Hydrology and Earth System Sciences, 4(1), *23–34* (2000) © EGS

Flood frequency estimation by continuous simulation (with likelihood based uncertainty estimation)

David Cameron[1], Keith Beven[2], Jonathan Tawn[3] and Pamela Naden[4]

[1]The Environment Agency, Tyneside House, Skinnerburn Road, Newcastle Business Park, Newcastle Upon Tyne, NE4 7AR, UK
[2]Institute of Environmental and Natural Sciences, Lancaster University, Lancaster, LA1 4YQ, UK
[3]Department of Mathematics and Statistics, Lancaster University, Lancaster, LA1 4YF, UK
[4]Institute of Hydrology, Crowmarsh Gifford, Wallingford, Oxfordshire, OX10 8BB, UK
e-mail for corresponding author: david.cameron@environment-agency.gov.uk

Abstract

A continuous simulation methodology, which incorporates the quantification of modelling uncertainties, is used for flood frequency estimation. The methodology utilises the rainfall-runoff model TOPMODEL within the uncertainty framework of GLUE. Long return period estimates are obtained through the coupling of a stochastic rainfall generator with TOPMODEL. Examples of applications to four gauged UK catchments are provided. A comparison with a traditional statistical approach indicates the suitability of the methodology as an alternative technique for flood frequency estimation. It is suggested that, given an appropriate choice of rainfall-runoff model and stochastic rainstorm generator, the basic methodology can be adapted for use in many other regions of the world.

Keywords: Floods; Frequency; TOPMODEL; Rainfall-runoff modelling

Introduction

With many areas of the world suffering from the effects of flood damage in recent years, it is clear that there remains a need for the reliable estimation of flood peaks with given return period. In the past, these have often been derived through the use of purely statistical or event-based hydrological modelling tools (e.g. Rodriguez-Iturbe and Valdes, 1979 Hebson and Wood, 1982; Diaz-Granados *et al.*, 1984; Bras *et al.*, 1985; Sivapalan *et al.*, 1990; Bradley and Potter, 1992; Michaud and Sorooshian, 1993; Gupta and Dawdy, 1995; Hosking and Wallis, 1997). Recently, with the availability of more powerful computing facilities, this problem has been approached through the use of continuous simulation (e.g. Beven, 1986, 1987; Calver and Lamb, 1996; Blazkova and Beven, 1997; Lamb, 1999; Cameron *et al.*, 1999).

Continuous simulation is used here to denote the prediction of a continuous discharge time series through the use of a hydrological model. Rainfall inputs may either be observed, or generated from a stochastic rainfall model. The simulated flow series can be analysed using traditional flood frequency techniques (e.g. annual maximum or peaks over threshold analysis). This approach has an advantage over the earlier derived distribution techniques (e.g. Eagleson, 1972) in that soil moisture conditions are continuously accounted for by the rainfall-runoff model. If it is accepted that the hydrological model used provides satisfactory simulations, then the resulting flood frequency estimates are therefore perhaps more hydrologically meaningful.

A recent paper by Cameron *et al.* (1999) demonstrated the use of a continuous simulation methodology for the purpose of flood frequency estimation within an uncertainty framework. This methodology utilised the rainfall-runoff model TOPMODEL, together with a stochastic rainstorm generator, in order to produce estimates of flood events of long return period. This was applied to the 21 years of hourly flow and catchment average rainfall data available for the Wye catchment, Plynlimon, Wales, UK. Good reproduction of the observed hourly annual maximum flood peaks was demonstrated. Reasonable estimates of longer return period floods (e.g. 100 years) were also obtained. In addition, a consistency between the parameterisation of TOPMODEL for both hourly annual maximum peaks and continuous hourly hydrograph simulation was shown.

In what follows, applications to four further gauged UK catchments are presented. The uncertainties involved in the resulting flood frequency estimates are considered, and the findings are compared with those of a traditional statistical approach.

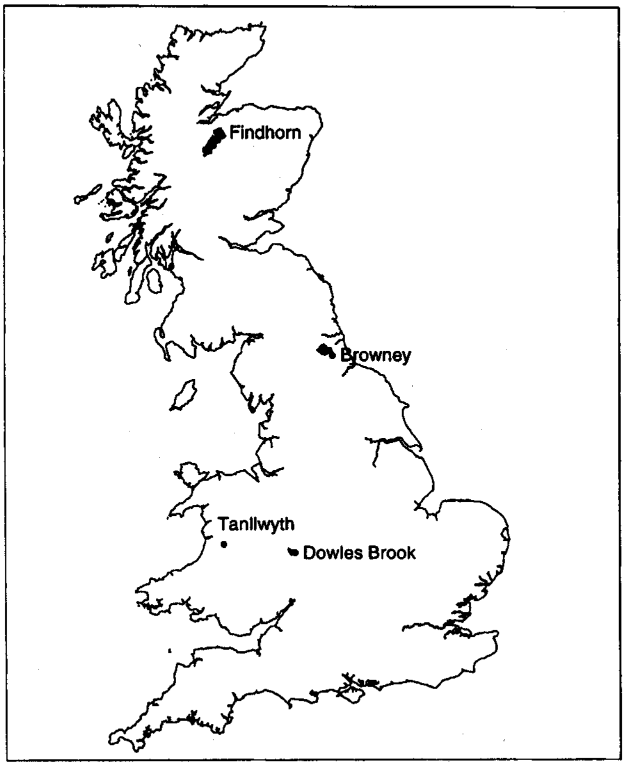

Fig. 1. *Locations of the four UK study catchments.*

Four gauged catchments in the UK

Four UK catchments, of varied location, data record, and catchment characteristics were selected for analysis (Fig. 1 and Table 1). These are the Tanllwyth (mid-Wales), the Browney (north-east England), the Findhorn (north-east Scotland) and Dowles Brook (west England). Daily potential evapotranspiration estimates were obtained from the nearest MORECS synoptic sites (Thompson *et al.*, 1981).

Although there is an extensive instantaneous annual maximum peak record available for each catchment, the quantity of hourly flow and catchment average hourly rainfall (CAHR) data is quite varied (Table 1). This ranges from 5 years (Dowles Brook) to 16 years (Tanllwyth). Moreover, these periods of hourly data are not consistent across the four catchments. In addition, the CAHR data were derived (by the UK's Institute of Hydrology, prior to the current study) using hourly raingauge records to distribute daily catchment rainfall totals (obtained by averaging over a large number of daily raingauges). This procedure featured the infilling of missing sections of data (achieved through the temporal disaggregation of appropriate daily data through the use of at-site mass curves, Huff, 1967; Lamb and Gannon, 1996). As a result, there are differences in the quality of the CAHR data between the four sites. For example, the calculation of the Tanllwyth's CAHR record included the use of data obtained from 4 hourly raingauges located within a small catchment area (0.9 km^2). Data from only 4 hourly raingauges within a much larger catchment area (415.6 km^2) were available for inclusion in the calculation of the Findhorn's CAHR record.

The hydrological model

Full details of TOPMODEL may be found in Beven *et al.* (1995), and Beven (1997), so only a brief summary is outlined here.

