

Calibration of a conceptual rainfall-runoff model for flood frequency estimation by continuous simulation

Robert Lamb

Institute of Hydrology, Oxfordshire, England, United Kingdom

Abstract. An approach is described to the calibration of a conceptual rainfall-runoff model, the Probability Distributed Model (PDM), for estimating flood frequencies at gauged sites by continuous flow simulation. A first step was the estimation of routing store parameters by recession curve analysis. Uniform random sampling was then used to search for parameter sets that produced simulations achieving the best fit to observed, hourly flow data over a 2-year period. Goodness of fit was expressed in terms of four objective functions designed to give different degrees of weight to peaks in flow. Flood frequency results were improved, if necessary, by manual adjustment of parameters, with reference to peaks extracted from the entire hourly flow record. Although the primary aim was to reproduce observed peaks, consideration was also given to finding parameter sets capable of generating a realistic overall characterization of the flow regime. Examples are shown where the calibrated model generated simulations that reproduced well the magnitude and frequency distribution of peak flows. Factors affecting the acceptability of these simulations are discussed. For an example catchment, a sensitivity analysis shows that there may be more than one set of parameter values well suited to the simulation of peak flows.

1. Introduction

Flood magnitudes and return periods may be estimated by the frequency analysis of continuous synthetic flow data generated by a rainfall-runoff model. This approach, which may be contrasted with design flood estimation on the basis of the simulation of individual events, has been demonstrated by a number of authors using various hydrological models [see, for example, *Bras et al.*, 1985; *Beven*, 1987; *Bradley and Potter*, 1992; *Calver and Lamb*, 1996; *Blazkova and Beven*, 1997]. Use of a model with continuous water balance accounting has the attractive feature that dynamic factors affecting runoff production may be represented implicitly. In particular, a conceptual rainfall-runoff model can simulate the changing antecedent moisture conditions that determine the storm runoff response for a given rainfall event. The problem of base flow estimation, present in event-based design flood approaches, is also avoided. Continuous simulation of flows also avoids any need to associate design flood estimates with a specific design storm. Flood estimates may instead be derived indirectly from continuous rainfall records or synthetic data supplied by a rainfall generator that is considered able to reproduce actual rainfall distributions.

The continuous simulation method has the potential to be generalized. For certain catchments a suitable model may already have been identified. Assuming that rainfall and potential evaporation data can be obtained, simulation results then depend on the parameters of the chosen hydrological model. Estimation of these parameters is therefore important, and this, in turn, has implications for choice of model structure.

Model parameters may be estimated for a gauged site by calibration. For an ungauged site, estimates might be obtained by some method on the basis of the translation or regionalization of values established at gauged catchments (with an ex-

pected reduction in confidence). In either case, it is appropriate to choose a relatively simple, standardized model structure with few parameters. In the case of calibration using observations, it has been argued that rainfall and flow data do not support the identification of meaningful values for a large number of parameters [e.g., *Jakeman and Hornberger*, 1993]. For ungauged sites it is likely that estimation of parameters on the basis of values established by calibration at gauged catchments will require standardization of model structure (see, for example, regionalized applications of a simple, general model structure by *Post and Jakeman* [1996] and *Sefton and Boorman* [1997]).

This paper presents examples of results from the most extensive application, to 40 gauged catchments, of the continuous simulation method in Great Britain. It is part of a wider study which aims to develop the method for ungauged catchments by relating model parameter values calibrated at the gauged sites to physical catchment properties. For the reasons outlined above, a standardized rainfall-runoff model was chosen, namely, a simple formulation of the Probability Distributed Model (PDM) of *Moore* [1985, 1993], which has been used in real-time flood warning applications [e.g., *Institute of Hydrology*, 1995; *Moore et al.*, 1994].

This study advances previous work using the continuous simulation method by exploring the model calibration problem specifically in this context. The availability of a relatively large number of sites with good, continuous hourly data means that it has been possible to test calibration procedures according to a range of criteria and present results for a few sites that illustrate more typical issues. Main issues addressed are the usefulness of different objective functions for numerical calibration, the ability to satisfy different calibration criteria jointly to increase confidence in model predictions, and the non-uniqueness of parameter estimates.

2. PDM Rainfall-Runoff Model

The PDM has a simple, conceptual structure, shown in Figure 1. Rainfall first enters a soil moisture store, where the

Copyright 1999 by the American Geophysical Union.

Paper number 1999WR900119.
0043-1397/99/1999WR900119\$09.00

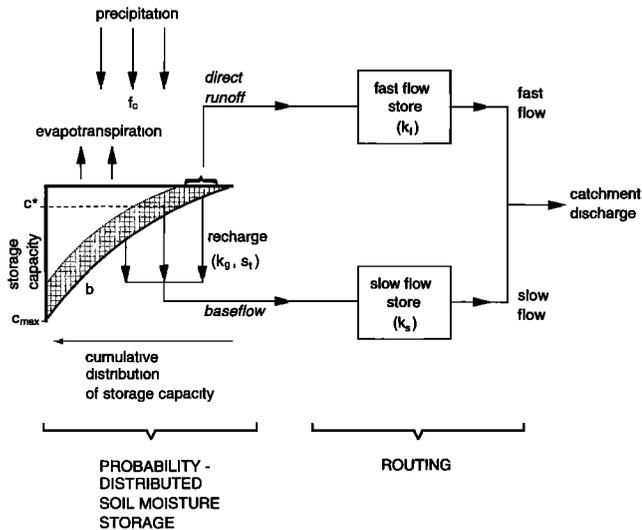


Figure 1. Structure of the Probability-Distributed Model (PDM).

distribution of storage capacity over the catchment is described by a specified function. The soil moisture store permits runoff to respond nonlinearly to rainfall inputs by varying the proportion of the catchment, generating fast, “direct” runoff according to the amount of water stored in the soil. Direct runoff is routed via a fast flow store before contributing to the total catchment runoff. The remaining rainfall enters the soil moisture store, which is depleted by evaporation. Drainage from the soil moisture store (representing “recharge”) takes place via a simple delay function and generates base flow after being routed by a “slow flow” store.

The main components of the PDM are fairly generic stores that may be characterized using a variety of different functions to describe the soil storage capacity distribution, the drainage of the soil moisture store, and the routing stores [*Institute of Hydrology, 1996*]. A particular choice of functions may be best suited to any given catchment. However, for this study, the choice was determined by the need for a standardized model with a small number of parameters. Given that a number of different configurations of the PDM are available, and have been used in practice, the formulation chosen for this study will be described briefly, highlighting the parameters to be estimated.

The specific soil moisture storage capacity c [L] was assumed to be described by a Pareto distribution, for which the distribution function F is

$$F(c^*) = Pr(c \leq c^*) = 1 - [1 - (c^*/c_{\max})]^b, \quad (1)$$

where c_{\max} [L] is the maximum storage capacity in the catchment and the parameter b [–] controls the variability of storage capacity. If c^* is the storage capacity depth below which saturation occurs, then $F(c^*)$ gives the proportion of the catchment contributing to direct runoff. *Moore* [1985, 1993] describes how the state of the soil moisture store is integrated over the distribution at each time step to calculate c^* .

Rainfall inputs to the soil moisture store were multiplied (at every time step) by a rainfall “correction” parameter f_c [–], introduced notionally to compensate for errors arising in the estimation of spatially averaged rainfall from point gauge data. Losses due to evaporation were calculated as a linear function of potential evaporation, estimated by the U.K. Meteorological

Office at synoptic sites, and the status of the soil moisture store. Drainage to the slow flow path, d [$L T^{-1}$], was also represented by a linear function of soil moisture storage s [L] such that

$$d = (s - s_t)/k_g, \quad (2)$$

where the parameters are a time constant k_g [T] and a threshold storage s_t [L] below which there is no drainage. Direct runoff from the proportion of the catchment where storage capacity has been exceeded is routed through the fast flow store, which was assumed to be described by the linear relationship

$$q_f = s_f/k_f, \quad (3)$$

where the specific fast flow q_f [$L T^{-1}$] is the contribution of the store to catchment runoff, s_f [L] is the store content, and the parameter k_f [T] is the time constant of the store. The slow or base flow component, q_s [$L T^{-1}$], of the total runoff was assumed to be routed through an exponential store such that

$$q_s = \exp(s_s/k_s), \quad (4)$$

where s_s [L] is the depth of storage and k_s [L] is the decay parameter of the store. Total runoff q was computed as the sum of q_f and q_s .

