Progress in Physical Geography 29, 3 (2005) pp. 392–410

Flood frequency analysis: assumptions and alternatives

R. Kidson[1,*] and K.S. Richards[2]

[1]Trinity College, Cambridge CB2 1TQ, UK
[2]Department of Geography, University of Cambridge, Downing Place, Cambridge CB2 3EN, UK

Abstract: Flood frequency analysis (FFA) is a form of risk analysis, yet a risk analysis of the activity of FFA itself is rarely undertaken. The recent literature of FFA has been characterized by: (1) a proliferation of mathematical models, lacking theoretical hydrologic justification, but used to extrapolate the return periods of floods beyond the gauged record; (2) official mandating of particular models, which has resulted in (3) research focused on increasingly reductionist and statistically sophisticated procedures for parameter fitting to these models from the limited gauged data. These trends have evolved to such a refined state that FFA may be approaching the 'limits of splitting'; at the very least, the emphasis was shifted early in the history of FFA from predicting and explaining extreme flood events to the more soluble issue of fitting distributions to the bulk of the data. However, recent evidence indicates that the very modelling basis itself may be ripe for revision. Self-similar (power law) models are not only analytically simpler than conventional models, but they also offer a plausible theoretical basis in complexity theory. Of most significance, however, is the empirical evidence for self-similarity in flood behaviour. Self-similarity is difficult to detect in gauged records of limited length; however, one positive aspect of the application of statistics to FFA has been the refinement of techniques for the incorporation of historical and palaeoflood data. It is these data types, even over modest timescales such as 100 years, which offer the best promise for testing alternative models of extreme flood behaviour across a wider range of basins. At stake is the accurate estimation of flood magnitude, used widely for design purposes: the power law model produces far more conservative estimates of return period of large floods compared to conventional models, and deserves closer study.

Key words: extreme floods, flood frequency analysis, power law, self-similar.

I Introduction

Flooding is one of the most pervasive natural hazards to impact negatively upon the activities of human beings, and requires various responses including construction (downstream flood defences), forecasting (for warning and evacuation), and land-use management (upstream catchment-scale changes of land use and runoff characteristics). To some degree, all of these require

*Author for correspondence. Current address: Corporate Licensing Unit, Metropolitan Water Sharing Directorate, Office of the Metropolitan Region, NSW Department of Infrastructure, Planning and Natural Resources (DIPNR), PO Box 651, Level 1, 308 High Street, Penrith, NSW 2751, Australia

 10.1191/0309133305pp454ra

hazard or risk assessment in the form of a flood frequency analysis (FFA) (Dunne and Leopold, 1978). This is traditionally most evident in the design of major engineering structures such as dams, flood embankments and bridges, where the accuracy of these methods has a profound significance for economic investment (Yixing *et al.*, 1987). It is common for a 'standard of service' to be defined for a flood defence structure, which is often to provide protection against an event with a specified return period, such as the 100-year flood. Since discharge data for most catchments have been collected for periods of time significantly less than 100 years, the estimation of the 'design discharge' (design stage, or water level) necessarily requires a degree of extrapolation, which in turn demands curve-fitting to the existing data.

All methods of FFA are thus methods of extrapolation. Extrapolation requires the fitting of a model, and here a limitation of FFA is apparent: the fitting of any model requires an *a priori* assumption about the underlying distribution generating flood events. Not only is this not known for extreme hydrological events beyond the observed record, but it is untestable within human timescales (Klemes, 1989). Yet, a model must be fitted, if predictions are to be made. The literature is dominated by one particular approach to modelling: the use of a range of more-or-less skewed, relatively complex, and often theoretically unjustified probability distributions, whose parameters are estimated from the data in the observational record. However, a recent alternative has been to examine the use of simple power law (PL) models which carry theoretical implications about self-similarity in the distribution of flood magnitudes (Malamud and Turcotte, 1999).

There are, however, some additional tools for effectively extending the instrumented gauging record. These include (1) using physically based models including rainfall-runoff modelling in continuous simulation mode (Beven, 1987); (2) combining data from several gauges in a regionalization exercise (GREHYS, 1996) and (3) incorporating historical and palaeoflood information into the instrumented record (House *et al.*, 2002; Baker, 2003). This review will focus on the implications of the third of these methods, and in particular will consider the use of historical and palaeoflood data and its implications for the two strategies of curve-fitting noted above. The more recent use of PL distributions has in part been justified when pre-instrumental flood evidence suggests that the most extreme events are outliers relative to the conventionally fitted curves.

II The mechanics of flood frequency analysis

FFA model fitting involves three main steps: data choice, model choice and a parameter estimation procedure (Bobee, 1999). These intricacies are discussed here for two reasons: (1) to illustrate the proliferation of techniques at each step, and (2) in order to highlight the uncertainty consequences of reliance upon any one technique at each step. Yen 2002: proposes a very useful scheme in which to consider uncertainty (Table 1). In this scheme, the modelling process is presented as a hierarchy. Progression along the modelling process entails assumptions at each step of the process, and each of these steps thus has the capacity to introduce error. If unquantified, this error may cascade down

Table 1 Yen's (2002) hierarchical uncertainty scheme

	Uncertainty type	Sensitive to
1	Natural uncertainty	Nonstationary conditions
2	Model uncertainty	Choice of model
3	Parameter uncertainty	Fitting technique; goodness-of-fit test
4	Data uncertainty	Data choice; accuracy of observed/gauged data
5	Operational uncertainty	Human errors/decisions

into the model predictions; this can have dire consequences in the case of FFA. Awareness of these issues is not often demonstrated in the FFA literature; yet, challenging some of these assumptions represents the best opportunity for advancement of the science of FFA. This paper addresses points (2)–(4) of Yen's scheme.

1 Data choice

Classically, FFA is performed on the series of annual maximum discharge values recorded at a specific river gauging section. However, an alternative is to use a form of partial series, the most common being the peaks over threshold (POT) method (Hosking and Wallis, 1987), where every event over a given threshold is included in the analysis (Figure 1). POT is now a cornerstone technique for the incorporation of historical and palaeoflood data, since it offers inclusion of the continuous instrumented gauged record and the discrete historic data within the same modelling framework. However, in conventional FFA, it is common practice to apply both methods in order to determine the difference that data choice decisions make for prediction. In particular, it is known that results are sensitive to choice of threshold (e.g., Adamowski *et al.*, 1998; Tanaka and Takara, 2002), with greater confidence when the threshold is close to the discharge for which a return period estimate is required (Hosking and Wallis, 1986a). Where it is possible, a reliable scheme of investigation would include a range of thresholds in order to determine this sensitivity.

2 Model choice I

A limited number of models have traditionally been employed in the analysis of flood peak

