STOCHASTIC PROCESSES IN ONE-DIMENSIONAL SERIES: AN INTRODUCTION

K. S. Richards

ISSN 0 306 6142
1S81 0 86094 030 6

Printed by Mid-Anglia Litho

CATMOG

(Concepts and Techniques in Modern Geography)

CATMOG has been created to fill a teaching need in the field of quantitative methods in undergraduate geography courses. These texts are admirable guides for the teachers, yet cheap enough for student purchase as the basis of classwork. Each book is written by an author currently working with the technique or concept he describes.

1. An introduction to Markov chain analysis - L. Collins
2. Distance decay in spatial interactions - P.J. Taylor
3. Understanding canonical correlation analysis - D. Clark
4. Some theoretical and applied aspects of spatial interaction shopping models - S. Openshaw
5. An introduction to trend surface analysis - D. Unwin
6. Classification in geography - R.J. Johnston
7. An introduction to factor analytical techniques - J.B. Goddard & A.Kirby
8. Principal components analysis - S. Daultrey
9. Causal inferences from dichotomous variables - N. Davidson
10. Introduction to the use of logit models in geography - N. Wrigley
11. Linear programming: elementary geographical applications of the transportation problem - A. Hay
12. An introduction to quadrat analysis - R.W. Thomas
13. An introduction to time-geography - N.J. Thrift
14. An introduction to graph theoretical methods in geography - K.J. Tinkler
15. Linear regression in geography - R. Ferguson
16. Probability surface mapping. An introduction with examples and Fortran programs - N. Wrigley
17. Sampling methods for geographical research - C. Dixon & B. Leach
18. Questionnaires and interviews in geographical research - C. Dixon & B. Leach
19. Analysis of frequency distributions - V. Gardiner & G. Gardiner
20. Analysis of covariance and comparison of regression lines - J. Silk
21. An introduction to the use of simultaneous-equation regression analysis in geography - D. Todd
22. Transfer function modelling: relationship between time series variables - Pong-wai Lai
23. Stochastic processes in one-dimensional series: an introduction - K. S. Richards
24. Linear programming: the Simplex method with geographical applications - J.E. Killen
25. Directional statistics - G.L. Gaile & J.E. Burt (in preparation)

CONCEPTS AND TECHNIQUES IN MODERN GEOGRAPHY NO. 23

STOCHASTIC PROCESSES IN ONE-DIMENSIONAL SERIES; AN INTRODUCTION

by

K. S. Richards

Department of Geography, University of Hull

CONTENTS

ACKNOWLEDGEMENTS

To Jim Thompson, Rob Ferguson and John Thornes for helpful criticisms of the first draft, Wendy Munday for drawing the diagrams, and Heather, Janet and Jill for coping with the manuscript.

I INTRODUCTION

One-dimensional time or distance series with equally spaced observations are common in geography. Climatic, hydrologic, economic and demographic data are all ordered with respect to time, and transect measurements along a line characterise spatial variation in one dimension. Any of these series may be regarded as having been generated by a *stochastic process,* which is a probabilistic model describing the sequential relationship between terms in a series. The distinction between purely probabilistic models, which merely describe the probability of occurrence of a given event, and stochastic models, which deal with systems developing in time or space according to a probabilistic law and hence involve consideration of the sequence of events, is clearly explained by Chow and Kareliotis (1971). A stochastic model describes the manner in which a series fluctuates over time or distance, the oscillations being a reflection of a certain pattern of *serial correlation* between values in the series. The theory of stochastic processes, and the methods of identification of the model underlying or generating a particular series, are becoming increasingly important in geographical data analysis, especially as it is recognised that the many data series exploited by geographers are not uncorrelated, as is assumed by the conventional statistical techniques hitherto applied (Unwin, 1977).

(i) Types of data and forms of analysis

The theory of stochastic processes becomes more complex if data are sampled at irregular intervals; thus *discrete* equally spaced observations are assumed in this introductory monograph. Many geographical series are in fact *continuous,* and discrete series must be obtained from them. This is achieved in one of two ways. The continuous series may be *sampled,* or *aggregated.* A sampled series consists of values read from a continuous series at equally spaced points, whereas an aggregated series involves accumulation and averaging within equal intervals of time. The example of the river bed-profile, used below to illustrate the techniques of stochastic modelling, is a sampled series, and the streamflow series is aggregated. Clearly no amount of analysis can provide information about fluctuations in the continuous series over intervals less than the length of the sampling interval; thus, no hydrological processes operating over time scales less than five days can be inferred from the analysis of the streamflow series. Some series are intrinsically discrete, and are associated with a set of discrete states. For example, in successive time periods a family may be defined as living in the suburbs, the countryside, or a town. The probability of moving to a given state from an initial state may be defined, and the set of probabilities compiled as a transition probability matrix. The appropriate means of analysis of such intrinsically discrete data is by Markov chains (Collins, 1975).

Two modes of analysis of the discrete approximations to continuous series are available. Analysis in the *frequency domain* involves decomposing a series into sine and cosine waves of different frequency and amplitude using harmonic or spectral methods (Rayner, 1971). Six years of monthly temperature data at a mid-latitude meteorological station will be dominated by a waveform of frequency $K = 6$, and the wavelength of this cycle is given by the series length ($6 \times 12 = 72$ months) divided by the frequency. In less clear-cut examples, the object of frequency analysis is to identify those frequencies (wavelengths) which contribute most to the variance of the series. The frequency concept may be meaningless in spatial data (Granger, 1969), although Kendall (1973, p.5) accepts the validity of frequency analysis of one-dimensional distance series. However, spectral methods are less useful for modelling series, for example for predictive purposes, where it is preferable to define the stochastic model in terms of a *time domain* analysis. Such an analysis investigates the relationship of values from one term to the next as exemplified by *serial correlations*. This monograph is therefore restricted to the analysis of one-dimensional series, either time series or distance series measured along a line, in which discrete measurements of a continuous variable are available, and for which analysis in the time (space) domain is appropriate. Some general comments are made in Chapter VIII about frequency analysis and the relation between continuous series and their discrete approximations.

(ii) Fundamental concepts of time domain series analysis

Three basic concepts underlie this form of analysis. First, the technique of serial correlation, or autocorrelation, may be used to demonstrate dependency between values in a series at different lags. Second, any observed series is a *realisation* of a particular stochastic model. Thirdly, different stochastic models have characteristic autocorrelation patterns, so that by calculating autocorrelation coefficients for an observed series, it is possible to identify the type of generating stochastic process.

(a) Autocorrelation

Autocorrelation is a simple technique which is an extension of conventional correlation analysis. The only important complication is that autocorrelation coefficients can only be interpreted successfully if the series from which they are derived is stationary. This means essentially that the series must be free of trend in mean and variance; the stationarity assumption is discussed more fully in Chapter II. A conventional correlation coefficient is given by

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \ \Sigma y^2}} \qquad (1.1)$$

where x, y are deviations from the respective sample means $\bar{X}$, $\bar{Y}$. Autocorrelations are calculated for a single variable in a one-dimensional series. Given $x_t = x_1, x_2, x_3, \ldots, x_N$, again as deviations from the series mean, the lag one autocorrelation is calculated by pairing x_t with x_{t+1}, thus:

$$(x_1, x_2)\ (x_2, x_3)\ (x_3, x_4)\ \ldots\ (x_{N-1}, x_N).$$

The lag four serial correlation is obtained by pairing x_t with x_{t+4}, thus:

$$(x_1, x_5)\ (x_2, x_6)\ (x_3, x_7)\ \ldots\ (x_{N-4}\ \ x_N).$$

As the lag increases, the number of pairs in the calculation of the autocorrelation coefficient decreases, and conventionally the series length should be at least N = 50, and the maximum lag k should be N/4. To estimate the autocorrelation coefficients for a stationary series, we calculate

$$r_k = \frac{\Sigma x_t x_{t+k}}{\sqrt{\Sigma x_t^2 \ \Sigma x_{t+k}^2}} \qquad \text{for lags } k = 1, 2, 3 \ldots N/4 \qquad (1.2)$$

and a plot of the autocorrelations r_k against lag k is the *autocorrelation function* or *correlogram*. In fact, for N > 50, the two sums of squares in the denominator of (1.2) are virtually equivalent, so the equation becomes

$$r_k = \frac{\Sigma x_t \ x_{t+k}}{\Sigma x_t^2} \qquad \text{for lags } k = 1, 2, 3 \ldots N/4 \qquad (1.3)$$

which indicates that serial correlation is even simpler than conventional correlation.

(b) Series as realisations of stochastic models

There are two main types of stochastic model, autoregressive (AR) and moving average (MA) models; these are considered in detail in Chapter III. The simplest AR model is

$$x_t = \phi_1 \ x_{t-1} + e_t \qquad (1.4)$$

which simply states that the value of the series at a given point in time, t, is a constant (ϕ_1) times the preceding value, plus a random component (e_t). The meaning of autoregression is evident, since this relationship is a lagged regression of the series on itself. This is a *first-order* autoregression; higher order models involve autoregression on successively more remote preceding values. The simplest MA model is

$$x_t = e_t + \theta_1 \ e_{t-1} \qquad (1.5)$$

in which a series is a linear weighted sum of present and past random values.

In both types of model, the random component is important because it enables a specific model to generate an infinite number of *realisations*. Consider the model

$$x_t = -0.70 \ x_{t-1} + e_t \qquad (1.6)$$

The 10 values of e_t (e_1 to e_{10}) in row 'a' of Table 1 cause this model to generate the series x_1 to x_{10} in row 'b' from an initial value of x_0 = 1.0, whereas the alternative sequence of e_t values in row 'c' generate the series x_t in row 'd' from the same initial value. Both x_t series are realisations of the same model, and that in row 'd' most clearly displays the characteristic pattern of reversing sign expected in a series generated by a first-order autoregressive model with negative coefficient, such as that of equation 1.6. That this is not so evident in the x_t series of row 'b' is a reflection of the particular random sequence of e_t values.

Table 1. EXAMPLES OF SERIES GENERATED BY THE STOCHASTIC MODEL OF EQUATION 1.6

	t=	1	2	3	4	5	6	7	8	9	10
a	e_t	+0.3	-0.5	-0.1	+0.2	+0.1	-0.4	+0.1	-0.1	+0.5	+0.2
b	x_t	-0.40	-0.22	-0.05	+0.24	-0.07	-0.35	+0.35	-0.35	+0.75	-0.33
c	e_t	+0.2	+0.5	-0.1	+0.1	-0.4	+0.1	+0.2	-0.1	-0.5	+0.3
d	x_t	-0.50	+0.85	-0.70	+0.59	-0.81	+0.67	-0.27	+0.09	-0.56	+0.69

(c) Identification of the stochastic model generating an observed series

Although a given model may generate an infinite number of series, it is possible to identify the model generating a *particular* series because each type of model generates series with a characteristic pattern of autocorrelation. Thus, given a stationary series of N > 50 observations, equation 1.3 can be used to calculate autocorrelations r_k for lags up to k = N/4, and the plot of rk against k (the *correlogram)* may be examined in order to suggest the nature of the underlying generating stochastic model. Box and Jenkins (1970) emphasise a strategy for stochastic model building which is followed in essence by this monograph, and requires the following sequence:

Observation - in which the discrete series is measured by sampling, or by aggregating.

Description - in which the mean and variance of the series are estimated, and a plot illustrates its basic structure.

Identification - in which the autocorrelation function (and the partial autocorrelation function) are used to identify the underlying generating process. This stage is considered in Chapter IV.

Estimation - in which the parameters of the identified model are estimated from the data. This is dealt with in Chapter V.

Diagnostic checking - in which the fit of the model to the observed data is assessed; any failure to fit satisfactorily requires reassessment of the identification and estimation stage. Checking methods are outlined in Chapter VI.

(iii) Geographical applications of series modelling

These techniques have been applied most frequently to time series (Box and Jenkins, 1970; Chatfield, 1975). Examples of time series analysis in economic geography are common (Bennett, 1974), although some investigations have used cross-correlations between time series in different regions to evaluate leads and lags and introduce a spatial element (Bassett and Tinline, 1970; Cliff et al, 1975). Hydrology is a fertile ground for time series analysis (Matalas, 1963; Carlson et *al,* 1970; Edwards and Thornes, 1973). However, in geomorphology analysis of one-dimensional spatial series is well-established, particularly in studies of river bed profiles and meanders (Melton, 1963; Nordin and Algert, 1966; Richards, 1976a; Ferguson, 1975; Thakur and Scheidegger, 1970).

The objectives of series analysis in classic control engineering texts (Box and Jenkins, 1970) are in forecasting and control; a future value of a series is predicted and feedback or feed forward control schemes are designed to minimise variation of that value from a target value. If series analysis in geography has alternative objectives, this may present the methodological problems which are considered in Chapter VII; in particular, description and explanation of series behaviour may be more important than prediction and control.

a) Description. The stochastic process is a statistical model describing series behaviour in a concise, quantitative manner. It can be used to define objectively parameters which are difficult to measure subjectively; for example, gravel bed streams have a quasi-oscillatory variation of bed elevation involving high points (riffles) and low points (pools), and the second order autoregressive model fitted to sampled elevation series provides a basis for objectively estimating the wavelength of the riffle-pool oscillation (Richards, 1976a).

b) Explanation. The stochastic model in itself provides no explanation of causal mechanisms, but may relate to additional, theoretical information to give a quantitative, mathematical summary of a qualitative explanation of the structure of a series. The relationship between model form and theory is clearer in physical geographical examples where a physical mechanism can often be found (see Chapter VII, section (i) for a brief summary of the mechanisms associated with stochastic models fitted to the two examples of series analysed in the text).

c) Forecasting. Forecasting is a relevant objective in time series analysis; in a distance series observed at a particular time, the (n 1)th value is usually measurable at that time. Examples where forecasting is applicable include studies of climatic series (e.g. the analysis of air pollution data by Bennett et al, 1976), and of economic series (unemployment and population data; Bennett, 1974; 1975). Stochastic models fitted to streamflow series are often used to generate synthetic data series to augment the relative shortage of such data for resource planning purposes (Quimpo, 1968).

d) Control. Although practical application of stochastic modelling in geography is unlikely to involve computer production control systems, control policies may be adopted to prevent undesirable forecast values of unemployment, air pollution, or water quality (Edwards and Thornes, 1973) from arising.

(iv) Prerequisites

The bivariate normal regression model underlies much of the theory of stochastic processes, and as prerequisites for reading this monograph it is suggested that some understanding of this, and of the mechanics of correlation and regression (Ferguson, 1978), is useful. Additionally, proficiency in elementary algebraic manipulation is desirable (matrix algebra, however, is not required).

II DATA PREPARATION

Given a series of data values sampled at discrete, equally spaced points through time or along a line (not necessarily straight), some preliminary operations are necessary before a stochastic process can be identified. The series must be *STATIONARY* before analysis. A stationary series is one in which the population mean (P_t) and variance (at) are constant with time. Furthermore, the covariance between pairs of values X_t and X_{t+k}, separated by a lag of k time units, depends only on the magnitude of the lag k and not on the absolute values of time t. This covariance is termed *autocovariance,Yk.* In qualitative terms, the stationarity assumption implies that the law that generates the data is constant over time or distance (Granger, 1969, p 2).

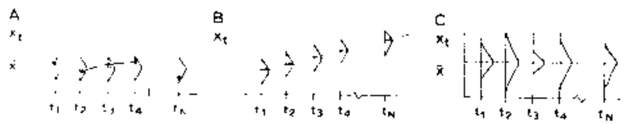

Figure 1. Stationary and non-stationary processes.
(A) A stationary process with constant mean and variance - a realisation of the process is sketched.
(B) A non-stationary process with trend in the mean, but constant variance.
(C) A non-stationary process with constant mean but time-dependent variance.

