

## UTILIZZO DI CRITERI DI PROSSIMITÀ NELL'ANALISI REGIONALE DEL DEFLUSSO ANNUO

A. Viglione, P. Claps, F. Laio

Dipartimento di Idraulica, Trasporti ed Infrastrutture Civili, POLITECNICO DI TORINO - Torino  
e-mail: alberto.viglione@polito.it, pierluigi.claps@polito.it, francesco.laio@polito.it

Parole chiave: Regionalizzazione, Geomorfologia, Deflusso Annuo, Cluster Analysis.

### SOMMARIO

*Per la valutazione della disponibilità idrica in ambito regionale si usano modelli probabilistici di stima del deflusso annuo, corredati da criteri utili a stabilire la variabilità spaziale della grandezza in esame. Il modello di analisi di frequenza regionale maggiormente utilizzato è quello della grandezza indice, schema basato sull'ipotesi che le distribuzioni di probabilità della grandezza in differenti siti appartenenti ad una regione statisticamente omogenea siano identiche a meno di un parametro di scala. Mentre la determinazione del valore indice è usualmente affidata a metodi consolidati di regressione multipla, le tecniche proposte per la determinazione delle regioni omogenee sono piuttosto complesse e per lo più basate su procedure che comportano scelte soggettive. In questo lavoro si propone un criterio oggettivo di raggruppamento dei bacini idrografici in regioni omogenee, nel quale si usano come variabili discriminanti alcune grandezze morfologiche e climatiche dei bacini. Nello spazio multidimensionale costruito con tali grandezze morfoclimatiche, vengono calcolate le differenze tra i valori assunti da queste nei diversi bacini. Tali differenze vengono interpretate come "distanze" tra i bacini e vengono poi correlate a distanze caratteristiche tra le rispettive curve di distribuzione di probabilità adimensionalizzate (curve di crescita). La significatività della correlazione tra tali distanze è valutata attraverso il test di Mantel. Dall'applicazione del test emerge un insieme di indicatori morfoclimatici significativi nei riguardi della discriminazione della forma delle distribuzioni di probabilità adimensionali. Tali indicatori sono successivamente utilizzati nella formazione delle regioni omogenee, attraverso tecniche di Cluster Analysis. La procedura proposta è stata messa a punto su un gruppo di 47 bacini idrografici appartenenti alle regioni Piemonte e Valle d'Aosta, le cui caratteristiche fisico-climatiche variano in modo sostanziale passando dalle aree alpine a quelle più temperate del Piemonte meridionale.*

### 1 INTRODUZIONE

Per molti problemi pratici dell'idrologia delle acque superficiali, come la gestione delle risorse idriche o lo studio degli eventi di piena, è importante poter far affidamento su informazioni relative ai deflussi che siano nello stesso tempo accurate e diffuse sul territorio. I deflussi, tuttavia, sono grandezze cumulate che non possono essere interpolate nello spazio con le tecniche usate per le grandezze puntuali, quali precipitazione o temperatura. Inoltre le serie storiche dei deflussi misurati sono spesso brevi al punto da rendere il campione inadeguato ai fini dell'inferenza statistica su base locale.

Per ovviare a questi inconvenienti si utilizzano tecniche statistiche di regionalizzazione del dato idrologico, che presuppongono di utilizzare, ai fini dell'inferenza statistica, opportuni raggruppamenti dei dati provenienti da più stazioni idrometriche, oltre che di sfruttare il legame tra grandezze morfoclimatiche facilmente determinabili e la variabile di interesse.

Uno dei primi contributi in questo ambito è quello di Dalrymple (1960) che propose la procedura nota come metodo della "piena-indice". Alla metodologia di Dalrymple si ispirano applicazioni quali il progetto VAPI (VALutazione delle Portate in Italia, url <http://caronte.gndci.cs.cnr.it/GNDCI/>

Vapi.htm) e l'FEH ("Flood Estimation Handbook", *Robson & Reed* (1999)) orientate all'analisi regionale di frequenza delle piene. La variabile presa in considerazione in questo lavoro è il deflusso annuo, ossia il volume d'acqua che transita in una sezione di un corso d'acqua in un anno. Le applicazioni dell'analisi regionale a grandezze medie annue o stagionali sono molto meno numerose di quelle relative alle piene. Riguardo al deflusso annuo, *Vogel & Wilson* (1996) danno una breve rassegna dei metodi, soffermandosi sui lavori effettuati negli Stati Uniti. Per quanto riguarda l'Italia ricordiamo i lavori di *Ferraresi et al.* (1988), di *Claps & Mancino* (2002) per la Basilicata e di *Brath et al.* (2004) per la Romagna.

L'obiettivo di questo lavoro è sia metodologico, mirato all'individuazione di una procedura oggettiva e robusta per la regionalizzazione delle curve di frequenza del deflusso annuo, che applicativo, relativamente all'implementazione dell'analisi regionale nel territorio piemontese e valdostano. Nel successivo Paragrafo 2 vengono appunto descritti in dettaglio gli strumenti metodologici messi a punto per l'analisi regionale dei deflussi annui, mentre nel Paragrafo 3 si riportano i risultati dell'applicazione alle regioni Piemonte e Valle d'Aosta.