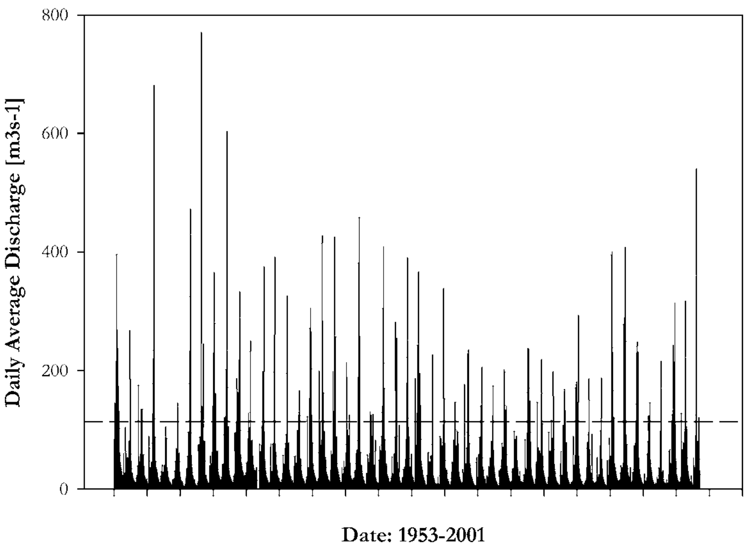

Figure 1 A partial duration series (PDS)/peaks over threshold (POT) approach applied to daily discharge records for the 49-year gauged record of the Mae Chaem river, northern Thailand. The minimum peak annual discharge over this period is 114 m^3s^{-1} (dashed line), and this forms the threshold for inclusion of flow peaks in the PDS (peaks must also be independent; a one-month separation of peaks is conventionally used as an eligibility criterion). In the case of these Mae Chaem records, this yields a PDS with n = 78

data for FFA, but there are a large number of variants on these. The simpler class of these conventional FFA models consists of 2-parameter functions that can be fitted analytically, such as the log normal and Gumbel extreme value type 1 (Gumbel EV1) double-exponential model (see Table 2). The two parameters represent location and shape, and the mean and variance of the sample population are employed (e.g., the annual series of discharges for a gauging station) with the method of moments (see below). The log Pearson type III (LP3) and generalized extreme value (GEV) models belong to a class of 3-parameter models which cannot be fitted analytically. The former is based on the gamma distribution. In both cases the three parameters represent location, shape and scale, and again, when the method of moments is used, are based broadly on the mean, variance and skewness of the distribution. The 2-parameter models have the advantage of simplicity and ease of fit; however the 3-parameter models, with the additional scale parameter, are regarded as having the flexibility to fit a larger number of catchments' records (NERC, 1999). Consideration of scale is also an acknowledgement that most annual series are significantly skewed.

Some lesser-known models include members of the gamma distribution family (e.g., the bivariate gamma distribution; Yue, 2001), the Weibull (Heo *et al.*, 2001a; 2001b), the extreme value family (e.g., trivariate extreme value; Escalante-Sandoval and

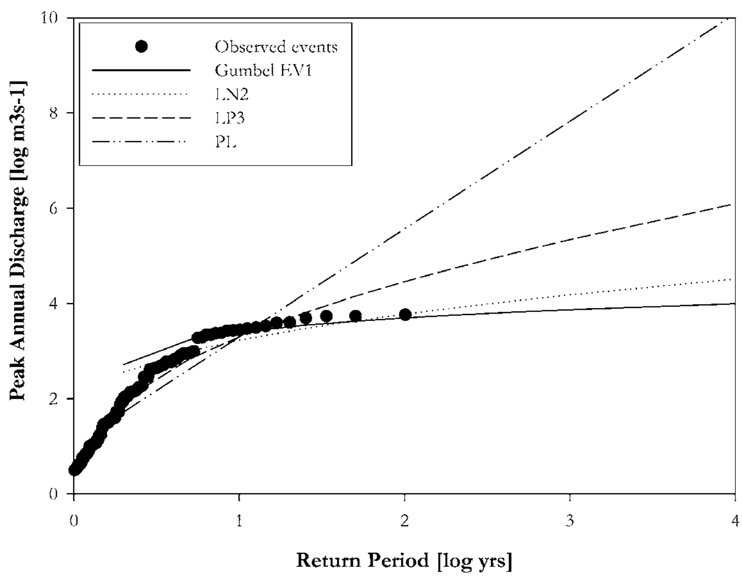

Figure 2 Four FFA models fitted to a synthetically generated record of 'annual floods' randomly sampled from a log normal distribution, and plotted in double-logarithmic space. Three conventional models are shown: the Gumbel EV1, the two-parameter log normal (LN2) and the log Pearson type III (LP3). All these distributions were fitted using the method of moments; the LP3 was fitted with the Water Resources Council 1981: skew coefficient. Also shown is a power law (PL) model fitted using simple least squares regression to the observed data. The three conventional distributions exhibit a concave-down shape in log-log space, contrasting with the PL which is a straight line

Table 2 Probability density functions (pdfs) and/or cumulative density functions (cdfs) for a selection of FFA distributions (adapted from Stedinger *et al.*, 1993, and Ramachandra Rao and Hamed, 2000)

Normal	$f(x)=\frac{1}{\sqrt{2\pi\sigma_x^{\ 2}}}\exp\left[-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2\right]$ pdf
Log normal (2-parameter) (LN2)	$f(x)=\frac{1}{x\sqrt{2\pi\sigma_y^{\ 2}}}\exp\left[-\frac{1}{2}\left(\frac{\ln(x)-\mu_y}{\sigma_y}\right)^2\right]$ pdf y = ln(*x*)
Pearson type 3	$f(x)=\lvert\beta\rvert\left[\beta(x-\epsilon)\right]^{\alpha-1}\frac{\exp\left[-\beta(x-\epsilon)\right]}{\Gamma(\alpha)}$ pdf $\Gamma(\alpha)$ is the Gamma function
Log Pearson type III (LP3)	$f(x)=\lvert\beta\rvert\left\{\beta\left[\ln(x)-\epsilon\right]\right\}^{\alpha-1}\frac{\exp\left\{-\beta\left[\ln(x)-\epsilon\right]\right\}}{x\Gamma(\alpha)}$ pdf
Exponential	$f(x)=\beta\exp\left[-\beta(x-\epsilon)\right]$ pdf $F(x)=1-\exp\left\{-\beta(x-\epsilon)\right\}$ cdf
Gumble EV1	$f(x)=\frac{1}{\alpha}\exp\left[-\frac{x-\epsilon}{\alpha}-\exp\left(-\frac{x-\epsilon}{\alpha}\right)\right]$ pdf $F(x)=\exp\left[-\exp\left(-\frac{x-\epsilon}{\alpha}\right)\right]$ cdf
Generalized extreme value (GEV)	$F(x)=\exp\left\{-\left[1-\frac{\kappa(x-\epsilon)}{\alpha}\right]^{1/\kappa}\right\}$ cdf
Weibull	$f(x)=\left(\frac{k}{\alpha}\right)\left(\frac{x}{\alpha}\right)^{k-1}\exp\left[-\left(\frac{x}{\alpha}\right)^k\right]$ pdf $F(x)=1-\exp\left[-\left(\frac{x}{\alpha}\right)^k\right]$ cdf
Generalized Pareto (GP)	$f(x)=\left(\frac{1}{\alpha}\right)\left[1-\kappa\frac{(x-\epsilon)}{\alpha}\right]^{1/\kappa-1}$ pdf $F(x)=1-\left[1-\kappa\frac{(x-\epsilon)}{\alpha}\right]^{1/\kappa}$ cdf

(*continued*)

Table 2 *continued*

Generalized logistic (GL)		$f(x)=\frac{1}{\alpha}\left[1-k\left(\frac{x-\epsilon}{\alpha}\right)\right]^{\left(\frac{1}{k}-1\right)}\left[1+\left\{1-k\left(\frac{x-\epsilon}{\alpha}\right)\right\}^{\frac{1}{k}}\right]^{-2}$ pdf $F(x)=\left[1+\left\{1-k\left(\frac{x-\epsilon}{\alpha}\right)\right\}^{\frac{1}{k}}\right]^{-1}$ cdf
Power law (PL)	$f(x)=Cx^{-\alpha}$	

Raynal-Villasenor, 1994), and the logistic distribution family (Ahmad *et al.*, 1988). The generalized pareto distribution (GDP) is a specialized model for POT-type data (Hosking and Wallis, 1987), and is based on the Poisson (binomial) distribution. Examples of the more commonly used models are provided in Table 2 and illustrated in Figure 2.

