NONPARAMETRIC REPRESENTATION AND SIMULATION OF HYDROLOGIC TIME SERIES

by

C.A. McGilchrist

Department of Statistics
University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series # 945
August 1974

NONPARAMETRIC REPRESENTATION AND SIMULATION OF HYDROLOGIC TIME SERIES

C.A. McGilchrist
University of North Carolina

Summary: A representation of a time series which may be seasonal but has no other nonstationary component is described. Such a representation may be used to simulate long periods of record and is applied specifically to hydrologic time series.

KEY WORDS: nonparametric, time series, hydrology

1. *INTRODUCTION*

Extensive records of rainfall and runoff are available throughout the world. Sometimes these records are amalgamated into yearly readings such as yearly averages, maxima or minima; sometimes into monthly readings but, more rarely, the daily readings are easily available for analysis. Particularly with runoff there is a strong dependence between successive observations so that time series methods are usually considered to be applicable.

The available time series methods are those which have been developed for applications in other fields and two such methods are Box-Jenkins models and spectral analysis. The former approach is given in Box and Jenkins [1] and the latter in Jenkins and Watts [7] while a specific illustration of spectral analysis applied to the Orinoco River is given by Rodriguez-Iturbe [17]. The usual method of application is to take the logarithm of the hydrologic variable and proceed on the assumption that the transformed variable is close enough to Gaussian to make inferences drawn from the procedure fairly robust.

However, within the hydrologic literature there has arisen serious doubt about whether such models are appropriate. Firstly different groups of hydrologists have widely varying opinions about the upper tail of the distribution of the hydrologic variable and secondly there is no agreement about the type of dependence structure which exists between successive values of the hydrologic variable. Each of these difficulties is discussed more fully in subsequent sections.

Because of the difficulty of making robust inferences particularly in the upper tail of the distribution of the hydrologic variable a nonparametric approach is suggested here. It is possible to simulate record from this robust nonparametric representation and to use such a simulation to answer a large proportion of questions asked by hydrologists in design and planning problems.

2. *UPPER DISTRIBUTION TAIL*

A danger in fitting parametric models to data is that these models may not describe the true probability distribution throughout the entire range of the variable. A fitting procedure will usually fit the model which is closest to the observed pattern of variable values where the observed values are dense but the fit will often be less good on the tails. This problem has always been recognized by hydrologists who have used a variety of sensible techniques for avoiding too much bias in their estimates. However, in one particular case great reliance has been placed on a probability model and this occurs in extreme value theory.

Original work on extreme value theory was done by Fisher and Tippett [5] and its application to hydrology is chiefly due to Gumbel [6]. A recent account of extreme value theory is given by David [2]. Broadly this theory separates out distributions into one of three types:- (i) exponential-tailed (ii) Cauchy-tailed (iii) finite upper bound. For each type the theory considers the distribution of the largest observation in a random sample of n observations. As $n \to \infty$, the asymptotic distribution for case (i) is referred to as the Gumbel distribution and for case (ii) the Weibull distribution. Case (iii) is more complex and the reader is referred to David. Thus for any exponential-tailed distribution the asymptotic distribution of the largest value is the Gumbel distribution. It is then argued that this probabilistic result makes a substantial contribution to the estimation of hydrologic parameters.

For example, if rainfall distributions have an exponential tail then yearly maxima should have a Gumbel distribution to a good approximation. Thus 20 observations of yearly maxima could be used to estimate the two parameters of a Gumbel distribution and hence obtain the "once in 100 years rainfall" by estimating the upper 1% point. A blow to this theory was the finding that sample size n must in some cases be very large before convergence to the predicted asymptotic distribution took place. A detailed investigation of this convergence is in Dronkers [3]. However, as pointed out by Huxham and McGilchrist [9] it is necessary only that the extreme value distribution converge to the asymptotic distributional type rather than the exact predicted asymptotic distribution, before successful use of the theory can be made.

A far more serious criticism of the use of Gumbel's distribution has been made by Mandelbrot [11] and Mandelbrot and Wallis [13] who claim that upper tails in both hydrologic and economic time series distributions are often of of the Cauchy type. This is supported in a graphical type analysis given by McGilchrist [14]. Should the hydrologist then be using the Weibull distribution rather than Gumbel's distribution? A further point is that if we are to make any sensible choice between asymptotic distribution forms, it must be possible to test whether a distribution tail is of one type or another. Since data have finite values and over a finite interval it is possible to approximate well any distribution by a member of each of the three types, some statisticians would contend that a test for a distribution belonging to any one of the three types is impossible. This contention strikes at one of the cornerstones of statistics, viz, it should be possible to accept one and reject one hypothesis in any comparison of two conflicting statistical hypotheses given sufficient sample data. If this cannot be done here then there is little to support use of any extreme value theory.

To add to confusion we now mention the other band of people who say that every hydrologic distribution must have a finite upper bound because we live in a finite environment. Nothing further will be said about this point of view, except that they must, of course, be right.

The conclusion of this section is that extreme value theory must be regarded as having a shakier application than was originally supposed.

