing off occurs is considerably in excess of n_m, the value of the lag at which the autocorrelation ρ_k becomes effectively zero. Allowing for slight instability in some of the estimated expected values, slight biases downwards are noticeable in $\tilde{E}[S^2]_n$ and in $\tilde{E}[\hat{\rho}_1]_n$ for small n, while $\tilde{E}[K]_n$ is relatively constant with n. However, because K is biased, this does not necessarily mean that h is constant with n. Nevertheless, K is known to be unbiased in the neighbourhood of 0.7, so the approximation to dfGn is probably

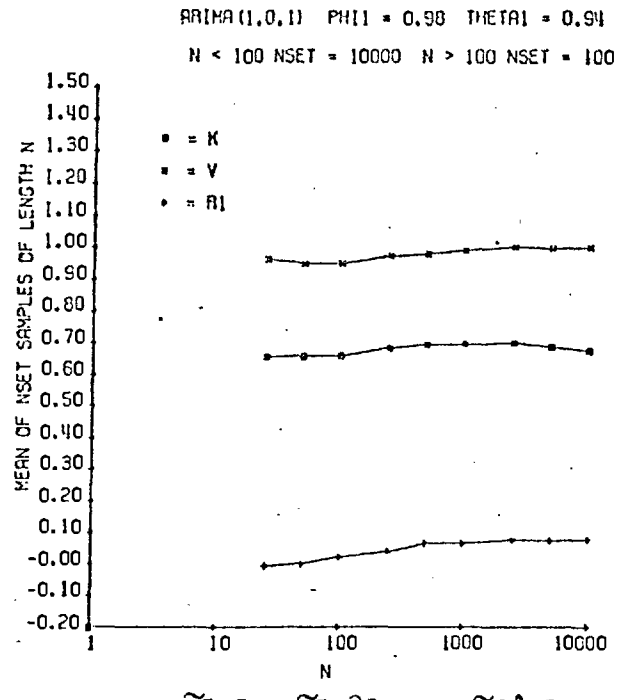


Figure (3.19) Values of $\widetilde{E}[K]_n$, $\widetilde{E}[S^2]_n$ and $\widetilde{E}[\hat{\rho}_1]_n$ as a function of n for an ARIMA (1,0,1) process with $\varphi = 0.98$ and $\theta = 0.94$. Note that S^2 is designated as V, ρ_1 as R1 and n as N. Each point is based on NSET realizations, where NSET = 10000 for $n \leq 100$ and NSET = 100 for $n > 100$.

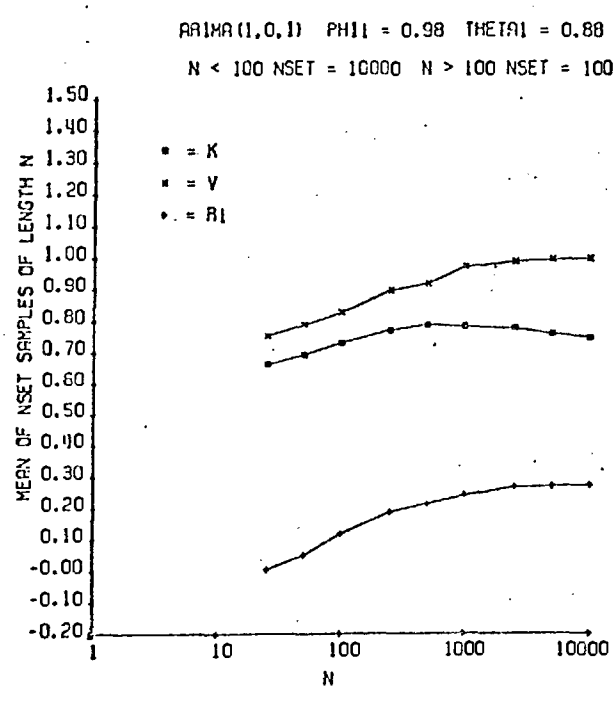


Figure (3.20) Values of $\widetilde{E}[K]_n$, $\widetilde{E}[S^2]_n$ and $\widetilde{E}[\hat{\rho}_1]_n$ as a function of n for an ARIMA (1,0,1) process with $\varphi = 0.98$ and $\theta = 0.88$. Note that S^2 is designated as V, ρ_1 as R1 and n as N. Each point is based on NSET realizations, where NSET = 10000 for $n \leq 100$ and NSET = 100 for $n > 100$.

quite reasonable. The results for $\varphi = 0.98$ and $\theta = 0.88$ (figure (3.20)) illustrate that higher values of $\tilde{E}[K]_n$ are now obtained, and that the biases in S^2, $\hat{\rho}_1$ and K are more severe which is a direct consequence of the increase in low frequencies. Because of the difficulty of assessing the bias in K, little can be inferred about the behaviour of h for the process. The biases in $\tilde{E}[\hat{\rho}_1]_n$ and $\tilde{E}[S^2]_n$ decrease rather slowly with n. A further decrease in θ to 0.76 yields a somewhat more extreme process with $\rho_1 = 0.630$; from figure (3.21) the biases in $\tilde{E}[\hat{\rho}_1]_n$ and $\tilde{E}[S^2]_n$ are extremely severe and have barely disappeared at n = 2500. Slightly higher values of K are observed than for $\theta = 0.88$; for n = 100 and $\theta = 0.76$, $\tilde{E}[K]_n = 0.807$ while for n = 100 and $\theta = 0.88$, $\tilde{E}[K]_n = 0.729$. A type 1 approximation to dfGn with h = 0.9 and M = 500 would probably produce a similar plot, although nothing definite can be said of the quality of the approximation to dfGn afforded by the ARIMA (1,0,1) process in this case. The experiments carried out in section (3.3.4) suggest the quality of the approximation diminishes when φ exceeds 0.95, particularly for large values of ρ_1.

The results of a further set of experiments for $\varphi = 0.92$ and $\theta = 0.82$, 0.76, 0.60 are presented in figures (3.22 - 3.24). Inspection of figure (3.22) corresponding to $\varphi = 0.92$ and $\theta = 0.82$ shows that $\tilde{E}[K]_n$ is relatively constant when $n \leq 500$, and is close to 0.7. As K is known to be unbiased in the neighbourhood of h = 0.7, the ARIMA (1,0,1) process provides a good approximation to dfGn in this case. Comparison of a process which has $\varphi = 0.92$ and $\theta = 0.76$ (figure (3.23)) with a process with $\varphi = 0.98$ and $\theta = 0.88$ (figure (3.20)) on the basis of approximately equal values of ρ_1 shows that as φ decreases, so too does $\tilde{E}[K]_n$, and the biases in $\tilde{E}[S^2]_n$ and $\tilde{E}[\hat{\rho}_1]_n$, with the latter statistics attaining their asymptotic values much more quickly. Figure (3.24) based on $\varphi = 0.92$ and $\theta = 0.60$ shows that as θ decreases, the turnover point in the plot of $\tilde{E}[K]_n$ against n tends to occur at smaller values of n. Nevertheless,

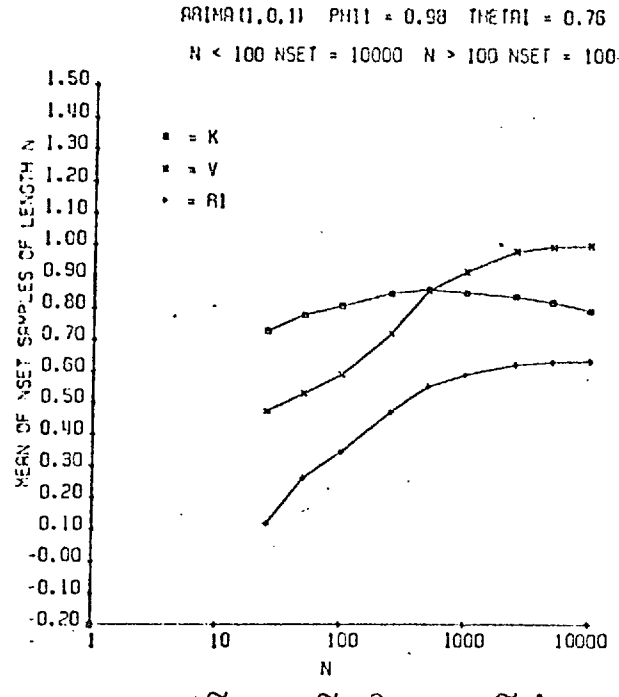


Figure (3.21) Values of $\tilde{E}[K]_n$, $\tilde{E}[S^2]_n$ and $\tilde{E}[\hat{\rho}_1]_n$ as a function of n for an ARIMA (1,0,1) process with $\varphi = 0.98$ and $\theta = 0.76$. Note that S^2 is designated as V, ρ_1 as R1 and n as N. Each point is based on NSET realizations, where NSET = 10000 for $n \leq 100$ and NSET = 100 for $n > 100$.

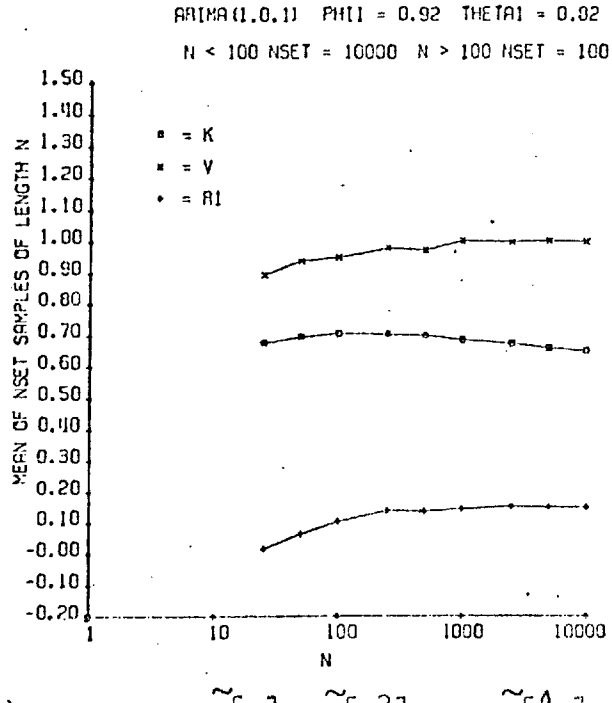


Figure (3.22) Values of $\tilde{E}[K]_n$, $\tilde{E}[S^2]_n$ and $\tilde{E}[\hat{\rho}_1]_n$ as a function of n for an ARIMA (1,0,1) process with $\varphi = 0.92$ and $\theta = 0.82$. Note that S^2 is designated as V, ρ_1 and R1 and n as N. Each point is based on NSET realizations, where NSET = 10000 for $n \leq 100$ and NSET = 100 for $n > 100$.

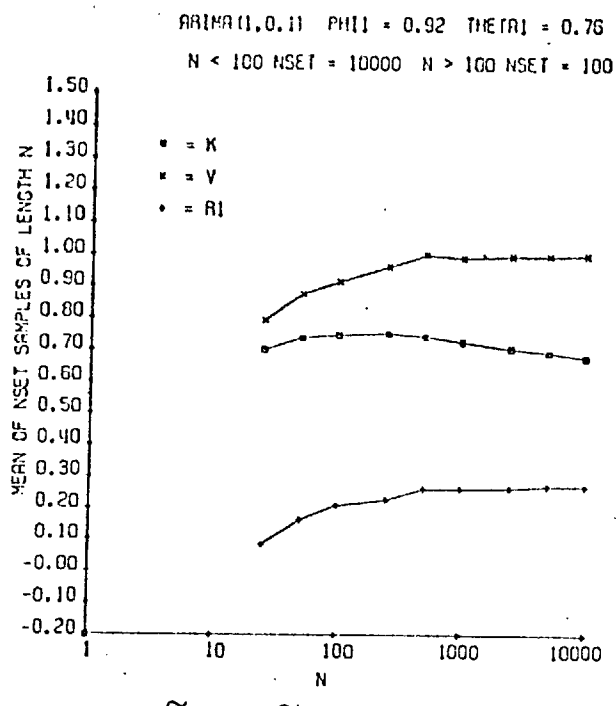


Figure (3.23) Values of $\tilde{E}[K]_n$, $\tilde{E}[S^2]_n$ and $\tilde{E}[\hat{\rho}_1]_n$ as a function of n for an ARIMA (1,0,1) process with $\varphi = 0.92$ and $\theta = 0.76$. Note that S^2 is designated as V, ρ_1 as R1 and n as N. Each point is based on NSET realizations, where NSET = 10000 for $n \leq 100$ and NSET = 100 for $n > 100$.

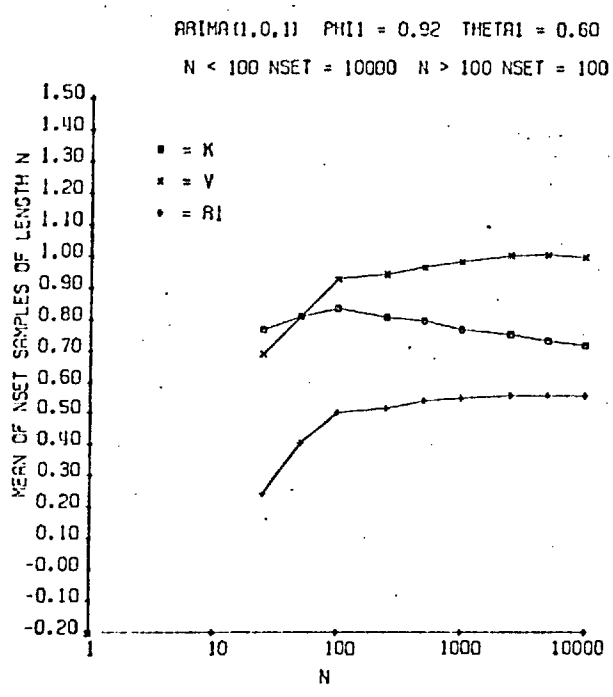


Figure (3.24) Values of $\tilde{E}[K]_n$, $\tilde{E}[S^2]_n$ and $\tilde{E}[\hat{\rho}_1]_n$ as a function of n for an ARIMA (1,0,1) process with $\varphi = 0.92$ and $\theta = 0.60$. Note that S^2 is designated as V, ρ_1 as R1 and n as N. Each point is based on NSET realizations, where NSET = 10000 for $n \leq 100$ and NSET = 100 for $n > 100$.

the process is still adequate for generating synthetic sequences of lengths used in design situations.

A further decrease in φ results in a smaller effective memory (figure 3.25), and the ARIMA (1,0,1) process starts to exhibit some of the characteristics of the lag-one Markov process, and $\tilde{E}[K]_n$ is now a monotonic decreasing function of n. A type 1 approximation to dfGn with a value of h of the order of 0.6 would probably produce a similar plot to figure (3.25) for values of n up to about 500. Nevertheless, comparison of figure (3.26) which corresponds to $\varphi = 0.82$ and $\theta = 0.58$ with a plot for a lag-one Markov process (figure (3.27)) on the basis of approximately equal values of ρ_1 illustrates that higher values of $\tilde{E}[K]_n$ are observed for the ARIMA (1,0,1) process, with the biases in $\tilde{E}[S^2]_n$ and $\tilde{E}[\hat{\rho}_1]_n$ disappearing more slowly, emphasising the presence of low frequencies.

For a selection of values of φ and θ, tables of $\tilde{E}[K]_n$ and $\tilde{E}[\hat{\rho}_1]_n$ have been abstracted for sample sizes n = 25, 50, 100, while $E[S^2]_n$ is given by equation (3.27). Within the parameter space covered, the long-term properties of the ARIMA (1,0,1) process should be close to those of a type 1 approximation to dfGn for the values of n considered. The values of $\tilde{E}[K]_n$ and $\tilde{E}[\hat{\rho}_1]_n$, each derived from 10000 samples of size n are presented in tables (3.2 - 3.7). Even for such a large number of realizations some slight instability is still noticeable in some of the estimates. For sample size 25, the range of K values which can be modelled for all the selected values of φ and θ is approximately 0.65 - 0.80 while for sample sizes 50 and 100, the corresponding ranges are 0.65 - 0.85 and 0.65 - 0.87, respectively. These somewhat restricted ranges are a consequence of the nature of the estimator K ; nevertheless, the majority of observed values of K and ρ_1 should fall within the bounds of the tables. In the following section the use of the tables is illustrated in formulating the ARIMA (1,0,1)

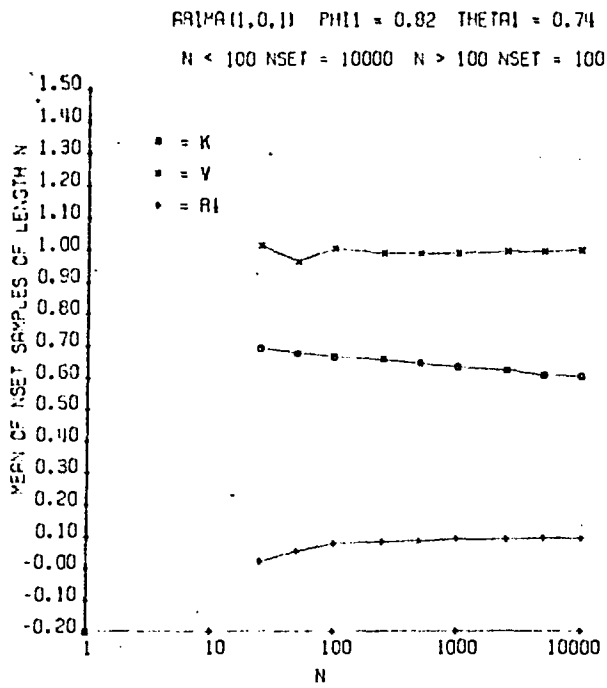


Figure (3.25) Values of $\tilde{E}[K]_n$, $\tilde{E}[S^2]_n$ and $\tilde{E}[\hat{\rho}_1]_n$ as a function of n for an ARIMA (1,0,1) process with $\varphi = 0.82$ and $\theta = 0.74$. Note that S^2 is designated as V, ρ_1 as R1 and n as N. Each point is based on NSET realizations, where NSET = 10000 for $n \leq 100$ and NSET = 100 for $n > 100$.

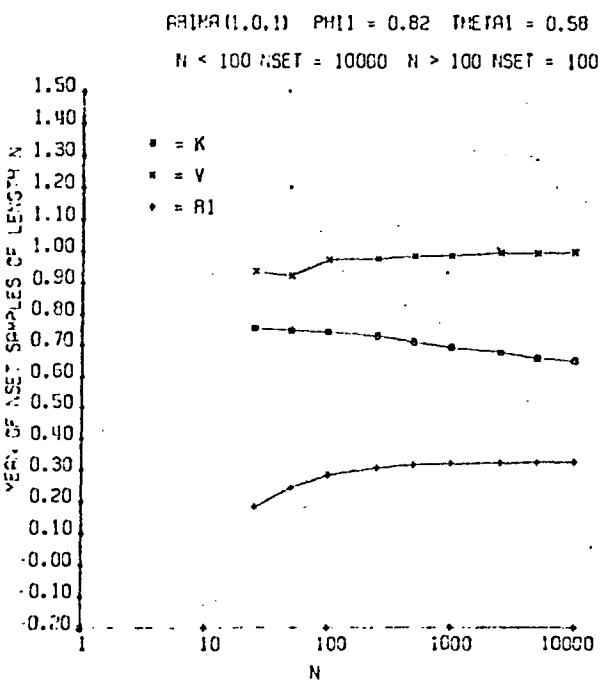


Figure (3.26) Values of $\tilde{E}[K]_n$, $\tilde{E}[S^2]_n$ and $\tilde{E}[\hat{\rho}_1]_n$ as a function of n for an ARIMA (1,0,1) process with $\varphi = 0.82$ and $\theta = 0.58$. Note that S^2 is designated as V, ρ_1 as R1 and n as N. Each point is based on NSET realizations, where NSET = 10000 for $n \leq 100$ and NSET = 100 for $n > 100$.

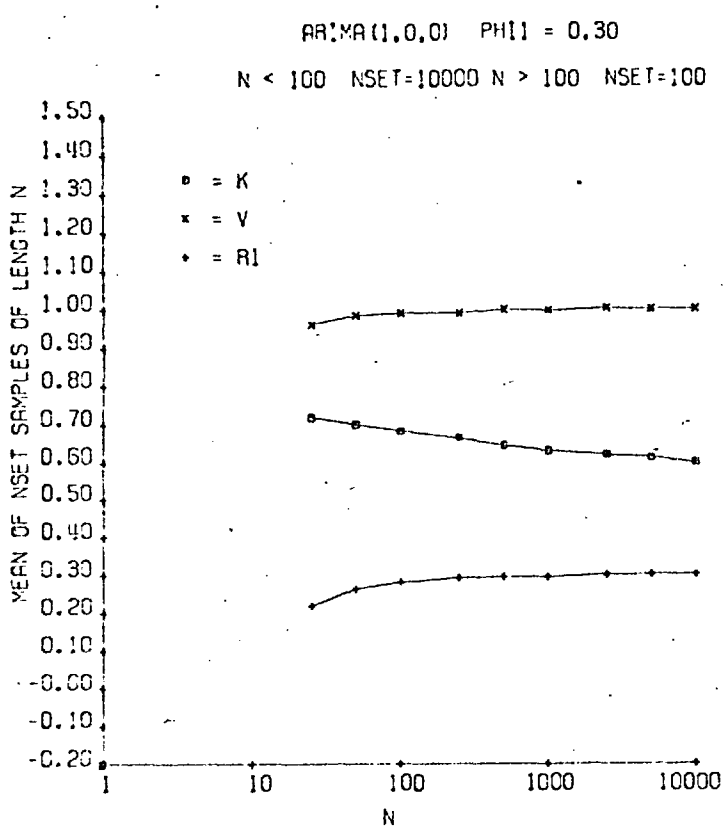


Figure (3.27) Values of $\tilde{E}[K]_n$, $\tilde{E}[S^2]_n$ and $\tilde{E}[\hat{\rho}_1]_n$ as a function of n for an ARIMA (1,0,0) or lag-one Markov process with $\varphi = \rho_1 = 0.30$. Note that S^2 is designated as V, ρ_1 as R1 and n as N. Each point is based on NSET realizations, where NSET = 10000 for $n \leq 100$ and NSET = 100 for $n > 100$.

$\varphi = 0.96$							
		$\tilde{E}[K]_n$			$\tilde{E}[\hat{\rho}_1]_n$		
θ	ρ_1 \ n	25	50	100	25	50	100
0.92	0.058	0.654	0.661	0.657	-0.005	0.005	0.025
0.88	0.146	0.683	0.679	0.698	-0.005	0.038	0.080
0.84	0.250	0.681	0.704	0.724	0.060	0.083	0.132
0.80	0.357	0.693	0.737	0.767	0.069	0.160	0.227
0.76	0.457	0.718	0.752	0.781	0.099	0.196	0.264
0.72	0.545	0.735	0.776	0.811	0.144	0.259	0.366
0.68	0.620	0.756	0.796	0.827	0.219	0.310	0.444
0.64	0.683	0.763	0.818	0.843	0.243	0.382	0.504
0.60	0.734	0.791	0.823	0.855	0.300	0.424	0.567
0.56	0.776	0.803	0.825	0.861	0.353	0.475	0.621
0.52	0.810	0.801	0.852	0.866	0.377	0.565	0.647

Table (3.2) Values of $E[K]_n$ and $E[\hat{\rho}_1]_n$ for selected values of φ and θ

$\varphi = 0.92$							
		$\tilde{E}[K]_n$			$\tilde{E}[\hat{\rho}_1]_n$		
θ	ρ_1 \ n	25	50	100	25	50	100
0.88	0.049	0.657	0.664	0.654	0.001	0.012	0.028
0.84	0.114	0.687	0.680	0.686	0.005	0.046	0.079
0.80	0.189	0.686	0.705	0.709	0.072	0.093	0.123
0.76	0.269	0.699	0.735	0.745	0.082	0.160	0.208
0.72	0.349	0.725	0.751	0.756	0.116	0.199	0.240
0.68	0.426	0.740	0.782	0.783	0.169	0.269	0.332
0.64	0.496	0.772	0.783	0.800	0.218	0.309	0.390
0.60	0.560	0.773	0.796	0.803	0.273	0.364	0.437
0.56	0.616	0.774	0.820	0.825	0.285	0.432	0.516
0.52	0.665	0.794	0.825	0.828	0.335	0.467	0.532

Table (3.3) Values of $\tilde{E}[K]_n$ and $\tilde{E}[\hat{\rho}_1]_n$ for selected values of φ and θ

$\varphi = 0.88$							
		$\tilde{E}[K]_n$			$\tilde{E}[\hat{\rho}_1]_n$		
θ	ρ_1 \ n	25	50	100	25	50	100
0.84	0.046	0.659	0.665	0.651	0.005	0.014	0.029
0.80	0.102	0.689	0.678	0.676	0.012	0.050	0.078
0.76	0.176	0.688	0.703	0.696	0.080	0.097	0.117
0.72	0.233	0.703	0.728	0.728	0.091	0.158	0.196
0.68	0.302	0.729	0.745	0.737	0.126	0.198	0.225
0.64	0.370	0.742	0.772	0.764	0.176	0.262	0.309
0.60	0.435	0.771	0.773	0.778	0.223	0.302	0.363
0.56	0.495	0.771	0.788	0.782	0.278	0.356	0.407
0.52	0.550	0.774	0.808	0.802	0.291	0.416	0.481

Table (3.4) Values of $\tilde{E}[K]_n$ and $\tilde{E}[\hat{\rho}_1]_n$ for selected values of φ and θ

$\varphi = 0.84$							
		$\tilde{E}[K]_n$			$\tilde{E}[\hat{\rho}_1]_n$		
θ	ρ_1 \ n	25	50	100	25	50	100
0.80	0.044	0.660	0.663	0.647	0.007	0.016	0.029
0.76	0.096	0.690	0.676	0.668	0.016	0.052	0.077
0.72	0.154	0.688	0.699	0.687	0.086	0.099	0.113
0.68	0.214	0.705	0.720	0.714	0.098	0.156	0.189
0.64	0.277	0.725	0.746	0.733	0.142	0.203	0.237
0.60	0.325	0.753	0.749	0.747	0.180	0.245	0.290
0.56	0.409	0.753	0.766	0.753	0.240	0.298	0.332
0.52	0.454	0.761	0.784	0.773	0.254	0.355	0.406

Table (3.5) Values of $\tilde{E}[K]_n$ and $\tilde{E}[\hat{\rho}_1]_n$ for selected values of φ and θ

$\varphi = 0.80$							
		$\tilde{E}[K]_n$			$\tilde{E}[\hat{\rho}_1]_n$		
θ	ρ_1 \ n	25	50	100	25	50	100
0.75	0.055	0.666	0.657	0.649	-0.006	0.022	0.041
0.70	0.119	0.685	0.694	0.678	0.056	0.079	0.098
0.65	0.188	0.708	0.698	0.708	0.114	0.103	0.166
0.60	0.260	0.736	0.727	0.719	0.139	0.192	0.227
0.55	0.331	0.737	0.750	0.741	0.175	0.247	0.289
0.50	0.400	0.756	0.764	0.746	0.263	0.315	0.343

Table (3.6) Values of $\tilde{E}[K]_n$ and $\tilde{E}[\hat{\rho}_1]_n$ for selected values of φ and θ

$\varphi = 0.75$							
		$\tilde{E}[K]_n$			$\tilde{E}[\hat{\rho}_1]_n$		
θ	ρ_1 \ n	25	50	100	25	50	100
0.70	0.054	0.666	0.655	0.645	-0.004	0.023	0.041
0.65	0.115	0.684	0.689	0.671	0.059	0.079	0.096
0.60	0.179	0.707	0.691	0.698	0.117	0.102	0.161
0.55	0.246	0.733	0.719	0.707	0.141	0.189	0.219
0.50	0.313	0.734	0.740	0.728	0.177	0.240	0.277
0.45	0.377	0.751	0.754	0.732	0.264	0.307	0.329

Table (3.7) Values of $\tilde{E}[K]_n$ and $\tilde{E}[\hat{\rho}_1]_n$ for selected values of φ and θ

process to maintain the required statistical resemblance between historic and synthetic sequences.

3.3.6 Formulation for Synthetic Hydrology

For the purpose of generating synthetic sequences, the ARIMA (1,0,1) process may be formulated as

$$X_t - \mu_x = \varphi(X_{t-1} - \mu_x) + \sigma_x \sigma_\varepsilon(\varepsilon_t - \theta\ \varepsilon_{t-1}) \qquad (3.29)$$

where μ_x and σ_x are the mean and standard deviation of the process, respectively. The term ε_t is an independent random variable with zero mean and unit variance, and σ_ε is defined by

$$\sigma_\varepsilon^2 = \frac{(1 - \varphi^2)}{(1+\theta^2-2\varphi\theta)} \qquad (3.30$$

which ensures that X_t will have variance σ_x^2.

If the process X_t is required to have a skewness γ_x then this may be achieved through replacing the term ε_t in equation (3.29) by an independent random variable η_t which may be defined by the Wilson-Hilferty transformation as follows :

$$\eta_t = \frac{2}{\gamma_\eta}\left[1 + \frac{\gamma_\eta\,\varepsilon_t}{6} - \frac{\gamma_\eta^2}{36}\right]^3 - \frac{2}{\gamma_\eta} \qquad (3.31)$$

where ε_t is NIP(0,1). The random component η_t will have zero mean, variance unity and skewness γ_η . However, for values of γ_η in excess of about 3.0, MacMahon and Miller (1971) have found that the transformation given by equation (3.31) is inadequate. A modified transformation has been proposed by Kirby (1972) which overcomes the problem. The skewness of the process X_t, γ_x, is related to the skewness of the term η_t as follows :

$$\gamma_x = \left[\left[\frac{(1 - \theta^3 + 3\varphi\theta^2 - 3\varphi^2\theta)}{(1 - \varphi^3)}\right] \Big/ \left[\frac{(1 + \theta^2 - 2\varphi\theta)}{(1 - \varphi^2)}\right]^{3/2}\right]\gamma_\eta \qquad (3.32)$$

Estimates of μ_x, σ_x, γ_x, ρ_x and h can be derived from a historic sequence of length n ; the question then arises as to which type of resemblance is required between synthetic and historic sequences. As outlined in section (1.2.4), two types of resemblance are possible, Type A and Type B. Both types of resemblance will now be considered.

(i) Resemblance of type A

In this case, resemblance is to be maintained between the estimates $\hat{\mu}_x$, $\hat{\sigma}_x^2$, $\hat{\rho}_x$, $\hat{\gamma}_x$ and $\hat{h}$ derived from a historic sequence of length n, and the corresponding estimates which would be obtained from a synthetic sequence of infinite length, i.e. $\hat{\mu}_x$, $\hat{\sigma}_x^2$, $\hat{\gamma}_x$, $\hat{\rho}_x$ and $\hat{h}$ assume the role of population quantities. However, the population value of h for the ARIMA (1,0,1) process is 0.5 so that resemblance of type A cannot be maintained in terms of h ; this is true of any finite memory approximation to fGn. However, resemblance in terms of an estimate of h can be maintained within finite synthetic sequences. The overall procedure is as follows :

(a) Derive estimates $\hat{\mu}_x$, $\hat{\sigma}_x^2$, $\hat{\rho}_x$ and K from a historic sequence of length n.

(b) From tables (3.2) - (3.7), identify values of φ and θ such that

$$\tilde{E}[K]_n \approx K$$

and

$$\rho_1 \approx \hat{\rho}_x$$

where ρ_1 is defined by equation (3.24). Some interpolation will invariably be involved here.

(c) The estimated skewness $\hat{\gamma}_x$ is then used to define an estimate of γ_η using equation (3.32). The term η_t can then be generated using equation (3.31) or the modified transformation suggested by Kirby (1972).

(d) The quantities φ, θ, $\hat{\mu}_x$ and $\hat{\sigma}_x$ are then incorporated into equation (3.29), whence synthetic flows may be generated.

(ii) Resemblance of type B

In this case, resemblance is maintained entirely in terms of parameters estimated from historic and synthetic sequences of length n. As

the small sample properties of $\hat{\gamma}_x$ are unknown, it will be excluded from consideration. If necessary, resemblance of type A can be maintained for this parameter.

The overall procedure is as follows :

(a) Derive estimates $\hat{\mu}_x$, $\hat{\sigma}_x^2$, $\hat{\rho}_x$ (using equation 3.28) and K from a historic sequence of length n.

(b) From tables (3.2) - (3.7), identify values of φ and θ such that

$$E[K]_n \approx K$$

and $$E[\hat{\rho}_1]_n \approx \hat{\rho}_x$$

Again, some interpolation will invariably be involved here.

(c) By re-writing equation (3.27) as

$$E[s^2]_n = \sigma_x^2 f(n, \rho_1, \varphi) \qquad (3.33)$$

where ρ_1 and φ have been defined under (b), and equating $E[s^2]_n$ with $\hat{\sigma}_x^2$, an unbiased estimate of σ_x^2 may be defined as

$$\hat{\sigma}_x^{*2} = \hat{\sigma}_x^2 / f(n, \rho_1, \varphi) \qquad (3.34)$$

(d) The quantities $\hat{\mu}_x$, $\hat{\sigma}_x^*$, φ and θ are then incorporated into equation (3.29), whence synthetic flows may be generated.

Resemblance of type B would appear to be the more logical type to achieve ; however, only consideration of the benefits and regrets associated with maintaining each type of resemblance can throw light on which resemblance is "best" to achieve.

Because of the nature of the estimator K, the estimates ρ_1 and K may occasionally fall outside the range of $\tilde{E}[\hat{\rho}_1]_n$ and $\tilde{E}[K]_n$ values in tables (3.2) - (3.7), thus preventing the exact required resemblance from being maintained. In such cases, only approximate resemblance can be maintained. In addition, estimates of ρ_1 and h may be in conflict with each other, such that the required match cannot be achieved, even though each parameter

falls within the bounds of tables (3.2) - (3.7). In such cases, resemblance could probably be achieved in terms of either parameter but not both.

3.3.7 A log-normal ARIMA (1,0,1) process

If synthetic flows conforming to a log-normal distribution are required, then the process $Y_t = \ln(X_t - a)$, where X_t is log-normally distributed with a lower bound a is assumed to be a Gaussian ARIMA (1,0,1) process. The relationships between the mean μ_y and standard deviation σ_y of the process Y_t and the lower bound and first three moments of the distribution of X_t are given by equations (2.7) - (2.9). Thus estimates of μ_x, σ_x^2 and ρ_x may be used to solve for estimates of μ_y, σ_y and a. If the process Y_t is assumed to be generated by an ARIMA (1,0,1) process with parameters within the parameter space covered by tables (3.2) - (3.7), then long-term persistence will also exist in X-space. The parameters φ and θ of the process Y_t may be defined using either of the procedures previously outlined to achieve resemblance of type A or type B ; in either case, K is derived from a sample of the process Y_t and an estimate of the lag-one autocorrelation of the flows in Y-space, ρ_y, is required which may be defined from the following equation (Appendix 3. 1)

$$\rho_x = \frac{e^{\sigma_y^2 \rho_y} - 1}{e^{\sigma_y^2} - 1} \qquad (3.35)$$

where ρ_x is replaced by its sample estimate. The estimate of σ_y^2 derived from solving equations (2. 7) - (2. 9) may be corrected for bias using the procedure outlined previously for the variance σ_x^2, if resemblance of type B is desired.

While resemblance of type B can be maintained for certain parameters in Y-space, there is no guarantee that such resemblance will be maintained in X-space because of the non-linear transformations relating Y_t and X_t.

The observed estimate of K in X-space may not be preserved as Mandelbrot and Wallis (1969e) have noted that estimates of h derived from a type 1 approximation to dfGn are not invariant under highly non-linear transformations.

3.4 Impact of Long-term Persistence on Reservoir Storage Design

While the lag-one Markov and ARIMA (1,0,1) processes can display very distinct long run effects, it is of primary interest to a designer to know under what design conditions long-term persistence becomes important. For the design of a storage reservoir, it is to be expected that as level of development and length of design period decrease, then so too will the importance of long-term persistence. For a design period of 100 years, Wallis and Matalas (1972) have shown through simulation experiments that, for levels of development greater than 0.8, approximations to dfGn yield significantly greater storage requirements than those corresponding to the lag-one Markov process. Similar experiments to those of Wallis and Matalas (1972) are carried out here to provide a further comparison of a formal approximation to dfGn with the ARIMA (1,0,1) process. In addition, the experiments are extended to cover design periods of 25 and 50 years.

