HIERARCHICAL APPROACH FOR REGIONAL FLOOD FREQUENCY ANALYSIS

Mauro Fiorentino
Dipartimento di Difesa del Suolo, Università della Calabria.
C/da S. Antonello, 87040 Montalto Uff. (Cosenza), Italy.

Salvatore Gabriele
Consiglio Nazionale delle Ricerche-IRPI.
Via Verdi 1, 87030 Rende (Cosenza), Italy.

Fabio Rossi
Dipartimento di Idraulica, Gestione delle Risorse Idriche ed Ingegneria Ambientale.
Via Claudio 21, 80125 Napoli, Italy.

Pasquale Versace
Dipartimento di Difesa del Suolo, Università della Calabria, and CNR-IRPI. Cosenza, Italy.

ABSTRACT. In this paper a hierarchical approach for regional flood frequency analysis is proposed. It is characterized by three investigation levels: i) identification of homogeneous regions wherein the skewness coefficient is assumed not to vary from site to site; ii) identification of homogeneous sub-regions wherein the coefficient of variation is assumed to be constant too; iii) use of regression models to relate a location parameter of the flood distribution to climatic and physiographic factors.

The hierarchical approach is herein presented with the use of the two-component extreme value (TCEV) distribution. The validity of the presented approach is demonstrated using twenty-eight Italian annual flood series (AFS). In particular, it is shown that the sample skewness coefficient exhibits an observed variability comparable with the theoretical sampling variability, the latter derived by computer simulation experiments from a regionalized TCEV distribution considered as representative of the entire region in question. Instead, for the same AFS's, the observed variance of the coefficient of variation is nearly twice the one derived from the sampling experiments.

V. P. Singh (ed.), Regional Flood Frequency Analysis, 35-49.
I. INTRODUCTION

Regionalization techniques are used in flood frequency analysis mainly to estimate T-year flood flow \( Q \) at ungauged river basin sites and also to improve site-specific estimates based on limited single site data. In an implicit or explicit manner, regional analysis is based on a flood distribution which describes how \( Q \) varies with the return period \( T \). It is also based on a regionalization model which associates annual flood characteristics with physiographic and climatic causative factors by a number of relationships equal to the number of flood distribution parameters.

The analysis can be greatly simplified since in large climatic areas the moment-ratios of the annual flood distribution - e.g. the coefficients of variation and skewness - appear to be non-significantly variable and practically independent of the physiographic factors, except perhaps the basin area. This empirical evidence has suggested the use of regional flood frequency distribution representing the average ratios of the floods at various return periods to a scaling index flood, e.g. the mean annual flood. To obtain estimates of the T-year flood at ungauged sites in the region, one can use this regionalized distribution, usually referred to as growth curve, together with regression models, the latter being used for estimation of the index flood. The approach referred to as the index flood method was first introduced in the U.S.A. by the Geological Survey (Dalrymple, 1960; Riggs, 1973).

Furthermore a procedure for identifying homogeneous regions is required. A realistic regionalization model was proposed by Matalas and Gilroy (1968), who adopted the classical analysis of the variance into two components representing different sources of random variation. Regionalization was then considered to extend lengths of the historical sequences and led to a reduction in time sampling errors, though at the price of introducing space disturbance errors. In addition, the reduction of the time sampling variance as the number of stations increases is limited by the correlation among concurrent flood flows.

The transfer of regional information is then affected by three factors, viz. time sampling variance, space disturbance variance, and interstation correlation. However, these play a different role as the order of the moment increases from the mean to the coefficient of skewness. It is known that the sampling variability increases with the order of moments, while in this paper the ratio of space disturbance variance to the sampling variance will be shown to decrease. Furthermore, Stedinger (1983) showed that the influence of the interstation correlation decreases passing from the mean to the skewness.

