SOME ASPECTS OF QUEUEING AND STORAGE PROCESSES

A thesis in partial fulfilment of the
requirements for the degree of Master of Science
in statistics at
Massey University

Trevor Milton CRAIG

March 1986

ABSTRACT

In this study the nature of systems consisting of a single queue are first considered. Attention is then drawn to an analogy between such systems and storage systems. A development of the single queue viz queues with feedback is considered after first considering feedback processes in general. The behaviour of queues, some with feedback loops, combined into networks is then considered. Finally, the application of such networks to the analysis of interconnected reservoir systems is considered and the conclusion drawn that such analytic methods complement the more recently developed mathematical programming methods by providing analytic solutions for sub systems behaviour and thus guiding the development of a system model.

CONTENTS

ERRATA

Page 53 line 2 should read as follows:
'the reservoir running dry during the given period [o,T]. Extending the interval to [0,T'] indicates that a dam'

In the appended lists of references for chapters 2 and 4, the name PHATARFOD, R.M. is misspelt as PHATARFOD, E.M. and as PHATAFOD, R.M.

CHAPTER 1 - QUEUEING SYSTEMS

1.1 INTRODUCTION

When customers arrive at a station where a particular service is offered, a queue of customers can form if the demand for service exceeds the ability of the server to supply the service immediately. A queueing system is formed when one or more such demand/supply structures operate in conjunction with one another. Although queueing systems often appear in which people are the customers, e.g. the queue formed by people waiting for service from a teller in a bank, the same demand/supply structure can be recognised in more diverse systems. One such example is a telephone system in which the placing of a telephone call corresponds to the arrival of a 'customer' and the 'service' provided is the provision of a telephone circuit for the duration of the call. When the call is completed the caller vacates the system by freeing the circuit for use by other callers. If the latter had rung during the initial call, they would have formed a queue waiting for the circuit to become free, or would have balked on finding the line was 'busy'. If they had found the line was busy for longer than they cared to wait, they could have cancelled their call, thus reneging from the queue. The terms customer and server are thus often applied figuratively. The rapid development of telephone systems at the turn of the century led to a need by telephone engineers for rules for determining the number of connecting lines, operators, etc, in order to handle the demand adequately but in an economical way. Until this time, rules of thumb based on experience had been applied, but the increasing demand for telephone services and complexity in telephone systems necessitated a more systematic approach. This was initiated in 1917 by A.K. Erlang who used probability distributions to describe the variation in the number of calls arriving/unit time and the variation in the length of these calls. He was thus able to determine a distribution function for the number of calls waiting and the distribution of waiting times. The probability distributions used by Erlang were either negative exponential or constant. Later work extended his results to other distributions for which the mathematical treatment proved to be much less tractable. Saaty (1961) gives details of this earlier development and an extensive bibliography up to 1961. Kleinrock (1976) also covers this material as an introduction to more recent work on the behaviour of more elaborate queueing systems viz computer networks.

The large variety of queueing networks as well as, within each queue, the infinity of combinations of interarrival time distributions, service time distributions, queue disciplines and number of servers has led to a large queueing theory literature (see 1.6). The development of digital computers has enabled systems, which because of their complexity were difficult to treat analytically, to be studied by simulation methods. Again, because of the variety of models and methods to be considered, a large literature on computer simulation has developed to which reference will be made later. In this study use is made of analytic and simulation methods to investigate the behaviour of a hydro-storage system. This is modeled as a queue-network incorporating a feedback mechanism between separate queues.

Although arrival and service processes in real queueing systems are generally found to be stochastic in nature, it is useful to consider first the limiting case in which the timing of future events is known.

1.2 DETERMINISTIC QUEUEING MODELS

Queueing models in which the interarrival times and service times are constant (i.e. deterministic) are useful as approximations to real queues in which, for a period at least, the variation in the interarrival and service times is limited. Such models, by being conceptually simpler than stochastic models, also enable a clearer view of the interaction between the arrival and service streams to be obtained. Figure 1 below illustrates a queueing system which is empty at time t=o and which has first-in-first-out (FIFO) queueing discipline. Also the arrival rate is less than the service rate (in order that the queue length should not keep increasing) and as each service is completed a new one is begun. Then clearly the number in the system n(t) = (no. of arrivals in (o,t]) - (no. of services completed in (o,t])

$$= [\lambda t] - [\mu t - \frac{\mu}{\lambda}] \quad \text{where } [x] = \text{integer part of } x,\ x \geqslant o$$

If the system size is limited to k-1 say, then this equation is valid only until time t_i where $n(t_i) = k$. Any customers arriving until the end of the current service will balk and the system size will remain at k-1. At the time of the next service completion, n(t) will drop to k-2, unless an arrival occurs at the same instant in which case n(t) remains at k-1. Arrival and service completion events coincide if and only if $^1/_\lambda$ is a multiple of $^1/_\mu$

i.e.

$$n(t) = \begin{cases} 0 & (t < 1/\lambda) \\ [\lambda t] - [\mu t - \mu/\lambda] & (1/\lambda \leq t < t_i) \\ k-1 & (t \geq t_i) \end{cases} \qquad (1.1)$$

Since in the example shown on Figure 1, $\frac{1}{\lambda} = 4$, $\frac{1}{\mu} = 8$ and k=5 Equation (1.1) shows that in this case $t_i = 32$ and so for $t \geq 32$ the system is in a steady state.

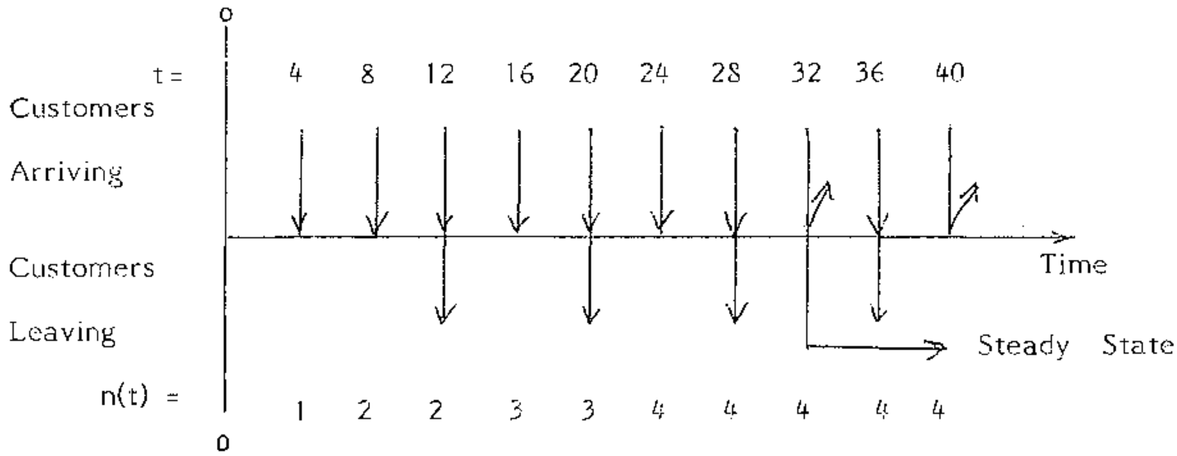


Figure 1

In addition to n(t), the number in the system at time t, another important measure of the queue performance is the waiting time experienced by customers forced to join the queue. Writing $W_q^{(n)}$ for the time spent in the queue by the n^{th} customer, $S^{(n)}$ for the same customer's service time and $T^{(n)}$ for the time between the arrivals of the n^{th} and $(n+1)^{th}$ customers the following recurrence relation exists between these values:

$$W_q^{(n+1)} = \begin{cases} W_q^{(n)} + S^{(n)} - T^{(n)} & (W_q^{(n)} + S^{(n)} - T^{(n)} > 0) \\ 0 & (W_q^{(n)} + S^{(n)} - T^{(n)} \leq 0) \end{cases} \qquad (1.2)$$

This is illustrated in Figure 2 and holds whether the times $S^{(n)}$ and $T^{(n)}$ are deterministic or stochastic:

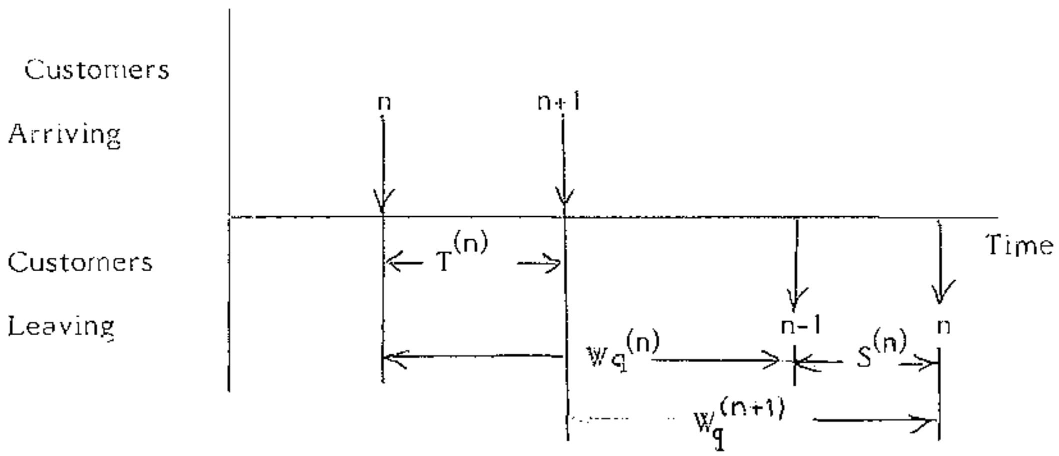


Figure 2

In the present example for n<8 equation (1.2) yields $W_q^{(n)}$ as follows:

$S^{(n)} = 8$ and $T^{(n)} = 4$ so that $W_q^{(n+1)} = W_q^{(n)} + 4$

i.e. $\Delta W_q^{(n)} = W_q^{(n+1)} - W_q^{(n)}$

$= 4$

so $W_q^{(n)} = 4n-4$

since $W_q^{(1)} = 0$

If n ≥ 8 each arrival (which does not balk) finds k-2 customers already in the system and each requiring a service time $S^{(n)} = 8$ thus if k=5,

$$W_q^{(n)} = \begin{cases} 4(n-1) & n < 8 \\ 24 & n \geq 8 \end{cases}$$

In the case in which $1/\mu$ is not a multiple of $1/\lambda$ a diagramatic method as used in Figure 1 reveals that the system size undergoes a cyclic pattern of change, the length of the cycle being equal to the least common multiple of $1/\mu$ and $1/\lambda$. Further complications can be introduced by having the system start off in a non-empty state or by changing the queue discipline, etc. Exact solutions are always obtainable by graphing however since all factors are deterministic. As the number of complications increases or the queues are combined into networks, graphical methods become more cumbersome and methods of approximation, to be considered later in connection with stochastic queueing models, become appropriate.

1.3 GRAPHICAL REPRESENTATION OF CUMULATIVE FUNCTIONS

Although the diagrams in section 1.2 clarify the relationships between individual arrival and service events, the nature of cumulative processes over a period of time is less clear. Three cumulative processes are of particular interest:

A(t) = Cumulative quantity or number to arrive by time t,

D(t) = Cumulative quantity or number to enter service by time t,

$D^*(t)$ = Cumulative quantity or number to have left service by time t;

all of these functions are monotonic non-decreasing.

Clearly Q(t) = quantity in queue or queue length at time t

$= A(t) - D(t)$

and S(t) = quantity or number in service

$= D(t) - D^*(t)$

so that $A(t) \geqslant D(t) \geqslant D^*(t)$ since $Q(t) \geqslant o$, $S(t) \geqslant o$

These relationships hold for any queue discipline or number of servers or number of servers and are illustrated in Figure 3 for the example in Section 1.2.

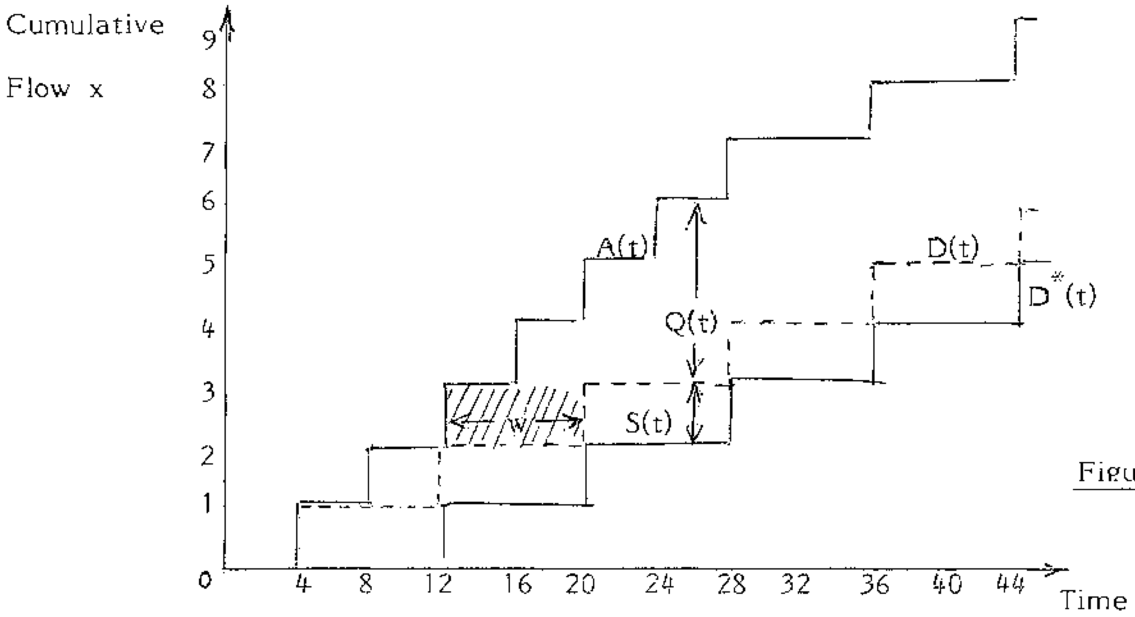


Figure 3

For FIFO queue discipline the horizontal distance marked w in Figure 3 represents A(12)-D(20) i.e. w = time spent in the queue by the 3rd customer. The height of the shaded rectangle is 1 unit, so that waiting time accumulated by all customers up to time t is equal numerically to the area between the A(t) and D(t) curves. For non FIFO queue disciplines or queues with more than one server, the simple interpretation of the length marked w in Figure 3 may not be valid and an alternative approach is needed. This is achieved by considering x as the independent variable and constructing as in Figure 4 a graph of $\Delta(x)$ where $\Delta(x)$ is the time of departure from the queue of x^{th} cumulative arrival, i.e. for $j-1 < x < j$, $\Delta(x)$ = departure time from the queue of the j^{th} arrival whereas $D^{-1}(x)$ was the time of the x^{th} cumulative departure. Thus x is now the label given to the customer on arrival.

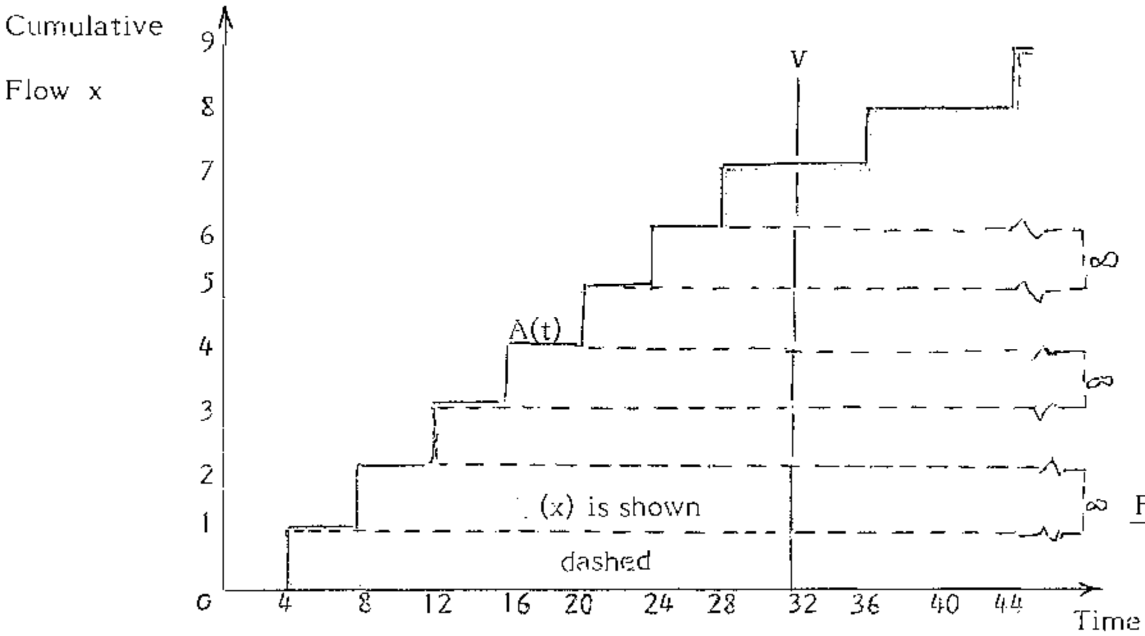


Figure 4

The shape of $\Delta(x)$ depends on the queue discipline and if this is FIFO the graphs $\Delta(x)$ and $D^{-1}(x)$ are identical. In Figure 4 the graph of $\Delta(x)$ corresponds to a last-in-first-out (LIFO) queue discipline applied to the queue graphed in Figure 1. In this case the 2nd, 4th and 6th customers wait in the queue forever for service, all other new arrivals arriving just at the instant a service is completed and so having priority to enter service.

Thus $$\Delta(x) - A^{-1}(x) = \begin{cases} o \text{ if } x \neq 2, 4 \text{ or } 6 \\ \infty \text{ if } x = 2, 4 \text{ or } 6 \end{cases}$$

corresponding to the horizontal measurement w in Figure 3.

A vertical line V drawn at t=32 in Figure 4 indicates that the cumulative total remaining in the queue at time t is now the sum of possibly several segments rather than just one (i.e. Q(t)) as in Figure 3. The same method applies to service completions if there is more than one server. Clearly, graphs such as Figures 3 and 4 can be drawn for stochastic as well as deterministic queues if the arrival and departure times of each customer are known. However, even if only the cumulative functions A(t) and D(t) or (x) are known, useful averages representing the queue behaviour can be found from such graphs. Evaluating the area between the graphs of A(t) and D(t) in Figure 3 using vertical strips and then horizontal strips illustrates Little's formula $L = \lambda w$ (J.D.C. Little, "A Proof for the Queueing Formula : $L = \lambda w$" Operations Research 9 383-387, 1961.) i.e. average queue length = arrival rate x average queue time/customer since in the example $3 = \frac{1}{8} \times 24$. Before considering these graphical methods in the fluid approximation of queues, it is necessary to consider the influence of the introduction of random variation into the arrival and service streams.

1.4 PROBABILISTIC DESCRIPION OF ARRIVAL AND SERVICE PROCESSES

The state of a queueing system in which arrival and/or service times vary in a random way, is not precisely predictable at future moments in time. The probability of a queue having a particular length at a particular time t in the future can be related to the probability of it changing from one length to another during a given interval as follows:

Let $P_t(\ell_1)$ = the prob. that the queue has length ℓ_1 at time t

Let $T_t(\ell_1,\ell_2,a)$ = the prob. of changing from length ℓ_1 to ℓ_2 during the interval $[t,t+a)$,

then by the theorem of total probability,

$$P_{t+a}(\ell_2) = \sum P_t(\ell_1)T_t(\ell_1,\ell_2,a) \quad \text{[summing over all lengths } \ell_1] \tag{1.3}$$

The probability $T_t(\ell_1,\ell_2,a)$ will depend on the number of arrivals and departures during $[t,t+a)$. In turn the number of arrivals will depend on the length of the interval and the time at which the last customer arrived before t, the start of the interval. Similarly, the number of departures will depend on the length of the interval and the length of time any services in progress at the start of the interval had been going on. This dependancy makes solving (1.3) difficult except for particular distributions of arrivals and service times. It is useful therefore to consider the case in which this dependancy is absent.

Let f(t) be the p.d.f. of the interarrival intervals

and $F(t) = \int_0^t f(u)du$ i.e. the d.f. and let $p(t, \delta t)$

be the probability of an arrival during $(t, t+(\delta t)$

$$\text{Then } p(t, \delta t) = \frac{\text{the proportion of the dist. in the interval } (t\ t+\delta t)}{\text{the proportion of the dist. in excess of } t}$$

$$\text{then } p(t, \delta t) = \frac{f(t)\,\delta t}{1-F(t)}$$

If now $P(t, \delta t)$ is independent of t, then $P(t, \delta t) = Kt$ say, where K is a constant.

$\therefore\ K\delta t = f(t)\ \delta t/(1-F(t))$

and hence $f(t) = Ke^{-Kt}$

As it can be shown (Parzen (1962)) that the negative exponential distribution is the only continuous p.d.f. with this Markov property, the central importance of this p.d.f. in queueing theory is apparent. Implied in the above derivation are the following three assumptions which define a Poisson process $\{N(t), t \geq 0\}$

(i) prob. of an arrival during the interval $[t,t+\delta t) = \lambda\delta t + o(\delta t)$

i.e. $p(t, \delta t) = \lambda\delta t + o(\delta t)$ where λ is a constant independant of N(t) and $\lim_{\delta t \to 0} \frac{o(\delta t)}{\delta t} = 0$

(ii) prob. of more than one arrival during $[t, t+\delta t) = o(\delta t)$

(iii) the number of arrivals in non-overlapping intervals are statistically independant. i.e. the process has independant increments. With these assumptions, the probability $P_n(t)$ of $n(n \geq 0)$ arrivals during a time interval of length t can be found by the following method which appears often in queueing theory.

$$P(t+\delta t) = \Pr\{n \text{ arrivals in } t \text{ and zero in } \delta t\}$$
$$+ \Pr\{n-1 \text{ arrivals in } t \text{ and } 1 \text{ in } \delta t\}$$
$$+ \Pr\{n-2 \text{ arrivals in } t \text{ and } 2 \text{ in } \delta t\}$$
$$+ \dots + \Pr\{0 \text{ arrivals in } t \text{ and } n \text{ in } \delta t\}, \quad n \geq 1 \qquad (1.4)$$

by (i), (ii) and (iii) (1.4) becomes

$$P_n(t + \delta t) = P_n(t)[1 - \lambda\delta t + o(\delta t)] + P_{n-1}(t)[\lambda\delta t + o(\delta t)] + o(\delta t) \qquad (1.5)$$

where the last term represents $P_n\{n-j \text{ arrivals in } t \text{ and } j \text{ in } \delta t;\ 2 \leq j \leq n\}$

If n=o then $P_o(t + \delta t) = P_o(t)[1-\lambda\delta t - o(\delta t)]$ (1.6)

Rewriting (1.5) and (1.6) and taking limits as $\delta t \to 0$ gives

$$\frac{dP_o(t)}{dt} = -\lambda P_o(t)$$

and

$$\frac{dP_n(t)}{dt} = -\lambda P_n(t) + \lambda P_{n-1}(t), \quad n \geq 1$$

The general solution of these linear first order differential equations using the boundary conditions $P_o(0) = 1$ and $P_n(0) = 0$, $n \geq 1$ is

$$P_n(t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t} \qquad n \geq 0 \qquad (1.7)$$

From (1.7) it follows that $E[N(t)] = \lambda t$ i.e. N(t) has a mean arrival rate λ. That the Poisson process has stationary increments i.e. for $t > s$, $N(t) - N(s)$ and $N(t+h) - N(s+h)$ are identically distributed, is seen by noting that assumption (iii) i.e. independent increments implies that there is no loss of generality if N(s) and N(s+h) are assumed to be zero. If the above derivation is carried out under assumptions (i), (ii) and (iii) the same formula results for N(t) as for N(t+h).

