

**Theme:** Big-data, knowledge and water data management

**Subtheme:** C1 Data-mining techniques

## **A Classic Hydroinformatic Problem - Floods**

**James E Ball<sup>1</sup>**

<sup>1</sup> School of Civil and Environmental Engineering, University of Technology Sydney, Australia

*Corresponding author: james.ball@uts.edu.au*

**Abstract.** While previous editions of ARR have served the engineering profession well, a number of issues have necessitated the production of a new edition. These issues include the many recent developments in knowledge about flood producing processes, the increased computational capacity and data manipulation available to engineers, and the rapidly expanding body of information about climate change. There is a need, therefore, to produce a new edition of ARR. As part of the development of this new edition, it has been necessary to review the methods used and the implications of assumptions necessary for implementation of these methods. An outcome of this review has been recognition of changes in design flood estimation since development of the digital computer and the subsequent development of hydroinformatics. This has led to recasting design flood estimation as a problem in hydroinformatics. Presented herein is the background to and a discussion of this concept.

**Keywords:** Data, Design, Floods

### **1 Introduction**

Design flood estimation remains a problem for many professionals involved in the management of rural and urban catchments. Advice is required regarding design flood characteristics for many design problems including the design of culverts and bridges necessary for cross drainage of transport routes, the design of urban drainage systems, the design of flood mitigation levees and other flood mitigation structures, design of dam spillways, and many environmental flow problems.

While the flood characteristic of most importance depends on the nature of the problem under consideration, but typically it is one of the following:

- Flood flow rate – typically, it is the peak flow rate of the flood hydrograph that is the desired design flood hydrograph characteristic;
- Flood level – similar to the flood flow rate, it is the peak flood level during the flood hydrograph that is the commonly desired design flood hydrograph characteristic;
- Flood rate of rise – this design flood characteristic is a concern when planning is undertaken for operational floods;
- Flood volume – this design flood characteristic becomes a concern when the design flood volume is a major factor in the design problem. This situation occurs when storage of a significant portion of a flood hydrograph is used as part of a flood management system; or
- System failure – the usual design flood problem is located at a single point. There are a number of design problems, however, where the issue becomes one of multiple points within a system. Typical examples of these problems include urban drainage systems where the individual components of the system are not statistically independent which is a common assumption, and transportation routes with multiple cross drainage structures of one or more river systems.

Design objectives are most commonly specified using risk-based criteria, and thus the focus is on the use of methods that provide estimates of flood characteristics for one or more specified Annual

Exceedance Probabilities (AEP). Within Australia, Australian Rainfall and Runoff [12] is the primary guideline document for prediction of design floods. While the 1987 edition of Australian Rainfall and Runoff and its predecessors served the engineering profession and the general community well, a number of developments have arisen that necessitate the production of a new edition. These developments include the many recent advances in knowledge regarding flood processes, the increased computational capacity available to the profession, and the rapidly expanding body of information about climate change and the influence of potential future climate states on the relationship between flood magnitude and flood frequency. Consideration of these developments resulted in the decision by Engineers Australia to prepare another edition of Australian Rainfall and Runoff [2].

An outcome of this revision has been recognition of changes in approaches to design flood prediction and the relevance to design flood estimation of hydroinformatics. This has led to recasting design flood prediction as a problem in hydroinformatics. The purpose of this paper is to provide an overview of the concept that design flood estimation is a problem in hydroinformatics.

## 2 Approaches for Estimation of Design Flood Quantiles

### 2.1 General

A fundamental basis of design flood estimation is that the desired flood characteristics need to be interpreted from a statistical viewpoint. This need arises from design objectives being specified in a risk-format; as previously noted, common design objectives are the estimation of the flood level or flood flow for a given AEP. This statistical interpretation contrasts with analysis of flood characteristics for a historical event where a deterministic viewpoint is appropriate.

