Filtered Poisson Processes as Models for Daily Streamflow Data

by

Gideon Weiss

Thesis submitted for the

Diploma of Membership of Imperial College

and the degree of

Doctor of Philosophy in the University of London

September 1973

This work is dedicated to

the memory of

Oded Levin

Abstract

Synthetic streamflow data, generated by the use of pseudo-random numbers from a stochastic model, are employed in simulation studies of hydrological problems. Some models for daily streamflow are proposed.

A prominent feature of daily streamflow data is the existence of rapid rises and slow recessions in the flow. Such behaviour may be exhibited by filtered Poisson processes. A physical interpretation of using such processes as models for streamflow is given.

Existence of rapid rises and slow recessions in a process implies that, unlike a Gaussian process, it is not time reversible. Time reversibility of stochastic processes is discussed and a statistical test for it is derived. It is shown that time reversibility of discrete mixed moving average autoregressive processes is a unique property of Gaussian processes.

Particular examples of filtered Poisson processes are shot noise processes which are first order autoregressive and have gamma distributed marginals. Shot noise processes are described in detail, including the discussion of multivariate shot noise processes, integrated shot noise processes and shot noise processes with periodic structure.

Sample path properties of shot noise processes such as the distribution of maxima and minima and expected crossing times are obtained. It is shown that the Ornstein Uhlenbeck process can be described as a limit of a series of shot noise processes, and expected first passage times for it are derived using this

representation.

Several models for multisite daily streamflow based on the shot noise process are proposed, and methods of parameter estimation and synthetic data generation are given. In this context, a method for handling non positive definite sample correlation matrices is suggested. The generation of synthetic daily data by disaggregation of monthly data is also discussed.

The application of some of the models to East Anglia data is described and the numerical results are analysed.

Acknowledgements

I am deeply indebted to my supervisor, Professor D.R. Cox, for his invaluable guidance and encouragement at all stages of this research. I am also deeply indebted to Professor T. O'Donnell and Mr. P.E. O'Connell for their supervision of a great part of this research. I wish to thank Mr. J.R. Sexton, Mr. P. O'Neill, Dr. M.D. Godfrey, Dr. M.J. Hall and Mr. A. Buishand for their constant interest and valuable help. My thanks are due to Dr. A.M. Herzberg for comments on the form of this thesis. I also thank Mrs. L. Rice for her efficient typing of the manuscript.

This research was carried out under contract to the Water Resources Board of England and Wales. I am grateful to the Board for their financial support and wish to thank the Director and the Staff of the Board who made this research possible.

Table of Contents

List of Tables

List of Figures

INTRODUCTION

The present work was carried out at Imperial College as part of a research contract with the Water Resources Board of England and Wales. The contract called for study and development of new methods of analysing and extending streamflow and other hydrological data, in order to obtain maximum information for the assessment of available water resources.

In this context multisite streamflow data were studied, and the following problems looked at: (a) development of stochastic models to describe multisite daily streamflow; (b) estimation of model parameters to fit daily streamflow data; (c) generation of synthetic data sequences; and (d) assessment of the success of such synthetic data in reproducing properties of historical data.

In particular it was required to develop models which can reproduce cross correlations and low flow properties of multisite daily streamflow. It was, however, realised that one of the main features of daily streamflow is that it consists of a series of rapid rises followed by slow recessions; it was decided to use models with this feature.

The models developed here seem to be adequate for data of some British streams that were studied. It is hoped that these models are also suitable for different streamflow regimes, and that similar methods will be suitable for other types of hydrological data.

The order in which the work is presented is described below:

Chapter 1 includes background material on modelling of hydrological data. A few comments are made on the use of simulation and of synthetic data generation to solve design and operation problems in hydrology. This provides the context in which the present work is to be used.

A discussion of models for monthly streamflow follows. Some of the problems encountered in modelling daily streamflow, e.g. seasonality of data and dependence in multisite flow, as well as methods for parameter estimation have been treated in the modelling of monthly streamflow. Existing approaches to these problems are reviewed. Finally special features of daily streamflows and of models for them, are discussed.

Chapter 2 contains a description of some classes of general stochastic processes which seem suitable for modelling streamflow. Filtered Poisson processes and linear filtered Poisson processes are defined and described. The use of such processes as models for daily streamflow is interpreted in hydrological terms. The "shot noise" process, on which the proposed models for daily streamflow are based, is cited as an example of a linear filtered Poisson process.

Chapter 3 introduces the concept of time reversibility of stationary stochastic processes. The existence of rapid rises and slow recessions in daily streamflow is a manifestation of time irreversibility. A statistical test for time reversibility is described. A theoretical result is given, showing that, for a restricted class of linear processes, time reversibility is a unique property of Gaussian processes.

In Chapter 4, properties of the shot noise process are developed in detail, through the use of linear operations.

Chapter 5 considers sample path properties of the shot noise process, which include behaviour of maxima and minima and expected crossing times. These properties may help in the investigation of low flow and flood periods. It is shown that the Gaussian first order autoregressive process can be obtained as a limit of a sequence of shot noise processes. This limiting approach is used incidentally to obtain results on expected crossing times of that Gaussian process.

Chapter 6 describes several models for daily streamflow based on the shot noise process. The results of Chapter 4 are used to estimate the parameters of these models.

Chapter 7 gives numerical results, obtained by applying two of the models of Chapter 6, namely the single shot noise model and the double shot noise model, to streamflow data from East Anglia.

Chapter 1

GENERAL BACKGROUND

1.1 Simulation Techniques in Hydrology

Four major categories of variables play a role in hydrology: (a) natural variables which include precipitation, streamflow, ground water levels and physical characteristics of drainage basins and aquifers; (b) demand variables which describe water requirements in space and time; (c) operational variables which include storage sizes, order and timing of storage construction and pumping and distribution policies; and (d) economic variables which include construction and operation costs, penalties for unsatisfied demands and environmental benefits and damages. Hydrological design can often be viewed as a decision problem in which natural variables and some demand variables are the constraints, other demand variables and operational variables are the controls, and the optimisation of economic variables is sought.

An example of such a problem, which can be solved analytically, is found in the theory of storage (Moran, 1959). More usually however, the formulation of hydrological problems is too complicated to allow analytical solutions. It may, however, be possible to formulate the interrelation between the variables in a simulation program. By fixing the constraints and varying the controls through several simulation runs the economic variables can then be optimised.

The stochastic character of natural variables, such as streamflow, implies that the quantities which have to be optimised are expected values of the economic variables. Simulation runs merely provide an estimate of those expectations. Hence, long realizations of the natural variables are required to obtain reliable answers. Such long realizations can be provided by synthetic data, that is data generated by random or pseudo-random numbers, from a model fitted to historical data.

In practice, simulation provides only approximate solutions. Simplifying assumptions have to be made about the interrelations of the different variables and some variables are ignored. The variables themselves are also approximated: demand variables are obtained by socio-economic predictions and economic variables, such as penalties, have to be determined subjectively. The use of synthetic data introduces two types of approximation errors; errors in the model which is fitted and estimation errors in the model parameters.

In this work some models of streamflow, which are used to generate synthetic data, are proposed. These synthetic data are to be used in simulation studies of regional water supply in Britain. Sensitivity studies carried out at the Water Resources Board of England and Wales suggest that such simulations should be on a daily basis; see Jamieson, Radford and Sexton (1973). Similar studies show the importance of cross correlations between daily flows at different sites, and of low flow properties, to regional water supply regulation. Therefore, preservation of these properties by the proposed models is desired.

In particular, a simulation of the Wash Estuary area is now being developed at the Water Resources Board (Sexton and Jamieson 1973a, 1973b), and the data used in Chapter 7 for deriving numerical results were drawn from that region.

1.2 Models for Monthly Streamflow

Models for the generation of synthetic monthly streamflow data have been in use for some time. In this section a short review of monthly streamflow models is presented, followed by a discussion of points which these models have in common with daily streamflow models. The generation of synthetic monthly streamflow apparently originated in the Harvard Water Program (Thomas and Fiering, 1962). The methods used in that project have been extended (Fiering, 1964; Matalas, 1967) and programmed for a computer (Young and Pisano, 1968).

These models are based on the multivariate Gaussian first order autoregressive process, which is formulated (Matalas, 1967) as

$$\underset{\sim}{Z}(t) = A\,\underset{\sim}{Z}(t-1) + B\,\underset{\sim}{\varepsilon}(t), \tag{1.1}$$

where $\underset{\sim}{Z}(t) = \{Z_1(t),\ldots,Z_K(t)\}$. In (1.1) $Z_i(t) \sim N(0,1)$, $\underset{\sim}{\varepsilon}(t) \sim N(0,I)$ and $\underset{\sim}{\varepsilon}(t)$ form a sequence of independent vector variables, with $\underset{\sim}{\varepsilon}(t)$ independent of $\underset{\sim}{Z}(t-1)$. The matrices A and B are related to the covariance matrices $M_0 = E\{\underset{\sim}{Z}(t)\underset{\sim}{Z}(t)'\}$ and $M_1 = E\{\underset{\sim}{Z}(t)\underset{\sim}{Z}(t-1)'\}$ by the following equations.

$$\begin{aligned} A &= M_1 M_0^{-1}\,, \\ BB' &= M_0 - M_1 M_0^{-1} M_1'\,. \end{aligned} \tag{1.2}$$

The application of this process to build models for monthly streamflow involves the following problems.

<u>Scaling</u>: Let the streamflow to be modelled be $\underset{\sim}{X}(t) = \{X_1(t),\ldots X_k(t)\}$, with $E\{X_i(t)\} = \mu_i$ and $Var\{X_i(t)\} = \sigma_i^2$. Then $\underset{\sim}{X}(t)$ is obtained from $\underset{\sim}{Z}(t)$ in (1.1) by

$$X_i(t) = \sigma_i Z_i(t) + \mu_i\,. \tag{1.3}$$

The use of (1.2) and (1.3) ensures that $X_i(t)$ has mean μ_i and variance σ_i^2, and that the lag zero correlation matrix of $\underset{\sim}{X}(t)$ is M_0 and the lag one correlation matrix of $\underset{\sim}{X}(t)$ is M_1.

Seasonality: The observed streamflow is subject to seasonal variations. Therefore $\underset{\sim}{X}(t)$ must be a process with periodic structure, with a period of one year. If the serial and cross correlation structure of $\underset{\sim}{X}(t)$ is assumed to be stationary, and only the means and variances change seasonally, $\underset{\sim}{X}(t)$ can be obtained from (1.3) by substituting $\mu_i = \mu_i(t)$ and $\sigma_i = \sigma_i(t)$ (Young and Pisano, 1968). A similar procedure can be employed for some of the daily models proposed in this work; see sections 4.4 and 6.5.

A different approach to seasonality is to estimate each of the model parameters separately for each season (Thomas and Fiering, 1962; Beard, 1965). This approach was also employed here for some models; see Chapter 6. A difficulty which arises with this approach is that transient effects, in the passage from one season to another, distort the results. An elegant method for estimating parameters of an autoregressive process with periodic structure is presented by Jones and Brelsford (1967).

Transformations: The assumption that the streamflow $\underset{\sim}{X}(t)$ has a normal distribution is rarely valid, since streamflow is non negative, and is as a rule positively skewed. An assumption which has been made in some studies is that by transforming streamflow as $\underset{\sim}{Z}(t) = f\{\underset{\sim}{X}(t)\}$, the transformed process $\underset{\sim}{Z}(t)$ has a normal distribution and can be fitted by (1.1). Suggested transformations are square root or logarithmic transformations (Fiering, 1964; Young and Pisano, 1968) or a transformation based on an assumption that $\underset{\sim}{X}(t)$ has a Pearson Type III distribution (Beard, 1965). This approach cannot be used with data which is not time reversible, and is therefore unsuited to daily streamflow; see section 2.6 and Chapter 3.

A different approach to obtain a model for non normal distributions is to assume that the innovations vector $\underset{\sim}{\varepsilon}(t)$ in (1.1) has a non normal distribution. The difficulty here is that the relation between the distributions of $\underset{\sim}{X}(t)$ and of $\underset{\sim}{\varepsilon}(t)$ is complicated. In essence this approach is employed in the models proposed here.

Estimation of Model Parameters: A method of parameter estimation, which is generally accepted in the modelling of monthly streamflow, is that of preserving suitable properties of the historical data (Matalas, 1967). To formulate this method, let $\theta_1, \ldots, \theta_m$ be the parameters of the model. Let $a_1, \ldots, a_m$ be random values of some properties of the streamflow, with observed values in the historical sample $\tilde{a}_1, \ldots, \tilde{a}_m$. The estimates of the parameters, $\tilde{\theta}_1, \ldots, \tilde{\theta}_m$, are obtained by solving the equations

$$E(a_i|\tilde{\theta}_1,\ldots,\tilde{\theta}_m) = \tilde{a}_i \quad (i = 1,\ldots m). \tag{1.4}$$

This procedure ensures that the properties $a_1, \ldots, a_m$ of synthetic data generated by the model will have $\tilde{a}_1, \ldots, \tilde{a}_m$ as expected values.

To illustrate the use of this method, let the distribution of the streamflow $X(t)$ be given by $\log X(t) \sim N(\mu,\sigma^2)$, and let the historical data be $X_1,\ldots, X_N$. Two properties which are usually of importance in design problems are the average flow and the variance of the flow. Thus it is useful to preserve the sample mean $\bar{X} = \sum_{t=1}^{N} X_t/N$, and the sample variance $S^2 = \sum_{t=1}^{N} (X_t-\bar{X})^2/N$. To preserve these the estimates $\tilde{\mu}$ and $\tilde{\sigma}^2$ must satisfy

$$\begin{aligned} \bar{X} &= \exp\{\tfrac{1}{2}\tilde{\sigma}^2+\tilde{\mu}\}, \\ S^2 &= \exp\{2(\tilde{\sigma}^2+\tilde{\mu})\} - \exp\{\tilde{\sigma}^2+2\tilde{\mu}\}. \end{aligned} \tag{1.5}$$

It should be noted that the maximum likelihood estimates for μ and σ^2, based on the sufficient statistics $\sum_{t=1}^{N} (\log X_t)$ and $\sum_{t=1}^{N} (\log X_t)^2$, are

$$\hat{\mu} = \sum_{t=1}^{N} \log X_t / N ,$$
$$\hat{\sigma}^2 = \sum_{t=1}^{N} (\log X_t)^2 / N - \hat{\mu}^2 . \tag{1.6}$$

Hence the estimates $\tilde{\mu}$ and $\tilde{\sigma}^2$ entail a loss of efficiency. On the other hand, $\tilde{\mu}$ and $\tilde{\sigma}^2$ preserve $\overline{X}$ and S^2, which, if the model were only approximate, could be badly distorted through the use of $\hat{\mu}$ and $\hat{\sigma}^2$. In other words, the proposed method minimises the bias arising from an incorrect model at the expense of the efficiency of estimation when the model is correct.

For the model described by (1.1) - (1.3), one can preserve the sample lag zero and lag one correlation matrices, as well as the sample means and sample variances of the data.

Positive Definiteness: To estimate the matrix B in (1.1), the sample estimates of M_0 and M_1, the lag zero and lag one correlation matrices of $\underset{\sim}{X}(t)$, are substituted in (1.2). The estimated matrix B will be real valued if the estimated right hand side of (1.2), BB', is positive definite or positive semi-definite. For data consisting of equi-length vectors with no missing values, this can be ensured by proper choice of the estimators of M_0 and M_1 (Valencia and Schaake, 1973). If some of the data are missing, or if transformations of the data are employed, it is quite possible that the estimated BB' will not be positive definite or semi-definite. In such cases some changes in the estimation of BB' have to be made. This problem was studied in relation to hydrological models by Fiering (1968) and by Crosby and Maddock (1970). The difficulty was also encountered here and a method of overcoming it is proposed in section 6.4.

Roesner and Yevjevich (1966) attempt to verify the applicability of models based on (1.1) for monthly streamflow; 137 series of monthly streamflow and 219 series of monthly rainfall data are examined.

Moreau and Pyatt (1970) attempt to fit a fourth order autoregressive model to weekly streamflow, and similar techniques are used by Quimpo (1968) and by Payne, Newman and Kerri (1969) to model daily streamflow.

Long Term Persistence Models: In the design of high yield storage reservoirs, the crucial design variable is R(t), the range of the cumulative flows over the design period t. It was found by Hurst (1951, 1956) and by Mandelbrot and Wallis (1969b), through the use of long records of annual streamflows and of related geophysical phenomena, that the sample rescaled range R(t)/S(t) (i.e. the range divided by the estimated standard deviation of the data) behaves as t^h for large t, where h, the Hurst coefficient, is typically 0.7. On the other hand, Mandelbrot and Wallis (1969c) have stated that for the process $\underset{\sim}{Z}(t)$ defined by (1.1), R(t)/S(t) behaves asymptotically as $t^{½}$. Several models, which give a theoretical value of $h > ½$, or which approximate $h > ½$ for finite but large values of t, have been proposed by Mandelbrot and Wallis (1968, 1969a), Mandelbrot (1971), O'Connell (1971), Rodriguez-Iturbe, Mejia and Dawdy (1972), Mejia, Rodrigues-Iturbe and Dawdy (1972) and Garcia, Dawdy and Mejia (1972).

For the study of regional water supply problems on a daily basis, the behaviour of R(t)/S(t) on an annual time scale is of secondary importance. Nevertheless, it seems that techniques similar to those described by Mandelbrot (1971) can be adapted to the models proposed here.

Models for Floods and Intermittent Streamflow: Streamflow in semi-arid rivers consists of long periods of zero flow and short periods of high flow. Models for intermittent flow and models for flood peaks are proposed in papers by Yakowitz (1972), Todorovic and Zelenhasic (1968) and Zelenhasic (1970). Although some superficial resemblance exists between models employed in these papers and models proposed here, more research will be needed to implement the models proposed here to semi-arid streams.

The models proposed here are based on the shot noise process which is a special case of a filtered Poisson process. Filtered Poisson processes and the shot noise process in particular, were suggested as models in a discussion of daily streamflow models by Bernier (1970).

1.3 Evaluation of Streamflow Models

The evaluation of a model normally consists of statistical tests on whether the data are consistent with the probabilistic assumptions of the model. Such tests can be carried out by comparing the original, historical data with synthetic data produced from the model. Theoretically, by producing enough synthetic data and carrying out comparisons of enough statistics, all the assumptions of the model can be tested in this way. In practice, however, there has to be a limit to the number of tests. If the object of the model is to provide synthetic data for simulation studies, then comparisons are made on properties which are known to affect the outcome of the simulation. Thus in synthetic data generation the stress is not on testing the assumptions underlying the model, but on testing properties which are problem orientated. This is a subjective way of evaluating the model, since for different problems different tests may be used and different conclusions may be drawn.

In particular, the method of estimating parameters so as to preserve some properties of the data, as discussed in section 1.2, is subjective since the choice of preserved properties is problem orientated. There are, however, two objective points to consider when choosing preserved properties. One is that these properties should be estimated by statistics which make good use of the information in the data. For example, the sample mean will be for particular models, a satisfactory statistic and will contain more information than, say, a single extreme value of the data. Thus it is usually preferable to preserve the mean than to preserve an extreme value. A second consideration is that preservation of a standard set of properties could form a good basis for comparing different models.

Synthetic data will automatically agree with the historical data on properties preserved through the estimation. A crude way of comparing other properties is as follows. A number of synthetic sequences, of the same length as the historical record, are generated and the statistics which measure the property that one wishes to compare are calculated, for both the historical and the synthetic sequences. The test then consists of observing how well the historical value lies within the range of the synthetic values.

There follows a discussion of properties which are important in the study of streamflow and for which comparison of historical and synthetic data should be made.

Basic Properties: These include the mean, the standard deviation and the first serial correlation coefficient of the flow at each site, and the lag zero cross correlation coefficients between sites. These properties can be estimated quite efficiently by the standard sample estimates, and will usually be chosen to be preserved by the model.

Sequential Dependence: This includes serial correlation coefficients of various lags, and the spectral density of the process. These properties will be used especially to test whether a first order autoregressive model is appropriate for the data.

Long Term Persistence Properties: Long term persistence is measured by the Hurst coefficient h. The choice of $h > \frac{1}{2}$ is made on the basis of geophysical records longer than the normally available streamflow data. The value of h affects the behaviour of the cumulative range, and the frequency of very long low flow periods.

Distribution Properties: These include the skewness, the higher moments and the empirical marginal distribution of the streamflow. These properties will be used to perform goodness of fit tests or to suggest a transformation of the data. Skewness may be preserved by some models (Beard, 1965).

Extreme Values: Both the distribution and the serial structure of extreme values are of interest. The distribution of high flows is especially important for flood control studies. The lengths of low flow periods are important in supply regulation.

Aggregation Properties: When a model is fitted to daily data, it is important that monthly averages resulting from the model agree with monthly averages of the historical data. This is partly achieved here by preserving in addition to the daily mean, the daily standard deviation, and the daily first serial correlation coefficient, also the monthly standard deviation and the monthly first serial correlation coefficient; see sections 6.7 - 6.10.

A different approach is to generate monthly data, and to disaggregate it into daily flows; see Schakke, Ganslaw, Fothergill and Harbaugh (1972) and Valencia and Schaake (1973). This approach is discussed in sections 4.5 and 6.11.

Visual Inspection: Examination of daily hydrographs may suggest properties of interest, which are not quantified by any of the properties discussed above. Thus models which generate hydrographs similar in appearance to observed streamflow are desirable.

Physical Justification: Some properties of a stochastic model may be based on known physical properties of streamflow. Such properties give the model a firmer basis than that provided by statistical comparison of synthetic and historical data alone.

1.4 Guidelines for Daily Multisite Streamflow Models

A very prominent feature of daily streamflow is the existence of rapid rises and slow recessions in the flow. These are clearly observable in graphical representations of the flow (see Fig. ~~2~~). 4

This appearance of daily streamflow suggests a model in which a sequence of random inflows occurs at random times, each of which causes a rise followed by a recession in the streamflow. This contrasts with Gaussian models, which arise from a random input of minute quantities all the time.

A class of processes exhibiting rapid rises and slow recessions are filtered Poisson processes introduced in Chapter 2. The process chosen as a basis for modelling streamflow, the shot noise process, is a first order autoregressive process belonging to that class. It has three parameters, and is of comparable complexity to the Gaussian first order autoregressive process, discussed in section 1.2.

The models based on the shot noise process are continuous time processes, and represent the instantaneous, continuous time, streamflow. Daily streamflow data, however, are discrete, and are obtained as averages of several measurements during each day. It is therefore assumed that the historical data consists of consecutive daily averages of the continuous-time flow. Similarly synthetic data is generated as consecutive daily averages of the continuous-time process.

Observation of multisite daily streamflow shows that rises and recessions often occur simultaneously at different sites. This is incorporated into the proposed models as the source of cross correlation.

Another property of streamflow is that it consists of both surface runoff and groundwater discharge. This is manifested in the data by the fact that recessions do not as a rule decay towards zero flow. To represent this, a model consisting of the sum of two shot noise processes is proposed.

Chapter 2

PROBABILISTIC BACKGROUND

2.1 Filtered Poisson Processes

In this section and in sections 2.2 - 2.4 filtered Poisson processes and linear filtered Poisson processes are discussed. These processes seem to offer a good choice of models for daily streamflow, since the property of rises and recessions in the flow, mentioned in section 1.4, follows from their definition. The shot noise process, on which the proposed models are based, is defined in section 2.5 as an example of a linear filtered Poisson process. The hydrological interpretation of choosing linear filtered Poisson processes as models for daily streamflow is given in section 2.6.

The definition of filtered Poisson processes is based on Parzen (1964, section 4.5). For additional material on filtered Poisson processes, see Rice (1944, 1945), Laning and Battin (1956, section 4.1) and Gilbert and Pollack (1960).

Definition 2.1: Let $\{N(t)\}$, $t > 0$, be a Poisson process of rate υ. Let $0 \le \tau_1 \le \ldots \le \tau_m \ldots$ be the times of the events of $\{N(t)\}$. Let $\{Y_m\}$ be a sequence of independent identically distributed variables, which are independent of $\{N(t)\}$. Let $w(t,\tau,Y)$ be some real function. A filtered Poisson process $X(t)$ is given by

$$X(t) = \sum_{m=1}^{N(t)} w(t,\tau_m,Y_m) \, . \tag{2.1}$$

Thus $X(t)$ is the sum of the values at t of those pulses $w(t,\tau,Y)$ which occured at τ, $0 < \tau < t$. Each individual pulse, $w(t,\tau_m,Y_m)$, has a random starting point τ_m, and a random value Y_m determines its size and shape. Note that since the summation (2.1) is over events in $(0,t)$, one may assume $w(t,\tau,Y) = 0$, for $\tau>t$. The sequences $\{\tau_m\}$ and $\{Y_m\}$

constitute the random part of $X(t)$ and are characterised by υ and by the distribution $F(y)$, with density $f(y)$ if that exists, of the $\{Y_m\}$.

The properties of $X(t)$ are determined by the joint characteristic functions of $X(t_1),\ldots X(t_n)$ for any n and any $t_1,\ldots,t_n$, denoted by $\phi(u_1,\ldots,u_n;t_1,\ldots,t_n)$. The following theorem (Parzen, 1964, section 4.5) relates these characteristic functions to υ, $F(y)$ and $w(t,\tau,Y)$. The proof is outlined here because various points in it are used in following sections.

Theorem 2.1

Let $X(t)$ be a filtered Poisson process as in (2.1). Then for any $t_2 > t_1 > 0$,

$$\phi(u,v;t_1,t_2) = \exp[\upsilon\int_0^{t_1} E\{e^{iuw(t_1,\tau,Y)+ivw(t_2,\tau,Y)}-1\}d\tau$$

$$+ \upsilon\int_{t_1}^{t_2} E\{e^{ivw(t_2,\tau,Y)}-1\}d\tau]. \tag{2.2}$$

Proof

By definition,

$$\phi(u,v;t_1,t_2) = E\{e^{iuX(t_1)+ivX(t_2)}\}.$$

The expectation can be written down conditionally on $N(t_2)$ and on $\tau_1 = s_1,\ldots, \tau_n = s_n$. By the joint independence on $N(t)$, the conditional expectation is

$$\phi(u,v;t_1,t_2|n,s_1\ldots,s_n) = \prod_{i=1}^{n} E\{e^{iuw(t_1,s_i,Y_i)+ivw(t_2,s_i,Y_i)}\}$$

where, in accordance with (2.1), one may set $w(t,\tau,Y) = 0$ for $\tau > t$.

The event times $\tau_1,\ldots\tau_n$, conditional on $N(t) = n$, form an ordered sample from a uniform distribution on $(0,t_2)$. Removal of the conditions $\tau_1 = s_1,\ldots,\tau_n = s_n$ yields

$$\phi(u,v;t_1,t_2|n) = \frac{1}{t_2^n}[\int_0^{t_1} E\{e^{iuw(t_1,\tau,Y)+ivw(t_2,\tau,Y)}\}d\tau$$

$$+ \int_{t_1}^{t_2} E\{e^{ivw(t_2,\tau,Y)}\}d\tau]^n \quad .$$

The final condition, $N(t_2) = n$, is removed by

$$\phi(u,v;t_1,t_2) = \sum_{n=0}^{\infty} \frac{(\upsilon t_2)^n e^{-\upsilon t_2}}{n!} \phi(u,v;t_1,t_2|n) \; ,$$

thus giving (2.2).

Several useful results follow directly from (2.2). By setting $v = 0$, one obtains $\phi(u;t)$, the characteristic function of $X(t)$.

$$\phi(u;t) = \exp[\upsilon\int_0^t E\{e^{iuw(t,\tau,Y)}-1\}d\tau] \; . \tag{2.3}$$

By taking derivatives of $\log\{\phi(u,v;t_1,t_2)\}$, one obtains

$$E\{X(t)\} = \upsilon\int_0^t E\{w(t,\tau,Y)\}d\tau \; , \tag{2.4}$$

$$\mathrm{var}\{X(t)\} = \upsilon\int_0^t E\{w(t,\tau,Y)^2\}d\tau \; , \tag{2.5}$$

$$\mathrm{cov}\{X(t_1),X(t_2)\} = \upsilon\int_0^{t_1} E\{w(t_1,\tau,Y)w(t_2,\tau,Y)\}d\tau \quad (t_2>t_1). \tag{2.6}$$

Similarly the general (n,m) joint cumulant of $X(t_1)$ and $X(t_2)$ is obtained as

$$K_{nm}\{X(t_1),X(t_2)\} = \upsilon\int_0^{t_1} E\{w(t_1,\tau,Y)^n w(t_2,\tau,Y)^m\}d\tau \cdot (t_2>t_1). \tag{2.7}$$

In definition 2.1, $X(t)$ is defined as the sum of the values at time t of all the pulses which have started in $(0,t)$, and $X(0)$ is defined as zero. Thus $X(t)$ in (2.1) is a transient filtered Poisson process.

