Hydrology and Earth System Sciences, 3(2), *205–212* (1999) © EGS

Hydrology & Earth System Sciences

The FORGEX method of rainfall growth estimation III: Examples and confidence intervals

D. S. Faulkner and D. A. Jones

Institute of Hydrology, Wallingford, Oxfordshire, OX10 8BB, UK
e-mail for corresponding author: duncan.faulkner@jeremybenn.co.uk

Abstract

This paper illustrates the performance of the FORGEX method of rainfall growth estimation. Results are presented for three regions of the United Kingdom: the East Midlands, north-west England and south-west England. Focused rainfall growth curves are compared between regions and between different sites within each region. Typical growth curve shapes are discussed with reference to the climate of each region. Daily growth curves are derived from a large number of records of annual maximum rainfalls. A smaller number of hourly annual maximum series is available for estimating sub-daily rainfall growth curves.

Rainfall growth rates are compared with the results of a widely used method. The present method allows more local and regional variation in growth rates. The new growth rates are higher for durations of 1 and 2 days in parts of south-west England, but lower for moderate return periods at some focal points in the north-west. In the East Midlands, the new 1-hour growth rates are considerably higher for long return periods.

Confidence limits for growth rates are derived by bootstrapping. This is accomplished by fitting a large number of growth curves to resampled sets of rainfall data.

Introduction

Previous papers (Stewart *et al.*, 1999; Reed *et al.*, 1999) have described the background and structure of the FORGEX method of rainfall frequency estimation. This paper presents a selection of results and a comparison with those of the *Flood Studies Report* (FSR), NERC (1975). It also describes the derivation of confidence limits for growth rates.

Results are presented for three study regions, which illustrate the variety of extreme rainfall regimes in the UK. The three regions are:

- The East Midlands. There are many more long computerized records of sub-daily data for this region than for elsewhere in the UK. Growth rates for all durations are known to be particularly high here. Convective rainfall is more significant than in the other regions studied, which are further west.
- South-west England. This region has featured in several previous studies of rainfall frequency (Bootman and Willis, 1977; Reed and Stewart, 1989; Clark, 1991). Alternative methods of rainfall frequency estimation have been developed previously for this region to overcome the perceived shortcomings of the FSR.
- North-west England. This is a region with varied topography, including the Pennines and the Lake District hills. Design rainfalls for 1-day rainfall are thought to be overestimated by the FSR method (Dales and Reed, 1989).

Different types of results and comparisons are presented for each region, in order to explore the features described above.

Figures 1 to 3 show the focal points in each region. These points are chosen to give an indication of local variations in growth rates. Most focal points are settlements, but two are reservoirs. All are sites for which design rainfalls are likely to be required at some time.

Availability of rainfall data

The analysis was restricted to records comprising at least ten annual maxima. The locations of daily raingauges are shown by the small dots on Figs. 1 to 3. Daily records are plentiful, but gauge networks providing hourly annual maxima are much sparser. The highest densities of hourly records are in London and the English Midlands; for example, for the focal point at Rutland Water, there are 238 hourly raingauges within the (maximum) 200 km radius used by FORGEX.