The close association between the Poisson process and the exponential distribution can be seen from (1.7) as follows. If T is the random variable 'time between successive arrivals' then

$$Pr\left\{T \geq t\right\} = Pr\left\{\text{zero arrivals in time } t\right\} = P_o(t) = e^{-\lambda t}$$

Letting A(t) be the d.f. of T, it follows that

$$A(t) = Pr\left\{T \leq t\right\} = 1-e^{-\lambda t}$$

thus T has the exponential distribution with mean $^1/\lambda$, which is intuitive since the mean arrival rate is λ. Conversely it can be shown that if the interarrival times are independent and have the same exponential distribution, then the arrival rate follows the Poisson distribution. This Poisson/exponential arrival process is sometimes referred to as completely random arrivals. This is because of the following property of a Poisson process. Given that k arrivals have occured during an interval [0,T] the k times $T_1 < T_2 < \ldots < T_k$ at which the arrivals occured are distributed as the order statistics of k uniform random variables on [0,T]

This is shown as follows:

Writing $f_{T_1,T_2,\ldots,T_k}(t_1,t_2,\ldots,t_k|k \text{ arrivals in } [0,T])\, dt_1\, dt_2 \ldots dt_k \equiv f_T(t|k)dt$

$$= Pr\left\{t_1 \leq T_1 \leq t_1 + dt_1, \ldots, t_k \leq T_k \leq t_k + dt_k | k \text{ arrivals in } [0,T]\right\}$$

$$= \frac{\lambda dt_1 e^{-\lambda dt_1} \ldots \lambda dt_k e^{-\lambda dt_k} \quad e^{-\lambda(T-dt_1-dt_2-\ldots-dt_k)}}{(\lambda T)^k e^{-\lambda T}/k!}$$

which reduces to $f_T(t/k) = k!/T^k$ which is the joint density function of the order statistics of k random variable on [0,T]

The arrival process considered above can also be used to describe the service pattern if in assumptions (i) - (iii), the term arrival is replaced by service and if the probabilities are conditioned on the system being non-tempty. In addition, the basic Poisson/exponential process can be generalised in several ways which include (a) truncating the infinite range (b) allowing to depend on t, i.e. the process becomes non-homogeneous (c) allowing that more than one occurence in t has probability greater than o(t) i.e. a batch process.

More general renewal processes than the Poisson process can be used to describe arrival/ service patterns as can non-renewal processes. These will be considered later where appropriate but the Poisson process often appears in the description of a queueing system either directly or in an imbedded process or as a first approximation. The reason for this is not just the tractable properties of this process but also because many real-life processes obey at least approximately the requirements (i) - (iii) listed previously. Information theory provides an additional argument. This is that the information content for the distribution f(x), defined as $\int_0^\infty f(x)\log f(x)dx$, is lease for the exponential function and as such provides a conservative description of arrival and service patterns.

WAITING TIME AND BUSY TIME DISTRIBUTIONS

These features of queue behaviour are, unlike the measures of effectiveness, dependant on the queue discipline. It is also noted that the time a ficitious customer would have to wait were he to arrive at an arbitrary point in time, i.e. the virtual waiting time, has a steady state distribution equal to that of the waiting time of an actual customer iff the input is Poisson. For the present model the waiting time distribution $W_q(t)$ is part discrete and part continuous.

$$W_q(t) = \begin{cases} 1-\rho & t=0 \\ 1-\rho e^{-\mu(1-\rho)t} & t>0 \end{cases}$$

The mean waiting time (via a Riemann-Stieltjes integration) is $W_q = \frac{\lambda}{\mu(\mu-\lambda)}$

Similarly for the total time spent in the system,

$$W(t) = (\mu-\lambda)e^{-(\mu-\lambda)t} \quad t>0$$

$$\text{and} \quad W = E[T] = \frac{1}{\mu-\lambda}$$

These results exemplify again Little's formula

$$L_q = \lambda W_q \quad \text{or} \quad L = \lambda W$$

which is valid under much less stringent restrictions than those of the present model. Since a busy period continues as long as there is at least one item in the system $P_0(t)$ is seen to be the d.f. of the busy period and $P_0'(t)$ the p.d.f..

giving $P_0'(t) = \dfrac{2\sqrt{\mu\lambda}\, e^{-(\lambda+\mu)}\, I_1(2\sqrt{\mu\lambda}\, t)}{t}$ and also

$$E[T_{Busy}] = \frac{1}{\mu-\lambda}$$

1.5 PROBALISTIC DESCRIPTION OF QUEUES

The simplest probabilistic queuing model is the single server model with exponential interarrival and service times and having a first-in-first-out (FIFO) queue discipline.

To find an equation relating the state probabilities

$P_n(t) = \Pr\{n \text{ in the system at time } t\}$ a similar method to that used to obtain (1.4) is employed.

Applying the assumptions set out in § 1.4 gives

$$P_n(t+\delta t) = P_n(t)(1 - \lambda\delta t - \mu\delta t) \quad P_{n+1}(t)(\mu\delta t) + P_{n-1}(t)(\lambda\delta t) + o(\delta t),\ n \geq 1$$

Similarily $P_o(t+\delta t) = P_o(t)(1-\lambda\delta t) + P_1(t)(\mu\delta t) + o(\delta t)$

Rewriting and taking limits as $\delta t \to o$ gives

$$\frac{dP_n(t)}{dt} = -(\lambda+\mu)P_n(t) + \mu P_{n+1}(t) + \lambda P_{n-1}(t),\ n \geq 1$$

$$\frac{dP_o(t)}{dt} = -\lambda P_o(t) + \mu P_1(t) \qquad (1.8)$$

Bailey (1956) gave the following solution to (1.8)

$$P_n(t) = e^{-(\lambda+\mu)t}\Big[(\mu/\lambda)^{(i-n)/2} I_{n-i}(2\sqrt{\lambda\mu}\,t) + (\mu/\lambda)^{(i-n+1)/2} I_{n+i+1}(2\sqrt{\lambda\mu}\,t)$$
$$+ \left(1-\frac{\lambda}{\mu}\right)\left(\frac{\lambda}{\mu}\right)^{n} \sum_{\ell=n+i+2}^{\infty} \left(\frac{\mu}{\lambda}\right)^{\ell/2} I_\ell(2\sqrt{\lambda\mu}\,t)\Big],\ n \geq 1 \qquad (1.9)$$

where i is the number in the system at time t=o, and

$I_n(y) = k^{-n} J_n(ky)$ where $J_n(y)$ is the regular Bessel function. Using the assymptiotic approximation $I_n(y) \sim \dfrac{e^y}{\sqrt{2\pi y}}$ it can be shown that as $t \to \infty$, $P_n(t) \to (1 - \frac{\lambda}{\mu})(\frac{\lambda}{\mu})^n$ (1.10)

The symbol ρ is often used to represent the rate which although dimensionless is often given in 'erlangs' in honour of A.K. Erlang. Thus (1.9) becomes $P_n = (1-\rho)\rho^n, \rho<1]$

Considering the complexity of (1.9) and its derivation, in this the simplest of probabilistic queueing models, it is fortunate that it is often the limiting distribution which is of most interest. By taking the limit as $t \to \infty$ in (1.8), the resulting equations can be used to determine (1.10) directly. There are a number of ergodic theorems which consider the existence of steady-state solutions but, as in this case, the conditions under which a queueing process is ergodic often becomes apparent from other considerations.

Having determined the steady-state probability distribution of the system size, representative characteristics of the system called measures of effectiveness can be calculated. These include the expected number in the system (L) and the expected number in the queue, (L_q) (i.e. customers actually waiting - excluding the customer being served).

$L = E[N] = \sum_{n=0}^{\infty} n P_n = \frac{\rho}{1-\rho}$ Similarly, $L_q = E[N_q] = \frac{\rho^2}{1-\rho}$

Also of interest is the expected size of non-empty queues i.e. $L'_q = E[N_q \mid N_q \neq 0]$

Let $P'_n = Pr\{\text{n in the system} \mid n \geq 2\}$

then $L'_q = \sum_{n=2}^{\infty} (n-1) P'_n = \frac{1}{1-\rho}$

The classification of more general queue models was facilitated by a notation due to Kendall.

1.6 KENDALLs NOTATION

D.G. Kendall (1953) introduced a notation later modified to the following form A|B|m|K|L to clarify different queue models. In the notation A and B represent the distribution function of the inter arrival time and the service time respectively. m represents the number of servers, K the system capacity (queue plus service) and L the size of the customer population. The arrival and service streams are considered to be sequences of random variables having independent and identical distributions. Morse (1957) discusses sampling of a queueing system to obtain the A and B distributions and describes the Erlang and hyper-exponential distributions which provide a reasonable representation of sampling distributions found in practice. The letters M = Markov (i.e. Poisson process), G = general distribution, D = deterministic are used in positions A and B in the notation.

The moments and often the distributions of many of the random variables which are involved in the description of queue behaviour of many different queue models have been determined. The complexity of the derviation of these results tends to increase as either A or B or both differ from a Markov process. Brief references to this work follows.

Page (1972) considers various queueing models which have Erlang, Poisson or deterministic inter arrival or service time distributions. Graphs and tables of system variables are provided. Some comment is made about priority queues with reference to Jaiswal (1968) for a more complete discussion. Takacs (1961) in addition to batch arrival processes discusses the application of queueing models to particle counting and considers queues with infintiely many servers. Riordan (1962) includes a discussion on virtual waiting time and queue disciplines other than FIFO.

Morse (1957) considers the derviation of sampling distributions from arrival and service processes. Erlang and hyper-exponential distributions are used as models and useful tables of these distributions are provided. Saaty (1961) considers the ergodic properties of queues and in addition to a study of queues with Poisson and non-Poisson input and service processes, provides an interesting discussion of less common queueing models. These include cooperating parallel channels and cyclic queues. A final chapter indicates the wide range of problems to which queueing models have been applied including semi-conductor noise, hospitals and the demand for medical care, as well as an introduction to dams and storage systems which is further considered in Chapter 2. Kosten (1973) considers the M|G|m model under several different queue disciplines and queue size restrictions.

An introduction to computer simulation methods includes a comparison of simulation and analytic methods in investigating the behaviour of queues. Gross and Harris (1974) provide a systematic coverage of Markovian queue models. They then procede to a discussion of semi-Markovian and ergodic processes in queues with general arrival and/or service processes as well as the use of approximation methods with such models. A chapter on computer simulation and simulation languages is followed by a final chapter detailing a case study involving queueing theory and simulation which provides and example of optimizing a queueing model using a cost critereon. Kleinrock (1975) Volume 1 provides a fuller discussion of the more general models G|M|m and G|G|1 and in Volume 2 considers queue networks and their application to time-shared computers

NEWELL (1972, 1982) has considered queueing models in an engineering context emphasising approximation and graphical methods. Borǒkov (1976) has presented general and unified mathematical treatments of a quite wide class of queueing models. Kingman (1966) discusses an algebriac approach to generalising treatment of the G|G|m queue model for m 1.Syski (1967) discusses the Pollaczek method in queueing theory and Prahbu (1974 discusses another technique of wide application in queueing theory, the Wiener-Hopf method. In addition to developing unifying treatments and general methods of analysis, more recent papers have treated topics in the optimising of queue operation, queueing network theory, simulation of queues as well as the analysis of queueing models arising in real systems. Reference to papers on these and related topics can be found in the followng and similar publications. Management Science, Operations Research, Information, ORSNZ, Journal of Advanced Probability, Journal of Applied Probability, Naval Research Logical Quarterly Journal. In the present context approximation methods and computer simulation are of particular relevance and are considered below.

1.7 APPROXIMATION METHODS, BOUNDS AND INEQUALITIES

In those cases in which an exact expression can be found for an expected value or the distribution of a variable of interest in a queueing model, it is often difficult to evaluate numerically and further, the assumptions about the conditions under which the result was derived may be difficult to verify. Consequently considerable effort has been spent in developing approximations, bounds and inequalities which are robust to changes from underlying assumptions and which are relatively quick to calculate. In practical situations the control of a queueing system becomes most critical when the traffic density is greatest i.e. as $\rho \to 1$. Kleinrock (1975) Volume 2 gives a discussion of results obtained for the G|G|1 queue model in the heavy traffic case. Central to to these results is that the waiting time distribution is approximately exponentially distributed with mean wait given by: $(\sigma_a^2 + \sigma_b^2)/2(1-\rho)\bar{t}$ where σ_a^2, σ_b^2 are the variances of the interarrival and service times respectively, ρ is the traffic density and $\bar{t}$ is the mean interarrival time. It is also shown that this mean wait forms an assymptotically sharp upper bound for the mean wait in any G|G|1 queue. The case for a lower bound on the mean waiting time is less clear cut and the results obtained depend on the nature of the input process. Bounds on the tail of the distribution of waiting time are given and bounds on the mean waiting time for the

G|G|m (m 1) model are derived. All of these results are approximations and bounds for the exact solution. An alternative procedure is to find exact solutions to an approximation of the original problem. As was noted in § 1.2 the basic recurrence relation for waiting times

$$W^{(n+1)} = \max\left[0, W^{(n)} + S^{(n)} - T^{(n)}\right]$$

holds for deterministic or stochastic arrival and service processes. Consequently an approximation to the system behaviour can be obtained by approximating the stochastic processes which control the operation of the system. A 'first-order' approximation is obtained by replacing the stochastic processes by their possibly time-dependent averages. This is called the fluid approximation method. By allowing each stochastic process to be represented by both its mean and its variance a 'second-order' approximation is obtained. This is termed a diffusion approximation. These methods enable the elementary deterministic methods outlined previously to facilitate the analysis of queues involving complex stochastic processes. To apply these approximation methods it is necessary first to estimate the appropriate parameters from the relevant stochastic processes. Gross and Harris (1974) discuss those aspects of statistical inference which relate to parameter estimation in queueing sytems. Computer simulation of queues provides another method of investigating queueing systems. As simulation amounts to statistical sampling, the approximations and bounds outlined above are helpful also in drawing inferences from the sample statistics produced by the computer simulation. (Fishman, 1974). An application of these methods is described in Chapter 5.

CHAPTER 2 - STORAGE

2.1 INTRODUCTION

One area in which queueing models have been applied is in the study of systems such as dams and inventories, all of which have in common the feature of having a storage facility. The general similarity of the structure of a storage system with its inflow, storage facility and outflow, and that of a queueing system with its inflow of customers, a waiting line and/or service station and the outflow stream of serviced customers is clear. Moran (1954) first formulated the probability theory of storage systems and in his monograph Moran (1959), results obtained up to 1958 are described. Prahbu (1964) discusses time-dependent results in storage theory and includes a bibliography. Ghosal (1970) studies explicitly the similarities between queueing and storage systems and also appends a bibliography. Storage systems in which the input stream variables are not independent have been studied in connection with hydro dams since river flows often show short term correlation. Papers by Pakes (1973) and Phatarfod(1973) give references. Prahbu (1980) presents an overview of stochastic processes involved in storage systems. As in queueing theory the large variety of models and methods has led to a large literature on the probability theory of storage systems. The application of those results appropriate to the network of queues here being considered, is discussed in the following sections.

2.2 AN ANALOGY BETWEEN STORAGE MODELS AND QUEUEING MODELS

Real storage systems often feature interconnected reservoirs with multiple inputs and outputs but in this section only a simple system is considered. This has a single reservoir with storage level Z_t at time t where t is initially taken to be discrete. x_t is the input to the reservoir just after the level Z_t was recorded and y_t is the release from the reservoir at the end of the time period (t,t+1). Then the recurrence relation :$Z_{t+1} = [Z_t+x_t-y_t]^+$ (2.1) follows from the conservation of flow. It is noted that if Z_o and the sequences $\{x_t\}$ and $\{y_t\}$ are known deterministically then $\{Z_t\}$ can also be determined exactly. In general however either or both of the sequences $\{x_t\}$ and $\{y_t\}$ are stochastic or have a stochastic component so that Z_t becomes a random variable.

That $\{Z_t\}$ forms a random walk process with a reflecting barrier at 0 will be considered later. In this section the analogy between (2.1) and equation (1.2) is considered. First it is noted that if the reservoir has a maximum capacity k then (2.1) becomes

$$Z_{t+1} = \min [k, Z_t + x_t - y_t]^+ \qquad (2.2)$$

As noted in Chapter 1, the following recurrence relations hold for a single-server queue under 'first-come first-served' queue discipline:

$$w_{n+1} = [w_n + s_n - t_n]^+ \qquad (2.3)$$

or if waiting customers reneg after waiting a time k

$$w_{n+1} = \min [k, w_n + s_n - t_n]^+ \qquad (2.4)$$

Also if Q_t represents the queue length, A_t the number of arrivals during (t, t+1), B_t the number served during (t, t+1) then

$$Q_{t+1} = [Q_t + A_t - B_t]^+ \qquad (2.5)$$

The correspondences between equations (2.2), (2.3) and (2.5) are set out in Table (2.1)

Table 2.1

Notation	**Storage Systems**	**Waiting Time**	**Queue Size**
Z_t	Storage level at the beginning of the interval (t, t+1)	Waiting time of t^{th} customer	Queue size at arbitrary time t
x_t	Input at the beginning of the interval	Service time of the t^{th} customer	Arrivals during interval (t, t+1)
y_t	Release at the end of the interval	Inter arrival time between t^{th} and $(t+1)^{th}$ customers	Departures after service during interval (t, t+1)

A similar analogy with other variables occuring in queueing models such as the queue size at the time of the t^{th} departing customer or the bulk service model in which y_t represents the number of customers at the time t. In some situations however caution is needed when drawing an analogy between a queueing model and a storage model. For example dams in series do not have an analogue for queues in series.

In equations (2.1) to (2.5) t has been discrete t=0,1,2, ... The distribution function of Z_t may be continuous if x_t and y_t are continuous random variables but will be discrete if x_t and y_t are discrete. The time t can also be considered to be continuous in which case the continuous analogue of equation (2.2) can be written as

$$Z(t+\delta t) = \min [K, Z(t) - \delta X(t) - \delta Y(t)]^{+} \quad (2.6)$$

where $\delta x(t)$ is the inflow during $(t, t+\delta t)$ and $\delta y(t)$ is the outflow during the same period.

Having observed the above analogy between queueing and storage systems, it is appropriate to investigate the behaviour of storage systems with the results and methods established in queueing theory in mind. Three aspects of the description of a storage system are of particular interest.

(i) given the distribution functions of x_t and y_t, determine the distribution function of Z_t

(ii) given the distribution functions of x_t and y_t and the initial value of i.e. $Z_0 = y$ to find the probability that Z_t reaches zero before reaching full capacity and also the distribution of the time over which Z_t is non-zero.

(iii) given the distribution functions of x_t and y_t, to determine the optimal capacity k according to some cost criterion or alternatively given the distribution of x_t and the system capacity k determine an optimal release rule y_t according to some cost critereon.

Early work on these and related problems was restricted to models in which simplifying assumptions such as independent input variables were made. As the theory of storage systems has developed, models which more closely represented real systems have been studied. In the following sections the application of this theory to hydro storage systems in relation to items (i) to (iii) is considered.

2.3 STORAGE MODELS WITH INDEPENDENT INPUT

The storage model in which the input random variables are identically and independantly distributed (i.i.d) provide a first approximation to some hydro storage systems if the length of the time interval is about a year (Bhat and Gani (1965)). This enables single server queueing models to be used to represent, approximately, hydro storage systems.

It is noted that equation (2.2) involves the random variables x_t, y_t only as a difference $u_t = x_t - y_t$ say, so that this model could be represented by the notation $[Z_t, u_t, k]$. In item (i) above, the distributions of x_t and y_t are given (i.e. u_t is known in distribution) and the problem is to determine the distribution of Z_t. Consequently it is of interest to find out if a limiting distribution of Z_t exists as $t \to \infty$. Two cases arise, as is frequently the case in storage theory, depending on whether k is finite or infinite. Lindley (1952) showed that if $k = \infty$, the limiting distribution $F(y) = \lim_{t \to \infty} F_t(y)$ of Z_t exists if $E[u_t] < 0$. If k is finite F(y) exists if $E[u_t] < \infty$ (Kendall (1951), Finch (1958, 1960)).

The distribution function of Z_t for various u_t distributions have been reported in the queueing literature. Also the development of general methods for deriving distributions in models of the GI|G|1 type have been reported. Kingman (1962, 1966), Finch (1961) discuss the use of Spitzer's identity in studying this model. Another general method is the use of Wiener-Hopf decomposition.

