Hydrological Sciences - Journal - des Sciences Hydrologiques, 35, 2, 4/1990

An appraisal of the "region of influence" approach to flood frequency analysis

DONALD H. BURN

Department of Civil Engineering, 342 Engineering Building,
University of Manitoba, Winnipeg, Manitoba, Canada R3T 2N2

Abstract Regional flood frequency entails the pooling of data from sites within a defined region to enhance the estimation of at-site quantiles. Conventional regionalization techniques normally identify a fixed set of stations forming a contiguous region. An approach to regional flood frequency analysis that involves each site having a potentially different set of stations included for the at-site estimation of extremes was compared with a more traditional regionalization technique. The characteristics of the stations identified as being of relevance for the purposes of at-site estimation using the two approaches were contrasted and also the extreme flow values obtained were compared. The results indicated that the region of influence approach results in a group of stations with greater homogeneity than was the case for the regionalization technique and also leads to extreme flow estimates which are more accurate.

Une évaluation de la méthode "d'influence régionale" pour l'analyse des fréquences de crues

Résumé L'analyse régionale de fréquence des crues nécessite le regroupement de données provenant de sites pris dans une région définie pour améliorer l'estimation des quantiles pour le site intérêssé. Normalement, les techniques conventionelles de régionalisation identifient un nombre fixe de stations formant une région définie. Une méthode d'analyse régionale de fréquence des crues qui implique chaque site ayant potentiellement un nombre différent de stations utilisées pour l'estimation des extrémes pour le site intérêssé a été comparée à une technique de régionalisation plus traditionelle. Les caractéristiques des stations identifiées comme étant valables pour l'estimation au site intérêssé en utilisant les deux méthodes ont été contrastées, et également les valeurs extrêmes de débit obtenues on été comparées. Les résultats indiquent que la méthode d'influence régionale donne un groupe de stations plus homogène que dans le cas de la technique de régionalisation, et abouti aussi à des éstimations de débits extrêmes plus précises.

INTRODUCTION

The problem of determining the probability that a selected flow value will be

Open for discussion until 1 October 1990

exceeded at a given location on a river is a topic of fundamental interest to engineers in the water resources field. The accurate estimation of the relationship between extreme flow events and the associated recurrence interval (the so-called *Q-T* relationship) is therefore an issue that has received considerable attention in the literature.

A common difficulty that has been experienced in the estimation of the *Q-T* relationship arises from the lack of a sufficient quantity of data to characterize properly the functional relationship. This difficulty is especially acute for the long return intervals that are of primary concern in the design of many hydraulic structures. In an attempt to compensate for an insufficient temporal characterization of the extreme flow behaviour, regional flood frequency analysis has been used as a means of substituting spatial data for temporal data. Regional flood frequency can be used to enhance the at-site estimation of extreme flow events at gauged sites, or to predict extreme flow probabilities for ungauged sites, using a regional growth curve. The emphasis in this paper is on the application to gauged sites.

A need to identify regions that are homogeneous with respect to pertinent basin parameters arises within the regional flood frequency procedure. The importance of regional homogeneity has been addressed by several researchers (Greiss & Wood, 1981; Lettenmaier *et al.*, 1987; Wiltshire, 1986; Burn, 1988). Although the homogeneity of a region can be increased by decreasing the number of stations included in the region, doing so means that the amount of information used in the extreme flow estimation is reduced. This trade-off between quality information and an increased quantity of data has been discussed by Burn (1988).

The present paper describes an alternative methodology for effecting the information transfer from surrounding stations for the at-site estimation of extreme flows. The methodology employed (which appears to have been first suggested by Acreman & Wiltshire (1987) and Acreman (1987)) is referred to herein as the region of influence (ROI) approach. The premise of the technique is that each site should be allowed to have a unique set of stations which constitutes the "region" for the site. Thus there is no need for boundaries between regions nor is there a need for all sites in a particular area to use the same number of stations in the estimation of at-site extreme flows.

METHODOLOGY

The foundation of the proposed technique is the identification of a region of influence for each gauging station consisting of the set of gauged sites that are in close proximity to the candidate station. Proximity is measured by the Euclidean distance in a p-dimensional attribute space where the attributes are measures pertinent for the identification of stations with a similar extreme flow response. The distance metric used is thus defined as:

$$D_{jk} = \left[\Sigma_{i=1}^{p} \left(C_j^i - C_k^i \right)^2 \right]^{1/2} \tag{1}$$

where D_{jk} is the Euclidean distance from site j to site k, p is the number of attributes included in the distance measure (i.e. the dimensionality of the problem), and C_j^i is a standardized value for the measure of attribute i for site j. The standardization of the attributes involves dividing the raw data by the standard deviation of data calculated for attribute values from a total of *NS* stations. The standardization process eliminates the units from each attribute and reduces any differences in the range of values amongst the attributes. This procedure is invoked to avoid the introduction of bias due to scaling differences for the attributes. The distance value from equation (1) will thus provide a measure of how close each station is to every other station (i.e., a symmetric *NS* by *NS* matrix of distance measures results).