Figure 1 distinguishes stationary from non-stationary processes involving changes of mean, or 'level', and variance. An individual time series is seen from this diagram to be a set of sampled values from the probability distributions of the phenomenon measured at time t = 1, 2, 3.... N. The mean and variance of the stochastic process underlying, or generating, this individual time series are estimated from the mean X and variance S^2 of the sample time series itself;

$$\bar{X} = \frac{1}{N} \sum_{t=1}^{N} X_t \quad (2.1)$$

$$S^2 = C_o = \frac{1}{N} \sum_{t=1}^{N} (X_t - \bar{X})^2 \quad (2.2)$$

These equations explain the importance of stationarity; the assumption is that N values, one from each of N time periods, are equivalent to a sample of N values from the probability distribution at a single point in time. If mean and variance change through time, this assumption no longer applies. The same is true of estimates of autocovariance at different lags, k up to a maximum lag, L (equal to N/4);

$$C_k = \frac{1}{N} \sum_{t=1}^{N-k} (X_t - \bar{X})(X_{t+k} - \bar{X}), \; k = 1,2...L \quad (2.3)$$

N - k may seem a more logical divisor in equation 2.3, and is used in an alternative estimator of the autocovariance (Chapter VII). Dividing the sum of N - k lagged products by N clearly underestimates autocovariances at large lags, which has a practical advantage of discouraging spurious interpretation. With series of N > 50, and maximum lags L « N, if attention is concentrated on autocovariances and autocorrelations at small lags, the divisor N is satisfactory. Since the autocorrelation at lag k is simply the autocovariance at that lag divided by the variance of the series (equation 1.3), it follows that correlations, as well as covariances, are independent of absolute time.

Before identifying and estimating the stochastic process generating the time (distance) series, the data are manipulated to comply with the stationarity assumption; normalisation and trend removal are particularly important procedures.

(i) Normalisation

Some series will benefit from conventional transformation procedures, to ensure approximation of the distribution of data values to normality. For example, Box and Jenkins (1970, p. 178-197) analysed the series of Wolfer sunspot numbers from 1770 to 1869. The mean of this series is 46.9, and the standard deviation 37.2. The minimum recorded value is zero (1810), and the maximum is 154 (1778). The distribution is therefore non-normal and right-skewed, and an initial transformation is desirable in order to approach normality and stabilisation of variance. One important implication of the logarithmic transformation is that it converts a non-linear, multiplicative relationship between the terms of a series into a linear, additive one (see Chapter VII, section iii).

(ii) Trend Removal

If the series is stationary it is usual to express the data in the form of a series of deviations from the mean, that is

$$x_t = X_t - \bar{X} \qquad (2.4)$$

However, many series are non-stationary in that they exhibit trends in mean and/or variance. Both trends must be removed before the techniques of serial correlation are applied to the residuals from the trend(s). The problem can be illustrated with reference to stream bed elevations, which decline downstream along the profile, while experiencing an irregular oscillation about that trend as a result of the riffle-pool sequence. As the stream gets larger downstream, the amplitude and spacing of that oscillation may increase, so that the mean and variance of the series both trend in the downstream direction. Some of these changes may be abrupt, at major tributary junctions; this type of trend is difficult to remove by conventional methods.

At the trend removal stage in the analysis, deterministic and stochastic components in the data may be considered. True deterministic series are rare, but many include elements of both. Yevjevich (1971), for example, suggests that total stream discharge, Q_T , can be considered as the sum of a deterministic and a stochastic element, with the former being a decreasing proportion of Q_T as forward prediction becomes more remote. In more general terms, a long series of monthly streamflow averages could be written

$$Q_T = Q_R + Q_C + Q_S + Q_E \qquad (2.5)$$

where total discharge is the sum of components due to temporal trend (Q_R , a regression trend), seasonal cyclic variations above and below the trend value (Q_C), a stochastic component described by one of the models outlined below (Q_S) and an unpredictable, residual error component (Q_E) which is a normal, independently distributed random variable with zero mean and constant variance (NID $(0,\sigma_e^2)$). There are two deterministic trend components in this case - the regression and the sinusoidal trends of Q_R and Q_S . The additive nature of the components of equation 2.5 is indicated by Figure 2.

The choice of trend removal technique will be dictated by the nature of the data and the purpose of analysis. Trend may be treated as deterministic or stochastic.

(a) Deterministic trends

In geomorphological distance series, deterministic trends may be identified. Bed-profile elevations decline progressively downstream, and the gradient of this trend may influence stream power and sediment transport processes, so that the long profile trend is deterministic in the sense that it partly controls processes responsible for the oscillation of the riffle-pool sequence. In a short reach of a few hundred metres a linear or low

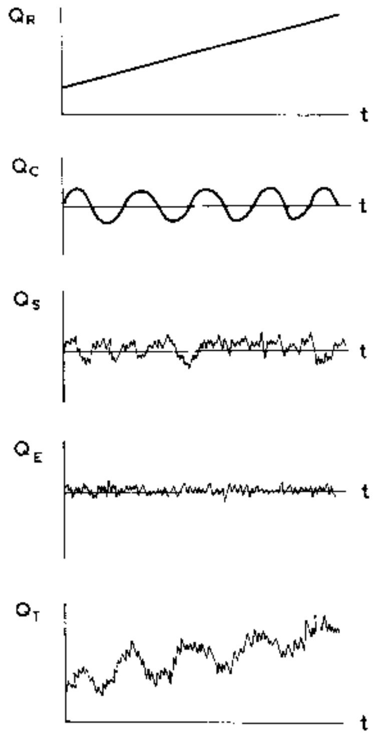

Figure 2. $Q_T = Q_R + Q_C + Q_S + Q_E$ - the composition of total discharge includes a trend (Q_R), an annual cycle (Q_C), a residual stochastic process (Q_S), and a final random element (Q_E).

order polynomial regression trend may be subtracted from the initial bed elevation data, while a long series of several kilometres may demand a logarithmic long profile curve. Usually, such trend removal produces a residual series with minimal low-frequency variance (Granger and Hatanaka, 1964, p.136).

In meteorological and hydrological data sets, a deterministic oscillation may exist which reflects seasonal variation of temperature, rainfall or runoff. Annual temperature cycles, of fairly constant amplitude and wave-length, are controlled by the annual variation in the solar elevation. A deterministic waveform may be subtracted, and a stochastic process identified in the residual series. Monthly discharge data often exhibit trends in mean and variance, with higher and more variable winter discharges. A convenient way of removing the seasonal trends in mean and variance involves the transformation of monthly discharge Q_t to X_t by

$$X_t = \frac{Q_t - M_\tau}{S_\tau} \qquad (2.6)$$

where M_τ and S_τ are the mean and standard deviation of discharges for the month τ. The counter t runs through the whole data series (i.e. from month

1 to 60 for 5 years of data), while T counts through the twelve months of the year then repeats. The twelve expected monthly means M_τ are represented by the sum of harmonics

$$M_\tau = \frac{1}{12} \sum_{\tau = 1}^{12} \bar{Q}_\tau + \sum_{K=1}^{6} A_K \cos \frac{2\pi K\tau}{12} + \sum_{K=1}^{6} B_K \sin \frac{2\pi K\tau}{12} \tag{2.7}$$

where the $\bar{Q}_\tau$ are the twelve monthly means. The monthly standard deviations are generalised by a similar expression (Yevjevich, 1971; Anderson, 1975). Thus under certain circumstances, harmonic analysis may identify a deterministic trend, which is subtracted from the data before serial correlation is applied to the residuals. It may be necessary to remove a regression trend before a (seasonal) sinusoidal trend. The least squares method employed to fit sinusoidal functions is outlined by Rayner (1971, p.12-25).

Polynomial or harmonic regression trends, although considered deterministic, both involve estimation of parameters which are estimates, in the statistical sense, of population values. This presents a methodological difficulty in this aspect of the analysis.

(1) Stochastic 'trends'

A stochastic 'trend' is a type of non-stationary behaviour in which random changes in level (mean) and slope occur in the data series. Box and Jenkins (1970) favour the method of differencing to render such a series stationary and amenable of analysis. This produces a transformation Z_t from the original X_t by

$$Z_t = X_t - X_{t-1} \tag{2.8}$$

Further differencing may be undertaken if Z_t is still non-stationary, i.e.

$$Z'_t = Z_t - Z_{t-1} \tag{2.9}$$

If a stationary stochastic process can be fitted to the series Z_t, the original series X_t can be recreated from

$$X_t = Z_t + X_{t-1} \tag{2.10}$$

given an initial value for X_t. Some problems of conventional linear regression trend removal are illustrated by these differencing and cumulating methods. If a random series (Z_t), which is NID ($0,\sigma^2$), is cumulated by equation 2.10, the result is a stochastic sequence with an apparent trend. Removal of this trend by linear regression, and serial correlation of the residuals, may yield significant serial correlations despite the random nature of the initial data.

This method of achieving stationarity permits identification of pure stochastic autoregressive, moving average or mixed models. If a series requires differencing before the stationarity assumption is met, the model generating the series is referred to as an *integrated* model. A general representation of such a model is provided in the abbreviation ARIMA (p,d,q), where p is the order of the autoregressive component, d is the degree of differencing required to achieve stationarity, and q is the order of the

moving average element. ARIMA (2, 1, 0) indicates a second order integrated autoregressive process in a series requiring one differencing. The only slight drawback in this method is the progressive data loss with each differencing, a problem which may be serious in short data sets. Nevertheless, in practice it is rare for d to exceed 2.

(iii) Examples - the riffle-pool sequence, and annual streamflow variation

The first data series employed in this monograph as an example is a distance series of bed elevations along a bed-profile of a riffle-pool stream. The example is a reach on the River Fowey, Cornwall, illustrated in Figure 3.

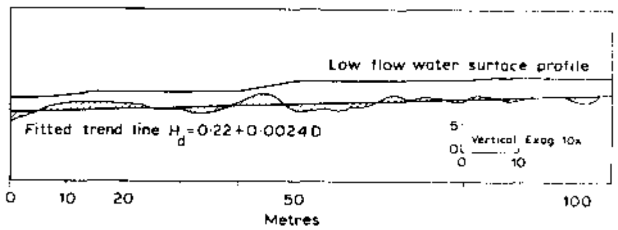

Figure 3. Bed profile series, River Fowey, Cornwall.

A downstream trend was generalised by the linear regression

$$H_d = 0.22 + 0.0024\ D \tag{2.11}$$

fitted to 54 pairs of values of distance (from D = 0 to D = 106 m upstream) and elevation (H_d) above zero datum at the downstream end. Heights were measured, in metres, at equally spaced intervals (2m) giving about 12 data points per oscillation (Nordin and Algert, 1966, p.101).

There is no *a priori* reason to assume a regular wavelength or amplitude for the riffle-pool 'cycle', so removal of a harmonic trend would be unjustifiable. Deterministic waveforms do not adequately describe the irregular bed profile oscillation. Furthermore, removal of a harmonic trend as well as the regression trend demands storage of at least four parameters - intercept and coefficient of the regression, and sine and cosine coefficients. Thus, Table 2a simply lists the regression residuals, from the upstream end moving downstream, since this is the direction in which any process link would operate. This residual series will be investigated further in order to identify the generating stochastic process.

The second series to be analysed consists of a set of 73 pentad streamflow means for the same river at the Trekeivesteps gauging station. Each value represents the mean of 6 years of data for the particular 5-day period involved. Thus 6 years of daily data (a total of 219? values) are reduced to the somewhat more manageable proportions listed in Table 2b. This method has two important effects which must be emphasized. First, it suppresses the evidence for any trend over the 6 year period, and second,

Table 2. DATA TABULATION

A. Bed elevations, River Fowey, Cornwall. Regression residuals (m).

Upstream end

0.013	0.019	-0.107	-0.085	0.047	0.031	0.039	0.020	-0.008	-0.037
0.005	0.070	0.063	-0.022	-0.009	0.024	0.028	0.032	0.010	0.096
0.065	-0.038	-0.051	-0.075	-0.121	-0.114	-0.109	-0.130	0.086	0.004
0.177	0.212	0.131	0.053	-0.032	-0.063	-0.079	-0.106	-0.064	-0.027
-0.031	-0.018	0.060	0.061	0.061	0.090	0.094	0.100	0.113	0.071
0.006	-0.049	-0.116	-0.218		Downstream end				

B. Pentad streamflow averages and residuals from the annual cycle (cumecs)

River Fowey, Trekeivesteps gauge, Oct 1 1958 to Sept 30 1964.

Oct 1-5

2.21	2.24	1.61	1.47	1.49	1.60	2.31	2.36	2.30	2.89
0.76	0.71	-0.00	-0.22	-0.28	-0.25	0.39	0.37	0.24	0.77
2.60	2.83	2.51	2.82	2.53	2.24	2.24	2.31	2.74	2.61
0.42	0.60	0.24	0.51	0.18	-0.13	-0.15	-0.10	0.33	0.20
1.97	1.89	2.42	2.42	2.10	2.31	1.85	1.62	1.19	1.59
-0.43	-0.05	0.50	0.08	-0.20	0.05	-0.36	-0.54	-0.91	-0.45
1.49	1.49	1.71	2.04	1.57	1.56	1.54	1.50	1.47	1.57
-0.48	-0.41	-0.12	0.29	-0.10	-0.03	0.03	0.07	0.13	0.31
1.84	1.60	1.32	1.30	1.11	0.99	0.81	0.79	0.72	0.72
0.66	0.49	0.29	0.34	0.22	0.16	0.04	0.08	0.05	0.10
0.64	0.57	0.52	0.42	0.41	0.52	0.47	0.75	0.54	0.45
0.05	0.01	-0.01	-0.10	-0.10	0.02	-0.04	0.23	0.00	-0.11
0.54	0.46	0.59	0.58	0.71	0.71	0.60	0.67	0.73	0.69
-0.06	-0%17	-0.09	-0.15	-0.07	-0.14	-0.31	-0.31	-0.32	-0.44
0.72	0.62	1.10	(Sept 26-30)						
-0.49	-0.67	-0.27							

it considerably reduces the variance of the series actually being analysed. This latter point is analogous to the problems that arise if a linear regression is fitted to a set of measurements made at fixed values of X; if all the original data are employed, the correlation coefficient will be less than that which is obtained if the analysis is conducted on the set of mean values for each X. Because of this reduction in variance, the annual cycle emerges very strongly in a harmonic analysis of the data in Table 2b, with 80% of the variance being explained by a wave of frequency K = 1. Thus the first stage in analysis of these data is the removal of this deterministic cyclic component and calculation of the residuals, which are also listed in the Table and are subsequently analysed in order to identify the generating stochastic process. Figure 4 shows the original series and the fitted cycle, and the evidence for non-randomness among the residuals is clear.

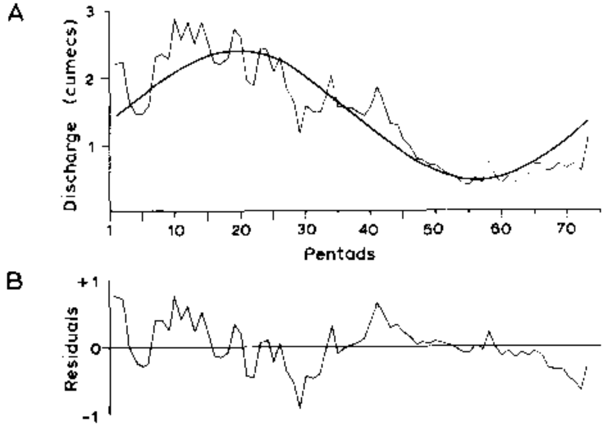

Figure 4. Pentad streamflow series for River Fowey at Trekeivesteps (A) and series of residuals from the annual cycle, (B)

The general procedure in analysis of any series begins by assessing the trend components, perhaps using regression methods as in the bed profile example (Table 2a). Secondly, a cyclic component is identified if appropriate, and is subtracted from the residuals from the trend (cf. Table 2b). Thirdly, the residuals from trend and cycle are analysed by the techniques described in the following chapters, and a stochastic process is identified. The appropriate AR or MA model is then subtracted to obtain a final residual series; if this is a random series, the AR/MA model has been correctly identified, but any remaining sequential correlation reflects a failure to identify the appropriate stochastic process in the third stage of the analysis. This progressive break-down of the series into its components is shown in Figure 2.

III TYPES OF STOCHASTIC PROCESS

The procedures outlined in the previous section indicate some of the techniques available for producing a stationary time or distance series. Either by estimating the mean and expressing the series as deviations from the mean, or by calculating residuals from polynomial or sinusoidal regressions, or by differencing the initial data, we are left with a series whose mean is zero, and whose values represent normally distributed deviations about that mean. It is now possible to consider the problem of identifying within that sequence of observations a stochastic process which expresses the relationships between successive values in the series. However, before examining the practical techniques for achieving this end, it is necessary to outline the main types of stochastic process.

All stationary, normal stochastic processes may be considered to be generated by a series of random 'shocks', however much dependency exists between successive values in the time/distance series which they represent. The observed time series may be viewed as the output from a *linear system* which is itself described by the stochastic process equation, and which transforms the random input shocks into the observed output. This is illustrated in Figure 5.