To define a stationary filtered Poisson process, let $\{N(t)\}$ be a Poisson process, now over the whole real line, and let $w(t,\tau,Y) = w(t-\tau,Y)$. For an arbitrary T, consider the process $X(t)$ defined as a sum over all events in $(-T,T)$, i.e.

$$X(t) = \sum_{m=N(-T)}^{N(T)} w(t-\tau_m, Y_m).$$

Using the same steps as in the proof of Theorem 2.1, one obtains for this process

$$\phi(u,v;t_1,t_2) = \exp[\upsilon\int_{-T}^{T} E\{e^{iuw(t_1-\tau,Y)+ivw(t_2-\tau,Y)}-1\}d\tau].$$

Similar expressions can be found for $\phi(u_1,\ldots,u_n;t_1,\ldots,t_n)$ for every n and every $t_1,\ldots,t_n$. If a continuous limit of $\phi(u_1,\ldots,u_n;t_1,\ldots,t_n)$ exists as $T \to \infty$, then

$$X(t) = \sum_{m=-\infty}^{\infty} w(t-\tau_m,Y_m) \tag{2.8}$$

is a well defined stationary process, and satisfies

$$\phi(u,v;t,t+s) = \exp[\upsilon\int_{-\infty}^{\infty}E\{e^{iuw(\tau,Y)+ivw(\tau+s,Y)}-1\}d\tau]. \tag{2.9}$$

From (2.9) expressions for $\phi(u;t)$, $E\{X(t)\}$, $Var\{X(t)\}$, $cov\{X(t),X(t+s)\}$ and $K_{nm}\{X(t),X(t+s)\}$ of the stationary process $X(t)$ are easily obtained. These expressions are similar to those given by (2.3) - (2.7) for the transient process, with the only difference that the limits of integration are $(-\infty,\infty)$. Note that with $X(t)$ defined by (2.8), the value of $X(t)$ is determined not only by past events but also by the values of $w(t-\tau_m,Y_m)$ for $\tau_m > t$.

2.2 Nonhomogeneous Filtered Poisson Processes

Streamflow and other hydrometeorological processes are not stationary, but are at best processes with a periodic structure. It is therefore necessary to investigate nonstationary filtered Poisson processes.

A simple generalisation is to take a nonhomogeneous Poisson process $\{N(t)\}$ with rate $\upsilon(t)$, and events at $0 \leq \tau_1 \leq \ldots \leq \tau_m \ldots$. With the events is associated a sequence of variables $Y_m = Y(\tau_m)$, $(m=1,2,\ldots)$, where $\{Y_m\}$ are jointly independent, and where the distribution of Y_m depends on the time τ_m of the event but does not depend on $\{N(t)\}$ otherwise. The pulse shape is given by $w\{t,\tau_m,Y(\tau_m)\}$.

On defining, similarly to (2.1),

$$X(t) = \sum_{m=1}^{N(t)} w\{t,\tau_m,Y(\tau_m)\}, \tag{2.10}$$

it can be shown that for the nonhomogeneous process $X(t)$,

$$\phi(u,v;t_1,t_2) = \exp\Big(\int_0^{t_1} \upsilon(\tau)E[e^{iuw\{t_1,\tau,Y(\tau)\}+ivw\{t_2,\tau,Y(\tau)\}}-1]d\tau$$

$$+\int_{t_1}^{t_2} \upsilon(\tau)E[e^{ivw\{t_2,\tau,Y(\tau)\}}-1]d\tau\Big). \tag{2.11}$$

The proof of (2.11) follows essentially the same steps as the proof of Theorem 2.1. The only changes are that now

$$P\{N(t) = n\} = \frac{\{\Lambda(t)\}^n e^{-\Lambda(t)}}{n!},$$

where

$$\Lambda(t) = \int_0^t \upsilon(\tau)d\tau,$$

and the conditional distribution of $\tau_1 \leq \ldots \leq \tau_n$ given $N(t) = n$ is

$$f(\tau_1 \leq \ldots \leq \tau_n | n) = n!\,\frac{\upsilon(\tau_1)\ldots\upsilon(\tau_n)}{\{\Lambda(t)\}^n}.$$

If $N(t)$ and the variables $Y(\tau)$ are defined over the whole real line, the definition of $X(t)$ may be extended to

$$X(t) = \sum_{m=-\infty}^{\infty} w\{t,\tau_m,Y(\tau_m)\}, \tag{2.12}$$

provided that the sum converges.

For processes with periodic structure it is fairly easy to check whether a limit exists, and in section 4.4 use is made of such processes.

2.3 Multisite Filtered Poisson Processes

To model multisite streamflow, multivariate filtered Poisson processes are required.

Definition 2.1 can be extended to a definition of a multivariate process $\underset{\sim}{X}(t) = \{X_1(t),\ldots,X_k(t)\}$ as follows. Let $\{N(t)\}$ be a single Poisson process with rate υ, and events at $\ldots \leq 0 \leq \tau_1 \leq \ldots \leq \tau_m \ldots$ With each event τ_m, a vector $\underset{\sim}{Y}_m$ is associated, such that $\{\underset{\sim}{Y}_m\}$ forms a sequence of independent identically distributed random vectors, and $\{\underset{\sim}{Y}_m\}$ are independent of $\{N(t)\}$. Each vector $\underset{\sim}{Y}_m$, $\underset{\sim}{Y}_m = \{Y_{1m},\ldots,Y_{Lm}\}$, may have dependent components, and L may be different from k. Let $\underset{\sim}{w}(t,\tau,\underset{\sim}{Y}) \doteq \{w_1(t,\tau,\underset{\sim}{Y}),\ldots,w_k(t,\tau,\underset{\sim}{Y})\}$ be a vector of k real functions. Then a multivariate filtered Poisson process $\underset{\sim}{X}(t)$ is defined by

$$X_i(t) = \sum_{m=1}^{N(t)} w_i(t,\tau_m,\underset{\sim}{Y}_m), \quad (i=1,\ldots k). \tag{2.13}$$

With $\phi(\underset{\sim}{u},\underset{\sim}{v};t_1,t_2)$ denoting the joint characteristic function of $\underset{\sim}{X}(t_1)$ and $\underset{\sim}{X}(t_2)$ $(t_2>t_1)$ it can be shown, using the same steps as in the proof of Theorem 2.1, that

$$\phi(\underset{\sim}{u},\underset{\sim}{v},t_1,t_2) = \exp[\upsilon\int_0^{t_1}E\{e^{i\underset{\sim}{u}'\underset{\sim}{w}(t_1,\tau,\underset{\sim}{Y})+i\underset{\sim}{v}'\underset{\sim}{w}(t_2,\tau,\underset{\sim}{Y})}-1\}d\tau$$

$$+\upsilon\int_{t_1}^{t_2}E\{e^{i\underset{\sim}{v}'\underset{\sim}{w}(t_2,\tau,\underset{\sim}{Y})}-1\}d\tau]. \tag{2.14}$$

In particular (2.14) can be used to obtain the cross covariance between $X_i(t_1)$ and $X_j(t_2)$, $(t_2>t_1)$, for $i,j = 1,\ldots,k$

$$\text{cov}\{X_i(t_1),X_j(t_2)\} = \upsilon\int_0^{t_1}E\{w_i(t_1,\tau,\underset{\sim}{Y})w_j(t_2,\tau,\underset{\sim}{Y})\}d\tau. \tag{2.15}$$

The definition of $\underset{\sim}{X}(t)$ can be extended to the stationary case as in section 2.1.

Two special cases of (2.13) are of interest. For the first, let each event τ_m have associated with it a random vector $\underset{\sim}{G}_m = \{G_{1m},\ldots,G_{km}\}$

and a random vector $\underset{\sim}{Y}_m = \{Y_{1m},\ldots,Y_{km}\}$, and let the pulses be of the shape $w_i(t,\tau_m+G_{im},Y_{im})$. Thus the start of the pulse associated with the event τ_m, for the i^{th} component of $\underset{\sim}{X}(t)$, is at τ_m+G_{im}. This means effectively that the events behave like events in a multivariate Poisson process of a type discussed by Cox and Lewis (1972).

For the second special case, let each event τ_m have associated with it a random vector $\underset{\sim}{I}_m = \{I_{1m},\ldots,I_{km}\}$ and a random vector $\underset{\sim}{Y}_m = \{Y_{1m},\ldots,Y_{km}\}$, where the values of $\underset{\sim}{I}_m$ are all zeroes or ones and $\underset{\sim}{I}_m$ is independent of $\underset{\sim}{Y}_m$. Let the pulses be $w_i(t,\tau_m,I_{im},Y_{im})$, where

$$w_i(t,\tau_m,I_{im},Y_{im}) = \begin{cases} w_i(t,\tau_m,Y_{im}) & (I_{im}=1) \\ 0 & (I_{im}=0) \end{cases} . \tag{2.16}$$

In this case events occur in effect simultaneously for a subset of the components $X_1(t),\ldots,X_k(t)$. Thus $X_i(t)$ will be marginally a filtered Poisson process with event rate $\nu P(I_{im}=1)$. This type of process is used in section 4.2.

2.4 Linear Filtered Poisson Processes

A linear process $X(t)$ is defined as

$$X(t) = \int_{-\infty}^{\infty} w(t-\tau)\,dY(\tau), \tag{2.17}$$

where $Y(t)$ is a homogeneous additive process, a process with stationary independent increments, and $w(t)$ is some real function. The stochastic integral (2.17) converges in the sense of mean square if

$$E\{Y(t)\}^2 < \infty ,$$

$$\int_{-\infty}^{\infty} |w(t)|^p dt < \infty \qquad (p=1,2) ;$$

see Doob (1953, sections 9.2 and 11.9). The increments of a homogeneous additive process, $Y(t+s)-Y(t)$ $(s>0)$, have a log characteristic function given by

$$\psi(u,s) = i\gamma us + s\int_{-\infty}^{\infty}\{e^{i\lambda u}-1 - \frac{i\lambda u}{1+\lambda^2}\}\frac{1+\lambda^2}{\lambda^2}\, dG(\lambda), \tag{2.18}$$

with $G(\lambda)$ a bounded nondecreasing function.

Two special cases of the function $G(\lambda)$ give rise to important classes of linear processes. The first case is

$$G(\lambda) = \begin{cases} \sigma^2 & (\lambda \geq 0) \\ 0 & (\lambda < 0) \end{cases}. \tag{2.19}$$

In this case $Y(t+s)-Y(t)$ has a normal distribution, and the process $Y(t)$ is a Brownian motion. The use of this process as the additive process in (2.17) gives rise to Gaussian linear processes.

For the second case, let $F(\lambda)$ be a distribution function, and let

$$dG(\lambda) = \upsilon\frac{\lambda^2}{1+\lambda^2}\, dF(\lambda). \tag{2.20}$$

In this case $Y(t)$ is a compound Poisson process. The use of this $Y(t)$ as the additive process in (2.17) gives rise to a filtered Poisson process, with $N(t)$ having event rate υ, the variables $\{Y_m\}$ having distribution $F(Y)$, and the pulse shape being of the special form $w(t,\tau,Y) = Yw(t-\tau)$. Thus, linear filtered Poisson processes are defined either by

$$X(t) = \sum_{m=-\infty}^{\infty} Y_m w(t-\tau_m) \tag{2.21}$$

or by (2.17), (2.18) and (2.20).

2.5 The Shot Noise Process as an Example of a Linear Filtered Poisson Process

The process central to the subsequent work is in many ways the simplest example of a linear filtered Poisson process, and is called the shot noise process. It is defined by

$$X(t) = \sum_{m=1}^{N(t)} Y_m e^{-b(t-\tau_m)}, \tag{2.22}$$

where N(t) has event rate υ, b is a constant decay rate, $b > 0$, and the variables Y_m are exponentially distributed, with mean θ, i.e. their density function is

$$f(y) = \frac{1}{\theta} e^{-y/\theta} ; \tag{2.23}$$

see Bartlett (1957) and Bernier (1970).

A detailed examination of the shot noise process X(t) is given in Chapter 4. In the present section the characteristic function of X(t) is derived, to illustrate the theory in the previous sections.

By (2.3)

$$\phi(u;t) = \exp\left[\upsilon\int_0^t E\{e^{iuye^{-b\tau}}-1\}d\tau\right]. \tag{2.24}$$

The characteristic function of the exponential variable Y, is:

$$\phi_Y(u) = \frac{1}{1-i\theta u} .$$

The substitution of $\phi_Y(u)$ into the integrand of (2.24) yields

$$\phi(u;t) = \exp\left[\upsilon\int_0^t \left\{\frac{1}{1-i\theta ue^{-b\tau}} -1\right\}d\tau\right],$$

and performing the integration one obtains

$$\Phi(u;t) = \left(\frac{1-i\theta ue^{-bt}}{1-i\theta u}\right)^{\upsilon/b}. \tag{2.25}$$

Equation (2.25) gives the characteristic function of X(t) for the transient process defined in (2.22). As $t \to \infty$, the characteristic function for the stationary process is obtained

$$\phi(u;t) = \left(\frac{1}{1-i\theta u}\right)^{\upsilon/b} . \tag{2.26}$$

Thus the marginal distribution of the stationary shot noise is gamma with parameters $\frac{1}{\theta}$ and $\frac{\upsilon}{b}$, i.e.

$$f(x) = \frac{(1/\theta)^{\upsilon/b} x^{\upsilon/b-1}}{\Gamma(\upsilon/b)} e^{-x/\theta}. \tag{2.27}$$

2.6 Streamflow as a Stochastic Process

In the following chapters, processes based on the shot noise process are used as models for daily streamflow. In this section, an interpretation is given in hydrological terms, of the assumption that streamflow is a linear filtered Poisson process.

Catchments can be considered as systems to which rainfall is the input and streamflow is the output (Dooge, 1959). In this context the input is assumed to contain all the randomness, i.e. for a given input the output is deterministic.

If a linear filtered Poisson process is assumed to model the streamflow, then the sequence of variables $\{Y_m\}$ at the Poisson times $\{\tau_m\}$ constitutes the random input, while $w(t-\tau)$ describes the system.

Before examining $\{\tau_m\}$ and $\{Y_m\}$ as inputs it should be noted that rainfalls do not behave like such inputs, since they exhibit serial dependence, and since rainstorms are shaped like pulses. It is thus necessary to postulate an extended system in which $\{Y_m\}$ and $\{\tau_m\}$ are assumed to be some climatic events forming the input from which rainfall and streamflow result. Some studies of daily rainfalls show that they may be Markovian (Gabriel and Neumann, 1962), which fits this picture. Since only the output of the system, the streamflow, is studied here, the assumption of an extended system presents no difficulties.

Two assumptions are made when $\{Y_m\}$ and $\{\tau_m\}$ are chosen as the input. The first assumption is that the input is discrete, that is during a time interval dt the input dY(t) is either zero, or with probability νdt a noninfinitesimal random quantity Y. This seems to be borne out by the observation, on a daily basis, of rapid rises and slow recessions in the streamflow, and by the appearance of daily and hourly rainfalls as series of dry periods interspersed by fairly short rainstorms. The assumption that dY(t) is Gaussian white noise, that is $dY(t) \sim N(0,\sigma^2 dt)$, is certainly untenable, although it may account for a part of the input.

The second assumption is that $\{\tau_m\}$ are Poisson events and $\{Y_m\}$ are independent and independent of the Poisson porcess N(t). These properties make Y(t), the accumulated input, into an additive process. It is quite conceivable that the times between climatic events are more complicated, e.g. renewal processes, and that Y_m depends on the interval length (τ_{m-1},τ_m). This would give rise to filtered renewal processes, (Takács, 1956) and introduce short term dependencies in the input. An additive input is also contested by Mandelbrot and Wallis (1968) who suggest a fractional noise model introducing very long serial dependence in the input.

The assumption that the input is compound Poisson implies for the streamflow that the times between rises in the flow are serially independent, are independent of the magnitudes of the rises, and are exponentially distributed. Tests of these assumptions are performed on some streamflow data in section 7.3, with results which bear out only some of the assumptions.

The deterministic system producing the streamflow is characterised by the pulse shape $w(t,\tau,Y)$, on which three assumptions are made. It is assumed that $w(t,\tau,Y) = 0$ for $\tau > t$, i.e. that the streamflow depends on past input only. It is assumed further that $w(t,\tau,Y) = Yw(t,\tau)$, i.e. that the system is linear. Finally it is assumed that

$w(t,\tau) = w(t-\tau)$, i.e. that the system is stationary. Together with the assumptions on the input these assumptions lead to a linear filtered Poisson process (see section 2.4). Thus

$$X(t) = \int_{-\infty}^{t} w(t-\tau)dY(\tau) \tag{2.28}$$

where $Y(t)$ is the compound Poisson process defined by $\{Y_m\}$ and $\{\tau_m\}$.

Note that $X(t)$ is nonstationary if the additive input is non-homogeneous. The shape of $w(t-\tau)$ completely determines the system, and since an additive input $Y(t)$ is assumed, $w(t-\tau)$ also determines the autocorrelation function of $X(t)$.

Linearity of the system is equivalent to the assumption that for any two input realisations, $Y_1(t)$ and $Y_2(t)$, with respective outputs $X_1(t)$ and $X_2(t)$, and for any constants a_1 and a_2, the input $a_1Y_1(t) + a_2Y_2(t)$ produces the output $a_1X_1(t) + a_2X_2(t)$. Stationarity of the system is equivalent to the assumption that the input $Y_1(t+T)$ produces the output $X_1(t+T)$, for any T. Both assumptions do not hold for the relation between rainfall and streamflow, but they are quite widely accepted as approximations.

The function $w(t-\tau)$ is known in hydrology as the instantaneous unit hydrograph, IUH. This is defined as the limit of the streamflow resulting from a rainfall of 1 inch, uniform over the catchment area and over a time period, as the time period tends to zero. The IUH can be estimated, for example, through harmonic analysis of rainfall and streamflow (O'Donnell, 1966). A catchment can also be modelled conceptually as an interlinked system consisting of simple linear systems e.g. reservoirs and channels, to yield a much simplified IUH. Some such simple systems and their resulting IUH are listed (Kraijenhoff, 1966).

Single Linear Reservoir. It is assumed that the output $X(t)$ is proportional to the level of storage S(t), with proportionality constant b. Then, for accumulated input Y(t),

$$X(t) + \frac{1}{b}\frac{d}{dt}X(t) = \frac{d}{dt}Y(t), \tag{2.29}$$

by which

$$X(t) = \int_{-\infty}^{t} be^{-b(t-\tau)}dY(\tau), \tag{2.30}$$

and so

$$w(t-\tau) = be^{-b(t-\tau)} \quad (t>\tau). \tag{2.31}$$

This is the system assumed for the shot noise process (see section 2.5), and for the monthly streamflow model (1.1).

Two linear reservoirs in series yield

$$w(t-\tau) = \frac{b_1 b_2}{b_2-b_1}\{e^{-b_1(t-\tau)} - e^{-b_2(t-\tau)}\} \quad (t>\tau), \tag{2.32}$$

where b_1 and b_2 are the reservoir constants.

A cascade of K identical linear reservoirs yields

$$w(t-\tau) = \frac{b^K(t-\tau)^{K-1}}{(K-1)!}e^{-b(t-\tau)} \quad (t>\tau), \tag{2.33}$$

Two linear reservoirs in parallel with proportional inputs. This system has IUH

$$w(t-\tau) = (1-\theta)b_1e^{-b_1(t-\tau)} + \theta e^{-b_2(t-\tau)} \quad (t>\tau), \tag{2.34}$$

where b_1 and b_2 are the reservoir constants, and where a proportion $1-\theta$ of the input enters the first reservoir and a proportion θ of the input enters the second; in section 6.10 such a model is discussed.

A linear channel with lag T yields

$$w(t-\tau) = \delta\{(t-\tau)-T\}. \tag{2.35}$$

The final assumption made in defining the shot noise process as a model for streamflow, is that the variables $\{Y_m\}$ are exponentially distributed. This assumption is made mainly for mathematical convenience. Although it can be given a physical interpretation, based on the lack of memory of the exponential distribution, no empirical evidence exists to support it. Consequences of this assumption are used in section 4.5 and in Chapter 5.

Chapter 3

TIME REVERSIBILITY

3.1 Time Reversibility of Linear Stochastic Processes

The rapid rises followed by slow recessions observed in daily streamflow imply that if the time direction is reversed a different process is obtained, i.e. streamflow is not a time reversible process. In this chapter the property of time reversibility is investigated. Some of the results obtained are of general interest, and are unrelated to the rest of the work. Time reversibility is, however, used to derive a statistical test of whether daily streamflow may or may not be fitted by a Gaussian process, as discussed in section 3.2.

Definition 3.1. A stationary process $X(t)$ is said to be *time reversible* if for any T, any n, and any $t_1,\ldots,t_n$, the joint distribution of $\{X(t_1),\ldots,X(t_n)\}$ and the joint distribution of $\{X(T-t_1),\ldots,X(T-t_n)\}$ are identical.

A definition of second order time reversibility, involving only second moment properties is superfluous, since the covariance function of a stationary or second order stationary process is symmetrical. One could however consider K^{th} order time reversibility, for moments up to order K $(K>2)$. Because the covariance function is symmetrical, one has:

Theorem 3.1

Let $X(t)$ be a stationary Gaussian process, i.e. all finite dimensional distributions of $X(t)$ are multivariate normal. Then $X(t)$ is time reversible.

By Theorem 3.1, it is obvious that if streamflows are not time reversible, then they cannot be represented by a Gaussian process.

Moreover, let $X(t)$ be a stationary process, and let H be a one to one function, and let, for every t,

$$Y(t) = H\{X(t)\} . \tag{3.1}$$

Then obviously $Y(t)$ is time reversible if and only if $X(t)$ is time reversible. Therefore if the streamflow $X(t)$ is not time reversible, then no transformation of the type (3.1) can change $X(t)$ into a Gaussian process.

3.2 A Statistical Test for Time Reversibility

In the previous section, it was shown that if a stationary process $X(t)$ is not time reversible, then neither $X(t)$ nor processes obtained from $X(t)$ by a transformation of the type (3.1) can be Gaussian processes. In this section a statistical test for time reversibility is described. This test can thus be used to decide whether data should be fitted by a Gaussian process.

The test is based on crossing times of a stationary process $X(t)$. Let $X(t)$ be a process whose sample paths are, with probability one, continuous apart from a countable number of discontinuity points at which right hand and left hand limits exist. The following definition of crossing times is slightly different from that given by Cramér and Leadbetter (1967, Chapter 10).

Definition 3.2. Let $X(t)$ be a sample path of the process and let W be a fixed level. Then $X(t)$ has an upcrossing of W at t' if in every left neighbourhood of t' $(t'-\varepsilon,t')$, there exists t for which $X(t) < W$, and in every right neighbourhood of t', $(t',t'+\varepsilon)$, there exists t for which $X(t) > W$. A downcrossing of W at t' is defined similarly.

Let W_1 and W_2, $(W_2>W_1)$, be two fixed levels. Crossing times between the two levels are defined as follows:

Definition 3.3. A sample path $X(t)$ has an upcrossing from W_1 to W_2 in an interval (t',t''), where $t'' \geq t'$, if: (a) there is an upcrossing of W_1 at t'; (b) there is an upcrossing of W_2 at t''; and (c) for all t such that $t' < t < t''$, $W_1 \leq X(t) \leq W_2$. A downcrossing from W_2 to W_1 is defined similarly.

Let $X(t)$ be observed over a time T. A sequence of intervals (t_i',t_i'') of upcrossings from W_1 to W_2 and of intervals (s_i',s_i'') of downcrossings from W_2 to W_1 is obtained. Without loss of generality, let

$$0 \leq t_1' \leq t_1'' \leq s_1' \leq s_1'' \leq \dots s_n' \leq s_n'' \leq T$$

and let $T_1,\dots,T_n$ and $S_1,\dots,S_n$ be the lengths of the intervals, $T_i = t_i'' - t_i'$ and $S_i = s_i'' - s_i'$; see Fig. 1. The test for time reversibility is based on the sequences $\{T_i\}$ and $\{S_i\}$.

For a stationary process $X(t)$, $\{T_i\}$ are identically distributed random variables and $\{S_i\}$ are identically distributed random variables. The distributions of $\{T_i\}$ and $\{S_i\}$ are determined by the finite joint distributions of $X(t)$. If $X(t)$ is time reversible, it is easy to see that $\{T_i\}$ and $\{S_i\}$ are identically distributed. If $X(t)$ is a Markovian process then $\{T_i\}$ and $\{S_i\}$ are all independent.

Let $X(t)$ be stationary and Markovian. The test to be performed is between

H_0 : $X(t)$ is time reversible

H_1 : $X(t)$ is not time reversible.

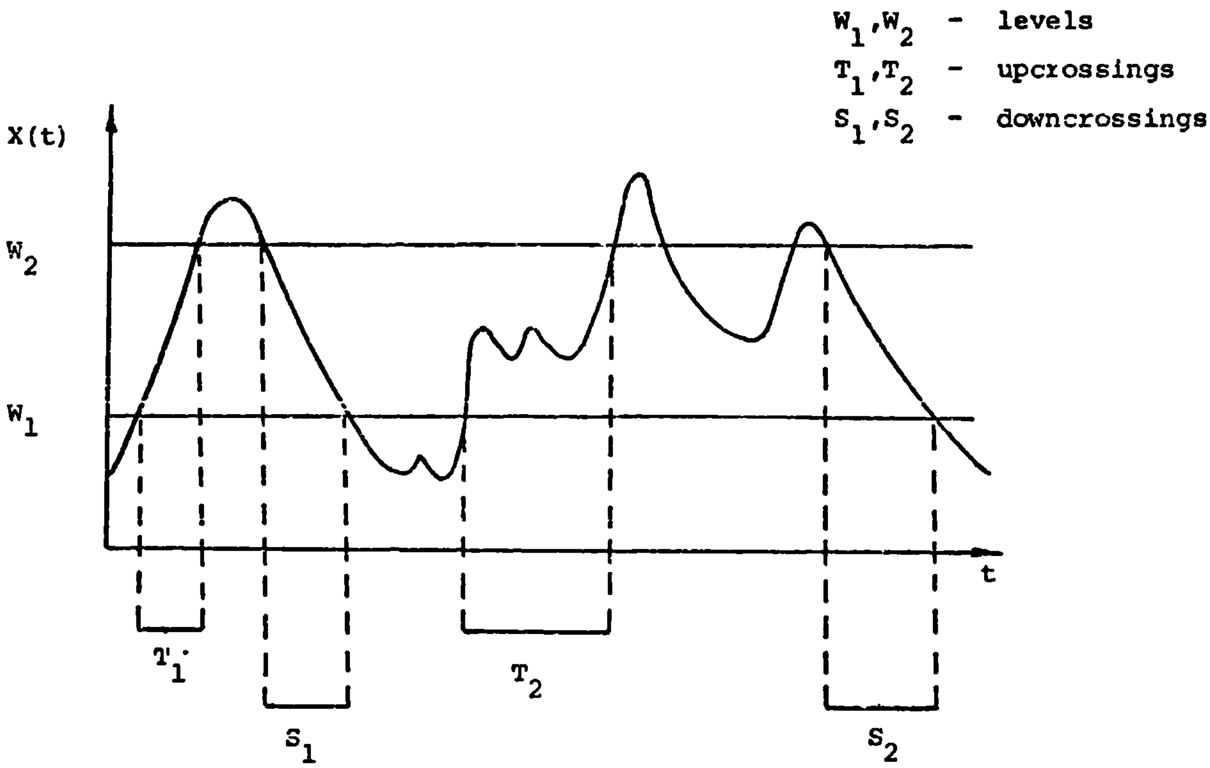

Figure 1. Crossing Times Between Two Levels

Under both H_0 and H_1, $T_1,\ldots,T_n$ and $S_1,\ldots,S_n$ form two independent random samples. Under H_0 the two samples come from the same distribution. Thus H_0 can be tested against H_1 by testing

H_0' : $T_1,\ldots,T_n$ and $S_1,\ldots,S_n$ are samples from the same distribution

H_1' : $T_1,\ldots,T_n$ and $S_1,\ldots,S_n$ are samples from different distributions

Various nonparametric tests can be used for H_0' against H_1'.

There follow some comments on the test.

The assumption that $X(t)$ is Markovian ensures that $T_1,\ldots,T_n$, $S_1,\ldots,S_n$ are independent. If, however, the times between crossing intervals are long, the dependence between T_i, S_i and T_{i+1} will often be small, even in non-Markovian processes. The degree of dependence can be regulated by the choice of levels W_1 and W_2.

The assumption that $X(t)$ is stationary is important for the test. In fact the test is probably also a good test against trend in the data.

The levels W_1 and W_2 can be chosen as fixed percentiles of the data. If this is done, the test is invariant under monotone transformations of the data. Thus, if for example percentiles P_1 and P_2 of the data are chosen, the same test values will be obtained before or after a logarithmic transformation of the data is employed. The question of optimal choice of P_1 and P_2 was not studied.

Several pairs of levels, giving non overlapping intervals, may be chosen, and a separate test can be performed for each pair. If the process is Markovian, then these tests are independent.

Finally, it should be remembered that the data available for a test will be discrete and thus the crossing times will have to be approximated by interpolation.

3.3 A Characterisation of Some Gaussian Linear Processes

In section 3.1, it was shown that Gaussian processes are time reversible. In this section the question, unrelated to streamflow modelling, is posed as to whether a linear time reversible process can be anything but a Gaussian process. It seems that the answer to this question is no. The theorem given here proves this assertion for a limited class of linear processes.

