A joint probability approach to flood frequency estimation using Monte Carlo simulation

T. R. Kjeldsen, C. Svensson and D. A. Jones
Centre for Ecology & Hydrology, Maclean Building, Crowmarsh Gifford, Wallingford, OX10 8BB, UK

Email of corresponding author: trkj@ceh.ac.uk

Abstract

In the UK, flood estimation using event based rainfall-runoff modelling currently assigns pre-defined design values to the input variables which control the size of the flow events, apart from the rainfall magnitude which is treated as a random variable. The use of design values, rather than allowing the variables to be described by their full probability distribution, is a practical simplification but may lead to biases in the output flood magnitudes. The present study simulates a large number of flow events using sets of input variables from distributions fitted to observed event data, taking into account seasonality. These simulated datasets are used for running a rainfall-runoff model, and a frequency analysis is applied to the peaks of the output flow hydrographs. The simulated inputs are the rainfall intensity and duration, and the soil moisture deficit (SMD) and initial river flow at the beginning of the rainfall event. An inter-event arrival time is simulated so that a series of events is obtained. The initial conditions of SMD and river flow of each event are made dependent on the (simulated) time elapsed since the previous event, and on the SMD at the end of the previous event.

Accepted for publication in Proc. of the BHS Third International Symposium: Role of Hydrology in Managing Consequences of a Changing Global Environment, Newcastle University, Newcastle upon Tyne, United Kingdom, 19-23 July 2010.

http://www.ceg.ncl.ac.uk/bhs2010/

1. Introduction

Eagleson (1972) provided a foundation for estimating a flood frequency relation in the absence of streamflow records by deriving it from the density functions for climatic and catchment variables. Since then, joint probability methodologies have been investigated in a number of studies (e.g. Arnaud and Lavabre, 2002; Rahman *et al*., 2002; Paquet *et al*., 2006). The method presented in this paper is a prototype for a novel statistical approach for modelling extreme fluvial flood events by combining an existing rainfall-runoff model with a joint probability description of the major flood-producing variables.

The current UK standard method for event-based flood modelling is the revitalised FSR/FEH rainfall-runoff method (Kjeldsen *et al.*, 2006), which is an improvement of the model first presented in the Flood Studies Report (FSR) (NERC, 1975). For design-event modelling, the rainfall and other inputs are not provided by a stochastic model but instead are set to carefully selected values which supposedly lead to a flood event of a given return period. Despite their widespread use these models have not been designed through consideration of the joint probabilities of the different flood-producing variables such as soil moisture content and rainfall depth, duration and profiles. Instead, justification for the values chosen for model inputs relies to some extent on intuition and on comparing the estimated flows with estimates from other sources, in cases where these are available. The joint probability approach presented here has the potential to allow the eventual creation of a methodology notionally similar to the design-event method but it would be one with a firmer theoretical justification. The catchment of the river Blyth at Hartford Bridge (National River Flow Archive flow gauge number 22006) in north east England is used to illustrate the method and results presented in this paper.

2. Simulation strategy

The method generates a string of individual events over a pre-defined time period. The hydrograph for each event is simulated using the PDM model (e.g. Moore, 2007) with boundary conditions generated using a set of pre-defined stochastic models. A separate stochastic model defines the inter-event arrival time (IEAT), i.e. the time between the end of an event and the onset of the next event.

The required boundary conditions to be generated from stochastic models are:

- Rainfall duration (D) [hours]
- Rainfall intensity (I) [mm h^{-1}]
- Soil moisture deficit (SMD) at onset of event [mm]
- Initial flow (qs) [m^3 s^{-1}]
- Inter-event arrival time (IEAT) [hours]

In addition, the soil moisture deficit at the end of each flood event, SMD^*, is required, but this quantity is derived from the output of the PDM model, and it does not require a separate stochastic model. The sequential relationship between the boundary conditions and the simulated flood hydrographs is illustrated in Figure 1.