TOPMODEL is a simple semi-distributed model of catchment hydrology that predicts storm runoff from a combination of variable saturated surface contributing area and subsurface runoff (e.g. Beven 1986, 1987; Quinn and Beven, 1993). The dynamics of the contributing area for rapid runoff as the catchment wets and dries are based on a quasisteady state analysis. As with all other TOPMODEL applications (see Beven, 1997), the topographic index $\ln(a/\tan\beta)$ is used as an index of hydrological similarity, where a is the area draining through a point, and $\tan\beta$ is the local surface slope. The use of this form of topographic index implies an effective transmissivity profile that declines exponentially with increasing storage deficits. Calculation of the $\ln(a/\tan\beta)$ index was achieved using a modified version of Quinn *et al.*'s (1995) multiple flow direction algorithm taking into account the presence of river pixels in the digital terrain map (Cameron *et al.*, 1999).

Evapotranspiration losses are controlled by potential evapotranspiration and storage in the root zone with the parameter *SRMAX* (effective available water capacity at the root zone). The potential evapotranspiration estimation routine uses the same seasonal sine curve as Beven (1986, 1987) and Blazkova and Beven (1997) with a single mean hourly potential evapotranspiration parameter. This was derived directly from the available daily potential evapotranspiration as calculated for the nearby synoptic sites by MORECS (Thompson *et al.*, 1981).

In this study, TOPMODEL is driven by observed hourly catchment rainfall inputs for the purposes of flood frequency estimation. Longer return period estimates are then obtained through the coupling of TOPMODEL with a stochastic rainfall model parameterised separately for each catchment.

The stochastic rainfall model

Full details of the stochastic rainfall model are provided in Cameron *et al.* (1999), so only a brief summary is given here.

The stochastic rainfall model is based on the available data and generates random rainstorms via a Monte Carlo sampling procedure. The model characterises a storm in

Table 1. Details of the four gauged UK catchments used in this study.

	Catchment (UK hydrometric station number)			
	Tanllwyth (54090)	Browney (24005)	Findhorn (7001)	Dowles Brook (54034)
Location and grid reference	Mid-Wales 22 (SN) 843 876	NE England 45 (NZ) 259 387	NE Scotland 28 (NH) 826 337	W England 32 (SO) 768 764
Original data source	Institute of Hydrology	Former NRA[a] (Northumbrian Region)	Former RPB[b] (Highland Region)	Former NRA[a] (Severn-Trent Region)
Period of daily MORECS PET	1988–1992	1985–1992	1985–1992	1988–1992
Period of inst. peak data	1973–1992	1954–1995[c]	1960–1994	1971–1995[c]
Period of hourly flow and CAHR[d]	1974–1989	1985–1992	1985–1990	1988–1992
Length of hourly data record (yrs)	16	8	6	5
No. hourly raingauges	4	14	4	3
Catch. area (km^2)	0.9	178.5	415.6	40.8
Physical characteristics	Afforested; shales and grits.	Coal measures.	Blanket peat.	Afforested; sandstones and marls.
Mean annual rainfall (mm)	2554	744	1262	734
Mean annual runoff (mm)	2086	301	1037	300
Mean annual flood (m^3s^{-1})	1.1	37.6	234.9	13.1
Q10 (m^3s^{-1})	0.146	3.523	30.900	0.955
Q50 (m^3s^{-1})	0.028	0.979	7.898	0.160
Q95 (m^3s^{-1})	0.006	0.306	2.046	0.034

[a] National Rivers Authority – now part of the Environment Agency in England and Wales, UK.
[b] River Purification Board – now part of the Scottish Environmental Protection Agency, UK.
[c] Partially incomplete record.
[d] Catchment average hourly rainfall – calculated by the Institute of Hydrology prior to this study.

Table 2. Rainfall model duration classes and number of observed rainstorm events obtained for each catchment.

	Tanllwyth	Browney	Findhorn	Dowles Brook
Duration Class				
1 hr	3490	1635	1182	1056
2–3 hr	1980	718	749	377
4–10 hr	1652	655	574	183
11–16 hr	254	53	117	16
17–27 hr	165	21	49	2
28–62 hr	39	2	1	1[a]
≥63 hr	47	1	1[a]	1

[a] Assumed number of events if duration class "empty"

Table 3a. Maximum likelihood (ML) estimates and *behavioural* parameter ranges for the GPD shape parameter (κ).

CDF	Tanllwyth		Browney		Findhorn		Dowles Brook	
	ML	Range	ML	Range	ML	Range	ML	Range
1 hr	−0.1901	−0.3528:−0.0926	−0.1167	−0.2458:−0.0754	−0.1161[a]	−0.1585:−0.0856	−0.0837[a]	−0.1200:−0.0609
2–3 hr	−0.0908[a]	−0.1128:−0.0741	−0.1837	−0.3747:−0.0663	−0.5687	−0.8636:−0.1694	−0.3762	−0.8654:−0.0708
4–10 hr	−0.0802[a]	−0.1003:−0.0652	−0.1164[a]	−0.1586:−0.0879	−0.1851	−0.3656:−0.1019	−0.5003	−1.2349:−0.0443
11–16 hr	−0.4617	−0.6483:−0.2226	−0.0741[a]	−0.1217:−0.0485	−0.1091	−0.3169:−0.0385	−1.1995	−3.2301:−0.0042
17–27 hr	−1.0180	−1.2702:−0.8026	−0.5897	−0.6537:−0.1183	−0.2167	−0.2363:−0.0701	n/a	n/a
28–62 hr	−1.1451	−2.3921:−0.7862	n/a	n/a	n/a	n/a	n/a	n/a
≥63 hr	−0.3837	−0.4731:−0.1108	n/a	n/a	n/a	n/a	n/a	n/a
Duration	n/a	n/a	−0.1211[a]	−0.1373:−0.1072	−0.1211[a]	−0.4716:−0.1613	n/a	n/a

Upper bounding required.

Table 3b. Maximum likelihood (ML) estimates and *behavioural* parameter ranges for the GPD scale parameter (σ).

CDF	Tanllwyth		Browney		Findhorn		Dowles Brook	
	ML	Range	ML	Range	ML	Range	ML	Range
1 hr	0.5731	0.4481: 0.7346	0.5089	0.4007: 0.6853	0.5615[a]	0.4139: 0.7668	0.4047[a]	0.2944: 0.5805
2–3 hr	0.3701[a]	0.3021: 0.4599	0.4088	0.2828: 0.5978	0.7085	0.4510: 1.0333	0.5593	0.2888: 1.0545
4–10 hr	0.2446[a]	0.1986: 0.3058	0.3547[a]	0.2679: 0.4836	0.4556	0.3106: 0.6177	0.3528	0.1413: 0.7665
11–16 hr	0.6564	0.4864: 0.8729	0.2347[a]	0.1535: 0.3852	0.2429	0.1681: 0.2981	0.3688	0.0952: 0.9973
17–27 hr	2.7988	2.2146: 3.4990	0.7682	0.3759: 0.8061	0.4045	0.2526: 0.3352	n/a	n/a
28–62 hr	2.3913	1.8048: 4.9969	n/a	n/a	n/a	n/a	n/a	n/a
≥63 hr	0.4960	0.2713: 0.6689	n/a	n/a	n/a	n/a	n/a	n/a
Duration	n/a	n/a	0.4635[a]	0.4106: 0.5255	0.4635[a]	0.4710: 0.6929	n/a	n/a

Upper bounding required.

terms of a mean storm intensity, duration and inter-arrival time, in addition to a storm profile component. A "storm" is defined as any event with a minimum intensity of 0.1 mm at an hour, a minimum duration of 1 hr and a minimum inter-arrival time also of 1 hr. In general, this accounts for 99% of the CAHR in each catchment. This definition was therefore used to extract a series of storm events directly from each observed CAHR series.