3.4.1 Simulation Experiments

In comparing design results such as reservoir storages obtained from different generating processes, considerable care must be exercised. An obvious approach is to select equal population parameters (e.g. mean, variance, skewness, lag-one autocorrelation etc.) for each generating process, and compare design storages evolved from synthetic sequences of length equal to the design period. However, because different small sample biases operate for each generating mechanism, such studies may

result in fallacious conclusions as to the impact of generating mechanism on a design. The following example illustrates this possibility.

Markov and ARIMA (1,0,1) generating processes are to be compared on the basis of the following population parameters :

$$\mu = 10$$

$$\sigma = 3$$

$$\rho_1 = 0.3$$

which yield typical values of $C_v = \sigma/\mu$, the coefficient of variation, and ρ_1 for annual streamflow. Reservoir storages for a 100 year design period are to be used for comparison purposes. As various values of φ can yield $\rho_1 = 0.3$, a value of 0.88, which corresponds to a process exhibiting a moderate intensity of long-term persistence, will be used in the example.

For the lag-one Markov generating process, the expected values of estimates of σ^2 and ρ_1 in samples of size n are given by equations (2.91) and (2.88), respectively, which, for the example quoted, reduce to $E[S^2]_{100} = 8.91$, and $E[\hat{\rho}_1]_{100} = 0.278$. Correspondingly, for the ARIMA (1,0,1) process, $E[S^2]_n$ is given by equation (3.27), while an analytical result for $E[\hat{\rho}_1]_n$ does not exist. However, reference to table (3.4) shows that, for $\varphi = 0.88$, $E[\hat{\rho}_1]_{100} \approx 0.225$, while from equation (3.27), $E[S^2]_{100} = 8.59$. The overall effect of the larger biases in estimates of σ^2 and ρ_1 for the ARIMA (1,0,1) process might well be to suggest that, contrary to expectation, the lag-one Markov process would require larger storages over a 100 year design period. Consequently, valid comparisons can only be made on the basis of equal values of $E[S^2]_n$ (or $E[S]_n$) and $E[\hat{\rho}_1]_n$ for each process.

Accordingly, in comparing expected storage requirements for the ARIMA (1,0,1) and (1,0,0) or lag-one Markov processes, comparisons were

based as far as possible on equal values of $E[S]_n$ and $E[\hat{\rho}_1]_n$, so as to allow comparisons with the results of Wallis and Matalas (1972). Estimates of σ for the lag-one Markov process satisfy

$$E[S_m]_n = \sigma_m \, g(n, \rho_1) \tag{3.36}$$

and, for the ARIMA (1,0,1) process

$$E[S_a]_n = \sigma_a \, g(n, \rho_1, \varphi) \tag{3.37}$$

where $g(n, \rho_1)$ and $g(n, \rho_1, \varphi)$ are small sample bias factors for the standard deviations σ_m and σ_a respectively. If the analytic forms for $g(n, \rho_1)$ and $g(n, \rho_1, \varphi)$ were known, then for a given value of $E[S]_n$ for both processes, the corresponding population variances could be defined as

$$\sigma_m = E[S]_n / g(n, \rho_1) \tag{3.38}$$

and

$$\sigma_a = E[S]_n / g(n, \rho_1, \varphi) \tag{3.39}$$

However, in the absence of analytic forms for $g(n, \rho_1)$ and $g(n, \rho_1, \varphi)$, the following generating procedure was adopted.

For each process, synthetic sequences were generated with $\mu = 10$ and $\sigma = 3$ which represents a typical value of C_v for historic streamflow sequences. The lag-one autocorrelation ρ_1 was allowed to assume values which adequately covered the range 0 to 1 for both processes. For the ARIMA (1,0,1) process the parameter φ was used to control the intensity of long-term persistence, and was given 3 values, 0.70, 0.80 and 0.90. For each design sequence length, n = 25, 50 and 100, for each value of ρ_1 (and φ in the case of the ARIMA (1,0,1) process), 1000 sequences were generated, and the sequent peak algorithm (Appendix 3.2) was used to derive a reservoir design storage corresponding to each sequence, denoted by C_m in the case of the lag one Markov process and C_a in the case of the

ARIMA (1,0,1) process, for a given level of development. In applying the sequent peak algorithm, levels of development of 0.80, 0.90 and 0.99 were considered, with a uniform draft pattern over the design period. The following procedure was adopted for correcting the bias in the standard deviation.

For each process, approximate values of $g(n, \rho_1)$ and $g(n, \rho_1, \varphi)$ were defined as

$$g(n, \rho_1) = \bar{S}_m/\sigma \tag{3.40}$$

and

$$g(n, \rho_1, \varphi) = \bar{S}_a/\sigma \tag{3.41}$$

where $\bar{S}_m$ and $\bar{S}_a$ denote the average standard deviation for each process defined from 1000 samples for each value of n. However, separate experiments were not required to estimate $g(n, \rho_1)$ and $g(n, \rho_1, \varphi)$ beforehand so that they might be used to define the appropriate values of σ_m and σ_a from equations (3.38) and (3.39), respectively. If a scaling factor is applied to the flows to be operated on by the sequent peak algorithm, then the storage, C_m or C_a, which would result from the unscaled flows will have the same scaling factor applied because of the linearity of the operations in the sequent peak algorithm (Appendix 3.3). Hence, the mean storage corresponding to equal expected values of the standard deviation in samples of size n for each process may be defined as

$$\lambda_m = \bar{C}_m(\sigma/\bar{s}_m)$$

for the lag one Markov process, and

$$\lambda_a = \bar{C}_a(\sigma/\bar{s}_a)$$

where $\bar{C}_m$ and $\bar{C}_a$ are the mean storages calculated from a 1000 successive realizations of each process generated with the same population standard deviation σ.

The problem of different values of $E[\hat{\rho}_1]_n$ for each process corresponding to a given value of ρ_1 may be obviated through comparing λ_m and λ_a values graphically on the basis of equal $E[\hat{\rho}_1]_n$ values. Values of $E[\hat{\rho}_1]_n$ for various values of ρ_1 may be obtained from equation (2.88) while estimates of $E[\hat{\rho}_1]_n$ for the ARIMA (1,0,1) are available from the sampling experiments conducted in section (3.3.5). A comparison of λ_m and λ_a for n = 100 and $\alpha' = 0.99$ is presented in figure (3.28). It should be noted that λ_m and λ_a are plotted against $\tilde{E}[\hat{\rho}_1]_n$ rather than against ρ_1. While there is a one-to-one relation between λ_m and $E[\hat{\rho}_1]_n$ for the lag-one Markov process, different values of φ lead to a family of curves for λ_a, one for each value of φ. For the values of φ considered, λ_a is at all times greater than λ_m, even for small values of $E[\hat{\rho}_1]_n$, with the difference between λ_a and λ_m increasing with increasing φ. The parameter φ essentially controls the intensity of long-term persistence. As φ decreases for a given value of φ, ρ_1 increases, eventually attaining an upper limit of φ, when the process becomes Markovian. Hence the λ_m and λ_a curves converge eventually at a value of $\tilde{E}[\hat{\rho}_1]_n$ corresponding to a value of $\rho_1 = \varphi$. The clear separation between the λ_m and λ_a values is similar to that obtained by Wallis and Matalas (1972) for the filtered fractional noise approximation, which is further evidence of the suitable approximation to dfGn represented by the ARIMA (1,0,1) process. Obviously, for high levels of development, use of the correct generating process in a design situation is desirable.

If the level of development is dropped to 0.90 and 0.80 (figures (3.29) and (3.30)), the ratio of λ_m and λ_a does not decrease very much, suggesting that the effect of long term persistence on required storage is significant, even for a moderately high level of development of 0.80.

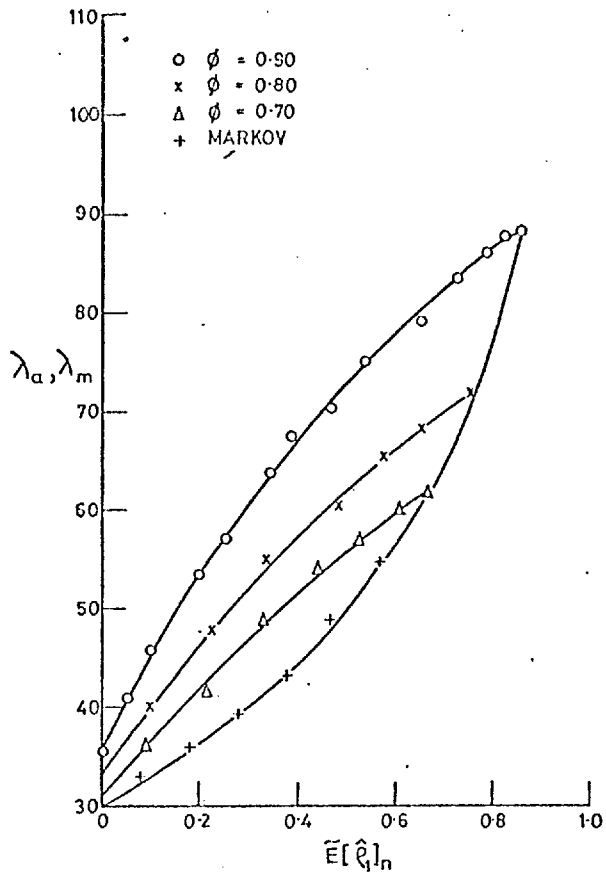


Figure (3.28) The average minimum storage for ARIMA (1,0,1) and lag-one Markov processes as a function of the approximate expected value of ρ_1, the lag-one autocorrelation in samples of size 100. Each point is based on 1000 realizations of length 100 with $\mu = 10$ and $E(S)_n = 3$ and α', the level of development, equal to 0.99.

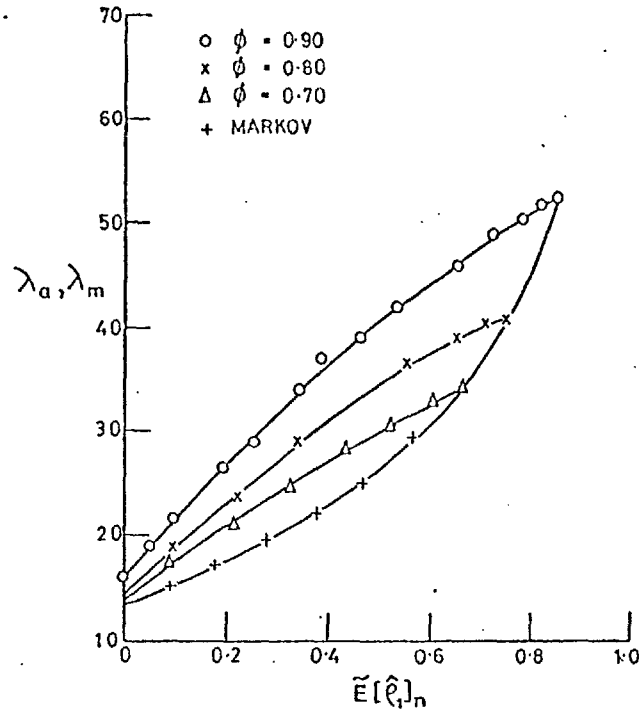


Figure (3.29) The average minimum storage for ARIMA (1,0,1) and lag-one Markov processes as a function of the approximate expected value of ρ_1, the lag-one autocorrelation in samples of size 100. Each point is based on 1000 realizations of length 100 with $\mu = 10$ and $E(S)_n = 3$ and α', the level of development, equal to 0.90.

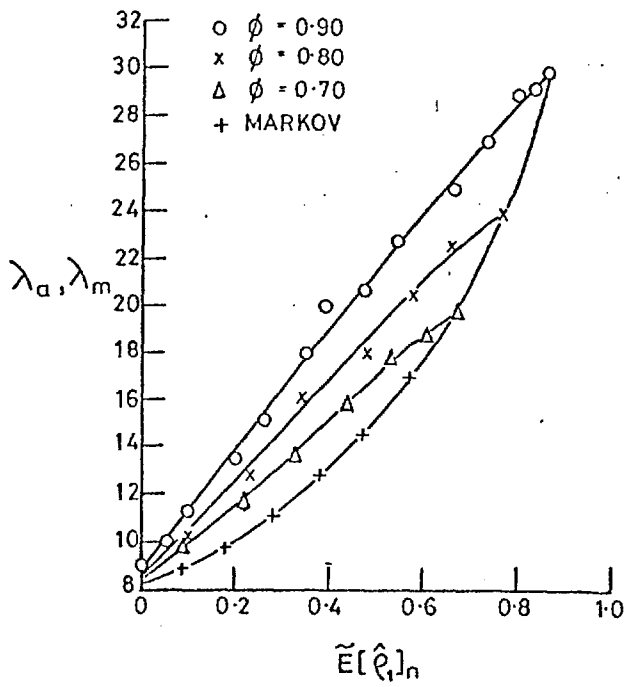


Figure (3.30) The average minimum storage for ARIMA (1,0,1) and lag-one Markov processes as a function of the approximate expected value of ρ_1, the lag-one autocorrelation in samples of size 100. Each point is based on 1000 realizations of length 100 with $\mu = 10$ and $E(S)_n = 3$ and α', the level of development equal to 0.80.

If the length of the design period is decreased to 50 years, the same pattern emerges as for n equal to 100 (figure (3.31)). The absolute magnitudes of λ_m and λ_a have obviously decreased relative to their values for n equal to 100, but the relative magnitudes of λ_m and λ_a have altered little. Decreasing the sample size still further to n equal to 25 results in a somewhat different picture (figure (3.32)). The separation between λ_a and λ_m ceases to be as distinct as for the larger sample sizes, a reflection of the fact that low frequencies can barely manifest themselves in such small sample sizes. Further, the λ_a curves for the higher values of φ tend to collapse on those curves corresponding to the lower values of φ. Apparently, the reason for this behaviour is that a considerable proportion of the C_a values tend to be zero for larger values of φ i.e. a reservoir is not needed. The reason for this rather surprising behaviour is that the small sample bias in the standard deviation in sample size 25 is so severe as to render the variability of the flows about the sample mean so small in a number of cases that the resulting C_a values are zero. Had the correction for bias in the standard deviation been applied to the flows before the C_a values were computed, the effect of the zero C_a values on the λ_a curves would have been diminished ; however, separate experiments would have been required to define $g(n, \rho_1, \varphi)$ beforehand. Again, the results presented in figure (3.32) emphasise the care that is needed in designing small sample experiments involving processes exhibiting long-term persistence. Results for levels of development less than 0.99 and sample sizes less than 100 showed that zero C_m and C_a values again played too dominant a role to allow representative λ_m and λ_a values to be presented.

In the foregoing experiments, the effects of length of design period, level of development and generating process on required reservoir storage

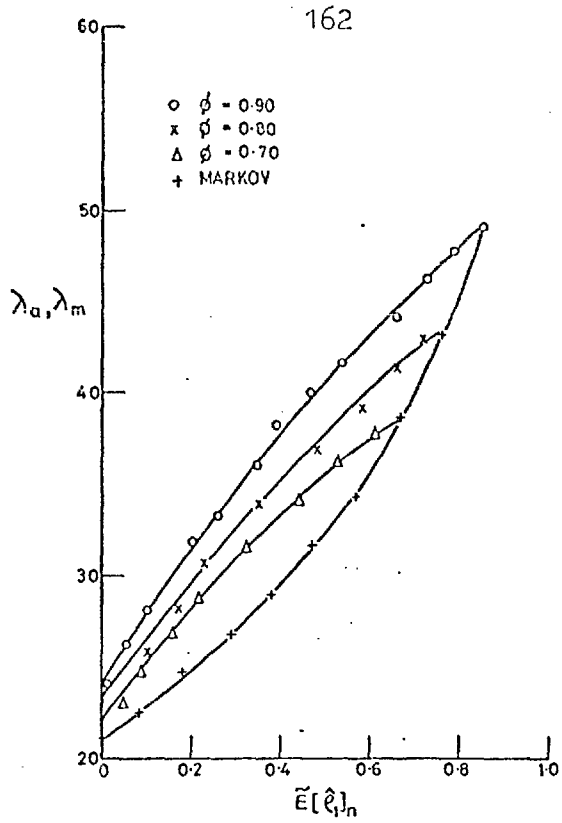


Figure (3.31) The average minimum storage for ARIMA (1,0,1) and lag-one Markov processes as a function of the approximate expected value of ρ_1, the lag-one autocorrelation in samples of size 50. Each point is based on 1000 realizations of length 50 with μ = 10 and $E(S)_n$ = 3 and α', the level of development, equal to 0.99.

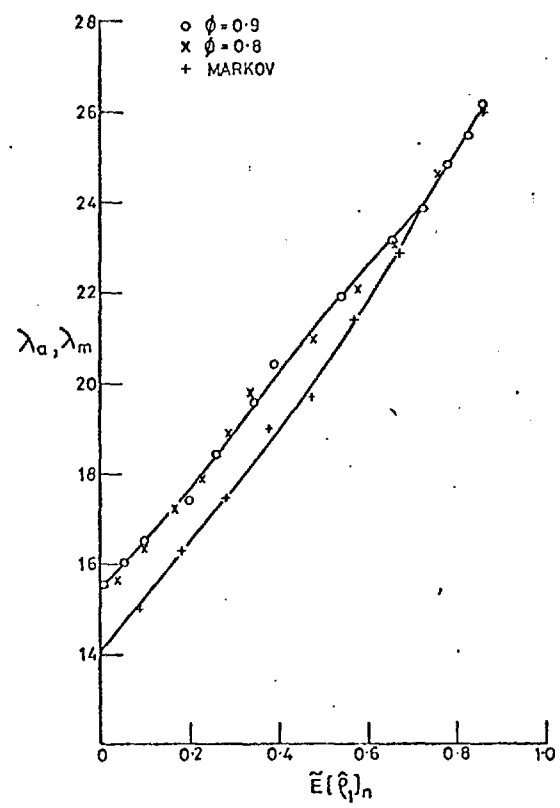


Figure (3.32) The average minimum storage for ARIMA (1,0,1) and lag-one Markov processes as a function of the approximate expected value of ρ_1, the lag-one autocorrelation in samples of size 25. Each point is based on 1000 realizations of length 25 with μ = 10 and $E(S)_n$ = 3 and α', the level of development, equal to 0.99.

have been investigated. The results of the experiments are relevant only to reservoir designs based on an annual time unit, and a level of variability of $C_v = 0.3$ has been chosen accordingly. As sample size decreases, long-term persistence becomes less important. However, for a design period of 100 years, the ratio λ_a / λ_m for any given value of $E[\hat{\rho}_1]_n$ does not vary much for the three levels of development considered ; presumably levels of development less than 0.8 would need to be considered before long-term persistence would cease to influence storage design. The experiments have also illustrated that care is required in designing such experiments where small sample properties differ markedly from the population properties of a process.

3.5 Higher Order ARIMA Processes

The long-term persistence of stationary higher order ARIMA processes, in particular, the ARIMA (2,0,1) and ARIMA (1,0,2) processes, have also been explored using similar experiments to those conducted in sections (3.3.4) and (3.3.5) ; however, lack of space does not allow a detailed description to be furnished here. Briefly, for each model, regions in the parameter space (as defined by stationarity and invertibility conditions) were identified which gave autocorrelation functions which decayed slowly, but with the restriction that the high frequency behaviour of the process should be similar to that of annual streamflow. The approximation to dfGn afforded by either process was not found to be significantly better than that provided by the ARIMA (1,0,1) process ; for the added complexity of an extra parameter, the only apparent advantage of the two higher order processes is to allow some extra flexibility in fitting estimated lag-one and lag-two autocorrelations.

In addition, some limited experiments were carried out on the behaviour of Hurst's law for some non-stationary ARIMA processes. While

such processes cannot be used to generate synthetic streamflows, the possibility that annual streamflow is generated by a non-stationary process cannot be ruled out. There are a vast number of ways in which a process can be non-stationary ; however, in the experiments which were conducted, only the cases of non-stationary ARIMA (0,1,1) and (0,1,2) processes were considered. While reasonable agreement with Hurst's law was achieved for small to moderate values of n (roughly < 400) a value of $h = 1$ then appeared to govern the behaviour of R/S for larger values of n, resulting in divergence from Hurst's law. In the absence of such behaviour among long geophysical records, the type of non-stationarity which characterizes the ARIMA (0,1,1) and (0,1,2) processes is hardly plausible for such records.

3.6. Summary

The long-term behaviour of the ARIMA (1,0,1) process has been examined using extensive simulation experiments, which showed that the process can be used to approximate the behaviour of the rescaled range R/S for long geophysical records. In addition, the process can model the short-term properties of such records, as measured by the lag-one autocorrelation coefficient. The small sample properties of estimates of the variance, lag-one autocorrelation and Hurst coefficient for sample sizes 25, 50 and 100 have been derived through simulation, illustrating that large disparities exist between the small sample and population expectations of those quantities in the presence of long-term persistence. The small-sample properties of the ARIMA (1,0,1) process have been used in formulating the process to maintain the required resemblance between historic and synthetic sequences. Some simple reservoir design experiments suggest that, for moderate to high levels of development, long-term persistence would influence reservoir storage sizes for design periods of 25 - 100 years.

Experiments with higher order stationary ARIMA models suggest that the complexity introduced by extra parameters does not provide any significant improvement over the long-term properties of the ARIMA (1,0,1) process.

Chapter 4

MULTISITE STOCHASTIC MODELS OF LONG-TERM PERSISTENCE

The design of a water resource system will invariably demand the generation of synthetic sequences at a number of sites within a river basin or river basins. If the historic sequences in such a situation are spatially uncorrelated, then an appropriate univariate generating mechanism may be used at each of the m sites within the system, and no representation of spatial correlation is necessary. However, significant cross-correlations will usually exist between historic events measured at neighbouring sites at the same time point. Such correlations are referred to as lag-zero cross-correlations between sites. Because of serial correlation at each individual site, significant correlations will also exist between sites where the flows at each site are lagged k time units with respect to all other sites where k = 1,2,3,...; such correlations are referred to as lag-k cross-correlations. The preservation of such cross-correlations between synthetic sequences is of primary importance, particularly in the design of storage systems, where the simulation of simultaneous periods of critical low flows at all sites must be attempted. Generating processes which preserve cross-correlations between sites in addition to the appropriate properties at individual sites are referred to as multivariate generating processes.

In section (4.1) the development of multivariate generating processes is reviewed, with particular attention paid to the multivariate lag-one Markov process, a multivariate fractional noise model, and distributional properties of the flows. The multivariate ARIMA (1,0,1) process is formulated in section (4.2), and an iterative computational procedure is found to be necessary for obtaining a general solution to

the matrix equations which specify the coefficients of the generating process. An analytic solution is also derived, shortcomings of the solution noted, and a modified multivariate ARIMA (1,0,1) process is developed which is suitable for generating synthetic flows preserving a measure of long-term persistence at each site. Distributional properties of multivariate ARIMA (1,0,1) flows are considered in section (4.3).

4.1 Development of Multivariate Models

The first attempt at preserving cross-correlations between synthetic sequences was perhaps that of Thomas and Fiering (1962), when they attempted to model observed cross-correlations between pairs of sites. If X_t and Y_t denote the flows at each of two sites, then the generating processes for X_t and Y_t are as follows:

$$X_t = \rho_x X_{t-1} + \sqrt{1 - \rho_x^2}\, \varepsilon_t \qquad (4.1)$$

$$Y_t = \rho_{xy} X_t + \sqrt{1 - \rho_{xy}^2}\, \delta_t \qquad (4.2)$$

The above strategy endeavoured to preserve estimates of the lag-zero cross correlation ρ_{xy} between X_t and Y_t and of the lag-one autocorrelations ρ_x and ρ_y at each site. However, while the procedure ensures that estimates of ρ_{xy} and ρ_x are preserved, the flows Y_t will have $\rho_{xy}\,\rho_x$ as lag-one autocorrelation coefficient rather than the observed estimate of ρ_y. Fiering (1964) pointed out this deficiency and presented a modified bivariate generating process which preserved ρ_x, ρ_y and ρ_{xy}.

While the above approach may be extended to generate synthetic flows at a third station from Y_t, and so on, the technique necessitates the selection of a key station X_t and as a result, all inter-station correlations will not be preserved. Fiering (1964) proposed a technique based on

principal components whereby all inter-station correlations might be preserved in addition to means, variances and lag-one serial correlations at individual sites. However, Matalas (1967) subsequently showed that while all cross-correlations between sites were indeed preserved, the serial correlations at individual sites were not, and suggested how the generating process might be modified to preserve lag-one serial correlations.

Beard (1965) presented a technique based on a multiple linear regression for generating synthetic sequences, where the dependent variable being generated is the flow at a particular site at time t, and the independent variables are the flows at times t, t-1,..., t-k at the remaining sites. A set of transformations are effected initially on the historic flows to assure normality; as a result only statistics of the transformed variates are preserved rather than those of the actual flows and consequently generated flows may not be representative of historic flows.

4.1.1 A Multivariate Autoregressive Process

Matalas (1967) presented a weakly stationary lag-one autoregressive process defined as

$$\underline{x}(t) = \underline{A}\,\underline{x}(t-1) + \underline{B}\,\underline{\varepsilon}(t) \qquad (4.3)$$

In equation (4.3), $\underline{x}(t)$ and $\underline{x}(t-1)$ are (m x 1) matrices whose elements are $x_i(t) = X_i(t) - \mu_i$ and $x_i(t-1) = X_i(t-1) - \mu_i$, respectively, where $i = 1,2,\ldots,m$ and m is the number of sites; $X_i(t)$ and $X_i(t-1)$ denote the flows for site i at times t and (t-1) respectively and μ_i is the mean flow at site i. The random component $\underline{\varepsilon}(t)$ is an (m x 1) matrix whose elements are independent of the elements of $\underline{x}(t-1)$. $\underline{A}$ and $\underline{B}$ are (m x m) matrices whose elements must be defined in such a way that multivariate synthetic sequences generated by equation (4.3) will resemble multivariate

historic sequences in terms of $\hat{\mu}_i$, $\hat{\sigma}_i$, $\hat{\gamma}_i$, $\hat{\rho}_i(1)$, $\hat{\rho}_{ij}(0)$ and $\hat{\rho}_{ij}(1)$, $i,j = 1,2,\ldots,m$, where $\hat{\sigma}_i$, $\hat{\gamma}_i$ and $\hat{\rho}_i(1)$ represent estimates of the standard deviation, skewness and lag-one autocorrelation at site i, and $\hat{\rho}_{ij}(0)$ and $\hat{\rho}_{ij}(1)$ represent estimates of the lag-zero and lag-one cross-correlations between sites i and j, $i,j = 1,2,\ldots,m$. Matrix notation permits a more concise presentation, and a lag-zero variance-covariance matrix $\underline{M}_o$ for the process is defined as follows:

$$\underline{M}_o = E[\underline{x}(t)\ \underline{x}^T(t)]$$

$$= \begin{bmatrix} \sigma_1^2 & \cdots & \rho_{ij}(0)\sigma_i\sigma_j & \cdots \\ \vdots & \sigma_2^2 & \vdots & \\ \rho_{ji}(0)\sigma_j\sigma_i & \cdots & \ddots & \\ \vdots & & & \sigma_m^2 \end{bmatrix}$$

where

$$E[x_i(t)\ x_j(t)] = \sigma_i^2 \qquad i = j$$

and $$E[x_i(t)\ x_j(t)] = \rho_{ij}(0)\sigma_i\sigma_j \qquad i \neq j$$

The lag-one covariance matrix is defined as follows:

$$\underline{M}_1 = E[\underline{x}(t)\ \underline{x}^T(t-1)]$$

$$= \begin{bmatrix} \rho_1(1)\sigma_1^2 & \cdots & \rho_{ij}(1)\sigma_i\sigma_j & \cdots \\ \vdots & \rho_2(1)\sigma_2^2 & \vdots & \\ \rho_{ji}(1)\sigma_j\sigma_i & \cdots & \ddots & \\ \vdots & & & \rho_m(1)\sigma_m^2 \end{bmatrix}$$

where

$$E[x_i(t)\ x_j(t-1)] = \rho_i(1)\ \sigma_i^2 \qquad i = j$$

and $$E[x_i(t)\ x_j(t-1)] = \rho_{ij}(1)\ \sigma_i\sigma_j \qquad i \neq j$$

If the process is standardized as

$$x_i(t) = \frac{X_i(t) - \mu_i}{\sigma_i}$$

then $\underline{M}_o$ becomes the lag-zero cross-correlation matrix of the flows

$$\underline{M}_o = \begin{bmatrix} 1 & \cdots & \rho_{ij}(0) & \cdots \\ \vdots & \ddots & & \\ \rho_{ji}(0) & & \ddots & \\ \cdots & & & 1 \end{bmatrix}$$

and $\underline{M}_1$ becomes the lag-one cross-correlation matrix

$$\underline{M}_1 = \begin{bmatrix} \rho_1(1) & \cdots & \rho_{ij}(1) & \cdots \\ \vdots & \rho_2(1) & & \\ \rho_{ji}(1) & & \ddots & \\ \cdots & & & \rho_m(1) \end{bmatrix}$$

In practice, estimates of the elements of $\underline{M}_o$ and $\underline{M}_1$ must be provided from the historic sequences, and $\underline{A}$ and $\underline{B}$ subsequently defined such that the generated multivariate synthetic sequences will resemble historic sequences in terms of the estimated $\underline{M}_o$ and $\underline{M}_1$. Note that as $\underline{M}_o$ is symmetric, only

$$\frac{m(m+1)}{2} + m^2$$

conditions may be imposed on the elements of $\underline{A}$ and $\underline{B}$. Matalas has shown that $\underline{A}$ and $\underline{B}$ may be derived as

$$\underline{A} = \underline{M}_1 \underline{M}_o^{-1} \tag{4.4}$$

$$\underline{B}\underline{B}^T = \underline{M}_o - \underline{M}_1 \underline{M}_o^{-1} \underline{M}_1^T \tag{4.5}$$

The elements of the matrix $\underline{B}$ are given by the solution of equation (4.5). As $\underline{B}\underline{B}^T$ is symmetric a unique solution does not exist for $\underline{B}$ and any matrix

$$\underline{B}^* = \underline{B}\,\underline{\Delta}$$

where $\underline{\Delta}$ is an (m x m) matrix such that $\underline{\Delta}\underline{\Delta}^T = \underline{I}$, where $\underline{I}$ is the identity matrix, will satisfy equation (4.5). The method of principal components was advocated by Matalas (1967) as a means of solving for the elements of $\underline{B}$. In a discussion on the generating process given in equation (4.3), Young (1968) suggested that, as $\underline{B}\underline{B}^T$ is symmetric, $\underline{B}^*$ might more conveniently be considered a lower triangular matrix, allowing simple recursive solutions for the elements of $\underline{B}^*$.

The matrices $\underline{A}$ and $\underline{B}$ may only be solved for in terms of the estimated $\underline{M}_o$ and $\underline{M}_1$ under certain conditions. Firstly, $\underline{M}_o$ must be positive definite; provided $\underline{M}_o$ is estimated from records of equal length, this condition will be satisfied. However, Fiering (1968) has pointed out that if records of unequal length are used to estimate $\underline{M}_o$, then $\underline{M}_o$ may not be positive definite; Crosby and Maddock (1970) have presented a technique based on monotone samples of data which ensures that the estimated $\underline{M}_o$ will be positive definite.

However, the fact that $\underline{M}_o$ is positive definite does not ensure that $\underline{B}\underline{B}^T$ will be positive definite (Matalas and Wallis, 1971a). Unless $\underline{B}\underline{B}^T$ is positive definite, the elements of $\underline{B}$, solved for by principal component techniques or by lower triangularisation will be complex

numbers and as a result the generated flows will themselves be complex numbers. The fact that $\underline{B}\underline{B}^T$ is not positive definite when computed from finite historic sequences may be attributed to the estimators used for the matrices $\underline{M}_o$ and $\underline{M}_1$. Crosby and Maddock (1970) and Valencia and Schaake (1972) have proposed estimators for $\underline{M}_o$ and $\underline{M}_1$ which ensure that $\underline{B}\underline{B}^T$ will always be positive definite when computed from finite data samples.

Matalas and Wallis (1971a) have shown that the condition that $\underline{B}\underline{B}^T$ be positive definite for the process given by equation (4.3) corresponds to a set of constraints imposed on the elements of $\underline{M}_o$ and $\underline{M}_1$. If the generating process of streamflow possesses correlations which lie outside these constraints then it is not entirely clear how the techniques of Crosby and Maddock (1970) and Valencia and Schaake (1972) can provide a positive definite estimate of $\underline{B}\underline{B}^T$ based on the hypothesis that the generating process is equation (4.3). Such a premise is based on the fact that in the limiting case of infinite historic sequences, the correlation of the real world generating process must converge to their respective population quantities which by definition lie outside the constraints for equation (4.3). Only if the estimators provided by Crosby and Maddock (1970) and Valencia and Schaake (1972) converge to population quantities which lie inside the constraints, will $\underline{B}\underline{B}^T$ be positive definite in the limiting case for equation (4.3).

4.1.2 A Multivariate Markovian Process

If the elements of the matrix A in equation (4.3) are defined in a particular fashion, then the autocorrelation function at each of the sites and the lag-k cross correlations between sites assume a lag-one Markovian form. Matalas (1967) has shown that if the matrix A is defined as a diagonal matrix with the lag-one serial correlations for

the n sites along the diagonal. i.e.

$$\underline{A} = \begin{bmatrix} \rho_1(1) & & & & & \\ & \rho_2(1) & & & & \\ & & \ddots & & & \\ & & & \ddots & & \\ & & & & \ddots & \\ & & & & & \rho_m(1) \end{bmatrix}$$

then the elements of $\underline{M}_1$ must be defined as

$$\rho_{ij}(1) = \rho_i(1)\rho_{ij}(0) \qquad i,j = 1,2,\ldots,m \qquad (4.6)$$

Such an approach precludes taking the inverse of $\underline{M}_o$ and $\underline{B}\underline{B}^T$ may now be defined as

$$\underline{B}\underline{B}^T = \underline{M}_o - \underline{A}\,\underline{M}_1^T \qquad (4.7)$$

As a result, only $\underline{M}_o$ and the lag-one serial correlations at all sites have to be estimated; provided the lag-one cross-correlations are not required to be preserved explicitly, then the amount of computation is reduced, $\underline{M}_o$ will be preserved as observed and the elements of $\underline{M}_1$ will be preserved as defined by equation (4.6). Hence multivariate synthetic sequences will resemble multivariate historic sequences in terms of $\hat{\mu}_i$, $\hat{\sigma}_i$, $\hat{\rho}_{ij}(0)$ and $\hat{\rho}_i(1)$ but not in terms of $\hat{\rho}_{ij}(1)$ for $i \neq j$. For higher lags the autocorrelation function at each of the sites may be shown to be

$$\rho_i(k) = \hat{\rho}_i(1)^{|k|} \qquad (4.8)$$

while the lagged cross-correlations between sites may be shown to be

$$\rho_{ij}(k) = \rho_i(k) \cdot \hat{\rho}_{ij}(0) \qquad (4.9)$$

Consequently the multivariate lag-one Markov process suffers from the

same deficiencies with respect to modelling h as the univariate Markov process. Even if $\underline{A}$ and $\underline{B}$ are defined via equation (4.4) and (4.5), the long-term properties of the process will be similar to those of the multivariate lag-one Markov process.

4.1.3 Multivariate Fractional Noise Models

As already noted, multivariate lag-one Markovian processes are characterized by values of h = 0.5 at each site. Consequently, Matalas and Wallis (1974) have proposed a multivariate filtered fractional Gaussian noise process as

$$x_i(t) = (h_i - 0.5) \sum_{\delta=pt-M_i}^{pt-1} (pt - \delta)^{h_i-1.5} \sum_{r=1}^{i} b_{ir}\varepsilon_r(\delta) \quad (4.10)$$

where

$$x_i(t) = X_i(t) - \mu_i$$

and $\varepsilon_r(t)$ is a normal independent random variable. By considering $E[x_i(t)\, x_j(t)]$ for i,j = 1,2,...,m then m(m + 1)/2 equations with m(m + 1)/2 unknowns may be solved recursively for the b's. As a result, multivariate synthetic sequences will resemble multivariate historic sequences in terms of estimates of μ_i, σ_i, h_i and $\rho_{ij}(0)$. However, Matalas and Wallis (1971b) have noted that different values of p, the filtering parameter, may not be assumed at each of the sites; otherwise the process will not be stationary with respect to the cross-correlations. Consequently, difficulty may be encountered in preserving estimates of $\rho_i(1)$ in conjunction with estimates of h_i at each site.

