# Non-negative Time Series and Shot Noise Processes as Models for Dry Rivers

by

Jane Luise Hutton

Thesis submitted for the

Diploma of Membership of Imperial College

and the degree of

Doctor of Philosophy in the University of London

August 1986

To

Mutti and Dad

Job 12:15

Abstract

Daily flow data are available on intermittent and ephemeral streams, which have periods without flow. The prominent features of these data are the occurrences of zero flows and the extreme variability of flow when present. A detailed description of data from Australia, Malawi and the United States of America is given, and the distinction between intermittent and ephemeral streams is examined.

Non-negative time series, or storage models, are investigated. The marginal distribution and various approximations to it are studied. Elaborations to include exact zeroes are discussed, as is the need to increase the variance of the distribution resulting from the basic model. The extension of earlier work on shot noise processes to include seasonality is considered.

Applications of the discrete and continuous time models to data from an intermittent stream on Malawi and an ephemeral stream in the United States of America are described. The numerical results from different forms of seasonal parameters are evaluated.

Acknowledgements

Professor Sir David Cox provided guidance and encouragement throughout this research, for which I am very grateful. Dr. David Jones gave a practical view of the problems addressed.

I would like to thank Antony, Charles, David, Marie and Patty, and others at Imperial College for stimulating discussions, not always statistical. My brother generously provided a comfortable flat. I am deeply indebted to my flatmates, particularly Liz, Priyan and Warren, and to Mary, Janet and Roy, for their patience, good humour and love.

The Natural Environment Research Council supported the work financially with a CASE Award, and the Committee of Vice-Chancellors and Principals awarded an Overseas Research Students Scholarship, both of which are gratefully acknowledged.

Profound thanks to my Father, who has made my years of study possible.

Table of Contents


| | | |
|---|---|---|
| Abstract | | 3 |
| Acknowledgements | | 4 |
| Table of Contents | | 5 |
| List of Tables | | 6 |
| List of Figures | | 8 |
| Chapter 1 | Introduction | 9 |
| Chapter 2 | Data analysis | |
| 2.1 | Introduction | 11 |
| 2.2 | Annual patterns of flow | 15 |
| 2.3 | Monthly patterns of flow | 34 |
| Chapter 3 | Storage models | |
| 3.1 | Introduction | 46 |
| 3.2 | Formulations | |
| | .1 Introduction | 49 |
| | .2 Deterministic description | 51 |
| | .3 Stochastic description | 55 |
| | .4 Seasonality | 56 |
| 3.3 | Derivation of Properties | |
| | .1 Basic properties, generating functions and likelihoods | 60 |
| | .2 Approximations to the marginal distribution | 66 |
| | .3 Results for truncated models | 78 |
| | .4 Comment on error models | 90 |
| | .5 Simulation results | 94 |
| Chapter 4 | Shot Noise Processes | |
| 4.1 | Introduction | 115 |
| 4.2 | Periodic Shot Noise Processes | 115 |
| 4.3 | Simulation results | 124 |
| Chapter 5 | Conclusions | 135 |
| References | | 138 |

List of Tables


| | | |
|---|---|---|
| 1. | Summary of data | 12 |
| 2. | Summary statistics for daily flow | 15 |
| 3. | Annual mean flows | 22 |
| 4. | Fit of three harmonics to unconditional mean daily flow | 25 |
| 5. | Fit of three harmonics to the probability of flow for daily data | 29 |
| 6. | Mean length of dry periods and endpoints in water year | 31 |
| 7. | Lower bounds for gradient of decreasing flow | 33 |
| 8. | Percentage of days without flow in each month | 34 |
| 9. | Number of months without flow in each year | 35 |
| 10. | Mean daily flow for each month; Malawi and USA | 36 |
| 11. | Mean daily flow for each month; Australia | 37 |
| 12. | Mean nonzero flow for each month; Malawi and USA | 38 |
| 13. | Coefficients of variation for flow in each month | 39 |
| 14. | Comparison of storage model and gamma distribution with an atom at zero | 71 |
| 15. | Comparision of shot noise model and mixture of gamma distributions | 74 |
| 16. | Regression of log(proportion of zeroes) on log($\epsilon$) for simulations of $S_{n+1} = (\rho S_n - \epsilon + I_n)^+$ | 82 |
| 17. | Observed and expected numbers of zeroes | 84 |
| 18. | Regression of log(proportion of zeroes) on log($\epsilon$) for simulations of $S_{n+1} = (\rho S_n - \epsilon + I_n)^+$ | 86 |
| 19. | Observed and expected numbers of zeroes | 87 |
| 20. | Regression of log(proportion of zeroes) on log($\epsilon$) for simulations of $S_{n+1} = (\rho S_n - \epsilon)^+ + I_n$ | 88 |

21. Harmonic fit of input probability, $\delta$, and input size, $\theta$ 95

22. Estimates of $\delta$ and $\theta$ for step function periodic input probability and size 96

23. Annual statistics of ten simulations of storage models M2C8, $\rho$=.6 99

24. Coefficients of variation for positive flow; M2C8, $\rho$=.6 100

25. Conditional mean daily flow, for each month; M2C8, $\rho$=.6 101

26. Annual statistics of ten simulations of storage models M2C8, $\rho$=.8 105

27. Coefficients of variation for positive flow; M2C8, $\rho$=.8 105

28. Conditional mean daily flow, for each month; M2C8, $\rho$=.8 106

29. Annual statistics of ten simulations of storage models AR3 108

30. Coefficients of variation for positive flow; AR3 109

31. Conditional mean daily flow, for each month; AR3 111

32. Values of the mean, cv and skew of a seasonal shot noise process and the ratio of the third cumulants of the shot noise process and a gamma with the same first cumulants 120

33. Annual statistics of ten simulations of shot noise models M2C8 127

34. Coefficients of variation for positive flow; M2C8 129

35. Conditional mean daily flow, for each month; M2C8 130

36. Statistics of ten simulations of shot noise model; AR3 132

List of Figures


| | | | |
|---|---|---|---|
| 1 | M2C8 | Daily flow over two years | 17 |
| 2 | AR3 | Daily flow over two years | 18 |
| 3 | M2C8 | Log(coefficient of variation) vs log(area) | 20 |
| 4 | AR3 | Log(coefficient of variation) vs log(area) | 20 |
| 5 | M2C8 | Annual mean flows | 23 |
| 6 | AR3 | Annual mean flows | 23 |
| 7 | Fitz: | Annual mean flows | 24 |
| 8 | M2C8 | Mean daily flows | 27 |
| 9 | AR3 | Mean daily flows | 27 |
| 10 | M2C8 | Five day means | 28 |
| 11 | AR3 | Five day means | 28 |
| 12 | M2C8 | Exponential probability plots | 43 |
| 13 | AR3 | Exponential probability plots | 44 |
| 14 | Fitz | Exponential probability plots | 45 |
| 15 | Probability plots for storage model simulations; $\lambda$=1 | | 69 |
| 16 | Estimated daily input rate and size; $\rho$=.6 | | 97 |
| 17 | M2C8 | Storage model simulation - MS1; daily flow | 103 |
| 18 | AR3 | Storage model simulation - Step2; daily flow | 114 |
| 19 | M2C8 | Shot noise simulation - SNS1; daily flow | 128 |
| 20 | AR3 | Shot noise simulation - Step2; daily flow | 134 |

# 1. Introduction.

The work described in this thesis was carried out at Imperial College and the Institute of Hydrology, Wallingford. The Institute of Hydrology provided flow data on dry rivers. The expression "dry river" covers both ephemeral and intermittent streams. An ephemeral stream flows only immediately after rain, usually for a few days; most of the year it is dry. When an ephemeral stream is flowing it is influent, i.e. contributing to groundwater. Such streams are found in arid and semi-arid zones. Intermittent streams occur in temperate regions; they flow during the rainy season and tend to dry up during the dry season. An intermittent stream acts both as an influent and an effluent stream - i.e. flow is augmented from groundwater - according to the season.

The Institute of Hydrology required methods to analyse and model these data. This study deals with time series and stochastic processes. Particular features are that the process spends appreciable time at zero and that when nonzero the flow is very variable. Models for ephemeral streamflow are reviewed in Kisiel, Duckstein and Fogel (1971). The stochastic models use random variables generated from distributions assumed for properties such as the start of the rainy season, the number of

flow events and volume of flow. Some models include deterministic recession. Lane, Diskin and Renard (1971) consider input-output relationships. The stochastic models are generalized by Diskin and Lane (1972) to allow dependence on stream-basin characteristics. Lee (1975) includes seasonality in this framework. Srikanthan and McMahon (1980) assess six procedures for generating monthly flows of ephemeral streams. Abdulrazzak and Morel-Seytour (1983) study input-output relationships. A model which combines deterministic recession with a Markov process for inputs is given by Yakowitz (1973). In Peebles, Smith and Yakowitz (1981) recession is determined from a leaky reservoir formulation. No work has been found specifically on intermittent streams.

Both discrete time and continuous time approaches have been studied. In chapter two the data are summarised; characteristics of daily and monthly flow are given. Chapter three describes non-negative time series or storage models. In chapter four seasonal forms of shot noise processes are analysed. Results from the simulation of three seasonal models are given. Recommendations on the use of the various models are made in chapter five.

# 2. Data Analysis

## 2.1 Introduction.

The Institute of Hydrology provided daily flow data on eleven dry rivers. There are records from four intermittent streams in Malawi and four ephemeral streams in the United States of America. The remaining three records are Australian.

The main characteristics of the rivers are given in Table 1. The Malawian and American rivers will be referred to by the codes given in the table, e.g. M1K1 or AR1, and the Australian records by name. The records are fairly short in length, generally 17 to 30 years. Within each country there is considerable overlap in the years for which data are given. The records for Fitzroy and Nogoa are unusually long for dry river data, 61 and 43 years respectively. there is considerable overlap in the years for which data are available. The data from U.S.A. and Australia were given in 'water years', i.e. beginning on 1st October. The dry season usually ends just after the beginning of the water year. The Malawi data were re-ordered to be in water years, beginning on 1st October.

Table1.

*Summary of data.*

| | | Catchment area: km.$^2$ | Calendar Years. | Number of years. |
|---|---|---|---|---|
| Malawi | | | | |
| 1K1 | Tomali | 1680.0 | 1952-75 | 24 |
| 2C8 | Naisi | 75.0 | 1959-75 | 17 |
| 5D2 | Bua | 6790.0 | 1954-75 | 22 |
| 5D3 | Mtiti | 233.0 | 1958-75 | 18 |
| USA | | | | |
| AR1 | 09 510 100 | 11.6 | 1965-81 | 17 |
| AR2 | 09 505 350 | 367.8 | 1961-81 | 21 |
| AR3 | 09 513 800 | 215.7 | 1961-81 | 21 |
| NM4 | 08 400 000 | 681.3 | 1952-81 | 30 |
| Australia | | | | |
| Todd | 006 009 | 452.0 | 1953-80 | 28 |
| Fitzroy | GS 130 003 | 132140.0 | 1922-82 | 61 |
| Nogoa | GS 130 201 | - | 1920-62 | 43 |

The catchment areas vary widely; the range for Malawi is 80 to 7000 km$^2$. The mean daily flows, given that flow is positive, of the Malawi rivers have the same ranks as the catchment areas. M5D3 is a tributary of M5D2, so the M5D3 catchment is a subsection of that of M5D2. The larger catchment reduces the variation in flow; M5D2 has smaller coefficients of variation and skewnesses. The American catchments range from 10 to 700

$km^2$. The conditional mean flows for these rivers also reflect the catchment sizes.

The flow data are given in various units. The Malawi data is the mean of two daily readings, given in cubic metres per second (cumecs). The rating curves, which are used to convert measurements of water depth into rates of flow, changed during the records for all the rivers. M1K1 and M2C8 are reasonably defined, whereas M5D2 and M5D3 are adequately recorded for low flows, but less reliable for high flow. Extreme flows will usually be above the level for which the rating curves have been calibrated, and the values for the corresponding flows will be found by extrapolation. The peak values of the Malawi flow data are probably the result of cyclones having been driven inland. Extreme values might be influential in the estimation of parameters for theoretical models, and the basic descriptive statistics.

The U.S.A. data are in cubic feet per second. The method of collecting the data is not explained. The absence of missing values might indicate that some values have been estimated. The U.S.A. rivers have some very high values (e.g. 13000.0 for NM4). Todd flows are in cumecs. Fitzroy and Nogoa are given as daily volume in megalitres. The water

depth gauges used for Fitzroy and Nogoa were changed once in each record. Eleven years have missing data in the Todd record, with approximately 100 to 300 observations missing for each of these years. Fitzroy has missing data in the first and last years of the record; the lack of missing data in the remaining years might again indicate estimated data. Nogoa has missing data estimated as total monthly flow for some months, which are not specified.

## 2.2 Annual patterns of flow.

Table 2 shows the percentage of days for which the observations are zero. As expected, the intermittent rivers generally have a lower

Table 2

*Summary statistics*

| | | | | | Conditional on non-zero flow. | | |
|---|---|---|---|---|---|---|---|
| | % 0 | Mean | CV | Skew | Mean | CV | Skew |
| M1K1 | 46.6 | 3.18 | 3.35 | 11.83 | 5.96 | 2.35 | 9.18 |
| M2C8 | 16.7 | .70 | 2.51 | 6.36 | .76 | 2.25 | 5.88 |
| M5D2 | 13.2 | 16.68 | 1.87 | 2.56 | 19.24 | 1.70 | 2.33 |
| M5D3 | 10.9 | .81 | 5.88 | 21.63 | .91 | 5.53 | 20.46 |
| AR1 | 24.1 | 1.01 | 6.16 | 15.51 | 1.33 | 5.34 | 13.55 |
| AR2[1] | 71.4 | 34.46<br>32.29 | 5.44<br>3.29 | 46.08<br>4.08 | 120.39<br>113.11 | 2.78<br>1.54 | 27.33<br>2.39 |
| AR3 | 72.2 | 9.37 | 6.24 | 10.36 | 33.73 | 3.17 | 5.37 |
| NM4[2] | 98.5 | 2.47 | 51.80<br>23.31 | 96.58<br>38.58 | 168.92 | 6.18<br>2.31 | 11.64<br>3.61 |
| Todd | 87.9 | .50 | 11.05 | 21.19 | 4.18 | 3.72 | 7.34 |
| Fitz | 4.4 | 14099.79 | 4.42 | 10.03 | 14749.90 | 4.32 | 9.81 |
| Nogoa | 46.5 | 1545.70 | 7.20 | 15.25 | 2891.13 | 5.22 | 11.17 |

1. Statistics in small type are for 13100,3310 replaced by 1310,331.
2. Statistics in small type are for 13000 replaced by 1300 .

proportion of zero flows than ephemeral streams. Fitzroy is an intermittent stream, with one or no dry period each year. Nogoa has

almost half its record zero, and the daily data as presented suggest that it behaves as an intermittent or ephemeral stream depending on whether a given year has high or low rainfall. Two years of daily flows of M2C8 and AR3 illustrate the difference between an intermittent and an ephemeral stream, see figures 1 and 2.

The coefficients of variation of the daily flow range from 1.8 to 51.8. The coefficients are larger for the ephemeral streams, as are the skewnesses. NM4 has almost all observations zero and a few very high values; this is reflected in the large coefficient of variation and skewness. In the AR2 data there are two consecutive values which are outliers, and there is one on the NM4 record. Reducing these values by a factor of ten or a hundred gives flows which are of the same order as the surrounding observations. The adjusted cv and skewnesses are more compatible with those of the remaining rivers. Replacing 13000. by 1300. in NM4 reduces the cv and skewness to 23. and 39. respectively. The skewness of AR2 , 46.08, which seems large, reduces to 4.80 if the extreme data values, (13100.,3310.) are replaced by (1310.,331.) or by (131.,331.). However, there was no way of checking whether the data had been corrupted. These values clearly have a large influence on the statistics and analyses with and without the adjustments are considered.

Figure 1: *M2C8 Daily flow over two years, Oct. 1963 to Sept. 1965*

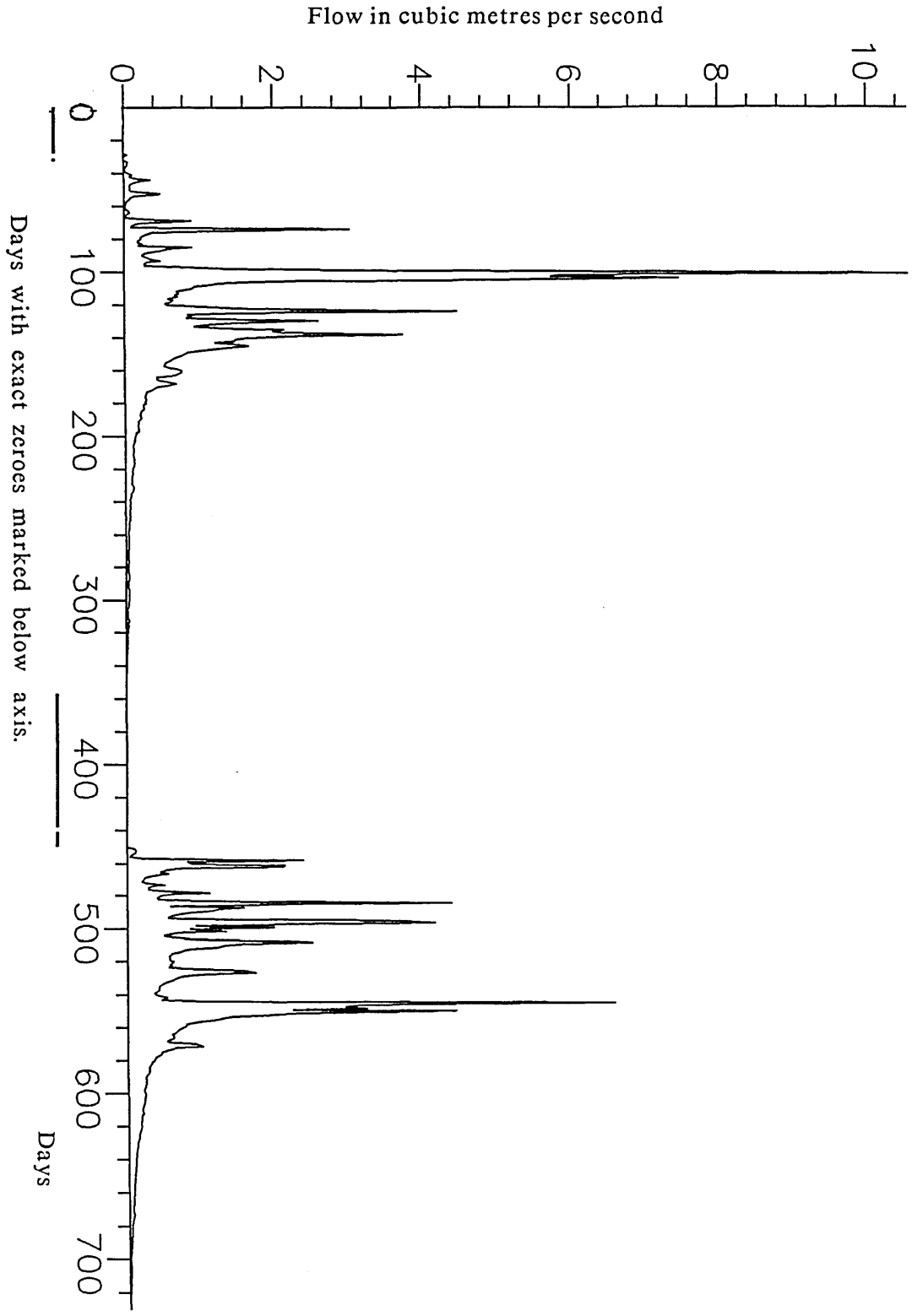
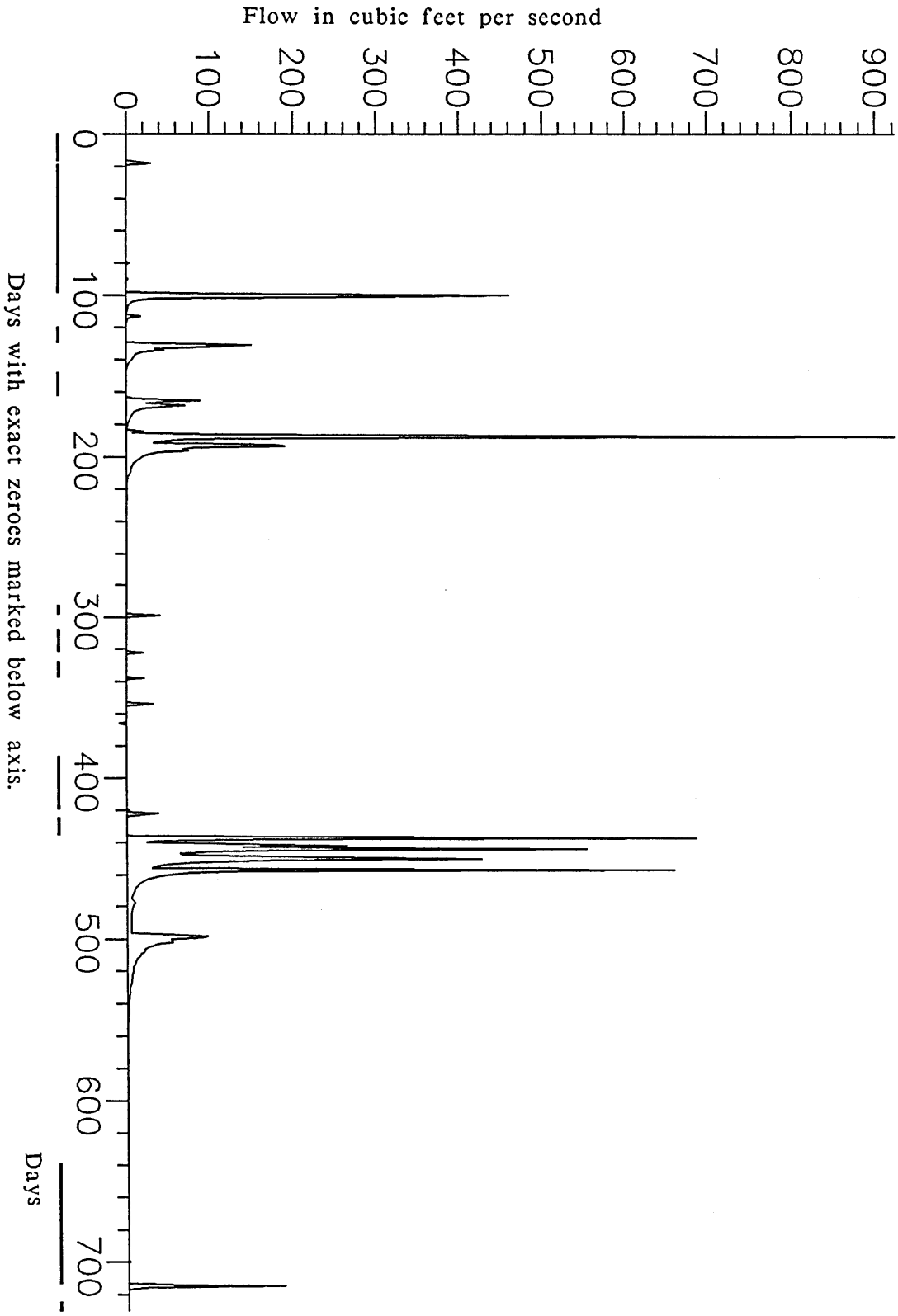


Figure 2 *AR3 Daily flow over two years, Oct. 1965 to Sept. 1967*

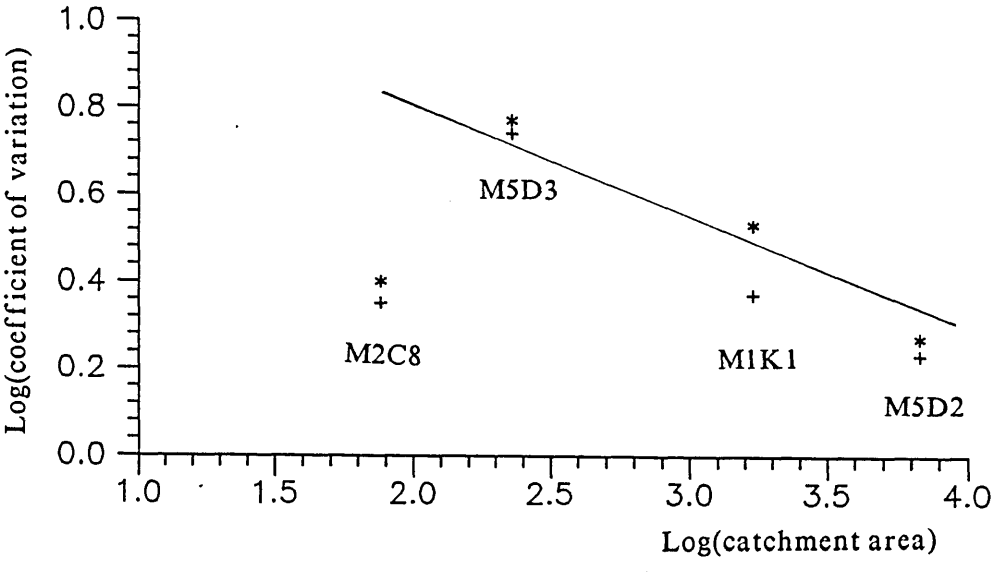


The coefficients of variation of positive flows lie in the range 1.7 to 6.2, with similar ranges for intermittent and ephemeral streams. Plots of the logarithm of cv against the logarithm of area shows near collinearity for M1K1, M5D2 and M5D3, for both conditional and unconditional cv, see figure 3. If the cv for the amended data is used for NM4 but AR2 is unaltered, the plot for conditional cv is also approximately linear, see figure 4. The gradients of the fitted lines are roughly $-\frac{1}{4}$ . The cv decreases as the catchment area increases, more or less as the inverse of the fourth root of the area, which is the square root of the putative length of the river. If the contributions to flow are thought of as occurring randomly along the river, the number, n, say, of random contributions would increase directly with the length. The coefficient of variation of the mean of n independent, identically distributed random variables decreases as $n^{-\frac{1}{2}}$, which concurs with the above.

The skewnesses of positive flow are much smaller than the unconditional skewnesses for AR2, AR3, NM4, and Todd, the rivers with most dry days. The full range is 2.3 to 27.3, with M5D3 and original AR2 the largest. The ranges for the different countries overlap considerably.

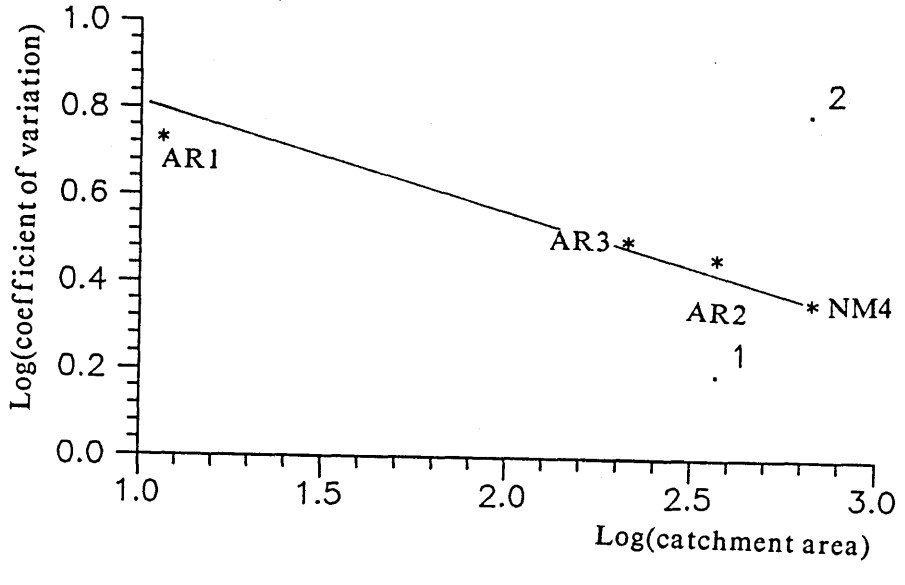
The annual mean flows of all the rivers are very variable, with no

Figure 3 Malaysi *Log(coefficient of variation) vs log(area)*




+ conditional, * unconditional

Figure 4 USA *Log(coefficient of variation) vs log(area)*




, * conditional

1. With adjustment 13100.,3310. to 1310.,331.
2 Without adjustment 13000., to 1300.

obvious structure in the variation, see table 3 and figures 5 to 7. There is some similarity in the occurence of dry and wet years within the Malawi data and within the U.S.A. data. The number of days without flow in each year is also very variable. The Malawi rivers have one main dry period each year. The U.S.A. rivers and Todd have periods of flow separated by dry periods. NM4 never has more than fourteen days with flow in a year. Todd flows for less than 70 days in most years, but in 1974 and 1976 for 292 and 271 days. The corresponding mean annual flows are large. There was high rainfall between 1973 and 1976, the effect of which on the hydrology of the region is discussed in Verhoeven (1977) ; the low flow volume and relatively few days with flow in 1975 are not explained. Fitzroy has a dry period in about a third, 23, of the years recorded. Nogoa appears to flow most of the year in some years and to have several dry periods in other years. These dry periods may be an artefact of the estimation of missing data; the estimated monthly flow seems to be given on one day and the values for the rest of the month set to zero.

