Australian Rainfall & Runoff

DISCUSSION PAPER: MONTE CARLO SIMULATION TECHNIQUES

FINAL REPORT

AR&R D2

MAY 2013

ENGINEERS AUSTRALIA
Water Engineering

ENGINEERS
AUSTRALIA
Water Engineering

Engineers Australia
Engineering House
11 National Circuit
Barton ACT 2600

Tel: (02) 6270 6528
Fax: (02) 6273 2358
Email:arr_admin@arr.org.au
Web: www.engineersaustralia.org.au

AUSTRALIAN RAINFALL AND RUNOFF
DISCUSSION PAPER: MONTE CARLO SIMULATION TECHNIQUES

FINAL

MAY, 2013

Discussion Paper Topic Monte Carlo Simulation Techniques	**AR&R Discussion Paper Number** AR&R D2
Date 6 May 2013	**ISBN**
Authors Rory Nathan Erwin Weinmann	**Verified by**

ACKNOWLEDGEMENTS

This discussion paper is the result of a significant amount of in kind hours provided by Engineers Australia Members. This discussion paper was prepared at the request of the AR&R revision team.

ENGINEERS
AUSTRALIA
Water Engineering

FOREWORD

AR&R Revision Process

Since its first publication in 1958, Australian Rainfall and Runoff (AR&R) has remained one of the most influential and widely used guidelines published by Engineers Australia (EA). The current edition, published in 1987, retained the same level of national and international acclaim as its predecessors.

With nationwide applicability, balancing the varied climates of Australia, the information and the approaches presented in Australian Rainfall and Runoff are essential for policy decisions and projects involving:

- Infrastructure such as roads, rail, airports, bridges, dams, stormwater and sewer systems;
- Town planning;
- Mining;
- Developing flood management plans for urban and rural communities;
- Flood warnings and flood emergency management;
- Operation of regulated river systems; and
- Estimation of extreme flood levels.

One of the major responsibilities of the National Committee on Water Engineering of Engineers Australia is the periodic revision of A&RR. As part of the revision process, there is a need to consult with the profession to provide early advice on likely changes to AR&R and to receive feedback on these potential changes.

For this reason, the AR&R Revision Team will publish Discussion Papers outlining concepts and methods that are being considered for inclusion in AR&R. Feedback from the profession on these concepts and methods will form an essential aspect of the final decision regarding their inclusion or exclusion in AR&R.

This Discussion Paper deals with Monte Carlo approaches to design event simulation. Outlined in the paper are details of both the concept and available information for its implementation. Discussion on both aspects of the paper is welcomed and encouraged.

Mark Babister
Chair Technical Committee for
ARR Research Projects

Assoc Prof James Ball
ARR Editor

PROJECT TEAM

- Rory Nathan, SKM
- Erwin Weinmann

EXECUTIVE SUMMARY

Current practice for estimation of design floods is typically based on the “design event” approach, in which all parameters other than rainfall are input as fixed, single values. Considerable effort is made to ensure that the single values of the adopted parameters are selected with the objective of ensuring that the resulting flood has the same annual exceedance probability as its causative rainfall.

Monte Carlo simulation offers an alternative to the design event method. This approach recognises that any design flood characteristics (e.g. peakflow) could result from a variety of combinations of flood producing factors, rather than from a single combination. The approach mimics “mother nature” in that the influence of all probability distributed inputs are explicitly considered, thereby providing a more realistic representation of the flood generation processes.

This report describes the practical implementation of Monte-Carlo techniques for flood estimation. The discussion focuses on the manner in which the current (ARR) guidelines and available design information can be used to take into account the natural variability of the inputs, and presents the concepts in a manner that can be implemented in a spreadsheet. While it would be possible to directly employ these techniques with existing models, this guidance should also be found helpful to those using Monte Carlo frameworks that are available in the public domain.

TABLE OF CONTENTS

1. Introduction

1.1. Hydrologic Applications of Monte Carlo Simulation

Monte Carlo simulation is a technique that uses a large number of random samples to find solutions to physical problems that cannot otherwise be easily solved. The first systematic development of Monte Carlo methods derives from work on the atomic bomb during the second world war, though a number of isolated and undeveloped instances have been reported (Hammersley and Handscomb, 1964). The name was coined to capture the random properties of the roulette wheel as most famously played in the city of the same name in Monaco.

In modelling hydrologic systems, there are two sources of random variation that may affect the estimated system outputs: (i) the *natural temporal and spatial variability* of climate and catchment factors being modelled, and (ii) the random variation resulting from unavoidable *uncertainty* in the definition of the model structure, the model inputs and in the estimated model parameters. Monte Carlo simulation has been widely used to determine the impacts of model and parameter uncertainty on simulation results; these are generally expressed in the form of confidence limits on hydrologic estimates (Section 6).

Monte Carlo simulation of natural variability of climate and catchment factors has also been used to solve hydrological problems for many years. An early hydrological example occurred before the advent of computers and involved the use of a deck of cards to simulate the stochastic processes of the Nile floodwaters for estimating yield (Hurst et al, 1965, as reported by McMahon and Mein, 1986). The approach is ideally suited to the solution of many hydrological problems which involve both a stochastic and a deterministic component in the manner described by Laurenson (1974). In the context of flood estimation the *stochastic component* comprises the (notionally random) occurrence of rainfall, its temporal and spatial distribution, and antecedent conditions such as losses and initial reservoir level. These random factors are represented in the simulation by probabilistic sampling of model inputs (and possibly parameter values) from their respective distributions. The *deterministic component* is the transformation (via an appropriate hydrologic model) of rainfall into rainfall excess, and its subsequent routing through the catchment, to yield a catchment outflow hydrograph. This transformation is deterministic in the sense that, for a given set of conditions and events of similar magnitude, the catchment response can be assumed to remain essentially unchanged. The deterministic catchment model will thus always predict the same outflow hydrograph for a given set of rainfall inputs, antecedent conditions, and parameter values.

1.2. Monte Carlo Simulation vs. ‘Design Event’ Approach

In essence, Monte Carlo methods provide a framework for simulating “mother nature”, where all important flood producing factors can be treated as stochastic variables, and the less important ones can be fixed. Most traditional rainfall-based flood estimation techniques are based on the *design event* approach, in which all parameter values and inputs other than rainfall are treated as fixed values. The application of these traditional methods generally involves the implicit assumption that the annual exceedance probability (AEP) of the flood is the same as its causative rainfall. To satisfy this *probability-neutral* assumption, it is necessary to select model inputs carefully to ensure that no probability bias is introduced in the transformation of rainfall to the flood characteristics of interest. While over the years a considerable body of experience has evolved to guide the selection of fixed parameter values, their adoption without the benefit of comparison with independent design flood estimates necessarily involves a degree of faith. When tested more rigorously it is usually found necessary to vary fixed inputs such as losses (and other factors) with the severity of the event to ensure that probability-neutrality is maintained. Kuczera et al (2003) provide a number of examples that highlight the deficiencies of the design event approach.

The design event approach also assumes that there is a *critical rainfall duration* that produces the design flood for a given catchment. This critical duration depends on the interplay of catchment and rainfall characteristics; it is not known *a priori* but is usually determined by trialling a number rainfall durations and then selecting the one that produces the highest flood peak (or volume) for the specific design situation.

In the most general Monte Carlo simulation approach for design flood estimation, *rainfall events of different duration* are sampled stochastically from their distribution. The simulated design floods are then weighted in accordance with the observed frequency of occurrence of rainfall events of different durations that produced them. This avoids any positive bias of estimated flood probabilities which may be associated with the application of the critical rainfall duration concept (Weinmann et al., 2002).

The application of this generalised approach relies on the derivation of new design data for rainfall events that are consistent with a new probabilistic definition of storm ‘cores’ or complete storms (Hoang et al., 1999). Such new design rainfall data is currently not widely available, thus limiting the application of the generalised approach. Accordingly the description in Section 2 of the proposed generic Monte Carlo simulation approach for design flood estimation is still based on the concept of critical rainfall duration. This aspect is further discussed in Section 4.4.

1.3. Practical Considerations

One of the main attractions of Monte Carlo methods is that the modelling tools and hydrologic concepts involved are essentially identical to those used in traditional flood estimation approaches. Differences only arise in the manner in which the inputs are handled and the results analysed. Thus, whether the adopted model of choice is based on a simple runoff-coefficient, a loss-model combined with a unit hydrograph, or a more complex runoff-routing approach, the model can be applied in a Monte Carlo framework to better capture the natural variability of the flood producing and flood modifying factors.

Adapting models to run in a stochastic environment allows the practitioner to take advantage of existing design information and experience. If industry-accepted models are used, the vast body of experience and empirical relationships available concerning model configuration and parameterisation are still largely applicable. Less important parameters can be fixed at values deemed appropriate from previous experience and available design guidance, and average values of stochastically varying inputs can be expected to fall within a range previously considered for use in deterministic design.

The information available for specifying input distributions is often the same as that used to derive typical or average values of fixed inputs, and many inputs can easily be non-dimensionalised and pooled from regional observations. For example it is already common to represent temporal patterns in a dimensionless form (as provided in Section 2, Book II), and suitable empirical samples can easily be extracted from pluviograph data. Of course some thought and care is required to ensure that the distributions being used are relevant to the design problem of interest, and that correlations between inputs are handled appropriately, but the concepts involved in the preparation and treatment of design data are very similar to those used in traditional design practice.

1.4. Scope of Application

For many practical problems Monte Carlo simulation is well suited to deriving unbiased estimates of design flood characteristics. Allowing the important flood producing factors to vary in accordance with the variation found in nature obviates the need to identify fixed, probability-neutral, design inputs. As long as the governing distributions are based on processes relevant to the design events of interest, and they are adequately characterised and sampled, it is possible to substantially reduce the degree of uncertainty and bias involved in estimating the required probabilities of exceedance.

Monte Carlo simulation is particularly suitable in cases where design flood characteristics need to be determined at multiple locations within a system. Using the design event approach with fixed input values, the probability-neutral transformation from design rainfall to design flood would require a separate model run for each location of interest, each run using the appropriately selected input value for that location. With Monte Carlo simulation, the sampling of input values over a wide range ensures that the changing influence of different factors is automatically allowed for when moving from sites of interest in the upper catchment to sites near the catchment outlet.

There are, however, a number of problems in which the design objective is more easily accommodated by continuous simulation approaches. Predominately these involve systems which posses multiple flood controlling factors that are dependent on antecedent conditions and/or other correlated variables.

The following discussion of Monte Carlo techniques is narrowly focused on those analyses that are of most relevance to design flood estimation. A general and very accessible introduction to Monte Carlo methods can be found in Burgman (2005), and more comprehensive and practical guidance is provided in Vose (2000) and Saucier (2000); the latter reference includes C++ source code for a collection of various distributions of random numbers suitable for performing Monte Carlo simulations. Hammersley and Handscomb (1964) provide a more advanced theoretical treatment of the subject, and useful discussion on the advantages of using Monte Carlo methods to estimate design floods can be found in Weinmann et al (2002), Kuczera et al (2003), and Weinmann and Nathan (2004).

2. Overview of the Monte Carlo Simulation Process

The generic steps involved in undertaking a Monte Carlo simulation for design flood estimation are illustrated by the flowchart in Figure 1. In essence, there are three preparatory steps, followed by the actual Monte Carlo simulation and the construction of the derived flood frequency curve. In the following, these steps are briefly described.

(i) Selecting an appropriate flood event simulation model

The criteria for selecting an appropriate model are similar to those used with the traditional design event approach and are described in Chapters 4 and 6 of Book V, and Chapter 5 of Book VI. The selected model should be able to be run in batch mode with pre-prepared input files or be called from the Monte Carlo simulation application. Models with fast execution speeds are well suited to Monte-Carlo simulation; complex models with slow run-times can still be utilised, though generally they need to be invoked within a stratified sampling scheme (Section 4.3) to ensure that the simulations times are within practical constraints.

(ii) Identifying the model inputs and parameters to be stochastically generated

The stochastic representation of model inputs should focus on those inputs and parameters which are characterised by a high degree of natural variability and a non-linear flood response. Examples include rainfall temporal pattern, initial loss and reservoir storage content at the onset of a storm event. If the assessment indicates limited variability and essentially linear system response, then there may be little to be gained from applying a Monte Carlo simulation approach.

(iii) Defining the variation of inputs/parameters by appropriate distributions and correlations

The considerations and methods applied in this step are presented in detail in Section 5. The distributions used to generate the stochastic inputs can be defined by the use of specific probability distributions or else an empirical, non-parametric approach can be adopted (Section 3.3). Schaeffer (2002, 2004) adopts a strongly parametric approach to sampling a wide range of storm and catchment processes, and Rahman et al (2002[a,b]) provides examples in which both losses and temporal patterns are defined using a Beta distribution. Nathan et al (2003) and Nathan and Weinmann (2004) adopt a more empirical approach that is strongly based on the nature of design information used in the traditional design event method.

(iv) Monte Carlo simulation of flood events

The model is run *N* times where at each simulation step a set of *n* inputs are stochastically generated and the results recorded. Only those inputs that have a significant influence on the results need to be stochastically generated, and other inputs can be treated as fixed (usually average or median) values. Generally many thousands of simulations are required to adequately sample the inherent variability in the system, and thus for most practical problems some thought is required to minimise disc storage space and simulation times.

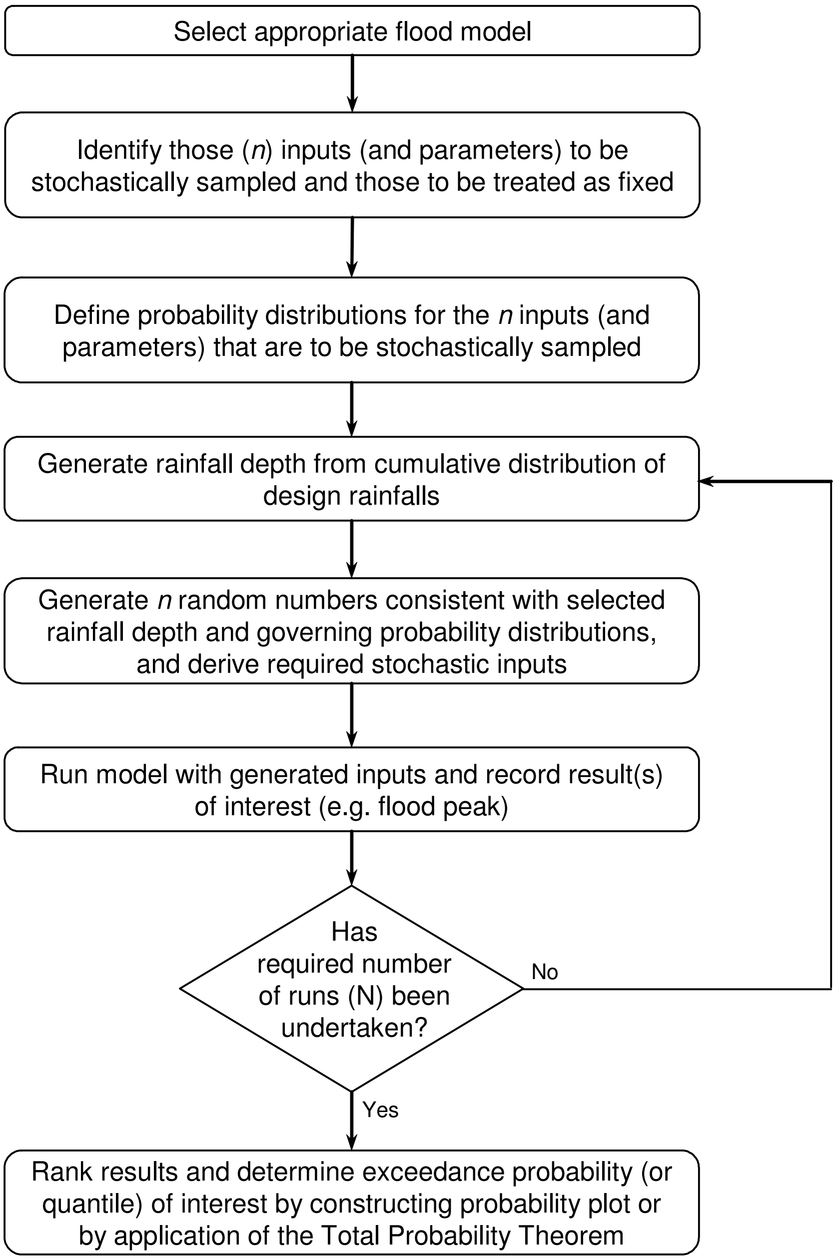


Figure 1 Flow chart of generic Monte Carlo simulation process..

