Climatic controls on the flood frequency distribution

V. IACOBELLIS[a], P. CLAPS[b] & M. FIORENTINO[b]

[a] *Dipartimento di Ingegneria delle Acque, Politecnico di Bari, via E. Orabona 4, I-70125, Bari, Italia - e-mail:* v.iacobellis@poliba.it
[b] *Dipartimento di Ingegneria e Fisica dell'Ambiente, Università della Basilicata, Contrada Macchia Romana, I-85100, Potenza, Italia*
e-mail: fiorentino@unibas.it

ABSTRACT

Statistical models for flood frequency analysis present refined estimation and validation techniques but do not provide ultimate solutions to when regional analyses are required. More than on specific probabilistic models, in these cases the emphasis is put on the homogeneity or variability patterns of the moments of annual maxima. In this paper the use of climatic ancillary data is shown to provide physically-based clues for understanding variability of parameters, and this proves to be particularly useful for parameters appearing in Poisson-derived distributions. With respect to regional analysis, the discriminating capability of a-priori climatic information is also compared here with a classical statistical approach based on a state-of-the art homogeneity measure. The latter is applied to L-moments of orders 2 to 4 and the former is based on an index of average dryness or humidity of basins. Application on a 10,000 km^2 region in Southern Italy involving 22 gauging stations shows that in the context of the index-flood approach the use of climatic information suggests objective criteria for aggregation of stations in homogeneous regions.

1. INTRODUCTION: CLIMATE-CONSISTENT IMPLICATIONS IN REGIONAL ANALYSIS

Statistical methods for regional flood frequency analysis are widely available in the literature and the proposed approaches generally aim to improve technical details related to the statistical techniques. In this research field the ultimate goal is to provide robust estimates of floods at high recurrence interval, with particular reference to basins with unavailable or unreliable observed flow data. As regards robustness, great efforts have been devoted to the development of statistical methods as much as possible insensitive to violations of some basic assumptions of the specific distributions. This task has been tackled by means of 3- to 5-parameter distributions, in which parameters are fitted with efficient, consistent and accurate estimation procedures.

On the other hand, the development of efficient statistical techniques do not reduce the critical importance of the regional analysis, that represents the ensemble of methods which supplement the inadequacy (or absence) of individual flood data samples with the joint analysis of observations from different stations belonging to a 'statistically homogeneous' region. Regional analysis is even more important in the context of application of probability distributions with 3 or more parameters, because of its beneficial effect of reducing the sampling error with respect to at-site estimates. This effect is even more important when flood risk evaluation is needed with regard to return periods that largely exceed the length of the observed records.

Objective methods for the selection of 'homogeneous' regions are still matter of investigation and different views on the approach to follow are available in the literature. Recent experiences on the development of 'pooling' methods (Institute of Hydrology, 1999) have not obscured the well known 'index flood' method (NERC, 1975), which defines areas with homogeneous (*i.e.* constant) dimensionless probability distribution $F(y)$, with $y=x/m$. A first-order parameter (the index flood m) represents the 'local' information and constitutes the scale factor of the distribution at a given site: $x(F) = m \cdot y(F)$. More important differences among approaches to regional analysis emerge in the degree of use of climatic, geological and morphological information for the selection of stations with common behavior (pooling) or that can be reasonably grouped in a homogeneous region (index-flood).

Considering in more detail this last point, it is important to recognise that climatic issues have recently gained significant attention from different viewpoints related to flood frequency analysis (see *e.g.* Farquharson *et al.*, 1992; Burn, 1997; Iacobellis *et al.*, 1998; Institute of Hydrology, 1999). These contributions point out the evidence of the role of climatic factors in explaining variability and shape of the flood frequency curves. The current status of research in this field concerns the implementation of methods that make practical use of climatic variables in the regional analysis of floods.

