Un’applicazione di analisi regionale delle piene in un’ampia regione climaticamente eterogenea

Iacobellis V.[1,2], Claps P.[1], Fiorentino M.[1], Gioia G.[2]
[1] *Dipartimento di Ingegneria e Fisica dell’Ambiente - Università della Basilicata*
[2] *Dipartimento di Ingegneria delle Acque - Politecnico di Bari*

SOMMARIO

Vengono affrontati alcuni problemi relativi alla determinazione della distribuzione di probabilità delle portata di piena nel contesto di un’analisi regionale. In particolare si sono applicati alcuni metodi recentemente proposti in letteratura che si innestano in fasi fondamentali del metodo della piena indice. Attualmente, infatti, nessuno dei modelli disponibili in letteratura può essere considerato chiaramente superiore agli altri ed universalmente accettato. Per l'applicazione delle procedure di stima vengono utilizzate le serie dei massimi annuali disponibili in una regione molto ampia, comprendente la Basilicata e la Puglia, caratterizzata da significative differenze climatiche al suo interno. L’analisi mostra come anche le tecniche proposte più di recente lasciano aperti ampi margini di arbitrarietà in alcuni passaggi fondamentali della procedura, quali ad esempio la scelta del numero di zone e sottozone omogenee, la definizione delle stesse o l’analisi di affidabilità dei dati disponibili.

1. INTRODUZIONE

La valutazione del rischio di piena rimane a tutt’oggi un problema aperto ove si consideri l’assenza di una metodologia universalmente accettata per il calcolo delle portate di piena a fissato tempo di ritorno.

Tale indeterminazione si spiega plausibilmente con la mancanza di un supporto teorico fisicamente basato che giustifichi forma e variabilità delle distribuzioni statistiche adottate. Bobée et al. [3] stigmatizzano tale situazione sottolineando che fra tutte le procedure ideate negli ultimi 40 anni, la mancanza di solide basi teoriche e l’impossibilità di individuare un metodo chiaramente superiore agli altri ha determinato un inevitabile distacco tra ricerca e pratica progettuale. Tale problema viene peraltro affrontato in un rapporto del WMO [39], in cui viene fornita una panoramica dettagliata dei metodi esistenti in letteratura, e delle relative prestazioni senza tuttavia riuscire a porre una parola definitiva in questo dibattito.

Il regime delle portate, come evidenziano Moon e Lall [29], dipende da un’ampia varietà di condizioni climatiche e geomorfologiche dei bacini interessati. Le serie dei massimi annuali possono essere condizionate da fattori che caratterizzano in maniera differente la coda della distribuzione. Klemes [26] attribuisce tali differenze ai diversi meccanismi che sono all'origine degli eventi di piena: precipitazioni di diversa natura, scioglimento delle nevi e concomitanza di tali fenomeni; queste difformità hanno provocato il proliferare di leggi di distribuzione di probabilità e di metodi di stima. Haktanir [20] in un'analisi comparativa di varie distribuzioni di frequenza coniugate con diversi metodi di stima dei parametri su dati osservati in Anatolia arriva a contare 29 diversi modelli.

La carenza del supporto teorico viene inoltre spesso esasperata dalla necessità di valutare il rischio relativo a portate il cui tempo di ritorno eccede di gran lunga la numerosità delle serie osservate. In altre parole le sole tecniche statistiche inferenziali non assicurano una corretta determinazione della coda della distribuzione in corrispondenza della quale i dati disponibili sono ovviamente scarsi. Gioia e Lo Presti [16] sottolineano la scarsa potenza dei classici test di adattamento.

Anche con riferimento ai metodi di stima, il metodo della massima verosimiglianza, asintoticamente efficiente per la stima dei parametri, perde in accuratezza se utilizzato su campioni di modesta dimensione [2]. Tutto ciò ha determinato la nascita di altri metodi di stima quali quelli dei momenti pesati in probabilità [17] e degli L-moments [21], mentre altri metodi, quali quello dei "momenti misti" [32, 31] e dei "momenti generalizzati" [1], offrono prestazioni decisamente inferiori.

In questo contesto hanno trovato rapido sviluppo le tecniche di regionalizzazione atte a compensare la scarsità dei dati disponibili mediante l'analisi di campioni creati accorpando le osservazioni effettuate all'interno di regioni statisticamente omogenee.

Tali metodologie hanno consentito l'adozione di distribuzioni ad almeno tre parametri quali la GEV (Generalized Extreme Value) [25], la TCEV (Two Component Extreme Value) [34, 35], la Wakeby [24] che, come mostrato da Cunnane [6], sono le uniche che consentono di ben riprodurre le caratteristiche statistiche delle serie storiche dei massimi annuali delle portate di piena. Esse hanno, però, contestualmente determinato l'insorgenza di nuovi problemi connessi ad esempio all'individuazione delle regioni omogenee, nonché all'inefficienza ed alla distorsione nella valutazione dei momenti di ordine superiore al secondo da campioni di scarsa numerosità.

Il GREHYS [18] in una recensione dei metodi attualmente più accreditati in letteratura, [ROI ("Region of Influence"); analisi delle correlazioni canoniche; analisi di corrispondenza e clusterizzazione ascendente gerarchica; metodo della portata indice mediante GEV-PWM; analisi regionali non parametriche, portata indice mediante Pareto Generalizzata, portata indice mediante Esponenziale Generalizzata, metodi regressivi, ecc.], osserva che in tutti questi metodi é necessario effettuare delle scelte soggettive e che le relative prestazioni sono comunque fortemente condizionate dalla particolare situazione in esame e dal tipo di dati disponibili, risulta quindi impossibile giungere a conclusioni di carattere generale. Brath *et al.* [4] osservano che le tecniche di regionalizzazione basate sulla distribuzione campionaria di coefficiente di variazione e coefficiente di asimmetria spesso inducono ad accettare l'ipotesi di omogeneità anche su vaste zone climaticamente eterogenee, mentre, i metodi basati sull'analisi multivariata dei dati (clusterizzazione, analisi discriminante) producono spesso risultati geograficamente e fisicamente incongruenti.