A time step length of 1 hour was adopted for modeling, following the conclusion of *Spijkers and Naden* [1994] that a daily time step would imply too much smoothing for the estimation of flood quantiles. Areally averaged rainfall inputs were calculated for each catchment using the method of *Jones* [1983b], which averages data from triangles of rain gauges surrounding points in a mesh superimposed on the catchment. Each gauge is weighted according to the inverse-square distance from a given mesh point. The algorithm also mixes information from daily and hourly gauges to maintain consistency between times when different hourly gauges may be operating.

3. Peaks Over Threshold Flood Frequency Analysis

Continuous hourly flow and rainfall data were acquired for each of the catchments used in this study (details of the data collection and quality control procedures are given by *Lamb and Gannon* [1996]). A record of ~10 years duration (1985–1995) was obtained for most of the sites. The relatively short length of these records, in terms of flood frequency studies, prompted the use of the peaks-over-threshold (POT) method to fit flood frequency distributions to the observed and simulated flow series. Single-site POT analyses have been investigated theoretically by a number of authors [*Davison, 1984; Davison and Smith, 1990; Smith, 1984; Wang, 1991b*]. Here the methods for POT analysis reported by *Naden* [1992] were followed.

For both simulated and observed flow series the magnitudes of the POT data were fitted using the generalized Pareto distribution (GPD), with the number of peaks per year assumed to correspond to a Poisson distribution. The combination of these two assumptions is equivalent to the use of the generalized extreme value distribution for annual maxima [*Smith, 1984; Wang, 1991b*]. Fitting was carried out using the method of probability weighted moments [*Hosking and Wallis, 1987*]. Rather than specifying an arbitrary flow threshold for the ex-

traction of peaks for each catchment, peaks were instead extracted at an average rate of 3 yr^{-1} (i.e., 30 peaks would be extracted from a period of 10 years, but not necessarily three peaks from every calendar year). To ensure that the extracted peaks represented independent events, a minimum separation period was imposed, specified for each catchment as three times a typical event time to peak.

It is perhaps worth emphasizing that the aim here is to investigate model calibration in the specific context of deriving a flood frequency curve by continuous simulation, with a consequent emphasis on peak flows. For this reason, no detailed analysis is presented to assess the suitability of the distributions fitted to the POT data, nor the implicit flow thresholds resulting from the average extraction rate of 3 peaks yr^{-1} . However, an important consideration is that exactly the same techniques were applied to both simulated and observed flow data in the frequency analysis. Also, *Naden* [1992, 1993] has presented evidence on the basis of the analysis of POT data from 826 stations that these techniques are likely to be reasonable for the purposes of this study, that is, single-site analyses in the United Kingdom for relatively short return periods.

4. Calibration Methods

In principle, calibration of a rainfall-runoff model using continuous data may take account of more information regarding runoff-generating processes than would be contained in a series of annual maximum or POT data alone. However, it may not always be possible for a single model structure and set of parameter values to reproduce all aspects of the flow regime to the same standard. Although greatest weight has to be given here to simulation of peak flows, confidence in simulated flood frequency distributions will be increased if other aspects of the simulated flow series are also realistic.

Several objective functions were therefore used in an automated procedure to assess model performance in simulating different aspects of the flow regime. In addition, two model parameters were fixed directly from the data by deterministic reasoning, and a final element of subjective assessment was included via manual parameter adjustment. This approach, combining automated and manual calibration using many objective functions, has been exploited in the past, for example, in the calibration of the U.S. National Weather Service flood forecasting model [*Brazil and Hudlow*, 1981; *Brazil*, 1988].

The methods adopted here are a straightforward, heuristic response to a multiple-objective problem, where compromise solutions are sought between several different aims. More formal approaches to hydrological model calibration as a multiple-objective problem are also possible [see, for example, *Yapo et al.*, 1998]. Recently, *Gupta et al.* [1998] have argued that the existence of multiple objectives is a fundamental aspect of the calibration problem for hydrological models. This may be particularly notable for simulations of several different output fluxes or "internal state" variables. The problem described in this paper is an example of a particular case where multiple objectives can also be readily identified for simulation of a single output variable.

The existence of several compromise solutions in calibration may be reflected in uncertainty about predicted variables. A second key aspect of model calibration is uncertainty arising because of data errors, interactions between model parameters, and "model error." Such issues have been discussed by *Beven* [1993], *Beven and Binley* [1992], *Jones* [1983a], *Klepper et*

al. [1991], *Kuczera* [1983, 1988], *Romanowicz et al.* [1994], *Spear and Hornberger* [1980], and *van Straten and Keesman* [1991]. An uncertainty framework was not adopted here, however, because of an initial requirement to obtain a single "most acceptable" simulation for each catchment for use in later work on spatial generalization.

4.1. Estimation of Storage Constants by Recession Curve Analysis

The PDM, as described above, has seven parameters to be calibrated, including the constants in the two routing stores (k_f and k_s). However, the parameters of a conceptual routing store can be estimated directly from flow recession curves, if these are assumed to correspond to drainage from the store during periods of negligible input (or recharge) and low evaporation. Under these assumptions the storage function may be solved with a continuity equation for an input of zero, resulting in an expression for the decay of discharge over time. This recession equation can be fitted to observed recession curves to estimate the parameters of the original storage function. For the linear store, equation (3), the recession equation for flow at time $t + T$, expressed in terms of the flow at time t , T time units earlier, is

$$q_{t+T} = q_t \exp[-(T/k_f)]. \quad (5)$$

The recession equation for the exponential store (equation (4)) is

$$q_{t+T} = q_t [1 + (q/k_s)T]^{-1}. \quad (6)$$

It is difficult to specify exact, objective criteria to identify periods of the record where recharge and evaporation can be assumed to have little effect on flows, but two useful guides will be low rainfall and low potential evaporation. An interactive, semiautomated system, described by *Lamb and Beven* [1997], was used to search for recession curves during such periods.

A difficulty in associating recession curve data with two or more conceptual stores is the need to separate the total flow into components related to each store. The procedure used [*Lamb and Beven*, 1997, Figure 2] was designed to exclude direct storm runoff from a base flow "master recession curve" without recourse to hydrograph separation formulae. The master curve is constructed by first sorting individual recessions into a sequence, arranged in ascending order of the discharge at the end of each recession period. Starting from the lowest flow rate, and working toward higher flows, the data from each individual recession are then transferred to the master curve, up to the flow rate at the end of the next recession in the sequence; at this point the remaining data from the first recession are discarded and flow data from the next recession are transferred to the master curve.

The base flow master recession was found to be consistent with the slow flow routing store (equations (4) and (6)) for most catchments. The parameter k_s was therefore estimated for each catchment from its base flow master recession. To estimate appropriate values for k_f , the recession data excluded from the base flow master recession curves were plotted for each catchment on a logarithmic scale so that periods corresponding to equation (5) would appear as straight lines. Taking care to exclude data immediately following peaks in the rainfall, the transformed recessions were compared and (5) was fitted to a period judged to be representative.

The recession analysis technique involves an element of sub-

jectivity. However, this was felt to be justifiable in order to reduce the number of parameters requiring calibration and also in view of other potential sources of uncertainty, both in the data and due to the simplified representation of the catchment in the model. Although approximate, careful recession curve analysis should at least help to ensure that derived storage parameters are capable, in combination with other parameters, of generating realistic flow simulations representative of the runoff responses of a catchment.

4.2. Uniform Random Sampling

Random values for each of the remaining five parameters (f_c , c_{max} , b , k_p , and s_t) were drawn from uniform distributions over specified ranges, where the extremes were set according to past experience [Institute of Hydrology, 1996]. Each of 10,000 sets of five randomly generated parameter values was supplied to the PDM, which was then run over a 2-year calibration period. A length of 2 years was chosen to avoid excessive run times (in the context of calibration for a relatively large number of sites). Although computationally more expensive than a direct search optimization procedure, uniform random sampling (URS) allows a more complete exploration of the parameter space and can be applied to each of a number of catchments using the same initial parameter ranges.