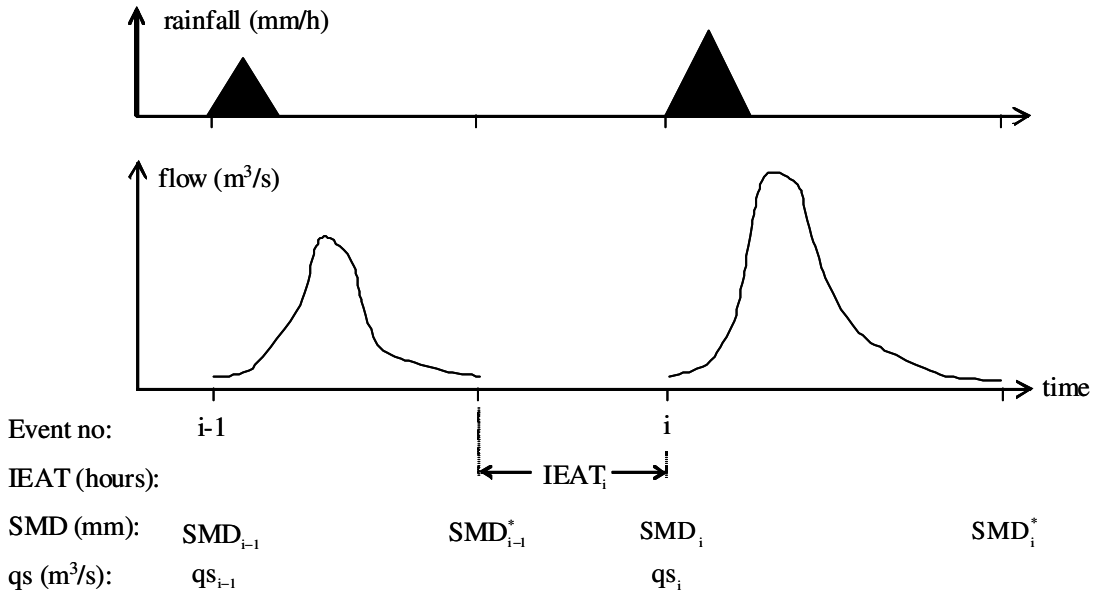

Figure 1 Sequential relationship between boundary conditions and flood hydrographs.

When generating the flood hydrographs, it is important to consider the appropriate conditional probability models for each of the boundary conditions. Notably, the SMD at the onset of an event depends on the SMD at the end of the previous event and the time elapsed between the end of the previous event and the start of the current event, IEAT. Also, each of the boundary conditions depends on the time of year (season) of occurrence of each particular event.

The conditional relationships between the boundary conditions are illustrated in Figure 2, where the upper line represents the boundary conditions that must be generated using stochastic models. On the lower line, the PDM model operator, *MPDM*, will generate output time series of flow and soil moisture deficit for each event. From these time series the maximum simulated flow, QMAX, and the soil moisture deficit at the end of the event, SMD*, can be extracted before moving on to the next event. Note that the arrows in Figure 2 only serve to illustrate the sequence of operations in the simulation, and they do not imply any sort of functional relationship.

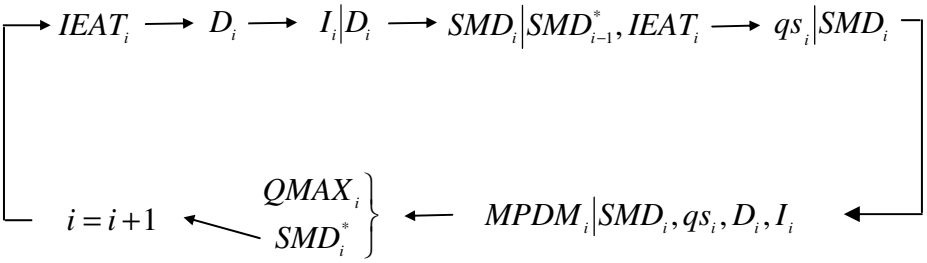

Figure 2 Simulation procedure and conditional relationship between boundary conditions.