The determination of a set of appropriate attributes to include in the distance measure is predicated on the data available for the network of sites. Although the choice of attributes requires engineering judgement, guidance for the selection of relevant attributes can be obtained from examining the correlation between potential attributes and measures of the at-site extreme flow. Attributes may be derived from extreme flow data (e.g. coefficient of variation, skewness) or may consist of measures of physical features of the basin (e.g. drainage area, soil type).

The next step in the process of identifying the region of influence involves selecting a threshold value that will function as a cut-off point for the distance measure. All stations that are a distance greater than the threshold value from the candidate site will be excluded from the region of influence for the site. The choice of a threshold value is somewhat analogous to the selection of the number of regions to divide a network of gauging stations into using conventional regionalization techniques. Larger threshold values will increase the number of sites included in the ROI, but the homogeneity of the set of stations can be expected to decrease. Conversely, a smaller threshold will result in an increase in the homogeneity of the stations included, but the information transfer will be decreased due to the smaller number of stations. The threshold value can be adjusted until an appropriate compromise is reached. A useful criterion for the selection of a threshold is the correlation between the candidate site and sites at or near the threshold value. If a target correlation is specified, the threshold can be chosen to minimize deviations between the at-threshold correlations and the target value for sites included in the ROI.

Since all of the stations included in the region of influence will not be equally close to the site for which the ROI is being determined, a weighting function is required to reflect the relative importance of each station in the estimation of the at-site extreme flows. The weighting function used was of the form:

$$WF_{jk} = 1.0 - \left[\frac{D_{jk}}{THL}\right]^n \qquad (2)$$

where WF_{jk} is the weighting for station k in the ROI for site j, *THL* is a parameter, and n is a positive constant. The effect of the parameter *THL* is to dictate the value of the weighting function for stations at the threshold. For this reason, the value of *THL* should logically be greater than or equal to the threshold value. If *THL* is equal to the threshold, then stations at the threshold will have no contribution to the determination of at-site extremes; larger values will increase the weighting of all stations included in the ROI. The value of the constant n will determine the rate of decrease of the weighting values as stations further away from the site (in terms of the distance measure) are considered. Using the procedure outlined above, the stations which constitute the region of influence for each site may be determined and the relative importance of each member of the ROI in estimating at-site extreme flow values may be ascertained.

With the determination of a region of influence for each site, it is possible to estimate at-site extreme flow values incorporating information from all stations that are members of the ROI. Several options (in the form of alternate estimators) exist for combining all of the available information with different flood frequency distributions. In the present work, two distributions were considered, namely the generalized extreme value (GEV) and the log-Pearson type III (LP3) distribution. These distributions were selected as representative distribution functions that have been found to provide satisfactory fit to extreme flows in situations where regional information is available. For each of the distribution functions, it would be possible to consider several estimators for the unknown parameter values. However, for conciseness, only one estimator per distribution will be presented and discussed herein. It is not anticipated that the conclusions reached in this work will be a function of either the estimator or the distribution employed. It would, however, certainly be possible to utilize other estimators or distribution functions within the context of the methodology described.

The GEV distribution function is defined as:

$$F(x) = \exp\left[-\left\{1 - g(x-\xi)/\alpha\right\}^{1/g}\right] \qquad \text{for } g \neq 0 \qquad (3a)$$

and

$$F(x) = \exp\left[-\exp\left\{-(x-\xi)/\alpha\right\}\right] \qquad \text{for } g = 0 \qquad (3b)$$

The three parameters can be estimated from three probability weighted moments (PWMs) which may be obtained as (after Hosking *et al.*, 1985):

$$M_r = \frac{1}{np}\,\Sigma_{i=1}^{np} p_i^r x_i \qquad r = 0,\ 1,\ 2 \qquad (4)$$

where $p_i = (i - 0.35)/np$ is the plotting position for the data point x_i and np is the number of data points. PWMs are calculated for each station and scaled values obtained through:

$$t_{1,k} = M_1^k/M_0^k \qquad k = 1, 2, ..., NS \qquad (5a)$$

$$t_{2,k} = M_2^k/M_0^k \qquad k = 1, 2, ..., NS \qquad (5b)$$

where the index k is used to denote the station number. PWMs for ROI are then calculated from the PWMs for the stations in the ROI as:

$$T_i^j = \Sigma_{k\in I_j} t_{i,k}\, np_k WF_{jk} / \Sigma_{k\in I_j} np_k WF_{jk} \qquad i = 1, 2 \qquad (6)$$

where I_j is the set of all stations in the ROI for site j, and np_k is the number of data points (years of record) at station k. The index j on the regional PWM indicates the site for which the PWM is calculated. It can be seen from the form of equation (6) that the PWMs from the individual stations are weighted through the weighting function value, WF_{jk}, reflecting the closeness of the station to the site, and also are weighted according to the number of years of record at the station, np_k.