Writing G(x) for $Pr\left\{u_t \le x\right\}$ for all t and $F_t(y)$ for $Pr\left\{Z_t \le y\right\}$ then when Z_t and u_t are independent and the u_t variables are mutually independent.

$$F_{t+1}(y) = -\int_0^{\infty} F_t(x)\, dG(y-x), \quad 0 \le y < \infty$$

and taking limits as as $t \to \infty$

$$F(y) = -\int_0^{\infty} F(x)\, dG(y-x), \quad 0 \le y < \infty \tag{2.7}$$

Lindley (1952) showed that the condition for a unique solution to (2.7) is that

$$E(u) = \int_{-\infty}^{\infty} x\, dG(x) < 0 .$$

As some lake inflows can be approximately represented by a gamma distribution it is useful to apply the Wiener-Hopf method to the model $M|E_p|1$. E_p symbolises the Erlang distribution of the service time x_t. If this is denoted by $G_1(x)$, then

$$dG_1(x) \equiv (\mu^p/(p-1)!)\, e^{-\mu x} x^{p-1} dx, \quad (\mu > 0, 0 \le x < \infty)$$

It can be shown (Ghosal (1970)) that the distribution of $u_t = (x_t - y_t)$ is given by

$$G(x) = Pr\left\{u_t \le x\right\}$$

$$= \int_z Pr\left\{X_t < x + Y_t\right\} Pr\left\{z \le Y_t \le z + dz\right\}$$

which becomes $$G(x) = \begin{cases} 1 - e^{-\mu x}\left\{1-\left(\mu/\lambda+\mu\right)^{p}\right\} \sum_{r=0}^{p-1} (\lambda+\mu)^{r+1} x^{r}/r! & (x>0) \\ e^{\lambda x}\left\{1-\left(\mu/\lambda+\mu\right)^{p}\right\} & (x \leq 0) \end{cases} \qquad (2.8)$$

From (2.8) $\psi(\theta)$ can be obtained where

$$\psi(\theta) = \int_0^{\infty} e^{i\theta x}\, dG(x) + \int_{-\infty}^{0} e^{i\theta x}\, dG(x) \qquad (2.9)$$

(2.9) yields

$$\phi(\theta) = 1 - \frac{p\lambda}{\mu}\left[1+\lambda\left\{1-\mu^{p}/(\mu-i\theta)^{p}\right\}/i\theta\right]^{-1} \qquad (2.10)$$

This yields*

$$F(y) = 1 - \sum_{r=1}^{p} A_r\, e^{-\theta_r y} \qquad (2.11)$$

Although the storage models with independent input variables have wide application, they provide only a first order approximation to real reservoir systems which is general have input flows which shown some degree of correlation as discussed in the next section.

2.4 CORRELATED INFLOWS

Recalling Moran's model of the discrete infinite dam is an appropriate starting point for introducing independence in the input variables. Let Z_t ($0 \leq Z_t \leq k-m$) denote the integer valued content of the dam at yearly epochs t = 0,1,2 ... and consider the case where initially the dam is not empty, i.e. $Z_0 > 0$. The input to the dam during the year (t, t+1) is an integer variable x_t = 0,1,2 ... and a release is made from the dam just before the end of the interval (t, t+1). The size of the release is the lesser of m where 0 < m k-m and the whole content of the dam. The dam content Z_{t+1}, just after the release, is given by

$$Z_{t+1} = \min\left\{Z_t + X_t, k\right\} - \min\left\{Z_t + X_t, m\right\}$$

* See PRABHU, N.U., 'Queues and Inventories' Wiley (1965) for detailed form of A_r, θ_r; $r = 1, \ldots, p$

An overlow of size $Z_t + x_t - k$ occurs during the year (t, t+1) if $Z_t + x_t > k$ during that time. As noted previously, the annual inputs x_t are assumed to form a sequence of independent and identically distributed random variables with probability distribution.

$p_i = \Pr\{x_t = i\}$ i = ,1,2 ... for all t

The sequence $\{Z_t\}$ thus forms a homogeneous Markov chain with most early work concentrating on determining stationary distributions.

$P_j = \lim_{t \to \infty} P_{uj}(t)$

and later the time dependent distribtuions

$P_{uj}(t) = \Pr\{Z_t = j \mid Z_0 = u\}$

Lloyd (1963, 1967) extended Morans model by considering an input sequence which formed a homogeneous Markov chain. Although this meant that $\{Z_t\}$ no longer formed a Markov chain, the joint process $\{Z_t, x_t\} = \{u_t\}$ and also $\{v_t\} = \{Z_t, x_{t-1}\}$ are bivariate Markov chains. Lloyd considered the stationary properties of $\{u_t\}$ whilst Gani (1968), Ali Khan (1967) developed the time-dependent properties of the $\{v_t\}$ process and derived the probability of the dam becoming empty for the first time, for k infinite and finite. Also the transient behaviour of the dam content Z_t was determined in these two cases with unit releases (m=1). These results were further extended by Phatarfod and Mardia (1973) who considered the influence, on the behaviour of dam processes, of the nature of the serial correlation in the Markov input chain. Pakes (1973) added to this discussion of particular Markov chains by considering limiting theorems for quite general Markov chain inputs. Phatarfod (1975) made a comparison between probabilitic and simulation approaches to the design and analysis of reservoirs. After drawing attention to the paper by Gould (1961) in which a numerical method based on Moran's model is used to study the behaviour of a reservoir taking into account seasonality, serial correlation of inflows and variable drafts, Phatarfod expresses the view that the analytic and simulation methods should be complementary. The former method giving a first approximation to a solution in a particular situation thus enabling an initial estimate of storage size and inflow parameters to be made.

Simulation can then be carried out more efficiently to investigate the detailed behaviour of the reservoir under varying factors such as those included in Gould's method. The paper considers two main questions: (1) the nature of the inflow distribution; (2) the effect of the parameters of the inflow distribution on storage size. In the discussion on storage size a useful approximation to the equilibrium distribution of reservoir content is obtained by making analogy with the random walk model.

2.5 RANDOM WALK ANALOGY

The model of a system denoted by $[Z_t, u_t, k]$ where u_t is a stochastic variable, can be regarded as a random walk of the process $\{Z_t\}$. Equation (2.2) $Z_{t+1} = \min[Z_t + u_t, k]^+$ implies that $Z_{t+1} \geq 0$ and that $Z_{t+1} \leq k$. The process is such that these barriers are reflecting rather than absorbing. Ghosal (1970) uses this analogy to consider item (ii) listed in Section (2.1). He shows that if v(y) = probability that, given the initial position of the particle is y ($o < y < k$), the walk $\{Z\}$ reaches the barrier 0 before reaching the barrier k, then $v(y) = F(k-y), 0 \leq y < k$ where F(y) is the limiting distribution of Z_t.
The second part of item (ii) viz the distribution of the time over which Z_t is non-zero is treated by analogy with the busy time distribution of the server in a single server queue. The third item listed in Section (2.1) which involved determining optimal storage size or release rules under given cost constraints are also considered by the same author. (These topics are considered later in Chapter 5).

Phatarfod (1975) uses Wald's (1947) identity and a duality relation as follows to obtain an approximation of the equilibrium distribution of the reservoir content. Considering first the case of independent and identically distributed random variables $\{x_i\}$ set $S_N = \sum_{i=1}^{N} x_i$ and let there be two absorbing barriers at $a > o$ and $b < o$ with the random walk starting at the origin. Let n be the smallest positive integer such that $S_n \geq a$ or $S_n \leq b$ then if the Laplace Transform of y is $L(\theta)$

$$E\left[e^{-\theta S_n} L(\theta)^{-n}\right] = 1 \quad \forall \theta : |L(\theta)| \geq 1 \qquad (2.11)$$

Wald (1947) shows that for a random variable which can take both positive and negative values, there is one and only one real $\theta_0 \neq 0$ such that $L(\theta_0) = 1$.
Substituting this value into (2.11) gives the probability of absorption at b. In the case of a reservoir of size k with constant draft m, the probability that the reservoir becomes empty before filling up having started at level u is given approximately by

$$P_u \approx \frac{1 - e^{-(K-u)\theta_0}}{e^{u\theta_0} - e^{-(K-u)\theta_0}}$$

where θ_0 is the non-zero solution of $L(\theta)e^{m\theta} = 1$, $L(\theta)$ being the LT of the stream flow distribution. Considering next two random walks, one between two absorbing barriers and a second between two reflecting barriers at 0 and k, the equilibrium distribution function of position of the second random walk F(x) is related to P_u by $P_u = F(k - u)$ so finally

$$F(x) = Pr\{\text{content of the reservoir} \leq x\}$$

$$= \frac{e^{x\theta_0} - 1}{e^{K\theta_0} - 1}$$

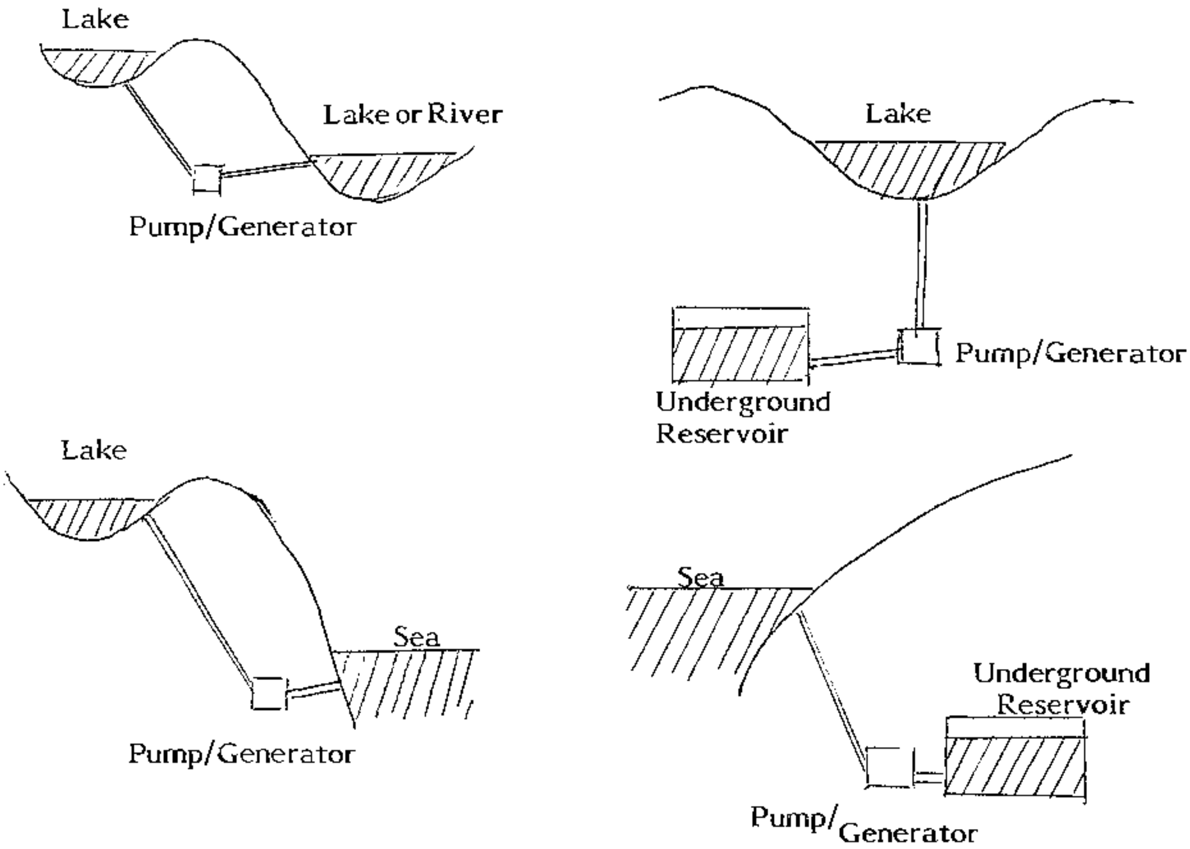
Phatarfod extends this approximation to Markovian input chains and evaluates its accuracy and application in serveral specific cases.

The reservoir and general storage models already considered have generally involved a stochastic process such as a river flow which at least to some extent was not controlled by the operator of the system. In some circumstances the ability to control, within certain limits, both the inflow to and outflow from a part of a system of reservoirs could help achieve a better overall performance by the system. This added flexibility is one feature of pumped storage systems which are considered next.

2.6 PUMPED STORAGE SYSTEMS

(i) Introduction

Pumped storage systems, in general terms, involve the use of externally generated power to increase the potential energy of that system. This is in order to utilise the stored energy at a time which proves, for whatever reason, to be to the greater advantage to the general system of which the storage system is part. Although a battery could be considered an example of pumped storage under this definition, the term is usually reserved for systems in which the working substance is a fluid such as water or compressed air. In principle however the battery is a storage system of potential energy. Its use in this capacity has been limited to smaller power systems because of the limited energy density/unit cost, so that even battery emergency supply systems of limited capacity are bulky and expensive. Regenerative braking systems in electric vehicles such as tramcars is another example of the widely used principle of potential energy storage. The largest scale application of this principle is found in electricity generating systems. In this case (as illustrated below) the gain in potential energy is obtained by raising the level of a quantity of water or increasing the pressure of a gas in a container. As considered below, the size of the storage system is determined partly by the physical features of the landscape and partly by the present and potential requirements of the power system it supplements.



Many Pumped Storage Configurations are Possible

Figure 1

(ii) Historical note.

Although generally associated with electrical supply systems, pumped storage systems were installed in the 18th century in Britain to supplement the operation of water wheels by evening out the supply of water. Early systems used horses to haul water back up hill but by 1774 a system installed by James Watt in Birmingham was using a steam driven pump. By the late 19th century small scale pumped storage systems were being used in conjunction with small hydro-electric stations run by private manufacturers in Europe. In this Century the scale of pumped storage schemes has been steadily increasing. In addition, the role of pumped storage schemes has widened as the complexity of power generating systems has increased.

(iii) Economic considerations.

The case for installing pumped storage in a largely thermal generating system is recognised since the pumped storage can handle short term fluctuations in demand more easily than can large thermal plant and can also help improve thermal plant factors by pumping when electricity demand is low. In a generating system which has a considerable proportion of hydro plant already installed the case for pumped storage or additional pumped storage needs to be established. Because it derives its energy from other sources, pumped storage must be evaluated by its effect on the long term fixed and running costs of the generating system as a whole. In such a system some of the advantages associated with hydro-turbine generation such as the ability to 'fine tune' to match demand or to upload or down load quickly and so act as spinning reserve will already be available in the system What the system might require is more thermal plant for dry year firming rather than more or the introduction of pumped storage for increased water utilization. Barr (1981) An initial economic assessment could be made on the basis of capital cost per kilowatt of pumped storage.

(iv) Pumped Storage Operation.

The two graphs that follow indicate a typical daily load curve and the corresponding storage reservoir level.

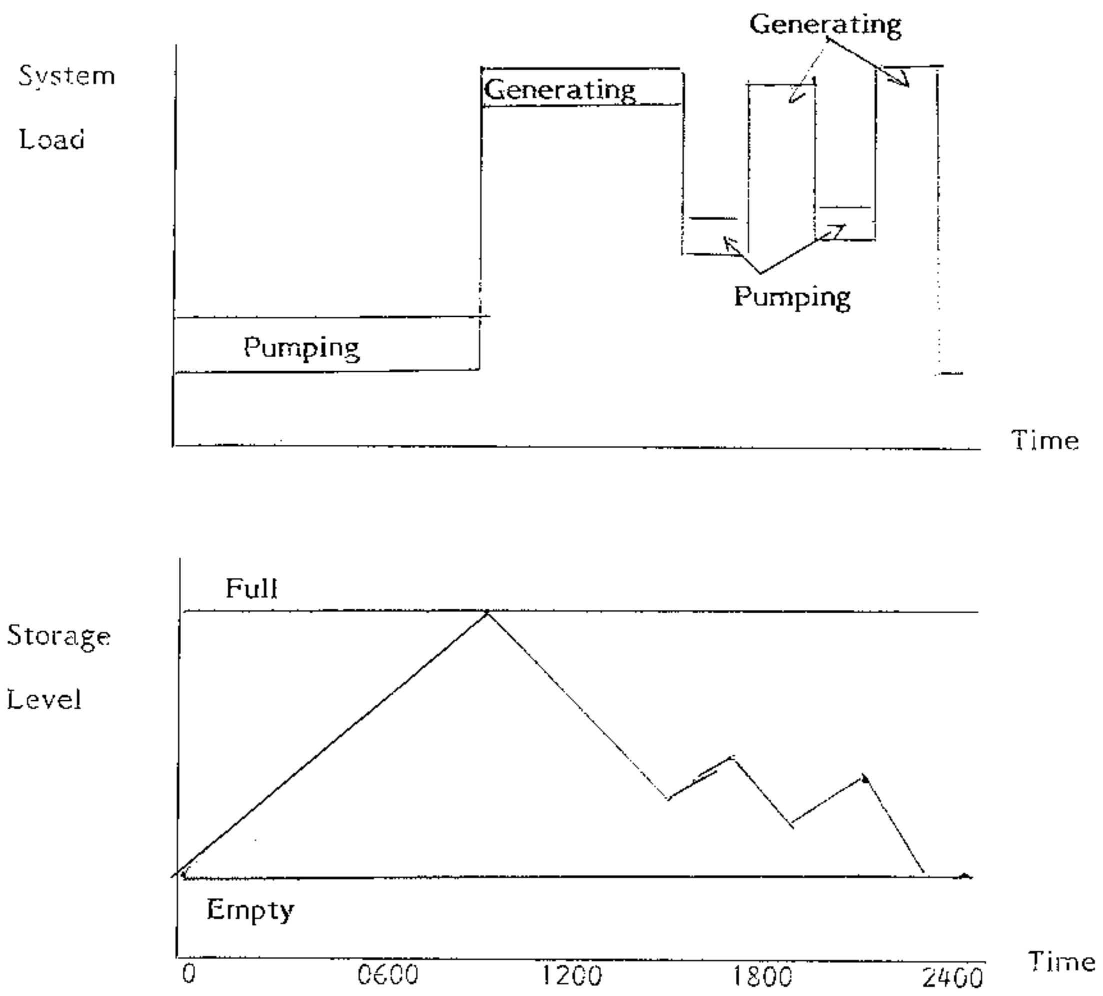


Figure 2

A third graph shown below indicates the effect of pumped storage on a daily (or similarly weekly if operated on a weekly cycle) load duration curve. Area A is larger than B because the effeciency of the storage cycle is less than unity. If the pumped storage was used to displace high cost thermal generation as shown in the graph above a load level Pg then the pumping level Pp is selected so that apart from operational considerations

$$\frac{\text{the incremental cost at Pp}}{\text{cycle efficiency}} \lessgtr \text{decremental cost at Pg}$$

The dashed curves show that the alteration to the load duration curve caused by the operation of pumped storage.

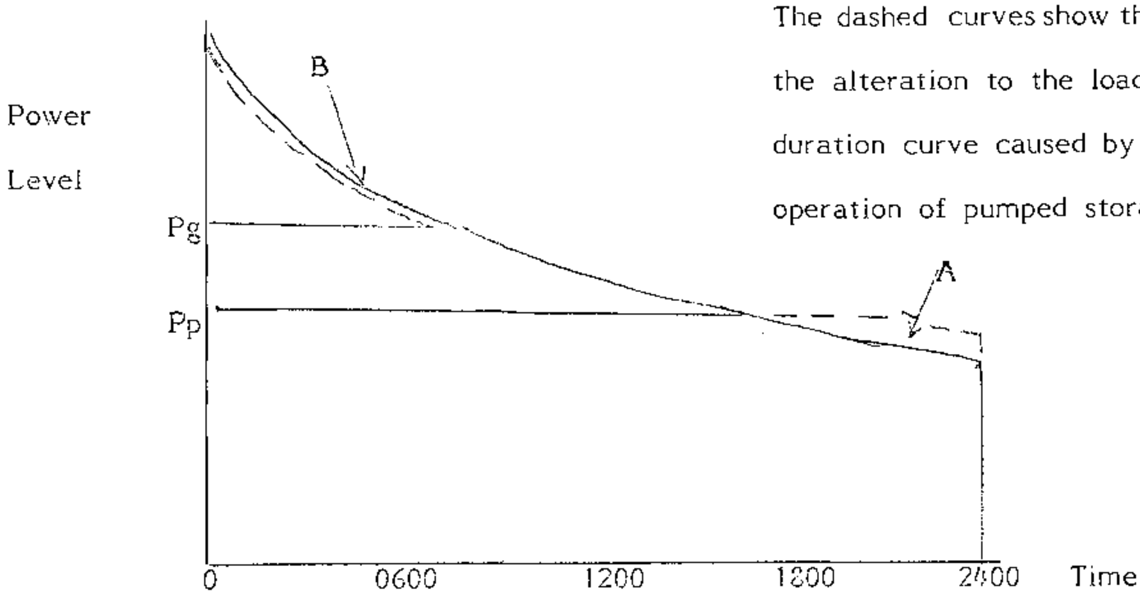


Figure 3

If a pumped storage scheme is operated on a daily cycle with a typical efficiency of about 75% maximum power generation is limited to about 10 hours if pumping and generating powers are equal and the system is closed. If however the cycle is several days e.g. a week long or if the upper reservoir has natural inflows, the system can be operated in a more flexible manner. In the present context a pumped storage scheme could be considered as an additional means whereby the operator of a reservoir with natural inflows and/or outflows could manipulate the reservoir level by pumping water to or from the reservoir. This manouver would be restricted in the longer term by the considerations outlined above but would provide more flexibility than obtainable by regulating outflows only.

Cochran (1979) discusses the development of pumped storage in power systems in greater detail. This aspect of pumping water to and from reservoirs is similar to the operation of a feedback mechanism and in the next chapter feedback systems and queues with feedback are considered.

CHAPTER 3 - SYSTEMS WITH FEEDBACK CONTROL

3.1 REPRESENTATION OF SYSTEMS

Often when real systems are studied, a model is developed to represent in some respects and to some degree the behaviour of the system. A model is thus an abstraction of reality and an attempt is made to include sufficient detail in the model, to ensure that the model behaviour resembles that of the real system to a degree appropriate for the purpose of the study. The question of the validity of the model must be kept in mind when using the model to draw inferences about the real system. Mathematical modelling can be considered in two categories: Analytic and Synthetic.

Analytic modelling consists of assembling and interconnecting the mathematical descriptions of the behaviour of the physical components of the model, such as the voltage and current relationships involved in a network of electrical components. Alternatively, synthetic modelling is the selection of mathematical relationships which appear to fit observed input-output data. Synthetic modelling will be considered after a review of analytic modeling. In either method a boundary must be decided upon to separate the system from its environment. An output variable from the system to the environment should not influence the environment to an extent that input variables from the environment to the system are altered. If this occurs the boundary should be enlarged so as to keep such 'feedback' effects within the system. Once the system is defined the various activities within itself can be modelled by various idealised components which exhibit similar behaviour. The next step in analytic modelling is to develop a mathematical model of the whole system using mathematical characterisations of the idealised components. Often in control theory lumped-parameter models are used. In these models there are two basic types of variable: through (or rate) variables and across (or level) variables. Through variables flow through two-terminal elements and have the same value at each terminal. Across variables have different values at each terminal. Common examples being electric current or flow rate and voltage or pressure respectively. Associated with each two-terminal element is one through and one across variable and at lease one elemental equation relating these variables. After the system has been modelled by an interconnection of two-terminal elements, two further types of rule relate the through and across variables. These are firstly a conservation of through variables where two-terminal elements connect and secondly, a rule for compalibility between across variables. Kirchoff's current node and voltage loop laws are examples. These equations may then be manipulated into a form convenient for the particular purpose of the study and the behaviour of the resulting model compared with that of the real system. Model modification may follow until the comparison is considered satisfactory.