The analysis of each storm event series revealed that there was a similar dependence of mean storm intensity upon storm duration to that shown in the previous Wye catchment study (see Cameron *et al.*, 1999). On the basis of similar mean storm intensities, this dependence was incorporated into the model through the subdivision of the observed storm event series into the seven duration classes defined in that study. The duration classes are: 1 hr, 2–3 hr, 4–10 hr, 11–16 hr, 17–27 hr, 28–62 hr, and ≥;63 hr (Table 2). At the sites where the comparatively short available data sample (e.g. 5 years at Dowles Brook) would have resulted in an "empty" class, it was assumed that that class contained only one storm for the purpose of defining a relative frequency.

For each duration class, an empirical cumulative distribution function (cdf) of log-transformed mean storm intensity was constructed directly from the events located within that class. Where appropriate, the upper tail of a given cdf was extrapolated via the fitting of a Generalised Pareto distribution (GPD). This was done in order to permit the generation of extreme storm events unrecorded within the available catchment storm series.

This procedure required a definition for an "upper tail". In order to maintain a consistency across the sites, the proportion between the number of storms in a given duration class' upper tail and the number of storms in that class as a whole was set to the proportion used in the earlier Wye study of Cameron *et al.* (1999). A further prerequisite for upper tail identification was that the number of storms in the upper tail was at least equal to the number of years in the shortest observed hourly record (5 years, Dowles Brook). In certain cases (e.g. Dowles Brook), this latter requirement limited the extrapolations to the cdfs of short-medium duration classes. However, this restriction did not inhibit the realistic simulation of the observed extreme rainfalls.

The initial GPD fits were obtained using maximum likelihood (Tables 3a and 3b). Where necessary, an upper bound, taken from the observed maximum UK rainfalls, was applied to the fit (Cameron *et al.*, 1999). This was done in order to prevent the generation of (physically) unrealistically high mean storm intensities at levels of very high non-exceedance probability. It introduces a dependency between the shape (κ) and scale (σ) parameters of the GPD and therefore does not increase the number of parameters required.

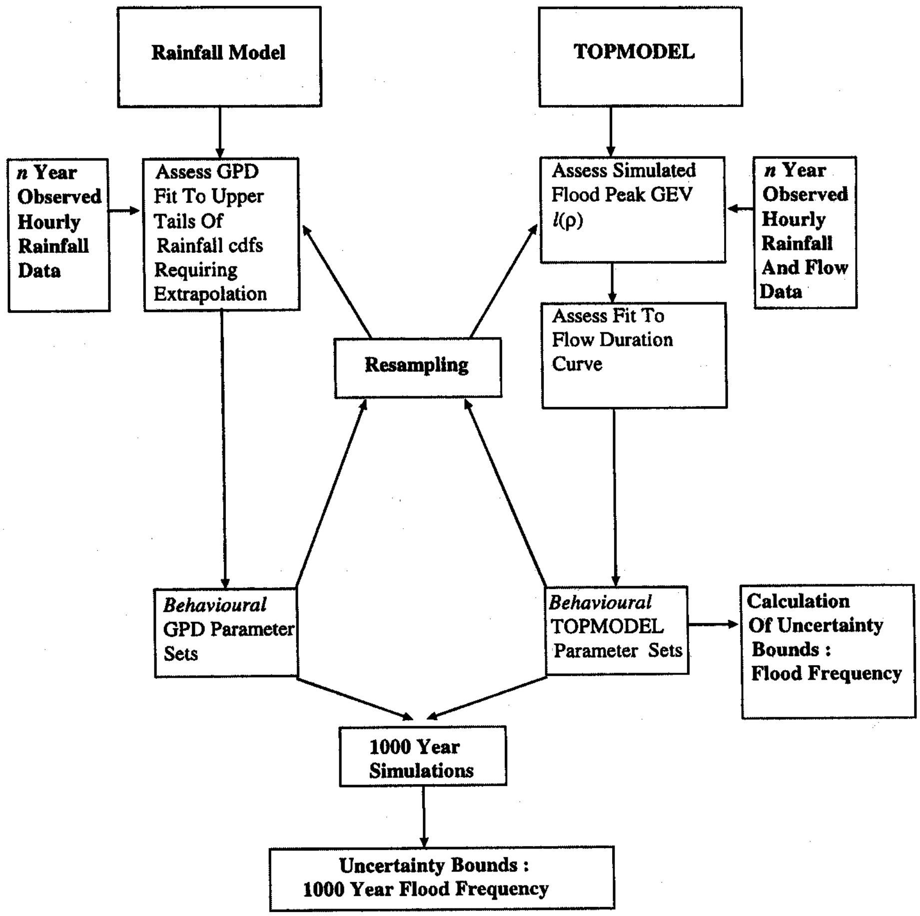

Fig. 2. *Flood frequency estimation by continuous simulation via GLUE TOPMODEL/stochastic rainfall model: sequence of events. (Adapted from Cameron et al., 1999).*

The storm duration and inter-arrival time characteristics derived from the observed event series of each catchment were also modelled using their empirical cdfs. In the case of inter-arrival time, it was assumed that the observed samples required no further extrapolation. However, for the Browney and the Findhorn, the limited sample of storms necessitated the extrapolation of the duration cdf. This was achieved via the fitting of a GPD, with an upper bound of 318 hrs (for consistency with the earlier Wye study, Cameron *et al.*, 1999), to the upper tail of that cdf (Tables 3a and 3b). The upper tail was defined as having a minimum threshold (or location parameter, u) of 7 hrs. This yielded 216 and 306 upper tail storms for the Browney and the Findhorn, respectively.

The final component of the model is a storm profile. In the earlier study of Cameron *et al.* (1999) of the Wye catchment, the observed 21 year rainstorm event series was utilised to provide an extensive database of storm profiles for each duration class. These were normalised by cumulative volume and total duration. A comparison of these profiles with those of the four study catchments revealed no significant differences. The normalised Wye profiles were therefore used for all four catchments. This provided a larger profile database than was available from the shorter periods of record of the four study catchments. During a model run, the Wye profiles were randomly selected in order to provide storm profiles for the simulated rainfall events.

The generalised likelihood uncertainty estimation (GLUE) framework

Every flood frequency estimate is subject to some degree of

Table 4. Initial ranges for the TOPMODEL parameter sets.

Parameter	Range
m (recession parameter)	0.0010:0.0450 m
S_{rmax} (max. root zone storage).	0.0010:0.2000 m
T_0 (transmissivity)	0.0001:8.0000 log
$STDT$ (standard deviation of transmissivity)	0.0001:10.0000 log

uncertainty. The major sources of this uncertainty in the continuous simulation approach include the limitations of the observed data series and the choice of rainfall and hydrological models (especially with respect to the model structures, and their calibration/validation). In this study, the Generalised Likelihood Uncertainty Estimation (GLUE) framework of Beven and Binley (1992) was used to assess this uncertainty (see also Beven, 1993; Freer *et al.*, 1996; Franks *et al.*, 1998; Cameron *et al.*, 1999).