4.1.4 Distribution of Flows

Resemblance between multivariate historic and synthetic sequences in terms of the elements of $\underline{M}_o$ and $\underline{M}_1$ may be achieved through any choice of distribution for the element $\underline{\varepsilon}(t)$ in equation (4.3) provided

$\varepsilon(t)$ is an independently distributed random variable with zero mean and unit variance. If, in addition, the distribution of $\varepsilon(t)$ is assumed normal then the distribution of the flows will be multivariate normal and the process will be strictly stationary.

If the skewness at each individual site, γ_i, is of interest, multivariate synthetic sequences may be generated which resemble multivariate historic sequences in terms of γ_i, $[i = 1,2,\ldots,m]$. Matalas and Wallis (1974) have shown how the skewness of the random element in equation (4.3) may be defined in order to achieve resemblance in terms of γ_i, $i = 1,2,\ldots,m$. The Wilson-Hilferty transform defined by equation (2.5) may be used to generate the random element $\varepsilon_i(t)$ at each site with the required skewness.

If a particular choice of distribution for the flows is of interest, then the log-normal distribution may be considered. Let

$$Y_i(t) = \ln (X_i(t) - a_i) \tag{4.11}$$

then the lag-one autoregressive generating process in terms of $Y_i(t)$ is defined as

$$\underline{y}(t) = A'\underline{y}(t-1) + B'\underline{\varepsilon}(t) \tag{4.12}$$

where $\underline{\varepsilon}(t)$ is NIP(0,1) and $y_i(t) = \dfrac{Y_i(t) - \mu'_i}{\sigma'_i}$. The matrices $\underline{A}'$ and $\underline{B}'$ must be defined from the correlation matrices $\underline{M}'_0$ and $\underline{M}'_1$ of the process $\underline{Y}(t)$. The elements of the matrices $\underline{M}'_0$ and $\underline{M}'_1$, may be defined as follows (Matalas, 1967). Let $\rho'_{ij}(0)$ and $\rho'_{ij}(1)$ denote typical elements of $\underline{M}'_0$ and $\underline{M}'_1$, which may be defined from the elements of $\underline{M}_0$ and $\underline{M}_1$ through the following equations

$$\rho_{ij}(0) = \frac{\exp\left[\sigma'_i\,\sigma'_j\,\rho_{i'j}(0)\right] - 1}{\left[\exp\left(\sigma'^2_i\right) - 1\right]^{\frac{1}{2}}\left[\exp\left(\sigma'^2_j\right) - 1\right]^{\frac{1}{2}}} \tag{4.13}$$

$$\rho_{ij}(1) = \frac{\exp[\sigma'_i \sigma'_j \rho_{i'j}(1)] - 1}{[\exp(\sigma'^2_i) - 1]^{\frac{1}{2}} [\exp(\sigma'^2_j) - 1]^{\frac{1}{2}}} \qquad (4.14)$$

Consequently, the elements of $\underline{M}'_0$ and $\underline{M}'_1$ may be defined through inverting equations (4.13) and (4.14). The mean, variance and skewness of the flows $X_i(t)$ at each site are defined by equations (2.7) - (2.9), from which the mean μ'_i, the variance σ'^2_i of $Y_i(t)$ and the lower bound a_i of the flows $X_i(t)$ at each site may be defined. If $\underline{\mu}'$, $\underline{M}'_0$ and $\underline{M}'_1$ are thus defined where $\underline{\mu}'$ is an (m x 1) matrix consisting of the means of the $Y_i(t)$, and A' and B' are calculated using equations (4.4) and (4.5) formulated in terms of $\underline{M}'_0$ and $\underline{M}'_1$, multivariate synthetic sequences will resemble multivariate historic sequences in terms of $\hat{\mu}_i$, $\hat{\sigma}^2_i$, $\hat{\gamma}_i$, $\hat{\rho}_{ij}(0)$ and $\hat{\rho}_{ij}(1)$ [i,j = 1,2,...,m], and the distribution of the flows will be multivariate log-normal.

If, however, the lag-one cross-correlations of the flows are not of interest, needless computation may be avoided by defining

$$\rho'_{ij}(1) = \hat{\rho}'_i(1) \cdot \hat{\rho}_{ij}'(0) \qquad (4.15)$$

Thus, $\underline{A}'$ will be a diagonal matrix with $\hat{\rho}'_i(1)$, [i = 1,2,...,m] along the diagonal and the process will be lag-one Markovian in Y-space. Note, however, that in general

$$\rho_{ij}(1) \neq \hat{\rho}_i(1).\hat{\rho}_{ij}(0)$$

4.2 The Multivariate ARIMA (1,0,1) Process

4.2.1 Formulation of Matrix Equations

Following the sign convention adopted for the univariate ARIMA (1,0,1) process the multivariate ARIMA (1,0,1) process may be formulated as

$$\underline{x}(t) - \underline{A}\,\underline{x}(t-1) = \underline{B}\,\underline{\varepsilon}(t) - \underline{C}\,\underline{\varepsilon}(t-1) \qquad (4.16)$$

where $\underline{x}(t)$ and $\underline{x}(t-1)$ are (m x 1) matrices whose elements are

$x_i(t) = (X_i(t) - \mu_i)$ and $x_i(t-1) = (X_i(t-1) - \mu_i)$ respectively, for $i = 1,2,\ldots,m$, where m is the number of sites. The terms $\underline{\varepsilon}(t)$ and $\underline{\varepsilon}(t-1)$ represent vectors of independent random variables at time points t and (t-1) respectively, and $\underline{A}$,$\underline{B}$ and $\underline{C}$ are (m x m) matrices of coefficients. In the univariate case, μ_i, σ_i, $\rho_i(1)$ and $\rho_i(2)$ are required to define the process; hence $\underline{M}_o$, $\underline{M}_1$ and $\underline{M}_2$ are required to define the elements of the matrices $\underline{A}$, $\underline{B}$ and $\underline{C}$ where $\underline{M}_o$ and $\underline{M}_1$ are as previously defined for the multivariate ARIMA (1,0,0) process and $\underline{M}_2$ is the lag-two covariance matrix defined as

$$\underline{M}_2 = \begin{bmatrix} \rho_1(2)\sigma_1^2 & \cdots & & \\ \cdots & \rho_2(2)\sigma_2^2 & \cdots & \\ & & \cdots & \rho_{ij}(2)\sigma_i\sigma_j \cdots \\ & \cdots \rho_{ji}(2)\sigma_j\sigma_i & \cdots & \\ & & & \cdots \\ & & & \cdots \rho_m(2)\sigma_m^2 \end{bmatrix}$$

However, if the elements of $\underline{x}(t)$ are defined as

$$x_i(t) = \frac{X_i(t) - \mu_i}{\sigma_i}$$

then the matrix $\underline{M}_2$ becomes the lag-two cross-correlation matrix

$$\underline{M}_2 = \begin{bmatrix} \rho_1(2) \cdots & & & \\ \cdots \rho_2(2) \cdots & & & \\ & & \cdots & \rho_{ij}(2) \cdots \\ \cdots \rho_{ji}(2) \cdots & & & \\ & & & \cdots \\ & & & \cdots \rho_m(2) \end{bmatrix}$$

In the following development, the matrices $\underline{M}_o$, $\underline{M}_1$ and $\underline{M}_2$ will,without

loss of generality, be taken to be the lag-zero, lag-one and lag-two correlation matrices. Re-writing equation (4.16) gives

$$\underline{x}(t) = \underline{A}\,\underline{x}(t-1) + \underline{B}\underline{\varepsilon}(t) - \underline{C}\underline{\varepsilon}(t-1) \qquad (4.17)$$

Postmultiplying in equation (4.17) by the (1 x m) matrix $\underline{x}^{T}(t)$ and taking expectations yields

$$E[\underline{x}(t)\,\underline{x}^{T}(t)] = \underline{A}E[\underline{x}(t-1)\,\underline{x}^{T}(t)] + \underline{B}E[\underline{\varepsilon}(t)\,\underline{x}^{T}(t)] - \underline{C}E[\underline{\varepsilon}(t-1)\,\underline{x}^{T}(t)] \qquad (4.18)$$

Defining

$$\underline{M}_{o} = E[\underline{x}(t)\,\underline{x}^{T}(t)] \qquad (4.19)$$

$$\underline{M}_{1} = E[\underline{x}(t)\,\underline{x}^{T}(t-1)] \qquad (4.20)$$

$$\underline{M}_{1}^{T} = E[\underline{x}(t-1)\,\underline{x}^{T}(t)] \qquad (4.21)$$

$$\underline{M}_{2} = E[\underline{x}(t)\,\underline{x}^{T}(t-2)] \qquad (4.22)$$

$$\underline{M}_{2}^{T} = E[\underline{x}(t-2)\,\underline{x}^{T}(t)] \qquad (4.23)$$

and noting that $E[\underline{\varepsilon}(t)\,\underline{\varepsilon}^{T}(t)] = \underline{I}$, where $\underline{I}$ is the identity matrix, the terms on the r.h.s. of equation (4.18) may be evaluated as follows:

$$\underline{A}E[\underline{x}(t-1)\,\underline{x}^{T}(t)] = \underline{A}\,\underline{M}_{1}^{T}$$

The second and third terms on the r.h.s. of equation (4.18) may be evaluated by taking the transpose of equation (4.17) and substituting for $\underline{x}^{T}(t)$ to give

$$\underline{B}E[\underline{\varepsilon}(t)\,\underline{x}^{T}(t)] = \underline{B}\left[E[\underline{\varepsilon}(t)\,\underline{x}^{T}(t-1)]\underline{A}^{T} + E[\underline{\varepsilon}(t)\,\underline{\varepsilon}^{T}(t)]\underline{B}^{T} - E[\underline{\varepsilon}(t)\,\underline{\varepsilon}^{T}(t-1)]\underline{C}^{T}\right]$$

$$= \underline{B}\underline{B}^{T}$$

and

$$\underline{C}E[\underline{\varepsilon}(t-1)\ \underline{x}^{T}(t)] = \underline{C}\Big[E[\underline{\varepsilon}(t-1)\ \underline{x}^{T}(t-1)]\underline{A}^{T} + E[\underline{\varepsilon}(t-1)\ \underline{\varepsilon}(t)]\underline{B}^{T} - E[\underline{\varepsilon}(t-1)\ \underline{\varepsilon}^{T}(t-1)]\underline{C}^{T}\Big]$$

$$= \underline{C}\underline{B}^{T}\underline{A}^{T} - \underline{C}\underline{C}^{T}$$

Hence equation (4.18) reduces to

$$\underline{M}_{o} = \underline{A}\underline{M}_{1}^{T} + \underline{B}\underline{B}^{T} - \underline{C}\underline{B}^{T}\underline{A}^{T} + \underline{C}\underline{C}^{T}$$

or

$$\underline{B}\underline{B}^{T} + \underline{C}\underline{C}^{T} = \underline{M}_{o} - \underline{A}\underline{M}_{1}^{T} + \underline{C}\underline{B}^{T}\underline{A}^{T} \qquad (4.24)$$

Postmultiplying in equation (4.17) by the (1 x m) matrix $\underline{x}^{T}(t-1)$ and taking expectations yields

$$E[\underline{x}(t)\ \underline{x}^{T}(t-1)] = \underline{A}E[\underline{x}(t-1)\ \underline{x}^{T}(t-1)] + \underline{B}E[\underline{\varepsilon}(t)\ \underline{x}^{T}(t-1)] - \underline{C}E[\underline{\varepsilon}(t-1)\ \underline{x}^{T}(t-1)]$$

Hence

$$\underline{M}_{1} = \underline{A}\ \underline{M}_{o} - \underline{C}\underline{B}^{T}$$

or

$$\underline{C}\underline{B}^{T} = \underline{A}\ \underline{M}_{o} - \underline{M}_{1} \qquad (4.25)$$

Postmultiplying in equation (4.17) by the (1 x m) matrix $\underline{x}^{T}(t-2)$ and taking expectations yields

$$E[\underline{x}(t)\ \underline{x}^{T}(t-2)] = \underline{A}E[\underline{x}(t-1)\ \underline{x}^{T}(t-2)] + \underline{B}E[\underline{\varepsilon}(t)\ \underline{x}^{T}(t-2)] - \underline{C}E[\underline{\varepsilon}(t-1)\ \underline{x}^{T}(t-2)]$$

or

$$\underline{M}_{2} = \underline{A}\underline{M}_{1}$$

Hence

$$\underline{A} = \underline{M}_{2}\underline{M}_{1}^{-1} \qquad (4.26)$$

Substituting for $\underline{A}$ in equation (4.25) and for $\underline{C}\underline{B}^T$ and $\underline{A}$ in equation (4.24) yields:

$$\underline{B}\underline{B}^T + \underline{C}\underline{C}^T = \underline{M}_o - \underline{M}_2\underline{M}_1^{-1}\underline{M}_1^T + \underline{M}_2\underline{M}_1^{-1}\underline{M}_o\underline{M}_1^{T-1}\underline{M}_2^T - \underline{M}_1\underline{M}_1^{T-1}\underline{M}_2^T \quad (4.27)$$

Hence the r.h.s. of equation (4.27) may be compiled into a symmetric matrix $\underline{S}$ which is a function of $\underline{M}_o$, $\underline{M}_1$ and $\underline{M}_2$:

$$\underline{B}\underline{B}^T + \underline{C}\underline{C}^T = \underline{S} \quad (4.28)$$

Equation (4.25) provides a further equation for $\underline{C}$ and $\underline{B}$ in terms of $\underline{M}_o$, $\underline{M}_1$ and $\underline{M}_2$ and may be re-written as

$$\underline{C}\underline{B}^T = \underline{T} \quad (4.29)$$

where $\underline{T}$ is a non-symmetric matrix which is a function of $\underline{M}_o$, $\underline{M}_1$ and $\underline{M}_2$. Equations (4.28) and (4.29) must now be solved simultaneously for the matrices $\underline{B}$ and $\underline{C}$. The matrices $\underline{M}_o$, $\underline{M}_1$ and $\underline{M}_2$ provide

$$\frac{m(m+1)}{2} + m^2 + m^2$$

conditions for defining the elements of the matrices $\underline{A}$,$\underline{B}$ and $\underline{C}$. $\underline{A}$ is non-symmetric and may be solved for via equation (4.26), as a result

$$\frac{m(m+1)}{2} + m^2$$

conditions remain to define the elements of $\underline{B}$ and $\underline{C}$. This suggests that a lower triangular form may be assumed for either $\underline{B}$ or $\underline{C}$ but not both, in order to ensure that the process will have $\underline{M}_o$, $\underline{M}_1$ and $\underline{M}_2$ as its correlation matrices.

In order that equations (4.28) and (4.29) may be solved to yield real valued coefficients for $\underline{B}$ and $\underline{C}$ the matrices $\underline{S}$ and $\underline{T}$ must satisfy certain conditions. Extra equations derived from equations (4.28) and (4.29) allow the specification of these conditions:

$$(\underline{B} + \underline{C})(\underline{B} + \underline{C})^T = \underline{B}\underline{B}^T + \underline{C}\underline{C}^T + \underline{C}\underline{B}^T + \underline{B}\underline{C}^T$$

$$= \underline{S} + \underline{T} + \underline{T}^T \qquad (4.30)$$

$$(\underline{B} - \underline{C})(\underline{B} - \underline{C})^T = \underline{S} - \underline{T} - \underline{T}^T \qquad (4.31)$$

Necessary and sufficient conditions that the matrices $(\underline{B} + \underline{C})$ and $(\underline{B} - \underline{C})$ be real valued are that the matrices $(\underline{S} + \underline{T} + \underline{T}^T)$ and $(\underline{S} - \underline{T} - \underline{T}^T)$ be positive semidefinite. Note that $(\underline{S} + \underline{T} + \underline{T}^T)$ and $(\underline{S} - \underline{T} - \underline{T}^T)$ are symmetric. If $(\underline{B} + \underline{C})$ is a real valued matrix and $(\underline{B} - \underline{C})$ is a real valued matrix then

$$(\underline{B} + \underline{C}) + (\underline{B} - \underline{C}) = 2\underline{B}$$

must be a real valued matrix. Similarly

$$(\underline{B} + \underline{C}) - (\underline{B} - \underline{C}) = 2\underline{C}$$

must be a real valued matrix. Consequently, both $\underline{B}\underline{B}^T$ and $\underline{C}\underline{C}^T$ must be positive semidefinite.

4.2.2 Iterative Solution of Matrix Equations

An analytic solution to equations (4.28) and (4.29) can be found through assuming lower triangular forms for the matrices $(\underline{B} + \underline{C})$ and $(\underline{B} - \underline{C})$ in equations (4.30) and (4.31), respectively, which implies lower triangular forms for both $\underline{B}$ and $\underline{C}$. As a result the process will not, in general, preserve the estimated $\underline{M}_o$, $\underline{M}_1$ and $\underline{M}_2$ matrices.

A general analytic matrix solution of equations (4.28) and (4.29) which permits a lower triangular form for either $\underline{B}$ or $\underline{C}$ has not been found possible and an iterative numerical approach has been adopted. Equation (4.29) may be rearranged as:

$$\underline{C} = \underline{T}^T(\underline{B})^{-1} \qquad (4.32)$$

$$\underline{C}^T = \left[(\underline{B})^{-1}\right]^T \underline{T}^T$$

$$= \underline{B}^{-1}\underline{T}^T$$

Substituting for $\underline{C}$ and $\underline{C}^T$ in equation (4.28) gives

$$\underline{B}\underline{B}^T + \underline{T}(\underline{B}^T)^{-1}\,\underline{B}^{-1}\,\underline{T}^T = \underline{S}$$

or

$$\underline{B}\underline{B}^T + \underline{T}(\underline{B}\underline{B}^T)^{-1}\,\underline{T}^T = \underline{S}$$

By letting $\underline{B}\underline{B}^T = \underline{U}$ an iterative solution for $\underline{U}$ may be defined as

$$\underline{U}_j = \underline{S} - \underline{T}\,\underline{U}_{j-1}^{-1}\,\underline{T}^T \qquad (4.33)$$

where $\underline{U}_j$ represents the value of $\underline{U}$ obtained on the j^{th} iteration, and is a function of $\underline{U}_{j-1}$, $\underline{S}$ and $\underline{T}$. An arbitrary form must be assumed for $\underline{U}$ initially; the identity matrix $\underline{I}$ may be used. Convergence occurs when $\underline{U}_j$ stabilizes. The matrix $\underline{U}$ may then be solved for $\underline{B}$ through assuming a lower diagonal form; as already noted $\underline{B}\underline{B}^T$ must be positive semidefinite to obtain a solution. The matrix $\underline{C}$ may then be obtained from equation (4.32).

Alternatively, a substitution may be made for $\underline{B}$ from equation (4.29) into equation (4.28) as follows

$$\underline{B} = \underline{T}^T(\underline{C}^{-1})^T \qquad (4.34)$$

$$\underline{B}^T = \underline{C}^{-1}\,\underline{T}$$

$$\underline{T}^T(\underline{C}^{-1})^T\,\underline{C}^{-1}\,\underline{T} + \underline{C}\underline{C}^T = \underline{S}$$

$$\underline{T}^T(\underline{C}\underline{C}^T)^{-1}\,\underline{T} + \underline{C}\underline{C}^T = \underline{S}$$

By letting $\underline{C}\underline{C}^T = \underline{V}$ an iterative solution for $\underline{V}$ may be defined as

$$\underline{V}_j = \underline{S} - \underline{T}^T\,\underline{V}_{j-1}^{-1}\,\underline{T} \qquad (4.35)$$

$\underline{V}$ may be solved for in a similar fashion to $\underline{U}$ and a lower triangular solution evolved for $\underline{C}$. $\underline{B}$ may then be derived from equation (4.34). Both iterative solutions are equivalent and either $\underline{B}$ or $\underline{C}$ may be assumed to be lower triangular without any loss of generality; in either case the process will have the estimated $\underline{M}_o$, $\underline{M}_1$ and $\underline{M}_2$ as correlation

matrices.

Problems may be encountered in solving either of equations (4.33) or (4.35) for $\underline{U}$ or $\underline{V}$, respectively. The convergence properties of the iterative procedure have been examined for a number of different cases and the following points have been noted:

(i) The convergence properties of equations (4.33) and (4.35) are virtually identical.

(ii) In certain cases neither $\underline{U}$ or $\underline{V}$ may converge to a solution, even though the matrices $(\underline{S} + \underline{T} + \underline{T}^T)$ and $(\underline{S} - \underline{T} - \underline{T}^T)$ are positive definite, and both $\underline{U}$ and $\underline{V}$ may oscillate from iteration to iteration. A damping coefficient λ may be inserted into equation (4.33) or (4.35) to define a new iterative procedure as

$$\underline{U}_j = \underline{S} - \lambda \underline{T}\, \underline{U}_{j-1}^{-1}\, \underline{T}^T \tag{4.36}$$

$$\underline{V}_j = \underline{S} - \lambda \underline{T}^T\, \underline{V}_{j-1}^{-1}\, \underline{T} \tag{4.37}$$

where $0.0 < \lambda < 1.0$

Convergence may now be achieved but the equations which have been solved are

$$\begin{aligned} \underline{B}\underline{B}^T + \lambda \underline{C}\underline{C}^T &= \underline{S} \\ \underline{C}\underline{B}^T &= \underline{T} \end{aligned} \tag{4.38}$$

$$\begin{aligned} \lambda \underline{B}\underline{B}^T + \underline{C}\underline{C}^T &= \underline{S} \\ \underline{C}\underline{B}^T &= \underline{T} \end{aligned} \tag{4.39}$$

whereas a solution for $\lambda = 1$ is required.

In cases where convergence is obtained with $\lambda = 1$, the values obtained for the coefficients of the matrices $\underline{B}$ and $\underline{C}$ may, under

certain conditions, be checked to ensure that the solution is correct and that the required correlations will be preserved. Provided that the preservation of the off-diagonal elements of the estimated $\underline{M}_2$ matrix is not required, the inversion of the estimate of the matrix $\underline{M}_1$ in equation (4.26) is avoided through defining $\underline{A}$ as a diagonal matrix.

$$\underline{A} = \begin{bmatrix} a_{11} & & & & & \\ & a_{22} & & & & \\ & & \ddots & & & \\ & & & \ddots & & \\ & & & & \ddots & \\ & & & & & a_{mm} \end{bmatrix} \qquad (4.40)$$

with elements a_{ii}, $i = 1,2,\ldots,m$ defined as

$$a_{ii} = \hat{\rho}_i(2)/\hat{\rho}_i(1) \qquad (4.41)$$

provided $\hat{\rho}_i(2) < \hat{\rho}_i(1)$. Thus, a_{ii} is in fact the moment estimate of the parameter φ_i of a univariate ARIMA (1,0,1) process at each site. Under these conditions the process will preserve lag-zero, lag-one and lag-two cross-correlations as follows (Appendix 4.1)

$$\overline{\rho}_{ij}(0) = \frac{1}{(1-a_{ii}a_{jj})} \left[-a_{ii} \sum_{k=1}^{m} c_{jk}\, b_{ik} + \sum_{k=1}^{m} b_{ik}\, b_{jk} \right. \qquad (4.42)$$

$$\left. - a_{jj} \sum_{k=1}^{m} c_{ik}\, b_{jk} + \sum_{k=1}^{m} c_{ik}\, c_{jk} \right]$$

$$\overline{\rho}_{ij}(1) = a_{ii}\, \overline{\rho}_{ij}(0) - \sum_{k=1}^{m} c_{ik}\, b_{jk} \qquad (4.43)$$

$$\overline{\rho}_{ij}(2) = a_{ii}\, \overline{\rho}_{ij}(1) \qquad (4.44)$$

Provided the matrices $\underline{B}$ and $\underline{C}$ have been correctly defined by the iterative procedure, $\overline{\rho}_{ij}(0)$, $\overline{\rho}_{ij}(1)$, will equal their corresponding

historic record estimates $\hat{\rho}_{ij}(0)$ and $\hat{\rho}_{ij}(1)$, respectively, while $\bar{\rho}_{ij}(2)$ will equal $\hat{\rho}_{ij}(2)$ only for $i = j$.

As far as modelling estimates of the Hurst coefficient at each site is concerned, a diagonal form for the matrix $\underline{A}$ is convenient. This permits the selection of a value of φ_i at each site such that estimates of $\rho_i(1)$ and h_i are preserved. The procedure recommended in section (3.3.6) for the selection of the φ_i values may be applied at each site on the basis of the observed estimates of $\rho_i(1)$ and h_i.

4.2.3 Analytic solution of the Matrix Equations

A less general solution for the matrices $\underline{B}$ and $\underline{C}$ may be obtained by assuming lower triangular forms for the matrices $(\underline{B} + \underline{C})$ and $(\underline{B} - \underline{C})$ in equations (4.30) and (4.31) respectively. Such a solution entails lower triangular forms in turn for $\underline{B}$ and $\underline{C}$. As a result, only

$$\frac{m(m+1)}{2} + \frac{m(m+1)}{2}$$

conditions are then available to define the

$$m^2 + \frac{m(m+1)}{2}$$

coefficients of the matrices $\underline{B}$ and $\underline{C}$. As a result, the process will have correlation matrices $\bar{\underline{M}}_o$, $\bar{\underline{M}}_1$ and $\bar{\underline{M}}_2$ which, in general, will not equal any arbitrary estimates of $\underline{M}_o$, $\underline{M}_1$ and $\underline{M}_2$ which satisfy conditions for the solution; however, under certain conditions $\bar{\underline{M}}_o$ will equal $\underline{M}_o$, and $\bar{\underline{M}}_1$ and $\bar{\underline{M}}_2$ will have diagonal elements which will equal the corresponding elements of $\underline{M}_1$ and $\underline{M}_2$.

From equations (4.25) and (4.27) the r.h.s. of equations (4.30) and (4.31) may be evaluated as

$$\underline{S} + \underline{T} + \underline{T}^T = \underline{M}_o + \underline{M}_2\underline{M}_1^{-1}\underline{M}_o^T\underline{M}_1^{-1T}\underline{M}_2 + \underline{M}_2\underline{M}_1^{-1}\underline{M}_o$$

$$+ \underline{M}_o\underline{M}_1^{-1}\underline{M}_2 - \underline{M}_1 - \underline{M}_1\underline{M}_1^{-1}\underline{M}_2^T - \underline{M}_1^T - \underline{M}_2\underline{M}_1^{-1}\underline{M}_1^T$$

$$= (\underline{I} + \underline{M}_2\underline{M}_1^{-1})\,\underline{M}_o\,(\underline{I} + \underline{M}_2\underline{M}_1^{-1})^T - \underline{M}_1\,(\underline{I} + \underline{M}_2\underline{M}_1^{-1})^T$$

$$- (\underline{I} + \underline{M}_2\underline{M}_1^{-1})\,\underline{M}_1^T$$

$$= (\underline{I} + \underline{A})\,\underline{M}_o\,(\underline{I} + \underline{A})^T - \underline{M}_1(\underline{I} + \underline{A})^T - (\underline{I} + \underline{A})\,\underline{M}_1^T \qquad (4.45)$$

$$\underline{S} - \underline{T} - \underline{T}^T = \underline{M}_o + \underline{M}_2\underline{M}_1^{-1}\underline{M}_o\underline{M}_1^{-1}\underline{M}_2^T - \underline{M}_2\underline{M}_1^{-1}\underline{M}_o$$

$$- \underline{M}_o\underline{M}_1^{-1}\underline{M}_2 + \underline{M}_1 - \underline{M}_1\underline{M}_1^{-1}\underline{M}_2^T + \underline{M}_1^T - \underline{M}_2\underline{M}_1^{-1}\underline{M}_1^T$$

$$= (\underline{I} - \underline{M}_2\underline{M}_1^{-1})\,\underline{M}_o\,(\underline{I} - \underline{M}_2\underline{M}_1^{-1})^T + \underline{M}_1(\underline{I} - \underline{M}_2\underline{M}_1^{-1})^T$$

$$+ (\underline{I} - \underline{M}_2\underline{M}_1^{-1})\,\underline{M}_1^T$$

$$= (\underline{I} - \underline{A})\,\underline{M}_o\,(\underline{I} - \underline{A})^T + \underline{M}_1(\underline{I} - \underline{A})^T + (\underline{I} - \underline{A})\underline{M}_1^T \qquad (4.46)$$

Equations (4.45) and (4.46) are now in computationally convenient form. Lower diagonal forms may now be assumed for the matrices $(\underline{B} + \underline{C})$ and $(\underline{B} - \underline{C})$, assuming that $(\underline{S} + \underline{T} + \underline{T}^T)$ and $(\underline{S} - \underline{T} - \underline{T}^T)$ are positive semidefinite. Thus,

$$\underline{\beta}\underline{\beta}^T = \underline{S} + \underline{T} + \underline{T}^T \qquad (4.47)$$

and

$$\underline{\beta}^*\underline{\beta}^{*T} = \underline{S} - \underline{T} - \underline{T}^T \qquad (4.48)$$

where $\underline{\beta}$ and $\underline{\beta}^*$ represent the lower diagonal forms assumed for $(\underline{B} + \underline{C})$ and $(\underline{B} - \underline{C})$. The matrices $\underline{B}$ and $\underline{C}$ may then be evaluated; a simple (2 x 2) case illustrates the procedure.

$$\underline{B} + \underline{C} = \begin{bmatrix} b_{11} + c_{11} & - \\ b_{21} + c_{21} & b_{22} + c_{22} \end{bmatrix} = \begin{bmatrix} \beta_{11} & - \\ \beta_{21} & \beta_{22} \end{bmatrix}$$

$$\underline{B} - \underline{C} = \begin{bmatrix} b_{11} - c_{11} & - \\ b_{21} - c_{21} & b_{22} - c_{22} \end{bmatrix} = \begin{bmatrix} \beta^*_{11} & - \\ \beta^*_{21} & \beta^*_{22} \end{bmatrix}$$

The elements of $\underline{B}$ and $\underline{C}$ may now be evaluated through simultaneously solving pairs of equations provided by $(\underline{B} + \underline{C})$ and $(\underline{B} - \underline{C})$ as follows

$$b_{11} + c_{11} = \beta_{11} \tag{4.49}$$

$$b_{11} - c_{11} = \beta^*_{11} \tag{4.50}$$

whence

$$b_{11} = \tfrac{1}{2}[\beta_{11} + \beta^*_{11}]$$

$$c_{11} = \tfrac{1}{2}[\beta_{11} - \beta^*_{11}]$$

The remaining coefficients of $\underline{B}$ and $\underline{C}$ may be solved for in a similar fashion.

If the matrix $\underline{A}$ in equation (4.16) is assumed to be a null matrix the multivariate ARIMA (1,0,1) process reduces to a multivariate ARIMA (0,0,1) process. Such an assumption will not be of great interest insofar as generating multivariate synthetic streamflow sequences is concerned; however, the nature of the restriction imposed on the correlation matrices of the process by lower triangular forms for both B and C can now be illustrated more easily.

(a) The Multivariate ARIMA (0,0,1) Process

The process is defined as

$$\underline{x}(t) = \underline{B}\,\underline{\varepsilon}(t) - \underline{C}\,\underline{\varepsilon}(t-1) \tag{4.51}$$

with the same notation as for the multivariate ARIMA (1,0,1) process. The matrices $\underline{B}$ and $\underline{C}$ must be defined so that the process will have $\underline{M}_o$ and $\underline{M}_1$ as its correlation matrices. The conditions under which $\underline{B}$ and $\underline{C}$ may be solved for are identical to those for the ARIMA (1,0,1) case as specified by the conditions imposed on the matrices $(\underline{S} + \underline{T} + \underline{T}^T)$ and

$(\underline{S} - \underline{T} - \underline{T}^{T})$. From equations (4.45) and (4.46)

$$\underline{S} + \underline{T} + \underline{T}^{T} = \underline{M}_o - \underline{M}_1 - \underline{M}_1^{T} \tag{4.52}$$

$$\underline{S} - \underline{T} - \underline{T}^{T} = \underline{M}_o + \underline{M}_1 + \underline{M}_1^{T} \tag{4.53}$$

where $\underline{S} = \underline{M}_o$ and $\underline{T} = -\underline{M}_1$. If $\underline{B}$ and $\underline{C}$ are solved for using the iterative procedure described for the ARIMA (1,0,1) case then the process will have $\underline{M}_o$ and $\underline{M}_1$ as its correlation matrices. If lower triangular forms are assumed for both $\underline{B}$ and $\underline{C}$ then the process may be shown to have $\underline{M}_o$ as its lag-zero correlation matrix and will have a lag-one correlation matrix $\underline{M}_1^*$ with diagonal elements equal to those of $\underline{M}_1$. A simple (2 x 2) case illustrates this fact. Let

$$\underline{M}_o = \begin{bmatrix} 1 & R \\ R & 1 \end{bmatrix} \tag{4.54}$$

$$\underline{M}_1 = \begin{bmatrix} \rho & \alpha \\ \beta & r \end{bmatrix} \tag{4.55}$$

without any loss of generality. The process may be shown (Appendix 4.2) to preserve the elements of $\underline{M}_o$ as

$$\begin{aligned} \rho_{11}(0) &= 1 \\ \rho_{22}(0) &= 1 \\ \rho_{12}(0) &= \rho_{21}(0) = R \end{aligned} \tag{4.56}$$

and the elements of $\underline{M}_1$ as

$$\begin{aligned} \rho_{11}(1) &= \rho \\ \rho_{22}(1) &= r \end{aligned} \tag{4.57}$$

$$\rho_{12}(1) = \frac{1}{4}\left[(2\beta + 2\alpha) - \frac{(R+\beta+\alpha)\sqrt{1-2\rho}}{\sqrt{1+2\rho}} + \frac{(R-\beta-\alpha)\sqrt{1+2\rho}}{\sqrt{1-2\rho}}\right]$$

$$\rho_{21}(1) = \frac{1}{4}\left[(2\beta + 2\alpha) - \frac{(R-\beta-\alpha)\sqrt{1+2\rho}}{\sqrt{1-2\rho}} + \frac{(R+\beta+\alpha)\sqrt{1-2\rho}}{\sqrt{1+2\rho}}\right] \qquad (4.58)$$

However, note that

$$\rho_{12}(1) + \rho_{21}(1) = \alpha + \beta \qquad (4.59)$$

Consequently the off-diagonal elements of $\underline{M}_1$ will not be individually preserved as sufficient conditions are not available to define all the elements of $\underline{M}_o$ and $\underline{M}_1$. Note that only

$$2\left[\frac{m(m+1)}{2}\right] = 6$$

conditions are imposed on the elements of $\underline{M}_o$ and $\underline{M}_1$. The foregoing results for the (2 x 2) case may be shown to be true for the general (m x m) case. For the multivariate ARIMA (1,0,1) process, it can be shown that lower triangular forms for both $\underline{B}$ and $\underline{C}$ do not permit the preservation of any of the estimated matrices $\underline{M}_o$, $\underline{M}_1$ and $\underline{M}_2$.

(b) A Modified Multivariate ARIMA (1,0,1) Process

Under certain conditions lower triangular forms can be assumed for both $\underline{B}$ and $\underline{C}$, and still yield an acceptable solution. If the matrix $\underline{A}$ is restricted to be a diagonal matrix with a constant, φ, $-1 < \varphi < 1$ along the diagonal,

$$\underline{A} = \begin{bmatrix} \varphi & & & & \\ & \varphi & & & \\ & & \ddots & & \\ & & & \ddots & \\ & & & & \varphi \end{bmatrix} \qquad (4.60)$$

then the process will have $\underline{M}_o$ as its correlation matrix, and will have

a lag-one correlation matrix $\underline{\bar{M}}_1$ with diagonal elements identical to $\underline{M}_1$(Appendix 4.3). Hence, if lag-zero cross-correlations and lag-one serial correlations are only of interest then the matrices $\underline{B}$ and $\underline{C}$ may be solved for without recourse to an iterative procedure. Further, the inverse of $\underline{M}_1$ is not required and the diagonal matrix $\underline{A}$ may be substituted for $\underline{M}_2\underline{M}_1^{-1}$ in equations (4.45) and (4.46).