(v) Construction of derived flood frequency curve

Once the required number of runs have been undertaken, it is necessary to analyse the results to derive the required exceedance probabilities. Where very simple models are used or the probabilities of interest are not extreme – more frequent than, say, annual exceedance probabilities (AEPs) of 1 in 100 – the results can be analysed directly by construction of a derived frequency curve (Section 4.2). Alternatively, in order to estimate rarer exceedance probabilities (or use more complex models with slow execution speeds) it will be necessary to adopt a stratified sampling approach, as described in Section 4.3.

Simple problems such as those involving the sizing of storage ponds or flood storage of tailings dams (Section 7.6, Book VI) can usually be solved using commercially available spreadsheet software. A number of proprietary "add-ins" that provide Monte-Carlo functionality are available for use with some spreadsheet programs, but while this added functionality certainly makes it easier to develop a solution, they are by no means necessary. Simple problems can often be solved by merely invoking cell formulae, but for more complex problems it is usually desirable to write customised “macro” functions to more easily describe the hydrological processes being simulated.

Direct sampling schemes (Section 4.2) can be developed with a minimum of programming skill by coupling command batch-file routines that run any flood event model with pre-prepared input files. Those with more advanced programming skills can develop schemes that invoke flood event models without the need for reading and writing computer disc files; the attraction of this is that simulation times are greatly reduced, which is a great benefit when undertaking more numerically-intensive simulations.

Section 5 provides more details on some of these steps in the context of design flood estimation, and Section 7 gives worked examples for a number of typical applications.

3. Stochastic Sampling

3.1. Inverse Transformation Approach

The method used to stochastically sample from the input distributions is the core algorithm used in Monte Carlo simulation. Once a suitable framework has been established additional model inputs and/or parameter values can be added to the sampling procedure as required.

The following two sections describe the *inverse transformation* approach that can be applied to either formally defined probability density functions, or else empirical "data-driven" distributions. The inverse transform approach is a simple and efficient technique for generating stochastic samples though other approaches (suited to the use of probability density functions that cannot be inverted) are described by Saucier (2000).

The basis of the inverse transformation approach is to generate the required probability density function $f(x)$ through uniform sampling of the inverse of the cumulative distribution function $F(x)$[1]. The two-step process for doing this is illustrated in Figure 2, and the algorithm can be summarised as follows:

1. Generate a random number (U) uniformly distributed between 0 and 1
2. Calculate the value (x) of the inverse of the cumulative density function $F^{-1}(U)$

This process is illustrated in Figure 2 for three random numbers. The first random number generates a value near the tail of the distribution, and the next two yield values that are more centrally tended. For illustration purposes the input random numbers (U) in Figure 2 are shown as being equally spaced, but on exit the transformed numbers are unequally spaced, in conformance with the adopted distribution. Inverse functions of a number of useful distributions (Normal, log-Normal, Beta, Gamma) are provided in standard spreadsheet software (see example in Section 7.1), and source codes for these and others (in Fortran, C++, and Pascal) are freely available (Saucier, 2000; Griffiths et al, 1985; Press et al, 1989, 1993; and Vetterling et al, 1993). If an empirical distribution is used then values can be interpolated from a look-up table comprised of values of the cumulative density function (Section 3.3d).

[1] $F(x)$ gives the probability P of x being less than a specified value, while the inverse of the cumulative density function $F^{-1}(P)$ allows direct determination of the value of x that corresponds to a given cumulative probability P.

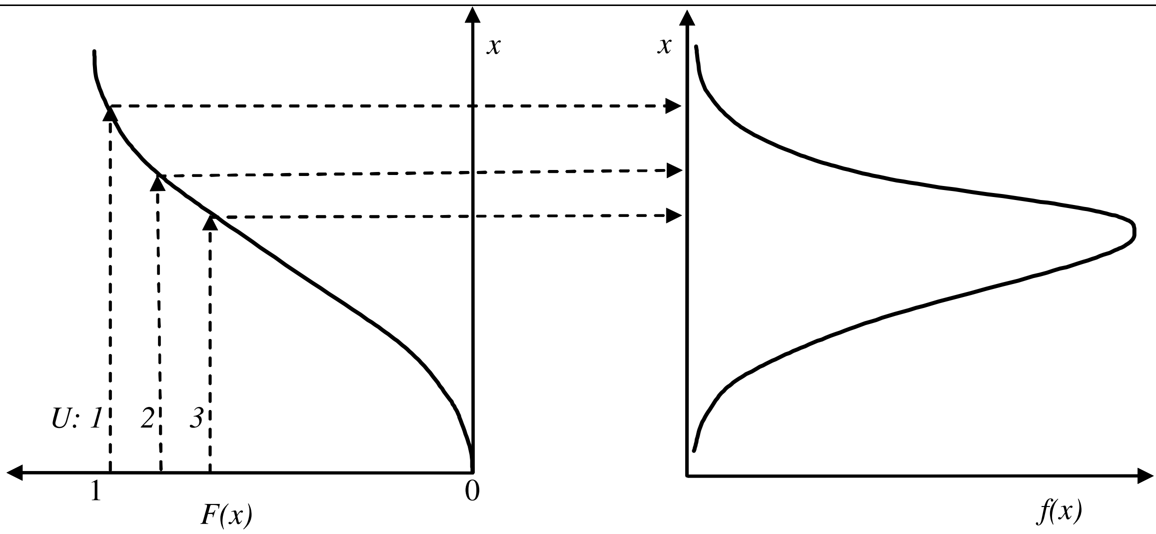


Figure 2 Illustration of the inverse transform method.

3.2. Random Number Generation

Any application of Monte Carlo analysis requires the generation of random numbers. Random numbers are required as the first step in application of the inverse transformation approach (as described in the preceding section), and in any scheme in which discrete inputs are sampled in random manner. Many algorithms have been developed for the generation of random numbers (or rather, pseudo-random numbers, as all algorithms exhibit non-randomness to varying levels of acceptability). Many software compilers provide random number generators as intrinsic functions, and a variety of source codes is available in the public domain (Press et al, 1993; Saucier, 2000; Griffiths et al, 1985). Non-programmers can generate a suitable sample of random numbers using commonly available spreadsheet programs, and if necessary these can be exported as a text file for use in another application.

3.3. Probability Density Functions

There is a large number of statistical distributions that can be used to represent variability in different types of hydrological processes and input uncertainty. The characteristics of a few distributions found to be useful in hydrology are presented below, but information on other distributions of potential use can be found in Saucier (2000), Vose (2000), and Maidment (1993).

(a) Uniform Distribution

This is the most simple of distributions but, as discussed in the preceding section, it is used to generate random numbers between 0 and 1 in the first step of the inverse transformation approach. In its discrete form it is also useful for randomly selecting inputs such as temporal and spatial patterns of rainfall.

The algorithm to generate a uniform random number between the limits x_{min} and x_{max} is:

1. Generate U = U(0,1)
2. Return $X = x_{min} + (x_{max}-x_{min})U$

where *U(0,1)* denotes a random number uniformly distributed between 0 and 1.

If a random discrete integer is required between the integer limits i_{min} and i_{max}, the algorithm is slightly modified as follows:

1. Generate $U = U(0,1)$
2. Return $X = i_{min} + ROUND\{(i_{max}-i_{min})U\}$

   where *ROUND{}* is a function that rounds a rational number to the nearest integer.

(b) Normal Distribution

The normal distribution is often used to describe the distribution of stochastic processes and thus it is extensively incorporated into Monte Carlo simulation schemes. Many stochastic processes in hydrology conform to the log-Normal distribution (that is they only take positive values and are skewed towards higher values),  and transforming the data beforehand into the logarithmic domain is a simple means of taking direct advantage of the Normal distribution. Another common use (described in Section 3.5 and provided as a worked example in Section 7.1) is to use the Normal distribution to generate errors associated with unexplained variance in regression-based generation of data. The *sum* of many random processes generally conforms to the Normal distribution (regardless of the distribution of individual processes) whereas the *product* of many random processes conforms to the log-Normal distribution. Details of the Normal distribution are provided in all statistics textbooks and thus will not be presented here. Source code for estimation of the cumulative Normal distribution is freely available (e.g. Press et al, 1993) and is available in spreadsheet software.

It is worth noting that, while few hydrologic variables conform to the normal distribution, many data sets can be transformed into the Normal domain by the Box-Cox transformation (Box-Cox, 1964); with this approach, a variate X can be transformed into the normal domain (Z) by the following equation:

$$Z = \frac{X^{\lambda} - 1}{\lambda} \qquad \text{Equation 1}$$

where λ is a parameter determined by trial and error to ensure that the skewness of the transformed distribution is zero. A noteworthy special case of this transformation arises when λ is set to zero, then the transformation is equivalent to taking logarithms of the data. Fitting the parameter λ is most easily achieved by optimisation or the use of "solver" routines that are commonly available in spreadsheet programs.

Illustration of the efficacy of the Box-Cox transformation is provided in Figure 3. Panel (a) of this figure is a "quantile-quantile" (Q-Q) plot that shows the distribution of annual streamflows with a skew coefficient of 1.1. A Q-Q plot is constructed by plotting the expected cumulative distribution (in this case the Normal) versus the observed distribution of data points, and the closer the points lie along the one-to-one line the more defensible the assumption of normality. After application of the Box-Cox transform (using a λ value of 0.21) the resulting distribution is found to be normally distributed, as illustrated in Figure 3(b). Thus, if it was desired to stochastically generate the annual streamflows, then this could be achieved by first applying the Box-Cox transform, then stochastically generating values using the Normal distribution, and finally applying the inverse transform to the results.

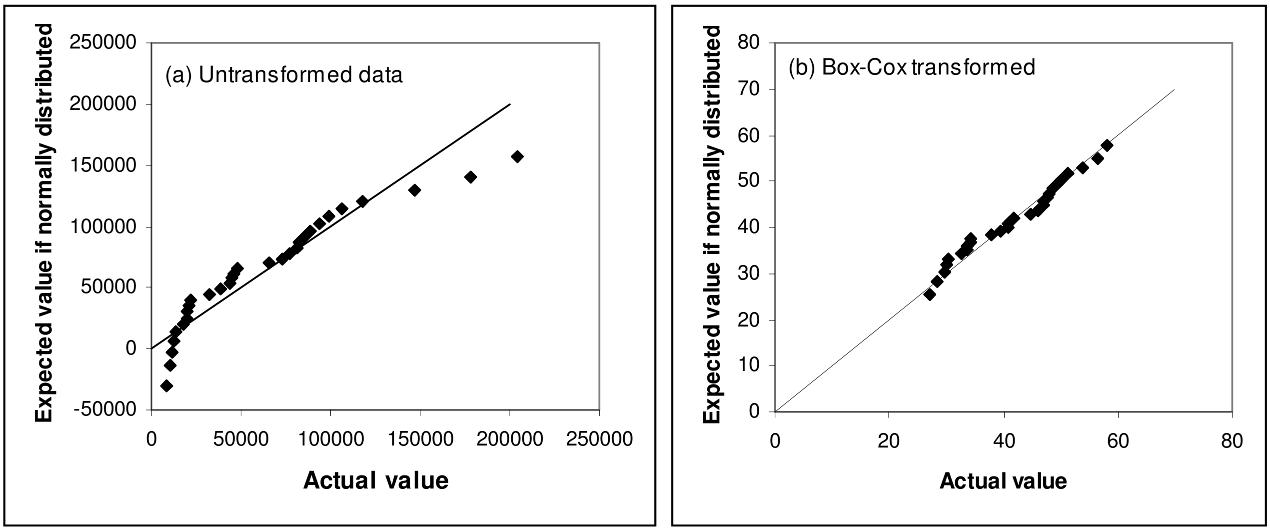


Figure 3 Illustration of application of the Box-Cox transform to achieve normality.

(c) Triangular Distribution

The triangular distribution has lower and upper limits (x_{min}, x_{max}), and a mode (c) that can be specified to yield either a symmetric or skewed distribution. It has no theoretical basis but easily accommodates speculative judgement involving estimation of the upper and lower bounds and the most likely value of a distribution. While its simplicity makes it a popular choice, its distributional properties may not be well suited many practical situations. It is a "thick-tailed" distribution which provides a moderately high chance that values will be generated from a region near the maximum and minimum limits. In addition, if a skewed distribution is selected, then the mean may be quite different from the mode; this is because the mean, which is determined simply as the arithmetic average of the three parameters ($=\{x_{min}+x_{max}+c\}/3$) is heavily influenced by the limiting value of the longest tail (either x_{min} or x_{max}), and this may be quite distant from the value of the mode (c).

The algorithm to generate a triangular random variate between the limits x_{min} and x_{max} (with a mode of c) is:

1. Generate $U = U(0,1)$
2. Return $X = \begin{cases} x_{min} + \sqrt{(x_{max} - x_{min})(c - x_{min})U} & \text{if } U \le (c - x_{min})/(x_{max} - x_{min}) \\ x_{max} - \sqrt{(x_{max} - x_{min})(x_{max} - c)(1 - U)} & \text{if } U > (c - x_{min})/(x_{max} - x_{min}) \end{cases}$

(c) Beta Distribution

The Beta distribution has the same pragmatic advantages of the triangular distribution but is similarly devoid of theoretical justification. In addition to capturing information about the minimum and maximum limits and the mode, the Beta distribution is able to accommodate a wide variety of shapes. The distribution has four parameters: two positive shape parameters (α and β) and an upper and lower bound (x_{min} and x_{max}). If the two shape parameters have the same value then the distribution is symmetric. Adoption of different values of the shape parameters yields a skewed distribution, where the greater the difference between them the more skewed the distribution. Higher values of shape parameters induces thinner tails and a greater central tendency; lower values yields thicker tails, to the point where adoption of $\alpha=\beta=1$ results in a uniform distribution. (In the extreme the distribution can yield a bimodal distribution where all the weight is in the tails – similar to the flattened letter "U" – but this has no known practical application in hydrology and thus is not shown). An illustration of the range of shapes achievable with different combinations of shape factors is presented in Figure 4.

The Beta distribution is most easily incorporated into Monte Carlo simulation using the inverse transformation approach (Section 3.1), and source code for estimation of the cumulative distribution is freely available (e.g. Press et al, 1993). The function can also be accessed using commonly available spreadsheet software and, if desired, this can be used to generate a look-up table and treated as an empirical distribution, as discussed below.