The three parameters of the GEV distribution may be estimated from:

$$c = \left(2T_1^j - 1\right) / \left(3T_2^j - 1\right) - \log(2)/\log(3) \qquad (7)$$

$$g = 7.8590c + 2.9554c^2 \qquad (8)$$

$$\alpha = M_0^j \left(2T_1^j - 1\right) g \Big/ \left\{\Gamma(1 + g) \left(1 - 2^{-g}\right)\right\} \qquad (9)$$

$$\xi = M_0^j + \alpha \left\{\Gamma(1 + g) - 1\right\} / g \qquad (10)$$

where it is implied that each parameter has an index j associated with it to indicate the appropriate site. The parameter estimator utilized above incorporates "regional" information for the estimation of the parameters g and α, and then uses these values and the at-site mean to calculate the value for the remaining parameter. This estimator is analogous to the GEV-1 estimator described by Lettenmaier *et al.* (1987). With the parameters calculated from the above equations, it is possible to estimate the at-site extreme flow value for any selected return interval, T, through:

$$x_T = \xi + \alpha/g \left[1 - \left\{- \log \left[1 - \frac{1}{T}\right]\right\}^g\right] \qquad (11)$$

where x_T is the estimate for the T-year flow value.

For the LP3 distribution, the three parameters of the distribution were estimated using the method of moments wherein information from the sites within the region of influence was included through the calculation of a generalized skew coefficient, $\bar{G}_j$, given as:

$$\bar{G}_j = \Sigma_{k \in I_j} \gamma_k \cdot WF_{jk} / \Sigma_{k \in I_j} WF_{jk} \tag{12}$$

where γ_k is the skew coefficient for station k, and all other symbols are as previously defined. A weighted skewness value for site j can then be calculated using (Thomas, 1985):

$$\hat{G}_j = \begin{cases} \bar{G}_j & \text{if } np_j \leqslant 25 \\ \left[\dfrac{np_j - 25}{75}\right] \gamma_j + \left[1 - \dfrac{np_j - 25}{75}\right] \bar{G}_j & \text{if } 25 < np_j \leqslant 100 \\ \gamma_j & \text{if } np_j > 100 \end{cases} \tag{13}$$

The three parameters of the distribution are calculated using the value of $\hat{G}_j$ and the first two at-site moments via standard procedures (Kite, 1977). The at-site extreme flow value for any return period, T, is estimated from:

$$y_T = \log x_T = \mu_y + K\sigma_y \tag{14}$$

where μ_y and σ_y are the mean and standard deviation of the logarithms of x and K is the Pearson frequency coefficient which is a function of the skewness and the return period.

In the example presented below, the ROI technique is compared with results from delineating regions using the procedure described by Burn (1988). The latter procedure, which will be referred to as the regionalization approach, is based on principal components analysis. The regionalization approach involves calculating the correlation matrix of the annual flow data from a common period for the gauging stations. Principal components are determined from the correlation matrix and the number of regions chosen is defined by the number of principal components required to explain adequately the total variance. The principal components are subsequently rotated to obtain a more equal distribution of the variation explained by each principal component. Each station is then assigned to a group corresponding to the rotated principal component with which the station has the largest correlation. Further details on the procedure are presented in Burn (1988).

ILLUSTRATIVE EXAMPLE

The region of influence approach outlined above was applied to a set of 91 streamflow gauging stations located in southern Manitoba, Canada. The data set comprised the annual extreme flows for all rivers in southern Manitoba with at least 10 years of record. The average number of years of record for the sites in the network was 28 with a range from 10 to 78 years. The mean drainage area for the stations included in the analysis was approximately 7300 km^2 with a median value of 572 km^2.

The ROI technique was contrasted with the regionalization approach on two bases. The first involved a comparison of the characteristics of the stations included in the at-site estimation of extreme flow values. It is desirable that the stations included have a high degree of similarity so that there is a transfer of comparable information. The second basis for comparison was the extreme flow values calculated for each site using both of the approaches with the two distribution function assumptions.

To apply the ROI approach, it was necessary to establish a set of station attributes to form the basis for the distance measure. In order to ensure that values for the attributes could be estimated for each of the stations in the data set, it was essential that the attributes be restricted to characteristics that are readily obtainable for a network of stations of the size and diversity examined herein. As a result, the attributes chosen included two measures of a statistical nature and two location (spatial) measures. The statistical attributes consisted of the coefficient of variation (*CV*) and the ratio of mean annual flow to drainage area (*QDA*) calculated from the annual extreme flow values at each site. Similarity in the values for these two parameters would indicate that the stations have similar *Q-T* relationships such that the parent distribution function for the stations is likely to have a comparable form. This likeness could result from a variety of mechanisms including correspondence in the physical characteristics of the contributing drainage area such as the slope, soil type, and vegetative cover.

The two location attributes used were measures of the longitude and latitude of each station relative to a reference location. The spatial attributes thus reflected the physical proximity of station pairs. It is to be expected that sites which are close together could exhibit similar extreme flow responses due to similarities in the causative precipitation events that act as input to the flow generation process.