3. *STRENGTH OF DEPENDENCE*

The autoregressive-moving average models of Box-Jenkins assume a Markovian structure and it is usually thought that most time series can be fitted by an autoregressive-moving average model with only a few parameters. It follows that observations from distant time points will be almost uncorrelated in these models. Such models have been applied in hydrologic time series.

On the other hand, stimulated by Hurst's [8] finding that the range of cumulated water flows did not behave asymptotically as suggested by Markovian models, Mandelbrot [11], Mandelbrot and Van Ness [12] and Mandelbrot and Wallis [13], have constructed and assessed processes called fractional Gaussian processes which have a long term dependence pattern.

Again, there is conflict and this time it concerns the analagous phenomenon of what is happening in the tails of the spectral density function. We suggest that this conflict again points to the use of robust nonparametric methods.

4. *THE PROBABILITY INTEGRAL TRANSFORMATION*

In distribution-free statistics applied to independent, identically distributed random variables, the model taken is that each observation X has a cumulative probability function $F(x)$, where F is not restricted to belong to any class of distribution functions. To avoid discussion of ties in initial theoretical development, it is usually assumed that F is continuous and has density $f(x)$. A basic transformation for much of distribution-free theory is

$$Z = F(X)$$

which results in Z variates being independent rectangular variates on $[0,1]$, denoted by $R[0,1]$.

Here we use identical theory. If X_t are observations in a stationary process with no seasonal component then

$$Z_t = F(X_t)$$

are distributed as R[0,1] but will not usually be independent. The model is completed by describing the dependence structure of the Z_t. It is worth commenting that in forming Z_t we have removed the problem of tail area probabilities, as discussed in section 2, from the variables Z_t and put it where it properly belongs— in the estimation of F.

Given an estimate of F we can transform any generated set of Z_t back to X_t. Prior information can be used in the estimation of F and smoothing techniques such as suggested by Durbin and Knott [4], Rosenblatt [18], Watson and Leadbetter [20] and many other authors can be used. A recent review paper of Wegman [19] describes these procedures. Although the X_t are here dependent the same smoothing can be used but distributional theory is altered.

In hydrologic time series, observations X_t do not usually have a constant distribution function since seasonal effects are often present. With monthly observations we let X_{ni} be the observation in month i of year n. Then

$$Z_{ni} = F_i(X_{ni}).$$

With daily or weekly observations it seems reasonable in most cases to assume F constant over each month. Again F_i can be estimated for each month and smoothing procedures used where appropriate.

5. *TEST FOR SEASONALITY*

The procedure for representation of a hydrologic process begins with a test for seasonality. Observations X_t are arranged in the form,

X_{ni} = observation in month i of year n. The observations within each year are then ranked to produce

Y_{ni} = rank of observation in month i within year n, and, if there are N years of record, then

$A_i = \sum_{n=1}^{N} Y_{ni} / N$ is the average rank of the i^{th} month.

Under the null hypothesis of no seasonality, permutations of ranks within each year should be equally likely and the usual Friedman test statistic becomes

$$T = N \sum_{i=1}^{12} (A_i - 6.5)^2/13.$$

When rankings within successive years are independent this statistic is distributed asymptotically as χ^2 with 11 d.f. Such independence could be achieved by taking every second year and it seems plausible that, since the same asymptotic distribution applies irrespective of the "actual number" of years, then the asymptotic distribution of T is robust to small dependences between successive rankings. We use this conjecture and base T on the total record.

As an example we take the short period of record for the Leigh River at

Mt. Mercer, Victoria, Australia. This river is in a winter rainfall region and the order of months in the hydrologic year is May $\to$ April. The record is given in Table 1 for the nine years May 1957 to April 1966. Table 1 gives monthly runoff in thousands acre feet in the first part of the table. Below are the ranks of monthly runoff within each year obtained by ranking separately observations in each year. Average monthly ranks and standard errors computed from the ranks within each month are given. The (standard error)^2 is simply the sample variance of ranks divided by number of years. For this record $T = 63.3$ and if $T > 19.68$ we decide that the stream is seasonal.

It is worth commenting here that, if we had found no seasonality in the distribution of monthly flows, we should have said the stream was nonseasonal.

This conclusion also implies that there is no seasonality in the dependence structure. Although we go on to test for seasonality in dependence structures, the conclusion of no seasonality from simply a look at marginal monthly distributions seems reasonable.

An interesting way of representing seasonality in precipitation has been given by Markham [16]. He uses mean monthly rainfalls to draw a vector diagram. For January he draws a vector at 000^o, 030^o for February, 060^o for March and so on (his actual angles are different but principle is the same). The first vector is drawn from the origin, the second from the end of the first vector etc., and lengths of vectors are (monthly mean rainfall)/(annual mean rainfall). The length of the resultant vector is an index of seasonality and the direction of the resultant indicates the month of maximal seasonal influence. We suggest here a similar procedure but using average monthly ranks instead of proportional monthly precipitation (or runoff). Figure 1 gives the results for the Leigh River at Mt. Mercer.