The URS method would not be recommended for an entirely automatic optimization scheme because it does not sample the parameter space in a highly efficient manner. Other studies [e.g., Duan *et al.*, 1992; Wang, 1991a] have used population-evolution strategies, such as the genetic algorithm approach, to refine parameter estimates. In the present case, however, URS was applied primarily as an easily implemented way to locate starting points for interactive "fine tuning" of parameters; this manual adjustment stage was carried out using the entire period of hourly data available for each site. For this reason also, testing of the sensitivity of the URS algorithm results to alternative calibration periods was not considered although all flow and rainfall records were scrutinized to avoid calibration periods that were thought to be unrepresentative of the longer-term data.

4.3. Objective Functions

Four objective functions were evaluated for each simulation. The functions were selected to give different degrees of weight to errors associated with predictions of peaks in the measured flow data. The function least explicitly weighted toward simulation errors around the observed peaks was the efficiency measure of Nash and Sutcliffe [1970] (NSE):

$$NSE = 1 - \left[\frac{\sum_{t=1}^n (Q_t - q_t)^2}{\sum_{t=1}^n (Q_t - \bar{Q})^2} \right], \quad (7)$$

where Q_t is the observed flow at time t , q_t is the simulated flow, and n is the number of time steps. The value of NSE increases as the similarity between simulated and observed flows increases and would be equal to one for an error-free simulation.

Greater weight may be given to errors in the simulation of high flows by multiplication of the absolute difference between observed and simulated flows with the observed flow, raised to a power p . This results in an objective function, denoted the sum of weighted absolute errors (WAE):

$$WAE = \sum_{t=1}^n Q_t^p |Q_t - q_t|. \quad (8)$$

The exponent p was set subjectively to equal 1.5, after trials on a number of catchments, to provide increased sensitivity to peaks in flow without becoming entirely insensitive to recession and low-flow periods.

Equations (7) and (8) give some weight to the simulation errors at low and intermediate flows. Two objective functions were also chosen to incorporate only the errors in simulating a POT series, extracted in the same way as for a frequency analysis. The first of these functions attempts to combine information about both the timing and magnitudes of the first J ranked POT peaks by considering each pair of observed and simulated peaks, in order of rank, as a vector in two dimensions (time and discharge). The error for the pair of peaks of rank j was expressed as:

$$\varepsilon_j = \sqrt{[(Q_j - q_j)/q^*]^2 + [(T_j - t_j)/t^*]^2}, \quad (9)$$

where Q_j is the magnitude of the observed peak, q_j is the magnitude of the simulated peak, T_j is the time of occurrence of the observed peak, and t_j is the time of the simulated peak. The differences between observed and simulated variables were standardized by q^* , the difference between the largest and smallest flow magnitudes in the observed POT series, and t^* , the difference between the time of occurrence of the earliest and latest of the observed POT series. Equation (9) was summed over J peaks to obtain the objective function value, termed POT-QT.

The fourth objective function was based only on the errors in the magnitudes of the first J ranked pairs of POT peaks, irrespective of timing. This function, POT-SAE, is given by

$$POT-SAE = \sum_{j=1}^J |Q_j - q_j|. \quad (10)$$

In applying (9) and (10), J was chosen to correspond to the POT extraction rule of an average rate of 3 peaks yr^{-1} .

4.4. Rejection of Poor Parameter Sets and Manual "Fine Tuning"

The value of the NSE statistic (equation (7)) will be zero for a simulated flow series equal to the mean of the observations at every time step. Given that (7) encompasses the entire range of flow magnitudes, and is therefore sensitive to hydrograph shape, there would be no point in using a conceptual runoff model having seven parameters when better overall performance could apparently be obtained by the mean of the observations. Any set of parameters giving rise to $NSE \leq 0$ in the URS procedure was therefore rejected.

Parameter sets corresponding to the best values of the four objective functions were then used in simulations of the entire period of record. Flood frequencies were plotted for observed and simulated flows. By examining these results along with time series plots and flow duration curves, one of the four parameter sets was selected as a candidate for manual adjustment, if necessary, to improve the fit of the simulated and observed flood frequency plots, while attempting to maintain acceptable results in terms of time series plots and flow duration curves.

5. Results

Results for all of the catchments in the study were inspected according to the criteria listed in the following headings and a small number of examples selected to illustrate specific issues that were noted, arising in various combinations, for other catchments.

5.1. Comparison of Simulated and Observed POT Series

A simple test of the continuous simulation method is to compare the flood quantiles of the simulated and observed flows. Comparisons can be made between the fitted frequency distributions or, more directly, between the POT magnitudes. Figure 2 shows POT data (symbols) and fitted frequency distributions (curves) for the Mellte at Pontneddfechan, a 65.8-km² catchment in Wales. These flood frequency results were calculated using simulated and observed flow data for the period January 1985 to May 1995. The heavy, solid line and solid circles are the results from analysis of the observed flow data; the remaining results correspond to simulated flows, generated using the parameter sets that achieved the best results in the URS procedure for each of the four objective functions, established by applying the URS procedure to the shorter period 1989–1990.

It is immediately apparent from Figure 2 that the two objective functions NSE and WAE have not been as useful as the functions based only on six POT peaks, extracted from the same period. In particular, the parameter set found using the NSE statistic results in a consistent, significant underestimation of peak flow magnitudes. The best estimate of the observed POT series has been obtained using the parameters found by POT-QT objective function. The frequency curves corresponding to the observed and POT-QT peaks seem to fit the point data fairly well although there are some deviations in both observed and simulated cases at the higher return periods.

For the Mellte catchment a good flood frequency result was obtained by the automated URS procedure. However, for most catchments (32 out of 40) some manual adjustment was needed. This was done by matching simulated and observed flood frequency data by eye and often achieved by adjusting the parameter f_c only. One such catchment was the 0.9-km² Tanllwyth, part of the Plynlimon catchment experiment [Hud-

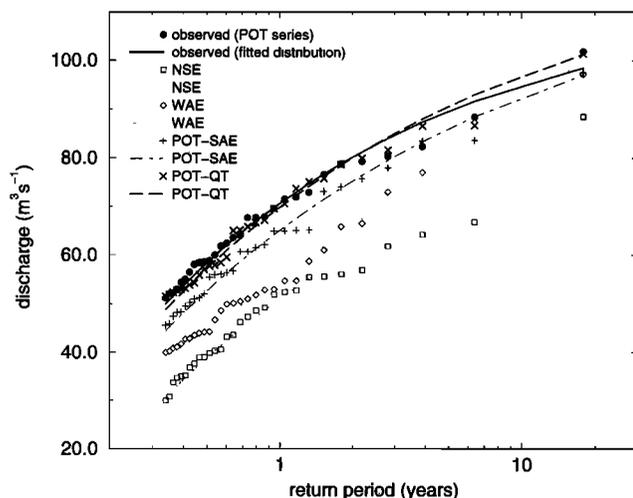


Figure 2. Observed and computed flood frequency data for the Mellte catchment.

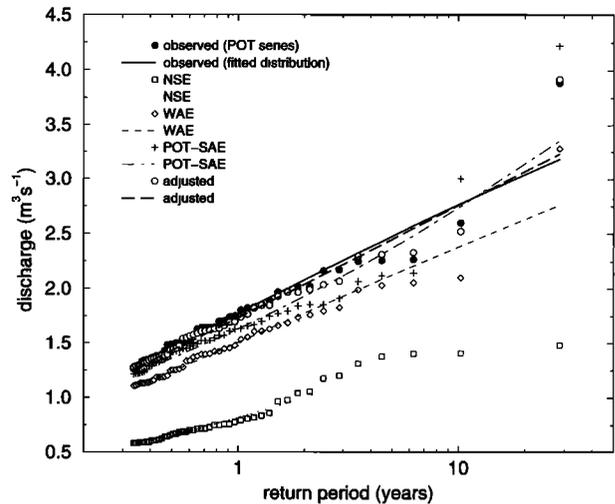


Figure 3. Observed and computed flood frequency data for the Tanllwyth catchment.

son and Gilman, 1993]. Data from the period 1988–1989 were used in the URS procedure. Flood frequencies were then calculated using the data from the period 1974–1989 and are shown in Figure 3. (Results have not been shown for the POT-QT function, as these coincided closely with those obtained by the WAE function but were associated with a less realistic flow time series.)

It can be seen that the NSE objective function again failed to identify a parameter set capable of reproducing the magnitudes of the largest peaks in the observed flow data. Results obtained using the POT-SAE function are close to the observed peaks in Figure 3. However, the parameter set found using the WAE function was chosen as a basis for final “tuning” because it gave rise to a fitted frequency distribution of similar slope to the observed data, appeared to generate the correct difference between the magnitudes of the two highest peaks, and generated a realistic overall flow time series. The adjusted parameters give rise to a good fit between the observed and simulated POT series (indicated in Figure 3 by solid and open circles, respectively).