The remaining parameters that must be selected for the region of influence approach are the threshold value for the distance measure, and values for the two parameters of the weighting function, *THL* and *n*. The threshold value will affect the number of stations included in the region of influence for each site. With larger values for the threshold, more stations will be included, but the similarity of the ensemble of stations will necessarily decrease. The choice of a value for the threshold is thus a decision that must be reached by weighting the trade-offs between an increased quantity versus a decreased quality of information. After a perusal of the distance measure matrix, and considering the above factors, the threshold value was set equal to 1.8. The estimation of at-site extremes will not likely be sensitive to the choice of the threshold value due to the nature of the weighting function. Since sites at or near the threshold receive a relatively small weighting in the extreme flow estimation procedure, the impact of adding or removing stations that are near the threshold will be small.

Judgement is also required in the selection of values for the weighting function parameters. The values chosen will implicitly specify the rate of decrease of the weighting function with an increase in the distance measure from a station to the reference site. After giving consideration to the value selected for the distance threshold, the values of *THL* and *n* were chosen to be 1.85 and 4 respectively. The particular values selected for the two parameters of the weighting function are not expected to unduly affect the estimation of at-site extremes providing that logical values are selected so that the prescribed weighting function shape is maintained.

RESULTS

The regionalization approach resulted in the division of the 91 stations into four regions, as shown in Fig. 1. The number of regions selected resulted from considering the trade-off between the increased quantity of information in each region resulting from a small number of larger regions and the increase in homogeneity of the regions that comes from a large number of smaller regions. Further details on the regionalization process invoked are contained in Burn (1988) to which interested readers are referred.

The statistical characteristics of the stations used to estimate an at-site extreme flow value for the regionalization option and the ROI approach are summarized in Table 1 for several locations. The locations included in Table 1 correspond to selected sites at which the number of stations in the site's region and the number of stations in the region of influence for the site do not differ by more than two stations. As such, the quantity of data (in terms of number of stations) should not be an issue and the stations used for the two methods can be compared with respect to the values for the statistical attributes. Similarity in the number of stations included in a site's ROI and region implies that the site is likely to be representative of regional characteristics. The sites presented in Table 1 are thus expected to be "average" stations in terms of attribute values. The ROI and regionalization approaches should result in similar at-site extreme estimates for sites of this nature. A comparison of regional and ROI attributes for unusual sites will be presented later in the paper.

Table 1 presents the average and the standard deviation of the *CV* and *QDA* for the stations included in each of the regions. For every site, the *CV* and *QDA* for the site and for the site's ROI are displayed along with the standard deviation for the parameters calculated from the stations in the ROI. Two observations from the results are:

(a) the average for the statistical attributes for the ROI stations were invariable closer to the at-site values than was the case for the regional stations; and

(b) the variability of the statistical attributes (as measured by the standard deviation) was less for the ROI stations than for the regional stations in every case but one.

While the above results considered sites where the number of stations included for the two approaches was similar, a congruence in the number of

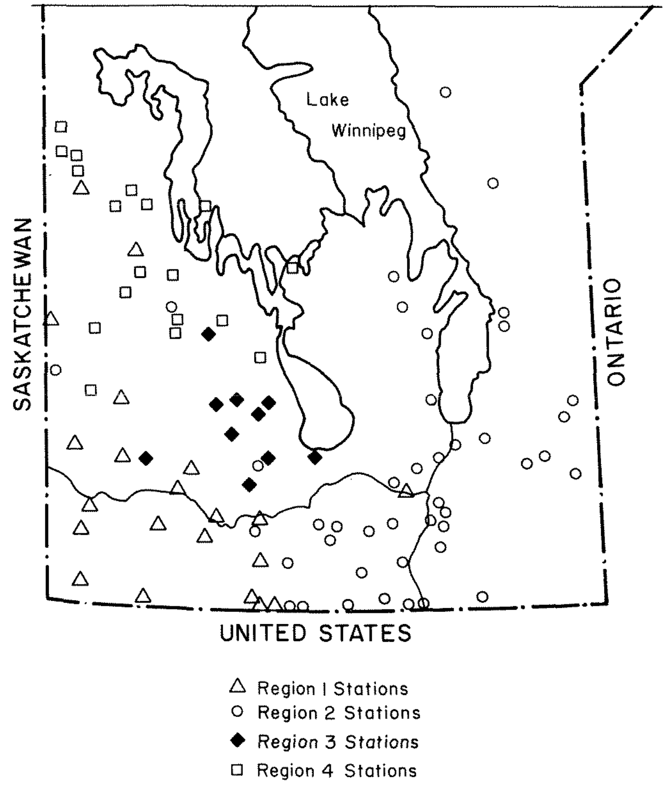

Fig. 1 Station locations for the groups defined using the regionalization procedure.

