INSTITUTE OF HYDROLOGY

# REGIONAL GROWTH CURVES.

by

M J STEVENS and P P LYNN

ABSTRACT

This report puts forward the reasons for pooling flood frequency data to obtain more stable estimates of floods of given return period than is otherwise possible. Some of the problems as well as the advantages of pooling by geographical regions are pointed out. The regions chosen in the Institute's Flood Studies Report were subjected to statistical tests from which it is shown conclusively that regional differences do exist although some curves are very similar. Implications for future work on regionalisation are pointed out.

REPORT NO 52

October 1978

## CONTENTS



## FIGURES



## TABLES

# 1. INTRODUCTION

Engineers commonly have to produce estimates of floods of given return periods for catchments which have little or no flow data. The Flood Studies Report (FSR), published by NERC, 1975, was largely concerned with providing methods to help them do this. One of these methods was developed from a statistical analysis of all available flow data in the UK. The statistical approach can be divided into two stages. The first stage is to estimate the typical size of flood on a catchment. The second stage concerns the variability of floods and in particular how much larger is a rare flood than a common one. The second aspect is the main subject of this report, but the first is dealt with briefly to set the later sections in context.

In the FSR the main data analysed were the annual maximum flows for the period of record at each station. The value chosen to index the typical size of flood was the mean annual maximum flow. In general this was calculated as the arithmetic average of the annual maxima but was sometimes modified by extending the record by correlation with nearby stations. For the ungauged catchment the mean annual flood needs to be related to characteristics of the catchment which can be read off maps etc. This was done by multiple regression and the equation to predict the mean annual flood ($\bar{Q}$) was

$$\bar{Q} = 0.0201\ AREA^{.94} STMFRQ^{.27} SOIL^{1.23} RSMD^{1.03} (1+LAKE)^{-0.85} S1085^{.16} \quad (1)$$

When residuals (differences between observed and estimated mean annual floods) from this equation were plotted on a map there were clear areas of overall under- or over-prediction. This led to the regionalising of the prediction equation so that different geographical regions had different multiplying constants at the beginning of the equation. Originally eleven regions were selected on the basis of hydrometric area boundaries but after tests of the significance of differences between the regions only six were eventually retained. One of the six, the Thames, Lee and Essex region, was found to be better predicted by a separate equation.

This type of regionalisation is of the 'discriminating' type, as opposed to the 'clustering' type. The discrimination approach is to ask the question:- 'Given that I have these regions, is the relationship between mean annual flood and catchment characteristics the same in them all?', if the answer is 'yes' then the regions are merged; if 'no' they remain separate. The clustering approach is to start with the basic data and choose groups of data points which are similar as shown by the data, not as shown by some external property such as geographical proximity. Only as many groups are chosen as are found to be distinct. It will be seen later that the regionalisation of the variability aspect of the floods has also been done on a 'discriminating' basis.

Having solved the first part of the problem, that relating to the typical flood size, there are several possible methods of approaching the second part, ie the variability of floods. Among these are:

(a) To produce regression equations similar to equation (1) for floods with a range of probability of occurrence (or return period). This method was not used in the FSR because the number of stations with sufficiently long record to reliably estimate floods of long return periods is small.

(b) To produce regression equations for higher moments of the annual maximum floods which can be used with the estimate of the mean annual flood and an assumed probability distribution to calculate floods of given return periods. This method, sometimes known as the frequency factor method (Chow, 1953), was tried during the Flood Studies but no worthwhile relationship between coefficient of variation and catchment characteristics could be found.

(c) The pooling of estimates of higher moments for a group of stations. The pooling estimate can then be used as in (b) to obtain floods of given return periods.

(d) The pooling of the annual maxima standardised by some value such as the mean annual flood. The pooling can take the form of a station year analysis or some more complicated method to overcome dependence between stations.

Approaches (c) and (d) are attempts to overcome the problem of estimating parameters from short records which leads to large sampling errors. By sacrificing the fit of the model at individual stations reliable estimates of regional parameters are sought. The approach chosen in the flood studies was (d) and is described in more detail in the next section.

## Pooled growth curves

The arithmetic mean of the annual maxima was chosen as the standardising factor. The relationship between the standardised floods ($Q/\bar{Q}$) and return period is called a growth curve. Having decided on a group of stations for which a pooled growth curve is required the steps in the derivation are:

(a) For each station rank the standardised flood values and assign probabilities to each. The probabilities used are those which, when converted to reduced variate values, have been shown to be most appropriate for the Gumbel of EVI distribution (see Flood Studies Report, vol. I.3.2]. For the ith smallest of the n annual maxima, the probability of non-exceedance ($F_i$) is

$$F_i = \frac{i-0.44}{n+0.12}$$

and the corresponding reduced variate value (y) is

$$y = -\ln(-\ln F_i)$$

If required, historical floods may be introduced at this stage by assigning appropriate probabilities and reduced variate values to them (Flood Studies Report vol I p 177).

(b) Average the y and $Q/\bar{Q}$ values for floods from all stations which have y values in the range from -2.0 to -1.5. Repeat this for all intervals of y of 0.5 units wide until all floods have been used. The pairs of averages of y and $Q/\bar{Q}$ form the pooled growth curve.

Note that if the standardised floods were all independent and from the same distribution it would be possible to regard all station-years of data as a single record. The above averaging procedure is valid even if the floods are not independent (which is of course the case), since all stations are subject to broadly similar meteorological conditions in a given year. It does however provide a growth curve to a return period which corresponds only to the largest 'y' value from the longest record. To extend the curve to higher return periods the following can be done.

(c) Form the stations into several groups so that no two stations in any group are close together. (In the Flood Studies Report four or five groups were formed from about 50 stations.) It can then be assumed that there is no dependence within each group.

(d) From each group select the four highest values of $Q/\bar{Q}$ and assign F and y values appropriate to the four highest in N observations where N is the number of station years in the group. Then average these high values as in (b) to provide points on the growth curve at the higher return periods.