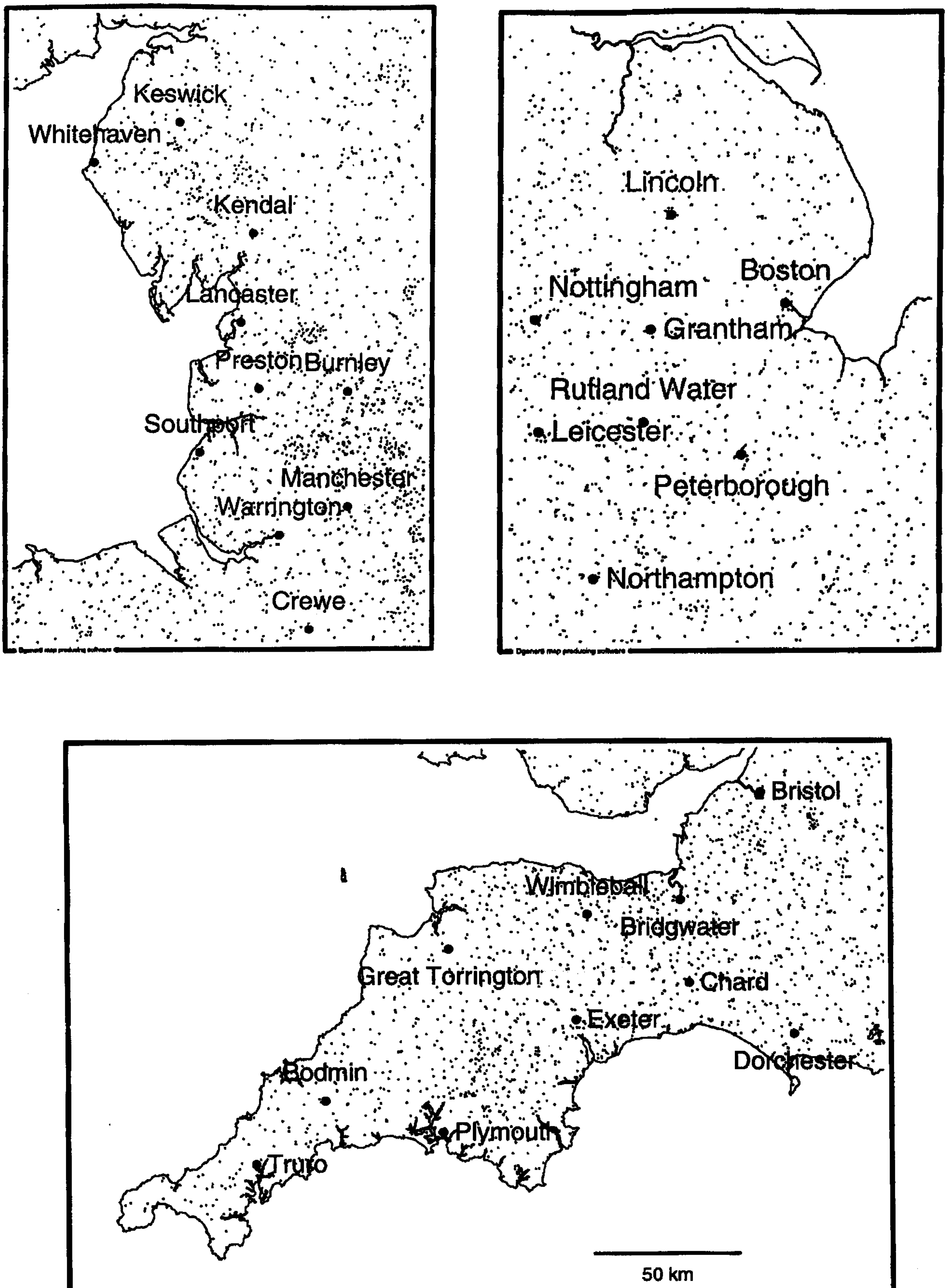

Fig. 1–3. *Focal points and daily raingauges.*

Results and discussion

EAST MIDLANDS

Figure 4 shows 1-hour growth curves for the eight focal points in the East Midlands. Hourly growth curves do not vary much between focal points. This could be because local variations are masked by the sparseness of hourly records, or it may reflect a genuine uniformity in the meteorological conditions which give rise to short-duration rainfall extremes. There is, however, an apparent regional trend, which is mirrored in daily growth curves. Over a wide range of return periods, the highest curve in Fig. 4 is the one focused on Lincoln. Other sites in the east of the region such as Grantham and Boston also have relatively high growth rates. The lowest growth rates over most of the return period scale are for Leicester, the westernmost focal point.

Figure 4 also includes a growth curve derived using the FSR procedure and standardized by the 2-year rainfall so that it can be plotted on the same scale as the focused growth curves. The FSR curve is taken from statistics for Rutland Water, but results for other locations in the region are very similar. The focused growth curves lie above the FSR curve, the difference rising at a return period of 100 years to approximately 35%. Put another way, a 1-hour rainfall of 4.5 times the median annual maximum is four times more likely to occur according to the focused growth curve than according to the FSR: a major difference.