Seasonality is evident in the occurence of the dry season during August to November for the Malawi rivers and Fitzroy, and in the clustering of flow events in the U.S.A. rivers and Nogoa. AR1, AR2, AR3 have more frequent and larger flow events from December to March. The

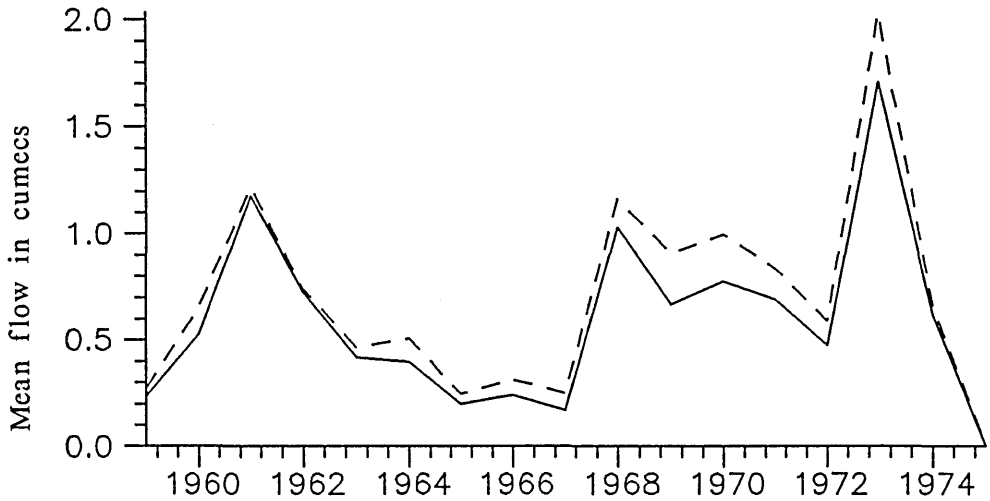
Table 3

*Mean annual flows for M1K1, M2C8, AR1, AR3 and Fitzroy*

| M1K1 | | M2C8 | | AR1 | | AR3 | | Fitzroy | |
|---|---|---|---|---|---|---|---|---|---|
| 4.16 | 5.80 | - | - | - | - | .49 | 59.67 | 58.28 | 58.28 |
| 4.78 | 7.21 | - | - | - | - | 2.89 | 8.95 | 51.87 | 51.87 |
| 7.38 | 8.39 | - | - | - | - | 1.66 | 18.37 | 151.08 | 159.39 |
| 5.80 | 8.13 | - | - | - | - | 3.39 | 8.67 | 3.99 | 3.99 |
| 1.31 | 3.41 | .24 | .27 | .62 | .96 | 11.51 | 18.94 | 6.79 | 7.44 |
| .15 | .67 | .53 | .65 | 1.85 | 2.84 | 16.41 | 24.54 | 36.00 | 43.95 |
| 1.12 | 2.65 | 1.18 | 1.22 | .07 | .10 | .35 | 1.15 | 26.48 | 130.92 |
| 3.03 | 4.50 | .72 | .74 | 1.53 | 1.97 | 6.82 | 13.88 | 148.31 | 159.60 |
| 13.51 | 16.67 | .42 | .46 | .33 | .39 | 2.19 | 5.13 | 2.27 | 2.56 |
| 1.28 | 3.12 | .40 | .51 | .48 | .72 | 2.18 | 11.05 | 49.74 | 72.33 |
| 4.37 | 6.34 | .20 | .25 | .03 | .04 | 2.17 | 14.95 | 274.80 | 303.03 |
| .61 | 1.43 | .24 | .31 | .01 | .01 | .14 | 5.71 | 86.23 | 91.74 |
| 1.86 | 3.62 | .17 | .25 | 2.90 | 2.92 | 27.54 | 42.56 | 26.31 | 29.82 |
| .00 | I | 1.03 | 1.16 | .08 | .11 | .01 | 1.13 | 539.32 | 539.32 |
| 1.63 | 3.15 | .67 | .91 | .17 | .23 | 1.56 | 63.31 | 147.94 | 422.94 |
| 1.27 | 3.33 | .78 | 1.00 | .35 | .54 | 3.69 | 33.72 | 430.56 | 430.56 |
| 2.88 | 7.46 | .69 | .83 | .03 | .05 | .12 | 8.92 | 176.88 | 176.88 |
| 1.09 | 2.81 | .47 | .59 | 2.31 | 3.24 | 29.75 | 96.10 | 244.16 | 244.16 |
| .79 | 2.21 | 1.71 | 2.03 | 3.12 | 3.28 | 37.59 | 63.23 | 188.46 | 188.46 |
| 6.83 | 9.47 | .62 | .66 | 3.17 | 3.17 | 46.27 | 105.18 | 35.13 | 35.13 |
| 1.07 | 1.96 | .00 | .00 | .08 | .08 | .01 | .72 | 95.16 | 111.70 |

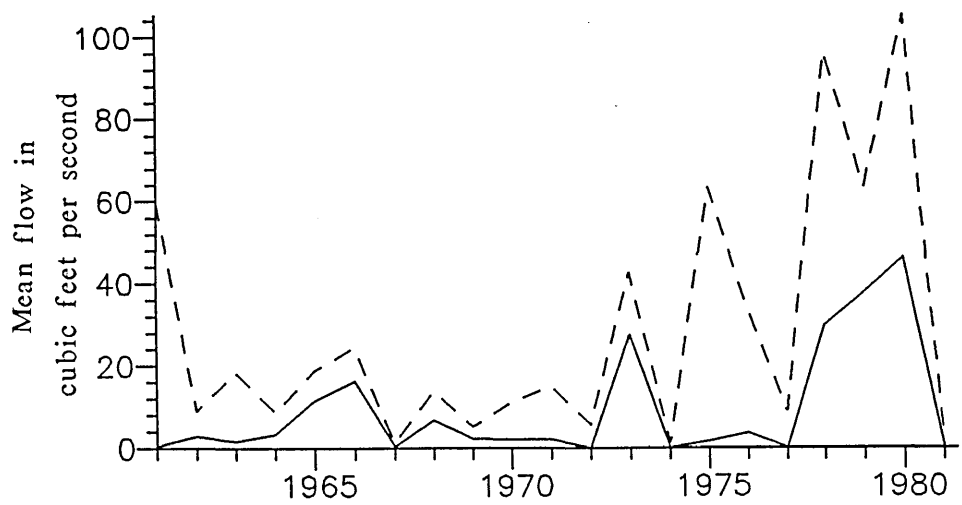
The 1st column for each river is mean of all flows, the 2nd is mean of nonzero flows.
On each line the Malawi means, and the USA and Ftizroy means are contemporaneous.

Figure 5 *M2C8 Annual mean flows, 1959 - 1975*



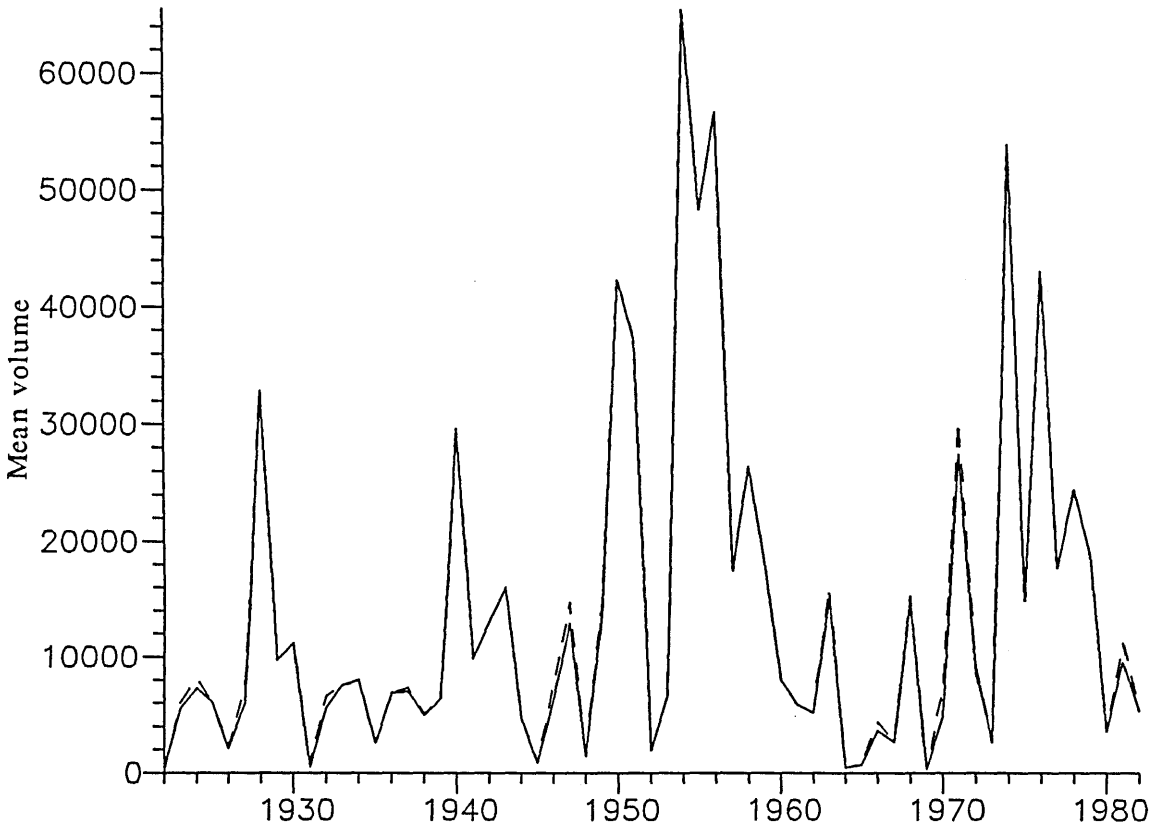

Solid line - mean of all flows Dotted line - mean of nonzero flows

Figure 6 *AR3 Annual mean flows, 1961 - 1982*




Solid line - mean of all flows Dotted line - mean of nonzero flows

Figure 7 *Fitzroy Annual mean flows, 1922 - 1982*




Solid line - mean of all flows  Dotted line - mean of nonzero flows

flow events in NM4 occur between May and September, and from December to April for Nogoa.

The first three harmonics were fitted to the mean daily flow, where the mean is taken over the years in the record conditional on the data's not being indicated as missing. Table 4 gives the percentage of total variation explained by the harmonics. The intermittent rivers vary

Table 4

*Fit of three harmonics to unconditional mean daily flow*

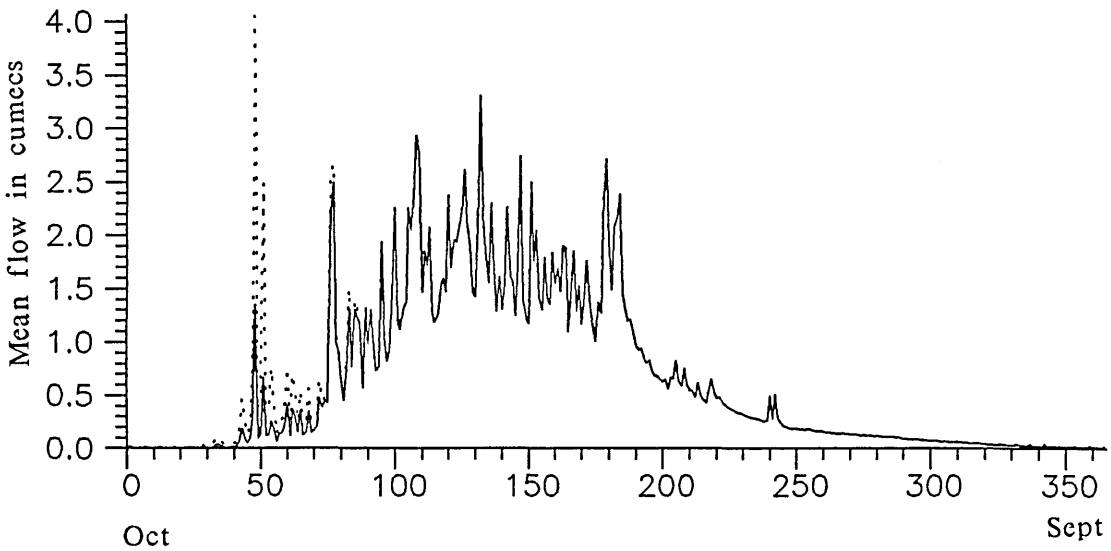
| | % of variation explained by harmonics | | | |
|---|---|---|---|---|
| | 1st | 2nd | 3rd | 1-3 |
| Malawi | | | | |
| 1K1 | 64 | 21 | 4 | 89 |
| 2C8 | 70 | 12 | 0 | 83 |
| 5D2 | 66 | 26 | 7 | 98 |
| 5D3 | 35 | 18 | 6 | 59 |
| USA | | | | |
| AR1 | 35 | 10 | 1 | 46 |
| AR2 | 38 | 10 | 8 | 57 |
| AR3 | 37 | 13 | 1 | 51 |
| NM4 | 2 | 1 | 1 | 3 |
| Australia | | | | |
| Todd | 13 | 4 | 21 | 20 |
| Fitzroy | 59 | 20 | 5 | 85 |
| Nogoa | 32 | 9 | 44 | 44 |

fairly regularly, with 60% to 98% of variation explained by the first three harmonics. The ephemeral streams show some regular variation, though less than the intermittent streams. NM4 is an exception; there is essentially no regular variation. Graphs of the mean daily flow have sharp peaks imposed on the basic periodic variation, see figures 8 and 9. The graph for M5D2 is, however, very smooth. Five day means were calculated for M2C8 and M5D3, with and without conditioning on positive flow. The graphs of these means also show large fluctuations, see figures 10 and 11. This again reflects the variability of the data, particularly the irregular occurence of large flows.

The results of fitting harmonics to the proportion of days over the years in the record with flow for daily data are given in table 5. Most rivers have more than 90% of variation explained by the first three harmonics. The exceptions are NM4 and Todd. The rivers with greater overall proportion of zero flows, M1K1, AR1, AR2, and AR3 have more variation explained by the first harmonics than the rivers with more days with positive flow. The distinction between intermittent and ephemeral streams is not clear from these statistics.

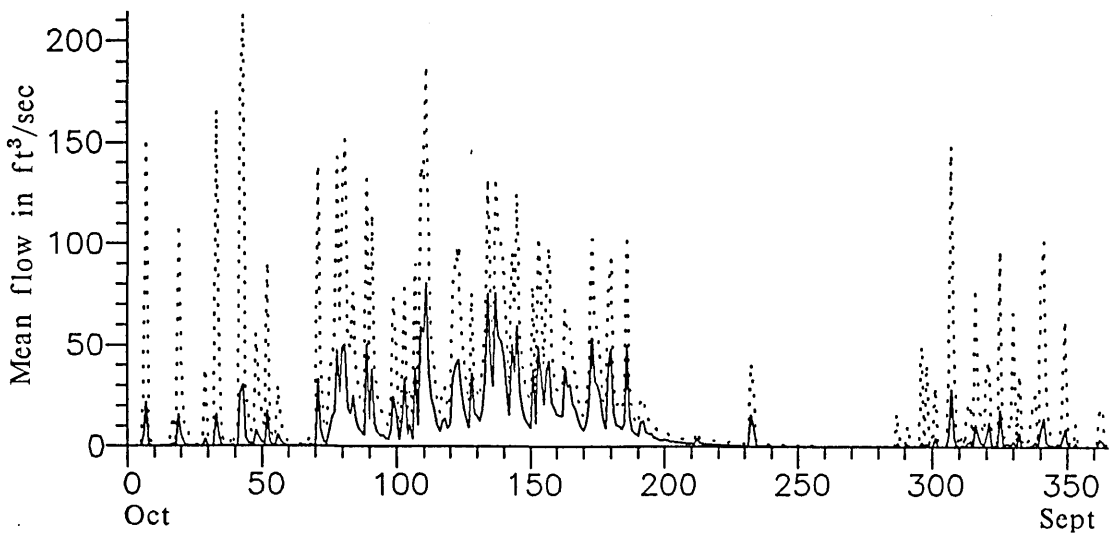
Coefficients of variation for daily flows range from .5 to 7.0. The

Figure 8 *M2C8 Mean daily flows*



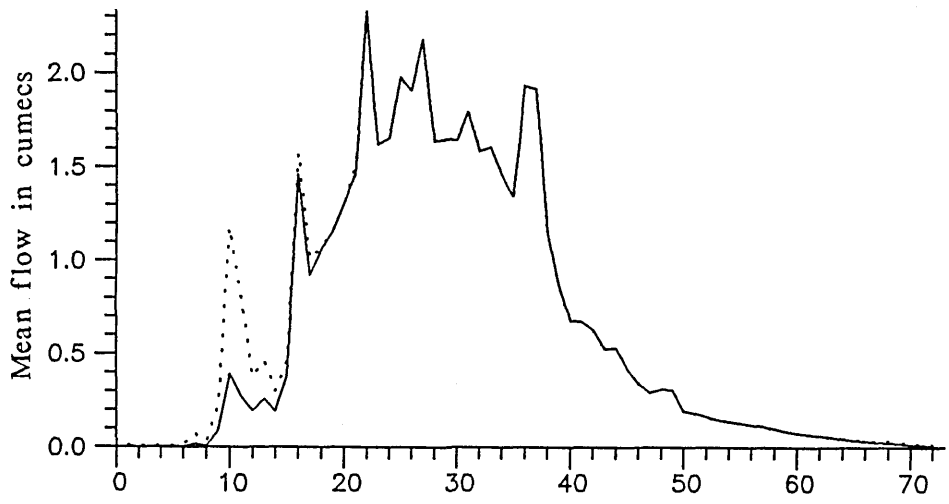

Solid line - mean of all flows Dotted line - mean of nonzero flows

Figure *9 AR3 Mean daily flows*



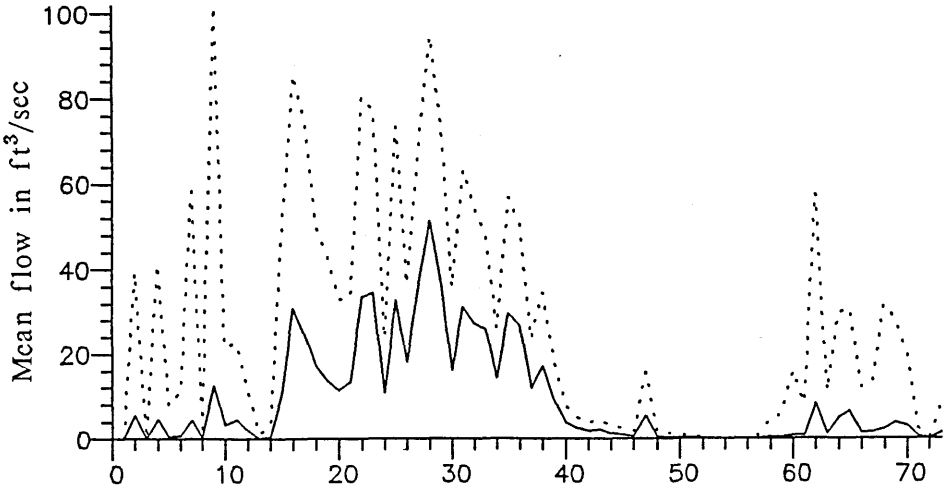

Solid line - mean of all flows Dotted line - mean of nonzero flows

Figure 10 *M2C8 Five day means.*




Solid line - mean of all flows Dotted line - mean of nonzero flows

Figure 11 *AR3 Five day means*




Solid line - mean of all flows Dotted line - mean of nonzero flows

Table 5

*Fit of three harmonics to the probability of flow for daily data*

| | | % of variation explained by harmonics | | | |
|---|---|---|---|---|---|
| | | 1st | 2nd | 3rd | 1-3 |
| Malawi | | | | | |
| | 1K1 | 83 | 13 | 2 | 98 |
| | 2C8 | 61 | 29 | 7 | 97 |
| | 5D2 | 61 | 27 | 9 | 97 |
| | 5D3 | 50 | 28 | 11 | 89 |
| USA | | | | | |
| | AR1 | 89 | 3 | 2 | 95 |
| | AR2 | 67 | 22 | 6 | 95 |
| | AR3 | 84 | 6 | 4 | 94 |
| | NM4 | 30 | 3 | 0 | 34 |
| Australia | | | | | |
| | Todd | 8 | 9 | 11 | 28 |
| | Fitzroy | 64 | 20 | 7 | 91 |
| | Nogoa | 88 | 8 | 4 | 94 |

lower values occur during the wet season or season with more flow events. The larger values, with greater fluctuations, occur during dry periods. This reflects variation in duration of the dry periods and occasional flow within these periods.

The length of the dry season in intermittent streams is of interest. There might be a few days with low flows between runs of zeroes or

isolated events before the continuous flow of the wet season begins.Two possible definitions for this length are:

1) Longest length: the longest consecutive run of zeroes from April to March of the following year.

2) Extreme length: the length from first to last zero in the year from April to March.

Table 6 gives the results for Malawi, AR1 and Fitzroy. AR1 is included as only 25% of days are without flow. The difference between these two definitions is most marked for AR1, 70 and 128 days, as there are several years with flows during the dry season. The results for the remaining rivers show that there is little to choose between the definitions because most years have only one run of zeroes. The standard deviations of the lengths and the endpoints are generally slightly smaller for the first definition, which is used hereafter. The results show that the length of the dry season is very variable, covering a wide range from zero. M1K1 has a much longer mean dry season, 161 days, than the other Malawi rivers. The coefficients of variation for AR1 and Malawi are between .5 and .6for longest length and slightly smaller for extreme length. Fitzroy has a short mean dry season, but high cv, .9, and the longest dry season is three times the mean length. The standard deviations of the beginnings of the dry seasons are similar to the mean lengths; the ends of the seasons are more

Table 6

*Mean length of dry periods and endpoints in water year.*

| *Longest* | AR1 | M1K1 | M2C8 | M5D2 | M5D3 | Fitz |
|---|---|---|---|---|---|---|
| Length | 69.5 | 161.5 | 59.1 | 63.9 | 50.8 | 49.0 |
| s.e. | 37.2 | 79.9 | 32.7 | 39.0 | 28.3 | 42.9 |
| c.v. | .54 | .50 | .55 | .61 | .56 | .88 |
| Beginning | 311.6 | 303.5 | 361.7 | 20.1 | 16.6 | 10.5 |
| s.e. | 43.0 | 70.7 | 24.2 | 31.0 | 28.7 | 45.6 |
| End | 16.2 | 99.5 | 55.8 | 84. 0 | 67.4 | 59.5 |
| s.e. | 50.2 | 26.8 | 20.1 | 17.0 | 6.5 | 18.4 |
| Max. No.[1] | 4 | 3 | 2 | 7 | 6 | 42 |
| Min. length | 26 | 44 | 12 | 3 | 10 | 3 |
| Max. length | 151 | 260 | 104 | 127 | 89 | 164 |

| *Extreme* | AR1 | M1K1 | M2C8 | M5D2 | M5D3 | Fitz |
|---|---|---|---|---|---|---|
| Length | 127.9 | 164.7 | 67.0 | 66.6 | 63.1 | 52.4 |
| s.e. | 51.9 | 75.9 | 34.2 | 37.5 | 30.2 | 44.9 |
| c.v. | .41 | .46 | .51 | .56 | .38 | .86 |
| Beginning | 277.2 | 306.8 | 361.7 | 20.1 | 12.4 | 9.0 |
| s.e. | 37.4 | 68.2 | 24.2 | 31.0 | 26.0 | 44.8 |
| End | 40.2 | 106.5 | 63.7 | 86.8 | 75.5 | 61.4 |
| s.e. | 32.8 | 24.5 | 22.5 | 17.3 | 15.9 | 18.6 |
| Min. length | 26 | 44 | 12 | 3 | 23 | 3 |
| Max. length | 208 | 260 | 112 | 127 | 114 | 177 |

1. Max. No. gives the maximum number of dry periods in any one year taken from 1st April to 31st March.

predictable. AR1 is an exception to this with the first definition; its dry period is better defined by the extremes, this distinguishes it from the intermittent streams. The dry season occurs roughly simultaneously in the Malawi records and Fitzroy, M1K1 beginning earlier and ending later.

As the dry periods are often preceeded by runs of small values, the lengths were found with zero defined to be less than .005 and.01 predictable. AR1 is an exception to this with the first definition; its dry period is better defined by the extremes, this distinguishes it from the intermittent streams. The dry season occurs roughly simultaneously in the Malawi records and Fitzroy, M1K1 beginning earlier and ending later.

As the dry periods are often preceeded by runs of small values, the lengths were found with zero defined to be less than .005 and.01 cumecs for the Malawi data which are given to three decimal places. The different thresholds make little difference to M1K1, M2C8 & M5D2. In M5D3 multiple 'dry' seasons result, and the length and its cv increase. There is no consistent change in the variances of the lengths or endpoints, and no advantage in this respect in using a threshold above zero.

Plots of flow against flow on the previous day show that the

correlation when flow is decreasing is generally high. The values of b such that most points for decreasing flow lie above flow(d)=b×flow(d-1) are given in the table 7.

Table 7.

*Lower bounds for gradient of decreasing flow.*

| M1K1 | M2C8 | M5D3 | AR1 | AR2 | AR3 | Nogoa |
|---|---|---|---|---|---|---|
| .5 | .5 | .85 | .67 | .5 | .5 | .33 |

There are some points close to the horizontal axis in the AR2 and AR3 plots; the points for AR4 are concentrated on or near the axes. Such points indicate abrupt rises from and decline to zero. The points for increasing flow for M5D2 mainly lie below flow(d)=1.1×flow(d-1), i.e. increases in flow are small. The points above this line are fairly close to the vertical axis, i.e. large increases occur when there is little flow.

## 2.3 Monthly patterns of flow.

The percentage of days with no flow taken over all years is given for each month in table 8. M2C8, M5D2, M5D3 and Fitzroy almost always flow throughout January to June. AR1 flows from January to April. The wettest season of M1K1 is February to March; days without flow occur during these months. Flow occurs more frequently from January to June in AR2 and AR3, whereas the few flow events of NM4 occur between May

Table 8

*Percentage of days without flow in each month.*

| | Malawi | | | | USA | | | | Australia | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Month | 1K1 | 2C8 | 5D2 | 5D3 | AR1 | AR2 | AR3 | NM4 | Tod | Fitz | No |
| October | 92 | 76 | 35 | 42 | 53 | 93 | 89 | 99 | 86 | 13 | 63 |
| November | 96 | 62 | 64 | 63 | 29 | 86 | 86 | 100 | 82 | 18 | 53 |
| December | 85 | 16 | 40 | 12 | 8 | 79 | 73 | 100 | 85 | 4 | 34 |
| January | 33 | 0 | 8 | 1 | 0 | 66 | 59 | 100 | 89 | 0 | 22 |
| February | 13 | 0 | 0 | 0 | 0 | 46 | 50 | 100 | 89 | 0 | 15 |
| March | 6 | 0 | 0 | 0 | 0 | 26 | 48 | 100 | 90 | 0 | 20 |
| April | 5 | 0 | 0 | 0 | 1 | 21 | 50 | 100 | 88 | 1 | 37 |
| May | 21 | 0 | 0 | 0 | 6 | 72 | 63 | 99 | 90 | 0 | 60 |
| June | 41 | 0 | 0 | 0 | 27 | 97 | 82 | 97 | 90 | 1 | 57 |
| July | 49 | 0 | 1 | 0 | 50 | 95 | 93 | 96 | 87 | 1 | 57 |
| August | 54 | 6 | 5 | 2 | 51 | 86 | 86 | 96 | 90 | 5 | 73 |
| September | 61 | 42 | 9 | 17 | 61 | 87 | 86 | 97 | 92 | 9 | 70 |

and October. The missing data in Todd and estimation of missing data for Nogoa mean that conclusions from these statistics are unreliable.

The minimum and maximum number of months without any flow in one year are given in table 9. This shows the greater variation in flow

Table 9

*Number of months without flow in each year.*

| | 1K1 | 2C8 | 5D2 | 5D3 | AR1 | AR2 | AR3 | NM4 | Tod | Fitz | Nog |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Min | 1 | 0 | 0 | 0 | 0 | 2 | 1 | 9 | 1 | 0 | 0 |
| Max | 12 | 3 | 2 | 3 | 4 | 10 | 10 | 12 | 12 | 3 | 9 |

from year to year in M1K1, AR2, AR3, Todd and Nogoa.

The maximum monthly mean flow in a year usually occurs in January, February or March for the Malawi rivers, clearly reflecting the climate of that region. The data for the U.S.A. rivers does not indicate a common rainfall distribution. Fitzroy and Nogoa have maximum flows in January to March.