In several instances, a national standardization of the approach to FFA is enforced; this may be to provide the semblance of objectivity, and protection against legal liability. The USA is a notable case, where the LP3 has been the official model since 1967 (NRC, 1988), to which data from all catchments are fitted for planning and insurance purposes. There is a correspondingly large body of research concerning this model (e.g., Vogel and McMartin, 1991; Wu-Boxian *et al.*, 1991; Fortin and Bobee, 1994; Ouarda and Ashkar, 1998; Chen *et al.*, 2002). By contrast, the UK endorsed the GEV1 distribution (NERC, 1999) up until 1999; the official distribution in this country is now the generalized Logistic (GL). There are several instances where a number of alternative models have been evaluated for a particular country, for example Kenya (Mutua, 1994), Bangladesh (Karim and Chowdhury, 1995), Turkey (Bayazit *et al.*, 1997) and Australia (Vogel *et al.*, 1993). One consequence of the adoption of a single standard model is the 'one size fits nobody' scenario, where optimal fitting to a specific catchment is precluded by a (possibly unsuited) national model–this may result in mediocre accuracy on flood prediction in any given basin.

Given the confusing range of models available, it is perhaps not surprising that there have also been calls to adopt an international standard for FFA (Bobee *et al.*, 1993). However, FFA results are obviously influenced by model choice, and the best means of quantifying this is through comparing a number of models (Haktanir and Horlacher, 1993). The proliferation of models is itself symptomatic of the weak theoretical basis in hydrology for the application of these FFA models. Chow 1954: presents perhaps the only convincing, hydrologically plausible theoretical argument justifying the application of a particular model for FFA. Chow demonstrated mathematically that multiplicative random natural events would generate a log normal distribution. Gumbel's extreme value theory (EVT: Gumbel, 1958) and this family of FFA models have a theoretical basis in closed-system statistical assumptions that may or may not be hydrologically valid. For the remaining models commonly applied, and most notably the official distributions in the USA and UK, theoretical justifications are weakly articulated, if at all, and theoretical considerations are rarely employed as a legitimate criterion to guide model selection in FFA. In the case of the officially mandated models, their original selection was based almost exclusively on the basis of empirical goodness-of-fit to the events of record: it is implicitly assumed that the best fit will provide the most reliable extrapolation.

Classically, the appropriateness of a given model is assessed via a goodness-of-fit test (e.g., Chowdhury *et al.*, 1991) and this is another limitation of FFA: different goodness-of-fit tests will favour different models, and not all models can be evaluated by the same test. NERC (1999) demonstrate a classic example where different goodness-of-fit measures (e.g., absolute error and least squares error) favour alternative models. The FFA analyst, therefore, must be aware of sensitivity to choice of goodness-of-fit test, and it is useful to employ several of these to test for sensitivity (e.g., Stedinger *et al.*, 1993). None of this supports the view that one of these traditional models deserves to be adopted as a universal choice.

3 Parameter estimation procedure

Having selected an *a priori* model, the next step is to identify the parameters required to fit the model to the selected data. This has usually been achieved using the method of moments (MOM), which is based on the statistical moments (i.e., mean, variance, skew) of the sample data. An alternative is the L-moment method (Hosking, 1990), now having wide application in regional analyses (Adamowski, 2000). L-moments define the characteristics of the sample discharge data based on combinations of the difference between two randomly selected events in the data. A third parameter-fitting technique is maximum likelihood (ML). This is a non-analytical technique where parameters are optimised via a search through parameter space, hence it is computationally demanding. ML offers a consistent, objective technique for parameter fitting where the analysis includes different types of data, and is particularly relevant for FFA where a systematic record is combined with categorical data (in the form of historical POT numbers) (Stedinger and Cohn, 1986).

A particular advantage of ML methods is that they can be applied to probability density functions (pdfs) that are multimodal or otherwise complex (i.e., without a strong central tendency, or deriving from the combination of events drawn from different distributions). ML has the flexibility to deal with information/data that deviate from statistical assumptions of normality (e.g., O'Connell *et al.*, 2002). As this characterizes much hydrological data, ML is a valuable tool. Furthermore, it is an approach that can assist in defining confidence limits on discharge or return period estimates. However, it is sensitive to the choice of goodness-of-fit test chosen for the optimization, and it is not possible to arrive at numerical solutions for all cases.

III Model choice: an alternative

The PL model (Table 2 and Figure 3) offers a simple alternative to the more complex probability models discussed above. The evidence supporting PL distributions for extreme natural events is growing, with examples including earthquakes, volcanic eruptions, landslides, avalanches and forest fires (see Turcotte, 1994; Scheidegger, 1997; Birkeland and Landry, 2002). Recent work in other fields of hydrology also demonstrates PL relationships for specific ranges of rainfall data (Gupta and Waymire, 1990; Hubert, 2001) and catchment runoff (de Michele *et al.*, 2002). A PL fit implies that discharge ratios are the same (hence 'self-similar') for given return period ratios, at least for a certain range of discharge magnitudes (scales), e.g., Q_{100}/Q_{10} equals Q_{10}/Q_{1}, where the scaling exponent is constant over this range. Self-similarity offers a plausible theoretical basis for flood frequency analysis.

A prominent application of self-similarity to FFA was presented by Malamud *et al.*, (1996), who demonstrated a close fit of the discharge/recurrence interval relationship of the extreme 1993 event on the Mississippi river with a PL model. Similar close fits were demonstrated with historic and palaeoflood data for the Colorado river. This was a

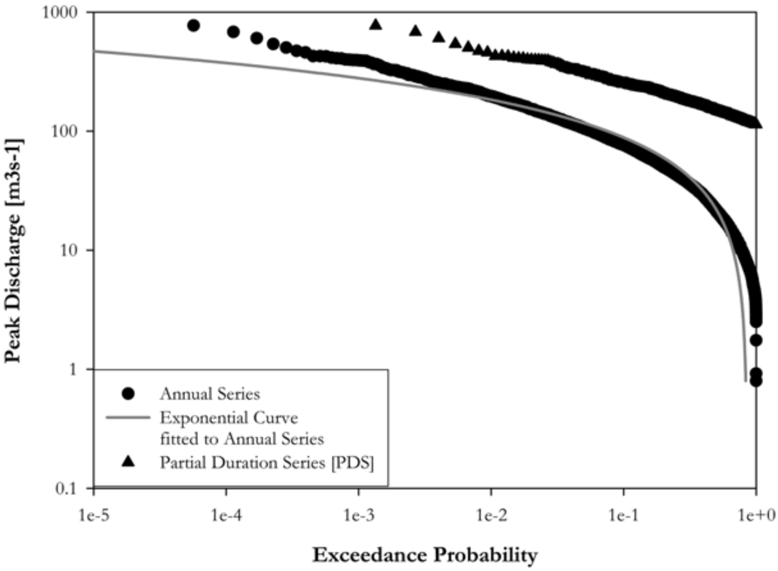

Figure 3 Power law tail behaviour in daily discharge data. For the annual series, there is a straight line divergence in plotting position of the observed events with low probability, relative to the tail of an exponential curve fitted to the same data. Also shown is the partial duration series (PDS). Use of the PDS tends to remove much curvature in space at high exceedance probabilities, relative to annual series plots, and often provides more compelling evidence of straight line (i.e., power law) behaviour in log-log space

significant outcome, as the PL model together with historic (pre-instrumental) flood data predict substantially larger discharges at a given recurrence interval. If a PL model is indeed the underlying probability model for flood behaviour on these rivers, then the estimated discharge of the industry-standard 100-year flood requires revision, with planning and design decisions needing to be correspondingly more conservative (Figure 4).

Improving the theoretical basis and documenting additional case studies of PL behaviour constitute important areas for future research. Alila and Mtiraoui 2002: assert that the selection of the most plausible distribution for flood frequency analysis should be based on hydrological reasoning as opposed to the sole application of the traditional statistical goodness-of-fit tests. If goodness-of-fit criteria, as applied to short-record annual series data, are not regarded as the best test of the applicability of a flood frequency model, this leaves few grounds for accepting any model *a priori* for a new catchment– unless there is a firmly established, generalized theoretical basis. A critical outstanding question in assessing the plausibility of the PL model is the scale at which PL behaviour applies; flood magnitudes over a period of, say 10 years may be generated by processes distinct from those which produce the 100-year or 1000-year floods, nonstationarity considerations notwithstanding.