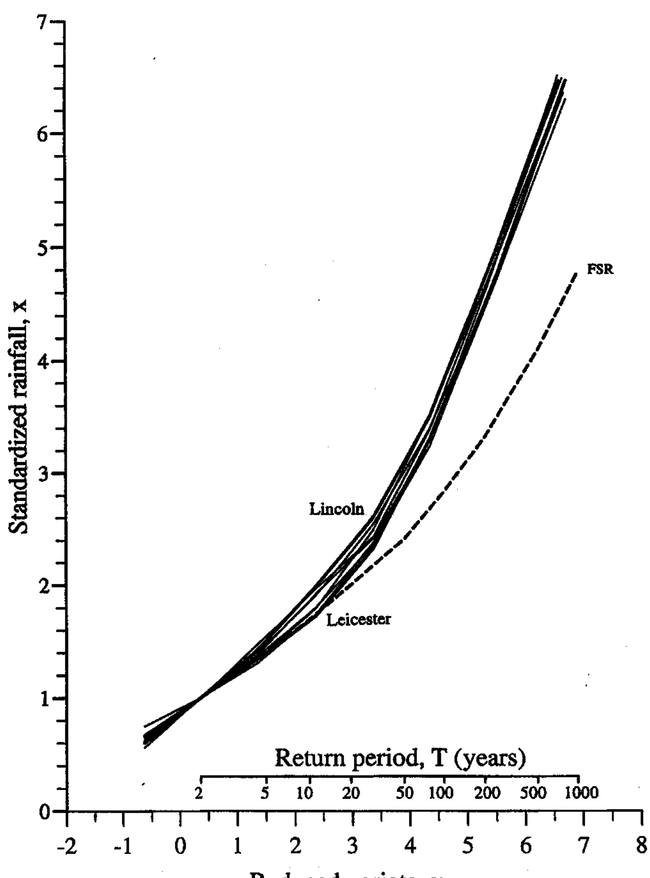

Fig. 4. *1-hour growth curves for focal points in the East Midlands.*

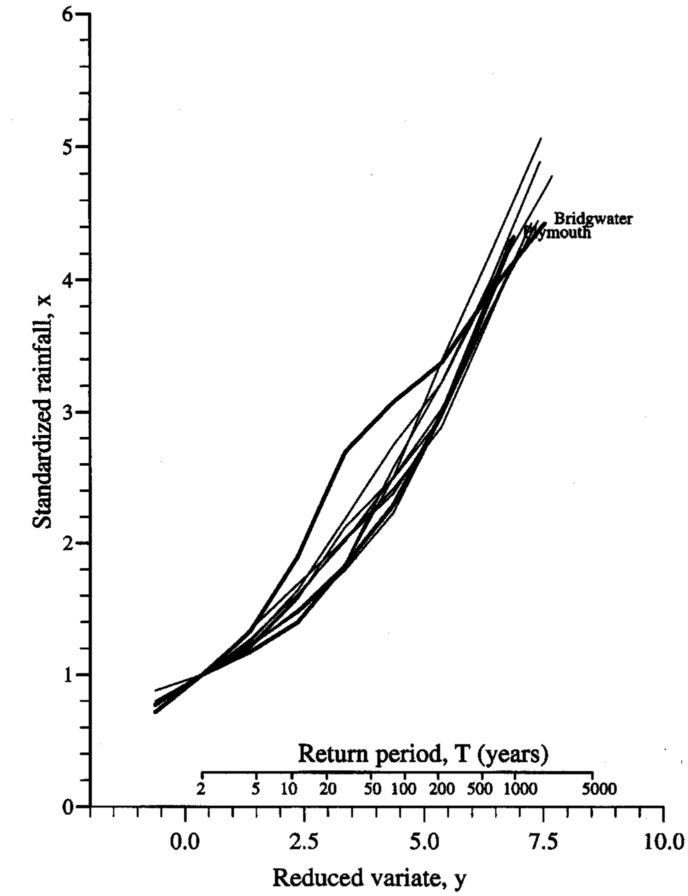

Fig. 5. *1-day growth curves for focal points in south-west England.*

SOUTH-WEST ENGLAND

Figure 5 compares growth curves for 1-day rainfall focused on ten sites in south-west England. At moderate return periods (10 to 100 years), the curve for Bridgwater is the highest by some margin. At longer return periods, other sites in the east of the region have the highest growth rates. Growth rates in the west, near the coast, at sites like Plymouth, are lower. The results reflect the occurrence of unusually high rainfall totals in the east of the region—particularly near Bridgwater and Dorchester—some of which are described by Bootman and Willis (1977). These authors attribute the anomaly near Bridgwater to the geography of the area; there is a central basin surrounded by hills, well suited to the development of thunderstorms.

The humped shape of the Bridgwater growth curve between the 10 and 100-year return periods occurs because this is the closest focal point to several of the historic notable rainfalls. These exceptional rainfalls were recorded by gauges which form a relatively small network around

