RC 20349 (89939) 1/22/96
Environmental Science 17 pages

Research Report

Regional frequency analysis of floods in central Appalachia

J. R. M. Hosking and J. R. Wallis

IBM Research Division
T. J. Watson Research Center
Yorktown Heights, NY 10598

IBM Research Division
Almaden • T.J. Watson • Tokyo • Zurich

Regional frequency analysis of floods in central Appalachia

J. R. M. Hosking and J. R. Wallis

Mathematical Sciences Dept.
IBM Research Division
P. O. Box 218
Yorktown Heights, NY 10598

Abstract. Data on annual maximum streamflow at 104 gaging stations in the Central Appalachia region of the United States are analysed by regional frequency analysis using L-moments. Each of the four stages in regional frequency analysis is illustrated: screening of the data, identification of homogeneous regions, choice of a frequency distribution, and estimation of the frequency distribution. Particular emphasis is given to identification of homogeneous regions, which is usually the most difficult stage of the analysis. The main features of the analysis are the use of cluster analysis to tentatively identify homogeneous regions, and constant attention to hydrological reasoning in order to ensure that regions obtained by statistical procedures are also physically meaningful.

1. Introduction

Smith (1992) analysed data on annual maximum streamflow at 104 gaging stations in the Central Appalachia region of the United States. Smith was interested in relating the first two moments of the at-site frequency distributions to the drainage area of the basins. We use the same set of sites to illustrate another approach to estimating the frequency of extreme events: regional frequency analysis using L-moments.

Regional frequency analysis augments the data from the site of interest by using data from other sites that are judged to have frequency distributions similar to that of the site of interest. A simple but effective method of regional frequency analysis is the index-flood procedure, which summarizes the data by a weighted average of statistics computed from the data for each site in a "region"; a regional frequency distribution is then computed from these regional average statistics. L-moments (Hosking, 1990) are particularly suitable statistics for summarizing the sites' data. This approach is briefly described by Hosking and Wallis (1993), whose notation and terminology we also use here. A fully detailed account of regional frequency analysis using L-moments will appear in Hosking and Wallis (1996).

This report aims to describe the practical steps involved in the application of regional frequency analysis. Each of the four stages identified by Hosking and Wallis (1993) is illustrated: screening of the data, identification of homogeneous regions, choice of a frequency distribution, and estimation of the frequency distribution. Particular emphasis is given to identification of homogeneous regions, which is usually the most difficult stage of the analysis. The main features of the present analysis are the use of cluster analysis to tentatively identify homogeneous regions, and constant attention to hydrological reasoning in order to ensure that regions obtained by statistical procedures are also physically meaningful.

2. Data

Data were obtained from "Hydrodata" CD-ROMs (Hydrosphere, 1993), which reproduce data from the U.S. Geological Survey's WATSTORE data files.

The study area includes parts of the Piedmont and Valley and Ridge physiographic provinces of Virginia and Maryland. Locations of the sites are shown in Figure 1. Drainage areas of the basins cover a very wide range, from $0.3\,\mathrm{mi}^2$ to $10\,000\,\mathrm{mi}^2$. Elevations of the gaging sites range from 10 ft to 2 000 ft above sea level. In addition to these variables, the data set contains the gage location (latitude and longitude) and the magnitude and date of the maximum instantaneous streamflow in each "water year", running from Oct. 1 through Sep. 30, for some or all of the years 1895–1991. Some variables that might be expected to influence the frequency distribution, such as mean annual precipitation over the basin, the underlying geology of the basin, and the extent to which the basin is forested or urbanized, are not available.

Two gage elevations are missing from the data set, and were estimated from maps. 26 streamflow values were missing and were ignored. 24 data points are marked as

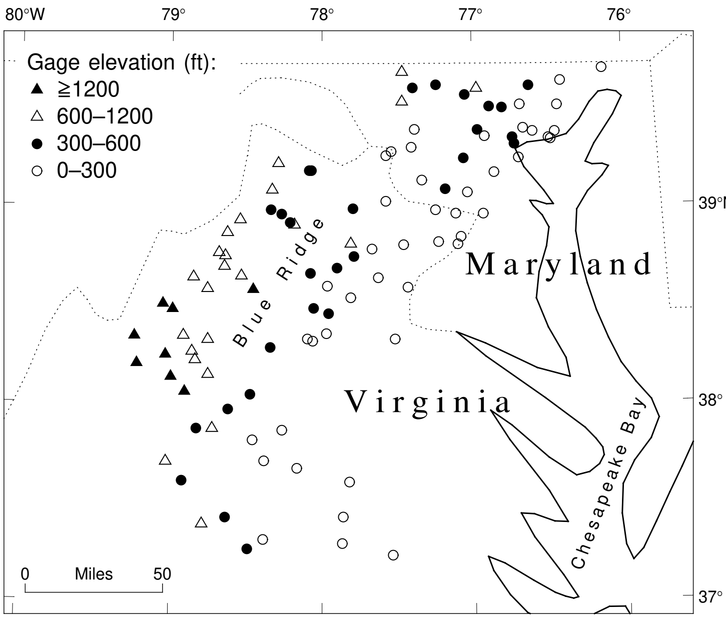

Figure 1. Streamflow gaging stations in central Appalachia.

being "historic values". All but one of these are separated from the main sequence of the gaged record by one or more missing years. We suspect that these values are present in the data only because their magnitude is exceptionally large: therefore, their inclusion in the data would prevent the data from being regarded as a random sample. Accordingly, they were excluded from the data set. The other historic value occurred at the start of a continuous sequence of gaged record and was retained. Site 01626000 has an observation in 1943 but no others until a continuous period of gaging started in 1953. The value for 1943 is unusually high, and appears to be a historic value that was not marked as such. We deleted it from the data set, again because we suspect that it is present only because its magnitude is exceptionally large.