The GLUE methodology rejects the concept of a single, global optimum parameter set and instead accepts the existence of multiple acceptable (or *behavioural*) parameter sets (Beven, 1993). In this study, a variant of Cameron *et al.*'s (1999) procedure for estimating flood frequency within the GLUE framework was used for each catchment. The procedure is illustrated in Fig. 2 and will now be summarised.

Five thousand TOPMODEL parameter sets, containing a fairly broad range of parameter values, are initially generated from independent uniform distributions. Four parameters are varied: the exponential scaling parameter (m), effective available water capacity of the root zone ($SRMAX$), mean log transmissivity of the soil at saturation of the surface ($\ln(T_0)$) and standard deviation of log transmissivity ($STDT$) (Table 4). Other parameters, such as those derived directly from the observed data (e.g. the mean hourly potential evapotranspiration parameter) are kept constant.

A single continuous simulation of the hourly observed series, utilising observed hourly rainfall inputs, is derived from each TOPMODEL parameter set using a 20 processor parallel Linux PC cluster available at the University of Lancaster, UK. Maximum likelihood is used to fit separate Generalised Extreme Value (GEV) distributions to the observed and individual simulated series of hourly annual maximum flood peaks. The performance of each parameter set is evaluated using the log likelihood function:

$$\begin{aligned} l(\rho) = \sum_{i=1}^{nrp} & - \log \alpha_s \\ & + (-1/k_s - 1) \, . \, \log\,[1 + k_s \, . \, (y_i - u_s)/\alpha_s] \\ & - [1 + k_s\,(y_i - u_s)/\alpha_s]^{-1/\mathrm{ks}} \end{aligned} \tag{1}$$

Where nrp is the number of annual maximum (ANNMAX) peaks in the GEV fit to the observed series with return periods of less than or equal to half of the observed hourly series length (nrp = 14, 6, 4, and 3, for the Tanllwyth, the Browney, the Findhorn and Dowles Brook, respectively), α_s, k_s and u_s are the scale, shape and location parameters of the GEV distribution fitted to the simulated series, and y_i is a peak extracted from the GEV distribution fitted to the observed series.

A parameter set is retained as *behavioural* if:

$$D_{pa} \leq TD \tag{2}$$

where D is the deviance calculated between the maximum value of $l(\rho)$ in the sample of five thousand parameter sets ($l(P)$), and the value of $l(\rho)$ for a given parameter set (pa), as:

$$D_{pa} = 2[l(P) - l(\rho)_{pa}] \tag{3}$$

and TD is a threshold deviance of 6.25 obtained from the χ^2 distribution at 3 degrees of freedom (for the GEV) and probability level $p = 0.9$. TD is constant for each catchment.

The parameter sets which are retained as *behavioural* under the flood peak criterion are also tested via a χ^2 statistic calculated between the observed and simulated flow duration curves. Thirteen points on the flow duration curve are used (*Q1*, *Q5*, *Q10–Q90*, *Q95* and *Q99*; where Q is a discharge and the associated value is the percentage of time that that discharge is equalled or exceeded over the course of the observed flow data series; see Table 1 for sample observed values), as:

$$\chi^2{}_{d,p} = \sum_{i=1}^{13} [(O_i - S_i)^2/S_i] \tag{4}$$

Where d is twelve degrees of freedom, $p = 0.9$, O_i is the observed percentage time spent beneath a given flow value, and S_i is the simulated percentage time spent beneath a given flow value. This yields a rejection threshold of 18.5 which is constant for each catchment. Parameter sets which provide simulations which meet, or fall below, this threshold are retained as *behavioural.*

This likelihood measure is used in preference to the continuous hourly hydrograph Nash and Sutcliffe (1970) efficiency because of the variable quality of the CAHR data and the consequent increase in timing errors.

To provide flood frequency estimates beyond the upper return period limit of the observed series, the stochastic rainfall generator is coupled with TOPMODEL. This requires the estimation of the rainfall model's GPD parameters prior to the runs with TOPMODEL and this is also achieved within the GLUE framework.

Five thousand GPD parameter sets are initially generated for each cdf upper tail requiring extrapolation. For the cases where the maximum likelihood estimate of the GPD fit requires an upper bound, σ is sampled from a uniform distribution within the range of three standard errors on either side of that estimate. Upper bounding is assumed and κ calculated. For the other cdf upper tails, both σ and κ are independently sampled from uniform distributions within the range of three standard errors on either side of the maximum likelihood estimate, and a parameter set is rejected as *non-behavioural* if the upper bound is exceeded. For these latter upper tails only, this more explicit form of bounding produces superior GPD fits to the data in comparison with those obtained through the first procedure.

The GPD parameter sets retained by each procedure are also evaluated in terms of providing a reasonable goodness-of-fit to the appropriate cdf upper tail. This is calculated

Table 5. Number of TOPMODEL parameter sets retained following rejection of *non-behavioural* parameter sets.

	Tanllwyth	Browney	Findhorn	Dowles Brook
Flood Peak	769	1457	2991	2641
Flow Duration Curve	727	1263	2892	520

using the log likelihood measure, $l(\theta)$, as:

$$l(\theta)_{du} = \sum_{i=1}^{np} - \log\sigma + (-1/\kappa - 1).\log[1 + \kappa.(x_i - u)/\sigma] \tag{5}$$

Where du is the particular cdf upper tail, κ is a shape parameter, u is a location parameter (or threshold), x_i is an event in the upper tail, np is the number of events in that tail, and x_i–u is an exceedance.

Rejection of the *non-behavioural* GPD parameter sets is achieved in an identical manner to that of the flood peak constraint. On a cdf upper tail basis, the deviance of a given value of $l(\theta)$ from the original maximum likelihood estimate is calculated and compared with a threshold deviance (Eqns. 2 and 3). A probability level of $p = 0.9$ with 2 degrees of freedom (for the GPD), yielding a threshold deviance value of 4.61, is utilised for the χ^2 distribution. The threshold deviance is consistent for each upper tail and for each catchment. Tables 3a and 3b contain the *behavioural* ranges of κ and σ, respectively.

A standard sample size of 1000 *behavioural* parameter sets for TOPMODEL, and for the rainfall model, are generated. For the catchments with greater than one thousand *behavioural* parameter sets retained from the initial sample of five thousand, a random sample of one thousand of those *behavioural* parameter sets is taken. For the other catchments, resampling is conducted in order to increase the numbers of *behavioural* parameter sets to the required level. In each case, the resampling procedure is identical to that of the original, with the exception that, for the purpose of efficiency, the new parameter sets are generated over parameter ranges which are consistent with those of the initial *behavioural* parameter sets. The new parameter sets are then evaluated as before.