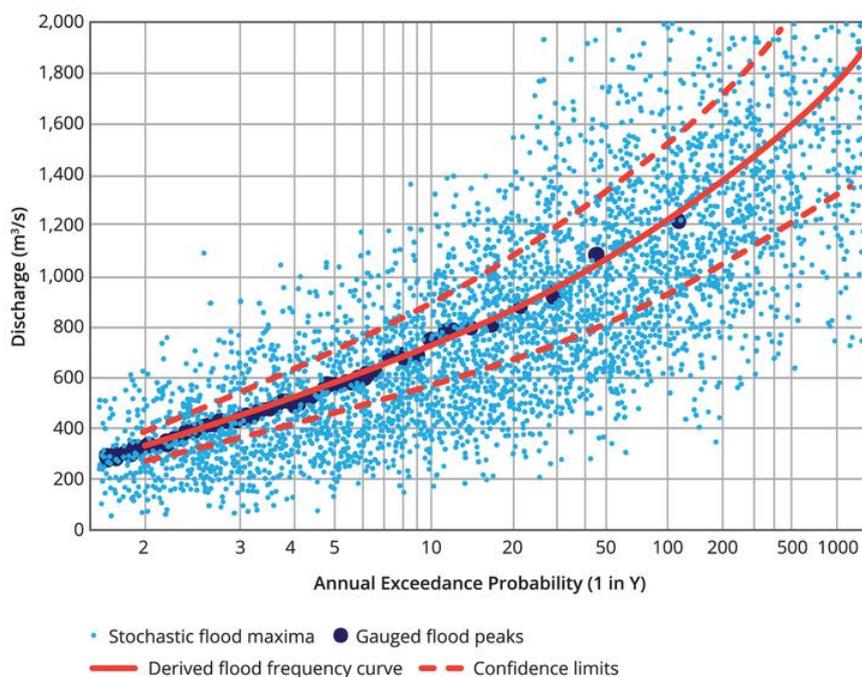


Figure 1. Stochastic Nature of Design Flood Estimation (after [2])

The general nature of the estimation problem is illustrated in Figure 1 where the recorded flows, and the resultant flood probability model and its confidence limits are shown. The computed confidence limits are conditioned on the assumed underlying probability model and it needs to be recognised that these confidence limits only represent the uncertainties most easily characterised. Also shown in this figure are the floods derived from 10,000 years of rainfall data with an assumed stationary climate. These flood peaks were developed through

simulation considering the many random natural and anthropogenic influences on flood generation. For development of Figure 1, the flood peaks are plotted against the AEP of the causative rainfall.

As implied through consideration of Figure 1, there are two alternative situations when design flood characteristics are required; these are:

- Historical data is available; and
- Historical data is not available.

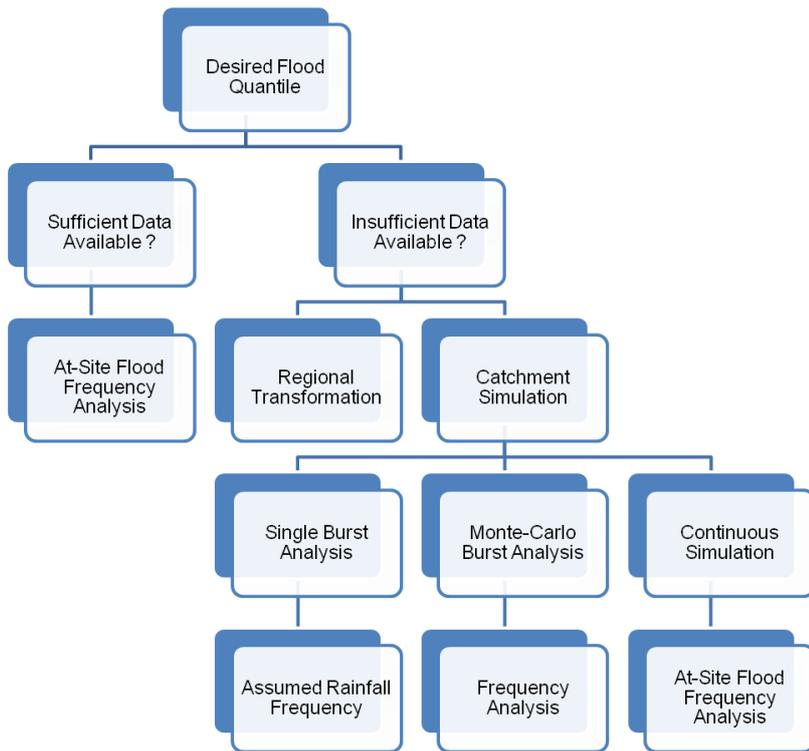


Figure 2 Design Flood Estimation Approaches

location in the catchment.

An important component of undertaking an FFA is the data used for identification of the flood probability model. As the true flood probability model is unknown, the choice of a model often is guided by the fit between the model and the available data; in other words, the flood probability model is calibrated to the available data.

Calibration of the flood probability model will be influenced by the quality of the data available. Data quality related to design flood estimation is discussed by [1]. They found, from analysis of 477 gauging stations in NSW that:

- Extrapolation required to convert recorded stages to flows at 16% of the gauging stations exceeded an order of magnitude;
- At 19% of the gauging stations, over 50% of the AMS data points were in the extrapolation zone of the stage-discharge relationship; and

A framework for consideration of these alternative situations and the alternative approaches for generation of data is shown in Figure 2. An important aspect of the flood data is the quality of the data available for analysis; it is possible to have data available but the quality of the data to be inadequate for the proposed analysis.