The mean monthly flows are given in tables 10 and 11. These statistics complement the description above, basically showing a steady increase to the maximum and then a somewhat more gradual decline. The

Table 10

*Unconditional mean daily flow for each month; Malawi and USA*

| | Malawi | | | | USA | | | |
|---|---|---|---|---|---|---|---|---|
| | 1K1 | 2C8 | 5D2 | 5D3 | AR1 | AR2 | AR3 | NM4 |
| Oct | .01 | .00 | .25 | .03 | .23 | 6.82 | 1.90 | 1.64 |
| Nov | .01 | .17 | .07 | .04 | .14 | 11.47 | 4.54 | .0 |
| Dec | .85 | .77 | 2.81 | .70 | 1.85 | 51.06 | 15.15 | 0[1] |
| Jan | 7.75 | 1.64 | 15.61 | 1.71 | 1.87 | 27.27 | 21.05 | 0 |
| Feb | 12.49 | 1.85 | 53.01 | 4.09 | 3.15 | 67.61 | 31.17 | 0 |
| Mar | 8.92 | 1.60 | 68.89 | 2.31 | 3.00 | 117.20 | 24.67 | 0 |
| Apr | 5.13 | .90 | 38.92 | .49 | 1.21 | 116.05 | 6.96 | 0 |
| May | 2.09 | .39 | 13.16 | .15 | .31 | 12.43 | 1.63 | 1.31 |
| June | 1.06 | .17 | 4.29 | .11 | .08 | .01 | .11 | 1.06 |
| July | .66 | .10 | 2.41 | .10 | .02 | .24 | .42 | 2.03 |
| Aug | .36 | .05 | 1.39 | .07 | .03 | 1.12 | 3.89 | 18.93 |
| Sep | .16 | .02 | .67 | .05 | .32 | 3.52 | 1.82 | 4.25 |

1. 0 represents exact zero.

value for August for NM4, 18.93, decreases to 6.35 for the adjustment mentioned in §2.2, replacing 13000. by 1300. The December value of AR2 changes from 51.06 to 28.37 with the substitution of (1310.,331.) for (13100.,3310.). This gives a generally increasing sequence from June to March, followed by a sharp decline.

The conditional mean flows, given in tables 11 and 12 show the

Table 11.

*Mean daily flow for each month; Austaralia*

| | *Unconditional.* | | | *Conditional.* | | |
|---|---|---|---|---|---|---|
| | Fitzroy | Nogoa | Todd | Fitzroy | Nogoa | Todd |
| Oct | 1902.37 | 479.74 | 1.54 | 2113.97 | 1278.40 | 5.57 |
| Nov | 2288.14 | 1601.39 | 1.36 | 2471.38 | 2775.00 | 5.88 |
| Dec | 14446.44 | 1859.33 | .80 | 14819.13 | 2767.90 | 8.77 |
| Jan | 26894.47 | 1884.58 | .30 | 26894.47 | 2603.54 | 5.02 |
| Feb | 61955.32 | 6438.09 | .37 | 61955.32 | 6928.87 | 4.67 |
| Mar | 31424.85 | 2211.25 | .13 | 31424.85 | 2373.13 | 2.68 |
| Apr | 17333.19 | 1775.72 | .13 | 17334.67 | 1928.53 | 1.52 |
| May | 5135.13 | 902.45 | .13 | 5135.63 | 1660.00 | 1.45 |
| June | 4105.78 | 381.32 | .05 | 4105.79 | 725.84 | .61 |
| July | 4133.80 | 1065.26 | .24 | 4203.90 | 1936.13 | 2.87 |
| Aug | 1716.93 | 139.46 | .39 | 1807.30 | 438.76 | 3.08 |
| Sep | 612.04 | 169.51 | .51 | 655.79 | 471.58 | 5.80 |

same seasonal pattern. The range of values is smaller than that of the unconditional flows for M2C8, M5D2, M5D3, AR1 and Fitzroy, as the months with larger means have continuous flow. The difference between unconditional and conditional flow indicates that both the number of days with flow and the volume of flow, given that it occurs, vary with the time of year. For example, the ratio of conditional to unconditional flow is greater for September than August in AR3, though the percentage of

days without flow is the same.

Table 12.

*Mean nonzero daily flow for each month; Malawi and USA*

| | Malawi | | | | USA | | | |
|---|---|---|---|---|---|---|---|---|
| | 1K1 | 2C8 | 5D2 | 5D3 | AR1 | AR2 | AR3 | NM4 |
| Oct | .06 | .01 | .34 | .04 | .36 | 33.91 | 14.36 | 70.91 |
| Nov | .19 | .44 | .14 | .08 | .22 | 61.43 | 35.60 | .0 |
| Dec | 6.51 | .89 | 3.92 | .83 | 1.85 | 196.69[1] | 41.43 | - |
| Jan | 10.77 | 1.59 | 15.46 | 1.71 | 1.87 | 77.27 | 49.50 | - |
| Feb | 13.73 | 1.85 | 53.13 | 4.09 | 3.12 | 107.58 | 54.38 | - |
| Mar | 9.69 | 1.60 | 68.89 | 2.31 | 3.00 | 155.52 | 33.30 | - |
| Apr | 5.34 | .90 | 38.92 | .49 | 1.22 | 128.86 | 12.42 | - |
| May | 2.23 | .39 | 13.16 | .15 | .33 | 28.34 | 3.21 | 85.96 |
| June | 1.59 | .17 | 4.29 | .11 | .08 | .27 | .78 | 26.42 |
| July | 1.22 | .10 | 2.41 | .10 | .04 | 4.02 | 12.53 | 33.56 |
| Aug | .78 | .05 | 1.45 | .07 | .04 | 6.98 | 26.18 | 297.97[2] |
| Sep | .40 | .03 | .74 | .05 | .55 | 13.92 | 15.85 | 95.87 |

1. If we change 13100.,3310. to 1310.,331. the value is 123.20
2. If we change 13000., to 1300. the value is 102.97

Coefficients of variation for mean monthly flows, unconditional and conditional are given in table 13. The unconditional cvs are mostly greater than one, i.e. flow is overdispersed. The conditional flows vary less from year to year. The cvs of the Malawi rivers are generally smaller than the cvs of the other rivers. There is some tendency for the cv to be smaller when flow is continual.

Table 13(a)

*Coefficients of variation for daily flow in each month.*

| | Malawi | | | | USA | | | | Australia | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Month | 1K1 | 2C8 | 5D2 | 5D3 | AR1 | AR2 | AR3 | NA4 | Fitz | Nog | No |
| October | 3.2 | 2.2 | 1.5 | 3.3 | 4.0 | 3.8 | 4.0 | 4.5 | 3.8 | 3.9 | 3.4 |
| November | 4.8 | 2.2 | 2.3 | 2.2 | 3.0 | 2.4 | 2.4 | - | 2.0 | 4.6 | 1.9 |
| December | 2.1 | .9 | 1.6 | 1.6 | 2.3 | 2.7 | 2.3 | - | 2.2 | 2.9 | 2.4 |
| January | 1.5 | .7 | 1.2 | 1.3 | 1.7 | 1.8 | 2.4 | - | 2.0 | 1.4 | 2.5 |
| February | 1.1 | .8 | .8 | 1.4 | 1.9 | 1.4 | 2.1 | - | 1.7 | 2.6 | 2.8 |
| March | 1.4 | .9 | .6 | 1.5 | 1.8 | .9 | 2.1 | - | 1.8 | 2.0 | 2.5 |
| April | 1.5 | 1.2 | .8 | 1.8 | 1.3 | 1.3 | 2.4 | - | 2.4 | 3.4 | 2.2 |
| May | 1.3 | 1.3 | .9 | 2.0 | 1.3 | 3.6 | 3.1 | - | 2.5 | 4.5 | 2.5 |
| June | 1.4 | 1.0 | .9 | 2.0 | 1.6 | 3.7 | 2.3 | 4.9 | 3.5 | 3.2 | 2.6 |
| July | 1.4 | 1.1 | .9 | 2.7 | 1.7 | 2.4 | 1.9 | 2.6 | 2.9 | 3.3 | 2.0 |
| August | 1.5 | 1.1 | 1.0 | 2.5 | 1.3 | 2.5 | 1.8 | 4.7 | 4.5 | 4.5 | 2.6 |
| September | 1.9 | 1.4 | 1.1 | 2.6 | 4.0 | 3.5 | 2.1 | 4.1 | 3.2 | 3.1 | 1.9 |

With the exception of Fitzroy, the monthly cvs take values in ranges much lower than the overall cvs; i.e. variation within months contributes less to the total than variation between seasons. The conditional coefficients are near to one for M1K1, M2C8 and M5D2 , with range .6 to 1.4. M5D3, the U.S.A. rivers and Todd has larger values and wider range, 1.0 to 3.3. The rivers with long records, Nogoa and Fitzroy, have conditional cvs in the range 1.7 to 4.3. These larger values might reflect climatic changes during

the collection of data.

Table 13(b)

*Coefficients of variation for conditional daily flow in each month.*

| | Malawi | | | | USA | | | | Australia | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Month | 1K1 | 2C8 | 5D2 | 5D3 | AR1 | AR2 | AR3 | NM4 | Fitz | Nog | To |
| October | .7 | .8 | 1.1 | 2.8 | 3.3 | 1.8 | 1.6 | 1.7 | 3.6 | 2.8 | 1.4 |
| November | 1.4 | 1.5 | 1.3 | 1.3 | 2.5 | 1.6 | 1.2 | - | 1.9 | 3.2 | 1.2 |
| December | .7 | .8 | 1.3 | 1.6 | 2.3 | 1.1 | 1.4 | - | 2.2 | 2.1 | 2.3 |
| January | 1.1 | .7 | 1.1 | 1.3 | 1.7 | 1.2 | 1.7 | - | 2.0 | 1.3 | 1.4 |
| February | 1.0 | .8 | .8 | 1.4 | 1.9 | .9 | 1.5 | - | 1.7 | 2.4 | 1.4 |
| March | 1.3 | .9 | .6 | 1.5 | 1.8 | .6 | 1.7 | - | 1.8 | 1.9 | 1.2 |
| April | 1.4 | 1.2 | .8 | 1.8 | 1.3 | 1.1 | 1.7 | - | 2.4 | 3.2 | .9 |
| May | 1.2 | 1.3 | .9 | 2.0 | 1.3 | 2.3 | 2.1 | 1.8 | 2.5 | 3.2 | 1.2 |
| June | .9 | 1.0 | .9 | 2.0 | 1.5 | 1.0 | 1.1 | 1.6 | 3.5 | 2.2 | .6 |
| July | .8 | 1.1 | .9 | 2.7 | 1.1 | 1.5 | 1.3 | 1.0 | 2.9 | 2.4 | .9 |
| August | .7 | 1.1 | 1.0 | 2.5 | .8 | 1.6 | 1.0 | 3.2 | 4.3 | 2.4 | 1.4 |
| September | 1.3 | 1.0 | 1.0 | 2.5 | 3.2 | 1.7 | 1.1 | 2.0 | 3.1 | 1.7 | 1.2 |

Figures 12 to 14 show plots of nonzero flows in February, May and September against exponential order statistics for M2C8, AR3 and Fitzroy; these months represent large, mid-range and small conditional mean flows. The plots for M2C8 are reasonably straight, apart from the two or three largest values . The flows in May for AR3, with small mean are also close to exponential except for the two largest values. The

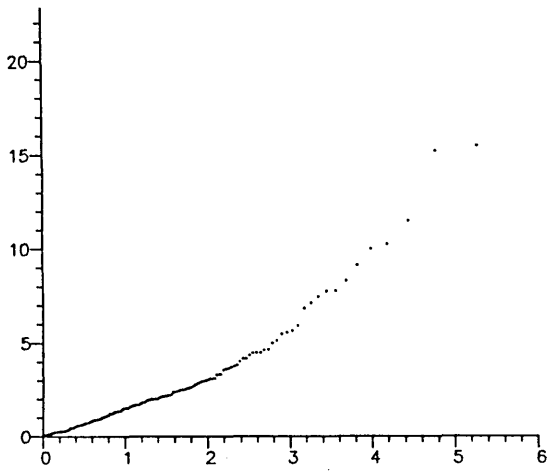
remaining plots show the distributions of flow are substantially longer tailed than the exponential.

This analysis shows that dry rivers are characterised both by periods without flow and by large variation in the flows. The difference between intermittent and ephemeral streams is seen in the pattern of zero observations. Intermittent streams have a dry and a wet season in the year. Ephemeral streams have dry periods throughout the year, with some clustering of flow events. In both cases there is considerable difference from year to year. The mean daily flows have large coefficients of variation and skewnesses, with ephemeral streams having greater values than the intermittent. This distinction is also seen when the statistics are calculated for each month. However, the variation is less within seasons than overall. Thus models for these data need to have periodic parameters for input rate and size. For intermittent streams these parameters probably need to be continuous, increasing to a maximum during March or April and declining to small values in November and October. The records are too short to determine changes in annual flow volume resulting from variation in the climate, but models need to allow for large variation in annual flows. The structure of any model must preserve the sharp increases in flow and more gradual declines; thus the models must be

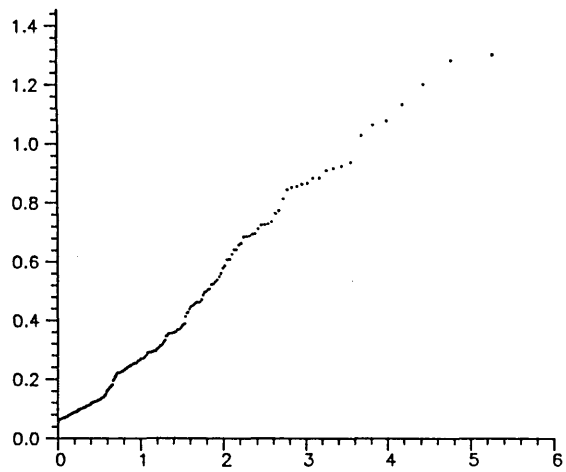
time-irreversible. It is of interest to find whether the difference between intermittent and ephemeral streams could be reproduced by the same model with different values for the parameters.

Figure 12 *M2C8 Exponential probability plots*

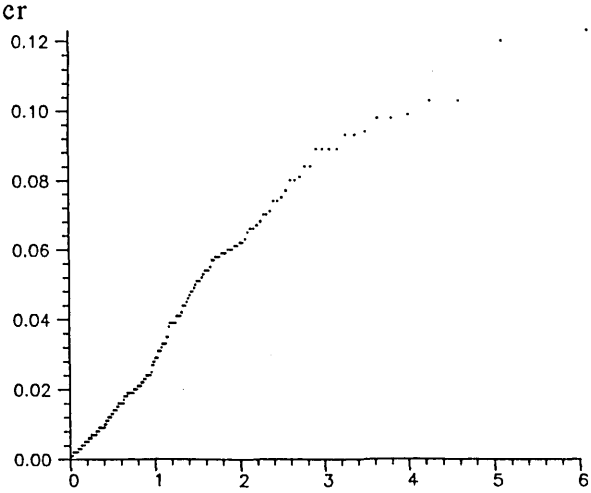
i) February




ii) May



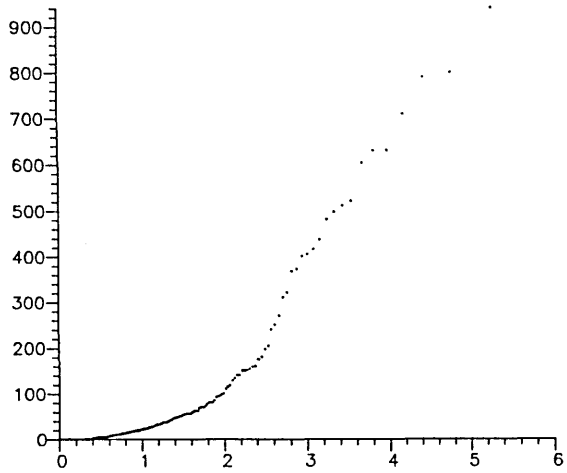

iii) September



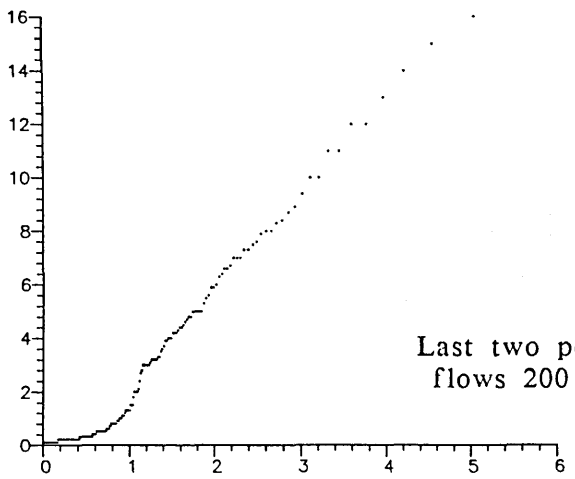

Vertical axes - ordered nonzero flows
Horizontal axes - exponential order statistics

Figure 13 *AR3 Exponential probability plots*

i) February

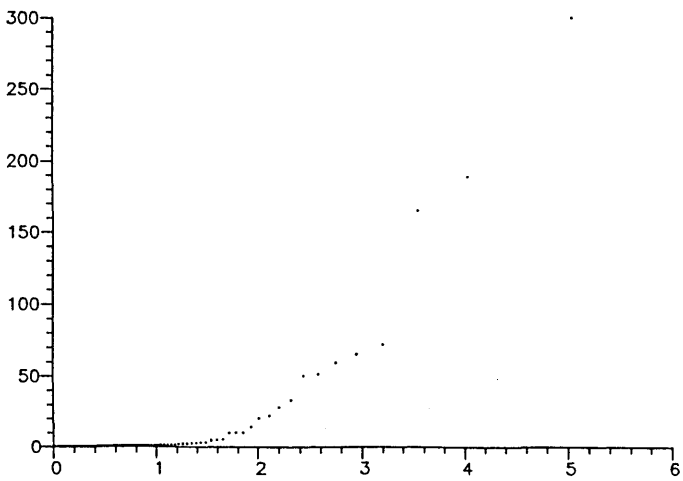



ii) May




Last two points omitted
flows 200 and 324

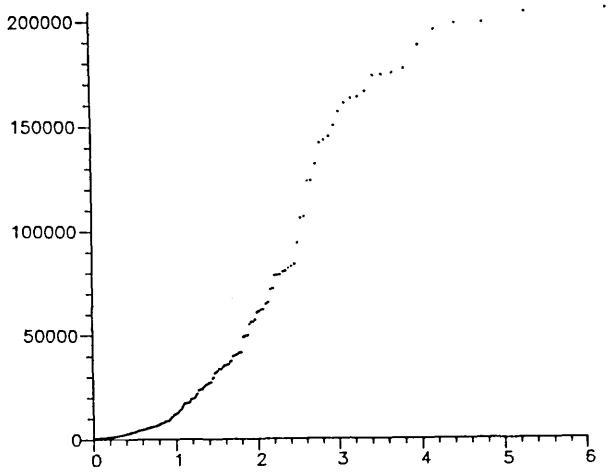
iii) September



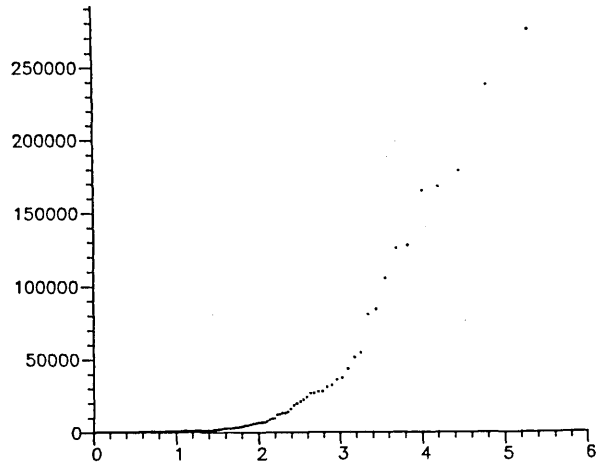

Vertical axes - ordered nonzero flows
Horizontal axes - exponential order statistics

Figure 14 *Fitzroy Exponential probability plots*

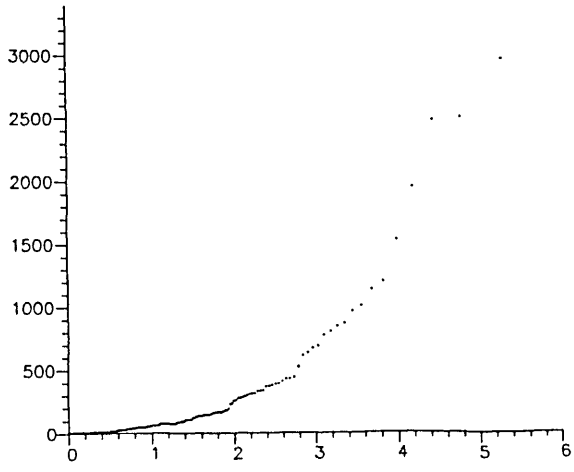
i) February




ii) May




iii) September




Vertical axes - ordered nonzero volumes
Horizontal axes - exponential order statistics

# 3. Storage Models

## 3.1 Introduction.

There is a wide range of possible models to describe the data and to simulate flow records. In the hydrological literature, models often incorporate several catchment and stream channel characteristics, evaporation rates and watertable levels in order to generate streamflow from rainfall. Such models require extensive data so that the various parameters can be estimated; in particular, rainfall as well as runoff data are usually needed to calibrate the system. Analysis of the equations involved is generally not easy. Rainfall data are not available for the rivers under consideration. However, models which might be descriptive of reality are preferable to purely arbitrary mathematical formulations. Storage models are proposed as relatively simple physically motivated models.

Storage models are based on one or more notional reservoirs with various inputs and outflows. The contents of the different reservoirs represent water retained within the catchment in a number of different notional states. Inputs might be rain or result from changes in the

watertable. Loss due to evaporation and streamflow can occur from any of the reservoirs. As the data consist only of flow levels, the inputs will have to be deduced from the flow, not from corresponding rainfall data. A simple formulation of the storage model has independent, identically distributed inputs to a single reservoir, and flow directly proportional to the volume of water stored

Storage models have been fairly widely discussed, particularly in the hydrological literature. Peebles, Smith and Yakowitz (1981) is the one paper using this concept for ephemeral streams. Their interest is in modelling the recession of a flash flood over a time span of a few hours. Differential equations for the flow rate are derived by regarding the stream channel as a reservoir from which water is lost both by outflow and through the streambed at rates which depend on the volume of water in the conceptual reservoir. There are three parameters to be estimated which characterise the particular river; then the differential equations must be solved numerically for given values of the initial storage.

The storage model is a non-negative time series. Properties of non-negative time series with specified marginal distributions are discussed in Gaver and Lewis (1980), Lawrance and Lewis (1980) and

several other papers by Gaver, Jacobs, Lawrance and Lewis. The problems addressed are different from those arising from a physically motivated storage model. Note that transforming the data also makes physical interpretation difficult. The logarithmic transformation is often used in hydrological work, but obviously the zeroes of ephemeral streamflow present problems.

## 3.2 Formulations

### 3.2.1 Introduction

The simplest model has a single reservoir. We consider stationary models before introducing seasonality. The water-balance equation, which accounts for all water entering and leaving the catchment, is

$$dS/dt=\{r-\epsilon-q(S)\}^{+},$$

where r and $\epsilon$ are rainfall and evaporation rates, assumed uniform over time and q is a variable run-off rate which is a function of the catchment storage, S. With the assumption of a linear relation between q and S, Lambert (1972) derives the outflow rate, q(t) and the incremental run-off volume. Lambert uses these purely deterministically. If we let the rainfall be a stochastic process, R(t), $\epsilon=0$, and take $q(S)=qS$, the solution of the differential equation with a fixed starting value, $S(0)=S_0$ is

$$S(t) = \int_0^t e^{-qz} R(t-z)\, dz + S_0 e^{-qt}$$

Letting the starting time tend to the infinite past gives

$$S(t) \simeq \int_0^{\infty} e^{-qz} R(t-z)\, dz$$

The mean of S(t) for a given process R(t) with mean $\mu_R$ can be found:

$$E[S(t)] = \int_0^\infty e^{-qz} \mu_R \, dz = \mu_R/q$$

as can other moments, including the autocorrelation function. However, as flow is recorded at intervals rather than in continuous time, we consider models formulated in discrete time with deterministic and stochastic components.

We set $\epsilon$ to zero for the initial discussion. Let the probability of there being no input on any given day, n, be $\delta$. The loss from the reservoir is the volume of water flowing out, $Q_n$, which is directly proportional to the volume in the reservoir, $S_n$. Inputs, $I_n$, are random variables, independently and identically distributed, and independent of $S_n$. The mathematical formulation is

$$S_{n+1}=S_n - Q_n+I_n .$$

Two cases can be distinguished in discrete time, depending on the order in which the input is added and the flow lost:

I. If input is added at the end of the interval, after the flow is lost, we have

$$Q_n= kS_n \text{ and } S_{n+1} = (1-k)S_n + I_n. \qquad (1)$$

II. If input is added at the beginning of the interval the flow is

$$Q_n = k(S_n + I_n) \text{ and } S_{n+1} = (1-k)S_n + (1-k)I_n \; . \qquad (2)$$

For the storage cases I and II differ only in the scale of the input, the input of II being (1-k) times smaller than the input of I. The relationships between consecutive flows is identical in form in the two cases, as shown below:

I $Q_{n+1} = kS_{n+1} = k\{(1-k)S_n + I_n\} = (1-k)Q_n + kI_n$ ;

II $Q_{n+1} = k(S_{n+1} + I_n) = k\{(1-k)S_n + (1-k)I_n + I_{n+1}\} = (1-k)Q_n + kI_{n+1}$ .

When flow is directly proportional to storage, the properties of the flow will correspond to those of the storage, provided the rescaling of inputs is taken into account.

3.2.2 Deterministic description

In the above formulations the storage, and hence the flow, will never be zero after the first input, as decay is geometric. Modifications are needed to ensure that flow can be exactly zero. There are two main approaches.

For a sequence , $\{S_n\}$, of values from one of these storage models where there are long runs without inputs the values will become very small. It might be reasonable to regard small values as negligible, and

physically difficult to measure. Long runs of constant small values might indicate standing water at the base of the gauge. Regarding values less than $\epsilon$, say, as zero is equivalent to integrating the marginal distribution over the range $(0,\epsilon)$. Thus we have a derived sequence, $\{T_n\}$, given by

$$T_n = \begin{cases} 0 & S_n \leq \epsilon \\ S_n & S_n > \epsilon \end{cases} .$$

The sequence $T_n = (S_n - \epsilon)^+$ is similar, but in this case, $T_n$ takes all values on the range zero to infinity. Let $F_S$ be the distribution function of the marginal (i.e. equilibrium) distribution of $S_h$. We have

$$F_T(t) = \begin{cases} F_S(\epsilon) & t = 0 \\ F_S(t+\epsilon) & t > 0 \end{cases} .$$

Thus we need to find the marginal and conditional distributions of $S_n$.

Alternatively, we can define a model which will lead to an atom of probability at the exact value zero. Including a constant loss over each interval will add a constant to the linear rate of decay of the basic model. Evaporation can be regarded as independent of storage, and clearly only the volume of water present can evaporate. We incorporate this idea, and ensure that negative values for storage do not arise. The modification will again depend on the order in which the inputs are added and the flow and evaporation are lost. This leads to six variations on the model. There is no

physical significance in the ordering, which arises because we are working in discrete time. The order in which these operations are performed is denoted by a triple, for example QEI defines flow subtracted before evaporation is lost, with input added subsequently. The formulae for the storage and flow are:

| | | | |
|---|---|---|---|
| QEI | $Q_n = kS_n$ | $S_{n+1} = \{(1-k)S_n - \epsilon\}^+ + I_n$ | (5) |
| QIE | $Q_n = kS_n$ | $S_{n+1} = \{(1-k)S_n - \epsilon + I_n\}^+$ | (6) |
| EQI | $Q_n = k(S_n - \epsilon)^+$ | $S_{n+1} = (1-k)(S_n - \epsilon)^+ + I_n$ | (7) |
| IQE | $Q_n = k(S_n + I_n)^+$ | $S_{n+1} = \{(1-k)(S_n + I_n) - \epsilon\}^+$ | (8) |
| EIQ | $Q_n = k\{(S_n - \epsilon)^+ + I_n\}$ | $S_{n+1} = (1-k)\{(S_n - \epsilon)^+ + I_n\}$ | (9) |
| IEQ | $Q_n = k(S_n + I_n - \epsilon)^+$ | $S_{n+1} = (1-k)\{S_n + I_n - \epsilon\}^+$ | (10) |

When $S_n$ is large, the pairs of formulations (5) and (6), and (9) and (10) then have identical behaviour. However, these formulations are distinct for given values of the parameters k, $\delta$, $\rho$ and $\epsilon$. Formulations (6) and (10) have higher probability of zero than (5) and (9) respectively; formulation (8) has the highest probability of zero. The relationships between successive flows for (5) and (6) are of the same form as those of the storages:

$$Q_{n+1} = \{(1-k)Q_n - k\epsilon\}^+ + kI_n \quad \text{and}$$

$$Q_{n+1} = \{(1-k)Q_n - k\epsilon + kI_n\}^+$$

respectively; this is not so for the other formulations. Whether there is any

importance in the difference between these two simplest formulations is considered in §3.3.3.