The above procedure produces an average growth curve for the region. In reality each station in the region will have a true growth curve which departs to a greater or lesser degree from this average. The average curve will fit less well the flood data at the individual stations than would a single curve for each station. However random sampling variations will be much less for the averaged curve and much greater confidence can be placed in this than could be placed in each individual station's curve. The use of the average curve implies that it is worthwhile to sacrifice the fit at individual stations to obtain a more stable estimate of the overall growth curve. It is implicitly assumed that standardisation by the mean reduces the flood frequency curves at each station to the most comparable set of growth curves.

## The regional growth curves

A pooled growth curve, as described above, was derived for each of the eleven geographical regions which had been used for mean annual flood regressions. A map of the regions is shown in Figure 1. Because of a lack of long records in Region 7, it was decided to pool Regions 6 and 7, thus giving a single curve for the two regions.

The curve for each region was summarised in parametric form, by fitting a curve corresponding to the General Extreme Value (GEV) distribution to the points $(Q/\bar{Q}, y)$. The appropriate curve is $Q/\bar{Q} = u + \alpha/k\ (1 - e^{-ky})$.

Thus the curve for any region may be described by all the three fitted parameters u, $\alpha$ and k. The ten curves are shown in Figure 2.

An averaged curve for the whole of Great Britain was obtained by pooling the data from all regions. Because of the large amount of data it is possible, by using the grouping procedure described above, to plot some points corresponding to high values of T, the return period. Thus, although the individual region curves extended only as far as return periods of approximately 500 years, the averaged curve could be extended out to T = 1000 years. The FSR recommended that the "Great Britain" curve should be used for estimating floods of return period greater than 500 years.

The regional growth curves published in the Flood Studies Report were derived without any analysis of differences or similarities between the curves. The remainder of this report presents the analyses that have since been done to examine these differences.

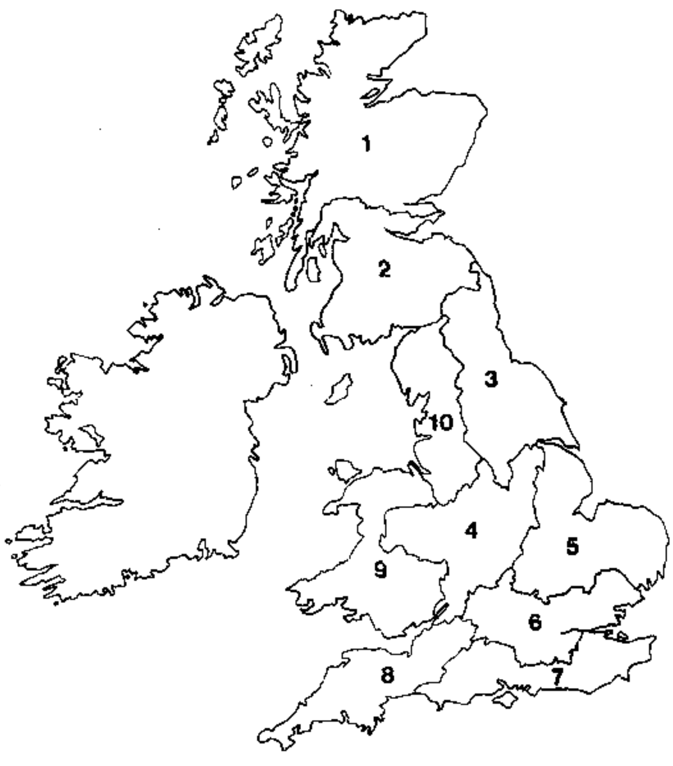


FIGURE 1 The geographical regions used in the Flood Studies Report

## 2. COMPARISON OF REGIONAL GROWTH CURVES

It can be seen from Figure 2 that there are quite large differences between some pairs of region curves, particularly at high return periods. The question to be answered is whether these differences are "real" (ie whether the average distribution of standardised annual maximum floods does differ from one part of the country to another) or whether the apparent differences could be due to sampling fluctuations. It is certain that the standard error of the $Q/\bar{Q}$ estimate at high return periods will be large, as there is necessarily very little data contributing to the estimate.

The Flood Studies Report points out that there is a hydrological argument for accepting differences between curves. A statistical analysis of the differences will now be described. (Region 11, Ireland, was not included in the analysis, and the Regions 6 and 7 were treated separately). Three separate methods of analysis were tried.