Theorem 3.2

Let X_t be a discrete time stationary mixed autoregressive moving average process, given by

$$X_t = \sum_{k=1}^{N} a_k X_{t-k} + \sum_{\ell=0}^{M} b_\ell \varepsilon_{t-\ell} , \tag{3.2}$$

with $\{\varepsilon_t\}$ an independent, identically distributed sequence, and ε_t independent of X_{t-s} $(s>0)$. Assume $N \neq 0$, or if $N = 0$, assume that for some ℓ $(\ell = 1,\ldots,M)$, $b_\ell \neq b_{M-\ell}$. If $X(t)$ is time reversible then ε_t is normally distributed.

The proof is divided into proofs of three special cases. Lemma 3.1 deals with the first order autoregressive process, Lemma 3.2 deals with the pure moving average process of order M, and finally in Lemma 3.3 the general case when $N \neq 0$ is covered. Lemma 3.1 is a consequence of the general case for $N \neq 0$, and is only brought in as an illustration, because of its simpler proof. To prove Lemma 3.1 the following characterisation of the normal distribution is used.

Characterisation 3.1 (Feller, 1971, Sections 3.4 and 15.8).
Let X_1 and X_2 be independent random variables. Let

$$Y_1 = a_{11}X_1 + a_{12}X_2,$$
$$Y_2 = a_{21}X_1 + a_{22}X_2, \tag{3.3}$$

where the transformation is non-trivial, i.e. the case $Y_1 = a_{11}X_1$ and $Y_2 = a_{22}X_2$ and the case $Y_1 = a_{12}X_2$ and $Y_2 = a_{21}X_1$ are excluded. If Y_1 and Y_2 are independent then X_1, X_2, Y_1 and Y_2 are all normally distributed.

Lemma 3.1

Theorem 3.2 holds for a first order autoregressive process

$$X_t = aX_{t-1} + \varepsilon_t. \tag{3.4}$$

Proof

Define

$$\eta_t = X_t - aX_{t+1}. \tag{3.5}$$

For every $s > 0$, consider

$$(\eta_t, X_{t+s}) = (X_t - aX_{t+1}, X_{t+s}),$$

$$(\varepsilon_{t+s}, X_t) = (X_{t+s} - aX_{t+s-1}, X_t).$$

By time reversibility, (X_t, X_{t+1}, X_{t+s}) and $(X_{t+s}, X_{t+s-1}, X_t)$ have the same joint distribution; this can be seen when replacying m by $2t + s - m$. Hence (η_t, X_{t+s}) and (ε_{t+s}, X_t) have the same joint distribution, and in particular η_t and X_{t+s} are independent for all $s > 0$.

Using (3.4) and (3.5) one can write

$$\varepsilon_{t+1} = (1-a^2)X_{t+1} - a\eta_t,$$
$$X_t = aX_{t+1} + \eta_t. \tag{3.6}$$

In (3.6) η_t and X_{t+1} are independent, and ε_{t+1} and X_t are independent. Hence by characterisation 3.1, all four variables are normally distributed.

To prove Lemmas 3.2 and 3.3, a more general characterisation of the normal distribution, of which Characterisation 3.1 is a special case, is used.

Characterisation 3.2 (Rao, 1966). Let $\underset{\sim}{Y}$ and $\underset{\sim}{Z}$ be two random vectors, with $\underset{\sim}{Y} = (Y_1,\ldots,Y_p)$ all independent, and $\underset{\sim}{Z} = (Z_1,\ldots,Z_q)$ all independent. Let A be an mxp matrix, and let B be an mxq matrix. Assume that the n^{th} column of A, $\underset{\sim}{A}_n$, is not proportional to any column of B. If the random vectors $A\underset{\sim}{Y}$ and $B\underset{\sim}{Z}$ have the same joint distribution, then Y_n is normally distributed.

Lemma 3.2

Theorem 3.2 holds for a moving average process,

$$X_t = \sum_{\ell=0}^{M} b_\ell \varepsilon_{t-\ell}, \tag{3.7}$$

if for some ℓ ($\ell=0,\ldots,M$), $b_\ell \neq b_{M-\ell}$.

The exclusion of the case for which $b_\ell = b_{M-\ell}$ for all ℓ is necessary, because in that case X_t is obviously time reversible, irrespective of the distribution of ε_t.

Proof

Consider the two vectors $\underset{\sim}{X}' = (X_t,\ldots,X_{t+M})$ and $\underset{\sim}{X}'' = (X_{t+M},\ldots,X_t)$, and the vector $\underset{\sim}{\varepsilon} = (\varepsilon_{t-M},\varepsilon_{t-M+1},\ldots,\varepsilon_t,\ldots,\varepsilon_{t+M})$. By (3.7) one can write

$$\begin{aligned} \underset{\sim}{X}' &= B_1\underset{\sim}{\varepsilon}\,, \\ \underset{\sim}{X}'' &= B_2\underset{\sim}{\varepsilon}\,, \end{aligned} \tag{3.8}$$

where

$$B_1 = \begin{bmatrix} b_M & b_{M-1} & \cdot & b_1 & b_0 & 0 & \cdot & \cdot & 0 \\ 0 & b_M & \cdot & \cdot & b_1 & b_0 & 0 & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & 0 & b_M & b_{M-1} & \cdot & b_1 & b_0 \end{bmatrix}, \tag{3.9}$$

$$B_2 = \begin{bmatrix} 0 & \cdot & \cdot & 0 & b_M & b_{M-1} & \cdot & b_1 & b_0 \\ 0 & \cdot & 0 & b_M & b_{M-1} & \cdot & \cdot & b_0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ b_M & b_{M-1} & \cdot & b_1 & b_0 & 0 & \cdot & \cdot & 0 \end{bmatrix}. \tag{3.10}$$

Because $b_0 \neq 0$ and $b_M \neq 0$ the $(M+1)^{th}$ column of B_1 is not proportional to any of the first M columns of B_2 or to any of the last M columns of B_2. By the assumption that for some ℓ, $B_\ell \neq B_{M-\ell}$, the $(M+1)^{th}$ column of B_1 is not proportional to the $(M+1)^{th}$ column of B_2. By the time reversibility $\underset{\sim}{X}'$ and $\underset{\sim}{X}''$ have the same joint distribution. Hence by Characterisation 3.2, ε_t is normally distributed.

<u>Lemma 3.3</u>

Theorem 3.2 holds for any finite autoregressive moving average process

$$X_t = \sum_{k=1}^{N} a_k X_{t-k} + \sum_{\ell=0}^{M} b_\ell \varepsilon_{t-\ell}, \tag{3.11}$$

where $N \neq 0$.

<u>Proof</u>

Define in a similar way to (3.4) and (3.5) two sequences of random variables

$$G_t = X_t - \sum_{k=1}^{N} a_k X_{t-k} = \sum_{\ell=0}^{M} b_\ell \varepsilon_{t-\ell}, \tag{3.12}$$

$$F_t = X_t - \sum_{k=1}^{N} a_k X_{t+k}. \tag{3.13}$$

By the time reversibility the pairs (G_t, F_t) and (F_t, G_t) have the same bivariate distribution. Moreover, for every r and s, the pairs (G_{t+s}, F_{t+r}) and (F_{t+r}, G_{t+s}) have the same distribution. To see the latter statement the time scale is to be reversed by letting m be replaced with 2t + s + r − m.

Let s > 0, and consider X_{t+s}. By (3.11) X_{t+s} can be written in terms of $X_{t-N}, \ldots, X_{t-1}$ and $\varepsilon_{t-M}, \ldots, \varepsilon_{t+s}$, as

$$X_{t+s} = \sum_{k=1}^{N} p_k X_{t-k} + \sum_{\ell=0}^{s+M} q_\ell \varepsilon_{t+s-\ell}. \tag{3.14}$$

Furthermore, the coefficients q_ℓ for $\ell = 0, \ldots, s$ are independent of s, and $q_\ell \to 0$ as $\ell \to \infty$. Similarly by (3.13) and (3.14), F_{t+s} can be written as

$$F_{t+s} = \sum_{k=1}^{N} u_k X_{t-k} + \sum_{\ell=0}^{s+N+M} v_\ell \varepsilon_{t+s+N-\ell}, \tag{3.15}$$

where v_ℓ for $\ell = 0, \ldots, s+N$ are independent of s, and $v_\ell \to 0$ as $\ell \to \infty$. Also, since X_t is not a pure moving average process, the sequence v_ℓ is infinite.

Choose now an index L such that $L \geq N$, $L \geq M$, $v_L \neq 0$ and for every $\ell \geq L-M$, $v_\ell \leq \frac{1}{2}\min(b_0, \ldots, b_M)$. Then F_{t+L-N} can be expressed as in (3.15) by

$$F_{t+L-N} = Y_1 + \sum_{\ell=0}^{M} v_{L-\ell} \varepsilon_{t+\ell} + Y_2, \tag{3.16}$$

where Y_1 is a function of $X_{t-N}, \ldots, X_{t-1}$ and of $\varepsilon_{t-M}, \ldots, \varepsilon_{t-1}$, and Y_2 is a function of $\varepsilon_{t+M+1}, \ldots, \varepsilon_{t+L}$. Clearly $Y_1, \varepsilon_t, \ldots, \varepsilon_{t+M}$ and Y_2 are all independent.

By (3.12) one can write

$$G_{t+M} = \sum_{\ell=0}^{M} b_{M-\ell} \varepsilon_{t+\ell}. \tag{3.17}$$

Denote by $\underset{\sim}{Y}$ the vector $(Y_1, \varepsilon_t, \ldots, \varepsilon_{t+M}, Y_2)$. Equations (3.16) and (3.17) can be written as

$$\begin{bmatrix} G_{t+M} \\ F_{t+L-N} \end{bmatrix} = A_1 \underset{\sim}{Y}, \tag{3.18}$$

or as

$$\begin{bmatrix} F_{t+L-N} \\ G_{t+M} \end{bmatrix} = A_2 \underset{\sim}{Y}, \tag{3.19}$$

where

$$A_1 = \begin{bmatrix} 0 & b_M & \cdots & b_0 & 0 \\ 1 & v_L & \cdots & v_{L-M} & 1 \end{bmatrix}, \tag{3.20}$$

$$A_2 = \begin{bmatrix} 1 & v_L & \cdots & v_{L-M} & 1 \\ 0 & b_M & \cdots & b_0 & 0 \end{bmatrix}. \tag{3.21}$$

The choice of the index L ensures that the second column of A_1, which is the column multiplying ε_t in (3.18), is not proportional to any column of A_2. Since the vectors (G_{t+M}, F_{t+L-N}) and (F_{t+L-N}, G_{t+M}) have the same bivariate distribution, it follows from Characterisation 3.2 that ε_t is normally distributed.

Chapter 4

THE SHOT NOISE PROCESS

4.1 Definition and Basic Properties

The shot noise process, which was defined in section 2.5 as an example of a linear filtered Poisson process, forms the basis of the models proposed here. The shot noise process was suggested as a model for daily streamflow in an expository paper by Bernier (1970); see also Bartlett (1957). In this Chapter a detailed discussion of properties of the shot noise process is given.

The transient shot noise process is defined by

$$X_0(t) = \sum_{m=N(0)}^{N(t)} Y_m e^{-b(t-\tau_m)}, \tag{4.1}$$

and the stationary shot noise process is defined by

$$X(t) = \sum_{m=-\infty}^{N(t)} Y_m e^{-b(t-\tau_m)}, \tag{4.2}$$

with $\{Y_m\}$ exponentially distributed, i.e.

$$f(y) = 1/\theta \; e^{-y/\theta}. \tag{4.3}$$

The shot noise process has three parameters, namely:

ν - the event rate of the process $\{N(t)\}$;

θ - the average jump height, the expectation of the Y_m;

b - the decay rate, $b > 0$, the parameter of the pulse shape.

The following properties of the shot noise process are easily obtained in any of three ways: (i) by considering $X(t)$ as a linear process; (ii) by using the fact that $X(t)$ has a marginal gamma distribution; or (iii) by using properties of filtered Poisson processes as given in Chapter 2.

(a) $\underline{X(t) \geq 0}$. This obvious property is important if $X(t)$ is used to model streamflow.

(b) The process $X(t)$ can be written as a *linear process* in continuous time,

$$X(t) = \int_{-\infty}^{t} e^{-b(t-u)} d\{Y(u)\}, \tag{4.4}$$

where $Y(t)$ is a compound Poisson process with rate ν and with an associated variable Y, exponentially distributed with mean θ; see section 2.4. The function $e^{-b(t-u)}$ is the instantaneous unit hydrograph of a single linear reservoir with a constant b; see section 2.6. On the other hand, $X(t)$ is a continuous time first order autoregressive process.

(c) Discrete Sampling of $X(t)$. From (4.2), when $X(t)$ is considered at discrete times, $t = \ldots,-1,0,1,\ldots,$

$$X(t) = e^{-b} \sum_{m=-\infty}^{N(t-1)} Y_m e^{-b(t-1-\tau_m)} + \sum_{m=N(t-1)}^{N(t)} Y_m e^{-b(t-\tau_m)}$$

$$= e^{-b} X(t-1) + E(t). \tag{4.5}$$

Thus $X(t)$ can be represented as a discrete time first order autoregressive process. The sequence $E(t)$ of innovation variables is a sequence of independent identically distributed variables, and $E(t)$ is independent of $X(t-s)$ $(s>1)$. Note, by (4.1), the innovation $E(t)$ is identical to the transient process $X_0(t)$ at $t = 1$.

(d) The Autocorrelation Function of $X(t)$ is by (4.4) or (4.5)

$$\rho\{X(t),X(t+s)\} = e^{-bs} \quad (s>0). \tag{4.6}$$

(e) The Distribution of $X(t)$. The characteristic function of $X(t)$ is given in (2.26) as

$$\phi(u;t) = \left(\frac{1}{1-i\theta u}\right)^{\nu/b}. \tag{4.7}$$

Thus $X(t)$ has a gamma distribution, with parameters $1/\theta$ and ν/b, and probability density function

$$f(x) = \frac{(1/\theta)^{\nu/b} x^{\nu/b-1}}{\Gamma(\nu/b)} e^{-x/\theta} \tag{4.8}$$

This distribution, and the Pearson Type III distribution, obtained from it by a simple shift of the origin, have been suggested for fitting streamflow; see Beard (1965).

(f) Let $X(t)$ be a stationary first order autoregressive process, and let $X(t)$ have a gamma distribution. That is, consider

$$X(t) = \rho X(t-1) + \varepsilon(t), \tag{4.9}$$

with $\varepsilon(t)$ independent of $X(t-1)$ and with $\phi_X(u;t) = (1-i\theta u)^{-\alpha}$. The characteristic functions of both sides of (4.9) satisfy

$$\phi_X(u;t) = \phi_X(\rho u;t-1)\phi_\varepsilon(u;t),$$

and hence

$$\phi_\varepsilon(u;t) = \frac{\phi_X(u;t)}{\phi_X(\rho u;t-1)} = \left(\frac{1-i\rho\theta u}{1-i\theta u}\right)^{\alpha}. \tag{4.10}$$

A comparison of (4.10) with (2.25) and (2.26) shows that $X(t)$ is a shot noise process.

(g) <u>Cumulants of $X(t)$</u>. From (2.7),

$$K_n\{X(t)\} = E(Y^n)\int_0^\infty (e^{-b\tau})^n d\tau, \tag{4.11}$$

and by (4.3)

$$E(Y^n) = n!\theta^n. \tag{4.12}$$

Therefore

$$K_n\{X(t)\} = (n-1)!\,\frac{\nu\theta^n}{b}. \tag{4.13}$$

(h) In particular the *mean, variance, skewness* and *kurtosis* of $X(t)$ are, by (4.13),

$$
\begin{aligned}
E\{X(t)\} &= \frac{\nu\theta}{b}, \\
\mathrm{Var}\{X(t)\} &= \frac{\nu\theta^2}{b}, \\
\gamma_1\{X(t)\} &= 2\sqrt{\frac{b}{\nu}}, \\
\gamma_2\{X(t)\} &= 6\,\frac{b}{\nu}.
\end{aligned}
\tag{4.14}
$$

(j) <u>The Joint Distribution of $X(t)$ and $X(t+s)$ $(s>0)$</u>. From (2.2) the joint characteristic function of $X(t)$ and $X(t+s)$ is

$$\phi(u,v;t_1,t_2) = \left\{\frac{1}{1-i\theta(u+ve^{-bs})}\right\}^{\nu/b} \left\{\frac{1-i\theta ve^{-bs}}{1-i\theta v}\right\}^{\nu/b}. \tag{4.15}$$

Note that the last factor is the characteristic function of the innovation in $(t,t+s)$, as in (4.10). Thus (4.15) is another way of writing (4.5). Note also that $\phi(u,v;t_1,t_2)$ is not symmetric in u and v, which is an indication of the time irreversibility of the process.

(k) <u>The Distribution of the Innovation Variable</u>. The characteristic function of the innovation $E(t)$ is

$$\phi(u) = \left(\frac{1-i\rho\theta u}{1-i\theta u}\right)^{\nu/b}, \tag{4.16}$$

where $\rho = e^{-b}$. Writing a binomial expansion of (4.16), one obtains

$$
\begin{aligned}
\left(\frac{1-i\theta\rho u}{1-i\theta u}\right)^{\nu/b} &= \left(\rho+\frac{1-\rho}{1-i\theta u}\right)^{\nu/b} \\
&= \sum_{n=0}^{\infty} \frac{\Gamma(\nu/b+1)}{\Gamma(\nu/b-n+1)\Gamma(n+1)}\,\rho^{\nu/b-n}(1-\rho)^n\left(\frac{1}{1-i\theta u}\right)^n.
\end{aligned}
\tag{4.17}
$$

This can be inverted directly to obtain

$$P\{E(t) = 0\} = \rho^{\nu/b} = e^{-\nu}, \tag{4.18}$$

$$f_{E(t)}(x) = \sum_{n=1}^{\infty} \frac{\Gamma(\nu/b+1)}{\Gamma(\nu/b-n+1)\Gamma(n+1)} \rho^{\nu/b-n}(1-\rho)^n \frac{(1/\theta)^n x^{n-1}}{\Gamma(n)} e^{-x/\theta} \tag{4.18}$$

The summation in (4.18) is over $1,\ldots,\nu/b$ only, when ν/b is an integer. Note that $P\{E(t) = 0\}$ is the probability that no event occurs in $(t-1,t)$. The expansion of (4.16) as a negative binomial series gives a different form of the density (4.18), see Bernier (1970).

4.2 Multisite Shot Noise Processes

To model multisite streamflow, a multivariate process $\underset{\sim}{X}(t)$ is needed. In this section, a process $\underset{\sim}{X}(t) = \{X_1(t),\ldots,X_k(t)\}$ is defined, such that for each site i $(i=1,\ldots,k)$, $X_i(t)$ is a shot noise process with parameters ν_i, θ_i, and b_i. The events of the various sites are all generated from a single Poisson process. Each event of this Poisson process may initiate simultaneous pulses at several of the sites, which may be correlated, and which may be separated by fixed lags.

Let $\{N(t)\}$ be a Poisson process with rate ν and with event times $\ldots 0 \le \tau_1 \le \ldots \le \tau_m \ldots$. With each event, in accordance with the discussion in section 2.3, three random vectors are associated, $\underset{\sim}{I}_m$, $\underset{\sim}{Y}_m$ and $\underset{\sim}{G}_m$. The sequences $\{\underset{\sim}{I}_m\}$, $\{\underset{\sim}{Y}_m\}$ and $\{\underset{\sim}{G}_m\}$, each consist of independent identically distributed vectors, the three sequences are independent, and they are independent of $\{N(t)\}$. The vector $\underset{\sim}{I}_m = (I_{1m},\ldots,I_{km})$ is a vector of zeroes and ones, and $I_{im} = 1$ indicates that a pulse is initiated by the event τ_m at site i. The vector $\underset{\sim}{Y}_m = (Y_{1m},\ldots,Y_{km})$ gives the jump heights at the different sites. The vector $\underset{\sim}{G}_m = (G_{1m},\ldots,G_{km})$ consists of random delays, with G_{im} the delay between the event time τ_m and the beginning of the pulse at site i. The distributions of the three vectors $\underset{\sim}{I}$, $\underset{\sim}{Y}$ and $\underset{\sim}{G}$

are considered below.

The indicators $\underset{\sim}{I}$ are chosen so that

$$P(I_i=1) = \nu_i/\nu \quad (i=1,\ldots,k). \tag{4.19}$$

Thus the events at site i are Poisson events with rate ν_i. By (4.19) it is necessary that $\nu \geq \nu_i$. The joint distribution of $\underset{\sim}{I} = (I_1,\ldots,I_k)$ can be chosen in many ways and the choice used here is described in section 6.3. In particular, the joint distribution gives rise to values of ν_{ij} $(i\neq j, i,j=1,\ldots,k)$, the rates of events common to each pair of sites.

The jumps $\underset{\sim}{Y}$ are chosen with a joint density function $f(y_1,\ldots,y_k)$ such that the marginal densities are

$$f(y_i) = 1/\theta_i e^{-y_i/\theta_i} \quad (i=1,\ldots,k), \tag{4.20}$$

i.e. each Y_i is distributed exponentially with mean θ_i. The particular joint distribution chosen is as follows.

Let $\underset{\sim}{Z} = (Z_1,\ldots,Z_k)$ be a multivariate normal vector, with $Z_i \sim N(0,1)$, and with a correlation matrix $R = \{r_{ij}\}$. Let $\underset{\sim}{Z}_1$ and $\underset{\sim}{Z}_2$ be independent, and each of them distributed as $\underset{\sim}{Z}$. The random vector $\underset{\sim}{Y}$ is defined by

$$Y_i = \tfrac{1}{2}\theta_i(Z_{i1}^2+Z_{i2}^2) \quad (i=1,\ldots,k). \tag{4.21}$$

With this definition of $\underset{\sim}{Y}$, the joint characteristic function of Y_i and Y_j $(i\neq j, i,j=1,\ldots,k)$, is

$$\phi(u,v) = \frac{1}{1-iu\theta_i-iv\theta_j-(1-r_{ij}^2)u\theta_i v\theta_j}, \tag{4.22}$$

and the correlation coefficient between Y_i and Y_j is

$$C_{ij} = \rho(Y_i,Y_j) = r_{ij}^2. \tag{4.23}$$

The delay variables $\underset{\sim}{G} = (G_1,\ldots,G_k)$ can be chosen to give quite general models. Because of the difficulty of verifying and of estimating such models a very simplified form was chosen, i.e. the various G_i ($i=1,..,k$) are constant. Thus the pulses at the different sites are assumed to have only fixed lags between them. When these lags are given, one can shift the time scales at the different sites, and so without loss of generality $G_i = 0$ ($i=1,\dots,k$) is chosen.

An event of $\{N(t)\}$ at time τ_m gives rise to pulses at the various sites given by

$$w_i(t,\tau_m,I_{im},Y_{im}) = \begin{cases} Y_{im}e^{-b_i(t-\tau_m)} & (I_{im}=1), \\ 0 & (I_{im}=0); \end{cases} \tag{4.24}$$

where $t > \tau_m$ and $i = 1,\dots,k$.

By (2.15), the cross-covariance between $X_i(t)$ and $X_j(t+s)$ $(s>0)$ is

$$\begin{aligned} \mathrm{cov}\{X_i(t),X_j(t+s)\} &= \nu_{ij}E(Y_iY_j)\int_0^\infty e^{-b_it}e^{-b_j(t+s)}dt \\ &= \nu_{ij}E(Y_iY_j)\frac{1}{b_i+b_j}e^{-b_js} \quad (s>0). \end{aligned} \tag{4.25}$$

Using (4.14) for the variance of $X_i(t)$ and $X_j(t+s)$, one obtains the cross correlation

$$\begin{aligned} \rho\{X_i(t),X_j(t+s)\} &= \frac{\nu_{ij}}{\sqrt{\nu_i\nu_j}}\,\frac{E(Y_iY_j)}{\sqrt{E(Y_i^2)E(Y_j^2)}}\,\frac{2\sqrt{b_ib_j}}{b_i+b_j}e^{-b_js} \\ &= \frac{\nu_{ij}}{\sqrt{\nu_i\nu_j}}\,\tfrac{1}{2}(C_{ij}+1)\,\frac{2\sqrt{b_ib_j}}{b_i+b_j}e^{-b_js} \quad (s>0). \end{aligned} \tag{4.26}$$

The different factors on the right hand side of (4.26) have the following interpretation. The first expresses the proportion of simultaneous events at both sites. The second arises from the cross correlation between jumps at simultaneous events. Thus these two factors express the dependence between the inputs at the two sites. The third factor expresses the dependence arising from similarity in the pulse shapes, i.e. the system responses, at the two sites. Thus the first three factors give the lag zero cross correlation between the two sites. The final factor is the serial correlation at the second site over the interval (t,t+s).

4.3 Averaging of the Shot Noise Process

In section 1.4 it was noted that daily streamflow data consist of daily averages of the continuous streamflow process. The choice between the assumption that daily data are sampled instantaneous values, and the assumption that they are averages over a day, makes a substantial difference. This is particularly so when the serial correlation is high; see section 7.6 for a numerical example. In this section, the process obtained by averaging the shot noise process X(t) over a fixed time unit of length T is discussed. The averaged process is thus defined by

$$Z(t) = \frac{1}{T}\int_{t-T}^{t} X(\tau)d\tau, \tag{4.28}$$

where X(t) is a shot noise process with parameters ν, θ and b.

As in (4.5), X(t) can be written as

$$X(t) = e^{-bT}\sum_{m=-\infty}^{N(t-T)} Y_m e^{-b(t-T-\tau_m)} + \sum_{m=N(t-T)}^{N(t)} Y_m e^{-b(t-\tau_m)}.$$

On performing the integration in (4.28), Z(t) is obtained as

$$Z(t) = \frac{1}{bT}(1-e^{-bT})\sum_{m=-\infty}^{N(t-T)} Y_m e^{-b(t-T-\tau_m)} + \frac{1}{bT}\sum_{m=N(t-T)}^{N(t)} Y_m\{1-e^{-b(t-\tau_m)}\}. \tag{4.29}$$

Hence Z(t) is itself a linear filtered Poisson process, with pulse shape given by

$$w(t) = \begin{cases} \frac{1}{bT}(1-e^{-bt}) & (0\le t<T), \\ \frac{1}{bT}(1-e^{-bT})e^{-b(t-T)} & (T\le t). \end{cases} \tag{4.30}$$

The following properties of Z(t) are obtained directly by using the equations of Chapter 2. Equation (4.30) and most of the following discussion can also be derived by considering X(t) as a continuous time linear process as given by (4.4), and applying (4.28).

(a) The Characteristic Function of Z(t). Using (2.3), one obtains

$$\phi(u;t) = \exp[\nu\int_0^\infty E\{e^{iuYw(t)}-1\}dt]$$
$$= \exp[\nu\int_0^T\{\frac{1}{1-\frac{i\theta u}{bT}(1-e^{-bt})}-1\}dt+\nu\int_T^\infty\{\frac{1}{1-\frac{i\theta u}{bT}(1-e^{-bT})e^{-b(t-T)}}-1\}d$$

Thus

$$\phi(u;t) = [e^{-bT}\{1-\frac{i\theta u}{bT}(1-e^{-bT})\}]^{-\frac{\nu}{b}\frac{i\theta u/bT}{1-i\theta u/bT}} \tag{4.31}$$

(b) Cumulants of the Averaged Shot Noise.

To calculate the cumulants of Z(t), (2.7) is used, which involves the integral

$$\int_0^\infty\{w(t)\}^n dt = \frac{1}{(bT)^n}\{\int_0^T(1-e^{-bt})^n dt+(1-e^{-bT})^n\int_T^\infty e^{-nb(t-T)}dt\}$$

$$= \frac{1}{b}\frac{1}{(bT)^n}\{bT-\sum_{k=1}^{n-1}\frac{1}{k}(1-e^{-bT})^k\}.$$

Thus the n^{th} cumulant of Z(t) is, by (2.7) and (4.12),

$$K_n\{Z(t)\} = (n-1)!\frac{\nu\theta^n}{b}\frac{n}{(bT)^n}\{bT-\sum_{k=1}^{n-1}\frac{1}{k}(1-e^{-bT})^k\} \tag{4.32}$$

In particular, the mean, variance and skewness of Z(t) are given by

$$E\{Z(t)\} = \frac{\nu\theta}{b},$$

$$\mathrm{var}\{Z(t)\} = \frac{\nu\theta^2}{b}\frac{2\{bT-(1-e^{-bT})\}}{(bT)^2}, \tag{4.33}$$

$$\gamma_1\{Z(t)\} = \sqrt{\frac{b}{\nu}}\,\frac{3\{bT-(1-e^{-bT})-\frac{1}{2}(1-e^{-bT})^2\}}{\sqrt{2\{bT-(1-e^{-bT})\}^3}}\,.$$

As expected, the skewness of $Z(t)$ is smaller than that of $X(t)$, given in (4.14).

(c) The Autocovariance and Autocorrelation Function of $Z(t)$.