The family of models with a constant loss to evaporation includes zero flow as an intrinsic part of the model, whereas regarding very small readings as standing water or errors which are really zero is perhaps less appealing. The second might be easier analytically and in simulation. Evaporation is known to vary seasonally, independently of the amount of water in the reservoir, but the specification of what is to be neglected would be fixed.

The model can be further elaborated by requiring the volume of storage to reach a certain level v,say, before there is any outflow, i.e.

$$Q_n = k(S_n - v)^+ .$$

This introduces another constant into the formulations; for example, (6) becomes

$$S_{n+1} = (S_n - Q_n - \epsilon)^+ + I_n \qquad = \{S_n - k(S_n - v)^+ - \epsilon\}^+ + I_{n+1} .$$

This has the effect of increasing the probability of flow being zero.

As there are a few extreme flows in the data, a model with a

finite reservoir might be useful. If the input increased the volume of water above its upper limit u, say, all the excess over u would be lost in flow immediately. The formulations I and II differ slightly, I being

$$Q_n = \begin{cases} k(S_n + I_n) & S_n + I_n < u \ , \\ ku + (S_n + I_n - u) & S_n + I_n \geq u \ . \end{cases}$$

This gives a greater rate of decay at peak volume, which is observed in the data.

3.2.3 Stochastic description

The explicit expression to be used for the inputs is

$$I_n = \begin{cases} 0 & \text{with probability } \delta \ , \\ Y_n & \text{with probability } (1-\delta) \ , \end{cases}$$

where $Y_n$ is a continuous positive random variable. The random variables $\{Y_n\}$ are independent and independent of $I_n$ . If we let $Y_n$ be exponentially distributed with mean $\lambda^{-1}$ , then equation (1) is the model studied in Gaver and Lewis (1980), with $\rho = 1-k$. Gaver and Lewis required $X_n$ to have exponential marginal distribution, which is obtained by setting $\rho = \delta$ . This constraint on the parameters has no physical interpretation, and the properties of the model for $\rho \neq \delta$ are required. Other distributions for $Y_n$ might well be used.

Another feature of these models is smoothly decaying runs when there are no inputs. Gaver and Lewis (1980) comment that this implies that parameter estimation is straightforward. However, hydrological data do not have sequences of flow decaying geometrically. This may be due to measurement error or to natural sources of variation. Variation can be added to the model by defining a second sequence which is a function of the original sequence. As the data are constrained to be non-negative, the error structure must retain this feature. This suggests using a multiplicative error, which also preserves zeroes. Thus, given a sequence $\{X_n\}$, for storage or flow, the observed sequence $\{W_n\}$ is given by $W_n=X_nZ_n$, where $\{Z_n\}$ are i.i.d. non-negative random variables centred at one, and independent of $\{X_n\}$. It is convenient to work with the form $W_n=X_n/Z_n$. A suitable distribution for the error must be chosen and both the parameters of the underlying model and those of the error variable need to be estimated.

### 3.2.4 Seasonality

The models must be seasonal to reflect the variation during the year in the rate and size of inputs, and the occurence of dry periods. Any combination of parameters could be made seasonal. We first consider

models with one or two time dependent parameters, and find whether the data can be adequately simulated.

The data suggest using a step function for the probability of no input on day n,

$$\delta(n) = \begin{cases} 1 & B \leq n \bmod 365 \leq E \\ D & \text{otherwise} \end{cases}$$

Thus there is a dry season during which there is no input and a wet season. This implies that the beginning, B, and end, E, of the dry season would be estimated, and the remaining parameters would be derived from the data in the wet season. As a simple step function may not capture the wide variation in length of dry season and volume of flow, further variation could be added by making d or B and E random variables, with B and E having the same or different variances. Clearly, flow could continue beyond B, and would begin some time after E.

The obvious continuous functions to use for periodic parameters are sinusoids. As the parameters are non-negative, use has been made in hydrological literature of exponentials of sinusoids. For example, let

$$\delta(t) = d \exp \{ \alpha \cos(\omega t + \phi) \}$$

where $\omega = 2\pi/365$ if t is measured in days, and the value of $\alpha$ determines the range of $\delta$. This can be approximated by

$$\delta(t) = d \{ 1 + \alpha \cos(\omega t + \phi) \}$$

which might be easier to manipulate. The input size parameter, $\lambda^{-1}$, could be a step function taking the value zero, or continuous; the decay parameter must be positive. The size parameter is the obvious second parameter to make periodic, as rainfall amounts vary over the year, whereas the physical characteristics of the catchment remain broadly similar. Potential evaporation rates are known to exhibit at least some seasonality.

If $\lambda$ were the sole periodic parameter in the model, a step function with $\lambda^{-1}(t)$ zero for some t would also give a dry and a wet season. Having two parameters periodic introduces wide variation in the patterns of flow between years, and simulations suggest that this model will be sufficient to reproduce the main features of the data.

A storage formulation provides a simple model for daily flows, which has a deterministic and a stochastic part. The model is extended in various ways to include nonlinearity so that the process can take the value zero, or have an upper bound. The stochastic element is generalised to

increase the variance of the sequence and to introduce seasonality.

## 3.3 Derivation of properties

This section describes the statistical properties of the models formulated in §3.2. Exact results are given for the simplest form and then various approximations are described. The discussion assumes stationarity.

### 3.3.1 Basic properties and generating functions

The series $\{S_n\}$ for the simplest form, (1), is a Markov process: the distribution of $S_n$ given $S_m=s$, where $m<n$, is clearly independent of any observation prior to $S_m$. The autocorrelation function of $\{S_n\}$ is

$$\rho(S_n,\ S_{n+m}) = (1-k)^m \qquad m= 0, 1, 2, 3, \ldots .$$

The series is not time reversible, i.e. the joint distributions of $\{S_{t(1)},S_{t(2)},\ldots,S_{t(n)}\}$ and $\{S_{N-t(1)}, S_{N-t(2)}, \ldots, S_{N-t(n)}\}$ for any t(i) and N are not identical. The sharp rises and gradual recessions of the data confirm that the empirical process is not time reversible.

The Markov structure suggests using the relation

$$f_{\underline{S}}(s_1, \ldots, s_n \mid s_0) = \prod_{i=1}^{N} f_{S_i}(s_i \mid s_{i-1})$$

to express the likelihood. We let the input be as specified in §3.2.3

$$I_n = \begin{cases} 0 & \text{with probability } \delta \cdot \\ Y_n & \text{with probability } 1-\delta. \end{cases}$$

The probability density function of $S_{n+1}$ given $S_n$ is

$$f_{S_{n+1}}(s \mid S_n=s_n) = \delta\partial(s-\rho s_n) + (1-\delta)\, f_Y(s-\rho s_n)\, I_{(0,\infty)}(s-\rho s_n) ,$$

where $\rho = 1-k$ ; $\rho$ is used in the following when convenient. The range of $S_{n+1}$ is $[\rho s_n, \infty)$ , and $S_{n+1}$ takes the value $\rho s_n$ if and only if there is no input. The indicator function for a set A is denoted $I_A(\cdot)$ ; the Dirac delta function is denoted $\partial(\cdot)$. The likelihood is thus

$$L(\underline{\theta}, \underline{S} \mid s_0) = \prod_{i=1}^{N} \{ \delta\partial(s_i-\rho s_{i-1}) + (1-\delta) f_Y(s_i-\rho s_{i-1})\, I_{(0,\infty)}(s_i-\rho s_{i-1}) \} ,$$

where $\underline{\theta}$ is a parameter vector. This likelihood has many singularities which arise because inputs are recorded as positive values on the real line, but there are days without input, that is, with exactly zero input, and flow is $\rho$ times the flow on the previous day. The function which indicates whether there is an input is dependent on the decay parameter. This form of the likelihood is not useful for statistical analysis, as the log--likelihood is not simple.

An alternative form for the conditional p.d.f. is

$$f_{S_{n+1}}(s \mid S_n = s_n) = \delta^{I_{\{0\}}(s-\rho s_n)} \{(1-\delta) f_Y(s-\rho s_n)\}^{I_{(0,\infty)}(s-\rho s_n)}$$

for $\rho s_n \leq s < \infty$, which leads to the likelihood

$$L(\underline{\theta}, \underline{S} \mid S_0) = \delta^{\Sigma I_{\{0\}}(s_i - \rho s_{i-1})} \times \{(1-\delta) f_Y(s_i - \rho s_{i-1})\}^{\Sigma I_{(0,\infty)}(s_i - \rho s_{i-1})}$$

where the summations in the indices are from i=1 to infinity. The log-likelihood is

$$l(\theta, S \mid s_0) = \{\sum_{i=1}^{n} I_{\{0\}}(s_i - \rho s_{i-1})\} \log \delta + \{\sum_{i=1}^{n} I_{(0,\infty)}(s_i - \rho s_{i-1})\} \times \{\log(1-\delta) + \log f_Y(s_i - \rho s_{i-1})\}$$

If we take $\rho$ fixed, the maximum likelihood estimator of $\delta$ is

$$\hat{\delta} = n^{-1} \sum_{i=1}^{n} I_{\{0\}}(s_i - \rho s_{i-1}) ,$$

the proportion of days with no input. This does not depend on the distributional form of the inputs. The exponential distribution, with p.d.f. $f_Y(y) = \lambda e^{-\lambda y}$, is used in the analysis below, as it is the simplest positive distribution. The maximum likelihood estimator of $\lambda$, with $\rho$ fixed is

$$\hat{\lambda} = m^{-1} \sum_{i=1}^{n} (s_i - \rho s_{i-1}) , \qquad m = \sum_{i=1}^{n} I_{(0,\infty)}(s_i - \rho s_{i-1}) .$$

The maximum likelihood estimator of $\rho$ is exact with probability one; it is the value of $S_i/S_{i-1}$ of which more than one occurs exactly. This suggests finding all such values and averaging them; however, the model gives no

obvious way in which this should be done. It is the mixed nature of the inputs which necessitates the use of indicator functions. This particular simplification of reality leads to an estimator of the correlation which is uninformative.

The moment generating function of $S_n$ can be used to investigate the marginal distribution. We denote the moment generating function by

Note: $\{S\}$ is a linear AR(1) series.

$$M_S(t) = E(e^{tS}) = M_{\rho S}(t)\, M_I(t)$$

$$= M_S(\rho t)\, M_I(t)\,. \qquad (1)$$

Now $M_I(t)=\delta + (1-\delta)M_Y(t)$ and taking Y exponentially distributed with mean $\lambda^{-1}$ we get

$$M_I(t) = (\lambda-\delta t) / (\lambda-t)\,. \qquad (2)$$

Substituting this into (1) gives

$$M_S(t) / M_S(\rho t) = (\lambda-\delta t) / (\lambda-t)\,. \qquad (3)$$

Equating $\rho$ and $\delta$ gives $M_S(t) = \lambda / (\lambda-t)$, i.e. $S_n$ has an exponential marginal distribution when the probability of zero input equals the lag-one correlation : see Gaver and Lewis (1980). As we wish to allow $\rho \neq \delta$, we examine (3) further. Substitution of $\rho t$ for t yields

$$M_S(\rho t) / M_S(\rho^2 t) = (\lambda- \rho\delta t) / (\lambda-\rho t)$$

and hence

$$M_S(t) / M_S(\rho^2 t) = (\lambda-\rho\delta t)\,(\lambda-\delta t) / \{\,(\lambda-\rho t)\,(\lambda-t)\,\}\,.$$

This leads to the finite product

$$M_S(t) / M_S(\rho^n t) = \prod_{i=1}^{n} (\lambda - \delta\rho^{i-1} t) / (\lambda - \rho^{i-1} t) .$$

As k is the proportion of storage which is lost in flow, we have $0 < \rho < 1$; if we let n tend to infinity we have $M_S(0) = 1$ in the denominator, so

$$M_s(t) = \prod_{i=1}^{\infty} (\lambda - \delta\rho^{i-1} t) / (\lambda - \rho^{i-1} t) .$$

The cumulant generating function is

$$K_S(t) = \log M_S(t) = \sum_{i=1}^{\infty} \{ \log(1 - \delta\rho^{i-1} t/\lambda) - \log(1 - \rho^{i-1} t/\lambda) \} .$$

The cumulant generating function is defined for $t < \lambda$ and for $|t| < \lambda$ we can substitute the series expansion for $\log(1+x)$ and change the order of summation to get

$$K_S(t) = \sum_{j=1}^{\infty} (t/\lambda)^j j^{-1} (1-\delta^j) (1-\rho^j)^{-1} . \qquad (4)$$

The cumulants are the coeffients of $t^r/r!$:

$$C_r(S) = (r-1)! (1-\delta^r) / \{\lambda^r (1-\rho^r)\} . \qquad (5)$$

An alternative derivation of the cumulants follows by noting that the cumulants of I are $C_j(I) = (j-1)! (1-\delta^j) \lambda^{-j}$. Substitution of this expression in $C_r(S) = \rho^r C_r(S) + C_r(I)$ yields

$$C_r(S) = C_r(I) / (1-\rho^r) = (r-1)!\,(1-\delta^r) / \{\lambda^r (1-\rho^r)\} .$$

In particular, the mean and variance of S are

$$\mu_s = (1-\delta) / \{\lambda(1-\rho)\} = (1-\delta) / (\lambda k),$$

$$\sigma_s^2 = (1-\delta^2) / \{\lambda^2(1-\rho^2)\} = (1-\delta^2) / [\lambda^2\{1-(1-k)^2\}] .$$

As we would expect, the mean is directly proportional to the size and probability of input, and inversely proportional to the fraction, k, of storage which is lost to flow. The coefficient of variation depends on k and $\delta$ as

$$cv_s = [(1+\delta)(2-k) / \{k(1-\delta)\}]^{1/2} .$$

The alternative formulation in §3.2,

$$S'_{n+1} = (1-k) S'_n + (1-k) I_n ,$$

has cumulants with an additional factor $(1-k)^r=\rho^r$,

$$C_r(S') = (r-1)!\,(1-\delta^r)\,\rho^r / \{\lambda^r (1-\rho^r)\} .$$

As the input random variable is stochastically smaller, the cumulants are decreased; the cumulants for variables standardized to unit mean are the same for both formulations.

The cumulant generating function can, in general, be computed for given parameter values. Numerical techniques could then be used to find the moment generating function and density. However,

investigating the behaviour of the density for a range of parameter values would involve numerical inversion for a large number of points of the density for each combination of parameters.

### 3.3.2 Approximations to the marginal distribution

The series for the cumulant generating function, (4) , does not have a closed form, although it is convergent for $|t| < \lambda$ . We consider the limiting distribution as d=1-δ and k tend to zero in fixed proportion. This corresponds to observing the storage at increasingly frequent intervals, so that the probability of input and the amount of outflow within an interval decrease towards zero. Let $d/k=\mu$ and consider

$$C_r(S) = (r-1)!\{1-(1-d)^r\} / [\ \lambda^r\{1-(1-k)^r\}\ ]$$

as $k \to 0$. De l'Hopital's rule gives $\lim_{k\to 0} C_r(S) = (r-1)!\ \mu/\lambda^r$. Thus the limiting distribution is Gamma $(\mu,\lambda)$, i.e. the probability density function is

$$f(s) = \lambda^{\mu} s^{\mu-1} e^{-\lambda s} / \Gamma(s).$$

This is the marginal distribution of the continuous time shot noise process, with Poisson process of event times, exponential event sizes and constant

decay parameter. The shape parameter, $\mu$, is the ratio of the rate of the Poisson process to the decay parameter. The scale parameter, $\lambda$, is that of the exponential inputs. Weiss (1973) derives the characteristic function of this shot noise process. Brill (1979) derives this marginal distribution for a dam in continuous time, with Poisson arrivals, exponential inputs and release rate proportional to dam levels.

We next consider approximations for fixed intervals between observations. We substitute $\delta = 1 - d$ in the formula for the rth cumulant, (5), and expand the expression:

$$C_r(S) = (r-1)!\{1-(1-d)^r\} / [\lambda^r \{1-(1-k)^r\}]$$

$$= (r-1)!\, rd\,\{1-(r-1)\,d/2 + o(d)\} / [\lambda^r rk\,\{1-(r-1)\,k/2 + o(k)\}]$$

To order d and k the series are those of exponentials in $-(r-1)d/2$ and $-(r-1)k/2$ for the numerator and denominator:

$$C_r(S) \simeq (r-1)!\,\lambda^{-r}\,(d/k)\,e^{-(r-1)d/2}\,e^{(r-1)k/2} \qquad (7)$$

In the data the correlations between observations are high and the proportion of days with input is low, so k and $1-\delta$ are small, and the approximation to the exponential series is reasonable. Rewriting (7) in the form

$$C_r(S) \simeq (r-1)!\,\{\lambda\, e^{\frac{1}{2}(d-k)}\}^{-r}\,(d/k)\,e^{\frac{1}{2}(d-k)} \qquad (8)$$

we see these are the cumulants of a gamma random variable with shape parameter $d/ke^{\frac{1}{2}(d-k)}$ and scale $\lambda e^{\frac{1}{2}(d-k)}$. When d and k are close in value, this is close to exponential, reducing to the exact distribution when d and k are equal. If we substitute $d=k+\epsilon$ ,i.e. $\delta=\rho-\epsilon$, in the expression for the parameters, the shape parameter is $(1+\epsilon/k)e^{\epsilon/2}$ and the scale parameter is $\lambda e^{\epsilon/2}$. The deviation from the exponential for $\epsilon>0$ is an increase in the shape and scale parameters; the mean increases. If the mean is scaled to unity, the higher cumulants decrease exponentially with $\epsilon$.

As we are interested in runs at zero, another approximation to consider is an atom of probability at zero, plus a suitably weighted gamma random variable. The empirical marginal distribution found by simulation is compatible with this, see figure 15. The probability plots, which are based on 300 points, clearly show an atom of probability at zero. The approximating random variable, say X, has the following probability density function:

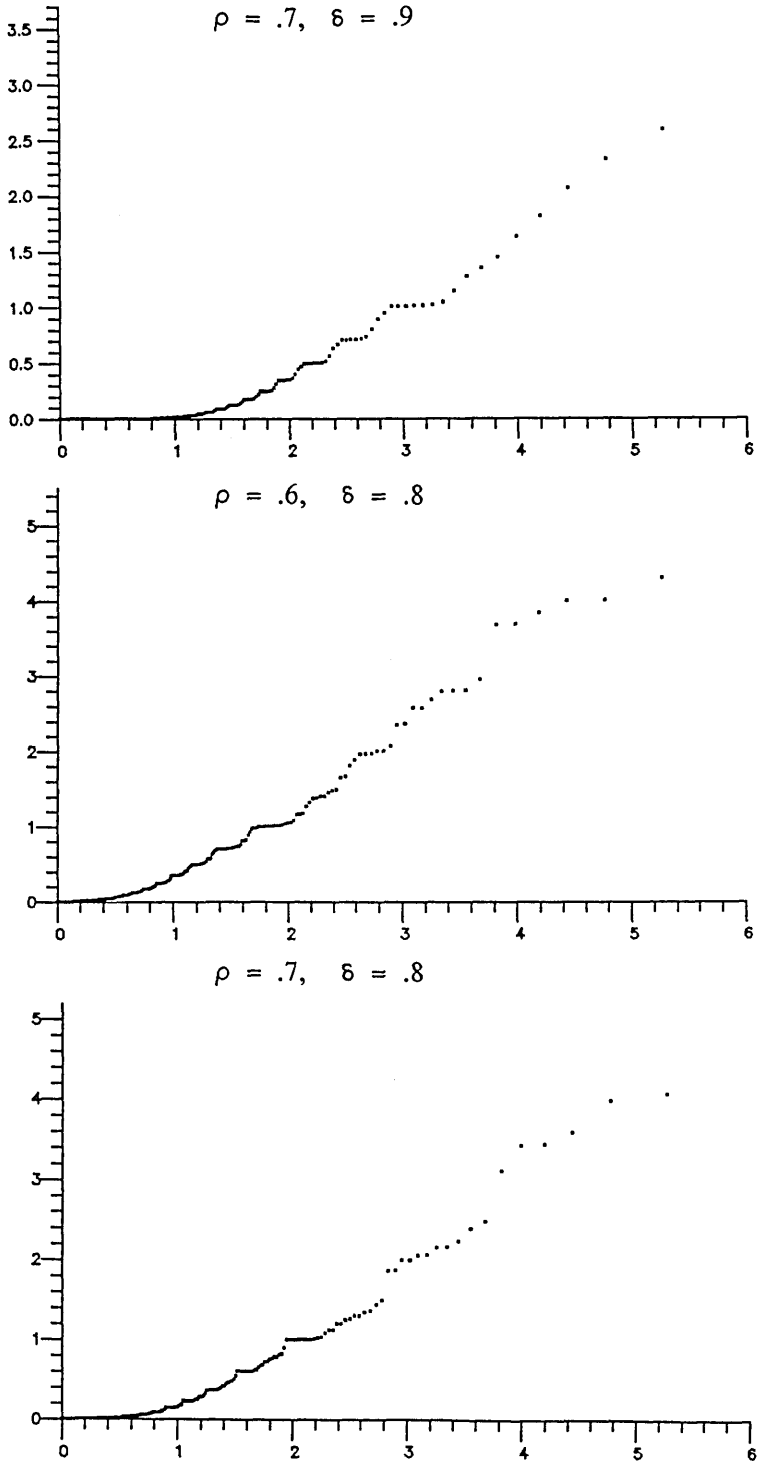
$$g_X(x) = (1-q)\ \delta(x) + q\eta^{\alpha}x^{\alpha-1}e^{-\eta x}\{\Gamma(\alpha)\}^{-1} I_{(0,\infty)}(x)$$

The moment and cumulant generating functions are

$$M_X(t) = (1-q) + q\{\eta(\eta-t)\}^{\alpha} = (1-q)\ [\ 1 + \{q/(1-q)\}\{\eta/(\eta-t)\}^{\alpha}\ ] \qquad \text{and}$$

$$K_X(t) = -\sum_{r=1}^{\infty}(q^r/r)[\ 1 + q(1-q)^{-1}\{\eta/(\eta-t)^{\alpha}\}\ ] + \log .$$

Figure 15 *Probability plots for storage model simulations; λ=1.*




Vertical axes - ordered flows, including zeroes
Horizontal axes - exponential order statistics

This does not yield a concise formula for the cumulants, which must be calculated directly.

There are three parameters to be found. We equate the first three moments of the storage distribution to those of X and find the ratio of the fourth moments to see how close the distributions are. This was done for a wide range of values of the storage parameters. The ratio is constant with respect to $\lambda$ ; $\eta$ is directly proportional to $\lambda$, with the constant of proportion a function of $\delta$ and k. In general the fit was good, the ratio being unity to two decimal places, see table 14. When $\delta$, the probability of zero input, is near one and k, the proportion lost in flow, is small, q is very near one and the ratio is one to three significant digits. Effectively there is then no atom at zero and the approximation is that of the gamma$\{(1-\delta/k),\lambda\}$ of the related shot noise process (6). With k small and $\delta$ not near one, the value of q is one and the approximation reduces to the gamma (8). The ratio deviates from one by up to 6% and with $k \geq .1$, say, q may be greater than one. For this combination of parameter there is no valid approximation of this form. The atom at zero is sizeable when the probability of no input, $\delta$, is near one and the loss in outflow, k, is not small, $k \geq .2$. This accords with intuition ; for this range of values the ratio is within 4% of one. The ratio increases with k, i.e. the difference between

Table 14.

*Comparison of storage model and gamma distribution with an atom at zero.*

| $\delta$ | k | $\lambda$ | q | $\alpha$ | $\eta$ | Ratio of 4th moments |
|---|---|---|---|---|---|---|
| .92 | .02 | .01 | 1.000 | 1.124 | .010 | 1.000 |
| .92 | .10 | .01 | 1.000 | .792 | .010 | 1.000 |
| .98[1] | .02 | .01 | 1.000 | 1.000 | .010 | 1.000 |
| .98 | .10 | .01 | .992 | .194 | .010 | 1.001 |
| .3 | .02 | .01 | 1.000 | 53.115 | .015 | .936 |
| .3 | .10 | .01 | 1.002 | 10.043 | .014 | .939 |
| .5 | .02 | .01 | 1.000 | 32.926 | .013 | .975 |
| .5 | .10 | .01 | 1.002 | 6.262 | .013 | .977 |
| .92 | .2 | .01 | .987 | .382 | .009 | 1.002 |
| .92 | .8 | .01 | .205 | .402 | .008 | 1.041 |
| .98 | .2 | .01 | .924 | .099 | .009 | 1.003 |
| .98 | .8 | .01 | .056 | .361 | .008 | 1.042 |

1. Exponential.

k and $\delta$ decreases. For fixed k, decreasing $\delta$ decreases the ratio very slightly.

This approximation shows that the model $(S_n-\epsilon)^+$ in §3.2 will have the required atom at zero for suitable $\epsilon$ and appropriate subspace of the parameter space. As mentioned in §3.2, $\epsilon$ might be interpreted as the level

below which the gauge does not record. The properties of this random variable will depend on the conditional distribution of S given that S is greater than $\epsilon$. As the distribution of S is not known explicitly, this conditional distribution will have to be based on an approximation.

It was shown that the limiting distribution as k and 1-6 tend to zero with fixed ratio is gamma{(1-6)/k,$\lambda$}. To examine the departure from this distribution, we write the cumulant generating function of equation (4) as

$$K_S(t) = \sum_{r=1}^{\infty} (t/\lambda)^r [\, \mu + \{1-(1-\mu k)^r\}/\{1-(1-k)^r\} - \mu \,]/r \ .$$

Expansion of the binomial terms leads to

$$K_S(t) \simeq \log(1-t/\lambda)^{-\mu} + \sum_{r=1}^{\infty} (t/\lambda)^r \tfrac{1}{2}(r-1)\,\mu(1-\mu)k/r$$
$$= \{-\mu + \tfrac{1}{2}\mu(1-\mu)k\} \ln(1-t/\lambda) + \tfrac{1}{2}\mu(1-\mu)kt/(t-\lambda)$$

The first term is that of a gamma{$\mu-\frac{1}{2}\mu(1-\mu)k,\lambda$}. The second term is not recognisable as the cumulant generating function of a known distribution. However, if we exponentiate to get the moment generating function and let $a=\frac{1}{2}\mu(1-\mu)$, we get

$$M_S(t) \simeq \{\lambda/(\lambda-t)\}^{\mu-ak} e^{akt/(\lambda-t)} \qquad (9)$$

It is convenient for the next few lines to rewrite (9) as a Laplace transform,

$$f^*(s) = \{\lambda/(\lambda+s)\}^{\mu-ak} \exp\{-ak+\lambda ak/(\lambda+s)\}$$

$$= L\{e^{-\lambda t} h(t);s\},$$

where

$$h^*(s) = (\lambda/s)^{\mu-ak} \exp(\lambda ak/s).$$

Thus the inverse Laplace transform is

$$f(t)= \lambda^{\mu-ak} e^{-\lambda t} (\lambda ak)^{-\frac{1}{2}\nu} I_\nu\{2(\lambda ak)^{\frac{1}{2}} t^{\frac{1}{2}}\},$$

where $\nu = \mu-\frac{1}{2}\mu(1-\mu)k-1$ and $b=ak\lambda$, see Erdelyi [1954, (§5.5)], where $I_\nu(\cdot)$ is a Bessel function, Abramowitz and Stegun[1964 , (§9.6)]. Note that $\nu>-1$:

$$\mu-\tfrac{1}{2}(1-\mu)k \geq \mu-\tfrac{1}{2}\mu(1-\mu) > 0,$$

because $k\leq 1$.