Such a varying regional behavior of the statistical parameters suggested to Fiorentino et al. (1985) a modified version of the index flood approach. Instead of using two levels of analysis, the first one dealing with ratio of the T-year flood to the scaling value and the second with the location parameter, regional analysis can be hierarchically broken down into three distinct levels, one each for the following statistical parameters: coefficient of skewness, coefficient of variation, and location value. A homogeneous region with respect to the coefficient of skewness is firstly identified. Within this region more homogeneous sub-
regions with respect to the coefficient of variation can be discriminated to reduce large space disturbance errors. This approach, herein referred to as hierarchical approach, optimizes the transfer of regional information.

In this paper a regionalization estimation procedure is presented. It is based on the two-component extreme value (TCEV) distribution, on the maximum likelihood (ML) estimation method, and on two different averaging procedures for regional estimation. However the hierarchical approach is suitable for application with different distributions and different estimation procedures.

2. REGIONALIZATION MODELS

A conceptual framework to transfer regional information was articulated by Matalas and Gilroy (1968). Suppose that there are k gauged sites in a region, each with n years of record. Let \( \alpha \) be a parameter or a moment-ratio of the annual flood distribution. The simplest regionalization model assumes that in a given region \( \alpha \) is not significantly affected by climatic and physiographic factors. However, this does not mean that \( \alpha \) is identical at any site. Since presumably there are many unaccounted for causative factors at work, it is reasonable to assume for the basin \( j \) (\( j=1,2,\ldots,k \))

\[
\alpha_j = \alpha_0 + \delta_j
\]  

(1)

where \( \alpha_0 \) is a regional parameter equal to the mean of the \( \alpha \)'s and \( \delta_j \) represents a random space disturbance error with mean \( \text{E}(\delta)=0 \) and variance \( \text{VAR}(\delta)=\omega^2 \), the latter being equal to that of the \( \alpha_j \)'s. At station \( j \) let us suppose that an unbiased estimate \( \tilde{\alpha}_j \) of \( \alpha_j \) can be obtained from data and is given as

\[
\tilde{\alpha}_j = \alpha_0 + \gamma_j
\]  

(2)

where the random variable \( \gamma_j \) is a time sampling error having \( \text{E}(\gamma_j)=0 \) and \( \text{VAR}(\gamma_j)=\omega_2^2 \) which is equal to that of the \( \tilde{\alpha}_j \)'s. Combining (1) and (2) we get

\[
\hat{\alpha}_j = \alpha_0 + \delta_j + \gamma_j
\]  

(3)

Assuming that the variance \( \sigma_j^2 = \sigma^2 \), i.e. it is the same for all sites, and that \( \delta_j \) and \( \gamma_j \) are independent, the mean and variance of the \( \hat{\alpha}_j \)'s over both the time and space are

\[
\text{E}[\hat{\alpha}_j] = \alpha_0 \quad ; \quad \text{VAR}[\hat{\alpha}_j] = \sigma^2 + \omega^2 = V
\]  

(4)

The simplest regionalization model then contains three parameters : \( \alpha_0, \omega^2, \sigma^2 \). Estimates of these parameters can be obtained from the \( \hat{\alpha}_j \)'s. Particularly an estimate of \( \alpha_0 \) is obtained by
\[ \hat{\alpha}_0 = \frac{1}{k} \sum_{j=1}^{k} \hat{\alpha}_j / k \]  

(5)

while estimates of \( \omega^2 \) can be derived by

\[ \omega^2 = \nu - \sigma^2 \]  

(6)

where

\[ \nu = \frac{1}{k} \sum_{j=1}^{k} (\hat{\alpha}_j - \alpha_0)^2 / (k-1) \]  

(7)

and the sampling variance \( \sigma^2 \) can be evaluated by analytical computation or computer simulation experiments.