These are calculated by the use of (2.6), and involve the integral $\int_0^\infty w(t)w(t+s)dt$, with $w(t)$ the pulse shape given by (4.30). By straightforward integration

$$\mathrm{cov}\{Z(t),Z(t+s)\} = \begin{cases} \dfrac{\nu\theta^2}{b}\,\dfrac{1}{(bT)^2}\,\{2b(T-s)+e^{-b(T+s)}+e^{-b(T-s)}-2e^{-bs}\} & (0\leq s\leq T), \\[2ex] \dfrac{\nu\theta^2}{b}\,\dfrac{1}{(bT)^2}\,(1-e^{-bT})^2 e^{-b(s-T)} & (s>T). \end{cases} \qquad (4.34)$$

Hence, for $s > T$,

$$\rho\{Z(t),Z(t+s)\} = \frac{(1-e^{-bT})^2}{2\{bT-(1-e^{-bT})\}}\,e^{-b(s-T)} \qquad (s>T). \qquad (4.35)$$

As expected, because $Z(t)$ is obtained from $X(t)$ by averaging over T, $\rho\{Z(t),Z(t+s)\}$ decreases like e^{-bs}, for $s > T$.

(d) Averaging of Multisite Shot Noise Processes

Let $\underset{\sim}{X}(t) = \{X_1(t),\ldots,X_k(t)\}$ be a multisite shot noise process, as described in section 4.2. The averaged multisite process $\underset{\sim}{Z}(t)$ is obtained from $\underset{\sim}{X}(t)$ by defining $Z_i(t)$ $(i=1,\ldots,k)$, as the average of $X_i(t)$ over a period T. The process $\underset{\sim}{Z}(t)$ is a multivariate filtered Poisson process, with the same multivariate input as $\underset{\sim}{X}(t)$, and with pulse shapes $w_i(t)$ $(i=1,\ldots,k)$, given by (4.30), with decay rates b_i.

The cross correlation between $Z_i(t)$ and $Z_j(t)$ is composed of a factor expressing the dependence between the inputs of sites i and j, and a factor expressing the dependence due to similarity between $w_i(t)$ and $w_j(t)$. The first factor is the same as for $X_i(t)$ and $X_j(t)$, the non averaged processes. The second factor is $\int_0^\infty w_i(t)w_j(t)dt$, and its calculation is straightforward. If $\rho\{X_i(t),X_j(t)\}$ denotes the cross correlation of the non averaged process (see (4.26)), then

$$\rho\{Z_i(t),Z_j(t)\} = \rho\{X_i(t),X_j(t)\}\,\frac{d_i+d_j}{2\sqrt{d_i d_j}}\,, \tag{4.36}$$

where

$$d_i = \frac{2\{b_iT-(1-e^{-b_iT})\}}{(b_iT)^2}$$

The cross correlation between $Z_i(t)$ and $Z_j(t+s)$ $(s>0)$ is obtained similarly, by calculating $\int_0^\infty w_i(t)w_j(t+s)dt$. For $s > T$ one obtains

$$\rho\{Z_i(t),Z_j(t+s)\} = \rho\{X_i(t),X_j(t)\}\rho\{Z_j(t),Z_j(t+s)\}\sqrt{\frac{d_j}{d_i}} \tag{4.37}$$

where $\rho\{Z_j(t),Z_j(t+s)\}$ is the serial correlation of the averaged process $Z_j(t)$ at site j, given by (4.35)

4.4 Shot Noise Processes with Periodic Structure

Streamflow is subject to seasonal changes. Thus, to model streamflow, a nonstationary process, in fact a process with periodic structure with a period of one year, is required. In this section, shot noise processes whose parameters change periodically with a period T of one year are studied.

In general let the parameters of a shot noise process vary with time, that is let $\nu = \nu(t)$, $\theta = \theta(t)$ and $b = b(t)$. To obtain convergence, it is assumed that $\nu(t)$ and $\theta(t)$ are bounded, and $b(t)$ is bounded away from zero. By applying the results of section 2.2, the characteristic function of $X(t)$ is

$$\phi(u;t) = \exp\left[\int_{-\infty}^{t} \nu(\tau)\left\{\frac{1}{1-i\theta(\tau)ue^{-B(\tau,t)}}-1\right\}d\tau\right], \tag{4.38}$$

with

$$B(\tau,t) = \int_{\tau}^{t} b(\sigma)d\sigma.$$

A shot noise process with periodic structure is obtained when $\nu(t)$, $\Theta(t)$ and $b(t)$ are periodic with period T, in which case $\phi(u;t)$ and $X(t)$ will have period T.

In modelling streamflow by a shot noise process, it is assumed that the sequences $\{\tau_m\}$ and $\{Y_m\}$ of events and jumps form the random input to a linear system with impulse response e^{-bt}. It may be assumed that only the random input is subject to seasonal changes, while the system, and in particular the value of b, are time invariant. This case is investigated below.

First assume that b and Θ are constant, and that

$$\nu(t) = \nu\{1+\cos(\phi t-\Psi)\}, \tag{4.39}$$

with $\phi = \frac{2\pi}{T}$. Then the mean of $X(t)$ is calculated by (4.38),

$$\begin{aligned} E\{X(t)\} &= \int_{-\infty}^{t} \nu\{1+\cos(\phi t-\Psi)\}\ \Theta\ e^{-b(t-\tau)}d\tau \\ &= \frac{\nu\Theta}{b}\ \{1+\frac{b^2}{b^2+\phi^2}\cos(\phi t-\Psi) + \frac{b\phi}{b^2+\phi^2}\sin(\phi t-\Psi)\}. \end{aligned} \tag{4.40}$$

Therefore defining $\chi = \arctan\frac{\phi}{b}$, one obtains

$$m(t) = E\{X(t)\} = \frac{\nu\Theta}{b}\{1+\cos\chi\cos(\phi t-\Psi-\chi)\}. \tag{4.41}$$

Thus the mean has also a single cosine function of period T, with a phase lag between $m(t)$ and $\nu(t)$, as expected, because of the linearity of the system.

Assume now that b is constant, and that $\nu(t)$ and $\Theta(t)$ are periodic and can be written as, or can at least be approximated by,

$$\begin{aligned} \nu(t) &= \sum_{k=0}^{N} \nu_{1k}\cos k\phi t + \nu_{2k}\sin k\phi t, \\ \Theta(t) &= \sum^{N} \Theta_{1k}\cos k\phi t + \Theta_{2k}\sin k\phi t. \end{aligned} \tag{4.42}$$

It is then possible to write

$$\nu(t)\Theta(t) = \sum_{k=0}^{2N} c_k \cos(k\phi t-\Psi_k),$$
$$\nu(t)\Theta^2(t) = \sum_{k=0}^{3N} d_k \cos(k\phi t-\eta_k). \tag{4.43}$$

The expectation and the variance of the process, $m(t)$ and $\sigma^2(t)$, are given by

$$m(t) = \int_{-\infty}^{t} \nu(\tau)\Theta(\tau)e^{-b(t-\tau)}d\tau,$$
$$\sigma^2(t) = 2\int_{-\infty}^{t} \nu(\tau)\Theta^2(\tau)e^{-2b(t-\tau)}d\tau. \tag{4.44}$$

Substituting (4.43) into (4.44), integrals similar to (4.40) are obtained. It follows that

$$m(t) = \sum_{k=0}^{2N} \frac{c_k}{b}\cos\chi_k \cos(\phi t-\Psi_k-\chi_k),$$
$$\sigma^2(t) = \sum_{k=0}^{3N} \frac{d_k}{b}\cos\zeta_k \cos(\phi t-\eta_k-\zeta_k), \tag{4.45}$$

where

$$\chi_k = \arctan\frac{k\phi}{b}, \quad \zeta_k = \arctan\frac{k\phi}{2b}.$$

The simplicity of the relations between $\nu(t)$, $\Theta(t)$ and $m(t)$, $\sigma^2(t)$, expressed by (4.42), (4.43) and (4.45), lies in the fact that because of the linearity of the system, the phase lags χ_k and ζ_k, do not depend on the Fourier coefficients ν_{1k}, ν_{2k}, Θ_{1k} and Θ_{2k}, but only on T and b. Thus it is easy to solve for $\nu(t)$ and $\Theta(t)$ when $m(t)$ and $\sigma^2(t)$ are known.

The periodic structure of $\nu(t)$ and $\Theta(t)$ does not affect the serial correlation of $X(t)$, as is seen by

$$\begin{aligned}\mathrm{cov}\{X(t),X(t+s)\} &= \int_{-\infty}^{t} 2\nu(\tau)\Theta^2(\tau)e^{-b(t-\tau)}e^{-b(t+s-\tau)}d\tau \\ &= \sigma^2(t)e^{-bs}.\end{aligned} \tag{4.46}$$

The application of (4.46) presents a difficulty because it is not clear how $\mathrm{cov}\{X(t),X(t+s)\}$ should be estimated. A simple approach is to consider $\{X(t)-m(t)\}/\sigma(t)$ where $m(t)$ and $\sigma(t)$ are estimated first. Whether this approach is appropriate for the shot noise process or whether a better approach exists is unknown. This difficulty is encountered in section 6.5.

It is interesting to compare these results with those for a Gaussian linear process with a stationary system and a periodic input. In the Gaussian case the input is assumed to be $\sigma(t)dY(t) + m(t)$, where $dY(t)$ is stationary white noise, and $\sigma(t)$ and $m(t)$ are periodic. The periodicity of both $m(t)$ and $\sigma(t)$ affects the magnitude scale of the input, and is thus similar to the periodicity of $\Theta(t)$. There is no parallel to the periodicity of $\nu(t)$ which affects the time scale.

The extension of the result (4.45) to $Z(t)$, the process obtained from $X(t)$ by averaging over a period of length 1, is laborious. Approximate results, holding when $T \gg 1$, are obtained as follows:

$$E\{Z(t)\} = \int_{t-1}^{t} m(\tau)d\tau \simeq m(t); \tag{4.47}$$

$$\begin{aligned} \mathrm{var}\{Z(t)\} &= \int_{t-1}^{t}\int_{t-1}^{t} \mathrm{cov}\{X(\tau_1)X(\tau_2)\}d\tau_1 d\tau_2 \\ &\simeq \int_{t-1}^{t}\int_{t-1}^{t} \sigma(\tau_1)\sigma(\tau_2)e^{-b|\tau_1-\tau_2|}d\tau_1 d\tau_2 \\ &\simeq \sigma^2(t).\ \frac{2\{b-(1-e^{-b})\}}{b^2}, \end{aligned} \tag{4.48}$$

where $m(t)$ and $\sigma^2(t)$ are the periodic mean and variance of $X(t)$. Note the similarity between (4.48) and (4.33). In a similar way, corresponding to (4.46), one has approximately

$$\rho\{Z(t),Z(t+s)\} \simeq \frac{(1-e^{-b})^2}{2\{b-(1-e^{-b})\}}\ e^{-b(s-1)} \tag{4.49}$$

Finally, consider the case where $b = b(t)$ is periodic with period T, as well as $\nu(t)$ and $\Theta(t)$. In this case expressions for the properties of X(t) are difficult to obtain. One has however

$$\mathrm{cov}\{X(t),X(t+s)\} = \int_{-\infty}^{t} 2\nu(\tau)\Theta^2(\tau)e^{-B(\tau,t)}e^{-B(\tau,t+s)}d\tau$$

where

$$B(\tau,t) = \int_{\tau}^{t} b(s)ds.$$

Hence,

$$\mathrm{cov}\{X(t),X(t+s)\} = \mathrm{var}\{X(t)\}e^{-\int_t^{t+s} b(\tau)d\tau}.$$

Thus the serial correlation of X(t) is given by

$$\rho\{X(t),X(t+s)\} = \sqrt{\frac{\mathrm{var}\{X(t)\}}{\mathrm{var}\{X(t+s)\}}}\, e^{-\int_t^{t+s} b(\tau)d\tau}. \tag{4.50}$$

4.5 A Relation Between Instantaneous and Averaged Shot Noise

Let streamflow be modelled by a shot noise process X(t). In section 6.3 the properties of averages of X(t) were derived, to enable the study of daily and monthly data. It is of interest to study the behaviour of continuous streamflow and of daily data when monthly averages are given, and this behaviour is studied in this section.

Let X(t) be a shot noise process with parameters ν, Θ and b, and let

$$Z(t) = \int_{t-T}^{t} X(\tau)d\tau. \tag{4.51}$$

Define a process W(t) by

$$W(t) = Z(t) - \frac{1}{b} X(t-T) + \frac{1}{b} X(t). \tag{4.52}$$

It is easily seen that

$$W(t) = \frac{1}{b} \sum_{m=N(t-T)}^{N(t)} Y_m, \tag{4.53}$$

i.e. $W(t)$ is the sum of the total volume of all the pulses which occur in $(t-T,t)$. For large T, $W(t)$ is approximately equal to $Z(t)$.

The following discussion gives the behaviour of $X(t)$ in the interval $(t-T,t)$ conditional on the value of $W(t)$. The values of $X(t)$ in $(t-T,t)$ are completely determined by the values of the event times $t - T \leq \tau_1 \leq \ldots \tau_N \leq t$ and the jumps $Y_1, \ldots, Y_N$ in $(t-T,t)$, where N is the random number of events in $(t-T,t)$.

The distribution of N when $W(t) = w$ is given, is obtained first. The relations between N, T and W are summed up by

$$P(N=n|T) = \frac{(\nu T)^n}{n!} e^{-\nu T} \tag{4.54}$$

because the events form a Poisson process, and by

$$f_{W(t)}(w|N=n) = \frac{(\frac{b}{\theta})^n w^{n-1}}{(n-1)!} e^{-\frac{b}{\theta}w} \tag{4.55}$$

because $W(t)$ is given by (4.53), where Y_m are exponentially distributed.

Simple manipulations of conditional probabilities yield, for $W(t) = w \neq 0$

$$\begin{aligned} P(N=n|w,T) &= \frac{P\{N=n, W(t)=w|T\}}{P\{W(t)=w|T\}} \\ &= \frac{P\{W(t)=w|N=n,T\}P(N=n|T)}{\sum_{r=1}^{\infty} P(W(t)=w|N=r,T\}P(N=r|T)}, \end{aligned} \tag{4.56}$$

where $P\{W(t)=w|N=n,T\}$ is in fact a density function, and is given by (4.55). Substituting (4.54) and (4.55) in (4.56), one obtains

$$P(N=n|w,T) = \frac{\frac{(\frac{b}{\theta})^n w^{n-1}}{(n-1)!} e^{-\frac{b}{\theta}w} \frac{(\nu T)^n}{n!} e^{-\nu T}}{\sum_{r=1}^{\infty} \frac{(\frac{b}{\theta})^r w^{r-1}}{(r-1)!} e^{-\frac{b}{\theta}w} \frac{(\nu T)}{r!} e^{-\nu T}} . \tag{4.57}$$

Let $\lambda = \frac{\nu b}{\theta} w T$; then (4.57) can be written as

$$P(N=n|w,T) = \frac{\lambda^{n-\frac{1}{2}}}{n!(n-1)!} \cdot \{I_1(2\sqrt{\lambda})\}^{-1}, \tag{4.58}$$

where I_1 is the modified Bessel function of order 1, defined by

$$I_1(x) = \sum_{k=1}^{\infty} \frac{\{(\frac{1}{2}x)^2\}^{k-\frac{1}{2}}}{(k-1)!k!} .$$

Note that the assumption that the jumps $\{Y_m\}$ are exponentially distributed is essential here, in obtaining (4.55).

As mentioned before, the conditional behaviour of $X(t)$ given $W(t) = w$ is determined by the density $P(N=n,\tau_1,..,\tau_n,Y_1,...,Y_n|w,T)$. This density can be written as

$$P(N=n,\tau_1,...\tau_n,Y_1,...Y_n|w,T) = P(N=n|w,T) \cdot f(\tau_1,...\tau_n|n,T) f(Y_1,...,Y_n|n,w), \tag{4.59}$$

where the first factor is given by (4.58).

The density $f(\tau_1,...\tau_n|n,T)$ is simply the density of an ordered sample of size n from a uniform distribution over $(0,T)$. For the third factor, the density $f(Y_1,...,Y_n|n,w)$, consider

$$u_k = \frac{1}{b} \sum_{i=1}^{k} Y_i, \ (k=1,...,n). \tag{4.60}$$

By the assumption that Y_i are exponentially distributed, $u_1,...,u_{n-1}$ form an ordered sample of size (n-1) from a uniform distribution over $(0,w)$. These results completely determine the behaviour of $X(t)$ in a period of length T, conditional on w.

Let now $Z_1,...,Z_M$ be accumulated flows over M consecutive periods of length T. Synthetic values of the shot noise process

$X(t)$ over $(0,MT)$ can be generated so that the accumulated values of $X(t)$ over the M periods are approximately $Z_1,\ldots Z_m$, by assuming that $W_k = Z_k$, $(k=1,\ldots,M)$.

The problem of only approximating $Z_1,\ldots,Z_M$ is not alleviated when in addition to $Z_1,\ldots Z_M$, the flows at the beginning and end of each period, $X(0), X(T),\ldots X(MT)$, are given. Although in this case W_k can be found exactly, from (4.52), the distribution of $n, \tau_1,\ldots,\tau_n$ and $Y_1,\ldots,Y_n$ in the kth period will now depend on W_k and on $X\{(k-1)T\}$ and $X\{kT\}$.

Chapter 5

PROPERTIES OF SAMPLE PATHS OF SHOT NOISE PROCESSES

5.1 General Comments

In this Chapter, sample path properties of the shot noise process are discussed. These include distributions of maxima and minima of the process, and the times which the process spends above, below or between two levels. These properties are important in the study of streamflow becasue they characterise flood peaks and low flow period behaviour. The results which are obtained are, however, inapplicable to streamflow data, for three reasons. Firstly, the results are for the continuous process while the data consist of averaged values. Secondly, the results apply to the stationary shot noise process while streamflow is at best of periodic structure, and thirdly, the actual model employed for streamflow consists of a sum of two shot noise processes. More research may, however, lead to a way of using the results of this Chapter in these more complex situations.

Sample path properties are discussed by Cramér and Leadbetter (1967, Chapter 10), where some exact and some asymptotic results are obtained for Gaussian processes whose autocovariance functions have a continuous second derivative at zero. These results are not applicable to first order autoregressive processes, because their autocovariance functions do not have a continuous derivative at zero. Sample path properties of the continuous time Gaussian autoregressive process, also referred to in the literature as the Ornstein Uhlenbeck process, and of other diffusion processes, can be derived from diffusion equations as described by Darling and Siegert (1953) and by

Cox and Miller (1965, section 5.10). Both types of results are not applicable to the shot noise process, because sample paths of the shot noise process are discontinuous.

5.2 Maxima and Minima of the Shot Noise Process

Let X(t) be a shot noise process with parameters ν, θ and b. The paths of X(t) are, with probability one, continuous except for a countable number of points, i.e. the event times $\ldots,\tau_1,\ldots,\tau_m\ldots$, at which the limit from the left $X(\tau_m^-)$ and the limit from the right $X(\tau_m^+)$ exist and satisfy $X(\tau_m^+) > X(\tau_m^-)$. Let $\{L_m\}$ and $\{H_m\}$ denote the sequences $\{X(\tau_m^-)\}$ and $\{X(\tau_m^+)\}$, i.e. the heights just before and just after jumps, respectively. In this section the properties of $\{L_m\}$ and $\{H_m\}$ are studied.

For any fixed arbitrary t, there exists an m such that $\tau_m \le t < \tau_{m+1}$, and thus $H_m \ge X(t) > L_{m+1}$. It seems therefore that the maxima H, the minima L and the variable X(t) are stochastically ordered by $H \ge X(t) \ge L$. In fact L is distributed identically to X(t), i.e. $L \sim \gamma(1/\theta,\nu/b)$, while $H \sim \gamma(1/\theta,\nu/b+1)$.

To show this, consider t and τ_m. The time interval $T = t - \tau_m$ is exponentially distributed, with mean $1/\nu$; see Feller (1971, section 1.4) for a discussion of this point. The random variable T is independent of H_m. Given that $H_m = h$, for any value of T, $X(t) = he^{-bT}$. Thus X(t) = x is equivalent to $T = \frac{1}{b}\log\frac{h}{x}$. Hence the conditional density of X(t) given $H_m = h$ is

$$f_{X(t)|H}(x|h) = f_{X(t)|H}(he^{-bT}|h)$$

$$= \frac{1}{bx}f_T(\frac{1}{b}\log\frac{h}{x})$$

$$= \frac{\nu}{bx}\, e^{-\{\nu\frac{1}{b}\log\frac{h}{x}\}}$$

or

$$f_{X(t)|H}(x|h) = \frac{\nu}{b} \frac{x^{\nu/b-1}}{h^{\nu/b}} \qquad (0 \le x \le h). \tag{5.1}$$

Thus the unconditional density of X(t) is

$$f_{X(t)}(x) = \int_x^{\infty} \frac{\nu}{b} \frac{x^{\nu/b-1}}{h^{\nu/b}} f_H(h)\,dh, \tag{5.2}$$

where $f_H(h)$ is the density of H. However, by (4.8), $X(t) \sim \gamma(1/\Theta, \nu/b)$; therefore

$$\frac{(1/\Theta)^{\nu/b} x^{\nu/b-1}}{\Gamma(\nu/b)} e^{-x/\Theta} = \int_x^{\infty} \frac{\nu}{b} \frac{x^{\nu/b-1}}{h^{\nu/b}} f_H(h)\,dh$$

which is solved uniquely by

$$f_H(h) = \frac{(1/\Theta)^{\nu/b+1} h^{\nu/b}}{\Gamma(\nu/b+1)} e^{-h/\Theta}. \tag{5.3}$$

For every event τ_m, $H_m = L_m + Y$, where Y is the jump at τ_m, which is independent of L_m and exponentially distributed with mean Θ. Thus the condition

$$f_H(h) = \int_0^h \frac{1}{\Theta} e^{-(h-\ell)/\Theta} f_L(\ell)\,d\ell \tag{5.4}$$

yields an equation for $f_L(\ell)$, the probability density of L, which is solved by

$$f_L(\ell) = \frac{(1/\Theta)^{\nu/b} \ell^{\nu/b-1}}{\Gamma(\nu/b)} e^{-\ell/\Theta}. \tag{5.5}$$

The fact that L is distributed identically to X(t) rather than say $L \le X(t) < H$ or $X(t) \sim H$ is explained intuitively as follows. The variable $L = X(\tau_m^-)$ is determined only by events previous to τ_m, and the backward times to these events from τ_m are distributed identically to the backward times from any fixed t. Thus, $X(\tau_m^-)$ and X(t) are identically distributed for every stationary Poisson process with $w(t-\tau,Y) = 0$ for $\tau > t$. A rigorous proof of that statement can be obtained by following the steps of the proof of theorem 2.1 for $X(\tau_m^-)$.

Consider again the shot noise process. It is clear that the sequences $\{H_m\}$, $\{L_m\}$, and the combined sequence $\ldots L_1, H_1, \ldots, L_m, H_m, \ldots$ are Markovian. By (5.4), the conditional distribution of H_m given L_m is

$$f_{H_m|L_m}(h|\ell) = 1/\Theta e^{-(h-\ell)/\Theta} \quad (h \geq \ell) \tag{5.6}$$

and, by (5.1), the conditional distribution of L_{m+1} given H_m is

$$f_{L_{m+1}|H_m}(\ell|h) = \nu/b \; \frac{\ell^{\nu/b-1}}{h^{\nu/b}} \qquad (0 \leq \ell \leq h). \tag{5.7}$$

Equations (5.3), (5.5) (5.6) and (5.7), together with the Markovian property, completely characterise the behaviour of the maxima and minima of the shot noise process.

5.3 Upcrossings and Downcrossings of the Shot Noise Process

In section 3.2 upcrossings and downcrossings of a fixed level w by a sample path of a process X(t) were defined. If X(t) is a shot noise process, then an upcrossing at time t occurs if t is an event time with $X(t^-) < w$ and $X(t^+) > w$. A downcrossing at time s occurs if $X(s) = w$ and the process is continuous at s.

Let s be a downcrossing, i.e. $X(s) = w$. By the Markovian property, $X(t_1)$ $(t_1 > s)$ is independent of $X(t_2)$ $(t_2 < s)$. Let t be an upcrossing, then $X(t^+) > w$ and by the assumption that jumps are exponentially distributed, $\{X(t^+) - w\}$ is itself exponentially distributed. Hence by the Markovian property $X(t_1)$ $(t_1 > t)$ is independent of $X(t_2)$ $(t_2 < t)$.

Denote the sequence of downcrossings by $\ldots s_1 < \ldots < s_m \ldots$, and the sequence of upcrossings by $\ldots t_1 < \ldots < t_m \ldots$, so that

$$\ldots t_1 < s_1 < \ldots < t_m < s_m \ldots .$$

From the previous considerations $\{t_m\}$ and $\{s_m\}$ form a renewal process, and $\ldots . t_1, s_1, \ldots, t_m, s_m \ldots .$ form an alternating renewal process.

The renewal density of the upcrossing times, $\eta(t)$, is calculated as follows. Observe $X(t)$ in an interval $(t,t+dt)$. The probability that an event occurs during that interval is νdt. This event leads to an upcrossing if $X(t) < w$ and if the jump Y satisfies $X(t) + Y > w$. This happens with probability

$$\begin{aligned} P &= \int_0^w f_X(x) e^{-(w-x)/\theta} dx \\ &= \int_0^w \frac{(\frac{1}{\theta})^{\nu/b} x^{\nu/b-1}}{\Gamma(\nu/b)} e^{-x/\theta} e^{-(w-x)/\theta} dx \\ &= \frac{(\frac{1}{\theta})^{\nu/b} w^{\nu/b}}{\Gamma(\nu/b+1)} e^{-w/\theta}. \end{aligned}$$

Hence the renewal density $\eta(t)$ is

$$\eta(t)dt = \nu \frac{(\frac{1}{\theta})^{\nu/b} w^{\nu/b}}{\Gamma(\nu/b+1)} e^{-w/\theta} dt . \tag{5.8}$$

The expected time between renewals is $1/\eta(t)$ (Cox, 1962, sections 2.1 and 4.4). Thus the expected time between upcrossings is

$$E(t_m - t_{m-1}) = \{\nu \frac{(1/\theta)^{\nu/b} w^{\nu/b}}{\Gamma(\nu/b+1)} e^{-w/\theta}\}^{-1} \tag{5.9}$$

By (5.3) and (5.5), this can be written as

$$E(t_m - t_{m-1}) = \frac{1}{\nu\{P(H>w) - P(L>w)\}} , \tag{5.10}$$

where H and L are the maxima and minima variables of X(t). Furthermore, since $\ldots t_1, s_1, \ldots, t_m, s_m, \ldots$ form an alternating renewal process, $E(s_m - s_{m-1})$ is equal to $E(t_m - t_{m-1})$, and so

$$E(s_m - s_{m-1}) = \frac{1}{\nu\{P(H>w) - P(L>w)\}} . \tag{5.11}$$

Introduce the notation

$$\begin{aligned} U(w) &= E(s_m - t_m), \\ V(w) &= E(t_{m+1} - s_m). \end{aligned} \tag{5.12}$$

In the interval (t_m, s_m), the process X(t) is always above w, while in (s_m, t_{m+1}), X(t) is always below w. Thus U(w) is the expected length of an excursion of X(t) above w, while V(w) is the expected length of an excursion of X(t) below w. Consider all the time spent by the process above w and all the time spent below w, then

$$\frac{U(w)}{V(w)} = \frac{P\{X(t)>w\}}{P\{X(t)<w\}} . \tag{5.13}$$

On the other hand, from (5.10)

$$U(w) + V(w) = \frac{1}{\nu\{P(H>w) - P(L>w)\}} . \tag{5.14}$$

The relations (5.13) and (5.14) thus yield

$$U(w) = \frac{P\{X(t)>w\}}{\nu\{P(H>w) - P(L>w)\}} , \tag{5.15}$$

$$V(w) = \frac{P\{X(t)<w\}}{\nu\{P(H>w) - P(L>w)\}} . \tag{5.16}$$

Instead of calculating $\eta(t)$ as the renewal density between upcrossings, (5.8) can be obtained by calculating the renewal density of the times between downcrossings. This is a slightly more general way of obtaining $\eta(t)$, because the downcrossings form a renewal process even when the jumps Y are not exponentially distributed. In that case one obtains

$$\eta(t)dt = \nu w f_x(w)(b/\nu)dt \quad .$$

5.4 Crossing Times Between Two Levels

In section 3.2, crossing times between two levels were defined. In the present section, the expected values of these times and of some related variables for the shot noise process are discussed. Let w_1 and w_2 $(w_2 > w_1)$, be two fixed levels.

(a) Let t_1 be a downcrossing of w_2 and let t_2, $t_2 > t_1$, be the first downcrossing of w_1 following t_1. That is $t_2 - t_1$ is a first passage time from a downcrossing of w_2, to w_1. Denote its expectation by $R(w_2,w_1) = E(t_2-t_1)$.