Alternatively we can approximate the exponential term in (9) by $1+ak\{\lambda/(\lambda-t)-1\}$, which gives

$$M_S(t) \simeq \{\lambda/(\lambda-t)\}^{\mu-ak}(1-ak) + ak\{\lambda/(\lambda-t)\}^{\mu+1-ak}$$

This represents a mixture of two gammas, i.e. a random variable Z such that

$$Z = \begin{cases} W & \text{with probability } 1-ak \\ V & \text{with probability } ak \end{cases}$$

where $W \sim Ga(\mu-ak,\lambda)$ and $V \sim Ga(\mu+1-ak,\lambda)$. We need $\delta > \rho$ to ensure that $a > 0$. The mixture random variable, Z, and S have the same mean. The

variances and third cumulants of Z and S are within 1% for $d \leq k \leq .1$, and within 5% for $k \leq .4$, $\mu > .4$, see table 15. The approximation is good for a

Table 15

*Comparison of storage model and mixture of gamma distributions.*

ak - weight on second term ; $\mu$-ak - index of first gamma

| k | d | $\mu$=d/k | ak | $\mu$-ak | $\sigma_Z^2/\sigma_S^2$ | $\mu_{Z3}/\mu_{S3}$ |
|---|---|---|---|---|---|---|
| .02 | .01 | .500 | .003 | .498 | .999 | .999 |
| .08 | .01 | .125 | .004 | .121 | .998 | .996 |
| .08 | .04 | .500 | .010 | .490 | .999 | .998 |
| .10 | .01 | .100 | .005 | .096 | .998 | .994 |
| .10 | .05 | .500 | .013 | .488 | .998 | .997 |
| .35 | .15 | .429 | .043 | .386 | .977 | .957 |
| .35 | .30 | .857 | .021 | .836 | .994 | .992 |
| .40 | .15 | .375 | .047 | .328 | .968 | .939 |
| .40 | .20 | .500 | .050 | .450 | .973 | .954 |

wide range of d and k; the ratios of second and third cumulants are independent of $\lambda$. The second gamma has a small contribution. Convergence to the limit is not uniform over $0 < \mu \leq 1$.

A natural question to ask is whether this mixture of two gammas can be approximated to order k by a single gamma. Given

$$M(t,k) = (1-t\theta)^{-\mu+ak}\{1-ak+ak/(1-t\theta) + o(k)\}\ , \qquad (10)$$

with $\theta=\lambda^{-1}$, can we find constants b and c such that

$$M(t,k) = \{1-t(\theta + bk )\}^{-\mu+ck}\{1+0(k^n)\} \qquad \text{for } n\geq 1 \quad ? \qquad (11)$$

We equate the logarithms of (10) and (11),

$$ak\ln(1-t\theta) + \ln\{1-ak+ak/(1-t\theta) +o(k)\}$$

$$= ck\,[\ln(1-t\theta) + \ln\{1-b\,t/(1-t\theta)\}] + \ln\{1+o(k)\},$$

to get

$$a\ln(1-t\theta) + at\theta/(1-t\theta) = c\ln(1-t\theta) - cb/(1-t\theta).$$

Thus by setting $b=-\theta$ and $c=a$, we get

$$M(t,k) = \{1-t\theta(1-k)\}^{-\mu+ak}\{1+o(k^2)\}.$$

If in expanding the expression for the cumulant generating function we keep terms in $k^2$, we find the following further approximation to the moment generating function:

$$M'_S(t) = (1-t/\lambda)^{-\mu\{1-\omega(k)\}} \times$$

$$\{1-\chi(k) + \{\chi(k) - \psi(k)\}/(1-t/\lambda) + \psi(k)/(1-t/\lambda)^2 + o(k)\}$$

where $\omega(k) = \{\tfrac{1}{2}k(1-\mu) + k^2(1+3\mu-4\mu^2)/12\}$

$\chi(k) = \tfrac{1}{2}k(1-\mu) + \tfrac{1}{2}k^2\{\mu(1-\mu)\}$

$\psi(k) = k^2(1-3\mu-2\mu^2)/12$

Note $\omega(k) = \chi(k) + \psi(k)$, and that, for the density function of a gamma$(r,\upsilon)$ random variable, $f_G(x;r,\upsilon)$, the identity

$$(\upsilon x/r)\, f_G(x\,;\, r,\upsilon) = f_G(x\,;\, r+1,\upsilon)$$

holds. Thus we can write this approximation to the probablity density function of the storage as

$$f_S(s) \simeq f_G\{s,\mu(1-\omega),\lambda\} \times [1-\chi + (\chi-\psi)\lambda s/\{\mu(1-\omega)\} + \psi\lambda^2 s^2/\{\mu(1-\omega)\{\mu(1-\omega)+1\}\}]$$

where the dependence of $\omega$, $\chi$ and $\psi$ on k is suppressed. This is a polynomial expansion, with the coefficients of the powers of S of the same order in k. This suggests finding an orthogonal polynomial expansion. We need $c_n$ such that

$$f_S(s) = f_G(s;r,\upsilon)\{1 + c_2 L_2(s) + c_3 L_3(s) + \ldots\}$$

where the $L_n(.)$ are the generalized Laguerre polynomials, see Abramowitz and Stegun[1964, (eqn22.2.12)]. We would hope that $f_G(s;r,\upsilon)\{1+c_2L_2(s)+c_3L_3(s)\}$ is an adequate approximation. We find the expansion by equating the cumulants of S with those of orthogonal polynomial expansion. After some algebra we get

$$f_S(s) = f_G(s;\mu,\lambda)[1 + k(1-\mu)/\{(2-k)(1+\mu)\}L_2(s) + k(1-\mu)\{3-k(2-k)(2-\mu)\}/\{(2-k)(1+\mu)(2+\mu)(3-3k+k^2)\}L_3(s)]$$

$$L_2(s) = (\lambda s)^2 - 2(\mu+1)\lambda s + \mu(\mu+1)d$$

$$L_3(s) = -(\lambda s)^3 + 3(\mu+2)(\lambda s)^2 - 3(\mu+1)(\mu+2)\lambda s + \mu(\mu+1)(\mu+2)$$

The coefficient of the third order polynomial is order k; compare Gram-Charlier expansions in which terms decrease to zero irregularly. When $d < k$, i.e. $\rho < \delta$, the cumulants of the appropriately rescaled

distribution are smaller than those of the corresponding gamma, i.e. the tails are lighter.

The marginal distribution of the continuous time limit of the simple storage model is gamma. For the discrete time process, the marginal distribution can be approximated by a gamma distribution, with or without an atom of probability at zero. The approximation can be improved by taking a mixture of gamma or a finite orthogonal polynomial expansion.

### 3.3.3 Results for truncated models

We analyse two of the six formulations of the model which includes a constant loss, viz

$$S_{n+1} = \{(1-k)S_n - \epsilon\}^+ + I_{n+1} \qquad (11)$$

$$S_{n+1} = \{(1-k)S_n - \epsilon + I_{n+1}\}^+ \qquad (12)$$

These are the formulations which result in the same form for the flow as for the storage, whereas for the remaining formulations the relation between successive flows is more complicated than that between successive storages.

The models form Markov chains with a reflecting barrier at zero. However, as the distribution of step-size is state-dependent, the equilibrium distribution cannot be found from standard Wiener-Hopf equations. The distribution of this model is known for $\epsilon = 0$ and $\rho = \delta$. We wish to find how the marginal distribution is perturbed by the inclusion of $\epsilon$, and the two ways of truncating. Bounds can be given for the moments, using relations such as

$$E[S_{n+1}] \geq E[\rho S_n - \epsilon + I_{n+1}]$$

for lower bounds, and setting $\epsilon=0$ to give upper bounds. This leads to the following bounds for the mean and variance:

$$(1-\delta)/\{(1-\rho)\lambda\} - \epsilon/(1-\rho) \leq E[S] \leq (1-\delta)/\{(1-\rho)\lambda\};$$

$$\frac{2}{1-\rho^2}\left[\frac{\rho}{1-\rho}\left\{\frac{(1-\rho)}{\lambda} - \epsilon\right\}^2 + \frac{1-\delta}{\lambda(\lambda-\epsilon)}\right] - \frac{(1-\delta)^2}{\lambda^2(1-\rho)^2} \leq \text{var}(S)$$

$$\leq \frac{1}{(1-\rho)^2}\left[\frac{2(1-\delta)(1-\rho\delta)}{(1+\rho)\lambda^2} - \frac{(1-\delta)^2}{\lambda^2} + 2(1-\delta)\epsilon - \epsilon^2\right].$$

with a more complicated expression for cov $[S_n, S_{n+1}]$.

For formulation (12) we can write down the following equations for the continuous density and the atom of probability at zero.

$$f_{S_{n+1}}(s) = (\delta/\rho)f_{S_n}\{(s+\epsilon)/\rho\} + \Pr(S_n=0)(1-\delta)\lambda e^{-\lambda(s+\epsilon)} +$$

$$(1-\delta)\int_{0^+}^{(s+\epsilon)/\rho} f_{S_n}(u)\lambda e^{-\lambda(s-\rho u+\epsilon)}du \qquad (13a)$$

The first term on the right hand side refers to there being no input, the second to the preceding value being zero and the third term is the convolution of an input and positive $S_n$. Similarly

$$\Pr(S_{n+1}=0) = \delta\Pr(S_n=0) + (1-\delta)\Pr(S_n=0)(1-e^{-\lambda\epsilon}) + \delta\int_0^{\epsilon/\rho} f_{S_n}(u)du +$$

$$(1-\delta)\int_0^{\epsilon/\rho}\{1-e^{-\lambda(\epsilon-\rho u)}\} f_{S_n}(u)du. \qquad (13b)$$

We take as an initial estimate of the marginal distribution an atom at zero and an exponential distribution

$$f_S^{(0)}(s) = (1-p^{(0)})\ \gamma e^{-\gamma s} \qquad s>0$$

$$\Pr(S=0) = p^{(0)} .$$

We iterate using the equations (13); the first iteration of (13b) is

$$p^{(1)} = p^{(0)}\{1-(1-\delta)e^{-\lambda\epsilon}\} + (1-p^{(0)})\{1-e^{-\gamma\epsilon/\rho} + (1-\delta)\gamma(e^{-\lambda\epsilon}-e^{-\gamma\epsilon/\rho})/(\lambda\rho-\gamma)\}.$$

If we let $\theta=\lim_{n\to\infty} p^{(n)}$ we find

$$\theta = \frac{\lambda\rho - \gamma + (1-\delta)\gamma e^{-\lambda\epsilon} + (\delta\gamma - \lambda\rho)e^{-\gamma\epsilon/\rho}}{\lambda\rho - \gamma + (1-\delta)\lambda\rho e^{-\lambda\epsilon} + (\delta\gamma - \lambda\rho)e^{-\gamma\epsilon/\rho}} .$$

The constraint $\theta\leq 1$ implies $\gamma \leq \lambda\rho$. The first iteration of the continuous part of the distribution is

$$f_S^{(1)}(s) = \lambda(1-\delta)\{p^{(0)} - (1-p^{(0)})\gamma e^{\lambda s}/(\lambda\rho - \gamma)\}\ e^{-\lambda(s+\epsilon)} +$$

$$\{\gamma(1-p^{(0)})/\rho\}\ \{\delta + (1-\delta)\lambda\rho e^{\lambda s}/(\lambda\rho - \gamma)\}\ e^{-\gamma(s+\epsilon)/\rho} .$$

Letting $\gamma = \lambda$ or $\gamma = \lambda\rho$ simplifies these expressions ; the mixed form of the distribution ensures that further iterations become unmanageable. We can derive integral equations by letting $\theta = \Pr(S=0)$ without stating an initial estimate. We obtain

$$f_S(s) = (\delta/\rho)\, f_S\{(s+\epsilon)/\rho\} + (1-\delta)\lambda e^{-\lambda(s+\epsilon)}\{\theta + \int_0^{(s+\epsilon)/\rho} f_S(u)e^{\lambda\rho u}du\}$$

$$\theta = (1-\delta)^{-1}\{e^{\lambda\epsilon}\int_{0+}^{\epsilon/\rho} f_S(u)du - (1-\delta)\int_{0+}^{\epsilon/\rho} e^{\lambda\rho u} f_S(u)du\} .$$

The equations for formulation (11) are :

$$f_S(s) = (1-\delta)\lambda e^{-\lambda s} \times$$

$$\{ \theta + \int_0^{\epsilon/\rho} f_S(u)du + \int_{\epsilon/\rho}^{(s+\epsilon)/\rho} f_S(u)\, e^{-\lambda(\epsilon-\rho u)} du\} + (\delta/\rho) f_S\{(s+\epsilon)/\rho\}$$

$$\theta = \{\delta/(1-\delta)\} \int_0^{\epsilon/\rho} f_S(u)du \ .$$

Clearly we shall have to resort to approximations and numerical methods.

We investigate how the probability of $S_n$ being zero is affected by the perturbation $\epsilon$. Assume as a first approximation that the marginal distribution of

$$S_{n+1}^{(0)} = \rho S_n^{(0)} + I_n$$

is exponential with mean $\lambda^{-1}$. Then for

$$S_{n+1}^{(1)} = (\rho S_n^{(0)} - \epsilon + I_n)^+ ,$$

we find

$$\Pr(S_{n+1}^{(1)}=0) = 1 + \{(\rho-\delta)\, e^{-\lambda\epsilon/\rho} - (1-\delta)\, e^{-\lambda\epsilon}\}/(1-\rho) \ .$$

If $\rho=\delta$, which gives an exact exponential ($\lambda$) marginal distribution for $S^{(0)}$, $\Pr(S^{(1)}=0)$ reduces to $1-e^{-\lambda\epsilon}$. This approximation is

$$1 + (\rho-\delta)(1-\lambda\epsilon/\rho)/(1-\rho) - (1-\delta)(1-\lambda\epsilon)/(1-\rho) = \lambda\epsilon\delta/\rho$$

to order $\epsilon$, which is proportional to $\epsilon$. The adequacy of this expression was tested by simulating the model and fitting the proportion of zeroes. The NAG library random number generators were used in simulating $10^4$ observations for $\rho=\delta=.1(.1).9$ , and $\epsilon = .001(.001).01$ , $.01(.01).1$, $.1(.1)1.$ and $\lambda = 1.0$. A NAG routine was used to regress the logarithm of the proportion of zeroes on the logarithm of $\epsilon$ for each of the three runs of ten values of $\epsilon$ at each value of $\rho$. For the smallest ten $\epsilon$ values, the coefficient of log$\epsilon$ is near 1, see table 16. However, the intercept tends to increase with $\rho$,

Table 16.

*Regression of log(proportion of zeroes) on log($\epsilon$) for simulations of*

$S_{n+1} = (\rho S_n - \epsilon + I_n)^+$ $\quad$ $Pr(I_n = 0) = \rho$

log(prop of 0) = $c + \alpha \log(\epsilon)$ $\quad$ $\lambda = 1.0$

| $\epsilon$ | .001 (.001) .01 | | | .01 (.01) .1 | | | .1 (.1) 1.0 | | |
|---|---|---|---|---|---|---|---|---|---|
| $\rho$ | $\alpha$ | c | $e^c$ | $\alpha$ | c | $e^c$ | $\alpha$ | c | $e^c$ |
| .1 | .91 | -.35 | .71 | .97 | -.02 | .98 | .79 | -.36 | .70 |
| .2 | .97 | -.01 | .99 | .94 | .02 | 1.02 | .77 | -.31 | .73 |
| .3 | .93 | .02 | 1.02 | 1.02 | .34 | 1.40 | .74 | -.27 | .76 |
| .4 | 1.04 | .66 | 1.94 | 1.00 | .45 | 1.56 | .71 | -.22 | .80 |
| .5 | .89 | .10 | 1.11 | .91 | .37 | 1.45 | .65 | -.17 | .84 |
| .6 | 1.01 | .94 | 2.57 | .91 | .56 | 1.75 | .58 | -.14 | .87 |
| .7 | .96 | .90 | 2.47 | .94 | .88 | 2.41 | .53 | -.08 | .92 |
| .8 | .92 | 1.22 | 3.38 | .80 | .78 | 2.17 | .41 | -.04 | .96 |
| .9 | 1.05 | 2.57 | 13.07 | .68 | .89 | 2.44 | .25 | .0 | 1.00 |

consistently over a number of repetitions of these simulations. We examine the dependence of the approximation on $\rho$ and $\delta$ in greater detail. The result obtained from the further iteration

$$S^{(2)}_{n+1} = (\rho S^{(1)}_n - \epsilon + I_n)^+$$

is

$$\Pr(S^{(2)}_{n+1}=0) = 1 + (1-\delta)e^{-\lambda\epsilon} - \rho(1-\delta)^2 e^{-2\lambda\epsilon}/(1-\rho)^2 + \{(1-\rho)(1-\rho^2)\}^{-1} \times$$

$$\left\{(\delta-\rho)(\rho^2-\delta)e^{-\lambda\epsilon(\rho+1)/\rho^2} - (1-\delta)(1+\rho+\rho^2)(\delta-\rho)e^{-\lambda\epsilon(\rho+1)/\rho}\right\},$$

which increases with $\rho$ and with $\delta$. Setting $\rho=\delta$ gives

$$\Pr(S^{(2)}_{n+1}= 0) = 1 - (1-\rho)e^{-\lambda\epsilon} - \rho e^{-2\lambda\epsilon} = \lambda\epsilon(1+\rho) + o(\epsilon). \qquad (14)$$

The constant of proportionality increases with $\rho$ ; this gives a reasonable expected number of zeroes for the simulations for small $\rho$ but insufficient for $\rho$ near one, see table 17. A possible explanation for the inadequacy of these appproximations is that the expected numbers of zeroes is small, and the variation is random error of estimation.

In the simulations for the values of $\epsilon$ from .01 to .1 the index of $\epsilon$ tends to decrease and the proportionality constant tends to increase as $\rho$ increases. The constant is near one for the largest run of $\epsilon$ values but the index decreases from .8 with $\rho$. Of course, $\epsilon$ = 1.0 is not small compared

Table 17.

*Observed and expected numbers of zeroes in simulations.*

$S_{n+1} = (\rho S_n - \epsilon + I_n)^+$ ; $\lambda = 1.0$

| | $\epsilon = .001$ 1st approx : 10 | | $\epsilon = .01$ 1st approx: 100 | |
|---|---|---|---|---|
| $\rho$ | Observed | 2nd approx | Observed | 2nd approx |
| .1 | 12 | 11 | 117 | 110 |
| .2 | 12 | 12 | 138 | 120 |
| .3 | 23 | 13 | 122 | 130 |
| .4 | 16 | 14 | 138 | 140 |
| .5 | 28 | 15 | 220 | 150 |
| .6 | 20 | 16 | 289 | 160 |
| .7 | 32 | 17 | 289 | 170 |
| .8 | 57 | 18 | 547 | 180 |
| .9 | 66 | 19 | 1069 | 190 |

Second iteration approximation: $10^4 \times\{1-(1-\rho)e^{-\lambda\epsilon} - \rho e^{-2\lambda\epsilon}\}$

with $\lambda^{-1}=1.0$. Although the expected number of zeroes is larger, the approximation is not good even for $\epsilon = .01(.01).1$.

We consider a more general approximation with a gamma marginal. The truncation collapses a small part of the distribution on to zero

$$\Pr(S_n = 0) \propto \int_0^{\epsilon} y^{\alpha-1}\, dy \propto \epsilon^{\alpha}.$$

The shape parameter, $\alpha$, of a gamma distribution is the inverse of the squared coefficient of variation. The first two cumulants of the storage model determine the value of $\alpha$:

$$\alpha = \{(1-\rho)^2(1-\delta^2)\} / \{(1-\rho^2)(1-\delta)^2\} = \{(1+\rho)(1-\delta)\} / \{(1-\rho)(1+\delta)\}.$$

With $\delta$ fixed, $\alpha$ increases with $\rho$ and the expected number of zeroes decreases. With $\rho$ fixed and $\delta$ increasing the atom at zero increases. The constant of proportionality is $\{\alpha\Gamma(\alpha)\}^{-1}$. Simulating with $\rho \neq \delta$ shows that the approximation,

$$\Pr(S_n = 0) \approx c\epsilon^{\alpha} ,$$

is good for the index, $\alpha$, see table 18. The estimate of the constant of proportion, $c=\{\alpha\Gamma(\alpha)\}^{-1}$, varies between .9 and 15.0 for most simulations. The variation is not systematic and the t-values for the intercepts suggest that the variation is due to fitting from relatively short simulations.

For the formulation (11) the first approximation using an exponential marginal distribution is

$$\Pr(S^{(1)}_{n+1} = 0) = \delta(1 - e^{-\lambda\epsilon/\rho}) .$$

Comparing this with (13) shows that there is no difference to order $\epsilon$:

$$\Pr(S^{(1)}_{n+1} = 0) = \delta\lambda\epsilon/\rho .$$

Table 18.

*Regression of log (proportion of zeroes) on log(ϵ) for simulations of*

$S_{n+1} = (\rho S_n - \epsilon + I_n)^+$ $\quad Pr(I_n = 0) = \delta$

log(prop of 0) = c + α log(ϵ) $\quad \lambda = .5$

a = { (1-δ)(1+ρ) } / { (1+δ)(1-ρ) } , index of ϵ ; α estimated a

$\{\Gamma(\alpha+1)\}^{-1}$ , constant of proportion ; c estimated constant

ratio = c×Γ(α+1)

| | δ = .5 | | | | | δ = .8 | | | |
|---|---|---|---|---|---|---|---|---|---|
| ρ | a | α | c | ratio | ρ | a | α | c | ratio |
| .43 | .836 | .886 | 2.23 | 2.10 | .73 | .712 | .768 | 3.61 | 3.28 |
| .44 | .857 | .900 | 2.07 | 1.96 | .74 | .744 | .817 | 4.06 | 3.73 |
| .45 | .876 | .976 | 2.76 | 2.64 | .75 | .778 | .666 | 1.68 | 1.56 |
| .46 | .901 | .794 | 1.01 | .97 | .76 | .815 | .673 | 1.63 | 1.53 |
| .47 | .925 | .966 | 2.12 | 2.05 | .77 | .855 | .747 | 1.86 | 1.53 |
| .48 | .949 | .931 | 1.83 | 1.79 | .78 | .899 | .961 | 5.26 | 5.06 |
| .49 | .974 | .970 | 1.74 | 1.72 | .79 | .947 | 1.012 | 6.27 | 6.14 |
| .50 | 1.000 | 1.155 | 4.00 | 4.00 | .80 | 1.000 | 1.077 | 6.71 | 6.71 |
| .51 | 1.027 | 1.419 | 14.48 | 14.46 | .81 | 1.058 | .885 | 2.33 | 2.29 |
| .52 | 1.056 | .907 | .93 | .96 | .82 | 1.123 | 1.283 | 13.46 | 14.25 |
| .53 | 1.085 | .939 | 1.10 | 1.14 | .83 | 1.196 | 1.581 | 51.71 | 56.85 |
| .54 | 1.116 | 1.069 | 1.65 | 1.74 | .84 | 1.278 | 1.175 | 5.34 | 6.15 |
| .55 | 1.148 | 1.299 | 4.90 | 5.25 | .85 | 1.370 | 1.751 | 67.90 | 82.74 |
| .56 | 1,182 | 1.752 | 40.87 | 44.60 | | | | | |
| .57 | 1.217 | 1.101 | 1.36 | 1.52 | | | | | |

Simulations of this formulation show broadly similar results, though for ϵ=.1(.1)1. the approximation is completely inadequate, see table 20.

Table 19.

*Observed and expected numbers of zeroes in simulations.*

$S_{n+1} = (\rho S_n - \epsilon)^+ + I_n$ ; $\lambda = 1.0$

| | $\epsilon = .001$ 1st approx : 10 | | $\epsilon = .01$ 1st approx: 100 | |
|---|---|---|---|---|
| $\rho$ | Observed | 2nd approx | Observed | 2nd approx |
| .1 | 16 | 11 | 124 | 110 |
| .2 | 15 | 12 | 139 | 120 |
| .3 | 9 | 13 | 124 | 130 |
| .4 | 21 | 14 | 162 | 140 |
| .5 | 17 | 15 | 213 | 150 |
| .6 | 9 | 16 | 258 | 160 |
| .7 | 28 | 17 | 344 | 170 |
| .8 | 55 | 18 | 492 | 180 |
| .9 | 132 | 19 | 890 | 190 |

Second iteration approximation: $N \times\rho\{1-(1-\rho)e^{-\lambda\epsilon/\rho}-\rho e^{-2\lambda\epsilon/\rho}\}$

One more iteration of (11) gives

$$\Pr(S^{(2)}_{n+1}=0) = \delta\,[\,(\delta-\rho+1)-(1-\rho)e^{-\lambda\epsilon/\rho}-(1-\rho)^{-1}\{\rho(1-\delta)e^{-2\lambda\epsilon/\rho}+(\delta-\rho)e^{-\lambda\epsilon(1+\rho)/\rho^2}\},$$

If $\rho=\delta$, this simplifies to

$$\Pr(S^{(2)}_{n+1}=0) = \rho\,(1-(1-\rho)e^{-\lambda\epsilon/\rho}-\rho e^{-2\lambda\epsilon/\rho}) = \lambda\epsilon(1+\rho)+o(\epsilon),$$

which is the same as (14) to order $\epsilon$. The simulations again show an excess of zeroes for $\rho$ near 1, see table 19.

Table 20.

*Regression of log (proportion of zeroes) on log(ϵ) for simulations of*

$S_{n+1} = (\rho S_n - \epsilon)^+ + I_n$ $\quad$ $Pr(I_n = 0) = \rho$

log(prop of 0) = c + α log(ϵ) $\quad$ λ = 1.0

| ϵ | .001 | (.001) | .01 | .01 | (.01) | .1 | .1 | (.1) | 1.0 |
|---|---|---|---|---|---|---|---|---|---|
| ρ | α | c | $e^c$ | α | c | $e^c$ | α | c | $e^c$ |
| .1 | .85 | -.63 | .53 | .78 | -.89 | .41 | .19 | -2.22 | .11 |
| .2 | .93 | -.09 | .92 | .86 | -.40 | .67 | .31 | -1.52 | .22 |
| .3 | 1.15 | 1.11 | 3.02 | 192 | -.06 | .94 | .40 | -1.14 | .32 |
| .4 | .84 | -.38 | .68 4 | 185 | -.08 | .92 | .45 | -.88 | .42 |
| .5 | 1.13 | 1.31 | 3.69 | .92 | .28 | 1.33 | .45 | -.69 | .50 |
| .6 | 1.43 | 3.04 | 20.88 | .95 | .58 | 1.78 | .44 | -.53 | .59 |
| .7 | .99 | 1.03 | 2.81 | .86 | .57 | 1.77 | .43 | -.38 | .69 |
| .8 | .86 | .85 | 2.34 | .80 | .76 | 2.14 | .36 | -.23 | .79 |
| .9 | .87 | 1.62 | 5.067 | .73 | .99 | 2.70 | .25 | -.10 | 1.91 |

In both formulations of the model with constant loss, introducing non-linearity dramatically alters the marginal distribution in a way which is difficult to describe analytically. In practical terms it appears that trial and error is the best way to determine the value of ϵ which will give a satisfactory number of zeroes, rather than using an approximate analytical expression for the atom at zero.

Finally, we briefly consider the difference between the two

formulations

$$S_{n+1}=(\rho S_n - \epsilon)^+ + I_n \qquad \text{and} \qquad T_{n+1}=(\rho T_n - \epsilon + J_n)^+ .$$

For identical inputs T has a larger atom at zero than S. If $S_n$ is zero, then $S_{n+1}$ is positive if there is an input, whereas for $T_{n+1}$, $J_n$ must be greater than $\epsilon$. Let N be the number of days until there is a positive value, given that storage is zero, i.e.

$$N = \{m : S_{n+m}>0 \ \& \ S_{n+j}= 0 , j=1,\ldots, m-1 \mid S_n = 0 \}$$

N has a geometric distribution, with

$$\Pr(N= n) = (1-q)^{n-1} q , \qquad \text{where} \qquad q = \begin{cases} 1-\delta & \text{for S} \\ (1-\delta)e^{-\lambda\epsilon} & \text{for T} \end{cases}$$

As $e^{-\lambda\epsilon}<1$, the expected value of N, $(1-q)/q$, is greater for T than for S. We consider whether it is possible to define $\{I_n\}$, given $\{J_n\}$, so that the two formulations are identical. An underlying continuous time process would be truncated at zero for S, whereas T allows decline to continue until the next point at which an observation is taken. Thus the difference is relevant only if there can be much decline between two observations. Clearly for values away from zero we require the p.d.f.s of I and J to coincide. The atom at zero of J must be greater than that of I. Thus the continuous parts of J and I cannot have the same type of distribution (exponential or gamma) as the p.d.f.s must cross over near zero but coincide further out. Although the difference between the formulations is

small, the two are distinct, i.e. in principle, given sufficient data in discrete time it would be possible to discriminate between these two models.