The definition of the matrix $\underline{A}$ given by equation (4.60) means that the flows at each site are generated by a univariate ARIMA (1,0,1) process, with parameter φ and the preservation of estimates of ρ_1 and h at each site may be attempted using the procedure outlined in section (3.3.6). However, the fact that φ is constant from site to site represents a constraint on the preservation of observed K_i values at each site. From tables(3.2) - (3.7), $\tilde{E}[\hat{\rho}_1]_n$ and φ define $\tilde{E}[K]_n$ uniquely, and as estimates of ρ_1 will vary between sites, the value of $\tilde{E}[K_i]_n$ defined at each site by φ and $\hat{\rho}_i(1)$ may not necessarily be equal to the observed values of K_i at each site. However, reliable estimates of K are not available for individual sequences. If the observed values of K_i at each site are averaged over all sites, an average value $\bar{K}$ is obtained; using tables (3.2) - (3.7) a value of φ could then be chosen such that

$$\bar{K} = \frac{1}{m}\sum_{i=1}^{m} \tilde{E}[K_i]_n \qquad (4.61)$$

which means that an average measure of long-term persistence is preserved over the m sites, together with the estimated $\underline{M}_o$ matrix and lag-one autocorrelations. Such an approach might well be more reliable than modelling individual estimates of h at each site. If, however, the preservation of the observed values of K_i, i = 1,2,...,m, is required, this may be achieved using a constant value of φ, but

the observed estimates of $\rho_i(1)$ will not be exactly preserved. If the observed values of K_i and $\rho_i(1)$ are to be preserved at each site, then different values of φ are required at each site, and the matrices $\underline{B}$ and $\underline{C}$ must be solved for using the iterative procedure outlined in section (4.2.2) to ensure that the estimated $\underline{M}_o$ matrix will be preserved.

4.3 A Multivariate Log-Normal ARIMA (1,0,1) Process

In situations where the distribution of the flows at each site is assumed to be log-normal with lower bound a_i at each site, such that the process $Y_i(t) = \ln(X_i(t) - a_i)$ follows a multivariate normal distribution, the ARIMA (1,0,1) process in terms of $Y_i(t)$ may be written as

$$\underline{y}(t) = \underline{A}'\underline{y}(t-1) + \underline{B}'\underline{\varepsilon}(t) - \underline{C}'\underline{\varepsilon}(t-1) \qquad (4.62)$$

where $\underline{y}(t)$ and $\underline{y}(t-1)$ are vectors of standardized normal variates with elements $y_i(t) = (Y_i(t) - \mu_i')/\sigma_i'$ and $y_i(t-1) = (Y_i(t-1) - \mu_i')/\sigma_i'$ respectively, and μ_i' and σ_i' are the mean and standard deviation of the process $Y_i(t)$. The matrices $\underline{A}'$, $\underline{B}'$ and $\underline{C}'$ must be defined from the correlation matrices $\underline{M}_o'$, $\underline{M}_1'$ and $\underline{M}_2'$ of the process $\underline{Y}(t)$. The elements of the matrices $\underline{M}_o'$, $\underline{M}_1'$ and $\underline{M}_2'$, $\rho_{ij}'(0)$, $\rho_{ij}'(1)$ and $\rho_{ij}'(2)$, respectively, may be defined from the corresponding elements of the matrices $\underline{M}_o$, $\underline{M}_1$ and $\underline{M}_2$ through the following equations (Appendix 4.4).

$$\rho_{ij}(0) = \frac{[\exp(\sigma_i'\sigma_j'\rho_{ij}'(0)) - 1]}{[\exp(\sigma_i'^2) - 1][\exp(\sigma_j'^2) - 1]^{0.5}}^{0.5} \qquad (4.63)$$

$$\rho_{ij}(1) = \frac{[\exp(\sigma_i'\sigma_j'\rho_{ij}'(1)) - 1]}{[\exp(\sigma_i'^2) - 1][\exp(\sigma_j'^2) - 1]^{0.5}}^{0.5} \qquad (4.64)$$

$$\rho_{ij}(2) = \frac{[\exp(\sigma_i' \sigma_j' \rho_{ij}'(2)) - 1]}{[\exp(\sigma_i'^2) - 1][\exp(\sigma_j'^2) - 1]^{0.5}} \quad (4.65)$$

In the general case, estimates of the elements of $\underline{M}_0$, $\underline{M}_1$ and $\underline{M}_2$ may be inserted into equations (4.63-4.65) to yield estimates of the elements of $\underline{M}_0'$, $\underline{M}_1'$ and $\underline{M}_2'$. The quantities μ_i', σ_i' and a_i at each site may be defined from the mean, variance and skewness of $X_i(t)$ as outlined in section (3.3.7). The iterative procedure outlined in section (4.2.2) may then be used to solve for the matrices $\underline{A}'$, $\underline{B}'$ and $\underline{C}'$ in terms of $\underline{M}_0'$, $\underline{M}_1'$ and $\underline{M}_2'$. Synthetic flows will then conform to a multivariate log-normal distribution, and will resemble historic flows in terms of the temporal and spatial statistical properties of the flows, rather than in terms of the properties of the logarithms of the flows.

If a diagonal form is assumed for the matrix $\underline{A}'$, with elements φ_i along the diagonal, then the procedure outlined in sections (3.3.6) and (3.3.7) may be applied to the process $Y_i(t)$ to ensure that both short-term and long-term properties in Y-space are preserved at each site. Attention is thus confined only to preserving estimates of the elements of the matrix $\underline{M}_0$ which can be used to define the elements of $\underline{M}_0'$ from equation (4.63), and the diagonal elements of the matrix $\underline{M}_1$ which can be used to define the diagonal elements of $\underline{M}_1'$ from equation (3.35). Again, the assumption is made that the value of K_i preserved in Y-space at each site will be invariant under the non-linear transformation, which may not be a valid assumption.

4.4 Summary

A review of multisite stochastic models of annual streamflow has shown that no simple multisite model with the ability to model long-term persistence has appeared in the literature. A multisite ARIMA (1,0,1) process has been formulated which allows the preservation of short-term and long-term properties at each site in addition to the appropriate cross-correlations between sites. A simple iterative procedure can be used to solve the matrix equations in the general case. A less general solution can be obtained analytically which should be sufficient for the majority of applications. A multivariate log-normal ARIMA (1,0,1) process has also been developed.

Chapter 5

ESTIMATION OF THE PARAMETERS OF THE ARIMA (1,0,1) PROCESS

The approach outlined in section (3.3.6) represents an indirect method of estimating the parameters φ and θ of the ARIMA (1,0,1) process using estimates of ρ_1 and h. Such an approach seeks to fulfill the aims of synthetic hydrology in attempting to achieve the "correct" statistical resemblance between historic and synthetic sequences in terms of parameters which are thought to exert an important influence on the water resource system design process.

However, in the presence of long-term persistence, estimates of ρ_1 and h are biased, and probably do not yield very efficient estimates of φ and θ. Alternative methods of estimating the parameters of a general ARIMA (p,d,q) process have been proposed by Box and Jenkins (1970) which rely on more classical statistical methods and which are particularly suited for forecasting applications. The Box-Jenkins approach involves a series of techniques for model identification, parameter estimation and diagnostic checking in order to establish the appropriate ARIMA (p,d,q) model for the sample of data. In the context of water resource system design, a good approach would be one yielding (a) the model and (b) parameter estimates for that model which would minimise the expected loss and hence the risk associated with decisions to be taken by a designer on the basis of the observed parameter estimates. Due to the impossibility of assessing whether or not these specifications are satisfied, a general approach such as that of Box and Jenkins (1970) may be adopted which attempts to identify the correct model for the data and then derive maximum likelihood estimates of the parameters of the model.

A selection of observed time series from industry and economics have been employed by Box and Jenkins (1970) to illustrate their methodology. In general, each observed series contained between 200 and 300 observations, which perhaps allows some reliable conclusions to be drawn as to the correct model for the data. However, observed hydrological sequences of annual data, on average, may be only of the order of 30-50 years, and, in the presence of long-term persistence, the information content (Matalas and Langbein, 1962) of such sequences will be rather small. Consequently, the ability of the Box and Jenkins methodology to provide reliable guidance as to the appropriate model for annual streamflow cannot, perhaps, be readily accepted and merits some investigation. In addition, the properties of maximum likelihood estimates from small samples are largely unknown.

Section (5.1) reviews the moment and maximum likelihood parameter estimation techniques employed by Box and Jenkins (1970) and illustrates how such estimates may be obtained for the ARIMA (1,0,1) process. Some of the diagnostic tests for model inadequacy are reviewed in section (5.2) and the problem of type II errors is discussed. Some Monte Carlo sampling experiments are carried out in section (5.3) to investigate the power of some of the diagnostic tests described in section (5.2) and to investigate some of the properties of moment and maximum likelihood parameter estimates in small samples for the ARIMA (1,0,1) process. The results of some sampling experiments with higher order ARIMA models are reported briefly. In section (5.4) the ARIMA (1,0,1) model is fitted to some observed time series of tree ring indices and to some shorter series of annual stream flow data.

5.1 Parameter Estimation Techniques

The parameters of ARIMA models may be estimated using the method of moments or the method of maximum likelihood (ML). In general, moment estimates of parameters are easier to obtain, and do not require any assumption as to the distribution from which the data derives. On the other hand, moment estimates do not, in general, possess the desirable property of asymptotic efficiency which ML estimates possess. Nevertheless, the derivation of maximum likelihood estimates requires an assumption about the underlying distribution, and frequently entails the solution of some complicated non-linear equations.

While the large sample or asymptotic properties of moment and maximum likelihood estimates can generally be derived analytically, the extent to which these properties hold for small sample estimates is largely unknown. Some large sample properties of ML estimates of autoregressive and moving average parameters have been derived by Box and Jenkins (1970) who have shown that moment estimates of autoregressive (AR) parameters closely approximated fully efficient (asymptotically) ML estimates, while moment estimates of moving average (MA) parameters do not. Box and Jenkins suggest using moment estimates of AR and MA parameters as initial approximations to ML estimates which must be solved for using iterative numerical procedures. Little is apparently known of the properties of both moment and ML estimates of AR and MA parameters in small samples. However, from the simulations performed in section (3.3.5) estimates of ρ_1 for the ARIMA (1,0,1) process are biased in the presence of long-term persistence, which implies that estimates of ρ_k ($k > 1$) will also be biased. Consequently, moment estimates of the parameters of the ARIMA (1,0,1) process will be biased in small samples. A similar conclusion cannot as yet be drawn about maximum

likelihood estimates.

ARIMA models have already been applied to a small number of annual streamflow sequences, for the purposes of forecasting (Carlson, McCormick and Watts, 1970), and the validity of using such models for annual streamflow has been questioned (Mandelbrot, 1971b). The contention that forecasting is essentially a short-run problem (Carlson and Watts, 1971) as distinct from the generation of synthetic streamflows, where long-run effects can be important, may not be valid, as small sample estimates of the parameters of ARIMA models can be expected to suffer from the effects of long-term persistence.

5.1.1 Moment Estimates for the ARIMA (1,0,1) Process

Moment estimates of the sets of parameters $\underline{\varphi} = [\varphi_1, \varphi_2, \ldots, \varphi_p]$ and $\underline{\theta} = [\theta_1, \theta_2, \ldots, \theta_q]$ for the stationary ARIMA (p,o,q) or ARMA (p,q) process may be defined from estimates of the (p+q) autocorrelations $\rho_1, \rho_2, \ldots, \rho_{p+q}$. Jenkins and Watts (1968) suggest using the algorithm

$$\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_o} = \frac{\frac{1}{n}\sum_{i=1}^{n-k}(X_t-\bar{X})(X_{t+k}-\bar{X})}{\frac{1}{n}\sum_{i=1}^{n}(X_t-\bar{X})^2} \qquad (5.1)$$

for estimating ρ_k on the grounds that it provides an estimate with smaller mean square error than estimators with (1/n-k) in the numerator of equation (5.1). Having thus derived the estimated autocorrelations $\hat{\rho}_1, \hat{\rho}_2, \ldots, \hat{\rho}_{p+q}$ (or equivalently the estimated autocovariances $\hat{\gamma}_0, \hat{\gamma}_1, \ldots, \hat{\gamma}_{p+q}$) a general two-stage procedure which is essentially based on the two-stage specification of the process in section (3.1.5), is employed by Box and Jenkins (1970) for estimating the elements of the sets $\underline{\varphi}$ and $\underline{\theta}$. The elements of the set $\underline{\varphi}$ are estimated from the estimated autocovariances $\hat{\gamma}_{q-p+1}, \hat{\gamma}_{q-p+2}, \ldots, \hat{\gamma}_{q+1}, \hat{\gamma}_{q+2}, \ldots, \hat{\gamma}_{q+p}$ by

solving the set of p linear equations

$$\begin{aligned}
\hat{\gamma}_{q+1} &= \hat{\varphi}_1\hat{\gamma}_q + \hat{\varphi}_2\hat{\gamma}_{q-1} + \ldots + \hat{\varphi}_p\hat{\gamma}_{q-p+1} \\
\hat{\gamma}_{q+2} &= \hat{\varphi}_1\hat{\gamma}_{q+1} + \hat{\varphi}_2\hat{\gamma}_q + \ldots + \hat{\varphi}_p\hat{\gamma}_{q-p+2} \\
\ldots &\quad \ldots\ldots\ldots\ldots\ldots\ldots \\
\hat{\gamma}_{q+p} &= \hat{\varphi}_1\hat{\gamma}_{q+p-1} + \hat{\varphi}_2\hat{\gamma}_{q+p-2} + \ldots + \hat{\varphi}_p\hat{\gamma}_q
\end{aligned} \tag{5.2}$$

Using estimates of the members of the set $\underline{\varphi}$, the first (q+1) autocovariances $\hat{\gamma}'_k$ (k = 0,1,...,q) of the derived series

$$X'_t = X_t - \hat{\varphi}_1 X_{t-1} \ldots - \hat{\varphi}_p X_{t-p} \tag{5.3}$$

are calculated. The autocovariances $\hat{\gamma}'_0, \hat{\gamma}'_1, \ldots, \hat{\gamma}'_q$ are then used in place of $\hat{\gamma}_0, \hat{\gamma}_1, \ldots, \hat{\gamma}_q$ in equation (3.16) to solve iteratively for the parameter estimates of $\theta_1 \ldots \theta_q$, an estimate of the residual variance, σ_ε^2, also being obtained.

In obtaining the autocovariances $\hat{\gamma}'_0, \hat{\gamma}'_1, \ldots, \hat{\gamma}'_q$ it is not necessary to obtain the derived series X'_t, and subsequently estimate γ'_k, k = 0,1,...,q therefrom as Box and Jenkins (1970) have defined the relations between $\hat{\gamma}_k$ and $\hat{\gamma}'_k$ as:

$$\hat{\gamma}'_k = \sum_{i=0}^{p} \hat{\varphi}_i^2\hat{\gamma}_k + \sum_{i=1}^{p} (\hat{\varphi}_o\hat{\varphi}_i + \hat{\varphi}_1\hat{\varphi}_{i+1} + \ldots + \hat{\varphi}_{p-i}\hat{\varphi}_p)\, d_k \tag{5.4}$$

where k = 0,1,...,q

$$d_k = \hat{\gamma}_{k+i} + \hat{\gamma}_{k-i}$$

$$\varphi_o = -1$$

For the ARIMA (1,0,1) process, equations (5.3) and (5.4) reduce to

$$\hat{\gamma}_2 = \hat{\varphi}\,\hat{\gamma}_1 \tag{5.5}$$

and

$$\begin{aligned}
\hat{\gamma}'_o &= \hat{\gamma}_o(1 + \hat{\varphi}^2) - 2\hat{\varphi}\,\hat{\gamma}_1 \\
\hat{\gamma}'_1 &= \hat{\gamma}_1(1 + \hat{\varphi}^2) - \hat{\varphi}_1(\hat{\gamma}_2 + \hat{\gamma}_o)
\end{aligned} \tag{5.6}$$

respectively where the subscripts on φ_1 and θ_1 have been omitted. The parameter θ is then estimated through solving the following equations iteratively

$$\hat{\sigma}_\varepsilon^2 = \frac{\hat{\gamma}_0}{1 + \hat{\theta}^2}$$
$$\hat{\theta} = \frac{\hat{\gamma}_1}{\hat{\sigma}_\varepsilon^2} \qquad (5.7)$$

The iteration is commenced with $\theta = 0$ and an estimate of σ_ε^2, the residual variance, is obtained. $\hat{\sigma}_\varepsilon^2$ is then substituted into the second equation to yield an updated estimate of θ. The iteration is repeated until convergence takes place. The procedure converges linearly; however, quadratic convergence can be obtained through using a Newton-Raphson algorithm.

Equivalently, equation (3.25) can be solved for the parameter $\hat{\varphi}$ while equation (3.24) then reduces to a quadratic in θ which yields two estimates of $\hat{\theta}$; nevertheless, only one estimate will satisfy the invertibility condition $-1 < \theta < +1$. However, in the case of higher order ARIMA models, the more general two-stage procedure is necessary.

5.1.2 Maximum Likelihood Estimates for the ARIMA (1,0,1) Process

The limiting properties of Maximum Likelihood (ML) estimates are usually established for independent observations (Rao, 1965), but Whittle (1953) has shown that the theory may be extended to cover stationary time series. The likelihood function forms the basis whereby ML estimates may be obtained, and, for n random variables $Z_1, Z_2, \ldots, Z_n$, is their joint density $L(Z_1, Z_2, \ldots, Z_n; \xi_1, \xi_2, \ldots, \xi_k)$ which is a function of the vector of parameters $\underline{\xi} = \xi_1, \xi_2, \ldots, \xi_k$. In particular, if $Z_1, Z_2, \ldots, Z_n$ is a random sample from the density function $f(z;\underline{\xi})$ then the likelihood function is

$$L(Z_1, Z_2,\ldots,Z_n; \underline{\xi}) = f(Z_1;\underline{\xi}).\ f(Z_2;\underline{\xi})\ldots\ f(Z_n;\underline{\xi}) = \prod_{i=1}^{n} f(Z_i;\underline{\xi}) \tag{5.8}$$

and the maximum likelihood estimates of the parameters $(\xi_1, \xi_2,\ldots,\xi_k)$ are obtained by solving the set of k equations

$$\frac{\partial L(\xi_1, \xi_2,\ldots,\xi_k)}{\partial \xi_j} = 0 \qquad (j = 1,2,\ldots,k) \tag{5.9}$$

to yield the estimates $\hat{\xi}_1, \hat{\xi}_2,\ldots,\hat{\xi}_k$. Frequently, it is more convenient to work with the log-likelihood function

$$l(Z_1,Z_2,\ldots,Z_n;\underline{\xi}) = \ln[L(Z_1,Z_2,\ldots,Z_n;\underline{\xi})] \tag{5.10}$$

from which ML estimates of $\xi_1, \xi_2,\ldots,\xi_k$ are obtained by solving the equations

$$\frac{\partial l}{\partial \xi_j} = \frac{\partial(\ln L)}{\partial \xi_j} = \frac{1}{L}\left(\frac{\partial L}{\partial \xi_j}\right) = 0 \qquad (j = 1,2,\ldots,k) \tag{5.11}$$

The values of the ML estimates are those which maximize the probability of obtaining the given sample.

In the case of a stationary ARIMA (p,d,q) process, the parameters to be estimated are the mean μ, the variance σ^2 (or equivalently the residual variance σ_ε^2), the p AR parameters $\varphi_1, \varphi_2,\ldots,\varphi_p$ and the q MA parameters $\theta_1, \theta_2,\ldots,\theta_q$. The mean μ can be estimated as $\bar{X} = \frac{1}{n}\sum_{t=1}^{n} X_t$ and a new sequence of observations defined as

$$W_t = X_t - \bar{X} \tag{5.12}$$

with $E[W_t] = 0$. Alternatively, the mean μ may be included in the vector of parameters $\underline{\xi}$, which otherwise comprises p autoregressive parameters, q moving parameters and σ^2, or equivalently, σ_ε^2, the variance of the residual white noise ε_t. The term ε_t may now be defined as:

$$\varepsilon_t = W_t - \varphi_1 W_{t-1} - \varphi_2 W_{t-2} - \cdots - \varphi_p W_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} \tag{5.13}$$

However, for any particular $\underline{\xi}$, the starting values of ε_t cannot be calculated from equation (5.13), as the p values $W_{t-1}, W_{t-2}, \ldots, W_{t-p}$, denoted as $\underline{W}^*$, and q values $\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots, \varepsilon_{t-q}$, denoted as $\underline{\varepsilon}^*$, will be unknown. However, for a particular choice of $\underline{W}^*$ and $\underline{\varepsilon}^*$, the vector $\underline{\varepsilon}$ of length n may be calculated from equation (5.13). Assuming that the ε_t values are normally distributed the likelihood function is

$$L(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n) \simeq \sigma_\varepsilon^{-n} \exp\left[-\left(\sum_{t=1}^{n} \varepsilon_t^2 / 2\sigma_\varepsilon^2\right)\right] \tag{5.14}$$

whence the log-likelihood function, conditional on the choice $(\underline{W}^*, \underline{\varepsilon}^*)$, is:

$$l^*(\underline{\varphi}, \underline{\theta}, \sigma_\varepsilon) = -n \ln \sigma_\varepsilon - \frac{S^*(\underline{\varphi}, \underline{\theta})}{2\sigma_\varepsilon^2} \tag{5.15}$$

where $\underline{\varphi}$ and $\underline{\theta}$ denote the sets of parameters $\varphi_1, \varphi_2, \ldots, \varphi_p$ and $\theta_1, \theta_2, \ldots, \theta_q$, respectively and

$$S^*(\underline{\varphi}, \underline{\theta}) = \sum_{t=1}^{n} \varepsilon_t^2(\underline{\varphi}, \underline{\theta} \mid \underline{X}^*, \underline{\varepsilon}^*, \underline{\varepsilon}) \tag{5.16}$$

Hence the conditional likelihood function given by equation (5.15) involves the data only through the conditional sum of squares (equation 5.16), and contours of l^* for any fixed value of σ_ε in the space $(\underline{\varphi}, \underline{\theta}, \sigma_\varepsilon)$ are contours of S^*. Thus, values of the elements of $\underline{\varphi}$ and $\underline{\theta}$ which minimize S^* maximize l^*, and the resulting least squares estimates represent ML estimates.

Ideally, the unconditional likelihood function should be used for parameter estimation (Box and Jenkins, 1970), but suitable choices of $\underline{W}^*$ and $\underline{\varepsilon}^*$ allow a sufficient approximation to the unconditional likelihood function for moderate to large n by using the conditional likelihood function. One choice for the elements of $\underline{\varepsilon}^*$ and $\underline{W}^*$ would

be the unconditional expectations of ε_t, and W_t, respectively, which are zero. However, if the values of some of the autoregressive parameters lie near boundaries, then this approximation may not be sufficient. A more reliable procedure is to calculate the values of ε_t from equation (5.13) for ε_{p+1} onwards, setting previous ε_t values to zero. Hence the sum of squares S* will then be derived from n-p values of ε_t but the slight loss of information would be unimportant for long series (Box and Jenkins, 1970). However, for short series, the best approach is to work with the unconditional log-likelihood function.

The unconditional log-likelihood function for a sequence of n observations assumed to have been generated by an ARIMA (p,o,q) process is given as

$$l(\underline{\varphi}, \underline{\theta}, \sigma_\varepsilon) = f(\underline{\varphi}, \underline{\theta}) - n \ln \sigma_\varepsilon - \frac{S(\underline{\varphi}, \underline{\theta})}{2\sigma_\varepsilon^2} \qquad (5.17)$$

where $f(\underline{\varphi},\underline{\theta})$ is a function of $\underline{\varphi}$ and $\underline{\theta}$ (Box and Jenkins, 1970). The unconditional sum of squares function is given as

$$S(\underline{\varphi}, \underline{\theta}) = \sum_{t=-\infty}^{n} \left\{ \varepsilon_t \middle| \underline{\varphi}, \underline{\theta}, \underline{W} \right\} \qquad (5.18)$$

where $\left\{ \varepsilon_t \middle| \underline{\varphi}, \underline{\theta}, \underline{W} \right\} = E\left\{ \varepsilon_t \middle| \underline{\varphi}, \underline{\theta}, \underline{W} \right\}$ denotes the expectation of ε_t conditional on $\underline{\varphi}$, $\underline{\theta}$ and $\underline{W}$, and thus may be abbreviated to $\{ \varepsilon_t \}$. As $f(\underline{\varphi},\underline{\theta})$ is usually unimportant for moderate to large n, contours of $S(\underline{\varphi}, \underline{\theta})$ closely approximate contours of log-likelihood. Hence, least squares estimates obtained through minimizing $S(\underline{\varphi},\underline{\theta})$ in equation (5.18) will usually provide close approximations to ML estimates.

5.1.3 Maximizing the Likelihood Function

A popular procedure in obtaining ML estimates is to set the first derivatives of the log-likelihood function with respect to the parameters equal to zero, and to solve the resulting equations analytically or by

some numerical procedure. However, if the analytic form of the log-likelihood is intractable, recourse must be had to iterative numerical optimization techniques which may or may not necessitate the calculation of derivatives.

While the conditional expectation $\{\varepsilon_t\}$ is linear in the elements of the parameter set $\underline{\varphi}$, it may be shown to be non-linear in the elements of the set $\underline{\theta}$. Consequently, techniques which rely on $S(\underline{\varphi}, \underline{\theta})$ being quadratic in the parameters, such as linear least squares, are not strictly applicable. Box and Jenkins (1970) suggest how $\{\varepsilon_t\}$ may be suitable linearized, and how linear least squares techniques may then be applied iteratively to obtain ML estimates, provided reasonable initial guesses at the parameter values are available. However, more general optimization techniques for finding the greatest or least value of a function without calculating derivatives are now widely available (e.g. Rosenbrock, 1960), and may be readily applied to minimize the function $S(\underline{\varphi}, \underline{\theta})$. In calculating $S(\underline{\varphi}, \underline{\theta})$, conditional expectations are taken in equation (5.13) to yield the values of $\{\varepsilon_t | \underline{\varphi}, \underline{\theta}, \underline{W}\}$. As the recurrence relationship starts with t = 1 and proceeds forward, values W_{-j}, j = 0,1,2 are required to start off the forward recurrence relationship. In order to provide these values, Box and Jenkins define an ARIMA (p,o,q) process using a forward shift operator

$$F W_t = W_{t+1} \tag{5.19}$$

and define the value of the process W_t at time t in terms of a set of random shocks δ_t, δ_{t+1}, δ_{t+2},... whence

$$\varphi(F)W_t = \theta(F)\delta_t \tag{5.20}$$

which may be written in recurrent form as

$$\begin{aligned}\delta_t = {} & W_t - \varphi_1 W_{t+1} - \varphi_2 W_{t+2} - \ldots - \varphi_p W_{t+p} \\ & + \theta_1\delta_{t+1} + \theta_2\delta_{t+2} + \ldots + \theta_q\delta_{t+q}\end{aligned} \tag{5.21}$$

Equation (5.21) is used to provide the values W_{-j} $j = 0,1,2$ needed to start off the forward recurrence defined by equation (5.13). The procedure may be illustrated by considering an example with $n = 12$ successive values given in table (5.1) (Box and Jenkins, 1970).

t	1	2	3	4	5	6	7	8	9	10	11	12
W_t	2.0	0.8	-0.3	-0.3	-1.9	0.3	3.2	1.6	-0.7	3.0	4.3	1.1

Table 5.1

An ARIMA (1,0,1) process with $\varphi = 0.3$ and $\theta = 0.7$ serves to illustrate the calculations for which the recurrence relationships (equations (5.13) and (5.21)) reduce to

$$\varepsilon_t = W_t - 0.3W_{t-1} + 0.7\varepsilon_{t-1} \tag{5.22}$$

$$\delta_t = W_t - 0.3W_{t+1} + 0.7\delta_{t+1} \tag{5.23}$$

respectively. As $W_1, W_2, \ldots, W_n$ are distributed independently of $\delta_0, \delta_{-1}, \delta_{-2}, \ldots$ in equation (5.23) then

$$\{\delta_0\} = \{\delta_{-1}\} = \{\delta_{-2}\} = \ldots = 0$$

The series of observations W_t, t=1, n are first of all entered in table (5.2) and the iteration then proceeds as follows:

(i) Starting p steps from the end of the series, where $p = 1$, calculate $\delta_{11}, \delta_{10}, \ldots, \delta_1$. δ_{12} is set to zero, while $\delta_0, \delta_{-1}, \delta_{-2}$ are zero by definition.

(ii) Equation (5.23) is now rearranged as

$$W_t = \delta_t + 0.3W_{t+1} - 0.7\delta_{t+1} \tag{5.24}$$

whence values $\{W_0\}, \{W_{-1}\}, \{W_{-2}\}, \ldots, \{W_{-m}\}$ may be calculated until $\{W_{-m}\}$ has become sufficiently small.

(iii) Using equation (5.22) values $\varepsilon_{-m}, \varepsilon_{-m+1}, \varepsilon_{-m+2}, \ldots, \varepsilon_0, \varepsilon_1, \ldots, \varepsilon_n$ are calculated, assuming that $\{W_{-m}\}$ is effectively zero, which means that

$\{\varepsilon_{-j}\} \approx 0$ for $j > m - 1$. The unconditional sum of squares $S(\underline{\varphi}, \underline{\theta})$ $= S(\varphi,\theta) = S(0.3,0.7)$ is obtained through summing the squares of all the calculated $\{\varepsilon_t\}$ values. Thus, from table (5.2)

$$S(\varphi,\theta) = S(0.3,0.7) = \sum_{t=-4}^{12} \{\varepsilon_t\}^2 = 89.2 \qquad (5.25)$$

The successive entries in table (5.2) corresponding to steps (i) - (iii) have been suitable labelled.

t	$\{\varepsilon_t\}$	$\{W_t\}$	$\{\delta_t\}$
-4	-0.01	-0.01	0
-3	-0.04	-0.03	0
-2	-0.11	-0.09	0
-1	-0.36	-0.31	0
0	-1.20	-1.04	0
1	1.47	2.00	2.34
2	1.23	0.80	0.83
3	0.32	-0.30	-0.08
4	0.02	-0.30	0.18
5	-1.80	-1.90	-0.13
6	-0.39	0.30	2.66
7	2.84	3.20	4.74
8	2.63	1.60	2.89
9	0.66	-0.70	1.54
10	3.67	3.00	4.49
11	5.95	4.30	3.97
12	3.99	1.10	

(ii) (iii) (i)

Table 5.2

In general, the foregoing cycle will be sufficient to define $S(\varphi,\theta)$ accurately. However, the convergence of the procedure may be checked by applying a second iterative cycle.

(i) Using the value of $\{\varepsilon_{12}\} = 3.99$ computed from the previous cycle, and noting that $\{\varepsilon_j\}$, $j = n+1, n+2, \ldots, n+m$ equal zero as $\{\varepsilon_j\}$, $j = n+1, \ldots$ are independently distributed of $\underline{W}$, equation (5.24) may be used to calculate $\{W_j\}$, $j = n+1, n+2, \ldots, n+m$ until the $\{W_j\}$ values are essentially zero for $j > m - 1$.

(ii) Using equation (5.23) δ_{n+m} , δ_{n+m-1} , $\ldots$, δ_1 may then be evaluated

recurrently noting again that $\{\delta_{-j}\} = 0$ for $j = 1,2,\ldots$ The sum of squares may then be defined as

$$S(0.3,0.7) = \sum_{t=1}^{18} \delta_t^2 = 89.3 \qquad (5.26)$$

which compares closely with the value obtained from the first cycle. Table (5.3) contains the entries corresponding to the second cycle.

t	$\{\varepsilon_t\}$	$\{W_t\}$		$\{\delta_t\}$	
0				0	
1		2.00		2.34	
2		0.80		0.83	
3		-0.30		-0.09	
4		-0.30		0.17	
5		-1.90		-0.14	
6		0.30		2.64	
7		3.20		4.72	
8		1.60		2.86	(ii)
9		-0.70		1.49	
10		3.00		4.42	
11		4.30		3.89	
12	3.99*	1.10		-0.14	
13	0	-2.46		-2.84	
14	0	-0.74		-0.85	
15	0	-0.22	(i)	-0.26	
16	0	-0.07		-0.08	
17	0	-0.02		-0.03	
18	0	-0.01		-0.01	

* Initial entry from first cycle

Table 5.3

Box and Jenkins (1970) advocate graphical plotting and subsequent inspection of the likelihood function as important steps in obtaining ML estimates, and warn against the danger of obtaining incorrect results if an automated procedure is blindly followed in minimizing the sum of squares function $S(\underline{\varphi}, \underline{\theta})$. In certain situations, the likelihood function may have multiple maxima (Box and Draper, 1965), and may exhibit other characteristics such as sharp ridges and spikes. Multiple optima of approximately equal height suggest that a number of sets of values exist

for the parameters which might explain the data. Oblique ridges suggest interaction between the parameters, where parameter values considerably different from their ML values may correspond to practically identical values of the likelihood function.

5.1.4 Variances of ML Estimates

Expressions for the large-sample variances and covariances of ML estimates for the parameters of ARIMA (p,o,q) models have been derived by Box and Jenkins (1970). Assume that the log-likelihood function is quadratic, for example, over a 95% confidence region, and that a vector $\underline{\alpha}$ is defined to contain the (p+q) autoregressive and moving average parameters contained in the sets $\underline{\varphi}$ and $\underline{\theta}$. Thus, the ARIMA (p,o,q) model is, apart from the mean, which may be estimated beforehand, completely specified by the (p+q+1) parameters $\underline{\alpha}$ and σ_ε or, as previously denoted, $\underline{\xi}$. Then $l(\underline{\xi})$ may be expanded as

$$l(\underline{\xi}) = l(\underline{\alpha},\sigma_\varepsilon) \simeq l(\hat{\underline{\alpha}},\sigma_\varepsilon) + \tfrac{1}{2} \sum_{i=1}^{k} \sum_{j=1}^{k} d_{ij}(\alpha_i-\hat{\alpha}_i)(\alpha_j-\hat{\alpha}_j) \qquad (5.27)$$

where k = (p+q). Under the assumption that $l(\underline{\xi})$ is quadratic, the derivatives

$$d_{ij} = \frac{\partial^2 l(\underline{\alpha},\sigma_\varepsilon)}{\partial\alpha_i\,\partial\alpha_j} \qquad (5.28)$$

are constant. For large n, the influence of the term $f(\underline{\varphi},\underline{\theta})$ in equation (5.17) may be ignored, and $l(\underline{\alpha},\sigma_\varepsilon)$ will be approximately quadratic in $\underline{\alpha}$ if $S(\underline{\alpha})$ is. As already noted, $S(\underline{\alpha})$ will be essentially quadratic in $\underline{\alpha}$ if the conditional expectations $\{\varepsilon_t \mid \underline{\alpha}, \underline{W}\}$ are approximately locally linear in the elements of $\underline{\alpha}$.