6. *SUFFICIENCY*

It is well known that if $X_1, X_2, \ldots, X_n$ is a random sample from a distribution with cumulative probability function $F(x)$ then the sample distribution function,

$$\hat{F}(x) = (\text{number of } X \text{ values} \leq x)/n,$$

is sufficient for the sample. This is because $\hat{F}(x)$ determines the order statistics and the order of occurrence in a random sample is irrelevant information.

However, in a time series the order of occurrence describes the dependence structure and this order of occurrence may be described by a permutation of the first n positive integers (we are consistently ignoring the problem of ties). The combination of the sample distribution function and this permutation is sufficient in that it fully describes the data.

7. *DEPENDENCE MODEL FOR NONSEASONAL PROCESSES*

In this section we consider a dependence model for nonseasonal processes and generalise this to seasonal processes in the next section. One way in which the dependence structure can be expressed is to transform the stationary process X_t to

$$Z_t = F(X_t)$$

and then given conditional distributions of Z_t given $Z_{t-1}, Z_{t-2}, \ldots$. Series approximations for such models are given in McGilchrist and Huxham [15] and Huxham and McGilchrist [10]. However, when it comes to estimating parameters of these conditional models we must form $Z_t^* = F^*(X_t)$, where F^* is either the sample distribution function or some smoothed version of it. The Z_t^* estimate the Z_t but errors in estimation here may affect estimates of parameters in the conditional distributions.

A way to avoid this difficulty is as follows. The dependence structure may be described by the order of occurrence of the ranks of X_t, or equivalently of Z_t, and this order of occurrence does not depend on knowing F. The model for dependence then becomes the selection of a rank ordering from a class of rank permutations with the selection process governed by a probability measure placed on this class of permutations.

The approach may be made more graphically appealing if we define a permutation matrix $\underset{\sim}{H} = [h_{ij}]$, where

h_{ij} = 1, if j^{th} ordered observation immediately follows i^{th} ordered observation,

0, otherwise.

To make this clearer we illustrate with a simple example. Suppose the observations are

2.3, 4.9, 3.8, 9.5, 5.1, 3.9, 7.4

then the permutation of the first seven positive integers giving the order of occurrence is (1427536). The equivalent permutation matrix $\underset{\sim}{H}$ has

$$h_{14} = h_{42} = h_{27} = h_{75} = h_{53} = h_{36} = 1$$

and all other h_{ij} are zero giving

$$\underset{\sim}{H} = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

One row and one column do not have a 1 but all the rest do. If the process is regarded as circular i.e. 7.4 is considered to be followed by 2.3 then we would change h_{61} to 1 and then every row and column would have exactly one 1.

Corresponding to permutations being selected from a class of permutations under some probability measure we have $\underset{\sim}{H}$ considered as being selected from some class $\mathcal{H}$ of permutation matrices and probability measure P gives the probability that each matrix in $\mathcal{H}$ will be selected.

Letting T be the total number of observations, we may plot any permutation matrix $\underset{\sim}{H}$ by plotting row i at position i/T on the horizontal axis and column j at j/T on the vertical axis. It is this plot (which looks like a bivariate scatter diagram) that provides the idea for representing $\mathcal{H}$ and P.

If we let $\mathcal{H}$ be the class of all permutation matrices of order T, then the probability measure P can be described as follows. Take any random process $\{Y_t\}$ and, for each selection of T points, set out the equivalent matrix $\underset{\sim}{H}$ as defined previously. Each such random process gives a particular measure P defined on H and in the continuous random variable case no generality is lost by requiring that each Y_t is marginally R[0,1]. The technique we adopt is to find such a random process $\{Y_t\}$ which generates matrices $\underset{\sim}{H}$ which are usually "like" the observed $\underset{\sim}{H}$.

The representation we adopt for the process $\{Y_t\}$ is that Y_t are marginally R[0,1]. Let $g_1(y_t|y_{t-1})$ be the conditional density of Y_t given y_{t-1}, where g_1 is the density corresponding to distribution function G_1. Let

$$Y_{11} = Y_1, \quad Y_{1t} = G_1(Y_t|y_{t-1}), \; t = 2,3,\ldots, T.$$

The distribution of Y_{1t} given y_{t-1} is R[0,1] and is the same as the marginal distribution of Y_{1t}. Thus Y_{1t} and Y_{t-1} are independent. Since

$$Y_{1,t-1} = G_1(Y_{t-1}|y_{t-2})$$

then $Y_{1,t-1}$ and Y_{1t} are independent given y_{t-2}. Since y_{t-2} can be found from $Y_{11}, Y_{12}, \ldots, Y_{1,t-2}$ then Y_{1t} and $Y_{1,t-1}$ are independent given $Y_{11}, Y_{12}, \ldots, Y_{1,t-2}$.