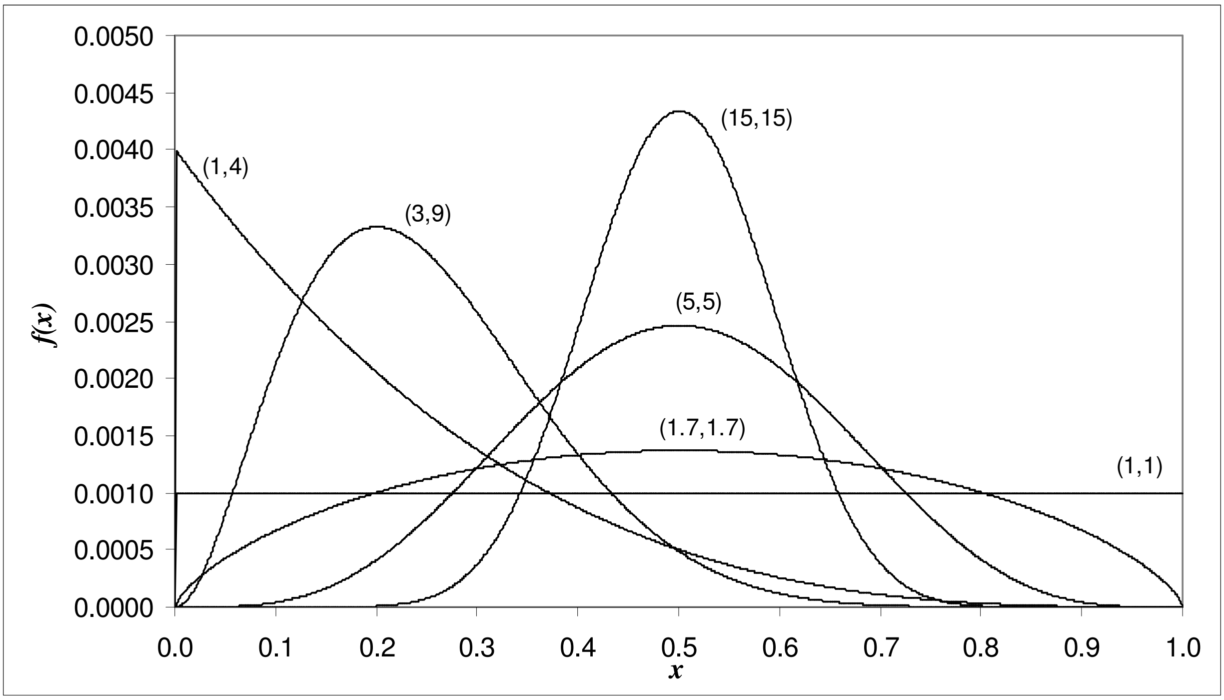


Figure 4 Illustration of various forms of the Beta distribution for the case where x_{min} and x_{max} are 0.0 and 1.0, respectively (figures in brackets indicate parameters α and β).

(d) Empirical Distributions

One very practical way of undertaking a Monte Carlo simulation is to sample from a given set of data. This is a fast and simple technique that can be used to take advantage of empirical data sets (such as losses and reservoir drawdown) in a more defensible manner than simple adoption of a single best estimate or representative value. It is also useful for sampling from "pragmatic" distributions, such as rainfall frequency curves based on application of the shape factor approach (Section 3.6.2, Book VI), which are not based on a theoretical distribution function.

The algorithm to construct and sample from an empirical distribution is as follows:

1. Sort empirical data into either ascending or descending order as appropriate, and assign a cumulative probability value to each. If there are *n* data values, then the largest data value (x_1) is assigned an exceedance probability $F(x_1)$, the second largest (x_2) is assigned an exceedance probability $F(x_2)$, and so on till the last value, represented by x_n and $F(x_n)$.

2. Generate $U = U(0,1)$
3. Locate interval i such that $F(x_i) \leq U < F(x_{i+1})$
4. Return $X = x_i + \dfrac{U - F(x_i)}{F(x_{i+1}) - F(x_i)}(x_{i+1} - x_i)$
5. Generate additional points by returning to step 2.

While simple to implement, the use of empirical distributions in Monte Carlo simulation does require care. Most importantly, it is necessary to ensure that the data sample being used is relevant to the whole range of conditions being simulated. For example if the data set is comprised of initial reservoir levels recorded over a short historic period, then these may not be relevant to the assessment of extreme flood risks under a different set of operating rules.

It is seen in step 4 of the above algorithm that values within each interval are obtained by linear interpolation. This is normally quite acceptable, though obviously the less linear the relationship between the data values and their corresponding exceedance probabilities the less defensible is such an approach. Accordingly, in some cases it is best to first transform the data and/or the exceedance probabilities assembled for step 1 of the algorithm. Many hydrological variables are approximately log-Normally distributed, and thus it is often desirable to undertake the interpolation in the log-Normal domain. To this end, the ranked data values are transformed into logarithms (it does not matter what base is used) and the exceedance probabilities are converted to a standard normal deviate (that is, the inverse of the standard normal cumulative distribution). Step 2 of the above algorithm would thus need to be replaced with $U = U(z_{min}, z_{max})$ where z_{min} and z_{max} represent the standard normal deviates corresponding to $F(x_1)$ and $F(x_n)$, ie the adopted limits of exceedance probability range.

Care is also required when sampling from the tails of the distribution. Empirical data sets are of finite size and, if the generated data are to fall between the upper and lower limits of the observed data, the cumulative exceedance probability of the first ranked value $F(x_1)$ should be zero, and that of the last ranked value $F(x_1)$ should be 1.0. Thus use of empirical data sets is appropriate for those inputs whose extremes of behaviour are not of great relevance to the output. Losses, for example, are zero bounded, and thus the difference in flood peak between a loss exceeded 95% of the time and that exceeded 99.999% of the time may well be of no practical significance. However, if an empirical approach is being used for the generation of rainfalls that are defined for exceedance probabilities ranging between 1 in 2 and 1 in 100, then it is inevitable that more than half the random numbers generated in step 2 of the above algorithm can be expected to lie outside the specified range of rainfalls. As long as the probability range of interest lies well within the limits specified, then rainfall values can be obtained by some form of appropriate extrapolation; however, if this approach is used then checks should be undertaken to ensure that the extrapolated values do not influence the results of interest.

3.4. Generating correlated variates

Many hydrologic variables are correlated and thus it is sometimes necessary to ensure that the adopted sampling scheme preserves the correlation structure of the inputs. A simple means of generating correlated variables is described by Saucier (2000). The approach is based on rotational transformation and the steps involved in generation of uniformly distributed variates can be stated as follows:

1. Independently generate two uniform random variates, $X = U(-1,1)$ and $Z = U(-1,1)$
2. Set $Y = \rho X + Z\sqrt{1-\rho^2}$ where □ is the required correlation between X and Z
3. Return: $x = (x_{min}+x_{max})/2 + X(x_{max}-x_{min})/2$

   $y = (y_{min}+y_{max})/2 + Y(y_{max}-y_{min})/2$

   where x_{min} and x_{max} are the lower and upper bounds of the first variate and y_{min} and y_{max} are the corresponding bounds of the other.

Application of the above algorithm is illustrated in Figure 5(a). The bounds along the x-axis are 5 and 130, and those along the y-axis (for the mid-point of the x distribution) are 30 and 75. The figure illustrates the results for the generation of 2000 correlated variates where the correlation coefficient (ρ) adopted is -0.7.

The above algorithm can easily be adapted to the generation of correlated variates that conform to some specified distribution. For the Normal distribution, the required algorithm is:

1. Independently generate two normal random variates with a mean of zero and a standard deviation of 1: $X = N(0,1)$ and $Z = N(0,1)$
2. Set $Y = \rho X + Z\sqrt{1-\rho^2}$ where ρ is the required correlation between X and Z
3. Return: $x = \mu_x + X.\sigma_x$

   $y = \mu_y + Y.\sigma_y$

   where μ_x and μ_y are the means of the two distributions and σ_x and σ_y are the required standard deviations.

Application of the above algorithm is illustrated in Figure 5(b). The input parameters to this example are $\rho=-0.7$, $\mu_x=70$ and $\sigma_x=10$, and $\mu_y=50$ and $\sigma_y=10$, and as before a total of 2000 correlated variates are generated. Any distribution could be used in lieu of the Normal distribution, or else the variates of interest could be transformed into the normal domain.

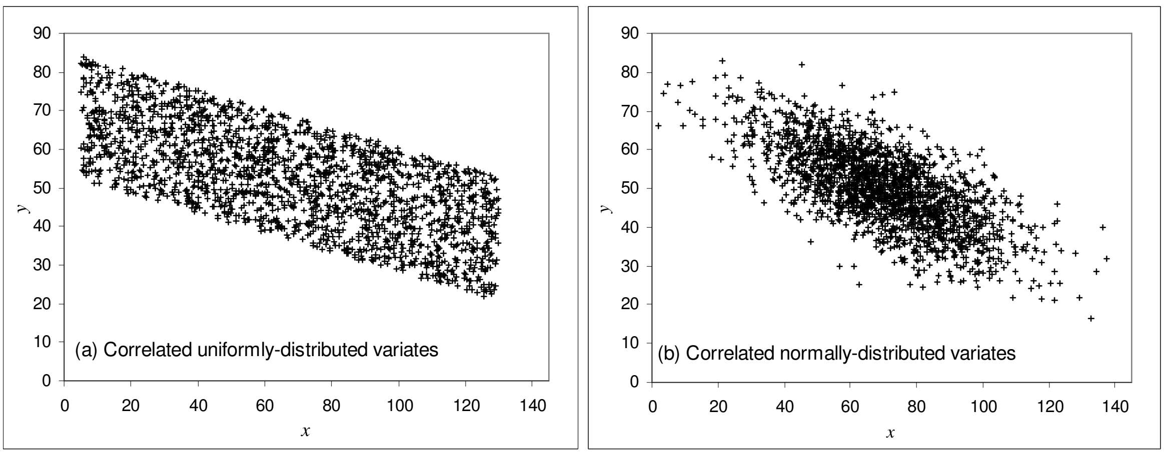


Figure 5 Illustration of generation of variables with a correlation of -0.7 based on (a) uniform and (b) normal distributions.

3.5. Use of Fitted Functions

Often the best means of capturing dependency between two variables is to fit a function between the two, and then to stochastically sample about the fitted relationship. The steps involved in this approach are summarised as follows:

1. Prepare a scatter plot with the primary (independent) variable of interest (x) plotted on the x-axis, and the dependent variable (y) on the y-axis.
2. Fit a suitable function that adequately captures the variation in the dependent variable as a function of the independent variable, $\hat{y} = f(x)$. It is easiest if the function is fitted in such a way that the residuals ($r = \hat{y}$-y) are normally distributed, which is a standard requirement of least squares regression.
3. Generate values of the independent variable by the most appropriate means, and for each value calculate the "best estimate" of the dependent variable ($\hat{y}$) from the fitted function.
4. Add to the "best estimate" a stochastic component (ε) which is generated from the distribution of residuals obtained from Step 2; that is, estimate y' = $\hat{y} + \varepsilon$. If the residuals obtained from Step 2 are found to be normally distributed then $\varepsilon = X\sigma$, where σ is the standard deviation of the residuals (r) and X is a random normal deviate, $N(0,1)$. If the residuals are not normally distributed then an alternative function must be used to ensure that the nature of scatter about the line of best fit is adequately preserved.

An example application of the above steps to the generation of two correlated variables is illustrated in Figure 6. The scatter plot of the variables is shown in panel (a). As is common with correlated hydrologic variables, the degree of scatter increases with magnitude, and thus in order to ensure that the residuals of the fitted function are normally distributed, the variables are both first transformed into the log domain, as shown in Figure 6(b). A non-linear function is fitted to the data, and a "quantile-quantile" plot (see 3.3b) is constructed to check that the residuals are normally distributed (Figure 6c). Figure 6(d) shows a stochastic sample of 2000 values. This sample is obtained by use of the fitted function shown in Figure 6(b) with the added "noise" obtained from the stochastic generation of normal residuals as summarised in Steps 3 and 4, above; the resulting values are than transformed back into the arithmetic domain. It is seen that the stochastic sample in Figure 6(d) exhibits the same correlation as the observed data in panel (a), with both the degree of non-linearity and non-constant variance adequately preserved.

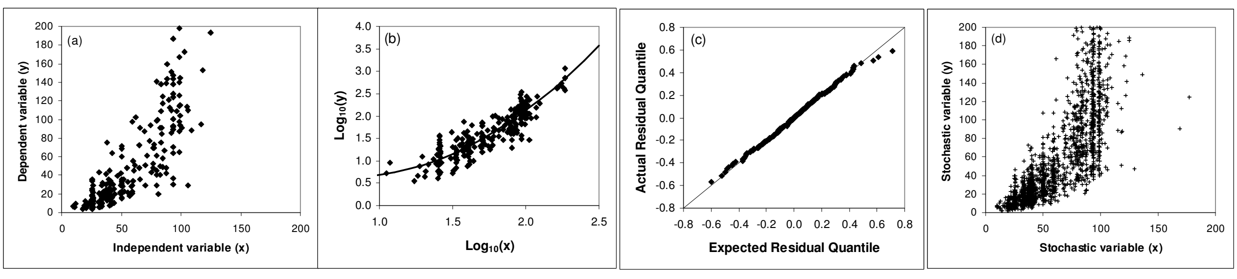


Figure 6 Illustration of the steps involved in the generation of a dependent variable from a fitted function.

3.6. Conditional Sampling

The preceding two sections provide a means for generating “well-behaved” variables that can be fitted to a suitable function or distribution. However, many correlated hydrologic variables are awkwardly distributed and their variability is dependent on some (often non-linear) function of their magnitude. A typical example of this type of correlation is the manner in which the level in an upstream reservoir is weakly dependent on the level in a downstream reservoir. The nature of one such dependence is shown by the large solid symbols in Figure 7, which is derived from the behaviour of two reservoirs located in south-eastern Australia. Such data is difficult to normalise or fit to probability distributions, and thus an empirical sampling approach can be used.

The approach that can be followed to stochastically sample from such a data set can be described as follows:

1. Identify the “primary” variable that is most important to the problem of interest, and prepare a scatter plot of the two variables with the primary variable plotted on the x-axis (as shown in Figure 7).
2. Divide the primary variable into a number of ranges such that variation of the dependent variable (plotted on the y-axis) within each range is reasonably similar; in the example shown in Figure 7 a total of seven intervals has been adopted as being adequate. This provides samples of the secondary variable that are conditional on the value of the primary variable.
3. Stochastically generate data for the primary variable using the empirical approach as described in Section 3.3(d).
4. Derive an empirical distribution of the dependent data for each of the conditional samples identified in Step 2 above (that is, undertake Step 1 of the empirical approach as described in Section 3.3(d) for each of the intervals); thus, for the example shown in Figure 7 a total of seven separate empirical distributions of upstream storage levels are prepared;
5. For each generated value of the primary variable, stochastically sample from the conditional distribution corresponding to the interval that it falls within; for example, if a downstream storage level of 1500 ML was generated in Step 3 above, then the empirical approach described in Section 3.3(d) is applied to the conditional distribution obtained from data occurring within the third lowest interval shown in Figure 7.

The results from application of the above procedure are illustrated in Figure 7 for 2000 stochastic samples (shown by the small “+” symbols). It is seen that the correlation structure in the observed data set is preserved reasonably well by this procedure.