The spatial mean \( \bar{\alpha}_0 \) of the \( \hat{\alpha}_j \)'s can be used as a regional estimator of \( \alpha_j \) at every site \( j \). The mean square error (MSE) of \( \bar{\alpha}_0 \), i.e. the variance of \( \bar{\alpha}_0 \) about \( \alpha_j \), was given by Matalas and Gilroy (1968)

\[ \text{MSE}[\bar{\alpha}_0] = E[(\bar{\alpha}_0 - \alpha_j)^2] = \sigma^2 / k_\alpha + \omega^2 (1 \pm 1 / k) \]  

(8)

where the sign plus or minus differentiates the ungauged from the gauged site \( j \). In (8) \( k_\alpha \) is the equivalent number of independent stations and can be put in the form:

\[ k_\alpha = k / [1 + (k-1) \beta_\alpha] \]  

(9)

where \( k \) is the number of stations and \( \beta_\alpha \) is the average interstation correlation coefficient of the \( \hat{\alpha}_j \)'s. Matalas and Benson (1961), Matalas and Gilroy (1968), and Stedinger (1983) showed that \( \beta_\alpha \) is more or less equal to \( \beta, \beta^2 \) and \( \beta^3 \), \( \alpha \) being equal to the mean, variance, and skewness respectively, and where \( \beta \) is the average interstation correlation of concurrent annual floods.

The regionalization leads to a space-time trade-off that is regulated by the relation (8). Increasing the number of sites, one obtains a reduction of time sampling variability which is partially counterbalanced by the space disturbance errors and is limited by interstation correlation. A criterion of minimizing MSE of the regionalized estimator can be used for identifying homogeneous regions together with cluster analysis techniques.

When the space disturbance variance \( \omega^2 \) is large, it can be reduced by dividing the region into more homogeneous sub-regions or introducing some climatic or physiographic factors which are able to account, at least partially, for \( \omega^2 \). In the latter case the simplest regionalization model of equation (1) becomes a regression model, and the MSE of the regionalized estimator is given by a regionalized expression of the relation (8) (Matalas and Gilroy, 1968).

In this paper it will be shown, using Italian data, that the skewness coefficient exhibits a smaller space disturbance variance than the coefficient of variation. Thereupon, and also because it is less af-
fected by correlation, more stations can be used to transfer information for estimating skewness or distribution parameters depending on skewness.

3. HIERARCHICAL REGIONALIZATION APPROACH USING THE TCEV DISTRIBUTION

3.1. Flood Distribution.

The TCEV distribution estimation (Rossi et al., 1984) constitutes a reasonable basis for a hierarchical-type regionalization model. This distribution originates from a phenomenological model of the flood process by considering the annual flood to be the maximum of a Poisson-distributed number K of independent and identically distributed random variables Z's, also independent of K (Versace et al., 1982). The distribution of Z variate was empirically assumed to be a mixture of two exponential components. Its probability density function (PDF) is

\[
f_X(x) = \exp(-\Lambda_1 e^{-x / \theta_1} - \Lambda_2 e^{-x / \theta_2})(\Lambda_1 e^{-x / \theta_1 + \Lambda_2 e^{-x / \theta_2}} ; x>0) \quad \theta_1 \quad \theta_2
\]

(10)

\(\Lambda_1 > \Lambda_2\) and \(\Lambda_2 > 0\) being the mean annual number of independent floods coming from the first and second component respectively. Since the finite probability that the value is exactly zero is negligible, PDF (10) is indistinguishable in practice from a continuous function also allowing for negative values, except perhaps in arid or semiarid basin sites, where small values of \(\Lambda\) are likely to be found.

In order to develop the regionalization model it is useful to introduce the reduced variate

\[
Y = \frac{X - \ln\Lambda_1}{\theta_1}
\]

(11)

having PDF which for \(x>0\) is

\[
f_Y(y) = \exp(-e^{-y - \Lambda_2 e^{-y / \theta_2}})(e^{-y + \Lambda_2 e^{-y / \theta_2}}) \quad \theta_2
\]

(12)

where

\[
\theta_2 = \theta_2 / \theta_1
\]

\[
\Lambda_2 = \Lambda_2 / \theta_1 / \theta_2
\]

(13)