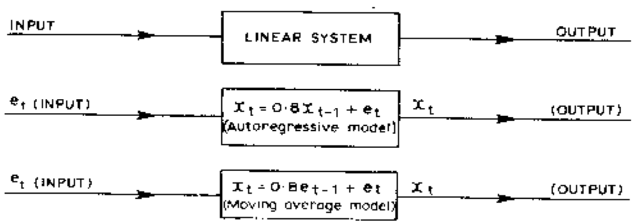

Figure 5. The stochastic process as a linear system transforming an independent, random input shock into an observed output time series.

The simplest example of this idea is the drainage basin hydrological process; the basin provides a complex physical mechanism for transforming an input sequence of precipitation, which may be random in quantity over space and time, into a (relatively) smoothly varying output series of discharge values, which may not be temporally random but may show a sequential correlation from day to day. The stochastic process represents a simple statement of the linear function which forms a black-box model of the transformation process. In Figure 6, this idea is extended further using the two stochastic processes shown in Figure 5. In each case, a normally distributed random variable 'e_t', with zero mean and constant variance σ_e^2, is sampled at time periods $t-1$, t, $t+1$ .... etc. The samples are random and uncorrelated; this emphasises a point already noted, namely that if it were possible to re-run time and draw different values at random from the distributions of the input variable 'e_t' at each time period, a different realisation of the same stochastic process could result. In the Figure, successive sampled values

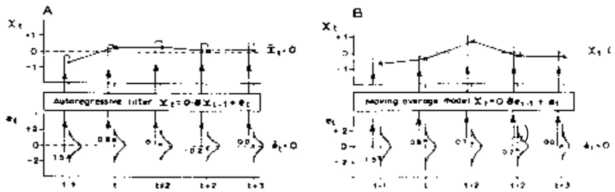

Figure 6. Transformation of input of an independent random normal variate to an observed time series by
(A) an autoregressive linear filter and
(B) a moving average linear filter (or stochastic process).

of the independent, random 'shock' are transformed into the observed output time series x_t by the linear function of the stochastic process, which may be either an AUTOREGRESSIVE or a MOVING AVERAGE model. In the former, the observed value of the time series is dependent on a prior observed value plus the present random shock (or 'error' term), whereas in the latter the observed value of the time series is dependent only on present and past shocks.

To enable comparison between the two diagrams, the input sequences are both drawn from a normal distribution with unit standard deviation, and the values are identical. These random series form the inputs to an auto-regressive process in Figure 6A and a moving average process in Figure 6B. The coefficients of these processes, both of which are first-order (showing dependency only on the previous values of e_t or x_t), are each 0.8 The autoregressive process demands dependency on the previous x_t value, so x_{t-2} is taken to be 1.0, whereas the moving average process implies dependency on the previous e_t value, and in this case (Figure 6B) e_{t-2} is taken to be 1.0. Thus x_{t-1} in Figure 6A is given by

$$x_{t-1} = 0.8x_{t-2} + e_{t-1}$$
$$= (0.8)\ (1.0) + (-1.5)$$
$$= -0.7$$

whereas x_{t+2} in Figure 6B is given by

$$x_{t+2} = 0.8\ e_{t+1} + e_{t+2}$$
$$= (0.8)\ (0.1) + (-0.2)$$
$$= -0.12$$

The most obvious difference between the two output series, x_t is that the autoregressive process produces a smoother output, with a higher degree of sequential correlation between successive values of the x_t series, as would be expected from the structure of the model. The importance of the probabilistic component in a stochastic process is also clear from the diagram. If sequential dependency was merely expressed as

$$x_t = 0.8\, x_{t-1}$$

for an autoregressive process, an x_t series of deviations from a zero mean would itself converge to zero; if x_1 was 1.0, x_2 would be 0.8, x_3, 0.64 and x_{12} would be 0.09. It is only the input of the random component that 'throws' the series from positive to negative values.

The two basic types of stochastic process introduced above - autoregressive (AR) and moving average (MA) - may seem rather different, but mathematically they are closely related. The first order models used in Figure 5 and 6 display dependency on the previous value of the random input series (the MA model) or the series itself (the AR model). In the MA model, the observed time series x_t is linearly dependent on a *finite* sequence of independent shocks, or error terms, and in the case of the first order model this extends only to the preceding shock. The AR model, however, implies an *infinite* linear dependence on previous shocks, but a finite dependence on previous values of the x_t series itself. This can be proved by successive substitution;

$$x_t = \phi_1\, x_{t-1} + e_t$$

defines the first order AR model, from which it follows that

$$x_{t-1} = \phi_1\, x_{t-2} + e_{t-1}$$

If this expression for x_{t-1} is substituted in the first, we get

$$x_t = \phi_1\, (\phi_1 x_{t-2} + e_{t-1}) + e_t$$

This process can be repeated for x_{t-2} and continued until eventually we arrive at

$$x_t = \sum_{j=0}^{\infty} (\phi_1^j\, e_{t-j}),$$

because $\phi_1^m\, x_{t-m}$ converges to zero as $m \to \infty$, since ϕ_1 lies between + 1 and -1. Thus the AR process is expressed as an infinite MA process. The stochastic processes with which we are concerned are therefore members of a single basic family of models, and the dualism between MA and AR processes is worth remembering because it is reflected in the different nature of their correlograms (see Chapter 4). Qualitatively, it effectively means that the AR process, with its dependency of values of x_t on previous values of this output series, has a long-term memory of the input shocks, whereas the MA process only has a short-term memory.

i) Pure random processes

The simplest stochastic process is one with no memory at all, with the observations of x_t (or x_d) obtained at discrete points over time or distance being mutually uncorrelated, independent values. This type of series is

often referred to as 'white noise'. Formally, this is defined as

$$x_t = e_t \tag{3.1}$$

using the approach adopted in Figures 5 and 6; the observed output time series simply being an untransformed representation of the random input series. A stationary, random series of this type will display no correlation between successive values, since the autocovariance, defined by equation 2.3, will be theoretically zero for all lags other than a lag of zero. Where the lag is zero, the series is being paired directly with itself (x_1 with x_1, x_2 with x_2 and so on) so that the covariance expression (equation 2.3) reduces to the familiar variance expression (equation 2.2), since k = 0, and the autocorrelation, which is simply the covariance divided by the variance, is 1. So for a random series, the autocorrelation r_k is 1 for k = 0, and 0 for all other lags. Further discussion of autocorrelation coefficients is deferred to the next section.

Although the random process may seem trivial, it is important in that the objective of series analysis is essentially to identify all those components of the series that are non-random, and successively remove these by subtraction from the original data, until the final residual series can no longer be 'explained' because of its pure random nature. This final series can only be described by its mean (usually zero) and variance, so that it represents the distribution of random input values e_t.

ii) Autoregressive processes

An autogressive process is, as the name implies, a model which is similar to a conventional simple or multiple linear regression except in that the variable x_t is regressed on previous values of itself, hence autoregression. Where the x_t are deviations from a mean, a general representation of an AR process is

$$x_t = \phi_1\, x_{t-1} + \ldots\ldots + \phi_p x_{t-p} + e_t \tag{3.2}$$

so that the value at t is a finite linear aggregate of previous values of the series plus the random input at that time. The extent of this aggregate is defined by the order of the process, p, beyond which lag further coefficients are zero. In fact, the total number of parameters required to characterise the process is p + 2, because the variance of the random error term (σ_e^2) and the mean of the series are required as well as the coefficients. If the x_t series are regression residuals, the regression parameters (trend and cycle components) are additional parameters which replace the mean; and if the x_t are formed by differencing, the degree of differencing is required before the series can be recreated.

The two main AR processes are the first and second order ones. The first order (Markov) process is

$$x_t = \phi_1\, x_{t-1} + e_t \tag{3.3}$$

If ϕ_1 is positive, a realisation of such a process will be relatively smooth, but if it is negative, successive values of the series will tend to be of opposite sign. For the process to be stationary, the value of ϕ_1 must be limited; if ϕ_1 is greater than 1, for example, x_t will tend to depart from zero by more than x_{t-1}, and the series will cumulate away from zero and hence become non-stationary. The random inputs (the e_t) will not be able to prevent this in the long term. Only if

$$-1 < \phi_1 < 1 \tag{3.4}$$

will the process be stationary, because the fractional coefficient will cause convergence towards zero. In fact, the value of ϕ_1 is equal to the lag one serial correlation coefficient (see below). The closer ϕ_1 is to -1 or +1 respectively, the more 'jagged' or 'smooth' the series will be, because the successive dependency on the previous value suppresses the influence of the random component more completely.

The second order (Yule) process is

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + e_t \tag{3.5}$$

Realisations of such a process again depend on the values and signs of the coefficients. For stationarity, limitations must be placed on the combinations of values; for instance, if $\phi_1 = 1.5$ and $\phi_2 = 1.0$, two successive positive values in the series would be cumulated in a weighted sum to give the next value, which would tend to deviate more markedly from zero than either of the initial two values. For the process to be stationary it can be shown that ϕ_1 and ϕ_2 must satisfy

$$\begin{aligned} \phi_2 + \phi_1 &< 1 \\ \phi_2 - \phi_1 &< 1 \\ -1 < \phi_2 &< 1 \end{aligned} \tag{3.6}$$

If both parameters are positive, but satisfy the inequalities of 3.6, the series will be smoother than an AR(1) process because cumulating the effects of two previous positive values of x will demand a larger input random 'shock' to 'throw' the series across to negative values. If ϕ_1 is positive, and ϕ_2 negative, and particularly if

$$\phi_1^2 + 4\phi_2 < 0 \tag{3.7}$$

the realisation of the process will be pseudo-oscillatory (i.e. a randomly varying cycle). This can be illustrated using the process $x_t = 1.0\,x_{t-1} - 0.5\,x_{t-2}$, and initially assuming the error term e_t to be zero. If this series is started with the values $x_1 = x_2 = 1.0$, then successive values are (from x_3 to x_9); 0.50, 0.00, -0.25, -0.25, -0.13, 0.00, 0.06. In the absence of the e_t, the series oscillates about zero and disappears; however, the added random error terms prevent the series from dying away to zero, but do not obscure the 'cyclic' element completely. This behaviour is only apparent if inequality (3.7) is satisfied; in the above example it is, because $\phi_1^2 = 1.0$ and $4\phi_2 = -2.0$.

iii) Moving average processes

Moving averages *sensu stricto* are commonly employed to smooth an irregular series of data; here, the moving average is a *linear smoothing filter*. For example, a moving average is usually symmetrical and centred on the value whose local mean is being estimated, as in $x_t' = (x_{t-2} + x_{t-1} + x_t + x_{t+1} + x_{t+2})/5$.

In this example, each value in the range over which the local mean is estimated is equally weighted (the weight being $\frac{1}{5}$) and the weights sum to 1.

The moving average stochastic process is similarly a transformation of one series into another, but the 'filter' involved is neither symmetrical and centred, nor do the weights sum to 1. A general representation of an MA process of order q is

$$x_t = e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \dots + \theta_q e_{t-q} \tag{3.8}$$

and again there are q + 2 parameters to be estimated from the observed series of data. This is clearly not a moving average in the sense defined above, but the term is now well established in relation to this type of stochastic process. The first and second order models are the most useful, and are respectively

$$x_t = e_t + \theta_1 e_{t-1} \tag{3.9}$$

and

$$x_t = e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} \tag{3.10}$$

The series x_t may be considered as a finite linear aggregate of present and past random shocks. Since there is no carry-over between successive values of the output series, a moving average process is stationary for any values of its parameters. However, limits are placed on the range of values taken by the parameters by the *invertibility condition.*

The invertibility condition is imposed to ensure that, when a model has been identified, its equation can be transposed in order to recover the series of 'shocks' (the e_t series) from the x_t values. Considering an MA (1) process (equation 3.9), we can write the e series in terms of the x series thus

$$e_t = x_t - \theta_1 e_{t-1}, \text{ and } e_{t-1} = x_{t-1} - \theta_1 e_{t-2} \quad \text{(etc.)}.$$

By successive substitution of progressively lagged expressions we obtain

$$e_t = x_t - \theta_1 (x_{t-1} - \theta_1 (x_{t-2} - \theta_1 e_{t-3}))$$

which, when continued to infinity, yields

$$e_t = x_t - \theta_1 x_{t-1} + \theta_1^2 x_{t-2} - \theta_1^3 x_{t-3} \dots \tag{3.11}$$

Only if the parameter θ_1 lies between +1 and -1 does this form a convergent series, permitting estimation of the e_t series by a finite weighted sum of the observed x_t's. It is often necessary to recover the e_t series in this way in order to estimate model parameters by examining the sum of squares of the e_t series for different parameter values (see Chapter V). Equation (3.11) can also be written

$$x_t = \theta_1 x_{t-1} - \theta_1^2 x_{t-2} + \theta_1^3 x_{t-3} - \dots + e_t \tag{3.12}$$

which shows that the *finite* MA process can be transposed into an *infinite* AR model. Again, it is obvious from (3.12) that this cannot be a realistic weighted sum if $\theta_1 > 1$, because the weight increases with the lag. Thus, in (3.11), and in the MA(1) model of (3.9), we require that

$$-1 < \theta_1 < 1.$$

The series in (3.12) is said to be invertible if θ_1 obeys the inequality of (3.13).

For a second order MA model to be invertible, the parameters, must obey

$$\theta_2 + \theta_1 < 1$$

$$\theta_2 - \theta_1 < 1 \qquad (3.14)$$

$$-1 < \theta_2 < 1$$

The dualism between AR and MA models is now clear. A finite AR model has been shown to be equivalent to an infinite MA model, and vice versa, by successive substitution. An AR model is automatically invertible, but limits are placed on the parameter values to ensure stationarity; these limits are the inequalities of (3.4) and (3.6) for first and second order models. An MA model is automatically stationary, but to be invertible requires limits to the parameter values as given by inequalities (3.13) and (3.14) for the first order and second order cases. Comparison of these constraints reveals that the stationarity conditions for an AR model are equivalent to the invertibility conditions for an MA model of the same order.

iv) Mixed autoregressive - moving average models

A fundamental objective in the modelling of series of data by stochastic processes is to identify the process which is *parsimonious* in its representation of the data; that is, which fits the data well with as few parameters as possible. Sometimes it may be found that a mixture of AR and MA components in the model provides a more parsimonious model than, say, an AR model on its own. Since a finite MA model is equivalent to an infinite AR model, for example, a data series which is generated by an MA process could not be parsimoniously represented by an AR model. This may mean that, in practice, we require mixed (ARMA) models such as the model

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \ldots + \phi_p x_{t-p} + e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \ldots + \theta_q e_{t-q} \qquad (3.15)$$

which is an ARMA model of order p,q. The ARMA (1,1) model is most common;

$$x_t = \phi_1 x_{t-1} + e_t + \theta_1 e_{t-1} \qquad (3.16)$$

which is stationary if $-1 < \phi_1 < 1$ and invertible if $-1 < \theta_1 < 1$.

IV MODEL IDENTIFICATION

(i) Serial correlation

When a series of data has been rendered stationary, the first requirement is that the underlying generating stochastic process is identified. Each of the models introduced in the previous chapter has a characteristic *correlogram*, or *autocorrelation function*, so the method of identification

involves estimation of the sample autocorrelation function (acf) from the data series, and its comparison with the theoretical acf's for the various possible models. The correlogram is simply a plot of the serial correlation against the lag.

Serial correlations are calculated from

$$r_k = \frac{\sum x_t x_{t+k}}{\sum x_t^2} = \frac{\text{Autocovariance at lag k}}{\text{Variance}} = \frac{C_k}{C_o} \qquad (4.1)$$

where C_o and C_k are given by equations (2.2) and (2.3). Equation 4.1 yields a sample estimate of the population autocorrelation coefficient, ρ_k; there may be underestimation of autocorrelation at large lags because the autocovariance of equation (2.3) uses the divisor N.