Comincia a farsi strada, di recente, l'idea che notevole influenza sulla diversità tra le diverse distribuzioni di probabilità che si osservano nella realtà, possa essere esercitata dal clima. Ad esempio Farquharson *et al.* [9] hanno analizzato dati relativi a zone aride e semi-aride di tutto il mondo (162 stazioni con pioggia totale annua minore di 600 mm in Africa, Iran, Giordania, Arabia Saudita, Russia, USA), confrontandoli con altri provenienti dalla Gran Bretagna, relativi a clima umido e temperato. Essi rilevano una significativa analogia tra le curve regionali delle zone aride, tutte caratterizzate da elevato coefficiente di variazione e forte asimmetria, a differenza delle curve ricostruite sui dati europei.

In ogni caso, ad oggi, le migliori stime possono essere ottenute solo facendo ricorso a metodi di analisi regionale. Tra questi, il metodo della piena indice [7, 30], che ne è il capostipite, considera omogeneo il comportamento delle serie storiche ottenute normalizzando i dati osservati mediante la media locale delle osservazioni. Tale ipotesi implicitamente assume che nelle regioni omogenee siano ovunque costanti i momenti di ordine superiore al primo. Questo metodo viene da alcuni considerato poco affidabile a causa della elevata variabilità del coefficiente di variazione [8, 27].

In particolare, Gupta e Dawdy [19] e Robinson e Sivapalan [33], studiando il comportamento dei picchi di piena in funzione delle caratteristiche geomorfologiche dei territori ed in particolare all'area del bacino, osservano che in generale i momenti delle portate di piena, hanno, in relazione all'area del bacino, un comportamento multifrattale e non scala invariante come implicitamente assunto nel metodo della piena indice. Al contrario Rosso *et al.* [37] affermano che i modelli di invarianza di scala semplice e multipla possono invece diventare parte integrante del metodo della piena indice come criterio per la individuazione delle zone omogenee. Infatti si possono considerare tali le regioni in cui è possibile osservare un comportamento monofrattale delle portate di picco rispetto all'area del bacino.

Nell'ambito dei metodi regionali, intanto, sia in Italia che all'estero hanno preso quota criteri di stima delle distribuzioni regionali di natura gerarchica, in cui si tiene conto che i diversi parametri caratteristici dei campioni osservati hanno differente variabilità spaziale [13]. In particolare, si assume che il coefficiente di asimmetria delle serie dei massimi annuali sia costante su aree più grandi rispetto al coefficiente di variazione che, a sua volta, varia in misura inferiore rispetto alla media delle osservazioni.

Questi modelli si avvalgono anche, con particolare riferimento a quelli basati sull'uso delle distribuzioni derivate dai modelli di *Poisson composti*, quali la TCEV, del significato fisico del parametro di dispersione, fortemente legato al numero medio di eventi di piena indipendenti. Ciò aiuta nella definizione delle regioni omogenee. Il modello di regionalizzazione basato sull'uso della TCEV è stato adottato in Italia per la redazione delle guide tecniche alla valutazione delle piene elaborate dal Gruppo Nazionale per la Difesa dalle Catastrofi Idrogeologiche (v. es: [5])

Esiste un'ampia letteratura sulla definizione delle prestazioni di distribuzioni di probabilità. In particolare, scelto un modello costituito dalla coppia di legge di distribuzione e metodo di stima dei parametri, tale modello si dice robusto, se produce stime dei quantili con bassa distorsione e piccolo errore standard anche se il campione osservato appartiene ad una popolazione diversa da quella assunta [6].

Le proprietà statistiche degli stimatori regionali basati sulla TCEV mostrano che i modelli regionali, se accompagnati da una buona definizione delle regioni omogenee, consentono di pervenire a stime di modesta incertezza [10, 12, 15]. Proprietà statistiche del tutto confrontabili vengono anche ottenute con i modelli di regionalizzazione basati sull'uso della GEV. In tale ottica già Lettenmaier *et al.* [28] e Haktanir [20], hanno evidenziato le buone prestazioni ottenute, con parametri stimati tramite i momenti pesati in probabilità (PWM), in termini di robustezza statistica. Al medesimo risultato sono giunti Gabriele e Arnell [14] comparando le prestazioni dei modelli GEV-PWM con TCEV-ML (Massima Verosimiglianza) e TCEV-PWM.

Il modello di analisi regionale utilizzato è dunque quello basato sulla GEV; ciò anche allo scopo di fornire elementi per un utile confronto con la procedura ufficialmente adottata dal GNDCI, basata sulla TCEV. In questo lavoro vengono inoltre approfondite alcune questioni legate all'analisi regionale, quali la scelta delle regioni omogenee e l'influenza del clima e della dimensione dei bacini idrografici sulla distribuzione di probabilità delle piene. A tal fine vengono utilizzate le serie dei massimi annuali disponibili in una regione molto ampia, comprendente la Basilicata e la Puglia, caratterizzata da significative differenze climatiche al suo interno.

2. MODELLO PROBABILISTICO E METODI DI STIMA

L'espressione della GEV è :

$$F(x) = exp\left[-\left\{1-k(x-u)/a\right\}^{1/k}\right] \qquad k \neq 0 \tag{1}$$

$$F(x) = exp\left[-exp\left\{-(x-u)/a\right\}\right] \qquad k = 0 \tag{2}$$

in cui u ed a sono rispettivamente parametro di posizione e scala mentre k è parametro di forma. Per $k = 0$ si ottiene la distribuzione asintotica del massimo valore del I tipo (EV I) nota anche come legge di Gumbel, mentre rispettivamente per $k < 0$ e $k > 0$ si ottengono una EV II (Fréchet) ed una EV III (Weibull) e la distribuzione risulta rispettivamente limitata superiormente o inferiormente da $u+a/k$.

I PWM sono definiti, nella loro espressione più generale, come:

$$M_{p,r,s} = E\left[X^p\{F(X)\}^r\{1-F(X)\}^s\right] \tag{3}$$

dove $F(X)$ è la funzione di probabilità cumulata di x. Di fatto le espressioni che più frequentemente si usano sono riferite ad un solo indice (r):

$$\beta_r = M_{1,r,o} = E\left[X\{F(X)\}^r\right] \tag{4}$$

In alternativa ai PWM, Hosking [21] suggerisce l'uso degli L-moments, che sono definiti come:

$$\lambda_r = E\left[XP^*_{r-1}\{F(X)\}\right] \tag{5}$$

dove $P^*_r(.)$ è l'operatore polinomio di Legendre. Gli L-moments si possono esprimere anche come combinazioni lineari dei PWM:

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

mentre i rapporti tra gli L-moments si definiscono come :

$$\tau_r = \lambda_r / \lambda_2\,; \quad r = 3, 4, ... \tag{7}$$

Gli L-moments possono essere stimati tramite le:

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

ove x_j, per $j = 1,...,n$ è il campione ordinato in senso crescente ed n è il numero di dati del campione; b_r e l_r sono stimatori indistorti di β_r e λ_r, mentre $t=l_2 / l_1$ e $t_r=l_r / l_2$ sono stimatori consistenti ma non indistorti di τ e τ_r.