Now if $g_2(y_{1t}|y_{1,t-2})$ is the conditional density of Y_{1t} given $y_{1,t-2}$ corresponding to distribution function $G_2(y_{1t}|y_{1,t-2})$ we let

$$Y_{21} = Y_{11}, \; Y_{22} = Y_{12}. \; Y_{2t} = G_2(Y_{1t}|y_{1,t-2}), t = 3,4, \ldots, T.$$

It is apparent that given $Y_{11}, Y_{12}, \ldots, Y_{1,t-2}$ then

$$Y_{2,t-1} = G_2(Y_{1,t-1}|y_{1,t-3}), \quad Y_{2t} = G_2(Y_{1t}|y_{1,t-2})$$

are independent. It is also apparent that Y_{2t} and $Y_{1,t-2}$ are independent and hence $Y_{2t}, Y_{2,t-2}$ are independent given $Y_{21}, Y_{22}, \ldots, Y_{2,t-3}$.

The process may be continued by forming $Y_{3t} = G_3(Y_{2t}|y_{2,t-3})$, where G_3 is the conditional distribution function of Y_{2t} given $y_{2,t-3}$. Eventually

$$Y_{dt} = G_d(y_{d-1,t}|y_{d-1,t-d})$$

and if $Y_{d-1,t}$ and $Y_{d-1,t-d}$ are independent then G_d is $R[0,1]$ and the Y_{dt} formed are exactly the same as the $Y_{d-1,t}$. We call d the *order of dependence*.

It should be noted that it is necessary to model only the order of the Y_{it} by the $Y_{i+1,t}$ at each stage in the transformation process. In our estimation process we choose the order of dependence and the functions G_i so that the $\{Y_t\}$ process generates matrices $\underset{\sim}{H}$ usually "similar to" to observed $\underset{\sim}{H}$.

8. *DEPENDENCE MODEL FOR SEASONAL PROCESSES*

For seasonal processes we must consider the possibility that the dependence structure alters with time. This results in twelve separate permutation matrices $\underset{\sim}{H}_i = [h_{i;jk}]$, where

$h_{i;jk} = 1,$ if j^{th} ordered observation of month i is followed by
k^{th} ordered observation of month $i+1$

$= 0,$ otherwise

For $i = 12$, it is understood that the subsequent month is the first month and the following observation is in the next year.

Each separate permutation matrix can be treated as in the previous section, however it does seem worthwhile to test whether there is any significant seasonality of the dependence structure even though there may be seasonality in the marginal monthly distributions. In cases where there is no major seasonality of the dependence structure we amalgamate the twelve separate observed $\underset{\sim}{H}$ matrices.

9. *SEASONALITY OF DEPENDENCE PATTERN*

Although a hydrologic time series X_{ni} may be seasonal in that monthly distributions are not the same, the dependence structure may not necessarily be seasonal. Let $\underset{\sim}{H}_i$, $i = 1,2, \ldots,12$ be the permutation matrices described in section 8.

Graphs of such matrices for the Barwon River at Walgett, New South Wales, Australia using 49 years of record 1902-1950 are given in Figure 2. (To save space only the Jan→Mar graphs are given). The question that arises is whether or not these transition pictures represent only random differences from one another.

The way in which differences are most likely to occur is that some transition patterns are more dispersed than others. It is this aspect of seasonality of dependence that we test. For each pair of consecutive months we record

$$V_{ni} = |Z_{n,i+1} - Z_{ni}|$$

and we then rank these positive differences to obtain

$$S_{ni} = \text{rank of } V_{ni}.$$

Letting $\bar{S}_{.i} = \sum_{n=1}^{N} S_{ni}/N$ and $U = \sum_{i=1}^{12} (\bar{S}_{.i} - 6N-\frac{1}{2})^2 /(12N+1)$ then U is the Kruskal-Wallis test statistic and is compared to the χ^2 distribution with 11 d.f.

Although this test is not sensitive to all aspects of seasonal differences in dependence pattern, it is sensitive to the pattern most likely to apply and should detect any seasonality that remains after the effect of different marginal monthly distributions is removed. Ranks of positive differences, average ranks resulting in $U = 18.5$ are given for Leigh River at Mt. Mercer in Table 2. A significant value of U at 5% level is 19.68. It is noticeable that although the stream is strongly seasonal, the dependence pattern is at most weakly seasonal. In what follows we assume that there is negligible seasonality in the dependence structure, although the methods can be adjusted to allow for such seasonality.

10. *ORDER OF DEPENDENCE*

As described in section 3, the possibility that long term dependences exist has caused some anxiety in the application of Markovian models to hydrology. Consternation increases when we see that the dependence structure $\underset{\sim}{H}$ can always be represented by a first order model in the following way. We choose the describing process $\{Y_t\}$ to be

(i) Y_t is discrete uniform on i/T, $i = 1,2,\ldots, T$,

(ii) $G_1(Y_t|Y_{t-1})$ is the discrete distribution given by the plot of $\underset{\sim}{H}$ with Y_{t-1} as horizontal axis and Y_t as vertical.

Indeed, this process always generates the observed dependence matrix $\underset{\sim}{H}$ (when circularity is taken).