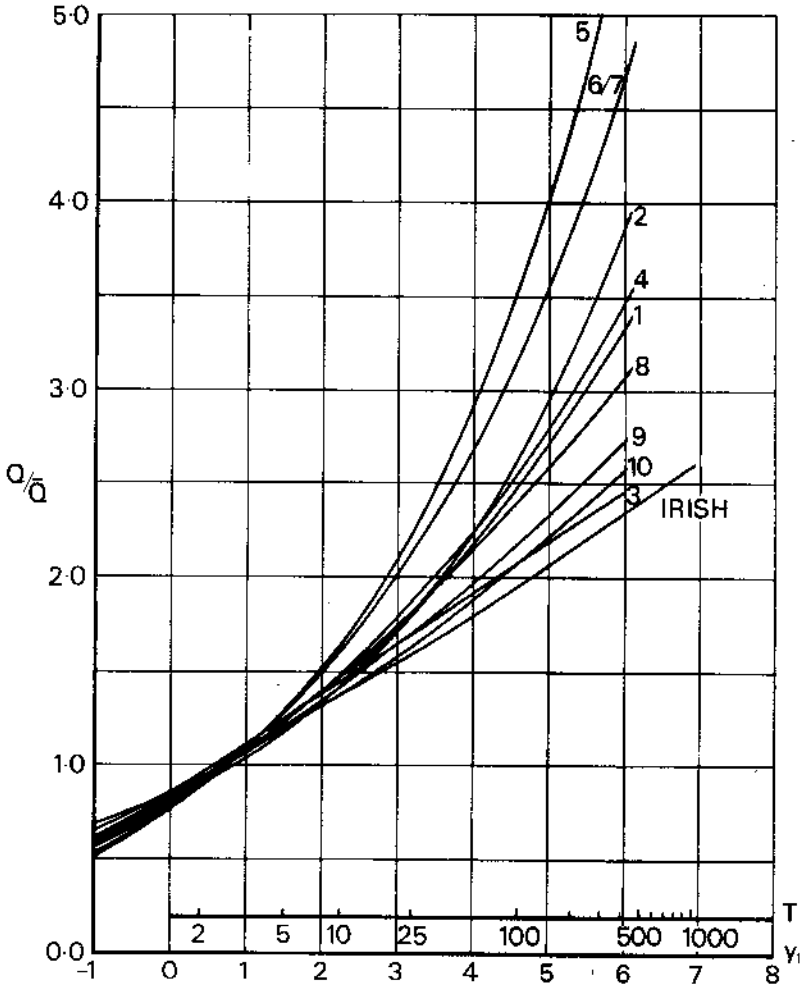


FIGURE 2 The Regional Growth Curves derived in the Flood Studies Report

Method A: Analysis of Variance

The region curves are to be interpreted as average curves for the region, and individual stations may depart from the average line. It is natural to examine whether this variation is large by comparison with the difference between two region curves.

Therefore the "within region" variation, which is the variation amongst the $Q/\bar{Q}$ values in a single region which were averaged and plotted at one y-value, is compared with the "between region" variation, which is the variation amongst the $Q/\bar{Q}$ estimates for different regions which were plotted in the same y-interval. This can be done using the Analysis of Variance (ANOVA).

The ANOVA tests the hypothesis that two or more regions have the same average $Q/\bar{Q}$ value in this y-interval. The test can be repeated for successive y-intervals, and conclusions about the similarity of the set of regions under test drawn for the results over the whole range of y.

Specifically, consider a single y-interval, and compare N regions ($2 \leqslant N \leqslant 10$). Let the $Q/\bar{Q}$ values in the jth region which plot in this y-interval be $x_{ij}$, for $i = 1,\ldots,n_j$, where $n_j$ is the number of values in the jth region and $j = 1,\ldots.N$. Let the total number of values in the y-interval be M,

$$\text{ie} \quad M = \sum_{j=1}^{N} n_j$$

Then the point plotted on the curve for the jth region is the regional mean, $\bar{x}_j$.

$$\bar{x}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} x_{ij}$$

The total (over all N regions) within region sum of squares is WSS.

$$WSS = \sum_{j=1}^{N} \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2 \quad \text{which has (M-N) degrees of freedom.}$$

The within region mean square, WMS, is given by $WMS = WSS/(M-N)$. The between region sum of squares is BSS.

$$\text{BSS} = \sum_{j=1}^{N} n_j (\bar{x}_n - \bar{x})^2 \quad \text{where } \bar{x} \text{ is the overall mean}$$

$$\text{ie } \bar{x} = \frac{1}{M} \sum_{j=1}^{N} \sum_{i=1}^{n_j} x_{ij} = \frac{1}{M} \sum_{j=1}^{N} n_j \bar{x}_j .$$

The BSS has (N-1) degrees of freedom, so the between region mean square is BMS = BSS/(N-1).

The hypothesis to be tested is that the $X_{ij}$ all come from a distribution which has the same mean (ie the $Q/\bar{Q}$ value in this interval should be the same for all regions). If the hypothesis is true, the ratio BMS/WMS has an F distribution with parameters N-1 and $\Sigma n_j - N$.

Let f(0.05) be the value from the F distribution with these parameters which has a probability of only 0.05 of being exceeded. Then the hypothesis is rejected if the ratio BMS/WMS exceeds f (0.05).

*Notes on the application of the method*

When considering the 'within region' variation, a distinction should be made between the variation of all the $Q/\bar{Q}$ values in the region which plot in a given y-interval, and the 'between stations' variation. The difference is due to the fact that several floods from the same station may be plotted in the same y-interval. The 'between stations' variation is the variation between the station mean $Q/\bar{Q}$ values. It is preferable to use this quantity as the 'within region' variation, because of the large amount of dependence between floods from one station with adjacent y-values.

Points obtained by the grouping procedure, or from historical data, were not included. In practice, they could not have been analysed because they fall into the high y-intervals which do not contain enough points for comparison by ANOVA. Although there are points with y-values ranging from -2.0 to +5.5, only those intervals from -1.5 to +3.5 or +4.0 contained enough points for the analysis.

Finding the variance of a set of $Q/\bar{Q}$ values which plot at different y-values in an interval means that this estimate of the variance may be slightly too large. This effect may be appreciable in the calculation of WMS, but not in the calculation of BMS, since the $\bar{x}_j$ all correspond to a mean y-value which is always very close to the centre of the interval. Therefore, the obtained ratio BMS/WMS may be slightly too small, and the test errs on the side of not rejecting the hypothesis quite often enough. This gives extra confidence in the result of the test when it shows that two curves are different.

The assertion that the variance ratio BMS/WMS has an F distribution is a consequence of an assumption that a sample from a GEV distribution which plots in a given y-interval is normally distributed about the GEV curve, and that the variance of the normal distribution does not change much among the different GEV distributions which are used to describe the regions. This assumption seems to be justified except for very high y-values, where the skewness of the distribution of floods is likely to affect the distribution of the errors.

It is difficult to draw precise conclusions on the basis of the whole set of variance ratios obtained in a single comparison. They cannot be considered as an independent sample from an F-distrubution, partly because there is dependence between successive intervals, and partly because the number of degrees of freedom (ie the parameters of the F-distribution) change between intervals. It is possible, however, to assess similarities between regions in a subjective way by looking at the results over the whole range of y.

*Results*

A large number of combinations of regions were examined, though not all possible ones since this would have been prohibitive in time and unnecessary. The results are shown in Table I. For each set of regions, the variance ratio obtained on each y-interval is given, together with the critical value, f (0.05) of the appropriate F distribution. Intervals where the hypothesis that the regions are the same is rejected are denoted by a cross, and those where it is not rejected remain blank. Hence a row of blanks indicates a set of regions which appear to be very similar to one another.