With these modifications, the data set contains a total of 4456 streamflow observations, with sample sizes at the 104 sites varying from 19 to 97. The at-site sample L-CV ranges from 0.22 to 0.71, and the at-site sample L-skewness ranges from 0.14 to 0.74, with one site having the exceptionally low value -0.18. The average value for L-CV is 0.42 and for L-skewness is 0.44. These are high values, compared with other U.S. annual maximum streamflow data, and indicate that the frequency distributions are highly skew.

3. Initial screening of data

The L-CV and L-skewness of the data are shown in Figure 2. Treating the entire set of 104 sites as a single region, the discordancy statistic D_i of Hosking and Wallis

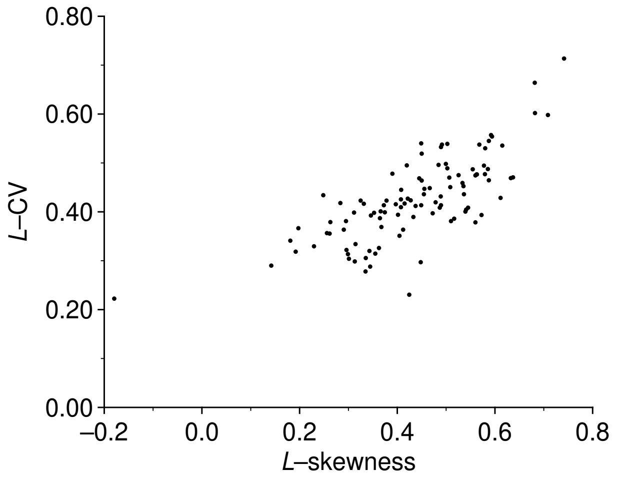

Figure 2. L-moment ratios for the Appalachian streamflow data.

(1993) was calculated for each site.

One site stands out as being particularly discordant, with $D_i = 13.5$. This is the site with negative L-skewness. The data for this site, site 01624800, are shown in Figure 3, together with the data for the site's nearest neighbor, site 01626000. There is good qualitative agreement between the sites: at both, the data values for 1968, 1981 and 1988 are among the lowest and those for 1969 and 1986 are high. To some extent the negative L-skewness at site 01624800 occurs because the period of gaging included three years that had particularly low maximum flows. However, there are also physical reasons that may explain the pattern of the data at this site. The geology of the basin consists of fractured quartzite sandstone overlying dolomite, resulting in a Karstlike hydrology: the basin is very porous and flood response to even the most extreme storm events is slow. In consequence, the frequency distribution of annual maximum flows at the gaging site should be less skew than at sites with impervious geology and a fast response time to extreme storms. It is not clear whether this site is geologically so atypical of the rest of the data that it should be excluded from the analysis. We therefore decided to retain the site in the data set. As noted below, this makes little difference to the results of the regional frequency analysis.

Other checks on the data were inconclusive. There are no obvious trends or changes in level apparent in the data. Some sites have unusually high sample L-CV and L-skewness, but the data appear to be correct: at the four sites with L-CV greater than 0.6, the data contain a single high outlier occurring on 20 August 1969 or 21 June 1972. These dates coincide with hurricanes (Camille and Agnes, respectively) that passed over or near the study area.

One odd feature of the data concerns two sites on the Rapidan River. Site 01665500 is upstream of site 01667500. A plot of the data for these sites, Figure 4, shows that in one year, 1961, the maximum flow at the upstream site exceeded that at

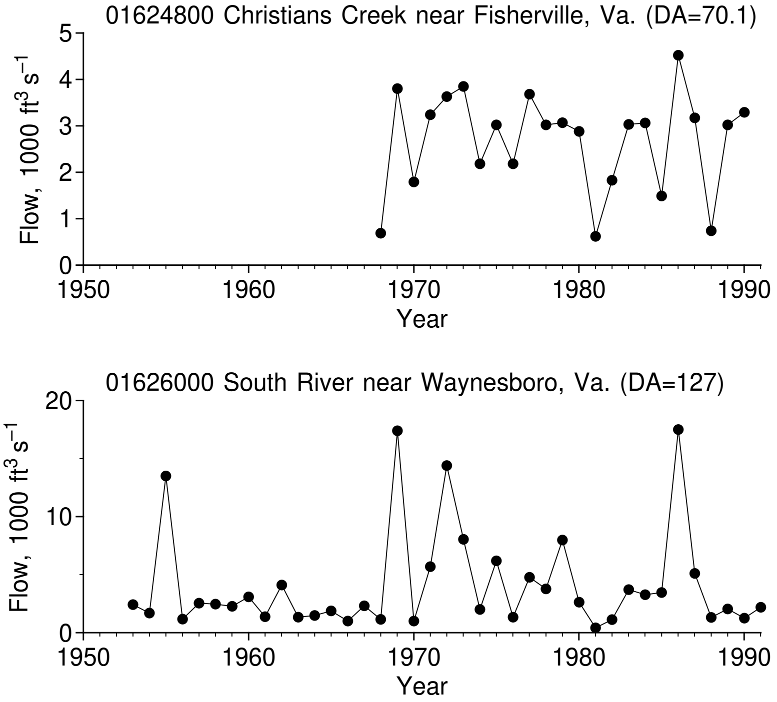

Figure 3. Annual maximum flows for site 01624800 and its nearest neighbor. DA denotes drainage area, in square miles.