Using a Monte Carlo approach, a string of events can be generated for a time period of a given duration (e.g. ten years), and several of these time periods can be generated from which the mean flood statistics and the associated standard deviations can be

estimated. Because the output of the method consists of entire flood hydrographs, flow characteristics such as total volume and hydrograph shape can be estimated. However, this paper presents results for the flow peak only.

3. Stochastic models of boundary conditions

3.1 Selection of observed events for model fitting

A stochastic model has to be specified for each of the five boundary conditions listed in Section 2 above. The models have all been fitted to a dataset consisting of on average 10 events per year. For the example in this paper, the river Blyth at Hartford Bridge, this means a total of 170 events occurring from January 1985 to December 2001.

The observed events were selected based on rainfall data, using a trial-and-error approach to ensure that large river flow events as well as large rainfalls were selected. Continuous hourly series of catchment average hourly rainfalls (CAHR) were used to define rainfall events, for which subsequently the associated river flow peak, as well as the SMD and river flow at the start and end of the rainfall event, were extracted (all at an hourly resolution). The SMDs were derived through continuous simulation using a version of the PDM, with hourly series of CAHR and monthly MORECS evaporation data as input.

The initial rainfall event selection required at least one hourly rainfall total within the event to exceed a certain threshold (5% of the 2-year return period 1-hour CAHR), and a running mean rainfall over a certain number of hours (75% of the catchment response time) to exceed a lower threshold. This rainfall event definition resulted in a very large number of events, from which on average 10 events per year were selected based on the largest total event rainfall. Time series of rainfall, river flow and SMD were plotted, and selected events marked, to allow a visual inspection of whether the largest river flow peaks in each year were among the selected events.

Looking at each variable separately, there is no serial (rank) correlation (at the 5% significance level) between the rainfall totals of the selected events, whereas the SMDs and initial flows of successive events do show dependence. For cross-variable dependence, the initial flow and the SMD at the start of each event show a strong linear relationship in log space.

3.2 Rainfall duration and intensity

The selection of rainfall events results in a lower bound for both event duration, D, and depth, P. Thus, to enable the Monte Carlo method to reproduce the characteristics of the observed flood events, a set of marginal distributions were adopted for intensity and duration. For duration, the shortest possible event is 1 hour, corresponding to the resolution of the rainfall data. For rainfall depth, the lower bound was defined as the minimum observed depth, P_{min}. Consequently, shifted duration and depth are defined as

$$D' = D - 1 \quad (1)$$

$$P' = P - P_{\min} \quad (2)$$

resulting in a transformed intensity defined as

$$I' = \frac{P'}{D'} = \frac{P - P_{\min}}{D - 1} \quad (3)$$

The transformed duration and intensity are uncorrelated (using Spearman's rank correlation). The marginal distributions are modelled using an exponential distribution for the intensity and a gamma distribution for the duration. This represents an initial model that may need to be improved to take into account a possible decrease in variability of the intensity with increasing duration.

Figure 3 shows histograms of transformed rainfall duration (x-axis) and intensity (y-axis) for the Blyth at Hartford Bridge catchment, distinguishing between summer and winter events. Separate models have been defined for each season and the model parameters are shown in Table 1. Figure 3 includes scatter plots for summer and winter of the observed (transformed) intensity and duration (coloured points). The grey crosses on Figure 3 represent stochastically generated rainfall events based on 100 Monte Carlo samples of 16 water years (1 October – 30 September), and suggest a reasonable correspondence between the observations and the simulated values.

A complete rainfall event will also require consideration of the temporal profile of the event. At this stage, a triangular profile has been adopted for each event.