Equation (10) can then be written as

\[
f_X(x) = \exp(-\Lambda_1 e^{-x / \theta_1} - \Lambda_2 e^{-x / \theta_2})(\Lambda_1 e^{-x / \theta_1 + \Lambda_2 e^{-x / \theta_2}} ; x>0) \quad \theta_1 \theta_2
\]

(14)
The coefficients of skewness \( \gamma_1 = \frac{\mu_3}{\mu_2^{3/2}} \) of TCEV distribution depends on \( \theta_1 \) and \( \Lambda_2 \) while its coefficient of variation \( \gamma = \frac{\mu_2^{1/2}}{\mu} \) depends on \( \theta_1, \Lambda_2 \) and \( \Lambda_1, \mu_2 \) and \( \mu_3 \) being the second and third central moment respectively. The mean \( \mu \) depends on all the four parameters of the distribution. Expressions of the moments were given by Beran et al. (1986).

3.2. Regionalization Model.

Using the TCEV distribution, and according to the considerations outlined in the Introduction, a hierarchical regionalization approach can be developed in the following manner (Fiorentino et al., 1985).

At the first level an extensive region, wherein the theoretical coefficient of skewness \( \gamma_1 \) is assumed to be constant, is selected. Within this region dimensionless parameters \( \theta_1 \) and \( \Lambda_2 \) are also constant and the distribution of \( Y \) variate is therefore identical at any site.

The second level assumes that homogeneous sub-regions with respect to \( \gamma \) exist within the region where \( \gamma_1 \) is constant. In each sub-region, parameters \( \theta_1 \), \( \Lambda_2 \), and \( \Lambda_1 \) are assumed not to change from site to site, and therefore the ratio of \( X \) variate to the index flood is supposed to be identically distributed regardless of the site.

Finally at the third level the hypothesis that the index flood is correlated with some climatic and physiographic factors of the basin is made.

3.3. Estimation Procedure.

Estimation of parameters is strictly related to the regionalization model and differs from level to level.

Four parameters of the TCEV distribution can be estimated from a site-specific data set (it can be thought of as the zeroth level of regionalization) using ML estimation method (Rossi et al., 1984). However, estimates of \( \theta_1 \) and \( \Lambda_2 \) obtained in this manner are too uncertain, especially in case very short samples are used. This prompts one to invoke the use of regionalization techniques, at least for estimation of \( \theta_1 \) and \( \Lambda_2 \), it also influences the choice of the regionalization procedure.

The method proposed by Fiorentino and Gabriele (1985), and currently used to estimate \( \theta_1 \) and \( \Lambda_2 \), is an iterative ML procedure, being essentially a station-year method which, as well known, is sensitive to the interstation correlation effects. Nevertheless this method was chosen because it does not invoke an explicit averaging between site-specific estimates of \( \theta_1 \) and \( \Lambda_2 \). Furthermore it can be noted that (1) estimation of \( \theta_1 \) and \( \Lambda_2 \) implicitly involves estimation of moments with order larger than two, and (2) correlation between moments progressively decreases as the order increases.

More precisely, the first level of the regional estimation procedure consists of carrying out the following steps using \( k \) annual flood series (APS) each with \( n_j \) \((j=1,2,\ldots,k)\) years of record:

a) for each series, assign tentative values to parameters \( \theta_1 \) and \( \Lambda_1 \);
b) transform all the data by (11), and form a pooled sample with the so obtained y's. The length of this sample is

\[ N = \sum_{j=1}^{k} n_j \]

c) estimate \( \theta_z \) and \( \Lambda_z \) using (12): \( \theta_z, \Lambda_z : \ln L_4 = \max \), where

\[ L_4 = \prod_{i=1}^{N} f_y(y_i ; \theta_z, \Lambda_z) \]

d) for each series estimate \( \theta_1 \) and \( \Lambda_1 \) using the ML method constrained to the values of \( \theta_z \) and \( \Lambda_z \) derived at step c): \( \theta_1, \Lambda_1 : \ln L_2 = \max \), where

\[ L_2 = \prod_{i=1}^{n_j} f_x(x_i ; \theta_1, \Lambda_1 | \theta_z, \Lambda_z) \]

and \( f_x(x_i) \) is given by (14);

e) repeat steps b) to d) till estimates of \( \theta_z \) and \( \Lambda_z \) no longer change.