*Conclusions*

The comparison of all ten regions rejects the hypothesis that they all come from the same distribution on every y-interval but the highest. It is therefore concluded that there is a real difference in the distribution of annual maximum floods in different parts of the country. However, it should be noted that regions cannot be distinguished on the basis of high y-value points alone, since these contain such a large sampling variance within regions.

Certain pairs of regions were not distinguished on any interval by the test. They are: 1 and 3, 3 and 9, 7 and 8. 4 and 8, and 9 and 10 were only distinguished on one interval.

Large groups of regions which were not distinguished are: 1, 3, 9 and 10, and 4, 6 and 8. (But note that the hypothesis was rejected on one interval for 3, 9 and 10).

The hypothesis was rejected in all comparisons which included Region 2.

The overall conclusions may be summarised thus:

TABLE 1 ANOVA and Chi-squared tests of the hypotheses that groups of regions have identical annual maximum distributions. On each interval a cross denotes rejection of the hypothesis and a blank denotes acceptance.

| Regions Compared | -1.5 -1.0 | -1.0 -0.5 | 0.5 0.0 | 0.0 0.5 | 0.5 0.0 | 1.0 1.5 | 0.5 2.0 | 2.0 2.5 | 2.5 3.0 | 3.0 3.5 | 3.5 4.0 | 4.0 4.5 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | Y-Intervals | | | | | | | |
| | 4.79 | 12.72 | 5.26 | 3.22 | 3.47 | 6.11 | 3.96 | 4.66 | 4.27 | 1.08 | | | Variance Ratio |
| All | 1.95 | 1.90 | 1.90 | 1.90 | 1.90 | 1.90 | 1.90 | 1.90 | 1.90 | 1.95 | | | f(0.05) |
| | x | x | x | x | x | x | x | x | x | | | | Accept Hypothesis? |
| | .00 | 1.42 | .22 | 1.23 | 1.39 | 0.22 | .00 | .65 | .41 | 1.98 | 1.98 | .94 | |
| 1, 3 | 4.00 | 3.95 | 3.95 | 3.95 | 3.96 | 3.95 | 3.96 | 4.00 | 4.05 | 4.10 | 4.96 | 5.59 | |
| | 4.34 | 1.27 | .98 | .34 | .00 | .30 | 2.18 | .16 | 1.71 | .03 | .02 | .10 | |
| 9, 10 | 4.1 | 3.95 | 3.95 | 3.95 | 3.95 | 3.95 | 4.02 | 4.03 | 4.04 | 4.41 | 4.75 | 5.32 | |
| | x | | | | | | | | | | | | |
| | 1.72 | .67 | .59 | .77 | 1.09 | .55 | 3.55 | 1.80 | 1.78 | .24 | .07 | .33 | |
| 3, 9, 10 | 3.15 | 3.06 | 3.06 | 3.06 | 3.07 | 3.06 | 3.11 | 3.12 | 3.12 | 3.28 | 3.68 | 3.98 | |
| | | | | | | | x | | | | | | |
| | 1.20 | .93 | .64 | 1.69 | 2.42 | .35 | 2.10 | 2.46 | 1.29 | .75 | .37 | .31 | |
| 1,3,9,10 | 2.72 | 2.66 | 2.66 | 2.66 | 2.66 | 2.66 | 2.67 | 2.67 | 2.71 | 2.77 | 3.05 | 3.29 | |
| | 1.70 | .61 | .00 | .38 | 1.61 | .28 | 1.61 | 1.52 | .19 | .30 | .15 | 1.40 | |
| 3, 9 | 4.08 | 3.94 | 3.94 | 3.94 | 3.95 | 3.94 | 3.98 | 4.00 | 4.00 | 4.22 | 5.12 | 5.99 | |
| | 4.61 | 2.49 | 1.47 | 1.95 | 3.84 | 5.27 | 5.85 | 2.55 | 1.10 | 1.63 | .07 | | |
| 2,3,9,10 | 2.47 | 2.37 | 2.38 | 2.38 | 2.38 | 2.39 | 2.40 | 2.42 | 2.44 | 2.51 | 2.76 | | |
| | x | x | | | x | x | x | x | | | | | |

Analysis of Variance

| Regions Compared | Chi-squared value | Degrees of Freedom | C(0.05) | Accept Hypothesis ? |
|---|---|---|---|---|
| All | 364.89 | 144 | 170 | x |
| 1, 3 | 16.84 | 18 | 28.87 | |
| 9, 10 | 15.73 | 18 | 28.87 | |
| 3, 9, 10 | | | | |
| 1,3,9,10 | 57.61 | 54 | ~73 | |
| 3, 9 | | | | |
| 2,3,9,10 | 110.71 | 64 | ~83 | x |

Chi-Squared Test

TABLE 1 contd.