IV Mixed generating mechanisms

It is notable in Figure 4 that, while the PL distribution fits the extremes well, it fails to fit the lower range of annual flood maxima.

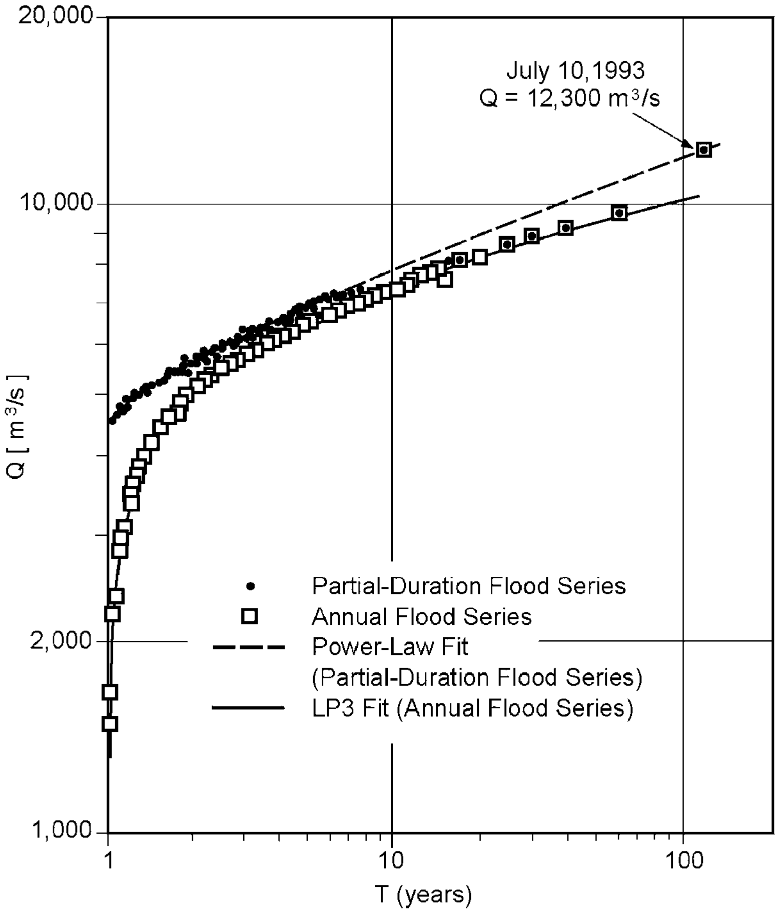

Figure 4 Annual maximum discharges (Q) of the Mississippi River at Keokuk, Iowa, for the period 1879–1995, plotted on log-log axes against return period (T). The annual series is fitted with a log Pearson type III distribution (LP3), and the partial duration series with a power law model. The 1993 flood discharge has a return period of about 100 years according to the power law, but a return period in excess of 1000 years according to the LP3 model
Source: after Malamud *et al.* (1996). Reprinted with permission of the authors.

Self-similar behaviour in a wide range of phenomena exhibit a *phase transition* (Schertzer *et al.*, 1993), and PL scaling may only be manifest above this. This implies that the PL may be a model best used for large events above a fairly high threshold. In turn, this has the consequence that PL behaviour may be difficult to detect in short records, if only a small number of floods have exceeded the phase transition threshold. Malamud *et al.*, 1996: utilized the partial duration series as a means of overcoming this issue.

Many studies of self-similar processes are *multifractal*, i.e., exhibit different PL scaling exponents over different scale ranges. This is significant for FFA because different flood generating mechanisms may give rise to different scaling exponents over specific ranges. This recalls the approach developed by Waylen and Woo (1982) and Waylen (1985) , which focuses on the concept of mixed generating mechanisms for floods. Their studies in Canada demonstrated that floods caused by winter west-coast rain could be identified and separately modelled in the flood record from those caused by a distinct process, spring snowmelt, and this study is elegant in its simplicity (Figure 5). Recent studies have further demonstrated the problematic and fallacious nature of attempting to fit a single distribution to flood data generated by separate identifiable climatic processes (Murphy, 2001; Alila and Mtiraoui, 2002). The practice is questionable of identifying more complex multiparameter pdfs to describe flood series that may include subsets generated by distinct processes, each of which may have a characteristic flood distribution which is adequately defined by a simple log normal or PL distribution. Nevertheless, attempts have been made to model such joint distributions using variants of the traditional models (e.g., the Gumbel mixed model of Yue *et al.*, 1999, or the multivariate extreme value with Gumbel marginals of Escalante-Sandoval, 1998).

The mixed generating mechanism concept is closely related to a form of nonstationarity in the flood record, and of course a critical assumption of all FFA is that of statistical stationarity (Stedinger, 2000) and independence of individual flood events. An annual flood series may include several El Niño/La Niña phases, with the flood magnitude pdfs for these different climatic phases being noticeably different, as was classically demonstrated in Australia with the identification of 'drought dominated regimes' (DDRs) and 'flood dominated regimes' (FDRs) (Warner, 1987). More recently this pattern has been interpreted as a multidecadal nonstationarity associated with modes in the Inter-decadal Pacific Oscillation (Figure 6) (Kiem *et al.*, 2003). With this recognition, it is possible for FFA studies to be used as evidence for climatic nonstationarity (Franks, 2002) following the temporal partitioning of data. This of course reduces the sample size for the fitting of individual pdfs, and thereby makes the use of multiparameter extreme-event distributions more difficult by increasing the uncertainty of their parameter estimates. This suggests that there are practical reasons (of parsimony) for favouring multiple simple distributions, in addition to the possibility that there may be better theoretical grounds for assuming that the annual flood record is more likely to be generated by mixed, simple distributions than by single, complex ones.

V The role of historical and palaeoflood data

Palaeoflood studies have traditionally been utilized for testing assumptions about the stationarity of the flood record over very long (i.e., millennial) timescales (particularly the existence of distinct climatic phases with different flood-generating capabilities; Ely, 1997). Studies of long (Holocene-scale) flood records demonstrate distinct temporal clusters of flood events (Wohl *et al.*, 1994; Ely *et al.*, 1996), reflecting the climatic history of the region in question. However, another possibility is the utilization of palaeoflood studies over more modest timescales (e.g., 100–200 years) for the purpose of testing the self-similar hypothesis in catchments with a short instrumented record. This also represents the timescale of most relevance for design structures. This approach was utilized in the following case study from northern Thailand (Kidson *et al.*, 2005).