The nature of the variables and equations used in the description of the system provides a way of classifying systems. Stochastic systems are those which contain parameters or variables which must be described in a probabilistic manner. As already indicated such systems can sometimes be conveniently approximated by a deterministic model. A further distinction is made between systems defined for continuous time and those defined at discrete intervals only, such as in sampling or digital systems. If one of the elemental equations is non-linear then the system as a whole is described as non-linear. In addition, if such an equation has a time dependent parameter or form, the system is said to be time-varying. Finally a system with external inputs acting is called non-homogeneous whereas in a homogeneous system the system behaviour is determined by the internal initial conditions. The type of system being considered has an influence on the form of the mathematical model of the system which is generally of one of two forms.

(a) input-output equations (b) equations which describe the internal behaviour of the system as well as the input-output equations. For example, the input-output equations of linear systems with constant coefficients can be conveniently represented in the form of Laplace transforms. Another form of representation involves a minimum set of the variables used in the elemental equations (or functions of these). The use of this minimum set of variables is called the state-space approach and this method of representing the system is widely used in a modern control theory, as it is particularly suitable for use in computer programmes. The development of the state-space approach has provided a uniform method of analysis of systems of arbitrary order, linear or non-linear with time-varying or constant coefficients. It has also led to new developments in the theory of optimization of system behaviour. The use of computer algorithms and numerical integration methods has also enabled the study of much larger systems than was possible with the older graphical and Laplace Transform-Inverse Transform techniques.

Association with computer methods and the state variable representation distinguishes modern control theory from the classical approach, however the flexibility of representing a system in terms of generalised coordinates is well established in Hamiltonian mechanics. In the next section aspects of classical control theory are considered.

3.2 CLASSICAL CONTROL THEORY

In classical control theory much attention was given to the analysis of linear, constant coefficient models. Many systems can be approximated over part of their operating range and their parameters assumed constant over a finite interval so that these models found wide application. Attempts to develop methods for non-linear or time-varying systems such as phase-plane methods or describing functions were generally limited to low-order systems and also to systems with only a few inputs and outputs (usually one of each only). Laplace transforms and frequency domain methods are well suited to the analysis of linear, constant coefficient systems. Most of the systems considered in classical control theory contained a feedback path as indicated in Figure 3.1

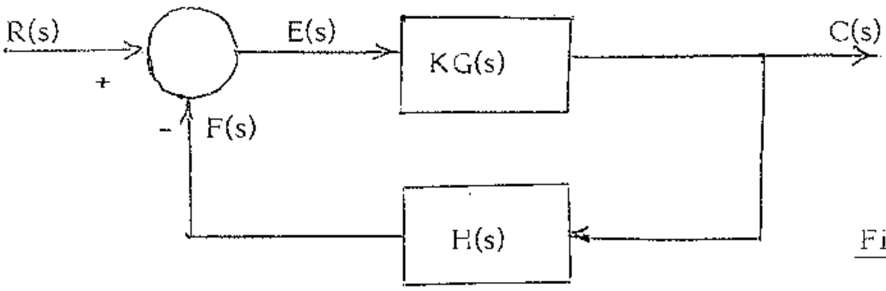


Figure 3.1

At the summing junction in Figure 3.1 R - HC = E. Also the relation between C and E is KGE = C. Eliminating E gives

$$C = KGR/(1 + KGH) \qquad (3.1)$$

$$\text{so that } E = R/(1 + KGH)$$

KG(s) is called the forward transfer function, H(s) the feedback transfer function and KG(s)H(s) the open-loop transfer function. Such feedback paths imbue the system with the following benefits:

1. Deviations in the output C from that specified by the input R can be countered. Such deviations might be due to external disturbances or to changes in the system parameters.

2. Achieving desired transient and steady-state response is made easier.

To counter these positive aspects of negative feedback are the following disadvantages:

a. Care is needed that the system with feedback does not become unstable;

b. System gain is reduced as well as extra precision parts being required in the feedback path.

To enable the merits of various models to be compared classical control theory sought to develop measures of performance which could easily be applied to feedback systems. Also methods of analysing such systems in terms of these measures, were sought as well as methods of improving the systems performance if necessary. Equation 3.1 in the form

$$C(s) = \frac{KG(s)R(s)}{1+KG(S)H(s)}$$

indicates the difficulty in comparing different feedback models since taking the inverse Laplace transform of the right-hand side to obtain C(t) may be difficult except for particular forms of the denominator and also R(s) may take an infinite number of forms. Instead of seeking a complete analytical solution to (3.1) classical theory uses only certain properties which C should have as the basis for comparison, without actually solving for C. The methods used, developed before the availability of computers, stress graphical techniques. Also, the choice for R(s) is limited to a few standard forms such as ramps, step functions and sinusoids.

Four measures of performance were determined to be indicators of good design in a control system. These are (1) stability, (2) steady-state accuracy, (3) satisfactory transient response, and/or (4) satisfactory frequency response. These are considered in order.

(1) Stability refers to the ability of the system output not to grow without bound due to a bounded input, initial condition or an unwanted disturbance. For linear constant coefficient systems the stability depends only on the location of the roots the denominator of (3.1) i.e. of $1 + KG(s)H(s) = 0$. If any roots have positive real parts, the system is unstable. Routh's criterion, Root locus and Bode plots are amongst methods used to determine stability.

(2) Steady state accuracy requires that the error signal E(t) tends to a value regarded as being sufficiently small as $t \to \infty$. The final value theorem enables this requirement to be evaluated without the use of inverse transforms since $\lim_{t \to \infty} \{E(t)\} = \lim_{s \to 0} \{sE(s)\}$

(3) Satisfactory transient response refers to the behaviour of the output when sudden changes occur in the input. There should be no excessive overshoot, the oscillations following the overshoot should be of acceptable amplitude and frequency as well as duration. On the other hand the speed of response to a change in the input should be satisfactory. These responses depend on the location of the closed-loop poles, i.e. the roots of $1+KG(s)H(s) = o$ and their proximity to the stability boundary which is most easily determined by the root locus method.

(4) Satisfactory frequency response refers to the behaviour of the system in response to different input frequencies. The bandwidth and frequency at maximum gain are measures of this response. The Bode and Nyquist plots give information regarding the frequency response of open-loop systems. This information can then be transferred to a Nichol's chart to determine closed-loop frequency response.

Once the system responses have been determined, the system behaviour can be altered by changing the systems parameters or structure. These changes include:

(a) altering the gain parameter K

(2) altering the feedback path characteristics

(3) altering the structure of the system

(4) adding compensation networks

Changes (2) and (4) are referred to as compensation techniques. These can be applied in an analytic or synthetic way. In the analytic approach, experience suggests an appropriate alteration to try. The performance of the new system is evaluated using the root-locus or other graphical methods. This process is continued until the system converges to a satisfactory performance level. To synthesise the desired performance, a closed loop transfer function corresponding to that performance is determined. The problem then becomes one of designing a feedback system with this particular transfer function. Similar techniques have been developed for sampled and digital systems. These techniques can be extended to multiple loops and/or multiple inputs and outputs. The methods soon become cumbersome as the system complexity increases.

Mayr (1970) provides an interesting account of the development of the concept of the feedback loop from antiquity upto the invention in 1788 of the centrifugal governor by James Watt. Watt's invention helped to familarise the concept and practice of feedback design and subsequently development of both aspects accelerated.

3.3 STATE-SPACE REPRESENTATION

State variables comprise a minimum set of system variables $\{x_i\}$ which summarize the status of the system in the following sense: if at any time t_o, the values of the state variables $x_i(t_o)$ are known, then the output $y(t_1)$ and the values of the state variables $x_i(t_1)$ can be uniquely determined for any time $t_1 > t_o$ provided $u_{[t_o, t_1]}$ i.e. the input function is known. Advantages of the state-space approach include the following:

(1) The set of first order differential or difference equations resulting from the application of this method would be easier to solve than the original equations relating system variables if these equations involved higher order derivatives or differences

(2) The first order equations resulting from the state-space representation enables the use of simplifying matrix notation.

(3) The inclusion of the initial conditions of a system in the analysis of control systems, which is difficult with conventional techniques, is implicit in the state-space approach.

(4) The method has wide application in the study of control systems including most non-linear, time-varying, stochastic and sampled-data systems.

(5) The state-space approach is convenient when system synthesis is being used to study a control system.

An appropriate example appropriate in the present context is shown in Figure 3.2

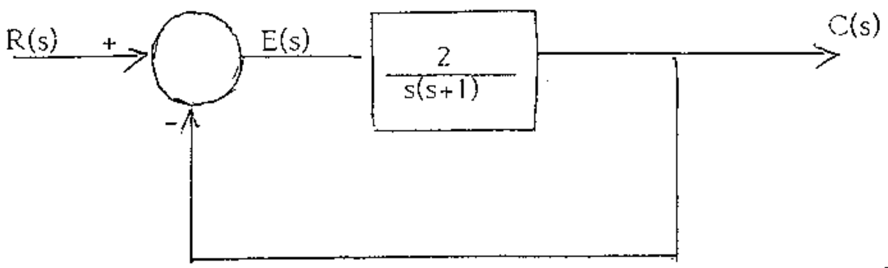


Figure 3.2

Applying (3.1) gives

$$\frac{C(s)}{R(s)} = \frac{2/s(s+1)}{1 + 2/s(s+1)} = \frac{2}{s^2 + s + 2}$$

In the time domain this becomes

$$\frac{d^2c}{dt^2} + \frac{dc}{dt} + 2c = 2r$$

By defining the state variables as $x_1 = c$, $x_2 = \dot{c}$ the system can be described by two first order differential equations

$$\dot{x}_1 = x_2 = \dot{c}$$

$$\dot{x}_2 = -2x_1 - x_2 + 2r$$

These equations can be written in the following vector notation

$$\dot{\underset{\sim}{x}} = P\underset{\sim}{x} + Br$$

$$\underset{\sim}{c} = L\underset{\sim}{x}$$

where $P = \begin{bmatrix} 0 & 1 \\ -2 & -1 \end{bmatrix}$, $B = \begin{bmatrix} 0 \\ 2 \end{bmatrix}$, $\underset{\sim}{X} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$, $L = [1, 0]$

It is noted that a state variable may not be observable or measurable. Observability is considered later in the section on optimal control. The set of state variables representing the system is not unique. Other sets system variables or functions of them can satisfy the definition given of state variables. Also various methods for obtaining the state variables from the given description of the system are available, which one is most appropriate depends on the nature of the system being considered and the form of its description. Brogan (1974) and Shinners (1978) provide a general discussion of these topics and include bibliographies.

3.4 LINEAR SYSTEMS

The example in Figure 3.2 is a lumped parameter, linear, continuous-time system. The most general form of the state-space equations for this type of system is

$$\dot{\underset{\sim}{x}} = A(t)\underset{\sim}{x} + B(t)\underset{\sim}{u}(t) \qquad (3.2)$$

$$\underset{\sim}{y}(t) = C(t)\underset{\sim}{x} + D(t)\underset{\sim}{u}(t) \qquad (3.3)$$

Two types of problem arise in the study of such systems. The first is to determine the nature of the input $\underset{\sim}{u}(t)$ in order to achieve the desired state $\underset{\sim}{x}(t)$ or output $\underset{\sim}{y}(t)$ behaviour. The second problem is to analyse the response of the system in terms of $\underset{\sim}{x}(t)$ and $\underset{\sim}{y}(t)$ when $\underset{\sim}{u}(t)$ is a known function. Since in (3.3) $\underset{\sim}{y}(t)$ is related algebraically to $\underset{\sim}{x}(t)$ and $\underset{\sim}{u}(t)$, the analysis concentrates on determining $\underset{\sim}{x}(t)$. The solution of the time-varying homogeneous case of (3.2), i.e. $\dot{\underset{\sim}{x}} = A(t)\underset{\sim}{x}$ (3.4) must be unique for every $\underset{\sim}{x}(t_0)$ if $\underset{\sim}{x}$ is a valid state variable vector. This imposes constraints on the nature of A(t).

A sufficient (but not necessary) condition for a unique solution to (3.4) is for A(t) to have continuous elements. If this unique solution exists it is termed the fundamental solution matrix U(t) where $\dot{U}(t) = A(t)U(t)$ and $U(t_o) = I_n$, A having dimension nxn. The solution of the non-homogeneous case (3.2) can be shown to be

$$\underset{\sim}{x}(t) = U(t)U^{-1}(t_0)\underset{\sim}{x}(t_0) + \int_{t_0}^{t} U(t)U^{-1}(\tau)B(\tau)\underset{\sim}{u}(\tau)d\tau$$

Discrete time systems arise if a continuous system is approximated for the purpose of computer simulation or if the input consists of sampled data or if the system naturally discrete in nature. In these cases the methods for solving the state variable equations closely resembles the methods used in the continuous case. Two important properties of control systems are controllability and observability. These are defined for linear systems as follows:

Controllability: a linear system is controllable at t_o if it is possible to find an input function u(t) which will transfer the initial state $\underset{\sim}{x}(t_o)$ to zero at some finite time t_1.

Observability: a linear system is said to be observable at t_o if $\underset{\sim}{x}(t_o)$ can be determined from a knowledge of the output function $\underset{\sim}{y}(t)$ for $t \varepsilon [t_o, t_1]$ where $t_o \leq t_1 < \infty$.

Several criteria have been determined to enable which linear systems have these properties to be found from the nature of the matrices A,B,C,D of (3.2) and (3.3) and these are of assistance when considering the design of linear feedback systems. As noted in the previous section on classical control theory, designing a feedback system involves either altering the forward transfer function or by altering the feedback transfer function or adding extra feedback paths, in order to obtain the required system response. Different system adjustments or alterations can be combined to induce behaviour not obtainable from any single change. Sometimes it is desirable for a particular input to a system to affect only a single output or group of outputs. Such a system is said to be decoupled. A system with m inputs and m outputs is decoupled if the transfer function H(s) is diagonal and non-singular. If the number of inputs and outputs are not equal partial decoupling may occur. Although as just outlined, a general theory of behaviour and design of linear systems has been developed, general analytic solutions of particular non-linear systems are often not possible. In the next section methods used in the study of non-linear systems are considered. Brogan (1974) and Shinners (1978) provide a general reference and non-linear systems More recent developments appear in IEEE Transactions on Automatic Control.

3.5 NON-LINEAR SYSTEMS

In general real systems are linear only over a limited range of their operation. This however is often quite acceptable, e.g. the operation of a small signal amplifier may be linear only in a region about is quiescent operating point - the non-linearity introduced by the finite voltage and current levels available from the power supply do not affect the normal operation of the amplifier. In other situations non-linearity is introduced to counter the effects of other undesirable non-linearities or because a non-linear characteristic best suits the operation of the system.

Figure 3.3 shows four types of input/output plots characteristic of many non-linear systems. These types of behaviour do not occur in linear systems.

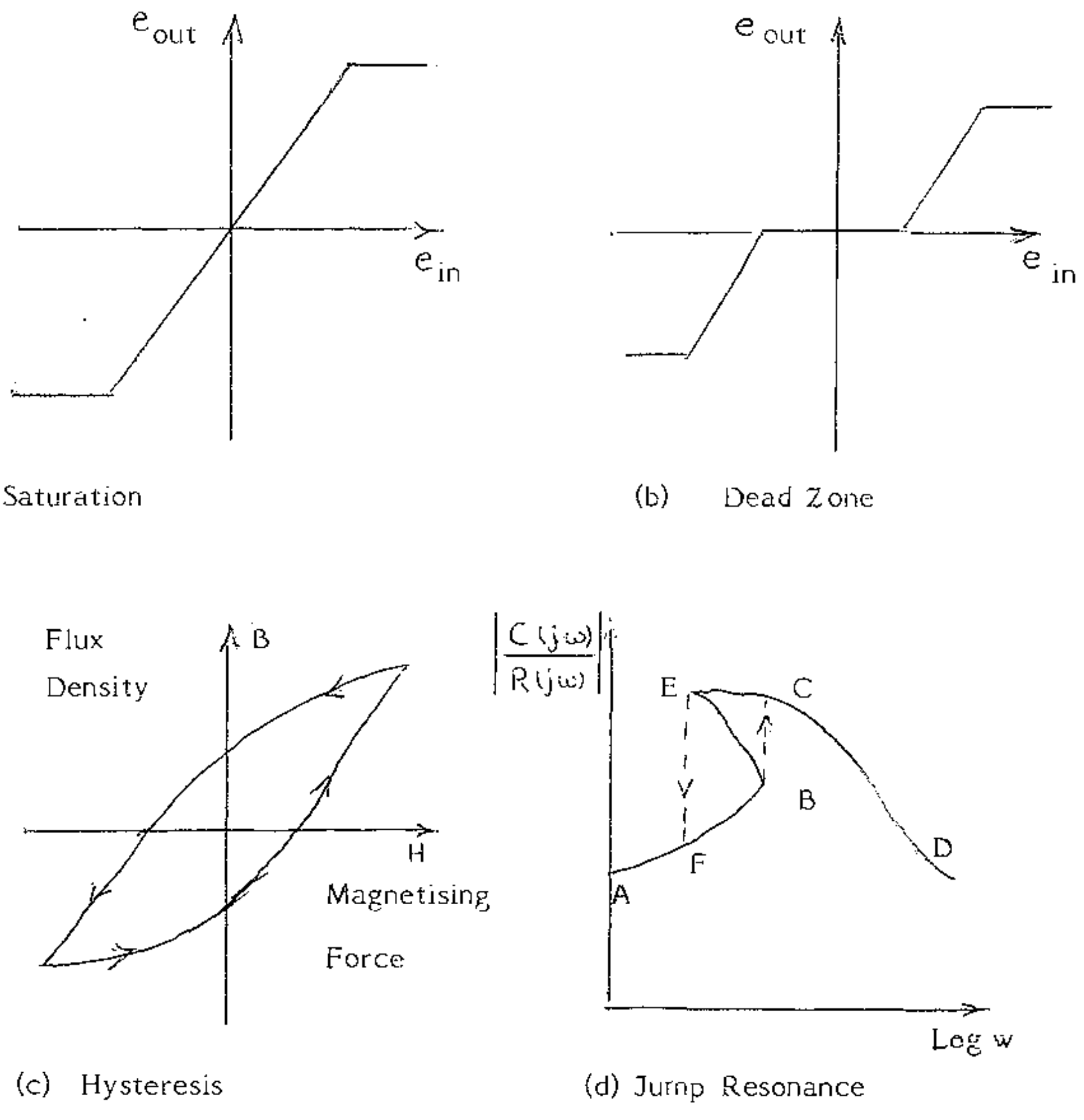


Figure 3.3

Figure 3.3 (d) shows jump resonance which can occur in second or higher order non-linear systems. As ω increases the operational path is from A to B followed by a jump to C then on to D. If ω is now decreased the path is from D to E followed by a jump to F then back to A. Subharmonic generation occurs in non-linear systems. In general if sinusoidal signals f_1, f_2 are added and their sum applied by a non-linear device, the output contains frequency components given by $af_1 + bf_2$ where a,b are integers. In addition to the above propeties peculiar to non-linear systems there are several properties of linear systems which non -linear systems do not have. Super position, i.e. if $r_1(t)$, $r_2(t)$ are inputs, $C_1(t)$, $C_2(t)$ the corresponding outputs then $a_1r_1(t) + a_2r_2(t)$ has the output $a_1c_1(t) + a_2c_2(t)$ does not apply in non-linear systems. Stability of linear systems dpends only on the systems parameters. The stability of non-linear systems, however, depends on the initial conditions and the nature of the input signal as well as the system parameters. In linear systems instability results in an output that grows without bound either exponentially or in an oscillatory manner with an exponential envelope. In non-linear systems instability may result in a constant amplitude output such as that from a saturating amplifier, the waveform depending possibly, in part, on the input waveform.

Several methods have been developed for analysing non-linear systems. Their suitability in a particular case depends on the order of the system being analysed and on the degree of non-linearity. The object of the analysis of non-linear systems is to detect the presence and effects of the behaviour diagrammed in Figure 3, i.e. limit cycles, soft and hard excitation, hysteresis, jump resonance and sub-harmonic generation. As in the linear case, the response to specific input functions must also be determined.

Linearising approximations based on Taylor's expansion are appropriate if the deviation from linearity is not too large. Another approach, the describing-function method concerned with discovering limit cycles,simplifies the problem by assuming a sinusoidal input and that the frequency of the only significant component in the output is that of the input signal. In the piecewise-linear approach, non-linearities are approximated by several linear stages. The phase-plane method is very useful when analysing the response of second-order non-linear systems. The question of the stability of non-linear systems was first considered in the late 19th century by Lyapunov using generalisations of energy concepts. In 1959, Popov introduced a method which was later modified to develop the generalised circle critereon for non-linear system stability.

Non-linear systems for which the above methods are difficult to apply or inappropriate can be investigated using computer numerical techniques, However, the resulting solution is generally difficult to extend or apply to similar systems with different parameters or inputs.

3.6 SYSTEM IDENTIFICATION

In the analytic modelling discussed on the previous sections it was assumed that the form and parameters of the system being modelled were known. Often however this information is not immediately available and so must be determined by measurement on the system. The study of the identification of a suitable model to represent a particular real system has developed considerably over the last two decades. Astrom and Eykhoff (1971) and Gustavsson (1977) provide extensive surveys of the literature. Box and Jenkins (1970) introduced an iterative procedure for system identification which is generally available as a computer library programme. Box and MacGregor (1973) draw attention to the need to take into account the effect of feedback on the behaviour of a closed loop system when carrying out system identification, since by applying open loop procedures incorrect models may result and lack of fit not be detected. McLeod (1982) discusses the practical application of the Box Jenkins method to univariate stochastic processes and transfer function analysis. Kendall (1973) provides a comparison of several types of modelling methods, drawing attention to the underlying assumptions inherent in each procedure. Chatfield (1980) in an introduction to time series analysis provides a comprehensive reference list. This includes a reference to Priestley (1971) 'Fitting relationships between time series' and (1978) 'Non-linear models in time series analysis'.

Akaike (1984) discusses the application of multivariate autoregressive models and the use of the TIMSAC (Time Series and Control) computer program. This approach was motivated by the difficulty of analysing the behaviour of an industrial process incorporating a feedback mechanism Akaike (1972). Similar estimation of parameters and distributions is needed in the analysis of queueing models. Gross and Harris (1974) consider this statistical aspect of queueing theory and give further references. These aspects of system identification are considered further in Chapter 5.

3.7 OPTIMAL CONTROL THEORY

The discussion on the design of feedback control systems in sections (3.4) and (3.5) indicated that the design procedure itself was of a feedback loop nature since if the analysis of a control system indicated that its performance was not as required, an adjustment or alteration was made to the system followed by an analysis of the resulting system. Optimal control theory seeks to avoid this convergence procedure and develop a system which is optimal by design according to some measure of performance. Wiener (1949) developed the concept of optimum design based on optimising a performance criterion. The first application of this concept to the design of control systems was by McDonald (1950). Two major developments on the theory of optimal design were dynamic programming Bellman (1959, 1962) and the maximum principle Pontryagin (1959, 1960).