In short, at the first level two regional estimates, of \( \theta_z \) and \( \Lambda_z \), and two site-specific estimates constrained to them for each series, are attained.

At the second level, regional estimation of \( \Lambda_1 \) is carried out by introducing the parameter \( CV_1 \) which is obtained by the transformation

\[ CV_1 = 0.557/(\log\Lambda_1 + 0.251) \]

and whose estimator is less variable and biased than that of \( \Lambda_1 \) (Fiorentino and Gabrielle, 1985). More precisely, average of the site-specific estimates of \( CV_1 \) is used to give regional estimate of \( \Lambda_1 \) by inverting relationship (15).

At this level only the parameter \( \theta_1 \) assumes different values from site to site in the sub-region investigated. Its estimates are obtained using the ML method and constraining the solution to the regional estimates of \( \theta_z \), \( \Lambda_z \), and \( \Lambda_1 \):

\[ \theta_1 : \ln L_3 = \max \), where

\[ L_3 = \prod_{i=1}^{n_j} f_x(x_i ; \theta_1 | \theta_z, \Lambda_z, \Lambda_1) \]

At the third level regression models are used to achieve index flood estimates at ungauged sites.
4. STATISTICAL PROPERTIES OF THE REGIONALIZED TCEV-ML ESTIMATOR

FOR \( \sigma^2 = \hat{\beta} = 0 \)

The regional flood-frequency estimation procedure using the TCEV distribution was assessed in two previous papers (Fiorentino and Gabriele, 1985; Arnell and Gabriele, 1986) exploiting Monte Carlo experiments. The hypothesis that floods come from a homogeneous world was made throughout the simulation works. That is disturbance effects caused (1) by floods drawn from a population different from that of others and (2) by data errors were left out of account. Moreover spatial (interstation) correlation effects were not simulated in the experiments. Fiorentino and Gabriele (1985) gauged the accuracy in estimating parameters and quantiles with increasing exploitation of the regional estimation procedure levels. Synthetic series were generated from a TCEV distribution, with theoretical coefficients of skewness \( \gamma_1 = 2.75 \) and of variation \( \gamma = 0.59 \), which was considered as a reasonable "parent" of the central and southern Italian annual floods.

The fitting of the TCEV distribution to a site-specific data set was first simulated. Then the performance of the regional estimator at both first and second level was assessed. Quantile estimators were shown to be little biased regardless of the amount of regional information utilized, though the bias increased with return period. Root mean square error was also used as a measure of performance. It was standardized to ensure compatibility between different experiments by \( \text{RMSE} = \left( \frac{1}{E(X-x)^2} \right)^{\frac{1}{2}} \), where \( X \) and \( x \) are the estimated and population value respectively of either at-site quantile or one of the parameters as appropriate. As regards the quantile estimates, RMSE values are shown in Table 1. It can be seen that, as expected in an idealistic situation of having homogeneous data, the RMSE drastically decreased everywhere as regional information was acquired. It is worth emphasizing the reasonable low value of the RMSE of estimates, even of upper quantiles, once the second level of the regionalization procedure is exploited. On the other hand, admittedly, quantile estimation based upon a single data set appears very uncertain.

<table>
<thead>
<tr>
<th>F</th>
<th>Regionalization level</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9</td>
<td>0.15 0.10 0.08</td>
</tr>
<tr>
<td>0.98</td>
<td>0.27 0.12 0.09</td>
</tr>
<tr>
<td>0.99</td>
<td>0.31 0.14 0.10</td>
</tr>
<tr>
<td>0.998</td>
<td>0.38 0.16 0.12</td>
</tr>
<tr>
<td>0.999</td>
<td>0.40 0.17 0.13</td>
</tr>
</tbody>
</table>

Table 1 - RMSE of quantile estimates.