For each catchment, the one thousand *behavioural* TOPMODEL and GPD parameter sets are used to calculate likelihood weighted uncertainty bounds for both the observed series length and the one thousand year flood frequency simulations. The former requires the TOPMODEL likelihood weights only and are calculated using the likelihood $L(\rho)$, defined as the exponential of $l(\rho)$, under the assumption that each likelihood is equivalent to a relative probability. The one thousand year simulations require the use of a combined measure (CM), which assumes equal weightings between the rainfall and TOPMODEL parameter sets:

$$CM = \exp\{l(\rho) + 1/nd.\sum_{i=1}^{nd}[l_s(\theta)_i]\} \tag{6}$$

Where nd is the number of cdfs requiring extrapolation (e.g. seven), and $l_s(\theta)$ is the rescaling of each value of $l(\theta)$ such that they share a common scale with $l(\rho)$.

A standard procedure (Freer *et al.*, 1996; Cameron *et al.*, 1999) is used to calculate the uncertainty bounds. $L(\rho)$ is used for the observed series length flood frequency simulations, and CM for the 1000 year flood frequency simulations. In each case, this involves the rescaling of the likelihood weights over all of the *behavioural* simulations in order to produce a cumulative sum of 1.0. A cdf of discharge estimates is constructed for each ANNMAX peak using the rescaled weights. Linear interpolation is used to extract the discharge estimate appropriate to cumulative likelihoods of 0.05, 0.5, and 0.95. This allows 90% uncertainty bounds, in addition to a median simulation, to be derived.

Following this procedure, it is also possible to make a

Table 6. Parameter ranges of *behavioural* TOPMODEL parameter sets.

	Tanllwyth	Browney	Findhorn	Dowles Brook
m **(m)**	0.0063:0.0394	0.0043:0.0331	0.0010:0.0424	0.0045:0.0128
S_{rmax} **(m)**	0.0010:0.1996	0.0012:0.1998	0.0010:0.1998	0.0142:0.1999
T_0 **(log)**	0.0001:7.9924	0.0120:7.9874	0.0072:7.9874	0.0133:7.9949
$STDT$ **(log)**	0.0047:9.9576	0.0043:9.9989	0.0043:9.9851	0.0018:9.9913

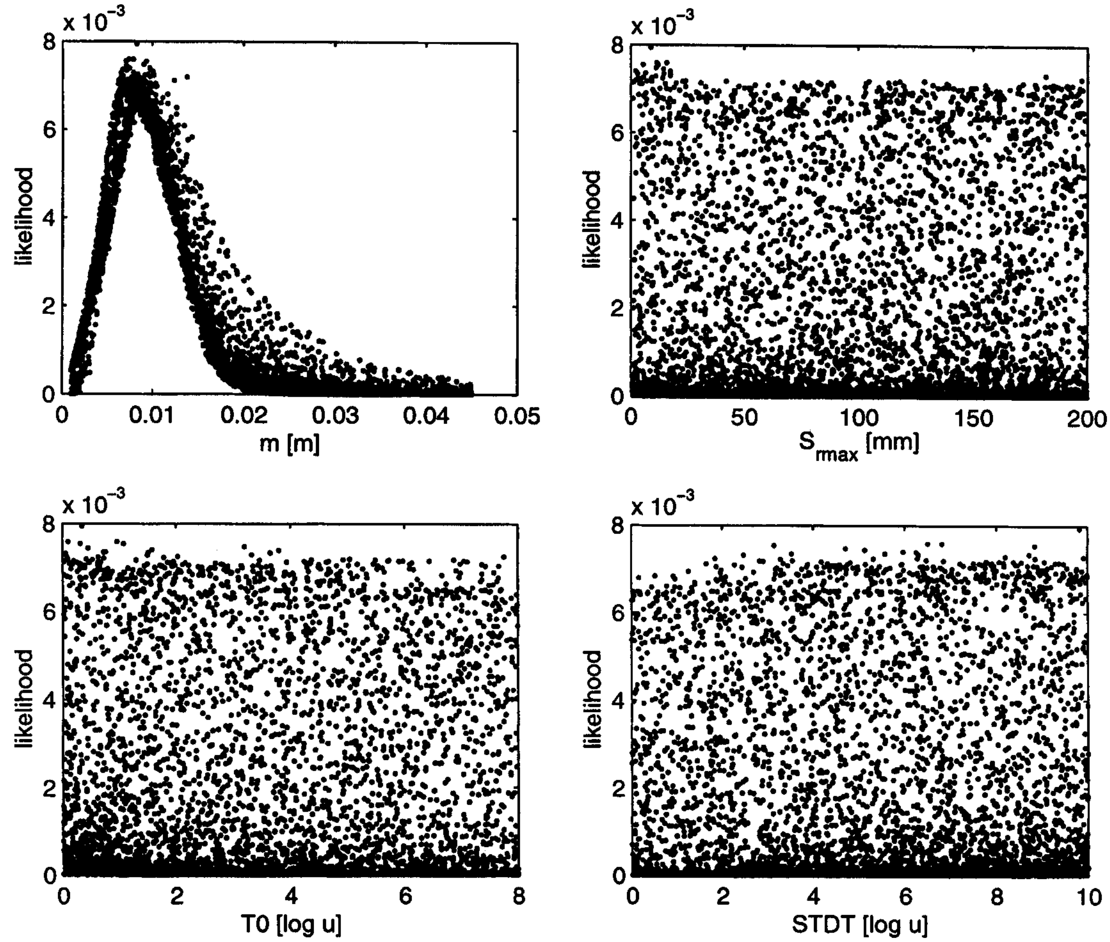

Fig. 3. *Example dotty plot for the Findhorn catchment showing the initial 5000 TOPMODEL parameter values vs L(ρ). Behavioural parameter sets are located at the top.*

comparison between the simulated hourly flood peak estimates and those of the extensive observed instantaneous peak record available at each site.

Results and discussion

Table 5 details the results of the initial flood frequency runs using the observed hourly rainfall series for each catchment. Given the initial five thousand parameter sets, it can be seen that the system of constraint is effective in rejecting *non-behavioural* parameter sets. It is notable, however, that the flood peak measure is not such a strong constraint for the catchments with shorter hourly data records. For example, 2991 of the initial 5000 TOPMODEL parameter sets generated for the Findhorn (which has 6 years of hourly data available) are retained on a flood peak basis. This contrasts with the 769 TOPMODEL parameter sets retained for the Tanllwyth (16 years of hourly record). Clearly the number of years of hourly record available is an important control upon the effectiveness of $l(\rho)$ at the chosen level of probability ($p = 0.9$).

For three of the four catchments (Tanllwyth, Browney and Findhorn), the flow duration curve measure rejects relatively few of the TOPMODEL parameter sets which have been accepted as *behavioural* under the flood peak likelihood measure. For example, 1263 (of 1457 flood peak *behavioural*) parameter sets are retained on a flow duration curve basis for the Browney (Table 5). This suggests that, given the likelihood measures and rejection thresholds used, there is a strong consistency between flood peak and flow duration curve parameterisation for these catchments (see also Cameron *et al.*, 1999). This finding also applies to the Dowles Brook, although it is apparent that the number of parameter combinations in the initial sample which are acceptable on the basis of both likelihood measures is much smaller than for the other three catchments (520 parameter sets of the initial sample are retained on a flow duration curve basis, Table 5). This result will be returned to below. Overall therefore, the constraining strategy is effective in retaining parameter sets which yield acceptable simulations of hourly ANNMAX flood peaks in a hydrologically convincing manner.