To calculate $R(w_2,w_1)$, let t be the first crossing of $w_1 + \Delta$ following t_1, so that $t_1 < t < t_2$. Clearly

$$R(w_2,w_1) = R(w_2,w_1+\Delta) + R(w_1+\Delta,w_1). \tag{5.17}$$

The passage time from $w_1 + \Delta$ to w_1 is at least

$$dt = \frac{1}{b}\log\left(\frac{w_1+\Delta}{w_1}\right) \quad .$$

If an event occurs during $(t,t+dt)$, then an expected time $U(w_1) + o(\Delta)$ will elapse before the process downcrosses $w_1 + \Delta$ again. An event during $(t,t+dt)$ will happen with probability $\nu dt + o(dt)$. Thus,

$$R(w_2,w_1) = R(w_2,w_1+\Delta) + \frac{1}{b}\log\frac{w_1+\Delta}{w_1}\{1+\nu U(w_1)\} + o(\Delta). \tag{5.18}$$

Rearranging (5.18), dividing by Δ, and letting $\Delta \to 0$, one obtains

$$\frac{-\partial R(w_2,w_1)}{\partial w_1} = \frac{1}{bw_1}\{1+\nu U(w_1)\}. \tag{5.19}$$

Under the initial condition that $R(w_2,w_2) = 0$, (5.19) yields

$$R(w_2,w_1) = \frac{1}{b}\log\frac{w_2}{w_1} + \frac{\nu}{b}\int_{w_1}^{w_2} \frac{U(w)}{w}\,dw\ . \tag{5.20}$$

(b) Let t_1 be an upcrossing of w_1, and let t_2, $t_2 \geq t_1$, be the first upcrossing of w_2 following t_1. That is $t_2 - t_1$ is the first passage time from an updrossing of w_1, to w_2. denote its expectation by $S(w_1,w_2) = E(t_2-t_1)$.

To calculate $S(w_1,w_2)$, let t be the first crossing of $w_2 - \Delta$ following t_1, so that $t_1 \leq t \leq t_2$. The passage time from $w_2 - \Delta$ to w_2 is zero, i.e. $t = t_2$, unless $w_2 - \Delta < X(t^+) < w_2$. This means that the jump at t crosses $w_2 - \Delta$ but not w_2, and since the jumps are exponential this happens with probability $\Delta/\theta + o(\Delta)$. In this case an expected time $V(w_2) + o(\Delta)$ elapses before the process upcrosses $w_2 - \Delta$ again. Hence

$$S(w_1,w_2) = S(w_1,w_2-\Delta) + \frac{\Delta}{\theta}V(w_2) + o(\Delta), \tag{5.21}$$

from which it follows that

$$\frac{\partial S(w_1,w_2)}{\partial w_2} = \frac{1}{\theta}V(w_2).$$

Thus, by $S(w_1,w_1) = 0$, one has

$$S(w_1,w_2) = \frac{1}{\theta}\int_{w_1}^{w_2} V(w)\,dw. \tag{5.22}$$

(c) Let $J(w_2,w_1)$ be the expected crossing time from w_2 to w_1 and let $O(w_1,w_2)$ be the expected crossing time from w_1 to w_2 as defined in section 3.2. Define also $G(w_1,w_2)$ to be the expected length of an excursion of $X(t)$ above w_1 which does not cross w_2, and define $H(w_2,w_1)$ to be the expected length of an excursion of $X(t)$ below w_2 which does not cross w_1. The expressions $J(w_2,w_1)$, $O(w_1,w_2)$ $G(w_1,w_2)$ and $H(w_2,w_1)$ are expectations of the same random times as $R(w_2,w_1)$, $S(w_1,w_2)$, $U(w_1)$ and $V(w_2)$, under the condition that the process $X(t)$ remains between w_1 and w_2. Using the same arguments as in (a) and (b), one obtains

$$J(w_2,w_1) = \frac{1}{b}\log\frac{w_2}{w_1} + \frac{\nu}{b}\int_{w_1}^{w_2} \frac{G(w,w_2)}{w}\, dw, \tag{5.23}$$

$$O(w_1,w_2) = \frac{1}{\theta}\int_{w_1}^{w_2} H(w,w_1)dw. \tag{5.24}$$

(d) An upcrossing of w_1 is followed first either by a downcrossing of w_1, or by an upcrossing of w_2. Let $p(w_1,w_2)$ be the probability of the latter event. Similarly, let $q(w_2,w_1)$ be the probability that a downcrossing of w_2 is followed by a downcrossing of w_1 before the next upcrossing of w_2.

To calculate $p(w_1,w_2)$, the sequence of upcrossings of w_1, $\ldots t_1 < \ldots < t_m \ldots$, is observed. The upcrossings form a renewal process, with expected renewal time $U(w_1) + V(w_1)$, as shown in section 5.3. A new renewal process is formed by thinning the sequence of renewal times $\{t_m\}$. The renewal times of the new process, $r_1 < \ldots < r_m \ldots$, are defined as $r_1 = t_1$, and for every $m \geq 1$, r_{m+1} is the first upcrossing of w_1 following r_m, such that $X(t)$ crosses w_2 in $[r_m, r_{m+1})$. Note that the definition is unique for $m = 2, 3, \ldots$, but not for $m < 2$. It is easy to see that the expected renewal time of this new process is

$$E(r_{m+1}-r_m) = S(w_1,w_2) + U(w_2) + R(w_2,w_1) + V(w_1). \tag{5.25}$$

The new process counts the number of upcrossings of w_1, starting from t_1, which are followed by upcrossings of w_2. Thus

$$p(w_1,w_2) = \frac{U(w_1)+V(w_1)}{S(w_1,w_2)+R(w_2,w_1)+U(w_2)+V(w_1)} \quad . \tag{5.26}$$

Similarly, one obtains

$$q(w_2,w_1) = \frac{U(w_2)+V(w_2)}{S(w_1,w_2)+R(w_2,w_1)+U(w_2)+V(w_1)} \tag{5.27}$$

(e) Excursions of the process $X(t)$ above the level w_1 can be divided into two groups. With probability $1 - p(w_1,w_2)$, $X(t)$ does not cross w_2, and the expected length of the excursion is $G(w_1,w_2)$. With probability $p(w_1,w_2)$, $X(t)$ crosses w_2, and then the expected length of the excursion is easily seen to be $O(w_1,w_2)+U(w_2)+R(w_2,w_1)$. Hence

$$U(w_1) = \{1-p(w_1,w_2)\}G(w_1,w_2)+p(w_1,w_2)\{O(w_1,w_2)+U(w_2)+R(w_2,w_1)\}. \tag{5.28}$$

Similarly

$$V(w_2) = \{1-q(w_2,w_1)\}H(w_2,w_1)+q(w_2,w_1)\{J(w_2,w_1)+V(w_1)+S(w_1,w_2)\}. \tag{5.29}$$

(f) From (5.28) and (5.29) $G(w_1,w_2)$ and $H(w_2,w_1)$ can be expressed in terms of $O(w_1,w_2)$ and $J(w_2,w_1)$. Substituting these expressions into (5.23) and (5.24), one obtains

$$J(w_2,w_1) = \frac{1}{b}\log\frac{w_2}{w_1} + \frac{\nu}{b}\int_{w_1}^{w_2}\left[\frac{U(w)}{1-p(w,w_2)} - \frac{p(w,w_2)}{1-p(w,w_2)}\right.$$
$$\left.\cdot\{O(w,w_2)+U(w_2)+R(w_2,w)\}\right]\frac{dw}{w}, \tag{5.30}$$

$$O(w_1,w_2) = \frac{1}{\theta}\int_{w_1}^{w_2}\left[\frac{V(w)}{1-q(w,w_1)} - \frac{q(w,w_1)}{1-q(w,w_1)}\right.$$
$$\left.\cdot\{J(w,w_1)+V(w_1)+S(w_1,w)\}\right]dw. \tag{5.31}$$

Substituting (5.31) into (5.30), one obtains an equation for $J(w_2,w_1)$,

$$J(w_2,w_1) = \frac{1}{b}\log\frac{w_2}{w_1} + \frac{\nu}{b}\int_{w_1}^{w_2}\{\frac{U(w)}{1-p(w,w_2)} - \frac{p(w,w_2)}{1-p(w,w_2)} \tag{5.32}$$

$$\cdot(\frac{1}{\theta}\int_{w}^{w_2}[\frac{V(x)}{1-q(x,w)} - \frac{q(x,w)}{1-q(x,w)}\{J(x,w)+V(w)+S(w,x)\}]dx$$

$$+ U(w_2) + R(w_2,w))\}\frac{dw}{w}.$$

The calculation of values of $J(w_2,w_1)$ from (5.32) would be laborious. It is necessary to tabulate $U(w)$ and $V(w)$ from incomplete gamma functions, then to tabulate S and R using $U(w)$ and $V(w)$, and only then to solve (5.32) numerically.

5.5 A Gaussian Process as a Limit of Shot Noise Processes

Let $X(t)$ be a shot noise process with parameters ν, $\theta = 1$ and b. Let $\alpha = \sqrt{\nu/b}$ and define

$$Z(t) = \frac{X(t)-\alpha^2}{\alpha} \tag{5.33}$$

Throughout this section, b is fixed while ν and consequently α vary. For every value of α, $E\{Z(t)\} = 0$, $\mathrm{var}\{Z(t)\} = 1$, and $Z(t)$ is a first order autoregressive process with autocorrelation function $e^{-b|s|}$, see (4.6) and (4.14). Thus for any $s > 0$

$$Z(t) = e^{-bs}Z(t-s) + \varepsilon_s(t). \tag{5.34}$$

By the use of (4.7) and (4.16), the characteristic functions of $Z(t)$ and $\varepsilon_s(t)$ are obtained

$$\phi_Z(u) = e^{-iu\alpha}(\frac{1}{1-iu/\alpha})^{\alpha^2}, \tag{5.35}$$

$$\phi_\varepsilon(u) = e^{-iu\alpha(1-e^{-bs})}(\frac{1-iue^{-bs}/\alpha}{1-iu/\alpha})^{\alpha^2}. \tag{5.36}$$

If $\alpha \to \infty$ then

$$\lim \phi_Z(u) = \exp\{-\tfrac{1}{2}u^2\}, \tag{5.37}$$

$$\lim \phi_\varepsilon(u) = \exp\{-\tfrac{1}{2}u^2(1-e^{-2bs})\}, \tag{5.38}$$

i.e. both the variable $Z(t)$ and the innovation variable $\varepsilon_s(t)$ tend in distribution to normal variables. From (5.34), (5.37) and (5.38), it is easy to check that every finite vector $\{Z(t_1)\ldots Z(t_n)\}$ tends in distribution to a multivariate normal vector $\{Z^*(t_1),\ldots,Z^*(t_n)\}$. The limiting process $Z^*(t)$ is a Gaussian first order autoregressive process, and $Z^*(t)$ satisfies $E\{Z^*(t)\} = 0$, $\mathrm{var}\{Z^*(t)\} = 1$ and $\rho\{Z^*(t),Z^*(t+s)\} = e^{-b|s|}$, as does $Z(t)$.

In sections 5.3 and 5.4 sample path properties of the shot noise process $X(t)$ were discussed, and expectations of some random interval lengths, defined between various crossing times, were calculated. The same random intervals are equally well defined for $Z(t)$ and for $Z^*(t)$, using the definitions of crossing times given in section 3.2. To avoid confusion the notation used here is, for example,

$$J(w_2,w_1),\quad J_z(z_2,z_1) \text{ and } J^*(z_2,z_1)$$

for the expected downcrossing time between two levels for the shot noise process $X(t)$, the rescaled process $Z(t)$ and the Gaussian process $Z^*(t)$, respectively. The levels z_1 and z_2 $(z_2>z_1)$, are now two fixed levels for the processes $Z(t)$ and $Z^*(t)$. Since $Z(t)$ is only a rescaled version of $X(t)$, one has

$$J_z(z_2,z_1) = J(\alpha z_2+\alpha^2,\alpha z_1+\alpha^2) \tag{5.39}$$

with similar relations for the other expressions discussed in sections 5.3 and 5.4. As $\alpha \to \infty$, one may calculate the limit

$$J^*(z_2,z_1) = \lim_{\alpha\to\infty} J_z(z_2,z_1). \tag{5.40}$$

For a rigorous proof that this limit is indeed the corresponding property of the Gaussian process $Z^*(t)$ the concept of weak convergence could probably be used, although this was not looked into here.

Two differences between the shot noise process $X(t)$ and the Gaussian process $Z^*(t)$, are relevant to the present discussion. First, for $Z^*(t)$, with probability one, in every finite neighbourhood of a crossing of a level z, there are other crossings of z. Thus every crossing is, in effect, both a downcrossing and an upcrossing. As a result, one would expect to have

$$\begin{aligned} U^*(z) = V^*(z\ ) = G^*(z_1,z_2) = H^*(z_2,z_1) = 0, \\ p^*(z_1,z_2) = q^*(z_2,z_1) = 0. \end{aligned} \tag{5.41}$$

Also the expressions $R^*(z_2,z_1)$ and $S^*(z_1,z_2)$ are now simply expected first passage times between z_1 and z_2.

The second difference is that $Z^*(t)$, unlike $X(t)$, is time reversible, and that $Z^*(t)$ for each t is a symmetric variable. Hence one has

$$R^*(z_2,z_1) = S^*(-z_2,-z_1) \tag{5.42}$$

$$J^*(z_2,z_1) = O^*(z_1,z_2) \tag{5.43}$$

In the following discussion, limiting operations of the type (5.40) are applied to equations of sections 5.3 and 5.4, to obtain results for $Z^*(t)$.

The first expression to be examined is the expected excursion length above a level, $U(w)$. By (5.9) and (5.15), for the shot noise process X(t)

$$U(w) = \frac{1}{b}\frac{\Gamma(\alpha^2)}{w^{\alpha^2}e^{-w}} \; P\{X(t)>w\}. \tag{5.44}$$

Hence, for Z(t),

$$U_z(z) = \frac{1}{b}\frac{\Gamma(\alpha^2)}{(\alpha z+\alpha^2)^{\alpha^2}e^{-(\alpha z+\alpha^2)}} \; P\{Z(t)>z\}. \tag{5.45}$$

Using Stirling's formula, one has

$$\begin{aligned}\frac{\Gamma(\alpha^2)}{(\alpha z+\alpha^2)^{\alpha^2}e^{-(\alpha z+\alpha^2)}} &= \frac{(\alpha^2)^{\alpha^2-\frac{1}{2}}e^{-\alpha^2}\sqrt{2\pi}\{1+o(\frac{1}{\alpha^2})\}}{e^{-\alpha^2}e^{-\alpha z}(\alpha^2)^{\alpha^2}(1+\frac{z}{\alpha})^{\alpha^2}} \\ &= \frac{\sqrt{2\pi}}{\alpha}\;\frac{e^{\alpha z}}{(1+\frac{z}{\alpha})^{\alpha^2}}\{1+o(\frac{1}{\alpha^2})\}.\end{aligned} \tag{5.46}$$

As $\alpha \to \infty$,

$$\lim \frac{e^{\alpha z}}{(1+\frac{z}{\alpha})^{\alpha^2}} = e^{z^2/2} \tag{5.47}$$

Also, as $\alpha \to \infty$,

$$P\{Z(t)>z\} = 1 - \Phi(z),$$

where $\Phi(z)$ is the cumulative standard normal distribution.

Hence although as predicted in (5.41), $\lim_{\alpha\to\infty} U_z(z) = 0$, one has

$$\lim_{\alpha\to\infty}\alpha U_z(z) = \frac{1}{b}\sqrt{2\pi}e^{z^2/2}\{1-\Phi(z)\}. \tag{5.48}$$

Similarly

$$\lim_{\alpha\to\infty}\alpha V_z(z) = \frac{1}{b}\sqrt{2\pi}e^{z^2/2}\Phi(z). \tag{5.49}$$

Using (5.48) and (5.49), one can obtain

$$\lim_{\alpha\to\infty}\frac{U_z(z_1)+V_z(z_1)}{U_z(z_2)+V_z(z_2)} = \frac{\frac{1}{\sqrt{2\pi}}e^{-z_2^2/2}}{\frac{1}{\sqrt{2\pi}}e^{-z_1^2/2}}. \tag{5.50}$$

The meaning of (5.50) is that the ratio between the (infinite) numbers of times that $Z^*(t)$ crosses z_1 and z_2 is equal to the ratio of the densities. Note that by (5.9) this is not the case for the shot noise process.

The expected first passage time of $Z^*(t)$ from z_2 to z_1 is obtained from [see (5.20)]

$$R_z(z_2,z_1) = \frac{1}{b}\log\left(\frac{z_2\alpha+\alpha^2}{z_1\alpha+\alpha^2}\right) + \int_{z_1}^{z_2} \frac{\alpha^2}{z\alpha+\alpha^2}\,\alpha U_z(z)\,dz. \tag{5.51}$$

Letting $\alpha \to \infty$, by (5.48), one has

$$R^*(z_2,z_1) = \lim_{\alpha\to\infty} R_z(z_2,z_1) = \frac{1}{b}\int_{z_1}^{z_2}\sqrt{2\pi}\; e^{z^2/2}\;\{1-\Phi(z)\}dz. \tag{5.52}$$

Similarly one obtains from (5.22)

$$S^*(z_1,z_2) = \frac{1}{b}\int_{z_1}^{z_2}\sqrt{2\pi}\; e^{z^2/2}\Phi(z)\,dz. \tag{5.53}$$

Note the agreement with (5.42).

Consider the expression for $p(w_1,w_2)$ in (5.26). By (5.48), (5.49), (5.52) and (5.53),

$$\lim_{\alpha\to\infty} \alpha p_z(z_1,z_2) = \frac{\sqrt{2\pi}\; e^{z_1^2/2}}{\int_{z_1}^{z_2}\sqrt{2\pi}\; e^{z^2/2}dz} \tag{5.54}$$

and similarly

$$\lim_{\alpha\to\infty} \alpha q_z(z_2,z_1) = \frac{\sqrt{2\pi}\; e^{z_2^2/2}}{\int_{z_1}^{z_2}\sqrt{2\pi}\; e^{z^2/2}dz}\;. \tag{5.55}$$

Using the limiting results (5.49), (5.53) and (5.55), and using (5.43), one can obtain by substituting into (5.31) that

$$J^*(z_2,z_1) = \int_{z_1}^{z_2}\Big[\frac{1}{b}\sqrt{2\pi}\, e^{z^2/2}\Phi(z) - \frac{\sqrt{2\pi}\, e^{z^2/2}}{\int_{z_1}^{z}\sqrt{2\pi}\, e^{y^2/2}dy} \cdot \{J^*(z,z_1) + \int_{z_1}^{z}\frac{1}{b}\sqrt{2\pi}\, e^{y^2/2}\Phi(y)dy\}\Big]dz. \tag{5.56}$$

Taking derivatives with respect to $z_2 = z$, and fixed z_1, one obtains a first order linear differential equation for $J^*(z,z_1)$, having the solution

$$J^*(z_2,z_1) = \frac{1}{b}\,\frac{1}{\int_{z_1}^{z_2}\gamma(z)dz}\int_{z_1}^{z_2}\{\beta(z)\int_{z_1}^{z}\gamma(y)dy-\gamma(z)\int_{z_1}^{z}\beta(y)dy\}dz \tag{5.57}$$

where

$$\gamma(z) = \sqrt{2\pi}\, e^{z^2/2}$$

$$\beta(z) = \sqrt{2\pi}\, e^{z^2/2}\Phi(z).$$

Equations (5.52) and (5.57) express the first passage time from z_2 to z_1, and the first passage time from z_2 to z_1 conditional on remaining within (z_1,z_2) respectively, as obtained by a limiting procedure. The validity of these expressions for the continuous time Gaussian first order autoregressive process can be verified by solving the appropriate diffusion equations directly, as described by Darling and Siegert (1953) and by Cox and Miller (1965, section 5.10).

Chapter 6

DAILY STREAMFLOW MODELS

6.1 The Single Shot Noise Model

In this Chapter, several models for daily streamflow, based on the shot noise process, are presented. The procedure of fitting the models to data and generating synthetic data from the models is discussed. Throughout the Chapter, the data $X_1,\ldots,X_N$ are assumed to consist of daily averages of a continuous process. The description of each model includes the following: (a) the continuous process which models the continuous streamflow is defined; (b) the properties of the data chosen to be preserved by the model are listed; (c) the estimates used for the preserved properties are described; (d) estimation equations are set up for the model parameters which preserve the desired properties (see section 1.2); (e) the solution of the equations is outlined; and (f) the generation of synthetic data is described.

For the first model, the single shot noise model, the streamflow is assumed to be a continuous-time stationary shot noise process X(t). The properties preserved are the mean, the standard deviation and the first serial correlation coefficient of the data, which are estimated by

$$\tilde{\mu} = \frac{1}{N}\sum_{i=1}^{N} X_i, \tag{6.1}$$

$$\tilde{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{N} X_i^2 - \tilde{\mu}^2, \tag{6.2}$$

$$\tilde{\rho} = \frac{\sum_{i=1}^{N-1} X_i X_{i+1} - \frac{1}{N-1}\sum_{i=1}^{N-1} X_i \sum_{i=2}^{N} X_i}{\sqrt{\left\{\sum_{i=1}^{N-1} X_i^2 - \frac{1}{N-1}\left(\sum_{i=1}^{N-1} X_i\right)^2\right\}\left\{\sum_{i=2}^{N} X_i^2 - \frac{1}{N-1}\left(\sum_{i=2}^{N} X_i\right)^2\right\}}} \tag{6.3}$$

The estimates of ν, θ and b, the parameters of the shot noise process, have to satisfy (see 4.33 and 4.35)

$$\tilde{\mu} = \frac{\nu\theta}{b} \quad , \tag{6.4}$$

$$\tilde{\sigma}^2 = \frac{\nu\theta^2}{b} \; \frac{2\{b-(1-e^{-b})\}}{b^2} \quad , \tag{6.5}$$

$$\tilde{\rho} = \frac{(1-e^{-b})^2}{2\{b-(1-e^{-b})\}} . \tag{6.6}$$

The decay rate b is found first, by solving (6.6) numerically, and (6.4) and (6.5) are then solved for ν and θ. A unique solution exists if $0 < \tilde{\rho} < 1$.

6.2 Generation of Synthetic Shot Noise Data

Let X(t) be a shot noise process with parameters ν, θ and b. Let X(t), (t=0,1,2,...) denote the values of the continuous process, and let X_t, (t=1,2,...), denote the averages of $X(\tau)$ over unit (daily) intervals, $t-1 < \tau < t$. Generation of synthetic data is based directly on

$$X(t+1) = e^{-b}X(t) + \sum_{m=N(t)}^{N(t+1)} e^{-b(t+1-\tau_m)} Y_m \quad , \tag{6.7}$$

$$X_{t+1} = \frac{1}{b}(1-e^{-b})X(t) + \sum_{m=N(t)}^{N(t+1)} \frac{1}{b}\{1-e^{-b(t+1-\tau_m)}\}Y_m ; \tag{6.8}$$

see (4.5) and (4.29). The first term in (6.7) and (6.8) is the contribution of the past, while the second term is the contribution of events in (t,t+1).

The generation starts from an initial value X(0). This initial value is used to generate X(1) and X_1. Then X(1) is used to generate X(2) and X_2 and so on. Assume $X_1,\ldots,X_t$ and X(t) have been generated. Then X(t+1) and X_{t+1} are obtained by the following algorithm.

(a) The first term of (6.7) and of (6.8) is calculated, from $X(t)$, to initiate the values of $X(t+1)$ and X_{t+1}. After that $X(t)$ is discarded.

(b) The time of the last event in $(t-1,t)$ need not be stored, because the events form a Poisson process. Event times are initiated by setting $m = 0$ and $\tau_m = t$.

(c) The next event τ_{m+1} is generated as $\tau_{m+1} = \tau_m + I$, where I is a random number from an exponential distribution with mean $1/\nu$.

(d) If $\tau_{m+1} > t + 1$, all the events in $(t,t+1)$ have been exhausted and the generation of $X(t+1)$ and X_{t+1} is complete.

(e) For $\tau_{m+1} < t + 1$, Y_{m+1} is generated as a random number from an exponential distribution with mean Θ.

(f) The contribution of the $(m+1)^{th}$ event, at τ_{m+1}, with jump Y_{m+1} is calculated as in (6.7) and (6.8) and added to $X(t+1)$ and X_{t+1}.

(g) Set m to $m + 1$ and return to (c).

The generation requires only random numbers from exponential distributions, which are easy to create. With slight alterations this algorithm serves also for generating data with the models discussed in the following sections.

6.3 Multisite Single Shot Noise Model

The streamflow at K sites is assumed to be a multivariate stationary shot noise process, $\underset{\sim}{X}(t) = \{X_1(t),\ldots,X_K(t)\}$, as defined in section 4.2. For each $i = 1, \ldots, K$, $X_i(t)$ is itself a shot noise process. Therefore ν_i, Θ_i and b_i can be estimated to preserve at each site $\tilde{\mu}_i$, $\tilde{\sigma}_i^2$ and $\tilde{\rho}_i$ as in section 6.1. In addition

ρ_{ij}, the cross correlation between the daily flows at each pair of sites $(i,j=1,\ldots,K)$, is to be preserved. The cross correlations are estimated by

$$\tilde{\rho}_{ij} = \frac{\frac{1}{N}\sum_{n=1}^{N} X_{ni}X_{nj} - \tilde{\mu}_i\tilde{\mu}_j}{\sqrt{\tilde{\sigma}_i^2\tilde{\sigma}_j^2}} \tag{6.9}$$

Using (4.26) and (4.36), the equation

$$\tilde{\rho}_{ij} = \frac{\nu_{ij}}{\sqrt{\nu_i\nu_j}}\,\frac{C_{ij}+1}{2}\cdot\frac{2\sqrt{b_ib_j}}{b_i+b_j}\,\frac{b_j^2\{b_i-(1-e^{-b_i})\}+b_i^2\{b_j-(1-e^{-b_j})\}}{2b_ib_j\sqrt{\{b_i-(1-e^{-b_i})\}}\sqrt{\{b_j-(1-e^{-b_j})\}}} \tag{6.10}$$

is set up, with unknowns ν_{ij} and C_{ij}, where ν_{ij} is the joint event rate of the two sites, and C_{ij} is the correlation between simultaneous jumps at the two sites. On setting

$$\nu_{ij} = \min(\nu_i,\nu_j) \tag{6.11}$$

C_{ij} can be determined from (6.10). The choice of ν_{ij} as in (6.11) is made for mathematical convenience. It means that the events at the site with the smaller event rate are all simultaneous with events at the other site.

To generate synthetic data, the algorithm of section 6.2 is used with the following modification. Assume $\nu_1 > \ldots > \nu_K$; then events are generated with rate ν_1, and the probabilities ν_{i+1}/ν_i are used to decide at which sites jumps occur. For these sites simultaneous jumps are generated with the correlations C_{ij}, using the distribution described by (4.21).

When (6.10) is solved, C_{ij} may turn out to be negative, or greater than 1, and both cases are unacceptable. In the first case one can set $C_{ij} = C_{min} \geq 0$, and solve (6.10) for $\nu_{ij} < \min(\nu_i,\nu_j)$. In the second case, C_{ij} is set to some value $C_{max} \leq 1$, and the exact preservation of $\tilde{\rho}_{ij}$ is abandoned.

After these modifications, the matrix $C = \{C_{ij}\}$ has values $0 \leq C_{ij} \leq 1$. To generate simultaneous jumps from (4.21), multivariate normal vectors have to be generated, with correlation matrix $D = \{d_{ij}\}$ where $d_{ij} = \sqrt{C_{ij}}$. The matrix D may, however, be non positive definite. In this case, D will be modified as described in section 6.4, at the expense of the exact preservation of $\tilde{\rho}_{ij}$.

6.4 The Modification of Non Positive Definite Matrices

Let $R = \{r_{ij}\}$ be a symmetric matrix, with $r_{ii} = 1$ and $-1 \leq r_{ij} \leq 1$. The matrix R can be written as $R = UVU'$, where U is an orthogonal matrix and V is a diagonal matrix with real diagonal elements $v_1 \geq v_2 \cdots \geq v_n$. The matrix R is positive definite or semi-definite if $v_n \geq 0$.

Define a matrix $R^* = \{r^*_{ij}\}$ by

$$r^*_{ij} = \begin{cases} d+\dfrac{r_{ij}-d}{1+a}, & (i \neq j), \\ 1, & (i=j), \end{cases} \tag{6.12}$$

where $-1 < d < 1$, and

$$a = \begin{cases} \dfrac{|v_n|}{1-d} & (v_n<0), \\ 0 & (v_n \geq 0). \end{cases} \tag{6.13}$$

The matrix R* can be written as

$$R^* = \frac{ad}{1+a} E + U\{\frac{1}{1+a} V + \frac{a(1-d)}{1+a} I\} U', \tag{6.14}$$

where E is a matrix with all elements equal to unity. It is easily seen from (6.14) that R* is positive definite or semi-definite, and that $r^*_{ii} = 1$ and $-1 < r^*_{ij} < 1$. Thus R* can be used as a correlation matrix of a multisite normal distribution.