Statistical properties of the regionalized TCEV-ML estimators were compared with those of other regionalization procedures by Arnell and Gabriele (1986), who investigated the robustness of competing estimators with respect to changes in parent distribution. Herein the
The real world of floods is assumed to have a TCEV distribution. Three regional estimation algorithms were selected for comparison. The first two follow a regionalization procedure based on the probability-weighted moment (PWM) estimation method for the Wakeby (WAK) distribution and the generalized extreme value (GEV) distribution, which are described in Wallis (1980) and in Hosking et. al. (1984) respectively. The third one is an up-to-date computerized version of the graphical procedure which was employed in the 1975 U.K. Flood Studies Report (NERC, 1975). As quoted by the Authors this version slightly differs from that developed by Hosking et. al. (1985). It will be referred to as FSR.

Some of the results obtained by Arnell and Gabriele (1986) are herein shown.

<table>
<thead>
<tr>
<th>Quantiles</th>
<th>Growth factors</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F</td>
</tr>
<tr>
<td>GEV-PWM</td>
<td>0.95</td>
</tr>
<tr>
<td></td>
<td>0.98</td>
</tr>
<tr>
<td></td>
<td>0.99</td>
</tr>
<tr>
<td></td>
<td>0.998</td>
</tr>
<tr>
<td></td>
<td>0.999</td>
</tr>
<tr>
<td>WAK-PWM</td>
<td>0.95</td>
</tr>
<tr>
<td></td>
<td>0.98</td>
</tr>
<tr>
<td></td>
<td>0.99</td>
</tr>
<tr>
<td></td>
<td>0.998</td>
</tr>
<tr>
<td></td>
<td>0.999</td>
</tr>
<tr>
<td>FSR</td>
<td>0.95</td>
</tr>
<tr>
<td></td>
<td>0.98</td>
</tr>
<tr>
<td></td>
<td>0.99</td>
</tr>
<tr>
<td></td>
<td>0.998</td>
</tr>
<tr>
<td></td>
<td>0.999</td>
</tr>
<tr>
<td>TCEV</td>
<td>0.95</td>
</tr>
<tr>
<td></td>
<td>0.98</td>
</tr>
<tr>
<td></td>
<td>0.99</td>
</tr>
<tr>
<td></td>
<td>0.998</td>
</tr>
<tr>
<td></td>
<td>0.999</td>
</tr>
</tbody>
</table>

Table 2 - RMSE of quantile estimates (second level of the regionalized TCEV-ML procedure)
Two different TCEV distributions were taken into account as parent. The former was considered as representative of the U.K. flood experience (γ₁ = 3.5, γ = 0.40) (WORLD-1) the latter was chosen to be the one which has been referred above (γ₁ = 2.75, γ = 0.59; WORLD-2).

Table 2 refers to the RMSE of quantile and growth factor estimates, growth factor being the ratio of quantile to the index flood. It can be inferred that in comparison to the other estimators, the TCEV-ML procedure performs better for at-site quantiles than for growth factor.

With regard to quantile estimators, TCEV procedure exploited the advantage that the population was a TCEV-type world, and in both the experiments, whereas it was outperformed by the competing estimators only in the growth factor estimation for WORLD-2. By the way, something needs to be said about small sample properties of ML estimators of θγ and Λγ. In fact θγ reflects (RMSE=0.153), the good performance of quantile regional estimators, whereas the RMSE of Λγ (≈ 0.717) is undoubtedly unsatisfactory. As a consequence, this may suggest that rather better statistical properties of quantile and specifically of growth factor regional estimators could be attained by enhancing the estimates of Λγ.

5. IMPACT OF THE REGIONALIZATION MODEL WITH DATA

The use of regionalization techniques requires hydrologists to check the suitability of available data in order to ensure homogeneity between flood series.