Obviously we do not want a model which generates only the observed dependence structure and to generalise we are involved in smoothing $\underset{\sim}{H}$ to estimate distribution functions G_1, G_2, ... It seems likely that the type of smoothing we do may well affect our ideas about the order of model required. Let (ξ_i,η_i), $i = 1,2,\ldots, T$ represent points in the plot of $\underset{\sim}{H}$ on the unit square. To estimate G_1 we assume that $G_1(Y_t|y_{t-1})$ varies slowly with y_{t-1} and that G_1 looks like a smoothed $\underset{\sim}{H}$. Let us divide up the range of y_{t-1} into τ intervals; then

$$\hat{G}_1[y_t|y_{t-1}\epsilon(\tfrac{j-1}{\tau}, \tfrac{j}{\tau})] = [\text{number of } (\xi_i,\eta_i) \text{ with } \xi_i\le y_t, \eta_i\epsilon(\tfrac{j-1}{R},\tfrac{j}{\tau}]/m_j,$$

where m_j is the number of $\eta_i \epsilon (\frac{j-1}{\tau}, \frac{j}{\tau})$. The estimate of G_1 can then be used to form $Y_{1t} = \hat{G}_1(Y_t|y_{t-1})$ and Y_{1t} used to form Y_{2t} in a similar manner. The process continues until Y_{dt} is thought to be independent of $Y_{d,t-d-1}$ as judged by a 2x2 contingency table test.

Details of the procedure are illustrated by application to the runoff record for the Barwon River at Walgett (described in section 9). Assuming non-seasonality of dependence pattern we begin with the observed permutation of ranks. Since there are 49 years of record there are 588 observations and the order of occurrence is given by a permutation of the first 588 integers. The observed permutation matrix $\underset{\sim}{H}$ is plotted on the unit square by dividing each integer in the permutation by 588 and plotting every pair of consecutive values. Imagine then a unit square with 587 points on it. The square is sub-divided in 15^2 smaller squares by dividing up the range of each unit axis into 45 equal parts and the frequency with which points fall in these small squares is given in Table 3.

Table 3 gives an idea of what an estimate of G_1 looks like since it is essentially the bivariate histogram. Actually $\hat{G}_1$ is formed by using intervals on the horizontal axis only and forming a sample distribution function for each interval. With this estimate of G_1, Y_{1t} are formed and Table 4 then gives a similar picture to Table 3 but using Y_{1t}, $Y_{1,t-2}$. Since the most likely association is high with high and low with low values, a 2x2 contingency table with equal frequencies in each row and column seems appropriate for testing the independence of Y_{1t}, $Y_{1,t-2}$. The observed frequencies in such a table are

	$Y_{1,t-2} \leq 0.5$	$Y_{1,t-2} > 0.5$
$Y_{1t} \leq 0.5$	152	142
$Y_{1t} > 0.5$	142	152

giving a χ^2 value of 0.68. In this case one order of dependence seems adequate.

11. *SIMULATION OF A HYDROLOGIC PROCESS*

eginning with a generator for independent $R[0,1]$ variates giving $\{Y_{dt}\}$ we may use estimated G_i, $i = 1,2,\ldots,d$ to transform back to $Y_{d-1,t}, Y_{d-2,t} \ldots Y_{1t}, Y_t$. The Y_t obtained are marginally $R[0,1]$ and have the estimated dependence structure.

It is not necessary to obtain $\underset{\sim}{H}$ and then generate a fresh set of $R[0,1]$ variates since we may simply use the Y_t. Then

$$Y_t = F^*(X_t)$$

using some smooth estimator F^* gives generated observations X_t.

REFERENCES

[1] Box, G.E.P. and Jenkins, G.M.(1970) Time series analysis; forecasting and control. Holden-Day, San Francisco.

[2] David, H.A.(1970) Order statistics. Wiley, New York

[3] Dronkers, J.J.(1958) Approximate formulae for the statistical distributions of extreme values *Biometrika* 45, 447-470.

[4] Durbin, J. and Knott, M.(1972) Components of Cramer-von Mises statistics I, *J. Roy. Statist. Soc. B.*, 34, 290-307.

[5] Fisher, R.A. and Tippett, L.H.C.(1928) Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proc. Camb. Phil. Soc.* 24, 180-190.

[6] Gumbel, E.J.(1958) The statistics of extremes. Columbia University Press, N.Y.

[7] Jenkins, G.M. and Watts, D.G.(1968) Spectral analysis and its applications. Holden-Day, San Francisco.

[8] Hurst, H.E.(1957) Long term storage capacity of reservoirs. *Trans. Amer. Soc. Civil Engs.* 116, 770-799.

[9] Huxham, S.H. and McGilchrist, C.A.(1969) On the extreme value distribution for describing annual flood series. *Water Resources Research* 5, 1404-5.

[10] Huxham, S.H. and McGilchrist, C.A.(1974) Extreme values from a stationary Markov sequence. submitted to *Ann. Statist.*

[11] Mandelbrot, B.R.(1972) Statistical methodology for nonperiodic cycles: from the covariance to R/S analysis. *Annals of Economic and Social Measurement* 1, 259-290.

[12] Mandelbrot, B.R. and Van Ness, J.W.(1968) Fractional Brownian motions, fractional noises and applications. *SIAM Review* 10, 422-437.