## 2.2 Flood Frequency Analysis

As shown in Figure 2, where sufficient historical data is available, estimation of the desired flood parameter can be undertaken using at-site flood frequency analysis (FFA) methods (see, for example, [7], and [8]). FFA approaches are procedures that use recorded and related flood data to identify the underlying probability model of flood peaks at a given

- Extrapolation for out-of-bank flows at 76% of the gauging stations was based on gauging data collected only for channel flows.

Consideration of these outcomes illustrates the importance of data management if extraction of the information content in the available historical data is to be maximised.

From a hydroinformatic viewpoint, the implementation of an FFA approach is premised on data-mining the available historical data to enable prediction of the likelihood of hazard magnitudes. This data-mining is achieved by calibrating a statistical model to the available historical data. As the calibration of the statistical model is influenced by the reliability of the available historical data which is a data management problem and a fundamental component of all hydroinformatic problems.

### 2.3 Regional Flood Methods

Where adequate historical data is not available, it is necessary to use approaches that generate relevant data for estimation of the design quantile. Techniques can be divided into two categories, namely Regional Transformation techniques (sometimes referred to as Regional Flood Methods) and Catchment Simulation techniques (sometimes referred to as rainfall-based techniques).

The philosophical basis of Regional Flood Methods (RFMs) is the transfer of flood frequency characteristics from a group of gauged catchments to the location of interest. Hence, these methods are examples of data-driven approaches.

According to [4], most common methods applied for development of RFMs include various forms of the Rational Method (see, for example, [10]), the Index Flood Method, and Regression based techniques (see, for example, [6] and [11]). Irrespective of the development methodology employed, the development process consists of:

- (i) Formation of Regions – this step involves identification of regions where the flood data from the available streamflow gauging stations can be pooled for analysis; and
- (ii) Development of Regional Estimation Equations – this step involves derivation of the prediction equations to be used within a region; depending on the method employed, this step may involve development of the equation form or just the equation parameters.

An important step in the development of any RFM is the formation of the regions from which the desired flood characteristics will be extracted. These regions can be based on geographic proximity or on similarity in catchment attributes. Additionally, a region can be fixed, having a definite common boundary for all sites within it, or it can be formed around the ungauged catchment of interest (i.e. the location where flood quantile estimation is desired), using the nearest stations in geographic or catchment attributes space. Irrespective, regions must satisfy explicitly or implicitly the assumption of ‘*regional homogeneity*’.

The decision on what constitutes a homogeneous region for the purposes of Regional Flood Frequency Estimation depends on the methods used, more specifically on the extent to which differences in flood characteristics can be expressed through parameters in the regionalisation method. Nonetheless, identification of homogeneous regions is premised on data-mining the available historic flood data and the catchments where that data has been collected.

The second step in development of a RFM is the development of the prediction equations. While this step is variable in the approaches employed, a minimum requirement is the identification of parameters for one, or more, equations. Calibration of the parameters for these equations is required to ensure suitability of the predictions. Like fitting the statistical model in FFA, this calibration requires data-mining of the available historic data.

As outlined in the preceding discussion, development of a RFM is premised on data-mining the historical flood data and the catchment characteristics to ensure reliable transfer of flood characteristics. Development of suitable and reliable data-mining approaches for the identification of hydrologically similar regions, the prediction equations and their parameters is a hydroinformatic problem.

## **2.4 Catchment Modelling**

The third approach for prediction of the desired design flood quantile is the use of catchment simulation to generate the data that would have been recorded if catchment monitoring had been in place for the event, or sequence of events, being considered. When this approach is employed to generate the information, there are numerous alternative approaches that can be employed inclusive of:

- Single Burst (either the peak burst or a total event) techniques;
- Monte-Carlo techniques (see, for example, [14]); or
- Continuous simulation techniques.

In general, application of catchment simulation from any one of these three categories requires an analysis of the resultant predictions for estimation of the quantiles of the resultant flood characteristic. The form of the post-modelling analysis, however, is dependent on the form of modelling undertaken; for example, the post-modelling analysis for many approaches using single burst modelling consists of the assumption that the rainfall probability is transferred to the flood probability. This is the AEP Neutral assumption implicit in many guidelines; see, for example, [2].