A re-examination of the whole flood data base is currently being done in Italy. Since one of the aims of this paper is to show the different spatial heterogeneity of the skewness and variation coefficients, the local factors which can affect skewness and variability of the AFS's have been examined, particularly: i) lateral spilling immediately upstream the flood gauging station; ii) effect of the basin area; iii) ephemeral streams in semiarid areas.

Flood series recorded downstream of a river branch tends to be bounded at the upper end when a significant and systematic lateral spilling of high flows occurs at the upstream branch. This leads to outlying low values of both coefficients of variation and skewness. In some cases, however, it was impossible to get sufficient information to discard an AFS showing a suspicious upward curvature of the right-hand tail when plotted on extremal probability paper. Several of these series exhibiting an evident upper bound (a typical example is shown in Fig. 1) represent the basins having drainage area greater than 3000 Km², which according to central and southern Italy geomorphology correspond to rivers self-formed mostly in alluvial flat plains.

The basin drainage area does not seem to affect the coefficients of variation CV and skewness G of the Italian AFS's, except for areas greater than 3000 Km² where lower coefficients of variation are generally found. It is difficult to recognize if this is due to the upstream spilling phenomenon or to climatic reasons, although with regard the latter hypothesis, there is the observation that the annual number Λ = Λ₁ + Λ₁ of independent floods tends to increase with the area.
Fig. 1 - Observed CDF of a typical AFS whose largest values are affected by lateral spilling phenomena.

Fig. 2 - Observed CDF of a typical AFS recorded in semiarid regions.
Anyway, for the sake of simplicity, basins with area larger than 3000 Km² have not been considered so far.

Coefficient of variation tends to increase from humid to arid regions. In Italy, sub-regional median values of CV range from 0.40 to 0.67 except in semi-arid regions, where it reaches unity. AFS's recorded in that region, because of the scanty average annual number of floods, lead to a non-negligible probability of having zero, or practically zero values of annual maximum flow. Clearly, this contrasts with the use of the proposed ML estimation procedure which assumes that the PDF is undistinguishable from a continuous function also allowing for negative values. These AFS's were not considered, either. A typical example of such AFS, recorded in the Sardinia Island, is shown in Fig. 2.

According to the observations outlined above, a data base for inferring the regionalization model was formed by selecting, amongst the AFS's recorded at stations operating in Italy on non-Alpine streams, twenty-eight AFS's having an average length nearly equal to 39, and with a total of 1091 station-year data. The AFS's having less than 35 years of record were not taken into account, as suggested by Arnell and Gabriele (1986) for the first level of the regional estimation procedure.

A sensitivity analysis demonstrated that the selected 28 AFS's were not significantly affected by small changes such as the removal of individual data sets.

TCEV-ML estimation of parameters was carried out using the procedures detailed above and assuming that the parameters θ*, λₙ, and λ₄ have constant values regardless of the site in the entire region in question, θ₀ (and thus any location parameter) being variable from site to site. The following regional estimates of θ*, λₙ, and λ₄ were attained: θ₀=2.654, λₙ=0.350, λ₄=12.15.

Errors depending on neglecting interstation correlation effects are likely insignificant, for the 28 AFS's are quite scattered over the entire region and neither do cover a homogeneous time duration.

Theoretical sampling distribution of the sample coefficients of skewness and variation of the regionalized TCEV distribution with the attained regional values of θ*, λ₄ and λ₄ were derived by exploiting Monte Carlo experiments.

Ten-thousand independent synthetic series were generated with their lengths distributed in the manner of the 28 AFS's. Coefficients of skewness G and variation CV of each series were calculated. Experiment-derived mean and variance of G and CV worked out to be the values listed in Table 3 where they are compared with the respective values calculated using the 28 AFS's.

One may note that the observed variance of G is almost completely accounted for by the sampling variability. This means that the contribution of the spatial disturbance is so little that there is no evidence to reject the hypothesis that the region wherein the 28 AFS's have been recorded is homogeneous with respect to the skewness.