The basic problem of optimal control theory consists of choosing the input $\underset{\sim}{u}$ to a control system so that the performance of the system is optimum with respect to some performance criterion.

The structure of the problem can be set out as follows:

(a) There is a controlled process with state variables $\underset{\sim}{x}$, input $\underset{\sim}{u}$ at time t which are related in the form $\underset{\sim}{\dot{x}} = f(\underset{\sim}{x}, \underset{\sim}{u}, t)$ (3.5)

(b) There are restrictions on the input $\underset{\sim}{u}$ and/or the state variables $\underset{\sim}{x}$

(c) A reference signed $\underset{\sim}{r}$ which corresponds to the desired output from the system

(d) A performance criterion having the following general form

$$S = \int_{t_0}^{T} G(\underset{\sim}{c}(t), \underset{\sim}{u}(t), \underset{\sim}{r}(t), t)\, dt \qquad (3.6)$$

The integrand is referred to as the loss function and represents a measure of instaneous change from ideal performance. The optimal control problem can now be stated as finding the input u which subject to (3.5) and any initial and final boundary conditions will minimise S. If $\underset{\sim}{u}$ is determined as a function of the initial state and other system parameters the control is said to be open-loop. If $\underset{\sim}{u}$ is determined as a function of the current state then the control is said to be closed-loop.

The importance of the concept of controllability now appears for if the system is completely controllable, there is at least one control u which will transfer any initial state to any desired final state, provided this is an admissable control in terms of the constraints specified in the problem. If the system is not controllable no optimum control exists.

Dynamic programming provides an efficient means for sequential decision making. It is based upon Bellman's principal of optimality that an optimal policy has the property that whatever the initial state and initial decision are, the remaining decisions must constitute an optimal policy with regard to the state resulting from the first decision. The outcome of this principle is that the selection of an optimal choice from amongst all possible choices can be carried out as a sequence of optimal choices between fewer possibilities. Shinners (1978) sets out the close relationship between (3.6), dynamic programming, the calculus variations and Pontryagin's maximum principle. Gumowski and Mira (1968) provide a more complete treatment of optimisation in control theory. Brogan (1974) includes a comparison of the practicality of these procedures. Ross (1983) discusses stochastic dynamic programming. The application of these methods is considered later in regard to hydro-electric systems.

3.8 QUEUES WITH FEEDBACK

An example of a system with a feedback mechanism is found in queues with retrials. This is represented in diagram in Figure (3.4).

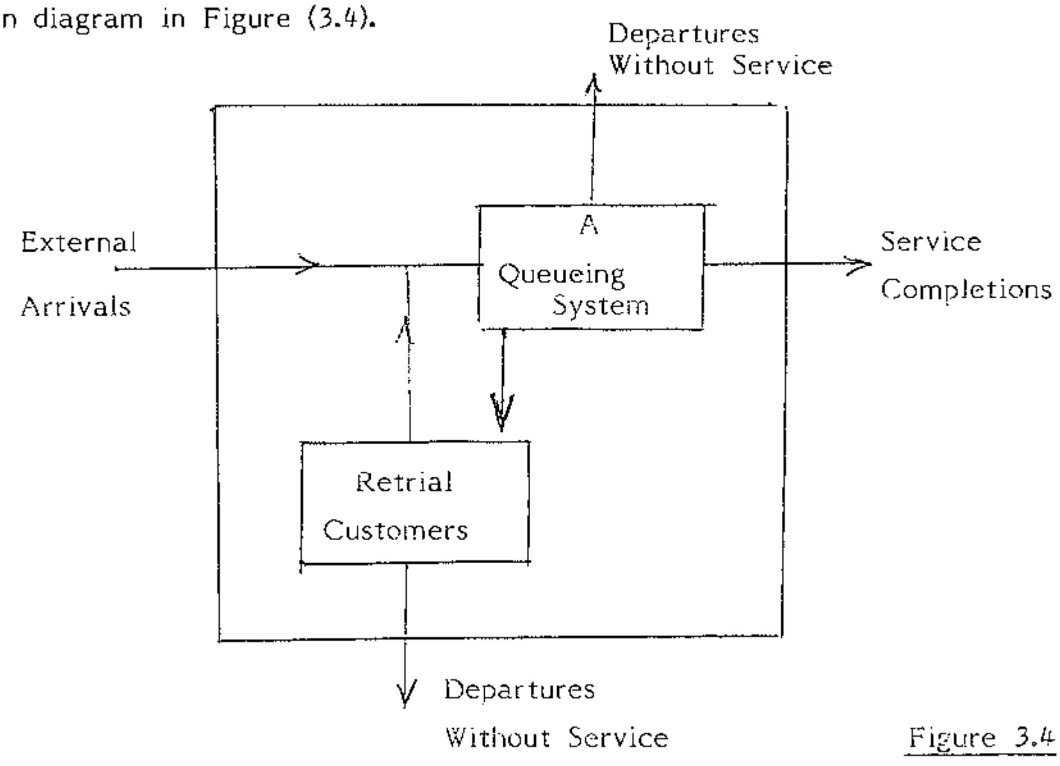


Figure 3.4

Queues with feedback or retrials occur frequently in practice. Computer time-sharing networks and assembly line inspection sites are examples. In these cases the customer is generally returned immediately to the waiting line.

This is termed instantaneous feedback. If however the customer is delayed before being allowed to re-enter the waiting line, the more general case of delayed feedback results. Foley and Disney (1983) provide a discussion of delayed M-server feedback queues and derive results which are valid for more general delay mechanisms. Kulkarni (1983) also considers delayed feedback and derives a general result for delayed queue models similar to Little's formula. Feedback queues with instantaneous feedback have been studied the following papers which consider in general M/M/1 or M/G/1 queues. Takacs (1962), Disney and D'Avignon 1975), Hannibalsson and Disney (1977), Disney, McNickle and Simon (1980), Disney (1981), Disney, Konig and Schmidt (1984), Ali and Neuts (1984), Hunter (1983, 1984). Queues with feedback will be considered in connection with networks of queues in the next chapter.

CHAPTER 4 - NETWORKS OF QUEUES

4.1 INTRODUCTION

In previous chapters the behaviour of individual queues was considered. In practice however, queues are often found to operate in networks with the output of one queue forming an input to another queue and so on. Examples of networks of this kind are found in communications systems, air traffic control systems, computer time-sharing systems, maintenance and repair facilities, hospitals and elsewhere. Erlang (1917) and Engset (1918) initiated the study of networks of queues in regard to telephone systems. Two papers by Jackson (1957, 1963) laid the foundation for much of the more recent work in queueing network theory. A convenient terminology for describing queue networks is as follows: each simple service system, i.e. a service centre, the associated service time process, the waiting capacity and the queue discipline, is termed a node. It is then assumed that these nodes are connected by arcs over which customers flow without delay. It is also assumed that each node has an arrival process consisting of arrivals from outside the network (exogenous arrivals) and/or arrivals from other nodes. In addition there is associated with this collection of nodes and arcs as set of rules called a switching process, which serves to direct customers through the nodes from the entrance, as an exogenous arrival, to an exit as a departure. Another familiar example of suc a queueing network is the job shop in which an item passes through several stages of machining, each stage forming a simple queue, until its departure on completion of all the machining operations. To describe the behaviour of a queueing network vector notation is convenient. Thus the queue length process of the network is defined as:

$$N = \left\{N_1(t), N_2(t), \ldots, N_J(t);\ t \geq 0\right\}$$ where $N_j(t)$, $j = 1, \ldots, J$ is the

queue length at node j at time t. Embedded queue length process such as the length of queues at all nodes at the arrival time of the m^{th} customer are also of interest. Let W_{mj} denote the waiting time at the j^{th} node of the m^{th} arrival at that node. Then if this customer traverses nodes $1, 2, \ldots, \ell$ the total waiting time

$$W_m \equiv W_{m1} + W_{m2} + \cdots + W_{m\ell}$$

The W_{mj} are not generally independent for a fixed m. $\{W_m, m = 1,2 \ldots\}$ is the waiting time process. Similarly, the sojourn time process $\{S_m, m = 1,2, \ldots\}$ is the sequence of times S_m where $S_m = S_{m1} + S_{m2} + \ldots + S_{m\ell}$ is the sojourn time of the m^{th} customer passing through nodes $1,2, \ldots, \ell$ Again the S_{mj} are generally not independent for a fixed m. In simple queues there are three main processes representing customer flows; arrival, output and departure processes. In queue networks each node has these processes so that representation of customer flows is more complex for a queue network. Disney and Konig (1985) give an extensive review of these random processes for a variety of network models and include a comprehensive reference list to recent work.

4.2 AN OPEN NETWORK MODEL WITH EXPONENTIAL SERVICE TIMES

In this network there are a finite number of nodes and any exogenous demand processes are assumed to be Poisson and all arriving customers are of the same type. The arrival processes are independent of one another and of the service processes. As all customers eventually leave the network, the network is said to be open. Conversley, networks with no arrivals and no departures are said to be closed.

Let $u = \{1,2, \ldots, J\}$ be the set of nodes, $J < \infty$. Each node is assumed to have infinite waiting capacity (i.e. $L_j = \infty, j=1, \ldots, J$) and to have service times which are independently and exponentially distributed with parameter $\mu_j(n_j) > 0$ for $n_j > 0$ Also $\mu_j(0) = 0$. Thus the service rate depends on the queue size n_j. The exogenous arrival process at node j has parameter λ_j and is independent of the arrival process at any other node. Upon passing through node j the customer is directed either to another node or to leave the network by a switching (or routing) process. This process is modelled by introducing a random variable Y_m and $Y_m = j$ if node j is the m^{th} node visited by a particular customer. It is assumed that the switching process $\{Y_m; m = 1,2 \ldots\}$ is a homogeneous finite Markov chain which guides every customer through the network. Denoting the outside of the network as node Δ the following define the switching probability for the chain:

$$Pr[Y_m = k \mid Y_{m-1} = j] = p(j,k) \qquad K = 1,2, \ldots, J;\ j = 1,2,\ldots,J$$

$$p(k) = Pr[Y_m = \Delta \mid Y_{m-1} = k] = 1 - \sum_{j=1}^{J} p(k,j),\ k = 1,2,\ldots,J$$

$$Pr[Y_m = \Delta \mid Y_{m-1} = \Delta] = 1$$

To simplify the model assume initially that $p(j,j) = 0$, $j=1,2,\ldots,J$

It can be shown that the traffic equations

$$\alpha_j = \lambda_j + \sum_{k=1}^{J} \alpha_k p(k,j), \quad j = 1,2,\ldots,J$$

have a unique solution $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_J)$, $\alpha_j > 0$ where α_j is the arrival rate at node j (λ_j being the arrival rate of exogenous arrival processes). Disney and Konig (1985) show that for this model the queue length process N has the stationary distribution Π given by $\Pi(n) = \prod_{j=1}^{J} \Pi_j(n_j)$ (4.1)

$$\text{where} \quad \Pi_j(n_j) = b_j \alpha_j^{n_j} \Big/ \prod_{r=1}^{n_j} \mu_j(r)$$

$$\text{and} \quad b_j = \sum_{i=0}^{\infty} \alpha_j^i \Big/ \prod_{r=1}^{i} \mu_j(r)$$

(4.1) indicates that the distribution $\Pi(n)$ is just the product of the nodal probabilities $\Pi_j(n_j)$ so that the elements within the vector process are mutually independent for each time t. This does not mean however, that the overall queueing processes N at time t_1 and N at time t_2 are independent.

A feature of network behaviour is that the combination of several input and output processes at a node can result in a net queueing process at that node quite distinct in character from that of its component processes. As an example assume in the above model that $\mu_j(n_j) = \mu_j \ \forall n_j > 0$ and $\mu_j(0) = 0$, $j = 1,2,\ldots,J$. Then $\Pi_j(n_j) = (1-\rho_j)\rho_j^{n_j}$, $n_j = 0,1,2,\ldots$ where $\rho_j = \alpha_j/\mu_j < 1$.

This equation has the same form as that for the probabilities of queue lengths for an $M/M/1$ queue viz $\Pi(n) = (1-\rho)\rho^n$, $n = 0,1,2,\ldots$ and $\rho = \frac{\lambda}{\mu} < 1$.

This same parallel can be shown to hold for more general input processes. As in the theory of single queues the above model can be elaborated in many ways both by altering the nature of the nodes and also the switching process. In addition different classes of arrivals can enter the network at different nodes resulting in a multiclass queueing network. As the complexity of the network model increases so often does the derivation of numerical evaluation of analytic results. Consequently there has been considerable interest in developing bounds and approximations to the behaviour of networks of queues. Another approach to dealing with the complexity often found in particular applications has been computer simulation.

The comments by Phatarfod (1973) noted previously regarding the inter-play between simulation and analysis are relevant here also since the method of analysing the output of a simulation is influenced by the nature of the processes being simulated. In addition, the example given above indicates the usefulness of an awareness of the flow processes within a network of queues when setting up a simulation model, since these processes may influence the choice of the type of node at a particular junction in the model. In the presentcontext networks of queues with feedback are particularly relevant and are considered next.

4.3 FEEDBACK QUEUEING NETWORKS

A pumped storage system can be considered as an example of a queue with delayed feedback since water arriving at the lower reservoir is pumped to the upper reservoir and then at an appropriate time in the future is released back to the lower reservoir, generating power in the process. If the upper reservoir had an inflow and controlled outflow separate from the pumped storage inlet and outlet, then provided that the storage volume of the upper reservoir was an order of size greater than the volume of water pumped, the usual operation of the upper reservoir and its function as the upper pumped storage reservoir could be considered as independent, at least to a first approximation. This approximation would become less valid as the level of the upper reservoir approached either its maximum or minimum. The hydro system shown in Figure 1 could then be approximated by that shown in Figure 2.

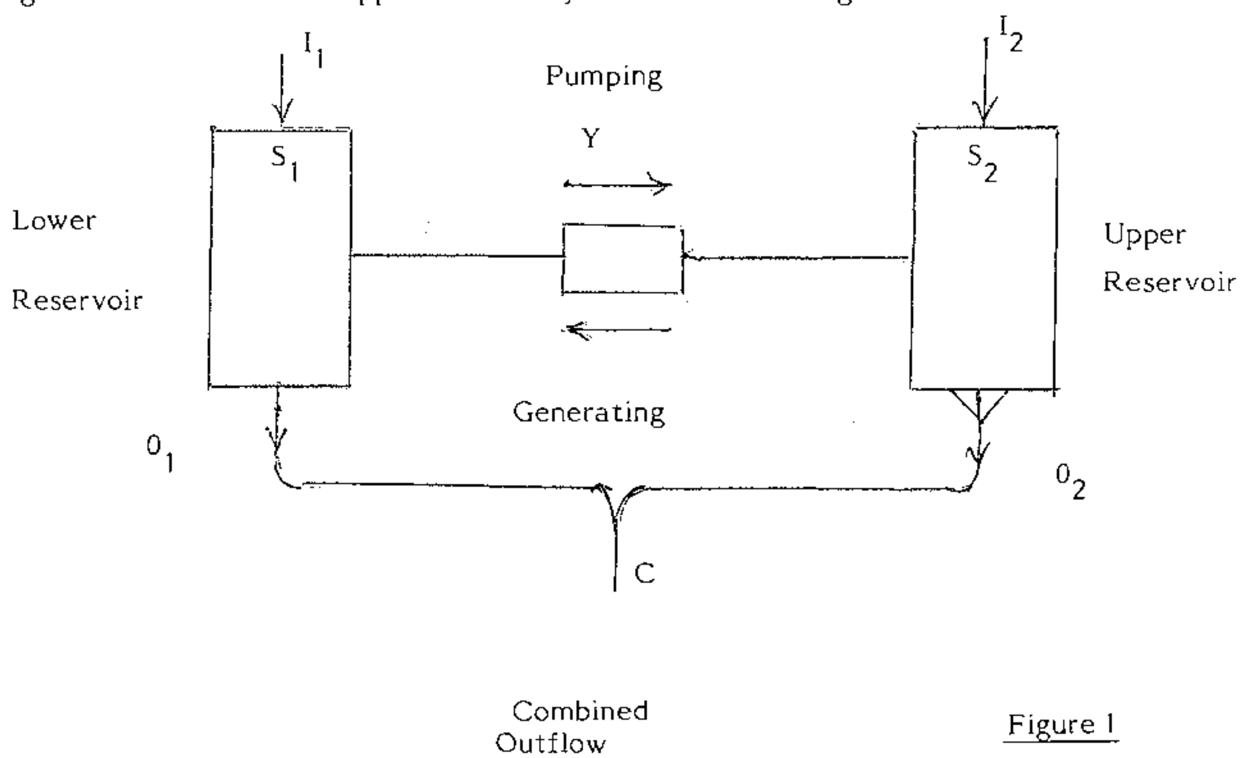


Figure 1

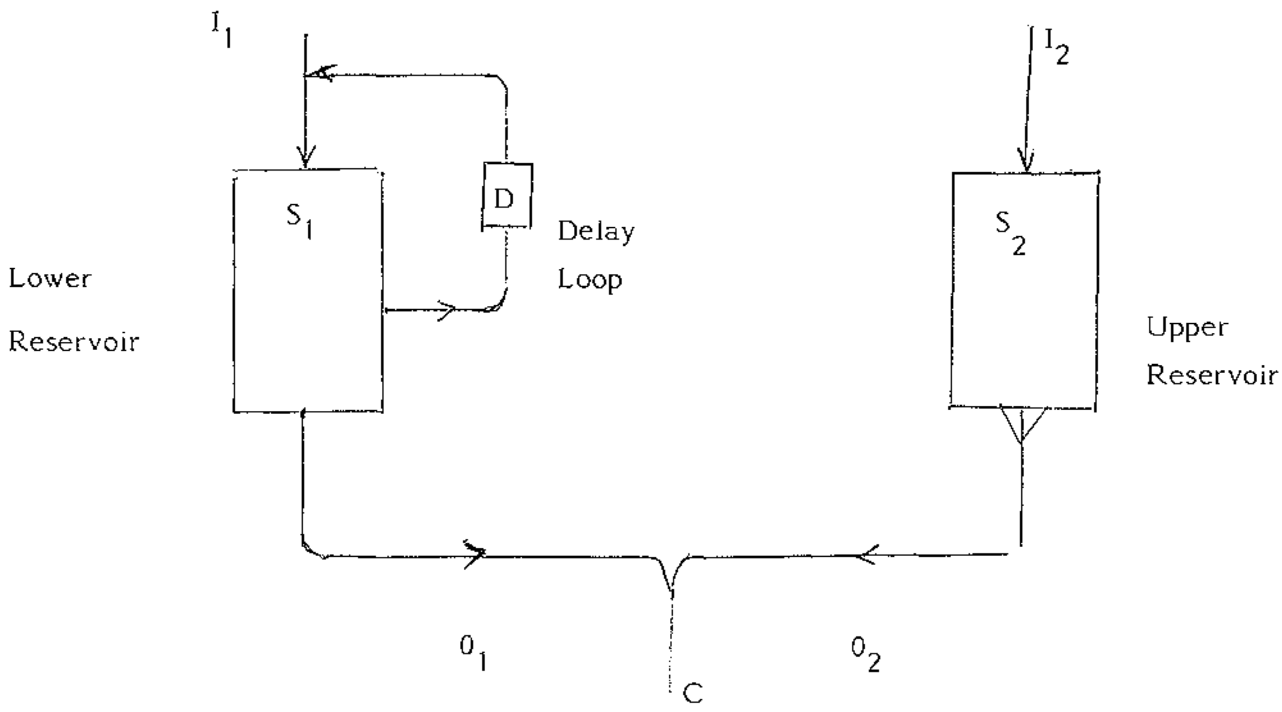


Figure 2

If there were powerstations downstream of the confluence C, one of the purposes of controlling outflow O_2 would be to avoid spilling at these stations. A second reason would be to store water for times when inflows I_1, I_2 were seasonally low. As noted in Chapter 1 the derivation and evaluation of distribution functions for time dependent queues is often a lengthy procedure and in the present application, the changing state of the system may invalidate the basis of the calculation. McNickle (1984) has indicated that simulations carried out on 2 and 3 node closed and open queue networks showed correlation in input and output streams before a steady state was reached. Consequently some care is needed when applying the queue-storage analogy described in Chapter 2 to networks of reservoirs. However, during periods in which conditions are approximately steady state the results of the theory of networks of queues can be usefully applied. Disney and Konig (1985) reference several such applications. Reference is also made to a study by Latouche (1979) on queues with correlated interarrival times. Two advantages of treating a system as a network of queues is the facility with which the varying nature of different parts of the system can be treated.

Thus a heavy traffic diffusion approximation for sojourn times in Jackson networks Reiman (1982) or the other heavy traffic approximation methods outlined in Chapter 1 might be appropriate for modelling heavy flows below the confluence point C in the previous figure. This modular approach also facilitates the use of simulation languages such as Simscript or Gasp which are particularly suited to handling the book-keeping aspects of multiple queue models.

Several papers have been published in which analytic solutions to the two stream problem in which only one stream is controlled. This is the same situation shown in Figures 1 and 2 except that the models do not include a feedback loop. Gani (1969) considers anextension of the above model by including possibly more than two tributaries and considering control dams either on a tributary or on the main river after its confluence with the tributary. This paper and that of Anis (1970) generally consider inflows which are mutually independent and without auto-correlation. Barnett and Phatarfod (1973) introduced cross-correlation between the inflows but these were still assumed to be serially independent. Barnett (1975) treats the inflows as independent Markov chains. Further generalisation is obtained by incorporating cross-correlation between the inflow streams. The introduction of the feedback loop to these models is made easier by the assumption of minimal disturbance on the upper reservoir behaviour by the pumped storage operation and that this operation is of a largely deterministic nature. The mean of the Markov chain inflow would then be increased or decreased according as water was released from or pumped to the upper reservoir. This type of operation was illustrated in Figure 2 of Chapter 2. If however, the pumped storage was used as spinning reserve or frequency control its operation might well be more intermittent. The need to be able to adjust the operation of a system quickly in response to changing circumstances has led to the development of computer programmes for system on-line control and planning purposes. These programmes are considered in the next chapter and exemplify the comment in Phatarfod (1975) concerning the complementary nature of simulation and analytic methods in system modelling.