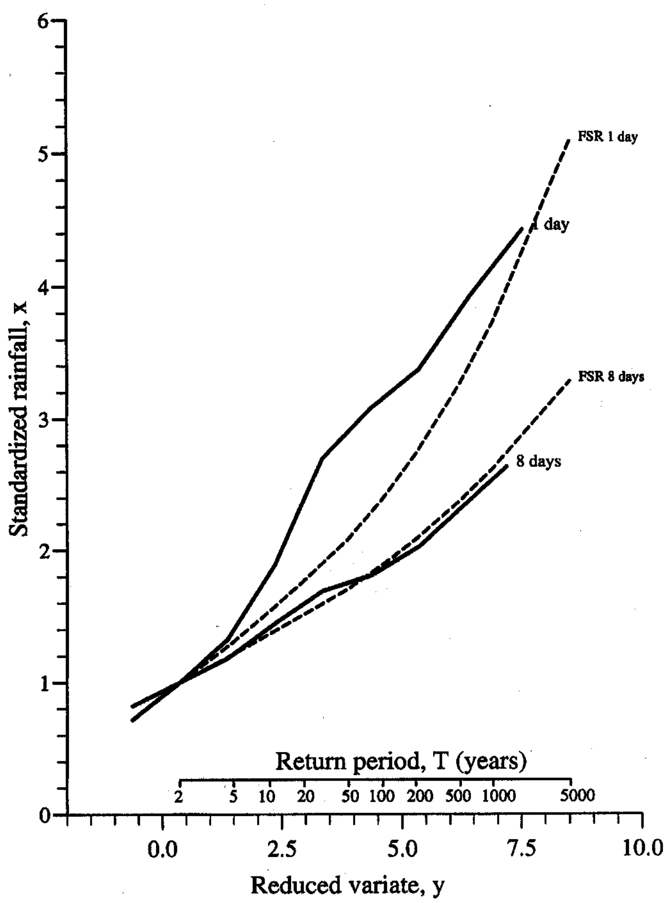

Fig. 6. *Comparison of FSR results with growth curves focused on Bridgwater.*

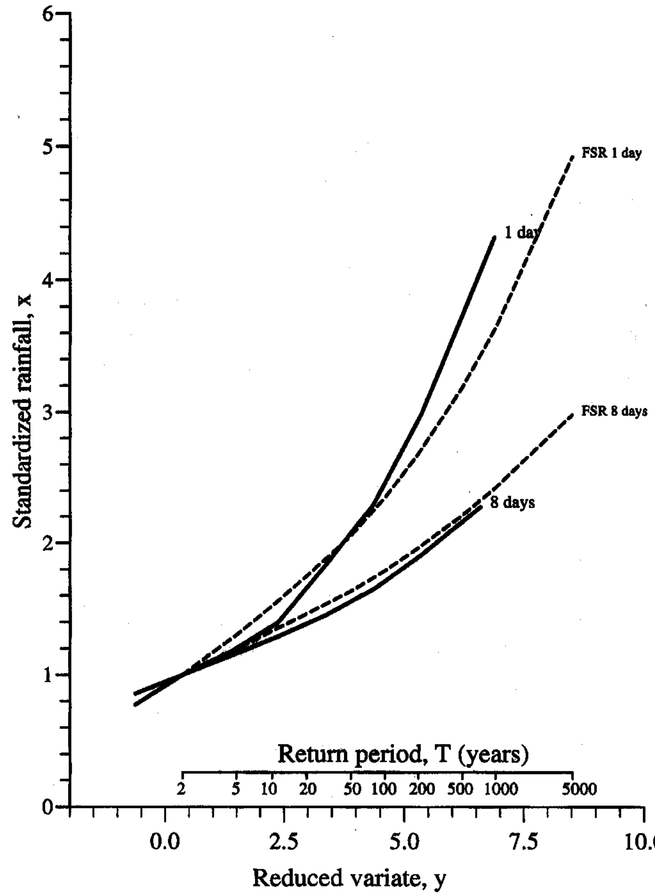

Fig. 7. *Comparison of FSR results with growth curves focused on Plymouth.*

the focal point, and so they plot at lower return periods at Bridgwater than they do for the other focal points.

Focused growth curves are compared with FSR results in Fig. 6 (for Bridgwater) and Fig. 7 (for Plymouth). Both figures show 1 and 8-day curves. At both focal points, 1-day growth rates are higher than those given by the FSR. The difference is greatest at Bridgwater, where for a return period of 100 years, the new growth rate is 35% higher. Put another way, a 1-day rainfall of three times the median annual maximum is five times more likely according to the focused growth curve than according to the FSR. At Plymouth, the differences are much smaller. The 8-day growth curves are close to the FSR results at both sites. These findings reinforce the concerns of Bootman and Willis (1977) that the FSR method underestimates extreme rainfalls around Bridgwater. Figures 6 and 7 expose the lack of local variation in FSR rainfall growth curves.

NORTH-WEST ENGLAND

Figure 8 compares growth curves for 1-day rainfall focused on ten sites in north-west England, ranging from Crewe (a lowland site) in the south of the region to Keswick (in the Lake District hills) to the north. There is a north-south trend in growth rates, with the highest curves being for Crewe, Manchester, Warrington and Burnley. The lowest curves are those focused on Kendal, Keswick and Whitehaven in the north.

The growth curves in north-west England lie below those in the south-west. The curves for the north-west have a marked kink at a return period of 80 years, so that the final segments have a steeper slope. It is possible that this shape reflects the more mixed nature of extreme rainfall in the north-west, with many of the extreme 1-day rainfalls being associated with frontal events but there still being scope for exceptional totals arising from convective events.