| Regions Compared | -1.5 -1.0 | -1.0 -0.5 | -0.5 0.0 | 0.0 0.5 | 0.5 1.0 | 1.0 1.5 | 1.5 2.0 | 2.0 2.5 | 2.5 3.0 | 3.0 3.5 | 3.5 4.0 | 4.0 4.5 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | Y-Intervals | | | | | | | |
| | 12.75 | 7.45 | 4.35 | 5.83 | .35 | 10.52 | 12.07 | 5.40 | .55 | 4.38 | .62 | | Variance Ratio |
| 1, 2 | 4.06 | 3.94 | 3.94 | 3.94 | 3.94 | 3.94 | 3.98 | 3.99 | 4.06 | 4.15 | 4.96 | | f(0.05) |
| | x | x | x | x | | x | x | x | | x | | | Accept Hypothesis? |
| | 8.55 | 18.75 | 13.20 | 3.09 | 4.38 | 16.87 | 12.36 | 5.46 | 2.63 | 2.83 | .31 | | |
| 2, 4 | 4.08 | 3.95 | 3.95 | 3.96 | 3.95 | 3.95 | 4.00 | 4.00 | 4.05 | 4.16 | 7.71 | | |
| | x | x | x | | x | x | x | x | | | | | |
| | 4.16 | 14.61 | 14.56 | 1.87 | .28 | 2.08 | 12.56 | 5.35 | 2.67 | 9.13 | .01 | | |
| 2, 8 | 4.22 | 3.95 | 3.95 | 3.95 | 3.96 | 3.95 | 4.02 | 4.03 | 4.03 | 4.45 | 7.71 | | |
| | | x | x | | | | x | x | | x | | | |
| | .02 | .00 | .10 | .00 | 4.2 | 3.34 | .12 | .08 | .24 | .14 | 16.98 | | |
| 4, 8 | 4.18 | 3.95 | 3.95 | 3.95 | 3.98 | 3.95 | 4.04 | 4.08 | 4.08 | 4.32 | 18.51 | | |
| | | | | | x | | | | | | | | |
| | 4.00 | 10.69 | 8.52 | 1.36 | 3.08 | 6.90 | 7.14 | 3.33 | 1.68 | 1.96 | .20 | | |
| 2,4,8 | 3.20 | 3.06 | 3.06 | 3.06 | 3.07 | 3.06 | 3.11 | 3.12 | 3.12 | 3.28 | 5.79 | | |
| | x | x | x | | x | x | x | x | | | | | |
| | 1.35 | 3.46 | .07 | 0.04 | 2.34 | 1.68 | .85 | 1.00 | .18 | .22 | .26 | | |
| 4,6,8 | 3.20 | 3.92 | 3.92 | 3.92 | 3.93 | 3.92 | 3.98 | 3.98 | 4.00 | 4.17 | 3.98 | | |
| | .39 | .12 | .50 | 2.79 | 4.25 | 3.90 | 1.18 | 2.46 | 1.58 | .13 | | | |
| 4,7,8 | 3.31 | 3.11 | 3.11 | 3.11 | 3.12 | 3.11 | 3.14 | 3.21 | 3.17 | 3.44 | | | |
| | | | | | x | x | | | | | | | |

Analysis of Variance

| Regions Compared | Chi-squared value | Degrees of Freedom | C(0.05) | Accept Hypothesis ? |
|---|---|---|---|---|
| 1, 2 | 48.43 | 16 | 26.3 | x |
| 2, 4 | 84.50 | 16 | 26.3 | x |
| 2, 8 | 71.26 | 16 | 26.3 | x |
| 4, 8 | 20.47 | 16 | 26.3 | |
| 2,4,8 | 119.6 | 32 | ∿ 45 | x |
| 4,6,8 | | | | |
| 4,7,8 | 41.40 | 32 | ∿ 45 | |

Chi-Squared Test

TABLE 1 contd.

| Regions Compared | -1.5 -1.0 | -1.0 -0.5 | -0.5 0.0 | 0.0 0.5 | 0.5 1.0 | 1.0 1.5 | 1.5 2.0 | 2.0 2.5 | 2.5 3.0 | 3.0 3.5 | 3.5 4.0 | 4.0 4.5 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | Y-Intervals | | | | | | | |
| | .60 | .17 | .48 | 3.53 | 1.32 | 1.97 | 1.93 | 3.09 | 1.44 | .03 | | | Variance Ratio |
| 7, 8 | 4.75 | 4.00 | 4.00 | 4.00 | 4.06 | 4.02 | 4.12 | 4.35 | 4.12 | 7.71 | | | f(0.05) |
| | | | | | | | | | | | | | Accept Hypothesis? |
| | 3.66 | 7.04 | 6.81 | 4.64 | 3.56 | 4.51 | 4.46 | 4.76 | 3.00 | 1.84 | | | |
| 2,4,7,8 | 2.80 | 2.67 | 2.67 | 2.67 | 2.68 | 2.67 | 2.70 | 2.72 | 2.72 | 2.88 | | | |
| | x | x | x | x | x | x | x | x | x | | | | |
| | .02 | 4.91 | 1.13 | 4.85 | 2.24 | 5.16 | .26 | 2.90 | 1.43 | .02 | | | |
| 6,7 | 4.30 | 3.98 | 3.98 | 3.98 | 3.99 | 3.98 | 4.09 | 4.08 | 4.12 | 4.84 | | | |
| | | | | x | | x | | | | | | | |
| | 1.02 | 2.87 | .44 | 2.18 | 2.91 | 3.36 | 1.12 | 2.26 | 1.13 | .18 | | | |
| 4,6,7,8 | 2.88 | 2.67 | 2.67 | 2.67 | 2.68 | 2.67 | 2.79 | 2.79 | 2.80 | 2.91 | | | |
| | | x | | | x | x | | | | | | | |
| | 6.84 | 10.66 | 1.09 | 2.61 | .14 | 5.22 | 2.15 | .09 | .85 | .90 | 1.29 | | |
| 4,5 | 4.13 | 3.96 | 3.96 | 3.96 | 3.97 | 3.96 | 4.00 | 4.05 | 4.05 | 4.32 | 5.32 | | |
| | x | x | | | | x | | | | | | | |
| | 1.17 | .77 | 1.28 | 2.57 | 4.03 | 10.28 | 6.70 | .52 | .03 | 4.20 | .11 | | |
| 5,6 | 4.17 | 3.95 | 3.95 | 3.95 | 3.96 | 3.95 | 4.02 | 4.03 | 4.03 | 4.67 | 4.60 | | |
| | | | | | x | x | x | | | | | | |
| | .71 | 4.31 | .87 | 2.92 | 4.52 | 9.90 | 4.78 | 2.31 | 1.05 | 2.43 | | | |
| 5,6,7 | 3.28 | 3.07 | 3.08 | 3.07 | 3.08 | 3.08 | 3.13 | 3.15 | 3.14 | 3.98 | | | |
| | | x | | | x | x | x | | | | | | |

Analysis of Variance

| Regions Compared | Chi-squared value | Degrees of Freedom | C(0.05) | Accept Hypothesis ? |
|---|---|---|---|---|
| 7, 8 | 19.51 | 16 | 26.3 | |
| 2,4,7,8 | | | | |
| 6,7 | 14.42 | 16 | 26.3 | |
| 4,6,7,8 | 56.18 | 48 | ∿ 65 | |
| 4,5 | 20.91 | 16 | 26.3 | |
| 5,6 | 24.17 | 16 | 26.3 | |
| 5,6,7 | 39.2 | 32 | ∿ 45 | |