Table 6 details the parameter ranges of the one thousand

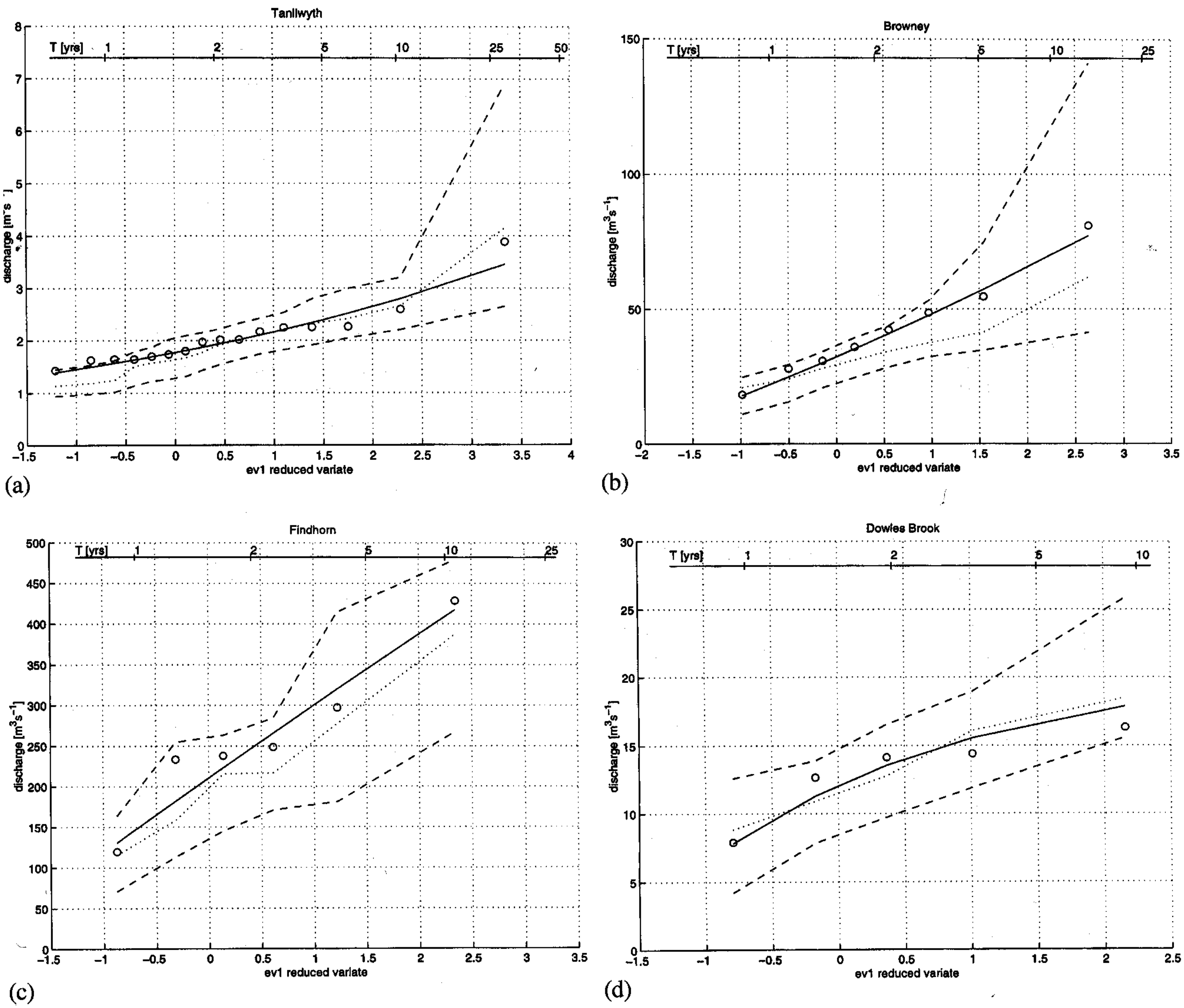

Fig. 4. *90% uncertainty bounds calculated from the annual maximum peaks obtained from 1000 behavioural TOPMODEL parameter sets run with observed hourly rainfall data. a: Tanllwyth. b: Browney. c: Findhorn. d: Dowles Brook. Circles – observed hourly peaks; solid line – GEV distribution fit to the observed hourly peaks; dashed lines – 90% uncertainty bounds; dotted line – median simulation. Return periods (T) are also included.*

behavioural TOPMODEL parameter sets for each catchment. A comparison with Table 4 yields some interesting similarities and differences. For instance, it can be seen that the majority of the *behavioural* parameter ranges (e.g. those of *SRMAX*, T_0 and *STDT*) are consistent with those of the original sampling ranges (Table 4) for the Tanllwyth, the Browney and the Findhorn. The main exception occurs with respect to the *m* parameter, which is important in controlling the rate of contributing area expansion and the hydrograph recession response.

It is also possible to examine this finding visually. This can be done by plotting the individual parameter values against the appropriate likelihood measure. An example of such a plot, for the Findhorn, is illustrated in Fig. 3. This contains the original five thousand parameter sets as plotted against $L(\rho)$ (the *behavioural* parameter sets are in the upper region of each plot). These dotty plots represent a projection of the n-dimensional parameter response surface onto a single parameter axis. Each dot represents a sample from the (complex) likelihood measure surface resulting from a model run with a specific parameter set. It can be seen that, for this catchment, and given the choice of likelihood measure, only the *m* parameter appears to demonstrate peaked behaviour (this is located between approximately 0.005 m and 0.010 m). The other three parameters exhibit a relatively flat upper limit, although *SRMAX* does display a small peak (between approximately 1 mm and 10 mm). Indeed, although such plots will not reveal any parameter interactions, they do reveal that *behavioural* values are located almost across the full spread of the respective parameter ranges. Similar results for $L(\rho)$ are also obtained for the other three catchments.