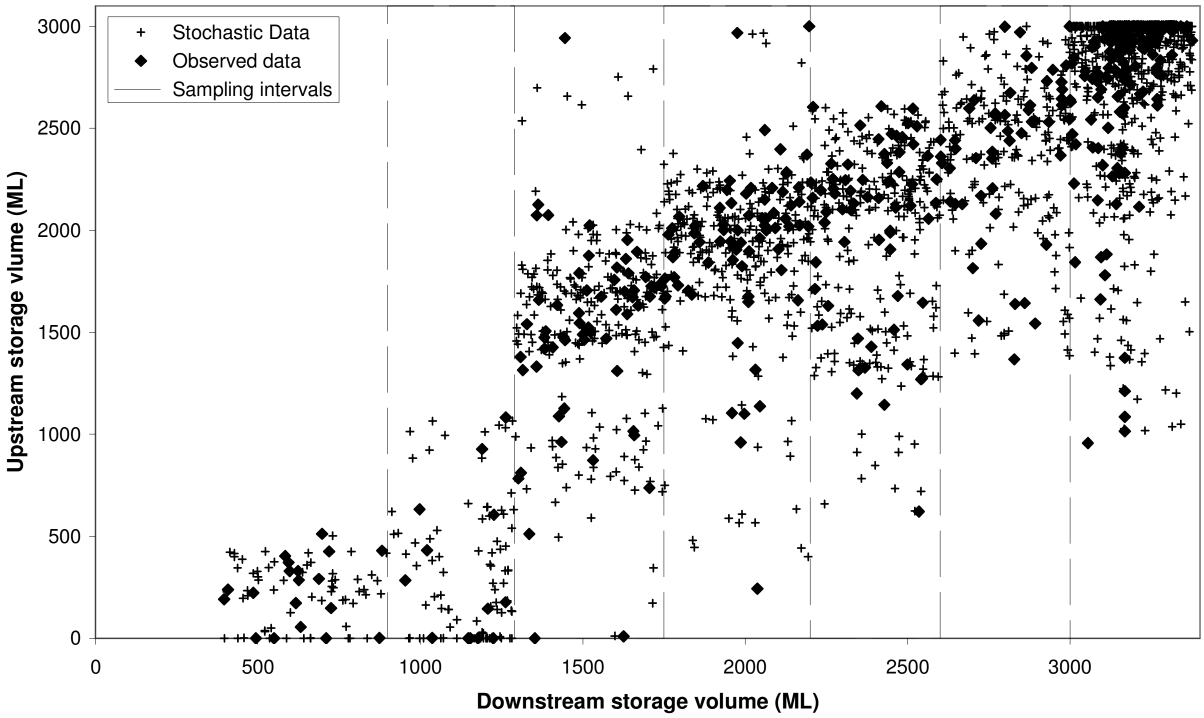


Figure 7 Illustration of conditional empirical sampling in which the storage volume in an upstream dam is correlated with the volume in a downstream dam (2000 correlated values are stochastically generated based on information contained in 500 observations).

4. Estimation of the Derived Frequency Distribution

4.1. Selection of Method

There are broadly two approaches to constructing a derived frequency distribution from the Monte Carlo simulation results: either the simulation employs direct sampling and the results are analysed using non-parametric frequency analysis methods, or else a stratified sampling scheme can be employed and the results analysed by application of the Total Probability Theorem. The decision regarding which approach to use is largely a practical one. It is always necessary to experiment with many different model parameters, model configurations, and design scenarios, and simulation times of more than an hour or so soon become impractical.

The first approach – where the results are derived by direct sampling and analysed using traditional frequency analysis – is the most straightforward to implement. It is well suited to the analysis of problems that can be computed quickly, or else to more complex problems in which the probability range of interest is limited to reasonably frequent events. As a rule of thumb, the number of simulations required is around 10 to 100 times the largest average recurrence interval of interest. That is, if the rarest event of interest is the “thousand year” event (more correctly speaking an event with an annual exceedance probability of 0.001), then it will be necessary to generate between 10000 to 100000 stochastic samples in order to derive a stable result. This approach requires little programming skill, and an adequate scheme can be formulated by developing a command “batch” file that runs your standard model of choice many times with the required range of design inputs.

The second approach does require more effort to implement and, although it is ideally implemented using customised software, it can still be formulated using a “batch” file approach. The benefit of this additional programming effort is that the number of runs required to estimate the exceedance probability of rare events is considerably fewer; indeed the algorithm can be designed so that a similar number of runs is required regardless of the range of probabilities of interest.

Both schemes can be implemented using standard spreadsheet programs; for example the analysis of long-term behaviour of tailings dams can be easily implemented in a spreadsheet. However, for most design applications it is preferable to develop a scheme that takes advantage of existing flood estimation models. There are two reasons for this: first, the model adopted can be selected from the range of proven and accepted flood models, and second the nature of the design inputs (such as losses and temporal patterns) are similar to that used in traditional design event approaches. Rahman et al (2002[a]) provide an example of the first approach described above in which the URBS model is run as a batch process and the results analysed using frequency analysis methods; other examples applied to flood estimation are also available (Wark, 1982; Grayson et al, 1996; Nazarov, 1999; Rahman et al, 2002[b]). The stratified sampling scheme described below was first developed for the estimation of AEP-neutral extreme floods (Nathan et al 2002, 2003) and is implemented in version 5 of the RORB model (Laurenson et al, 2005).

Further information on these two approaches is provided in the next two sections.

4.2. Direct Sampling

The general framework to be used here is that illustrated in Figure 1. The number of model simulations (*N*) should be selected so that the design objective of interest is largely unaffected by the undertaking of additional runs; in general this is around 10 to 100 times the reciprocal of the exceedance probability of interest. The results from each run retained for further analysis are dependent on the design objective of interest, but in the context of flood estimation these are usually related to flood peak, flood volume, or maximum water level estimated at one or more locations.

4.2.1. Sampling of Design Rainfall Depth

For flood applications the primary variable to be generated is the rainfall depth for a given duration. For rainfall depths up to an annual exceedance probability of 1 in 500 the algebraic procedures presented in Section 1.3 of Book II can be implemented. For rarer rainfalls the simplest approach is to sample directly from the cumulative distribution of design rainfalls. The steps involved in this are as follows:

1. Prepare a function that relates rainfall depth to its annual exceedance probability. This is most easily achieved by interpolating between pairs of rainfall depths and corresponding exceedance probabilities, or else by adopting a suitable function. The former approach is applicable to estimating Rare rainfalls if the “shape-factor” approach is adopted (as described in Section 3.6.2 of Book VI), where interpolation is best undertaken in the arithmetic-normal probability domain. The latter approach is applicable if the parabolic interpolation function of Siriwardena and Weinmann (1998) is used (Section 3.6.3, Book VI).
2. Generate a random number that is uniformly distributed between 0 and 1; this number represents the annual exceedance probability of the rainfall event being simulated.
3. Use the relationship developed in step 1 to find the rainfall depth associated with the exceedance probability generated in step 2.

This sample value of design rainfall depth is then used with stochastically sampled values of other inputs/parameters for a simulation model run, and the procedure repeated *N* times.

4.2.2. Frequency analysis of simulated flood characteristics

The results output from the Monte Carlo simulation are most easily analysed by non-parametric frequency analysis. Using flood peaks as an illustration, the steps involved can be summarised as follows:

1. Sort the *N* simulated peaks in order of decreasing magnitude.
2. Assign a rank (*i*) to each peak value; 1 to the highest value, 2 to the next highest, and so on, down to rank N.
3. Calculate the plotting position (*p*) of each ranked value using either the Weibull (Equation 1) or the Cunnane (Equation 2) formulae:

$$p = \frac{i}{N+1}$$ **Equation 2**

$$p = \frac{i-0.4}{N+0.2}$$ **Equation 3**

4. If the design focus is on estimating the *probability* of a given flood magnitude then the Weibull formula (Equation 1) should be used as this provides an unbiased estimate of the exceedance probability of any distribution. Alternatively, if the focus is on the *magnitude* associated with a given exceedance probability then the Cunnane formula (Equation 2) is preferred as this provides approximately unbiased quantiles for a range of distributions.
5. Construct a probability plot of the ranked peaks against their corresponding plotting positions. The plot scales should be chosen so that the frequency curve defined by the plotted values is as linear as possible. In many hydrological applications the ranked values are plotted on an arithmetic or log scale and the estimated exceedance probabilities (the plotting positions) are plotted on a suitable probability scale. Most popular spreadsheet programs do not include probability scales and thus, for probability plots conforming approximately to the Normal or log-Normal distribution, it is necessary to convert the probabilities to their corresponding standard normal cumulative distribution values. Alternatively, for probability plots conforming approximately to the exponential distribution, the reciprocal of the exceedance probabilities (the average recurrence interval) can be plotted on a logarithmic scale.
6. The magnitude associated with a given exceedance probability (if the Cunnane plotting position is used) or else the exceedance probability associated with a given magnitude (if the Weibull plotting position is used) can be read directly from the probability plot. For convenience, a suitable smoothing function can be fitted to the plotted values in the region of interest to simplify the estimation of design values. The function is used merely to interpolate within the body of the plotted points and thus, as long as there is no bias in the fit, it matters little what function is used (polynomial functions are quite suitable).

An application of this approach to simulate water levels in a simple storage is illustrated in Figure 8. Each point on the graph represents the maximum water level reached in one year of simulation, and a Weibull plotting position has been used to estimate the individual exceedance probabilities. The derived frequency curve is based on 5000 simulations and it is seen that the average recurrence interval of the storage reaching a level of 160 is estimated as about 70 years, corresponding to an annual exceedance probability of 0.014.

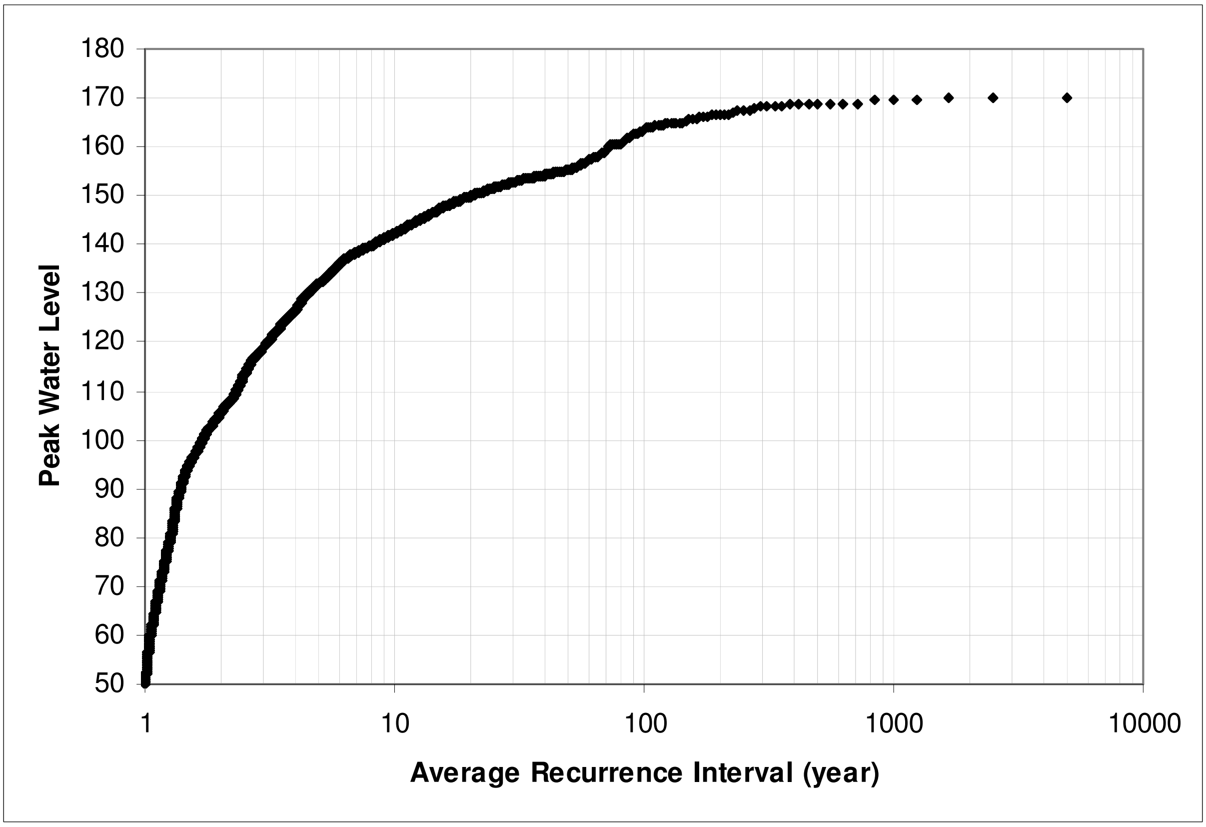


Figure 8 Illustration of derived frequency curve of peak water levels obtained using direct simulation.

4.3. Stratified Sampling

While the above approach is straightforward, it is computationally inefficient. For the example illustrated in Figure 8 only around 500 of the 5000 simulations provide any information about events with exceedance probabilities rarer than 1 in 10. That is, the vast majority of computational effort is expended on deriving results for the range of exceedance probabilities that is of least interest. This inefficiency is of little concern when using simple models with parsimonious outputs and fast simulation speeds. However, as the data processing becomes more complicated and execution speeds increase, simulation times and data storage requirements quickly pose significant practical problems.

Adoption of a stratified sampling approach ensures that the computational effort is always focused on the region of interest and, if the simulation scheme is configured carefully, then it will usually be possible to apply Monte Carlo simulation to most practical problems.

The approach follows the same logic as represented in the flow chart of Figure 1, however an additional step is introduced into the generation of design rainfalls, and the results are analysed in a different fashion. The conceptual steps in this approach are illustrated in Figure 9. The following description is based on that developed by Nathan et al (2002, 2003) to analyse the conversion of extreme rainfalls to floods, though the same approach could be applied to any stochastic-deterministic system.

First the rainfall frequency curve is divided into N intervals uniformly spaced over the standardised normal probability domain (Detail A in Figure 9). Typically 50 intervals should suffice, though care is required to ensure that there is adequate sampling over the region of most interest. (Non-uniform intervals could be adopted in which finer resolution is adopted for critical areas of interest and, if desired, the divisions could be based on an appropriate discretisation of the rainfall depth distribution, or whatever the primary variable of interest is). Within each interval *M* rainfall depths are stochastically sampled and for each rainfall depth a model simulation is undertaken using an appropriate set of stochastic inputs (Detail B in Figure 9). The number of simulations specified in each interval (*M*) is dependent on the number of inputs being stochastically generated and their degree of variability, but in general it would be expected that between 50 and 200 simulations should be sufficient to adequately sample from the range of associated inputs.

The model results are recorded for all simulations taken in each interval (Detail C in Figure 9). These results are aggregated using the Total Probability Theorem (see below) to yield expected probability estimates of the flood frequency curve. In all, if the rainfall frequency curve is divided into 50 intervals and 200 simulations are undertaken in each interval, a total of 10000 runs is required. The same number of simulations could be used whether the upper limit of exceedance probability is 1 in 1000 or 1 in 10^6, and it is merely necessary to ensure that a representative number of combinations is sampled within each rainfall range of interest.