***Table 1** Comparison of regional and ROI characteristics*

Region	*Regional*		*Site*			*ROI*	
	CV	*QDA*		*CV*	*QDA*	*CV*	*QDA*
1	*1.11*	*0.024*	*55*	*1.23*	*0.010*	*1.16*	*0.020*
	(0.24)	*(0.017)*				*(0.12)*	*(0.014)*
2	*0.84*	*0.044*	*7*	*0.91*	*0.041*	*0.92*	*0.041*
	(0.23)	*(0.030)*				*(0.13)*	*(0.020)*
3	*0.89*	*0.051*	*21*	*0.97*	*0.113*	*0.92*	*0.091*
	(0.09)	*(0.035)*				*(0.15)*	*(0.022)*
4	*0.81*	*0.050*	*59*	*0.75*	*0.031*	*0.81*	*0.036*
	(0.23)	*(0.036)*				*(0.17)*	*(0.019)*

Note: terms in parentheses indicate the parameter standard deviation.

stations is not a requirement and indeed this illustrates one of the advantages of the ROI approach. For the regionalization technique, the same number of stations is used in the estimation of extremes at each site in the region. However, the same set of sites will have a range of values for the number of stations included in the at-site analysis when the ROI approach is used. This phenomenon is illustrated in Table 2(a) which presents the number of stations in each region as well as the mean and the range for the number of stations for the same group of sites under the ROI approach. The range of values for individual sites using the ROI approach reflects the fact that some sites are more similar to the surrounding stations than are other sites. Thus a site that is in essence an outlier (in a statistical sense) will have comparatively few stations included in its region of influence. Although this implies that the precision of the at-site extreme flow estimates is likely to be low, this is not unreasonable since the site is essentially different from the majority of the sites.

In Table 2(b), the number of station-years of record utilized at the sites with each of the two approaches is presented. The mean values for the station-years of record for the ROI stations exhibit less variability amongst the regions than is the case for the regional values. This would indicate that the ROI approach perhaps uses the available data more efficiently. The actual magnitudes of the values of station-years for the ROI approach could of course be altered by adjusting the threshold value for the distance measure such that a greater or lesser number of stations are included in the ROI's.

***Table 2** Comparison of amount of information included for regionalization and ROI approach*
(a) Number of stations

Region number	*Regionalization approach*	*ROI technique mean*	*range*
1	*21*	*22*	*8-39*
2	*42*	*24*	*1-50*
3	*10*	*29*	*9-47*
4	*18*	*14*	*2-34*

(b) Station-years of record

Region number	*Regionalization approach*	*ROI technique mean*
1	*603*	*582*
2	*1200*	*666*
3	*237*	*773*
4	*486*	*377*

Figure 2 shows that different amounts of information are used in the at-site estimation for different stations with the ROI approach. The plot of the frequency histogram for station-years of record for the 91 sites with the ROI approach indicates the range that exists for the amount of information included in the various regions of influence. A distribution for the number of station-years of record is a meritorious feature in that it is anticipated that different sites will have different degrees of similarity with the remaining

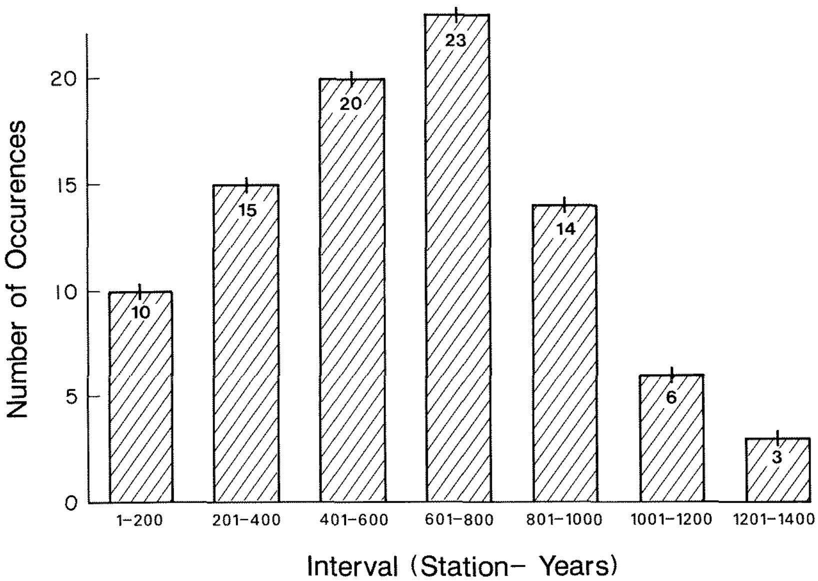

Fig. 2 Frequency histogram for the number of station-years included in the regions of influence.

stations. This differential likeness should lead to differing numbers of stations being of relevance in the at-site extreme flow estimation.

In addition to the summary comparison of the two approaches presented above, certain anomalies in the results were noted and will be discussed. Figure 3 presents the ROI for two pairs of sites: one pair from region 2, and one pair from region 4. As can be discerned from Fig. 3, there are dramatic differences in the respective ROI's in spite of the fact that in each case both stations in the pair are close together and come from the same region (as delineated by the regionalization technique). Although sites 39 and 46 have almost identical ROI's, the regions of influence for sites 24 and 58 have only one station in common. An explanation for this result can be found in Table 3 which summarizes the characteristics of the four sites. Although both station pairs exhibit similarities in drainage area and years of record, the congruence in the statistical attributes is markedly different. While sites 39 and 46 have similar values for *CV* and *QDA*, stations 24 and 58 have very different values for both attributes.