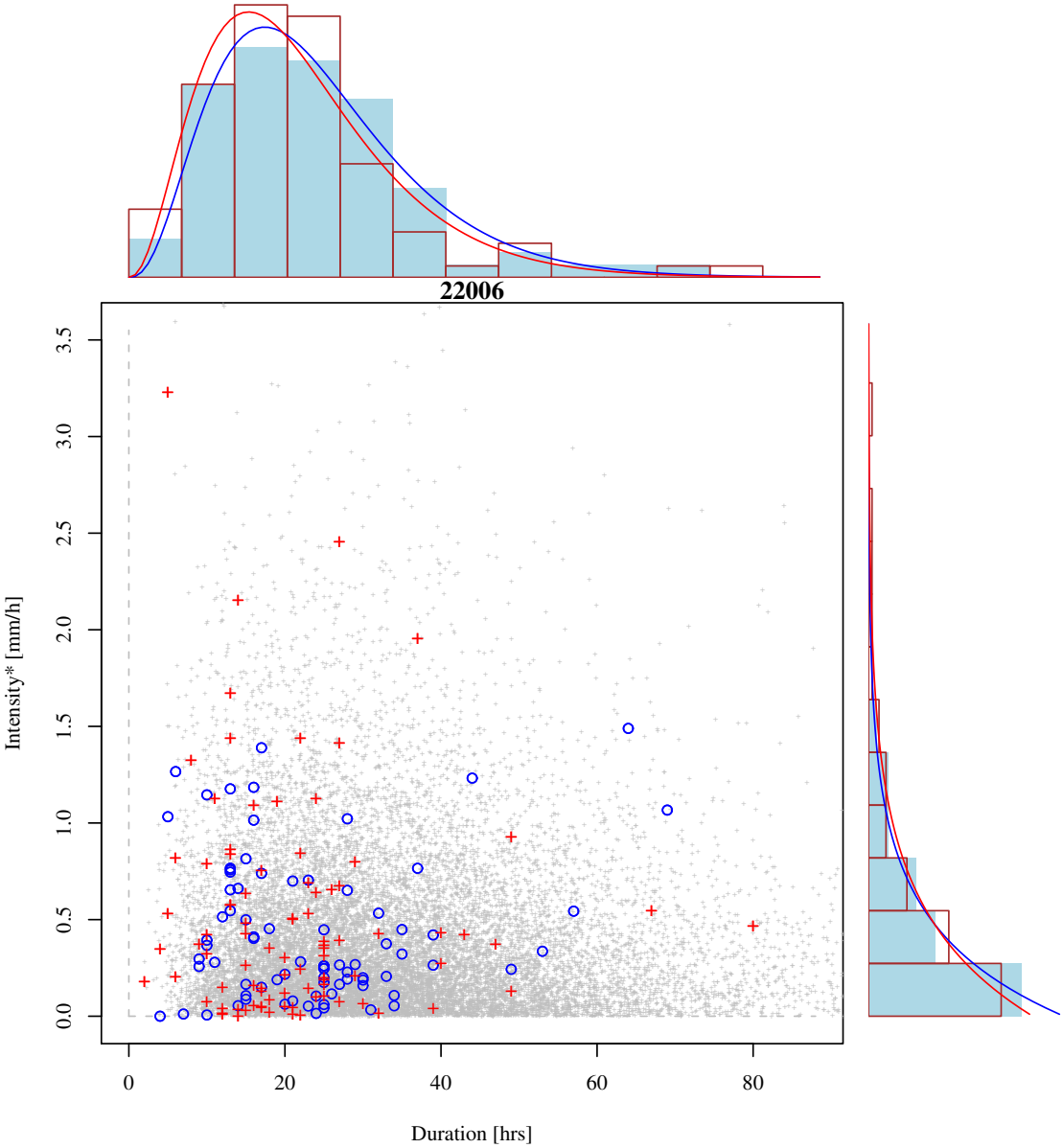

Figure 3 Marginal distributions of transformed observed rainfall duration, *D*', (x-axis) and intensity, *I*', (y-axis) for summer (red) and winter (blue) rainfall events. Simulated events are shown as grey crosses.

Table 1 P_{min}, number of events and model parameters for distributions of transformed duration, *D*' (gamma), and intensity, *I*' (exponential).