Chi-Squared Test

TABLE 1 contd.

| Regions Compared | Y-Intervals -1.5 -1.0 | -1.0 -0.5 | -0.5 0.0 | 0.0 0.5 | 0.5 1.0 | 1.0 1.5 | 1.5 2.0 | 2.0 2.5 | 2.5 3.0 | 3.0 3.5 | 3.5 4.0 | 4.0 4.5 | | Chi-squared value | Degrees of Freedom | C(0.05) | Accept Hypothesis |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1.99 | 4.44 | .56 | 2.13 | 3.42 | 6.63 | 2.23 | 1.90 | .93 | .61 | | | Variance Ratio | | | | |
| 4,5,6,7,8 | 2.51 | 2.44 | .44 | 2.44 | 2.44 | 2.44 | 2.45 | 2.47 | 2.47 | 2.64 | | | f(0.05) | 83.22 | 64 | ∿ 83 | ? |
| | | x | | | x | x | | | | | | | Accept Hypothesis? | | | | |
| 3,5 | | | | | | | | | | | | | | 49.19 | 16 | 26.3 | x |
| 5,10 | | | | | | | | | | | | | | 71.70 | 17 | 27.59 | x |
| 6,7,8 | | | | | | | | | | | | | | 40.63 | 32 | ∿ 45 | |
| 1,2,3 | | | | | | | | | | | | | | 76.72 | 32 | ∿ 45 | x |
| 1,9 | | | | | | | | | | | | | | 17.79 | 16 | 26.3 | |

Analysis of Variance

Chi-Squared Test

(a) Regions 1, 3, 9, 10 behave very similarly.

(b) Region 2 does not exhibit much similarity with any other region.

(c) 6, 7 and 8 are very similar. So are 7 and 8. 5 shows some similarity with all of these. The set 4, 5, 6, 7, 8 shows as much similarity as 4, 6, 7, 8.

It seems that these 5 regions cannot be considered as identical, but they are certainly all quite similar to one another.

Method B: Non parametric tests

It was felt that an analysis which did not depend on the assumed Gumbel plotting positions would be beneficial. For this purpose, all the standarised annual maximum floods $Q/\bar{Q}$ which have been recorded in a given region (not including historical data) were regarded as sample points from a single regional distribution. Then, considering the data from each of the 10 regions as 10 samples, two or more samples could be compared, using non-parametric statistical tests to decide whether or not they come from one (unspecified) distribution.

The major assumption made in the application of the tests is that all floods in a region are independent. This is not the case, but it was felt that since the amount of dependence is likely to be small, and since the effect of small dependence on the results of the tests can to some extent be judged, such an analysis was still worthwhile.

It is required to test whether two or more samples differ in any respect whatever, ie in mean, variance, skewness, etc. (Mean is expected to be the same since each station record is standardised to have mean one). The appropriate statistical tests are, therefore:

(a) The Chi-Squared Test, for two or more samples.

(b) The Kolmogorov-Smirnov test, for two samples only.

(a) *The Chi-squared test*

This tests the hypothesis that k samples come from identical distributions. The range of $Q/\bar{Q}$ values is divided into r intervals. The criterion for choosing the sizes and number of intervals is described below. (n.b. The 'intervals' used here should not be confused with those used in the ANOVA, which were intervals in the range of y, not $Q/\bar{Q}$).

The test statistic is

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{k} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

where $O_{ij}$ is the number of floods in the ith interval of the jth sample, and $E_{ij}$ is the number expected to be in that interval if the hypothesis is true. r is the number of intervals in the range.

$E_{ij}$ is calculated as $\left(\sum_{t=1}^{r} O_{tj}\right) \times \left(\sum_{t=1}^{k} O_{it}\right)/N$

where N is the total number of floods in all samples.

Then under the hypothesis, $\chi^2$ has an approximately Chi-squared distribution with $(r-1) \times (k-1)$ degrees of freedom. The hypothesis is rejected if $\chi^2 > C\ (0.05)$, the value which has a probability of only 0.05 of being exceeded.

The criterion for interval size is that there should be as many intervals as possible, subject to the restriction that $E_{ij}$ should not be less than 1 for any i or j, and that fewer than 20% of the $E_{ij}$'s should be less than 5. It was found that for most groups of regions a satisfactory set of intervals was one interval for $Q/\bar{Q}$ less than 0.5, fifteen intervals of width 0.1 up to 2.0, and one interval for $Q/\bar{Q}$ greater than 2.0.

### (b) *The Kolmogorov-Smirnov test*

This tests the hypothesis that two samples come from identical distributions. A cumulative frequency distribution is constructed as follows. The range of $Q/\bar{Q}$ values is divided into intervals. For convenience, these are at first taken to be the same as for the Chi-squared test. If the hypothesis is not rejected, the test was repeated, using smaller intervals.

Denote the interval boundaries by $\{X_j\}$. Let $S_i(X_j)$ = the proportion of values in sample i which are less than or equal to $X_j$, i=1,2. The test statistic is $D = \max_{X_j} |S_1(X_j) - S_2(X_j)|$.

The hypothesis is rejected if $D \geq d\ (0.05)$ which is the value in the sampling distribution of D which has a probability of 0.05 of being exceeded.

*Results*

The tests were performed, as for the ANOVA, on many different pairs and groups of regions. The results of the Chi-squared test are displayed in Table 1, so that they may be compared with the ANOVA results. Table 2 gives the results of the Kolmogorov-Smirnov test. It is reassuring to note that, for pairs of regions when both tests were used, they always supported each other, ie when the Kolmogorov-Smirnov test rejected the hypothesis, then the Chi-squared test did also, and vice versa.

Also, for the Kolmogorov-Smirnov test, reducing the interval width beyond the 0.1 used initially did not greatly increase the value of D, and if the hypothesis was not rejected under the first interval scheme, neither was it rejected for smaller intervals.