The transformation defined in (6.12) gives $R^* = R$ if R itself is positive definite or semi-definite. Otherwise the transformation has the effect of decreasing $r_{ij} - d$ for all the off-diagonal elements by a constant ratio, in other words, the off-diagonal elements shrink towards d. Two natural choices of d are d = 0 or d is the average of the diagonal elements of R, that is

$$d = \frac{2}{n(n-1)} \sum_{i<j} r_{ij}. \tag{6.15}$$

The technique of shrinking the off-diagonal elements towards a value d is reminiscent of the empirical Bayesian techniques described by Stein (1962), with d in (6.15) serving as an empirical estimate of the mean of a prior distribution.

6.5 Seasonally Changing Shot Noise Models

It is assumed that the streamflow is a shot noise process with parameters $\nu(t)$, $\Theta(t)$, $b(t)$ which change periodically with a period T of one year.

A crude way of fitting such a model to data is to divide the year into seasons, e.g. months, and assume that within each season the parameters are approximately constant. For each season these parameters are then estimated as in sections 6.1 and 6.3. This procedure produces a bias, because each season begins with a value distributed with the parameters of the previous season. A numerical example of this transient bias is given in section 7.8.

For a single site it is possible to use the results of section 4.4, if the decay rate is assumed to be a constant, i.e. $b(t) = b$. The model is fitted as follows. The daily mean and variance of the daily streamflow data are estimated as periodic functions $\tilde{m}(t)$ and $\tilde{\sigma}^2(t)$ and fitted with finite harmonic series. The data are standardised using $\tilde{m}(t)$ and $\tilde{\sigma}^2(t)$, and the serial correlation coefficient of the standardised data $\tilde{\rho}$, is calculated. The decay rate b is obtained from $\tilde{\rho}$ by solving (6.6). Equations (4.43) to (4.49) are then used to solve for $\nu(t)$ and $\Theta(t)$. As mentioned in section 4.4, it is not clear, however, whether the standardisation of the data to estimate $\tilde{\rho}$ is a good procedure for the shot noise process.

The generation of synthetic data for these models is done using the algorithm of section 6.2, where the parameters can be changed seasonally.

6.6 A Shot Noise Model Preserving Skewness

The continuous shot noise process has a gamma distribution. Let μ_x, σ_x and γ_x be the mean, standard deviation and skewness of the continuous shot noise process; then

$$\gamma_x = 2\,\frac{\sigma_x}{\mu_x}\ . \tag{6.16}$$

Thus the shot noise model of section 6.1 can only be fitted to streamflows whose skewnesses obey a restrictive rule.

In this section a modified model which can preserve a wider range of skewness values is discussed. The continuous streamflow $X(t)$ is now assumed to be a shot noise process with event rate ν, decay rate b and jumps which have a gamma distribution with parameters $1/\theta$ and α, and a density

$$f(y) = \frac{(1/\theta)^{\alpha} y^{\alpha-1}}{\Gamma(\alpha)}\, e^{-y/\theta}. \tag{6.17}$$

It is easy to see by (4.11) that the cumulants of the continuous process $X(t)$ are

$$\begin{aligned} \mu_x &= \frac{\nu\alpha\theta}{b} \\ \sigma_x^2 &= \frac{\nu\alpha(\alpha+1)\theta^2}{2b} \\ \kappa_3\{X(t)\} &= \frac{\nu\alpha(\alpha+1)(\alpha+2)\theta^3}{3b} \end{aligned} \tag{6.18}$$

and hence that

$$\gamma_x = \frac{4}{3}\,\frac{\alpha+2}{\alpha+1}\,\frac{\sigma_x}{\mu_x}\,. \tag{6.19}$$

Because $\alpha > 0$ it follows that

$$\frac{4}{3}\,\frac{\sigma_x}{\mu_x} < \gamma_x < \frac{8}{3}\,\frac{\sigma_x}{\mu_x}\,. \tag{6.20}$$

The model parameters ν, α, θ and b, can be estimated to preserve the mean, variance, first serial correlation coefficient and skewness of daily averaged data. This is done by solving

$$\begin{aligned}
\tilde{\mu} &= \frac{\nu\alpha\theta}{b}\,, \\
\tilde{\sigma}^2 &= \frac{\nu\alpha(\alpha+1)\theta^2}{b}\;\frac{\{b-(1-e^{-b})\}}{b^2}\,, \\
\tilde{K}_3 &= \frac{\nu\alpha(\alpha+1)(\alpha+2)\theta^3}{b}\;\frac{\{b-(1-e^{-b})-\frac{1}{2}(1-e^{-b})^2\}}{b^3} \\
\tilde{\rho} &= \frac{(1-e^{-b})^2}{2\{b-(1-e^{-b})\}}\,,
\end{aligned} \tag{6.21}$$

where $\tilde{K}_3$ is the third cumulant of the data, estimated by

$$\tilde{K}_3 = \frac{1}{N}\sum_{i=1}^{N}(X_i-\tilde{\mu})^3\,. \tag{6.22}$$

Note that the equation involving b and $\tilde{\rho}$ is identical to (6.6). Therefore, the same value of b is derived from the data for the shot noise model of section 6.1 and for the present model.

To generate synthetic data from the model with gamma distributed jumps, the algorithm of section 6.2 is used, with the modification that jumps are generated from a gamma distribution. Pseudo-random numbers from a gamma distribution can be generated by a method described by Jöhnk (1964) and by Bánkövi (1964).

A wider range of skewness values can be preserved if a fifth parameter q is added and the continuous streamflow is assumed to be $X(t) + q$, where $q > 0$ is a minimum flow.

6.7 The Double Shot Noise Model

Attempts to fit the single shot noise model, described in sections 6.1 - 6.5, to some daily streamflow data revealed serious discrepancies between the model and the historical data. The discrepancies fall into two categories, and arise because the decay rate b, which is calculated to preserve the daily serial correlation coefficient $\tilde{\rho}$, is too large. The first kind of discrepancy is in aggregate behaviour of monthly averages of the data. With a typical value of $\tilde{\rho} = 0.8$, b is calculated to be approximately 0.3. For this value of b, the modelling process will have a serial correlation between consecutive monthly averages of approximately 0.05. This contrasts with the value of 0.5 typically obtained from monthly average flows in the historical data, i.e. the real process has much more long term dependence than the model predicts. Similar discrepancies arise when comparing the variances of monthly averages predicted by the model with those found in the data; see sections 7.5 and 7.6 for a numerical example.

The second kind of discrepancy is found when comparing visually historical daily streamflow with synthetic streamflow produced by the model; see figures 4 and 5. In the historical data recessions decay very slowly once low flows are reached, while in the synthetic data the decay towards zero is very fast, and the synthetic flow approaches zero frequently. Also, more rises and recessions are observed in the synthetic data than in historical data.

Both kinds of discrepancies suggest the addition of another random component to the single shot noise process. This second component is a stochastic process with small variance but a much higher serial correlation function. Thus it may account for an increasing part of the variance and the serial correlation of averaged flows, as the averaging period increases, and may alleviate the first kind of discrepancy. It may also prevent fast decay when low flows are reached. Such a second component has a sound physical basis, because it imitates the behaviour of ground water discharge in streamflow.

A natural choice for this component is that it also is a shot noise process. Thus the streamflow X(t) is modelled by

$$X(t) = X_1(t) + X_2(t), \tag{6.23}$$

where $X_1(t)$ and $X_2(t)$ are shot noise processes, with parameters ν_1, θ_1, b_1 and ν_2, θ_2, b_2 respectively, and with $b_2 < b_1$. The double shot noise model is defined by (6.23) with the additional assumption that $X_1(t)$ and $X_2(t)$ are independent. This assumption of independence does not correspond to the physical interpretation of $X_1(t)$ as surface runoff and $X_2(t)$ as groundwater discharge. A model based on a more realistic assumption is discussed in section 6.10.

The double shot noise model is fitted so as to preserve properties of both daily and monthly average flow. Denote by T_1 the period of one day and by T_2 the period of one month. The following equations relate the six model parameters to $\tilde{\mu}$ the average of the flows, to $\tilde{\sigma}_1^2$ and $\tilde{\sigma}_2^2$ the variances of daily and of monthly flows respectively, and to $\tilde{\rho}_1$ and $\tilde{\rho}_2$ the first serial correlation coefficients of daily and of monthly flows, respectively:

$$\tilde{\mu} = \frac{\nu_1\theta_1}{b_1} + \frac{\nu_2\theta_2}{b_2} \tag{6.24}$$

$$\tilde{\sigma}_1^2 = \frac{\nu_1\theta_1^2}{b_1}\,\frac{2\{b_1T_1-(1-e^{-b_1T_1})\}}{(b_1T_1)^2} + \frac{\nu_2\theta_2^2}{b_2}\,\frac{2\{b_2T_1-(1-e^{-b_2T_1})\}}{(b_2T_1)^2} \tag{6.25}$$

$$\tilde{\sigma}_2^2 = \frac{\nu_1\theta_1^2}{b_1}\,\frac{2\{b_1T_2-(1-e^{-b_1T_2})\}}{(b_1T_2)^2} + \frac{\nu_2\theta_2^2}{b_2}\,\frac{2\{b_2T_2-(1-e^{-b_2T_2})\}}{(b_2T_2)^2} \tag{6.26}$$

$$\tilde{\rho}_1 = (1-\alpha_1)\,\frac{(1-e^{-b_1T_1})^2}{2\{b_1T_1-(1-e^{-b_1T_1})\}} + \alpha_1\,\frac{(1-e^{-b_2T_1})^2}{2\{b_2T_1-(1-e^{-b_2T_1})\}} \tag{6.27}$$

$$\tilde{\rho}_2 = (1-\alpha_2)\,\frac{(1-e^{-b_1T_2})^2}{2\{b_1T_2-(1-e^{-b_1T_2})\}} + \alpha_2\,\frac{(1-e^{-b_2T_2})^2}{2\{b_2T_2-(1-e^{-b_2T_2})\}} \tag{6.28}$$

where

$$\alpha_1 = \frac{\nu_2\theta_2^2}{b_2}\,\frac{2\{b_2T_1-(1-e^{-b_2T_1})\}}{(b_2T_1)^2} \Big/ \tilde{\sigma}_1^2 \tag{6.29}$$

$$\alpha_2 = \frac{\nu_2\theta_2^2}{b_2}\,\frac{2\{b_2T_2-(1-e^{-b_2T_2})\}}{(b_2T_2)^2} \Big/ \tilde{\sigma}_2^2. \tag{6.30}$$

The relations (6.24) - (6.28) yield five equations for the six parameters of the model. Thus one parameter has to be estimated by some other method before the equations for the other five parameters may be solved. In the present work ν_1 is estimated by visual inspection of the data, in an attempt to preserve the rate at which rises and recessions occur.

6.8 The Solution of the Double Shot Noise Model Equations

Simple manipulations of (6.24) - (6.28) give two equations for b_1 and b_2. Defining, for $i, j = 1, 2$,

$$d_{ij} = d(b_i,T_j) = 2\{b_iT_j-(1-e^{-b_iT_j})\}/(b_iT_j)^2, \tag{6.31}$$

$$e_{ij} = e(b_i,T_j) = (1-e^{-b_iT_j})^2/(b_iT_j)^2, \tag{6.32}$$

one has

$$\rho_1\sigma_1^2 = \frac{\sigma_1^2d_{22}-\sigma_2^2d_{21}}{d_{22}d_{11}-d_{12}d_{21}}\,e_{11} + \frac{\sigma_2^2d_{11}-\sigma_1^2d_{12}}{d_{22}d_{11}-d_{12}d_{21}}\,e_{21} \tag{6.33}$$

$$\rho_2\sigma_2^2 = \frac{\sigma_1^2 d_{22} - \sigma_2^2 d_{21}}{d_{22}d_{11} - d_{12}d_{21}}\, e_{12} + \frac{\sigma_2^2 d_{11} - \sigma_1^2 d_{12}}{d_{22}d_{11} - d_{12}d_{21}}\, e_{22}. \tag{6.34}$$

Once these equations are solved numerically for b_1 and b_2, the other parameters are obtained easily enough from (6.24) - (6.26).

The equations (6.33) and (6.34) were investigated in detail, but no complete picture as to the existence and uniqueness of a solution was reached. The following three conditions were found to be sufficient for the existence of a solution.

(a) For $T_2 > T_1$

$$\sigma_2^2\rho_2 T_2^2 > \sigma_1^2\rho_1 T_1^2. \tag{6.35}$$

This condition is necessary for any process with a positive autocorrelation function.

(b) Let b_{10}, b_{20} and b_0 be three decay rates of three single shot noise processes, such that

b_{10} preserves the daily serial correlation ρ_1,

b_{20} preserves the monthly serial correlation ρ_2,

b_0 preserves the ratio between monthly variance and daily variance σ_2^2/σ_1^2.

The decay rates b_{10} and b_{20} exist for any $0 < \rho_1 < 1$, $0 < \rho_2 < 1$. The decay rate b_0 exists for any $\sigma_2^2 < \sigma_1^2$, if condition (a) holds. All three values are uniquely determined. The second condition is

$$b_{10} > b_0 > b_{20}. \tag{6.36}$$

(c) When in equation (6.33) one lets $b_1 \to \infty$, an equation for the slow decay rate alone is obtained, whose solution is denoted by b_{21}. similarly when $b_1 \to \infty$ in (6.34) an equation for the slow decay rate is obtained, whose solution is denoted by b_{22}. When one

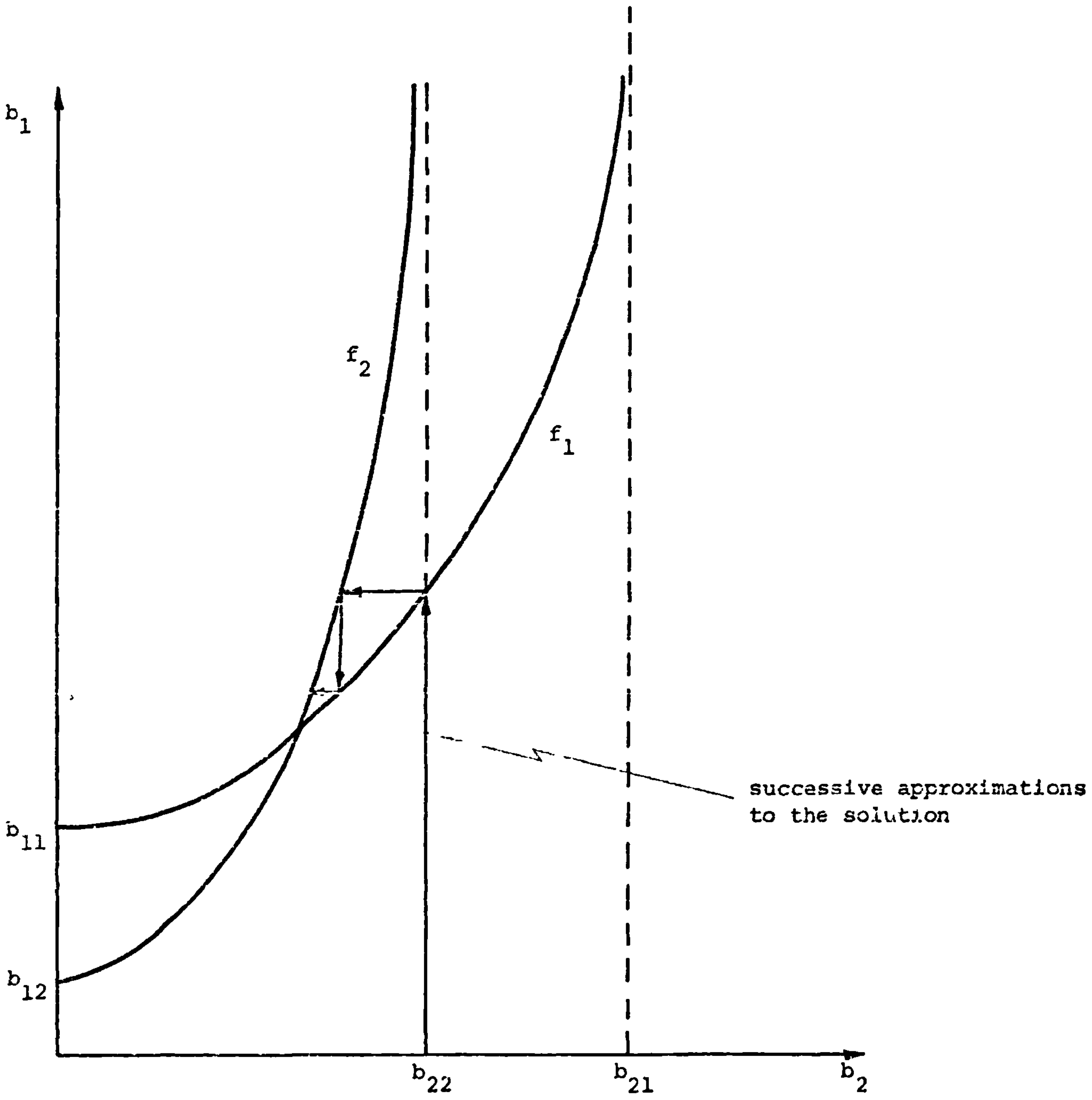

Figure 2. Solution of the Equations $\rho_1\sigma_1^2 = f_1(b_1,b_2)$ and $\rho_2\sigma_2^2 = f_2(b_1,b_2)$ for the Decay Rates b_1 and b_2, for the Double Shot Noise Model

lets $b_2 \to 0$ in (6.33) and (6.34), one obtains two equations for fast decay rates, whose solutions are denoted by b_{11} and b_{12} respectively. The solutions b_{21}, b_{22}, b_{11} and b_{12} exist and are unique if conditions (a) and (b) hold. The third condition is

$$b_{11} > b_{12}\ , \quad b_{21} > b_{22}. \tag{6.37}$$

If (6.37) holds then equations (6.33) and (6.34) describe two relations, $\rho_1\sigma_1^2 = f_1(b_1,b_2)$ and $\rho_2\sigma_2^2 = f_2(b_1,b_2)$, which are represented in figure 2. It is easy to see that a solution exists in this case. The solution can be obtained numerically by solving (6.32) and (6.33) alternately, as shown in figure 2.

It is not known whether the solution is unique and whether all three conditions are necessary.

6.9 Seasonal Multisite Double Shot Noise Models

As seen in sections 6.7 - 6.8, the stationary double shot noise process is quite complicated to handle practically. In adapting it to multisite seasonal data some simplifying assumptions are made.

To fit the process to seasonal data, it is assumed that the model parameters change slowly, and the crude approach of dividing the data into seasons, (e.g. months), and of estimating a set of parameters separately for each season is adopted. This procedure is not very good, because the problem of bias, caused by the transient effect of the initial value of each season, is quite substantial for the slow process, with its very high serial correlation.

The multisite double shot noise process $\underset{\sim}{X}(t) = \{X_1(t),\ldots,X_K(t)\}$, used to model multisite daily streamflow within each season, is defined by

$$X_i(t) = X_{1i}(t) + \mu_i X_2(t) \quad (i=1,\ldots,K) \tag{6.38}$$

where $X_{1i}(t)$ are fast shot noise processes, with parameters ν_{1i}, θ_{1i} and b_{1i} for each site $i = 1,\ldots K$, μ_i is the average flow at site i, and $X_2(t)$ is a single, common, slow shot noise process with parameters ν_2, θ_2 and b_2, where $b_2 < b_{1i}$ $(i=1,\ldots,K)$.

The assumption of a common slow process at all sites cuts down on the number of parameters and simplifies the cross correlation structure of the model. If all sites are in the same region then, since $\mu_i X_2(t)$ represents the groundwater discharge at site i, the assumption of a common process i.e. of proportional groundwater discharges at all sites, may be reasonable.

To fit the multisite double shot noise process to streamflow data, sample estimates of μ_i, σ_{1i}^2, σ_{2i}^2, ρ_{1i}, ρ_{2i} and ρ_{ij}, the daily lag zero cross correlation between the sites are calculated for i, j = 1, ..., K. The parameters are estimated in three steps.

(a) <u>Estimation of the parameters of the slow process $X_2(t)$.</u>

The averages of $(\frac{\sigma_{1i}}{\mu_i})$, $(\frac{\sigma_{2i}^2}{\sigma_{1i}^2})$, ρ_{1i} and ρ_{2i} over all the sites are calculated. In addition, a single value ν_1 is estimated by visual inspection as the rate at which rises and recessions occur in the region. These values, together with $\mu = 1$, are used to solve equations (6.24) - (6.28), and to obtain ν_2, θ_2 and b_2.

(b) <u>Estimation of the parameters of the fast processes $X_{1i}(t)$.</u>

For i = 1, ..., K the previously calculated values of ν_2, θ_2 and b_2 are substituted into equations (6.24), (6.25) and (6.27). Equations for ν_{1i}, θ_{1i} and b_{1i} in terms of μ_i, σ_{1i} and ρ_{1i}, similar to (6.4) - (6.6) are obtained and solved for ν_{1i}, θ_{1i} and b_{1i}, $(i=1,\ldots,K)$.

(c) Estimation of cross correlation parameters.

The cross correlation between each pair of sites is given by

$$\rho_{ij} = C_1 \rho\{X_{1i}(t), X_{1j}(t)\} + C_2 \cdot 1 \tag{6.39}$$

where 1 is the cross correlation between the slow processes at the two sites, C_1 and C_2 are weights obtained from the known values of the daily variances, and $\rho\{X_{1i}(t), X_{1j}(t)\}$ is the cross correlation between the daily values of the fast processes at the two sites. $\rho\{X_{1i}(t), X_{1j}(t)\}$ is calculated from (6.39), and the cross correlation parameters between the fast processes are obtained as in section 6.3.

This procedure attempts to preserve exactly only $\tilde{\mu}_i$, $\tilde{\sigma}^2_{1i}$, $\tilde{\rho}_{1i}$ and $\tilde{\rho}_{ij}$, i.e. the daily properties. The monthly properties $\tilde{\sigma}^2_{2i}$ and $\tilde{\rho}_{2i}$, and the predetermined value of ν_1 are only preserved approximately. They may be preserved accurately if $\frac{\sigma_{1i}}{\mu_i}$, $\frac{\sigma^2_{2i}}{\sigma^2_{1i}}$, ρ_{1i} and ρ_{2i} are similar for all sites. For a numerical example, see section 7.7.

At the various steps of the fitting procedure, cases may exist where there is no solution or where a solution is unacceptable. Ad hoc methods of modifying the equations to achieve solutions at the cost of the accuracy in the preservation of some of the properties were formulated for all these cases, and incorporated in a computer program written for the fitting of the model.

6.10 A Second Order Shot Noise Model

In this section the streamflow is modelled by a second order linear filtered Poisson process X(t). The input to the process is the same as for the single shot noise process, i.e. Poisson events with rate ν accompanied by exponentially distributed jumps with mean Θ. The pulse shape is given by

$$w(t) = (1-\alpha)e^{-b_1 t} + \alpha e^{-b_2 t} \quad (t>0) \tag{6.40}$$

where $b_1 > b_2$. This pulse shape corresponds to a system of two linear reservoirs in parallel, which can be represented by a second order linear differential equation between the input and the output, see section 2.6. In physical terms the two reservoirs represent the surface runoff and the groundwater discharge.

Using the notation $\Theta_1 = (1-\alpha)\Theta$ and $\Theta_2 = \alpha\Theta$, one can write

$$X(t) = X_1(t) + X_2(t) \tag{6.41}$$

where $X_1(t)$ and $X_2(t)$ are shot noise processes with parameters ν, Θ_1, and b_1 and ν, Θ_2, b_2 respectively. In the present model the two processes have proportional inputs, as opposed to independent inputs for the double shot noise process of section 6.7. It should be noted though, that for $b_1 >> b_2$ the cross correlation between $X_1(t)$ and $X_2(t)$ is small.

Using the techniques of sections 4.1 - 4.3 one can easily set up equations for the estimation of the parameters of X(t). Employing the notation of section 6.7, one has, for an averaging period T,

$$\mu = \nu\Theta\left(\frac{1-\alpha}{b_1} + \frac{\alpha}{b_2}\right), \tag{6.42}$$

$$\begin{aligned}\sigma_T^2 = \nu\Theta^2\Big[&\left\{\frac{(1-\alpha)^2}{b_1} + \frac{2(1-\alpha)\alpha}{b_1+b_2}\right\}\frac{2\{b_1T-(1-e^{-b_1T})\}}{(b_1T)^2} \\ &+ \left\{\frac{\alpha^2}{b_2} + \frac{2(1-\alpha)\alpha}{b_1+b_2}\right\}\frac{2\{b_2T-(1-e^{-b_2T})\}}{(b_2T)^2}\Big]\end{aligned} \tag{6.43}$$

$$\rho_T\sigma_T^2 = \nu\theta^2[\{\frac{(1-\alpha)^2}{b_1} + \frac{2(1-\alpha)\alpha}{b_1+b_2}\} \frac{(1-e^{-b_1T})^2}{(b_1T)^2} + \{\frac{\alpha^2}{b_2} + \frac{2(1-\alpha)\alpha}{b_1+b_2}\} \frac{(1-e^{-b_2T})^2}{(b_2T)^2}]. \tag{6.44}$$

Equation (6.42) and equations (6.43) and (6.44) written for T_1 and T_2, the periods of one day and of one month respectively, yield five equations for the five parameters of the model, and enable the preservation of μ, σ_1^2, σ_2^2, ρ_1 and ρ_2. A notable fact, which emerges from (6.43) and (6.44), is that the equations for b_1 and b_2 are the same as for the double shot noise model. Thus, the same values of b_1 and b_2 are derived from the data for both models.

6.11 Disaggregation of Monthly Flows

In section 4.5 the disaggregation of accumulated flows of a shot noise process X(t) over a period T was described. It was shown that when a monthly accumulated flow is given, a sequence of daily flows can be generated from the shot noise process X(t), with approximately the given monthly total flow. The practical purposes to which this technique may be applied, are discussed in the present section. The approximate nature of the technique is ignored.

For the first use of disaggregation, assume that the streamflow behaves like a shot noise process X(t), with parameters ν, θ and b which are known, or which have been estimated from daily data. Then a historical monthly record can be used to generate synthetic daily data.

A second application is as follows. Assume that daily and monthly data are examined, and that daily records show agreement with a shot noise model X(t), but monthly records exhibit a different behaviour, e.g. they are fitted by a fractional Gaussian noise process with Hurst coefficient $h > \frac{1}{2}$. Then synthetic monthly data may be generated first, from the monthly model. These monthly data may then be disaggregated into daily flows, as though they consisted of average values from the shot noise process fitted to the daily data. Thus an intermediate model, between the daily and the monthly models is obtained. A shortcoming of such a model is that it is not stationary. To see this, let $X_1,\ldots,X_T$ and $X_{T+1},\ldots,X_{2T}$ be daily data, produced by disaggregation of $Y_1 = \sum_{t=1}^{T} X_t$ and $Y_2 = \sum_{t=T+1}^{2T} X_t$. Then $(X_1,\ldots,X_T)$ and $(X_{T+1},\ldots,X_{2T})$ have the same joint distribution, but $(X_1,\ldots,X_T)$ and $(X_2,\ldots,X_{T+1})$ do not as a rule have the same joint distribution.

Chapter 7

APPLICATION OF SHOT NOISE MODELS

7.1 Sources of the Data

The shot noise models proposed in Chapter 6 were applied to daily data from several sites on the rivers Nene and Great Ouse in the Wash Estuary area of East Anglia. The data were supplied by the Water Resources Board of England and Wales. A general description of the sites is given in table 1, and their locations are given in figure 3, More details of the sites can be found in publications of the Water Resources Board (1968, 1971).

The records include eleven consecutive complete years of concurrent data of all sites over the period 1953-1963, and the average daily flows over that period were used to study the applicability of the models. Throughout this Chapter the unit of measurement is m^3/sec.