Comparisons with the FSR results highlight the different shapes of the growth curves. Figure 9 shows 1 and 8-day growth curves focused on Kendal. For moderate return periods (10 to 100 years), the 1-day growth rate is lower than by the FSR method, by up to 16%. The curves from the two methods cross, so that FORGEX gives higher results than the FSR for longer return periods. The 8-day rainfall growth factors also appear to be underestimated by the FSR at long return periods. The differences are less marked for sites in the south of the region.

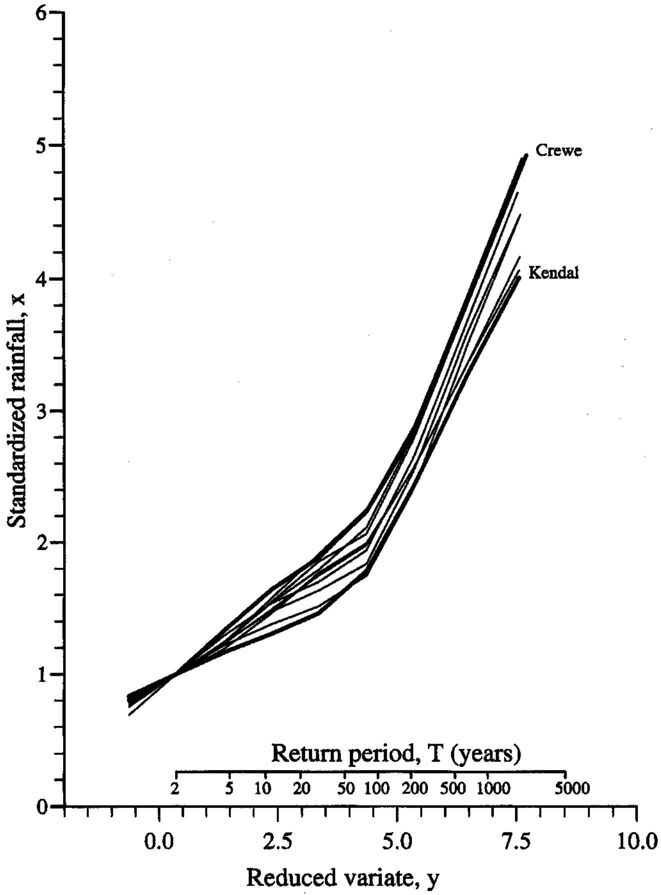

Fig. 8. *1-day growth curves for focal points in north-west England.*

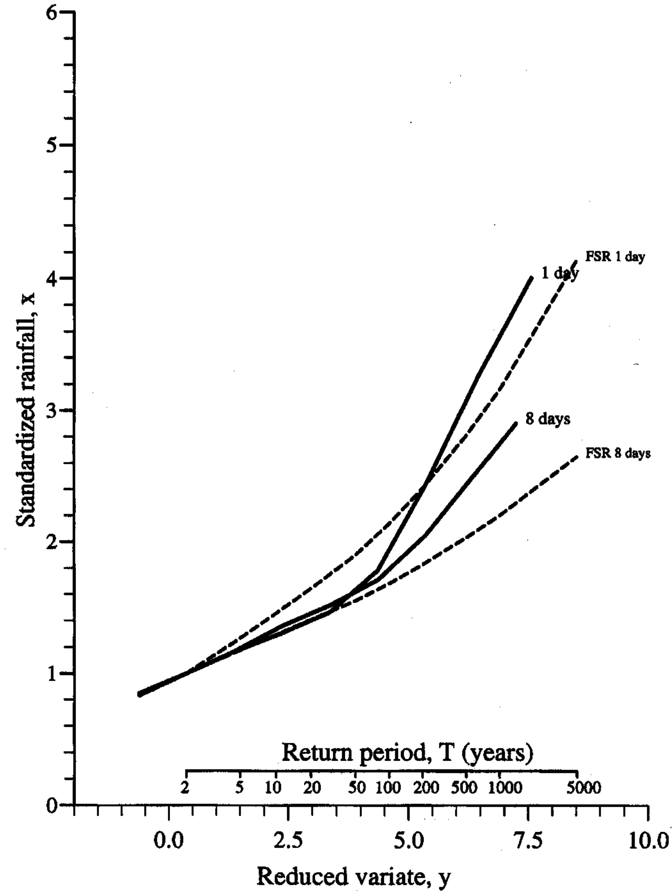

Fig. 9. *Comparison of FSR results with growth curves focused on Kendal.*

The results for moderate return periods match those of Dales and Reed (1989) who found that observed 1-day rainfalls are overestimated by the FSR in the north-west.

Confidence limits

BACKGROUND

Confidence limits give an indication of the range of values in which the growth rate can be expected to lie. In the absence of an infinitely long rainfall record consistent with the current climate, the true growth rate is unknown.