TABLE 2 Kolmogorov-Smirnov Test of the hypothesis that two regions have identical annual maximum distributions

| Regions Compared | Interval Width = 0.1 | | | Interval Width = 0.05 | | |
|---|---|---|---|---|---|---|
| | D | d(0.05) | Accept Hypothesis? | D | d(0.05) | Accept Hypothesis? |
| 1 and 3 | 0.026 | 0.072 | | 0.045 | 0.072 | |
| 9 and 10 | 0.031 | 0.078 | | 0.031 | 0.078 | |
| 3 and 5 | 0.097 | 0.078 | x | | | |
| 5 and 10 | 0.125 | 0.08 | x | | | |
| 5 and 6 | 0.054 | 0.078 | | 0.059 | 0.078 | |
| 6 and 7 | 0.081 | 0.113 | | 0.095 | 0.113 | |
| 7 and 8 | 0.091 | 0.127 | | 0.091 | 0.124 | |
| 4 and 8 | 0.034 | 0.091 | | 0.037 | 0.091 | |
| 1 and 2 | 0.085 | 0.075 | x | | | |
| 2 and 4 | 0.210 | 0.078 | x | | | |
| 2 and 8 | 0.143 | 0.091 | x | | | |
| 4 and 5 | 0.052 | 0.08 | | 0.052 | 0.08 | |
| 1 and 9 | 0.036 | 0.073 | | 0.036 | 0.073 | |

*Effect of non-independence*

When considering all the floods from a given region as a single sample from some distribution there will in fact be a certain amount of dependence between them, although the between-region dependence will be negligible.

As an approximation to the effect on the Chi-squared test, assume that the number of floods observed in interval i of region j is reduced by a factor p. i.e. $O_{ij}' = pO_{ij}$ where $p < 1$.

$$\text{Then also } E'_{ij} = pE_{ij} \text{ so } \frac{(O'_{ij} - E'_{ij})^2}{E'_{ij}} = p\,\frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

and $\chi^2$ is reduced by a factor p also. The number of degrees of freedom is not affected.

Thus the consequence of assuming independence is to obtain a slightly inflated $\chi^2$-value, which might lead to rejection of the hypothesis when this was not justified. But on examining the results it is found that the hypothesis was only rejected by a very large margin, and a small reduction in the $\chi^2$ value would not have affected this.

For the Kolmogorov-Smirnov test, if the effect of dependence is to reduce the number of floods in each sample by a factor p, then assuming the same amount of dependence along the range of $Q/\bar{Q}$, the proportion of floods in each sample less than $X_j$, $S_i(X_j)$, is unaffected and hence D is not altered. However, d(0.05) is increased by a factor $1/\sqrt{p}$, which again may lead to unjustified rejection of the hypothesis. But again the margins were large and for small amounts of dependence, the decisions would not be altered.

It is concluded that the non-independence does not invalidate the results of the two tests.

*Conclusions*

(1) The Chi-squared test rejects conclusively the hypothesis that all ten regions are the same.

(2) Pairs of regions which are not distinguished are: 1 and 3, 9 and 10, 1 and 9, 5 and 6, 7 and 8, 4 and 8.

(3) The Chi-squared test does not distinguish between regions 1, 3, 9 and 10.

(4) Various combination of the region 4, 5, 6, 7 and 8 are not distinguished, although for three or more of these regions the $\chi^2$-value is quite high, ie close to the rejection value C(0.05). It is concluded, as for the ANOVA, that the five regions are similar but cannot be considered to be identical.

(5) Region 2 does not appear to be similar to any other region.

### (c) *Focus on the high floods: the median test*

It was thought possible that not enough weight had been given, in the non-parametric tests, to the existence in some regions of very high floods. They have so far only been considered as members of the set of floods with $Q/\bar{Q}$ over 2.0. The Chi-squared test could not be adjusted to take account of them because of the requirement for $E_{ij}$ to be greater than 1 for all i and j. The Kolmogorov-Smirnov test is not sensitive to differences between the regions in this range of $Q/\bar{Q}$, since the differences, though important, are relatively small.

An attempt was made to examine the high floods more closely by considering only those floods which have $Q/\bar{Q}$ greater than 2.0. These will be referred to as the 'high' floods. The number of high floods in each region ranges from 6, in regions 2 and 10, to 21 in region 6.

If all regions have the same distribution, the mean high flood will be the same for all regions. If, as is now to be expected, the regions differ, then the mean high flood will also differ (although the mean of $Q/\bar{Q}$ over its whole range is always 1).

The differences in central tendency of groups of values can be analysed using the median test. Briefly, the median test for k samples is as follows. First, the median of all the values from the k regions is found. Then for each region, the number of values above and below the combined median is counted. The hypothesis that the medians for all regions are the same is tested using either the Chi-squared test, if there is a sufficient number of values, or the Fisher test if there are only a few values.

However, the test was not found to be very discriminating. It did display a difference between regions, but did not, for example, distinguish region 1 from any other region. It was found that if the regions were split into the groups 4, 5, 6, 7, 8 and 1, 2, 3, 9, 10, the test exhibited similarities between the regions in the same group, and differences between most pairs of regions consisting of one from each group. The test does not, therefore, contradict any of the previous conclusions, but does not seem to be strong enough to prove or disprove any new conjectures.

Method C: Simulation

This method differed from the first two in that, rather than looking for differences in the data, an attempt was made to analyse the sampling properties of the GEV distribution.

A collection of simulated station records was generated from the GEV distribution with parameters corresponding to the GB curve. Sets of stations were pooled and analysed in exactly the same way as the real data, to produce simulated 'region curves'. A large number of these were generated. The idea was to examine this 'sample' of region curves, to see if the variation between real region curves could be accounted for as sampling variance amongst curves from the same distribution.

However, a number of problems were encountered. Firstly, a marked effect due to standardisation by the mean of a station record was noticed. The simulated region curves were all considerably 'flatter' than the curve from which they had been generated, ie the $Q/\bar{Q}$ values corresponding to high y-values were very much smaller than the values in the original distribution. This effect had been expected, since the variance of a sample is necessarily reduced by dividing by the sample mean, but the magnitude of the effect was surprising and made deductions about the variation amongst curves difficult.