Under the regionalization approach, the same set of stations would be used for the estimation of the at-site extreme flow at site 24 and site 58, since both sites are in region 4. The results presented above indicate that this is clearly inappropriate and the ROI approach can thus be expected to provide a more effective information transfer. It should be noted that the example of sites 24 and 58 is an extreme case. The average *CV* and *QDA* for the stations in region 4 fall between the values for the two sites such that the stations could be regarded as outliers that will not be well described by the regional characteristics. It should also be noted that the similarity in the ROI's for sites 39 and 46 was exceptionally good. In most cases, the

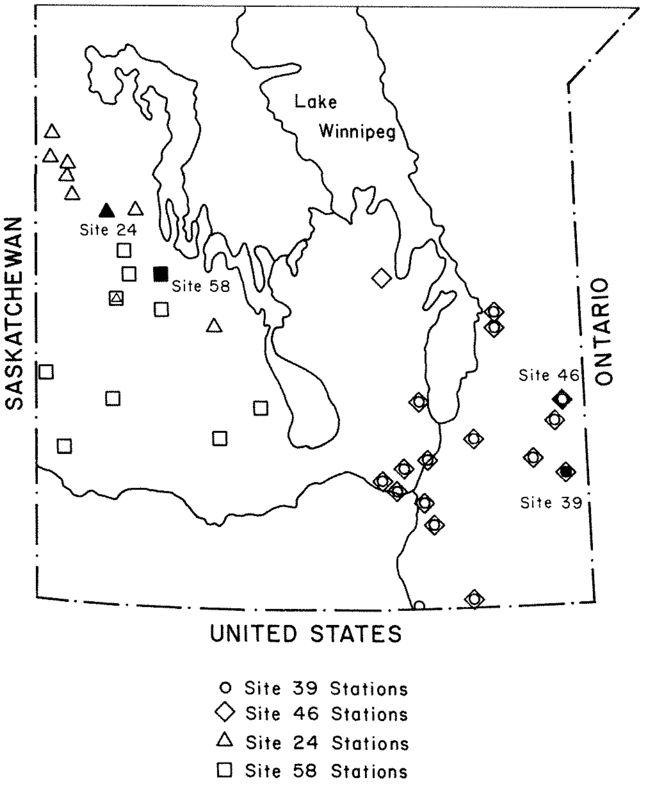

Fig. 3 *Stations included in the region of influence for sites 24 and 58 and sites 39 and 46.*

Table 3 *Station characteristics for selected sites*

Station	*River*	*Drainage area (km^2)*	*Years of record*	*CV*	*QDA*
39	*Whiteshell*	*883*	*25*	*0.54*	*0.017*
46	*Bird*	*1070*	*24*	*0.60*	*0.019*
24	*Pine*	*210*	*31*	*0.52*	*0.062*
58	*Fishing*	*262*	*35*	*1.28*	*0.033*

agreement noted between pairs of stations would lie somewhere between the two cases presented.

The final comparison on a station basis is embodied in Fig. 4 which presents the ROI and the regional stations for site 58. As can be discerned from the Figure, there are substantial differences in the station membership for site 58 with the two approaches. It may be recalled that site 58 was identified above as being somewhat of a regional outlier and it is therefore not surprising to observe the behaviour depicted in Fig. 4.

Site 58 differs from many of the surrounding stations in region 4 in

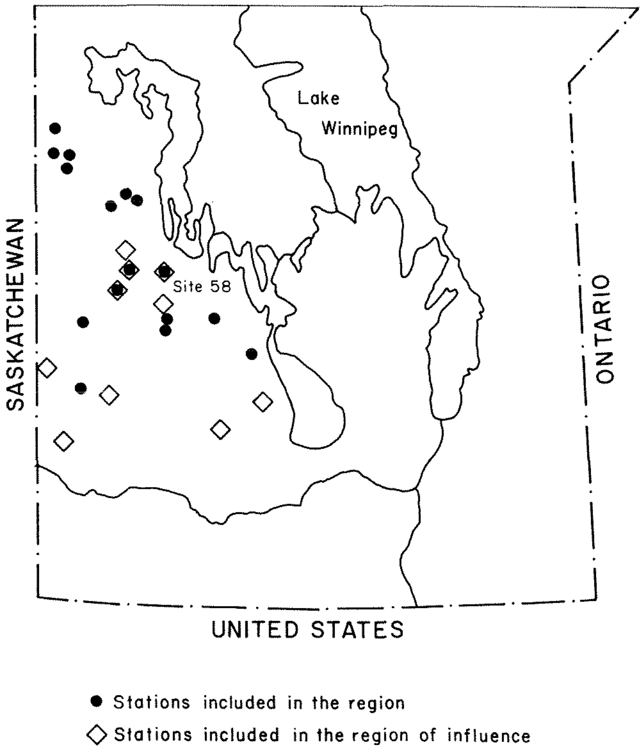

Fig. 4 A comparison of the region of influence and the region for site 58.

terms of the statistical attributes (i.e. CV = 1.28, QDA = 0.033 versus regional averages of CV = 0.81, QDA = 0.050). It is thus to be expected that the ROI for site 58 will contain fewer stations than are included in region 4, as is in fact observed. It would further be anticipated that the statistical attributes of the stations in the ROI would more closely resemble the values for site 58 than is the case for the stations in region 4. The average values for the attributes for the stations in the ROI are: CV = 1.06, QDA = 0.029, which indicates that the stations in the ROI have the desired statistical characteristics. The ROI and region for site 58 are statistically different (at the 5% level of significance) in terms of the CV and QDA values. In addition, the stations in the ROI constitute a more homogeneous group than is the case for the regional stations. It is also interesting to note, that of the ten stations in the ROI for site 58, only three are members of region 4. This illustrates a further advantage for the ROI approach for sites that differ from many other stations, in that information from stations identified as belonging to other regions may also be incorporated in the at-site extreme flow estimation.