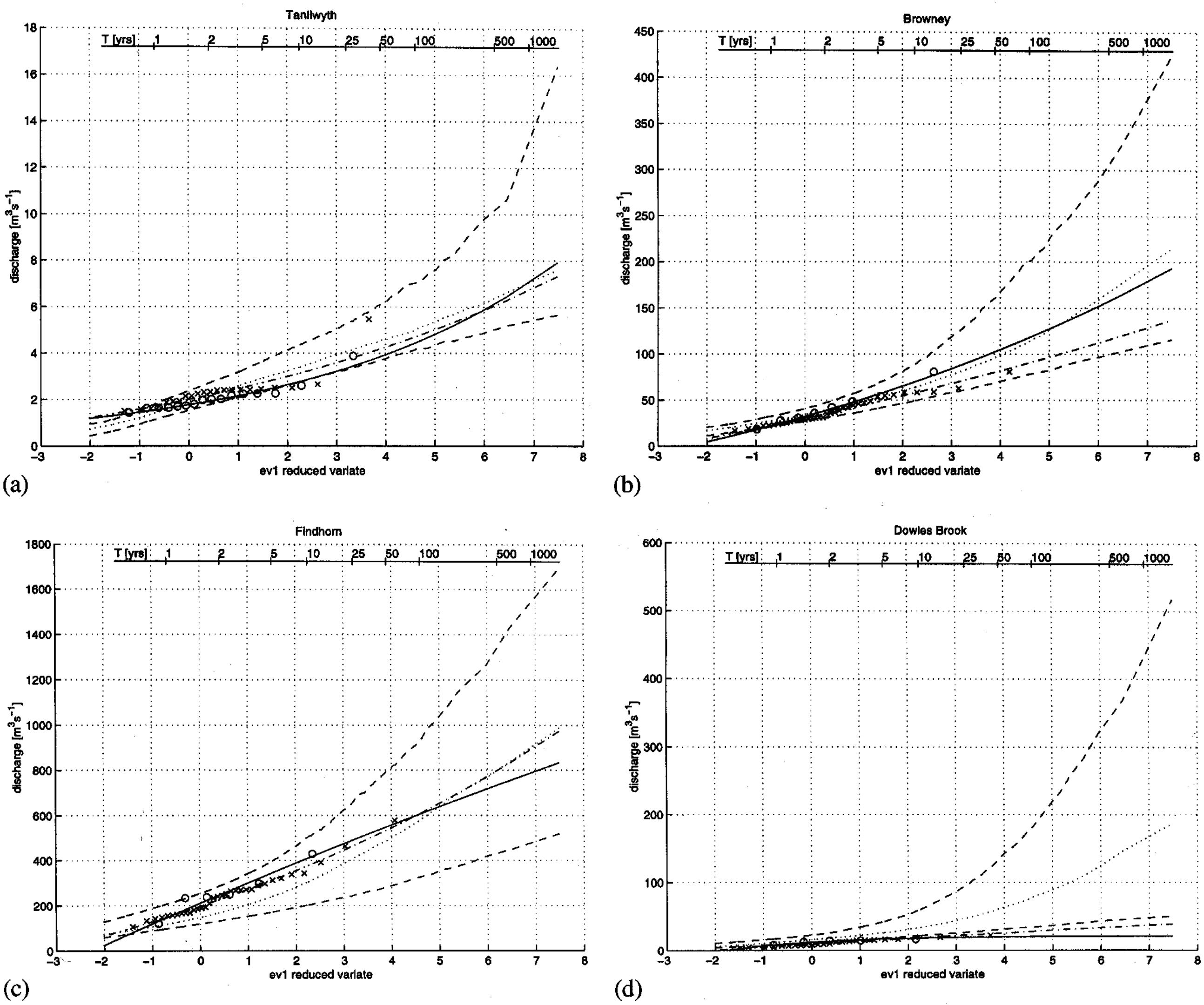

Fig. 5. *90% uncertainty bounds calculated from the annual maximum peaks obtained from 1000 behavioural TOPMODEL/rainfall parameter sets with 1000 year simulation length. a: Tanllwyth. b: Browney. c: Findhorn. d: Dowles Brook. Circles – observed hourly peaks; solid line – GEV distribution fit to the observed hourly peaks; crosses – observed instantaneous peaks; dash-dot line – GEV distribution fit to the observed instantaneous peaks; dashed lines – 90% uncertainty bounds; dotted line – median simulation. Return periods (T) are also included.*

Following the application of the flow duration curve constraint, the Dowles Brook results (Tables 5 and 6) demonstrate a reduction in the ranges of the *m*, *SRMAX* and T_0 parameters. This can perhaps be explained by the catchment's physical characteristics. Dowles Brook possesses a more permeable geology (sandstone, Table 1) than the other three catchments (e.g. the Tanllwyth is situated upon shales and grits). Consequently, it is perhaps unsurprising that the modelling of this catchment's flow duration curve characteristics can only be achieved successfully using a limited range of parameter values since the subsurface zone has a stronger control on the form of the hydrograph.

Figures 4a-d illustrate the 90% likelihood weighted flood frequency uncertainty bounds derived for the one thousand *behavioural* TOPMODEL parameter sets for each catchment using observed hourly rainfall inputs. The observed hourly ANNMAX peaks (circles), a GEV distribution fitted to those observed peaks (solid line), and a median simulation (dotted line), are also included. In general, it can be seen that the model performs well. For all four catchments there is good "bracketing" of the observed peaks and associated fitted GEV distributions.

Figures 5a–d detail the likelihood weighted uncertainty bounds obtained from the 1000 year simulations for each catchment. These were derived from the one thousand *behavioural* TOPMODEL and rainfall parameter sets. Both the observed hourly and the more extensive series of instantaneous ANNMAX peaks (and associated fitted GEV distributions) are included in each case. The instantaneous peak series are not directly comparable to the hourly timestep simulations (Beven, 1987), especially for small

Table 7. Forms of the GEV distributions fitted to the observed hourly and instantaneous ANNMAX peaks at each site.

	Tanllwyth	Browney	Findhorn	Dowles Brook
Hourly	EV2	EV2	EV3	EV3
Instantaneous	EV2	EV1	EV2	EV1

flashy catchments such as the Tanllwyth. However, they can be used as independent data sets for testing the validity of the simulations. The differences between the uncertainty bounds (with respect to the period of the observed hourly series) of these 1000 year simulations and those of the earlier observed rainfall-driven simulations (Figs. 4a-d) occur as a result of different rainfall realisations (see Cameron *et al.*, 1999).

For the Tanllwyth, Browney and Findhorn (Figs. 5a–c), the bounding of both observed series and their associated fitted GEV distributions is quite reasonable. Only a limited number of peaks (e.g. between return periods of approximately 6.5 and 14 years, or plotting positions of 1.8 to 2.6, on Fig. 5a) lies outwith the bounds. The estimates of long return period floods (e.g. 100 years or a plotting position of 4.6) are subject to wider uncertainty bounds for the catchments with the shortest records (e.g. the Findhorn, Fig. 5c).

The Dowles Brook results (Fig. 5d) are also reasonable with respect to the observed hourly peaks. However, beyond a plotting position of approximately 2.0 (or an 8 year return period), the 5% uncertainty bound is somewhat higher than the corresponding fitted GEV distribution. This also applies to the observed instantaneous peaks and their fitted GEV distribution. It is likely that these results stem from the short period of observed hourly record available for use in constraint for this catchment (Table 1). The assumption that the GPD rainfall constraint is consistent across different duration classes and across catchments may also be a contributing factor.

Finally, it is worth considering the forms of the GEV distributions fitted to the hourly and instantaneous peak data at each site (Table 7). It is interesting to note that, with the exception of the Tanllwyth (which possesses a relatively long observed hourly record, Table 1), there is a change in form between the short record of hourly peaks and the much longer record of instantaneous peaks. This is particularly significant for the Findhorn and the Dowles Brook, where the apparent EV3 hourly response has given way to the more extreme EV2 and EV1 distributions, respectively. This finding highlights the difficulty of accurate estimation of long return period events from limited data using traditional statistical techniques. While regional analysis (e.g. Hosking and Wallis, 1997) may be of assistance, it is probable that the continuous simulation techniques used in this study represent a more hydrologically meaningful methodology for the estimation of flood frequency at a particular site.

Conclusions

This paper has applied a variant of the continuous simulation methodology developed by Cameron *et al.* (1999) for hourly annual maximum flood frequency estimation to four gauged catchments in the UK. This is done within an uncertainty framework that avoids the idea that there is an optimal set of model parameter values. The methodology uses simulations based on both observed rainfalls and a stochastic rainstorm generation model, for extension to long return periods, to drive the rainfall-runoff model TOPMODEL. The strategy for uncertainty estimation incorporates conditioning on rainfall statistics, annual maximum flood peaks and flow duration curve simulation.