### 3.3.4 Comment on error models

In §3.2 a multiplicative error was introduced. The intention is to add variation to the smooth decay and eliminate non-singularities which make estimation awkward. Let $\epsilon=0$ for this discussion. The model $Y_n=S_n e^{Z_n}$ with $Z_n \sim N(0,\tau^2)$ , which is equivalent to $Y_n=S_n Z_n$, $Z_n$ lognormal, and with $S_n$ having a gamma distribution (shape $\alpha$, scale $\lambda$) does not yield a simple transformation for $Y_n$. Neither exact results nor approximations with $Z_n$ having gamma or Weibull distributions are useful. The moments and cross-moments can be calculated as $\{S_n\}$ and $\{Z_n\}$ are assumed to be independent. We can without loss of generality require $E[Z_n] = 1$ , so the mean of $Y_n$ is identical to that of $S_n$. The moments are:

$$E[Y_n] = E[Z_n]\,E[S_n] = \mu_s ,$$

$$\mathrm{Var}[Y_n] = \sigma_s^2 + \mu_s^2\sigma_z^2 ,$$

$$\mathrm{Corr}(Y_n, Y_{n-1}) = \rho\,\sigma_s^2 / \{\sigma_s^2 + \mu_s^2\,\sigma_z^2\} \text{ and}$$

$$E[(Y_n - \mu_y)^3] = \gamma_s\gamma_z + \mu_s\gamma_z(3\sigma_s^2 + \mu_s^2) + \gamma_s(3\sigma_z^2+1) + 6\mu_s\sigma_s^2\sigma_z^2 .$$

where $\gamma_x = E[(X - \mu_x)^3]$ . Clearly the coefficient of variation of Y is

greater than or equal to that of S :

$$cv^2(Y) = cv^2(S) + cv^2(Z) \geq cv^2(S) .$$

A simpler result may be derived if the transformation is taken in the form $Y_n=S_n/Z_n$. For $Z_n$ distributed as a gamma random variable with shape $\beta$ and scale $\eta$, the p.d.f. of Y is

$$f_Y(y) = (\lambda/\eta)^{\alpha}\, \eta^{\alpha+\beta}\, y^{\alpha-1} / \{ B(\alpha,\beta)\, (\lambda y + \eta)^{\alpha+\beta} \} .$$

The variable $V = \beta\lambda Y/\alpha\eta$ has an $F_{2\alpha,\, 2\beta}$ distribution. The mean of Z and $Z^{-1}$ are one when $\beta=\eta$. The mode is near one and the variance small when $\beta$ is large. The density of Y simplifies to

$$f_Y(y) = (\lambda/\beta)^{\alpha}\, y^{\alpha-1} / [\, B(\alpha,\beta)\, \{1+(\lambda y/\beta)\}^{\alpha+\beta} \,]$$

with mean $\alpha\beta/\{\lambda(\beta-1)\}$ and variance $\alpha\beta^2(\alpha+\beta-1) / \{\lambda^2(\beta-1)^2\,(\beta-2)\}$. As $\beta$ increases, $2\lambda Y$ tends in distribution to a $\chi^2_{2\alpha}$ variable. Thus the distribution of Y tends to that of S as $\beta$ increases, as is clear on general grounds.

The Markovian structure is lost when the error is included, and the likelihood is no longer of simple form. If the conditional distribution $F_{Y_{n+1}\,|\,Y_n,Y_{n-1}}(y_{n+1}\,|\,y_n\,,\, y_{n-1})$ were reasonably close to $f_{Y_{n+1}\,|\,Y_n}(y_{n+1}\,|\,y_n)$, the likelihood could be approximated by the product $\Pi_n$ $f_{Y_{n+1}\,|\,Y_n}(y_{n+1}\,|\,y_n)$. In order to use the m.g.f.s to find the conditional

distributions, we must find the joint m.g.f. of $\log S_n$ and $\log S_{n+1}$. However, this m.g.f. does not have a closed form, even for the special case $\rho = \delta$.

$$M_{\log S_n, \log S_{n+1}}(t,r) =$$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [s_{n+1}^{t} s_n^{r} \{\rho\partial(s_{n+1}-\rho s_n) + (1-\rho)\lambda\epsilon^{-\lambda(s_{n+1}-\rho s_n)} I_{(\rho s_n,\infty)}(s_{n+1})]$$

$$\times \lambda\epsilon^{-\lambda s_n} I_{(0,\infty)}(s_n)\, ds_{n+1} ds_n$$

$$= \rho^{t+1}\Gamma(t+r+1)\lambda^{-t-r} + (1-\rho)\int_{-\infty}^{\infty} s_n^{r} e^{-\lambda(1-\rho)s_n}\Gamma(t+1,\rho s_n)\, I_{(0,\infty)}(s_n) ds_n$$

using the notation of Abramowitz and Stegun (1964) for $\Gamma(a,x)$.

The alternative is to transform directly from the density functions, finding first the joint densities $f_{Y_{n+1},Y_n,Y_{n-1}}(y_{n+1},y_n,y_{n-1})$ and $f_{Y_n,Y_{n-1}}(y_n,y_{n-1})$, and then deriving the conditional distributions. The conditional distributions can be rescaled to be independent of $\beta$. The behaviour of the rescaled distribution as $\beta$ tends to infinity would then be examined to see whether the likelihood can be regarded as nearly Markovian.

First note that the joint distribution of $S_n$ and $S_{n-1}$ is, in the special case $\rho=\delta$,

$$f_{S_n,S_{n-1}}(s_n,s_{n-1}) = f_{S_n|S_{n-1}}(s_n|s_{n-1})\, f_{S_{n-1}}(s_{n-1})$$

$$= [\rho\partial(s_n-\rho s_{n-1}) + (1-\rho)\lambda e^{-\lambda(s_n-\rho s_{n-1})} I_{(\rho s_{n-1},\infty)}(s_n)]\lambda e^{-\lambda s_{n-1}} I_{(0,\infty)}(s_n).$$

As $Z_n$ and $Z_{n-1}$ are independent of $S_n$, $S_{n-1}$, the density of the four variables is the product of the denstities of $Z_n$ , $Z_{n-1}$ and the above joint density. This is then transformed to the joint density of $Z_n$ ,$Z_{n-1}$, $Y_n$ and $Y_{n-1}$. The resulting density is again a function of the incomplete gamma. Hence an alternative method must be used to estimate the parameters. The first three empirical moments and the correlation can be equated with their theoretical values to determine the four parameters. This can be done for either the multiplicative error, or the F distributed variable. The multiplicative error uses the exact moments, whereas the variable Y is based on a gamma approximation to S.

### 3.3.5 Fitting seasonal storage models

In §3.3.1 it was shown that, for the case where $\epsilon=0$, the maximum likelihood estimates of $\delta$ and $\theta=\lambda^{-1}$ for given $\rho$ are the observed proportion of days without input and the mean input size respectively. Section 4.3 discusses two approaches to estimating the lag-one correlation of the data. In this section we examine how the value chosen for $\rho$ affects the estimates of $\delta$ and $\theta$, and consider simulations of four seasonal models, with parameters estimated from M2C8.

Two of the seasonal models are based on step functions for $\delta$ and $\theta$; the functions are either zero or positive, see §3.2.4. pg 56 The simpler formulation has fixed end points for $\delta(t)$, which are estimated by the means of the first and last days of each year for which the subsequent days had a greater level of flow. The second formulation had random end points with the same standard deviation and the values of $\delta(t)$ and $\theta(t)$, when not equal to one and zero respectively, also random, with different standard deviations. Normal random variables were used. Table 22 gives estimates of these values, $\hat{\delta}$ and $\hat{\theta}$ for the step functions. The estimated standard deviation from year to year of the mean input rate and size are given. The values of $\hat{\theta}$ decrease as $\rho$ increases, as expected. The

Table 21.

*Harmonic fit of input probability, δ and input size, θ; M2C8.*

*Percentage of variation explained by harmonics.*

| | δ | | | | θ | | | |
|---|---|---|---|---|---|---|---|---|
| ρ | 1st | 2nd | 3rd | 1-3 | 1st | 2nd | 3rd | 1-3 |
| .1 | 62 | 29 | 7 | 97 | 27 | 3 | 0 | 30 |
| .5 | 68 | 23 | 5 | 96 | 29 | 2 | 0 | 32 |
| .6 | 70 | 20 | 5 | 95 | 26 | 2 | 0 | 29 |
| .7 | 75 | 16 | 3 | 94 | 24 | 3 | 0 | 27 |
| .8 | 79 | 12 | 9 | 93 | 24 | 3 | 0 | 27 |
| .9 | 76 | 11 | 1 | 88 | 25 | 4 | 0 | 29 |

*Coefficients of sine and cosine functions.*

| | δ | | | | | θ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| ρ | $\delta_0$ | $\delta_{1,1}$ | $\delta_{1,2}$ | $\delta_{2,1}$ | $\delta_{2,2}$ | $\theta_0$ | $\theta_{1,1}$ | $\theta_{1,2}$ | $\theta_{2,1}$ | $\theta_{2,2}$ |
| .1 | .17 | .28 | .12 | .14 | .15 | .48 | -.31 | .62 | -.17 | -.16 |
| .5 | .21 | .21 | .17 | .15 | .03 | .41 | -.31 | .43 | -.08 | -.13 |
| .6 | .23 | .28 | .20 | .14 | .03 | .37 | -.26 | .42 | -.09 | -.11 |
| .7 | .26 | .26 | .23 | .13 | .03 | .35 | -.22 | .42 | -.12 | -.10 |
| .8 | .31 | .31 | .27 | .11 | .04 | .34 | -.19 | .44 | -.16 | -.09 |
| .9 | .41 | .41 | .27 | .06 | .02 | .36 | -.19 | .50 | -.19 | -.12 |

estimated coefficient of variation is constant at .7 for the range of ρ used, although the mean input size is estimated from differing numbers of inputs. As ρ increases, the estimates $\hat{\delta}$ ~~decrease~~ increase, while ~~as do~~ the estimated standard deviations and cv's decrease. These standard errors are used in generating

the parameters for each year. The other two seasonal models were based on sinusoidally varying input size and probability. The estimates of $\delta$ and $\theta$ were found for each day by considering data over years and harmonic series with three terms were fitted. The first formulation used the first harmonic fit for $\delta(t)$ and $\theta(t)$ and the second used the first two harmonics. The results show that the harmonic fit to $\theta(t)$ is less good than for $\delta(t)$, see table 21 and figure 16. The estimates of $\theta(t)$ is based on few nonzero observations for some t, and fluctuates widely.

Table 22.

*Estimates of $\theta$ and $\delta$ for step function periodic rate and size ; M2C8.*

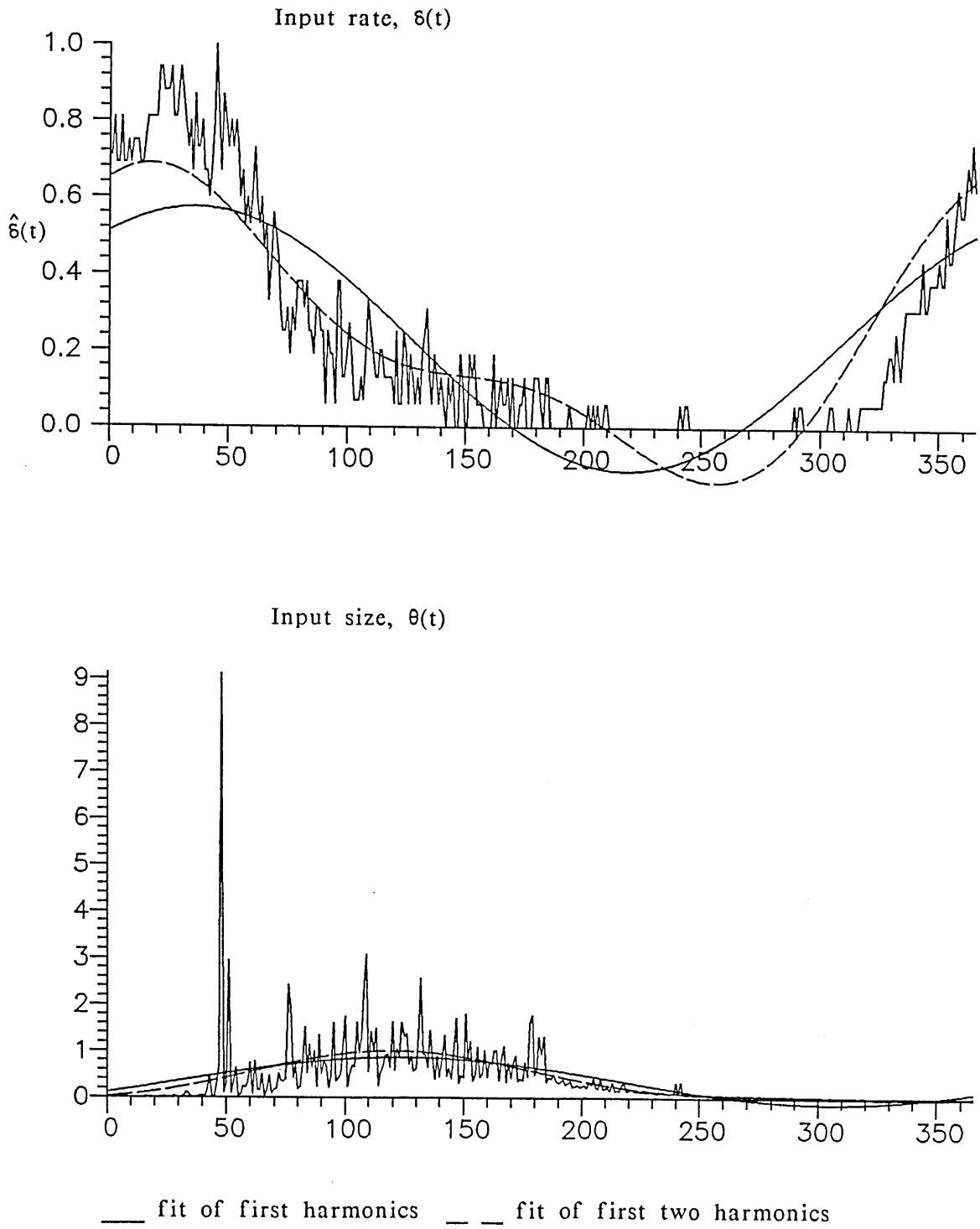
| $\rho$ | $\hat{\theta}$ | $sd(\hat{\theta})$ | $\hat{\delta}$ | $sd(\hat{\delta})$ |
|---|---|---|---|---|
| .1 | .647 | .440 | .041 | .072 |
| .5 | .408 | .276 | .091 | .089 |
| .6 | .362 | .247 | .115 | .087 |
| .7 | .322 | .220 | .149 | .079 |
| .8 | .290 | .197 | .201 | .073 |
| .9 | .287 | .196 | .310 | .057 |

The input parameter, $\delta$ , is estimated by

$$\delta(t) = \begin{cases} \hat{\delta} & t \bmod(365) \in (39,342) \\ 0 & \text{otherwise} \end{cases}$$

For the simulations, the value $\rho = .6$ was chosen, fairly arbitrarily, as near that of the correlation in the shot noise model of §4.3.

Figure 16 *Estimated daily input rate and size; ρ=.6*



The parameter values are as given in tables 21 and 22 ; when $\hat{\delta}(t)$ or $\hat{\theta}(t)$ were negative, they were set to zero, rather than constraining the estimates to be positive. This will, in general, give a greater probabilty of zero flows. The value of .01 was used for $\epsilon$ in the two step function variants, denoted "Step1" and "Step2" and for the sinusoidal variant with one harmonic, "MS1". The other sinusoidal variant, "MS2", had $\epsilon$=.005. These values were chosen by examing some simulations for a range of $\epsilon$, cf. §3.3.3. Annual and monthly statistics, with their standard errors, from ten simulations of seventeen years of data for the four variants are given in tables 23 to 25.

The mean flows for MS1 and MS2 are large and as MS1 has almost twice the number of zeroes, the conditional mean is even further from the historical value than that of MS2. MS2 has greater cv and skewness than MS1, as expected, but still far smaller than that of the data. The means are better preserved by Step1 and Step2. There are slightly too few zeroes in Step1 ; Step2 has a wide variation in the percentage of zeroes, which includes the historical value. The cv and skewness of Step1 are small ; it is surprising that Step2 does not have larger cv and skewness, though the standard errors of these are much larger than for the other models. The monthly cvs are far smaller than the historical values for Step1, MS1 and

Table 23.

*Results of ten simulations for M2C8.*

| | | | Conditional | | | |
|---|---|---|---|---|---|---|
| | Mean daily flow | % of flow at 0 | Mean daily flow | Stan dev. | Coef var | Skew |
| 2C8 | .63 | 16.7 | .76 | 1.70 | 2.25 | 5.88 |
| MS 1 | .704 | 26.28 | .955 | .911 | .955 | 1.634 |
| s.d. | .024 | .31 | .029 | .032 | .014 | .083 |
| MS 2 | .723 | 16.80 | .873 | 1.060 | 1.214 | 1.914 |
| s.d | .015 | .42 | .021 | .027 | .013 | .091 |
| Step1 | .642 | 15.44 | .759 | .453 | .599 | 1.296 |
| s.d. | .006 | .11 | .007 | .009 | .007 | .067 |
| Step2 | .634 | 23.79 | .834 | .731 | .880 | 2.044 |
| s.d. | .097 | 6.83 | .105 | .080 | .058 | .499 |
| Err1 | .789 | 26.21 | 1.065 | 1.132 | 1.064 | 2.435 |
| s.d. | .021 | .24 | .031 | .050 | .025 | .235 |
| Err2 | .884 | 26.14 | 1.196 | 1.484 | 1.240 | 3.578 |
| s.d. | .012 | .35 | .018 | .038 | .029 | .505 |

MS2, but nearer these values for Step2. Both sinusoidal variants reproduce the pattern of increase and decrease in observed conditional monthly flows. The step variants do not show this because the values of $\hat{\delta}$ and $\hat{\theta}$ are constant in any given year.

Table 24.

*Coefficients of variation for positive flow.*

| | Oct | Nov | Dec | Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2C8 | .8 | 1.5 | .8 | .7 | .8 | .9 | 1.2 | 1.3 | 1.0 | 1.1 | 1.1 | 1.0 |
| MS1 | .33 | .31 | .28 | .26 | .24 | .16 | .20 | .21 | .26 | .0 | .0 | .71 |
| s.d. | .05 | .07 | .06 | .05 | .05 | .05 | .0 | .03 | .03 | - | - | .09 |
| MS2 | .43 | .30 | .27 | .21 | .19 | .21 | .20 | .19 | .21 | .20 | .43 | 1.28 |
| s.d. | .09 | .05 | .05 | .03 | .06 | .03 | .0 | .03 | .03 | .07 | .08 | .32 |
| Step1 | .0 | .24 | .21 | .19 | .20 | .20 | .20 | .20 | .21 | .21 | .20 | .32 |
| s.d. | - | .05 | .03 | .06 | .05 | .0 | .0 | .05 | .03 | .03 | .0 | .06 |
| Step2 | .81 | .66 | .66 | .63 | .63 | .63 | .63 | .67 | .61 | .63 | .66 | .70 |
| s.d. | .16 | .19 | .11 | .08 | .08 | .07 | .11 | .12 | .06 | .08 | .10 | .14 |
| Err1 | .37 | .37 | .30 | .25 | .22 | .21 | .21 | .20 | .29 | .0 | .0 | .50 |
| s.d. | .08 | .07 | .0 | .05 | .04 | .06 | .03 | .07 | .06 | - | - | .07 |
| Err2 | .35 | .34 | .33 | .31 | .20 | .22 | .23 | .23. | 31 | .0 | .0 | .51 |
| s.d. | .09 | .05 | .07 | .06 | .05 | .04 | .05 | .05 | .06 | - | - | .14 |

Two sets of simulations of MS1 with flows divided by an error were performed, see §3.3.4. The intention was to increase the variance and skewness of the data while preserving the seasonal structure. The first set had a gamma(10,10) error and is denoted "Err1". The mode of the gamma probability function is at .9 and the flows are increased by 11%. The cv and skewness for the annual data and the monthly cvs are slightly larger than those of MS1, but all are still less than the historical values. The next

Table 25(a).

*Conditional mean daily flow, for each month.*

| | Oct | Nov | Dec | Jan | Feb | Mar |
|---|---|---|---|---|---|---|
| 2C8 | .009 | .444 | .876 | 1.591 | 1.853 | 1.597 |
| [illegible] 1 | .251 | .526 | .909 | 1.320 | 1.703 | 1.766 |
| s.d. | .029 | .040 | .088 | .098 | .780 | .043 |
| [illegible]2 | .081 | .322 | .993 | 1.972 | 2.262 | 1.726 |
| s.d. | .008 | .043 | .048 | .101 | .760 | .062 |
| Step1 | .0 | .724 | .781 | .771 | .783 | .782 |
| s.d. | - | .049 | .052 | .045 | .028 | .034 |
| Step2 | .676 | .787 | .849 | .821 | . 864 | .821 |
| s.d. | .270 | .134 | .105 | .087 | .116 | .101 |
| Err1 | .308 | .632 | 1.003 | 1.515 | 1.867 | 1.985 |
| s.d. | .023 | .060 | .088 | .100 | .134 | .101 |
| Err2 | .311 | .627 | 1.105 | 1.718 | 2.245 | 2.206 |
| s.d. | .037 | .058 | .058 | .081 | .089 | .083 |

set, with a gamma(5,5) error, "Err2", has flows inflated by 25%. The cv and skewness of the nonzero daily flows are the largest of all six models, but the historical value is larger by more than three standard errors. The monthly cvs are still considerably lower than the data: there is too little variation in conditional monthly mean flows from year to year in the simulations. A gamma(β,β-1), with the mode of its density at one, might be a better choice of error variable if a small β is needed to give a widely

Table 25(b).

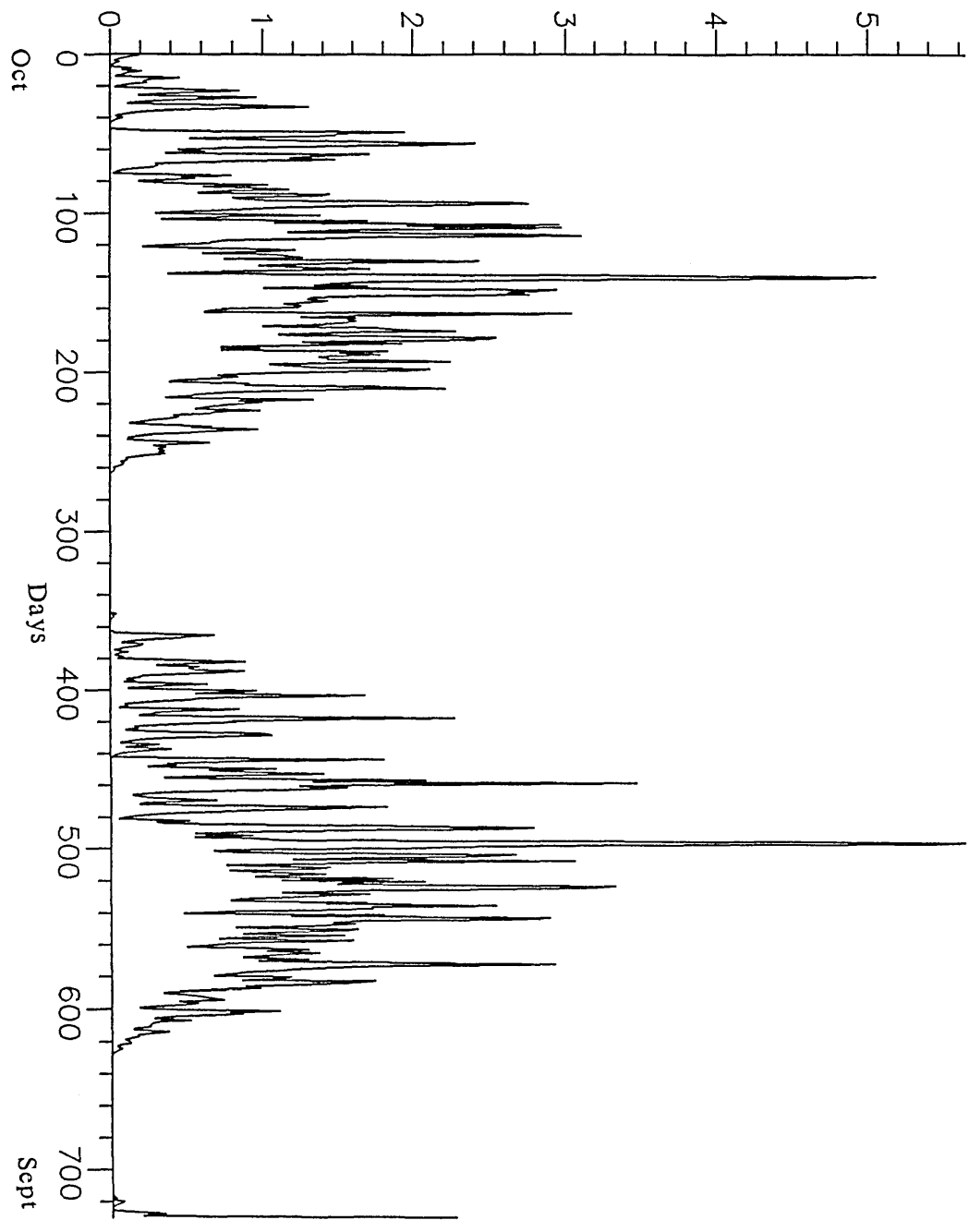
*Conditional mean daily flow, for each month.*

| | Apr | May | June | July | Aug | Sep |
|---|---|---|---|---|---|---|
| 2C8 | .901 | .391 | .165 | .098 | .047 | .024 |
| MS1 | 1.291 | .660 | .167 | .020 | .003 | .086 |
| s.d. | .070 | .038 | .012 | - | - | .020 |
| MS2 | .907 | .386 | .138 | .064 | .018 | .034 |
| s.d. | .032 | .013 | .004 | .003 | .001 | .016 |
| Step1 | .771 | .761 | .771 | .765 | .778 | .492 |
| s.d. | .035 | .037 | . 035 | .038 | .017 | .042 |
| Step2 | .825 | .831 | .850 | .847 | .676 | .578 |
| s.d. | .107 | .095 | .119 | .111 | .127 | .193 |
| Err1 | 1.395 | .724 | .175 | .0 | .0 | .080 |
| s.d. | .066 | .041 | .011 | - | - | .006 |
| Err2 | 1.586 | .803 | .198 | .0 | .0 | .092 |
| s.d. | .108 | .055 | .016 | - | - | .012 |

spread distribution. Multiplying by a gamma(β,β) error would reduce the mean flows to nearer the historical values. Figure 17 shows two years from one simulation of MS1. The seasonal pattern is preserved. However, the graph has at least twice as many peaks as figure 1, and decline in flow at the end of the wet season is steeper than in figure 1.

Step1, with five parameters, is clearly inadequate. The choice

Figure 17 *M2C8 Storage model simulation - MS1 ; daily flow*



between MS1, with seven parameters, and Step2, with eight parameters depends on the relative value placed on the variablity. MS2 preserves the monthly flows well, but has eleven parameters. MS1 is probably the most useful model; the estimation procedure needs to be improved. Including an error, and therefore another, abitrarily chosen, parameter in the MS1 formulation allows the variability to be increased while maintaining the particular pattern of seasonality. The shot noise model with sinusoidal periodic function, with seven parameters, preserves the pattern of monthly flows, and has greater variation. However, it has a high proportion of zeroes. The estimation and simulation for MS1 are simpler than that for the shot noise model, which is an important practical consideration.

A further series of simulations was done with $\rho$=.8 for M2C8. Ten seventeen year long sequences of daily flow were generated from the models MS1, MS2, Step1 and Step2, with the parameters as given in tables 20 and 21. The statistics from these simulations are presented in tables 26 to 28. The means of all flows for MS1 and MS2 are further from the historical values than for $\rho$=.6. Those of Step1 and Step2 are less and greater than, respectively, the observed means, whereas the previous values were the same as the observed. The proportion of zeroes is similar

Table 26.

*Annual statistics of ten simulations of storage models for M2C8; ρ=.8*

| | | | Conditional | | | |
|---|---|---|---|---|---|---|
| | Mean daily flow | % of flow at 0 | Mean daily flow | Stan dev. | Coef var | Skew |
| Obs. | .63 | 16.7 | .76 | 1.70 | 2.25 | 5.88 |
| MS1 | 1.070 | 26.37 | 1.451 | 1.270 | .874 | 1.107 |
| s.e. | .023 | .32 | .029 | .037 | .017 | .082 |
| MS2 | 1.044 | 19.81 | 1.301 | 1.427 | 1.097 | 1.546 |
| s.e. | .020 | .39 | .022 | .034 | .014 | .087 |
| Step1 | .458 | 15.66 | .545 | .360 | .663 | 1.391 |
| s.e. | .009 | .09 | .011 | .008 | .009 | .069 |
| Step2 | .933 | 22.04 | 1.202 | .894 | .749 | 1.398 |
| s.e. | .125 | 5.71 | .134 | .087 | .084 | .354 |

Table 27.