In this paper, we demonstrate the practical applicability of a climatic index in driving the choice of homogeneous regions in a regional analysis of floods in Southern Italy. More detailed relations found between this index and L-moments of the samples provide significant clues for the development of integrated physically-based flood distributions (*e.g.* Iacobellis and Fiorentino, 2000) that inherently make use of results of regional analysis without recurring to the index-flood method.

2. PROBABILISTIC MODEL AND PARAMETER ESTIMATION

A number of distributions, characterized by three or more parameters, have been developed for flood frequency analysis, such as the GEV (*Generalized Extreme Value*, Jenkinson**,** 1955), the TCEV (*Two Component Extreme Value*, Rossi et al., 1984) and the Wakeby (Houghton, 1978). These distributions are able, better than traditional distributions (*e.g.* Gumbel), to reproduce the statistical behavior of annual maximum flood data characterized by high positive skewness and, in many cases, by thick tail (Cunnane, 1986). The use of these distributions requires long record of hydrological observations and suggests the adoption of regional statistical analysis.

Remarkable advances have been achieved regarding statistical tools which accompany the use of the above distributions, particularly with regard to parameter estimation procedures. For the purposes of this paper, the GEV distribution will be considered, with parameters estimated through the Probability Weighted Moments (PWM) and L-moments. The General Extreme Value (GEV) distribution is expressed by:

$$F(x) = \exp\left\{-\left[1 - k(x-u)/a\right]^{1/k}\right\} \tag{1}$$

where u and a are respectively the position and the scale parameters while k is a shape parameter. For $k = 0$ equation (1) becomes the Extreme Value type I (EV I or Gumbel), while for $k < 0$ and $k > 0$ the EV type II (Fréchet) and the EV type III (Weibull) are obtained respectively; the lower bound of the former and the upper bound of the latter are obtained by $u+a/k$.

The GEV can also be derived as the distribution of the annual maximum of events (with Poisson-distributed arrival rate) exceeding a fixed threshold, with magnitude following a generalized Pareto distribution. In general, for independent events with probability of non-exceedance $G(x)$ the *cdf* of annual maximum values exceeding a threshold with annual rate Λ_q is expressed by:

$$F(x) = \exp\left\{-\Lambda_q\left[1 - G(x)\right]\right\} \tag{2}$$

When $G(x)$ is a generalized Pareto, equation (2) becomes:

$$G(x) = 1 - \left\{1 - k\left[(x-x_0)/\alpha\right]^{1/k}\right\} \tag{3}$$

that is a GEV distribution, with parameters a and u expressed as function of Λ_q, α and x_0 with relations (Stedinger *et al.*, 1992):

$$u = x_0 + \frac{\alpha(1-\Lambda_q^{-k})}{k}\;; \qquad a = \alpha\,\Lambda_q^{-k} \tag{4}$$

When the variable x represents the annual maximum of flood peaks, Λ_q represents the mean annual number of independent flood events and can be estimated by means of equations (4) as a function of the threshold x_o. When flood peaks are of much higher magnitude than the base flow, as commonly observed in impermeable basins, this threshold can be considered equal to zero.

Probability Weighted Moments (PWM) were defined by Greenwood et al. (1979), as:

$$\beta_r = M_{1,r,o} = E\left[X\{F(X)\}^r\right] \tag{5}$$

Hosking (1986, 1990) expressed the L-moments as a linear combinations of PWM:

$$\lambda_{r+1} = \sum_{k=0}^{r} p^*_{r,k}\beta_k \quad \text{with} \quad p^*_{r,k} = (-1)^{r-k}\binom{r}{k}\binom{r+k}{k} \tag{6}$$

First- and second-order L-moments can be interpreted as a measure of position and of scale, respectively. L-moment ratios:

$$\tau = \lambda_2 / \lambda_1\;; \qquad \tau_r = \lambda_r / \lambda_2\;; \quad r = 3, 4, \ldots \tag{7}$$

have the meaning of coefficient of variation, skewness and kurtosis: τ, τ_3 and τ_4, are also called *L-cv* , *L-ca* and *L-k*.