The (k x k) matrix $I(\underline{\alpha})$ whose elements are defined as $-E[d_{ij}]$ is referred to as the information matrix for the parameters $\underline{\alpha}$, where the expectation is taken over the distribution of $\underline{W}$. For a given value of σ_ε,

the variance-covariance matrix $V[\hat{\underline{\alpha}}]$ for the ML estimates for large samples is given by

$$V[\hat{\underline{\alpha}}] \approx - [\underline{D}]^{-1} \qquad (5.29)$$

where, from equation (5.17), the elements of $\underline{D}$ are approximated by

$$E[d_{ij}] \simeq - \frac{E[R_{ij}]}{2\sigma_\varepsilon^2} \qquad (5.30)$$

where

$$R_{ij} = \frac{\partial^2 S(\underline{\alpha}, \underline{W})}{\partial \alpha_i \, \partial \alpha_j} \qquad (5.31)$$

An estimate of σ_ε^2 is provided by

$$\sigma_\varepsilon^2 = \frac{S(\hat{\underline{\alpha}})}{n} \qquad (5.32)$$

while, within large samples, estimates of σ_ε^2 and $\underline{\alpha}$ are uncorrelated. Consequently, the variance-covariance matrix may now be estimated as

$$V[\hat{\underline{\alpha}}] \simeq 2\sigma_\varepsilon^2 [\underline{R}]^{-1} \qquad (5.33)$$

Hence, approximate confidence regions may be calculated from the standard errors of the parameters. For the particular case of the ARIMA (1,0,1) process, the variance-covariance matrix may be shown to be (Box and Jenkins, 1970)

$$V[\hat{\varphi}, \hat{\theta}] \simeq \frac{1 - \varphi\theta}{n(\varphi - \theta)^2} \begin{bmatrix} (1-\varphi^2)(1-\varphi\theta) & (1-\varphi^2)(1-\theta^2) \\ (1-\varphi^2)(1-\theta^2) & (1-\theta^2)(1-\varphi\theta) \end{bmatrix} \qquad (5.34)$$

The large-sample variances of $\hat{\varphi}$ and $\hat{\theta}$ are thus

$$\mathrm{Var}[\hat{\varphi}] = \frac{(1-\varphi\theta)^2 (1-\varphi^2)}{n(\varphi - \theta)^2} \qquad (5.35)$$

$$\mathrm{Var}[\hat{\theta}] = \frac{(1-\varphi\theta)^2 (1-\theta^2)}{n(\varphi - \theta)^2} \qquad (5.36)$$

5.2 Goodness-of-fit Tests

Once a stochastic model has been fitted to a time series through estimating the elements of the set of parameters $\underline{\xi}$, goodness-of-fit tests may be applied to check the adequacy of the model. In the context of ARIMA models, this procedure is called diagnostic checking by Box and Jenkins (1970). A wide variety of goodness of fit tests have been developed for stochastic models, with tests for the adequacy of an autoregressive scheme receiving considerable attention (Quenouille, 1949 ; Whittle, 1952). Diagnostic checking is frequently applied to the residual term ε_t, which, if the model is adequate, should be completely random. Some tests are now considered here, a number of which have been employed by Box and Jenkins (1970).

5.2.1 The Anderson Test

A well known test for randomness of a time series is Anderson's (1942) test of significance based on the estimated lag-one autocorrelation coefficient, $\hat{\rho}_1$. For a random normal process, $\hat{\rho}_1$ is normally distributed with mean $-\frac{1}{(n-1)}$ and variance $\frac{n-2}{(n-1)^2}$. Consequently, confidence limits for $\hat{\rho}_1$ are given by

$$C.L.(\hat{\rho}_1) = [-1 \pm Z_\alpha \sqrt{n-2}\,]/(n-1) \qquad (5.37)$$

where Z_α is a standard normal deviate corresponding to a probability level α. If $\hat{\rho}_1$ falls outside the confidence limits, then the hypothesis that $\rho_1 = 0$ is not accepted. As the test is based on ρ_1, it is predisposed towards detecting departures from randomness attributable to high frequency effects, and does not represent an adequate test for low frequency effects. In the presence of low frequencies, estimates of ρ_1 will be severely biased, and consequently the power of the Anderson test will be low as evidenced by a large percentage of type II errors (Wallis and Matalas, 1971).

5.2.2 Cumulative Periodogram Test

A test based on the integrated periodogram derived through a harmonic analysis of the residual term ε_t has been given by Box and Jenkins (1970). The periodogram of a time series ε_t, (t = 1, ..., n) where n is odd, may be defined as

$$I(f_i) = \frac{n}{2}(a_i^2 + b_i^2) \tag{5.38}$$

where i = 1,2, ..., r and r = (n-1)/2. The harmonic coefficients a_i and b_i are given by

$$a_i = \frac{2}{n}\sum_{t=1}^{n} \varepsilon_t \cos 2\pi f_i t$$

$$b_i = \frac{2}{n}\sum_{t=1}^{n} \varepsilon_t \sin 2\pi f_i t \tag{5.39}$$

where $f_i = i/n$. If n = 2r is even, then equations (5.38) and (5.39) apply for i = 1,2, ..., (r-1), but

$$a_r = \frac{1}{n}\sum_{t=1}^{n} (-1)^t \varepsilon_t$$

$$b_r = 0 \tag{5.40}$$

and

$$I(f_r) = I(0.5) = na_r^2 \tag{5.41}$$

The summation of the periodogram ordinates $I(f_i)$ is related to the sample variance S^2 estimated with a divisor of $(1/n)$ as

$$nS^2 = \sum_{t=1}^{n} (\varepsilon_t - \bar{\varepsilon})^2 = \sum_{i=1}^{r} I(f_i) \tag{5.42}$$

where $\bar{\varepsilon} = \frac{1}{n}\sum_{t=1}^{n} \varepsilon_t$

As the power spectrum p(f) for white noise has a constant value $2\sigma_\varepsilon^2$ over the frequency range 0 - 0.5 cycles, the cumulative spectrum for white noise may be defined as

$$P(f) = \int_0^f p(g)dg \qquad (5.43)$$

and is a straight line joining (0,0) and $(0.5, \sigma_\varepsilon^2)$. Hence a standardized cumulative spectrum may be defined as $P(f)/\sigma_\varepsilon^2$ and forms a line joining (0,0) and (0.5,1). By analogy, the cumulative periodogram $C(f_j)$ with range (0 - 1) may be defined as

$$C(f_j) = \frac{\sum_{i=1}^{j} I(f_i)}{nS^2} \qquad (5.44)$$

If the correct model were known together with the population values of the parameters, then the series ε_t would constitute a pure white noise and a plot of $C(f_j)$ against f_j would be scattered around a line joining the points (0,0) and (0.5,1). Systematic deviations from this line would represent a suggestion of model inadequacy. Even if the correct model were applied, only estimates of the parameters will be available and the ε_t values obtained, $\hat{\varepsilon}_t$, will themselves represent estimates of ε_t. However, for large samples, the periodogram of the ε_t will have similar properties to those of the ε_t (Box and Jenkins, 1970).

The probability relationship between the cumulative periodogram $C(f_j)$ and the integrated spectrum $P(f)/\sigma_\varepsilon^2$ is identical to that between the empirical cumulative frequency function and the cumulative distribution function. Hence a Kolmogorov-Smirnov test may be applied to $C(f_j)$ to detect departures from the theoretical form $P(f)/\sigma_\varepsilon^2$. However, the test is approximate as only $\hat{\varepsilon}_t$ values rather than ε_t values may be calculated, even if the correct model were known. Confidence limits may be placed on the line defined by $P(f)/\sigma_\varepsilon^2$ such that for a purely random series,

$C(f_j)$ would cross them $\alpha\%$ of the time. The confidence limits are drawn at a distance $\pm K_\alpha / g$ where $g = (n-2)/2$ for n even and $g = (n-1)/2$ for n odd. Values of K_α are given by Box and Jenkins (1970).

5.2.3 Autocorrelation Test

The estimated autocorrelation function ρ_k, $k = 0,1, \ldots$ may be used to check departures from randomness in the ε_t values after a model has been fitted to an observed series. However, the sampling properties of the $\hat{\rho}_k$ values are difficult to quantify, as neighbouring estimates $\hat{\rho}_k$ will in general be highly correlated. Only if the parameter values of the correct model are known exactly will the calculated ε_t values constitute a white noise series, whence the $\hat{\rho}_k$, $k = 1,2, \ldots$ will be uncorrelated and approximately normally distributed with mean zero and variance $1/n$, and hence, standard error $1/\sqrt{n}$. However, as only estimates of the ε_t will be available, and the correct model will be unknown, the application of confidence limits based on a standard error $1/\sqrt{n}$ applied to the $\hat{\rho}_k$ values may be inappropriate. Durbin (1969) has shown that if the ε_t are derived from an AR(1) process with parameter φ, the variance of $\hat{\rho}_1$ is φ^2/n which may be considerably less than $1/n$, and the use of $1/\sqrt{n}$ as a standard error could underestimate the variability of the $\hat{\rho}_k$, particularly for low k.

If the first K autocorrelations $\hat{\rho}_k$ ($k = 1,2, \ldots, K$) are derived from a series of ε_t values obtained by fitting any ARIMA (p,o,q) process to an observed sequence of length n, and if the fitted model is appropriate, the quantity

$$Q = n \sum_{k=1}^{K} \hat{\rho}_k^2 \qquad (5.45)$$

is approximately distributed as χ^2 $(K - p - q)$ (Box and Jenkins, 1970).

If the model is inappropriate then the average values of Q will tend to be inflated. Hence an approximate test may be applied be testing observed Q values against the appropriate percentage points of the χ^2 distribution. One of the apparent problems associated with the test is the choice of a suitable value for K.

5.2.4 Type II Errors and Power of Tests

The tests for randomness of the ε_t values discussed in the preceeding sections (5.2.1), (5.2.2) and (5.2.3) may all be applied at some preselected level of significance. Associated with each test is a power, defined as $1 - \omega$ where ω is the proportion type II errors. In general, as α increases, ω decreases and vice versa. Usually, ω cannot be determined analytically for any particular test. As type II errors with respect to autocorrelation may prove more costly in hydrologic applications than type I errors, a value of α should ideally be chosen such that the expected loss resulting from either a type I or type II error is minimal. Invariably, a value of 0.05 is chosen for α in hydrologic studies, but such a level may not be warranted in applying tests for detecting long-term persistence, where the economic losses associated with type II errors may be high.

The power of the tests already described may be assessed through sampling experiments. The ARIMA (1,0,1) process is of particular interest, and a large number of sequences exhibiting varying degrees of persistence and of lengths usually encountered in hydrology may be generated, and ω estimated as the proportion of times that a test fails to detect the existence of persistence.

5.3 Sampling Experiments

5.3.1 Power of Tests for Independence

Initially, some small sample experiments were carried out in order to assess the performance of the cumulative periodogram (CP) test presented

in section (5.2.2) in detecting the non-randomness which characterizes long-term persistence. The simple and widely used Anderson (A) (1942) test for independence was used as a basis for comparison, and the relative powers of the two tests compared as follows.

For sequences of length 25, 50 and 100, 1000 realizations were generated by an ARIMA (1,0,1) process with $\varphi = 0.90$ and for each of a range of values of θ yielding a suitable coverage of values of ρ_1 in the range 0.0 to 0.9. Each sequence was subjected to the A and CP tests both applied at a 5% level of significance, and the proportion of sequences identified as non-random noted for each value of ρ_1. Hence the proportion of type II errors may be obtained and plotted against ρ_1. The A test was found to have greater power for the sample sizes considered; this is illustrated for sample sizes 25 and 100 in figure (5.1). Consequently, in the presence of long-term persistence, the CP test performs poorly. This latter finding would appear to be in line with those of Mandelbrot and Wallis (1969b) who found that the sample spectrum was rather insensitive to long-term persistence. As previously noted by Wallis and Matalas (1971) the Anderson test lacks power because estimates of ρ_1 are biased downwards in addition to the fact that ρ_1 measures high frequency behaviour only. Nevertheless, in this instance, the Anderson test, with its computational simplicity is apparently superior to the cumulative periodogram test which involves greater computation. The CP test may well perform better in detecting other types of non-randomness.

The effect of sample size on type II errors for the A test is more clearly illustrated in fig. (5.2) for an ARIMA (1,0,1) process with $\varphi = 0.90$, which may be considered as a moderately persistent process. However, for small samples, the proportion of type II errors is high, even for moderately large values of ρ_1, while for small values of ρ_1 the power of the test is almost zero, with the proportion of type II errors

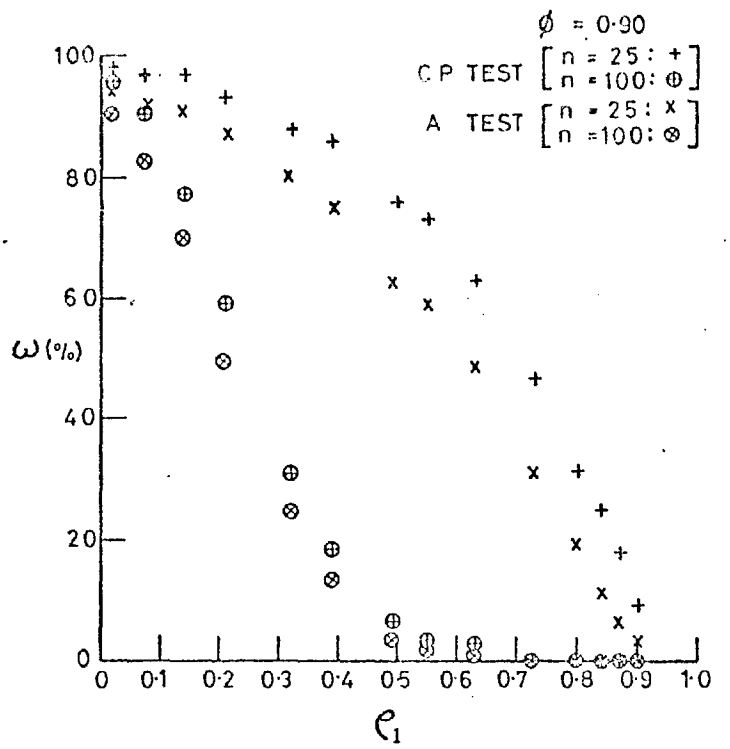


Figure (5.1) Power of CP and A tests, expressed as the percentage of type II errors in 1000 samples of size n, with n = 25 and 100, for an ARIMA (1,0,1) process with $\varphi = 0.90$. The tests were applied at a level of significance of $\alpha = 0.05$.

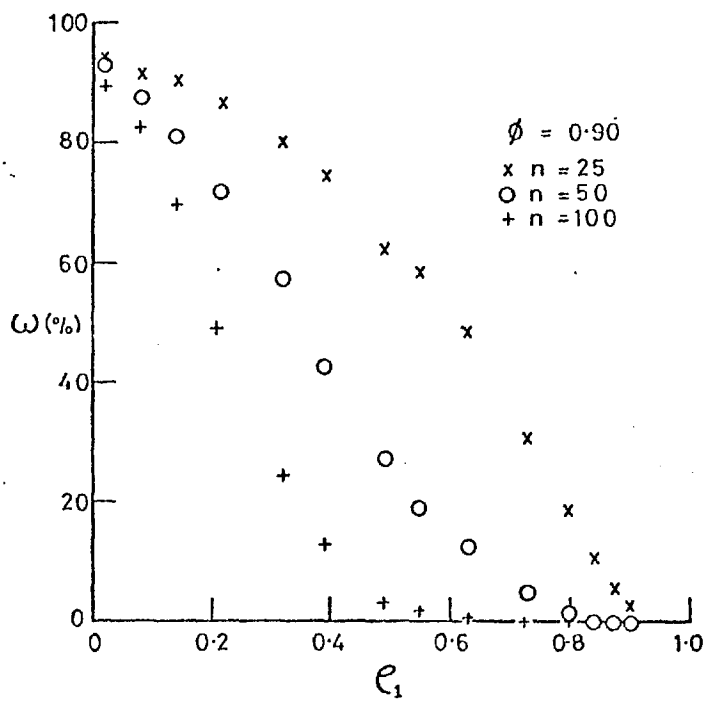


Figure (5.2) Power of A test, expressed as the percentage of type II errors in 1000 samples of size n, with n = 25, 50 and 100, for an ARIMA (1,0,1) process with $\varphi = 0.90$. The test was applied at a level of significance of $\alpha = 0.05$.

approaching $(1 - \alpha)$. For a given value of ρ_1, a decrease in φ represents a decrease in long-term persistence with a consequent decrease in the proportion of type II errors as illustrated in figure (5.3). Type II errors for the ARIMA (1,0,1) process will always exceed those for the lag-one Markov process for a given value of ρ_1, the latter process being the limiting case of the ARIMA (1,0,1) with $\varphi = \rho_1$.

Figures (5.4)-(5.6) illustrate the effect of level of significance, α, on type II errors for the ARIMA (1,0,1) process. As is to be expected, an increase in the level of significance leads to a decrease in type II errors. In applying tests of significance, hydrologists tend to adhere to a level of significance of 0.05, with little regard for what the effect of such a level may have on the decision-making process. Suppose that an observed record is available which could have come from an independent process ($\rho_1 = 0$) or, for example, an ARIMA (1,0,1) process with a particular value of ρ_1. If, when tested for randomness, the record is pronounced persistent when $\rho_1 = 0$, then a type I error with an associated loss l_1 results. Likewise, if the record is pronounced random when $\rho_1 > 0$, then a type II error with an associated loss l_2 ensues. If type I errors occur with probability p_1 and, if type II errors occur with probability p_2, then the overall expected losses $E(l_1)$ and $E(l_2)$ associated with type I and type II errors are $p_1 l_1$ and $p_2 l_2$, respectively, and are represented schematically as a function of α in figure (5.7).

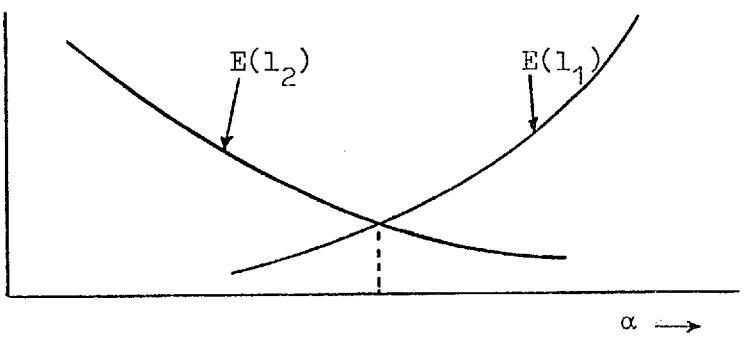


Figure (5.7)

A value of α for which the expected loss associated with both types of error is minimal would appear to be desirable.

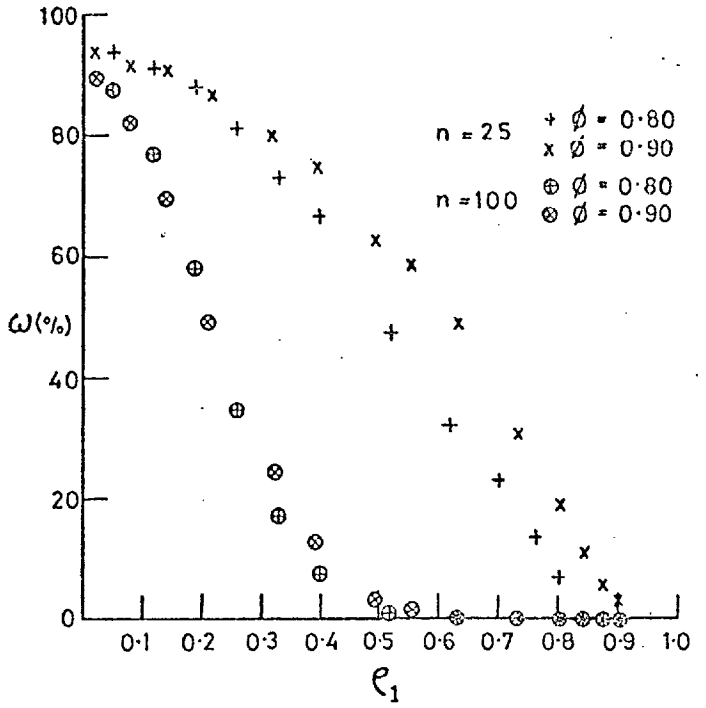


Figure (5.3) Power of A test, expressed as the percentage of type II errors in 1000 samples of size n with n = 25 and 100, for an ARIMA (1,0,1) process with φ = 0.90 and 0.80. The test was applied at a level of significance of α = 0.05.

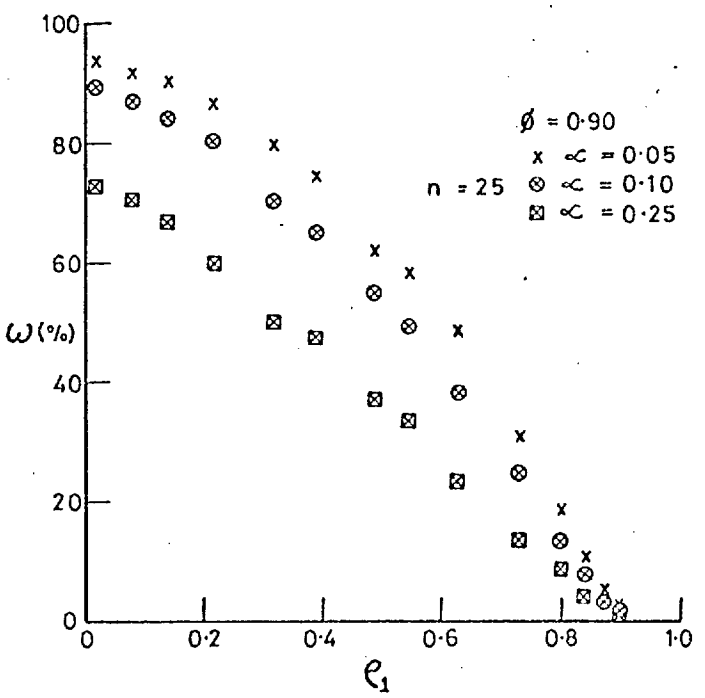


Figure (5.4) Power of A test, expressed as the percentage of type II errors in 1000 samples of size 25 for an ARIMA (1,0,1) process with φ = 0.90. The test was applied at levels of significance of α = 0.05, 0.10 and 0.25.

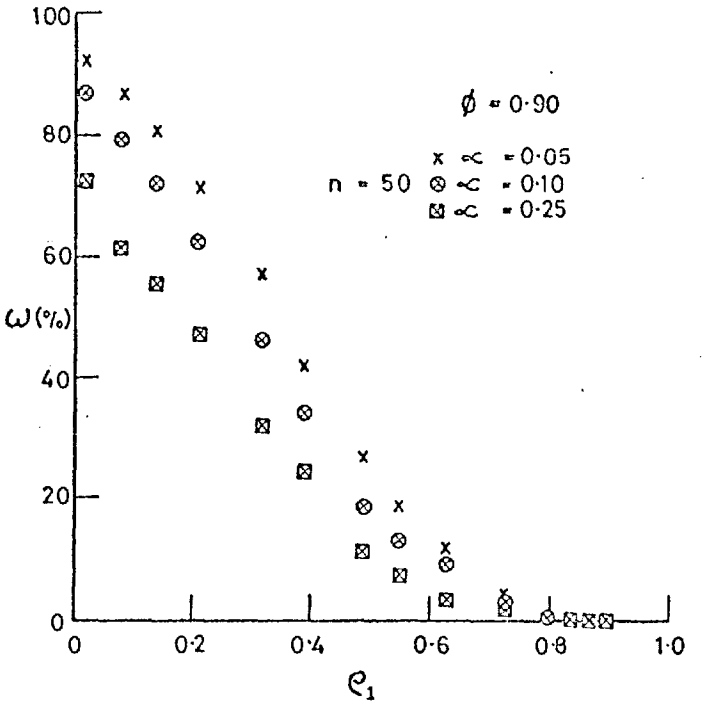


Figure (5.5) Power of A test, expressed as the percentage of type II errors in 1000 samples of size 50 for an ARIMA (1,0,1) process with $\varphi = 0.90$. The test was applied at levels of significance of $\alpha = 0.05$, 0.10 and 0.25.

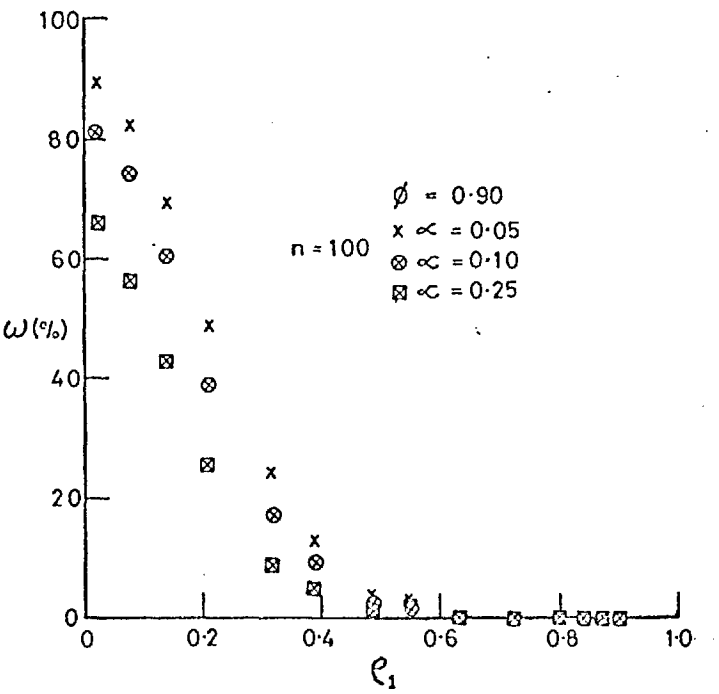


Figure (5.6) Power of A test, expressed as the percentage of type II errors in 1000 samples of size 100 for an ARIMA (1,0,1) process with $\varphi = 0.90$. The test was applied at levels of significance of $\alpha = 0.05$, 0.10, and 0.25.

5.3.2 Moment and ML Estimates of φ and θ from Synthetic Data

The purpose of the experiments performed herein was to study the performance of the moment and ML techniques discussed in sections (5.1.1) and (5.1.2) when applied to sample sizes of less than 100 generated by an ARIMA (1,0,1) process with various values of φ and θ. Initially, it was hoped to obtain some of the sampling properties of moment and ML parameter estimates, in order to assess the applicability of some of the "large sample" results derived theoretically by Box and Jenkins (1970) and given in section (5.1.4). In the face of the large variability synonymous with long-term persistence, the excessive computation time required for ML estimation prevented such an assessment, as a sufficient number of realizations could not be generated to obtain a stable estimate of the expectation and standard error of each parameter. However, some of the difficulties associated with deriving moment and ML estimates from small samples were revealed, and some evidence of the ability of a particular model to identify itself using the Box-Jenkins methodology was obtained.

The moment estimates of the parameters φ and θ denoted as $\hat{\varphi}^{(m)}$ and $\hat{\theta}^{(m)}$ were used to provide initial estimates for the iterative ML technique, and the initial value of the unconditional sum of squares $S(\varphi,\theta)$ was computed as outlined in section (5.1.3). The derivation of ML estimates, denoted as $\hat{\varphi}^{(1)}$ and $\hat{\theta}^{(1)}$ involves a two-dimensional search for the minimum of the function $S(\varphi,\theta)$. A number of general non-linear estimation algorithms exist, such as that of Marquardt (1963) which has been used by Box and Jenkins (1970). A reliable technique, due to Rosenbrock (1960), which does not require the calculation of derivatives, was, however, employed here. Preference for the Rosenbrock technique resulted from previous experience with the algorithm, which performed

satisfactorily on some difficult problems (O'Connell et al., 1970). A brief description of the algorithm is given in Appendix (5.1).

As a stationary, invertible ARIMA (1,0,1) process was used to generate synthetic samples, the search for the minimum of the function $S(\varphi,\theta)$ was confined within the limits defined for the parameters by the stationarity and invertibility conditions resulting in a constrained search in 2-dimensional space. Suitable transformations were employed (Appendix 5.2) which enabled solutions in the constrained space to be derived. Further numerical problems encountered in obtaining ML solutions are also discussed in Appendix (5.2). The computer program was initially checked through reproducing ML parameter estimates found by Box and Jenkins (1970) for a number of time series from business and industry.

(a) Sampling experiments

For the experiments involving the ARIMA (1,0,1) process, two sets of parameter values were chosen, and these are given in table (5.4)

Set	φ	θ
1	0.90	0.80
2	0.80	0.50

Table 5.4

For each set, 500 realizations of size n were generated with n = 25, 50, 100 and 250. The moment estimation procedure described in section (5.1.1) was then applied to each sequence, and assuming that estimates within the admissible parameter space were found, maximum likelihood estimates were derived using moment estimates as starting values for the non-linear search routine used to minimize $S(\varphi,\theta)$. If the moment estimates did not lie within the admissible parameters space, either an ML solution was not sought, or the following strategy was followed.

Assume that a moment estimate of φ with absolute value greater than unity has been found. Nevertheless, the possibility that an ML solution exists cannot be ruled out. However, if a value of φ equal to or greater than unity is used in computing the sum of squares function $S(\varphi,\theta)$, the values $\{W_0\}, \{W_{-1}\}, \{W_{-2}\}, \ldots, \{W_{-m}\}$ will form a divergent series. Hence, the search must at all times be confined within the limits $-1 + \varepsilon < \varphi < 1-\varepsilon$, where ε is some small positive quantity. Details of the limits imposed are given in Appendix (5.2). Consequently, any search for an ML solution must commence within these limits. If an admissible moment estimate was not available for either φ or θ, a starting value with absolute value $|1-\delta|$ where $\delta > \varepsilon$, was chosen for initiating the non-linear search, with the sign of the initial value determined from the sign of the out-of-bounds moment estimate. Details of this rather empirical procedure are given in Appendix (5.2). In addition, if the ML solution was found to lie on or near a boundary, the solution was rejected as detailed in Appendix (5.2).

(b) Results of sampling experiments

Tables (5.5) - (5.9) give the results of the sampling experiments carried out with the ARIMA (1,0,1) process. Table (5.5) gives the number of moment and maximum likelihood estimates of φ and θ obtained from 500 realizations for parameter set 1, together with the means and variances of those estimates. The percentage of moment solutions found ranges from approximately 50% for sample size 25, to 58% for sample size 250. Figures (5.8) and (5.9) illustrate one of the reasons why only a percentage of solutions may be expected. The parameter space $-1 < \varphi < 1$, $-1 < \theta < 1$ illustrated in figure (5.8) maps into the corresponding space for ρ_1 and ρ_2 in figure (5.9). However, in random sampling, moment estimates of ρ_1 and ρ_2 may range between +1 and -1.

n	25	50	100	250
No. of Mom. Sol.	247	249	261	291
No. of ML Sol.	188	209	240	288
Mean $\hat{\varphi}^{(m)}$	0.186	0.224	0.364	0.528
Variance $\hat{\varphi}^{(m)}$	0.219	0.219	0.193	0.126
M.s.e. $\varphi^{(m)}$	0.729	0.676	0.480	0.264
Mean $\hat{\varphi}^{(1)}$	0.205	0.293	0.420	0.677
Variance $\hat{\varphi}^{(1)}$	0.234	0.251	0.244	0.158
M.s.e. $\hat{\varphi}^{(1)}$	0.717	0.619	0.474	0.208
Mean $\hat{\theta}^{(m)}$	0.098	0.072	0.227	0.389
Variance $\hat{\theta}^{(m)}$	0.203	0.194	0.186	0.124
M.s.e. $\hat{\theta}^{(m)}$	0.696	0.724	0.514	0.293
Mean $\hat{\theta}^{(1)}$	0.062	0.129	0.287	0.555
Variance $\hat{\theta}^{(1)}$	0.171	0.220	0.238	0.159
M.s.e. $\hat{\theta}^{(1)}$	0.716	0.670	0.501	0.219
Mean $\hat{\rho}_1^{(m)}$	0.096	0.164	0.152	0.161
Variance $\hat{\rho}_1^{(m)}$	0.072	0.034	0.017	0.006
M.s.e. $\hat{\rho}_1^{(m)}$	0.074	0.035	0.017	0.006
Mean $\hat{\rho}_1^{(1)}$	0.158	0.192	0.164	0.163
Variance $\hat{\rho}_1^{(1)}$	0.096	0.041	0.018	0.006
M.s.e. $\hat{\rho}_1^{(1)}$	0.096	0.044	0.019	0.007

Table (5.5) Properties of moment and ML estimates for 500 samples of size n from an ARIMA (1,0,1) process with $\varphi = 0.90$, $\theta = 0.80$ and $\rho_1 = 0.14$. If a moment solution was not found, an ML solution was not sought.

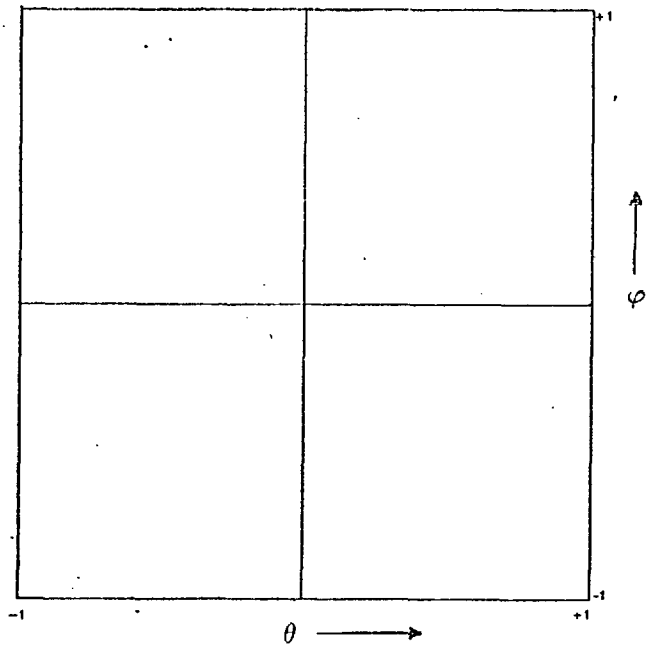


Figure (5.8) Parameter space for a stationary invertible ARIMA (1,0,1) process.

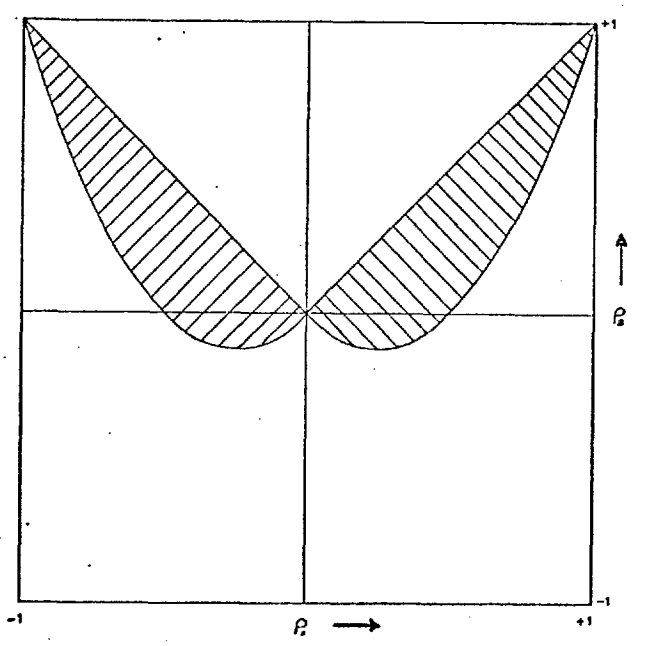


Figure (5.9) Feasible region (shown hatched) for ρ_1 and ρ_2 for the ARIMA (1,0,1) process with parameter space $-1 < \theta < +1$, $-1 < \varphi < +1$.

Hence, correlation constraints exist for the ARIMA (1,0,1) process which manifest themselves particularly strongly for small samples; as the sample size increases, the sample estimates tend to converge on their population values which of course, lie within the constraints. While ρ_1 and ρ_2 are not estimated explicitly by the ML technique, the correlation constraints still operate as the admissible parameter space for φ and θ still exists.

The fitting problem is aggravated whenever $|\theta|$ approaches $|\varphi|$. In this situation, the generated series are approaching white noise, for which case estimation of φ and θ is obviously impossible. When $\varphi = 0.90$ and $\theta = 0.80$ (corresponding to $\rho_1 = 0.14$) the high frequency fluctuations overwhelm the low frequency movements, which tend to emerge only for very large samples. In addition, values of φ close to positive unity tend to yield sequences which will be classified as non-stationary, as there is no clear division between stationary and non-stationary behaviour when autoregressive parameters lie near stationarity boundaries.