## 2 METODI

Lo scopo dell'analisi di frequenza di una variabile aleatoria è la stima della sua distribuzione di probabilità a partire da dati osservati. Nel caso considerato in questo studio, la variabile aleatoria è il deflusso annuo  $D$ . L'obiettivo che ci si pone è la determinazione di  $D(F)$ , ovvero il quantile con probabilità  $F$  della curva di frequenza di  $D$  o, ancora, il valore di  $D$  che ha probabilità  $F$  di non essere superato.

Per sezioni idrografiche sprovviste di misure o con pochi anni di osservazioni si deve far ricorso a metodi di analisi regionale. In questo lavoro ci si è ispirati alla metodologia della piena-indice, utilizzando l'approccio basato sugli  $L$ -momenti proposto da *Hosking & Wallis* (1997). Benché le tecniche descritte siano state proposte nell'ambito dell'analisi di frequenza delle piene, esse possono essere facilmente estese ad altri tipi di variabili idrologiche. Tali tecniche si basano sull'ipotesi che le distribuzioni delle piene in differenti siti appartenenti ad una "regione" siano identiche a meno di un parametro di scala, la grandezza-indice, che riflette le caratteristiche climatiche, morfometriche e geologiche di ogni bacino considerato.

In questo lavoro utilizzeremo il termine deflusso indice per indicare il valore caratteristico del deflusso annuo. Sebbene si possa usare qualsiasi parametro di scala della distribuzione come deflusso indice, i valori più frequentemente utilizzati in letteratura sono la media (v.es. *Hosking & Wallis*, 1997) o la mediana (v.es. *Robson & Reed*, 1999) campionarie. Da un precedente studio condotto dagli autori (*Viglione et al.*, 2006) si evince che, per variabili aleatorie caratterizzate da asimmetria bassa, la stima della media è meno distorta di quella della mediana. È quindi preferibile usare la media campionaria come valore indice per il deflusso annuo, considerata la sua bassa asimmetria. Indicheremo con  $D_m$  la media teorica del deflusso annuo, con  $\tilde{D}_m$  quella campionaria in sezioni monitorate e con  $\hat{D}_m$  quella stimata in sezioni senza dati.

La curva di frequenza di  $D$ , adimensionalizzata rispetto al deflusso indice  $x(F) = D(F)/D_m$ , verrà invece chiamata "curva di crescita". Anche in questo caso distingueremo la curva teorica, quella campionaria e quella stimata utilizzando i simboli  $x(F)$ ,  $\tilde{x}(F)$  e  $\hat{x}(F)$ .

L'analisi di frequenza regionale del deflusso annuo viene suddivisa in:

- stima regionale del deflusso indice  $D_m$  (Paragrafo 2.1);
- stima regionale della curva di crescita  $x(F)$  (Paragrafo 2.2).

Per quanto riguarda la prima parte dello studio di regionalizzazione si è ricercato un modello che legasse  $D_m$  alle caratteristiche morfoclimatiche dei bacini per i quali sono disponibili misure idrometriche. A questo scopo si sono utilizzati metodi di regressione lineare multipla.

Per la seconda parte si è valutata la necessità di dividere il territorio in regioni statisticamente omogenee, nelle quali sia possibile adottare un'unica  $x(F)$ . Il termine regione può far pensare ad un insieme di siti geograficamente vicini, ma la vicinanza geografica non comporta necessariamente la somiglianza in distribuzione di frequenza, soprattutto per variabili aggregate come il deflusso. È ragionevole invece identificare le regioni sulla base di quelle caratteristiche (associate ai bacini idrografici) che si pensa influenzino la forma della distribuzione di frequenza. La scelta di regioni geograficamente disperse ha anche il vantaggio di limitare la correlazione tra i dati che si utilizzano per la stima delle curve di crescita, permettendo un'inferenza statistica più rigorosa (*Hosking & Wallis*, 1997).

## 2.1 Stima regionale del deflusso indice

Per la stima di  $D_m$  si sono valutati diversi modelli di regressione lineare multipla, del tipo:

$$D_m = \beta_0 + \beta_1 M_1 + \beta_2 M_2 + \dots + \beta_{p-1} M_{p-1} + \varepsilon, \quad (1)$$

$$D_m = \alpha M_1^{\beta_1} M_2^{\beta_2} \dots M_{p-1}^{\beta_{p-1}} \varepsilon, \quad (2)$$

$$D_m^\lambda = \beta_0 + \beta_1 M_1 + \beta_2 M_2 + \dots + \beta_{p-1} M_{p-1} + \varepsilon, \quad (3)$$

dove  $M_i$  sono parametri morfoclimatici (eventualmente trasformati) e  $\beta_i$  sono i coefficienti della regressione. Si noti che anche l'Equazione (2) può essere resa lineare nei coefficienti utilizzando una trasformazione logaritmica. Per la stima dei coefficienti delle Equazioni (1), (2) e (3) si è utilizzata la tecnica dei Minimi Quadrati Ordinari (v.es. *Montgomery et al.*, 2001).