L-moments unbiased estimators are expressed by:

$$l_{r+1} = \sum_{k=0}^{r} p^*_{r,k} b_k \quad \text{with} \quad b_k = n^{-1}\sum_{j=1}^{n}\frac{(j-1)(j-2)\ldots(j-k)}{(n-1)(n-2)\ldots(n-k)}x_j \tag{8}$$

where x_j, $j = 1,\ldots,n$ is the ordered finite sample and n is the observation length, b_k and l_r are unbiased estimators of respectively β_k and λ_r while $t=l_2 / l_1$ and $t_r=l_r / l_2$ are consistent but not unbiased estimators of τ and τ_r. L-moments may also be estimated by means the plotting positions of the sample (Hosking, 1993).

Ultimately, estimates of GEV parameters are obtained by means of L-moments as:

$$\begin{aligned} k &= 7.8590\,c - 2.9554\,c^2 \\ a &= l_2\,\mathrm{k} / \{\Gamma(1+k)(1-2^{-k})\} \\ u &= l_1 + a\{\Gamma(1+k) - 1\}/k \end{aligned} \tag{9}$$

with $c=2/(3+t_3)-ln(2)/ln(3)$.

The relations above show that the shape parameter k depends only on *L-ca*, the scale parameter a depends on *L-ca* and *L-cv* and the position parameter u depends on *L-ca*, *L-cv* and the mean l_1. As a consequence, in a regional framework it is possible to perform a hierarchical estimation procedure, assuming constant *L-ca* within a first-level homogeneous region and *L-cv* as constant within sub-regions defined at a second-level of application of the procedure (Fiorentino *et al*., 1987, Gabriele and Arnell, 1991).

Application of PWM and L-moments in regional frequency analysis produce robust and accurate quantile estimates (Rossi and Villani, 1992, Vogel, 1993). Regional estimates of a, u and k are obtained by means of equations (9), through the respective regional estimates of L-moment ratios, found as weighted averages of the at-site estimates with weights equal to the recorded sample lengths.

In the following section, the GEV probabilistic model will be applied in a regional context. In this case, selection of homogeneous regions will be supported by a heterogeneity measure suggested by Hosking and Wallis (1993). The measure was defined by application of Monte Carlo simulation performed with the GEV and the four-parameter Kappa parent distributions.

On the other hand, the capability of a climatic index to provide meaningful grouping of stations will be also assessed and the heterogeneity measure by Hosking and Wallis (1993) will be used to analyze the groups obtained in terms of climatic indicators. The representation of basin climatic characteristics is obtained through an index of average water balance, as the Thornthwaite (1948) climatic index, defined as:

$$I = \frac{h - E_p}{E_p} \tag{10}$$

where h is the mean annual rainfall depth and E_p is the mean annual potential evapotranspiration. For the purposes of this paper a very simple E_p formula (Turc, 1961) was considered:

$$E_p = 320 + 25\,t + 0.05\,t^3 \tag{11}$$

which provides annual evapotranspiration E_p in mm based on the average annual temperature t in °C.

3. APPLICATION

Annual maximum flood records, with more than 15 data, related to 22 gauging stations in Southern Italy have been analysed in the context of a regional analysis. Table 1 reports the values of some physical parameters of the related basins as well as the main statistical features of the historical series.

Table 1. Main characteristics of basins and data series analysed. The variable *I* represents the Thornthwaite climatic index. N represents the record length.