Figure 4 shows flood frequency results, evaluated over the period 1985–1993 for the Frome at Ebley Mill, a 198-km² catchment in the west of England that has a significant groundwater component in the flow regime. Here, results obtained using the automated URS procedure (applied to the period 1987–1988) are shown only for the WAE and POT-SAE objective functions, which performed best. These results were not considered acceptable, and manual adjustment was therefore required. Although the POT-SAE objective function gave rise to better estimates of flood peak magnitudes over the duration of the full data record, inspection of the distribution of the POT series, when plotted against return period, suggests that the parameter set identified by the WAE function might be a better basis for manual adjustment using the POT data. A simple increase in the value of f_c , from 0.79 to 0.97, resulted in the much improved estimated flood quantiles, indicated by the open circles in Figure 4.

5.2. Comparison of Simulated and Observed Flood Frequency Distributions

In the results shown in Figures 2, 3, and 4, there is a good agreement between the best set of simulated POT peaks and

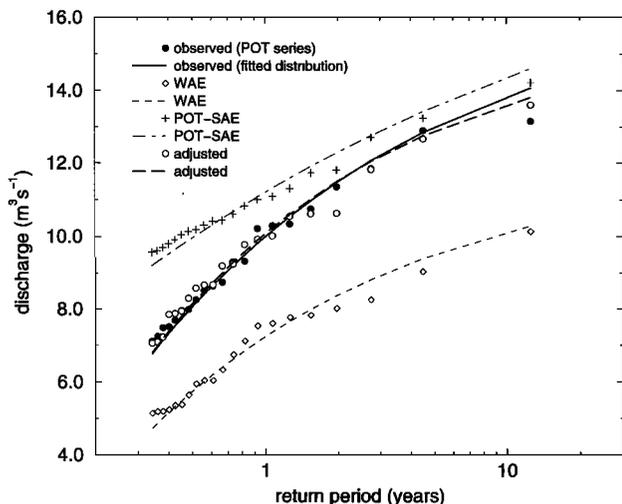


Figure 4. Observed and computed flood frequency data for the Frome catchment.

the observed peaks. It follows that there is also close agreement between the frequency distributions fitted to the observed and simulated peaks, suggesting that the corresponding parameter sets are useful for the simulation of flood frequency characteristics. However, it is important to note that confidence in calibrated model parameters that give rise to well-fitting flood frequency distributions may also depend on how well each distribution (simulated or observed) characterizes the respective flood peaks.

For the Tanllwyth catchment (Figure 3) the highest magnitude event is not close to, but clearly has an influence on, the fitted frequency curve, for both the observed data and the simulated flows. However, the fitted frequency curves appear to be consistent with the distribution of the many lower-return period peaks. In this case, given observations only, there could be doubts about the value of the fitted frequency curve or about the accuracy of the highest peak in the observed record. Doubts about the accuracy of the observed peaks may be dispelled, to some extent, by the ability of the PDM to simulate flow peaks of similar magnitude corresponding to the same storm events that generated the observed flows. Extrapolation of the fitted frequency curves would nonetheless require some care.

Good agreements can also be obtained between frequency distributions fitted to the observed and simulated flow peaks despite significant differences between the two POT series. An example is shown in Figure 5, where flood frequency distributions have been plotted for the River Ythan at Ellon, a 523-km² catchment in Scotland. The results of the numerical URS procedure suggested that there would be problems in obtaining an acceptable set of parameter values capable of reproducing the two largest events in the available record. This was confirmed during attempts to adjust the parameter sets manually. No combination of parameter values could be found that seemed capable of correctly reproducing the magnitudes of both the lower and the higher return period peaks extracted from the observed flow data.

Agreement can also be obtained between the frequency curves fitted to observed and simulated POT peaks because of interaction between the errors in simulating certain flood peaks, as shown in Figure 5. Such a result leads to questions over the reliability of the flood estimates obtained by analysis

of the simulated flows. One might accept the simulation results because they are reasonably consistent with the frequency distribution of most of the (lower-magnitude) POT peaks. However, given that the frequency curve fits the point data far less well in the simulated case than in the observed, the apparent fit between the two frequency curves should be treated cautiously. It is possible that the hydrological model is simulating flood peaks that represent a sample from the same underlying population that the real data are drawn from, but a poor fit between flood peaks and the theoretical distribution makes this difficult to test without exploring other flood frequency distributions. This type of result also illustrates that direct comparison of fitted flood frequency distributions for model calibration purposes could produce misleading results.

5.3. Quality of the Flow Time Series

In cases where the automated URS procedure was able to identify a set of parameters capable of generating acceptable flood frequency characteristics, it was found that the simulated flow time series appeared to be subjectively "realistic." For example, the time series data for the Mellte catchment, used to obtain the frequency results of Figure 2, are shown in Figure 6. The first plot shows the series of areally averaged rainfall data for the period 1985–1995. Below are the observed flows and the flows simulated using the PDM, with the parameter set selected from the URS procedure by the POT-QT objective function. There is a fair degree of correspondence between observed and simulated flows, failing mainly in generating peaks of intermediate magnitude and in the rather flat base flow response during winter periods.

The flow peaks in Figure 6 have been annotated in order of rank, down to the fifth largest peak. It can be seen that although there is agreement between the magnitude of the peaks plotted in Figure 2, the events associated with each pair of ranked peaks are not generally the same in the observed and simulated flow data. For instance, the largest peak in the observed series occurs in 1986, whereas the largest simulated flow is associated with an event at the end of 1994. Such differences in the timing of flow peaks that are of equivalent return period

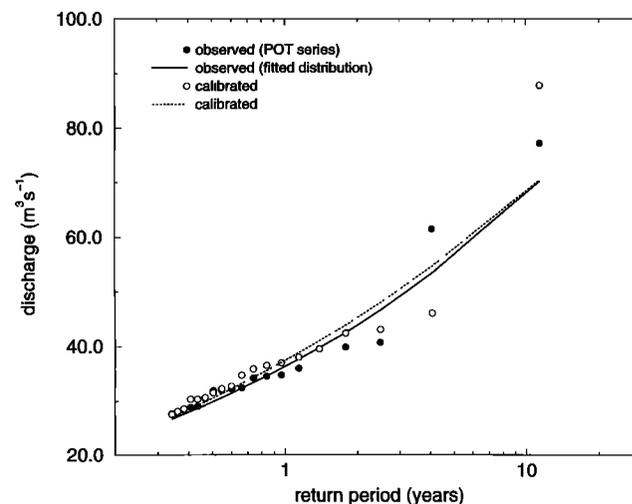


Figure 5. Observed and computed flood frequency data for the Ythan catchment, showing the interacting effect on fitted frequency curves of errors in the two highest-ranked peak flow data.