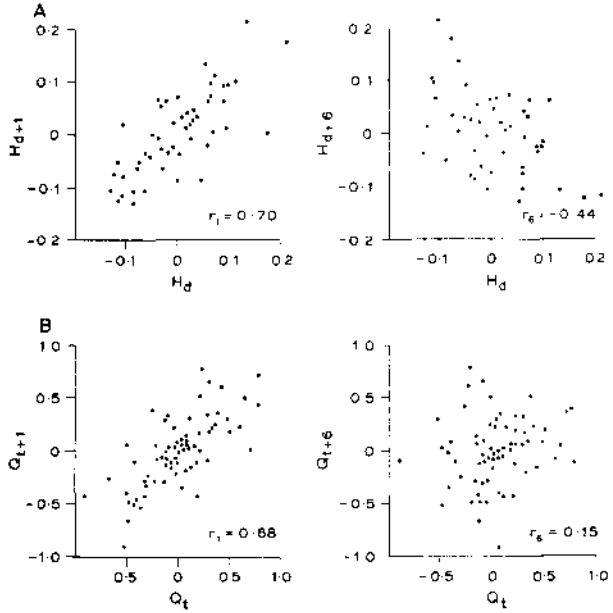

Figure 7. Scatter diagram showing
(A) bed elevation residuals H_{d+1} and H_{d+6} plotted against H_d and
(B) residuals from the streamflow annual cycle, Q_{t+1} and Q_{t+6} against Q_t.

Figure 7 illustrates some examples of scatter diagrams of the data listed in Table 2, with the appropriate autocorrelation coefficients. For both the bed elevation residuals and the streamflow residuals there is clearly dependency between successive values in the series. However, after a lag of

6 in the streamflow series, little correlation remains, whereas the direction of correlation has reversed in the bed elevation series.

(ii) Significance testing

In conventional correlation analysis, the significance of the coefficient is tested before conclusions are drawn about the relationship between two variables. Similarly, it is useful to have a test of significance for the autocorrelation coefficient, because although the general shape of the correlogram is important in the identification procedure, it is often the case that the theoretical correlogram cuts off after a certain lag, so that it is useful to be able to test that correlations beyond that lag are effectively zero. The usual approach is to assume the null hypothesis that the series is a random, uncorrelated series, and calculate the standard error. It can be shown that for a series of N random observations from the same distribution, the expected mean autocorrelation $r_k = -1/N$, and the r_k are normally distributed with an expected variance which is approximately $1/N$. The 95% confidence limits are therefore $-1/N \pm 1.96/\sqrt{N}$. For moderate sample sizes of $N > 50$, the mean r_k is usually taken as zero and the confidence limits taken as $2/\sqrt{N}$. Correlations which lie outside these limits are taken to be significantly different from zero, although it is important to note that if 20 lags are employed in the correlogram (assuming N = 80 at least), we would expect one apparently significant correlation even for a random series. In interpreting the correlogram, therefore, we need to consider the size and lag of a significant coefficient, particularly in relation to possible physical explanations. Testing the autocorrelation coefficients is slightly more complex than the foregoing explanation implies. We begin by assuming that a random process underlies the series, and test the lag one correlation using the standard error as defined above. If r_1 is significant, the series is not a random one, so it would be unrealistic to test r_2 using the same null hypothesis, because in fact r_2 will be unlikely to be zero if r_1 is significant. For example, if the underlying process is AR(1), successive values of the series are correlated; if x_t is correlated to x_{t+1}, and x_{t+1} to x_{t+2}, then some correlation is to be expected between x_t and x_{t+2}. Thus, r_2 must be tested using

$$S.E.\ (r_2) = \frac{1}{\sqrt{N}} \left[1 + 2\,(r_1^{\,2})\right]^{\frac{1}{2}} \qquad (4.2)$$

and, in more general terms, if $r_1, \ldots, r_q$ have been accepted as significant, each successive serial correlation is tested by

$$S.E.\ (r_k) = \frac{1}{\sqrt{N}} \left[1 + 2\,(r_1^{\,2} + r_2^{\,2} + \ldots + r_q^{\,2})\right]^{\frac{1}{2}}, \text{ for } k > q \qquad (4.3)$$

Thus to test the significance of r_6, the standard error incorporates the values of r_k for k = 1 to q, where q, in this case is 5.

(iii) The partial autocorrelation function

A second guide to the nature of the underlying generating process is the partial autocorrelation function (pacf). Partial autocorrelations are complete analogues of partial correlations in conventional multiple regression. If we compare a three-variable multiple regression

$$X_1 = a_{1.23} + b_{12.3} X_2 + b_{13.2} X_3 + e$$

and a second order AR process

$$X_t = \phi_{21} x_{t-1} + \phi_{22} x_{t-2} + e_t,$$

where ϕ_{pj} signifies the jth coefficient in a p-order AR process, the coefficients are of equivalent meaning. For example, $b_{13.2}$ is the partial regression coefficient which measures the relationship between X_1 and X_3 with X_2 controlled. Similarly, ϕ_{22} is the coefficient which measures the relationship of x_t to x_{t-2} after the relationship of x_t to x_{t-1} has been accounted for. In fact, it is a direct measure of the excess correlation between x_t and x_{t-2} not accounted for by ϕ_{11}, the coefficient in the lower order AR(1) process. The pacf is simply a plot of the values of ϕ_{pp} against the order of the process, p, that is, it is the set of last coefficients in AR processes of ascending order. The pacf is an aid to identification because if the generating process is AR(p), the value of ϕ_{kk} for k = p + 1 will be zero, because there is no residual excess correlation to be accounted for by this next higher order process. In order to test the values of ϕ_{kk} for significance, we can use the fact that if we hypothesise that a process is of order p, the partial autocorrelations of order p + 1 and higher are approximately random normally distributed variables with standard error

$$S.E.\ (\phi_{kk}) = \frac{1}{\sqrt{N}} \qquad (4.4)$$

for $k \geqslant p + 1$.

(iv) Theoretical acfs and pacfs for various processes

For most of the processes described in the previous chapter, it is possible to derive the theoretical acf and pacf after some algebraic manipulation. In the following section, some indication of the approach used to produce the theoretical functions is given, but the reader is referred to a text such as Box and Jenkins (1970) or Chatfield (1975) for further elaboration. The main objective is to illustrate the acf and pacf forms required for comparison with observed functions in practical investigations. Table 3 provides a summary of the necessary information.

The pure random process (white noise) has a simple acf. By definition there is no correlation between successive terms in a random series. The covariance, γ_k is zero for all lags k, and the acf is

$$\begin{aligned} \rho_k &= 1 \text{ for } k = 0 \\ \rho_k &= 0 \text{ for } k \neq 0 \end{aligned} \qquad (4.5)$$

where the ρ_k are population autocorrelations. Sample autocorrelations r_k may not be zero, but are expected to lie within the 95% confidence limits, as defined above. Obviously ρ_0 is 1, because even a random series yields a correlation coefficient of unity if it is correlated with itself with zero lag.

TABLE 3 : SUMMARY OF PROPERTIES OF AR, MA, ARMA (1,1) MODELS

Model	Equation	acf	pacf	Stationarity conditions	Invertibility conditions
AR(1)	$x_t = \phi_1 x_{t-1} + e_t$	Decays exponentially as lag increases	Cuts off - only ϕ_{11} non-zero	$-1 < \phi_1 < +1$	Always invertible
AR(2)	$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + e_t$	Decays with increasing lag-mixture of damped sinusoidals/exponentials	Cuts off - only ϕ_{11}, ϕ_{22} non-zero	$\phi_2 + \phi_1 < 1$ $\phi_2 - \phi_1 < -1$ $-1 < \phi_2 < 1$	Always invertible
MA(1)	$x_t = e_t + \theta_1 e_{t-1}$	Finite-cuts off after ρ_1	Infinite-exponential decay	Always stationary	$-1 < \theta_1 < 1$
MA(2)	$x_t = e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2}$	Finite-cuts off after ρ_2	Infinite-mixture of damped sinusoidals/exponentials	Always stationary	$\theta_2 + \theta_1 < 1$ $\theta_2 - \theta_1 < 1$ $-1 < \theta_2 < 1$
ARMA (1,1)	$x_t = \phi_1 x_{t-1} + e_t + \phi_1 e_{t-1}$	Decays with increasing lag-damped sinusoids and exponentials	Decays - mixture of damped sinusoids and exponentials	$-1 < \phi_1 < 1$	$-1 < \theta_1 < 1$

a) Autoregressive process

Derivation of theoretical expressions for the acfs of autoregressive processes is relatively straightforward, and involves generating a set of fundamental equations - the Yule-Walker equations - which are also used to provide estimates of the parameters of the AR model. Consequently, the algebra is outlined briefly here. If we begin by considering an AR process of order p;

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + e_t$$

and we multiply throughout by x_{t-k}, we obtain

$$x_t x_{t-k} = \phi_1 x_{t-1} x_{t-k} + \phi_2 x_{t-2} x_{t-k} + \dots + \phi_p x_{t-p} x_{t-k} + x_{t-k} e_t \quad (4.6)$$

The expected values of the terms in x in (4.6) are autocovariances; that of $x_t x_{t-k}$ is γ_k, the autocovariance of the series x at lag k, and that of $x_{t-2} x_{t-k}$ is γ_{k-2}. The expectation of $x_{t-k} e_t$ is zero, because there is no covariance between the x series at a lag $k > 0$ and the random e series. Thus (4.6) establishes that

$$\gamma_k = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2} + \dots + \phi_p \gamma_{k-p} \quad (4.7)$$

and this may now be divided through by γ_o, the variance of series x_t, to give

$$\rho_k = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2} + \dots + \phi_p \rho_{k-p}, \text{ for } k > 0 \quad (4.8)$$

Thus, we have a general difference equation which forms the basis for the evaluation of theoretical acfs (ρ_k being the terms in the acf for lags $k > 0$), and is the source of the Yule-Walker estimation for values of $\phi_1 \dots \phi_p$ (see chapter V).

First order Autoregressive process

In this process, p = 1, so equation (4.8) reduces to

$$\rho_k = \phi \rho_{k-1} \quad (4.9)$$

The expected lag one autocorrelation coefficient, ρ_1 is therefore

$$\rho_1 = \phi_1$$

since ρ_o is 1. By successive substitution in equation (4.9), the whole acf can be generated. $\rho_o = 1$, $\rho_1 = \phi_1$, therefore $\rho_2 = \rho_1 \rho_1 = \rho_1^2$ (substituting ρ_1 for ϕ in 4.9), and $\rho_3 = \rho_1 \rho_2 = \rho_1 \rho_1 \rho_1 = \rho_1^3 = \phi_1^3$. Thus the acf of an AR (1) process is

$$\rho_k = \phi_1^k \quad (4.10)$$

which implies an exponential decrease of the acf with increasing lag, with ρ_k always positive if ϕ_1 is positive, but switching sign from lag to lag and starting with negative ρ_1 if ϕ_1 is negative. The exponential decay is more rapid if ϕ_1 (and hence, ρ_1) is nearer zero. These acfs are illustrated in Figure 8A. The pacf for a first order AR process has only one non zero term, ϕ_{11}, after which it cuts off to zero.

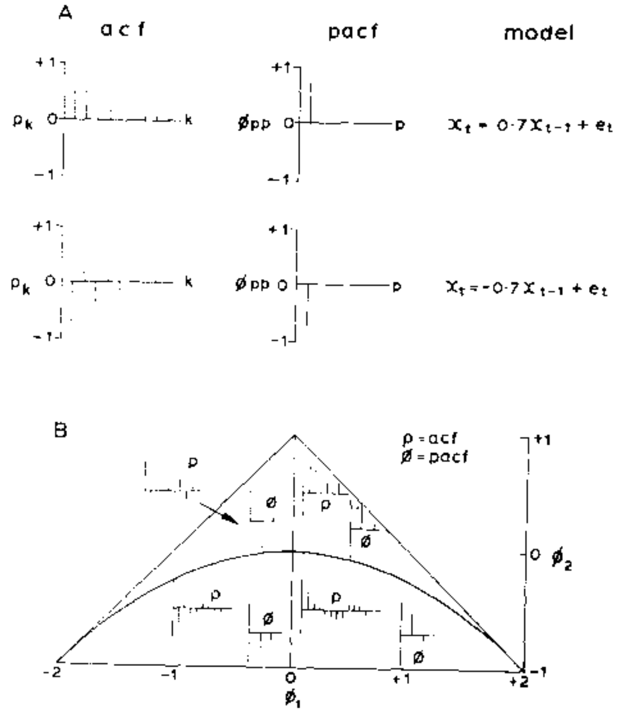

Figure 8 (A) Theoretical acf and pacf for AR(1) processes with $|\phi_1|$ = 0.7 and (B) Permitted values for ϕ_1 and ϕ_2 in AR(2) processes; inset are theoretical acf and pacf for processes with $|\phi_1|$ = 0.7 and $|\phi_2|$ = 0.2 (After Box and Jenkins, 1970, p.59).

Second order Autoregressive process

The order of process is now p = 2, so equation 4.8 becomes

$$\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2} \qquad (4.11)$$

and again ρ_0 = 1. The successive terms $\rho_1, \rho_2 \ldots \rho_k$ are in this case only obtained directly from the Yule-Walker equations, which are produced from 4.8 by substituting k = 1, 2, 3 ... in this equation. The first two Yule-Walker equations for p = 2 are (note that $\rho_1 = -\rho_1$)

$$\rho_1 = \phi_1 + \phi_2\rho_1$$

$$\rho_2 = \phi_1\rho_1 + \phi_2$$

from which

$$\rho_1 = \phi_1 / (1 - \phi_2), \text{ for example.}$$

The precise form of the acf and pacf depends on the signs and values of ϕ_1 and ϕ_2. Figure 8B illustrates the triangular region (the 'parameter space') within which possible combinations of ϕ_1 and ϕ_2 can occur for stationary AR(2) processes. This is divided into four areas by $\phi_1 = 0$ (i.e. negative and positive values of ϕ_1) and $\phi_1^2 + 4\phi_2 = 0$, which separates the region in which the combination of ϕ_1 and ϕ_2 gives the process pseudo-cyclic behaviour. In the two upper regions, exponential decay of the acf dominates, and the pacf is therefore seen to be an important tool in distinguishing AR(1) from AR(2) processes of this type, because in the latter case it cuts off after ϕ_{22}. In the lower two regions, the acf oscillates as a damped sine wave, starting with ρ_1 negative if ϕ_1 is negative.

(b) Moving average processes

All of the acfs for the AR process are seen to die away slowly to zero, being either exponential or damped oscillations. This is to be expected in that the generating process implies dependency on prior values of the observed series x_t. For an MA process, there is no dependency on prior values of x_t, only on a finite sequence of prior random shocks e_t. Intuitively, therefore, we would expect the acf of an MA process to cut off after a lag equal to the order of the process (the extent of backward dependency). In general form, the derivation is more complex, so the algebraic manipulation required is illustrated with reference to the MA(1) process only.

First order Moving Average process

In the MA(1) process, $x_t = e_t + \phi_1 e_{t-1}$. The expected variance of the x_t series is γ_0 (i.e. autocovariance at zero lag), which simply requires the squares of the x_t values, since the process has zero mean. Expected values are denoted by E( ), thus

$$\gamma_0 = E(x_t^2)$$

$$= E\left[(e_t + \theta_1 e_{t-1})^2\right]$$

$$= E\left[e_t^2 + 2\theta_1 e_t e_{t-1} + \theta_1^2 e_{t-1}^2\right]$$

Since e_t and e_{t-1} are uncorrelated, the expectation of their product is zero; furthermore e_t^2 and e_{t-1}^2 are both the expected variance of the zero mean e_t process. Thus

$$\gamma_0 = (1 + \theta_1^2)\sigma_e^2 \qquad (4.12)$$

By similar reasoning, autocovariances at lags k > 0 can be defined. For k = 1

$$\gamma_1 = E(x_t x_{t-1})$$

$$= E\left[(e_t + \theta_1 e_{t-1})(e_{t-1} + \theta_1 e_{t-2})\right]$$

$$= E\left[e_t e_{t-1} + \theta_1 e_{t-1}^2 + \theta_1 e_t e_{t-2} + \theta_1^2 e_{t-1} e_{t-2}\right]$$

$$= \theta_1 \sigma_e^2 \qquad (4.13)$$

The last statement follows because $e_t e_{t-1}$, $e_t e_{t-2}$ and $e_{t-1} e_{t-2}$ are all expected to be zero. Finally, for $k = 2$ and any other integer >1 (i.e. $>p$, the order of the model)

$$\gamma_2 = E(x_t x_{t-2})$$
$$= E\left[(e_t + \theta_1 e_{t-1})(e_{t-2} + \theta_1 e_{t-3})\right]$$
$$= 0$$

Thus the acf for an MA(1) process is γ_k/γ_o which is

$$\rho_k = \begin{cases} \theta_1 / (1 + \theta_1^2), & \text{for } k = 1 \\ 0, & \text{for } k \geq 2 \end{cases} \qquad (4.14)$$

Given the parameter-space for θ_1, solution of the expression for ρ_1 is possible because although a given value of ρ_1 is associated with two alternative values of θ_1 there being two roots to the quadratic equation $\rho_1 = \theta_1/(1 + \theta_1^2)$, only one of these values of θ_1 is acceptable for an invertible process. By substituting values of $\theta_1 = +1$ and -1 in equation (4.14), it will be seen that ρ_1 must lie between $+0.5$ and -0.5. Figure 9A shows acf and pacf for two MA(1) models, with positive and negative coefficients. The acf cuts off after lag one; the pacf decays exponentially. The pacf for an MA process can be calculated by substituting the values of ρ_k, $k = 1, \ldots L$, in the general equation of (4.8) for processes of increasing order p. Note the dualism between Figure 9A, and Figure 8A.