#	Station	Area (Km^2)	N	Mean (m^3/s)	Cv	Ca	L-cv	L-ca	L-k	*I*
1	Atella at Ponte sotto Atella	176	45	61	0.57	0.93	0.31	0.22	0.18	0.16
2	Ofanto at Rocchetta S.Antonio	1111	52	457	0.57	0.45	0.33	0.11	0.06	0.17
3	Arcidiaconata at Ponte Rapolla-Lavello	124	32	45	0.64	0.81	0.36	0.18	0.09	-0.04
4	Venosa at Ponte S.Angelo	263	34	56	1.18	2.16	0.53	0.55	0.35	-0.17
5	Carapelle at Carapelle	715	36	284	0.57	1.28	0.30	0.28	0.18	-0.23
6	Cervaro at Incoronata	539	53	216	0.58	0.63	0.33	0.17	0.07	-0.19
7	Celone at S.Vincenzo	92	15	32	0.61	1.14	0.33	0.27	0.23	-0.06
8	Celone at Ponte F.S. Foggia-S.Severo	233	39	46	0.72	2.34	0.34	0.31	0.32	-0.24
9	Vulgano at Ponte Troia-Lucera	94	18	75	0.80	0.38	0.47	0.14	-0.03	-0.09
10	Salsola at Casanova	44	18	46	0.74	1.12	0.40	0.12	0.21	-0.18
11	Casanova at Ponte Lucera-Motta	57	16	27	0.82	1.19	0.44	0.36	0.17	-0.14
12	Salsola at Ponte Foggia-S.Severo	455	40	76	0.54	0.27	0.31	0.08	0.00	-0.27
13	Triolo at Ponte Lucera-Torremaggiore	56	16	35	0.70	0.40	0.41	0.15	-0.04	-0.25
14	S.Maria at Ponte Lucera-Torremaggiore	58	15	18	0.92	0.89	0.51	0.34	0.05	-0.28
15	Bradano at S.Giuliano	1657	17	507	0.79	0.94	0.44	0.23	0.16	-0.17
16	Bradano at Ponte Colonna	462	32	202	0.76	1.15	0.41	0.32	0.10	-0.08
17	Basento at Menzena	1382	24	401	0.63	1.47	0.33	0.28	0.27	0.08
18	Basento at Gallipoli	853	38	353	0.63	2.16	0.31	0.30	0.23	0.28
19	Basento at Pignola	42	28	35	0.43	1.06	0.23	0.25	0.21	0.70
20	Agri at Tarangelo	511	32	189	0.38	0.71	0.22	0.15	0.13	0.47
21	Sinni at Valsinni	1140	22	555	0.56	2.25	0.27	0.35	0.31	0.57
22	Sinni at Pizzutello	232	19	262	0.25	0.70	0.14	0.18	0.15	1.26

L-moment ratios were computed for all stations: Figure 1 displays the comparisons of at-site estimates of *L-cv*, *L-ca* and *L-k*, that can suggest considerations useful for the regional analysis (Hosking and Wallis, 1993).

Estimated L-moments are highly scattered and the dispersion is mainly due to sample variability. The only station which behaves differently is that of Venosa at Ponte Sant'Angelo, identified by the point #4. We do not have enough information to explain the reasons of this specific outcome. However, to better address the objectives of this analysis we have excluded the station #4 from the estimating procedure of regional parameters.

Looking at the graphs in Figure 1 it is also possible to suppose the existence of a wide homogeneous region with reference to the L-ca moment at the first regionalization level. On the other hand, with regard to the second order L-moment, L-cv, we can assume the existence of two homogeneous sub-regions, as delineated by the horizontal line in Figure 1.

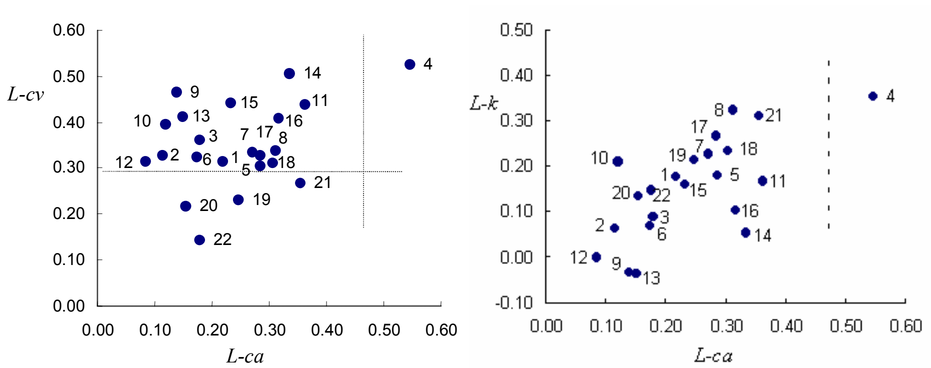

Figure 1. Diagram of L-moment ratios.