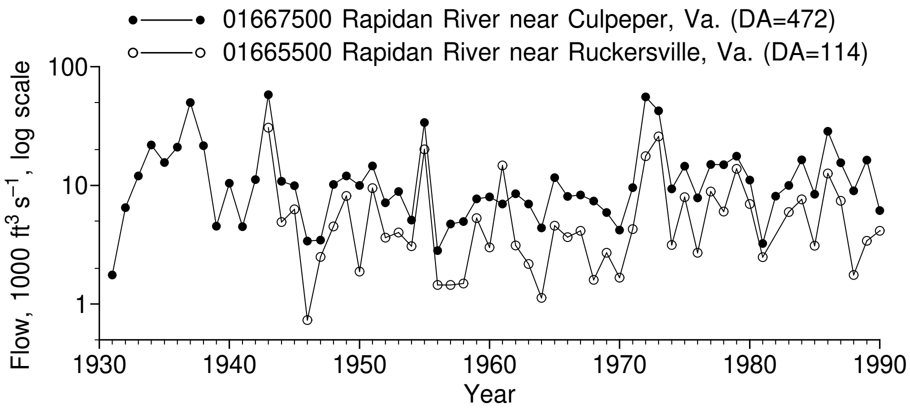

Figure 4. Annual maximum flows for two sites on the Rapidan River.

the downstream site. Although physically possible, such an occurrence is sufficiently odd that it would be worth checking the validity of the data for that year at these sites. Lacking easy access to sufficiently detailed data for such a check, we retained both sites in the data set.

4. Formation of regions

Hosking and Wallis (1993) defined a heterogeneity measure H based on the spread of sample L-CVs among the sites in a region, and suggested that regions be classified as "acceptably homogeneous" if $H < 1$, "possibly heterogeneous" if $1 \le H < 2$, and "definitely heterogeneous" if $H \ge 2$. Treating the entire set of 104 Appalachian gaging sites as a single region, the heterogeneity measure was calculated to be $H = 2.08$. The entire set of sites is therefore not far from being homogeneous, or at worst "possibly heterogeneous". Nonetheless, we emphatically reject the possibility of performing regional frequency analysis with the entire set of sites being treated as a single region. The main reason is that the theory and practice of hydrology imply that the frequency distribution is likely to depend on the drainage area of the basin. Regional frequency analysis should therefore be applied only to regions whose basins cover a fairly small range of drainage area. A further point is that in regional frequency analysis there is little to be gained by using regions containing more than about 20 sites. A reasonable starting point for regional frequency analysis would therefore be a subdivision of the set of sites, according to their drainage areas, into groups of not much more than 20.

The set of sites was accordingly divided into four groups of approximately equal size according to the sites' drainage areas. We term these groups Regions A–D. The discordancy, heterogeneity and goodness-of-fit measures defined by Hosking and Wallis (1993) were calculated for each region's data. A summary of the results is contained in Table 1. The heterogeneity measure H indicates that homogeneous regions were achieved for sites with drainage area greater than $60\,\mathrm{mi}^2$, but that the regions containing the smaller basins were heterogeneous. Other attempts to define homogeneous regions based on drainage area alone gave similar results.

It appears that, at least for the smaller basins, the frequency distribution of annual maximum streamflow is determined by more factors than drainage area alone. Accordingly, more information was introduced into the procedure for forming regions. The four available site characteristics, drainage basin area, gage elevation, gage latitude and gage longitude, were used in a cluster analysis procedure. Nonlinear transformations were applied to two of the variables: a logarithmic transformation to drainage basin area and a square root transform to gage elevation. These transformations give a more symmetric distribution of the values of the site characteristics at the 104 sites, reducing the likelihood that a few sites will have site characteristics so far from the other sites that they will always be assigned to a cluster by themselves, and, in our judgement, give a better correspondence between differences in site characteristics and the degree of hydrologic dissimilarity between different basins. All four variables

Table 1. Summary of initial regions for Appalachian streamflow data. Regions are defined in terms of drainage area alone. Units of drainage area are square miles. H is the heterogeneity statistic of Hosking and Wallis (1993). "Acceptable fit" indicates which of five distributions gave a value $|Z| \leq 1.64$ for the goodness-of-fit statistic Z defined by Hosking and Wallis (1993). The five distributions were generalized logistic (GLO), generalized extreme-value (GEV), lognormal (LN3), Pearson type III (PE3) and generalized Pareto (GPA).

Region	Range of area	Number of sites	Reg. ave. t	Reg. ave. t_3	H	Acceptable fit
A	<10	26	0.423	0.402	2.88	GLO GEV
B	10–60	28	0.450	0.472	1.70	GLO GEV
C	60–200	27	0.430	0.459	0.03	none
D	>200	23	0.387	0.419	0.29	GLO GEV

Table 2. Transformation of site characteristics.
s.d.(X) denotes the standard deviation of the 104 values of the site characteristic X.

Site characteristic, X	Cluster variable, Y
Drainage basin area (mi^2)	$Y = \log(X) \times 3/\text{s.d.}(X)$
Gage elevation (ft)	$Y = \sqrt{X}/\text{s.d.}(\sqrt{X})$
Gage latitude (deg)	$Y = X/\text{s.d.}(X)$
Gage longitude (deg)	$Y = X/\text{s.d.}(X)$

were then standardized by dividing by the standard deviation of their values at the 104 sites. Finally, the drainage basin area variable was multiplied by 3 to give it an importance in the clustering procedure equal to that of the other variables together. The transformed variables used in the cluster analysis are listed in Table 2. The transformation and weighting of the variables involves subjective decisions whose ultimate justification is the physical plausibility of the regions that are ultimately obtained from the clustering procedure.

Cluster analysis was performed using Ward's method (Ward, 1963; Mojena, 1988). This is an "agglomerative hierarchical" clustering procedure: initially each site is a cluster by itself, and clusters are then merged one by one until all sites belong to a single cluster. The assignment of sites to clusters can be determined for any number of clusters, and there is no formal measure of an "optimal" number of clusters. Choice of a suitable number of clusters is therefore subjective. The number of sites in a cluster should be large, to obtain the maximum benefit of regionalization, but the range of drainage areas of the sites in a cluster should not be too large, for otherwise it would not be reasonable to expect the clusters to be homogeneous. For the Appalachian streamflow sites we judged that seven clusters would be an appropriate number. The clusters obtained by Ward's method were adjusted using the k-means algorithm of Hartigan and Wong (1979), which yielded clusters that were a little more compact in the space of cluster variables.