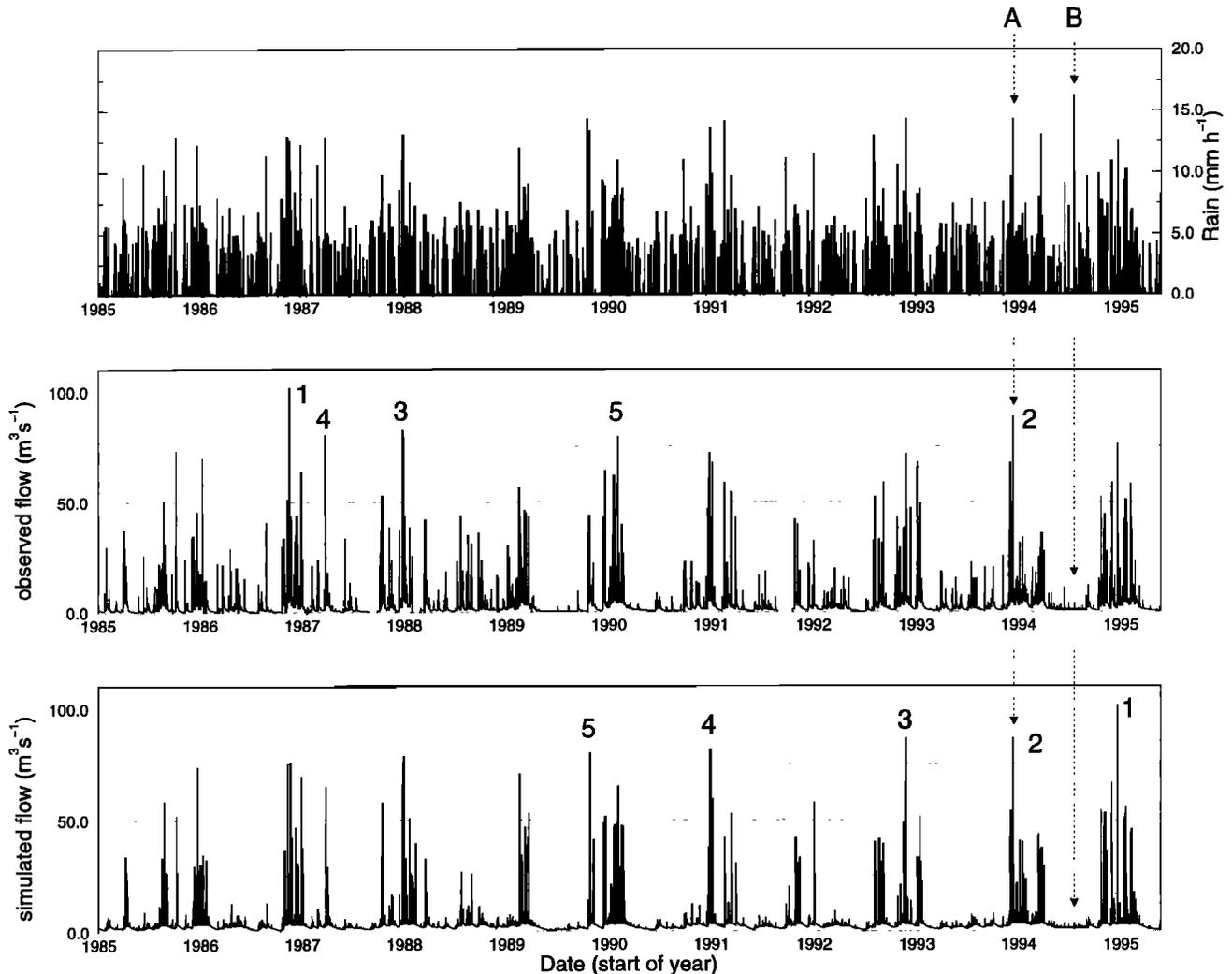


Figure 6. Time series showing observed rainfall, observed flow, and simulated flow data for the Mellte catchment. Flow peaks are ranked in order of magnitude. Note the differing responses of flow to rainfall at times A and B.

are not ideal. However, good reproduction of every peak over a 10-year period would be a remarkable result.

Another feature of the results in Figure 6 is indicated by the annotations A and B. A relatively large peak occurs in both observed and simulated flow series at time A, associated with a storm event that was preceded by a wet spell. At time B, there was another large rainfall event, but this followed a drier spell and caused negligible response in the observed flows, presumably because of drier antecedent conditions in the catchment. This effect has also been reproduced by the PDM, which correctly simulates the flow response at time B. Good representation of the nonlinearity in the rainfall-runoff transformation can also be seen in Figure 7, which shows the flow series simulated for the Tanllwyth catchment using the manually adjusted parameter set (see Figure 3 for the corresponding frequency distributions). Consider here the differing responses of the observed and simulated flows to the rainfall events of similar magnitude at times A and B. Peak flows in Figure 7 are again annotated by rank. (It can be seen that the largest peaks in the observed and simulated series both correspond to the same event, as discussed earlier.)

Inspection of flow time series offers some indication of the

overall realism of a simulation used for frequency analysis. It is to be expected that simulated flows may fail to reproduce in exact detail the observed data because of combinations of errors in the model and the recorded rainfall or flows. However, frequent, relatively minor errors in the simulation of a long, hourly flow series are not necessarily grounds for rejecting the simulation. In the present context, the aim is a realistic overall characterization of the flow regime, with priority being given to flood peaks.

In addition to time series plots, flow duration curves can also be used to check the overall quality of a given simulation. A contrast in model performance can be seen in Figure 8, which shows observed and simulated flow duration curves for the Tanllwyth and Frome catchments, where the flow data have been standardized by the median observed discharge in each case. The difference in the nature of the catchments is apparent from Figure 8. The Tanllwyth catchment is much flashier and has a variable flow regime, whereas the Frome catchment is influenced (but not dominated) by groundwater processes and, consequently, has a much flatter flow duration curve. In both cases, the simulated flows give rise to flow duration curves that are consistent with the observations. For both catchments

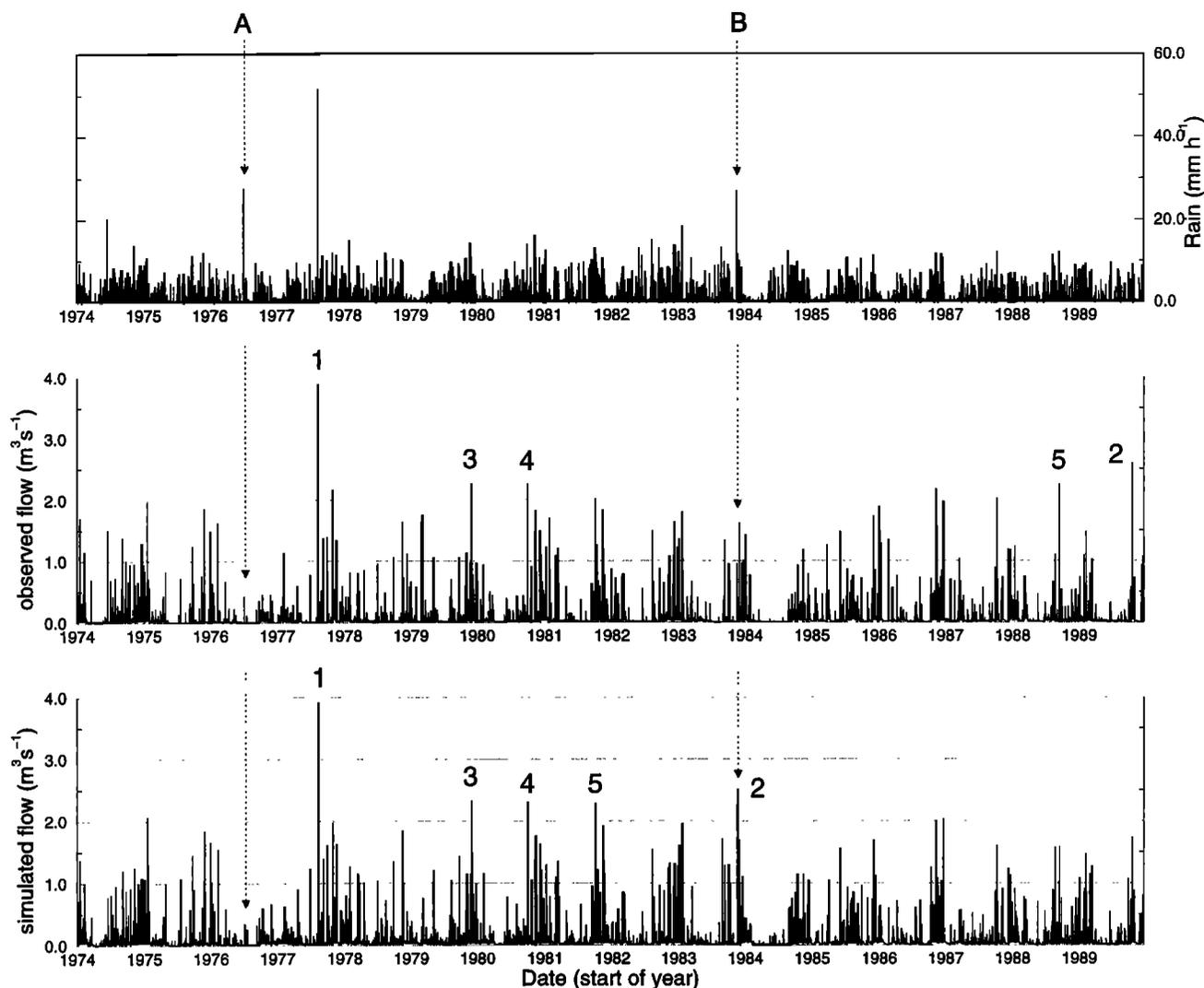


Figure 7. Time series showing observed rainfall, observed flow, and simulated flow data for the Tanllwyth catchment. Annotations are as for Figure 6.

the simulated flow duration curves suggest a tendency to overestimate. However, errors in the simulated flow quantiles, relative to the median observed flow, are much greater for the Frome catchment.