This grouping does not look particularly intuitive, unless one uses considerations related to the average climate of the observed basins. In particular, basins with L-cv greater then 0.3, which constitute the first subregion, are placed in the North-western sector of the geographic region under study and are characterized by climate ranging from semiarid to humid-subhumid. The last four basins, placed in the South-eastern sector, have more humid climate and present dense vegetation and high annual rainfall.

The above hypotheses for definition of homogeneous regions were then verified by means of the heterogeneity measure H_i by Hosking and Wallis (1993). Monte Carlo simulation was performed with both GEV and four-parameter kappa parent distributions (see also Iacobellis et al., 1997) with results shown in Table 2. Using a GEV parent and deriving H_i for moments of order $i-1$ (*e.g.* variance $\rightarrow$ $i=1$), we obtained values of $H_i < 1$, indicating homogeneous regions, for all parameters in both sub-regions.

Table 2. Heterogeneity measures for the two sub-regions examined.

sub-region	H_1	H_2	H_3	H_1	H_2	H_3
	parent kappa			*parent* GEV		
Semi-arid	1.81	1.17	1.84	**0.83**	**0.21**	**0.81**
Humid	**0.81**	**-0.23**	**-0.60**	**0.97**	**-0.14**	**-0.45**

The GEV parameters on the two sub-regions, estimated with a hierarchical procedure by means of equations 9, share the regional ***k*** value (obtained at the first level):

$$\boldsymbol{k} = -0.078$$

while the two independent regional set of values ***a*** and ***u*** (second level) were estimated as:

$\boldsymbol{a} = 0.47$	$\boldsymbol{u} = 0.69$	semi-arid sub-region
$\boldsymbol{a} = 0.29$	$\boldsymbol{u} = 0.81$	humid sub-region

The two regional GEV curves are displayed in Figure 2 along with the observed data. From the figure it is worth noting that the CDF of the semi-arid zone is steeper then that of the humid zone, consistently with what observed by Farquharson *et al.* (1992), and that the scale parameter is quite close to the value 0.5, indicated as representative of semi-arid basins in the world.

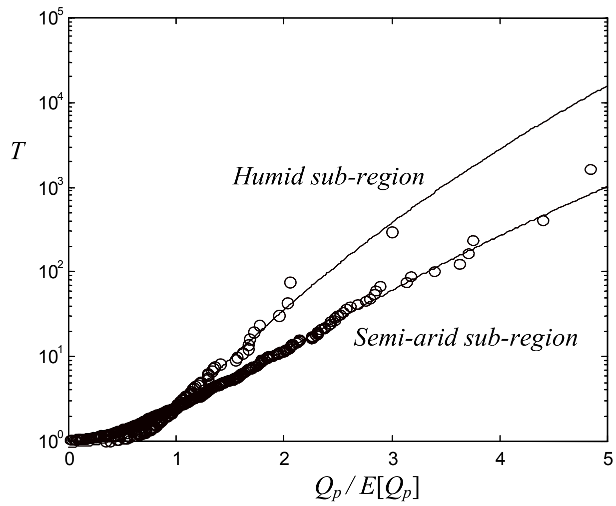

Figure 2. Regional dimensionless GEV CDF's for the two regions considered. T is the return period, in years

A deeper view of relations between climate and flood parameters is shown in Figure 3, where it is possible to observe a clear trend between dryness, represented by I, and the estimated values of L-cv. A process-type interpretation of this trend can be given through representation of at-site estimates of Λ_q

versus the index I (Figure 4). Estimates of Λ_q were obtained at site through equations (9) and (4) but using $\boldsymbol{k} = -0.078$ obtained in the first level of the regional analysis

Figure 4 shows clearly that dry basins (I>0) do not present significant variability of the second moment and tend to support a homogeneous-region approach. Conversely, humid basins show a marked trend with L-cv and the average number of flood events. This constitute the basis for justifying additional efforts in trying to link the variability of second-order parameters to the physical characteristics of the basins.