Figure 5 shows the final clusters on a graph whose axes are the transformed drainage basin area and gage elevation variables. Clusters are numbered 1–7 in increasing order of average drainage basin area for the sites in the cluster. Region 1 contains three sites with very small drainage area. Regions 2, 4, 5 and 7 span the range of drainage area from 2 to 10000 mi^2, and contain mostly low-elevation sites. Regions 3 and 6 contain mostly high-elevation sites. These distinctions are not exact, because the remaining variables, latitude and longitude, also affect the clustering. Their effect can be seen in Figure 6, which shows the geographical location of the sites in each cluster. Regions 2 and 5 lie predominantly in the north-east, and Regions 3 and 6 in the west, of the study area.

This set of clusters is not completely satisfactory, because we lack data on some useful site characteristics: urbanization is one example. In an urban area much of the land is covered by an impervious layer of concrete or asphalt, and stream channels are often artificially constructed and straighter than natural channels. An urban basin's flood response to storms, and the distribution of its maximum streamflows, is therefore likely to be different from that of an otherwise comparable basin in a rural area less affected by man's control of the environment. One difference between urban and rural basins is in their specific mean annual flood, defined to be the mean annual maximum streamflow divided by the drainage area of the basin. Figure 7 is a plot of specific mean annual flood for the Appalachian basins, identifying the points that correspond to basins in the Baltimore-Washington urban corridor. Most of these small urban basins have large values of specific mean annual flood compared to other basins with similar drainage area. One other site also has a large specific mean annual flood. This site, 02032700, Schenks Branch at Charlottesville, Va., is a small basin (area 1.3 mi^2) in a fairly large city; we regard it too as an urban basin. These urban basins are mostly in Region 2, with some others in Regions 1 and 4. It may be worthwhile to form a region containing just the urbanized basins, but we lack sufficiently precise information on the extent of urbanization in different basins to be sure of this.

Further adjustments to these clusters may be desirable, but would require further study of the reliability of the data from different gaging sites, detailed knowledge of local factors that might affect the climate or hydrology of each drainage basin, and more data on other site characteristics that could be used as the basis for forming regions. For the purposes of this example, however, we are content to take the clusters obtained from Ward's method and the k-means algorithm as the regions for use in frequency analysis.

The discordancy, heterogeneity and goodness-of-fit measures of Hosking and Wallis (1993) were calculated for each region's data. Summary statistics for the regions

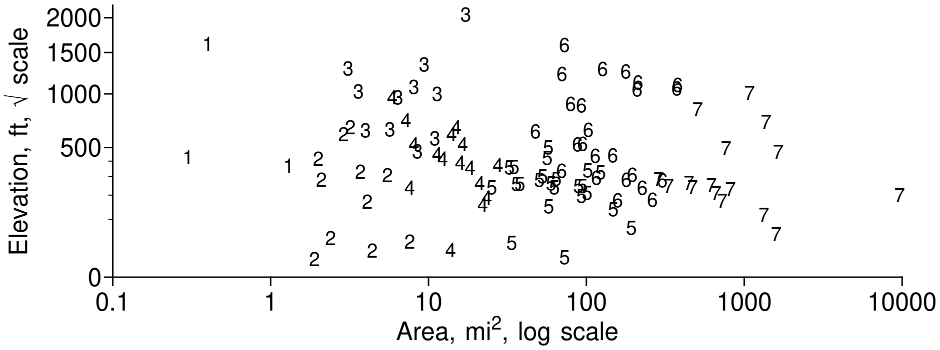

Figure 5. Drainage basin area and gage elevation for clusters of streamflow gaging stations in central Appalachia.

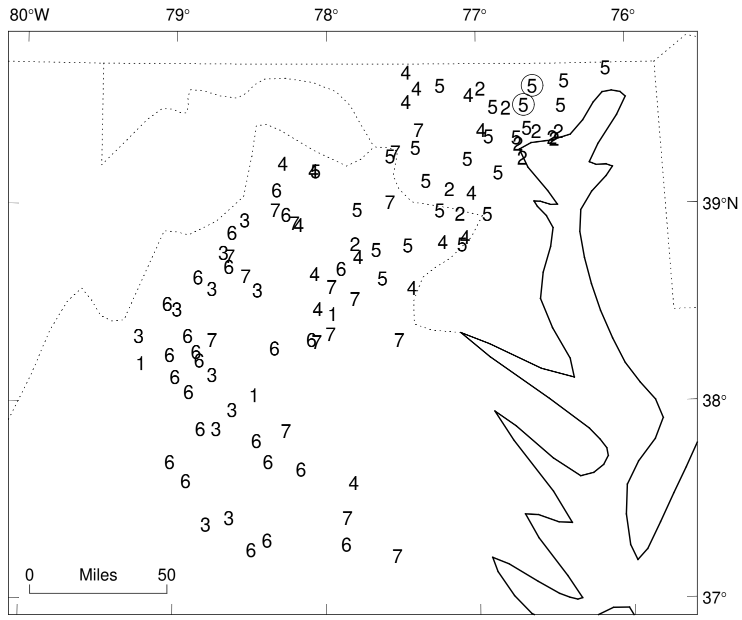

Figure 6. Location of the sites in each cluster of streamflow gaging stations in central Appalachia. The two circled sites are discussed further in the Section 7.