7.2 Time Reversibility Tests

Visual inspection of the daily data confirms that the flows have rapid rises followed by slow recessions, see figure 4. In Chapter 3, the concept of time reversibility was discussed and it was noted that the rapid rises and slow recessions of the flows are an expression of time irreversibility. This feature necessitates modelling of streamflow by non-Gaussian processes. In section 3.2, a statistical test for time reversibility was described. In this section, that test is applied to confirm the visual observation. To study the performance of the test, the same test is also applied to two sequences of synthetic random numbers. The first sequence was generated from a Gaussian first order autoregressive

Table 1: Description of sites used in the application of the models

Station Number	Site	National Grid Reference	Drainage Area Km^2	Mean Gauged Discharge m^3/sec	Annual Rainfall mm	Runoff mm	Period of daily records
32/1	Nene at Orton	TL 166 972	1630	8.29	626	161	1941-1970
32/2	Willow Brook at Fotheringhay	TL 067 933	89.6	0.69	609		1945-1965
32/3	Harpers Brook at Old Mill Bridge	SP 983 799	74.3	0.39	626	177	1939-1964
32/4	Ise at Harrowden Old Mill	SP 898 715	194	1.30	642	203	1945-1964
32/6	Kislingbury Branch at Upton Mill	SP 721 592	223	1.24	662	180	1940-1967
32/7	Brampton Branch at St Andrews Mill	SP 747 617	233	1.29	655	191	1939-1965
32/8	Kislingbury Branch at Dodford	SP 627 607	107	0.52	672	160	1946-1965
33/5	Great Ouse at Thornborough Mill	SP 736 353	389	2.42	668	201	1952-1965

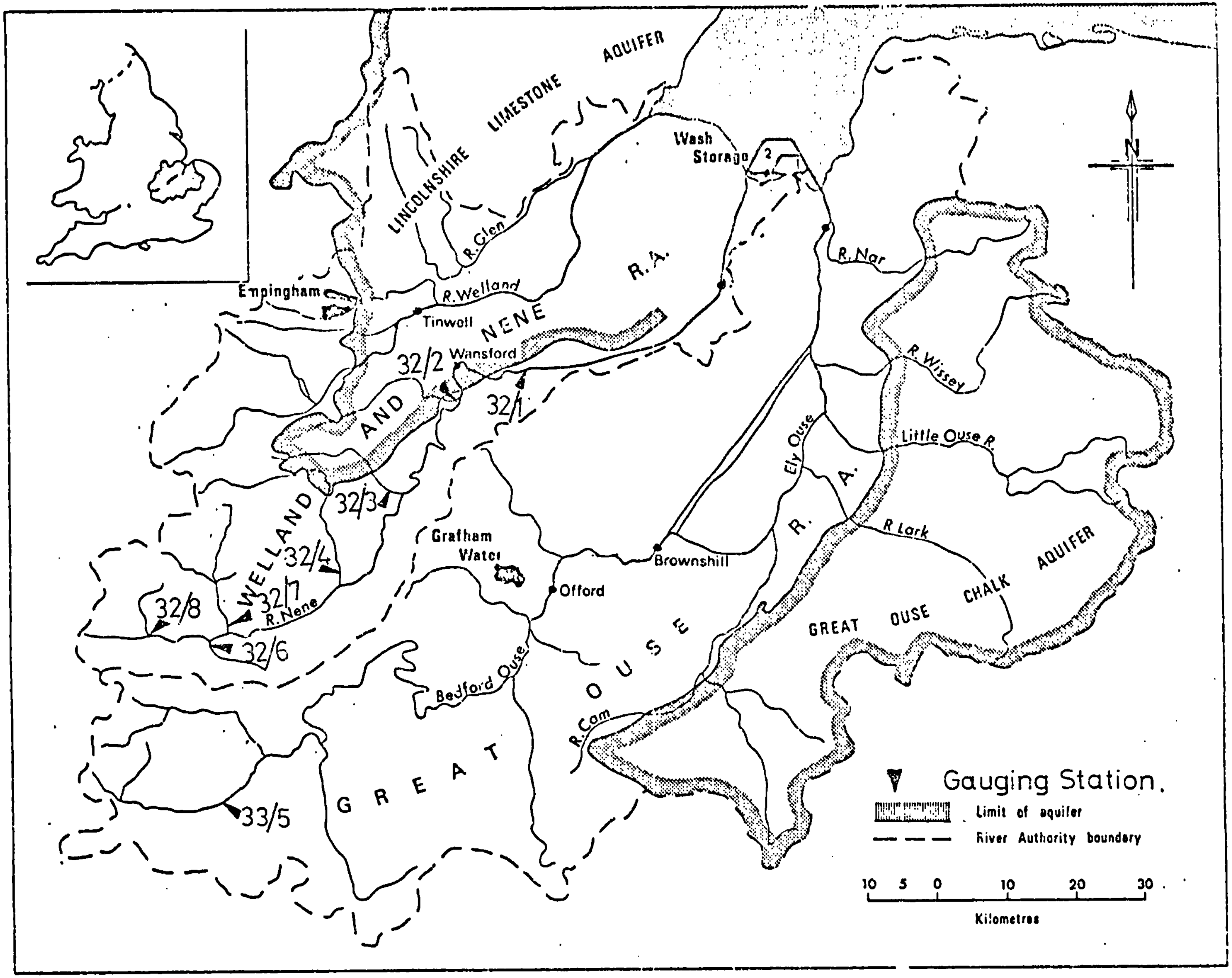

Figure 3. Locations of Gauging Stations in the Wash Estuary

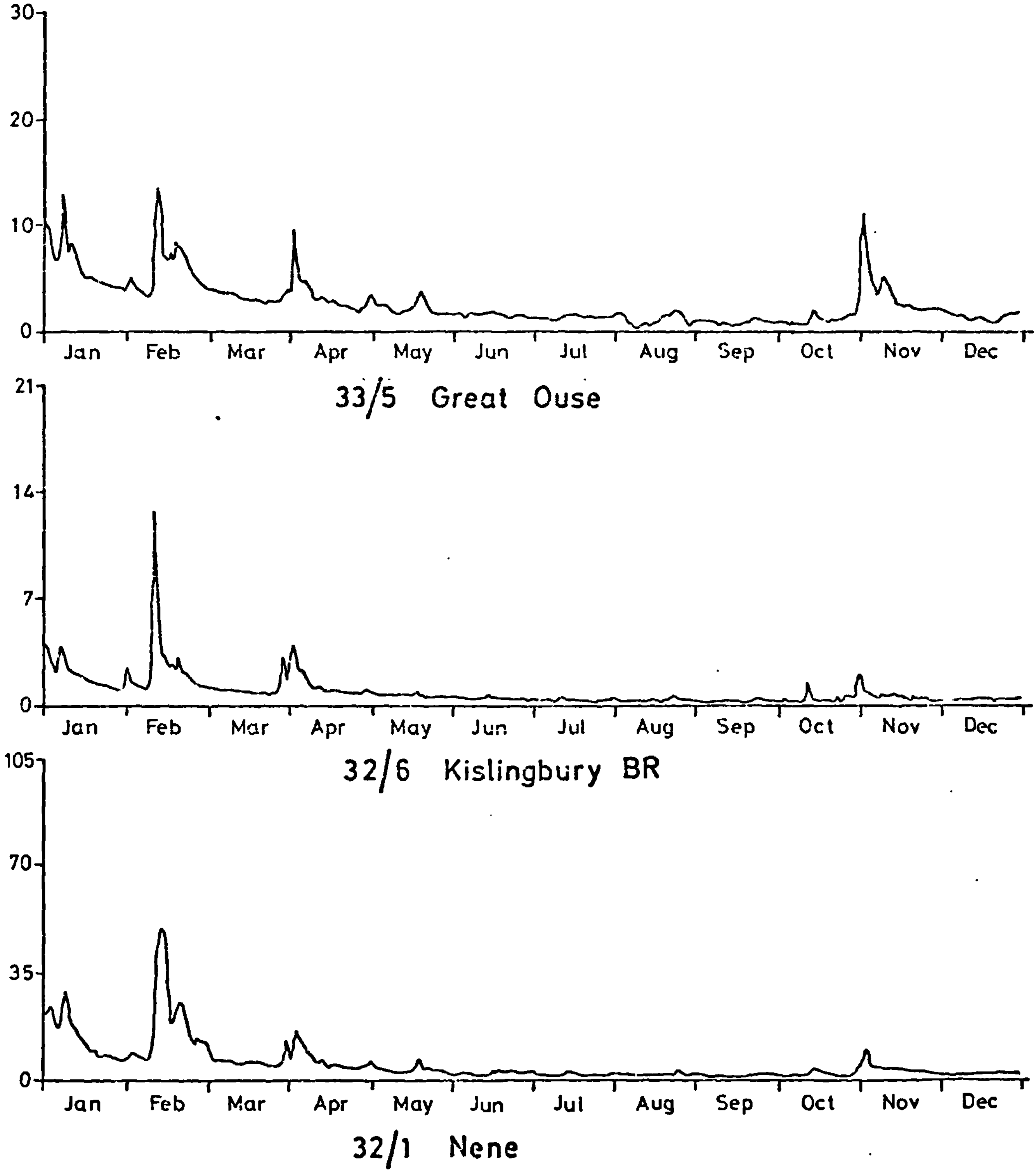

Figure 4. Daily streamflows, East Anglia, 1953

Table 2: Time reversibility tests

<table>
<tr><th rowspan="2">Source</th><th rowspan="2">Sequence length</th><th colspan="2">Levels</th><th rowspan="2">Number of crossing intervals</th><th colspan="3">Wilcoxon two sample test</th></tr>
<tr><th>%</th><th>Value</th><th>range of statistic, mean ± standard deviation</th><th>test value</th><th>significance</th></tr>
<tr><td rowspan="4">Historical</td><td rowspan="4">730</td><td>25%</td><td>0.357</td><td rowspan="2">9</td><td rowspan="2">0 - 81
40.5 ± 11.3</td><td rowspan="2">43</td><td rowspan="2">-</td></tr>
<tr><td rowspan="2">median</td><td rowspan="2">0.663</td></tr>
<tr><td rowspan="2">14</td><td rowspan="2">0 -196
98.0 ± 21.8</td><td rowspan="2">177</td><td rowspan="2">+</td></tr>
<tr><td>75%</td><td>1.098</td></tr>
<tr><td rowspan="4">Gaussian</td><td rowspan="4">730</td><td>25%</td><td>-0.088</td><td rowspan="2">18</td><td rowspan="2">0 - 324
162.0 ± 31.6</td><td rowspan="2">171</td><td rowspan="2">-</td></tr>
<tr><td rowspan="2">median</td><td rowspan="2">0.837</td></tr>
<tr><td rowspan="2">12</td><td rowspan="2">0 - 144
72.0 ± 17.6</td><td rowspan="2">83</td><td rowspan="2">-</td></tr>
<tr><td>75%</td><td>1.724</td></tr>
<tr><td rowspan="4">Shot noise</td><td rowspan="4">730</td><td>25%</td><td>0.021</td><td rowspan="2">5</td><td rowspan="2">0 - 25
12.5 ± 4.8</td><td rowspan="2">25</td><td rowspan="2">+</td></tr>
<tr><td rowspan="2">median</td><td rowspan="2">0.200</td></tr>
<tr><td rowspan="2">6</td><td rowspan="2">0 - 36
18.0 ± 6.2</td><td rowspan="2">36</td><td rowspan="2">+</td></tr>
<tr><td>75%</td><td>0.959</td></tr>
</table>

process which is Markovian and time reversible and the second was generated from a shot noise process, which is Markovian but is not time reversible.

The historical data tested consisted of the daily flows of the River Nene at Orton, for the two years 1953-1954. The flows were standardised to cancel the seasonal effects, by dividing them by the appropriate standard deviations for each month.

For each sequence two tests were performed; the two tests compare upwards and downwards crossing times between two pairs of levels. The levels for the first test are the lower quartile and the median of the data, and the levels for the second test are the median and upper quartile. The results of the tests are summarised in table 2.

The Wilcoxon two sample test is used to compare the crossing times (see Hájek 1969, section 3.10). A median test and a Kolmogorov-Smirnov test were also performed with virtually identical results. It should be noted that for the Gaussian sequence the test value is almost at the mean of the distribution while for the shot noise sequence it is almost at the extreme tail of the distribution. As for the historical data, in the 25% - median region no indication of time irreversibility was found. In the region median - 75%, however, time irreversibility is found, and so the test confirms that daily flows are not time reversible.

7.3 Analysis of Rises and Recessions in the Data

In this section, rises and recessions observed in the data are analysed to see whether they agree with the assumption that streamflow arises from a shot noise process. In a continuous realisation of a shot noise process, each event causes a vertical jump, and the recessions are shaped as exponentially decreasing curves, all with the same decay rate. The time intervals between jumps are independent and exponentially distributed; the jumps are independent and exponentially distributed; and the intervals and the jumps are independent.

The following sequences are defined from the data, by observing the sequence of days $t_1, t_2, \ldots$ for which $X(t_m) > X(t_m - 1)$:

$I_1, I_2, \ldots$ is the sequence of intervals defined by $I_m = t_m - t_{m-1}$.

$Y_1, Y_2, \ldots$ is the sequence of rises defined by $Y_m = X(t_m) - X(t_m - 1)$.

$B_1, B_2, \ldots$ is the sequence of decay rates, defined by $B_m = \log\{X(t-1)/X(t)\}$, for days for which $X(t-1) > X(t)$.

If streamflow is indeed a shot noise process and if the sampling period Δ is small, i.e. $e^{-b\Delta} \simeq 1$ and $\nu\Delta \simeq 0$, then these three sequences tend to the intervals between events, the jumps, and the constant decay rate, respectively.

The sequences $\{I_m\}$, $\{Y_m\}$ and $\{B_m\}$ were obtained from the record of daily streamflow of the Nene at Orton, over the period 1941-1969. Only the January flows were used, so that stationarity may be assumed. In table 3 the statistics which were calculated for the three sequences are given.

The following inferences are drawn from table 3.

(a) The interval lengths are serially independent.

(b) The interval lengths are independent of the rises at the beginning of each interval.

(c) The rises have a small but positive serial correlation.

(d) The decay rates are far from constant.

(e) Positive correlation exists between the decay rates in each interval, and both the interval length and the rise at the beginning of the interval.

Table 3: Analysis of observed recessions, Nene at Orton, 1941-1969, January

Series	Sample Size	Mean	Standard Deviation	Serial Correlation	Cross Correlations		
					I	Y	B
Intervals I	397	2.26	2.12	0.034*	1	- 0.041	0.42
Rises Y	397	4.70	7.42	0.18	- 0.041	1	0.12
Decay rates B	502	0.14	0.16	0.21	0.42	0.12	1

*For $N = 397$ the correlation ρ is non zero with a significance of 5% if $|\tilde{\rho}| > 0.09$

Table 4: Comparison between historical flows and flows with constant decay rate, Nene at Orton 1941-1969, January

	Mean	Standard Deviation	Skewness	Serial Correlation					
				$\rho(1)$	$\rho(2)$	$\rho(3)$	$\rho(4)$	$\rho(5)$	$\rho(10)$
X(t)	16.98	16.62	1.94	.939	.836	.720	.603	.498	.293
XS(t)	17.08	16.82	1.82	.951	.869	.775	.678	.588	.377

(f) In a separate test, it was found that the interval lengths do not come from an exponential distribution.

Of these conclusions only (a) and (b) tend to confirm the model. Two possible explanations for the disagreement between the model and the conclusions (c) - (f) may be given. The first is that the averaging period is too long for the series I, Y and B to approximate the continuous streamflow intervals, jumps and decay rates. The second is that possibly the streamflow behaves like a filtered Poisson process with a pulse shape which has a short but gradual rise. In this case the sequences defined are not very meaningful. The second point seems to be confirmed by visual inspection of the data; see figure 4.

To test the implication of assuming an exponential decay as the pulse shape, a sequence XS(t) is defined as follows. The sequence XS(t) has the same rises and the same intervals between rises as X(t), and for any day t for which $X(t) < X(t-1)$, $XS(t) = XS(t-1)e^{-b}$, where b is a constant. The value of b is chosen to preserve the mean of X(t). In table 4, a comparison between X(t) and XS(t) is given for the same data as in table 3. There seems to be good agreement between the two sequences.

7.4 Cross Correlations Between Sites

The cross correlations between the eight sites over the period 1953-1963 were calculated, for several lags. For each pair of sites, the lag corresponding to the maximum cross correlation was found. The calculations were performed separately for the flows within each calendar month. The resulting lags for January and for July are given in table 5.

These lags correspond in the model to lags between jumps of

Table 5: Lags between sites at which maximum cross correlation is obtained, East Anglia 1953-1963

Preceding / Succeeding Site	January								July							
	32/1	32/2	32/3	32/4	32/6	32/7	32/8	33/5	32/1	32/2	32/3	32/4	32/6	32/7	32/8	33/5
32/1		- 1	- 1	- 1	- 1	- 1	- 1	- 1		- 2	- 4	- 3	- 4	- 4	- 3	- 2
32/2			0	0	0	0	0	0			- 1	- 1	0	- 2	0	0
32/3				0	0	0	0	1				0	0	0	0	2
32/4					0	0	0	0					0	0	0	1
32/6						0	0	1						0	0	2
32/7							0	1							0	2
32/8								1								1
33/5																

simultaneous events at the various sites, which are assumed to be fixed; see section 4.2. From table 5 an ordering of the starts of pulses at the various sites can be obtained, and the lags can be estimated. The lags between jumps need not be integer in the model but in practice they are assumed to be so. Note that contradictions in the ordering might arise from matrices of the type given in table 5. However, no contradictions arose here.

For January, from table 5, one can see that events at site 32/1 precede the events at all other sites, while events at site 33/5 precede the events of most of the other sites, and the events at the other six sites seem to be simultaneous. An appropriate choice of fixed lags seems to be no lag for sites 32/1 and 33/5, and a lag of one day for the remaining sites. For July an appropriate choice is no lag for site 32/1, a lag of two days for sites 32/2 and 33/5, a lag of three days for sites 32/4 and 32/8, and a lag of four days for sites 32/3, 32/6 and 32/7.

As suggested in section 4.2, the data of the various sites can be shifted, so as to obtain zero lags between all sites, before estimating the model parameters. This was not carried out here.

7.5 Application of the Single Shot Noise Model

The single shot noise model described in sections 6.1 - 6.5 was applied to the East Anglia data, 1953 - 1963, and synthetic data was generated from the model. Some of the results are presented in tables 6 and 7. Table 6 summarises the fitting of model parameters for the river Nene at Orton. The estimates of the three data properties preserved, and the estimates of the parameters ν, θ and b are given for each month. Note that ν and θ are subject to

Table 6: Parameter fitting of the single shot noise model, Nene at Orton 1953-1963

Month	μ	σ	$\rho(1)$	ν	Θ	b	$1/\nu$
Jan	18.63	13.52	0.831	0.494	10.76	0.285	2.05
Feb	16.80	10.23	0.809	0.795	6.93	0.328	1.26
Mar	13.23	9.62	0.804	0.573	7.80	0.338	1.75
Apr	8.22	5.16	0.819	0.705	3.59	0.308	1.42
May	4.67	2.91	0.623	1.536	2.31	0.758	0.65
Jun	4.19	4.67	0.619	0.488	6.62	0.770	2.05
Jul	3.44	5.28	0.681	0.213	9.81	0.607	4.70
Aug	3.16	2.03	0.728	1.029	1.52	0.497	0.97
Sep	3.08	3.12	0.725	0.418	3.71	0.503	2.39
Oct	4.91	5.46	0.794	0.257	6.82	0.358	3.89
Nov	11.37	10.78	0.740	0.450	11.90	0.471	2.22
Dec	13.38	11.34	0.820	0.384	10.62	0.305	2.60

Table 7: Comparison of synthetic and historical data, single shot noise model, Nene at Orton 1953-1963

Month	Mean	Standard Deviation	Skew	ρ(1)	ρ(2)	ρ(3)	Ï
Historical Data							
Jan	18.63	13.52	1.83	0.831	0.607	0.450	2.20
Feb	16.80	10.23	1.71	0.809	0.635	0.458	2.22
Mar	13.23	9.62	2.34	0.804	0.537	0.396	2.30
Apr	8.22	5.16	2.77	0.819	0.605	0.513	2.73
May	4.67	2.91	5.09	0.623	0.390	0.345	2.13
Jun	4.19	4.67	5.01	0.619	0.389	0.226	2.20
Jul	3.44	5.28	6.77	0.681	0.459	0.327	2.16
Aug	3.16	2.03	2.83	0.728	0.421	0.200	2.21
Sep	3.08	3.12	4.98	0.725	0.449	0.261	2.10
Oct	4.91	5.46	4.43	0.794	0.622	0.426	2.10
Nov	11.37	10.78	2.42	0.740	0.462	0.286	2.21
Dec	13.38	11.34	2.11	0.820	0.574	0.366	2.07
Synthetic Data							
Jan	17.90	13.76	1.66	0.822	0.594	0.414	2.96
Feb	16.23	10.26	1.17	0.830	0.596	0.412	2.61
Mar	13.92	9.04	1.03	0.793	0.567	0.397	2.74
Apr	8.51	5.64	1.40	0.841	0.646	0.500	2.67
May	4.81	3.08	1.42	0.640	0.295	0.121	2.35
Jun	4.39	4.68	1.80	0.608	0.285	0.150	3.02
Jul	3.76	5.26	2.42	0.668	0.353	0.232	3.92
Aug	3.20	2.08	1.09	0.739	0.443	0.247	2.32
Sept	3.22	3.29	1.81	0.753	0.470	0.284	3.01
Oct	5.12	5.54	1.96	0.773	0.534	0.354	3.56
Nov	10.90	9.95	1.74	0.714	0.430	0.270	2.69
Dec	12.78	11.44	1.86	0.836	0.609	0.439	2.93

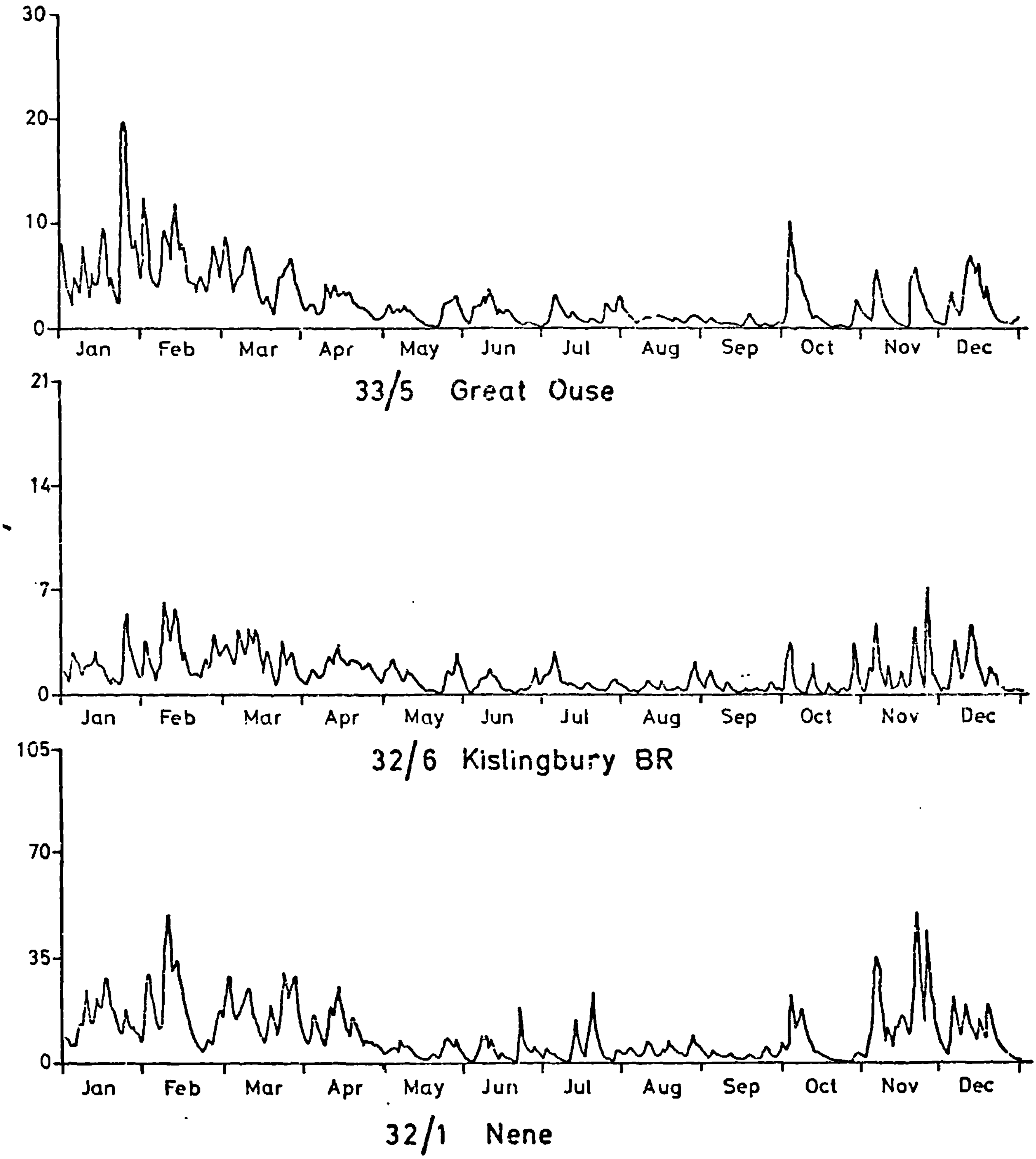

Figure 5. Single shot noise model, synthetic daily data.

large fluctuations between months, which seems to be a shortcoming of the estimation method.

In table 7, the historical data are compared with a sequence of thirty-five years of synthetic data. The values of μ, σ and $\rho(1)$ are preserved by the model and any differences are due to sampling errors, and, to a lesser extent, to transient effects between months.

Quite good agreement is obtained for the skewness for some of the months and for $\rho(2)$ and $\rho(3)$, the lag 2 and lag 3 serial correlation coefficients. The last column in table 7 lists the averages of the sequences of interval lengths, defined in section 7.2, and the agreement between the two sequences seems reasonable. The interval lengths should also be compared with the column listing the expected times between jumps, $1/\nu$, in table 6.

Figure 5 illustrates synthetic flows for a period of one year at three of the sites. Comparison with figure 4 shows that the synthetic data has too many rises and recessions, and that the flow decays too rapidly towards zero.

7.6 Examination of the Serial Correlation of Streamflow

Table 8 lists in the first column some sample values of serial correlations for data from the Nene at Orton, 1953 - 1963. The values $\rho(1)$ to $\rho(10)$ are serial correlation coefficients of daily flows in January, for lags of one to ten days. The value ρ_T is the first serial correlation coefficient of the standardised series of monthly flows over the period 1953 - 1963. The other four columns list the theoretical values of these coefficients obtained by four different models.

Table 8: Serial correlation coefficients of historical data and of some fitted models, Nene at Orton 1953-1963, January

Sampling Interval	Correlation Coefficient	Historical Data	A.R.1	A.R.2	Single Shot Noise	Double Shot Noise
Daily Data	ρ (1)	.831	.831	.831	.831	.831
	ρ (2)	.607	.690	.607	.644	.639
	ρ (3)	.450	.574	.416	.499	.512
	ρ (4)	.345	.477	.275	.387	.425
	ρ (5)	.259	.396	.178	.300	.364
	ρ (6)	.179	.329	.114	.233	.323
	ρ (7)	.123	.274	.072	.181	.294
	ρ (8)	.048	.227	.045	.140	.272
	ρ (9)	.0017	.189	.028	.109	.256
	ρ (10)	.0223	.157	.017	.084	.244
Monthly Data	ρ_T	.51	.108	.0445	.0662	.44

Let $\{X_t\}$ $(t=1,\ldots,N)$, denote daily streamflow. The models are

(a) A discrete time first order autoregressive process

$$X_t = 0.831X_{t-1} + \varepsilon_t \tag{7.1}$$

which preserves the first daily serial correlation coefficient $\rho(1)$.

(b) A discrete time second order autoregressive process

$$X_t = 1.055X_{t-1} - 0.270X_{t-2} + \varepsilon_t \tag{7.2}$$

which preserves the first and second daily serial correlation coefficients $\rho(1)$ and $\rho(2)$.

(c) A single shot noise model which preserves $\rho(1)$, i.e. X_t is assumed to be the average of $X(t)$ over $(t-1,t)$, where

$$X(t) = e^{-0.285s}X(t-s) + \eta_s(t). \tag{7.3}$$

(d) A double shot noise model which preserves $\rho(1)$ and preserves $\rho_T = 0.44$, i.e. X_t is assumed to be the average of $X_1(t) + X_2(t)$ over $(t-1,t)$ where

$$\begin{aligned} X_1(t) &= e^{-0.41s}X_1(t-s) + \zeta_s(t), \\ X_2(t) &= e^{-0.023s}X_2(t-s) + \xi_s(t). \end{aligned} \tag{7.4}$$

The value $\rho_T = 0.44$ is preserved rather than the sample value of 0.51, because of the fitting procedure; see section 6.9.

Under the assumption that X_t is a discrete time autoregressive process, a statistical test based on the sample values of $\rho(1)$ and $\rho(2)$ rejects model (a), and thus model (b) or a higher order model is needed; for details of the test see Jenkins and Watts (1968, section 5.4).

The same test can be applied as an approximate test based on the sample values of $\rho(1)$ and $\rho(2)$, for models (c) and (d), and both are found acceptable.

The three models (a), (b) and (c), however, are unacceptable if the sample value of ρ_T, the monthly first serial correlation coefficient is considered. This confirms the need for the double shot noise model as discussed in section 6.7.

7.7 Estimation of the Parameters of the Double Shot Noise Model

The double shot noise model was fitted to six of the sites, by the method described in section 6.9. The properties of the data considered for preservation are μ the mean flow, σ_1^2 the variance of daily flows, σ_2^2 the variance of monthly flows, ρ_1 the first serial correlation coefficient of the daily flows and ρ_2 the first serial correlation coefficient of the monthly flows. The sample values of μ, σ_1, σ_2 and ρ_1 were estimated separately for each site and for each month. The sample values of ρ_2 were estimated separately for each site, using standardised monthly flows.

Table 9 lists the values obtained for σ_1/μ, σ_2^2/σ_1^2, ρ_1 and ρ_2, and their averages over all six sites. These averages together with an estimated value of ν_1 were used to estimate ν_2, θ_2 and b_2, the parameters of the slow process, for each month. Note that the variability of the values in table 9, between the various sites, is quite small, which justifies the use of the same slow process at all sites.