Gli L-moments possono anche essere stimati usando dei PWM calcolati in base alla frequenza cumulata (plotting position) dei dati costituenti il campione [23]:

$$l_r = n^{-1} \sum_{j=1}^{n} P_{r-1}^{*}\left(p_{j:n}\right) x_{j:n} \tag{9}$$

ove $p_{j:n}$ è il vettore delle plotting position del campione ordinato $x_{j:n}$. Questi ultimi stimatori, nel seguito contrassegnati da un asterisco (es. *L-ca*),* pur non presentando prestazioni generalmente brillanti, sono particolarmente raccomandati per la stima dei parametri della GEV, quando la plotting position è espressa come:

$$p_{j:n} = \frac{j - 0.35}{n} \tag{10}$$

Gli L-moments forniscono ottime prestazioni in termini di robustezza e accuratezza della stima dei parametri specialmente in ambito di regionalizzazione, al pari dei PWM dai quali sono derivati [36, 38]. Rispetto ad essi inoltre offrono il vantaggio di una più facile interpretazione statistica come misura dei parametri di forma delle distribuzioni. In particolare, λ_1 è la media, ovvero una misura di posizione, λ_2 é una misura di scala, τ_3 e τ_4 (*L-ca* ed *L-k*) sono rispettivamente misure di asimmetria e kurtosi, τ (*L-cv*) é l'analogo del coefficiente di variazione.

In termini di L-moments i parametri della GEV possono essere stimati come:

$$\begin{aligned} &\mathrm{c} = 2/\left(3+\mathrm{t}_3\right) - \ln(2)/\ln(3) \\ &\mathrm{k} = 7.8590\mathrm{c} - 2.9554\mathrm{c}^2 \\ &\mathrm{a} = \mathrm{l}_2\ \mathrm{k} / \left\{\Gamma(1+\mathrm{k})\left(1-2^{-\mathrm{k}}\right)\right\} \\ &\mathrm{u} = \mathrm{l}_1 + \mathrm{a}\left\{\Gamma(1+\mathrm{k}) - 1\right\}/\mathrm{k} \end{aligned} \tag{11}$$

in cui si nota che *k* (il parametro di forma) dipende solo da *L-ca*, *a* (parametro di scala) dipende da *L-ca* ed *L-cv*, *u* (parametro di posizione) da *L-ca*, *L-cv* e dalla media delle osservazioni l_1.

Per la stima dei valori regionali ***a***, ***u*** e ***k*** si adottano nelle (11) le medie pesate delle stime, pesate sulla dimensione del campione, dei rapporti τ, τ_3 e τ_4 tra gli L-moments. Si determina così la distribuzione regionale, da cui si ricavano le distribuzioni nei diversi siti riscalando i quantili rispetto al valore medio locale.

Il metodo regionale così definito è già stato ampiamente applicato sia nella formulazione originale, sia in maniera gerarchica (v. es. [14]).

3. DETERMINAZIONE DELLE ZONE OMOGENEE

Tenuto conto dello scopo esplorativo di questo lavoro, si farà riferimento, per la definizione delle zone omogenee, piuttosto che a metodi consolidati comunemente in uso in Italia [11, 36], a tecniche che si possono considerare più sperimentali.

Sulla base degli L-moments, Hosking e Wallis [22] individuano vari criteri di analisi tra cui una misura di eterogeneità per l'identificazione delle regioni omogenee. Nel seguito ci soffermeremo su questa e sul metodo suggerito da Brath *et al.*[4], derivato dalla teoria di Gupta e Dawdy [19], sulla invarianza di scala semplice e multipla delle portate di piena in relazione all'area del bacino sotteso. Questi due criteri per la scelta delle regioni omogenee saranno applicati al caso di studio. Se ne premette una sintetica descrizione rimandando ai lavori citati per una più completa esposizione.

In una regione omogenea, per definizione, i valori teorici degli L-moments sono costanti in tutti i siti, i valori osservati saranno tuttavia diversi a causa della variabilità campionaria. La varianza campionaria di *L-cv*, ritenuta da Hosking e Wallis il parametro più significativo ai fini della individuazione delle regioni omogenee, viene calcolata ancora come media pesata rispetto al numero di osservazioni nelle *N* diverse stazioni:

$$V_1 = \sum_{i=1}^{N} n_i \left(t^{(i)} - \bar{t}\right)^2 / \sum_{i=1}^{N} n_i \qquad (12)$$

Mediante la generazione ripetuta di dati appartenenti ad una regione omogenea, con serie generate di lunghezza pari a quelle osservate, si possono calcolare media e deviazione standard di *V* e, conseguentemente, l'indice:

$$H_1 = \left(V_1 - \mu_V\right) / \sigma_V \qquad (13)$$

Confrontando la dispersione osservata con quella simulata H_1 misura il grado di eterogeneità in un gruppo di siti osservati e verifica che essi possano essere considerati appartenenti ad una regione omogenea. Hosking e Wallis suggeriscono i valori di $H<1$ come campo di accettazione dell'ipotesi di regione omogenea, $1<H<2$ per regioni possibilmente eterogenee, $H>2$ per regioni certamente eterogenee.

A tale criterio tuttavia non si conferisce la validità di un test dell'ipotesi statistica in quanto la valutazione del livello di significatività sarebbe inficiata dall'assenza nelle serie generate della correlazione presente nelle serie osservate e dalla incertezza nella scelta della *parent* ovvero della distribuzione "vera", da utilizzare nelle generazioni.

Hosking e Wallis suggeriscono per la generazione dei dati una legge tipo kappa [22, 23] a 4 parametri, di espressione:

$$F(x) = \left[1 - h\left\{1 - k(x-u)/a\right\}^{1/k}\right]^{1/h} \qquad (14)$$

legge che include come casi particolari la GL (Generalized Logistic) per $h = -1$, la GEV per $h \rightarrow 0$ e la GP (Generalized Pareto) per $h = 1$. Nel caso applicativo discusso più avanti noi utilizzeremo come *parent* sia una GEV sia una legge kappa.