Second Order Moving Average Process

Using the same approach as that used above, it is found that the variance (γ_o) of an MA(2) process is $\gamma_o = (1 + \theta_1^2 + \theta_2^2)\,\sigma_e^2$, and the acf is

$$\rho_1 = \theta_1 (1 + \theta_2) / (1 + \theta_1^2 + \theta_2^2)$$

$$\rho_2 = \theta_2 / (1 + \theta_1^2 + \theta_2^2) \qquad (4.15)$$

$$\rho_k = 0 \text{ for } k \geq 3$$

showing a cut off after lag two. Figure 9B shows the invertibility region for the θ coefficients, and the acf and pacf shapes for the four possible MA(2) processes. Again, where an AR(p) process has an infinite acf and finite pacf, and MA(q) process (q = p) has a finite acf and infinite pacf.

Figure 9. (A) Theoretical acf and pacf for MA(1) processes with $|\phi_1| = 0.7$ and (B) Permitted values for θ_1 and θ_2 in MA(2) processes; inset are theoretical acf and pacf forms. Note the direction of the axes. (After Box and Jenkins, 1970, p.73).

Figure 10. Theoretical acf and pacf forms for mixed ARMA (1,1) models with various parameter combinations (ϕ_1 and θ_1). After Box and Jenkins, (1970, p. 78).

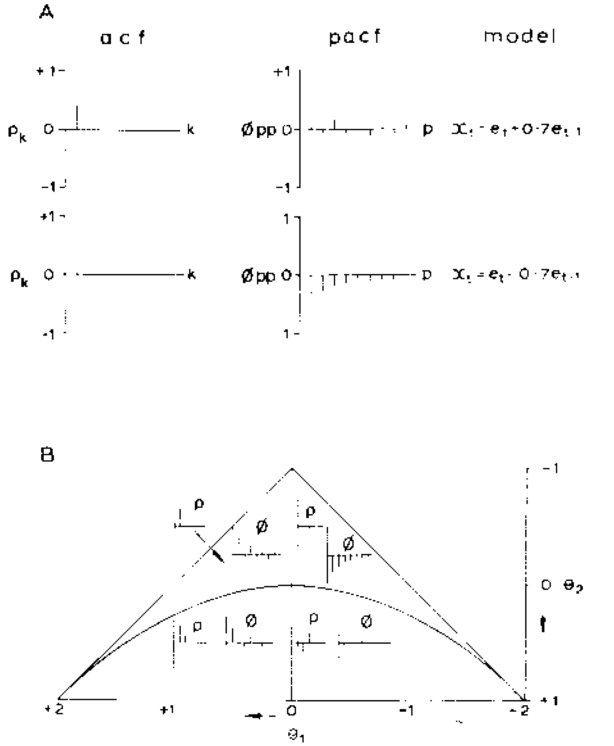

Figure 9

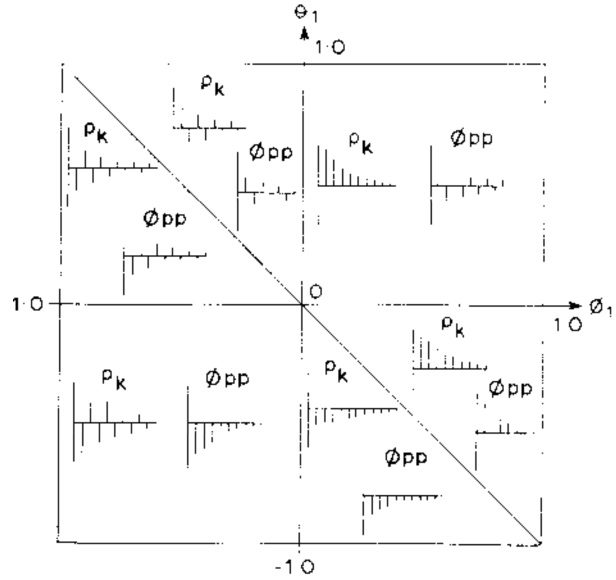

Figure 10

(C) First order mixed model

For the ARMA(1,1) process, the acf can be derived by using the principles outlined above for the MA(1) process. It is found that in this case the value of ρ_1 depends on θ_1 and ϕ_1 in the model (defined in equation 3.16), but that $\rho_2 = \phi_1\rho_1$, and ρ_k for $k \geq 2$ exhibits exponential decay. The acf decays exponentially from the starting value ρ_1, smoothly if ϕ_1 is positive, but with alternating sign if ϕ_1 is negative. Figure 10 illustrates the range of combinations of acf and pacf for ARMA (1,1) processes, depending on the parameter values and signs.

(v) Empirical examples

Figure 11 illustrates the correlograms (acf's) and pacf's for the two data series introduced in Chapter 2. Autocorrelations were calculated using equation (4.1), and the approximate 95% confidence intervals on the acfs were obtained from twice the standard error of equation (4.3). For example, the standard error against which r_3 is tested in the bed profile correlogram is

$$\text{S.E.} = 1/\sqrt{54}\,.\left[1 + 2\,(0.703^2 + 0.323^2)\right]^{\frac{1}{2}}$$
$$= 0.202$$

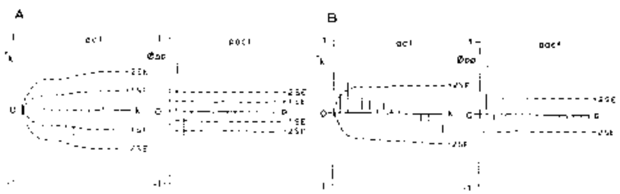

Figure 11. Observed correlograms (acf's) and pacf's for

(A) the bed profile residual series and

(B) the streamflow residual series.

Table 4 lists all terms in the acfs and pacfs, with their standard errors. The terms in the pacf were calculated by substituting the sample autocorrelations r_k in the recursive formulae of equation (5.7), which are explained in the next chapter. Their standard errors were obtained from equation (4.4).

The bed profile data (Table 2a) produce an oscillating acf with a (just) significant value of r_6 (see Figure 7A). The pacf cuts off after two significant terms. The 'cycle' in the data implied by the oscillation of the acf is over approximately 13 lags (26m) which agrees with roughly 4 riffle-pool sequences in the 106m reach (Figure 3). Examination of the summary of acf and pacf forms in Table 3 suggests that this series is generated by an AR(2) process. The streamflow data present difficulties in that the acf shows a broadly exponential decay, but rather than trending to zero, the autocorrelations become increasingly negative (but not

TABLE 4 : SAMPLE ACF AND PACF FOR EMPIRICAL EXAMPLES, WITH STANDARD ERRORS.

Bed-profile Residuals			Streamflow Residuals		
Terms in acf	Standard error	Terms in pacf	Terms in acf	Standard error	Terms in pacf
0.70	0.14	0.70	0.68	0.12	0.68
0.32	0.19	-0.33	0.43	0.16	-0.08
0.03	0.20	-0.04	0.29	0.18	0.04
-0.22	0.20	-0.12	0.22	0.18	0.03
-0.38	0.21	-0.08	0.17	0.19	0.01
-0.43	0.22	-0.06	0.15	0.19	0.04
-0.42	0.23	-0.07	0.15	0.19	0.03
-0.37	0.25	-0.07	0.11	0.19	-0.03
-0.29	0.26	-0.04	0.03	0.19	-0.04
-0.12	0.26	0.01	-0.02	0.19	-0.03
0.08	0.26	0.06	-0.07	0.19	-0.04
0.21	0.26	0.02	-0.13	0.19	-0.06
0.24	0.27		-0.18	0.20	-0.05
			-0.23	0.20	-0.07
			-0.25	0.20	-0.04
		SE = 0.14	-0.20	0.21	0.01
			-0.20	0.21	-0.04
			-0.20	0.21	
					SE = 0.12

significantly so) after a lag of 13. The probable explanation for this behaviour is provided by Figure 4; the residuals from the annual cycle suggest a six month 'cycle', which is often found in addition to the annual cycle in meteorological and hydrological data (Craddock, 1956). This occurs because the annual cycle is not a symmetrical sinusoidal variation, and higher frequency harmonics remain in the series of residuals. A K = 2 (six month wavelength) harmonic was not significant, so was not subtracted; nevertheless, this oscillation remains strong enough to influence the acf of the residual series by creating slight non-stationarity. However, the pacf cuts off after ϕ_1, so a tentative initial identification of an AR(1) model seems appropriate.

V MODEL ESTIMATION

The iterative method of model building (Box and Jenkins, 1970, p. 18-19) demands that the parameters of the identified model are estimated and the model is checked for adequacy of fit to the data before it is used further, for example in forecasting. Initial identification is achieved by comparison of the sample correlogram and pacf with the theoretical ones defined in the previous chapter. These theoretical acf and pacf are derived by assuming the underlying model, and obtaining expressions for the variance and auto-covariance function of series generated by this model; the procedure for AR models is illustrated in equations (4.6) and (4.8), for MA models in (4.12) and (4.13). Conversely, it is possible to express the ϕ parameters of an AR model or the θ parameters of a MA model in terms of the population autocorrelations ρ_k. The simplest example is the AR(1) process, for which it has been shown that $\phi_1 = \rho_1$. By substituting the sample autocorrelation estimates r_k in these expressions, estimates can be obtained for the model parameters; for an AR(1) process, $\phi_1 = r_1$. This simple substitution yields satisfactory parameter estimates for AR processes; however, the estimates obtained in this way for MA processes are statistically inefficient and are usually only employed as initial approximations.

(i) Autoregressive Estimation

Since an autoregressive process is simply a regression of values of a series on previous values, it is not surprising to find that least-squares regression methods provide convenient parameter estimates. The parallelism between conventional regression and autoregression is easily demonstrated. For a first order AR model, the parameter ϕ_{11} is equivalent to the regression coefficient b_{12} in a simple bivariate regression;

Simple regression	Autoregression
$X_1 = a_{1.2} + b_{12}X_2 + e$	$x_t = \phi_{11}x_{t-1} + e_t$

In the simple regression, $b_{12} = r_{12}(S_1/S_2)$, being the product of the correlation coefficient between X_1 and X_2 and the ratio of their standard deviations. For the autoregressive case, given large N, the standard deviations of x_t and x_{t-1} are equivalent, and the correlation is the lag one autocorrelation ρ_1. Thus

$$\phi_{11} = \rho_1 \frac{S_1}{S_2} = \rho_1$$

and the estimate of ϕ_{11} is obtained by using r_1 as the estimate for ρ_1. A second order autoregression is equivalent to a three variable multiple regression of the form $X_1 = a_{1.23} + b_{12.3}X_2 + b_{13.2}X_3 + e$. Here, the partial regression coefficient $b_{12.3}$ may be written in terms of the partial correlation coefficient, thus:

$$b_{12.3} = (r_{12.3})\frac{S_{1.3}}{S_{2.3}}$$

where $r_{12.3}$ is the correlation of X_1 with X_2 with X_3 controlled, and $S_{1.3}$ is a first order standard deviation (that is, the standard error of regression of X_1 on X_3). This can be expressed (Yeomans, 1968, 201-9, 226-9) thus;

$$b_{12.3} = \frac{r_{12} - r_{13}\,r_{23}}{\sqrt{1 - r_{13}^2}\,\sqrt{1 - r_{23}^2}} \cdot \frac{S_{1.3}}{S_{2.3}}$$

$$= \frac{r_{12} - r_{13}\,r_{23}}{\sqrt{1 - r_{13}^2}\,\sqrt{1 - r_{23}^2}} \cdot \frac{S_1\sqrt{1 - r_{13}^2}}{S_2\sqrt{1 - r_{23}^2}}$$

In the AR(2) process $x_t = \phi_{21}x_{t-1} + \phi_{22}x_{t-2} + e_t$, the parameter ϕ_{21} is equivalent to $b_{12.3}$. However correlations such as r_{12} and r_{23} are in this case lag one autocorrelations, and r_{13} is a lag two autocorrelation. The variances of x_t, x_{t-1}, x_{t-2} are equivalent, so $S_1 = S_2 = S_3$. Thus

$$\phi_{21} = \frac{\rho_1 - \rho_2\rho_1}{\sqrt{1 - \rho_2^2}\,\sqrt{1 - \rho_1^2}} \cdot \frac{\sqrt{1 - \rho_2^2}}{\sqrt{1 - \rho_1^2}}$$

$$= \frac{\rho_1\,(1 - \rho_2)}{1 - \rho_1^2} \qquad (5.1)$$

and, using similar reasoning,

$$\phi_{22} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} \qquad (5.2)$$

Estimates of ϕ_{21} and ϕ_{22} are thus obtained by substituting the values of r_1 and r_2 for ρ_1 and ρ_2 in these equations (5.1 and 5.2).

These estimates are in fact equivalent to the Yule-Walker estimates obtained from equation (4.8), and have been introduced here simply to emphasize the close relationship between conventional regression and auto-regression. In practice the Yule-Walker equations may not yield parameter estimates equivalent to least squares estimates because in finite series the assumption that the standard deviations of x_t and x_{t-1} (etc) are equal is not strictly true because of end effects. However, the Yule-Walker equations provide a convenient means of estimating parameters of successively higher order AR models. Equation (4.8) is an expression for the acf of a general AR process of order p;

$$\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2} + \ldots + \phi_p\rho_{k-p} \text{ for } k > 0. \qquad (4.8)$$

If values of k = 1, 2, ..., p are substituted in this equation, the full set of Yule-Walker equations is obtained;

$$\rho_1 = \phi_1 + \phi_2\rho_1 + \dots + \phi_p\rho_{p-1}$$
$$\rho_2 = \phi_1\rho_1 + \phi_2 + \dots + \phi_p\rho_{p-2}$$
$$\vdots \qquad \vdots \qquad \vdots \qquad \vdots$$
$$\rho_p = \phi\rho_{p-1} + \phi_2\rho_{p-2} + \dots + \phi_p \tag{5.3}$$

For an AR(1) process, p = 1 and the set reduces to

$$\rho_1 = \phi_1$$

For an AR(2) process, p = 2 and it reduces to

$$\rho_1 = \phi_1 + \phi_2\rho_1$$
$$\rho_2 = \phi_1\rho_1 + \phi_2 \tag{5.4}$$

which, when solved for ϕ_1 and ϕ_2, give equations (5.1) and (5.2). From the first of the simultaneous equations in (5.4), $\phi_1 = \rho_1 - \phi_2\rho_1$, which when substituted in the second gives

$$\rho_2 = \rho_1(\rho_1 - \phi_2\rho_1) + \phi_2$$
$$= \rho_1^2 - \phi_2\rho_1^2 + \phi_2$$
$$\rho_2 - \rho_1^2 = \phi_2(1 - \rho_1^2)$$
$$\frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} = \phi_2$$