Bootstrap methods have been developed relatively recently (Efron, 1979). They enable the derivation of confidence limits and the use of significance tests in situations where the underlying statistical population is unknown or where an analytical solution is impractical. Bootstrapping is based on the generation of many resamples, which are selected from the original sample. This sample is used as the distribution from which the resamples are chosen randomly with replacement, i.e. with each value being returned to the original sample after it has been chosen, so that it may be chosen again.

In regional rainfall frequency estimation there are two alternative samples to work with: the years for which data are held or the sites for which they are held. After trials with both options, it was elected to carry out the resampling in the time domain: there is more scope for variation between annual maxima in different years than at different sites, due to spatial dependence. It is easy to imagine an immense rainfall occurring in a year which is just before records start, or just after they end. It is less likely that an isolated storm would be totally missed by the dense (daily) raingauge network in the UK.

METHOD

The elements used for resampling are the individual years of data across all sites. Resampling takes place across the entire network rather than at individual gauges, to preserve the spatial dependence which is required when forming the network maximum series. The basic method to find the $100(1-2\alpha)\%$ confidence interval is as follows:

1. Draw a resample from the N years providing rainfall data;
2. Use the resulting at-site records to derive a focused growth curve in the usual way;

3. Repeat the above steps B times ($B = 199$, see Appendix 2) to give B growth curves;
4. For various return periods, find the bootstrap residuals E_i which are the deviations of each new growth rate G_i from the original sample growth rate G_{sam}, i.e. $E_i = G_i - G_{sam}$;
5. Rank the deviations in ascending order and find E_m and E_n where $m = \alpha(B + 1)$ and $n = (1 - \alpha)(B + 1)$. This means using the 5th and 195th values out of 199 for 95% confidence limits (see Appendix 2);
6. Construct the confidence interval for the unknown growth rate as $(G_{sam} - E_n, G_{sam} - E_m)$.

The assumptions and theory behind this method are treated more fully in Appendix 1.

In practice, the more efficient method of *balanced re-sampling* (e.g. Fisher, 1993) was used. The principle of this method is to ensure that each year occurs equally often overall among the B bootstrap samples. This is implemented by creating a vector of length BN consisting of the N years of record repeated B times. This array is then randomly re-ordered, and divided into slices of length N, to obtain B bootstrap samples.

TESTS AND MODIFICATIONS

Most published applications of bootstrap methods involve a relatively simple statistic such as the mean of a sample. The use of resampled years of record to find confidence limits for growth rates has more layers of complexity, so the method was extensively tested and several modifications were introduced.

The tests included an investigation of the influence of the mean record length in a resample. These mean record lengths differed considerably between resamples, although this was not seen to cause any systematic variation in growth curves. One reason for the variation in mean record length was the small proportion of daily raingauges supplying records from before 1961. These few records are much longer than the others, and so there are many years for which relatively few gauges provide data. This feature of the computerized dataset was preserved by resampling the periods before and after 1961 separately. GREHYS (1996) used a similar approach.

Another test examined the minimum record length in the resamples. To avoid estimating the standardizing variable (*RMED*, the median annual maximum) from records shorter than ten years, it was decided to standardize each record before resampling. This is not an ideal solution, because the median of each resampled series of standardized annual maxima will not necessarily be exactly equal to 1.

RESULTS

Confidence limits for 1-day growth rates at Leicester in the East Midlands are shown in Fig. 10. The dashed lines connect the results for different return periods. The variation of the confidence limits with return period is a feature of the FORGEX method. It reflects in particular the relative contributions of pooled and network maximum points at different return periods. For example, the limits are relatively narrow for return periods around 200–500 years, where the growth curve is fitted only to network maximum points (see Fig. 4 of the previous paper, Reed *et al.* (1999)). Similar features are found for confidence limits at other sites.