Un indice analogo ad H_1, che chiameremo H_j ($j = 2,3$), viene inoltre utilizzato anche sulle misure di dispersione relative agli *L-ca* ed *L-k*, definendo le varianze:

$$V_2 = \sum_{i=1}^{N} n_i \left[\left(t^{(i)} - \bar{t}\right)^2 + \left(t_3^{(i)} - \bar{t_3}\right)^2\right]^{0.5} / \sum_{i=1}^{N} n_i \qquad (15)$$

$$V_3 = \sum_{i=1}^{N} n_i \left[\left(t_3^{(i)} - \bar{t_3}\right)^2 + \left(t_4^{(i)} - \bar{t_4}\right)^2\right]^{0.5} / \sum_{i=1}^{N} n_i \qquad (16)$$

Come già evidenziato da Zrinji e Burn [40] tali grandezze si prestano naturalmente per la regionalizzazione in senso gerarchico.

Nella teoria formulata da Gupta e Dawdy [19] la proprietà di invarianza di scala multipla o comportamento multifrattale della portata di piena massima annuale in funzione dell'area del bacino, si sostiene che i quantili a tempo di ritorno assegnato (q_T), variano con l'area del bacino sotteso (A) con leggi di potenza del tipo:

$$q_{T,A} = q_{T,1} A^{h(T)} \qquad (17)$$

ove $q_{T,1}$ è il quantile relativo ad un bacino di area unitaria.

Tale proprietà, tradotta in termini di momenti centrali ordinari della portata di piena, diventa per il generico momento di ordine *k*:

$$E\left[\left(q_A\right)^k\right] = A^{n(k)}E\left[\left(q_1\right)^k\right] \tag{18}$$

dove E[.] è l'operatore valore atteso, q_A e q_1 sono le portate relative rispettivamente ad un bacino di area A e di area unitaria, *n(k)* , detta funzione di dispersione, è una funzione monotona in *k*. In conseguenza di ciò i parametri che dipendono dai momenti *cv* e *ca*, assumono a loro volta andamento variabile in funzione dell'area del bacino, a meno che la funzione di dispersione non assuma andamento lineare con k. Si ha cioè comportamento monofrattale quando *n(k) = kn(1).*

In tal caso, in termini di PWM si ha:

$$\beta_k = \beta_{k,1}A^m \tag{19}$$

E' evidente che nell'ipotesi di comportamento monofrattale delle portate di piena i parametri *cv* e *ca*, nonché *L-cv* ed *L-ca*, sono costanti e indipendenti dall'area, pertanto le zone caratterizzate da tale comportamento possono considerarsi omogenee ai fini dell'analisi regionale.

4. APPLICAZIONE

Sono state analizzate le serie storiche dei massimi annuali dei colmi di piena registrati in 22 stazioni con almeno 15 anni di osservazione, di cui 14 in Puglia e 8 in Basilicata (tab. I). Altre stazioni in cui sono state riscontrate forti inconsistenze nei dati [5] sono state escluse dall'analisi. Le statistiche dei dati osservati sono riportate in tabella II.

I grafici di figura 1 mostrano i valori campionari di *L-cv*, *L-ca* ed *L-k* stimati mediante le (8). Usando la (9) si ottengono risultati simili che in questa applicazione sono stati utilizzati per la stima dei parametri della GEV.

Il calcolo della misura di eterogeneità, effettuato mediante generazione di dati distribuiti secondo GEV e secondo una legge kappa a 4 parametri, contestualmente all'analisi dei grafici di figura 1, porta ad escludere dalle successive elaborazioni la stazione sul Venosa a Ponte Sant'Angelo. Ulteriori indagini sui dati osservati in questa stazione hanno confermato l'impressione che vi possano essere stati problemi di misura, ci si riserva tuttavia di effettuare ulteriori approfondimenti.

Le misure di eterogeneità calcolata sulle 21 restanti stazioni forniscono:

H_1 = 5.04, H_2 = 0.73 e H_3 = 0.43 con *parent* GEV.

H_1 = 4.35, H_2 = 0.72 e H_3 = -0.01 con *parent* kappa.

L'indice H_1, che dipende solo da *L-cv,* è sempre superiore ad 1, denotando una marcata eterogeneità e quindi l'impossibilità di individuare una regione omogenea al II livello. H_2 , ed H_3, invece, assumono sempre valori minori di 1, suggerendo la possibile esistenza di una regione omogenea al I livello, ovvero in termini di *L-ca* ed *L-k*.

Viene pertanto formulata l'ipotesi di suddividere le 21 stazioni in due sottoregioni come rappresentato con linee tratteggiate in figura 1. Altre suddivisioni potrebbero essere prese in considerazione sulla base dei diagrammi osservati, anche se per i valori di *L-cv* maggiori di 0.3 i dati relativi alle stazioni del Bradano e medio-basso Basento sembrano bene integrati con quelli delle stazioni pugliesi, individuando una zona che denomineremo semi-arida. Al di sotto della soglia indicata si raccolgono le stazioni relative ai bacini dell'Agri, del Sinni e dell'alto Basento, che si possono considerare parte di una zona più ricca di vegetazione e di precipitazioni, che designeremo come zona umida.

Codice	Stazione	Codice	Stazione
1	Atella a P.te sotto Atella	12	Salsola a P.te Foggia-S.Severo
2	Ofanto a Rocchetta S.Antonio (Monteverde Sc.)	13	Triolo a P.te Lucera-Torremaggiore
3	Arcidiaconata a P.te Rapolla Lavello	14	Canale S.Maria a P.te Lucera-Torremagg.
4	Venosa a P.te S.Angelo	15	Bradano a S.Giuliano
5	Carapelle a Carapelle	16	Bradano a P.te Colonna
6	Cervaro ad Incoronata	17	Basento a Menzena
7	Celone a S.Vincenzo	18	Basento a Gallipoli
8	Celone a P.te F.S. Foggia-S.Severo	19	Basento a Pignola
9	Vulgano a P.te Troia-Lucera	20	Agri a Tarangelo
10	Salsola a Casanova	21	Sinni a Valsinni
11	Casanova a P.te Lucera-Motta	22	Sinni a Pizzutello