1 An example from northern Thailand

The Mae Chaem river is a tributary of the Ping river to the southwest of Chiang Mai, joining the Ping at the town of Hot. In its lower reaches, the Mae Chaem passes through the Ob Luang gorge (Figure 7), and in the gorge there are caves which have trapped woody debris in extreme floods

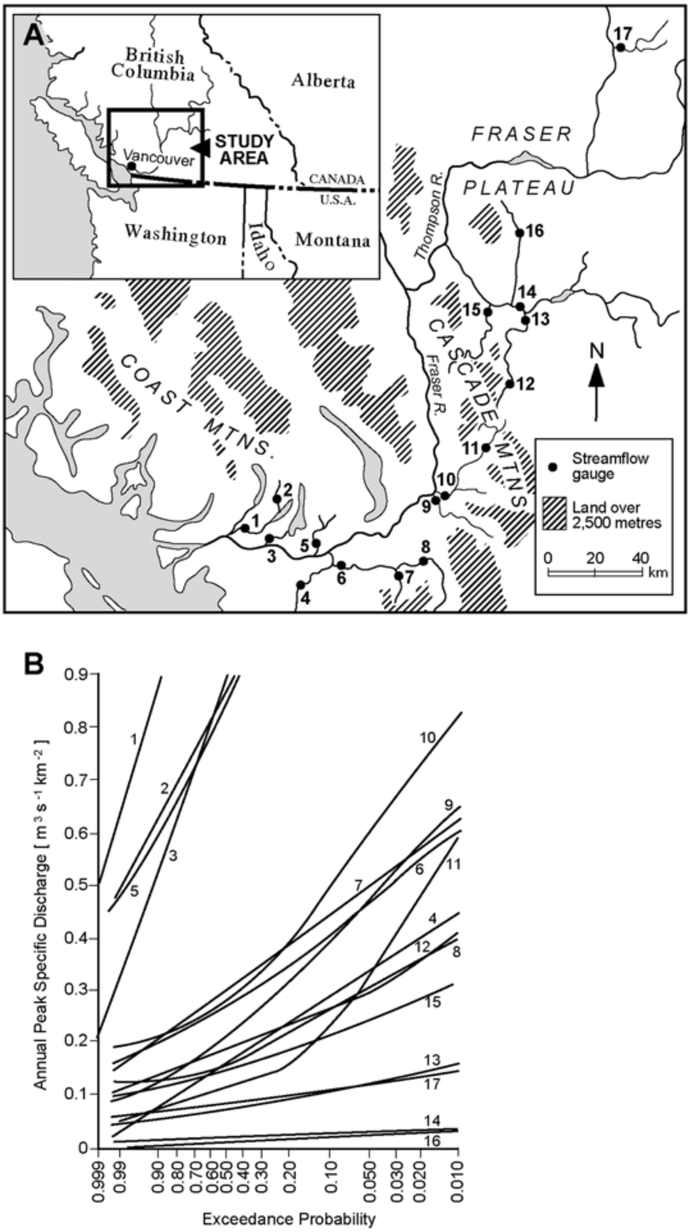

Figure 5 (a) Gauging stations on the Fraser River, British Columbia, and (b) annual flood exceedance probability curves for the numbered stations, plotted on Gumbel probability paper. Part (b) shows that flood frequency distributions are linear on this probability paper when there is a single generating process (and generating distribution), with the coastal rainfall-generated floods showing a steeper curve than the inland snowmelt floods. In a transitional region where the flood record is derived from both sources (and distributions), the curves reflect this combination
Source: after Waylen, (1985). Reprinted with permission of the author.

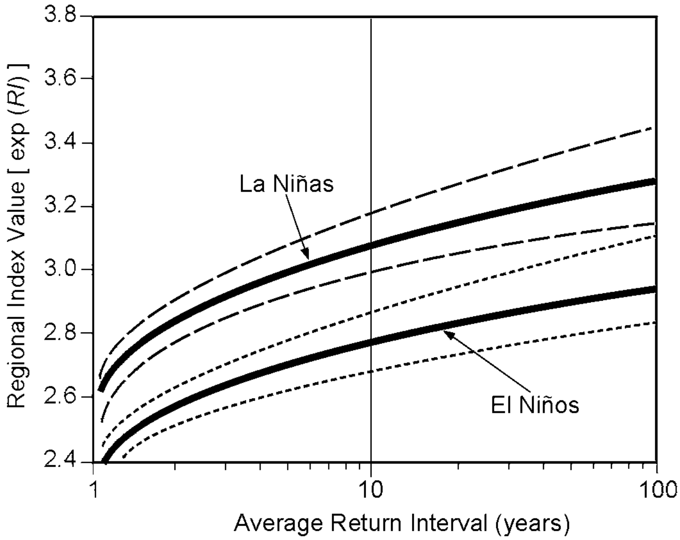

Figure 6 Regional flood frequency index curves for New South Wales, Australia, showing the expected values and the 90% confidence interval under El Niño and La Niña conditions
Source: after Kiem *et al.* (2003). Reprinted with permission of the authors.

predating the instrumental record. There have been large gauged floods in recent years, notably in August 2001, and it has been possible to survey the water level of this event and to calibrate a 1D hydraulic model to define the roughness coefficient applicable in events of this magnitude (Kidson *et al.*, unpublished data). The highest level flood deposits in the caves provide four palaeo-stage indicators (PSIs), and the 1D model with a roughness coefficient as suggested by the modelling of recent large flood events predicts a discharge of 2420 m^3s^{-1} for the water level implied by these PSIs. Dating of this pre-instrumental event in the gorge (using historical records, oral history, dendrochronology and radiocarbon dating) suggests that it occurred early in the twentieth century, and a return period of 84 years has been assigned to this event. The details of this investigation are reported elsewhere (Kidson *et al.*, 2005).

If FFA is undertaken for the instrumental flood record alone, and conventional pdfs (e.g., log Pearson type III, Gumbel EV1 and two-parameter log normal models, all fitted with the method of moments; the WRC (WRC, 1981) skew coefficient was also used in the case of the LP3) are used to estimate the discharge of an event with a return period of 84 years, the estimates are 1005, 1012 and 1040 m^3s^{-1}, respectively. They thus severely underestimate the discharge of the event associated with the PSIs in the gorge. By contrast, a PL model fitted by reduced major axis (RMA) regression to the gauged record yields an 84-year discharge estimate of 2479 m^3s^{-1}, which is a close match to the observed extreme event (Figure 8). In this application, the extreme event is not included in the model-fitting process; this data point is reserved in order to assess the predictive success of the models. This represents a methodological departure from conventional FFA, where all data points are used for the model calibration process, leaving none with which to assess the model. Hence, the predictive quality of conventionally fitted models remains untested.

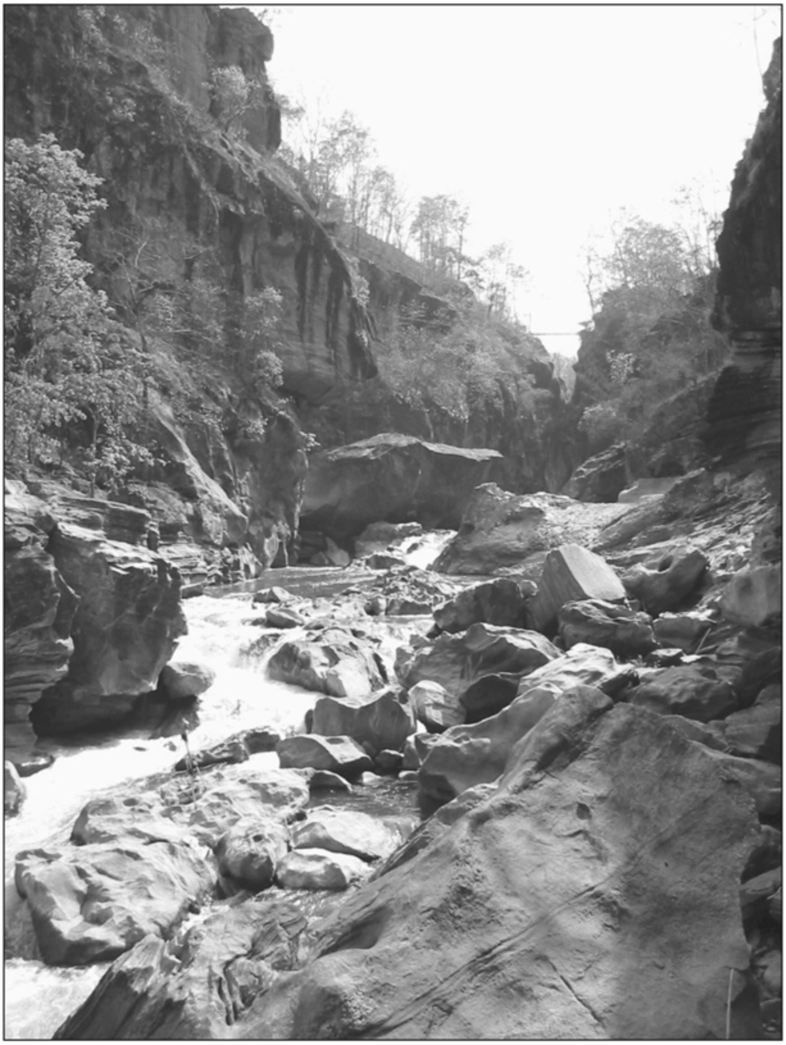

Figure 7 The Ob Luang gorge, lower Mae Chaem river, under low-flow conditions

The consequence of this finding in the case of the Mae Chaem is significant, in that the Gumbel EV1 model is the official basis for FFA in Thailand, where the maximum record lengths tend to be of the order of 50 years. However, in this catchment it appears to underpredict the magnitude of the 84-year event significantly, with inevitable consequences for structural design and planning.