CHAPTER 5 - OPTIMAL CONTROL OF RESERVOIR SYSTEMS

5.1 INTRODUCTION

Many optimization methods have been applied to finding solutions to different aspects of the problem of finding an optimal operating policy of a multireservoir, multipurpose hydro system. One of the reasons for this diversity is that there are several competing factors in the objective function and that the relative importance of these changes with the time scale of the optimization. Moder and Elmaghraby (1978) provide a discussion of short term and long term load forecasting methods. Because of the long lead time before a power station can be commissioned (10 to 12 years approximately) the uncertainty of the conditions that far into the future means that load predictions must be expressed as a distribution about a mean value. Shorter term load predictions take into account features such as recent weather patternsand recent demand. These different time scales influence operating policy since supply authorities are generally charged with the responsibility of providing power at minimum cost but at aprescribed level of security. Thus to use extensive hydro generation during periods of low inflow would mean low short term production costs but would jeopardize longer term security of supply as hydro storage was drawn down to save on thermal fuel bills. Earlier in the development of power systems each power station was connected directly to its load such as a mining operation or a town. As demand grew it became necessary to link power stations as well as building new larger stations since even these units were inadequate to supply on their own the larger demand centres. There were several advantages of this grid system over the former isolated supply/load method. These include fewer generators need to be held in reserve. Similarly, fewer generators need be held as spinning reserve to cope with any sudden changes in demand. Also the most economical plant could be used at times of lower demand and/or for base load, reserving older or less economical plant for times when demand exceeded the capacity of the lower order plant. In order to avoid transmission losses each region has approximately enough generating capacity to supply its own needs on average.

One disadvantage of such a grid system is the increased current that can be supplied to a fault, thus requiring larger circuit breakers for fault protection. For the system operator of a simple system of two generators of 50 MW and 70 MW capacity, the problem arises of how to allocate load to each generator if the total load is less than the total capacity 120 MW. Moder and Elmaghraby (1978) give an example in which the load is 60 MW and indicate that it was shown as early as 1933 using a Lagrangian multiplier method that the two machines should be run at an output at which their incremental costs ($/MW.hr) were equal. If however, power has to be transmitted on the grid to a distant load centre, it is more economical to generate power at the distant destination at a slightly higher incremental cost than to transmit power which although cheaper to produce is subject to transmission loss. This leads to a constraint of the form

$$\sum_{\text{(all units)}} \text{Power} = \text{Load} + \text{Losses} \qquad (5.1)$$

In a fixed voltage system since loss power = current^2 x resistance and power = voltage x current, doubling the power means doubling the current but in turn quadrupling the power loss. An adaptation of the no loss method described above, called the loss formula method (Kirchmayer 1958) enables the approximation to the optimal level load dispatching problem to be found when line losses are included. To obtain an exact solution to (5.1) involves solving equations (5.2), (5.3) while minimising fuel costs.

$$\text{Power} - \text{Load} = \sum_{\text{(all lines at the nodes)}} \text{Lineflows} \qquad (5.2)$$

$$\text{Vars} - \text{Loadvars} = \sum_{\text{(all lines at the nodes)}} \text{Linevars} \qquad (5.3)$$

where lineflows are the power flows away from a node in the network, vars are the reactive power generated at a node, loadvars are the reactive power demanded at a node and linevars are the reactive power that flows away from a node. Thus (5.2) states that the real power summation at a node must be zero and similarly (5.3) requires the summation of reactive power at a node to equal zero.

The problem thus has the form of a trans-shipment problem and a variety of mathematical programming methods have been advanced of which Sasson et al (1971) provides an assessment. The loss formula method has the advantage of being quick to apply but lacks the versatility of the MP methods Happ (1974).

In the above systems it was assumed that fuel was available for power generation whenever it was needed. Whilst in a thermal system this is a fair approximation (at least until fossil or nuclear fuels became scarcer), most renewable sources of energy have limited reserves at any given time so that (5.2) and (5.3) became time dependent. The objective function is still to minimise the fuel cost over the given period but in addition consideration must be given to the comparison of the time period and the natural cycle of storage in the hydro reservoirs. Papers which consider this time dependent aspect of the dispatch problem are considered in the following sections. A further aspect of the influence of time on the solution of (5.2), (5.3) is the variation in the daily load curve. Shutting down or starting up large thermal generating units can take many hours depending on design so that at a particular time an optimal dispatch solution may have a more expensive thermal plant running while a more economical plant is shut down for use at a later and more extended period. In a system with both hydro and thermal stations a sudden loss of generating capacity can be accommodated by keeping suitable hydro generators on standby at a lower cost than equivalent thermal units. Transmission line failure however, can be a more difficult problem to handle as each outage has a unique effect on the system, requiring a network flow calculation to identify and compensate for various approaches to modelling system security have been developed. Shen and Laughton (1970). Intentional outages of plant for maintenance provides another constraint when seeking an optimum dispatching solution. Other references to operating problems can be found in papers on the following topics:

Spinning reserve Anstine et al 1963, Billinton and Jain (1972), Bubenko and Anderson (1973), Scheduling of pumped storage plants by LP and DP methods Akiyama and Sekine 1961, Cobian (1971); Optimal Control theory applied to scheduling Fosha and Elgerd (1970), Bechart and Kwatny (1972), Yu and Moussa (1972); Load redistribution by altering network connections Couch and Morrison (1972); Sekine and Kawakawiz (1972). The Energy Plan prepared annually by the Ministry of Energy includes detailed discussion of planning and forecasts for all sectors of energy production and consumption in New Zealand.

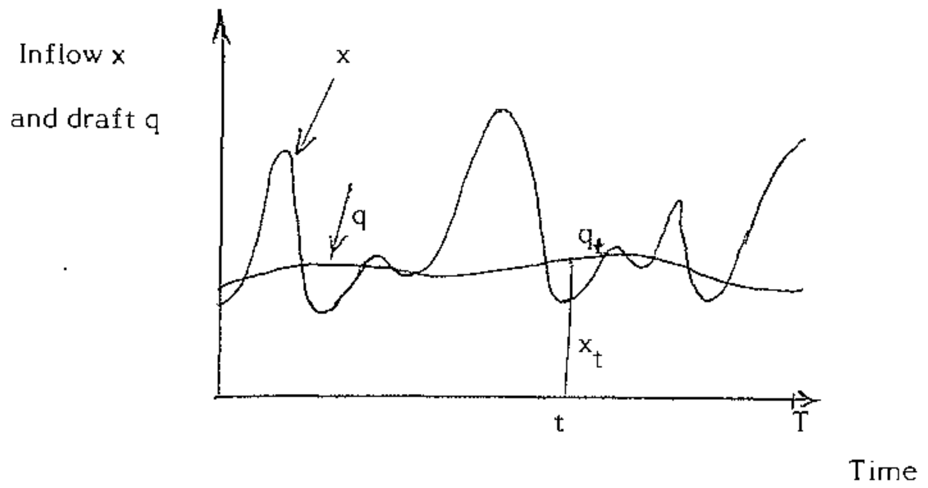
In particular details of the demand for electricity and plans to meet this projected demand for the next 15 year period are given. In the present context this information assists in making an assessment of the relevance of a pumped storage installation in the South Island to be guaged. This aspect is considered further in the next chapter.

5.2 METHODS OF OPTIMIZING DAM OPERATION

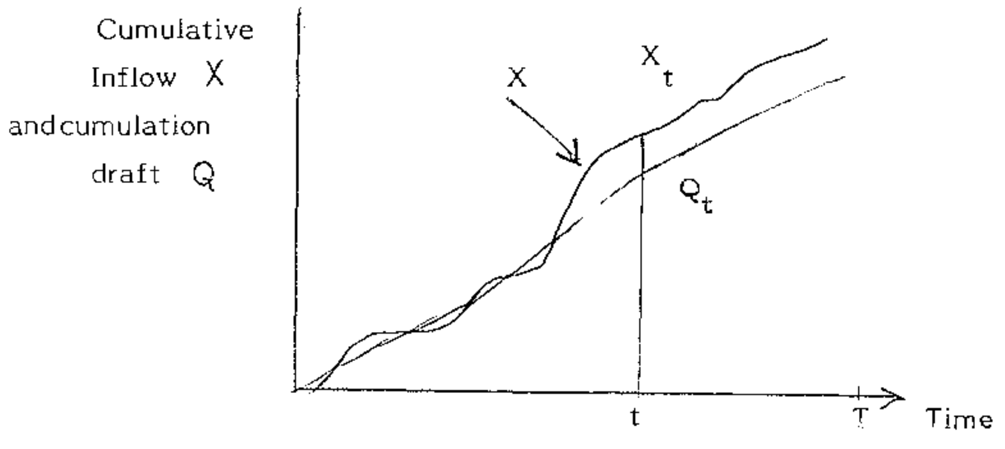
Since dams are expensive structures to build there are economic pressures for the widest use to be made of such a facility upon its completion. Tavares (1982) suggests that papers on the optimization of dam operation can be classified into three approaches (A) establishment of general operating rules; (B) the search for an optimal numerical solution for each set of data, using mathematical programming; and (C) estimation of best strategies by the application of statistical methods to operational results obtained by numerical experimentation or optimization. Historically, approach (A) was first and includes the early papers by Malterre (1912) and Rippl (1883) and others on the use of the mass curve of inflows and gives the optimal regulation if the greatest possible equalisation of outflows is desired and if complete knowledge of future inflows is known. Also, in this category are the models, mentioned earlier, by Moran, Lloyd, Phatarfod and others involving stochastic inflows together with a variety of release rules. In approach (B) mathematical programming methods are used, in particular linear programming (LP) and dynamic programming (DP) have been widely applied. Both of these methods can require a large amount of computer memory and a variety of techniques have been developed to overcome this limitation. Approach (C) involves the following steps
(a) solving the problem of determining the optimal release for serveral samples of inflow time series which are assumed to be deterministically known (usually, dynamic programming is used for this purpose).
(b) using statistical techniques (usually linear regression) to estimate the relationship between the optimal release and the system state. Askew et al (1971) describes this technique which proves to be more suitable for risk analysis than optimization studies as it is difficult to identify the functional form of the optimal releases policy. Method (B) has the limitation that the validity of the solution obtained is restricted to the data analysed whilst method (A) although obtaining the functional form of the optimal operating policy does so under operating assumptions of limited validity.

Tavares (1982) uses stochastic dynamic programming to develop analytic results and an algorithm for determining optimal releases from a dam with stochastic inflows. Tavares draws attention to the paper by Klemes (1979) which discusses storage mass-curve analysis and shows its close connection with DP and LP approaches. Klemes provides a useful introduction to early reservoir optimisation methods and concludes that although the mass storage-curve method has strict applicability to a rather special type of optimization problem, its current relative obscurity compared with LP and DP methods, with their greater generality and flexibility should be rectified. He suggests that the mass storage-curve technique has an intuitiveness which is lacking in the extended computations involved in mathematical programming methods. The mass storage-curve method was introduced by Rippl (1883) and as can be seen in the diagrams below is related to the fluid approximation of queues mentioned in Chapter 1.

In the following diagrams: $Z_t = \int_0^T (x - q)\, dt = X_t - Q_t$



(a)



(b)

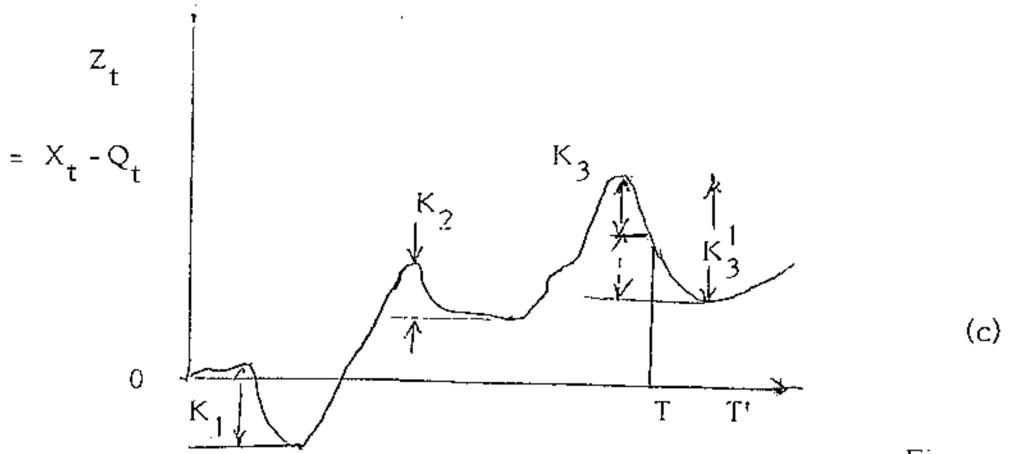


Figure 5.1

The heights marked K in Figure 5.1(c) represent the storage necessary to prevent the reservoir running dry during the given period [0,T'] indicates that a dam constructed with capacity K_3 would, with the inflow and draft curves shown, run dry. Klemes goes on to show the development of Rippl's mass-curve method to the streched-thread method shown in Figure 5.2 (due to Varlet (1923)) and relates this to recent mathematical programming methods.

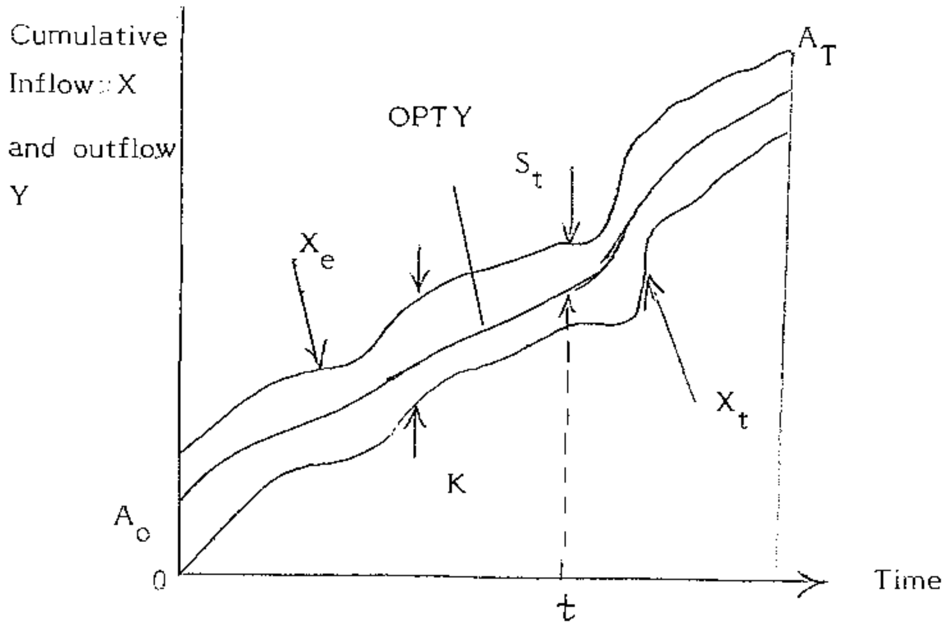


Figure 5.2

In Figure 5.2 K represents the reservoir capacity and mass inflow curves X_e and X_f are separated vertically by the value of K, X_e corresponding to X in Figure 5.1(b). Between and points A_o at t = o and A_T at t = T the optimum release mass-curve OPTY is shown to be the shortest path between these points constrained by the curves X_e and X_f.

Also mentioned in Chapter 1 were second order approximations to queueing models, one being the diffusion model. Bather (1968) discusses a diffusion model for the control of a dam. This models the inflow as a Wiener process but requires that the output process also be continuous. Conditions under which optimal release policies can be formulated are derived for a given input process. Other earlier development for optimal reservoir control was carried out by Thomas and Watermayer (1962) using a queueing model with independant inflows. Loucks (1968) extended this work to include serial correlation in the inflows. About this time mathematical programming methods were beginning to be applied to single as well as multiple reservoir systems. Often however, the size of the computational problems was such that numerous programming shortcuts and model simplifications were needed.

A recent paper by Boshier Manning and Read (1983) set out considerations determining the current hydro release policy in New Zealand and indicates that at that time a single reservoir model was used. A discussion of the influence of increasing diversity in thermal generation and fuel prices on scheduling procedures indicates that a more detailed approach than the 'Rule Curve' method used during the early 1970's was required. The effectiveness and efficiency of three different methods are discussed as in the extension of these methods to a two reservoir model. Individual hydro lake releases are determined on a basis of the likelihood of spill occuring. One advantage of pumped storage is the ability to reduce spilling but at the price of the power for pumping. This price depends on the current state of the rest of the system as well as expected future states. Models thus need to be sufficiently detailed and sufficiently short term to enable an evaluation of a proposed pumped-storage scheme to be made. In a paper by Boshier and Lermit (1977) a network flow method is used to determine optimum reservoir management over a time period of one to two years. The model is similar in concept to the networks of queues considered earlier and consists of a set of nodes interconnected by arcs representing the transfer of energy from one place to another or from one time or form to another.

The nodes represent energy at a particular time and in a particular form. An interesting result from this study is the high marginal value of additional water stored at Lake Hawea suggesting that although a net user of energy, a pumped-storage system at Hawea could be viable. By using such a method to estimate the marginal cost of energy at different times of the year and for peak and non-peak periods an estimate of the value of a pumped storage system could be estimated by developing a model as described in Chapter 4 to represent Lakes Hawea/Wanaka with pumped storage and then valuing the energy saved by reduced spillage at Roxburgh and Clyde. This decomposition of the optimality problem is however likely to produce a sub optimal solution since the amount of spill at Roxburgh and Clyde still depend, in part, on how the rest of the system is operated. In a paper by McMahon et al (1980) a feasability study of the introduction of pumped storage to an existing hydro system consisting of three dams and power stations in series with a dummy reservoir to represent pumped storage is described. This model also uses past Power station operating experience as a guide in the simulation studies. Considerable effort has been expended over the last three decades on developing models to obtain optimal operating schedules for systems of different configurations, different objectives and with optimization time periods varying from real-time to several years. Recent papers have provided useful surveys of this development each emphasizing one or more of the divisions of interest previously listed. These papers include the following Stedinger et al (1983) on Screening Models; Grygier et al (1985) on Algorithms for Optimising Hydropower Systems; Pereira et al (1985) on Stochastic Optimization using a decomposition approach; Marino et al (1985) discuss a dynamic model for multireservoir operation which treats inflows as a multivariate autoregressive process; Marino et al (1983) discusses optimal reservoir management from a reliability of freedom from flood and drought viewpoint.

5.3 A PRACTICAL EXAMPLE AND CONCLUSION

The Clutha river and the three main lakes which it drains viz Lakes Wanaka, Hawea and Wakatipu were first investigated for hydro power generation in 1904 but at that time it was concluded that development of the acknowledged potential would be difficult and expensive. In 1944 a further survey was carried out and in 1949 the present site of Roxburgh power station was approved by the Government, the first two generators being commissioned in 1956. With the present construction of the Clyde dam and the proposed construction of Queensberry and Luggate dams further upstream, the value of water stored in Lake Hawea (which has a dam at its outlet to control outflow) increases considerably. Lake Wakatipu has slightly less storage capacity than Lake Wanaka and is controlled by gates that were built in connection with gold mining in the Kawarau river into which the lake discharges. Lake Wanaka is uncontrolled due to the Lake Wanaka Protection Act. The Environmental Impact Report on 'Upper Clutha Valley Development' prepared by the Minister of Works and Development (April 1985) suggests in section 3.6.6 that only limited amounts of power would be available from a pumped storage scheme on Lakes Hawea or Wanaka. Appendix 24 of the same publication notes the attractiveness of a pumped storage scheme in the region between Lakes Hawea and Wanaka called The Neck. The Lake Wanaka Protection Act might be eased if it could be shown that periodic flooding of Wanaka township could be lessened in extent by providing this additional outlet. Alternatively, two or three pumping/generating stations could be established on the Hawea Flats between the outlet of Lake Hawea and the confluence of the Hawea and Clutha rivers. The report goes on to suggest that such schemes might become cost effective at a time when the peaking power or other benefits of pumped-storage were at a higher premium. In either of the regions proposed for pumped storage schemes the head available is approximately 65 metres. Using the formula KW 8.6 QH where Q is the flow rate (m^3/s) and H is the head available (m), the flows required for corresponding powers is listed over page.

KW	Q
25,000	44.7
50,000	89.4
75,000	134.2
100,000	178.8

The addition of water to Lake Wanaka at a rate of 100 cumecs for 12 hours would raise the level of the lake by 3.2 cm. For Lake Hawea the figure would be 2.24 cm (using CMD/M = 1550.9 for Wanaka and 2233.8 for Hawea). Thus a daily cycle or perhaps a weekly cycle pumped storage scheme might prove to be environmentally acceptable. The DC link restraint noted in the report appended, concerning the value of Lake Wanaka storage, may be removed if the 1984 Energy Plan considerations regarding increasing the DC link capacity are implemented. To evaluate the economic viability a comparison of the present system needs to be made with the system augmented by the pumped storage scheme with both systems evaluated under a range of load, inflows and fuel costs. Recent programming developments such as those reported in Marno et al (1985) may lessen the amount of computation involved.
Several papers concerning computer simulation procedures were published in the early 1960's. These papers (Ehrenfeld and Ben-Tuvia (1962) compare the efficiency of different procedures by simulating a simple queue; Conway (1963) and Cramer (1964) discuss the planning of a digital simulation are simulated by two recent papers which consider the planning and interpretation of a computer simulation Friedman (1984), Law (1983). Fishman (1978) provides a general text on discrete event simulation.
In conclusion, the analysis of reservoir systems can be carried out using queueing networks and related methods but the methods lack the flexibility of the more recently developed mathematical programming techniques. The results of the latter methods require careful statistical interpretation, however, and the computational methods involved lack the intuitive approach of the more direct network and flow apporaches. The latter by providing analytic solutions to the behaviour of smaller sub systems, enable parameters and constraints to be evaluated for insertion into the simulation program.

APPENDICES

1. This chart shows the location and type of power stations in New Zealand. The long transmission lines from South Island hydro stations to main North Island load centres as well as the total concentration of thermal stations in the North Island are also made clear. Source: Information Services, Electricity Division.

2. This chart shows a closer view of the upper Clutha catchment area, Lakes Hawea, Wanaka and Wakatipu and the Clutha and Kawarau rivers. The Energy Plan should be consulted for current information on the location and investigation of proposed hydro developments in the region. Source: Information Services, Electricity Department.

3. This topographical map of the region surrounding Lakes Hawea and Wanaka and their tributaries shows the areas known as The Neck and Hawea Flats which have been considered as technically feasible sites for pumped storage installations. Source: Lands and Survey Department.

4. This summary of a report prepared by the author in 1982 whilst in the employ of the Electricity Division, considered some aspects of controlling the outflow of Lake Wanaka (at present uncontrolled) and is included here to enable a comparison to be made with the measure of control obtained by introducing pumped storage. The views and conclusions expressed in the report do not and have not been representative of official Electricity Division policy.