[13] Mandelbrot, B.R. and Wallis, J.R.(1968) Noah, Joseph and operational hydrology *Water Resources Research*, 4, 909-919.

[14] McGilchrist, C.A.(1972) A test for exponential tail v finite upper bound or Cauchy tail. *Res. Sect. Bulletin(NSW) Stat. Soc.* 3, 1-10.

[15] McGilchrist, C.A. and Huxham, S.H.(1972) A note on some simple time series estimates. *Technometrics* 14, 223-227.

[16] Markham, C.G.(1970) Seasonality of precipitation in the United States, *Ann. Assn. of American Geographers* 60, 593-597.

[17] Rodríguez-Iturbe, I.(1968)A modern statistical study of monthly levels of the Orinoco River. *Bull. International Assn. of Scientific Hydrology* 13, 25-41.

[18] Rosenblatt, M.(1956) Remarks on some nonparametric estimates of a density function.*Ann.Math.Statist.*42,1815-1842.

[19] Wegman, E.J.(1972) Nonparametric probability density estimation: I. A summary of available methods. *Technometrics* 14, 533-546.

[20] Watson, G.S. and Leadbetter,M.R.(1963) On estimating a probability density I. *Ann. Math. Statist.* 34, 480-491.

FIGURE 1

SEASONALITY OF LEIGH RIVER AT MT. MERCER

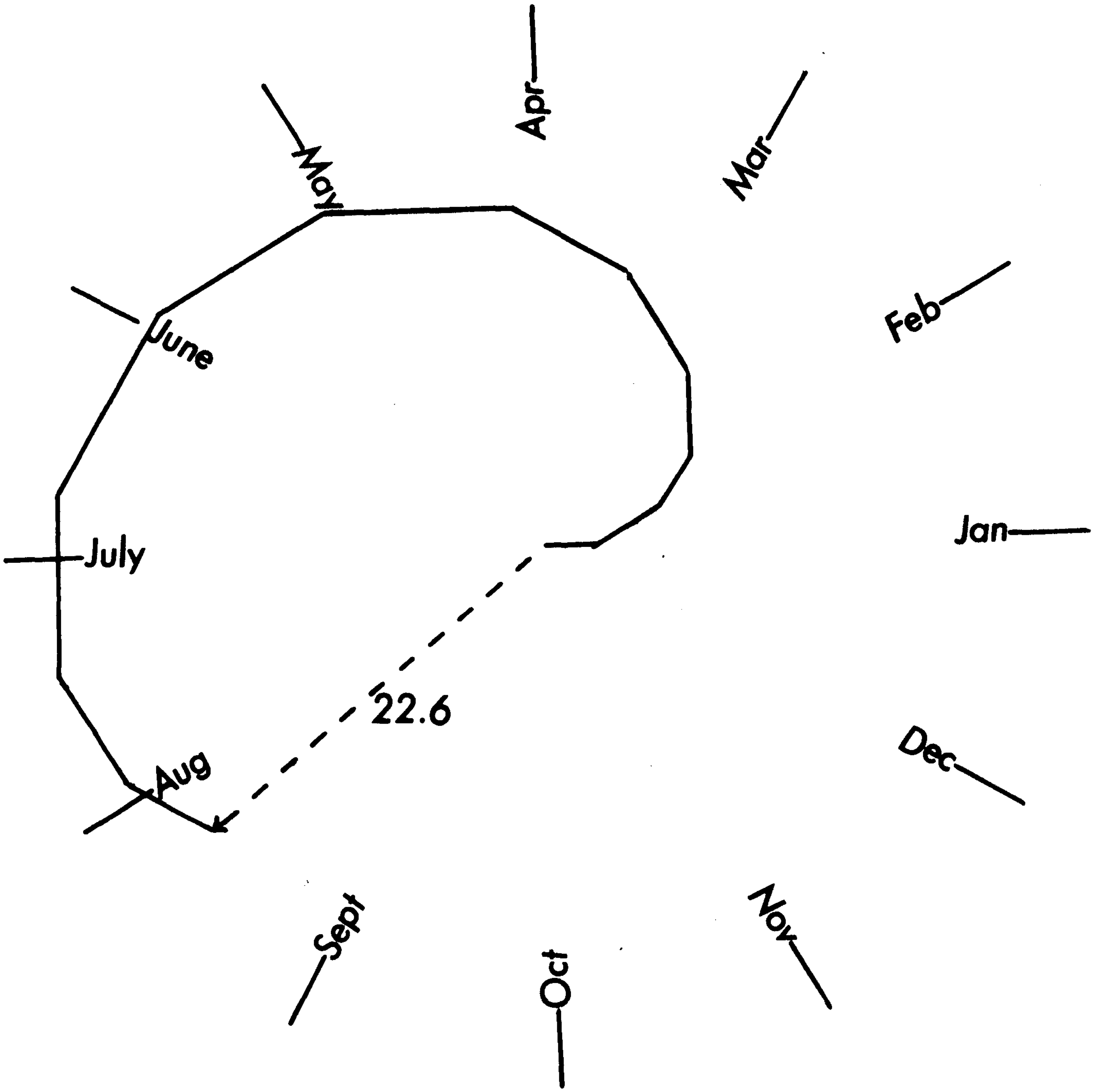

FIGURE 2

SEASONALITY OF DEPENDENCE STRUCTURE FOR BARWON RIVER AT WALGETT

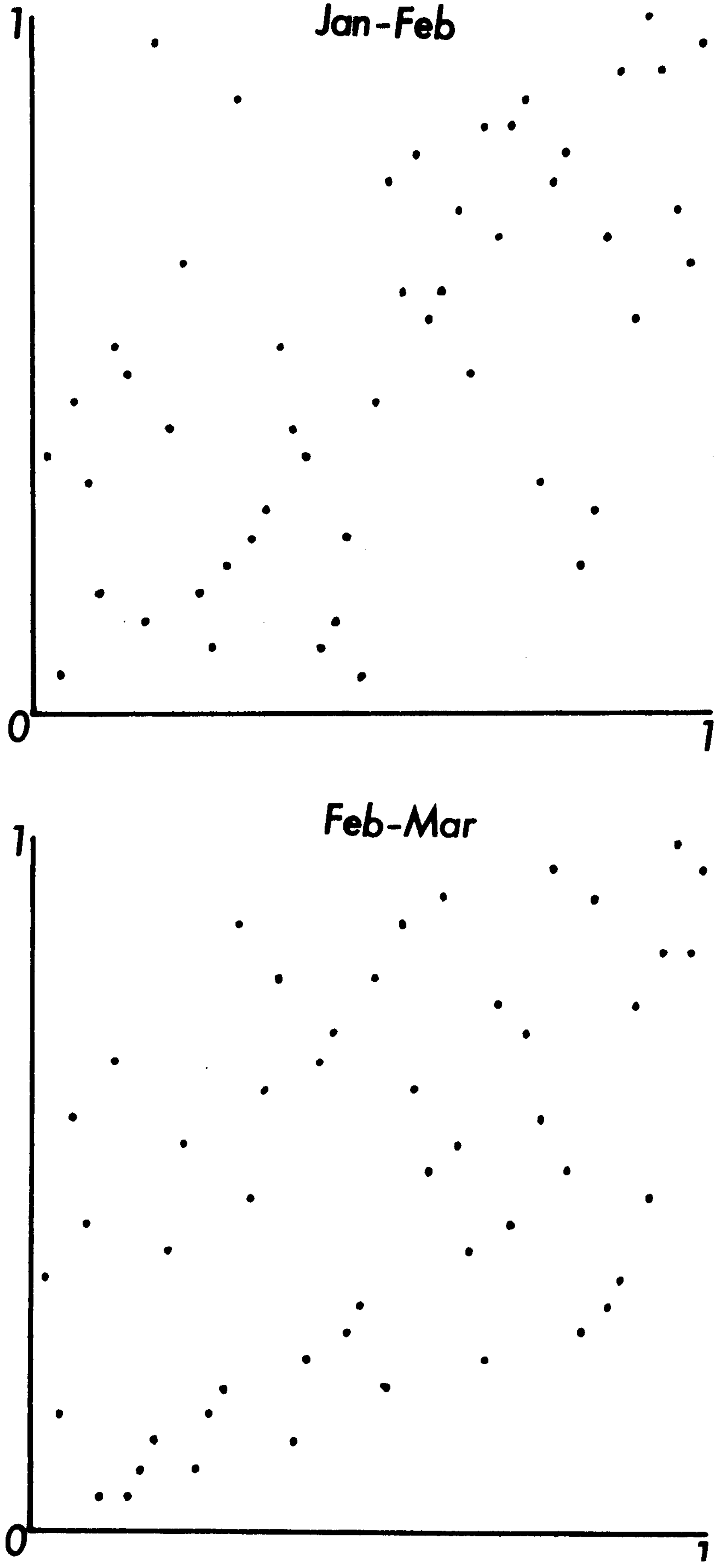

TABLE 1. MONTHLY RUNOFF OF LEIGH RIVER AT MT. MERCER.

May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.	Jan.	Feb.	Mar.	Apr.
0.72	0.67	0.95	1.28	2.51	6.38	0.75	0.49	0.37	0.49	0.34	0.38
1.02	0.49	1.25	9.17	7.36	13.00	1.41	0.72	0.37	0.68	0.77	0.63
0.58	0.80	0.93	2.86	1.76	4.69	0.57	1.17	0.34	0.41	0.39	3.65
19.00	10.70	28.70	16.80	33.40	6.14	3.95	0.88	0.45	0.41	0.89	1.32
0.77	1.57	5.65	6.43	2.77	1.22	0.61	0.60	0.43	0.34	0.57	0.43
0.79	1.73	2.48	4.53	10.10	9.09	1.10	0.88	8.62	1.27	0.48	0.45
3.80	3.71	21.00	10.60	11.00	2.01	0.69	0.41	0.35	0.62	0.34	1.12
0.92	1.90	18.60	14.70	17.30	20.70	3.69	1.07	0.66	0.32	0.37	0.82
0.71	0.97	4.02	22.50	4.85	0.84	0.97	1.07	0.34	1.14	0.93	0.53