2 Implications of threshold exceedance
It appears that the PL model may be applicable for very extreme events, but not for events around the mean annual flood; therefore the important question of appropriate methods of fitting PL models deserves closer research. For long records (e.g., 100 years), the threshold above which PL behaviour begins may be visible in the gauged record, permitting a regression relation using these

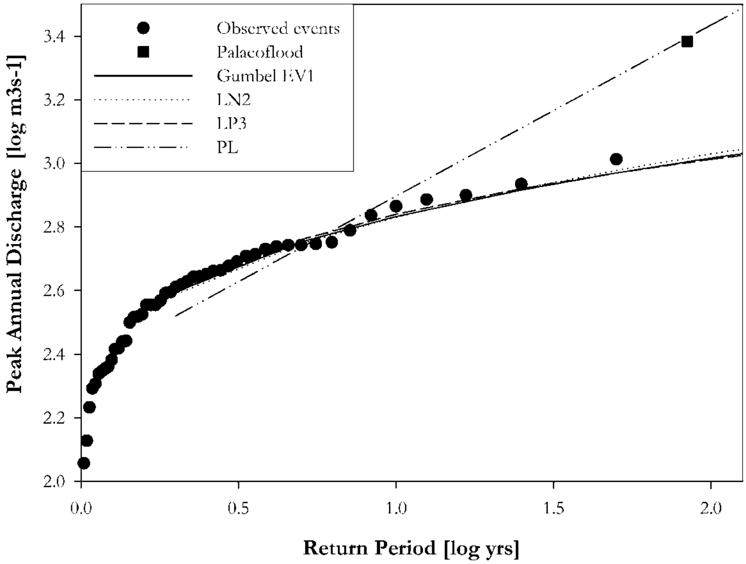

Figure 8 Four models fitted to the gauged annual floods of the Mae Chaem river; the reconstructed palaeoflood is also plotted, and the success of each of the models in predicting the palaeoflood (this event was *not* included in the model fitting process) are graphically illustrated

data–however, this may not be the case for records of short length, and other techniques and supplementary data may be required. This increases the importance of historical and palaeoflood information, since this is generally available in the form of peaks over threshold (POT). Often, the exact magnitude of an event is unknown, but is it is known to have exceeded a specific threshold: indeed, as Stedinger and Cohn 1986: note, 'historical floods are observed because their magnitude exceeds some threshold of perception'.

One of the positive consequences of the application of sophisticated statistics to FFA has been the generation and refinement of methods which usefully incorporate palaeoflood and historical information into FFA models. To take the POT example, this censored form of information, where no data are available on the flows below the threshold, poses a major statistical challenge (Kroll and Stedinger, 1996). Inclusion of historical and palaeoflood information to extend a gauging record typically 'fills in' the ungauged portion of the historic period with an appropriate number of replicates of the below-threshold portion of the systematic record. Various methods of achieving this have been proposed. The simplest is the probability weighted moments technique, advocated by IACWD (1982). However, Stedinger and Cohn (1986) demonstrated that a ML-based technique was superior: recently Cohn *et al.*, (2001) have proposed an expected moments (EM) technique as an alternative where application of ML poses numerical problems (Ouarda *et al.*, 1998). All of these techniques are designed to reduce systematic bias in estimates in the low-probability range. However, these are also examples of relatively abstruse statistical procedures which might be redundant or require revision within a self-similar modelling framework.

Hosking and Wallis (1986a; 1986b) questioned the utility of incorporating palaeoflood and historical information in FFA, demonstrating that it depends on the number of model parameters and the degree of discharge error. Frances *et al.*, (1994) also qualified its utility in terms of the relative magnitudes of the length of the systematic record, the length of the historic period, the perception threshold and the return period for which an estimate is required. Some of these qualifications also become less relevant if the extremes represented by palaeofloods are considered within a self-similar modelling approach. Several studies have now demonstrated the value of noninstrumental flood information in FFA (Stedinger and Cohn, 1987; Jin and Stedinger, 1989; Webb *et al.*, 1994; Ouarda *et al.*, 1998; Williams and Archer, 2002). In most cases these studies have been formulated within an existing paradigm of FFA based on the use of conventional extreme event distributions. It can be seen that these same data are of especial value in seeking to test self-similar hypotheses in flood behaviour; in a sense, palaeoflood and historic data represent more viable and efficient data sources on which to conduct such tests than a meagre increase in the length of the continuous instrumented record, which may yield no floods of sufficient magnitude to add value to a PL model. It is possible that the noninstrumental evidence of censored, and therefore particularly extreme, events may lead to a paradigm shift in which self-similar models play a greater role.

VI Discussion and conclusion

FFA analyses incorporate assumptions at each stage of the modelling process. An important one is the assumption of a high degree of accuracy in the estimation of discharges. It is known that considerable error is introduced, both in measurement in the instrumental record, and in modelling of palaeoflood discharges. Yet this assumption is the basis for the considerable body of literature addressing statistically detailed procedures focused on model parameter fitting and testing. The most fundamental assumption is that of the conceptual model, of which the self-similar PL approach represents a major alternative to conventional distributions.

Future research can usefully address these respective sources of uncertainty, at the most fundamental level challenging basic assumptions, and at the most practical level aiming to reduce the confidence intervals placed around flood recurrence estimates (NRC, 1999). The recent history of FFA has seen a range of innovative approaches that reflect the methods required in historic and palaeoflood studies. For example, the concept of the palaeohydrologic bound (Levish, 2002; O'Connell *et al.*, 2002) has emerged recently, which is the time interval during which a particular discharge is not exceeded, ascertained through dating the geomorphic surfaces that have not been inundated. This offers a more reliable means of defining upper confidence limits on flooding for major structures such as reservoirs. Also, Vogel *et al.*, (2001) have introduced the concept of the record-breaking flood – an event which exceeds all previous events, and this represents a new application of a branch of mathematics distinct from extreme value theory. These approaches represent a promising avenue in which to test alternative modelling frameworks (including the self-similar one), and there are thus good grounds for suggesting that the noninstrumental observation of the occurrence of extreme floods offers the best potential for a paradigm shift in FFA.

The history of FFA has been distinguished by the employment of increasingly sophisticated statistical techniques to satisfy the demand for rigorous curve fitting. This has increased the robustness of estimates and the treatment of uncertainty – but only embedded within the broader model assumptions. National mandates for particular models may be said to have reduced the tendency to challenge the conventional models, and has instead facilitated work within this paradigm. These methods have been flexible enough to incorporate the mixture of flood information that may be available for a single catchment: site-specific, regional, historical and palaeoflood information. However, concomitant theoretical advance in FFA has been somewhat limited; for example, Singh and Strupczewski (2002) have noted that the fitting of flood frequency models is largely a statistical exercise somewhat divorced from hydrological input. Given the many implicit assumptions in FFA, the safest approach lies in testing and citing the sensitivity of a given analysis to a range of possible techniques at each stage of the modelling process.