For the scheme illustrated in Figure 9, the expected probability that a flood peak (*Q*) exceeds a particular value *q* can be calculated from:

$$p(Q > q) = \sum_i p[Q > q \mid R_i]\, p[R_i] \qquad \text{Equation 4}$$

where the term $p[R_i]$ represents the probability that rainfall occurs within the interval *i*, and the term $p[Q>q|R_i]$ denotes the conditional probability that the flood peak *Q* generated using a rainfall depth from within this interval R_i exceeds *q*. The term $p[R_i]$ is simply the width of the probability interval under consideration (this will be different for each of the *N* intervals considered), and $p[Q>q|R_i]$ can be calculated merely as the proportion of exceedances, *n*, in the sample of *M* simulations within interval *i* (ie as *n/M*). A representative value of *R* can be used for all *M* simulations within the interval, though a smoother frequency curve can be obtained if *R* is sampled with the interval using a uniform distribution.

In order to ensure that the total probability domain is sampled, it is necessary to treat the first and last intervals differently from the intermediate ones. The issue here is that the full extent of the end intervals is not being adequately sampled, and on the assumption that these boundary intervals are distant from the probability region of interest, we can estimate their contribution to the total probability in a pragmatic fashion. For the last interval $p[R_1]$ is evaluated as the exceedance probability of its lower bound, and for the first interval it is evaluated as the non-exceedance probability of its upper bound. Also, for the first interval $p[Q>q|R_1]$ is replaced by the geometric mean of $p[Q>q|R_1^*]$ and, say, 0.1 x $p[Q>q|R_1^*]$, where R_1^* is the rainfall value at the upper bound of the interval. Similarly, for the last interval the term $p[Q>q|R_N]$ is replaced by the geometric mean of $p[Q>q|R_N^*]$ and 1.0, where R_N^* is the rainfall value at the lower bound of the interval. Thus, we are assuming for the lowest interval that as the frequency of the rainfall event becomes very high the likelihood that the flow threshold is exceeded trends towards a very low value, in this case taken as one tenth the probability of $p[Q>q|R_1^*]$; and for the uppermost interval we assume that the likelihood of the threshold being exceeded trends towards a value of 1.0 (ie a certainty). The geometric mean is used in place of the arithmetic mean as here we are assuming a highly non-linear variation over the interval.

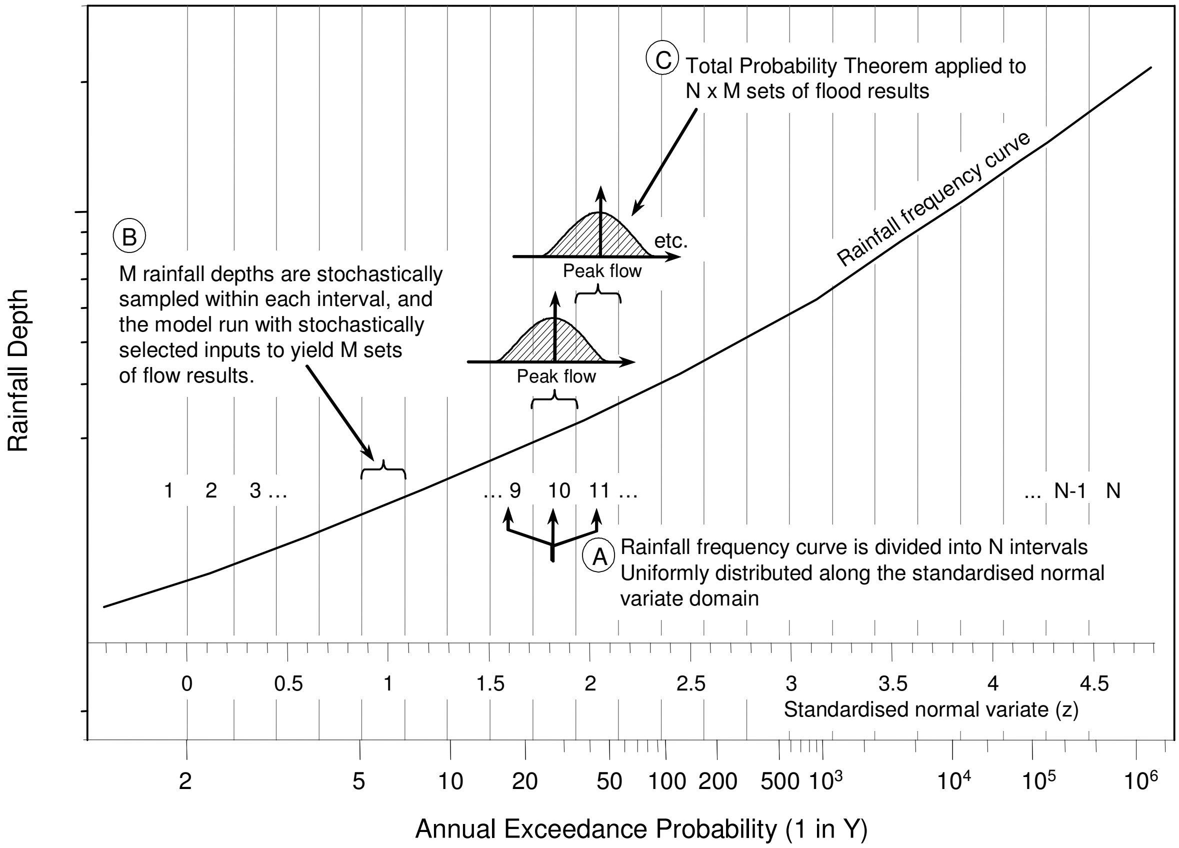


Figure 9 Illustration of manner in which stratified sampling is applied to rainfall frequency curve.

4.4. Identification of Critical Duration

The procedures described in the previous two sections can be applied to any type of rainfall input for which the necessary distributional information is available. The simplest approach – and one that is most consistent with the available design information – is to undertake the simulations for a *range* of storm durations and then adopt the envelope of the results; that is, the envelope curve is constructed by extracting the maximum value at each exceedance probability of interest. This is the same approach that is used with the traditional design event method. However the use of Monte-Carlo simulation provides a major benefit in that each frequency curve is derived from a wide range of flood producing factors. Accordingly, the irregularities between storm duration, event magnitude, and catchment area that can occur due to specific combinations of temporal patterns and other hydrometeorological characteristics are generally removed in the sampling process.

Of course, the rigorous solution to this problem is to stochastically sample the *duration* of the storm event. Hoang et al (1999) have successfully explored this approach, and while conceptually attractive, it does require the use of design inputs which are not currently available. There is good evidence that the inputs required for this approach are strongly linked to existing design information, but the manner in which these dependencies vary with location have not been quantified.

4.5. Validation of Modelling Process

It is essential that any Monte-Carlo scheme be assessed by the inspection of suitable diagnostic checks. Use of a sound conceptual basis does not guarantee that the adopted parameterisation adequately minimises bias and produces results that are consistent with either historical observations or physical reasoning.

A necessary but not always sufficient check is to ensure that the generated output distributions conform to the required input distributions. To this end plots of exceedance percentiles or cumulative distributions should be compared to check that the input and output distributions are consistent, particularly over the range of probabilities near the design objective of interest. Examples of some suitable diagnostic checks are provided in a worked example presented in Section 7.1. Any correlated variables should also be inspected to ensure that the nature and degree of dependence is adequately represented (as illustrated in Figure 7).

If at all possible, the performance of the simulation scheme should be tested against other independent evidence. The most suitable checks are to compare results with flood frequency estimates derived using (usually) the statistical analysis of annual maxima (as described in Book IV), or else other regional estimates of the required flood quantiles. Figure 10 provides an example of a useful check on the application of a Monte-Carlo scheme to simulate the effects of seasonal flood producing factors and a varying initial reservoir storage. Given that the simulation results fall well within the confidence limits of the fitted flood frequency distribution, it can be concluded that they are consistent with the observed flood data and may provide a suitable basis to extrapolate the flood frequency curve to more extreme events.

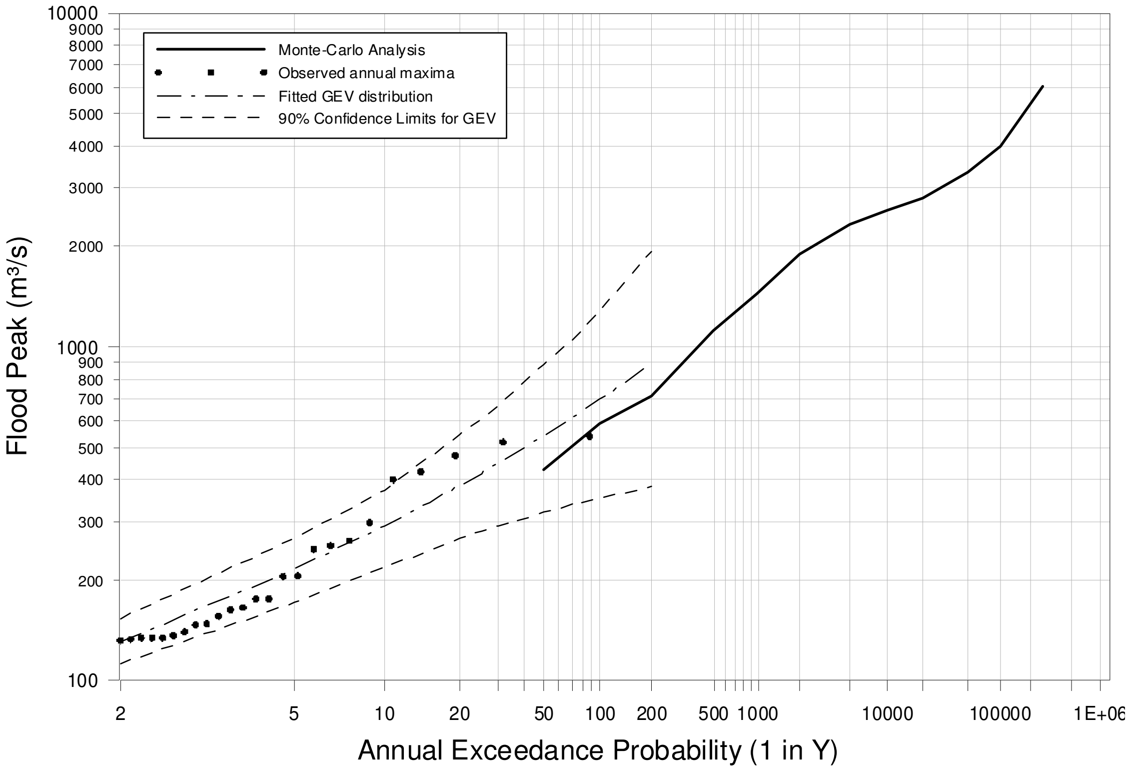


Figure 10 Comparison of flood frequency curve fitted to historical maxima with Monte-Carlo simulation results (assuming all gates operating under current conditions).

5. Flood Estimation Applications

The foregoing sampling methods can be applied to a wide variety of practical design problems. Worked examples of some typical sampling schemes are provided in Section 5; the following provides a few generic examples of applications relating to flood estimation.

5.1. Sampling temporal patterns

Flood estimation is traditionally undertaken using a single temporal pattern of "average variability" as provided in Section 2 of Book II. These single design temporal patterns have been derived from historical samples of temporal patterns; rather than averaging these patterns to provide a single design input (and thereby risk the introduction of bias, particularly with non-linear models) Monte Carlo analysis provides the means to use all patterns in the historical sample. Thus, in concept, the same temporal patterns used to construct the single design temporal patterns can be used as a sample for Monte-Carlo simulation.

A set of historical temporal patterns can be sampled using the discrete form of the uniform distribution (Section 3.3a). Suitable data sets of temporal patterns are provided for extreme long duration storms in the south-east and tropical regions of Australia by Meighen and Minty (1998) and Walland et al (2003), and Jordan et al (2005) provide a suitable set of patterns for use with short duration storms. The RORB model (Laurenson et al, 2005) provides a facility for extracting suitable sets of temporal patterns from pluviograph data. An illustration of the variability in an historical sample of temporal patterns compared to a fixed design pattern is shown in Figure 11. In this particular example it is seen that the fixed design pattern is more rear-loaded than the majority of the historical patterns, and it is also not representative of the sample's central tendency.

There is some evidence that temporal patterns become more uniform as the severity of the event increases (Pilgrim et al, 1969; Nathan, 1993; Jordan et al, 1995), and thus it may be appropriate to conditionally sample the temporal patterns based on the magnitude of the rainfall depth. This is best achieved by applying the uniform discrete sampling scheme described in Section 3.2 to samples of temporal patterns censored by an appropriate range of rainfall depths. A practical example of this conditional sampling approach has been incorporated into the Monte-Carlo simulation scheme in RORB (Laurenson et al, 2005).

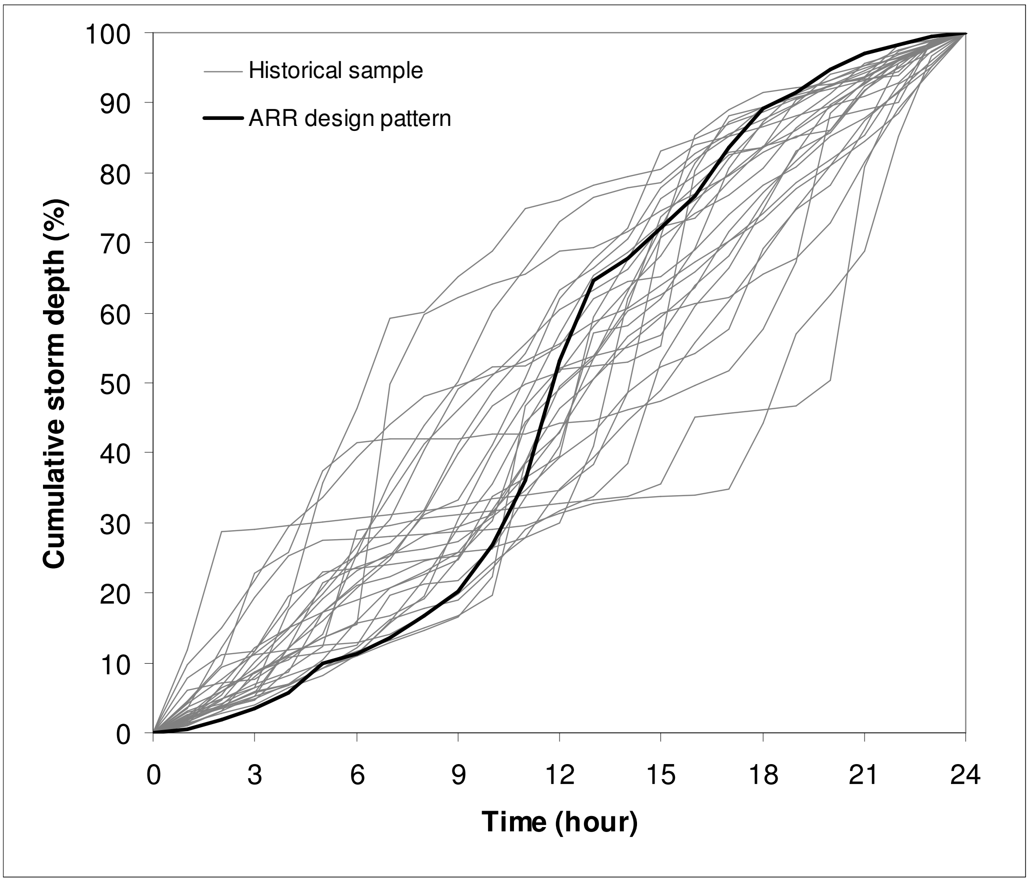


Figure 11 Cumulative plot of 24 hour temporal patterns from 30 largest storms extracted from Melbourne Regional Office pluviograph station with an ARI less than 25 years), and corresponding Zone 1 pattern for ARI < 25 year from Section 2, Book II).