	Winter	*Summer*
Duration, (shape parameter)	3.6516	3.3398
Duration, (scale parameter)	6.4628	6.4607
Intensity, (scale parameter)	0.4323	0.5145
Lower bound of depth, P_{min}	13.44	13.35
Number of events	80	90

3.3 Inter-event arrival time

Histograms of IEAT for summer and winter events are shown in Figure 4. An exponential distribution and a two-parameter gamma distribution were fitted to each of the two datasets using the method of moments, and the associated probability density functions have been plotted in Figure 4. There is little difference between the one- and two-parameter distributions, and the one-parameter exponential distribution was selected: the model parameters for this model for the two seasons are shown in Table 2.

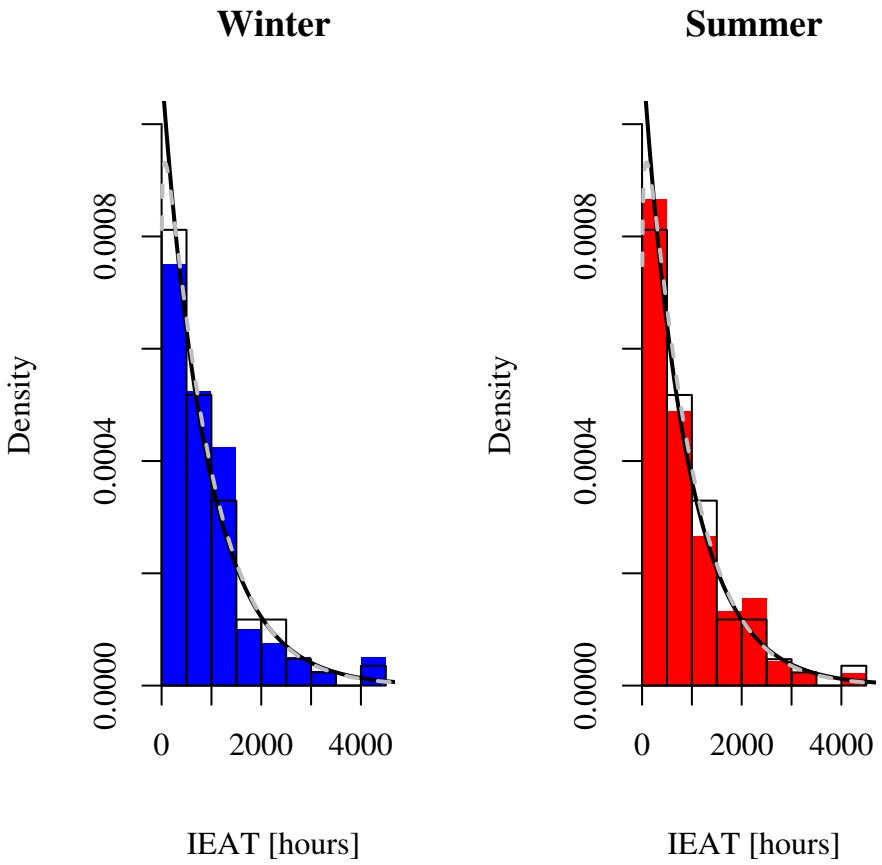

Figure 4 Histograms of IEAT for the winter (blue) and summer (red) seasons, showing the probability density functions of the exponential distribution (solid black line) and the gamma distribution (grey broken line). The black-line histograms represent the entire dataset, i.e. without seasonal considerations.

Table 2 Parameters for the IEAT model (exponential distribution) for the winter and summer seasons.

	Winter	*Summer*
Scale	910.95	881.10

3.4 Soil moisture deficit

A model for generating an appropriate value of soil moisture deficit at the onset of an event has been defined as a typical value of SMD at a particular time of year combined with a deviation from this typical value depending on the time elapsed since the end of the previous event, and the SMD at the end of the previous event.