From table (5.5), the mean values of $\hat{\varphi}^{(m)}$, $\hat{\theta}^{(m)}$, $\hat{\varphi}^{(1)}$ and $\hat{\theta}^{(1)}$ illustrate that both moment and ML estimates are biased. Convergence of the sample estimates to their respective population values is slow because of the closeness of θ to φ. The mean $\hat{\varphi}^{(1)}$ and $\hat{\theta}^{(1)}$ values display smaller biases than the mean values of $\hat{\varphi}^{(m)}$ and $\hat{\theta}^{(m)}$, with the bias in $\hat{\varphi}^{(1)}$ and $\hat{\theta}^{(1)}$ disappearing more rapidly with increasing sample size. Occasional reversals of this general trend may be explained by the lack of stability of the estimates.

Inspection of the estimated variances of moment and ML estimates in table (5.5) shows that ML estimates tend to have marginally larger variances which is in contrast with the asymptotic properties of ML estimates. While this result might be explained by the fact that the estimated variances of the ML estimates is based on a different (and

samller) number of estimates than the corresponding moment figure for the smaller sample sizes, the fact that the estimated variances for sample size 250 are essentially based on the same number of estimates tends to reject such a hypothesis. Nevertheless, it should be remembered that the results cannot be considered stable, and that the sampling experiments had as a principal aim the assessment of the ability of the moment and ML fitting techniques to fit the "correct" model. The mean and variance of the estimated lag-one autocorrelations derived from moment and ML estimates are also given in table (5.5) and these values tend to reflect the biases and variances of the parameter estimates themselves. The fact that some of the mean values of $\hat{\rho}_1^{(m)}$ and $\hat{\rho}_1^{(1)}$ tend to be larger than the population value of $\rho_1 = 0.14$ may be explained by the preference of the ARIMA (1,0,1) process for fitting positive correlations as evidenced by figure (5.9). However, the sampling distribution of $[\hat{\rho}_1]_n$ must of necessity encompass a large proportion of negative values in this case.

In situations where decisions are to be based on the estimated parameter values, the estimation procedure which minimises the expected loss accruing from the decisions is desirable. In cases where the loss function is quadratic and defined, for example, for the parameter θ as $(\theta - \hat{\theta})^2$, then the estimator which minimises the expected loss is that which minimises the mean square error (m.s.e) $E(\theta - \hat{\theta})^2$. The m.s.e. may be expanded as

$$E(\theta - \hat{\theta})^2 = \text{Var}(\hat{\theta}) + [\theta - E(\theta)]^2 \qquad (5.46)$$

which is the sum of the variance of the estimate and bias squared. The m.s.e. has been computed for the moment and ML parameter estimates in table (5.5), showing that ML estimates of φ and θ yield smaller m.s.e. than moment estimates. The value of m.s.e. is a convenient indicator in situations where consideration of bias and variance together cannot lead to any clear recommendation for an estimator. However, only

tentative conclusions can be drawn here, as the values of m.s.e. in table (5.5) are not, again, strictly comparable because of the different number of moment and ML estimates used for computing the values of m.s.e.

The results for parameter set 2 are given in table (5.6). As the value of φ is not as close to the boundary as for set 1 of parameter values, and the difference between φ and θ is larger, the fitting situation has improved with a larger proportion of solutions. Again, ML solutions are apparently not always available when moment solutions are. The bias in the ML estimates is again smaller than that for the moment estimates but by a relatively smaller amount. In general, the convergence to the population parameters is better over the range of sample size considered, while the overall variability of estimates is less than in the case of parameter set 1. However, in contrast to the results for set 1, the variances of the ML estimates are smaller than those of the moment estimates, apart from occasional reversals in this general trend, which may be attributed to instability. In general, the values of m.s.e. are smaller for the ML estimates than for the moment estimates. The estimates of ρ_1 derived from the parameter estimates $\hat{\varphi}^{(1)}$ and $\hat{\theta}^{(1)}$ have apparently slightly larger variance than those derived through the moment estimates $\hat{\varphi}^{(m)}$ and $\hat{\theta}^{(m)}$.

(c) Diagnostic checks

While the Box-Jenkins methodology for model fitting consists of identification, estimation and diagnostic checking, only the estimation procedures have been used so far, the correct model being known a priori. In these circumstances, diagnostic checking may be considered superfluous, as the residuals should closely resemble white noise. The low power of the cumulative periodogram test in the presence of long-term persistence has already been noted, but the test may well be useful in detecting departures from randomness which result from high

n	25	50	100	250
No. of Mom. Sol.	319	357	394	460
No. of ML Sol.	270	339	391	460
Mean $\hat{\varphi}(m)$	0.359	0.544	0.655	0.751
Variance $\hat{\varphi}(m)$	0.186	0.105	0.058	0.018
M.s.e. $\hat{\varphi}(m)$	0.375	0.171	0.079	0.020
Mean $\hat{\varphi}(1)$	0.420	0.550	0.673	0.768
Variance $\hat{\varphi}(1)$	0.159	0.120	0.055	0.011
M.s.e. $\hat{\varphi}(1)$	0.303	0.183	0.071	0.012
Mean $\hat{\theta}(m)$	0.057	0.227	0.357	0.455
Variance $\hat{\theta}(m)$	0.174	0.116	0.075	0.030
M.s.e. $\hat{\theta}(m)$	0.370	0.191	0.095	0.032
Mean $\hat{\theta}(1)$	0.062	0.199	0.362	0.465
Variance $\hat{\theta}(1)$	0.137	0.125	0.065	0.016
M.s.e. $\hat{\theta}(1)$	0.329	0.216	0.084	0.017
Mean $\hat{\rho}_1(m)$	0.314	0.364	0.367	0.388
Variance $\hat{\rho}_1(m)$	0.048	0.023	0.012	0.006
M.s.e. $\hat{\rho}_1(m)$	0.055	0.024	0.013	0.006
Mean $\hat{\rho}_1(1)$	0.380	0.407	0.387	0.396
Variance $\hat{\rho}_1(1)$	0.060	0.027	0.013	0.006
M.s.e. $\hat{\rho}_1(1)$	0.060	0.027	0.013	0.006

Table (5.6) Properties of moment and ML estimates for 500 samples of size n from an ARIMA (1,0,1) process with $\varphi = 0.80$ $\theta = 0.50$ and $\rho_1 = 0.40$. If a moment solution was not found, an ML solution was not sought.

frequency or periodic movements.

The autocorrelation and cumulative periodogram tests reported on in section (5.2) were incorporated into the experiment reported in table (5.6). Hence, for each sample for which an admissible ML solution was found, both tests were applied to the final set of residuals at a 5% level of significance. If either test yielded a significant result then the ML solution was rejected; obviously such a test has larger power than either test applied individually. The results are given in table (5.7), with the moment results quoted in table (5.6) presented for comparison purposes. Comparison of table (5.7) and table (5.6) shows that the tests reduce only slightly the percentage of ML solutions found, as might be expected. As the level of significance of the overall test should be greater than 0.05, then the percentage of rejections is somewhat smaller than would be expected due to chance alone. However, a larger number of samples would be necessary before any firm conclusions could be drawn.

(d) ML estimates when moment estimates unobtainable

Rather than abandon the search for an ML estimate when a moment solution was not found, a new set of ML results were generated through starting the search for an ML solution in the neighbourhood of the boundary (or boundaries) which the unacceptable moment solution lay closest to. If the search for an ML solution terminated too close to the boundary, the result was rejected. For details of the procedure reference can be made to Appendix (5.2). Table (5.8) gives the results for this strategy for parameter set 1, together with the moment results already presented in table (5.5), but reproduced here for comparison purposes. The proportion of ML solutions found has now increased significantly, ranging from 66% to 99% over sample size. While the bias in the mean $\hat{\theta}^{(1)}$

n	25	50	100	250
No. of Mom. Sol.	319	351	394	460
No. of ML Sol.	266	333	381	452
Mean $\hat{\varphi}^{(m)}$	0.359	0.544	0.655	0.751
Variance $\hat{\varphi}^{(m)}$	0.186	0.105	0.058	0.018
M.s.e. $\hat{\varphi}^{(m)}$	0.375	0.171	0.079	0.020
Mean $\hat{\varphi}^{(l)}$	0.425	0.550	0.673	0.767
Variance $\hat{\varphi}^{(l)}$	0.160	0.121	0.055	0.011
M.s.e. $\hat{\varphi}^{(l)}$	0.301	0.184	0.071	0.012
Mean $\hat{\theta}^{(m)}$	0.057	0.227	0.357	0.455
Variance $\hat{\theta}^{(m)}$	0.174	0.116	0.075	0.030
M.s.e. $\hat{\theta}^{(m)}$	0.370	0.191	0.095	0.032
Mean $\hat{\theta}^{(l)}$	0.063	0.199	0.362	0.465
Variance $\hat{\theta}^{(l)}$	0.138	0.127	0.064	0.016
M.s.e. $\hat{\theta}^{(l)}$	0.329	0.218	0.083	0.017
Mean $\hat{\rho}_1^{(m)}$	0.314	0.364	0.367	0.388
Variance $\hat{\rho}_1^{(m)}$	0.048	0.023	0.012	0.006
M.s.e. $\hat{\rho}_1^{(m)}$	0.055	0.024	0.013	0.006
Mean $\hat{\rho}_1^{(l)}$	0.383	0.408	0.386	0.396
Variance $\hat{\rho}_1^{(l)}$	0.060	0.027	0.013	0.006
M.s.e. $\hat{\rho}_1^{(l)}$	0.060	0.027	0.013	0.006

Table (5.7) Properties of moment and ML estimates for 500 samples of size n from an ARIMA (1,0,1) process with $\varphi = 0.80$, $\theta = 0.50$ and $\rho_1 = 0.40$. If a moment solution was not found, an ML solution was not sought ; if an ML solution was found, tests were applied to the residuals.

n	25	50	100	250
No. of Mom. Sol.	247	249	261	291
No. of ML Sol.	332	415	456	494
Mean $\hat{\varphi}^{(m)}$	0.186	0.224	0.364	0.528
Variance $\hat{\varphi}^{(m)}$	0.219	0.219	0.193	0.126
M.s.e. $\hat{\varphi}^{(m)}$	0.729	0.676	0.480	0.264
Mean $\hat{\varphi}^{(1)}$	0.183	0.283	0.414	0.699
Variance $\hat{\varphi}^{(1)}$	0.288	0.345	0.335	0.170
M.s.e. $\hat{\varphi}^{(1)}$	0.802	0.726	0.571	0.210
Mean $\hat{\theta}^{(m)}$	0.098	0.072	0.227	0.389
Variance $\hat{\theta}^{(m)}$	0.203	0.194	0.186	0.124
M.s.e. $\hat{\theta}^{(m)}$	0.696	0.724	0.514	0.293
Mean $\hat{\theta}^{(1)}$	0.065	0.154	0.305	0.590
Variance $\hat{\theta}^{(1)}$	0.236	0.297	0.300	0.159
M.s.e. $\hat{\theta}^{(1)}$	0.776	0.714	0.545	0.203
Mean $\hat{\rho}_1^{(m)}$	0.096	0.164	0.152	0.161
Variance $\hat{\rho}_1^{(m)}$	0.072	0.034	0.017	0.006
M.s.e. $\hat{\rho}_1^{(m)}$	0.074	0.035	0.017	0.006
Mean $\hat{\rho}_1^{(1)}$	0.137	0.157	0.143	0.153
Variance $\hat{\rho}_1^{(1)}$	0.079	0.035	0.019	0.007
M.s.e. $\hat{\rho}_1^{(1)}$	0.079	0.035	0.019	0.007

Table (5.8) Properties of moment and ML estimates for 500 samples of size n from an ARIMA (1,0,1) process with $\varphi = 0.90$, $\theta = 0.80$ and $\rho_1 = 0.14$. If a moment solution was not found the search for an ML estimate was started in the neighbourhood of a boundary as described in section (5.3.2).

and $\hat{\varphi}^{(1)}$ has not changed appreciable, the variance of the ML estimates has increased appreciably over the figures in table (5.5) as reflected by the estimated variances. Hence, while the procedure guarantees that a larger proportion of ML solutions will be found, the solutions are apparently more variable. The values of m.s.e. for $\hat{\varphi}^{(1)}$ and $\hat{\theta}^{(1)}$ are also larger than those for $\hat{\varphi}^{(m)}$ and $\hat{\theta}^{(m)}$. The reasons for this increased variability have not been closely investigated. Multiple minima for $S(\varphi,\theta)$ might exist as observed by Box and Draper (1965) but this would appear unlikely as the samples were known to be generated by an ARIMA (1,0,1) process. Some close inspection of the behaviour of the function $S(\varphi,\theta)$ over the admissible parameter space when moment solutions do not exist may provide some insight.

As in the case of paramater set 1, a search for an ML solution was initiated even when a moment solution was not available for parameter set 2, and the results, together with the corresponding moment results from table (5.6) are presented in table (5.9). The biases in the ML results have decreased slightly over those shown in table (5.6). While the variability of the estimates has again increased, the increase over the corresponding figures in table (5.6) tends to disappear with increasing sample size. As a result, moment estimates may be more efficient for sample sizes 25 and 50 but less efficient for sample sizes 100 and 250. Indeed, apart from sample size 25, the values of m.s.e. for $\hat{\varphi}^{(1)}$ and $\hat{\theta}^{(1)}$ in table (5.9) are less than the corresponding figures in table (5.6), and, apart from occasional reversals which might be attributed to sampling variability, are always less than the moment values. However, tentative conclusions can only be drawn because of the disparity in the number of estimates contributing to the values of m.s.e.

n	25	50	100	250
No. of Mom. Sol.	319	351	394	460
No. of ML Sol.	377	481	494	500
Mean $\hat{\varphi}^{(m)}$	0.359	0.544	0.655	0.751
Variance $\hat{\varphi}^{(m)}$	0.186	0.105	0.058	0.018
M.s.e. $\hat{\varphi}^{(m)}$	0.375	0.171	0.079	0.020
Mean $\hat{\varphi}^{(1)}$	0.403	0.580	0.697	0.772
Variance $\hat{\varphi}^{(1)}$	0.212	0.132	0.058	0.010
M.s.e. $\hat{\varphi}^{(1)}$	0.370	0.180	0.069	0.011
Mean $\hat{\theta}^{(m)}$	0.057	0.227	0.357	0.455
Variance $\hat{\theta}^{(m)}$	0.174	0.116	0.075	0.030
M.s.e. $\hat{\theta}^{(m)}$	0.370	0.191	0.095	0.032
Mean $\hat{\theta}^{(1)}$	0.099	0.265	0.405	0.475
Variance $\hat{\theta}^{(1)}$	0.181	0.135	0.067	0.017
M.s.e. $\hat{\theta}^{(1)}$	0.342	0.190	0.076	0.018
Mean $\hat{\rho}_1^{(m)}$	0.314	0.364	0.367	0.388
Variance $\hat{\rho}_1^{(m)}$	0.048	0.023	0.012	0.006
M.s.e. $\hat{\rho}_1^{(m)}$	0.055	0.024	0.013	0.006
Mean $\hat{\rho}_1^{(1)}$	0.334	0.380	0.373	0.391
Variance $\hat{\rho}_1^{(1)}$	0.068	0.029	0.016	0.007
M.s.e. $\hat{\rho}_1^{(1)}$	0.072	0.029	0.017	0.007

Table (5.9) Properties of Moment and ML estimates for 500 samples of size n from an ARIMA (1,0,1) process with $\varphi = 0.80$, $\theta = 0.50$ and $\rho_1 = 0.40$. If a moment solution was not found, the search for an ML estimate was started in the neighbourhood of a boundary as described in section (5.3.2).

5.3.3 Moment and ML Estimates for Higher Order ARIMA Models

A limited number of experiments similar to those reported in section (5.3.2) were conducted for the ARIMA (1,0,2) and (2,0,1) processes. Sets of parameters were chosen such that the autocorrelation function decayed slowly. Where possible, moment and ML estimates lying within the admissible parameter space were derived from synthetic sequences of length 25, 50, 100 and 250. In general, the fitting problem was found to be more acute than for the ARIMA (1,0,1) process, with smaller numbers of moment and ML solutions found under similar conditions. This suggests that, for moment estimates, the correlation constraints on the estimates of γ_0, γ_1, γ_2 and γ_3 imposed by the stationarity and invertibility conditions probably become more restrictive as the order of the model increases. It could be expected that ML estimates would also be more difficult to obtain.

In order to gain some insight into the effects of over and under-parameterization in fitting models, some incorrect model fitting was carried out. An ARIMA (2,0,1) model was fitted to sequences generated by itself and by an ARIMA (1,0,1) model, while an ARIMA (1,0,1) model was fitted to sequences generated by an ARIMA (2,0,1) model. For these experiments, an ML solution was not sought if no moment estimate was available. The results obtained when the simpler ARIMA (1,0,1) model was fitted to sequences generated by the ARIMA (2,0,1) model showed that a much larger number of both moment and ML solutions were found than when the ARIMA (2,0,1) model was fitted to sequences generated by itself. Hence, for small samples, the data will generally tend to suggest the simpler model.

When the ARIMA (2,0,1) model was fitted to sequences generated by the ARIMA (1,0,1) model, the number of moment and ML solutions obtained was considerably less than the number obtained in fitting the ARIMA (1,0,1)

model to sequences generated by itself. This suggests that, in small samples, the simplest model has a better chance of fitting the data, irrespective of the underlying model; such a result is perhaps, to be expected. In the experiments where the incorrect model was fitted to the generated sequences, no diagnostic tests were applied to the residuals because of the excessive computation time involved; however, it seems unlikely that the CP or Anderson test would indicate any non-randomness in the residuals because of the apparent low power of these tests in small samples.

5.4 Moment and ML Estimates of φ and θ from Historic Data

In the light of the results obtained in section (5.3.2), it appears that little can be learned about the applicability of the ARIMA (1,0,1) process to observed data if sample sizes less than 100 in length are available. Unfortunately, on an annual time scale, the majority of streamflow records rarely span more than 50 years. Nevertheless, the fact that streamflow records are sometimes spatially correlated suggests that if an ARIMA model is found to fit a sequence of observations from one stream, the same model may be found to adequately describe other streams with similar characteristics and governed by related climatic regimes. Hence, better results may be obtained in terms of model fitting than suggested by random sampling experiments of the type described in section (5.3.2). Further, the physical mechanism of streamflow suggests that positive values of ρ_1 are to be expected because of storage effects, which would favour the fitting of the ARIMA (1,0,1) model.

While streamflow records in excess of 100 years in length are rare, some considerably longer records of related geophysical phenomena are available. Rainfall fluctuations are probably closest to those of runoff but long rainfall records are also rather rare. Tree ring growth, for

example, is less closely related to runoff, but nevertheless can be expected to reflect long-term climatic movements. Long records of tree ring indices have been compiled and studied (Schulman, 1956; Matalas, 1962).

In applying the ARIMA (1,0,1) model to real data, 10 records of annual streamflow were abstracted from Yevjevich (1963), while ten records of tree ring indices were abstracted from Schulman (1956). The streamflow records ranged in length from 70 years to 150 years, while the length of the records of tree ring indices ranged from 301 to 669 years.

5.4.1 Results from Annual Streamflow Data

The values of mean annual discharge for the 10 selected rivers were abstracted from Yevjevich (1963). The name of each river, together with the name of the gauging station, the period of record, the record length, the catchment area in km^2 and the average annual discharge in m^3/s over the period of record are given in table (5.10). The records were selected primarily because of their length, and because they represent a wide range of hydrological regimes. The values of mean annual discharge given by Yevjevich (1963) represents natural or virgin average flow over a water year as far as possible. An index number has been assigned to each river in table (5.10) for ease of reference.

Estimates of γ_0, γ_1 and γ_2 were derived from each record using equation (5.1) and moment estimates of the parameters φ and θ of the ARIMA (1,0,1) model were then derived from equations (5.5) and (5.7). When moment estimates lay within the admissible parameter space, they were used as initial estimates of φ and θ in the search for the ML estimates, which were derived by minimising $S(\varphi,\theta)$ as reported in section (5.3.2). In the case of records where moment estimates lay outside the parameter space, the procedure described in section (5.3.2) and in Appendix (5.2) was used in an attempt to locate ML estimates within the admissible parameter space. Finally, the Anderson test described in section (5.2.1)

No.	River	Station	Period	Length	Catchment Area(km^2)	Mean Ann. Dis(m^3/s)
1	Gota, Sweden	Sjotorp-Vanersburg	1807-1957	150	46816.7	535.7
2	Nemunas, U.S.S.R.	Smalininkai	1811-1943	132	80030.7	545.1
3	St. Lawrence, U.S.A.	Ogdensburg	1860-1957	97	764565.0	6817.6
4	Tennessee, U.S.A.	Chattanooga	1874-1956	82	55425.8	1044.0
5	Neva, U.S.S.R.	Petrokrepost	1859-1935	76	271949.0	2589.3
6	Dnieper, U.S.S.R.	Dnieperpetrovsk	1881-1954	74	416988.4	1611.0
7	Goulburn, Australia	Murchison	1881-1954	73	10722.6	89.9
8	Kiewa, Australia		1885-1957	72	1165.5	20.7
9	Thames, England	Teddington	1883-1954	71	9873.0	62.9
10	Dal, Sweden	Norslund	1852-1922	70	24889.8	346.8

Table (5.10) Details of annual streamflow records selected for analysis from Yevjevich (1963). The records are tabulated in order of decreasing length.

was used to check the residuals for randomness.

The results of the model fitting are given in table (5.11). The index number assigned to each record in table (5.10) is used for reference purposes. The moment estimates of ρ_1 and ρ_2 are also presented. The confidence limits on $\hat{\rho}_1$ used to apply the Anderson test to the residuals, may also be applied to the estimates of ρ_1 for each of the observed records, suggesting that only records 1,2,3,and 5 are non-random. However, because of the low power of the Anderson test in the presence of persistence, the ARIMA (1,0,1) model was applied to all of the ten records.

Moment estimates of φ and θ within the admissible space were obtained for all the records, except in the case of rivers 5 and 7. In the case of river 5, inspection of the successive values of $\theta^{(m)}$ defined by equation (5.7) showed that the last value of $\hat{\theta}^{(m)}$ within the parameter space was - 0.987, so the initial values for the ML search were taken as $\hat{\varphi}^{(m)} = 0.025$ and $\hat{\theta}^{(m)} = -0.95$, resulting in an ML solution at $\hat{\varphi}^{(1)} = 0.275$ and $\hat{\theta}^{(1)} = -0.454$. In the case of river 7, the initial moment estimates used to start the ML search were $\hat{\varphi}^{(m)} = -0.751$ and $\hat{\theta}^{(m)} = -0.95$, resulting in ML estimates of $\hat{\varphi}^{(1)} = -0.525$ and $\hat{\theta}^{(1)} = -0.755$. Thus, the empirical procedure used to define initial values for the ML search in these cases appeared to work satisfactorily.

The percentage of moment and ML estimates obtained is perhaps greater than might be expected on the basis of the sampling experiments; this might be due to chance or to related climatic regimes. The overall moment and ML estimates of φ and θ are rather randomly scattered in the parameter space and are hardly indicative of long-term persistence, as few of the estimates lie in region 6 of figure (3.4). However, on the basis of the sampling experiments conducted in section (5.3.2) these results would not be inconsistent with the presence of long-term persistence. Hence, in general, no conclusions can be drawn on the basis

River No.	n	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\varphi}$ (m)	$\hat{\theta}$ (m)	$\hat{\varphi}$ (l)	$\hat{\theta}$ (l)	$\hat{\rho}_1$ for residuals with 95% con. limits
1	150	0.459	-0.004	-0.009	-0.673	0.157	-0.440	-0.005 (0.153, -0.167)
2	132	0.185	-0.008	-0.045	-0.239	0.066	-0.129	-0.010 (0.163, -0.178)
3	97	0.695	0.498	0.716	0.041	0.797	0.168	-0.011 (0.189, -0.209)
4	82	0.177	0.083	0.470	0.303	0.324	0.133	-0.037 (0.204, -0.229)
5	76	0.531	0.013	0.025	-	0.275	-0.454	0.018 (0.211, -0.238)
6	74	0.110	0.105	0.952	0.877	0.350	0.232	-0.014 (0.214, -0.242)
7	73	0.169	-0.127	-0.751	-	-0.525	-0.755	-0.005 (0.216, -0.243)
8	72	0.190	-0.085	-0.296	-0.675	-0.418	-0.888	0.008 (0.217, -0.245)
9	71	0.140	-0.012	-0.087	-0.232	-0.008	-0.153	-0.005 (0.218, -0.247)
10	70	0.101	-0.100	-0.990	-0.987	-0.306	-0.441	-0.001 (0.220, -0.249)

Table (5.11) Values of $\hat{\rho}_1$ and $\hat{\rho}_2$ and moment and ML estimates of φ and θ for the 10 rivers listed in table (5.10).

of the present evidence about the presence or absence of long-term persistence in these records. An exception, perhaps, is river 3, (the St. Lawrence) which has an unusually large value of ρ_1 and which would be expected to display strong long-run effects because of the very large storage in the Great Lakes.

Inspection of the estimates of the lag-one autocorrelations for the residuals obtained from the ML fitting procedure shows that they are highly insignificant according to the Anderson test. A large proportion are negative in accordance with theory.

5.4.2 Results from Annual Tree Ring Indices

The radial growth of trees may be measured by the widths of annual rings. The tree ring widths are characterized by a tendancy for the ring widths and the variation in the ring widths to decrease with the age of the tree (Matalas, 1962). Thus, the series of ring widths is a non-stationary time series. If R_t denotes the ring width for any year, and T_t denotes the trend component during that year, then the ring widths may be transformed into a stationary series of ring indices (Schulman, 1956) as

$$X_t = R_t/T_t \qquad (5.47)$$

Series of tree ring indices have been studied by Matalas (1962) and found to be non-random, with rather larger values of ρ_1 observed than for annual runoff or rainfall. Because of the lengths of the records of tree ring indices, this result could be explained by bias in estimates of ρ_1. Evidence of long-term persistence in tree ring indices is reflected in estimates of $h > 0.5$ presented by Mandelbrot and Wallis (1969d). The presence of long-term persistence could be attributed to long-term fluctuations in rainfall, to soil moisture carry-over and to biological properties of tree growth in the form of storage of food products from one year to the next.

The ARIMA (1,0,1) model was fitted to 10 records of tree ring indices abstracted from Schulman (1956). The records are referenced through the table numbers as in Schulman (1956), which are given in table (5.12), and which are ordered therein in terms of decreasing record length. Estimates of h for the records are reproduced from Mandelbrot and Wallis (1969d), and range from 0.69 to 0.91, suggesting that long-term persistence is present. Moment and ML estimates of φ and θ were obtained for the records of tree ring indices as for the annual streamflow records, and are presented in table (5.12). The residuals were checked for randomness using the Anderson test, and results are also presented in table (5.12).

In contrast to the streamflow records, all of the moment and ML estimates of φ and θ lie in region 6 of the parameter space given in figure (3.4), which is consistent with the presence of long-term persistence. None of the ML estimates of φ are greater than 0.80, which suggests that the intensity of long-term persistence is moderate, or that the estimates of $\varphi^{(1)}$ and $\theta^{(1)}$ are biased which would be consistent with a stronger intensity of long-term persistence. The estimated lag-one autocorrelations of the residuals are again highly insignificant according to the Anderson test. Bearing in mind the record lengths in question, there is reasonable evidence that the ARIMA (1,0,1) model is appropriate for describing the type of persistence present in the data.

5.5. Summary

The simulation experiments described in section (5.3.2) have illustrated the difficulty of obtaining representative results from small samples in the presence of long-term persistence, when the ARIMA (1,0,1) model tends to fit samples generated by itself on only a proportion of possible occasions. The main reasons for this are the existence of correlation

TABLE*	n	H	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\varphi}^{(m)}$	$\hat{\theta}^{(m)}$	$\hat{\varphi}^{(l)}$	$\hat{\theta}^{(l)}$	$\hat{\rho}_1$ for residuals, with 95% con. limits
36	669	0.770	0.418	0.281	0.669	0.317	0.729	0.391	0.010 (0.074, -0.077)
43	661	0.700	0.359	0.266	0.744	0.452	0.664	0.352	-0.010 (0.075, -0.078)
77	589	0.720	0.527	0.326	0.616	0.125	0.626	0.138	0.003 (0.079, -0.083)
79	582	0.720	0.416	0.214	0.523	0.121	0.517	0.121	0.001 (0.080, -0.083)
66	537	0.690	0.470	0.289	0.608	0.193	0.632	0.209	0.002 (0.083, -0.087)
76	447	0.780	0.494	0.404	0.823	0.454	0.762	0.362	-0.010 (0.191, -0.095)
30	411	0.750	0.356	0.247	0.687	0.388	0.687	0.384	-0.005 (0.094, -0.099)
85	378	0.910	0.600	0.414	0.690	0.142	0.770	0.280	0.033 (0.098, -0.104)
81	304	0.860	0.474	0.291	0.614	0.181	0.725	0.341	0.027 (0.109, -0.116)
73	301	0.710	0.191	0.172	0.904	0.774	0.669	0.484	-0.030 (0.110, -0.116)

* from Schulman (1956)

Table (5.12) Values of $\hat{\rho}_1$ and $\hat{\rho}_2$ and moment and ML estimates of φ and θ for 10 records of tree ring indices

constraints implied by the structure of the model, and the estimation techniques which pronounce sequences generated by a stationary invertible process as either non-stationary or non-invertible or both. ML estimates are apparently more difficult to obtain, and may be more variable in some cases than moment estimates, particularly when the parameters of the underlying process lie close to boundaries. Firm conclusions, however, cannot be drawn about the sampling properties of moment and ML estimates from the sampling experiments reported in section (5.3.2). However, the problem of correct model identification and fitting through applying the Box-Jenkins methodology to sample sizes of the order of streamflow sequence lengths would appear to be a very difficult task, particularly in the presence of long-term persistence. Only for much longer records, such as tree rings indices, can more reliable results be obtained.

Chapter 6

PHYSICAL CONSIDERATIONS AND CONCLUSIONS

In formulating models for generating synthetic flows, attention is invariably confined to the observed statistical characteristics of the data, as demonstrated in Chapters 3 and 4, with only cursary consideration being given to the physical processes which give rise to the observed historic flows. This may be explained by the fact that techniques of stochastic streamflow synthesis derive their origins within the fields of operations research and statistics, rather than within the field of hydrology. As a result, hydrologists, when developing models for generating synthetic flows, frequently lose sight of the physical context of the observed data, the main aim being the attainment of the correct statistical resemblance between historic and synthetic sequences, which can invariably be achieved to the exclusion of any physical considerations. Sometimes, this approach may be sufficient, although hydrologists have in the past tried to reconcile the probabilistic structure of a stochastic model with the physical structure of streamflow. (Yevjevich, 1963; Fiering, 1967). This latter approach would appear to be more desirable than a purely statistical approach, as results obtained from stochastic models with a sound physical basis are likely to be more acceptable to engineering hydrologists.

Section (6.1) discusses the physical basis for long-term persistence and fGn, while section (6.2) is concerned with a physical basis for the ARIMA (1,0,1) model. In section (6.3) the main conclusions of the thesis are presented and some suggestions for further work are put forward in section (6.4).

6.1 Physical Basis for Long-Term Persistence and Fractional Noise

If fractional Gaussian noise (fGn) is to be accepted as a realistic model of annual streamflow, then the hypothesis that streamflow possesses an infinite memory must be considered. In practice, sufficient evidence obviously does not exist to validate or invalidate this hypothesis; however, the evidence that is available suggests that the memory of streamflow (and of other geophysical processes) is at least as long as the longest records available. Hence, observed records do not negate the infinite memory hypothesis, which in physical terms suggests that events which occured in the very distant past exert a small but non-negligible influence on present events. In the case of streamflow, groundwater storage in a catchment undoubtedly induces some persistence but it is difficult to conceive how finite capacity aquifers could induce the long-term persistence implied by the fGn model. However, an explanation in terms of groundwater storage is not necessary as long rainfall records have also been found to exhibit the Hurst phenomenon.

It has long been recognised that long-term movements in temperature and precipitation have taken place in the past; such low frequency movements are frequently referred to collectively as "climatic change" by climatologists. Attempts to establish any regular pattern in these fluctuations have met with little success, as indeed have efforts to attribute climatic change to physical causes. If climatic change is governed by large scale oceanic and storage precesses, then long-term persistence and climatic change would be synonymous. However observed annual data support but do not prove the argument for a long-term persistence explanation of long-term climatic movements.

Attempts are frequently made to link climatic change with solar

activity, sometimes referred to as sunspot activity, and these have met with some measure of success, (King, 1973). However, the relationship between solar and meteorological activity is not simple, and no plausible mechanism has been advanced which can explain why solar activity should influence meteorological response. The suggestion is sometimes made that some "trigger" mechanism operates, and if this is the case, then climatic change on Earth would appear to be ruled by extraneous influences rather than by internal storage processes.

A possible alternative probabilistic explanation of the Hurst phenomenon could be some form of non-stationarity or time variance, as suggested by Nash (1971: personal communication). Hurst (1957) originally suggested an explanation of his findings in terms of non-stationarity in the mean. This raises the question as to whether it is reasonable to entertain the notion of a long-term mean for streamflow when river basins are known to be evolutionary systems. Again, this argument must also be applied to precipitation, where evolutionary influences are, however, less obvious.

A non-stationary ARIMA (0,1,1) model for Hurst's time series was examined briefly by O'Connell (1971) but some simulation experiments showed that the required behaviour in the rescaled range could not be achieved with this type of non-stationarity. In any case, such a model would be unsuitable for operational purposes (Watts, 1972).

More recently, Klemes (1974) has presented a thoughtful and thorough analysis of the possible underlying physical reasons for the Hurst phenomenon. In so doing, he has contended that non-stationarity may be the underlying reason, and notes that present techniques of time series analysis would suggest that long-term persistence existed when, in fact, true non-stationarity was the underlying reason for low

frequency effects. He has presented plots of ln(R/S) against ln(n) for simulations where the process possessed no memory and the mean and the period over which it operates were allowed to vary randomly in such a fashion that Hurst's law with $0.5 < h < 1$ was obeyed, thus lending weight to the non-stationarity hypothesis. However, for operational purposes, such models would be very difficult to implement, so that the flexibility of stationary approximations to dfGn for operational uses remains unchallenged. However, Klemes (1974) rightly points out that the long-term persistence explanation of the Hurst phenomenon may not be correct physically, and that, similar to the Ptolemaic planetary model, fGn and other models founded on the hypothesis of long-term persistence may fit observed patterns of behaviour very well without representing the correct physical explanation. However, it should be remembered that apparent observed non-stationarities may still be governed by long-term persistence.

6.2 Physical Basis for the ARIMA (1,0,1) Process

The questions raised about the plausibility of long-term persistence in section (6.1) also apply in the case of the ARIMA (1,0,1) model. However, the concept of an infinite memory does not arise, as the ARIMA (1,0,1) possesses an autocorrelation function which decays exponentially, but nevertheless, sufficiently slowly to allow an adequate approximation to dfGn as shown in section (3.3). Thus only a large but finite memory is hypothesised.

A comprehensive study of the causal factors contributing to dependence in annual streamflow has been made by Moss (1972a), who has formulated a physical model to describe the transfer mechanism which converts a stochastic input of precipitation into an output of streamflow.

In formulating his model, Moss has pointed out that the time-distribution of the inputs (precipitation) within each time unit affects significantly the magnitudes of the outputs (streamflow) and as a result, daily, monthly or annual streamflow volumes should not strictly be treated as discrete time processes. Moss uses the term discretized streamflow to describe the process resulting from summing or averaging continuous streamflow. The distinction between discretized processes and discrete time processes is discussed before describing the model for annual streamflow proposed by Moss (1972a).

6.2.1 Approximation of Continuous-Time Processes

In simulating synthetic streamflows, discrete time processes based on the increments of Brownian motion are invariably used to model what are strictly discretized processes as defined by Moss (1972a); the error introduced by this type of approximation has apparently not been extensively investigated. The increments of Brownian motion, defined as

$$\Delta B(t) = B(t) - B(t-1) \qquad (6.1)$$

do not take any account of the variation of B(t) over the unit time interval. However, as noted in section (2.5.1), because of self-similarity, the function B(t) is so locally erratic in its behaviour that it does not possess a derivative, which is in contrast with the smooth behaviour of continuous streamflow.