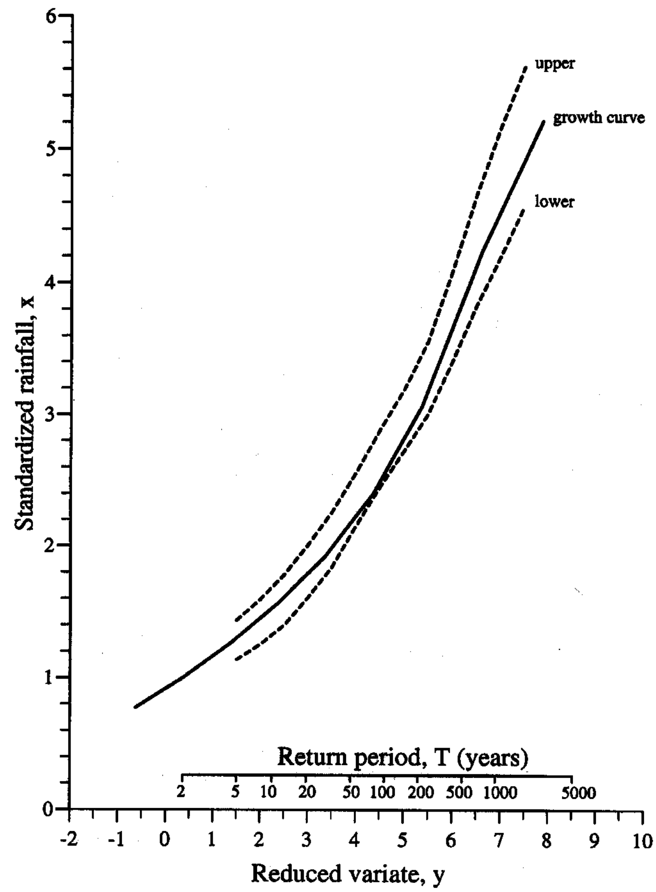

Fig. 10. *95% confidence limits for 1-day growth rates at Leicester.*

The upper limit is typically 0.4 to 0.6 units above the growth curve, and the lower limit varies from 0.01 to 0.3 below. The asymmetry is in part due to the curvature of the growth curve, which has no upper bound. This was confirmed by trial single-site frequency analyses for a range of UK *flood* peak data. It was found that the presence of one unusually large flood event is enough to pull up the fitted growth curve significantly, whereas the growth curve is not nearly as sensitive to an unusually small annual maximum.

Conclusions

The FORGEX method of rainfall frequency estimation can produce growth curves focused on any site in the UK.

By making efficient use of local and regional data, it can account for local and regional characteristics of extreme rainfall, and produce estimates of rare rainfalls, to return periods longer than 1000 years for daily durations at most focal points. It addresses a concern expressed about the results of the FSR in regions such as south-west England.

The technique of bootstrapping is a valuable tool for assessing confidence limits for growth rates. These give an idea of the limits of what we can know, given a finite period of record.

Acknowledgements

Rainfall records used to derive the results presented here were provided by the Met. Office and the Environment Agency. We are grateful for the suggestions provided by two referees.

References

Bootman, A.P. and Willis, A., 1977. *Extreme two-day rainfall in Somerset.* Wessex Water Authority.

Carpenter, J., 1999. Test inversion bootstrap confidence intervals. *J. Roy. Statist. Soc.* **B61**, 159–172.

Clark, C., 1991. A four-parameter model for the estimation of rainfall frequency in south-west England. *Meteorol. Mag.*, **120**, 21–31.

Dales, M.Y. and Reed, D.W., 1989. *Regional flood and storm hazard assessment.* Report No. 102, Institute of Hydrology, Wallingford, UK.

Davison, A.C. and Hinkley, D.V., 1997. *Bootstrap methods and their application.* Cambridge Univ. Press.

Efron, B., 1979. Bootstrap methods: another look at the jackknife. *Annals of Statistics,* **7**, 1–26.

Fisher, N.I., 1993. Some modern statistical techniques for testing and estimation. Chapter 8 in *Statistical analysis of circular data*, Cambridge Univ. Press, 199–218.

GREHYS, 1996. Inter-comparison of regional flood frequency procedures for Canadian rivers. *J. Hydrol.,* **186**, 85–103.

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

Reed, D.W., Faulkner, D.S. and Stewart, E.J., 1999. The FORGEX method of rainfall growth estimation, II: Description. *Hydrol. Earth System Sci.* 3, 197–203.

Reed, D.W. and Stewart, E.J., 1989. *Focus on rainfall growth estimation.* Proc. 2nd National Hydrol. Symp., Sheffield, September 1989, 3.57–3.65.

Shao, J. and Tu, D., 1995. *The Jackknife and Bootstrap.* Springer Series in Statistics. 516 pp.

Stewart, E.J., Reed, D.W., Faulkner, D.S. and Reynard, N.S., 1999. The FORGEX method of rainfall growth estimation, I: Review of requirement. *Hydrol. Earth Systems Sci.* 3, 197–203.

Appendix 1 Bootstrap confidence intervals

Confidence intervals give an indication of the range of values in which the true growth rate is expected to lie. The key assumption for deriving bootstrap confidence intervals is that the bootstrap *residuals*

$$E_i = G_i - G_{sam}$$

(where G_i is the estimate from bootstrap sample i and G_{sam} is the basic estimate from the sample data) are assumed to be representative of values drawn from the same distribution as the actual unknown residual $E = G_{sam} - G_{true}$. Shao and Tu (1995) use the term *hybrid bootstrap* for the present approach to constructing confidence intervals while, for Davison and Hinkley (1997), it is *the basic bootstrap*. It is a special case of the bootstrap-t method which uses an estimated scaling quantity fixed at unity. In addition, it is a test-inversion approach (Carpenter, 1999) and this provides the basis of the method described in Appendix 2 for prescribing which of the order statistics of the bootstrap sample to use in constructing the required confidence interval. An alternative approach, known as the *bootstrap percentile method*, is sometimes used (see later). This is described by Davison and Hinkley (1997, p203) as not working very well and it does not have a simple interpretation as a test-inversion method. It seems likely that the approach used here can be improved by using alternative specifications for the residuals, possibly including use of estimated scaling quantities, as for the bootstrap-t method, but these have not yet been evaluated.