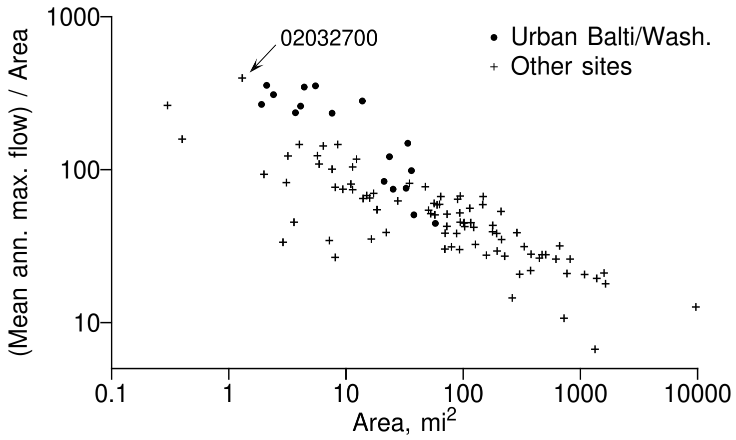

Figure 7. Specific mean annual flood for the Appalachian streamflow basins.

are contained in Table 3. The heterogeneity measure H indicates that homogeneous regions were achieved in Regions 5–7, containing sites with drainage area greater than 30 mi^2. Regions containing the smaller basins were not so successful, but should still be satisfactory for regional frequency analysis. Regions 2–4 are "possibly heterogeneous", with H values as high as 1.69, but should yield quantile estimates more accurate than those obtained from single-site frequency analysis.

Regions 6 and 7 each contain one discordant site. Region 6 contains the site with negative L-skewness, site 01624800, for which $D_i = 6.9$. As discussed previously, it is not clear whether this site should be deleted from the data set. To do so makes little difference to the results: Region 6′, defined as consisting of all sites in Region 6 except 01624800, is also homogeneous and its regional average L-moment ratios are very similar to those of Region 6. However, quantile estimates at the discordant site itself may be unreliable, and at-site estimation may be preferable there. In Region 7 site 02041500 has low L-CV, $t = 0.23$, and $D_i = 3.5$. Examination of the data shows no exceptional features, so the site was retained in the data set.

5. Choice and estimation of the frequency distribution

Figure 8 is an L-moment ratio diagram showing the regional average L-skewness and L-kurtosis of the seven final regions. Consistently with the results in Table 3, most of the points in Figure 8 lie close to the generalized extreme-value and generalized logistic lines. The lognormal distribution, used by Smith (1992), is acceptable in Regions 1–4, but is rejected for the regions containing large-area basins. If it is desired to use a single distribution for each region, the generalized logistic or generalized extreme-value, or a robust distribution such as the kappa or Wakeby, would be a reasonable choice.

Table 4 shows the quantiles of the regional frequency distributions, obtained by

Table 3. Summary of final regions for Appalachian streamflow data. Regions 1–7 are those obtained by cluster analysis. Region 6′ is Region 6 with site 01624800 excluded. Units of drainage area are square miles. H is the heterogeneity statistic of Hosking and Wallis (1993). "Acceptable fit" indicates which of five distributions gave a value $|Z| \leq 1.64$ for the goodness-of-fit statistic Z of Hosking and Wallis (1993). The five distributions were generalized logistic (GLO), generalized extreme-value (GEV), lognormal (LN3), Pearson type III (PE3) and generalized Pareto (GPA).

Region	Range of area	Number of sites	Reg. ave. t	Reg. ave. t_3	H	Acceptable fit
1	0.3– 1.3	3	0.348	0.331	–0.40	GLO GEV LN3 GPA PE3
2	1.9– 7.6	11	0.379	0.363	1.65	GLO GEV LN3
3	3.1– 17	11	0.498	0.471	1.62	GLO GEV LN3
4	5.9– 28	16	0.450	0.430	1.69	GLO GEV LN3 GPA
5	25– 190	23	0.424	0.491	0.24	none
6	47– 380	24	0.438	0.442	–0.03	GLO GEV
6′	47– 380	23	0.442	0.455	–0.71	GLO GEV
7	290–9600	16	0.376	0.420	0.87	GLO GEV

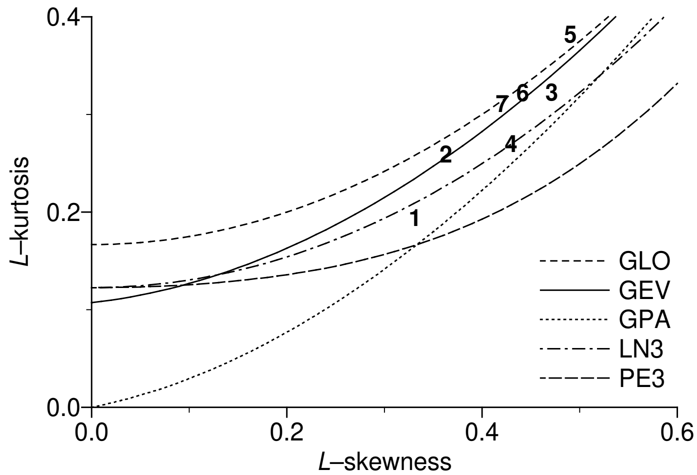

Figure 8. Regional average L-moments for the Appalachian streamflow regions. Curves are the L-skewness–L-kurtosis relations for five distributions: generalized logistic (GLO), generalized extreme-value (GEV), lognormal (LN3), Pearson type III (PE3) and generalized Pareto (GPA).

Table 4. Estimated quantiles for Appalachian streamflow regions. Fitted distribution: generalized extreme-value.