Mandelbrot and Wallis (1969c) have argued that fractional Brownian motion represents a suitable model for precipitation because of its considerable irregularity in continuous time; however, this argument cannot readily be applied to streamflow. Mathematical smoothing procedures can be introduced as artificial means of ensuring smoothness but such procedures hardly add to the physical plausability of models based on the increments of Brownian motion. The same criticism applies to models based on the increments of fractional Brownian motion,

although "smoothness" properties would not appear to be of primary importance in modelling long-term persistence. Rodriquez et al (1972) have argued for smoothness properties mainly from the viewpoint of applying crossing theory; the Broken Line process possesses certain properties of smoothness but it would appear to be difficult to give any physical interpretation to the parameters of the BL process or to its method of construction.

6.2.2 A Physical Model for Annual Streamflow

Moss (1972a) has postulated a model for discretized streamflow as follows. A stochastic input of effective rainfall (rainfall less evapotranspiration) is imposed on a physical system, the drainage basin; the resulting streamflow over a period is described in terms of three components: (i) direct runoff $r'_i P_i$ where r'_i represents the proportion of direct runoff for period i, (ii) baseflow caused by over period storage G_i and (iii) baseflow caused by within period precipitation F_i. A schematic representation of these components is given in figure (6.1)

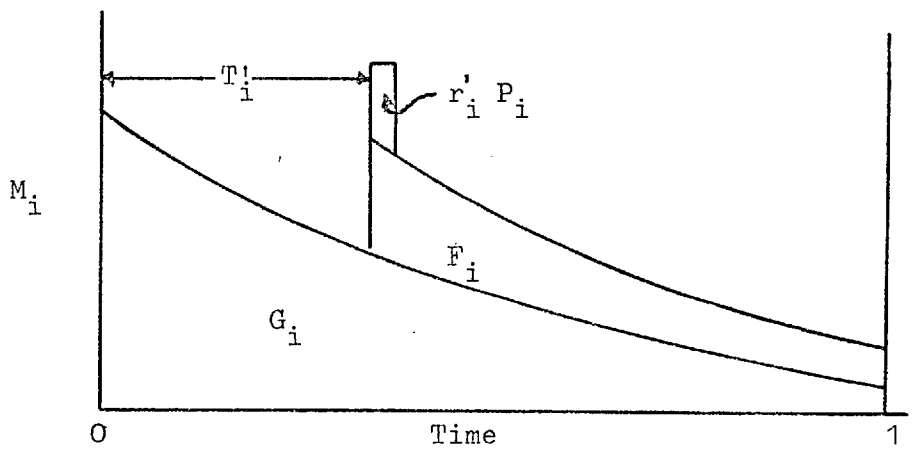


Figure (6.1) (after Moss, 1972a)

Total streamflow within a period is thus defined as

$$M_i = G_i + F_i + r'_i P_i \qquad (6.2)$$

In order to include the effect of the within-period distribution of precipitation, Moss (1972a) introduces an additional variable T'_i which represents the time of occurence of a storm with a magnitude

equal to the sum of the individual storms that occured in the i^{th} period and whose baseflow contribution within that period is equal in magnitude to the baseflow contributions of the individual storms. No correlations are assumed to exist between the magnitudes or between the times of occurence of the individual storms, and the magnitudes are treated as instantaneous pulses. Using a continuity equation and equation (6.2) Moss then shows that streamflow M_i in the current time period can be expressed as a function of precipitation in the current time period, and of streamflow and precipitation in the previous time period; thus, if precipitation is equated with the independent random term, the exact form of the ARIMA (1,0,1) process results. The parameters φ and θ are related to the parameters of the physical model proposed by Moss through the following equations

$$\varphi = e^{-k} \tag{6.3}$$

$$\theta = \frac{a'_{i-1}e^{k(T'_{i-1}-1)} - e^{-k}}{a'_i e^{k(T'_i-1)} - 1} \tag{6.4}$$

where $a'_i = (1-r'_i)$ and k is the baseflow recession constant characterizing a linear reservoir. Because of stationarity, Moss assumes the values of a'_i and T'_i to be constant between periods and equal to a' and T' respectively.

For a particular stream suited to the basic assumptions of the model, estimates of a', which represents the ratio of infiltration to effective precipitation and of k, the recession constant, were obtained by Moss (1972a) from a previous work and through conventional analysis. An estimate of the parameter T' was obtained from an analysis of precipitation and evapotranspiration records. For annual flows, comparison of the estimates of ρ_1 obtained using equations (6.3), (6.4)

and (3.24) with those obtained from the observed record showed poor agreement, no matter what month was taken to start the water year. Autocorrelation in the annual effective precipitation values, sampling errors and seasonality in the infiltration ratio r' were suggested as possible explanations (Moss, 1972b). A slight source of error might arise from the fact that equations (6.3) and (6.4) pertain to a discretized process while equation (3.24) pertains to a discrete process. However, sufficient verification of this model has not as yet been performed.

6.2.3 Justification for Parameter Space

The parameter space of the ARIMA (1,0,1) process was defined in section (3.3.1) as $-1 < \varphi < +1$ and $-1 < \theta < +1$. While stationarity is a necessary requirement insofar as generating synthetic flows is concerned, the form of the equation of the ARIMA (1,0,1) process

$$X_t = \varphi X_{t-1} + \varepsilon_t - \theta \varepsilon_{t-1} \qquad (6.5)$$

also suggests that the restriction $-1 < \varphi < +1$ is a sensible requirement for streamflow.

The restriction imposed on the parameter θ to ensure invertibility, as defined in section (3.3.1) requires some justification in the case of streamflow. Moss (1972a) has examined the invertibility restriction on θ from a physical standpoint, but found no obvious physical justification.

In section (3.1.2) the invertibility condition for a simple MA(1) process

$$X_t = \varepsilon_t - \theta \varepsilon_{t-1} \qquad (6.6)$$

was shown to be $-1 < \theta < +1$ which implies that if the process is written in the equivalent form of an infinite autoregression,

$$X_t = -\theta X_{t-1} - \theta^2 X_{t-2} - \theta^3 X_{t-3} \ldots\ldots + \varepsilon_t \qquad (6.7)$$

then invertibility ensures that the effects of past values of the process die out in a sensible manner. Nevertheless the process would still be stationary if $\theta > 1$.

The same argument may be shown to hold for the ARIMA (1,0,1) process which can also be written as an infinite autoregression. Using the backward shift operator, the process may be written as

$$(1 - B\varphi)X_t = (1 - B\theta)\varepsilon_t$$

or

$$\varepsilon_t = \left(\frac{1 - B\varphi}{1 - B\theta}\right)X_t$$

Expanding $1/(1-B\theta)$ gives

$$\varepsilon_t = (1-B\varphi)(1 + B\theta + B^2\theta^2 + \ldots)X_t$$

$$= (1 + B\theta + B^2\theta^2 + \ldots)X_t - (B\varphi + B^2\varphi\theta + B^3\varphi\theta^2 + \ldots)X_t$$

$$= X_t + (\theta-\varphi)X_{t-1} + (\theta^2-\varphi\theta)X_{t-2} + (\theta^3-\varphi\theta^2)X_{t-3} + \ldots$$

or

$$X_t = (\varphi-\theta)X_{t-1} + (\varphi\theta-\theta^2)X_{t-2} + (\varphi\theta^2-\theta^3)X_{t-3} + \ldots + \varepsilon_t \qquad (6.8)$$

The infinite series of coefficients $(\varphi-\theta)$, $(\varphi\theta-\theta^2)$, $(\varphi\theta^2-\theta^3)$,... forms a convergent series for $\theta < 1$ with no condition imposed on the value of the parameter φ, which gives an empirical justification for maintaining the invertibility condition.

6.3 Summary and Conclusions

The stochastic modelling of long-term persistence has been reviewed comprehensively in Chapters 1 and 2, with particular emphasis placed on the simulation of synthetic streamflows. In Chapter 3, a simple stochastic model with desirable short-term and long-term properties

was proposed and developed for simulation purposes, while the multivariate extension of the same model was derived in Chapter 4. Techniques for estimating the parameters of the model were examined in Chapter 5, while physical interpretations for the parameters and for long-term persistence in general have been considered in the preceding sections of this chapter.

The primary conclusions to be drawn from this thesis are:

1. While the discrete-time fractional Gaussian noise (dfGn) model has precisely the right mathematical structure for modelling Hurst's law, no simple adequate approximation to dfGn has been suitably developed for simulation purposes.

2. The observed long-term persistence in geophysical records can be successfully modelled by the simple ARIMA (1,0,1) process, as evidenced by the good agreement shown with Hurst's law within large simulated sequences. In addition, short-term persistence, as measured by estimates of the lag-one autocorrelation coefficient can also be modelled simultaneously.

3. For simulation purposes, the ARIMA (1,0,1) process has been formulated to maintain the required resemblance between historic and synthetic sequences in terms of both short-term and long-term properties. In achieving this resemblance, it is necessary to distinguish clearly between the small sample and population properties of the process.

4. Because of the computational ease associated with using the ARIMA (1,0,1) process to generate synthetic sequences, the process should be useful as a simulation tool and should prove attractive to engineers who wish to assess the effects of long-term persistence on water resource system performance.

5. A multisite version of the ARIMA (1,0,1) process has been developed for generating synthetic sequences which reproduce the observed temporal and spatial correlations. In contrast to other multisite models of long-term persistence, the application of the multisite ARIMA (1,0,1) model should not raise any major difficulties.

6. The reliable estimation of the parameters of the ARIMA (1,0,1) process from annual streamflow sequences of moderate length would appear to be a very difficult task. It can be concluded from the simulation experiments conducted in Chapter 5 that such sequences cannot be expected to provide much supporting evidence for the applicability of the ARIMA (1,0,1) model, or indeed of other models of long-term persistence.

7. A physical basis for the infinite memory hypothesis associated with the dfGn model has not as yet emerged. However, consideration of the physical processes which determine annual streamflow tends to suggest that the ARIMA (1,0,1) model has a reasonable physical basis. While observed properties of streamflow can be reproduced in synthetic sequences without reference to physical considerations, a model which also reflects physical reality should be more desirable.

8. The development of stochastic models for simulating synthetic flows would appear to be an important area for research in the light of the growing potential of simulation as a powerful tool in planning and operating water resource systems. The ARIMA (1,0,1) process should add to this potential and make the simulation of synthetic sequences with desirable short-term and long-term properties easier than has heretofore been possible.

6.4 Recommendations for Further Study

There would appear to be considerable scope for further research involving the ARIMA (1,0,1) process, both on properties of the process and on its application as a simulation tool in the planning of water resources. Specific recommendations can be made as follows:

(i). The properties of moment and maximum likelihood estimates of the parameters of the ARIMA (1,0,1) process in small samples require further exploration.

(ii). A Bayesian approach to the estimation of the parameter φ might prove more reliable than more classical approaches.

(iii). Because the Hurst coefficient h is theoretically defined as 0.5 for the ARIMA (1,0,1) process a simulation approach had to be adopted to define the expected value of h in small samples, and hence other small sample properties, such as the variance and lag-one autocorrelation could also be readily obtained. However, it might be possible to derive expected values for the standard deviation and lag-one autocorrelation analytically and thus to avoid the use of tables to achieve the desired resemblance between historic and synthetic sequences in terms of these properties.

(iv). The problem of model choice requires further study, and could proceed along two separate lines:

a) If the distribution of the rescaled range R/S in small samples could be derived for various processes, this should allow the formulation of more powerful tests of significance for long-term persistence and for discriminating between alternative models.

b) Bayesian decision theory could be applied where the choice

of model is ruled by the expected losses accruing from an incorrect choice of model. Such an approach could only be used where the form of the loss function is specifically known.

(v). The application of the ARIMA (1,0,1) process to monthly and daily streamflows has not been attempted here and could be achieved through applying the model directly or through disaggregating previously generated annual flows as suggested by Valencia and Schaake (1972).

(vi). Some investigations into the importance of long-term persistence in various design situations would be desirable. Synthetic sequences used in such investigations should perhaps maintain a memory of the historic record as suggested by Garcia et al (1972).

(vii). Long-term persistence, a term used to describe the occurence of long series of wet and dry years, may not be synonymous with autocorrelation as pointed out by Klemes (1974). Given that the Hurst phenomenon can be explained by a stochastic process characterized by time variance or autocorrelation, only physical considerations may ultimately decide which is the correct explanation, and some work is definitely necessary in this area. Indeed, the further advancement of stochastic techniques of streamflow synthesis can probably only come from a better understanding of the physical laws governing streamflow itself.

Appendix A2.1

The variance of the increments of fractional Brownian motion

$$\text{Var}\,[B_h(t+T) - B_h(t)] = \text{Var}\,[B_h(T) - B_h(0)]$$

$$= T^{2h}[\text{Var}\;\; B_h(1) - B_h(0)] \qquad \text{(A2.1.1)}$$

From equation (2.23)

$$\text{Var}\,[(B_h(1) - B_h(0)] = \frac{1}{\left[\sqrt{h+0.5}\right]^2}\;\text{Var}\left\{\int_{-\infty}^{0}[(1-u)^{h-0.5} - (-u)^{h-0.5}]dB(u)\right.$$

$$\left. + \int_0^1 (1-u)^{h-0.5}\,dB(u)\right\}$$

$$= \frac{1}{\left[\sqrt{h+0.5}\right]^2}\left\{\int_{-\infty}^{0}[(1-u)^{h-0.5} - (-u)^{h-0.5}]\;\text{Var}\,[dB(u)]\right.$$

$$\left. + \int_0^1 (1-u)^{h-0.5}\,\text{Var}\,[dB(u)]\right\}$$

$$= \frac{1}{\left[\sqrt{h+0.5}\right]^2}\left\{\int_{-\infty}^{0}[(1-u)^{h-0.5} - (-u)^{h-0.5}]^2\,du\right.$$

$$\left. + \int_0^1 (1-u)^{h-0.5}\,du\right\}$$

$$= \frac{1}{\left[\sqrt{h+0.5}\right]^2}\left\{\int_{-\infty}^{0}[(1-u)^{h-0.5} - (-u)^{h-0.5}]^2\,du + \frac{1}{2h}\right\}$$

(A2.1.2)

$$= V_h$$

As the integral on the r.h.s. of equation (A2.1.2) exists, V_h is finite and is a function of h. Thus, from equation (A2.1.1),

$$\text{Var}\,[B_h(t+T) - B_h(t)] = T^{2h}\,V_h$$

Appendix 3.1

Lag-one Autocorrelation for a Log-Normal Process

Let

$$Y_t = \ln (X_t - a) \qquad (A3.1.1)$$

where ln denotes $\log_e$ and a is the lower bound of the process X_t. Therefore

$$X_t = a + \exp(Y_t) \qquad (A3.1.2)$$

The lag-one autocorrelation of X_t is defined as

$$\rho_x = \frac{E[X_t X_{t-1}] - (E[X_t])^2}{E[X_t^2] - (E[X_t])^2} \qquad (A3.1.3)$$

The moment generating function for a normally distributed random variable with mean μ and variance σ^2 is defined, for real values of t, as (Parzen, 1960 ; p. 221)

$$M(t) = E[\exp(tx)]$$

$$= \exp(t\mu + \tfrac{1}{2}\sigma^2 t^2) \qquad (A3.1.4)$$

The moment generating function for jointly normal random variables X_1 and X_2 with means μ_1 and μ_2 and standard deviations σ_1 and σ_2 respectively and correlation ρ is defined as (Parzen, 1960 ; pp. 357-58)

$$M(t_1, t_2) = E[\exp(t_1 X_1) \exp(t_2 X_2)]$$

$$= \exp[t_1\mu_1 + t_2\mu_2 + \tfrac{1}{2}(t_1^2\sigma_1^2 + 2\rho\sigma_1\sigma_2 t_1 t_2 + t_2^2\sigma_2^2)] \qquad (A3.1.5)$$

for all real values of t_1 and t_2.

Thus, from equations (A3.1.2) and (A3.1.4) it can be deduced that

$$E[X_t] = a + \exp(\mu_y + \sigma_y^2/2) \qquad (A3.1.6)$$

$$E[X_t^2] = a^2 + 2a \exp(\mu_y + \sigma_y^2/2) + \exp[2(\mu_y + \sigma_y^2)] \qquad (A3.1.7)$$

$$\therefore \quad E[X_t^2] - (E[X_t])^2 = \exp[2(\mu_y+\sigma_y^2)] - \exp(2\mu_y+\sigma_y^2)$$

$$= \exp[2\mu_y+\sigma_y^2][\exp(\sigma_y^2) - 1] \quad \text{(A3.1.8)}$$

By letting $Y_t = X_1$ and $Y_{t-1} = X_2$ in equation (A3.5), and letting $t_1 = t_2 = 1$, it follows that

$$E[X_t X_{t-1}] = E[(a + e^{Y_t})(a + e^{Y_{t-1}})]$$

$$= a^2 + 2a \exp(\mu_y + \sigma_y^2/2)$$

$$+ \exp[2\mu_y + \sigma_y^2 \rho_y + \sigma_y^2] \quad \text{(A3.1.9)}$$

$$E[X_t X_{t-1}] - (E[X_t])^2 = \exp[2\mu_y + \sigma_y^2 \rho_y + \sigma_y^2] - \exp[2\mu_y + \sigma_y^2]$$

$$= \exp[2\mu_y + \sigma_y^2][\exp(\sigma_y^2 \rho_y) - 1]$$

(A3.1.10)

$$\therefore \quad \rho_x = \frac{(\exp[\sigma_y^2 \rho_y] - 1)}{(\exp \sigma_y^2 - 1)} \quad \text{(A3.1.11)}$$

The relationship between ρ_x and ρ_y given by equation (A3.1.11) apparently holds regardless of the process employed to maintain the correlation ρ_y between Y_t and Y_{t-1}. The approach could also be extended to higher lag autocorrelations in X and Y space.

Appendix 3.2

The sequent peak algorithm

The following formulation follows that of Wallis and Matalas (1972). Let $[x(j) : j \varepsilon J]$ and $[d(j) : j \varepsilon J]$ denote time sequence of flows and demands, where $J = [1,2,\ldots, n]$ refers to the set of equally spaced time points over which the reservoir is to operate. If there is partial development,

$$\sum_{j=1}^{n} x(j) > \sum_{j=1}^{n} d(j) \qquad \text{(A3.2.1)}$$

then there will be a sequence of wastes $[w(j) : j \varepsilon J]$, where

$$\sum_{j=1}^{n} w(j) = \sum_{j=1}^{n} x(j) - \sum_{j=1}^{n} d(j) \qquad \text{(A3.2.2)}$$

The total demand $D = \sum_{j=1}^{n} d(j)$ is related to the total flow by

$$D = \alpha' \sum_{j=1}^{n} x(j) \qquad \text{(A3.2.3)}$$

where $0 < \alpha' < 1$ denotes the level of development. Let the minimum capacity required to meet all demands be denoted by C_p. If $C(0)$ denotes the initial storage necessary to avoid storage deficiencies, the storage at the end of the j^{th} time period, $C(j)$, is given by

$$C(j) = \min_{j} [C_p, (x(j) - d(j) + C(j-1))] \qquad \text{(A3.2.4)}$$

and the waste water at time j, w(j) is given by

$$w(j) = \max_{j} [0, (x(j) - d(j) - C_p + C(j-1))] \qquad \text{(A3.2.5)}$$

Let

$$z(i) = \sum_{j=1}^{i} [x(j) - d(j)] \qquad i \varepsilon I = [1,\ldots,n] \qquad \text{(A3.2.6)}$$

Let the minimum and maximum values of z(i) be denoted by Z' and Z", respectively. The range of cumulative departures from the sample mean

or adjusted range R is defined as

$$R = Z'' - Z' \qquad (A3.2.7)$$

Hence, for $\alpha = 1$, $C_p = R$ and $C(0) = Z'$.

The solution for C_p may be obtained by the sequent peak algorithm (Thomas and Burden, 1963) as follows. In order to obtain the solution, the two time series $[x(j) : j \varepsilon J]$ and $[d(j) : j \varepsilon J]$ are assumed cyclic ; this strictly operational assumption is imposed to carry out the solution for C_p and has no physical significance. The solution requires the use of only two cycles ; in this case the time span for the two time series is defined as $J' = [1,\ldots,2n]$, where $x(j) = x(j+n)$ and $d(j) = d(j+n)$ for $j = 1,\ldots,n$. The solution for C_p is obtained through the following steps.

1. Calculate $x(j) - d(j)$ $\forall$ $j \varepsilon J'$
2. Calculate $z(i)$ $\forall$ $i \varepsilon I' = [1,\ldots,2n]$
3. In $[z(i) : i \varepsilon I']$ locate the sequence of peaks $[P_r : r \varepsilon R]$ where $R = [1,\ldots,m]$ such that $P_1 < P_2 < \ldots < P_m$.
4. Between sequent peaks locate the sequence of troughs $[T_s : s \varepsilon S]$ where $S = [1,\ldots,m-1]$
5. Form the sequence $[(P_s - T_s) : s \varepsilon S]$. The minimum design capacity C_p is given by

$$C_p = \max_s (P_s - T_s)$$

The values $C(j)$ and $w(j)$ $\forall$ $j \varepsilon J$ may be derived using equations (A3.2.4) and (A3.2.5), while $C(0)$ may be determined as follows. Set $C(0) = 0$, and using equation (A3.2.4) determine $C(j)$ $\forall$ $j \varepsilon J$. If $C(j) > 0$ $\forall$ $j \varepsilon J$ then $C(0) = 0$. If $C(q) < 0$ $\forall$ $q \varepsilon Q \leq J$, then $C(0) = \max_q |C(q)|$.

Appendix 3.3

Application of a constant in the sequent peak algorithm

If a constant β is applied to each of the flows $[x(j) : j \in J]$ and demands $[d(j) : j \in J]$, and if the level of development is maintained at α', then the steps in the sequent peak algorithm are as follows.

1. Calculate $\beta x(j) - \beta d(j) \; \forall \; j \in J'$
2. Calculate $\beta z(i) \; \forall \; i \in I'$
3. In $z(i) : i \in I'$ locate the sequence of peaks $[\beta P_r : r \in R]$ where $R = [1,\ldots,m]$ such that $\beta P_1 < \beta P_2 < \ldots < \beta P_m$
4. Between sequent peaks locate the sequence of troughs $[\beta T_s : s \in S]$ where $S = [1,\ldots,m-1]$
5. Form the sequence $[(\beta P_s - \beta T_s) : s \in S]$. The minimum design capacity C'_p is given by

$$C'_p = \max_s (\beta P_s - \beta T_s)$$

$$= \beta \max_s (P_s - T_s)$$

$$= \beta \, C_p$$

Appendix 4.1

Correlations Preserved by the Multisite ARIMA (1,0,1) Process

The case of 3 sites will be used to illustrate the derivation, which can then easily be generalized to the case of m sites. A diagonal form will be assumed for the matrix $\underline{A}$. In expanded matrix notation, equation (4.17) may be written as

$$\begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \end{bmatrix} = \begin{bmatrix} a_{11} & \cdot & \cdot \\ \cdot & a_{22} & \cdot \\ \cdot & \cdot & a_{32} \end{bmatrix} \begin{bmatrix} x_1(t-1) \\ x_2(t-1) \\ x_3(t-1) \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix} \begin{bmatrix} \varepsilon_1(t) \\ \varepsilon_2(t) \\ \varepsilon_3(t) \end{bmatrix} - \begin{bmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{bmatrix} \begin{bmatrix} \varepsilon_1(t-1) \\ \varepsilon_2(t-1) \\ \varepsilon_3(t-1) \end{bmatrix} \quad \text{(A4.1.1)}$$

The generating processes at each of the 3 sites are as follows:

$$x_1(t) = a_{11}x_1(t-1) + b_{11}\varepsilon_1(t) + b_{12}\varepsilon_2(t) + b_{13}\varepsilon_3(t) - c_{11}\varepsilon_1(t-1) - c_{12}\varepsilon_2(t-1) - c_{13}\varepsilon_3(t-1) \quad \text{(A4.1.2)}$$

$$x_2(t) = a_{22}x_2(t-1) + b_{21}\varepsilon_1(t) + b_{22}\varepsilon_2(t) + b_{23}\varepsilon_3(t) - c_{21}\varepsilon_1(t-1) - c_{22}\varepsilon_2(t-1) - c_{23}\varepsilon_3(t-1) \quad \text{(A4.1.3)}$$

$$x_3(t) = a_{33}x_3(t-1) + b_{31}\varepsilon_1(t) + b_{32}\varepsilon_2(t) + b_{33}\varepsilon_3(t) - c_{31}\varepsilon_1(t-1) - c_{32}\varepsilon_2(t-1) - c_{33}\varepsilon_3(t-1) \quad \text{(A4.1.4)}$$

Cross-multiplication by $x_2(t)$ in equation (A4.1.2) yields:

$$x_1(t)x_2(t) = a_{11}x_1(t-1)x_2(t) + b_{11}\varepsilon_1(t)x_2(t) + b_{12}\varepsilon_2(t)x_2(t) + b_{13}\varepsilon_3(t)x_2(t) - c_{11}\varepsilon_1(t-1)x_2(t) - c_{12}\varepsilon_2(t-1)x_2(t) - c_{13}\varepsilon_3(t-1)x_2(t) \quad \text{(A4.1.5)}$$

Application of the expectation operator in equation (A4.1.5) requires

the evaluation of the following terms:

$$E[x_2(t)\varepsilon_1(t)] = b_{21}$$

$$E[x_2(t)\varepsilon_2(t)] = b_{22}$$

$$E[x_2(t)\varepsilon_3(t)] = b_{23}$$

$$E[x_2(t)\varepsilon_1(t-1)] = a_{22}E[x_2(t-1)\varepsilon_1(t-1)] - c_{21} = a_{22}b_{21} - c_{21}$$

$$E[x_2(t)\varepsilon_2(t-1)] = a_{22}E[x_2(t-1)\varepsilon_2(t-1)] - c_{22} = a_{22}b_{22} - c_{22}$$

$$E[x_2(t)\varepsilon_3(t-1)] = a_{22}E[x_2(t-1)\varepsilon_3(t-1)] - c_{23} = a_{22}b_{23} - c_{23}$$

whence

$$\begin{aligned}\rho_{12}(0) = {} & a_{11}\rho_{21}(1) + b_{11}b_{21} + b_{12}b_{22} + b_{13}b_{23} \\ & - c_{11}(a_{22}b_{21} - c_{21}) - c_{12}(a_{22}b_{22} - c_{22}) \\ & - c_{13}(a_{22}b_{23} - c_{23})\end{aligned} \quad \text{(A4.1.6)}$$

In a similar fashion it may be shown that

$$\begin{aligned}\rho_{13}(0) = {} & a_{11}\rho_{31}(1) + b_{11}b_{31} + b_{12}b_{32} + b_{13}b_{33} \\ & - c_{11}(a_{33}b_{31} - c_{31}) - c_{12}(a_{33}b_{32} - c_{32}) \\ & - c_{13}(a_{33}b_{33} - c_{33})\end{aligned} \quad \text{(A4.1.7)}$$

$$\begin{aligned}\rho_{23}(0) = {} & a_{22}\rho_{32}(1) + b_{21}b_{31} + b_{22}b_{32} + b_{23}b_{33} \\ & - c_{21}(a_{33}b_{31} - c_{31}) - c_{22}(a_{33}b_{32} - c_{32}) \\ & - c_{23}(a_{33}b_{33} - c_{33})\end{aligned} \quad \text{(A4.1.8)}$$

The expression for $\rho_{ij}(0)$ in the general case of m sites can now be deduced from equations (A4.1.6) - (A4.1.8) as:

$$\begin{aligned}\rho_{ij}(0) = {} & a_{ii}\rho_{ji}(1) + \sum_{k=1}^{m} b_{ik}b_{jk} - a_{jj}\sum_{k=1}^{m} c_{ik}b_{jk} \\ & + \sum_{k=1}^{m} c_{ik}c_{jk}\end{aligned} \quad \text{(A4.1.9)}$$

The variances at each of the m sites are given by equation (A4.1.9) with j set equal to i, $[i = 1,2,\ldots,m]$.

Cross-multiplication by $x_2(t-1)$ in equation (A4.1.2) gives:

$$\begin{aligned} x_1(t)x_2(t-1) &= a_{11}x_1(t-1)x_2(t-1) + b_{11}\varepsilon_1(t)x_2(t-1) \\ &+ b_{12}\varepsilon_2(t)x_2(t-1) + b_{13}\varepsilon_3(t)x_2(t-1) \\ &- c_{11}\varepsilon_1(t-1)x_2(t-1) - c_{12}\varepsilon_2(t-1)x_2(t-1) \\ &- c_{13}\varepsilon_3(t-1)x_2(t-1) \end{aligned} \qquad (A4.1.10)$$

Application of the expectation operator in equation (A4.1.10) gives:

$$\rho_{12}(1) = a_{11}\rho_{12}(0) - c_{11}b_{21} - c_{12}b_{22} - c_{13}b_{23} \qquad (A4.1.11)$$

Similarly it may be shown that

$$\rho_{13}(1) = a_{11}\rho_{13}(0) - c_{11}b_{31} - c_{12}b_{32} - c_{13}b_{33} \qquad (A4.1.12)$$

$$\rho_{23}(1) = a_{22}\rho_{23}(0) - c_{21}b_{31} - c_{22}b_{32} - c_{23}b_{33} \qquad (A4.1.13)$$

The expression for $\rho_{ij}(1)$ in the general case of m sites can now be deduced from equations (A4.1.11) - (A4.1.13) as

$$\rho_{ij}(1) = a_{ii}\rho_{ij}(0) - \sum_{k=1}^{m} c_{ik}b_{jk} \qquad (A4.1.14)$$

The lag-one autocorrelations at each of the m sites are given by equation (A4.1.14) with j set equal to i, $[i = 1,2,\ldots,m]$.

Equations (A4.1.9) and (A4.1.14) may now be solved simultaneously to give

$$\begin{aligned} \rho_{ij}(0) = \frac{1}{(1 - a_{ii}a_{jj})} \Big[&-a_{ii}\sum_{k=1}^{m} c_{jk}b_{ik} + \sum_{k=1}^{m} b_{ik}b_{jk} \\ &- a_{jj}\sum_{k=1}^{m} c_{ik}b_{jk} + \sum_{k=1}^{m} c_{ik}c_{jk} \Big] \end{aligned} \qquad (A4.1.15)$$

whence $\rho_{ij}(1)$ may be calculated from equation (A4.1.14)

The lag-two autocorrelations may easily be shown to be:

$$\rho_{ij}(2) = a_{ii}\rho_{ij}(1) \qquad (A4.1.16)$$

Appendix 4.2

Correlations preserved by the Multisite ARIMA (0,0,1) Process

The case of 2 sites will be used to illustrate the nature of the constraints involved in assuming lower diagonal forms for both of the matrices $\underline{B}$ and $\underline{C}$. Equation (4.5.1) may be written in expanded form as

$$\begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \begin{bmatrix} b_{11} & 0 \\ b_{21} & b_{22} \end{bmatrix} \begin{bmatrix} \varepsilon_1(t) \\ \varepsilon_2(t) \end{bmatrix} - \begin{bmatrix} c_{11} & 0 \\ c_{21} & c_{22} \end{bmatrix} \begin{bmatrix} \varepsilon_1(t-1) \\ \varepsilon_2(t-1) \end{bmatrix} \qquad \text{(A4.2.1)}$$

Let the observed lag-zero and lag-one correlation matrices be denoted by

$$\underline{M}_o = \begin{bmatrix} 1 & R \\ R & 1 \end{bmatrix} \qquad \underline{M}_1 = \begin{bmatrix} \rho & \alpha \\ \beta & r \end{bmatrix}$$

From equations (4.52) and (4.53)

$$\underline{M}_o - \underline{M}_1 - \underline{M}_1^T = \begin{bmatrix} (1-2\rho) & (R-\alpha-\beta) \\ (R-\alpha-\beta) & (1-2r) \end{bmatrix}$$

$$= \underline{\beta}\,\underline{\beta}^T \qquad \text{(A4.2.2)}$$

$$\underline{M}_o + \underline{M}_1 + \underline{M}_1^T = \begin{bmatrix} (1+2\rho) & (R+\alpha+\beta) \\ (R+\alpha+\beta) & (1+2r) \end{bmatrix}$$

$$= \underline{\beta}^*\underline{\beta}^{*T} \qquad \text{(A4.2.3)}$$

Using lower triangularization equations (A4.2.2) and (A4.2.3) may be solved to yield the matrices $\underline{\beta}$ and $\underline{\beta}^*$ as

$$\underline{\beta} = \begin{bmatrix} \sqrt{1-2\rho} & 0 \\ \dfrac{R-\beta-\alpha}{\sqrt{1-2\rho}} & \sqrt{\dfrac{(1-2r)(1-2\rho) - (R-\beta-\alpha)^2}{(1-2\rho)}} \end{bmatrix}$$

$$\underline{\beta}^* = \begin{bmatrix} \sqrt{1+2\rho} & 0 \\ \dfrac{R+\beta+\alpha}{\sqrt{1+2\rho}} & \sqrt{\dfrac{(1+2r)(1+2\rho) - (R+\beta+\alpha)^2}{(1+2\rho)}} \end{bmatrix}$$

Recalling that

$$(\underline{B} + \underline{C})(\underline{B} + \underline{C})^T = \underline{\beta}\,\underline{\beta}^T \qquad (A4.2.4)$$

$$(\underline{B} - \underline{C})(\underline{B} - \underline{C})^T = \underline{\beta}^*\underline{\beta}^{*T} \qquad (A4.2.5)$$

the elements of $\underline{B}$ and $\underline{C}$ are formed by solving pairs of simultaneous equations

$$b_{11} = \tfrac{1}{2}\left[\sqrt{1+2\rho} + \sqrt{1-2\rho}\right] \qquad (A4.2.6)$$

$$c_{11} = \tfrac{1}{2}\left[\sqrt{1-2\rho} - \sqrt{1+2\rho}\right] \qquad (A4.2.7)$$

$$b_{21} = \tfrac{1}{2}\left[\frac{(R-\beta-\alpha)}{\sqrt{1-2\rho}} + \frac{(R+\beta+\alpha)}{\sqrt{1+2\rho}}\right] \qquad (A4.2.8)$$

$$c_{21} = \tfrac{1}{2}\left[\frac{(R-\beta-\alpha)}{\sqrt{1-2\rho}} - \frac{(R+\beta+\alpha)}{\sqrt{1+2\rho}}\right] \qquad (A4.2.9)$$

$$b_{22} = \tfrac{1}{2}\left[\sqrt{\frac{(1-2r)(1-2\rho) - (R-\beta-\alpha)^2}{1-2\rho}} + \sqrt{\frac{(1+2r)(1+2\rho) - (R+\beta+\alpha)^2}{1+2\rho}}\right] \qquad (A4.2.10)$$

$$c_{22} = \tfrac{1}{2}\left[\sqrt{\frac{(1-2r)(1-2\rho) - (R-\beta-\alpha)^2}{1-2\rho}} - \sqrt{\frac{(1+2r)(1+2\rho) - (R+\beta+\alpha)^2}{1+2\rho}}\right] \qquad (A4.2.11)$$

By applying expectation theory to equation (A4.2.1) it may be shown that

$$E\left[x_1^2(t)\right] = b_{11}^2 + c_{11}^2 \qquad (A4.2.12)$$

$$= \frac{1}{4}\left[1+2\rho + 1-2\rho + 2\sqrt{1-2\rho}\sqrt{1+2\rho} + 1-2\rho + 1+2\rho - 2\sqrt{1-2\rho}\sqrt{1+2\rho}\right]$$

$$= 1$$

on substitution from equations (A4.2.6) and (A4.2.7).