The error seen in the flow duration curves for the Frome catchment almost certainly reflects the difficulty of reproducing both a seasonal, groundwater-derived hydrograph and transient storm flow processes for this catchment using a relatively simple model structure. Even so, the simulated flow series was able to reproduce the main features of the observed series, which can be characterized as a seasonal base flow hydrograph with sharp, well-defined peaks superimposed. The problem represented in the simulated flow duration curve could have been due to poor estimates for the recession constants, leading to the simulated recessions being too flat, or to bias introduced by the value of f_c . However, inspection of the time series showed this not to be the case. Rather, the dynamics of recharge to the slow flow store appear to be the source of the base flow error. It may also be that an exponential function (equation (4)) is not the most appropriate form of slow flow routing store for a catchment such as at Frome.

5.4. Sensitivity of the Objective Functions to Joint Variations in Parameter Values

The URS approach can be used to examine the sensitivity of each objective function to variations in parameter values. It has to be stressed that each function evaluation is the product of a set of parameters; it is therefore difficult to draw conclusions about the marginal variations of objective function values with respect to any individual parameter. However, some information can be gained by plotting function values against the sampled values of each parameter. For the catchments discussed in this paper, it was found that many simulations gave rise to similar objective function values, where the individual parameters varied widely across the ranges sampled by the URS procedure. The results presented below are for the Tanllwyth catchment but are also typical of the other cases. The ability to find similar objective function values over wide ranges of each parameter was a feature of the results for all objective functions tested. This problem can be referred to as "equifinality," or "multiple local optima" in the objective function response surface, and has been noted for conceptual mod-

els in various contexts [e.g., Duan *et al.*, 1992; Beven, 1993; Gupta and Sorooshian, 1994; Zak *et al.*, 1997].

The sensitivity of the POT-SAE function, computed over a full evaluation period, is of primary importance, as this objective function is perhaps the most direct measure of the performance of a given simulation in the present context. To minimize the influence of length of record, the URS procedure was applied using the entire 16-year record for the Tanllwyth catchment. Two experiments were carried out, one in which the store parameters k_f and k_s were fixed by recession curve analysis (the "fixed-stores" case) and a second in which these parameters were allowed to vary (the "varied-stores" case).

Simulations in each experiment were divided into 10 classes, ranked in terms of the value of POT-SAE (from smallest total error to largest). The results for the best class in the varied-stores case are shown in Figure 9, where each point represents a single simulation, the abscissa indicates the value taken by each PDM parameter, and the best results lie at the upper margins of the graphs.

It is clear that simulations having similar values of POT-SAE can be obtained over the full range of values for some parameters. Parameter axes in Figure 9 indicate the limits of the ranges over which values were sampled, apart from the k_f store constant, which was sampled between the wider limits of 3.0 and 25.0 hours. Only for the k_f parameter did simulations not occur over the entire sampled range in the best performing class. Within the best class, only for k_f and f_c is there a suggestion that the marginal, minimum value of POT-SAE varies at all across the parameter range.

The dotted lines in Figure 9 indicate the parameter set that was obtained by manual adjustment to improve on the results of the original URS procedure, as applied to the 2-year calibration period (refer to Figure 3 for the corresponding flood frequency results). It is reassuring that the performance of this parameter set, in terms of the POT-SAE function, is approximately as good as any of the best five sets (shown by diamond symbols) obtained from the varied-stores URS experiment. However, there is no evidence that the manual and best five parameter sets are clustered around any globally optimum combination of values, except perhaps for the f_c and k_f parameters.

Figure 10 shows corresponding results for the fixed-stores

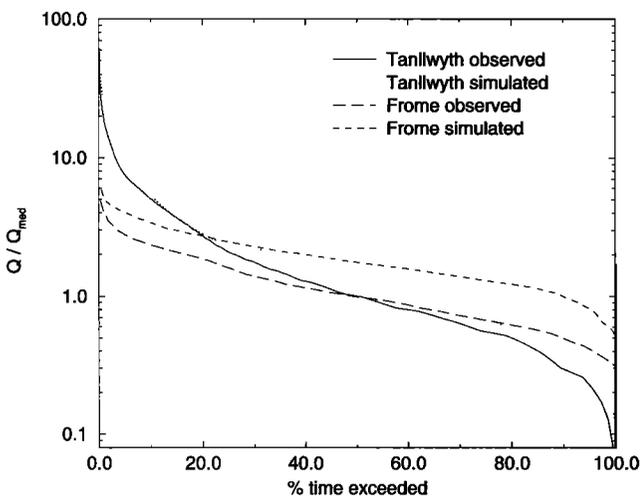


Figure 8. Flow duration curves for the Frome and Tanllwyth catchments standardized by the median observed flow in each case.

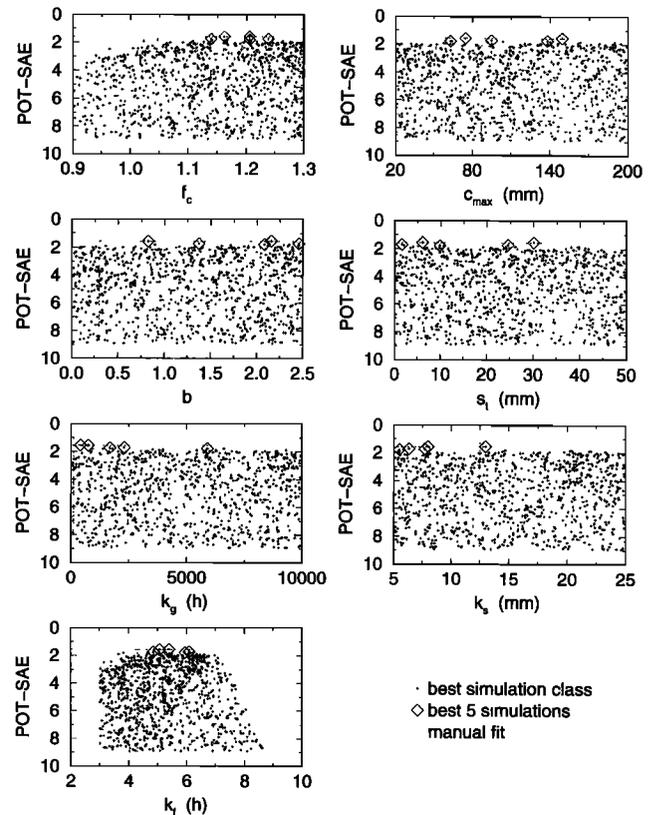


Figure 9. Results for the Tanllwyth catchment of the uniform random sampling of seven parameters of the PDM, where each simulation has been evaluated in terms of POT-SAE, the sum of the absolute errors on the POT series. Each small cross represents the outcome of a single simulation, diamonds denote the five best simulations, and the dotted lines show the manually adjusted parameter values.

URS experiment (note the change in POT-SAE axes scale). Here, there is greater evidence that the best simulations will be found clustered around certain values of f_c . The best five simulations and the manual parameter set also suggest small values at the margin for k_g . This is consistent with the flashy nature of the Tanllwyth catchment, where the exponential slow flow store is associated with lateral throughflow in the soil, rather than with deeper groundwater storage. However, the results for c_{max} illustrate that caution is needed in inferring marginal distributions of goodness of fit for individual parameters; the density of simulations in the best fraction of the class is greatest for larger values of c_{max} , but interactions with other parameters mean that a few simulations have almost equally good values of POT-SAE where c_{max} is less than 80 mm.