The values of ν_1 used in estimating ν_2, θ_2 and b_2 are listed in table 10. These values were estimated by counting visually the number of rises and recessions in the streamflows at the six sites. This method is too subjective, and it is suggested that ν_1 may be

Table 9: Sample estimates of properties used in fitting the double shot noise model, 6 sites in East Anglia 1953-1963

	Site / Month	32/1	32/3	32/4	32/6	32/8	33/5	Average
$\frac{\sigma_1}{\mu}$	Jan	.971	.921	.995	.910	.882	.911	.932
	Feb	.766	.736	.715	.807	.653	.683	.727
	Mar	.878	1.013	.903	.772	.770	.888	.871
	Apr	.725	.760	.541	.573	.623	.677	.650
	May	.864	1.012	.838	.858	.876	.894	.890
	Jun	1.346	1.019	1.132	.996	1.230	1.236	1.160
	Jul	1.950	1.399	2.066	.867	1.259	1.064	1.434
	Aug	.967	1.097	.870	.849	1.170	1.076	1.005
	Sep	1.160	1.693	1.130	1.035	1.524	1.147	1.282
	Oct	1.619	1.700	1.637	1.563	1.691	1.847	1.676
	Nov	1.505	1.498	1.420	1.471	1.346	1.416	1.443
	Dec	1.294	1.320	1.226	1.094	1.055	1.078	1.178
$\frac{\sigma_2^2}{\sigma_1^2}$	Jan	.442	.284	.326	.325	.334	.265	.329
	Feb	.368	.312	.370	.328	.335	.293	.334
	Mar	.315	.304	.311	.282	.346	.384	.324
	Apr	.248	.441	.302	.272	.292	.395	.325
	May	.478	.284	.281	.280	.377	.363	.344
	Jun	.315	.329	.270	.253	.201	.572	.323
	Jul	.381	.272	.292	.309	.262	.471	.331
	Aug	.358	.209	.284	.278	.235	.525	.315
	Sep	.239	.118	.288	.245	.185	.333	.235
	Oct	.528	.374	.546	.515	.565	.483	.502
	Nov	.602	.521	.532	.470	.578	.356	.510
	Dec	.571	.336	.427	.452	.469	.379	.439
ρ_1	Jan	.831	.694	.775	.724	.738	.719	.747
	Feb	.809	.786	.826	.777	.741	.822	.793
	Mar	.804	.656	.749	.673	.711	.771	.727
	Apr	.819	.855	.803	.798	.872	.798	.824
	May	.623	.672	.692	.721	.693	.766	.694
	Jun	.619	.591	.597	.674	.623	.733	.639
	Jul	.681	.634	.566	.600	.655	.778	.652
	Aug	.728	.595	.553	.562	.620	.755	.636
	Sep	.725	.567	.528	.564	.645	.748	.630
	Oct	.794	.729	.568	.498	.634	.810	.672
	Nov	.740	.732	.622	.570	.704	.815	.697
	Dec	.820	.636	.748	.652	.668	.777	.717
ρ_2		.509	.455	.520	.503	.499	.520	.501

Table 10: Values of ν_1, the event rate estimated by inspection, East Anglia 1953-1963

Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
ν_1	.20	.18	.18	.16	.16	.15	.15	.16	.16	.18	.18	.20

Table 11: Parameters of the double shot noise model, Nene at Orton 1953-1963

Month	Sample μ	Sample σ_1	Sample σ_2	Sample ρ_1	Sample ρ_2	ν_1	θ_1	b_1	ν_2	θ_2	b_2	Model σ_2	Model ρ_2
Jan	18.63	18.09	12.02	.831	.509	.111	31.3	.411	.0251	9.43	.0233	10.76	.440
Feb	16.80	12.87	7.80	.809	.509	.139	21.6	.449	.0448	4.06	.0180	7.36	.466
Mar	13.23	11.62	6.52	.804	.509	.122	21.4	.512	.0376	5.25	.0243	6.85	.464
Apr	8.22	5.96	2.97	.819	.509	.121	10.2	.372	.0320	1.13	.0074	3.25	.445
May	4.67	4.03	2.79	.623	.509	.234	10.7	1.689	.0437	2.08	.0286	2.30	.524
Jun	4.19	5.64	3.17	.619	.509	.089	21.3	1.207	.0230	3.31	.0291	2.89	.460
Jul	3.44	6.71	4.14	.681	.509	.045	28.8	.807	.0108	4.99	.0292	3.25	.388
Aug	3.16	3.05	2.28	.728	.509	.136	7.1	.896	.0348	1.70	.0285	1.80	.491
Sep	3.08	3.58	1.75	.725	.509	.158	7.4	.704	.0121	2.15	.0182	1.91	.503
Oct	4.91	7.04	5.77	.794	.509	.109	20.2	1.347	.0094	12.69	.0364	5.59	.487
Nov	11.37	17.10	13.28	.740	.509	.088	58.6	1.579	.0137	21.49	.0363	11.50	.483
Dec	13.38	17.31	13.08	.820	.509	.066	41.9	.596	.0184	16.01	.0338	11.06	.450

Table 12: Cross correlations between daily flows, comparison between sample values and values preserved by the double shot noise model. East Anglia, 1953-1963, January

Sample Values					
	32/1	32/3	32/4	32/6	32/8
32/3	.648				
32/4	.792	.817			
32/6	.706	.774	.921		
32/8	.721	.786	.912	.941	
33/5	.766	.682	.827	.737	.719
Model Values					
	32/1	32/3	32/4	32/6	32/8
32/3	.642				
32/4	.789	.718			
32/6	.675	.819	.739		
32/8	.675	.834	.727	.874	
33/5	.677	.777	.734	.812	.799

thought of as a value used to regulate the fitting of the model rather than as an estimate of some property of the data.

In the second step of the estimation procedure the values of μ, σ_1^2 and ρ_1 and the values of ν_2, θ_2 and b_2 were used to obtain ν_1, θ_1 and b_1, for each site and each month. Table 11 summarises this procedure for the Nene at Orton 1953 - 1963. For each month the sample values of μ, σ_1, σ_2, ρ_1 and ρ_2 and the six estimated parameters are listed. Note that the values of ν_1 which are shown in table 11 and which are used in the model are very different from the values of ν_1 in table 10. When these parameters are used, the model preserves the exact values of μ, σ_1 and ρ_1. The values of σ_2 and ρ_2 are preserved only approximately as can be seen from the last two columns of table 11.

The last step in fitting the model is the estimation of the cross correlation parameters. In table 12, the lag zero cross correlations of the daily flows at the six sites, are listed. The first part of the table shows the sample values obtained from the data. The second part shows the values actually preserved by the model. The differences between the two tables are due to inconsistencies arising in the fitting, as discussed in sections 6.3 and 6.4.

7.8 Analysis of Synthetic Data Produced by the Double Shot Noise Model

The parameters estimated in section 7.7 were used to generate synthetic data from the double shot noise process. Ten sequences of synthetic data were produced, each of which consists of eleven years of synthetic streamflow at six of the sites. This corresponds to the length of the historical record. Tables 13, 14 and 15 present a comparison between this synthetic data and the historical data. For each property which was compared, values for the single historical sequence, and for each of the ten synthetic sequences, were calculated. Tables 13, 14 and 15 contain the historical value, and the average, the minimum and the maximum of the ten synthetic values.

Table 13 contains a comparison of the means and of the standard deviations of the daily flows of the Nene at Orton, for the various months. These two properties are preserved exactly by the model for each month. The discrepancies between the synthetic data and the historical data are due to transient effects in passing from month to month.

To give an example of the transient effect, consider the months December and January. Let μ_1, μ_2 and μ be the means of the stationary fast shot noise process, of the stationary slow shot noise process and of the stationary combined double shot noise process, respectively. Let b_1 and b_2 be the decay rates for January, of the fast process and of the slow process, respectively. The fact that the initial value of January is generated by the parameters of December will yield a transient mean value for January, $E(\overline{X})$, which differs from the stationary mean by

$$\begin{aligned}\mu(\text{Jan}) - E(\overline{X}) &= \frac{1}{b_1T}(1-e^{-b_1T})\{\mu_1(\text{Jan}) - \mu_1(\text{Dec})\} \\ &\quad + \frac{1}{b_2T}(1-e^{-b_2T})\{\mu_2(\text{Jan}) - \mu_2(\text{Dec})\},\end{aligned} \tag{7.5}$$

where T is the period of one month. On substituting the estimated values of the parameters, from table 11, one obtains

$$b_1 = 0.411, \quad \frac{1}{b_1T}(1-e^{-b_1T}) = 0.08,$$

$$b_2 = 0.023, \quad \frac{1}{b_2T}(1-e^{-b_2T}) = 0.71,$$

Table 13: Comparison of historical data with ten sequences of synthetic data, generated by the double shot noise model, means and standard deviations of daily flows, Nene at Orton 1953-1963

Month	Mean				Standard Deviation			
	Historical Data	Synthetic Data			Historical Data	Synthetic Data		
		Average	Minimum	Maximum		Average	Minimum	Maximum
Jan	18.63	17.32	13.54	21.42	18.09	17.39	15.16	23.77
Feb	16.80	16.92	12.14	22.26	12.87	14.00	9.22	17.37
Mar	13.23	14.65	11.29	17.11	11.62	11.63	8.68	16.59
Apr	8.22	12.03	7.98	15.10	5.96	7.51	5.86	9.01
May	4.67	7.96	5.27	9.76	4.03	5.06	3.74	6.60
Jun	4.19	6.18	4.53	7.01	5.64	5.93	4.85	7.07
Jul	3.44	5.04	4.32	6.34	6.71	6.81	5.56	7.57
Aug	3.16	3.71	2.42	5.23	3.05	3.82	1.95	6.81
Sep	3.08	3.61	2.55	4.32	3.58	3.85	2.75	5.01
Oct	4.91	3.59	2.14	6.33	7.94	5.51	3.07	8.90
Nov	11.37	7.68	5.10	11.92	17.10	13.90	10.11	24.25
Dec	13.38	12.58	4.93	22.21	17.31	16.72	6.58	23.09

Table 14: Comparison of historical data with ten sequences of synthetic data generated by the double [illegible] noise model, [illegible] average flows of various durations, [illegible] at [illegible] [illegible]

Duration (months)	Historical Data	Synthetic Data		
		Average	Minimum	Maximum
1	1.182	0.282	0.033	0.697
2	1.229	0.767	0.128	1.485
3	1.293	1.343	0.909	1.937
6	1.744	2.211	1.807	2.798
12	3.415	4.146	2.424	5.357

Table 15: Comparison of historical data with ten sequences of synthetic data generated by the double shot noise model, various properties, Nene at Orton 1953-1963

January

	Historical Data	Model			Synthetic Data		
		Combined	Fast	Slow	Average	Minimum	Maximum
μ - mean daily flow	18.63	18.63	8.48	10.16	17.32	13.54	21.42
σ - standard deviation of daily flow	18.09	18.09	15.24	9.75	17.39	15.16	23.77
$\rho(1)$- serial correlation of daily flow, lag 1	0.831	0.831	0.768	0.985	0.749	0.717	0.805
$\rho(2)$- serial correlation of daily flow, lag 2	0.607	0.639	0.509	0.962	0.464	0.389	0.586
$\rho(3)$- serial correlation of daily flow, lag 3	0.450	0.512	0.337	0.940	0.268	0.124	0.450
$\sigma_{\bar{m}}$ - standard deviation of monthly flow	12.02	10.76	6.25	8.76	10.34	5.13	18.91
$\rho_{\bar{T}}$ - monthly serial correlation coefficient	0.509	0.440	0.044	0.641	0.495	0.300	0.649
ρ_{12} - cross correlation between sites 32/1 & 33/5	0.766	0.677	0.533	1.000	0.589	0.441	0.707
γ_1 - skewness of daily flow	1.83	2.38	3.48	1.92	1.85	0.79	2.98
highest average monthly flow	48.21				* 40.18	22.70	73.63
lowest average monthly flow	2.33				* 4.36	1.01	7.11
ν - event rate			0.111	0.0251			
θ - average jump height			31.3	9.43			
b - decay rate			0.411	0.0233			

* based on 9 sequences of synthetic data only

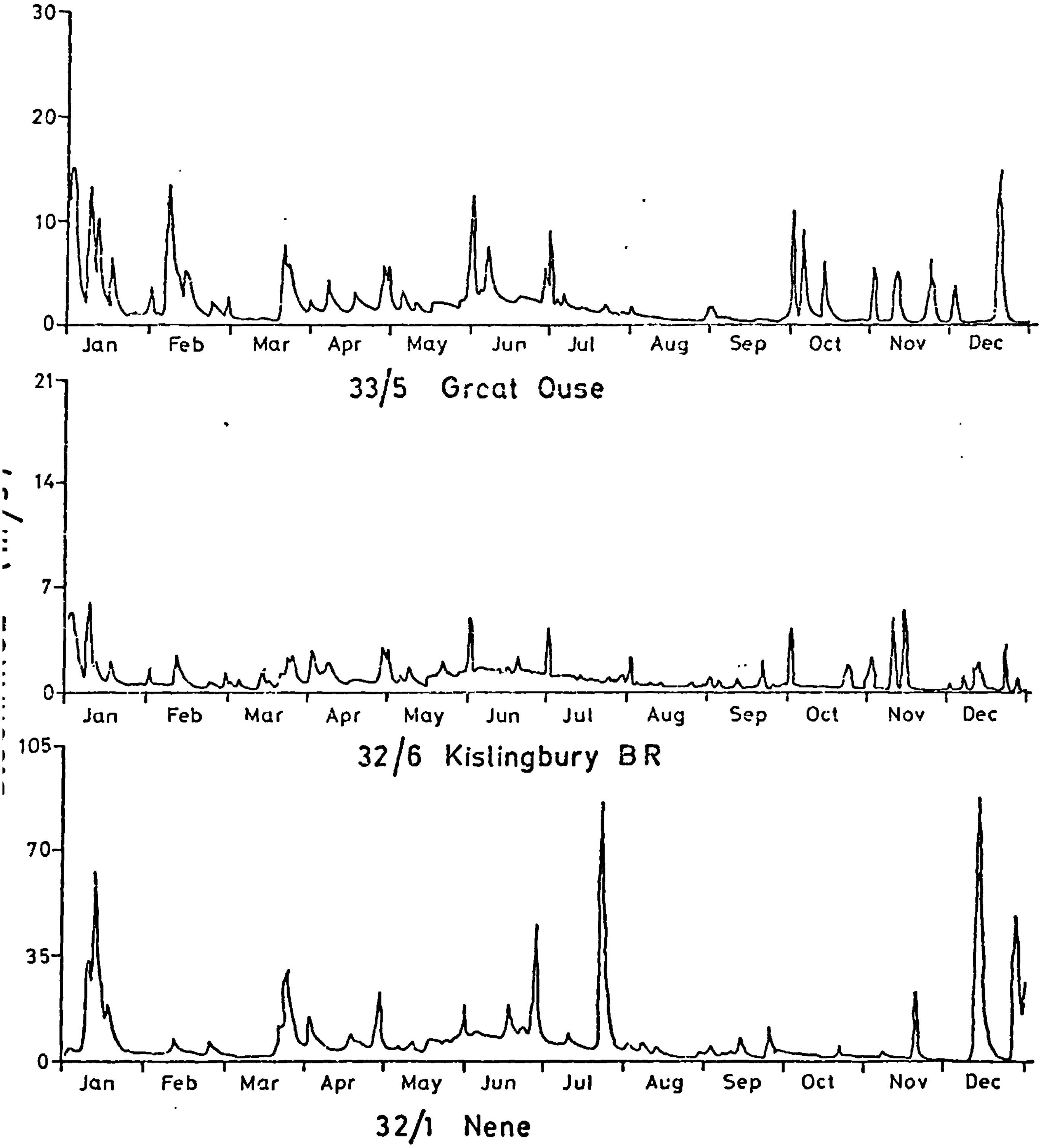

Figure 6. Double shot noise model, synthetic daily data.

$$\mu_1(\text{Jan}) - \mu_1(\text{Dec}) = 8.4 - 4.7 = 3.7$$

$$\mu_2(\text{Jan}) - \mu_2(\text{Dec}) = 10.2 - 8.7 = 1.5$$

and hence

$$\mu(\text{Jan}) - E(\overline{X}) = 0.08 \times 3.7 + 0.71 \times 1.5 = 1.36.$$

Note that most of the discrepancy is caused by the slow process.

Table 14 presents a comparison of periods of low flow occuring in the historical and the synthetic data. These were obtained by observing the sequences of average monthly flows. The comparison shows that for the shorter periods, e.g. the lowest single month in eleven years and the lowest period of two months, the synthetic data produces flows which are too low, while for the longer periods the synthetic flows seem too high. A different conclusion is reached, however, if for each eleven year sequence only the eleven average flows of January are considered. From Table 15, one can see that the lowest January in the record had a flow of 2.83, while for the ten synthetic sequences the lowest January was between 1.01 and 7.11, with a mean value of 4.36. Thus it seems that low flow periods are adequately modelled for January, but apparently not for all the other calendar months.

Table 15 presents a comparison of several properties of historical and synthetic streamflow of the Nene at Orton in January. The expected values of the properties according to the model, are also included. Reasonable agreement exists on all properties. Note however that the daily serial correlation coefficients as well as the daily cross correlation coefficient with site 33/5 appear too low in the synthetic data, when compared with the historical values or with the expected values of the model. This is probably because the estimates of these quantities are biased; see Kendall and Stuart (1968, section 48.3).

Figure 6 gives daily synthetic flows produced by the fitted model, for a period of one year, at three of the sites. Comparison with figure 4, shows similarity between the historical and the synthetic data. A large discrepancy does, however, exist within the summer months, for which the synthetic data exhibit too many high peaks.

The results obtained in the numerical application of shot noise models seem quite encouraging. In particular, the comparison of the historical data with the synthetic double shot noise data of January flows of the Nene at Orton is satisfactory. The computer programs used to fit the double shot noise model to the data are quite complex and include a number of ad hoc procedures. It is hoped that further experience in the use of the programs and some changes in the ad hoc procedures will result in a better fit for summer months as well. One such change might be made as follows. It is noted in table 11 that the daily serial correlation is lower in summer than in winter. On the other hand the monthly serial correlation is estimated as a single value for the whole year, because of lack of data. An ad hoc adjustment of the monthly serial correlations to be lower in summer than in winter, may improve the fitting of the model. The problem of fitting the model to seasonal data needs more consideration. The fitting might be improved by smoothing the estimates of the preserved properties or by smoothing the estimates of the parameters, on an ad hoc basis, or by applying the results of section 4.4. Another area for improvement is in the correction for bias of estimates of serial correlations.

The chief recommendation for the use of shot noise processes to model streamflow is that the various elements of the processes can be given a physical interpretation; see section 2.6. On that account, it seems that the second order shot noise model described in section 6.10 is preferable to the double shot noise model. The double shot noise model was chosen for the numerical application rather than the second order shot noise model because it was assumed, apparently wrongly, that the application of the latter is more complicated.

The assertion that shot noise models have a physical basis must be qualified on two accounts. First, the parameters of the models are estimated purely from the streamflow data, with no attempt to obtain estimates from physical parameters. Secondly, it is certain that the proposed models are only approximate models, representing a much more simplified system than the one creating real streamflow.

The fact that the models are only approximate, justifies the method of estimation employed, i.e. estimation of parameters so as to preserve some of the relevant data properties. Thus, estimation methods suited particularly to shot noise processes were not investigated.

The double shot noise model, proposed here to model daily monthly streamflow. This is necessitated by the fact that daily streamflow has complex properties which are averaged out in monthly data. On the other hand, daily streamflow records contain more information than monthly streamflow records, thus enabling the estimation of more complex models. The inherent impasse, of trying to find models which are realistic and estimable, will only be resolved gradually as more understanding of the physical process of streamflow is reached.

References

Bánkövi, G. (1964). A note on the generation of Beta distributed and Gamma distributed random variates. *Pub. Math. Inst. Hungarian Acad. Sci.* A 9, 555-562.

Bartlett, M.S. (1957). Contribution to discussion of papers by Dr. Gani and Mr. Kendall. *J. R. Statist. Soc.* B 19, 220-221.

Beard, L.R. (1965). Use of interrelated records to simulate streamflows. *Proc. Am. Soc. Civ. Engrs., J. Hydraul. Div.* 91, HY5, 13-22.

Bernier, J. (1970). Inventaire des modeles de processus stochastiques applicables a la description des debits journaliers des rivieres. *Rev. Inst. Int. Statist.* 38:1, 49-61.

Cox, D.R. (1962). *Renewal Theory.* London: Methuen.

Cox, D.R. and Lewis, P.A.W. (1972). Multivariate point processes. *Proc. 6th Berkeley Symp.* 3, 401-448.

Cox, D.R. and Miller, H.D. (1965). *The Theory of Stochastic Processes.* London: Methuen.

Cramér, H. and Leadbetter, M.R. (1967). *Stationary and Related Stochastic Processes.* New York: Wiley.

Crosby, D.S. and Maddock, T. (1970). Estimating coefficients of a flow generator for monotone samples of data. *Water Resources Res.* 6, 1079-1085.

Darling, D.A. and Siegert, A.J.F. (1953). The first passage problem for a continuous Markov process. *Ann. Math. Statist.* 24, 624-639.

Doob, J.L. (1965). *Stochastic Processes.* New York: Wiley.

Dooge, J.C.I. (1959). A general theory of the unit hydrograph. *J. Geophysical Res.* 64, 241-256.

Feller, W. (1971). *An Introduction to Probability Theory and Its Applications.* Vol. II, 2nd Edition, New York: Wiley.

Fiering, M.B. (1964). Multivariate techniques for synthetic hydrology. *Proc. Am. Soc. Civ. Engrs., J. Hydraul. Div.* 90, HY5, 43-60.

Fiering, M.B. (1968). Schemes for handling inconsistent matrices. *Water Resources Res.* 4, 291-297.

Gabriel, K.R. and Neumann, J. (1962). A Markov chain model for daily rainfall occurrence at Tel-Aviv. *Quart. J. R. Meteorological Soc.* 88, 90-95.

Garcia, L.E., Dawdy, D.R. and Mejia, J.M. (1972). Long memory

Resources Res. 8, 1100-1105.

Gilbert, E.N. and Pollack, H.O. (1960). Amplitude distribution of shot noise. *Bell Systems Tech. J.* 39, 333-350.

Hájek, J. (1969). *A Course in Nonparametric Statistics*, San Francisco: Holden-Day.

Hurst, H.E. (1951). Long-term storage capacity of reservoirs. *Trans. Am. Soc. Civ. Engrs.* 116, 776-808.

Hurst, H.E. (1956). Methods of using long-term storage in reservoirs. *Proc. Inst. Civ. Engrs.* 5(5), 519-590.

Jamieson, D.G., Radford, P. and Sexton, J.R. (1973). Hydrological design of water resources systems. *Water Resources Board Pub.* In press.

Jenkins, G.M. and Watts, D.G. (1968). *Spectral Analysis and Its Applications.* San Francisco: Holden-Day.

Jöhnk, M.D. (1964). Erzeugung von betaverteilten und gammaverteilten Zufallszahlen. *Metrika* 8, 5-15.

Jones, R.H. and Brelsford, W.M. (1967). Time series with periodic structure. *Biometrika* 54, 403-408.

Kendall, M.G. amd Stuart, A. (1968). *The Advanced Theory of Statistics.* Vol. 3, 2nd Edition, London: Griffin.

Kraijenhoff van de Leur, D.A. (1966). Runoff models with linear elements. In *Recent Trends in Hydrograph Synthesis.* Proc. Tech. Meeting 21, Committee Hydrological Res. T.N.O., The Hague, 31-64.

Lanning, J.H. and Battin, R.H. (1956). *Random Processes in Automatic Control.* New York: McGraw-Hill.

Mandelbrot, B.B. (1971). A fast fractional Gaussian noise generator. *Water Resources Res.* 7, 543-553.

Mandelbrot, B.B. and Wallis, J.R. (1968). Noah, Joseph, and operational hydrology. *Water Resources Res.* 4, 909-918.

Mandelbrot, B.B. and Wallis, J.R. (1969a). Computer experiments with fractional Gaussian noises, parts 1-3. *Water Resources Res.* 5, 228-267.

Mandelbrot, B.B. and Wallis, J.R. (1969b). some long-run properties of geophysical records. *Water Resources Res.* 5, 321-340.

Mandelbrot, B.B. and Wallis, J.R. (1969c). Robustness of the rescaled range R/S in the measurement of noncyclic long run statistical dependence. *Water Resources Res.* 5, 967-988.

Matalas, N.C. (1967). Mathematical assessment of synthetic hydrology. *Water Resources Res.* 3, 937-945.

Mejia, J.M., Rodriguez-Iturbe, I. and Dawdy, D.R. (1972). The broken line process as a potential model for hydrologic simulation. *Water Resources Res.* 8, 931-941.

Moran, P.A.P. (1959). *Theory of Storage*. London: Methuen.

Moreau, D.H. and Pyatt, E.E. (1970). Weekly and monthly flows in synthetic hydrology. *Water Resources Res.* 6, 53-61.

O'Connell, P.E. (1971). A simple stochastic modelling of Hurst's Law. *Proc. Int. Symp. Math. Models in Hydrology, Warsaw, Int. Ass. Sci. Hydrology,* 1, 327-358.

O'Donnell, T. (1966). Methods of computation in hydrograph analysis and synthesis. In *Recent Trends in Hydrograph Synthesis* Proc. Tech. Meeting 21, Committee Hydrological Res. T.N.O., The Hague, 65-103.

Payne, K., Newman, W.R. and Kerri, D. (1969). Daily streamflow simulation. *Proc. Am. Soc. Civ. Engrs., J. Hydraul. Div.* 95, HY4, 1163-1180.

Parzen, E. (1964). *Stochastic Processes*. San Francisco: Holden-Day.

Quimpo, R.G. (1968). Stochastic analysis of daily river flows. *Proc. Am. Soc. Civ. Engrs. J. Hydraul. Div.*, 94, HY1, 43-58.

Rao, G.R. (1966). Characterisation of the distribution of random variables in linear structural relations. *Sankhyā* A 28, 251-260.

Rice, S.O. (1944). Mathematical analysis of random noise. *Bell Systems Tech. J.* 23, 282-332.

Rice, S.O. (1945). Mathematical analysis of random noise. *Bell Systems Tech. J.* 24, 46-156.

Rodriguez-Iturbe, I., Mejia, J.M. and Dawdy, D.R. (1972). Streamflow simulation, a new look at Markovian models, fractional Gaussian noise, and crossing theory. *Water Resources Res.* 8, 921-930.

Roesner, L.A. and Yevjevich, V.M. (1966). Mathematical models for time series of monthly precipitation and monthly runoff. Hydrology paper no. 15, Colorado State Univ. Fort Collins.

Schaake, J.C., Ganslaw, M.J., Fothergill, J.W. and Harbaugh, T.E. (1972). Multivariate rainfall generator for annual, seasonal, monthly and daily events. *Proc. Int. Symp. Math. Modelling Techniques in Water Resource Systems,* Ottawa, 2, 437-460.

Sexton, J.R. and Jamieson, D.G. (1973a). Improved techniques for water resource systems design. *Symp. Design of Water Resources Projects with Inadequate Data,* Madrid, 1, 375-392.

Sexton, J.R. and Jamieson, D.G. (1973b). Hydrological design of the Wash Estuary storage. *Water Resources Board Publ.,* in press.

Stein, C. (1962). Confidence sets for the mean of a multivariate normal distribution. *J. R. Statist. Soc.* B 24, 265-296.

Takács, L. (1956). On secondary stochastic processes generated by recurrent processes. *Acta. Math. Hungarica* 7, 17-28.

Thomas, H.A. and Fiering, M.B. (1962). Mathematical synthesis of streamflow sequences for the analysis of river basins by simulation. In *Design of Water Resources Systems*, by Maass, A., et al., London: MacMillan, 457-493.

Todorovic, P. and Zelenhasic, E. (1968). The extreme values of precipitation phenomena. *Bul. Int. Ass. Sci. Hydrology* 13, 7-24.

Valencia, D.R. and Schaake. J.C. (1973). Disaggregation processes in stochastic hydrology. *Water Resources Res.* 9, 580-585.

Water Resources Board (1968). *Surface Water Yearbook of Great Britain, Supplement 1965*, H.M.S.O.

Water Resources Board (1971). *Surface Water Yearbook of Great Britain 1965-1966*, H.M.S.O.

Yakowitz, S. (1972). A statistical model for daily streamflow records with application to the Rillito river. *Proc. Int. Symp. Uncertainty in Hydrology and Water Resources Systems*, Tucson, Arizona, 1, 273-283.

Young, G.K. and Pisano, W.C. (1968). Operational hydrology using residuals. *Proc. Am. Soc. Civ. Engrs., J. Hydraul. Div.* 94, HY4, 909-923.

Zelenhasic, E. (1970). Theoretical probability distribution for flood peaks. Hydrology paper No. 42, Colorado State Univ., Fort Collins.