Similarly

$$E\left[x_2^2(t)\right] = b_{21}^2 + b_{22}^2 + c_{21}^2 + c_{21}^2 + c_{22}^2 \qquad (A4.2.13)$$

which, on substitution from equations (A4.2.8) - (A4.2.11) reduces to

$$E\left[x_2^2(t)\right] = \frac{1}{2}\left[\frac{(R-\beta-\alpha)^2}{1-2\rho} + \frac{(R+\beta+\alpha)^2}{1+2\rho} + \frac{(1-2r)(1-2\rho) - (R-\beta-\alpha)^2}{1-2\rho} + \frac{(1+2r)(1+2\rho) - (R+\beta+\alpha)^2}{1+2\rho}\right]$$

$$= 1$$

Again applying expectation theory in equation (A4.2.1) it may be shown that

$$E\left[x_1(t)\, x_2(t)\right] = E\left[x_2(t)\, x_1(t)\right] = b_{21}\, b_{11} + c_{21}\, c_{11} \qquad (A4.2.14)$$

$$E\left[x_1(t)\, x_1(t-1)\right] = -c_{11}\, b_{11} \qquad (A4.2.15)$$

$$E\left[x_2(t)\, x_2(t-1)\right] = -c_{21}\, b_{21} - c_{22}\, b_{22} \qquad (A4.2.16)$$

$$E\left[x_1(t)\, x_2(t-1)\right] = -c_{11}\, b_{21} \qquad (A4.2.17)$$

$$E\left[x_2(t)\, x_1(t-1)\right] = -c_{21}\, b_{11} \qquad (A4.2.18)$$

On substitution from equations (A4.2.6) - (A4.2.11), equations (A4.2.14) - (A4.2.18) reduce to

$$\rho_{12}(0) = \rho_{21}(0) = R \qquad (A4.2.19)$$

$$\rho_1(1) = \rho \qquad (A4.2.20)$$

$$\rho_2(1) = r \qquad (A4.2.21)$$

$$\rho_{12}(1) = \frac{1}{4}\left[(2\beta+2\alpha) - \frac{(R+\beta+\alpha)\sqrt{1-2\rho}}{\sqrt{1+2\rho}} + \frac{(R-\beta-\alpha)\sqrt{1+2\rho}}{\sqrt{1-2\rho}}\right] \qquad (A4.2.22)$$

$$\rho_{21}(1) = \frac{1}{4}\left[(2\beta-2\alpha) - \frac{\sqrt{1+2\rho}\,(R-\beta-\alpha)}{\sqrt{1-2\rho}} + \frac{\sqrt{1-2\rho}\,(R+\beta+\alpha)}{\sqrt{1+2\rho}}\right] \qquad (A4.2.23)$$

Hence, the observed lag-one cross-correlations will not be preserved, although their sum

$$\rho_{12}(1) + \rho_{21}(1) = \alpha + \beta$$

will be preserved. This result is also true in the general case f m sites.

Appendix 4.3

Correlations Preserved by the Modified Multisite ARIMA (1,0,1) Process

The case of 2 sites will again be used to illustrate the derivation. The modified multisite ARIMA (1,0,1) process is then written as

$$\begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \begin{bmatrix} \varphi & - \\ - & \varphi \end{bmatrix} \begin{bmatrix} x_1(t-1) \\ x_2(t-1) \end{bmatrix} + \begin{bmatrix} b_{11} & 0 \\ b_{21} & b_{22} \end{bmatrix} \begin{bmatrix} \varepsilon_1(t) \\ \varepsilon_2(t) \end{bmatrix} - \begin{bmatrix} c_{11} & 0 \\ c_{21} & c_{22} \end{bmatrix} \begin{bmatrix} \varepsilon_1(t-1) \\ \varepsilon_2(t-1) \end{bmatrix} \quad \text{(A4.3.1)}$$

Let the observed lag-zero and lag-one correlation matrices be denoted by

$$\underline{M}_o = \begin{bmatrix} 1 & R \\ R & 1 \end{bmatrix} \qquad \underline{M}_1 = \begin{bmatrix} \rho & \alpha \\ \beta & r \end{bmatrix}$$

From equations (4.45) and (4.46)

$$\underline{S} + \underline{T} + \underline{T}^T = \begin{bmatrix} (1+\varphi)^2 - 2\rho(1+\varphi) & (1+\varphi)^2 R - (\alpha+\beta)(1+\varphi) \\ (1+\varphi)^2 R - (\alpha+\beta)(1+\varphi) & (1+\varphi)^2 - 2r(1+\varphi) \end{bmatrix}$$

$$= \underline{W} = \underline{\beta}\,\underline{\beta}^T \quad \text{(A4.3.2)}$$

$$\underline{S} - \underline{T} - \underline{T}^T = \begin{bmatrix} (1-\varphi)^2 + 2\rho(1-\varphi) & (1-\varphi)^2 R + (\alpha+\beta)(1-\varphi) \\ (1-\varphi)^2 R + (\alpha+\beta)(1-\varphi) & (1-\varphi)^2 + 2r(1-\varphi) \end{bmatrix}$$

$$= \underline{W}^* = \underline{\beta}^*\underline{\beta}^{*T} \quad \text{(A4.3.3)}$$

Using lower triangularization, equations (A4.3.2) and (A4.3.3) may be solved for the coefficients of the matrices $\underline{B}$ and $\underline{C}$ as illustrated in Appendix (4.2) to give

$$b_{11} = (\beta_{11} + \beta^*_{11})/2 = (\sqrt{w}_{11} + \sqrt{w}^*_{11})/2$$

$$c_{11} = (\beta_{11} - \beta^*_{11})/2 = (\sqrt{w}_{11} - \sqrt{w}^*_{11})/2$$

$$b_{21} = (\beta_{21} + \beta^*_{21})/2 = \left(\frac{w_{21}}{\beta_{11}} + \frac{w^*_{21}}{\beta^*_{11}}\right)/2$$

$$c_{21} = (\beta_{21} - \beta^*_{21})/2 = \left(\frac{w_{21}}{\beta_{11}} - \frac{w^*_{21}}{\beta^*_{11}}\right)/2$$

$$b_{22} = (\beta_{22} + \beta^*_{22})/2 = \left(\sqrt{w_{22} - \beta^2_{21}} + \sqrt{w^*_{22} - \beta^{*2}_{21}}\right)/2$$

$$c_{22} = (\beta_{22} - \beta^*_{22})/2 = \left(\sqrt{w_{22} - \beta^2_{21}} - \sqrt{w^*_{22} - \beta^{*2}_{21}}\right)/2$$

From equation (4.42),

$$\rho_{12}(0) = \frac{1}{1-a_{11}a_{22}}\left[- a_{11}\, c_{21}\, b_{11} - a_{22}\, c_{11}\, b_{21} + b_{11}\, b_{21} + c_{11}\, c_{21}\right] \qquad \text{(A4.3.4)}$$

$$c_{21}\, b_{11} = \frac{1}{4}\left[w_{21} - \frac{w^*_{21}\sqrt{w_{11}}}{\sqrt{w^*_{11}}} + \frac{w_{21}\sqrt{w^*_{11}}}{\sqrt{w_{11}}} - w^*_{21}\right]$$

$$c_{11}\, b_{21} = \frac{1}{4}\left[w_{21} - \frac{w_{21}\sqrt{w^*_{11}}}{\sqrt{w_{11}}} + \frac{w^*_{21}\sqrt{w_{11}}}{\sqrt{w^*_{11}}} - w^*_{21}\right]$$

$$b_{11}\, b_{21} = \frac{1}{4}\left[w_{21} + \frac{\sqrt{w^*_{11}}\, w_{21}}{\sqrt{w_{11}}} + \frac{\sqrt{w_{11}}\, w^*_{21}}{\sqrt{w^*_{11}}} + w^*_{21}\right]$$

$$c_{11}\, c_{21} = \frac{1}{4}\left[w_{21} - \frac{\sqrt{w_{11}}\, w^*_{21}}{\sqrt{w^*_{11}}} - \frac{\sqrt{w^*_{11}}\, w_{21}}{\sqrt{w_{11}}} + w^*_{21}\right]$$

Noting that $a_{11} = a_{22} = \varphi$, equation (A4.3.4) reduces to

$$\rho_{12}(0) = \frac{1}{2(1-\varphi^2)}\left[- \varphi\,(w_{21} - w^*_{21}) + w_{21} + w^*_{21}\right] \qquad \text{(A4.3.5)}$$

Substitution for w_{21} and w^*_{21} from equations (A4.3.2) and (A4.3.3) gives

$$\rho_{12}(0) = \frac{1}{2(1-\varphi^2)}\left[2(1-\varphi^2)R\right]$$

$$= R$$

Similarly, it can be shown that the variances at each site will be preserved as unity. From equation (4.43) the lag-one autocorrelation at

site 1 is given as

$$\rho_{11}(1) = \varphi - c_{11}\, b_{11} \qquad \text{(A4.3.6)}$$

which, on substitution, reduces to

$$\rho_{11}(1) = \rho$$

Similarly,

$$\rho_{22}(1) = \varphi - c_{21}\, b_{21} - c_{22}\, b_{22} = r \qquad \text{(A4.3.7)}$$

Thus, the modified ARIMA (1,0,1) multisite process preserves the observed $\underline{M}_0$ matrix and the observed diagonal elements of the $\underline{M}_1$ matrix in the case of 2 sites.

The foregoing treatment can readily be extended to verify the result in the general case of m sites.

Appendix 4.4

Cross correlations for a multisite log-normal process

Let

$$Y_i(t) = \ln (X_i(t) - a_i)$$

where ln denotes $\log_e$, and a_i denotes the lower bound of $X_i(t)$ at site i. Thus,

$$X_i(t) = a_i + \exp\ Y_i(t) \qquad (A4.4.1)$$

The lag-zero cross-correlation between $X_i(t)$ and $X_j(t)$ is defined as

$$\rho_{ij}(0) = \frac{E\,[X_i(t)\ X_j(t)]\ - E\,[X_i(t)]\ E\,[X_j(t)]}{[E\,[X_i^2(t)]\ - (E\,[X_i(t)])^2]^{0.5}\,[E\,[X_j^2(t)]\ - (E\,[X_j(t)])^2]^{0.5}}$$

(A4.4.2)

Using the moment generating function defined by equation (A3.1.4) in Appendix (A3.1), it may be shown that

$$E\,[X_i^2(t)\ - (E\,[X_i(t)])^2] = \exp\,[2\mu_i' + \sigma_i'^2]\,[\exp(\sigma_i'^2) - 1] \qquad (A4.4.3)$$

$$E\,[X_j^2(t)\ - (E\,[X_j(t)])^2] = \exp\,[2\mu_j' + \sigma_j'^2]\,[\exp(\sigma_j'^2) - 1] \qquad (A4.4.4)$$

Using the moment generating function defined by equation (A3.1.5) in Appendix (A3.1) it may be shown that

$$\begin{aligned} E\,[X_i(t)]\,[X_j(t)] = {} & a_i\ a_j + a_j\ \exp(\mu_i' + \sigma_i'^2/2) \\ & + a_i\ \exp(\mu_j' + \sigma_j'^2/2) \\ & + \exp\left[\mu_i' + \mu_j' + \tfrac{1}{2}(\sigma_i'^2 + 2\rho_{ij}'(0)\ \sigma_i'\ \sigma_j' + \sigma_j'^2)\right] \end{aligned}$$

(A4.4.5)

whence

$$E\left[X_i(t)\,X_j(t)\right] - E\left[X_i(t)\right]E\left[X_j(t)\right]$$

$$= \exp\left[\mu'_i + \mu'_j + \tfrac{1}{2}(\sigma'^{2}_i + 2\,\rho'_{ij}(0)\,\sigma'_i\,\sigma'_j + \sigma'^{2}_j)\right]$$

$$- \exp\left[\mu'_i + \mu'_j + \sigma'^{2}_i/2 + \sigma'^{2}_j/2\right]$$

$$= \exp\left[\mu'_i + \mu'_j + \sigma'^{2}_i/2 + \sigma'^{2}_j/2\right].$$

$$\left[\exp\ \sigma'_i\,\sigma'_j\ \rho'_{ij}(0)\ - 1\right]$$

$$\therefore\ \rho_{ij}(0) = \frac{\exp\left[\sigma'_i\,\sigma'_j\ \rho'_{ij}(0)\ - 1\right]}{\left[e^{\sigma'^{2}_i} - 1\right]^{0.5}\left[e^{\sigma'^{2}_j} - 1\right]^{0.5}} \qquad \text{(A4.4.6)}$$

Again, equation (A4.4.6) would appear to be independent of the process used to model the cross-correlations $\rho'_{ij}(0)$. Similarly, it can be shown that

$$\rho_{ij}(1) = \frac{\exp\left[\sigma'_i\,\sigma'_j\ \rho'_{ij}(1)\ - 1\right]}{\left[e^{\sigma'^{2}_i} - 1\right]^{0.5}\left[e^{\sigma'^{2}_j} - 1\right]^{0.5}} \qquad \text{(A4.4.7)}$$

$$\rho_{ij}(2) = \frac{\exp\left[\sigma'_i\,\sigma'_j\ \rho'_{ij}(2)\ - 1\right]}{\left[e^{\sigma'^{2}_i} - 1\right]^{0.5}\left[e^{\sigma'^{2}_j} - 1\right]^{0.5}} \qquad \text{(A4.4.8)}$$

Appendix 5.1

Details of the optimization process

The optimization process developed by Rosenbrock (1960), with modifications by McConalogue of Imperial College, was employed to minimise the sum-of-squares function $S(\varphi, \theta)$. Some minor modifications were also introduced during the course of a previous study (O'Connell et al, 1970). The search geometry of Rosenbrock's original method remains unchanged but modifications have been made in the manner in which a minimum is found in each of the orthogonal directions. The search is also conducted in unconstrained space, and consequently constraints or transformations must be applied to the parameters to ensure that the minimum of $S(\varphi, \theta)$ lies in the stationary invertible region.

The initial directions searched correspond to the axes of the variables. When all the directions have been searched once, new directions are defined, one of which is the direction of advance during the first iteration (i.e. the vector joining the initial and final points) and the others are orthogonal to this. New searches are made in these directions and when new minima have been estimated the directions are redefined as before and so on.

The minimum along each direction is estimated by calculating $S(\varphi, \theta)$ at a series of points. At the start of each linear search, the variable is altered by 2 per cent and $S(\varphi, \theta)$ is computed again. If an initial failure (i.e. an increase in the value of $S(\varphi, \theta)$) is registered, the direction of search is reversed. If a success is indicated by a decrease in $S(\varphi, \theta)$, the last value of the parameter is altered by 3 per cent, then by 4.5 per cent and the magnification of the steps continues until a failure is registered. The minimum is predicted by quadratic interpolation of the three best values of $S(\varphi, \theta)$ using finite difference

approximations; if the estimation of the minimum is found to be within a certain tolerance the next direction is searched from this point. A search of all the current directions constitutes an iteration. When the function $S(\varphi, \theta)$ ceases to change significantly, a minimum is assumed to have been found, and the search is terminated by means of a convergence criterion. Convergence is assumed to have taken place when the change in $S(\varphi, \theta)$ between iterations is less than 1×10^{-6} of its current value.

Appendix 5.2

Constraints on the parameters φ and θ.

As pointed out in Appendix (5.1), the Rosenbrock search technique conducts an unconstrained search for a minimum (or maximum) in n-dimensional space. However, the search for the minimum of the function $S(\varphi, \theta)$ must be confined to the parameter space defined by the stationarity and invertibility conditions.

A number of techniques exist whereby constraints may be placed on parameters during an optimization search. A straightforward approach is to employ transformations (Box, 1966) which allow a search to be carried out in an unconstrained space which corresponds to the required constrained space. If the search for the minimum of a parameter v_i is to be kept between a lower limit g_i and an upper limit m_i then a transformation on the corresponding unconstrained parameter u_i is carried out as

$$v_i = g_i + (m_i - g_i) \operatorname{Sin}^2 u_i \qquad (A5.2.1)$$

Hence an unconstrained search may be carried out in u-space, while the value of v_i ranges between g_i and m_i according as $\operatorname{Sin}^2 u_i$ ranges between 0 and 1. Box (1966) has observed that for equation (A5.2.1) the neighbourhood of any point U in u-space maps into the neighbourhood of V in v-space, where U maps into V. While equation (A5.2.1) does not represent a 1 : 1 transformation, additional local optima cannot be introduced. The periodicity of optimal solutions in u-space should not cause any difficulty provided the optimization process does not take steps so large that it jumps from peak to peak.

As noted in section (5.3.1), parameter limits $-1 + \varepsilon < \varphi < 1 - \varepsilon$, where ε is some small positive quantity, are necessary for the autoregressive parameter φ to avoid numerical problems. For the sake of consistency, similar limits were imposed on the moving average parameter θ. A value

of $\varepsilon = 0.01$ was arbitrarily selected yielding limits

$$\varphi = -0.99 + 1.98 \text{ Sin}^2 u_1 \qquad \text{(A5.2.2)}$$

$$\theta = -0.99 + 1.98 \text{ Sin}^2 u_2 \qquad \text{(A5.2.3)}$$

As initial estimates of φ and θ are available, the inverses of equations (A5.2.2) and (A5.2.3) are obtained to define initial values for u_1 and u_2 to initiate the search.

In situations where no acceptable initial moment estimates are available for either φ or θ or both, an empirical approach to providing initial estimates is employed. For example, suppose that the initial moment estimates are $\hat{\varphi}^{(m)} = 1.56$ and $\hat{\theta}^{(m)} = 0.43$. The initial value of φ for the ML search is set to 0.95 while the value of θ remains at 0.43. These values are then used as initial estimates. A similar strategy is employed for the parameter θ. The search is initiated at a distance equal to 0.05 from the boundary in all cases in order to allow the search technique room to manouvre. Even if either or both the moment estimates are within the admissible region, but are closer than 0.05 to a boundary, the initial estimates are set to values within 0.05 of the boundary. Whenever both parameters are out of bounds, only the signs of those parameters can be used for guidance on initial estimates.

In order to guard against solutions being found on or very close to boundaries, which would probably indicate that the ML estimates lay out of bounds in any case, solutions were accepted only if both parameters were not any closer to a boundary than 0.015. Extensive inspections of computer printouts of parameter searches suggested that these rather arbitrary procedures worked satisfactorily.

Obviously most of the rules considered here are empirical but they should help to ensure that only valid solutions are found. Because of

the nature of the experiments conducted, detailed examination of each $S(\varphi, \theta)$ surface as advocated by Box and Jenkins (1970) was not possible, and a certain price may perhaps have been paid for the benefit of the automated procedure employed. However, in this situation the identity of the process generating the observations is known and perhaps the automated procedure may be justified on these grounds.

REFERENCES

AITCHISON, J., BROWN, J.A.C. (1957), The lognormal distribution with special reference to its use in economics, Cambridge, University Press, pp 176.

ANDERSON, R.L. (1942), Distribution of the serial correlation coefficient, Ann. Math. Stat., 13, 1-13.

ANIS, A.A. (1955), The variance of the maximum of partial sums of a finite number of independent normal variates, Biometrika, 42, 96-101.

ANIS, A.A. (1956), On the moments of the maximum of partial sums of a finite number of independent normal variates, Biometrika, 43, 79-84.

ANIS, A.A., LLOYD, E.H. (1953), On the range of partial sums of a finite number of independent normal variates, Biometrika, 40, 35-42.

BARNARD, G.A. (1956), Discussion of "Methods of using long-term storage in reservoirs", by H.E. Hurst, Proc. Inst. Civ. Engrs., 1, 552-553.

BARNES, F.B. (1954), Storage required for a city water supply, J. Instn. Engrs. Austral., 26, 198-203.

BARTLETT, M.S. (1966), An introduction to stochastic processes with special reference to methods and applications, 2nd Ed., Cambridge, University Press, pp. 362.

BEARD, L.R. (1965), Use of interrelated records to simulate streamflow, Proc. Am. Soc. Civ. Engrs., J. Hydraul. Div., 91(HY5), 13-22.

BENSON, M.A., MATALAS, N.C. (1967), Synthetic hydrology based on regional statistical parameters, Water Resour. Res. 3(4), 931-935.

BOX, G.E.P., DRAPER, N.R. (1965), The Bayesian estimation of common parameters from several responses, Biometrika, 52, 355-365.

BOX, G.E.P., JENKINS, G.M. (1970), Time series analysis : forecasting and control, San Francisco, Holden-Day Inc., pp. 553.

BOX, M.J. (1966), A comparison of several current optimization methods, and the use of transformations in constrained problems, Computer J., 9, 67-77.

CARLSON, R.F., MACCORMICK, A.J.A., WATTS, D.G. (1970), Application of linear random models to four annual streamflow series, Water Resour. Res. 6(4), 1070-1078.

CARLSON, R.F., WATTS, D.G. (1971), Reply to comments by B.B. Mandelbrot on 'Application of linear random models to four annual streamflow series' by R.F. Carlson, A.J.A. Maccormick and D.G. Watts, Water Resour. Res. 7(5), 1363-1364.

CLARKE, R.T. (1971), The use of the term 'linearity' as applied to hydrological models, J. Hydrol. 13, 91-95.

COX, D.R., MILLER, H.D. (1966), The theory of stochastic processes, London, Methuen and Co. Ltd., pp. 398.

CROSBY, D.S., MADDOCK, T. III (1970), Estimating coefficients of a flow generator for monotone samples of data, Water Resour. Res. 6(4), 1079-1086.

DITLEVSEN, O. (1971), Extremes and first passage times with applications in Civil Engineering, Ph.D. Thesis, Copenhagen, Tech. Univ. of Denmark.

DURBIN, J. (1969), Testing for serial correlation in least-squares regression when some of the regressors are lagged dependent variables, Econometrica, 37.

FELLER, W. (1951), The asymptotic distribution of the range of sums of independent random variables, Ann. Math. Stat., 22, 427-432.

FIERING, M.B. (1964), Multivariate techniques for synthetic hydrology, Proc. Am. Soc. Civ. Engrs., J. Hydraul. Div., 90(HY5), 43-60.

FIERING, M.B. (1967), Streamflow synthesis, London, The McMillan Press Ltd., pp. 139.

FIERING, M.B. (1968), Schemes for handling inconsistent matrices, Water Resour. Res. 4(2), 291-297.

FIERING, M.B. JACKSON, B.B. (1971), Synthetic streamflows, Water Resources Monograph 1, Washington, D.C., American Geophysical Union, pp. 98.

GARCIA, L.E., DAWDY, D.R., MEJIA, J.M. (1972), Long memory monthly streamflow simulation by a Broken Line model, Water Resour. Res. 8(4), 1100-1105.

HAZEN, A. (1914), Storage to be provided in impounding reservoirs for municipal water supply, Trans. Am. Soc. Civ. Engrs., 77, 1539-1669.

HURST, H.E. (1951), Long-term storage capacity of reservoirs, Trans. Am. Soc. Civ. Engrs., 116, 770-808.

HURST, H.E. (1956), Methods of using long-term storage in reservoirs, Proc. Instn. Civ. Engrs., 1, 519-543.

HURST, H.E. (1957), A suggested statistical model of some time series which occur in nature, Nature, 180, p. 494.

JACOBI, H.D., LOUCKS, D.P. (1972), Combined use of optimization and simulation models in river basin planning, Water Resour. Res., 8(6), 1401-1414.

JENKINS, G.M., WATTS, D.G. (1968), Spectral analysis and its applications, San Francisco, Holden-Day Inc., pp. 525.

KENDALL, M.G. (1954), Note on the bias in the estimation of autocorrelation, Biometrika, 42, 403-404.

KING, J.W. (1973), Solar radiation changes and the weather, Nature, 245, 443-446.

KIRBY, W. (1972), Computer-oriented Wilson Hilferty transformation that preserves the first three moments and lower bound of the Pearson Type 3 distribution, Water Resour. Res., 8(5), 1251-1254.

KLEMES, V. (1974), The Hurst phenomenon - a puzzle? Manuscript, pp. 53, to be published.

KRITSKY, S.N., MENKEL, M.F. (1970), The regularities of long-term fluctuations in streamflow as a component of the water balance. Proc. Int. Symposium on World Water Balance, Reading, Vol. 2, Int. Ass. Hydrol. Sci. Pub. 93, pp. 494-501.

LANGBEIN, W.B. (1956), Discussion of "Methods of using long-term storage in reservoirs", by H.E. Hurst, Proc. Instn. Civ. Engrs., 1, 565-568.

LLOYD, E.H. (1967), Stochastic reservoir theory, Advances in Hydroscience, ed. V.T. Chow, Vol. 4, pp. 281-339.

MAAS ET AL, (1962), Design of water resource systems, Cambridge, Mass., Harvard Univ. Press, pp. 660.

MANDELBROT, B.B. (1965), Une classe de processus homothetiques a soi ; application a la loi climatologique de H.E. Hurst, Compt. Rend. Acad. Sci. Paris, 260, 3274-3277.

MANDELBROT, B.B. (1971a), A fast fractional Gaussian noise generator, Water Resour. Res. 7(3), 543-553.

MANDELBROT, B.B. (1971b), Comments on 'Application of linear random models to four annual streamflow series' by R.F. Carlson, A.J.A. Maccormick, and D.G. Watts, Water Resour. Res. 7(5), 1360-1362.

MANDELBROT, B.B. (1972), Broken Line process derived as an approximation to fractional noise, Water Resour. Res. 8(5), 1354-1356.

MANDELBROT, B.B., VAN NESS, J.W. (1968), Fractional Brownian motions, fractional noises and applications, Soc. Ind. Appl. Math. Rev., 10(4), 422-437.

MANDELBROT, B.B., WALLIS, J.R. (1968), Noah, Joseph, and operational hydrology, Water Resour. Res. 4(5), 909-918.

MANDELBROT, B.B., WALLIS, J.R. (1969a), Computer experiments with fractional Gaussian noises. Part 1 - Averages and variances, Water Resour. Res. 5(1), 228-241.

MANDELBROT, B.B., WALLIS, J.R. (1969b), Computer experiments with fractional Gaussian noises. Part 2 - Rescaled ranges and spectra, Water Resour. Res. 5(1), 242-259.

MANDELBROT, B.B., WALLIS, J.R. (1969c), Computer experiments with fractional Gaussian noises. Part 3 - Mathematical appendix, Water Resour. Res. 5(1), 260-267.

MANDELBROT, B.B., WALLIS, J.R. (1969d), Some long-run properties of geophysical records, Water Resour. Res. 5(2), 321-340.

MANDELBROT, B.B., WALLIS, J.R. (1969e), Robustness of the rescaled range R/S in the measurement of non-cyclic long-run statistical dependence, Water Resour. Res. 5(5), 967-988.

MARQUARDT, D.W. (1963), An algorithm for least squares estimation of non-linear parameters, J. Soc. Ind. Appl. Math., 11, 431-441.

MATALAS, N.C. (1962), Statistical properties of tree ring data, Bull. Int. Ass. Sci. Hydrol., 7(2), 39-47.

MATALAS, N.C. (1966), Time series analysis, Water Resour. Res. 3(3), 817-829.

MATALAS, N.C. (1967), Mathematical assessment of synthetic hydrology, Water Resour. Res. 3(4), 931-935.

MATALAS, N.C., HUZZEN, C.S. (1967), A property of the range of partial sums, Proc. International Hydrology Symposium, Fort Collins, Colorado State University, Vol. 1, pp. 252-257.

MATALAS, N.C., LANGBEIN, W.B. (1962), Information content of the mean, Jour. Geophys. Res. 67(9), 3441-3448.

MATALAS, N.C., WALLIS, J.R. (1971a), Correlation constraints for generating processes. Proc. International Symposium on Mathematical Models in Hydrology, Warsaw, Int. Ass. Sci. Hydrol., in press.

MATALAS, N.C., WALLIS, J.R. (1971b), Statistical properties of multivariate fractional noise processes, Water Resour. Res. 7(6), 1460-1468.

MATALAS, N.C., WALLIS, J.R. (1974), Generation of synthetic flow sequences, Systems Approach to Water Management, Ed. A.K. Biswas, New York, American Elselvier, in press.

MCMAHON, T.A., MILLER, A.J. (1971), Application of the Thomas and Fiering model to skewed hydrologic data, Water Resour. Res., 7(5), 1338-1340.

MEJIA, J.M., RODRIGUEZ-ITURBE, I., DAWDY, D.R. (1972), Streamflow simulation. 2 - The broken line process as a potential model for hydrologic simulation,
Water Resour. Res. 8(4), 931-941.

MOODY, D.W. (1973), Application of multi-regional planning models to the scheduling of large-scale water resources systems development, Paper presented at Symposium on the Control of Water Resources Systems, Haifa, Israel, pp. 36, International Federation for Automatic Control.

MORAN, P.A.P. (1959), The theory of storage,
London, Methuen and Co. Ltd., pp. 111.

MORAN, P.A.P. (1964), On the range of cumulative sums,
Ann. Inst. Stat. Math., 16, 109-112.

MOSS, M.E. (1972a), Serial-correlation structure of discretized streamflow, Open-file Report, U.S. Geol. Surv., Water Resour. Div., Fort Collins, Colorado, pp. 83.

MOSS, M.E. (1972b), Response to General Report by R.T. Clarke on "Reduction in uncertainties in autocorrelation by the use of physical models" by M.E. Moss,
Proc. International Symposium on Uncertainties in Hydrologic and Water Resource Systems, University of Arizona, Tucson, Vol. 3, 1361-1362.

NORDIN, C., MCQUIVEY, R.S., MEJIA, J.M. (1972), Hurst phenomenon in turbulence, Water Resour. Res. 8(6), 1480-1486.

O'CONNELL, P.E. (1971), A simple stochastic modelling of Hurst's law,
Proc. International Symposium on Mathematical Models in Hydrology, Warsaw, Int. Ass. Hydrol. Sci., in press.

O'CONNELL, P.E., NASH, J.E., FARRELL, J.P. (1970), River flow forecasting through conceptual models : Part II - the Brosna catchment at Ferbane,
J. Hydrol. 10, 317-329.

O'CONNELL, P.E., WALLIS, J.R. (1973), Choice of generating mechanism in synthetic hydrology with inadequate data,
Proc. International Symposium on the design of water resources projects with inadequate data, Madrid, UNESCO/WMO, to be published.

PARZEN, E. (1960), Modern probability theory and its applications,
New York, John Wiley and Sons Inc., pp. 464.

PRABHU, N.U. (1965), Stochastic processes : basic theory and its applications, New York, The Macmillan Company, pp. 233.

QUENOUILLE, M.H. (1949), A large sample test for the goodness of fit of autoregressive schemes,
J. Roy. Stat. Soc., 110, 123-129.

RAO, C.R. (1965), Linear statistical inference and its applications, New York, John Wiley and Sons Inc., pp. 522

RIPPL, W. (1883), The capacity of storage reservoirs from water supply, Proc. Instn. Civ. Engrs., 71, 270-278.

RODRIGUEZ-ITURBE, I., MEJIA, J.M., DAWDY, D.R. (1972), Streamflow simulation. 1 - A new look at Markovian models, fractional Gaussian noise and crossing theory, Water Resour. Res. 8(4), 921-930.

ROEFS, T.G. (1968), Reservoir management ; the state of the art, Pub. no. 320-3508, IBM Scientific Center, Washington, D.C., pp. 85.

ROSENBROCK, H.H. (1960), An automatic method of finding the greatest or least value of a function, Computer Journal, 3, 175-184.

SCHULMAN, E. (1956), Dendroclimatic change in semi-arid America, Tucson, University of Arizona Press, pp. 142.

SEXTON, J.R., JAMIESON, D.G. (1973), Improved techniques for water resource systems design, Proc. International Symposium on the design of water resources projects with inadequate data, Madrid, UNESCO/WMO, to be published.

SLACK, J.R. (1972), Bias, illusion and denial as data uncertainties, International Symposium on Uncertainties in Hydrologic and Water Resource Systems, University of Arizona, Tucson, Vol. 1, pp. 122-132.

SLACK, J.R. (1973), I would if I could (Self-denial by conditional models), Water Resour. Res. 9(1), 247-249.

SOLARI, M.E., ANIS, A.A. (1957), The mean and variance of the maximum of the adjusted partial sums of a finite number of independent normal variates, Ann. Math. Stat., 28, 706-716.

SUDLER, C.E. (1927), Storage required for the regulation of streamflow, Trans. Am. Soc. Civ. Engrs., 91, 622-704.

TAYLOR, G.I. (1921), Diffusion by continuous movements, Proc. London Math. Soc., Ser. 2, 20, 196-211.

THOMAS, H.A., BURDEN, R.P. (1963), Statistical analysis of the reservoir yield relation, report, chap. 1, pp. 1-21, Cambridge, Mass., Harvard Water Resources Group.

THOMAS, H.A., FIERING, M.B. (1962), Mathematical synthesis of streamflow sequences for the analysis of river basins by simulation, A. Maas et al, "Design of Water Resource Systems", chap. 12, Cambridge, Mass., Harvard University Press.

VALENCIA, D., SCHAAKE, J.C. Jr. (1972), Disaggregation processes in stochastic hydrology,
Water Resour. Res., 9(3), 580-585.

WALLIS, J.R. (1971), Comment on "A simple stochastic modelling of Hurst's law" by P.E. O'Connell,
Proc. International Symposium on Mathematical Models in Hydrology, Warsaw, Int. Ass. Hydrol. Sci., in press.

WALLIS, J.R., MATALAS, N.C. (1970), Small sample properties of H and K-estimators of the Hurst coefficient h,
Water Resour. Res. 6(6), 1583-1594.

WALLIS, J.R., MATALAS, N.C. (1971), Correlogram analysis revisited,
Water Resour. Res. 7(6), 1448-1459.

WALLIS, J.R., MATALAS, N.C. (1972), Sensitivity of reservoir design to the generating mechanism of inflows,
Water Resour. Res. 8(3), 634-641.

WALLIS, J.R., O'CONNELL, P.E. (1972), Small sample estimation of ρ_1,
Water Resour. Res. 8(3), 707-712.

WALLIS, J.R., O'CONNELL, P.E. (1973), Firm reservoir yield : how reliable are historic hydrological records,
Bull. Int. Ass. Hydrol. Sci., 18(3), 347-365.

WATTS, D.G. (1972), Discussion on "Evaluation of seasonal time-series models : application to mid-west river flow data" by A.I. McKerchar and J. Delleur,
Proc. International Symposium on Uncertainties in Hydrologic and Water Resource Systems, University of Arizona, Tucson, Vol. 3, pp. 1407-1409.

WEISS, G. (1973), Filtered Poisson processes as models for daily streamflow data,
Ph.D. Thesis, London, Imperial College, University of London, pp. 138.

WHITTLE, P. (1952), Tests of fit in time series,
Biometrika, 39, 309-318.

WHITTLE, P. (1954), Estimation and information in stationary time series,
Arkov för Matematik, 2, 423-434.

YEVJEVICH, V.M. (1963), Fluctuations of wet and dry years : Part I - Research data assembly and mathematical models,
Hydrology Paper 1, Colorado State Univ., Fort Collins, pp. 55.

YEVJEVICH, V.M. (1967), Mean range of linearly dependent normal variables with application to storage problems,
Water Resour. Res. 3(3), 663-671.

YOUNG, G.K. (1968), Discussion of "Mathematical Assessment of Synthetic Hydrology, Water Resour. Res. 4(3), 681-682.

YULE, G.U. (1927), On a method of investigating periodicities in disturbed series, with special reference to Wölfer's sunspot numbers,
Phil. Trans., A226, 267.

SHORTENED LIST OF SYMBOLS AND ABBREVIATIONS

ARIMA	Abbreviation for 'autoregressive integrated moving average'
dfGn	Abbreviation for 'discrete-time fractional Gaussian noise'
$E(\cdot)$	Expectation operator
H	Estimate of h
h	Hurst coefficient
K	Estimate of h
n	Sample duration in discrete time
p	Order of autoregressive process
q	Order of moving average process
R, R_n	Adjusted range or range of cumulative departures from sample mean in discrete time
R_p	Range of cumulative departures from population mean in discrete time
R/S	Rescaled range in discrete time
S	Sample standard deviation
$S(\varphi,\theta)$	Sum of squares function for ARIMA (1,0,1) process
X_t	Discrete-time random variable at time t
γ_x	Skewness of random variable X_t
ε_t	Normally and independently distributed random variable at time t with zero mean and unit variance
θ	Parameter of first order moving average
μ_x	Population mean of random variable X_t
ρ, ρ_1	Lag-one autocorrelation coefficient
ρ_k	Lag-k autocorrelation coefficient
ρ_x	Lag-one autocorrelation coefficient of random variable X_t
$\rho_{ij}(k)$	Lag-k cross correlation coefficient between random variables $X_i(t)$ and $X_j(t)$
σ_x	Population standard deviation of the random variable X_t
φ	Parameter of first order autoregressive process