Nonetheless, the data generated through application of catchment simulation approaches is used to develop knowledge about the magnitude and probabilities of the flood characteristic of interest.

As noted above, the premise of a catchment model is the prediction of data that would have been recorded at the location (or locations) of interest for the desired catchment conditions if catchment monitoring was operational at that location (or locations). The desire of the catchment model, therefore, is that the physical processes influencing the transformation of rainfall to flood runoff are adequately represented by the catchment model; in other words, the responses of the physical processes in the true catchment are replicated by mathematical models in the catchment model.

This introduces two issues in the application of the catchment model; these issues are the mathematical representation of processes and the parameters necessary for solution of these mathematical representations. The first of these issues is related to the suitability of the selected mathematics for the catchment being simulated; this is an issue that can be addressed only by the modeller. The second issue can be recast as the calibration issue. While calibration of a catchment model is conceptually similar to the calibration necessary for application of FFA and development of a RFM, there are significant differences in implementation. These differences arise from the need to transform the input rainfall into the output flood runoff.



### 3 Conclusions

Within Australia, Australian Rainfall and Runoff [2] is the primary guideline document used for prediction of design flood quantiles. While the 1987 edition of Australian Rainfall and Runoff and its predecessors have served the engineering profession and the general community well, in the period since the release of the previous edition, developments have occurred leading to the decision by Engineers Australia to prepare another edition of Australian Rainfall and Runoff[2]. As part of this revision, the problem of estimating design flood quantiles has been recast as a problem in hydroinformatics.

Analysis of approaches and techniques used for design flood estimation highlighted the conceptual consistency with the definition of a hydroinformatic problem and the focus of the alternative approaches and techniques. This consistency confirms the proposal that design flood estimation is a problem in hydroinformatics.

### References

- [1] Ball, JE, (2016), Management of gauged data and its impact on the prediction of design flood flows, Proceedings IAHR-APD Colombo, Sri Lanka
- [2] Ball, JE, Babister M, Nathan R, Weeks W, Weinmann E, Retallick M, Testoni I, (Editors), (2016), Australian Rainfall and Runoff: A Guide to Flood Estimation, Commonwealth of Australia, ISBN 978-192529-7072
- [3] Ball, JE and Luk, KC, (1998), Modelling the spatial variability of rainfall over a catchment, ASCE, Journal of Hydrologic Engineering, 3(2):122-130.
- [4] Bates, B.C. (1994), 'Regionalisation of hydrological data: A review', Report 94/5, CRC for Catchment Hydrology, Monash University
- [5] Freni, G, Mannina, G, and Viviani, G, (2009), Uncertainty assessment of an integrated urban drainage model, Journal of Hydrology, 373(3-4):392-404
- [6] Haddad, K and Rahman, A, (2011), Regional flood estimation in New South Wales Australia using Generalised Least Squares Quantile Regression, ASCE, Journal of Hydrologic Engineering, 16(11):920-925.
- [7] Jin, M and Stedinger, JR, (1989), Flood frequency analysis with regional and historical information, Water Resources Research, 25(5):925-936.
- [8] Kuczera, G, (1999), Comprehensive at-site flood frequency analysis using Monte Carlo Bayesian inference, Water Resources Research, 35(5):1551-1558.
- [9] Kuczera, G, Kavetski, D, Franks, S, and Thyer, M, (2006), Towards a Bayesian Total Error Analysis of Conceptual Rainfall-runoff Models: Characterising Model Error using Storm-dependent Parameters, *Journal of Hydrology* , 331(1-2):161-177
- [10] Mittelstadt, GE, McDermott, GE and Pilgrim, DH, (1987), Revised flood data and catchment characteristics for small to medium sized catchments in New South Wales, Unpublished Master of Engineering Science Report, Department of Water Engineering, School of Civil Engineering, The University of New South Wales, Sydney, NSW.
- [11] Palmen, L.B. and Weeks, W.D. (2011), Regional flood frequency for Queensland using the quantile regression technique, Australian Journal of Water Resources, Volume15(1) pp.47-57.
- [12] Pilgrim, DH, (1987), Australian Rainfall and Runoff: A guide to flood estimation, The Institution of Engineers Australia, Barton, ACT.
- [13] Umakhanthan, K and Ball, JE, (2005), Rainfall Models for Catchment Simulation, Australian Journal of Water Resources, 9(1):55-67.
- [14] Weinmann, PE, Rahman, A, Hoang, TMT, Laurenson EM, and Nathan RJ, (2002), Monte Carlo simulation of flood frequency curves from rainfall - the way ahead, Australian Journal of Water Resources, IEAust, 6(1):71-80.