5.2. Sampling losses

Losses used to estimate rainfall excess are also traditionally treated as fixed values in the design event approach. A substantial amount of information is available on the distribution of initial and continuing losses (Hill et al, 1997; Ilahee, 2005), and this can be pooled to derive distributions that can be used directly in Monte-Carlo simulation. The concepts involved in this approach are described by Nathan et al. (2003) and are illustrated for initial loss in Figure 12. Empirical data and analyses presented by Hill et al (in prep) clearly support the assumption that while the *magnitude* of losses may vary between different catchments, the *shape* of the distribution does not. In other words, while it may be expected that typical (mean or median) loss rates vary significantly from one catchment to another, the degree of variation around this typical value is similar, and thus the likelihood of a catchment being in a *relatively* dry or wet state is similar for all catchments.

The concept of how the location of the loss distribution changes but not its shape is schematically illustrated in Figure 12. This approach is easily implemented using an empirical sampling scheme (comprised of pairs of exceedance probabilities and design values, Section 3.3d). The mean or median loss value applied in the traditional approach would generally be expected to provide a suitable estimate of the location parameter of the loss distribution. If there is evidence to suggest that this loss distribution may introduce bias into the flood estimates, the location parameter of the loss distribution could be adjusted..

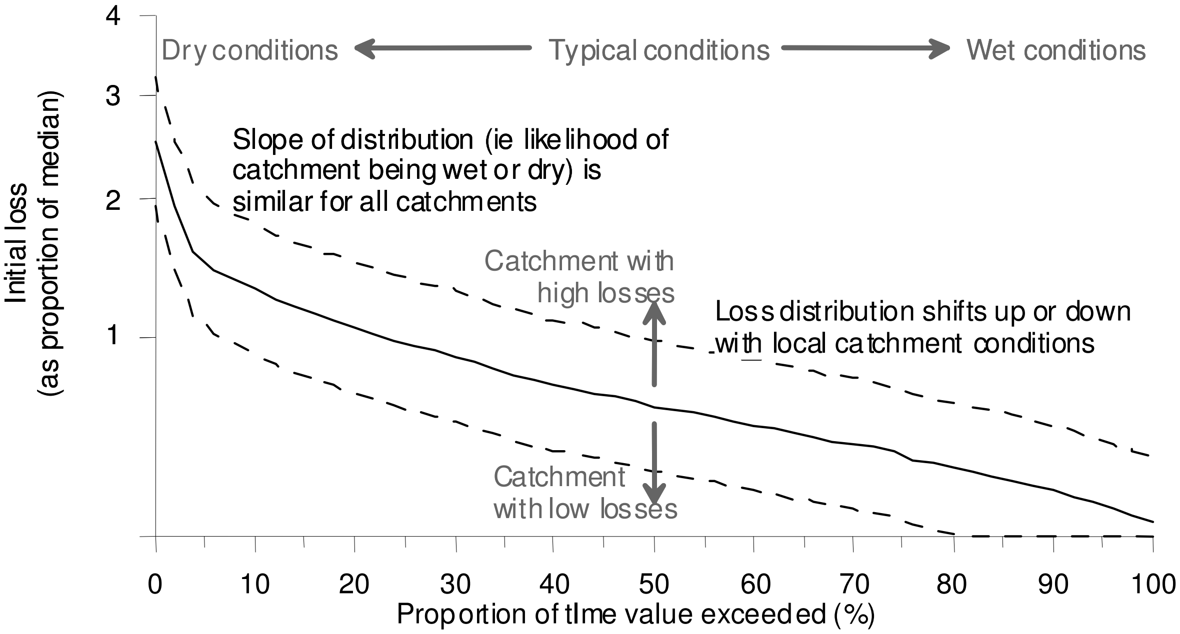


Figure 12 Schematic illustration of variation in location but not shape of initial loss distribution.

5.3. Antecedent Conditions

Antecedent conditions are often an important consideration in flood hydrology. Examples of this include selection of an initial starting condition in a reservoir, the rainfall that occurs prior to a design burst, initial snowpack conditions, and downstream tailwater controls caused by tidal influences. In some cases the antecedent conditions can be considered to be largely independent of the associated storm event, and thus they can be handled by any of the distributional or empirical approaches described in Section 3.3. However, it is not uncommon for these antecedent conditions to be correlated with storm events (Hill et al 2005 provide an example of an initial reservoir level that is conditional upon rainfall depth), in which case the dependence must be explicitly handled by one of the approaches described in Section 3.4, 3.5, or 3.6.

5.4. Concurrent Tributary Flows

The contribution of tributary flows is generally best handled by inclusion of the tributary catchment as a sub-area in the flood estimation model. However, there are occasions where it is either impractical or not desirable to do this, for example in those cases where the availability of data obviates the need for rainfall-based modelling, or a hydraulic model is being used to estimate flood levels some point downstream of the hydrologic model.

Where information on flood maxima is available on both the mainstream and the tributary, the most attractive approach is to separately fit flood frequency distributions to the individual sites, and then to combine them using correlated distributions, as described in Section 3.4. If floods are being estimated using regional information, then it may be preferable to adopt a more empirical approach, as described in Section 3.6.

5.5. Operational Factors

There are occasions where flood magnitude is dependent on stochastic operational factors, and there are a variety of approaches that can be considered.

One common example is the likelihood of spillway blockage or gate failure at a dam, or culvert blockage in a drainage system, during flood conditions (Nathan et al, 2003). The likelihood of such partial failure can be incorporated by assigning likelihoods to the different possible outcomes, and then undertaking the simulation by discrete sampling. For example, if a dam has three spillway gates, then the probability of one, two, or three gates failing can be computed using combinatorial statistics, or by development of a fault-tree. These conditions can be simulated by adoption of separate rating curves corresponding to the different failure conditions, where the likelihood of a given scenario is sampled using a discrete form of the empirical distribution (Section 3.6).

Another common example is the operation of power stations or some other anthropogenic process. Such processes can be handled stochastically by the use of operating rules combined with an error term (as described in Section 3.5), or else by the selection of a range of possible outcomes by discrete (possibly correlated) empirical sampling as described in Sections 3.3 and 3.6.

6. Characterising Uncertainty

The preceding discussion has focused on the use of Monte Carlo simulation to derive frequency distributions of a selected flood characteristic, allowing for natural variability in inputs. In addition to deriving best estimates, the simulation methods can also be used to assess the impact of *uncertainty* in inputs or model parameters on the derived frequency distributions. The steps involved in doing this may be summarised as follows:

1. Separate uncertainty in the description of model inputs from the natural variability of these inputs reflected in the Monte Carlo simulation procedures outlined previously.
2. Characterise the distribution of uncertainty in the salient model inputs (or model parameters) by use of an appropriate distribution and selection of its parameters.
3. Select one set of model inputs and model parameters that has been perturbed in accordance with the distributions determined in step 2, and undertake Monte Carlo simulation by direct or stratified sampling (Sections 4.2 or 4.3) to yield a single derived frequency distribution.
4. Repeat step 3 a sufficient number of times to adequately sample from the 'uncertainty' distribution of inputs and parameters (say, one hundred times), saving the results of each simulation run.
5. Rank all the results derived for each exceedance probability of interest (e.g. the 1:100 AEP flood event), and extract the required uncertainty limits; that is, if 100 sets of results are available, extract the 1:100 AEP flood event from each set of results, rank them, and the 5th largest and 5th lowest estimates represent the 90% confidence limits surrounding the 1:100 AEP result.

While the mechanics of the above method are straightforward to apply, separating uncertainty in input characteristics from natural variability and deriving the distributions used to characterise uncertainty in the inputs (Step 1) is a most difficult task. Not only are the limits and the nature of the distribution often difficult to ascertain, but the inputs may well be correlated and this also must be taken into account. In some cases (such as the adopted median of the loss distribution) the uncertainty distribution can be based on empirical data, but in other cases – such as the degree of non-linearity of flood response – such uncertainty must be based on pragmatic assumptions and the results must be regarded as speculative, or at best indicative, of the true uncertainty. In such cases the uncertainty analysis reduces to a sensitivity analysis, useful to answer 'what-if' questions.

One example of the pragmatic application of the above is presented in Figure 13. These results are based on 100 sets of Monte-Carlo simulations where the uncertainty associated with the following factors was considered:

- *Design rainfalls*: the uncertainty associated with the 1:2000 AEP event (derived using CRC-FORGE procedures, Section 3.5, Book VI) was assessed to be around ±10%, and the uncertainty in the AEP of the Probable Maximum Precipitation was assessed using the judgemental distribution proposed by Laurenson and Kuczera (1999; Section 3.6, Book VI); for convenience this was well reproduced by the Beta distribution using parameter values of $\alpha=\beta=2.6$, x_{min}=-2 and x_{max}=2.
- *RORB Routing parameter k_c* (Laurenson et al, 2005): uncertainty in the routing parameter k_c was characterised using a skewed Beta distribution to match the uncertainty in the spread of estimates obtained from reconciliation with flood frequency quantiles (as discussed in Section 6.4.2 of Book VI)

The uncertainty associated with the degree of non-linearity in flood response was not assessed due to its high dependence on the selected value of the catchment routing parameter. Uncertainty in other flood producing factors was not assessed as this was considered to be small relative to the two sources of uncertainty allowed for in the assessment.

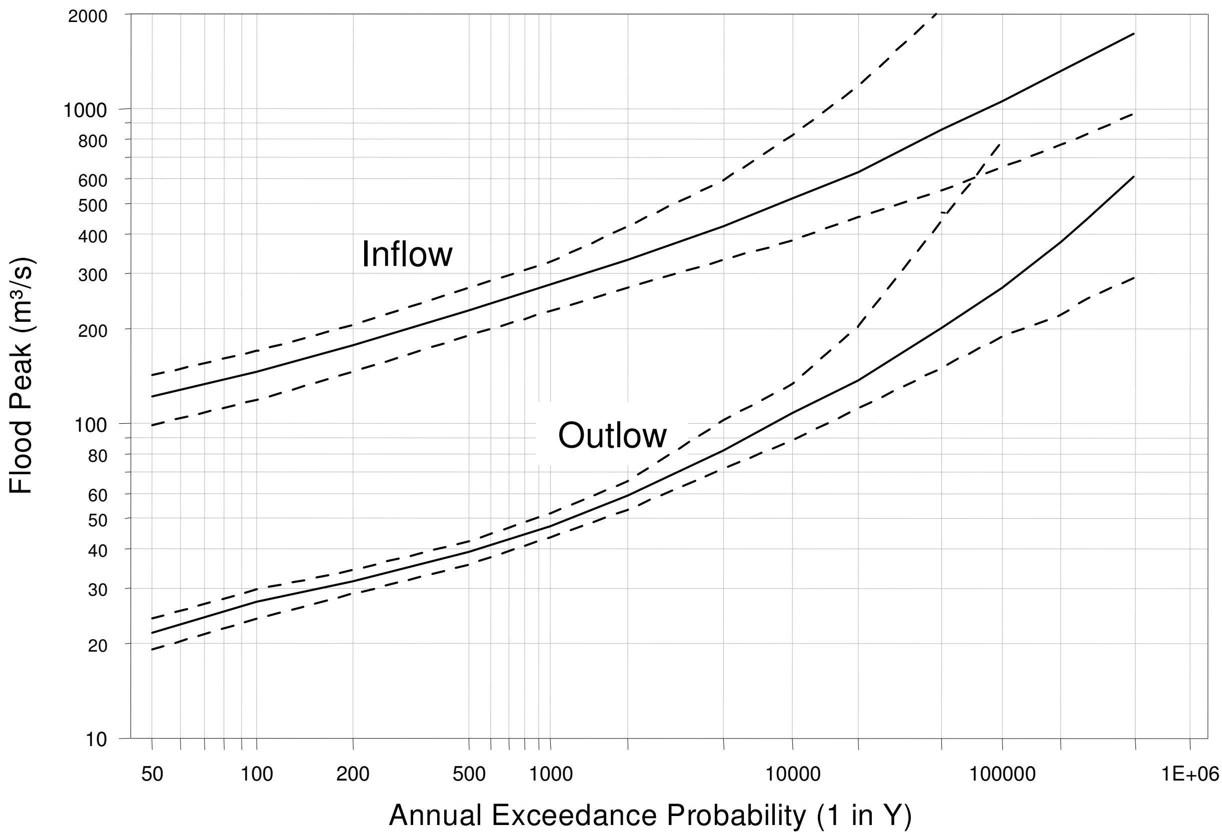


Figure 13 Illustration of the assessment of uncertainty in the derivation of inflow and outflow frequency curves for a dam with a fixed crest spillway.

7. Worked Examples

7.1. Generation of Correlated Time Series

This example illustrates the stochastic generation of a dependent time series in a manner that preserves its serial correlation. This type of problem commonly arises when using stochastic techniques to estimate overtopping floods for a tailings dam or some other system that involves simulation of a time series. It also illustrates generation of a correlated variable as would be required in application of Monte Carlo techniques to event-based flood estimation.

For the purpose of this example only 50 data items are considered, however in practice it would be expected that such an analysis would be based on a larger sample (containing many thousands of items, or multiple sets of many thousands). Tabulated values involved in the main computational steps are presented in Table 1. The nature of the relationship between the independent (X) and dependent (Y) variables of interest is shown by the solid symbols in the scatter plot of Figure 14. It is seen that there is a slight negative correlation between the two variables, and that there is a considerable degree of scatter about the fitted relationship.

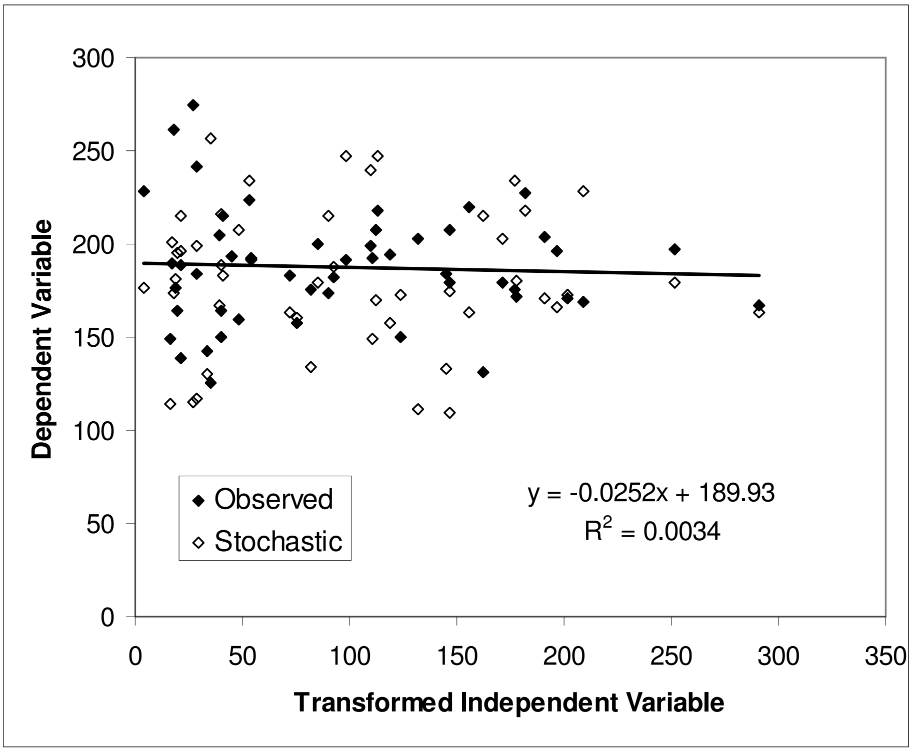
The first step in this example is to generate deterministic estimates of the dependent variable (Y') using the simple fitted relationship:

$$Y' = 0.02528.X + 189.93$$

This relationship is obtained as a simple least-squares regression fit to the data and is easily undertaken using a spreadsheet. The regression estimates for each value of the independent variable is provided in the fourth column of Table 1. The degree of scatter about the line is best characterised as the standard deviation of the residuals, that is the standard deviation of the difference between the observed and fitted values of the independent variable (Y'-Y). The calculation of the residuals is shown in the 5th column of Table 1, and the standard deviation is calculated to be 30.5 (as shown in the second last line of the table).