$$\ln\left[\frac{SMD_i}{S_{\max} - SMD_i}\right] = \mu(f) + \exp(-\theta_4 IEAT_i)\left(\ln\left[\frac{SMD^*_{i-1}}{S_{\max} - SMD^*_{i-1}}\right] - \mu(f^*)\right) + \varepsilon_i \tag{4}$$

where $S_{\max}$ is the total available soil moisture storage obtained through calibration of the PDM model. In the above expression, the term *f* is the time of year (as a fraction of a year) of the onset of event *i*, and *f** is the corresponding time at the end of event *i*-1. The typical soil moisture deficit value at any time of the year is treated using the mean value, $\mu(f)$, of the variable $\ln[SMD/(S_{\max} - SMD)]$, which is modelled as

$$\mu(f) = \theta_1 + \theta_2 \sin(2\pi f) + \theta_3 \cos(2\pi f). \tag{5}$$

The model error ε_i is assumed to follow a normal distribution, and to be independent and identically distributed, with a mean value of zero and constant variance σ^2.

The five free model parameters (θ_1 to θ_4 plus the model error variance σ^2) are estimated using a maximum-likelihood method. Figure 5 shows the residuals, ε_i, plotted against time of year, and the resulting model parameters are shown in Table 3.

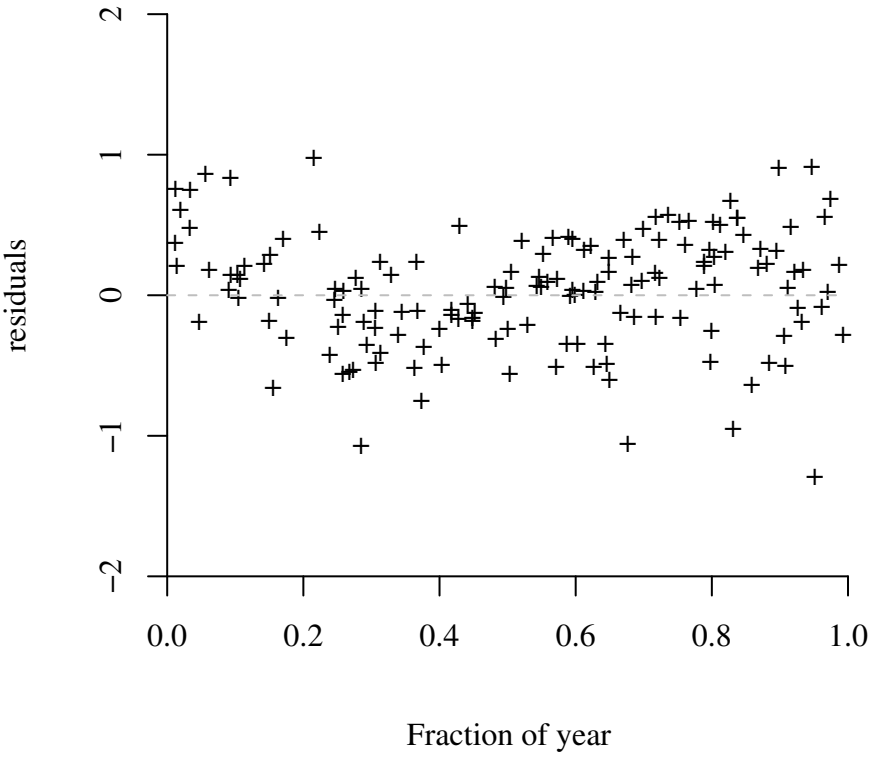

Figure 5 Residuals from SMD model fitted to events.

Table 3 Model parameters for the SMD model.

Coefficient	*Value*	*Std. error*	*t-value*	*p-value*
θ_1	1.3206	0.0576	-23.682	$< 2\times10^{-16}$
θ_2	0.1864	0.0721	-3.640	3.8×10^{-4}
θ_3	-0.9493	0.0704	-13.476	$< 2\times10^{-16}$
θ_4	0.0012	0.00001	85.070	$< 2\times10^{-16}$
σ^2	0.1682	0.0192	8.739	4.5×10^{-15}

3.5 Initial flow values

The initial flow, qs_i, is modelled as a function of the SMD at the onset of the event i and depends on the time of year as

$$\ln[qs_i] = \phi_0 + \phi_1 \ln[SMD_i] + \phi_2 \sin(2\pi f) + \phi_3 \cos(2\pi f) + \eta_i \quad (6)$$

where ϕ is a vector of model parameters and η is a set of errors that are independent and identically distributed, and follow a normal distribution. The model parameters are estimated using least-square regression, and the results are shown in Table 4.