At this point, it would seem appropriate to consider the two empirical examples. The bed profile data have been identified as a series generated by an AR(2) process. The parameters are therefore estimated using equations (5.1) and (5.2), substituting the sample autocorrelations given in Table 4 for the values of ρ_1 ρ_2. Thus

$$\phi_{21} = \frac{0.70\ (1 - 0.32)}{1 - (0.70)^2}$$
$$\simeq 0.933$$

and

$$\phi_{22} = \frac{0.32 - (0.70)^2}{1 - (0.70)^2}$$
$$= -0.333$$

which, allowing for rounding of values in the table, agrees with ϕ_{22} in the listed pacf values. The fitted model, with d subscripts to denote a distance series, is therefore

$$h_d = 0.93\ h_{d-1} - 0.33\ h_{d-2} + e_d \tag{5.5}$$

which is within the region of Figure 9B in which the AR(2) process exhibits pseudo-oscillatory behaviour. The streamflow data were identified as having been generated by an AR(1) process, which therefore takes the form

$$q_t = 0.68\ q_{t-1} + e_t \tag{5.6}$$

For higher order processes, values of $p > 2$ are substituted in the Yule-Walker equations of (5.3) and the resulting set of simultaneous equations is solved. In fact, a pair of recursive formulae (Durbin, 1960) may be used because they are easily adapted to hand and computer calculation. They tend to be susceptible to rounding errors, however. The equations, which are derived from (5.3) are

$$\phi_{p+1,\ j} = \phi_{pj} - \phi_{p+1,\ p+1}\ \phi_{p,\ p-j+1}, \quad j = 1, 2, \dots, p \tag{5.7a}$$

and

$$\phi_{p+1,\ p+1} = \frac{r_{p+1} - \sum_{j=1}^{p} \phi_{pj}\ r_{p+1-j}}{1 - \sum_{j=1}^{p} \phi_{pj}\ r_j} \tag{5.7b}$$

These may seem complex, but their use can be illustrated for the bed-profile data. For an AR(1) process, p = 1, and these equations permit estimation of the parameters of the next higher order model. Equation (5.7b) becomes

$$\phi_{22} = \frac{r_2 - \left[\phi_{11} r_1\right]}{1 - \left[\phi_{11} r_1\right]} = \frac{0.32 - \left[0.70.0.70\right]}{1 - \left[0.70.0.70\right]}$$

which is in fact equivalent to (5.2) because $\phi_{11} = r_1$, and gives $\phi_{22} = -0.333$. Again, for p = 1, Equation (5.7a) becomes

$$\phi_{21} = \phi_{11} - \phi_{22}\phi_{11} = 0.70 - (-0.333)(0.70) = 0.933$$

Now, the recursive method sets p = 2 and calculates parameters of the next higher order AR(3) process as follows. First ϕ_{33} is obtained from (5.7b);

$$\phi_{33} = \frac{r_3 - \left[\phi_{21}r_2 + \phi_{22}r_1\right]}{1 - \left[\phi_{21}r_1 + \phi_{22}r_2\right]} = \frac{0.03 - \left[(0.933 \times 0.32) + (-0.333 \times 0.70)\right]}{1 - \left[(0.933 \times 0.70) + (-0.333 \times 0.32)\right]} = -0.078$$

Rounding errors have produced a considerable discrepancy between this value and that tabulated in Table 4 under the pacf listing. However, since the purpose here is simply to illustrate the method of calculation, this value of ϕ_{33} is used in equation (5.7a), which is for j = 1, $\phi_{31} = \phi_{21} - \phi_{33}\phi_{22}$ = 0.933 - ((-0.078)(-0.333)) = 0.907

for j = 2, $\phi_{32} = \phi_{22} - \phi_{33}\phi_{21}$ = -0.333 - (-0.078 x 0.933) = -0.260

The third order model would therefore be

$$h_d = 0.907\ h_{d-1}\ \ -0.260\ h_{d-2}\ -0.078\ h_{d-3} + e_d$$

although since ϕ_{33} is non-significant in the pacf, this is not a parsimonious model and would not be fitted in preference to the second order model of equation (5.5). The extra explanation of h_d achieved by adding the third term is non-significant, and the relation of h_d to h_{d-3} is not worth incorporating in the equation, just as an independent variable with a non-significant partial correlation is not worth incorporating in a multiple regression.

(ii) Moving Average Estimation

In all problems of parameter estimation in statistics, it is necessary to specify the criterion to be employed. For example, in simple regression of the form $X_1 = a_{1.2} + b_{12}X_2 + e$, the usual criterion is that the values of $a_{1.2}$ and b_{12} are selected in order to minimise the sum of squares of the residuals, 'e'. This is the conventional least squares criterion. Even this can be modified, and $a_{1.2}$ set to an initial value (usually zero) so that, within this constraint, b_{12} alone is chosen to minimise the sum of squares Σe^2 (Brownlee, 1965, p.298). An alternative criterion is used in estimating the parameters of the reduced major axis regression (Till, 1974, p.99-103), which minimises the sum of the areas of right angled triangles formed between the regression line (as hypotenuse) and each point. In some cases maximum likelihood estimates are preferred, but if the error terms (e) are normally distributed these are equivalent to least squares estimates.

Estimation of autoregressive parameters is relatively straightforward for series of N> 50 observations, because the discrepancy between Yule-Walker estimates and least squares estimates which results from end effects is usually small. The *theoretical* equivalence between Yule-Walker and least squares estimators has been shown above. However, the equivalent equations for MA processes, which define the parameters in terms of autocorrelations, provide very inefficient estimates (with a high variance). They can only be used, therefore, to obtain an initial approximation. For example, equation (4.14) shows that, for an MA(1) process,

$$\rho_1 = \theta_1/(1 + \theta_1^2).$$

This can be rearranged to give a quadratic in the unknown, θ_1, if r_1 is substituted for ρ_1

$$\theta_1^2 - \frac{1}{r_1}\theta_1 + 1 = 0.$$

Using the conventional formula for the solution of quadratic equations, the two roots of this equation are

$$\frac{\frac{1}{r_1} \pm \sqrt{\frac{1}{r_1^2} - 4}}{2}$$

If, for example, r1 is -0.3, the two roots are -0.33 and -3.0; one is the reciprocal of the other, and the second, larger root is impossible for an invertible MA(1) process. Thus a first approximation to the value of 01 is -0.33. For MA(2) processes the appropriate initial estimating equations are given in (4.15).

Least squares estimates of the MA parameters are obtained by an iterative process in which the above initial estimates may be used as starting values, or initial approximations. If an MA(1) process has been identified, its parameter e1 is estimated by an iterative calculation of the sum of squares of the e_t series for different values of θ_1; that which gives the minimum sum of squares is then the required parameter value. The MA(1) process is

$$x_t = e_t + \theta_1 e_{t-1}$$

and the e_t values are simply obtained from

$$e_t = x_t - \theta_1 e_{t-1}$$

starting with $e_0 = 0$, since the expected value of e is its mean of zero. This need for an arbitrary initial value emphasises the importance of invertibility; only if θ_1 lies between +1 and -1 does the consequence of selecting an unrealistic starting value diminish with increased distance from the e_0. However, in the present example

$$e_1 = x_1$$

$$e_2 = x_2 - \theta_1 e_1 \text{ etc.}$$

Having identified the values of $e_1, e_2, \ldots, e_n$ for a given value of θ_1 the sum of squares of these residual values is calculated

(i.e. $\sum_{t=1}^{N} e_t^2$).

From a plot of the sum of squares as a function of θ_1, the 'least squares' parameter estimate is identifiable at the minimum of the function. If an MA(2) process is identified, two parameters are to be estimated. Again, values are chosen, the sum of squares of the 'e_t' series calculated, and these sums of squares plotted on a graph of θ_1 against θ_2. The least squares estimates are obtained by contouring this plot and locating the minimum. This method is in fact completely general; it may be used for AR processes as well, and is the method employed in estimation of parameters in mixed ARMA models.

As an example of the approach, the minimum sum of squares is located for varying ϕ_1 parameters in the AR(1) process fitted to the streamflow data of Table 2. The model in this case is $q_t = \phi_1 q_{t-1} + e_t$, so the values of e_t are obtained by $e_t = q_t - \phi_1 q_{t-1}$. Assuming $q_o = 0$ (again the expected value), the values are obtained thus, starting with $\phi_1 = -0.95$:

$$
\begin{aligned}
e_1 &= 0.76 - (-0.95)(0) = 0.76 \\
e_2 &= 0.71 - (-0.95)(0.76) = 1.43 \\
e_3 &= -0.00 - (-0.95((0.71) = 0.67
\end{aligned}
\tag{5.8}
$$

For this set of values of e_t, the sum of squares Σe_t^2 is calculated; this is the *conditional sum of squares*, conditional on the starting value $q_o = 0$. The calculation is repeated for $\phi_1 = -0.90, -0.85, \ldots, +0.90, +0.95$. The sums of squares Σe_t^2 for each ϕ_1 value are plotted in Figure 12, from which the least squares value of ϕ_1 is seen to be about 0.68; this agrees with the value derived earlier from r_1 as, indeed, it should since both methods in this instance evaluate the same least squares estimate.

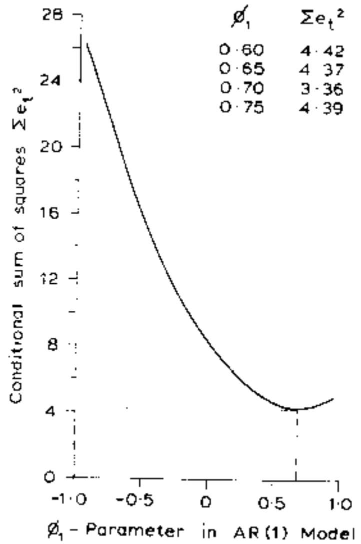

__Figure 12.__ Conditional sum of squares function for the streamflow residual data in Table 2.

VI MODEL CHECKING

Having identified the stochastic process which appears to generate an empirical series, and estimated its parameters, it is necessary to undertake some evaluation of the success of the fit of the model to the data. As in conventional regression analysis, this process largely involves consideration of the nature of the residuals from the model. Ferguson (1978, p.25), for example, shows that, having fitted a linear regression to a bivariate scatter plot, a plot of residuals against the independent variable X will show whether a linear relationship was appropriate by providing evidence of either a pattern or a non-linear relationship between the residuals and X. Some of the important checks have, however, already been discussed. In particular the calculation of standard errors for the terms in the acf and pacf is helpful in preventing *'overfitting'*. An overfitted model is one which is not parsimonious; it has an unnecessarily large number of parameters. Careful consideration of the pacf of an AR model will prevent overfitting because of the sharp cut-off after the 'p'th significant term, and the similar cut-off in the acf for MA processes prevents identification of excessively complex models. Box and Jenkins (1970, p. 249) argue that careful adherence to the identification - estimation - diagnostic checking sequence helps avoid overfitting because, for series of moderate length, it is natural to identify the simplest possible model commensurate with the appearance of the acf and pacf. Only when series of several thousand observations are being analysed are the standard errors likely to become small enough for more complex models to be entertained. Two types of further check on model adequacy are described below; the first is concerned with residual variance, the second with patterns of correlation within the residual series.

Before outlining these it is appropriate to consider the method of calculating residuals. If an AR(1) process is fitted to a data series, the residuals are obtained simply by subtracting the fitted value from the observation; since the model is

$$x_t = \phi_1 x_{t-1} + e_t$$

the residuals are

$$e_t = x_t - \phi_1 x_{t-1}$$

The procedure is therefore exactly the same as that required to calculate the et series in order to estimate model parameters by defining the conditional sum of squares function. If we consider the streamflow data in Table 2, the method of defining the first 3 and the last residuals from an AR(1) process is illustrated by equation 5.8. Estimation of e1 in this case was achieved by assuming q_o to be zero; otherwise, it is necessary to lose values at the beginning of the series, depending on the number of parameters in the fitted model. If the model is, in general terms, ARMA (p,q), then p + q values will be lost. For example, the bed profile model is

$$h_d = 0.93 h_{d-1} - 0.33 h_{d-2} + e_d$$

Unless h_o and h_{-1} are assumed zero, the first residual e we can estimate is e_3, which is $h_3 - 0.93h_2 + 0.33 h_1 = -0.120$ (since from Table 2, these values are h_3 = -0.107, h_2 = +0.019 and h_1 = +0.013). The last residual, e_{54}, is h_{54} - 0.93 h_{53} + 0.33h_{52} = -0.126 (h_{54} = -0.218; h_{53} = 0.116; h_{52} = -0.049). We should note that these are not true errors, but estimated ones, since the parameters in the fitted model are themselves estimates of population values.

(i) Residual variance

Where there is some doubt about the appropriate order of a model, a deliberate overfitting may be attempted in order to assess the progressive reduction of unexplained variance as each new term is added. The significance of the extra explained variance attributable to each new parameter can be considered by the method used by Chayes (1970) to compare trend surface models of successively higher order. This involves an analysis of variance which apportions the original sum of squares of the initial series into the sum of squares explained by each successive term and the remaining residual sum of squares. If the original sum of squares ($\Sigma(X_t - \bar{X})^2$) is S, the explained sum of squares due to the fitting of a first order AR process is R_1^2S, and that due to the second order AR model is R_2^2S: R_1 is a simple correlation, and R_2, R_3... are multiple correlation coefficients. Thus the additional explanation achieved by adding the second order term is ($R_2^2 - R_1^2$) S, and the residual sum of squares at this stage is (1 - R_2^2) S. To test the significance of the second order term, an analysis of variance table is established (Table 5).

TABLE 5 : ANALYSIS OF VARIANCE FOR COMPARISON OF AR MODELS

Source of Variation	Sum of Squares	D. F.	Mean Square	Variance Ratio
Total variation	S	N-1	S/N-1	
First order model	R_1^2S	1	R_1^2 S	
Second order model	R_2^2S	2	R_2^2S/2	
Improvement, AR(2) over AR(1)	(R_2^2-R_1^2)S	1	(R_2^2-R_1^2)S	$\frac{(R_2^2-R_1^2)(N-2-1)}{(1-R_2^2)}$
Residual after fitting AR(2)	(1-R_2^2)S	N-2-1	(1-R_2^2)S/N-2-1	

The variance ratio in the last column of Table 5 is the ratio of the mean squares for the improvement and residual sources of variation; the actual value of the sum of squares, designated S here, cancels and is therefore not required. It is tested using F tables, with 1 and N-p-1 degrees of freedom; if the variance ratio exceeds the tabulated value of F at p = 0.05, the 'improvement' variance estimate is significantly greater than the residual variance, and the AR(2) model can be considered an improvement over the AR(1). This approach is described in relation to conventional regression by Ferguson (1978, p.35).

If AR models are being considered, the last coefficient (ϕ_{pp}) fitted at each successively higher order is the partial correlation of x_t with x_{t-p}, p being the model order at each stage, assuming the x_t values are stationary and have constant variance. Thus ϕ_{pp}^2 is effectively a coefficient of extra determination (Ferguson, 1978, p.227) for the p th order model compared to the (p - 1) th order model. This makes it particularly easy to make comparisons between successively higher order AR models, and the method can be illustrated with reference to the bed-profile data; the pacf provides the necessary information and is listed in Table 4. After fitting a zero order AR model (i.e. no model at all), the explained sum of squares is 'S'. The coefficient of determination (R^2) value for the AR(1) model is ϕ_{11}^2, or r_1^2, which in this case is 0.494; thus the explained sum of squares due to the AR(1) model is (0.494) S, and the residual sum of squares at this stage is (1 - 0.494) S = (0.506) S. Part of this residual variation is explained by fitting an AR(2) process; the proportion is given by ϕ_{22}^2. The explained sum of squares due to the extra AR(2) term is therefore ϕ_{22}^2, multiplied by the residual sum of squares left after fitting the AR(1) model, or (0.114) (0.506) S = (0.058) S. The total R^2 at this stage is therefore (0.494) S + (0.058) S = (0.552) S, and the residual is (1 - 0.552) S = (0.448) S. Thus, the variance ratio required to test the significance of the ϕ_{22} term, and hence the improvement in explanation gained by moving from AR(1) to AR(2), is

$$\frac{\text{Improvement S.S.}}{\text{d.f.}} \div \frac{\text{Remaining residual S.S.}}{\text{d.f.}} = \frac{(0.058)\ S}{1} \div \frac{(0.448)\ S}{N-2-1} =$$

$$\frac{(0.058)(51)}{0.448} = 6.60$$

If this is compared with the critical F - ratio at p = 0.05 with 1,51 degrees of freedom it is found to be significant. Adding the third order term ϕ_{33}, we find an extra explained sum of squares of $(-0.036)^2$ (0.441) S = (0.0006) S. Total third order explained variation is therefore (0.552) S + (0.0006) S = (0.5526) S, and residual variation is (0.4474) S. Thus the variance ratio test comparing explanation by the AR(3) model with that for the AR(2) model is

$$\frac{(0.0006)(50)}{(0.4474)} = 0.07$$

which is clearly not significant at any relevant probability level.

The progressive reduction of unexplained variation which arises when additional terms are incorporated in the model can be calculated in this manner up to a high order, and the trend in the unexplained or residual sum of squares plotted as a function of the number of parameters in the model. In Figure 13 this plot is shown for the successive fitting of higher order AR processes to the bedprofile data; after the second term, there is little further reduction in residual sum of squares. Thus the evidence of the pacf, the F- test described above, and this plot all represent equivalent tests which lead to the conclusion that the AR(2) model is adequate. In practice, it would not be necessary to perform all these tests, since they all measure the same things.