*Coefficients of variation for positive flow; M2C8, ρ=.8*

| | Oct | Nov | Dec | Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Obs. | .8 | 1.5 | .8 | .7 | .8 | .9 | 1.2 | 1.3 | 1.0 | 1.1 | 1.1 | 1.0 |
| MS1 | .43 | .44 | .36 | .28 | .24 | .18 | .16 | .19 | .23 | .0 | .0 | .52 |
| s.e. | .07 | .10 | .10 | .04 | .05 | .04 | .05 | .06 | .07 | - | - | .06 |
| MS2 | .65 | .52 | .32 | .23 | .22 | .24 | .20 | .19 | .21 | .22 | .42 | .75 |
| s.e. | .14 | .10 | .06 | .05 | .04 | .05 | .0 | .06 | .03 | .06 | .08 | .17 |
| Step1 | .0 | .29 | .22 | .23 | .25 | .20 | .23 | .22 | .23 | .22 | .23 | .36 |
| s.e. | - | .06 | .04 | .05 | .05 | .0 | .07 | .04 | .05 | .06 | .05 | .07 |
| Step2 | 1.08 | .68 | .59 | .60 | .58 | .62 | .60 | .59 | .59 | .59 | .67 | .83 |
| s.e. | .20 | .15 | .14 | .11 | .11 | .15 | .12 | .12 | .10 | .09 | .17 | .16 |

for the different parameter values for MS1, MS2 and Step2 ; consequently the means of the positive flows are large. The cvs and skewnesses are reduced. The proportion of zeroes for Step1 is as in the data, giving a

Table 28(a)

*Conditional mean daily flow for each month; M2C8, ρ=.8*

| | Oct | Nov | Dec | Jan | Feb | Mar |
|---|---|---|---|---|---|---|
| Obs. | .009 | .444 | .876 | 1.591 | 1.853 | 1.597 |
| MS1 | .381 | .673 | 1.158 | 1.967 | 2.709 | 2.940 |
| s.e. | .041 | .045 | .067 | .112 | .099 | .112 |
| MS2 | .111 | .518 | 1.539 | 2.938 | 3.196 | 2.460 |
| s.e. | .013 | .068 | .125 | .231 | .112 | .080 |
| Step1 | .0 | .536 | .548 | .538 | .551 | .571 |
| s.e. | - | .034 | .026 | .032 | .036 | .034 |
| Step2 | .550 | 1.038 | 1.201 | 1.262 | 1.235 | 1.214 |
| s.e. | .175 | .269 | .157 | .146 | .182 | .132 |

| | Apr | May | June | July | Aug | Sep |
|---|---|---|---|---|---|---|
| Obs. | .901 | .391 | .165 | .098 | .047 | .024 |
| MS1 | 2.093 | .918 | .153 | .0 | .0 | .123 |
| s.e. | .101 | .019 | .009 | - | - | .018 |
| MS2 | 1.159 | .405 | .236 | .197 | .045 | .129 |
| s.e. | .054 | .405 | .011 | .008 | .004 | .062 |
| Step1 | .547 | .548 | .556 | .557 | .562 | .376 |
| s.e. | .021 | .018 | .0185 | .022 | .033 | .033 |
| Step2 | 1.279 | 1.222 | 1.264 | 1.213 | 1.121 | .757 |
| s.e. | .150 | .192 | .171 | .177 | .111 | .158 |

small conditional mean flow. The cvs and skewnesses are similar for the two parameterisations. The conditional monthly mean flows reflect the mean daily flows in whether they are greater or less than the observed. The monthly cvs of MS2, Step1 and the first six months of MS1 are generally slightly increased, and somewhat decreased for Step2 and the last six months of MS1. However, the estimated values for one parameterisation are within three standard errors of the other, but not of the observed values for the four variants. The standard errors of the statistics do not change with the change in parameters, nor does the basic seasonal structure. The values and dispersion of the flows are more satisfactory for $\rho$=.6 than for $\rho$=.8. The smaller correlation seems to be, in some sense, a better reflection of the data.

The two sinusoidal models and the step function model with Normal random variables for the endpoints and input parameter values were fitted to AR3, an ephemeral stream, taking $\rho$ to be .5. The estimates were:

MS1 : $\delta(t) = .75 + .210 \cos\psi t - .069 \sin\psi t$

$\theta(t) = 18.69 - .166 \cos\psi t + 13.777 \sin\psi t$

MS2 : $\delta(t) = .75 + .210 \cos\psi t - .069 \sin\psi t - .510 \cos 2\psi t + .016 \sin 2\psi t$

$\theta(t) = 18.69 - .166 \cos\psi t + 13.777 \sin\psi t - 1.522 \cos 2\psi t - 11.03 \sin 2\psi t$

| Step2: | B | E | δ | θ |
|---|---|---|---|---|
| Mean | 59.0 | 321.5 | .656 | 22.60 |
| Standard deviation | 52.3 | 52.3 | .296 | 24.68 |

The first series of simulations were performed with $\epsilon$=.01, the value used for the intermittent streams simulations. The results, given in tables 29 to 31, from these simulations are indicated by the suffix "a" , e.g. Table 29.

*Annual statistics of ten simulations of AR3*

| | | | Conditional | | | |
|---|---|---|---|---|---|---|
| | Mean daily flow | % of flow at 0 | Mean daily flow | Stan dev. | Coef var. | Skew |
| Observed | 9.37 | 76.70 | 33.73 | 107.07 | 3.17 | 5.37 |
| MS1a $\epsilon$=.01 | 10.446 | 21.378 | 13.291 | 21.161 | 1.590 | 3.324 |
| s.e. | .397 | 1.033 | .564 | .839 | .404 | .287 |
| MS2a $\epsilon$=.01 | 10.572 | 29.399 | 14.974 | 25.203 | 1.683 | 3.415 |
| s.e. | .328 | 1.264 | .382 | 1.196 | .056 | .325 |
| Step2a $\epsilon$=.01 | 7.490 | 72.802 | 26.980 | 67.665 | 2.563 | 5.950 |
| s.e. | 3.658 | 10.297 | 5.078 | 12.092 | .494 | 2.142 |
| MS1b $\epsilon$=6.0 | 6.169 | 71.315 | 21.503 | 24.444 | 1.137 | 2.438 |
| s.e. | .179 | .431 | .391 | .943 | .032 | .237 |
| MS2b $\epsilon$=6.0 | 6.743 | 74.360 | 26.298 | 30.894 | 1.175 | 2.611 |
| s.e. | .231 | .619 | .705 | .810 | .022 | .314 |
| Step2b $\epsilon$=.015 | 8.100 | 72.350 | 29.443 | 77.440 | 2.628 | 6.364 |
| s.e. | 1.561 | 3.889 | 5.161 | 16.710 | .299 | 1.189 |

MS1a. The mean of all daily flows for MS1a andMS2a are larger than the observed values. The proportions of zero flows are small - 20% and 30% - and the statistics for nonzero flow are less than the historical statistics. The monthly conditional means are bimodal, but the lower peak in MS1a

Table 30(a).

*Coefficients of variation for positive flow for AR3 simulations.*

| | Oct | Nov | Dec | Jan | Feb | Mar |
|---|---|---|---|---|---|---|
| Observed | 1.6 | 1.2 | 1.4 | 1.7 | 1.5 | 1.7 |
| MS1a $\epsilon$=.01 | .85 | .60 | .47 | .37 | .33 | .36 |
| s.e. | .18 | .08 | .09 | .05 | .05 | .07 |
| MS2a $\epsilon$=.01 | .70 | .56 | .44 | .37 | .32 | .32 |
| s.e. | .07 | .07 | .11 | .05 | .06 | .06 |
| Step2a $\epsilon$=.01 | .89 | .58 | .73 | .74 | .74 | .83 |
| s.e. | .52 | .27 | .33 | .47 | .46 | .54 |
| MS1b $\epsilon$=6.0 | .68 | .45 | .38 | .33 | .32 | .30 |
| s.e. | .23 | .05 | .09 | .05 | .06 | .05 |
| MS2b $\epsilon$=6.0 | .58 | .47 | .44 | .34 | .30 | .33 |
| s.e. | .14 | .07 | .08 | .08 | .05 | .08 |
| Step2b $\epsilon$=.015 | 1.43 | 1.03 | 1.08 | 1.24 | 1.28 | 1.21 |
| s.e. | .51 | .47 | .34 | .51 | .34 | .23 |

occurs in September, not August. The minimum for MS2a is in May, not June, and the dip in October in the sequence of means is an inadequate reflection of the data. The variability of these means is small. In order to

increase the proportion of zeroes in the sinusoidal models to roughly the same as in the data, a value of 6.0 was needed for $\epsilon$. The suffix "b" indicates where $\epsilon$=6.0 was used. The simulations with this $\epsilon$ have small mean flows; the mean, cv, skewness and monthly statistics for positive flows are small. MS1b does not pick up the peak in August. MS2b performs somewhat better than MS2a in the pattern of monthly means. The monthly conditional mean flows have slightly greater standard errors with the larger $\epsilon$; as the flows are truncated at zero, the larger flows are

Table 30(b).

*Coefficients of variation for positive flow for AR3 simulations.*

| | Apr | May | June | July | Aug | Sep |
|---|---|---|---|---|---|---|
| Observed | 1.7 | 2.1 | 1.1 | 1.3 | 1.0 | 1.1 |
| MS1a $\epsilon$=.01 | .35 | .36 | .41 | .57 | .84 | 1.18 |
| s.e. | .08 | .05 | .06 | .05 | .18 | .37 |
| MS2a $\epsilon$=.01 | .32 | .63 | 1.05 | .70 | .77 | .81 |
| s.e. | .06 | .13 | .42 | .18 | .28 | .17 |
| Step2a $\epsilon$=.01 | .70 | .85 | .85 | .76 | .99 | 1.14 |
| s.e. | .39 | .59 | .54 | .25 | .32 | .45 |
| MS1b $\epsilon$=6.0 | .34 | .38 | .51 | .62 | .61 | .66 |
| s.e. | .10 | .08 | .10 | .10 | .11 | .14 |
| MS2b $\epsilon$=6.0 | .43 | .0 | .91 | .68 | .78 | .60 |
| s.e. | .05 | - | .29 | .20 | .24 | .09 |
| Step2b $\epsilon$=.015 | .97 | 1.02 | 1.05 | .99 | 1.00 | .95 |
| s.e. | .32 | .29 | .24 | .23 | .19 | .39 |

Table 31(a)

*Conditional mean daily flow, for each month for AR3 simulations.*

| | Oct | Nov | Dec | Jan | Feb | Mar |
|---|---|---|---|---|---|---|
| Observed | 14.361 | 35.595 | 41.426 | 49.498 | 54.376 | 33.295 |
| MS1a $\epsilon$=.01 | 6.363 | 10.598 | 18.444 | 23.898 | 25.535 | 21.048 |
| s.e. | 1.138 | 1.549 | 1.728 | 1.780 | 1.400 | 1.192 |
| MS2a $\epsilon$=.01 | 4.977 | 6.453 | 13.457 | 26.542 | 36.362 | 28.472 |
| s.e. | .528 | .900 | 1.148 | 2.359 | 2.958 | 2.713 |
| Step2a $\epsilon$=.01 | 17.995 | 45.998 | 30.767 | 23.591 | 26.879 | 28.323 |
| s.e. | 16.227 | 35.126 | 12.783 | 16.517 | 8.077 | 11.745 |
| MS1b $\epsilon$=6.0 | 15.344 | 21.246 | 24.715 | 26.396 | 25.594 | 21.996 |
| s.e. | 2.178 | 1.848 | 2.471 | 1.468 | 1.122 | 1.759 |
| MS2b $\epsilon$=6.0 | 10.574 | 13.544 | 21.540 | 33.126 | 36.603 | 26.692 |
| s.e. | 1.338 | .918 | 1.844 | 2.409 | 2.459 | 1.680 |
| Step2b $\epsilon$=.015 | 19.546 | 31.670 | 29.158 | 30.541 | 30.568 | 28.767 |
| s.e. | 15.417 | 19.756 | 12.523 | 11.586 | 12.502 | 9.050 |

altered more than the smaller.

The annual statistics for Step2a are reasonably near the historical values. The standard errors are large; the seventeen year sequences differ considerably. Figure 18 illustrates two years of one run of Step2a. The difference in input rate and size from one year to the next is obvious. The number of peaks is several times greater than in figure 2. The monthly

Table 31(b).

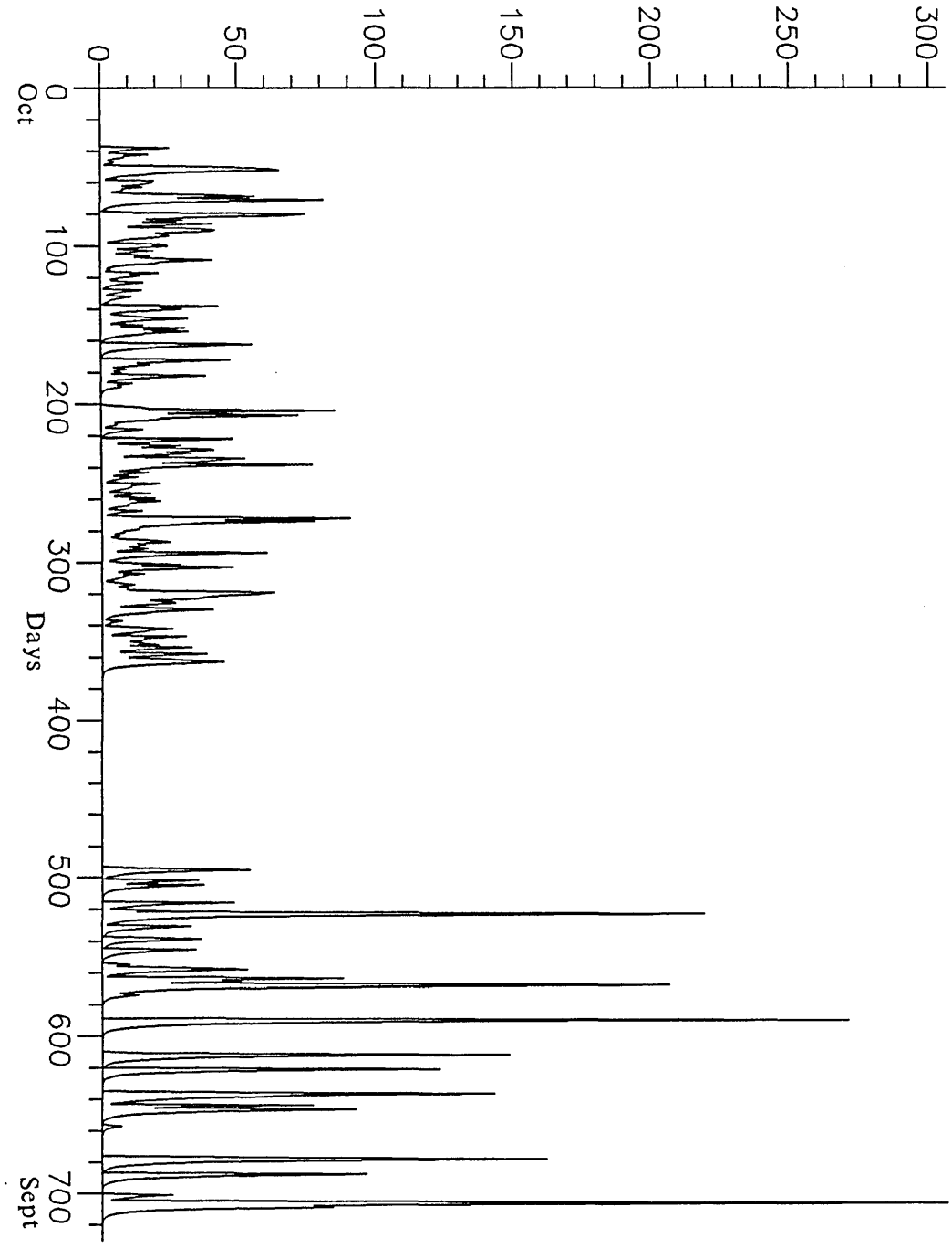
*Conditional mean daily flow, for each month for AR3 simulations.*

| | April | May | June | July | Aug | Sep |
|---|---|---|---|---|---|---|
| Observed | 12.416 | 3.216 | .779 | 12.525 | 26.180 | 15.848 |
| MS1a $\epsilon$=.01 | 6.785 | 3.008 | 1.944 | 2.014 | 3.284 | 21.048 |
| s.e. | 1.077 | .485 | .431 | .214 | .499 | 1.064 |
| MS2a $\epsilon$=.01 | 9.102 | .593 | 1.438 | 3.659 | 5.389 | 5.574 |
| s.e. | .611 | .083 | .282 | .494 | 1.415 | 1.008 |
| Step2a $\epsilon$=.01 | 23.683 | 24.634 | 23.795 | 21.673 | 21.808 | 16.314 |
| s.e. | 8.274 | 8.682 | 10.494 | 7.823 | 9.121 | 9.657 |
| MS1b $\epsilon$=6.0 | 14.116 | 8.856 | 5.202 | 4.767 | 6.613 | 10.137 |
| s.e. | .923 | .696 | .492 | .845 | 1.251 | 1.912 |
| MS2b $\epsilon$=6.0 | 10.927 | .118 | 3.631 | 9.554 | 14.200 | 13.012 |
| s.e. | 1.359 | .262 | 1.488 | 1.956 | 2.521 | 2.392 |
| Step2b $\epsilon$=.015 | 27.741 | 25.103 | 20.058 | 20.485 | 25.808 | 23.370 |
| s.e. | 6.318 | 6.957 | 5.173 | 6.589 | 8.542 | 9.623 |
| Snp step2 | 73.991 | 90.190 | 65.128 | 69.102 | 48.738 | 47.539 |
| s.e. | 60.010 | 50.948 | 46.213 | 44.943 | 16.768 | 37.514 |

means of nonzero flow do not, of course, reproduce the turning points of the data. The cvs are roughly half those of the data. Increasing $\epsilon$ from .01 to .015 left the annual statistics and conditional monthly means more or less unchanged; the variation over years of monthly flows increased towards the observed level. The standard errors decreased for the monthly statistics and most of the annual statistics. The monthly means have high

standard errors, ¼ to ¾ of their value. The standard errors for all the statistics are considerably larger than for the sinusoidal models. As ephemeral stream flow is very variable, this feature of the step function model is desirable; it is the more useful model.

Figure 18 *AR3 Storage model simulation - Step2; daily flow*



# 4 Shot Noise Processes

## 4.1 Introduction

It was shown in §3.3 that the limit of the discrete storage model as the interval between observations tends to zero is a shot noise process. Weiss (1973) discusses the properties of a particular shot noise process and its application in modelling perennial daily streamflow series. In order to use a shot noise model to represent intermittent streams, periodic variation is introduced to give dry seasons. The process must be aggregated to apply to daily readings. The physical interpretation of the shot noise process is similar to that of the non-negative time series models already discussed; however, the results of simulations differ, see §3.3.5 and §4.3.

## 4.2 Periodic Shot Noise Processes

The shot noise process of interest is defined by

$$X(t) = \sum_{m=N(0)}^{N(t)} Y_m e^{-b(t - \tau_m)} , \qquad (1)$$

where N(t) is a Poisson process with event rate $\nu$, b>0 is a decay rate and $Y_m$, associated with $\tau_m$, are independent and exponentially distributed

with mean $\theta$. The lower limit of the summation could be finite or infinite. The mean and variance are $\nu\theta b^{-1}$ and $\nu\theta^2 b^{-1}$; the correlation of X(t) and X(t+s) is exp(-bs). Weiss (1973) also gives the mean, variance and serial correlation for $\nu$ and $\theta$ varying sinusoidally with time. The serial correlation is given when b(t) is also periodic. The characteristic function of X(t) for the general case is

$$\phi(u;t) = \exp\left[\int_{-\infty}^{t} \nu(\tau)\left\{\frac{1}{1 - \theta(\tau)\, ui\exp\{-B(\tau,t)\}} - 1\right\}d\tau\right], \qquad (2)$$

where $B(\tau,t) = \int_{\tau}^{t} b(\sigma)\, d\sigma$, see Weiss (1973, 4.38). Weiss suggested a method for estimating $\nu(t)$ and $\theta(t)$ when expressed as a finite trigonometric series. The results of attempting to implement this estimation are discussed in §4.3. Weiss fitted a shot noise process with parameters estimated separately for each month; this has the disadvantage of introducing transient biases near the begining of each month.

We consider the behaviour of the marginal distribution of X(t) when the decay rate and input size are constant and the event rate is periodic:

$$\nu(t) = \alpha + \beta \cos(\psi t) \ , \qquad (3)$$

where $\psi = 2\pi/365$ to give a period of one year. The cumulant generating

function is

$$K(u;t) = \int_{-\infty}^{t} [\theta u e^{-b(t-s)} / \{1 - \theta u e^{-b(t-s)}\}] \nu(s) ds \qquad (4)$$

from (2). Substituting $\nu(t)$ and $t-s=\tau$ gives

$$K(u;t) = \int_{0}^{\infty} \frac{\theta u \exp(-b\tau)\, \nu(t-\tau)\, d\tau}{1 - \theta u \exp(-b\tau)}$$

$$= \frac{-\alpha}{b} \ln(1 - u\theta) + \int_{0}^{\infty} \frac{\beta \theta u \exp(-b\tau) \cos\{\psi(t-\tau)\}\, d\tau}{1 - \theta u \exp(-b\tau)} . \qquad (5)$$

This is defined for $u<1/\theta$, so $\theta u e^{-b\tau}/(1-\theta u e^{-b\tau})$ can be expanded as $\Sigma_{k=1}^{\infty}(\theta u)^k e^{-bk\tau}$. Noting that

$$\int_{0}^{\infty} e^{-ax} \cos(bx + c)\, dx = \frac{a \cos c - b \sin c}{a^2 + b^2} ,$$

(Gradsteyn and Ryshik,1965, eqn3.893), we find the second term in (5):

$$\int_{0}^{\infty} \sum_{k=1}^{\infty} (u\theta)^k e^{-kb\tau} \cos(\psi t - \psi\tau)\, d\tau$$

$$= \sum_{k=1}^{\infty} \frac{(u\theta)^k \cos\phi_k \cos(\psi t - \phi_k)}{kb} ,$$

where $\phi_k = \arctan(\psi/kb)$. This gives

$$K(u;t) = \sum_{k=1}^{\infty} (u\ \theta)^k \{\alpha + \beta \cos \phi_k \cos(\psi t - \phi_k)\}/(bk) \quad , \qquad (6)$$

and hence the cumulants are

$$K_r(t) = (r-1)!\ \theta^r \{\alpha + \beta \cos \phi_r \cos(\psi t - \phi_r)\} / b \ .$$

These are the sum of the cumulants of the stationary marginal distribution and a term due to the periodic form of the $\nu(t)$ The phase lag is expected as the system is linear and the wavelength is preserved. The lag decreases as the order of the cumulants increases because arctan is monotonic on $(0,\pi/4)$. The mean varies slightly behind the variance.

Standard trigonometric formulations are used to rewrite (6) as

$$K(u:t) = \sum_{k=1}^{\infty} (u\theta)^k \left[ \frac{\alpha}{bk} + \frac{bk\beta \cos \psi t}{b^2k^2 + \psi^2} + \frac{\beta\psi \sin \psi}{b^2k^2 + \psi^2} \right] .$$

Summing over the terms in $(u\theta)^k k^{-1}$ gives

$$K(u;t) = -b^{-1} (\alpha + \beta \cos \psi t) \ln(1-u\theta) + \beta\psi \sum_{k=1}^{\infty} \{(u\theta)^k \sin\psi t / (k^2 b^2 + \psi^2)\} \qquad (7)$$

$$- \beta\psi^2 b^{-1} \sum_{k=1}^{\infty} \{(u\theta)^k k^{-1} \cos\psi t / (k^2 b^2 + \psi^2)\} .$$

The first term is that of a gamma distribution with seasonally varying shape parameter $(\alpha+\beta\cos\psi t)/b$ and constant scale parameter, $\theta^{-1}$.This has mean $\theta(\alpha+\beta\cos\psi t) / b$. The second and third terms are convergent, but no

closed form was found. To gain some idea of the way in which the marginal distribution differs from that of a gamma, the ratio of the third cumulant of the seasonal shot noise process to that of a gamma with the same mean and variance was calculated for a range of values of the parameters and times. The range of parameter values covers the ranges of correlation and frequency of input events in the data. The ratio is independent of the scale parameter $\theta^{-1}$. The cumulants of the seasonal shot noise process are

$$K_1(t) = (\theta/b)\{\alpha + \beta\cos(\phi_1)\cos(\psi t - \phi_1)\}$$

$$K_2(t) = (\theta^2/b)\{\alpha + \beta\cos(\phi_2)\cos(\psi t - \phi_2)\}$$

$$K_3(t) = (2\theta^3/b)\{\alpha + \beta\cos(\phi_3)\cos(\psi t - \phi_3)\}$$

The third cumulant of the fitted gamma is $2K_2(t) / K_1(t)$ and hence the ratio is

$$R(t) = \frac{\{\alpha + \beta\cos\phi_1\cos(\psi t - \phi_1)\}\{\alpha + \beta\cos\phi_3\cos(\psi t - \phi_3)\}}{\{\alpha + \beta\cos\phi_2\cos(\psi t - \phi_2)\}^2} .$$

Dependence of this ratio on b is through the lags, $\phi_k$ , and on $\alpha$ and $\beta$ is through their ratio $\alpha/\beta$ , the ratio of the constant to the periodic components of the rate. The third moments of the seasonal shot noise process differ from those of the fitted gammas only by a few percent when $\alpha/\beta \geq 2$ . The scale parameters are also approximately equal and therefore the index of the fitted gamma varies with the mean of the shot noise

process. The coefficient of variation and skew vary inversely with the mean, see Table 31(a) and (b). Setting $\alpha = \beta$ gives the widest range of values for $\nu(t)$ , including $\nu(t) = 0$. As the mean of the seasonal shot noise

Table 31(a)

*Values of the mean, cv and skew of a seasonal shot noise process, and the ratio of the third cumulants of the shot noise process and a gamma with same first cumulants.*

*b - decay parameter* $\nu(t) = \alpha + \beta cos(\psi t)$ *- event rate*

| b = .51 , correlation = .6 | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $\alpha$ = .1 , $\beta$ = .05 | | | | | $\alpha$ = .2 , $\beta$ = .1 | | | |
| Mean | CV | Skew | Ratio | Day | Mean | CV | Skew | Ratio |
| .29 | 1.84 | 3.69 | 1.00 | 0 | .59 | 1.30 | 2.61 | 1.00 |
| .25 | 2.00 | 4.02 | 1.00 | 61 | .50 | 1.42 | 2.84 | 1.00 |
| .15 | 2.57 | 5.18 | 1.01 | 122 | .30 | 1.82 | 3.66 | 1.01 |
| .10 | 3.19 | 6.39 | 1.00 | 183 | .20 | 2.26 | 4.52 | 1.00 |
| .15 | 2.64 | 5.24 | .99 | 244 | .29 | 1.86 | 3.71 | .99 |
| .24 | 2.03 | 4.05 | 1.00 | 305 | .49 | 1.44 | 2.86 | 1.00 |

| b = .11 , correlation = .9 | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $\alpha$ = .1 , $\beta$ = .05 | | | | | $\alpha$ = .2 , $\beta$ = .1 | | | |
| Mean | CV | Skew | Ratio | Day | Mean | CV | Skew | Ratio |
| 1.35 | .86 | 1.72 | 1.00 | 0 | 2.71 | .61 | 1.21 | 1.00 |
| 1.19 | .91 | 1.84 | 1.01 | 61 | 2.38 | .64 | 1.30 | 1.01 |
| .74 | 1.13 | 2.33 | 1.03 | 122 | 1.49 | .80 | 1.65 | 1.03 |
| .46 | 1.45 | 2.95 | 1.01 | 183 | .93 | 1.03 | 2.09 | 1.01 |
| .63 | 1.28 | 2.50 | .97 | 244 | 1.26 | .91 | 1.77 | .97 |
| 1.08 | .98 | 1.92 | .98 | 305 | 2.15 | .69 | 1.36 | .98 |

Table 31(b).

$\nu(t) = \alpha \{1 + \cos(\psi t)\}$ - *event rate*

| $\alpha$ = .01 , b = .51 | | | | | | |
|---|---|---|---|---|---|---|
| Day | 100× Mean | CV | Skew | Ratio | 100× Index | Scale |
| 0 | 3.92 | 5.05 | 10.10 | 1.00 | 3.92 | 1.00 |
| 61 | 2.99 | 5.75 | 11.58 | 1.01 | 3.02 | 1.01 |
| 122 | 1.03 | 9.72 | 19.81 | 1.02 | 1.06 | 1.03 |
| 183 | .00 | 107.23 | 406.12 | 1.89 | .01 | 5.01 |
| 244 | .94 | 10.45 | 20.49 | .98 | .92 | .97 |
| 305 | 2.91 | 5.89 | 11.71 | .99 | 2.88 | .99 |

| $\alpha$ = .20 , b = .51 | | | | | | |
|---|---|---|---|---|---|---|
| Day | 100× Mean | CV | Skew | Ratio | 100× Index | Scale |
| 0 | .782 | 1.13 | 2.26 | 1.00 | .78 | 1.00 |
| 61 | .60 | 1.29 | 2.59 | 1.01 | .60 | 1.01 |
| 122 | .21 | 2.17 | 4.43 | 1.02 | .21 | 1.03 |
| 183 | .00 | 23.98 | 90.81 | 1.89 | .00 | 5.01 |
| 244 | .19 | 2.34 | 4.58 | .98 | .18 | .97 |
| 305 | .58 | 1.32 | 2.62 | .99 | .58 | .99 |

process decreases to zero, the third moment increases to about three times that of the gamma. The skewness of the fitted gamma also increases as the mean and index decrease. The shot noise process changes rapidly to being less skew than the gamma, as the mean increases from zero. The ratio is .9 at its minimum. In contrast, the skewnesses estimated from the historical

daily flow data are large when there is flow, and near or at zero for the dry periods.