A second basis for comparing the two methods for combining regional information with at-site information is through an estimation of the extreme flow values. Since the "true" extreme flow values at the sites are unknown, it is not possible to determine which of the estimation techniques is unequivocally the best. However, it will be possible to compare the results of using the two regional estimation approaches in order to evaluate indirectly the effectiveness of the information transfer. The work presented in this segment of the paper can be viewed as an attempt to move closer to a decision-space comparison of the two approaches.

The equations used for estimating at-site extreme flow values for the two distribution functions with the ROI approach have already been presented. To estimate at-site extremes within the regionalization framework, the stations used consisted of all stations in the region with each site receiving a weight of unity. Thus it was possible to use the same basic procedure as outlined previously, with a modified definition for the set I_j and the weighting factor, WF_{jk}.

Extreme flow values for various return intervals were estimated and an overall comparison of the regional and ROI values was obtained through the calculation of a test statistic defined as:

$$TS_T = \left[\frac{1}{NS} \Sigma_{i=1}^{NS} \left\{ \left(QR_T^i - QI_T^i \right) / \, QI_T^i \right\}^2 \right]^{1/2} \tag{15}$$

where TS_T is the value for the test statistic for the T year flow, QR_T^i and QI_T^i are the T year flows estimated for site i using the regional and ROI approaches respectively, and the summation is over all stations in the network. Values for TS_T were calculated for both distribution functions and for 25, 50, 100, and 200 year events. The results are presented in Table 4. As expected, the value of TS_T increased with an increase in the return period, and was also noted to be lower for the LP3 distribution than for the GEV at all return periods examined. This latter result may have arisen from differences in the manner in which information from surrounding stations is included in the at-site estimation procedure for the two distributions. The results of the analysis indicated that, on average, there is a reasonable agreement between the extreme flows calculated with the two regionalization approaches.

Although there was a fair agreement in the extreme flow estimates, as measured on a network average basis, individual sites exhibited a considerable range of congruence. Table 5 summarizes some of the characteristics for a selection of stations that displayed a substantial difference between the

Table 4 *Test statistic values for various return periods*

Return period (years)	*Test statistic value*	
	LP3	*GEV*
25	*0.113*	*0.130*
50	*0.153*	*0.179*
100	*0.193*	*0.231*
200	*0.232*	*0.287*

***Table 5** Station attributes for sites with statistically different regional versus ROI CV and QDA* values

Site	*Region number*	*Years of record*	*Number of stations*		*Site statistics*	
			Region	*ROI*	*CV*	*QDA*
16	*2*	*25*	*42*	*18*	*0.52*	*0.023*
28	*1*	*26*	*21*	*16*	*0.88*	*0.038*
29	*2*	*65*	*42*	*10*	*0.45*	*0.012*
38	*2*	*38*	*42*	*23*	*0.60*	*0.023*
39	*2*	*25*	*42*	*16*	*0.54*	*0.017*
40	*2*	*78*	*42*	*10*	*0.40*	*0.011*
46	*2*	*24*	*42*	*16*	*0.60*	*0.019*
58	*4*	*35*	*18*	*10*	*1.28*	*0.033*
64	*1*	*23*	*21*	*30*	*0.98*	*0.059*
82	*2*	*24*	*42*	*24*	*1.20*	*0.007*

extreme flow calculated on a regional and a ROI basis. All of the sites listed in Table 5 have statistically different (at the 5% level) regional *CV* and *QDA* values as compared to the ROI values. There are also an additional 17 sites with statistically different regional versus ROI *CV* values and a further 17 sites with statistically different *QDA* values. It is interesting to note that the majority of the sites in Table 5 have fewer stations included in their ROI than are in the region while, for site 64, the ROI has more stations. Fewer sites in the ROI would imply that there is a scarcity of similar information in the network of stations indicating that the site is somewhat of an outlier. Conversely, the situation for site 64 implies that there is similar information available, but not necessarily in the region to which the site has been assigned.

It is possible to infer which of the two approaches is providing a more accurate at-site extreme flow estimation by examining the average values for the statistical attributes for the ROI and the region for each site in Table 5. For all of the sites, the average value for the *CV* and *QDA* for the stations in the ROI was closer to the at-site statistics than was the case for the stations in the region. This does not prove that the ROI extreme flow estimates are superior since the intent in regional flood frequency analysis is to incorporate additional (and potentially different) information. For sites with a comparatively short record, any additional information may be meritorious, but for stations with even a modest data set length (say of the order of 25–30 years) it would be expected that the additional information should be quite similar to the at-site data. It can thus be presumed that the ROI estimates are closer to the "true" extreme flow values than are the estimates obtained from the regionalization procedure for the stations in Table 5.