Table 1 Computational steps for generation of correlated time series..

Time Increment (T)	Independent Variable (X)	Dependent Variable (Y)	Regression Estimate (Y')	Regression Residual (Y'-Y)	Random Normal Deviate N(0,1)	Stochastic sample of dependent variable
1	93	182	187.6	5.9	-0.430	174
2	17	189	189.5	0.1	0.729	201
3	41	215	188.9	-26.3	-0.686	183
4	39	204	188.9	-15.4	-0.396	167
5	29	184	189.2	5.0	0.628	199
6	98	191	187.5	-3.9	1.924	247
7	110	199	187.2	-11.5	0.803	240
8	18	261	189.5	-71.5	-1.116	174
9	27	275	189.3	-85.2	-2.169	115
10	29	241	189.2	-52.0	-1.417	117
11	4	229	189.8	-38.9	0.386	176
12	72	183	188.1	4.8	-1.211	163
13	291	167	182.6	15.5	0.024	163
14	75	158	188.0	30.0	-1.081	161
15	147	180	186.2	6.7	0.233	174
16	112	208	187.1	-20.5	-0.815	170
17	124	150	186.8	36.5	-0.036	172
18	111	192	187.1	-5.2	-1.473	149
19	197	197	185.0	-11.6	0.202	166
20	113	218	187.1	-31.1	2.221	248
21	182	227	185.3	-41.6	-0.170	218
22	45	194	188.8	-4.7	0.290	193
23	85	200	187.8	-11.9	-0.515	179
24	162	131	185.8	54.7	1.458	215
25	40	150	188.9	38.8	0.115	216
26	21	139	189.4	50.3	0.929	215
27	209	169	184.7	15.5	1.115	229
28	178	172	185.4	13.6	-0.918	180
29	90	174	187.7	13.7	1.638	215
30	177	175	185.5	10.4	0.833	234
31	145	184	186.3	2.7	-2.592	133
32	132	203	186.6	-16.4	-1.250	112
33	147	208	186.2	-21.8	-2.172	110
34	82	175	187.9	12.5	-0.691	134
35	20	164	189.4	25.4	0.675	195
36	53	224	188.6	-35.1	1.341	234
37	35	126	189.0	63.3	1.767	257
38	48	159	188.7	29.3	-0.401	208
39	19	177	189.5	12.6	-0.073	181
40	171	179	185.6	6.4	0.711	203
41	21	189	189.4	0.9	-0.193	196
42	119	194	186.9	-7.1	-1.022	157
43	191	204	185.1	-18.7	0.081	170
44	156	220	186.0	-34.0	-0.952	163
45	252	197	183.6	-13.1	0.443	179
46	54	192	188.6	-3.4	-0.164	192
47	202	171	184.8	13.9	-0.369	173
48	40	164	188.9	25.1	0.236	189
49	34	142	189.1	46.8	-2.451	130
50	16	149	189.5	40.3	-1.348	114
Average	97	**187**	187.5	0.0	-0.147	**182**
Std Dev	71	**31**	1.8	**30.5**	1.127	**37**
Ser Corr	0.15	**0.54**	0.153	0.538	0.207	**0.57**



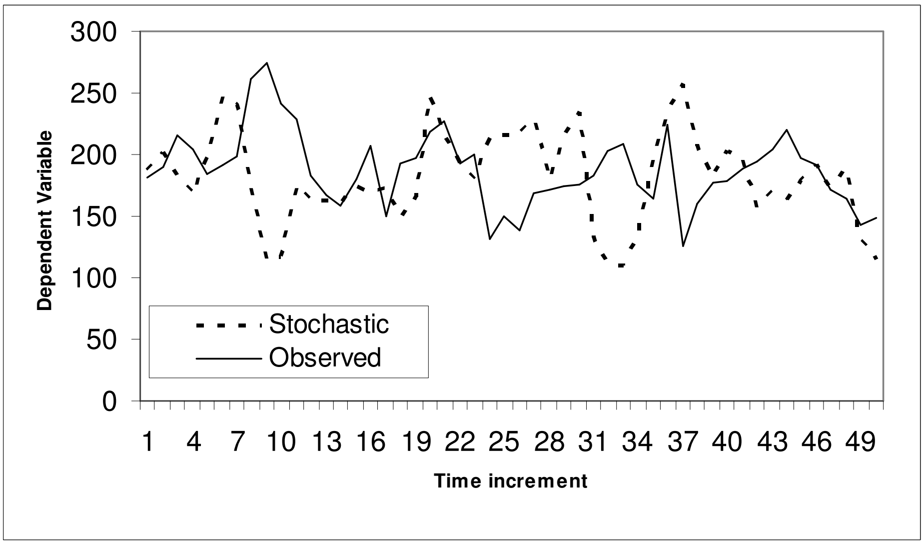


Figure 14 Scatter and time series plot of observed data and one stochastic sample.

The next step is to generate some random numbers. We will first check that the residuals are normally distributed about the fitted regression line (which is a necessary assumption of using least squares regression). This is done most easily using a "quantile-quantile" plot which is constructed by plotting the expected and observed cumulative distributions and assessing the degree of deviation. To this end, the residuals of the regression equation are ranked from lowest to highest and each value is assigned a plotting position using the Weibull formula (Equation 1), which provides an unbiased estimate of the exceedance probability of reach rank. The expected quantiles are computed using the inverse of the cumulative normal distribution for the corresponding exceedance probability. For example, the residual in the first row of the fifth column of Table 1 (5.90) is ranked 29th out of the sample of 50. Using the Weibull formula the exceedance probability of this rank is 29/(50+1)= 0.569, and the corresponding inverse of the cumulative normal distribution is 5.28 (using a mean and standard deviation of 0.0 and 30.9, as computed in the 2nd and 3rd last rows of the 5th column); the first (x,y) pair of the quantile-quantile plot is thus 5.28 and 5.90. Similar calculations are undertaken for each data point, and the completed plot is shown in Figure 15. It is seen that the residuals fall neatly along the 1:1 line except for the largest negative residual (represented as the lowest point in the lower left hand corner of the figure), and thus it is reasonable to assume that the residuals follow a normal distribution. If the deviation from the 1:1 line was marked then it would be necessary to apply some form of transformation to normalise the data (e.g. the Box-Cox power transform).

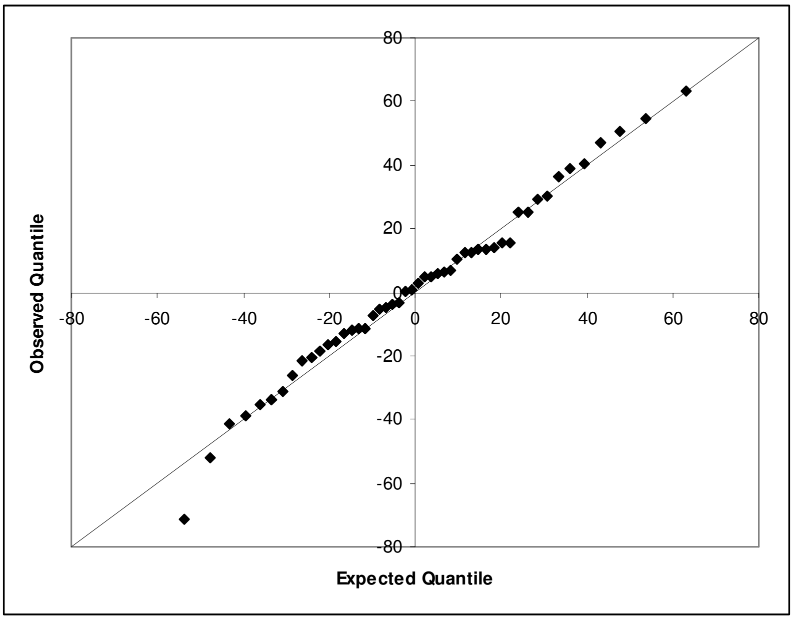


Figure 15 Quantile-quantile plot of regression residuals.

Having satisfied ourselves that the residuals are normally distributed, we can generate as many samples as we need. The process of doing this for a single sample of 50 values is shown in the last two columns of Table 1. First, a standard normal deviate is generated with a mean of zero and a standard deviation of one; this is easily done using standard functions found in proprietary spreadsheet programs whereby a random number is generated between 0 and 1, and then the inverse of the standard normal cumulative distribution (which has a mean of zero and a standard deviation of one) is then calculated using the random number as its input exceedance probability. By way of example, the first row of the 6th column of Table 1 is the result of first generating a random number of 0.3336 (not shown), and then calculating the corresponding inverse of the standard normal cumulative distribution, which is -0.430. The subsequent rows in the 6th column are merely the result of generating a further 49 random numbers and converting them to standard normal deviates in a similar fashion.

The very first stochastic estimate of the dependent variable (S_1) is calculated from:

$$\begin{aligned} S_1 &= Y_1' + s.z_1 \\ &= (-0.0252(93)+189.9)+ 30.5(-0.430) \\ &= 174 \end{aligned}$$

Where Y_1' is the regression estimate (the expected mean for the first independent variable, from Equation 1.6.1), s is the standard deviation of the residuals (30.5), and z_1 is the first standard normal deviate generated (in this case -0.430).

In order to preserve the observed serial correlation (r) from one estimate (S_i) to the next (S_{i+1}), the required dependency is introduced using Equation (XXX) where, for example, the second stochastic estimate is computed as follows:

$$
\begin{aligned}
S_2 &= Y_2' + s(r.z_1 + z_2\sqrt{(1-r^2)}) \\
&= (-0.0252(17)+189.9)+ 30.5(0.54(-0.430) + 0.729\sqrt{(1-0.54^2)}) \\
&= 201
\end{aligned}
$$

The calculations proceed in this fashion for as many stochastic estimates as required, and values for the first 50 such steps are shown in the last column of Table 1. A comparison of this first stochastic sample compared to the historic is shown as hollow symbols in the top panel of Figure 14, and a time series plot showing the required degree of serial correlation is shown in the lower panel.

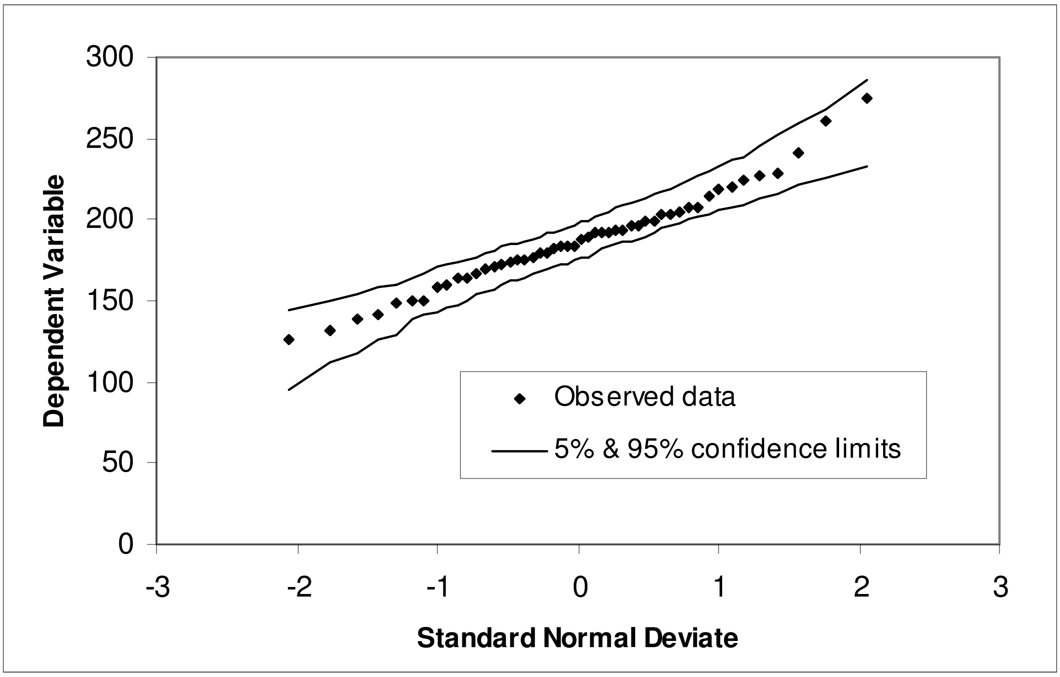


Figure 16 Distributional check on 100 generated replicates of 50 items.

When generating multiple replicates for design purposes it is important to check that the statistics of the samples are consistent with the original data. It is generally a trivial matter to ensure that the statistics of central tendency are preserved, but usually in hydrology it is the whole distribution of the sample (often particularly the tails) that is of importance. An appropriate check is illustrated in Figure 16 where the cumulative exceedance distribution of the observed dependent variable is shown as solid points (for convenience the probability scale is represented as a standard normal deviate). The confidence limits represent the 5% and 95% limits obtained from 100 samples each of 50 items, ie a total of 5000 values were generated. It is clear that the samples conform reasonably well to the distribution of the observed data, though it would be expected that additional checks (such as flood volume over different successive periods) would need to be undertaken relevant to the design objective of interest.

7.2. Sampling from an empirical distribution

In order to illustrate sampling from an empirical distribution we will use a cumulative distribution of rainfall depth, but the approach is applicable to any empirical distribution. The relationship between exceedance probability (expressed for convenience here as average recurrence interval) and rainfall depth is shown in the first and last columns of Table 2. The standard normal deviates (the inverse of the standard normal cumulative distribution) corresponding to the tabulated exceedance probabilities are shown in the second column of the table (a function to calculate these is commonly found in commercially available spreadsheet programs). The exceedance probabilities are first converted to standard normal deviates to help linearise the function, so that the use of linear interpolation (as explained below) does not introduce unnecessary errors.

Let us assume that we need to generate rainfalls to undertake a Monte Carlo simulation. At each step we generate a random number between 0 and 1, which represents the exceedance probability of the event. If at one step the random number generated is 0.0232, then the corresponding standard normal deviate is calculated to be 1.9917. From simple linear interpolation between the second and third columns of Table 2 the rainfall depth, R, corresponding to this exceedance probability is calculated to be:

$$R = 89 + \frac{1.9917 - 1.6449}{2.0538 - 1.6449} x(100 - 89)$$

$$= 98.3 \text{ mm}$$

A similar process is used at each step of the simulation, though checks should be made to ensure that random numbers generated outside the probability limits of the supplied distribution table are appropriately handled.

Table 2 Example of empirical cumulative distribution.