If the distribution is asymmetric, the test-inversion approach gives results which are the reverse of what the user may have expected, based on the incorrect interpretation that the resampled values G_i directly reflect uncertainty about where the unknown quantity, G_{true}, might be. Using test-inversion, if most of the range of resampled growth curves lies below the original estimated growth curve, this implies that the estimated growth curve from a given sample will usually lie below the true growth curve. Therefore, the confidence interval for the true growth curve will lie mainly *above* the estimated growth curve.

Mathematically speaking, if E_L and E_U are appropriate lower and upper percentage points of the unknown distribution of residuals, such that the probability:

$$\Pr\ (E_L \le E \le E_U) = 1 - 2\alpha$$

then

$$\Pr\ (E_L \le G_{sam} - G_{true} \le E_U) = 1 - 2\alpha$$

or

$$\Pr\ (G_{sam} - E_U \le G_{true} \le G_{sam} - E_L) = 1 - 2\alpha.$$

Thus $(G_{sam} - E_U, G_{sam} - E_L)$ is a confidence interval with coverage $1 - 2\alpha$.

The percentage points E_L and E_U are estimated by the empirical percentage points of the ordered bootstrap residuals, E_m and E_n:

$$(G_{sam} - E_n, G_{sam} - E_m).$$

This is the formula given in the main part of the paper. Substituting $G_m - G_{sam}$ for E_m where G_m is the mth smallest resampled growth rate, and similarly for E_n, the required confidence interval is obtained:

$$(2G_{sam} - G_n, 2G_{sam} - G_m).$$

Note that the upper percentage point is used to find the lower confidence limit. The bootstrap-percentile approach yields the interval (G_m, G_n).

Note that it is possible for this formula to give a lower confidence limit which dips below 1.0, which is unrealistic for growth rates at return periods longer than two years. However, this is unlikely to be a problem for the long return periods typically of interest. An alternative strategy would involve defining the residual as the ratio of growth rates. This would permit the use of a transformation to ensure that the lower confidence limit could not drop below 1.0.

Note also that it would be possible to use the adjusted estimator $G_{sam} - E_{med}$, where E_{med} is the median of the bootstrap residuals, as a bias-adjusted estimate of the growth rate, the bias having been estimated via the bootstrap.

Appendix 2 Number of resamples

It is not immediately clear how many resamples to use. Some authors recommend the use of the 6th highest and lowest values out of 200 to provide 95% confidence intervals. Here, the 5th highest and lowest out of 199 were adopted for the following reason.

The 95% confidence region consists of all those growth rates, g, for which there is not enough evidence to reject the null hypothesis at the 5% level that the variables $E = G_{sam} - g$ and the bootstrap residuals $E_i = G_i - G_{sam}$ have the same distribution, i.e. the null hypothesis is $g = G_{true}$. The hypothesis is rejected if E is in one of the tails of the distribution of the E_is.

Suppose there are B random samples, and they are ordered E_1, $E_2, \ldots E_B$. Under the assumption of a common distribution for E and E_i (i.e. the null hypothesis), E is equally likely to appear at any point in the ordered set of E_is. Each of the following occurrences is equally likely:

$$
\begin{aligned}
&E < E_1, \\
&E_1 < E < E_2, \\
&\ldots \\
&E_{B-1} < E < E_B, \\
&E_B < E.
\end{aligned}
$$

Each has probability $1/(B + 1)$. Therefore the probability of E occurring in one of the $2n$ most extreme regions is exactly $2n/(B + 1)$. This is the probability of rejecting the null hypothesis when it is actually true. When this is subtracted from one, it yields the probability that the random interval containing the other E_is does contain the true value. Hence the coverage probability of the confidence interval is $1 - 2n/(B+1)$.

When $B = 199$ and $n = 5$, this yields a coverage probability of 0.95. If resources permit the extra computation, values of $B = 1999$ and $n = 50$ could be used. Confidence intervals defined in the way prescribed here have the correct coverage probability (e.g. 95% of sets of basic sample data will lead to confidence intervals which do include the true value, given the assumption at the start of Appendix 1), regardless of the number of resamples, B. Use of a limited number of randomly selected resampled datasets to form the confidence intervals introduces unnecessary random variation into the results. Thus, a relatively large number is chosen in order to reduce this variation as far as is reasonable: this can be checked by repeating the resampling procedure a few times.