A similar, simple technique for assessing progressive variance reduction does not exist for either MA or mixed ARMA models. However, the estimation of their model parameters demands computation of the residual sum of squares. Thus if successively more complex models are fitted, with more parameters (and fewer degrees of freedom), the necessary information on the progressive reduction

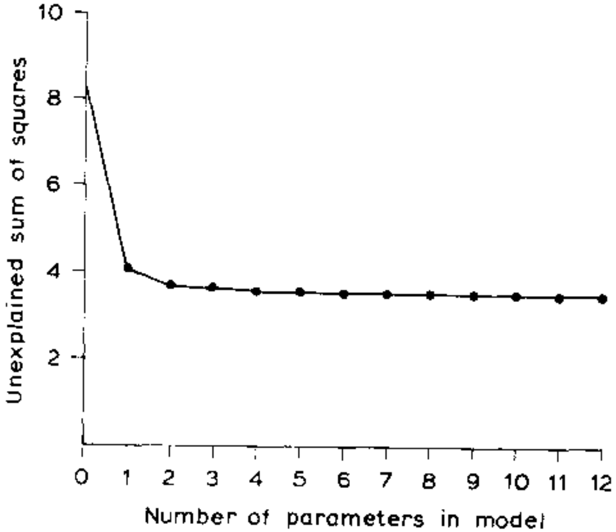

Figure 13. Progressive reduction in residual sum of squares after fitting AR processes from first to twelfth order to the bed profile data.

of unexplained variation arises at each stage. The residual sum of squares for an MA(q) model is the minimum conditional sum of squares which defines the parameter estimate(s).

(ii) Patterns among the residuals

In conventional regression, significance tests based on the F and t distributions demand normal, independent error terms (Poole and O'Farrell, 1971). Normality can be checked by examining the frequency distribution of the errors (Figure 14).

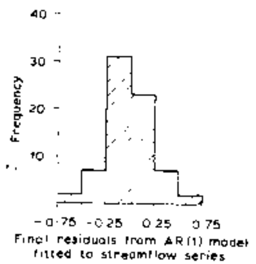

Figure 14. Histogram of residuals from the AR(1) process fitted to the streamflow series. The Kolmogorov-Smirnov test suggests that this distribution does not differ significantly from normality at p = 0.25.

However, care is necessary in applying the above F - test, since the assumption of independence may not apply to the set of residuals from the lower order model if this model is an inadequate fit to the original

series. Thus at each stage in the analysis one of the most important checking routines involves examination of the *series* of residuals from the fitted model for evidence of any remaining serial correlation. Technically, this is quite straightforward; the residuals are calculated as outlined above, and the correlogram (acf) for the resulting series of residual values is estimated. If the fitted model is the true underlying generating process, the residuals should behave as white noise, showing no autocorrelation at any lag (except that 1 in 20 autocorrelations may be expected to appear significant at p = 0.05). In Figure 15A, the acf of the series of residuals from the AR(1) process fitted to the streamflow data is plotted. None of the terms approach significance, the residual series may be regarded as random, and the AR(1) model accepted as a reasonable generating process.

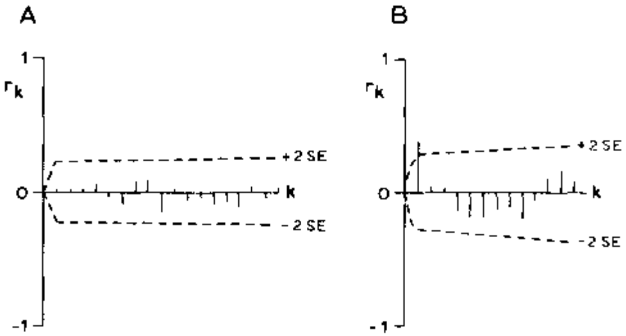

Figure 15. Correlograms of series of residuals from the AR(1) process fitted to

(A) the streamflow data and

(B) the bed profile data whose true generating process is an AR(2) model.

Thus no further terms are required in the model, and all we require is an estimate of the residual variance σ_e^2, which in this case is 0.0523 if the residuals are estimated unconditionally, starting at the second value of the original series, and 0.0597 if 73 residuals are estimated conditionally from an assumed starting value at q_o. With this information it is now possible to generate a streamflow series; normally distributed random deviates with a mean of zero and the estimated residual variance are drawn to represent the e_t values, and after an initial value of q_1 the sequence is created using $q_t = 0.68q_{t-1} + e_t$. In this case these are deviations from an annual cycle, so a series of 73 pentad streamflow values is modelled by adding these q_t values to the harmonic oscillation originally subtracted from the pentad streamflow values of Table 2, i.e.

$$Q_t = 1.458 - 0.009 \cos \frac{2\pi t}{73} + 0.954 \sin \frac{2\pi t}{73}$$

In Figure 15B, the acf of the residuals from an AR(1) model fitted to the bed profile series is plotted. The parameter of the model, $\phi_{11} = r_1 = 0.7$. An AR(2) process is the appropriate generating process in this case, and the acf in the diagram shows that the AR(1) model is inappropriate. It retains an element of cyclic variation, and r_1 is clearly significant. Residuals from an AR(1) process are therefore not random, uncorrelated variates, and the higher order process must be entertained.

The standard errors plotted in Figure 15 are calculated using equation (4.3). However, these estimates of the standard error may be excessively large for the autocorrelations at low lags. If we fit the correct model, with true population parameters, the autocorrelations r_k for the residual series would be uncorrelated, and would be distributed approximately normally about zero with standard error $1/\sqrt{N}$. However, we do not know the true parameter values and so we have to use estimates, which have confidence limits within which we hope the population parameter falls. The residual series we obtain by subtracting the *estimated* model from the observations is an estimated series, and the values of r_k for this series tend to be highly correlated with one another. This means that their theoretical variance is reduced in comparison to that of r_k values for a set of true, random residuals. For example, if a number of series is generated by an AR(1) process with a specific parameter value ϕ and AR(1) models are estimated for each and then subtracted to obtain residual series, the variance of the r_1 autocorrelations estimated for the residual series can be shown to be ϕ^2/N, which is less than $1/N$. This means that the conventional method of calculating standard errors produces an incorrect, excessively large standard error for residual autocorrelations. If a residual autocorrelation fails to appear significant when considered in relation to the confidence limit $\pm\ 2/\sqrt{N}$, this may not mean that it is non-significant. However, in Figure 15B, the value of r_1 for the residual series is clearly significant, because it considerably exceeds even the limit set by $+2/\sqrt{N}$.

Because of the difficulty of testing individual autocorrelations in the residual correlogram, its general pattern is equally useful in diagnostic checking. Furthermore, it is possible to test the whole correlogram by what Box and Jenkins (1970, p. 290) describe as a 'portmanteau lack of fit test'. If we have K autocorrelations for the residual series from an ARMA (p,q) process, then if the fitted model is appropriate,

$$Q = N \sum_{k=1}^{K} r_k^2$$

is distributed as χ^2 with K-p-q degrees of freedom. If the model is inappropriate, the values of Q will be inflated because of the relatively large autocorrelations. A simple test is therefore to calculate Q and compare it with values of χ^2 for the appropriate degrees of freedom; if $Q > \chi^2$ at $p = 0.05$, the fitted model may be considered inappropriate. However, even this test is not conclusive since Chatfield and Prothero (1973) have shown that it has poor power properties; they fitted four different ARMA models to the same data and found that the residual correlograms all gave non-significant values of Q. If this test is applied to the autocorrelations plotted in Figure 15, this problem is not apparent. For the residual acf of the streamflow model, $Q = (72)\ (0.0853) = 6.14$, with 17 degrees of freedom: $N = 72$ because residuals were estimated from e_2, and 18 autocorrelations were estimated after fitting a model of order $p = 1$. Compared with $\chi^2 = 27.6$ at $p = 0.05$ for 17 degrees of freedom, this value of Q is non-significant and implies acceptance of the fitted AR(1) model.

For the acf of residuals from an AR(1) model fitted to the bed-profile data, $Q = (52)(0.4504) = 23.42$, with 11 degrees of freedom. Here $N = 52$ because residual estimation loses 2 values from the beginning of the series, and degrees of freedom are $K - p - q = 13 - 2 - 0 = 11$. χ^2 at $p = 0.05$ with 11 degrees of freedom is 19.7; the Q value therefore suggests model inadequacy, as conclusively as one would expect given results of all the other tests comparing the relative fit of the AR(2) and AR(1) models to this data set.

VII PROBLEMS AND LIMITATIONS

Although conceptually the serial correlation technique which forms the basis of stochastic process identification is straightforward, there are numerous difficulties in the underlying mathematical statistics. Formulae which produce unbiased, efficient estimates of the required quantities are elusive, partly because progressive lagging of the series against itself results in a data loss. Thus there are alternative versions of the simple estimator of the autocovariance C_k given in equation (2.3). This estimating equation uses the divisor N, and results in a biased estimate of C_k and highly correlated successive values of C_k which can make interpretation of the correlogram difficult. One alternative estimator uses the divisor $N - k$, and has less bias; it is less efficient however, because estimates have a greater mean squared error. Additional problems arise in defining confidence limits and significance tests; this is evident in the discussion of standard errors for the autocorrelations, where non-independence of successive correlations necessitates a relatively complex testing procedure. These difficulties are further elaborated in texts such as Chatfield (1975) and Box and Jenkins (1970); the purpose of this chapter is to note some more general problems.

(i) Methodological limitations

The summary of objectives in Chapter 1 indicates the range of possible aims behind attempts to model series by stochastic generating processes. Much of the theory has developed in relation to forecasting and control objectives, but in geographical applications, and especially in studies of distance series, elucidation of the physical operation of the system under study is often of greater importance. Bearing in mind that models are fitted to sample realisations, a small difference between residual variances for two equally parsimonious models may prevent a clear choice being made between them. This is unlikely to be particularly critical for forecasting or control purposes, but would be an embarrassment in terms of physical explanation. Ideally, the fitted process should accord with underlying physical considerations. The storage element within the basin hydrological cycle adequately explains the carry-over effect implicit in the AR(1) model, aided by the phenomena of weather persistence (Grosswetterlagen) and singularities (Craddock, 1956). Yalin (1971) has argued theoretically that velocity pulsations along a straight, uniform channel are described by an AR(2) process, and this may be expected to interact with movable bed sediment in an initially uniform channel to create a bed profile which can also be modelled by an AR(2) process. Thus the models fitted to the two empirical examples in this monograph are physically explicable.

(ii) Non-stationarity

The stationarity requirement presents a practical and methodological problem. Known trends and periodicities have to be removed prior to serial correlation analysis, but this presupposes sufficient knowledge of the physical processes giving rise to these trends. Again, physical geographical series differ from human geographical data in that some underlying physical (and hence, perhaps, deterministic) process may be expected. However, Kisiel (1969, p. 84) notes that even here stationarity is relative; diurnal and annual deterministic variations may occur, but over long time scales it becomes more difficult to prove the existence of deterministic trends and cycles in hydrological or meteorological series. In contrast, the deterministic downward trend in an elevation distance series becomes more evident over greater distances. The answer would seem to be that pragmatism is necessary; the stationarity assumption must be satisfied, and the method employed depends on the type of data and the purpose of analysis. For forecasting, particularly with economic data, differencing may be appropriate, but for physical modelling where it may be desirable to remain in touch with the reality of the original data as much as possible, deterministic trends may be preferable unless it is more meaningful to model rates of change. In some cases the parameter structure of the model may be treated as variable in order to account for non-stationary behaviour of the data; this is suggested by Granger (1975, p. 205) and amplified in chapter VIII, section iv.

(iii) Non-linearity

The models considered in this monograph are all linear, although suitable initial transformation of the data may permit an extension of the methods employed to those relationships capable of linearisation through transformation. Where variance increases with the mean (for example, the problem of higher *and* more variable winter discharges) a logarithmic transformation may be a useful way of achieving stationarity of variance, and this immediately means that an autoregressive process underlying the series is non-linear, and the relation between successive terms in the series is a power function with multiplicative errors; this is because after the transformation we have

$$\log x_t = \phi_1 \log x_{t-1} + \log e_t$$

which in terms of the original variables is

$$x_t = x_{t-1}^{\phi_1} \; e_t$$

Non-linearities can often be treated as a form of non-stationarity, therefore, and avoided by suitable manipulation of the data before analysis.

(iv) Multivariate series analysis

Chow and Kareliotis (1970) bemoan the lack of application of time series analysis to composite hydrologic systems; most studies having been univariate. Multivariate series analysis is a complex problem (see Chapter VIII (ii)), and in order to achieve a model of the basin hydrological cycle they assume that each variable is generated by a stochastic process in which the errors are uncorrelated with any part of the stochastic process of one of the other

variables. They then proceed to analyse each variable separately, bringing them together in the basic hydrological equation

Storage change = Rainfall - Runoff - Evapotranspiration

Unfortunately their basic assumption may be unrealistic, and may hamper parameter estimation in the development of a composite model which is merely an assemblage of univariate models. Another illustration of a multivariate time series model is that of Stewart et al (1976), which predicts ozone levels (O) on day t, from temperature (T), radiation (I) and humidity (H) on that day, and the previous day's ozone level;

$$O_t = 0.68 O_{t-1}^{0.38} \; T_t^{0.37} \; I_t^{0.28} \; H_t^{-0.36}$$

This was fitted by stepwise multiple regression, so that the 2nd to 4th independent variables are effectively explaining the residuals from a log-transformed AR(1) model. However, true multivariate series analysis is more complex than either of these examples; some comments are made in VIII (ii), but a full explanation of transfer function modelling is given by Lai (1979).

(v) Sampling problems

Initially it might appear that model identification is heavily dependent on sampling interval, but this problem is not too severe. For example, incorrect model identification might be anticipated for a bed-profile actually generated by an AR(2) process which is pseudo-cyclic, if a large sample spacing was employed such that on average only two observations were made in each 'cycle'. However we need at least 50 observations, which demands 4-5 riffle-pool cycles at the recommended sampling frequency of 10-12 per 'cycle' (Nordin and Algert, 1966). At two observations per cycle, we need coverage of at least 25 successive riffle-pool sequences, which would mean that we should still identify the correct type of model because alternative observations would not necessarily be in riffle and pool locations, since the 'cycle' wavelength is irregular. Ferguson (1975) reinforces this point by noting that sample spacing has little effect on spectral estimates of meander wavelength. However, the actual coefficients of the model may vary somewhat in relation to sample spacing, which may make interpretation of the parameters difficult. For example, if an attempt is made to investigate the effects of possible controlling variables on the parameters of AR(2) models fitted to stream bed profiles, this will be frustrated if the relationship between sample spacing and mean wavelength is inconsistent. However, it may be difficult to maintain such consistency, which might demand very inconvenient sample intervals, during the process of field survey.

(vi) Re-creation of series

A hydrological objective of series analysis involves generating extra data with the same statistical characteristics as the observed record, in order to obtain longer artificial series to test water resource management proposals. Modelling of trend, cycle, and persistence components enables this to be achieved fairly successfully, but it is important to recognise that the artificially generated series does not recreate all characteristics of the true data. The streamflow data in Table 2 and Figure 4 illustrate this point; the pentad averages over six years suggest, for example, a group of positive residuals from the annual cycle between observations 7 and 15 (Oct 31 to Dec 14). These can only occur in the *averaged* data because of a consistent tendency for abnormally high streamflow to occur at this time. Similarly, a set of averaged pentad temperature values over several years would display positive temperature residuals (anomalies) at some times and negative anomalies at others (Craddock, 1956). The late May - early June positive anomaly and late June negative anomaly are examples. The stochastic process which describes persistence tendencies is unable to guarantee that the residual from the annual cycle would take a certain sign at a particular time. Hence, if the harmonic plus stochastic models were used to generate six years of artificial pentad data, and these were subsequently averaged, the residuals from the cycle would not display persistence because the positive and negative values would not have occurred at consistent times in each year. In fact, these averaged 'artificial' residuals would probably appear random and uncorrelated.