The results presented in Figures 9 and 10 indicate little marginal sensitivity of the objective function POT-SAE to most of the model parameters. However, these results are difficult to interpret unless also represented in terms of predictions of interest. Flood frequency results were therefore generated for the best five parameter sets (shown by diamond symbols in Figures 9 and 10), which are shown in Figure 11a for the fixed-stores case and Figure 11b for the varied-stores case. These results confirm the suggestion of Figures 9 and 10 that almost identical flood frequency results can be obtained by very different combinations of model parameters. One response to this may be that the apparent marginal insensitivity

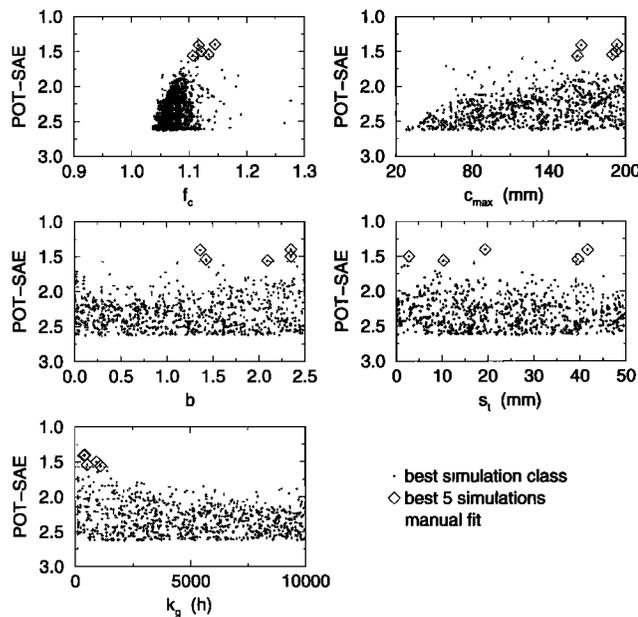


Figure 10. Results of uniform random sampling of PDM parameters with the routing store parameters fixed after recession curve analysis.

of the objective function, POT-SAE, indicates a robust model. However, this interpretation would fail to recognize that every simulation is a function of a set of parameters. For each value of a given parameter that appears in a good simulation, there are also many poor simulations.

It is also worth noting that the parameter sets giving rise to the similar flood frequency results in Figure 11 are not equivalent in terms of other aspects of the flow regime. Flow duration curves are shown in Figure 12 for the best five simulations from the fixed-stores and varied-stores experiments. It can be seen that none of these simulations performs as well as the manually adjusted parameters. Also, comparing the fixed-stores and varied-stores results, the fixed-stores simulations give rise to consistently better flow duration curves.

6. Conclusions

The results presented here suggest that calibrating a rainfall-runoff model using a simple efficiency measure based on the sum of the squared errors is likely to produce biased estimates of flood frequency distributions. However, when calibrated using a range of objective and subjective measures, a relatively simple, standardized model was able to generate acceptable simulated frequencies for catchments of different size and physiographic characteristics.

A conclusion from the URS experiments using 2-year simulation periods is that calibration in the flood frequency context requires significantly greater weight to be given to the simulation of peak flows than to lower-magnitude flows or, indeed, hydrograph shape. This is supported by a comparison of the differences between flood frequency results obtained using different objective functions, of which Figures 2, 3, and 4 are typical examples. In each case, the objective functions weighted more toward peak flows gave the best results. This is hardly surprising but focuses attention on the balance between good flood peak estimates and a realistic overall characterization of the flow regime. It would be ideal if good flood frequency results could always be obtained by a near-perfect simulation of the observed flows. However, experience sug-

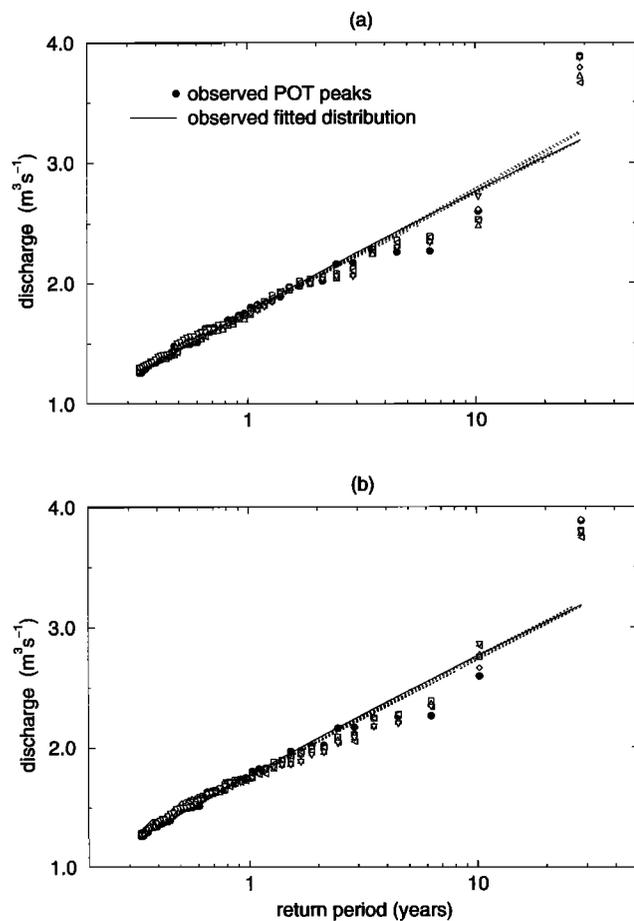


Figure 11. Flood frequency results for the best five simulations arising from uniform random sampling in (a) the fixed-stores case (see Figure 10) and (b) the varied-stores case (see Figure 9).

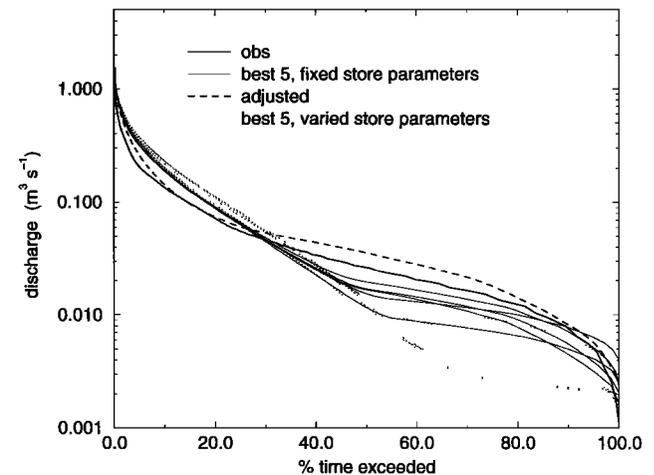


Figure 12. Flow duration curves for the Tanllwyth catchment showing the results obtained using manually adjusted parameters and results using the best five simulations from each of the two uniform random sampling experiments, corresponding to the flood frequency results in Figure 11.

gests that a conceptual rainfall-runoff model will often fall short of this ideal, especially for a long period of relatively fine time discretization.

One approach that should help ensure that simulated event hydrographs have a realistic form is the estimation of routing store constants prior to any trial-and-error calibration. This can be done, albeit approximately, by recession curve analysis. Such analysis, which could be carried out with even limited quantities of observed data, has the advantage that the derived parameter values are related directly to observations and are not affected by interaction with any other model parameters. This may improve the chance that the derived value can be related to some independent characteristics of the catchment, which will be useful if parameters are to be estimated for an ungauged site.

Simple graphical analysis of the sensitivity of objective functions to variation in parameter values suggested, for the Tallywh catchment (and other sites), that fixing routing parameters by recession curve analysis can help to constrain the values of other parameters. Although the scatterplots presented in Figures 9 and 10 do not reveal proper marginal probability distributions for model parameters, they do provide information about the sensitivity of any chosen objective function to parameter variations, provided it is remembered always that these variations are meaningful only in the context of the complete set of parameters corresponding to each objective function value.

Near-equivalence or "equifinality" of different parameter sets (in terms of one chosen objective function) and the existence of multiple objectives are two distinct sources of uncertainty in model predictions. Consideration of multiple objectives may, however, constrain the acceptable ranges for model parameters. Even simple subjective measures (such as visual inspection of flow duration curves) can be useful in selecting a suitable parameter set from a group of many sets that may produce almost identical flood frequency estimates. Further research would be useful to apply more objective methods for combining different calibration criteria in applying the continuous simulation method and to quantify the uncertainty about design flood estimates given multiple objectives.

Acknowledgments. This work was supported by funding from the U.K. Ministry of Agriculture, Fisheries and Food, project FD0404. Data were kindly supplied by the Environment Agency (England and Wales) and the Scottish Environmental Protection Agency. The author wishes to thank Ann Calver, Robert Moore, Duncan Reed, Beate Gannon, and Helen Davies for their comments and assistance. Additionally, the three anonymous reviewers are thanked for their constructive comments.