Tab. I. Stazioni osservate

codice	A (Km^2)	*n*	*media*	*std*	*cv*	*ca*	*L-cv*	*L-ca*	*L-k*	*L-cv**	*L-ca**	*L-k**
1	176	45	61	35	0.57	0.93	0.31	0.22	0.18	0.31	0.22	0.18
2	1111	52	457	260	0.57	0.45	0.33	0.11	0.06	0.33	0.12	0.08
3	124	32	45	29	0.64	0.81	0.36	0.18	0.09	0.36	0.19	0.11
4	263	34	56	66	1.18	2.16	0.53	0.55	0.35	0.52	0.53	0.33
5	715	36	284	161	0.57	1.28	0.30	0.28	0.18	0.30	0.29	0.19
6	539	53	216	125	0.58	0.63	0.33	0.17	0.07	0.33	0.18	0.08
7	92	15	32	19	0.61	1.14	0.33	0.27	0.23	0.33	0.27	0.22
8	233	39	46	33	0.72	2.34	0.34	0.31	0.32	0.34	0.31	0.31
9	94	18	75	60	0.80	0.38	0.47	0.14	-0.03	0.46	0.16	0.01
10	44	18	46	34	0.74	1.12	0.40	0.12	0.21	0.39	0.15	0.19
11	57	16	27	22	0.82	1.19	0.44	0.36	0.17	0.43	0.35	0.18
12	455	40	76	41	0.54	0.27	0.31	0.08	0.00	0.31	0.10	0.03
13	56	16	35	25	0.70	0.40	0.41	0.15	-0.04	0.40	0.18	0.02
14	58	15	18	17	0.92	0.89	0.51	0.34	0.05	0.49	0.33	0.09
15	1657	17	507	402	0.79	0.94	0.44	0.23	0.16	0.43	0.24	0.16
16	462	32	202	153	0.76	1.15	0.41	0.32	0.10	0.41	0.32	0.12
17	1382	24	401	252	0.63	1.47	0.33	0.28	0.27	0.33	0.28	0.26
18	853	38	353	221	0.63	2.16	0.31	0.30	0.23	0.31	0.30	0.23
19	42	28	35	15	0.43	1.06	0.23	0.25	0.21	0.24	0.25	0.22
20	511	32	189	73	0.38	0.71	0.22	0.15	0.13	0.22	0.16	0.15
21	1140	22	555	308	0.56	2.25	0.27	0.35	0.31	0.27	0.34	0.30
22	232	19	262	66	0.25	0.70	0.14	0.18	0.15	0.15	0.18	0.21

Tab. II. Statistiche relative alle serie osservate

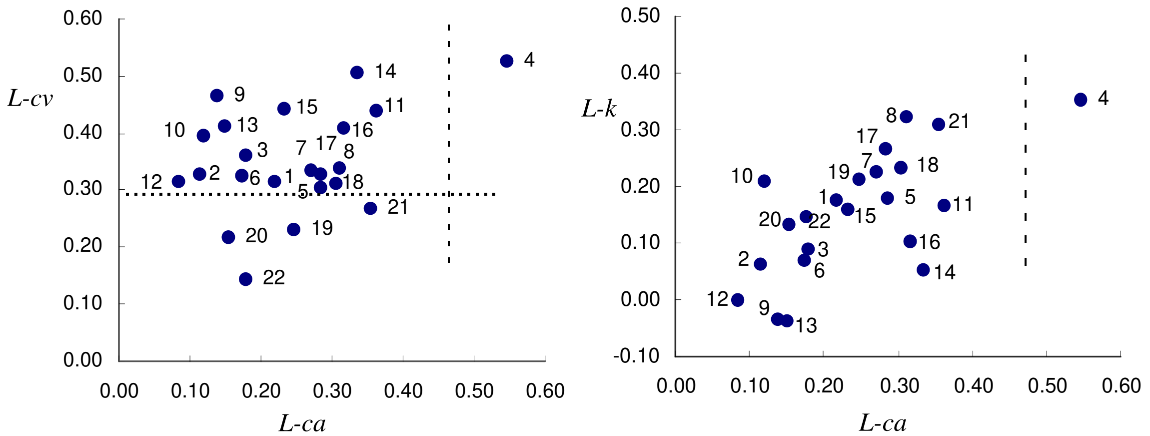

Fig. 1. Diagramma degli L-moments

La presenza di sole 4 stazioni in tale sottozona rende al momento solo indicativo questo raggruppamento. Ulteriori conferme vanno ricercate in una estensione dello studio ai territori Campani e Calabresi geograficamente contigui e climaticamente simili.

L'analisi di omogeneità condotta sulle due sottoregioni fornisce per l'insieme delle 17 stazioni appartenenti alla regione semi-arida:

$H_1 = 0.77$ con *parent* GEV; $H_1 = 1.43$ con *parent* kappa.

Per le rimanenti 4 stazioni appartenenti alla regione omogenea cosiddetta umida:

$H_1 = 0.99$ con *parent* GEV, $H_1 = 0.42$ con *parent* kappa.

L'indice di eterogeneità calcolato con *parent* kappa sulle stazioni della zona semi-arida assume un valore compreso tra 1 e 2, indicando la presenza di incertezza nei riguardi dell'omogeneità. Tale incertezza è legata, oltre che alla scelta della *parent*, al fatto di avere trascurato, nella generazione, l'autocorrelazione delle serie, la correlazione tra le serie nelle diverse stazioni e nell'avere assunto normalmente distribuito il campo delle varianze generate. Se tali ipotesi fossero vere, la "misura di eterogeneità" rappresentebbe un test statistico con soglia di rigetto pari ad 1.28 con livello di significatività del 10% [22].

Parallelamente si è condotta anche l'analisi di omogeneità svolta con il criterio suggerito da Brath *et al.* [4]. In figura 2 viene mostrato il diagramma di dispersione dei PWM, calcolati sulla zona unica costituita dalle 21 serie in esame, mentre in tabella III si possono leggere gli esponenti e i coefficienti di determinazione R^2 delle regressioni logaritmiche dei PWM ai primi 5 ordini rispetto all'area. I PWM sono stati stimati secondo la (8) e indicati con β_k, dove k è l'ordine del momento. Il parametro m riportato in tabella III è quello della (19). Si sono distinte le regressioni effettuate sulle 21 stazioni della zona omogenea di I livello, da quelle relative ai gruppi di stazioni ricadenti nelle due sottozone.