VIII FORECASTING AND OTHER EXTENSIONS AND APPLICATIONS

The classic application of these model building techniques is in forecasting, which involves making estimates of point and interval values at successive lead times/distances, defined as $x_t(j)$ where j is the extent of the lead. Point forecasting is straightforward, and employs the fitted model and four basic rules:

a) use actual x_t values when they are known;

b) use forecasted x_t values when predicting over larger lead times/distances;

c) use residuals as estimates of e_t values when they are known; and

d) use the expectation - zero - of the e_t values when they are unknown.

The effect of rule d is to cause convergence of forecasts from AR models to the observed series mean, an AR(1) forecast series converging exponentially and an AR(2) forecast series including a mixture of exponential decay and sinusoidal elements. For example, the bed profile series was generated by the model.

$$h_d = 0.93\, h_{d-1} - 0.33 h_{d-2} + e_d$$

and the 53rd and 54th observations were -0.116 and -0.218. Table 6, row a, lists the forecasts from $h_d(1)$ to $h_d(10)$, beginning with $h_d(1) = (0.93 \times -0.218) - (0.33 \times -0.116) + 0.0 = -0.164$. The unknown $e_d(1)$ is replaced by its zero expectation. For $h_d(2)$, this forecast value of $h_d(1)$ enters the model as the prior value (h_{d-1}). The forecast series is seen to be a damped sinusoid convergent on the zero mean of the initial series of residuals from the regression trend. Forecasting from an MA model depends on the prior shocks to the linear system; an MA(1) model is

$$x_t = e_t + \theta_1 e_{t-1}$$

TABLE 6. FORECASTS FROM THE AR(2) MODEL FITTED TO THE BED PROFILE SERIES

Lead times, j =	1	2	3	4	5	6	7	8	9	10
a) $x_t(j)$	-0.164	-0.81	-0.021	+0.007	+0.013	+0.010	+0.005	+0.001	-0.001	-0.001
b) 95% Confidence level	±0.110	±0.150	±0.161	±0.162	±0.162	±0.163	±0.163	±0.163	±0.163	±0.163

and this can be used to estimate the series of random shocks, as residuals from the model, by the method outlined in Chapter V. $x_t(1)$ depends on $e_t(1)$ which is unknown and set to zero, and on the last residual from the observed series, but $x_t(2)$ and subsequent forecasts are dependent on unknown e_t values which are set to zero, so these forecasts are also zero.

Interval forecasts, which estimate confidence limits above the forecast point value, are more difficult to make. The interval increases with increasing lead (as seen from the 95% confidence intervals in row b of Table 6), and is larger if the variance of the residuals from the fitted model (a) is large. The 95% interval for $x_t(1)$ is $1.96a_e$, but for forecasts further ahead the interval is defined by a series of weights which are dependent on the parameters of the fitted ARMA model; the definition of these is summarised by Box and Jenkins (1970, p.156).

In practice, forecasting is limited to only a few steps ahead with updating as soon as new observed values (and hence estimated e_t values) become available. Stationary stochastic models yield forecasts which converge exponentially (AR) or abruptly (MA) on the series mean, which is usually zero if the initial series is defined as deviations from a regression or a mean value. Of course, when forecasting any trend and cyclic fluctuation must also be built into the forecast; the point estimates of Table 6 must be added to the predicted bed elevations obtained by extrapolating the regression line of equation (2.11). Similarly, if the data to which a stochastic model was fitted were differenced first the forecasts must be cumulated according to equation (2.10), and this non-stationary stochastic model will exhibit a stochastic trend.

In geographical applications it may often be of more interest to use the model to simulate synthetic series rather than to make point forecasts. In this case, the model is used in conjunction with a randomly generated series of shocks drawn from a normal distribution with mean and variance of the residual series, e_t, as outlined briefly on p.45. Further, more sophisticated extensions and applications of the simple techniques outlined in this monograph are briefly considered below.

(i) Analysis in the frequency domain - spectral methods

The spectral density function is a plot of the contribution made to the total variance of a series by oscillatory components (cosine functions) in different frequency bands over a range of frequencies. It provides the same information as the correlogram, in that each stochastic process has a characteristic theoretical spectral density function, and is therefore simply an alternative visual representation of the data (Quimpo, 1968). The simplest derivation of the spectral density function is by a discrete Fourier transform of the autocovariance function (Chatfield, 1975). Theoretical spectra are derived in much the same way as theoretical correlograms; intuitively, for example, we can see that a random process should have a flat spectrum since no frequency dominates. A first order AR process with a positive ϕ_1 coefficient has an exponentially declining spectrum and if ϕ_1 is negative the spectrum increases exponentially to a maximum value at a high frequency. This is to be expected; the persistence of an AR(1) process with positive ϕ_1 gives dominance to low frequency variance, but the rapid oscillation of the AR(1) process with negative ϕ_1 yields a spectrum dominated by high frequency components. The spectrum of an AR(2) process such as that fitted to the bed-profile data would display a broad peak centred on the frequency $\omega = 2\pi/\lambda$ where λ is the wavelength of the pseudo-cyclic oscillation. Examples of theoretical spectra for different processes are shown by Box and Jenkins (1970).

(ii) Transfer functions - bivariate series

When two series are available that move jointly through time or over distance, the relationship between them may be summarised by the cross-covariance and cross-correlation functions. (Richards, 1976b). These are estimated much as the univariate functions, with

$$r_{xy}(k) = c_{xy}(k)/\sqrt{c_{xx}(0)\, c_{yy}(0)}$$

where the cross-covariance at lag k is divided by the square root of the product of the two variances. The cross-covariance is calculated for positive and negative lags, k, and for positive lags is estimated thus;

$$c_{xy}(k) = \frac{1}{N}\sum_{t=1}^{N-k}(X_t - \bar{X})(Y_{t+k} - \bar{Y}), \text{ for } k = 0, 1 \ldots L.$$

A problem which arises in examining the cross-correlation function (ccf) is that if both series have underlying generating processes the ccf may include spurious correlations, much as spurious correlations often arise when time or distance series are correlated directly without initial removal of trend (see Gregory, 1976, p.397). Thus estimation and interpretation is improved if both series are initially prewhitened; that is, the appropriate stochastic process is subtracted from each series before cross-correlation. Significant terms in the ccf can then be easily tested, the standard error being $1/\sqrt{N}$ since both transformed series are white noise. This enables identification of the *linear system* which is the time series analogue of a regression, and is often referred to as a *transfer function* of the form

$$Y_t = \sum_{k=0}^{\infty} h_k X_{t-k}$$

which is a *distributed lag* model in economics. With the Y_t and X_t series both prewhitened, the regression coefficients h_k, or *impulse response function*, are given by

$$h_k = c_{xy}(k)/S_x^2$$

The significant term(s) in the ccf indicated the lag(s) at which impulse response weights should be estimated. These techniques are covered in greater detail by Lai (1979), where this bivariate approach is extended via cross-spectral methods to multivariate time series analysis.

(iii) Continuous processes

Most series analysis proceeds using discrete data, which represent in most cases an approximation to a continuous process. For transfer functions of the type defined in equation (8.1), a close relation exists between the discrete transfer function and a continuous model. This is explored by Quimpo (1973), who illustrates that the conventional mathematical representation of the unit hydrograph as a linear time-invariant transformation of an input X to an output Y is

$$Y_t = \int_0^{\infty} h(k)\, X(t - k)\, du$$

which is the continuous form of equation (8.1). Ferguson (1976) shows that a linear differential equation of a given order describing a continuous univariate time/distance process may in general be represented in the discrete form by an autoregressive model of the same order, with the irregular but quasiperiodic oscillation of a meander pattern being represented by a second order differential equation and AR(2) model. This means that a truly scientific approach may be adopted in the analysis of processes developing through time or distance, with the theoretical model being developed in the form of a mathematical model (the differential equation) and tested using data which form a discrete approximation and permit the application of the identification and estimation methods described in this monograph.

(iv) Variable parameters

One form of non-stationary behaviour which may be modelled by an extension of the techniques outlined herein occurs when the parameters of a stochastic process change through the series so that, for example, ϕ in an AR(1) process is

$$\phi_t = B_1\phi_{t-1} + a_t$$

Estimation problems are severe in this type of analysis, but Bennett (1976) has succeeded in using this type of model to describe the marked change in parameter structure of AR models fitted to river bed profile data which exhibit a change in frequency of oscillation where pebble and cobble bed material breaks down to sand-sized material, with a parallel shift from riffle-pool bed oscillations to smaller scale sand-dune bedforms.

(v) Space-time modelling

A complete generalisation of series analysis can be achieved for spatio-temporal data compiled in a matrix of N regions by T time periods. If each cell value is expressed as a deviation from the overall spatio-temporal mean of all N x T observations, a spatio-temporal autocorrelation function may be estimated from

$$r_{zz}(k,s) = \sum_{i=1}^{N} \sum_{t=1}^{T} x_{ti} x_{t-k,i-s} \left(\sum_{i=1}^{N} \sum_{t=1}^{T} x_{ti}^2 \right)^{-1}$$

for any lag k in time and s in space (Bennett, 1975). A spatio-temporal autoregressive model (STAR) would be characterised as STAR (a,c) and would be expected to exhibit an acf with exponential decay and pacf with a cut off after the a'th time lag, and the c'th spatial lag (although this depends on the definition of the spatial lag). Bennett (1975) shows that population data for the British Northwest region from 1891-1971 can be modelled by a STAR (3,1) process which implies a considerable temporal carry-over effect of population change, but negligible diffusion of these effects over space. These models are in their infancy, and evidently involve difficulties because the spatial dependency operates in two dimensions and because stationarity may be impossible to define (Granger, 1975). Nevertheless, they represent the most advanced generalisation of the simple techniques described in this monograph, and will form the basis for sophisticated forecasting models.

IX BIBLIOGRAPHY

A. General texts

Box, G.E.P. and Jenkins, G.M., 1970, *Time series analysis - forecasting and control;* San Francisco, Holden-Day, 575p.

Brownlee, K.A., 1965, *Statistical theory in science and engineering;* 2nd Edition; New York, Wiley, 590p.

Chatfield, C., 1975, *The analysis of time series: theory and practice;* London, Chapman and Hall, 263p.

Granger, C.W.J. and Hatanaka, M., 1964, *Spectral analysis of economic time series;* Princeton, New Jersey, Princeton Univ. Press, 299p.

Kendall, M.G., 1973, *Time series;* London, Griffin, 197p.

Rayner, J.N., 1971, *An introduction to spectral analysis;* London, Pion, 174p.

Till, R., 1974, *Statistical methods for the earth scientist;* London, Macmillan, 154p.

Yeomans, K.A., 1968, *Applied statistics: statistics for the social scientist:* Volume 2; London, Penguin, 397p.

B. Technical references

Bassett, K. and Tinline, R.R., 1970, Cross-spectral analysis of time series and geographical research; *Area,* 2, 19-24.

Chatfield, C. and Prothero, D.L., 1973, Box - Jenkins seasonal forecasting: problems in a case study; *Journal Royal Statistical Society Series A,* 136, 295-336.

Chayes, F., 1970, On deciding whether trend surfaces of progressively higher order are meaningful; *Geological Society of America Bulletin,* 81, 1273-1278.

Cliff, A.D., Haggett, P., Ord, J.K., Bassett, K.A. and Davies, R.B., 1975, *Elements of spatial structure: a quantitative approach;* Cambridge, Cambridge Univ. Press, 258p.

Collins, L., 1975, *An introduction to Markov chain analysis;* Concepts and Techniques in Modern Geography no.1; Norwich Geo Abstracts, 36pp.

Durbin, J., 1960, The fitting of time series models; *Rev. Int. Inst.Stat.,* 28, 233-234.

Ferguson, R.I., 1978, *Linear regression in geography;* Concepts and Techniques in Modern Geography no.15; Norwich, Geo Abstracts, 44p.

Granger, C.W.J., 1969, Spatial data and time series analysis; Chapt. 1 in Scott, A.J., ed., *Studies in regional science;* London, Pion, 217p.

Granger, C.W.J., 1975, Aspects of the analysis and interpretation of temporal and spatial data; *The statistician,* 24, 197-210.

Gregory, S., 1976, On statistical myths and geographical fables; *Transactions Institute of British Geographers,* 1, 385-400.

Kisiel, C.C., 1969, Time series analysis of hydrologic data; *Advances in Hydroscience,* 5, 1-119.

Lai, P.W., 1979, *Transfer function modelling: relationship between time series variables;* Concepts and Techniques in Modern Geography no. 22; Norwich, Geo Abstracts, 49pp.

Poole, M.A. and O'Farrell, P.N., 1971, The assumptions of the linear regression model; *Transactions Institute of British Geographers,* 52, 145-158.

Quimpo, R.G., 1973, Stochastic and parametric hydrology; *Journal Hydraulics Division, Proceedings American Society of Civil Engineers,* 99, 461-470.

C. Applications

Anderson, M.G., 1975, Some statistical approaches towards physical hydrology in large catchments; Chapter 4 in Peel, R., Chisholm, M., and Haggett, p. eds., *Processes in Physical and Human Geography,* London, Heinemann, 417p.

Bennett, R.J., 1974, Process identification for time series modelling in urban and regional planning; *Regional Studies,* 8, 157-174.

Bennett, R.J., 1975, The representation and identification of spatio-temporal systems: an example of population diffusion in North West England; *Transactions Institute of British Geographers,* 66, 73-94.

Bennett, R.J., 1976, Adaptive adjustment of channel geometry; *Earth Surface Processes, 1,* 131-150.

Bennett, R.J., Campbell, W.J. and Maugham, R.A., 1976, Changes in atmospheric pollution concentrations; p. 221-235 in Brebbia, C.A. ed., *Mathematical models for environment problems,* London, Pentech Press.

Carlson, R.F., MacCormick, A.J.A. and Watts, D.G., 1970, Application of linear random models to four annual streamflow series; *Water Resources Research,* 6, 1070-1078.

Chow, V.T. and Kareliotis, S.J., 1970, Analysis of stochastic hydrologic systems; *Water Resources Research,* 6, 1569-1582.

Craddock, J.M., 1956, The representation of the annual temperature variatioi over central and northern Europe by a two-term harmonic form; *Quarterly Journal Royal Meteorological .Society,* 82, 275-288.

Edwards, and Thornes, J.B., 1973, Annual cycle in river water quality; a time series approach; *Water Resources Research,* 9, 1286-1295.

Ferguson, R.I., 1975, Meander irregularity and wavelength estimation; *Journal of Hydrology,* 26, 315-333.

Ferguson, R.I., 1976, Disturbed periodic model for river meanders; *Earth Surface Processes,* 1, 337-347.

Matalas, N.C., 1963, Autocorrelation of rainfall and streamflow minimums; *U.S. Geological Survey, Professional Paper* 434-B, 10p.

Melton, M.A., 1962, Methods for measuring the effect of environment factors on channel properties; *Journal of Geophysical Research,* 67, 1485-1490.

Nordin, C.F. and Algert, J.H., 1966, Spectral analysis of sand waves; *Journal of the Hydraulics Division, Proceedings American Society of Civil Engineers,* 92, 95-114.

Quimpo, R.G., 1968, Autocorrelation and spectral analysis in hydrology; *Journal of the Hydraulics Division, Proceedings American Society of Civil Engineers,* 94, 363-373.

Richards, K.S., 1976a, The morphology of riffle-pool sequences; *Earth Surface Processes,* 1, 71-88.

Richards, K.S., 1976b, Channel width and the riffle-pool sequence; *Geological Society of America Bulletin,* 87, 883-890.

Stewart, H.N.M., Sullivan, E.J. and Williams, M.L., 1976, Ozone levels in central London; Nature, 263, 582-584.

Thakur, T.R. and Scheidegger, A.E., 1970, Chain model of river meanders; *Journal of Hydrology,* 12, 25-47.

Unwin, D.J., 1977, Statistical methods in physical geography; *Progress in Physical Geography,* 1, 185-221.

Yalin, M.S., 1971, On the formation of dunes and meanders; *International Association for Hydraulic Research,* 14th Congress, Paris, Proc. Vol. 3, Paper C13, 1-8.

Yevjevich, V., 1971, Properties of river flows of significance to river mechanics; in Shen, H.W., ed., *River Mechanics vol.1;* Colorado State Univ. Press; p. 2.1-2.28.

This series, Concepts and Techniques in Modern Geography is produced by the Study Group in Quantitative Methods, of the Institute of British Geographers.
For details of membership of the Study Group, write to the Institute of British Geographers, I Kensington Gore, London, S.W.7.
The series is published by Geo Abstracts, University of East Anglia, Norwich, NR4 7TJ, to whom all other enquiries should be addressed.