	Nonexceedance probability									
Region	0.01	0.10	0.20	0.50	0.80	0.90	0.98	0.99	0.998	0.999
1	0.17	0.37	0.49	0.81	1.35	1.80	3.12	3.85	6.09	7.34
2	0.14	0.34	0.45	0.78	1.35	1.85	3.39	4.29	7.18	8.88
3	0.03	0.20	0.31	0.64	1.32	1.99	4.46	6.15	12.60	17.04
4	0.07	0.25	0.37	0.70	1.34	1.95	4.03	5.38	10.18	13.30
5	0.20	0.33	0.42	0.69	1.25	1.82	3.98	5.50	11.47	15.69
6	0.11	0.28	0.39	0.70	1.32	1.91	3.98	5.34	10.30	13.57
7	0.22	0.37	0.47	0.75	1.30	1.80	3.52	4.61	8.45	10.91

fitting generalized extreme-value distributions to each region's data using the method of regional L-moments. The quantiles for different regions are quite similar in the main body of the distribution; only for nonexceedance probabilities 0.98 and greater do marked differences become apparent. Regions 1 and 2 have relatively low upper-tail quantiles. Otherwise the upper-tail quantiles generally increase as the drainage area of the basin decreases, though Region 4 is an exception to this pattern.

6. Accuracy of estimation

Error bounds for the estimated regional growth curves were obtained by Monte Carlo simulation. In the simulations, data were generated for artificial regions that were homogeneous with generalized extreme-value frequency distributions at each site. The artificial regions had the same number of sites, the same record length at each site, and the same regional average L-moments, as the actual data. There is no evidence of inter-site dependence in Region 1, but inter-site dependence is present in the data for the other regions, the average correlation between sites varying from 0.30 in Region 2 to 0.67 in Region 5. The simulations for Regions 2–7 therefore included inter-site dependence, using the algorithm described by Hosking and Wallis (1988).

Error bounds were computed from the empirical quantiles of the distribution of the estimated regional growth curve. These can be obtained by calculating the ratio of the estimated regional growth curve to its true value, $\hat{q}(F)/q(F)$, averaging these ratios over the sites in the region and accumulating over the different realizations of the simulation procedure a histogram of the values taken by the ratio. For example, for a particular nonexceedance probability F it may be found that 5% of the simulated values of $\hat{q}(F)/q(F)$ lie below some value $\ell_{.05}(F)$ while 5% lie above some value $u_{.05}(F)$. Then 90% of the distribution of $\hat{q}(F)/q(F)$ lies within the interval

$$\ell_{.05}(F) \le \frac{\hat{q}(F)}{q(F)} \le u_{.05}(F)\,, \tag{1}$$

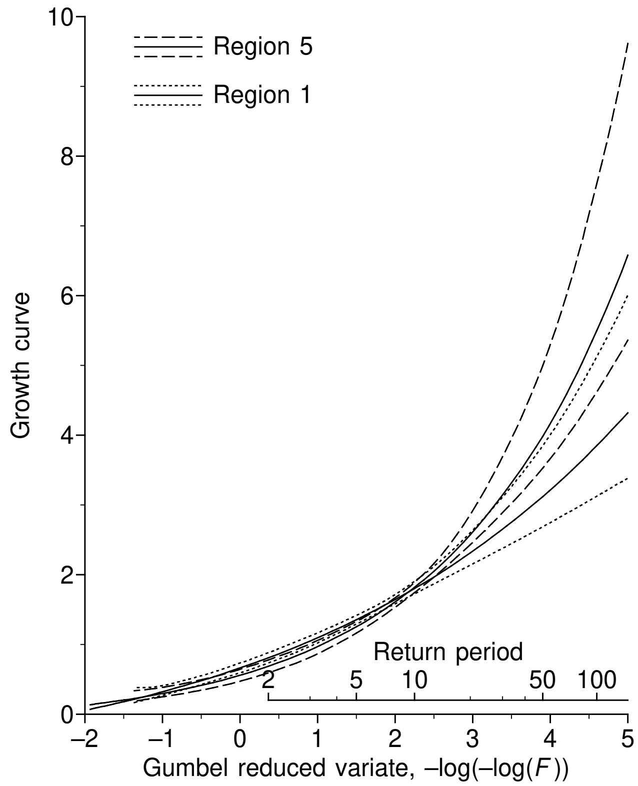

Figure 9. Estimated regional growth curves, with their 90% error bounds, for Appalachian streamflow regions 1 and 5.

and inverting this to express $q(F)$ in terms of $\hat{q}(F)$ gives

$$\frac{\hat{q}(F)}{u_{.05}(F)} \le q(F) \le \frac{\hat{q}(F)}{\ell_{.05}(F)} . \tag{2}$$

Expression (2) has the same form as a statistical confidence interval, but can be validly interpreted as one only if the distribution of $\hat{q}(F)/q(F)$ is independent of all of the parameters involved in the specification of the statistical model underlying the index-flood procedure; for the regional L-moment algorithm these parameters are the at-site means and the regional average L-moment ratios. In practice independence does not hold and confidence statements are at best approximate. Nonetheless, the interval (2) should give a useful indication of the amount of variation between true and estimated quantities. We refer to the limits $\hat{q}(F)/u_{.05}(F)$ and $\hat{q}(F)/\ell_{.05}(F)$ in (2) as the “90% error bounds” for $\hat{q}(F)$.

Results for Regions 1 and 5, which have the lowest and highest L-skewness of the six regions, are shown in Figure 9. Error bounds are given only for exceedance probabilities of 0.02 or greater: for extreme low quantiles the simulated estimates were sometimes negative and led to unreliable error bounds.