The apparent superiority of the ROI extreme flow estimates is further illustrated in Fig. 5, which presents the regional and ROI growth curves for site 16 with the GEV distribution. Also shown on Fig. 5 are the 25 observed annual extremes plotted using the Gringorten plotting position formula. The ROI growth curve obviously provides a better fit to the at-site data than does the regional curve. Similar results were obtained for other sites listed in Table 5.

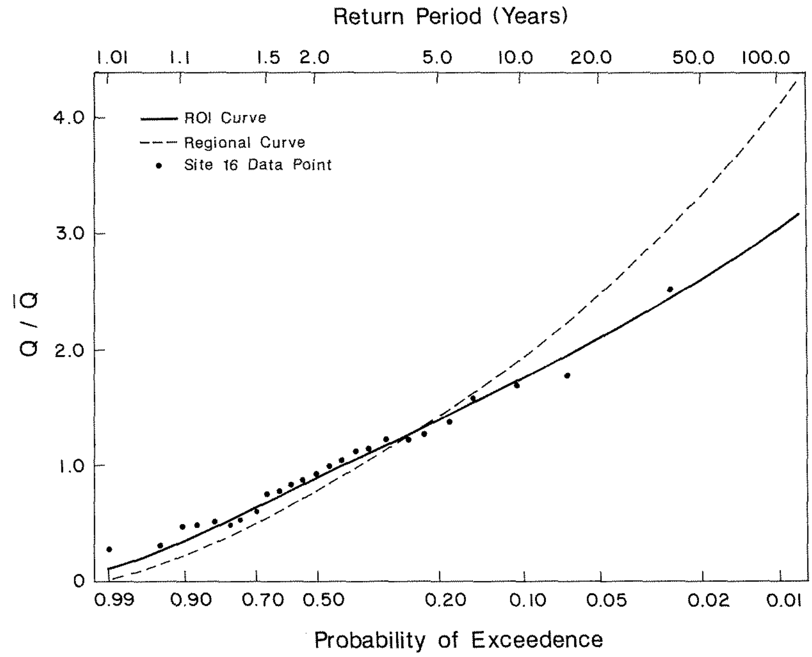

Fig. 5 GEV distributions for site 16.

SUMMARY AND CONCLUSIONS

The research described in this paper compared two techniques for combining regional information for flood frequency analysis. The region of influence approach was demonstrated to have several advantages over a conventional regionalization technique. The two approaches were compared based on the characteristics of the stations used in the at-site analysis, and on the extreme flow values estimated.

The region of influence approach and the regionalization technique were shown to produce comparable results for sites with "average" characteristics, but for unusual sites the advantages of the ROI approach were especially noteworthy. Many sites were identified for which the average attributes of the stations in the ROI for the site were statistically different from the corresponding regional average attributes. The superiority of the ROI approach is thus manifested in the identification of a set of stations with a closer concordance with the target location as well as the estimation of extreme flow values that are more accurate than results from the regionalization technique. In essence, the ROI approach can be viewed as a more flexible methodology for incorporating information from surrounding sites in the at-site extreme flow estimation process.

The ROI approach is a versatile procedure which can be combined with different extreme flow estimators. The opportunity exists for subjective inputs to the process through the selection of a threshold for the distance measure, the choice of attributes to use as a measure of similarity, and the definition of a weighting function to reflect the relative importance of stations. Engineering judgement can therefore be used to obtain an appropriate degree of homogeneity for the stations included in the region of influence.

REFERENCES

Acreman, M. C. (1987) *Regional Flood Frequency Analysis in the UK: Recent Research - New Ideas,* Institute of Hydrology, Wallingford, UK.

Acreman, M. C. & Wiltshire, S. E. (1987) Identification of regions for regional flood frequency analysis, *EOS* 68 (44), 1262. (Abstract)

Burn, D. H. (1988) Delineation of groups for regional flood frequency analysis; to appear in *J. Hydrol.*

Greiss, N. P. & Wood, E. F. (1981) Regional flood frequency estimation and network design, *Wat. Resour. Res.* 17 (4), 1167-1177.

Hosking, J. R., Wallis, J. R. & Wood, E. F. (1985) Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics* 27 (3), 251- 261.

Kite, G. W. (1977) *Frequency and Risk Analysis in Hydrology,* Water Resources Publications, Littleton, Colorado.

Lettenmaier, D. P., Wallis, J. R. & Wood, E. F. (1987) Effect of regional heterogeneity on flood frequency estimation, *Wat. Resour. Res.* 23 (2), 313-323.

Thomas, W. O. (1985) A uniform technique for flood frequency analysis, *J. Wat. Resour. Plan. Manag. ASCE* 111 (3), 321-337.

Wiltshire, S. E. (1986) Regional flood frequency analysis I: Homogeneity statistics, *Hydrol. Sci. J.* 31 (3), 321-333.

Received 16 May 1988; accepted 8 May 1989