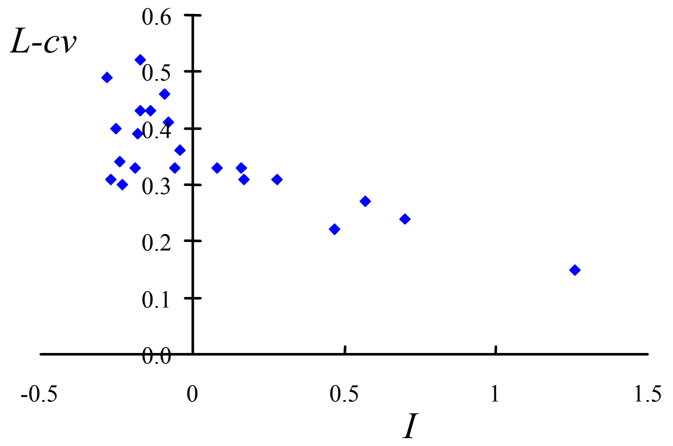

Figure 3. L-cv versus climatic index for all stations considered.

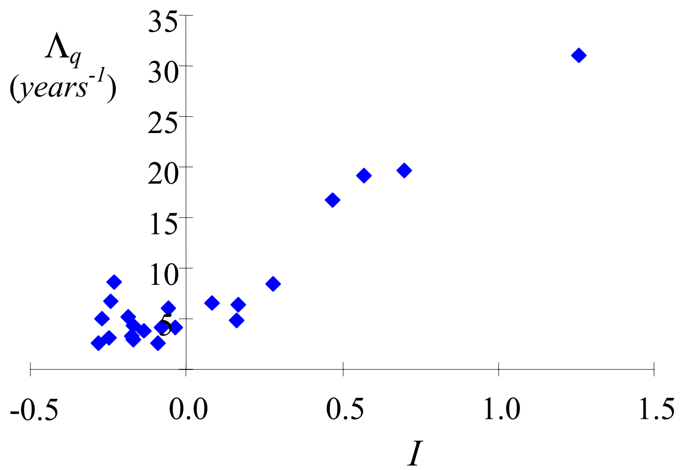

Figure 4. Mean annual number of flood events Λ_q versus climatic index.

4 FINAL REMARKS

The influence of climate in helping to select homogeneous regions for flood frequency analysis has been demonstrated here, using an efficient homogeneity measure proposed by Hosking and Wallis (1993) as the testing procedure. This result can have an interesting impact as a method that reduces the subjective judgment in the application of the index-flood approach to regional analysis. An additional outcome of the application shown in this paper concerns the attitude of the climatic index to cover a large range of values and to explain, at least in humid climates, the variability of the mean annual number of flood events, as well as of the second-order L-moment. Based on these results, the use of a climatic index (as a rough indicator of the average soil wetness) can represent an interesting starting point for introducing physically-based models of the spatial variability of parameters in the field of regional flood frequency analysis.

Acknowledgements

This work was supported by funds granted by MURST and CNR–GNDCI (project "VaPi-RiVers").

REFERENCES

Burn, D.H., Catchment similarity for regional flood frequency analysis using seasonality measures, *Jour. Hydrol.*, **202,** 212-230, 1997.

Cunnane, C., Review of statistical models for flood frequency estimation, *Proc. of the Int. Symp. on Flood Frequency and Risk Analysis*, Louisiana State University, Baton Rouge, U.S.A., 1986.

Farquharson, F. A., J. R. Meigh, and J. V. Sutcliffe, Regional flood frequency analysis in arid and semi-arid areas, *Jour. Hydrol.*, 138, 487-501, 1992.