   The three graphs each show a comparison over the period 1965-1980, of the uncontrolled level of Lake Wanaka and the level resulting from the construction of a dam across the lake outlet as at Lake Hawea. In Graph 1 the height to which the stored water could rise was unlimited. In Graph 2 the height of the stored water was limited to 280m above sea level. In Graph 3 this height was increased to 283m. In each graph the lower curve is the level of the uncontrolled lake. If the stored water reached these maximum levels the additional inflow was spilled.

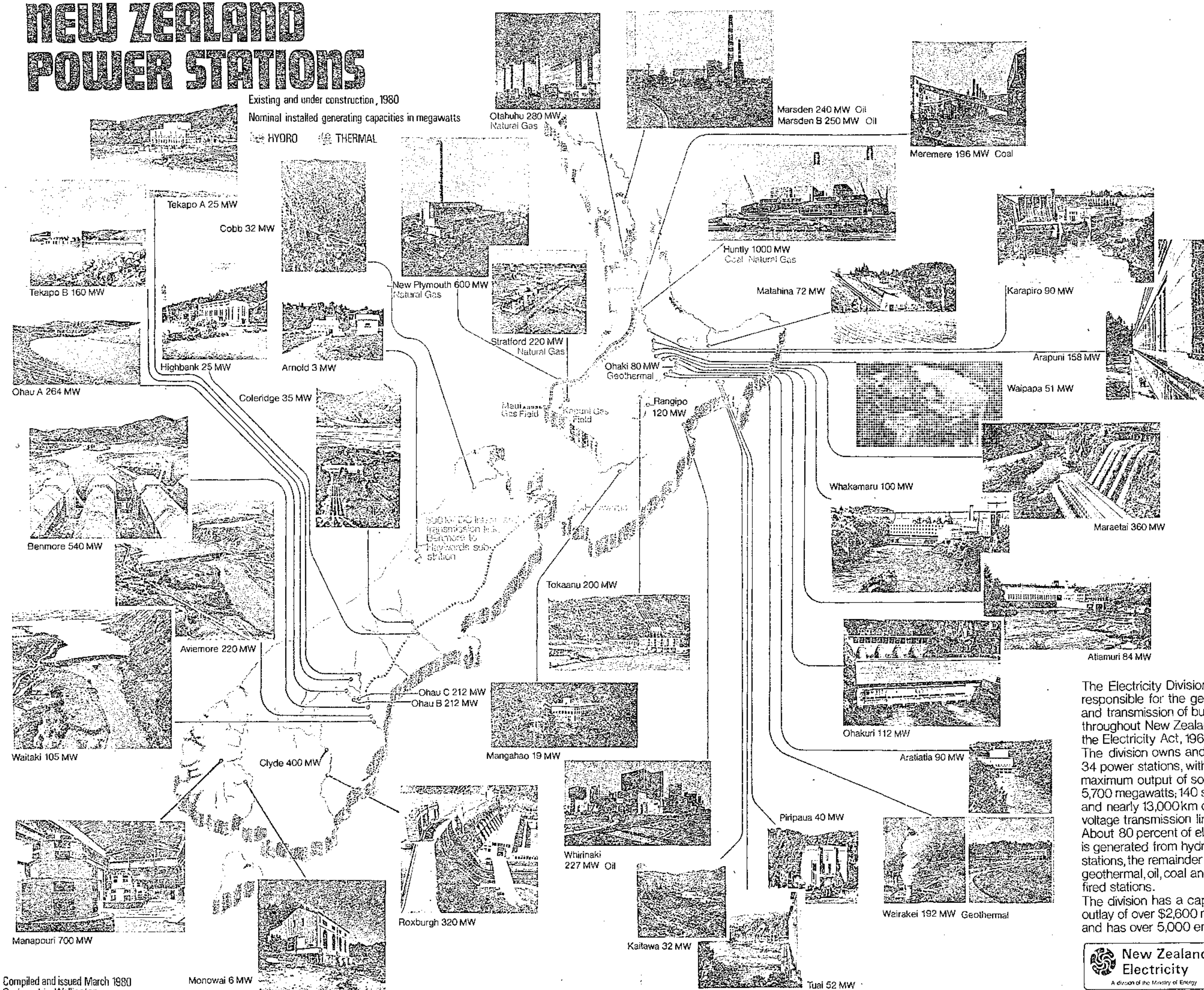
5. References.

NEW ZEALAND POWER STATIONS

Existing and under construction, 1980

Nominal installed generating capacities in megawatts

HYDRO THERMAL



The Electricity Division is responsible for the generation and transmission of bulk electricity throughout New Zealand under the Electricity Act, 1968.
The division owns and operates 34 power stations, with a maximum output of some 5,700 megawatts; 140 substations, and nearly 13,000 km of high voltage transmission lines.
About 80 percent of electricity is generated from hydro stations, the remainder from geothermal, oil, coal and gas fired stations.
The division has a capital outlay of over $2,600 million and has over 5,000 employees.

New Zealand Electricity

A division of the Ministry of Energy

Compiled and issued March 1980
Serigraphic, Wellington

CLUTHA VALLEY DEVELOPMENT

Proposals for Clutha Valley development were first put forward in 1904 when Mr P S Hay, a Public Works engineer, and Mr L M Hancock, an American electrical engineer, made a survey of the country's hydro resources and reported their conclusions to the Government:

"The expense of development here (the Clutha River) is too great and the difficulties too many to warrant work being done."

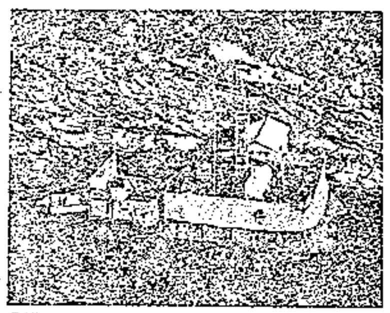
As engineering techniques improved and the need for electric power became greater, the potential of the Clutha Valley was reassessed. In 1944 a comprehensive survey of the Kawarau and Clutha Rivers was carried out. The Roxburgh power station was built by 1956, followed in 1959 by the Hawea dam.

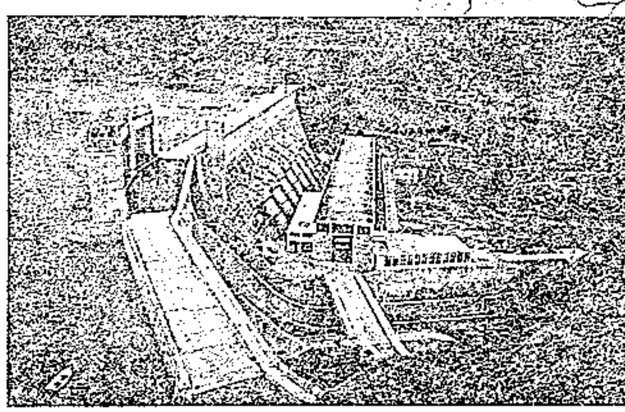
Present proposals for developing the Upper Clutha (Clutha and Kawarau Rivers) are based on studies in the 1960s and 1970s to investigate energy and other resources of the Clutha.

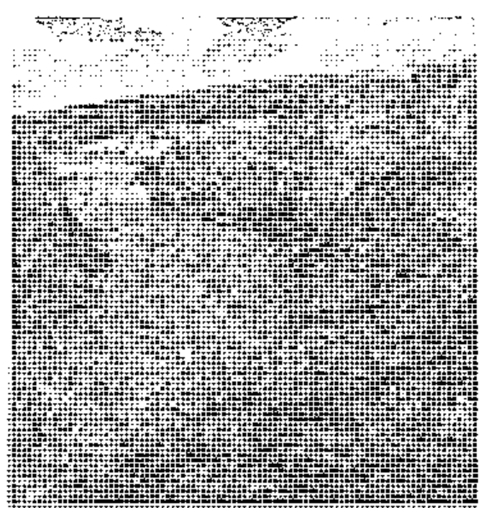
In 1973 the Government set up the Clutha Valley Development Commission to assess in more detail potential power schemes for the river and their impact on the local area.

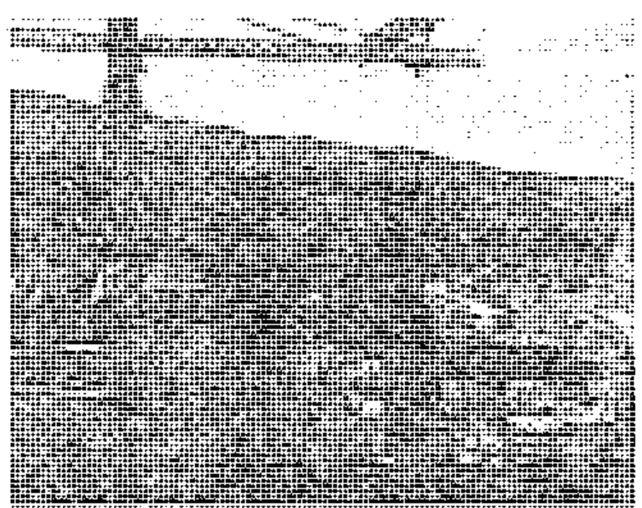
In 1976 a scheme was selected involving 5 power stations (3 on the Clutha River at Clyde, Luggate and Queensberry, and 2 on the Kawarau River).

Construction of Clyde power station began in 1982 and it is expected to be fully commissioned by 1990. Queensberry and Luggate power stations are planned for commissioning in the 1990s. Proposals for the Kawarau River are still in the investigation stages.

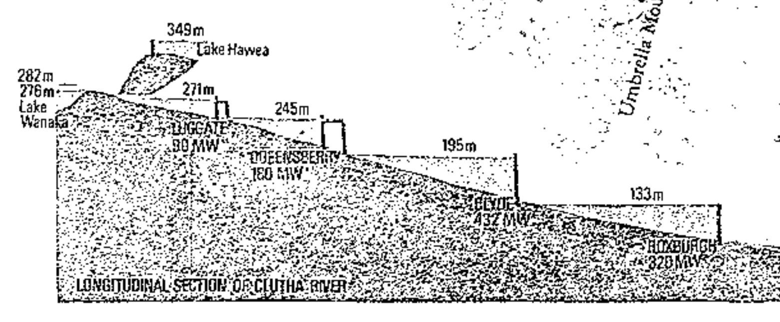

Drilling barge investigating geological conditions on the Kawarau River

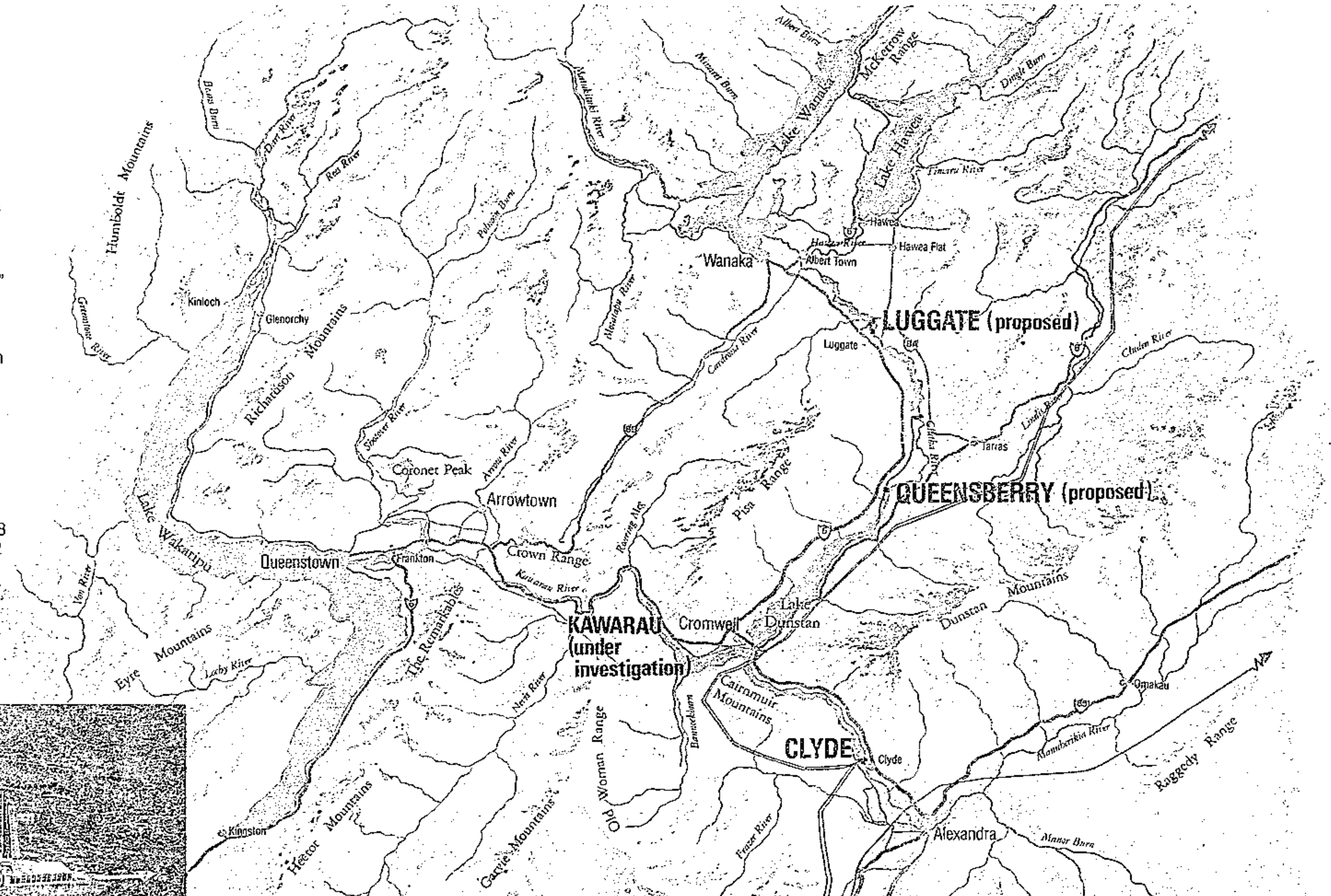

Roxburgh power station


Kawarau River


Construction at the Clyde dam site


Archaeological dig in the Kawarau Gorge





Tyrwhitt Peak
Harwich Island
Mt Gold
Stevensons Arm
Stevensons Island
THE PENINSULA
Crescent Island
Lookout Hill
Colquhouns Flat
West Wanaka
The Island
Roys Peninsula
Bishops Bay
Paddock Bay
Glendhu Bluff
Parkins Bay
Glendhu Bay
Glendhu
Motor Camp
MOTATAPU ROAD
WANAKA MT ASPIRING ROAD
Roys Peak T.V. Translator
Spear Grass Valley
Ridgeway
MAUNGAWERA RD
Rods Creek
Mt Burke
Mt Brown
Fork Farm
Dublin Bay
Bull Island
Beacon Point
BEACON POINT ROAD
Motor Camp
Ruby Island
Eely Point
Bremner Bay
AUBREY RD
ANDERSON RD
MT IRON ROAD
Mt Iron
Roys Bay
Hotel
Show Ground
Wanaka
Wanaka Airfield
GOLF COURSE RD
Cem.
Cardrona River
Sub Station
Albert Town
Pioneers Memorial
Mt Maude
Quartz Creek
Motor Camp
Control Hotel Dam
Hawea
CEMETERY ROAD
MUIR ROAD
Pumphouse
DOMAIN ROAD
Crosshill
T.V. Repeater
VHF Station
Maungawera
Horseshoe Bend
HAWEA RIVER
NEWCASTLE ROAD
Camp Hill
CAMP HILL ROAD
BUTTERFIELD ROAD
Hawea Domain
Hawea Flat
School
Hawea Downs
Lower Hawea Flat
LOACH RD
McLENNAN RD
McCARTHY RD
GLADSTONE ROAD
Ardmore
GRAYS RD
Lakeview
Mt Grand
BACK ROAD
HAWEA
Gladstone
Johns Creek
The Nook
Hospital Creek
Lagoon Valley
Lagoon Creek
Dam
SMALL RD
Coolah
CLUTHA RIVER

Burn or Nottingly River
Picnic site
Boat launching site
Wharf Creek
Big Hopwood Burn
Triplet Mt
Windy Point
M.W.D. Depot
Hut
00
Boundary Creek
10
Little Hopwood Burn
Terrace Peak 6650
McGregor Creek
5632
TEAT RIDGE
Sheepskin Creek
East Burn
Camp Peaks 5850
Waterfall Creek
Snag Bay
50
5675
Terrace Creek
Camp Creek
South Branch
Sentinel Peak 5943
The Four Peaks
Minaret Station
Minaret Bay
Cattle yards
Hunter Valley Station
Boat ramp
The Neck
HUNTER VALLEY ROAD
Picnic area
Kidds Bush
Bushy Point
Silver Island
The Peninsula
Dingle Burn
Dingle Burn Station
Waterfall
Silver Burn
HAWEA
Isthmus Peak 4650
Halls Creek
Rocky Point

15th January 1982

VALUE OF WANAKA STORAGE

PART 1

In this report I have used the computerised hydrology records to estimate the value of increased Wanaka storage. The method used was to take the spill recorded at Roxburgh during September to April of each yearly period and convert this to an increase in the level of Lake Wanaka. During the months May to August the water stored in Wanaka is released since during this part of the year demand is higher than during the summer months. In addition, the inflows during May to August are lower and the average spill is 14 cumec compared with an annual average spill of 85 cumec. The value of and the amount of water stored for winter release is dependent on the load at Roxburgh and this is determined by the way the rest of the power system is run throughout the year. In the first part of this report it was assumed that the present operating pattern would be continued but that during the winter period the load at Roxburgh would be increased by reducing thermal generation. In the second part of this report an LP model of the generating system was used to examine the ability of the system to utilise this additional power with particular regard to the capacity of the DC link. Table A shows the average spill at Roxburgh from October 1965 to April 1981 to be 85 cumecs. If sufficient storage was available to ensure that all of this water could be utilised from the Clutha with only Roxburgh operating at such times so that the value of the resulting generation was 4c/kwh, then an upper bound on the average value of the increased storage could be determined as follows:

$$\text{value of storage} = (85\ \text{cumec}/2.55\ \text{cumec/MW}) \times 0.04\ \text{c/kwh} \times 10^3 \times 8760$$

$$= \$11.68\ \text{M/annum}$$

A worst case can be considered by combining the maximum monthly spills from January to December the annual spill is found to be 223 m^3/s. This increases the \$11.68 M to \$30.64 M/annum. The savings from the completed Clutha system are considered below.

It is apparent from Graph 1 that if the maximum lake level is unrestricted very high lake levels can occur and it is more realistic to consider a maximum lake level of about 280m (see Graph 3). The table below shows the potential average annual savings resulting from the reduction in winter thermal operation due to the increase in Wanaka storage if northwards transmission is not constrained by the DC link.

Table 1

Maximum Control Level (m)	Average Annual Savings @ 4c/kwh
278	\$0.54 M
279	\$1.24 M
280	\$1.90 M
281	\$2.50 M
282	\$3.06 M
283	\$3.53 M

The tables over page show the monthly spill at Roxburgh at different maximum lake control levels. These show that the average spill is reduced from an historical average during the study period (01.09.65 to 31.08.80) of 85 m^3/s to approximately 68m^3/s with the maximum control level at 280m. The energy generated/cumec flow through the complete Clutha chain (Luggate, Queensberry, Clyde, Roxburgh) is approximately 3.9 times the generation/cumec flow through Roxburgh only.

SUMMARY:

If the flow from Lake Wanaka was restricted at times of spill at Roxburgh, with this flow being stored in Lake Wanaka during September to April and released during May to August, a useful reduction in winter thermal generation could be achieved when the DC link is not a constraint. During the storage period flow may have to be increased from Wanaka to ensure that the lake level did not rise above an acceptable level.

Table 1 enables the potential return on the water stored in Wanaka with various maximum control levels to be estimated. For example, if a maximum control level of 280m was acceptable to all interested parties, the value of the reduced thermal generation resulting from the increased Wanaka storage when the DC link is not a constraint would be $1.9 M/annum rising to $7.4 M/annum when all the Clutha stations have been completed. For comparison, a maximum control level of 280m allows a storage of 8150 CMD or approximately 1/3 the storage of Lake Hawea.

PART 2

An alternative approach to the question of storage at Lake Wanaka is to compare the fuel costs of thermal running with various levels of Wanaka storage in the recently developed LP regional model called "Regel". This model optimises the system operation in such a way as to minimise fuel costs subject to various operating constraints. The LP program showed that thermal running is not influenced by the availability of storage at Wanaka with the present DC link capacity for northwards transmission. To test the importance of this constraint, the model was run with the DC link capacity unbounded. Under this condition the northward power transmission rose to 1150 MW from the present limit of 540 MW. This indicates that not all of the additional generation at Roxburgh occuring with storage at Wanaka, would be transferable northwards via the DC link and that the savings on thermal running would be proportionately less than indicated in the study above. With the DC link unconstrained, comparison of the fuel costs with and without storage at Wanaka shows a saving of about $0.64 M per annum in a moderately wet year (1978) (assuming 8150 CMD storage at Wanaka). Because in practice the system is operated without the hindsight available to the LP program using historical data, this figure represents a minimum saving and is considered to be consistent with the savings calculated earlier in this report. The LP program indicates a reduction of about 0.5 PJ of Maui gas at Huntly, and a reduction in spillage at Roxburgh by 6300 CMD per annum.

CONCLUSION:

On the basis that the historical record of operation of the Clutha system provides a useful guide to future performance, this study indicates that a reduction in spill at Roxburgh could be achieved if Lake Wanaka were to be made available for hydro storage. Further discussion is required to determine whether a level of storage can be found which is environmentally acceptable and which provides an economic rate of return. The LP model indicates that in certain circumstances the DC link would act as a constraint on the amount of power which could be transmitted northwards and thus reduce the amount of thermal running. Even allowing for the overly optimistic results produced by the LP algorithm it is apparent that the DC link could be a limitation on the value of Wanaka storage.

Further study with projected 1990 demands which included an increased smelter load in the South Island and Ohaki and Marsden B thermal stations comissioned in the North Island, showed that with 1978 hydrology (moderately wet) a similar saving ($0.4 M) in thermal running resulted if the DC link was unconstrained. Imposition of the DC link constraint removed this saving.

In conclusion this study indicates that storage at Wanaka could be used to reduce spill at Roxburgh but that the DC link and increased S.I. high load factor demand place limitations on the value of the related potential generation for decreasing North Island thermal running.