In the sense that FFA for a catchment is ultimately information-dependent, it is impossible to identify which of the many FFA tools should be employed in a given instance. A state-of-the-art analysis is likely to entail a multidisciplinary approach, incorporating the results of both physically based (e.g., rainfall/runoff) modelling with detailed instrumental gauging data, supplemented with regional data, and historical and palaeoflood information (Stedinger, 2000). The cases where this full range can be applied in combination are limited. Nevertheless, in those catchments where there is evidence of a censored set of historic and palaeoflood water levels which can be converted into discharges, it is likely that flood frequency analyses may require radical revision as new theoretical concepts are increasingly employed. The self-similar PL model offers particular promise for the prediction of extreme events. Since individual catchments often do not supply a sufficient length of instrumented record on which to test alternative extreme event models, emphasis is placed on the relevance of historical, palaeoflood and regional evidence as a means to overcome the data limitations in single catchments and ultimately improve extreme flood prediction.

References

Adamowski, K. 2000: Regional analysis of annual maximum and partial duration flood data by nonparametric and L-moment methods. *Journal of Hydrology* 229, 219–31.

Adamowski, K., Geng-Chen-Liang and **Patry, G.G.** 1998: Annual maxima and partial duration flood series analysis by parametric and non-parametric methods. *Hydrological Processes* 12, 1685–99.

Ahmad, M.I., Sinclair, C.D. and **Werritty, A.** 1988: Log-logistic flood frequency analysis. *Journal of Hydrology* 98, 205–24.

Alila, Y. and **Mtiraoui, A.** 2002: Implications of heterogeneous flood-frequency distributions on traditional stream-discharge prediction techniques. *Hydrological Processes* 16, 1065–84.

Baker, V.R. 2003: Palaeofloods and extended discharge records. In Gregory, K.J. and Benito, G., editors, *Palaeohydrology–understanding global change,* Chichester: Wiley, 307–24.

Bayazit, M., Shaban, F. and **Onoz, B.** 1997: Generalized extreme value distribution for flood discharges. *Turkish Journal of Engineering and Environmental Sciences* 21, 69–73.

Beven, K. 1987: Towards the use of catchment geomorphology in flood frequency predictions. *Earth Surface Processes and Landforms* 12, 69–82.

Birkeland, K.W. and **Landry, C.C.** 2002: Power laws and snow avalanches. *Geophysical Research Letters* 29, 49.41–49.43.

Bobee, B. 1999: Extreme flood events valuation using frequency analysis: a critical review. *Houille Blanche* 54, 100–105.

Bobee, B., Cavadias, G., Ashkar, F., Bernier, J. and **Rasmussen, P.** 1993: Towards a systematic approach to comparing distributions used in flood frequency analysis. *Journal of Hydrology* 142, 121–36.

Chen, Y., Yu, H., Van-Gelder, P. and **Zhigui, S.** 2002: Study of parameter estimation methods for Pearson-III distribution in flood frequency analysis. *IAHS AISH* 271, 263–70.

Chow, V.T. 1954: The log-probability Law and its engineering applicatioin. Proceedings, American Society of Civil Engineers 80, Paper 536, 1–25.

Chowdhury, J.U., Stedinger, J.R. and **Li-Hsiung-Lu,** 1991: Goodness-of-fit tests for regional generalized extreme value flood distributions. *Water Resources Research* 27, 1765–76.

Cohn, T.A., Lane, W.L. and **Stedinger, J.R.** 2001: Confidence intervals for expected moments algorithm flood quantile estimates. *Water Resources Research* 37, 1695–706.

de Michele, C., La Barbera, P. and **Rosso, R.** 2002: Power law distribution of catastrophic floods. in Snorrason, A., Finnsdottir, H. and Moss, M., editors, *The extremes of the extremes: extraordinary floods*, vol. 271, Wallingford, IAHS, 277–82.

Dunne, T. and **Leopold, L.B.** 1978: Calculation of flood hazard. In Dunne, T. and Leopold, L.B., *Water in environmental planning*, San Francisco, WH Freeman and Co, CA: 279–391.

Ely, L.L. 1997: Response of extreme floods in the southwestern United States to climatic variations in the late Holocene. *Geomorphology* 19, 175–201.

Ely, L.L., Enzel, Y., Baker, V.R., Kale, V.S. and **Mishra, S.** 1996: Changes in the magnitude and frequency of late Holocene monsoon floods on the Narmada River, central India. *Geological Society of America Bulletin* 108, 1134–48.

Escalante-Sandoval, C. 1998: Multivariate extreme value distribution with mixed Gumbel marginals. *Journal of the American Water Resources Association* 34, 321–33.

Escalante-Sandoval, C.A. and **Raynal-Villasenor, J.A.** 1994: A trivariate extreme value distribution applied to flood frequency analysis. *Journal of Research* 99, 369–75.

Fortin, V. and **Bobee, B.** 1994: Nonparametric bootstrap confidence intervals for the log-Pearson type III distribution. In Zanetti, P., editor, *Computer techniques in environmental studies V, vol. II: environmental systems,* Billerica, MA: Computational Mechanics Publications, 351–58.

Frances, F., Salas, J.D. and **Boes, D.C.** 1994: Flood frequency analysis with systematic and historical or paleoflood data based on the two-parameter general extreme value models. *Water Resources Research* 30, 1653–64.

Franks, S.W. 2002: Identification of a change in climate state using regional flood data. *Hydrology and Earth System Sciences* 6, 11–16.

GREHYS, 1996: Presentation and review of some methods for regional flood frequency analysis. *Journal of Hydrology* 186, 63–84.

Greis, N.P. 1983: Flood frequency analysis: a review of 1979–1984. *Reviews of Geophysics and Space Physics* 21, 699–706.

Gumbel, E.J. 1958: *Statistics of extremes*. New York: Columbia University Press, 375 pp.

Gupta, V.K. and **Waymire, E.** 1990: Multiscaling properties of spatial rainfall and river flow distributions. *Journal of Geophysical Research* 95(D3), 1999–2009.

Haktanir, T. and **Horlacher, H.B.** 1993: Evaluation of various distributions for flood frequency analysis. *Hydrological Sciences Journal* 38, 15–32.

Heo, J.H., Salas, J.D. and **Boes, D.C.** 2001a: Regional flood frequency analysis based on a Weibull model: part 1. Estimation and asymptotic variances. *Journal of Hydrology* 242, 157–70.

— 2001b: Regional flood frequency analysis based on a Weibull model: part 2. Simulations and applications. *Journal of Hydrology* 242, 171–82.

Hosking, J.R.M. 1990: L-moments: analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society* B 52, 105–24.

Hosking, J.R.M. and **Wallis, J.R.** 1986a: Paleoflood hydrology and flood frequency analysis. *Water Resources Research* 22, 543–50.

— 1986b: The value of historical data in flood frequency analysis. *Water Resources Research* 22, 1606–12.

— 1987: Parameter and quantile estimation for the generalized Pareto distribution. *Technometrics* 29, 339–49.

House, P.K., Webb, R.H., Baker, V.R. and **Levish, D.R.** 2002: *Ancient floods, modern hazards: principles and applications of palaeoflood hydrology*, vol. 5, Washington, DC: American Geophysical Union, 385 pp.

Hubert, P. 2001: Multifractals as a tool to overcome scale problems in hydrology. *Hydrological Sciences* 46, 897–905.

Hurst, H.E., Black, R.P. and **Simaika, T.M.** 1965: *Long term storage*. London: Constable, 145 pp.

IACWD, (Interagency Advisory Committee on Water Data) 1982: Guidelines for determining flow frequency. In *Bulletin 17B*, Reston, UA: US Department of the Interior.

Jin, M. and **Stedinger, J.R.** 1989: Flood frequency analysis with regional and historical information. *Water Resources Research* 25, 925–39.