Secondly, no satisfactory measure of variation between curves was found. It was possible to look at individual y-intervals as in the ANOVA, and compare the variation amongst simulated points with that amongst the real region curves. When this was done, it was found that there was more variation in the real data than might have been expected if a single distribution was adequate for all regions, but no more exact statement could be made, since, due to the flattening effect mentioned above, the variation was about different means in the real and simulated cases. Thus the simulation exercise did not contribute anything further towards the problem of differences between regions. However, it demonstrated the rather worrying effect of standardisation by the mean, which may mean that the region curves ought to be steeper, for prediction purposes, that they are at present. This is to be the subject of further investigation.

## 3. CONCLUSIONS

Two successful types of test have been performed on the data from the ten regions. Method A, the analysis of variance, is dependent on the method of analysis of the data through the plotting position, y. Method B, the non-parametric tests, looks only at the samples of raw data. The results of both methods support each other.

There are probably not strong enough grounds for concluding that any two regions are identical. However, the similarities between regions could be very useful in pooling the data to obtain curves valid at higher return periods. Since it has been shown that it is not possible to use one curve for the whole country, serious error may be introduced by, for example, using the GB curve for region 5 for T > 500 years. A more sensible method would be to pool the data from regions 4, 5, 6, 7, 8 to obtain a curve which could be used for higher return periods for these regions. Similarly, a pooled curve could be derived for regions 1, 3, 9 and 10. Region 2 remains a problem. However since both geographically and in the appearance of its growth curve it is closer to 1, 3, 9 and 10 than to the other regions, it seems sensible to pool it with these regions. Thus two pooled curves are obtained, one for NW Britain and one for the SE, which could be used for higher return periods in the same way that it was previously recommended that the GB curve be used. The two curves are shown in Figure 3.

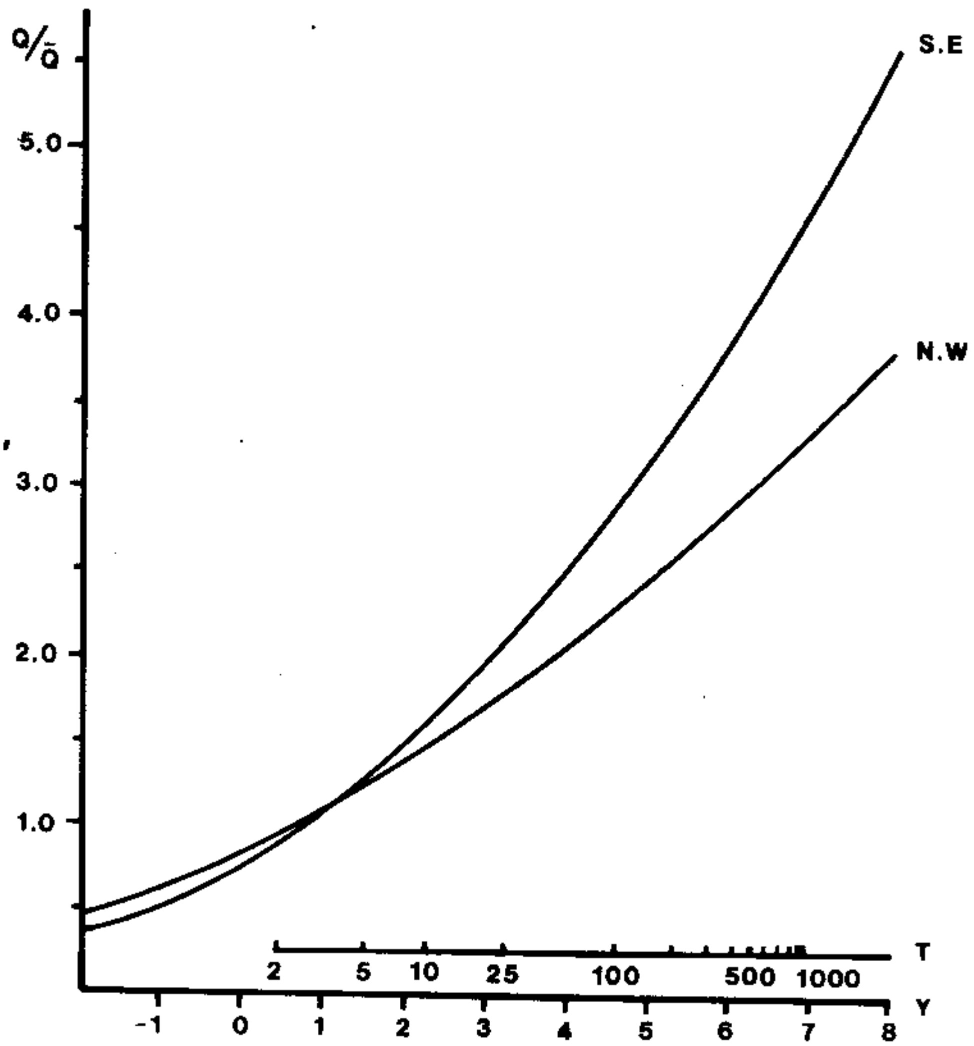


FIGURE 3 The pooled growth curves obtained by treating north-west and south-east Britain separately

The similarities and differences between some regions seem surprising when compared with the <u>fitted</u> growth curves shown in Figure 2. For example, the Region 1 curve looks very different from Region 3. This suggests that too much weight may have been given to points plotted at high return periods when fitting the GEV curves to the data. Inclusion of historical data also had some effect.

At present more annual maximum flood data are being collected, the intention being to re-analyse the total amount of data. The conclusions set out above provide a useful guide for future work.

## REFERENCES

NATURAL ENVIRONMENT RESEARCH COUNCIL (1975) Flood Studies Report, Volume I Hydrological Studies

CHOW, V T (1953) Frequency analysis of hydrologic data with special application to rainfall intensities. University of Illinois Engineering Experimental Station, Bulletin Series No. 414, 28-29