Table 31(c).

$\nu(t) = \alpha \{1 + cos(\psi t)\}$

| $\alpha$ = .01 , b = .11 | | | | | | |
|---|---|---|---|---|---|---|
| Day | 100× Mean | CV | Skew | Ratio | 100× Index | Scale |
| 0 | 17.96 | 2.37 | 4.71 | .99 | 17.80 | .99 |
| 61 | 14.71 | 2.56 | 5.24 | 1.02 | 15.24 | 1.04 |
| 122 | 5.81 | 3.90 | 8.45 | 1.08 | 6.57 | 1.13 |
| 183 | .21 | 10.82 | 37.94 | 1.75 | .85 | 4.15 |
| 244 | 3.53 | 5.70 | 10.41 | .91 | 3.08 | .87 |
| 305 | 12.44 | 2.91 | 5.61 | .96 | 11.81 | .95 |

| $\alpha$ = .20 , b = .11 | | | | | | |
|---|---|---|---|---|---|---|
| Day | 100× Mean | CV | Skew | Ratio | 100× Index | Scale |
| 0 | 3.59 | .54 | 1.05 | .99 | 3.56 | .99 |
| 61 | 2.74 | .57 | 1.17 | 1.02 | 3.05 | 1.04 |
| 122 | 1.16 | .87 | 1.89 | 1.08 | 1.31 | 1.13 |
| 183 | .04 | 2.42 | 8.48 | 1.75 | .17 | 4.15 |
| 244 | .71 | 1.27 | 2.33 | .91 | .62 | .87 |
| 305 | 2.49 | .65 | 1.26 | .96 | 2.36 | .95 |

We consider whether this seasonal shot noise process might be useful by examining the limiting behaviour of a gamma random variable, Z, say, with mean 1 and index $\beta$, i.e. variance $\beta^{-1}$. The variance and higher

moments of Z tend to infinity as $\beta$ tends to zero. We can see how rapidly the distribution tends to concentrate at zero by finding how $\epsilon$ decreases with $\beta$ to maintain a fixed area under the p.d.f, $f_Z(z)$, between zero and $\epsilon$ as $\beta$ tends to zero. Define $V_p(\beta)$ by

$$p = \int_0^{V_p(\beta)} f_Z(z)dz \ ,$$

If we let $u = \beta z$ and assume that $\beta V_p(\beta) \to 0$ as $\beta \to 0$, then expand $e^{-u}$ as $1+o(u)$ we get

$$p \simeq \{\beta \Gamma(\beta)\}^{-1} \{\beta V_p(\beta)\}^{\beta} + o(\beta^{\beta+1}).$$

Letting $\beta \to 0$ gives $p \simeq \{V_p(\beta)\}^{\beta}$ and hence

$$V_p(\beta) \simeq p^{1/\beta}$$

The assumption made to find this result is valid, as $\beta p^{1/\beta} \to 0$ as $\beta \to 0$. Thus, for small $\beta$, $\Pr(Z < p^{1/\beta}) = p$; $V_p(\beta) = (kp)^{1/\beta}$, $k > 0$ also gives a fixed value for $F_Z(V_p(\beta))$. For a fixed $\epsilon$, $\Pr(X < \epsilon)$ tends to one as $\beta$ tends to zero. Therefore in simulating, the large theoretical skewness for small values of the rate will not result in unrealistic simulations, as the simulated data will almost certainly be zero. The skewnesses for the larger values of the mean are of the same order as the historical values.

## 4.3 Fitting seasonal models

The methods and results of fitting three different seasonal shot noise processes are described in this section. The procedure suggested in Weiss (1973) for fitting with the rate parameter , $\nu$ (t) , and mean input, $\theta$ (t), assumed to be periodic with sinusoidal form is as follows:

1. Estimate the mean, m(t), and standard deviation ,$\sigma$(t) .

2. Standardize the data using m(t) and $\sigma$(t).

3. Estimate $\hat{\rho}$, the first serial correlation coefficient from the standardized data.

4. Find the estimate $\hat{b}$ from $\hat{\rho} = \frac{1}{2}(1-e^{-b}) / \{1-(1-e^{-b})\}$.

5. Calculate the phase lags, $\psi_k$ , $\eta_k$ .

6. Solve the equations

$$m(t) = \sum_{k=0}^{2N} c_k b^{-1} \cos\chi_k \cos(k\phi t - \psi_k - \chi_k) \qquad \text{and} \qquad (7a)$$

$$\sigma^2(t) = \sum_{k=0}^{3N} d_k b^{-1} \cos\zeta_k \cos(k\phi t - \eta_k - \zeta_k), \qquad (7b)$$

where

$$\chi_k = \arctan(k\phi/b), \quad \zeta_k = \arctan(\tfrac{1}{2}k\phi/b),$$

Weiss [1973, eqn.(4.45), with a minor correction], to find the coefficients and lags of the products $\nu(t)\theta(t)$ and $\nu(t)\theta^2(t)$, the mean & variance, respectively.

7. Solve the equations

$$\nu(t)\theta(t)= \sum_{k=0}^{2N} c_k \cos(k\phi t - \psi_k) \quad \text{and} \tag{8a}$$

$$\nu(t)\theta^2(t)= \sum_{k=0}^{3N} d_k \cos(k\phi t - \eta_k), \tag{8b}$$

Weiss [1973, eqn.(4.43)] to find $\nu(t)$ and $\theta(t)$.

The standard deviation must be divided by the factor $2\{b-(1-e^{-b})\}/b^2$, see Weiss [1973, eqn.(4.48)], if the data are averaged rather than instantaneous.

Implementing Weiss' procedure raises various problems. The periodic mean and standard deviation were estimated by their sample moments, and first order harmonic series were fitted, denoted $\tilde{m}(t)$ and $\tilde{\sigma}(t)$ . As the sample means and standard deviations are zero for some intervals, the fitted series takes negative values. The fitted series can be constrained to be non-negative; this gives at most one point at zero. The data were standardized using the transformation

$$Y(t) = \begin{cases} \{ X(t) - m(t) \} \ / \ \sigma(t) & \text{if} \quad \sigma(t) > 0 \\ X(t) & \text{if} \quad \sigma(t) = 0 \end{cases}$$

As the sample means and standard deviations are very variable, the correlation estimated from the standardized data is smaller than that of the raw data. Once the coefficients and lags of the products in step 6 are found, there are fifteen equation in six unknowns to be solved - step 7.

This system will be inconsistent. The inconsistency arises as the coefficients and lags in 7b are determined by those in 7a. Thus some form of constrained estimation is needed at step 6 or step 7. Further constraints must be introduced, as the final estimates of $\nu(t)$ and $\theta(t)$ must be non-negative. This is done by changing the parameterisation to

$$\nu(t) = \nu_0 + (\nu_1^2 + \nu_2^2)^{\frac{1}{2}} + \nu_1 \cos\phi t + \nu_2 \sin\phi t$$

with $\nu_0$ , $\nu_1$ and $\nu_2$ constrained to be positive; similarly for $\theta(t)$. A NAG routine for nonlinear least squares is used to solve for $\nu(t)$ and $\theta(t)$. In general this results in $\nu_0$ and $\theta_0$ being set to zero, so that the rate input and size are zero at their minima.

The above procedure was followed for M2C8, giving the following parameter estimates:

$$b = .543 \quad ; \quad \nu_0 = 0. \quad \nu_1 = -.030 \quad \nu_2 = -.069$$
$$\theta_0 = 0. \quad \theta_1 = 3.381 \quad \theta_2 = -1.571$$

Ten sequences of data were generated, each seventeen years long, the length of the M2C8 record. Results from these simulations are given in tables 33, 34 and 35, with the designation "SNS1". The c.v. and skew are reasonable, as is the pattern of increase and decrease of monthly flow means. The actual proportion of null flows, taken to be those less than .001, the accuracy to which the data are given, is too high by a factor of

three and the conditional mean is twice the observed value . The variation in conditional monthly flow is compatible with that of the historical series. The seasonal pattern is shown in the two years of synthetic data in figure 19. The excess of zeroes and rapid decay in comparison with figure 1 is evident.

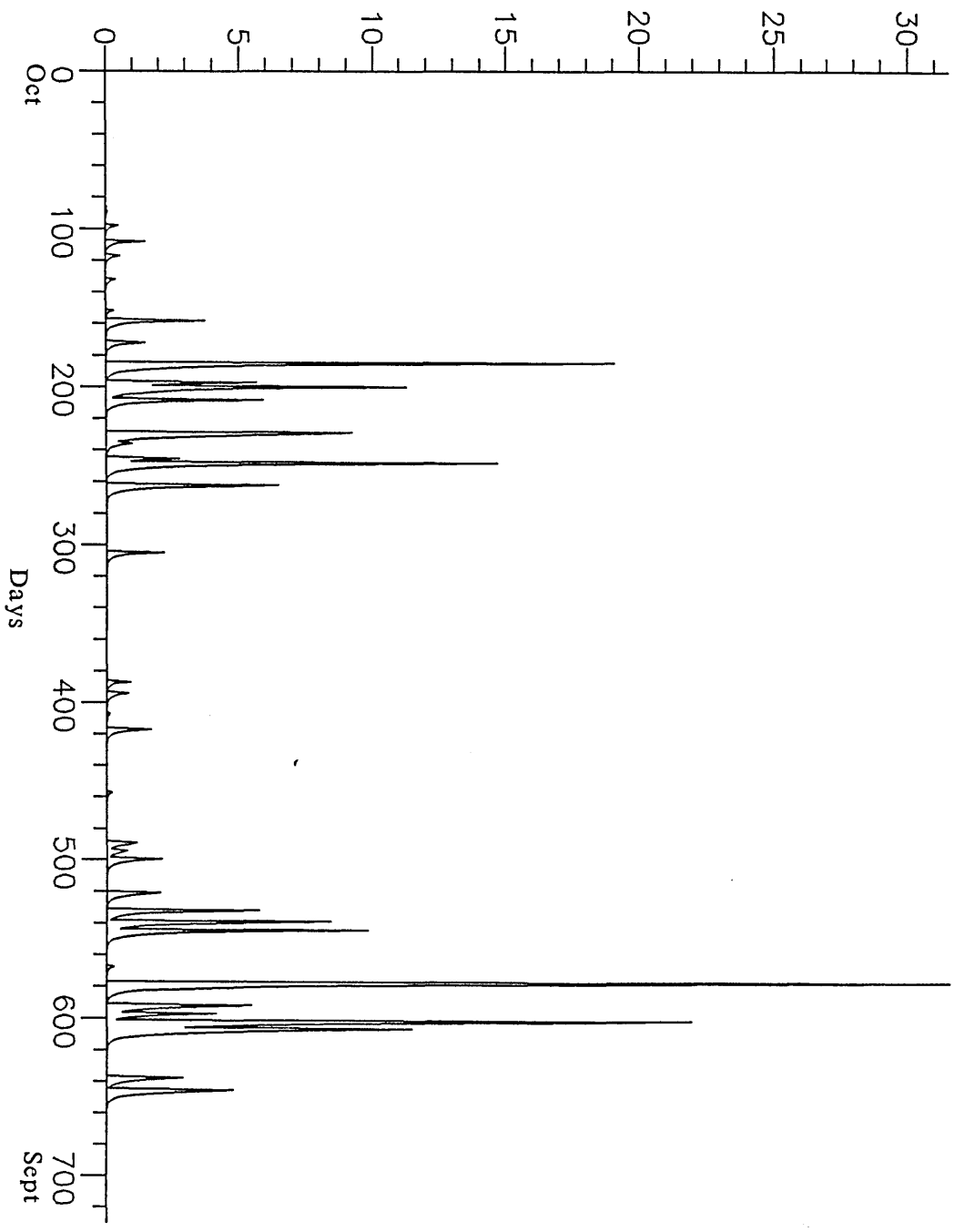
Table 33

*Annual statistics of ten simulations of shot noise for M2C8*

| | | | Conditional | | | |
|---|---|---|---|---|---|---|
| | Mean daily flow | % of flow at 0 | Mean daily flow | Stan dev. | Coef var | Skew |
| Obs. | .63 | 16.7 | .76 | 1.70 | 2.25 | 5.88 |
| SNS1 | .731 | 44.27 | 1.317 | 3.339 | 2.534 | 5.585 |
| s.e. | .046 | 1.69 | .096 | .254 | .068 | .493 |
| Step1 | 1.334 | 27.38 | 1.837 | 3.567 | 1.945 | 3.800 |
| s.e. | .053 | 1.02 | .066 | .127 | .058 | .439 |
| Step2 | .919 | 35.95 | 1.441 | 3.085 | 2.141 | 4.142 |
| s.e. | .097 | 2.21 | .068 | .160 | .021 | .239 |

As mentioned in §3.2, a simple form of periodic function is a step function. We let the event rate have this form, first with fixed jump points, and then with random end points. The end points were estimated by the means of the first and last days of each year for which there is an increase in flow level on the subsequent day. The parameters b, $\theta$ and $\nu$,

Figure 19 *M2C8 Shot noise simulation - SNS1; daily flow*



the value of the rate when nonzero, were then estimated from the flows on the days between the end points equating the mean, variance and first serial correlation coefficient for averaged data to the estimated values. The estimates are:

$$b = .786 \quad \theta = 5.260 \quad \nu(d) = \begin{cases} .163 & d \bmod 365 \in [39, 342] \\ 0 & \text{otherwise} \end{cases}$$

This model is denoted "Step 1" in the tables. The means are high , and the

Table 34.

*Coefficients of variation for positive flow; M2C8*

| | Oct | Nov | Dec | Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Obs. | .8 | 1.5 | .8 | .7 | .8 | .9 | 1.2 | 1.3 | 1.0 | 1.1 | 1.1 | 1.0 |
| SNS1 | 1.06 | .69 | .61 | .73 | .68 | .85 | .81 | 1.43 | 1.43 | .14 | .85 | .95 |
| s.e. | .44 | .16 | .09 | .11 | .10 | .29 | .19 | .39 | .48 | .44 | .24 | .18 |
| Step1 | .0 | .58 | .58 | .53 | .63 | .53 | .62 | .52 | .57 | .54 | .50 | .89 |
| s.e. | - | .09 | .09 | .11 | .29 | .08 | .18 | .08 | .13 | .10 | .11 | .20 |
| Step2 | 1.19 | .67 | .77 | .62 | .71 | .68 | .65 | .64. | .63 | .70 | .76 | .91 |
| s.e. | .28 | .20 | .28 | .08 | .14 | .11 | .12 | .24 | .13 | .12 | .19 | .13 |

cvs and skewnesses are small. The observed standard deviations of the end points are 28.05 and 21.31 respectively. As these are similar and non-negligible , we introduced more variation by making the end points random variables, with common standard deviation 24.9. The parameters

were estimated using all the flows between the first and last inputs in each year, and are:

$$b = .744 \quad \theta = 4.882 \quad \nu(d) = \begin{cases} .115 & d \bmod 365 \in [B,E] \\ 0 & \text{otherwise} \end{cases}$$

where B and E are assumed to have a normal distribution. This model is denoted "Step2" in the tables. The means are again large, though closer to

Table 35

*Conditional mean daily flow for each month; M2C8*

| | Oct | Nov | Dec | Jan | Feb | Mar |
|---|---|---|---|---|---|---|
| Obs. | .009 | .444 | .876 | 1.591 | 1.853 | 1.597 |
| SNS1 | .124 | .482 | 1.208 | 1.959 | 2.171 | 2.367 |
| s.e. | .048 | .096 | .167 | .222 | .419 | .590 |
| Step1 | .0 | 2.093 | 1.696 | 1.841 | 1.784 | 1.723 |
| s.e. | - | .163 | .350 | .383 | .242 | .348 |
| Step2 | 1.343 | 1.526 | 1.559 | 1.335 | 1.382 | 1.331 |
| s.e. | .479 | .300 | .224 | .296 | .210 | .204 |

| | Apr | May | June | July | Aug | Sep |
|---|---|---|---|---|---|---|
| Obs. | .901 | .391 | .165 | .098 | .047 | .024 |
| SNS1 | 1.402 | .869 | .151 | .768 | .255 | .035 |
| s.e. | .307 | .355 | .145 | .136 | .073 | .005 |
| Step1 | 1.900 | 1.781 | 1.839 | 1.781 | 1.850 | .867 |
| s.e. | .272 | .329 | .111 | .216 | .299 | .212 |
| Step2 | 1.426 | 1.336 | 1.398 | 1.377 | 1.369 | .944 |
| s.e. | .298 | .289 | .192 | .231 | .251 | .293 |

the historical values. The coefficients of variation and skew are larger than for "Step1", but still less than the original values. In both step models, the coefficients of variation for monthly conditional flows are too small. The conditional monthly means are too large, and do not show the appropriate pattern of increase and decrease over the year. This occurs because the rate and mean input size are constant when positive, and therefore the average flows are similar throughout the entire "wet" season.

The sinusoidal model, with seven parameters, gives more appealing results for M2C8 than the two step function models, with five and six parameters. If the method of estimating the parameters can be adjusted to give the right number of zero flows, then the sinusoidal model will be more useful for simulation of intermittent streams.

The second step function model was also fitted to AR3, an ephemeral stream. The end points, B and E, are normal with means 59.0 and 321.5, and common standard deviation 52.3. The remaining parameters are:

$$b = .820 \quad \theta = 435.424 \quad \nu(d) = \begin{cases} .021 & d \bmod 365 \in (B,E) \\ 0 & \text{otherwise} \end{cases}$$

The statistics from these simulations are given in table 36. The overall

Table 36

*Statistics of ten simulations of a shot noise model for AR3*

| | Mean daily flow | % of flow at 0 | Conditional Mean flow | Stan dev. | Coef var. | Skew |
|---|---|---|---|---|---|---|
| Observed | 9.37 | 76.70 | 33.73 | 107.07 | 3.17 | 5.37 |
| Snp step2 | 11.838 | 79.635 | 58.221 | 187.840 | 3.226 | 6.040 |
| s.e. | 1.775 | 2.592 | 6.104 | 20.964 | .115 | .796 |

*Coefficients of variation for positive flow*

| | Oct | Nov | Dec | Jan | Feb | Mar |
|---|---|---|---|---|---|---|
| Observed | 1.6 | 1.2 | 1.4 | 1.7 | 1.5 | 1.7 |
| Snp step2 | 1.15 | 1.04 | 1.34 | 1.14 | 1.17 | 1.54 |
| s.e. | .51 | .56 | .40 | .25 | .20 | .54 |

| | Apr | May | June | July | Aug | Sep |
|---|---|---|---|---|---|---|
| Observed | 1.7 | 2.1 | 1.1 | 1.3 | 1.0 | 1.1 |
| Snp step2 | 1.44 | 1.28 | 1.50 | 1.54 | 1.17 | 1.17 |
| s.e. | .46 | .27 | .73 | .51 | .27 | .39 |

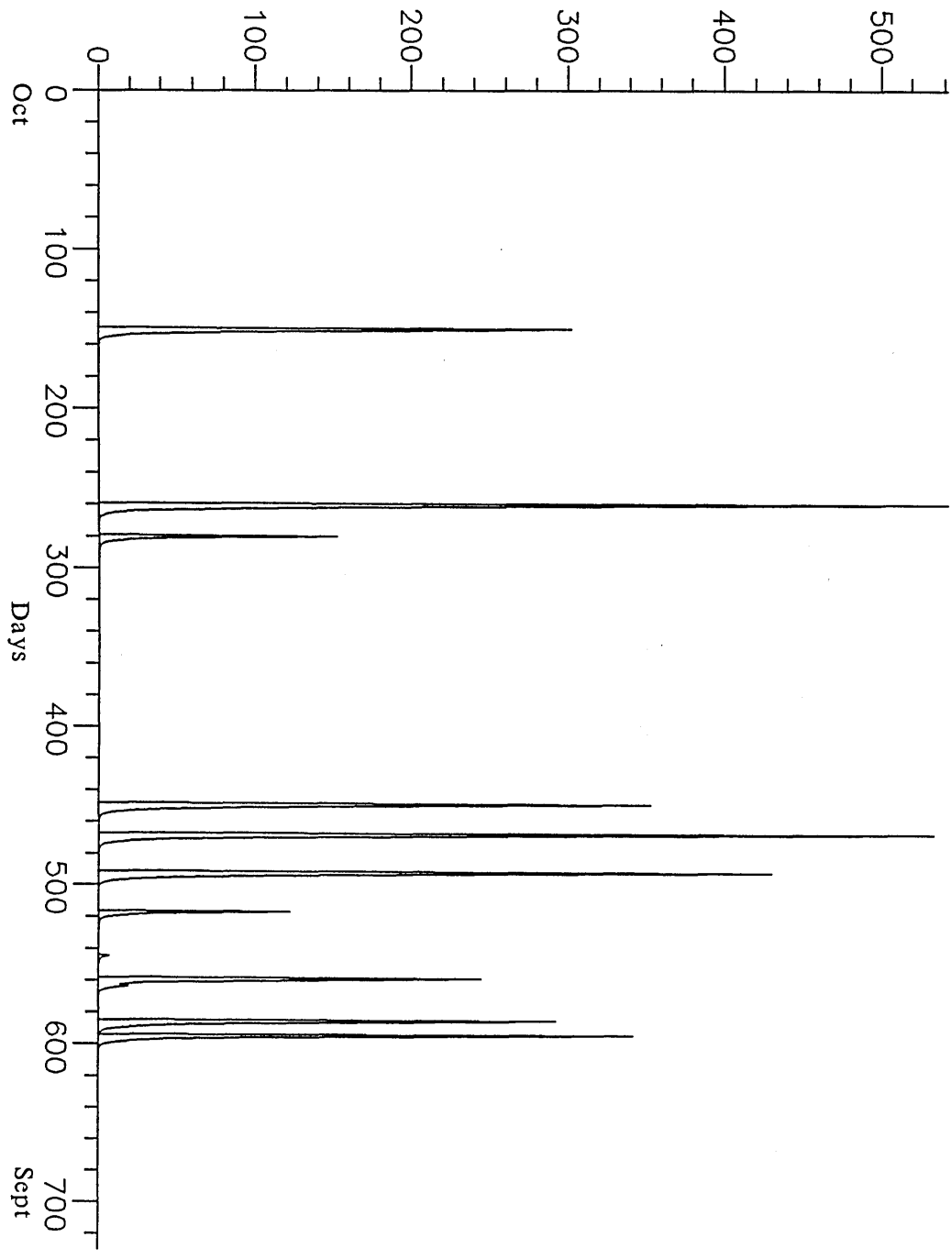
*Mean nonzero flows for each month*

| | Oct | Nov | Dec | Jan | Feb | Mar |
|---|---|---|---|---|---|---|
| Observed | 14.361 | 35.595 | 41.426 | 49.498 | 54.376 | 33.295 |
| Snp step2 | 67.916 | 61.760 | 81.192 | 71.115 | 72.384 | 85.190 |
| s.e. | 54.220 | 48.239 | 48.026 | 29.872 | 48.498 | 53.814 |

| | April | May | June | July | Aug | Sep |
|---|---|---|---|---|---|---|
| Observed | 12.416 | 3.216 | .779 | 12.525 | 26.180 | 15.848 |
| Snp step2 | 73.991 | 90.190 | 65.128 | 69.102 | 48.738 | 47.539 |
| s.e. | 60.010 | 50.948 | 46.213 | 44.943 | 16.768 | 37.514 |

statistics of mean daily flow, percentage of zeroes, conditional coefficient of variation and skewness are all larger than the historical values, but within two standard errors of them. The monthly conditional mean flows again do not show a pattern of increase and decrease over the year. In figure 20 there is some clustering of flows. The graph is more similar to that of the historical data in figure 2 than the simulation illustrated in figure 18 is. The historical data decline more gradually for small flows. The cvs of conditional flow in the shot noise simulations take values in the same range as the historical data. These monthly statistics have large standard errors, i.e. there is considerable variation between the seventeen year long simulations. As $\nu$ is small and $\theta$ large, events will be sparse and highly variable in size. It was noted that large variation is a characteristic of the data, and the step function variant is adequate for ephemeral streams.

Figure 20 *AR3 Shot noise simulation - Step2; daily flow*



# 5 Conclusions

Intermittent and ephemeral streamflow data present particular problems for statistical analysis. The statistics which characterise a river are the number of days without flow and the distribution of these days within the year, the mean size of flow and the variability of flow, which is large. The distinction between the two types of dry river is most obvious in the contrast between the single dry season of an intermittent stream and clusters of days without flow throughout the year for ephemeral streams. Ephemeral streams also tend to have flows which are more variable and skew than those of intermittent rivers. Stochastic models must reflect these features.

There are many ways in which a simple non-negative time series model can be generalized. However, theoretical results are non-trivial, especially once non-linearity is introduced. It is possible to deduce from approximations that the marginal distributions of variations on the basic storage model are close to gamma distributions, which are commonly used in hydrological statistics. Elaborations which introduce zero flows have a large influence on the distributional properties of the time series, even for small perturbations.

Simulations of seasonal storage models suggest that intermittent streams are best generated from models with smoothly varying parameters. It may be necessary to impose a multiplicative error structure on the sequence to increase the variance and skewness of the flows. The variability of ephemeral streams is better reproduced by seasonal parameters which are step functions with random end points and levels. Shot noise processes provide continuous time models for hydrological series. Again theoretical progress is difficult once seasonality is introduced and adjustments made to include zero flows. Synthetic data from seasonal shot noise processes shows similar advantages and disadvantages for continuous versus step functions for input rate and size. Intermittent streams are more adequately imitated by the smoothly varying parameters, whereas step functions are probably sufficient for ephemeral streams. In both cases estimates of input rate and size reflect the difference between the two types of dry river. The flows generated by a shot noise process with sinusoidal parameters have greater variation and skewness than those of sinusoidally based storage models. This advantage must be weighed against the more involved estimation and simulation procedure of the shot noise process.

Both storage models and shot noise processes have potential value

in summarizing data from dry rivers and generating flow sequences to aid in making decisions about the development of water resources. There is considerable scope for futher work on methods of fitting the models, whether analytic or computational, and in assessing the distributional properties of parameter estimates.

References.

**Abdulrazzak, M. J. and Morel-Seytour, H. J.** (1983). Recharge from an ephemeral stream following wetting front arrival to water-table. *Wat. Res. Res.* **19**, 194-200.

**Abramowitz, M. and Stegun, J. A.** (1964). *Handbook of mathematical functions with formulae, graphs and mathematical tables.* National Bureau of Standards Applied Mathematics Series, 55.

**Brill, P. H.** (1979). An embedded level crossing technique for dams and queues. *J. App. Prob.* **16**, 174-186.

**Diskin, M. H. and Lane, L. J.** (1972). A basinwide stochastic model for ephemeral stream runoff in south-east Arizona. *Bull. Int. Ass. Hyd. Sci.* **17**,61-76.

**Erdelyi, A.** (1954). *Tables of Integral Transformations, Vol 1.* McGraw-Hill.

**Gaver, D. P. and Lewis, P. A. W.** (1980). First order autoregressive gamma sequences and point processes. *Adv. Appl. Prob* **12** , 727-745.

**Gradshteyn, I. S. and Ryshik, I. M.** (1965) *Tables of integrals and, series and products.* Academic press.

**Kisiel, C. C., Duckstein, L. and Fogel, M. M.** (1971). Analysis of ephemeral flow in arid lands. *J. Hyd. Div. ASCE* **97** HY10 1699-1717.

**Lane, L. J. , Diskin, M. A. and Renard, K. G.** (1971). Input-output relationships for an ephemeral stream channel system. *J. Hydrol.* **13**,22-40.

**Lawrance, A. J. and Lewis, P. A. W.** (1980). The exponential autoregressive-moving average EARMA(p.q) process. *J. R. Statist. Soc.* B 42, 150-161.

**Lee, S.** (1975). Stochastic generation of synthetic streamflow sequenced in ephemeral streams. *Int. Ass. Hyd. Sci.* **Pub. no. 117**, 691-701.

**Peebles, R. W. ,Smith, R. E. and Yakowitz, S. J.** (1981). A leaky reservoir model for ephemeral stream flow recession. *Wat. Res. Res.* **17** , 628-636.

**Srikanthan, R. and McMahon, T. A.** (1980). Stochastic generation of monthly flows for ephemeral streams. *J Hydrol.* **47** , 19-40.

**Verhoeven, T. J.** (1977). The impact of high rainfall on an area within the Australian arid zone. *Hydrology Symposium 1977*. Brisbane, 28-30 June1977.

**Weiss, G.** (1973). Filtered Poisson processes as models for daily streamflow data. *Ph. D Thesis. Univ of London.*

**Yakowitz, S. J.** (1973). A stochastic model for dsily flows in an arid region. *Wat. Res. Res.* **9** 1271-1285.