Karim, M.A. and **Chowdhury, J.U.** 1995: A comparison of four distributions used in flood frequency analysis in Bangladesh. *Hydrological Sciences Journal* 40, 55–66.

Kidson, R., Richards, K.S. and **Carling, P.A.** 2005: Reconstructing the 1-in-100 year flood in northern Thailand. *Geomorphology*, in press.

Kiem, A.S., Franks, S.W. and **Kuczera, G.** 2003: Multi-decadal variability of flood risk. *Geophysical Research Letters* 30(2), 10.1029/2002GL015992, 7.1–7.4.

Klemes, V. 1989: The improbable probabilities of extreme floods and droughts. In Starosolszky, O. and Melder, O.M., editors, *Hydrology of disasters,* James and James, London: 43–51.

Kroll, C.N. and **Stedinger, J.R.** 1996: Estimation of moments and quantiles using censored data. *Water Resources Research* 32, 1005–12.

Levish, D.R. 2002: Paleohydrologic bounds – nonexceedance information for flood hazard assessment. In House, P.K., Webb, R.H., Baker, V.R., and Levish, D.R., editors, *Ancient floods, modern hazards: principles and applications of paleoflood hydrology,* Washington, DC: American Geophysical Union, 175–90.

Malamud, B.D. and **Turcotte, D.L.** 1999: Self-organized criticality applied to natural hazards. *Natural Hazards* 20, 93–116.

Malamud, B.D., Turcotte, D.L. and **Barton, C.C.** 1996: The 1993 Mississippi River flood: a one hundred or a one thousand year event? *Environmental and Engineering Geosciences* 2, 479–86.

Murphy, P.J. 2001: Evaluation of mixed-population flood-frequency analysis. *Journal of Hydrologic Engineering* 6, 62–70.

Mutua, F.M. 1994: The use of the Akaike information criterion in the identification of an optimum flood frequency model. *Hydrological Sciences Journal* 39, 235–44.

National Environment Research Council (NERC) 1999: *Flood Studies Report*, (in five volumes) Wallingford: Institute of Hydrology.

National Research Council (NRC) 1988: *Committee on techniques for estimating probabilities of extreme floods: estimating probabilities of extreme floods – methods and recommended research*. Washington, DC: National Academy Press.

— 1999: Improving American river flood frequency analyses. In *Committee on American river flood frequency: water science and technology board, committee on geoscience, environment and resources*, Washington DC: National Academy Press, 1–120.

O'Connell, D.R.H., Ostenaa, D.A., Levish, D.R. and **Klinger, R.E.** 2002: Bayesian flood frequency analysis with paleohydrologic bound data. *Water Resources Research* 38(5), 10.1029 2000WR000028.

Ouarda, T.B.M.J. and **Ashkar, F.** 1998: Effect of trimming on LP III flood quantile estimates. *Journal of Hydrologic Engineering ASCE* 3, 33–42.

Ouarda, T.B.M.J., Rasmussen, P.F., Bobee, B. and **Bernier, J.** 1998: Use of historical information in hydrologic frequency analysis. *Revue des Sciences de l'Eau* 11(Special Issue), 41–49.

Ramachandra Rao, A. and **Hamed, K.H.** 2000: *Flood frequency analysis*. Boca Raton, FL: CRC Press, 350 pp.

Scheidegger, A.E. 1997: Complexity theory of natural disasters; boundaries of self-structured domains. *Natural Hazards* 16, 103–12.

Schertzer, D., Lovejoy, S. and **Lavallee, D.** 1993: Generic multifractal phase transitions and self organised criticality. In Pendang, J.M. and Lejeune, A., editors, *Cellular automata: prospects in astrophysical applications* Singapore: World Scientific, 216–27.

Singh, V.P. and **Strupczewski, W.G.** 2002: On the status of flood frequency analysis. *Hydrological Processes* 16, 3737–40.

Stedinger, J.R. 2000: Flood frequency analysis and statistical estimation of flood risk. In Wohl E. E., editor, *Inland flood hazards: human, riparian and aquatic communities,* Cambridge: Cambridge University Press, 334–58.

Stedinger, J.R. and **Cohn, T.A.** 1986: Flood frequency-analysis with historical and paleoflood information. *Water Resources Research* 22, 785–93.

— 1987: Historical flood frequency data; its value and use. In Singh, V.P., editor, *Regional flood frequency analysis*, Proceedings of the International Symposium on Flood Frequency and Risk Analyses, Baton Rouge, USA, Baton Rouge, LN: D Reidel, 273–86.

Stedinger, J.R., Vogel, R.M. and **Foufoula-Georgiou, E.** 1993: Frequency analysis of extreme events. In Maidment, D., editor, *Handbook of hydrology,* New York: McGraw-Hill.

Tanaka, S. and **Takara, K.** 2002: A study on threshold selection in POT analysis of extreme floods. *IAHS AISH* 271, 299–304.

Turcotte, D.L. 1994: *Fractals and chaos in geology and geophysics*. Cambridge: Cambridge University Press, 221 pp.

Vogel, R.M., McMahon, T.A. and **Chiew, F.H.S.** 1993: Flood flow frequency model selection in Australia. *Journal of Hydrology* 146, 421–49.

Vogel, R.M., Zafirakou-Koulouris, A. and **Matalas, N.C.** 2001: Frequency of record-breaking floods in the United States. *Water Resources Research* 37, 1723–31.

Vogel, R.W. and **McMartin, D.E.** 1991: Probability plot goodness-of-fit and skewness estimation procedures for the Pearson type 3 distribution. *Water Resources Research* 27, 3149–58.

Warner, R.F. 1987: *The impacts of flood- and drought-dominated regimes on channel morphology at Penrith, New South Wales, Australia*. International Association of Hydrological Sciences (IAHS) Publication 168, 327–38.

Waylen, P.R. 1985: Stochastic flood analysis in a region of mixed generating processes. *Transactions of the Institute of British Geographers* NS 10, 95–108.

Waylen, P.R. and **Woo, M.K.** 1982: Prediction of annual floods generated by mixed processes. *Water Resources Research* 18, 1283–86.

Webb, R.H., Moss, M.E. and **Pratt, D.A.** 1994: Information content of paleoflood data in flood-frequency analysis. *EOS, Transactions of the American Geophysical Union* 75(44), 372.

Williams, A. and **Archer, D.** 2002: The use of historical flood information in the English Midlands to improve risk assessment. *Hydrological Sciences Journal* 47, 67–76.

Wohl, E.E., Fuertsch, S.J. and **Baker, V.R.** 1994: Sedimentary records of late holocene floods along the Fitzroy and Margaret Rivers, western Australia. *Australian Journal of Earth Sciences* 41, 273–80.

WRC, 1981: *Guidelines for determining flood frequency*. Washington, DC: US Water Resources Council.

Wu-Boxian, Hou-Yu, and **Ding-Jing,** 1991: The method of lower-bound to estimate the parameters of a Pearson type III distribution. *Hydrological Sciences Journal* 36, 271–80.

Yixing, B., Koung, T.Y. and **Hasfurther, V.R.** 1987: Evaluation of uncertainty in flood magnitude estimator on annual expected damage costs of hydraulic structures. *Water Resources Research* 23, 2023–29.

Yen, B. 2002: System and component uncertainties in water resources. In Bogardi, J.J. and Kundzewicz, Z.W., editors, *Risk, reliability, uncertainty and robustness of water resources systems*, Cambridge: Cambridge University Press, 133–42.

Yue, S. 2001: A bivariate gamma distribution for use in multivariate flood frequency analysis. *Hydrological Processes* 15, 1033–45.

Yue, S., Ouarda, T.M.B.J., Bobee, B., Legendre, P. and **Bruneau, P.** 1999: The Gumbel mixed model for flood frequency analysis. *Journal of Hydrology* 226, 88–100.