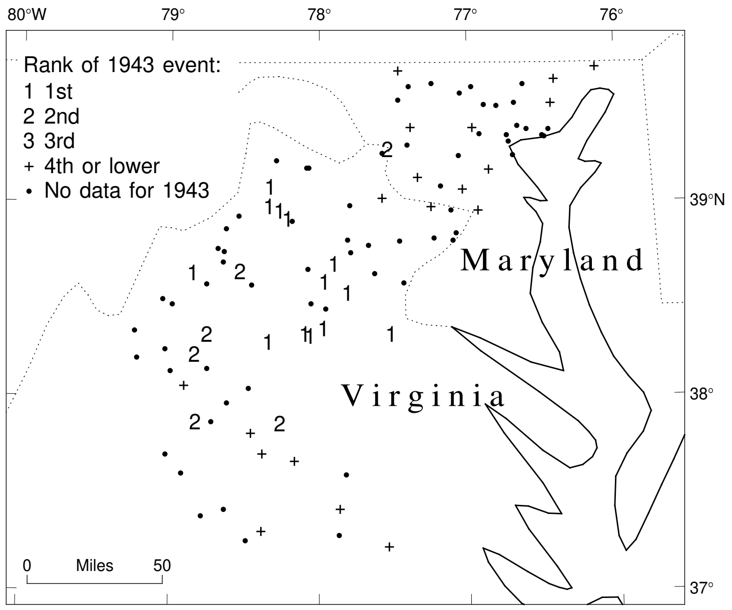

Figure 10. Rank of the maximum streamflow of water year 1943.

Although the regional growth curves for Regions 1 and 5 are fairly well separated, for other regions there is considerable overlap between the error bounds of the regional growth curves. This suggests that on statistical grounds there is little justification for treating the regions as distinct, and that they might as well be merged to form larger regions. This argument should be treated with caution, because the absence of statistically significant differences between regional growth curves may merely reflect an insufficiency of data. When the differences between the regional growth curves are consistent with the physical reasoning that led to the definition of the regions, there are grounds for belief that these differences are scientifically significant even though they may not be statistically significant.

7. A comparison of regional and at-site estimation

Central Appalachia is subject to occasional hurricanes and other large storms. Such events often give rise to high outliers in the data and large values of the sample L-CV and L-skewness, as noted in Section 3. Whether a high outlier is recorded is of course dependent on whether the streamgage was operating at the time of the flood event. Single-site frequency analysis can therefore give quantile estimates that are highly dependent on which large flood events happened to occur during the period of gaging.

As an example we consider the storm event of 15–16 October 1942 (we call it the "1943 event", since it occurred in water year 1943). The rank of this event in the data record of the 104 sites is shown in Figure 10. In much of northern Virginia the 1943

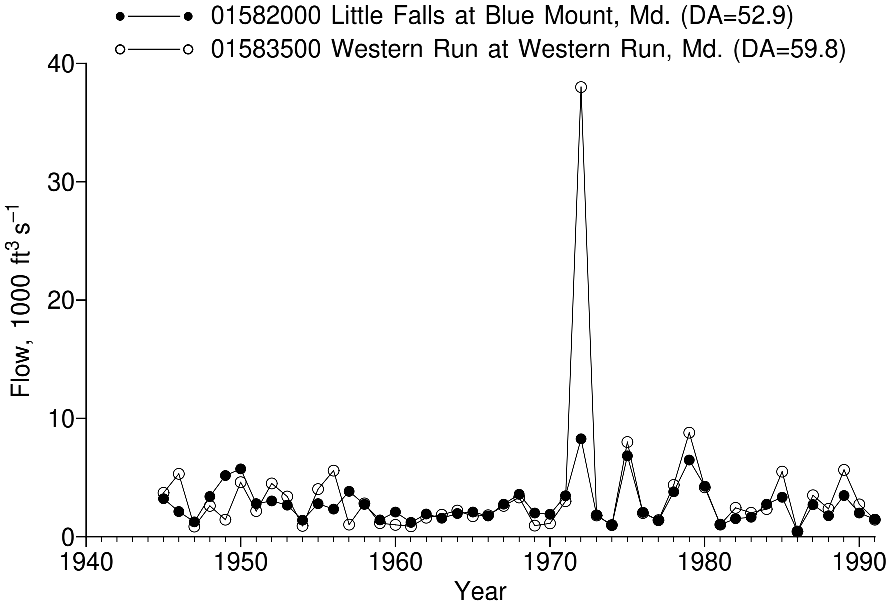

Figure 11. Annual maximum streamflow data for two sites in central Appalachian streamflow Region 5.

event was the largest flood on record, but many sites in this area were not active in 1943 and have no data for that year. At sites with no data for 1943 it seems plausible that the 1943 event, had it been recorded, would also have been among the largest on record; at these sites, therefore, single-site analysis is likely to underestimate the frequency of large events and give biased estimates of extreme quantiles of the flood frequency distribution. Regional frequency analysis, for a region that includes some sites that recorded the 1943 event and some that did not, is likely to give a less biased estimate of the frequency of flood events as extreme as that of 1943.

A reasoning behind regional frequency analysis is further illustrated by considering the spatial variability of storm events. When two sites are affected by the same storm, the streamflow resulting from the storm may be very different at the two sites although there are no obvious hydrological differences between the sites. As an example we give comparative frequency analyses for two neighboring sites: site 01582000, Little Falls at Blue Mount, Md., and site 01583500, Western Run at Western Run, Md. The sites are circled on Figure 6. The gaging sites are 7 mi apart and have similar elevations, 305 ft and 262 ft respectively; the basins also have similar drainage areas, 52.9 mi^2 and 59.8 mi^2 respectively. Both sites have gaged records from 1945 to 1991. Annual maximum streamflow data for the two sites are shown in Figure 11. The data for the two sites are generally of comparable magnitude except in 1972. At both sites in that year the annual peak was recorded on 22 June 1972, coincident with Hurricane Agnes, but the recorded peak flows, 8280 ft^{3}s^{-1} at Blue Mount and 38000 ft^{3}s^{-1} at Western Run, differ by a factor of 4.6. For the complete data samples, the L-CV is

Table 5. Estimated quantiles of annual maximum streamflow, in $ft^3 s^{-1}$, for Blue Mount and Western Run.