This does not apply to the coefficient of variation CV. In fact its observed variance is almost twice the experiment-derived variance which reproduces the expected sampling variability. Moreover, since the ratio of the time sampling variance to the observed variance is only
0.0127/0.0225=0.564, a division of the entire region into homogeneous sub-regions with respect to CV should be advocated.

<table>
<thead>
<tr>
<th></th>
<th>Observed</th>
<th>Experiment-derived</th>
</tr>
</thead>
<tbody>
<tr>
<td>CV</td>
<td>Mean</td>
<td>0.536</td>
</tr>
<tr>
<td></td>
<td>Variance</td>
<td>0.0225</td>
</tr>
<tr>
<td>G</td>
<td>Mean</td>
<td>1.431</td>
</tr>
<tr>
<td></td>
<td>Variance</td>
<td>0.634</td>
</tr>
</tbody>
</table>

Table 3 - Mean and variance of the coefficients of variation CV and skewness G for 28 observed Italian annual flood series and 10,000 experiment-derived series.

In other words the 28 AFS's exhibit a variability of G, though very large, but comparable with the theoretical one, whereas for the same series the variability of CV is much larger than that expected. This constitutes the basis for suggesting the use of the hierarchical approach. In fact, it enables us to utilize the information contained in a large number of AFS's over an extensive region for estimating the parameters for which the information provided by a specific series is particularly scanty. This approach exploits a smaller number of AFS's only at the next levels, when estimation of parameters for which a specific AFS provides a relatively larger amount of information, is being considered. The hierarchical approach can then perform better than the classical index flood method, which, in order to preserve the simultaneous homogeneity with respect to all the parameters to be estimated, is forced to exploit a limited number of AFS's for estimation of high-order moment-ratios also.

6. CONCLUSION

The hierarchical approach for regional flood frequency estimation herein proposed arises from the observation that the hydrologic homogeneity with respect to the shape parameters of the annual flood distribution can be hypothesized in large regions, while as regards the coefficient of variation the homogeneity is evident only in smaller areas. Therefore, for estimation of the shape parameter, one can use a larger number of data sets than for estimating the coefficient of variation. As a consequence, the hierarchical approach exploits, better than the classical index flood method, the hydrologic information.

A regional estimator which uses: i) the two-component extreme value distribution (TCEV), ii) the maximum likelihood (ML) estimation method, iii) a station-year type regionalization algorithm, is shown to account for the observed distribution of the skewness coefficient of a number of annual flood series (AFS) recorded at stations operating in Italy on
non-Alpine streams. For the same series the observed distribution of the coefficient of variation appears to be more variable than the theoretical sampling distribution. This suggests that the considered region should be broken down into different sub-regions to obtain valuable regional estimates of the dispersion parameter.

Computer simulation experiments showed that the regionalized TCEV-ML estimator exhibits good performance in estimating quantiles of the annual flood distribution, whereas the estimates of the parameters \( \theta \) and, especially \( \Lambda \), are relatively worse. It was shown that local disturbance factors can heavily affect the extreme values of floods, thus making AFS's exhibit anomalous shapes which are independent of the general characteristics of the basin. Some such shapes are frequently accounted for by the stream morphology, a typical example being constituted by those AFS's whose higher values are bounded because of the spilling phenomena occurring immediately upstream the gauging station. On the other hand it seems that these problems deserve more attention than those, largely discussed so far in the literature, devoted to ensure the efficiency and resistance of the estimators with respect to more or less unrealistic parent flood-distribution.

ACKNOWLEDGMENTS. This research has been supported by the Italian National Research Group for the Prevention of Hydro-Geological Disasters.

7. REFERENCES


Fiorentino, M., P.Versace, and F.Rossi, 'Regional flood frequency estimation using the two-component extreme value distribution'. Hydrological Sciences Journal, 30,51-64,1985.


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