Table 4 Model parameters for the initial flow model fitted to events.

Coefficient	*Value*	*Std. error*	*t-value*	*p-value*
ϕ_0	4.5703	0.3857	11.849	$<2\times10^{-16}$
ϕ_1	-1.3137	0.1095	-11.998	$<2\times10^{-16}$
ϕ_2	0.03243	0.0879	0.369	0.713
ϕ_3	-0.0527	0.1196	-0.411	0.660

The residual standard error for the model is 0.6691, and the fraction of variance explained is 0.715.

4. Results

The event-based PDM model used for the Monte Carlo simulation has the same parameters as the PDM that was calibrated to fit the continuous hourly time series of river flows, with hourly rainfall and evaporation data as input. The models for SMD and initial flow were also specified using series from the continuous simulation. The river flow peaks output from the joint probability analysis will therefore be compared with the river flow peaks extracted from the continuously modelled river flow series. Ideally, there should be a good agreement between the observed river flow peaks and the continuously simulated ones, but there is some discrepancy as the PDM parameter optimisation was carried out using a least sum of squares approach for the whole flow series. In practice, this means that apart from the flow peak, the calibration took into

account other considerations, such as the length and decay of recessions and the overall water balance.

Using the Monte Carlo framework, one hundred time series (or strings of events) were generated, each with a duration of 16 years (the same as the simulated record of complete water years). From each of the generated time series the annual maximum peak flows were extracted and plotted (grey lines in Figure 6). This Figure also includes the observed annual maximum peak flow series (crosses) and the annual maxima of the flows modelled using continuous simulation with the observed rainfall series (circles). Flows are plotted against return periods defined via Gringorten plotting positions on a Gumbel scale.