		Nonexceedance probability				
Method	Site	0.9	0.98	0.99	0.998	0.999
At-site	Blue Mount	4660	7570	9110	13600	16000
	Western Run	6430	15800	22990	54100	78000
Bulletin 17	Blue Mount	4680	7680	9270	13800	16300
	Western Run	6740	15800	21900	45100	60600
Regional	Blue Mount	4960	10900	15000	31300	42800
	Western Run	6390	14000	19300	40300	55100

$t = 0.30$ at Blue Mount, $t = 0.49$ at Western Run. Most of the difference is due to the 1972 data value: without it, the sample L-CVs would be $t = 0.28$ at Blue Mount, $t = 0.36$ at Western Run.

Table 5 gives the results of frequency analyses for the two sites. In the at-site analyses a generalized extreme-value distribution was fitted at each site using the method of L-moments. The estimated quantiles in the upper tail of the frequency distribution differ by factors of 1.4, 2.5 and 4.9 at nonexceedance probabilities $F = 0.9$, $F = 0.99$ and $F = 0.999$, corresponding to return periods of 10, 100 and 1000 years. The Bulletin-17 calculations were performed by the HEC-FFA program (U.S. Army Corps of Engineers, 1992), using a generalized skew coefficient of 0.6, this value being inferred from the map accompanying Bulletin 17B (U.S. Water Resources Council, 1981). The Bulletin-17 results are very similar to the at-site analysis based on L-moments, except that the estimates for Western Run at $F = 0.998$ and $F = 0.999$ are somewhat lower. The regional analysis used the regional growth curve of Region 5 in Table 4, to which both sites belong, and estimated the index flood at each site by the sample mean. Since the ratio of the sample means at the two sites is 1.29, the quantile estimates differ by this factor at all return periods.

These quantile estimates raise a critical question for frequency estimation of extreme events. Is the greater flood response of the Western Run basin a systemic property that can be expected to recur whenever future large storms pass over the study area, or are the different responses of the two basins a result of unpredictable local variations in precipitation intensity and duration, with future storms being as likely to cause greater runoff in one basin as in the other? At-site analysis is appropriate if the former explanation is correct, though even so it will be accurate only if the frequency of extreme events in the observed record matches the long-term rate of occurrence of such events. We believe, however, that differences so large as that illustrated in Figure 11 can rarely be ascribed with confidence to physical differences between such apparently similar basins. Regional frequency analysis is then the appropriate procedure. By using all available data from a set of physically similar sites, it will give a better overall estimate of the frequency of extreme events, thereby improving the accuracy of frequency estimation at every site.

8. Summary

This analysis of streamflow data does not pretend to be authoritative: if more data on site characteristics were available there would be a better prospect of identifying local variations in the climate or flood response of the basins in the study area. However, it illustrates some features that commonly occur in regional frequency analysis of environmental data.

When one site characteristic is known to have a strong effect on the frequency distribution, it makes sense to base the formation of regions on that site characteristic. Drainage area is such a characteristic for streamflow data, though in this example it became clear that drainage area alone was unable to explain the between-site variations in the frequency distribution of annual maximum streamflow. Cluster analysis closely tied to physical reasoning then led to the identification of other sources of variation between at-site frequency distributions. This approach should be widely applicable, though the patterns of dependence of the frequency distribution on site characteristics may be quite different in other parts of the world.

The main conclusion is that successful regional frequency analysis involves interaction between physical reasoning and statistical assessment of the regions and their estimated frequency distributions. Close study of the physical reasons for variation between sites is essential to the successful use of regional frequency analysis, since it enables reasonable regions to be obtained even when small samples or irregularities in the data make it difficult to define convincing regions by statistical methods alone.

Acknowledgement

We are grateful to J. A. Smith for sending us the data that he used in Smith (1992).

References

Hartigan, J. A. & Wong, M. A. (1979). Algorithm AS 136: A K-means clustering algorithm. *Applied Statistics*, **28**, 100–108.

Hosking, J. R. M. (1990). L-moments: analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society, Series B*, **52**, 105–124.

Hosking, J. R. M. & Wallis, J. R. (1988). The effect of inter-site dependence on regional flood frequency analysis. *Water Resources Research*, **24**, 588–600.

Hosking, J. R. M. & Wallis, J. R. (1993). Some statistics useful in regional frequency analysis. *Water Resources Research*, **29**, 271–281. Correction: *Water Resources Research*, **31** (1995), 251.

Hosking, J. R. M. & Wallis, J. R. (1996). *Regional frequency analysis: an approach based on L-moments*. To be published by Cambridge University Press.

Hydrosphere (1993). *Hydrodata CD-ROMs, vol. 4.0: USGS Peak Values.* Boulder, Colo.: Hydrosphere Data Products.

Mojena, R. (1988). Ward's clustering algorithm. In *Encyclopedia of Statistical Sciences, vol. 9*, edited by S. Kotz, N. L. Johnson and C. B. Read, pp. 529–532. New York: Wiley.

Smith, J. A. (1992). Representation of basin scale in flood peak distributions. *Water Resources Research*, **28**, 2993–2999.

U.S. Army Corps of Engineers (1992). HEC-FFA flood frequency analysis user's manual. *Report CPD-13*, Hydrologic Engineering Center, Davis, Calif.

U.S. Water Resources Council (1981). Guidelines for determining flood flow frequency. *Bulletin 17B*, Hydrology Committee, Washington, D.C.

Ward, J. H. (1963). Hierarchical groupings to optimize an objective function. *Journal of the American Statistical Association*, **58**, 236–244.