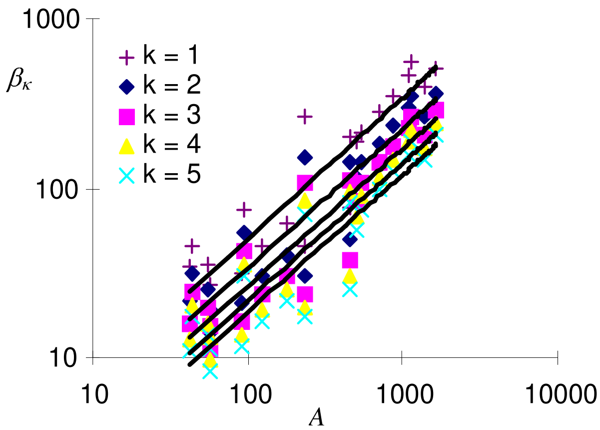

Fig. 2. Diagramma di dispersione

k	*m*	R^2	*m*	R^2	*m*	R^2
	zona unica		*zona semi-arida*		*zona umida*	
1	0.831	0.847	0.834	0.874	0.780	0.884
2	0.817	0.853	0.820	0.864	0.788	0.899
3	0.812	0.856	0.813	0.861	0.793	0.905
4	0.811	0.857	0.810	0.860	0.797	0.908
5	0.810	0.859	0.809	0.859	0.801	0.910

Tab. III. Parametri di interesse della regressione logaritmica

I risultati ottenuti tuttavia non evidenziano l'esistenza di due zone climaticamente eterogenee. Infatti l'analisi svolta su tutte le 21 stazioni darebbe credito all'ipotesi di esistenza di una zona unica di secondo livello. Il test, però, non sembra particolarmente potente se si osserva che dall'analisi separata delle due sottozone precedentemente identificate emerge la differenza tra gli esponenti *m* (relativi al 1° e 2° ordine, quelli cioè che interessano per l'analisi del comportamento di *L-cv*) della zona semi-arida e della zona umida. In realtà essendo *L-cv* molto sensibile al valore di *m* e non disponendo di strumenti per una verifica affidabile dell'uguaglianza degli *m* tra i diversi valori di *k*, il test lascia in questo caso ampi margini di incertezza. In definitiva, in questo lavoro, si è proceduto privilegiando l'ipotesi di una discriminazione tra la zona semi-arida e quella umida.

Pertanto la distribuzione GEV è stata stimata regionalmente, con criterio gerarchico, sui dati a disposizione raggruppati in una zona unica al I livello e in due sottozone al secondo. Si è così ottenuto un valore unico del coefficiente $\boldsymbol{k}$, stimato mediante le (11):

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

Per le due sottoregioni, vincolando la stima dei restanti parametri al valore regionale di $\boldsymbol{k}$, si sono ottenute le stime:

$\boldsymbol{a} = 0.47$ $\boldsymbol{u} = 0.69$ per la zona semi-arida

$\boldsymbol{a} = 0.29$ $\boldsymbol{u} = 0.81$ per la zona umida

Tali parametri danno luogo alle curve di crescita rappresentate nelle figure 3-5 in carta semilogaritmica, in cui *T* rappresenta il tempo di ritorno. In particolare nelle figure 3 e 4 sono mostrati i dati osservati, normalizzati in base al valor medio della serie corrispondente, e le due curve che si ottengono rispettivamente per la zona semi-arida e umida. Le curve in tratteggio in figura 3 sono le curve di crescita regionali, distribuite secondo una legge del tipo TCEV e stimate in [5], relativamente alle sole stazioni pugliesi. Le curve TCEV sono variabili in funzione di *cv* ed in figura 3 si sono rappresentate quelle relative ai *cv* minimo e massimo osservabili in regione. Nella stessa figura i punti relativi ai valori osservati contengono anche i dati del Venosa a P.te S. Angelo, non utilizzati per la stima dei parametri. Il buon adattamento della curva ai punti solleva dubbi sulla potenza del test di eterogeneità utilizzato. Va detto inoltre che anche la stima dei parametri regionali, come mostrato da calcoli qui non riportati, non risulta significativamente influenzata dall'utilizzo dei dati di detta sezione.

Dal confronto delle GEV regionali delle due sottozone (figura 5), risulta evidente la non trascurabile differenza nelle curve di crescita risultanti per le due zone.

4.1 Commento dei risultati

I risultati ottenuti offrono interessanti spunti di riflessione. In primo luogo appare evidente come l'analisi contestuale dei dati osservati in una regione ampia sia senz'altro di giovamento all'analisi regionale. Infatti, in tale regione è possibile riconoscere sia i caratteri di eterogeneità climatica, sia i caratteri di omogeneità idrologica all'interno delle sub-regioni, i cui confini non sono riconducibili a quelli amministrativi e, talvolta, nemmeno ai limiti di bacino idrografico. Va altresì rimarcato che i risultati ottenuti per la zona semi-arida sono molto significativi con particolare riferimento alla pendenza della curva di crescita e al relativo parametro di scala *a* che assume valore 0.47, molto prossimo al valor medio di quelli ritrovati da Farquharson *et al.* [9] analizzando un gran numero di serie storiche in regioni aride e semi-aride di tutto il mondo.