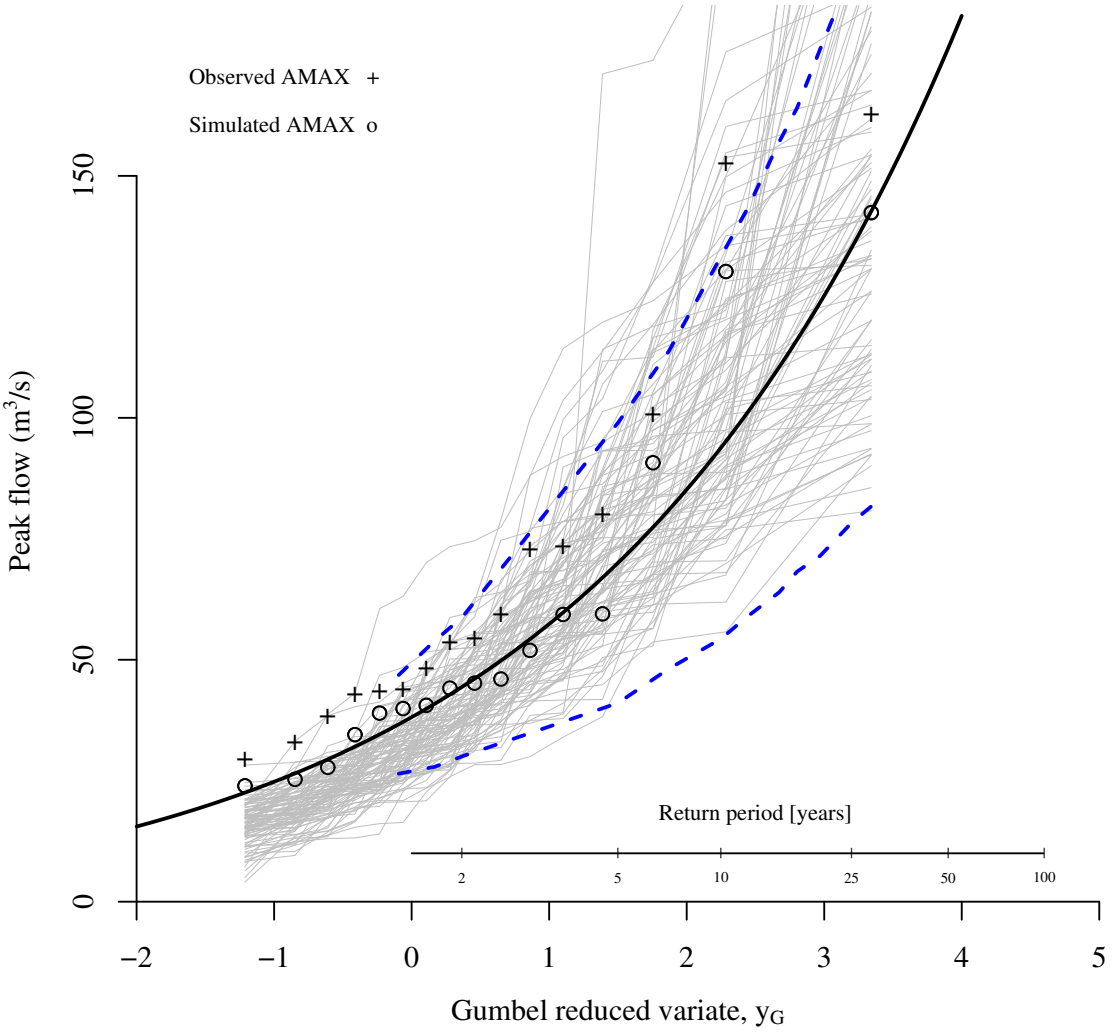

Figure 6 Comparison of observed and generated annual maximum peak flow

The thick black line in Figure 6 is the generalised extreme value (GEV) distribution fitted to the annual maxima from the continuously simulated flow series using a maximum-likelihood method, and the blue dashed lines delineate 95% confidence intervals for the true flow frequency curve of the simulated flows derived by bootstrapping using 999 replications. The results from the joint probability method (thin grey lines) agree well with the annual maxima from the simulated series (circles)

and the fitted GEV and confidence intervals. However, the observed annual maxima (crosses) are larger than those from the continuously simulated series. As a comparison, the at-site fitted frequency curve from the FEH statistical method would fit the observed maxima closely.

Acknowledgments

The work carried out for this study was funded by the Natural Environment Research Council's thematic programme Flood Risk from Extreme Events (FREE), grant number NE/F001037/1. Hourly river flow data was supplied by the Scottish Environment Agency (EA) and catchment average hourly rainfalls were calculated using hourly and daily rainfall data from the EA and from the UK Met Office, respectively. Both datasets were originally compiled for the Defra-funded project FD 2106. Monthly MORECS evaporation data were supplied by the UK Met Office.

References

Arnaud, P. and Lavabre, J. 2002. Coupled rainfall model and discharge model for flood frequency estimation. *Water Resour. Res*., **38**(6), 11-1 – 11-10.

Eagleson, P. S. 1972. Dynamics of flood frequency. *Water Resour. Res*., **8**(4), 878-898.

Kjeldsen, T. R., Stewart, E. J., Packman, J. C., Folwell, S. and Bayliss, A. 2006 *Revitalisation of the FSR/FEH Rainfall-Runoff Method*. R&D Technical Report FD1913/TR, Department of Environment Food and Rural Affairs, CEH Wallingford, 133pp.

Moore, R. J. 2007. The PDM rainfall-runoff model. *Hydrol. Earth Syst. Sci*., **11**(1), 483-499.

NERC 1975. *Flood Studies Report* (five volumes). Natural Environment Research Council, UK.

Paquet, E., Gailhard, J. and Garçon, R. 2006. Evolution de la métode du gradex: approche par type de temps et modélisation hydrologique. *La Houille Blanche*, **5**, 80-90.

Rahman, A., Weinmann, P. E., Hoang, T. M. T. and Laurenson, E. M. 2002. Monte Carlo simulation of flood frequency curves from rainfall. *J. Hydrol*., **256**, 196-210.