T Craig

TABLE A
Roxburgh Historical Spill (m^3/s)

YEAR	JAN	FEB	MAR	APR	MAY	JUN	JLY	AUG	SEP	OCT	NOV	DEC	MEAN
1965	?	?	?	?	?	?	?	0	0	4	177	249	?
1966	354	249	35	72	3	0	0	0	0	0	0	56	63
1967	99	263	267	227	182	0	0	48	0	0	92	628	150
1968	219	13	412	30	93	3	0	15	1	160	467	501	161
1969	251	41	44	0	0	0	0	0	339	21	0	220	81
1970	146	83	0	0	0	0	0	42	820	503	430	379	200
1971	101	0	0	0	0	43	0	0	4	437	219	165	82
1972	313	13	162	86	0	0	0	0	234	82	371	22	107
1973	0	1	0	0	0	0	0	0	0	35	405	0	36
1974	0	0	0	0	0	0	0	0	0	0	0	0	0
1975	0	21	91	553	175	28	0	71	9	26	24	151	96
1976	20	0	0	0	0	2	0	0	0	0	0	7	3
1977	0	0	0	0	0	0	0	0	0	6	0	0	0
1978	0	0	0	0	o	o	o	o	o	199	0	0	17
1979	0	0	25	0	62	0	0	0	0	0	0	634	61
1980	462	339	0	6	0	20	0	98	530	553	375	302	223
1981	1	154	308	75	?	?	?	?	?	?	?	?	?
MIN	0	0	0	0	0	0	0	0	0	0	0	0	0
MEAN	123	74	84	66	34	6	0	17	125	127	160	207	85
MAX	462	339	412	553	182	43	0	98	820	553	467	634	223

TABLE B
Roxburgh Spill (m^3/s) Lake Wanaka max. level 280m

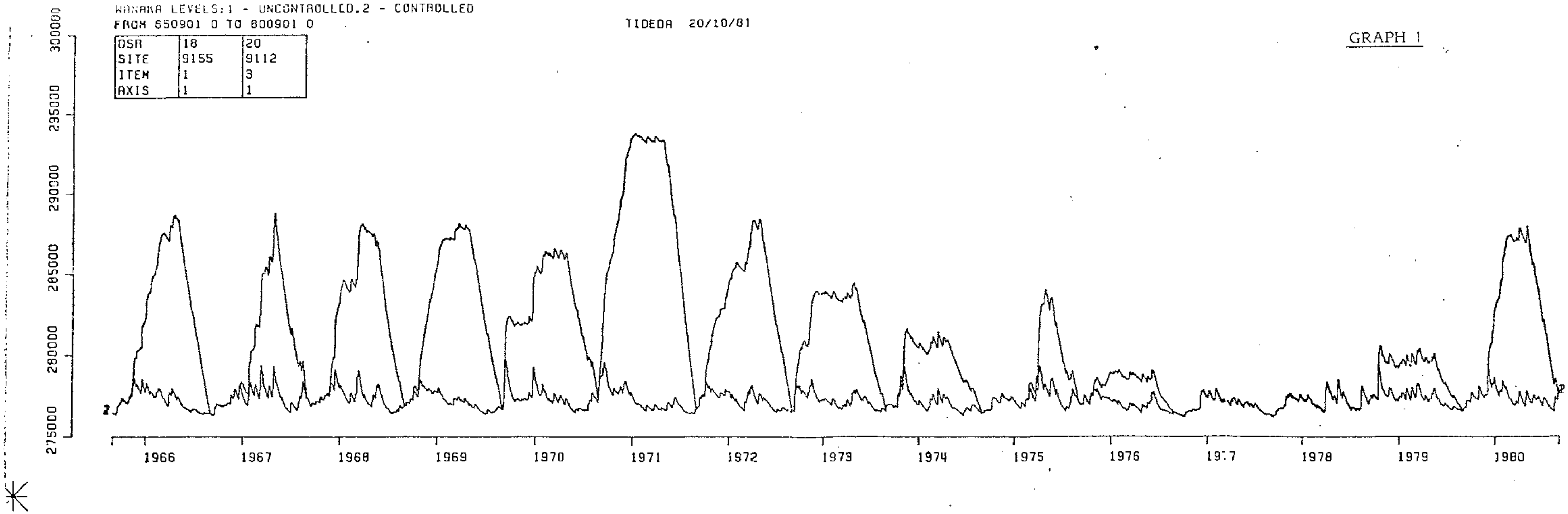
YEAR	JAN	FEB	MAR	APR	MAY	JUN	JLY	AUG	SEP	OCT	NOV	DEC	MEAN
1965	?	?	?	?	?	?	?	?	0	0	48	212	?
1966	336	271	32	69	0	0	0	0	0	0	0	11	59
1967	22	208	284	312	182	0	0	19	0	0	36	536	133
1968	204	4	420	13	124	0	0	10	0	204	425	490	159
1969	218	21	76	0	0	0	0	0	252	0	0	293	72
1970	101	29	13	0	0	0	0	102	829	447	443	349	193
1971	71	0	0	0	0	5	0	0	0	292	186	185	62
1972	277	5	232	48	0	0	0	0	266	58	366	2	104
1973	0	0	0	32	11	7	0	0	0	3	298	0	29
1974	0	0	0	0	0	0	0	0	0	0	0	0	0
1975	0	12	152	491	180	13	0	15	5	0	0	80	79
1976	7	0	0	0	0	0	0	0	0	0	0	0	1
1977	0	0	0	0	0	0	0	0	0	0	0	0	0
1978	0	0	0	0	0	0	0	0	0	91	0	0	8
1979	0	0	3	0	19	0	0	0	0	33	0	647	60
1980	464	267	21	18	2	26	0	60	?	?	?	?	?
MIN	0	0	0	0	0	0	0	0	0	0	0	0	0
MEAN	113	54	82	66	35	3	0	14	90	75	120	187	68
MAX	464	271	420	491	182	26	0	102	829	447	443	647	193

TABLE C

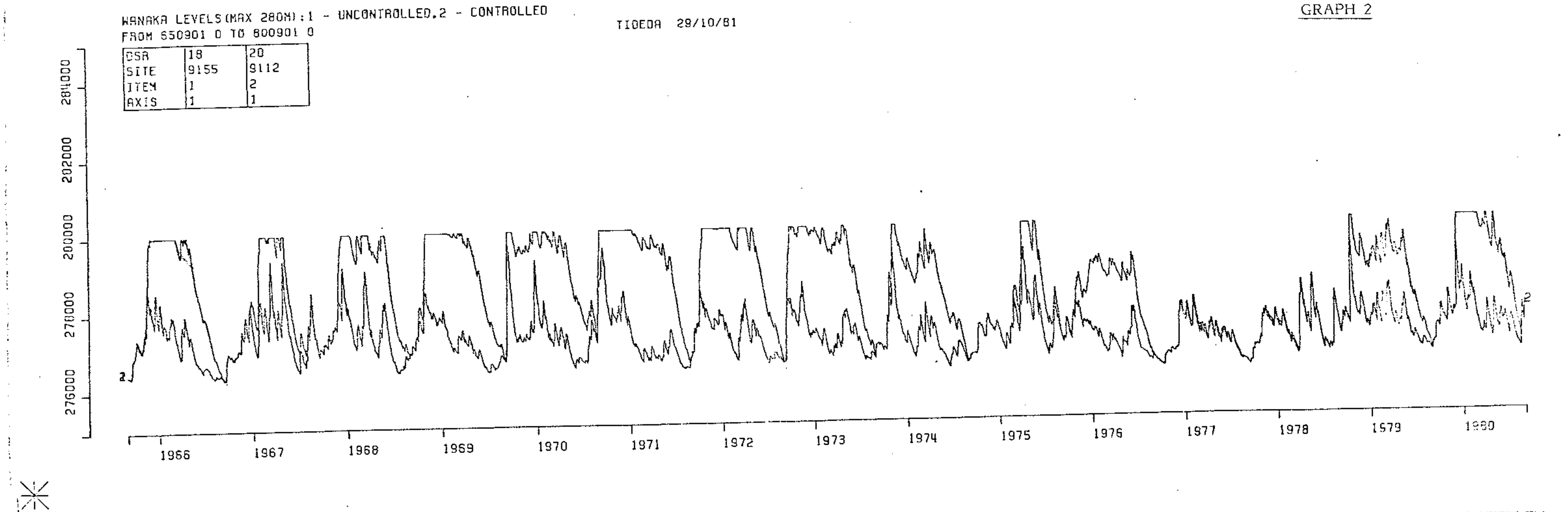
Roxburgh Spill (m^3/s) Lake Wanaka max. level 283m

YEAR	JAN	FEB	MAR	APR	MAY	JUN	JLY	AUG	SEP	OCT	NOV	DEC	MEAN
1965	?	?	?	?	?	?	?	?	0	0	24	72	?
1966	284	271	32	69	0	0	0	0	0	0	0	11	54
1967	14	64	206	312	216	0	0	45	0	0	36	320	102
1968	204	4	420	13	124	0	0	10	0	204	425	490	159
1969	218	21	76	0	0	0	0	0	72	0	0	250	54
1970	101	29	13	0	0	0	0	102	829	447	443	349	193
1971	71	0	0	0	0	9	0	0	0	199	88	157	44
1972	277	5	232	48	0	0	0	0	266	58	366	2	104
1973	0	0	0	32	46	32	0	0	0	3	178	0	24
1974	0	0	0	0	0	0	0	0	0	0	0	0	0
1975	0	0	62	493	135	32	0	34	5	0	0	80	70
1976	7	0	0	0	0	0	0	0	0	0	0	0	1
1977	0	0	0	0	0	0	0	0	0	0	0	0	0
1978	0	0	0	0	0	0	0	0	0	51	0	0	4
1979	0	0	3	0	12	0	0	0	0	0	0	512	45
1980	464	267	21	18	10	44	1	85	?	?	?	?	?
MIN	0	0	0	0	0	0	0	0	0	0	0	0	0
MEAN	109	44	71	66	36	8	0	18	78	64	104	149	61
MAX	464	271	420	491	216	44	1	102	829	447	443	512	193

GRAPH 1



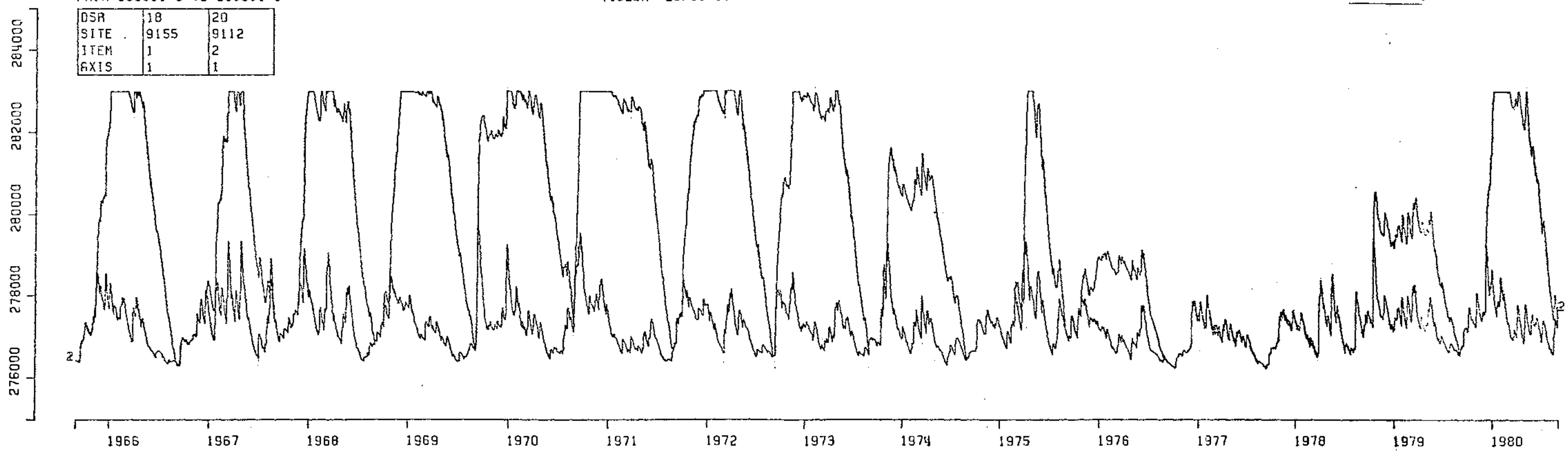
GRAPH 2



WANAKA LEVELS (MAX 283M): 1 - UNCONTROLLED, 2 - CONTROLLED
FROM 650901 0 TO 800901 0

TIDEDA 29/10/81

GRAPH 3



REFERENCES

CHAPTER 1

BOROVKOV, A.A., 1976. 'Stochastic Processes in Queueing Theory' Springer-Verlag

GROSS, D and HARRIS, C.M., 1974. 'Fundamentals of Queueing Theory' Wiley

JAISWAL, N.K., 1968. 'Priority Queues' Academic Press

KENDAL:, D.G., 1953. Ann. Math. Statist., Vol 24 pp 338-354

KINGMAN, J.F.C., 1966. 'On the Algebra of Queues' Methuen

KLIENROCK, L. 1976. "Queueing Systems' Volumes I, II Wiley

KOSTEN, L., 1973. 'Stochastic Theory of Service Systems' Pergamon Press

MORSE, P.M., 1957. 'Queues, Inventories and Maintenance' John Wiley

NEWELL, G.F., 1972, 1982. 'Applications of Qucueing Theory' Chapman and Hall

PAGE, E., 1972. 'Queueing Theory in OR' Butterworths

PARZEN, E., 1962. 'Stochastic Processes' Wiley

PRAHBU, N.U., 1974. Mathematical Methods in Queueing Theory Springer-Verlag

RIORDAN, J., 1962. 'Stochastic Service Systems' Wiley

SAATY, T.L., 1961. 'Elements of Queueing Theory' McGraw-Hill

SYSKI, R., 1967. Queueing Theory (Editor R Crvon) English Universities Press

TAKACS, L., 1961. 'Introduction to the Theory of Queues' Oxford University Press

CHAPTER 2

MORAN, P.A.P., 1954 Aust. J. App. Sci 5, 166-124

MORAN, P.A.P., 1959 'The Theory of Storage' Methvens

GHOSAL, A. 1970. 'Some Aspects of Queueing and Storage Systems' Springer-Verlag

COCHRAN, A.M. et al, 1979. Proc. IEE Vol 126, No. 5 pp433-438

PHATARFOD, E.M., 1975 J Hydrol., 30 : 199-217

GOULD, B.W., 1961. J. Inst. Eng., Aust., 33(12) : 405-416

WALD, A., 1947. Sequential Analysis Wiley

PHATAFOD, R.M. amd MARDIA, K.V., 1973. J. Appl. Probab, 10 : 166-180

PAKES, A.G., 1973. J. Appl. Prob., 10 : 317-329

BARR, H., 1981. 'Water flow variability and the Value of Oil and Gas Firming in a Hydro Thermal Generating System' D.S.I.R. App. Maths. Divn.

LINDLEY, D.V., 1952. Proc lamb. Phil. Soc, 48, 277-289

KENDALL, D.G., 1951. J. Roy. Statist. Soc., B13, 151-185

KINGMAN, J.F.C., 1966. J. App. Prob., 3, 285-326

FINCH, P.D., 1958 J. Roy. Stat. Soc. 520 182-186

FINCH, P.D., 1960. Biometrika 47, 45-52

FINCH, P.D., 1961. J. Aust. Math. Soc., 2, 217-228

PRAHBU, N.U., 1964. J. Appl. Prob., 1, 1-46

LLOYD, E.H., 1963. J. of Hydrology, Vol 1, pp 99-128

PRAHBU, N.U., 1980. 'Stochastic Storage Processes' Springer-Verlag

CHAPTER 3

MAYR, O., 1970. 'The Origins of Feedback Control' M.I.T. Press

BROGAN, W.L., 1974. 'Modern Control Theory' Quantum

SHINNERS, S.M., 1978. 'Modern Control System Theory and Application' Addison-Wesley

GUMOWSKI, I. and MIRA, C., 1968. 'Optimization in Control Theory and Practice' Cambridge University Press

ROSS, S.M., 1983. 'Introduction to Stochastic Dynamic Programming' University Press

KENDALL, M.G., 1973. 'Time-Series' Griffin

BOX, G.E.P. and JENKINS, G.M., 1970. 'Time-Series Analysis, Forecasting and Control' McGraw-Hill

CHATFIELD, C., 1980. 'The Analysis of Time Series : An Introduction' Chapman and Hall

PRIESTLY, M.B., 1971. Bull. Int. Statist. Inst., 44, Book 1, 295-324

PRIESTLY, M.B., 1978. The Statistician, 27, 159-176

PONTRYAGIN, L.J., 1959. USP Mat. Nauk 14, 3 (1959) 'Optimal Control Processes'

PONTRYAGIN, L.J., et al, 1960. 'The Maximum Principle' Izvestiya Akad. Nauk SSR, Ser, Mat. 24, 3(1960)

BELLMAN, R.E., 1959. 'Dynamic Programming' Princeton University Press

BELLMAN, R.E., 1962. 'Applied Dynamic Programming' Princeton Univeristy Press

BOX, G.E.P. and McGREGOR, J.F., 1973. Technometrics, Vol 16, No.3, Aug.1974

ASTROM, K.J. and EYKHOFF, P., 1971 Automatica Vol 7 pp 123-162

GUSTAVSSON, I., et al, 1977. Automatica Vol 13 pp 59-75

AKAIKE, H., et al, 1972. Automatica Vol 8 35-48

AKAITE, H., 1984 (Oct). Paper presented at workshop on Time Series and its Applications. Nat. University pf Singapore

McLEOD, G., 1982. 'Box Jenkins in Practice' G Jenkisn and Partners Ltd

GROSS, D. and HARRIS L.M., 1974. 'Fundamentals of Governing Theory' Wiley

WIENER, N., 1949. 'The Extrapolation, Interpolation, and Smoothing of Stationary Time Series' M.I.T.

McDONALD, D., 1950. Proc. National Electronics Conference, Vol 6 pp 400-421

FOLEY, R.D. and DISNEY, R.L. 1983. Adv. Appl. Prob 15, 162-182

KULKARNI, V.G., 1983. J. Appl. Prob. Vol 20, No 2, pp 380-389

TAKACS, L., 1962. The Bell System Technical Journal March 1963 pp 505-519

DISNEY, R.L. and D'AVIGNON, G.R., 1977. Management Science Vol 24 No. 2 pp 168-180

DISNEY, R.L. and HANNIBALSSON, I., 1977. Naval Res/ Logit. Quart. Vol. 24 pp 281-291

DISNEY, R.L., et al, 1980. Naval Res. Logist. Quart Vol. 27, pp 635-644

DISNEY, R.L., 1981. Naval Res. Logist. Quart Vol. 28, pp 679-684

DISNEY, R.L., et al. 1984. Adv. Appl. Prob. 16, pp 437-446

ALI, O.M.E. and NEUTS, M.F.. J. Appl. Prob. Vol 21, pp 404-413

HUNTER, J.J., 1983. Adv. Appl. Prob Vol 15 pp 349-375

HUNTER, J.J., 1984. Adv. Appl. Prob Vol 16 pp 422-436

CHAPTER 4

DISNEY, R.L. and KONIG, D., 1985. Siam Review Vol. 27 No. 3 pp 335-403

McNICKLE, D.C., 1984. NZOR Vol. 12 No. 2 pp 109-118

BARNETT, M.S., 1975. J. of Hydrology Vol. 26 pp 95-113

ANIS, A.A., 1970. J. of Hydrology Vol. 11 pp 363-378

GANI, J., 1969. Adv. Appl. Prob. Vol. 1 pp 90-110

BARNETT, N.S., ans PHATAFOD, R.M., 1973. J. of Hydrology Vo. 19 pp 171-184

PHATAFOD, R.M., 1975. J. of Hydrology Vol. 30 pp 199-217

REIMAN, M.I., 1982. App. Prob. - Computer Science : The Interface, Vol II pp 409-421

LA TOUCHE, G., 1979. Tech. Report 86, Université Libre de Bruxelles

JACKSON, J.R., 1957. Oper. Res., 5, pp 518-521

JACKSON, J.R., 1963. Management Science, 10, pp 131-142

ENGSET, T., 1918. E.T.Z., 31, PP 304-305 (In German)

ERLAND, A.K., 1917. Post Office Electrical Engineers' Journal, 10, pp 189-197

CHAPTER 5

MODER, J.J. and ELMAGHRABY, S.E., 1978. 'Handbook of Operations Research' Von Nostrand

FISHMAN, G.S., 1978. 'Principles of Discrete Event Simulation' Wiley

LAW, A.M., 1983. Operations Research, Vol. 31, No. 6, pp 983-1029.

MARINO, M.A., et al, 1983. Water Resources Research, Vol. 19, No. 3, pp 613-620

MARINO, M.A., et al, 1985. Water Resources Research, Vol. 21, No. 5, pp 619-630

BOSHIER, J.F., and LERMIT, R.J., 1977. NZOR Vol. 5, No. 2, pp 85-100

BOSHIER, J.F., et al, 1983. Trans., Vol. 10, No. 2/EMCH

WATERMEYER, P. and THOMAS, H.A., 1962. 'Queueing Theory and Reservoir Design'. The Annals of the Computation Laboratory of Harvard University, Vol. 31.

TAVARES, L.V., 1982. J. of Hydrology, Vol. 58, pp 237-250

BATHER, J.A., 1968. J. Appl. Prob. Vol. 5, pp 55-71

KLEMES, V., 1979. Water Resources Research, Vol. 15, No 2, pp 359-370

RIPPL, W., 1883. Minutes Proc. Inst. Civil. Eng., LXXI, 270-278

VARLET, H., 1923. See Klemes, V., 1979

MALTERRE, 1912. See Tavares, L.V., 1982

McMAHON, G.F., 1980. Journal of the Water Resources Planning and Management Division WRI March 1980 pp 239-254

FRIEDMAN, L.W. and FRIEDMAN, H.H., 1984. J. Statist. Comput. Simul., Vol. 19, pp 237-263

CRAMER, H., 1964. Technometrics, Vol 6, No 2, pp 133-158

CONWAY, R.W., 1963. Management Science, Vol. 10, No. 1, pp 47-61

EHRENFELD, S. and BEN-TUVIA, S., 1962. Technometrics, Vol. 4, No. 2, pp 257-275

PEREIRA, M.V.F. and PINTO, L.M.V.G., 1985. Water Resources Research, Vol. 21, No. 6, pp 779-792.

CRYGIER, J.C. and STEDINGER, J.R., 1985. Water Resources Research, Vol. 21, No. 1, pp 1-10.

STEDINGER, J.R. and SULE, B.F., 1983. Water Resources Research, Vol. 19, No. 6, pp 1383-1393.

ASKEW, A. et al, 1971. Water Resources Research, Vol. 7, No. 4, pp 819-826

LOUCKES, D.P., 1968. J. Sanit. Eng. Div. Amer. Soc. Civil Eng., 94 (SA4) pp 657-669

Other references in Chapter 5 are found in MODER, J.J. and ELMAGHRABY, S.E., 1978, Part II Chapter 6, pp 571-578.