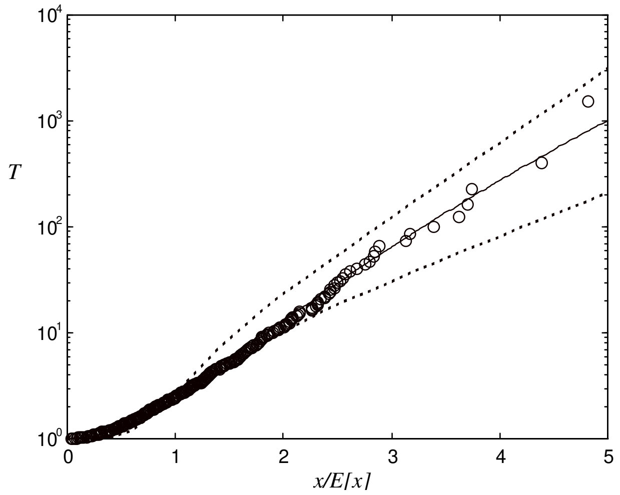

Figura 3. Dati osservati e curva di crescita della zona semi-arida

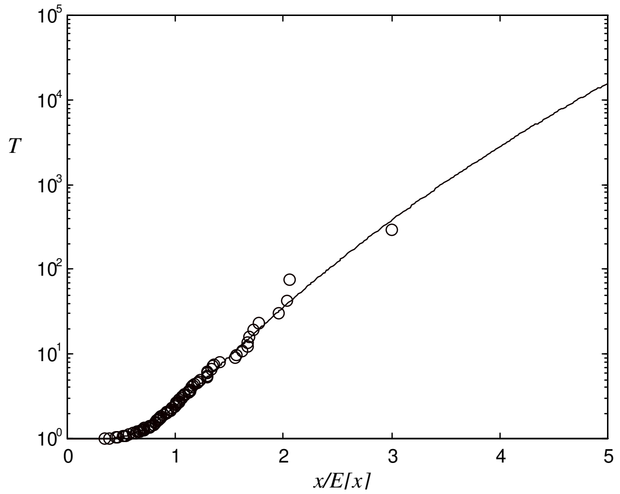

Figura 4. Dati osservati e curva di crescita della zona umida

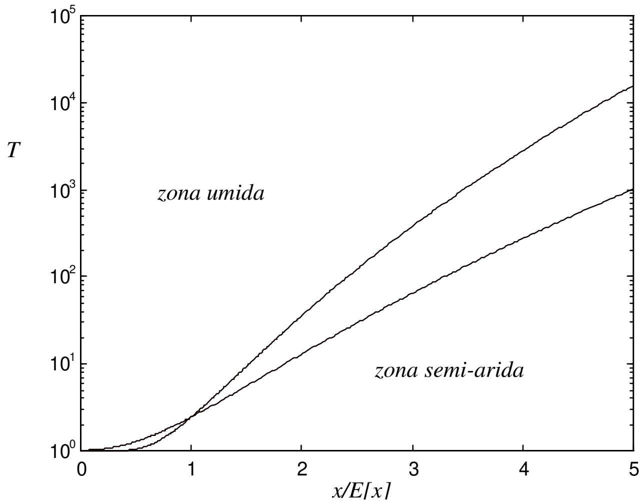

Figura 5. Curve di crescita delle due sottozone

Tuttavia rimangono aperti alcuni interrogativi di carattere generale sulle tecniche adottate per la stima regionale dei parametri e per la individuazione delle regioni omogenee. In particolare, l'ipotesi di base delle tecniche di regionalizzazione impone che il valore teorico dei parametri *L-cv* ed *L-ca* sia costante all'interno delle zone omogenee e che gli scostamenti dei valori osservati dalla media siano dovuti alla variabilità campionaria. Osservando i diagrammi di figura 1 si nota però che non è possibile individuare due zone a comportamento nettamente diverso. Esistono cioè certamente siti in cui si hanno valori elevati di *L-cv* ed altri in cui i valori sono decisamente più bassi, ma si registrano anche molti valori intermedi quasi a denotare una graduale evoluzione di *L-cv* sul territorio. Ciò suggerisce l'esistenza di una precisa variabilità spaziale che si sovrappone e si confonde con la variabilità campionaria e di cui i modelli esaminati non tengono alcun conto; al contrario essi sono costruiti con l'intento di interpretare la variabilità reale (spaziale e campionaria) esclusivamente come variabilità campionaria.

In tale situazione la scelta delle zone omogenee rimane piuttosto arbitraria a meno che non si tenga conto delle caratteristiche climatiche dei territori e della loro conformazione geografica. Sembra evidente quindi la necessità di superare le tecniche classiche di analisi regionale che non contemplano la possibilità che esista una variabilità spaziale (graduale) dei parametri. Considerazioni di questa natura costituiscono il necessario supporto all'analisi statistica e impongono la necessità di indagare ulteriormente sul significato fisico della variabilità spaziale dei dati e sui legami esistenti tra il clima, le caratteristiche fisiche dei bacini ed il comportamento statistico dei dati osservati.

5. CONCLUSIONI

Nel corso di questo lavoro si è cercato di sottolineare alcuni problemi connessi all'analisi di frequenza delle piene, già al centro di un intenso dibattito internazionale. Risulta infatti evidente come né le sempre più sofisticate tecniche analitiche, né le notevoli capacità di calcolo attualmente disponibili, siano ancora in grado di fornire modelli di calcolo oggettivi ed esaurienti.

A tal fine si è proceduto all'applicazione di tecniche di analisi regionale, più o meno consolidate, ai dati di piena osservati in un'ampia regione. Si è rilevato che, pur ricorrendo a metodi "oggettivi" restano inevitabilmente aperti ampi margini di arbitrarietà in alcuni passaggi fondamentali quali ad esempio la scelta del numero di zone e sottozone omogenee, la definizione delle stesse, l'analisi di affidabilità dei dati disponibili, ecc.

L'applicazione ai dati di Basilicata e Puglia, può considerarsi senza dubbio significativa ed i risultati conseguiti certamente offrono un contributo alla conoscenza dei fenomeni idrologici ivi presenti, come anche si è reso evidente il forte vantaggio connesso all'analisi contestuale dei dati disponibili nelle due regioni.

Resta tuttavia il dubbio che tali risultati positivi siano dovuti più alla conoscenza approfondita dei luoghi e dello loro caratteristiche climatiche e idrologiche, che non alla affidabilità del metodo utilizzato.

BIBLIOGRAFIA

[1]**Ashkar, F.; Bobée, B.; Leroux, D.; Morisette, D.;** The generalized method of moments applied to the generalized gamma distribution, *Stochastic Hydrol. Hydraul.*, 2, 161-174, 1988.

[2] **Ashkar, F.; Bobeé, B.; Rasmussen, P.; Rosbjerg, D.;** A perspective on the annual maximum flood approach to flood frequency analysis, in K. W. Hipel (eds.), *Stochastic and Statistical Methods in Hydrology and Environmental Engineering*, 1, 3-14, Kluwer Acad. Dordrecht, the Netherlands, 1994.

[3]**Bobée, G.; Cavadias, G.; Ashkar, F.; Bernier, J.; Rasmussen, P.;** Towards a systematic approach to comparing distribution used in flood frequency analysis, *Jour. of Hydrology*, 142, 121-136, 1993.