In applications of the methodology to the Tanllwyth, Browney, Findhorn and Dowles Brook catchments, good results were obtained for all four sites, particularly those with long observed hourly record lengths. The problem of accurate estimation of flood events of long return period (e.g. 100 years) via a traditional statistical approach, given limited data, was also illustrated. In particular, it was proposed that the continuous simulation approach yielded more hydrologically meaningful flood estimates than those obtained via the traditional techniques. Taken together with Cameron *et al.*'s (1999) findings for the Wye catchment, these results suggest strongly that the continuous simulation methodology used is acceptable as an alternative method for flood frequency estimation. Given a reasonable quantity of hourly rainfall and flow data, and an appropriate choice of rainfall-runoff and stochastic rainfall models, it is probable that this approach could be adapted easily for use in many other regions of the world.

Acknowledgements

The authors wish to thank the following authorities for the catchment data: the Northumbrian and Severn-Trent region NRAs (now part of the Environment Agency in England and Wales, UK), the Highland region RPB (now part of SEPA in Scotland, UK), and the Institute of Hydrology. The comments of two anonymous referees enhanced the clarity of the original manuscript. David

Cameron's contribution to this work was carried out at Lancaster University and the Institute of Hydrology under the NERC CASE studentship GT4/97/112/F.

References

Beven, K., 1986. Hillslope runoff processes and flood frequency characteristics. In A.D. Abrahams (Editor), *Hillslope Processes*, Allen and Unwin, pp 187–202.

Beven, K.J., 1987. Towards the use of catchment geomorphology in flood frequency predictions, *Earth Surf. Proc.* **12**, 69–82.

Beven, K.J., 1993. Prophecy, reality and uncertainty in distributed hydrological modelling, *Adv. in Water Resour.* **16**, 41–51.

Beven, K.J. (Editor), 1997. *Distributed Modelling In Hydrology: Applications of the TOPMODEL Concepts*, Wiley, Chichester UK.

Beven, K.J. and Binley, A., 1992. The future of distributed models: model calibration and uncertainty prediction, *Hydrol. Process* **6**, 279–298.

Beven, K., Lamb, R., Quinn, P., Romanowicz, R. and Freer, J., 1995. TOPMODEL. In, *Computer Models of Watershed Hydrology*, V.P. Singh (Ed.) Water Resources Publications, pp. 627–668.

Blazkova, S. and Beven, K.J., 1997. Flood frequency prediction for data limited catchments in the Czech Republic using a stochastic rainfall model and TOPMODEL, *J. Hydrol.* **195**, 256–278.

Bradley, A.A. and Potter, K.W., 1992. Flood frequency analysis of simulated river flows, *Water Resour. Res.* **28**, 2375–2385.

Bras, R.L., Gaboury, D.R., Grossman, D.S. and Vicen, G.J., 1985. Spatially varying rainfall and flood risk analysis, *J. Hydraul. Eng.* **111**, 754–773.

Calver, A. and Lamb, R., 1996. Flood frequency estimation using continuous rainfall-runoff modelling. *Phys. Chem. Earth* **20**, 479–483.

Cameron, D.S., Beven, K.J., Tawn, J., Blazkova, S. and Naden, P., 1999. Flood frequency estimation for a gauged upland catchment (with uncertainty), *J. Hydrol.* **219**, 169–187.

Diaz-Granados, M.A., Valdes, J.B. and Bras, R.L., 1984. A Physically Based Flood Frequency Distribution, *Water Resour. Res.* **20**, 995–1002.

Eagleson, P.S., 1972. Dynamics of food frequency. *Water Resour. Res.* **8**, 878–898.

Franks, S.W., Gineste, P., Beven, K.J. and Merot, P., 1998. On constraining the predictions of a distributed model: the incorporation of fuzzy estimates of saturated areas into the calibration process, *Water Resour. Res.* **34**, 787–798.

Freer, J., Beven, K. and Ambroise, B., 1996. Bayesian uncertainty in runoff prediction and the value of data: an application of the GLUE approach, *Water Resour. Res.* **32**, 2163–2173.

Gupta, V.K. and Dawdy, D.R., 1995. Physical Interpretations Of Regional Variations In The Scaling Exponents Of Flood Quantiles, in, *Scale Issues In Hydrological Modelling, Advancés In Hydrological, Process*, Kalma, J.D. and Sivapalan, M., (Eds.) 105–119, Wiley, Chichester, UK.

Hebson, C. and Wood, E.F., 1982. A derived flood frequency distribution. *Water Resour. Res.* **18**, 1509–1518.

Hosking, J.R.M. and Wallis, J.R., 1997. *Regional Frequency Analysis*, Cambridge University Press, Cambridge, UK.

Huff, F.A., 1967. Time distribution of rainfall in heavy storms. *Water Resour. Res.* **9**, 227–238.

Lamb, R.L., 1999. Calibration of a conceptual rainfall-runoff model for flood frequency estimation by continuous simulation, *Water Resour. Res.* **35**, 3103–3114.

Lamb, R. and Gannon, B., 1996. The establishment of a database of hourly catchment average rainfall and flow for flood frequency estimation by continuous simulation. *Ministry of Agriculture, Fisheries and Food Project FD0404 Milestone Report.*

Michaud, J. and Sorooshian, S., 1993. Comparison of simple versus complex distributed runoff models on a midsized semiarid watershed, *Water Resour. Res.* **30**, 593–605.

Naden, P.S., Crooks, S.M. and Broadhurst, P., 1996. *Impact of climate and land use change on the flood response of large catchments.* Proc. 31st MAFF Conference of River and Coastal Engineers, Keele, UK, July 1996, pp. 2.1.1–2.1.16.

Nash, J.E. and Sutcliffe, J.V., 1970. River flow forecasting through conceptual models 1. A discussion of principles. *J. Hydrol.* **10**, 282–290.

Quinn, P.F. and Beven, K.J., 1993. Spatial and temporal predictions of soil moisture dynamics, runoff, variable source areas and evapotranspiration for Plynlimon, mid-Wales, *Hydrol. Process.* **7**, 425–448.

Quinn, P.F., Beven, K., Chevalier, P. and Planchon, O., 1991. The prediction of hillslope flow paths for distributed hydrological modelling using digital terrain models, *Hydrol. Process.* **5**, 59–79.

Quinn, P., Beven, K.J. and Lamb, R., 1995. The ln $(a/\tan(\beta)$ index:how to calculate it and how to use it within the TOPMODEL framework, *Hydrol. Process.* **9**, 161–182.

Rodriguez-Iturbe, I. and Valdes, J.B., 1979. The geomorphologic structure of hydrologic response, *Water Resour. Res.* **15**, 1409–1421.

Sivapalan, M., Wood, E.F. and Beven, K.J., 1990. On Hydrologic Similarity 3. A dimensionless flood frequency model using a generalized geomorphological unit hydrograph and partial area generation, *Water Resour. Res.* **26**, 43–58.

Thompson, N., Barrie, I.A. and Ayles, M., 1985. The Meteorological Office Rainfall and Evaporation Calculation System MORECS, *Hydrological Memorandum No.* **45**, U.K Meteorological Office.