Average Recurrence Interval (year)	Annual exceedance probability	Standard normal deviate	Rainfall depth (mm)
2	0.5	0.0000	63
5	0.2	0.8416	73
10	0.1	1.2816	79
20	0.05	1.6449	89
50	0.02	2.0538	100
100	0.01	2.3264	113
200	0.005	2.5758	127

7.3. Construction of a Frequency Plot

To illustrate application of a simple Monte Carlo analysis, let us assume that the stochastic sample of dependent variables listed in the last column of Table 1 are generated flood peaks (in m^3/s), and that we wish to find the probability that a peak of 210 m^3/s is exceeded. It should be stressed that in practice this approach would normally involve running many thousands of simulations, but for simplicity it is assumed here that only 50 simulations have been undertaken.

To construct a derived frequency distribution the values in the last column of Table 1 need only be ranked from largest to smallest and assigned a plotting position. Since the objective of our design is to estimate the exceedance probability of a given flood peak, we use the Weibull plotting position (Equation 1). The rank one event is 257 m^3/s, and this is assigned an exceedance probability of (1/51=) 0.01961. The second highest value (248 m^3/s) is assigned an exceedance probability of (2/51=) 0.0392, and so on for all values in the sample.

The complete plot is shown in Figure 17, where for convenience the exceedance probabilities are shown as average recurrence intervals. Inspection of this plot shows that a flood with a peak of 210 m^3/s has an average recurrence interval of around 4 years. A more accurate estimate could be obtained by fitting a smoothing function (e.g. a polynomial) to the data points closest to the flood magnitude of interest. Use of 1000 rather than 50 simulations would be expected to result in a smoother derived frequency curve, and linear interpolation between adjacent data points would then generally be sufficient.

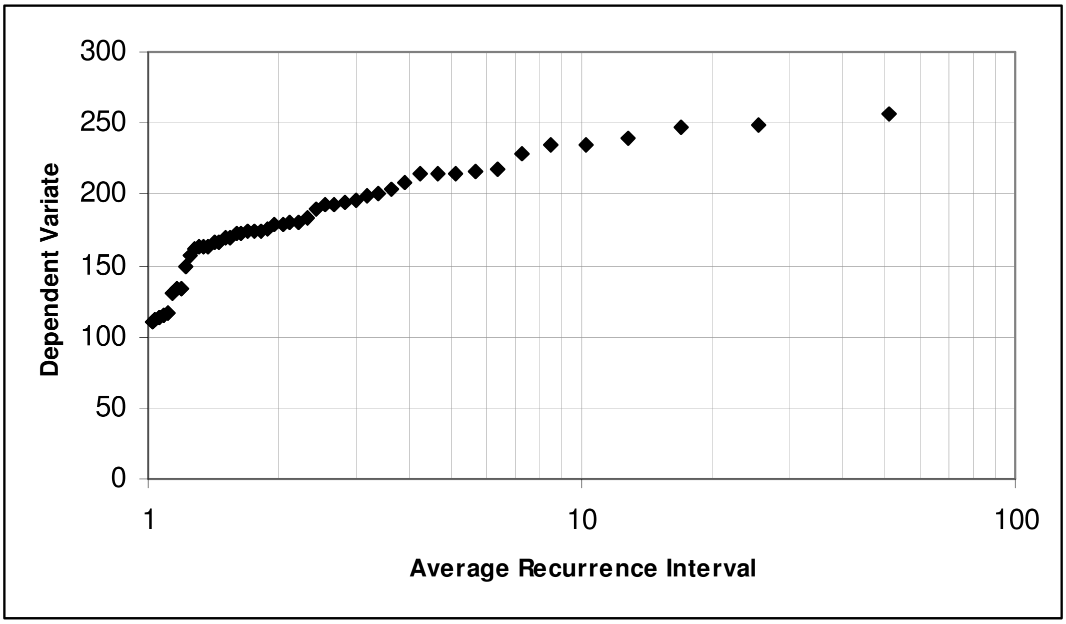


Figure 17 Example construction of a derived frequency curve.

7.4. Analysis of a Stratified Sample

The following provides an illustration of how to calculate expected probability estimates for a stratified Monte-Carlo simulation scheme. In this simplistic example it is assumed that the primary variable of interest (in this case rainfall) has been divided into ten intervals, and a flood model has been run 200 times within each interval (thus the total number of simulations undertaken is 2000). With reference to Figure 9, the number of intervals (*N*) is set to 10, and the number of simulations within each interval (*M*) is 200. Within a given interval a representative rainfall depth is selected, and the flood model is run using a stochastic sample of inputs (losses, temporal patterns, reservoir drawdown, etc) that are appropriate to the rainfall range under consideration. The results are collated, and the objective is now to estimate the probability that a flood event of a given magnitude (say, 50 m^3/s) is exceeded.

The calculations used are summarised in Table 3. The first column of this table lists the average recurrence interval (T) of the 11 bounds of the ten selected rainfall intervals, and the second column converts these to exceedance probabilities (*Y*), where $Y = 1/T$. The third column represents the probability that a rainfall occurs within the given interval, and these are simply calculated as the difference between the exceedance probabilities of the interval bounds; thus, for example, the probability that a rainfall occurs in the first interval is 0.9901-0.500=0.4901.

The fourth column of Table 3 summarises the number of times (*n*) out of the sample (*M*) of 200 that the flood model yields a flood peak greater than 50 m^3/s. As expected, it is seen that the number of times that this flood threshold is exceeded increases with rainfall depth, and the probability of this occurring in each rainfall interval is computed simply as *n/M* (as shown in the fifth column).

The flood exceedance probabilities shown in the fifth column of Table 3 are conditional upon the rainfall event occurring within the given interval, and thus estimates of the unconditional probabilities are derived by the product of the two probabilities, that is by the product of probabilities shown in columns three and five. For example, the likelihood that a rainfall with an average recurrence interval of between 2 and 5 years yields a flood peak in excess of 50 m^3/s is calculated as 0.30 x 0.01 = 0.001. The first and last intervals require special consideration as discussed in Section 4.3. Using the last interval as an illustration, the probability of the rainfall occurring within the last interval is replaced by the total exceedance probability of its lower bound, which is 0.0133. In addition, the conditional probability of 50 m^3/s being exceeded is replaced by the geometric mean of the conditional probability for the interval and 1.0; thus, the unconditional probability that a flood greater than 50 m^3/s is caused by a rainfall rarer than the "75-year" event is 0.0133x√(1.0x1.0)= 0.0133. The total probability that a flood in excess of 50 m^3/s occurs (from a rainfall depth of any magnitude) is calculated as the sum of all interval probabilities, and this is 0.0572, which is equivalent to an event with an average recurrence interval of around 17 years.

Table 3 Computational steps for calculation of expected probabilities.

Average Recurrence Interval of interval bound (T)	Cumulative exceedance probability of interval bound (Y)	Probability of interval width *p[R_i]*	Number of times in sample of 200 that event is exceeded (*n*)	Conditional probability that event is exceeded within given interval *p[Q>q|R_i]*	Probability that event is exceeded within given interval *p[Q>q|R_i]p[R_i]*
1.01	0.9901				
		0.5000*	0	0.00	0.0000
2.00	0.5000				
		0.3000	2	0.01	0.0030
5.00	0.2000				
		0.1000	10	0.05	0.0050
10.00	0.1000				
		0.0333	24	0.12	0.0040
15.00	0.0667				
		0.0167	48	0.24	0.0040
20.00	0.0500				
		0.0167	112	0.56	0.0093
30.00	0.0333				
		0.0083	170	0.85	0.0071
40.00	0.0250				
		0.0050	190	0.95	0.0048
50.00	0.0200				
		0.0067	200	1.00	0.0067
75.00	0.0133				
		0.0133*	200	1.00	0.0133
100.00	0.0100				
Expected probability that event is exceeded, p(Q>q):					0.0572

* Probability of first and last intervals represents full interval width to lower or upper bound

8. References

Box, G. E. P. and Cox, D. R. (1964): An Analysis of Transformations, *J Royal. Statistical Society*, 211-243, discussion 244-252.

Burgman, M. A. (2005): *Risks and decisions for conservation and environmental management*. Cambridge University Press, Cambridge.

Grayson, R.B., Argent, R.M., Nathan, R.J., McMahon, T.A. and Mein, R.G. (1996): *Hydrological recipes - Estimation techniques in Australian Hydrology*, Cooperative Centre for Catchment Hydrology, Monash University, 125 p.

Green, J., Walland, D., Nandakumar, N., Nathan, R. (2003) Temporal patterns for the derivation of PMP-DF and PMF estimates in the GTSM region of Australia, Proc. 28th Int. Hydrology and Water Res. Symp, Wollongong, pp. 1.97-1.104.

Griffiths, P. and Hill, I.D., Ellis Horwood (1985): *Applied Statistics Algorithms,* John Wiley and Sons, New York.

Hammersley, J.M. and Handscomb, D.C. (1964): *Monte Carlo Methods*, Methuen and Co, London; John Wiley and Sone, New York.

Hill, P.I., Mein, R.G., Weinmann, P.E. (1997): Development and testing of new design losses for South-Eastern Australia. *Proc 24th Hydrology and Water Resources Symposium Auckland*,

Hill, P.I., Sih, K., Nathan, R., Jordan, P. (2005): The extremes of tropical hydrology. *Proc ANCOLD conference*, Nov 21-22 2005, Perth.

Hill, P.I., Nathan, RJ, ?????. (in prep): Characterisation of regional loss information for probabilistic flood design.

Hoang, T.M.T., Rahman, A., Weinmann, P.E., Laurenson, E.M. and Nathan, R.J. (1999): Joint Probability Description of Design Rainfalls. Proc. Water 99 Joint Congress – Brisbane, Inst. Of Engineers, Australia, pp 379-384

Hurst, H.E., R.P Black & Y.M. Simaika (1965): *Long-Term Storage: An Experimental Study*. (Constable, London), 139pp

Ilahee, M. (2005): Modelling losses in flood estimation; thesis submitted to the School of Urban Development Queensland University of Technology in partial fulfilment of the requirements for the Degree of Doctor of Philosophy.

Jordan P, Nathan R., Mittiga L and Taylor B (2005): “Growth curves and temporal patterns for application to short duration extreme events”. *Aust J Water Resour* 9(1), 69-80.

Kuczera, G., Lambert, M., Heneker, T., Jennings, S., Frost, A., Coombes, P. (2003): Joint probability and design storms at the crossroads. Proc, 28th Int. Hydrology and Water Resour. Symp. 10-14 Nov, Wollongong.

Laurenson, E.M. (1974) Modeling of stochastic-deterministic hydrologic systems. *Water Resources Research* 10(1), 955-961.

Laurenson, E.M., Mein, R.G., and Nathan, R.J. (2005): RORB Version 5 Runoff Routing Program, User Manual, Monash University and Sinclair Knight Merz http://civil.eng.monash.edu.au/expertise/water/rorb.

Maidment, D.R. (Ed) (1993). *Handbook of Hydrology*, McGraw-Hill.

McMahon, T.A., and Mein, R.G. (1986): *River and Reservoir Yield*. Water Resources Publications, 368pp.

Meighen, J. and Minty, L.J. (1998): Temporal distribution of Large and Extreme design rainfall bursts over Southeast Australia, *Hydrology Report Series HRS Report No.* 5, Hydrometeorological Advisory Service, Bureau of Meteorology.

Minty, L.J. and Meighen, J. (1999): Development of temporal distributions of rainfall antecedent to large and extreme design bursts over southeast Australia, *Hydrology Report Series HRS Report No.* 6, Hydrometeorological Advisory Service, Bureau of Meteorology.

Nathan, R.J. "The derivation of design temporal patterns for use with the generalised estimates of Probable Maximum Precipitation", *I.E.Aust. Civil Eng. Trans.*, CE34, 2, 139-150, 1992.

Nathan, R.J., Weinmann, P.E. and Hill, P.I. (2002): Use of a Monte Carlo framework to characterise hydrological risk, *ANCOLD Bulletin* - Issue No. 122, 55-64.

Nathan, R., Weinmann, E., Hill, P. (2003) Use of Monte Carlo simulation to estimate the expected probability of Large to Extreme floods, *Proc. 28th Int. Hydrology and Water Res. Symp., Wollongong*, pp 1.105-1.112.

Nathan, R.J. and Weinmann, P.E. (2004): An improved framework for the characterisation of extreme floods and for the assessment of dam safety. *Hydrology: Science & Practice for the 21st Century, Vol 1. Proc. British Hydrol. Soc.*, London, 186-193.

Nazarov, A. (1999): The effect of initial drawdown on the risk of exceeding the spillway design floods of two dependent storages. *Proc. Water 99 Joint Congress, Brisbane, July 1999*, Institution of Engineers, Australia, 721-726.

Pilgrim DH, Cordery I, and French R. (1969): Temporal patterns of design rainfall for Sydney. *IEAust Civ. Eng. Trans.* Vol. CE11, No. 1, 9-14.

Press, W.H., Flannery, B.P, Teukolsky, S.A., Vetterling, W.T., (1989): Numerical recipes in Pascal: the art of scientific computing. Cambridge University Press.

Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P (1993): Numerical recipes in Fortran: the art of scientific computing. Cambridge University Press.

Rahman, A., Carroll, D and Weinmann, P.E. (2002[a]). Integration of Monte Carlo simulation technique with URBS model for design flood estimation. *Proc. 27th National Hydrology and Water Resources Symp., I. E. Aust.*, 20-23 May, Melbourne. CD-ROM Publication. 7 pp.

Rahman, A., Weinmann, P. E. and Mein, R.G. (2002[b]). The use of probability-distributed initial losses in design flood estimation. *Australian Journal of Water Resources*. 6(1), 17-30.

Saucier, R. (2000): Computer generation of statistical distributions. *Army Research Laboratory* ARL-TR-2168.

Schaefer MG and Barker BL (2002): Stochastic Event Flood Model (SEFM), Chapter 20, *Mathematical Models of Small Watershed Hydrology and Applications*, Water Resources Publications, LLC, VJ Singh editor.

Schaefer MG and Barker BL (2004): *Stochastic Event Flood Model (SEFM) - Users Manual*, MGS Engineering Consultants Inc, prepared for US Bureau of Reclamation Flood Hydrology Group.

Siriwardena, L. and Weinmann, P.E. (1998): A technique to interpolate frequency curves between frequent events and probable maximum events. *CRCCH Research Report* 98/9, Cooperative Research Centre for Catchment Hydrology, December 1998.

Vetterling, W.T., Teukolsky, S.A., Press, W.H. (1992): Numerical recipes example book (C). Press Syndicate of the University of Cambridge.

Vose, D. (2000): *Risk analysis: a quantitative guide*. John Wiley and Sons Ltd, Chichester.

Walland DJ, Meighen J, Xuereb K, Beesley C, and Hoang T. Revision of the generalised tropical storm method for estimating probable maximum precipitation, *HRS Report No. 8*, Hydrology Unit, Melbourne, 2003.

Wark, R.J. (1982): Flood design practice in W.A. *Proc. of the Workshop on Spillway Design*, Dept. of National Development and Energy, Australian Water Resources Council Conf.

Series No. 6, 223-233.

Weinmann, P.E. and Nathan, R.J. (2004): The continuing challenge of estimating extreme floods from extreme design storms. *Advances in Hydro-Science and Engineering, Volume VI, Proc. of the 6th Intern. Conf. on Hydro-Science and Engineering*, Brisbane May 31 – June 3, 2004.

Weinmann PE, Rahman A, Hoang TMT, Laurenson EM, Nathan RJ. (2002): Monte Carlo simulation of flood frequency curves from rainfall - the way ahead. *Aust J of Water Resources*, IEAust, 6(1): 71-80.