References

- Beven, K., Towards the use of catchment geomorphology in flood frequency predictions, *Earth Surf. Processes Landforms*, 12, 6–82, 1987.
- Beven, K., Prophecy, reality and uncertainty in distributed hydrological modelling, *Adv. Water Resour.*, 16, 41–51, 1993.
- Beven, K., and A. Binley, The future of distributed models—Model calibration and uncertainty prediction, *Hydrol. Processes*, 6, 279–298, 1992.
- Blazkova, S., and K. Beven, Flood frequency prediction for data limited catchments in Czech Republic using a stochastic rainfall model and TOPMODEL, *J. Hydrol.*, 195, 256–278, 1997.
- Bradley, A. A., and K. W. Potter, Flood frequency analysis of simulated flows, *Water Resour. Res.*, 28, 2375–2385, 1992.
- Bras, R. L., D. R. Gaboury, D. S. Grossman, and G. J. Vicen, Spatially varying rainfall and flood risk analysis, *J. Hydraul. Eng.*, 111, 754–773, 1985.
- Brazil, L. E., Multilevel calibration strategy for complex hydrological simulation models, Ph.D. dissertation, 217 pp., Colo. State Univ., Fort Collins, 1988.
- Brazil, L. E., and M. D. Hudlow, Calibration procedures used with the National Weather Service River Forecast System, in *Water and Related Land Resource Systems, IFAC Proc. Ser.*, edited by Y. Haimes and J. Kindler, pp. 457–466, Int. Fed. of Autom. Control, New York, 1981.
- Calver, A., and R. Lamb, Flood frequency estimation using continuous rainfall-runoff modelling, *Phys. Chem. Earth*, 20, 479–483, 1996.
- Davison, A. C., Modelling excess over high thresholds, with an application, in *Statistical Extremes and Applications*, edited by J. Tiago de Oliveira, pp. 461–482, D. Reidel, Norwell, Mass., 1984.
- Davison, A. C., and R. L. Smith, Models for exceedances over high thresholds, *J. R. Stat. Soc., Ser. B*, 52, 393–442, 1990.
- Duan, Q. S., S. Sorooshian, and V. K. Gupta, Effective and efficient global optimization for conceptual rainfall-runoff models, *Water Resour. Res.*, 28, 1015–1031, 1992.
- Gupta, V. K., and S. Sorooshian, A new optimization strategy for global inverse solution of hydrologic models, in *Computational Methods in Water Resources X*, edited by A. Peters, pp. 743–751, Kluwer Acad., Norwell, Mass., 1994.
- Gupta, H. V., S. Sorooshian, and P. O. Yapo, Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information, *Water Resour. Res.*, 34, 751–763, 1998.
- Hosking, J. R. M., and J. R. Wallis, Parameter and quantile estimation for the Generalised Pareto distribution, *Technometrics*, 29, 339–349, 1987.
- Hudson, J. A., and K. Gilman, Long-term variability in the water balances of the Plynlimon catchments, *J. Hydrol.*, 143, 355–380, 1993.
- Institute of Hydrology, Evaluation of FRONTIERS and local radar rainfall forecasts for use in flood forecasting models: Operational guidance note, *Natl. Rivers Auth. R&D Note 387*, 39 pp., Bristol, England, 1995.
- Institute of Hydrology, *A Guide to the PDM*, 40 pp., Oxfordshire, England, 1996.
- Jakeman, A. J., and G. M. Hornberger, How much complexity is warranted in a rainfall-runoff model?, *Water Resour. Res.*, 29, 2637–2649, 1993.
- Jones, D. A., Statistical analysis of empirical models fitted by optimisation, *Biometrika*, 70, 67–88, 1983a.
- Jones, S. B., The estimation of catchment average point rainfall profiles, *Inst. Hydrol. Rep. 87*, 34 pp., Inst. of Hydrol., Oxfordshire, England, 1983b.
- Klepper, O., H. Scholten, and J. P. G. van de Kamer, Prediction uncertainty in an ecological model of the Oosterschelde Estuary, *J. Forecasting*, 10, 191–209, 1991.
- Kuczera, G., Improved parameter inference in catchment models, 1, Evaluating parameter uncertainty, *Water Resour. Res.*, 19, 1151–1162, 1983.
- Kuczera, G., On validity of first-order prediction limits for conceptual hydrological models, *J. Hydrol.*, 103, 229–247, 1988.
- Lamb, R., and K. Beven, Using interactive recession curve analysis to specify a general catchment storage model, *Hydrol. Earth Syst. Sci.*, 1, 101–113, 1997.
- Lamb, R., and B. Gannon, The establishment of a database of hourly catchment average rainfall and flow for flood frequency estimation by continuous simulation, technical report to MAFF, 50 pp., Inst. of Hydrol., Oxfordshire, England, 1996.
- Moore, R. J., The probability-distributed principle and runoff production at point and basin scales, *Hydrol. Sci. J.*, 30, 273–297, 1985.
- Moore, R. J., Real-time flood forecasting systems: Perspectives and prospects, paper presented at British-Hungarian Workshop on Flood Defence, VITUKI, Budapest, Sept. 6–10, 1993.
- Moore, R. J., D. A. Jones, K. B. Black, R. M. Austin, D. S. Carrington, M. Tinnion, and A. Akhondi, RFFS and HYRAD: Integrated systems for rainfall and river flow forecasting in real-time and their application in Yorkshire, *Br. Hydrol. Soc. Occas. Pap.* 4, 12 pp., Br. Hydrol. Soc., London, 1994.
- Naden, P. S., Analysis and use of peaks-over-threshold data in flood estimation, in *Floods and Flood Management*, edited by A. J. Saul, pp. 131–143, Kluwer Acad., Norwell, Mass., 1992.
- Naden, P. S., Methods and techniques for peaks-over-threshold flood

- analysis, technical report to MAFF, 36 pp., Inst. of Hydrol., Oxfordshire, England, 1993.
- Nash, J. E., and J. V. Sutcliffe, River flow forecasting through conceptual models, 1, A discussion of principles, *J. Hydrol.*, 10, 282–290, 1970.
- Post, D. A., and A. J. Jakeman, Relationships between catchment attributes and hydrological response characteristics in small Australian mountain ash catchments, *Hydrol. Processes*, 10, 877–892, 1996.
- Romanowicz, R., K. J. Beven, and J. A. Tawn, Evaluation of predictive uncertainty in nonlinear hydrological models using a Bayesian approach, in *Statistics for the Environment 2: Water Related Issues*, edited by V. Barnett and K. Feridun-Turkman, pp. 297–317, John Wiley, New York, 1994.
- Sefton, C. E. M., and D. B. Boorman, A regional investigation of climate change impacts on UK streamflows, *J. Hydrol.*, 195, 26–44, 1997.
- Smith, R. L., Threshold methods for sample extremes, in *Statistical Extremes and Applications*, edited by J. Tiago de Oliveira, pp. 621–638, D. Reidel, Norwell, Mass., 1984.
- Spear, R. C., and G. M. Hornberger, Eutrophication in Peel Inlet, 2, Identification of critical uncertainties via Generalized Sensitivity Analysis, *Water Res.*, 14, 43–49, 1980.
- Spijkers, T., and P. Naden, Continuous rainfall-runoff modelling for flood estimation: Initial thoughts and data requirements, technical report to MAFF, 37 pp., Inst. of Hydrol., Oxfordshire, England, 1994.
- van Straten, G., and K. J. Keesman, Uncertainty propagation and speculation in projective forecasts of environmental change—A lake-eutrophication example, *J. Forecasting*, 10, 163–190, 1991.
- Wang, Q. J., The genetic algorithm and its application to calibrating conceptual rainfall-runoff models, *Water Resour. Res.*, 27, 2467–2471, 1991a.
- Wang, Q. J., The POT model described by the Generalised Pareto distribution with Poisson arrival rate, *J. Hydrol.*, 129, 263–280, 1991b.
- Yapo, P. O., H. V. Gupta, and S. Sorooshian, Multi-objective global optimization for hydrologic models, *J. Hydrol.*, 204, 83–97, 1998.
- Zak, S. K., K. Beven, and B. Reynolds, Uncertainty in the estimation of critical loads: A practical methodology, *Water Air Soil Pollut.*, 98, 297–316, 1997.

R. Lamb, Institute of Hydrology, Wallingford, Oxfordshire, OX10 88B, England, UK. (e-mail: r.lamb@mail.nwl.ac.uk)

(Received December 1, 1997; revised December 28, 1998; accepted April 7, 1999.)