RANK OF MONTHLY RUNOFF WITHIN EACH YEAR

May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.	Jan.	Feb.	Mar.	Apr.
7	6	9	10	11	12	8	4	2	5	1	3
7	2	8	11	10	12	9	5	1	4	6	3
5	6	7	10	9	12	4	8	1	3	2	11
10	8	11	9	12	7	6	3	2	1	4	5
7	9	11	12	10	8	6	5	2	1	4	3
3	7	8	9	12	11	5	4	10	6	2	1
9	8	12	10	11	7	5	3	2	4	1	6
5	7	11	9	10	12	8	6	3	1	2	4
3	6	10	12	11	4	7	8	1	9	5	2

AVERAGE MONTHLY RANKS

May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.	Jan.	Feb.	Mar.	Apr.
6.22	6.56	9.67	10.20	10.70	9.44	6.44	5.11	2.67	3.78	3.00	4.22

STANDARD ERRORS OF MONTHLY RANKS

May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.	Jan.	Feb.	Mar.	Apr.
2.44	2.01	1.73	1.20	1.00	3.00	1.67	1.90	2.83	2.68	1.80	2.95

TABLE 2. SEASONALITY OF DEPENDENCE FOR LEIGH RIVER AT MT. MERCER

RANKING OF POSITIVE DIFFERENCES IN SUCCESSIVE MONTHLY RANKS

17	1	18	19	80	48	49	50	20	81	2	102
103	51	52	3	67	21	68	22	53	4	54	82
55	56	23	24	69	70	107	108	57	25	97	5
6	7	26	27	83	84	85	58	71	86	8	87
28	29	59	30	31	9	32	72	88	89	90	73
33	60	34	74	35	36	10	75	11	91	61	98
12	13	62	37	92	14	63	64	76	93	99	38
39	15	16	40	41	42	43	44	106	65	77	94
66	45	95	100	78	96	79	104	105	46	101	47

AVERAGE RANK OF POSITIVE DIFFERENCE
OVER PAIRS OF SUCCESSIVE MONTHS

39.9	30.8	42.8	39.3	64	46.7	59.6	66.3	65.2	64.4	65.4	69.6

TABLE 3. CELL FREQUENCIES CORRESPONDING TO $\underset{\sim}{H}$ PLOT FOR BARWON RIVER AT WALGETT

0	0	0	0	1	0	1	0	1	2	2	5	7	7	21
0	0	0	1	0	1	0	2	1	4	1	7	5	9	5
0	0	1	2	2	2	0	0	4	2	3	6	5	5	4
0	0	1	1	1	0	3	9	3	4	5	8	4	2	7
0	0	0	2	0	4	3	1	3	2	3	5	4	3	6
0	0	0	3	1	2	4	5	5	4	3	4	2	2	1
0	0	1	3	1	4	2	7	4	3	3	2	4	2	0
1	1	2	8	7	4	7	6	2	2	4	2	1	1	0
0	3	3	7	5	5	2	1	1	3	2	3	0	0	1
0	2	8	3	4	2	4	2	2	2	2	2	2	0	1
0	2	1	7	2	3	4	7	3	2	1	0	1	2	1
4	6	9	5	4	4	2	2	3	4	4	0	0	1	0
4	6	7	5	2	3	0	3	1	0	1	2	1	0	1
5	13	1	0	4	0	3	3	2	1	2	0	0	2	0
22	2	2	1	2	2	1	0	1	1	0	2	0	0	0

Totals

36	36	36	48	36	36	36	48	36	36	36	48	36	36	48

TABLE 4. CELL FREQUENCIES FOR $Y_{1,t-2}$, Y_{1t}

FOR BARWON RIVER AT WALGETT

	8	2	6	2	2	2	3	2	1	3	3	4	2	4	5
	5	2	3	1	3	2	2	2	2	0	1	1	4	2	4
	1	0	4	2	2	2	2	6	4	5	2	3	5	4	3
	1	2	0	2	5	3	3	1	4	3	2	1	2	1	4
	1	2	4	4	0	3	3	3	2	4	1	0	3	2	2
	4	2	2	1	3	4	4	5	4	2	3	2	6	2	5
	1	2	1	1	3	1	2	4	2	3	3	2	5	1	3
	2	5	4	2	3	2	3	2	1	3	4	5	5	3	2
	1	4	3	5	2	0	2	4	1	3	3	1	1	0	4
	0	1	3	1	7	2	1	1	1	6	1	4	2	2	2
	2	2	5	5	4	3	3	3	3	4	3	1	3	3	5
	2	5	1	0	4	2	4	1	0	3	1	5	0	6	0
	3	1	3	2	5	5	0	5	5	4	3	3	3	0	3
	0	0	6	3	6	3	1	3	0	2	2	0	3	2	3
	3	4	0	3	1	0	1	3	4	4	2	2	1	2	4
Totals															
	34	34	45	34	49	34	34	45	34	49	34	34	45	34	49