[4]**Brath, A.; De Michele, C.; Galeati, G.; Rosso, R.;** Una metodologia per l'identificazione di regioni omogenee nel regime di piena. Applicazione all'Italia nord-occidentale, *L'Acqua*, 1, 17-26, 1997.

[5]**Claps, P.; Copertino, V.; Fiorentino, M.;** Analisi regionale dei massimi annuali delle portate al colmo di piena, in Copertino V. A. e Fiorentino M. (a cura di) *Valutazione delle piene in Puglia*, 211-246, DIFA-GNDCI, Potenza, 1994.

[6]**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.

[7]**Dalrymple, T.;** Flood-frequency analysis, *U. S. Geol. Surv. Water Supply Pap.,* 1543-A., 1960.

[8]**Dawdy, D. R.;** Variation of flood ratios with size of drainage area, *U. S. Geological Survey Prof. Pap.,* 424C, paper C36, 1961.

[9]**Farquharson, F. A. K.; Meigh, J. R.; Sutcliffe, J. V.;** Regional flood frequency analysis in arid and semi-arid areas, *Jour. of Hydrology*, 138, 487-501, 1992.

[10]**Fiorentino M.; Gabriele, S.;** Distribuzione TCEV: metodi di stima dei parametri e proprietà statistiche degli stimatori, Consiglio Nazionale delle Ricerche-IRPI, Geodata n.25, Cosenza, 1985.

[11]**Fiorentino M.; Versace, P.; Rossi, F.;** Regional flood frequency estimation using the two-component extreme value distribution, *Hydrological Sciences Journal*, 30 (1), 51- 64, 1985.

[12]**Fiorentino M.; Arora, K.; Singh, V. P.;** The two-component extreme value distribution for flood frequency analysis: derivation of a new estimation method, in *Stochastic Hydrology and Hydraulics*, 1(3), 1987.

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

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

[15]**Gabriele, S.; Villani, P.;** An investigation of the feasibility space of parameter estimation using POME and ML with reference to the TCEV distribution, V.P. Singh and M. Fiorentino (eds.), *Entropy and Energy Dissipation in Water Resources*, 193-203, Kluwer Acad. Dordrecht, the Netherlands 1992.

[16]**Gioia, G.; Lo Presti, M.;** Sulla individuazione della funzione di ripartizione di probabilità di serie storiche asimmetriche, Atti del 3° Seminario Nazionale *Sistemi di drenaggio urbano* , Vol. 1, Ancona, 1992.

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

[18]**Groupe de recherche en hydrologie statistique (GREHYS);** Presentation and review of some methods for regional flood frequency analysis, *Jour. of Hydrology*, 186, 63-84, 1996.

[19]**Gupta, V. K.; Dawdy D. R.;** Physical interpretations of regional variations in the scaling exponents of flood quantiles, *Hydrological Processes*, 9, 347-361, 1995

[20]**Haktanir, T.;** Comparison of various flood frequency distributions using annual flood peaks data of rivers in Anatolia, *Jour. of Hydrology*, 136, 1-31, 1992.

[21]**Hosking, J. R. M.;** L-moments: Analysis and estimation of distributions using linear combinations of order statistics, *J. R. Stat. Soc.,* Ser. B, 52, 105-124, 1990.
[22]**Hosking, J. R. M.; Wallis, J. R.;** Some Statistic usefol in Regional Frequency Analysis, *Water Resources Research*, 29 (2), 271-281, 1993.
[23]**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.
[24]**Houghton, J. C;** Birth of a parent: The Wakeby distribution for modeling flood flows, *Water Resources Research*, 14(6), 1105-1110, 1978.
[25]**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.
[26]**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.
[27]**Lettenmaier, D. P.; Potter, K. W.;** Testing flood frequency estimation methods using a regional flood generation model, *Water Resources Research*, 21(12), 1903-1914, 1985.
[28]**Lettenmaier, D. P.; Wallis, J. R.; Wood, E. F.;** Effect of Regional Heterogeneity on Flood Frequency Estimation, *Water Resources Research*, Vol. 23, N. 2, pp. 313-323, 1987.
[29]**Moon, Y.; Lall, U.;** Kernel quantile function estimator for flood frequency analysis, *Water Resources Research*, Vol. 30, N. 11, 3095-3103, 1994.
[30]**NERC - Natural Environment Research Council;** *Flood studies report,* Vol. I - Hydrologic studies, NERC, London, 1975.
[31]**Phien, H. N.; Hira, M. A.;** Log Pearson type 3 distribution: parameter estimation, *Jour. of Hydrology*., 64, 25-37, 1983.
[32]**Rao, D.V.;** Log Pearson type 3 distribution: method of mixed moments, *Jour. Hydraul. Div.*, ASCE, 111(7), 1043-1057, 1980.
[33]**Robinson, J. S.; Sivapalan, M.;** An investigation into the physical causes of scaling and heterogeneity of regional flood frequency, *Water Resources Research,* 33(5), 1045-1059, 1997
[34]**Rossi, F.; Versace, P.;** Criteri e metodi per l'analisi statistica delle piene, in: *Valutazione delle piene*, C.N.R. - P.F. Conservazione del Suolo n.165, 63-130, 1982.
[35]**Rossi, F.; Fiorentino, M.; Versace, P.;** Two component extreme value distribution for flood frequency analysis, *Water Resources Research*, 20(7), 847-856, 1984.
[36]**Rossi, F.; Villani, P.;** Regional methods for flood estimation, in G. Rossi (eds.) *Coping with floods*, NATO-ASI Series, Kluwer, 135-169, 1992.
[37]**Rosso, R.; Brath, A.; Burlando, P.; De Michele, C.; Mancini, M.;** Invarianza di scala del regime di piena, *L'Acqua*, 5, 15-23, 1996.
[38]**Vogel, R. M.; Thomas Jr., W. O.; McMahon, T. A.;** Flood-flow frequency model selection in southwestern United States, *Jour. of Water Resour. Plann. and Manag., ASCE,* 119(3), 353-366, 1993.
[39]**WMO (World Meteorological Organization);** *Statistical distributions for flood frequency analysis*, Operational Hydrology Report N. 33., WMO Publ. N 718, WMO, Geneva, Switzerland, 1989.
[40]**Zrinji, Z.; Burn, D. H.;** Regional flood frequency analysis with hierarchical region of influence, *Jour. of Water Resour. Plann. and Manag.*, *ASCE*, 122 (4), 245-252, 1996.