Fiorentino, M., S. Gabriele, F. Rossi, and P. Versace, Hierarchical approach for regional flood frequency analysis, in V. P. Singh (eds), *Regional flood frequency analysis*, 35-49, D. Reidel, Norwell, Mass, 1987.

Gabriele, S. and N. Arnell, A hierarchical approach to regional flood frequency analysis, *Water Resources Research*, 27(6), 1281-1289, 1991.

Greenwood, J. A., J. M. Landwehr, N. C. Matalas, and J. R. Wallis, Probability weighted moments: Definition and relation to parameters of several distributions expressable in inverse form, *Water Resour. Res.*, 15(5), 1049-1054, 1979.

Hosking, J. R. M., L-moments: Analysis and estimation of distributions using linear combinations of order statistics, *J. R. Stat. Soc.,* **B**, 52, 105-124, 1990.

Hosking, J. R. M., and J. R. Wallis, Some Statistic usefol in Regional Frequency Analysis, *Water Resources Research*, 29 (2), 271-281, 1993.
Hosking, J. R. M., *Fortran routines for use with the method of L-moments, version 3*, Res. Rep. RC20525, IBM Res., Yorktown Heights, N. Y., 1996.
Houghton, J. C., Birth of a parent: The Wakeby distribution for modeling flood flows, *Water Resources Research*, 14(6), 1105-1110, 1978.
Iacobellis V., P. Claps, G. Gioia and M. Fiorentino, An application of flood regional analysis in a wide, climatically heterogeneous, region, Proc. of 'Giornate di studio in onore del Prof. E. Orabona', (in italian), BIOS, 1997.
Iacobellis, V., P. Claps, and M. Fiorentino, On the dependence of flood distribution parameters from climate, in *Proc. XIV Conf di Idraulica e Costruzioni Idrauliche* (in Italian), vol II, 213-224, CUECM, Italy, 1998.
Iacobellis, V., and M. Fiorentino, Derived distribution of floods based on the concept of partial area coverage with a climatic appeal, *Water Resources Research*, **36**(2), 469-482, 2000.
Institute of Hydrology, *Flood Estimation Handbook,* Wallingford, Oxfordshire, UK, 1999.
Jenkinson, A. F., The frequency distribution of the annual maximum (or minimum) of meteorological elements, *Q. J. R. Meteorol. Soc. London*, 81, 158-171, 1955.
Klemes, V., Hydrological and engineering relevance of flood frequency analisys, Proc. Int. Symp. *on Flood Frequency and Risk Analysis*, Application of Frequency and Risk in Water Resources, Reidel, Dordrecht, 1-18, 1987.
Lettenmaier, D. P., J. R. Wallis, and E. F. Wood, Effect of Regional Heterogeneity on Flood Frequency Estimation, *Water Resources Research*, Vol. 23, N. 2, pp. 313-323, 1987.
NERC - Natural Environment Research Council; *Flood studies report,* Vol. I - Hydrologic studies, NERC, London, 1975.
Rossi, F., M. Fiorentino, and P. Versace, Two component extreme value distribution for flood frequency analysis, *Water Resources Research*, 20(7), 847-856, 1984.
Rossi, F., and P. Villani, Regional methods for flood estimation, in G. Rossi (eds.) *Coping with floods*, NATO-ASI Series, Kluwer, 135-169, 1992.
Stedinger, J.R., R.M. Vogel and E. Foufoula-Georgiou, Frequency analysis of extreme events. In Maidment D.R. (Ed.) *Handbook of Hydrology*, Cap. 18, McGraw-Hill, 1992.
Turc, L., Estimation of irrigation water requirements, potential evapotranspiration: a simple climatic formula evolved up to date, *Annals of Agron.*, 12, 13-14, 1961.
Vogel, R. M., W. O. Thomas Jr., and T. A. McMahon, Flood-flow frequency model selection in southwestern United States, *Jour. of Water Resour. Plann. and Manag., ASCE,* 119(3), 353-366, 1993.