Per ogni tipologia di regressione si sono confrontati i modelli ottenibili da tutte le combinazioni delle variabili morfoclimatiche considerate (eventualmente trasformate). Innanzitutto si sono esclusi quei modelli per i quali anche solo una delle variabili esplicative risultasse non significativa in base al test della  $t$  di Student all'1% (v.es. *Montgomery et al.*, 2001). Successivamente si è valutata la capacità descrittiva di ogni regressione tramite il coefficiente di determinazione corretto  $R_{adj}^2$  calcolato come:

$$R_{adj}^2 = 1 - \frac{(n-1) \sum_{i=1}^n (\tilde{D}_{m,i} - \hat{D}_{m,i})^2}{(n-p) \sum_{i=1}^n (\tilde{D}_{m,i} - \langle \tilde{D}_m \rangle)^2}, \quad (4)$$

dove  $n$  è il numero delle stazioni considerate,  $p$  il numero dei coefficienti stimati (vedi Equazioni (1), (2) e (3)),  $\tilde{D}_{m,i}$  e  $\hat{D}_{m,i}$  sono il deflusso medio annuo misurato e stimato alla stazione  $i$ , e  $\langle \tilde{D}_m \rangle$  è la media dei deflussi medi annui di tutte le stazioni considerate. In pratica si è calcolato  $R_{adj}^2$  per tutti i modelli possibili e si è scelto quello con  $R_{adj}^2$  maggiore.

Il coefficiente di determinazione non è però utilizzabile per il confronto tra modelli di natura diversa (Equazioni (1), (2) e (3)), per cui la scelta è stata ulteriormente definita con il metodo della cross-validazione. Il metodo consiste nell'analisi dei residui  $\hat{D}'_{m,i} - \tilde{D}_{m,i}$ , dove  $\hat{D}'_{m,i}$  è il valore stimato della  $i$ -esima variabile dipendente, basato però sulla regressione ottenuta utilizzando tutte le osservazioni eccettuata la  $i$ -esima. Si sono confrontati gli errori quadratici medi RMSE ("Root Mean Squared Error") ottenuti in cross-validazione, calcolati come:

$$\text{RMSE}_{cv} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{D}'_{m,i} - \tilde{D}_{m,i})^2}. \quad (5)$$

Si è quindi scelto il modello di regressione che fornisce l' $\text{RMSE}_{cv}$  minore.

## 2.2 Stima regionale della curva di crescita

La seconda fase del metodo regionale richiede la stima della curva di crescita del campione formato dal raggruppamento dei dati (adimensionalizzati) di più stazioni appartenenti ad una regione omogenea. È evidente che, data la complessità del territorio dell'area piemontese - valdostana, si deve affrontare il delicato problema della sua suddivisione in regioni omogenee all'interno delle quali possa essere utilizzata un'unica curva di crescita. La procedura utilizzata per la regionalizzazione di  $x(F)$  può essere suddivisa nei punti seguenti:

- determinazione delle variabili di classificazione;
- selezione della tecnica di raggruppamento;
- identificazione del numero di gruppi, tramite test di omogeneità;
- selezione dei modelli di distribuzione di frequenza con cui stimare le  $x(F)$  (model selection).

### 2.2.1 Identificazione delle variabili di classificazione

La determinazione delle regioni omogenee è stata effettuata tramite “cluster analysis”, anche detta analisi di raggruppamento. Il raggruppamento dei dati disponibili deve essere effettuato in maniera tale da rendere il più possibile simili le curve di crescita relative a serie di dati appartenenti allo stesso cluster. Non si possono tuttavia considerare, ai fini della formazione dei gruppi, statistiche direttamente riconducibili alle curve di crescita stesse, dal momento che risulterebbe impossibile assegnare ad una delle regioni un sito dove non si hanno dati (*Hosking & Wallis*, 1997). Bisogna invece considerare le variabili di bacino maggiormente correlate alle caratteristiche delle curve di crescita campionarie. La scelta iniziale delle variabili da considerare nell’analisi è di fondamentale importanza, in quanto la successiva suddivisione terrà conto solo di esse.

Bisogna innanzitutto decidere come caratterizzare le curve di crescita. A questo scopo si può considerare una delle statistiche dei campioni (es. il coefficiente di variazione o l’ $L$ -CV) come rappresentativa dell’intera curva di crescita e valutare la correlazione di questa statistica con le grandezze morfoclimatiche. Questo approccio tuttavia presuppone che la statistica considerata sia sufficientemente descrittiva dell’intera variabilità delle curve di crescita. In questo lavoro si è utilizzato un metodo più generale, legato alla valutazione di diversità, o distanza, tra le curve di crescita: la distanza tra le  $\tilde{x}(F)$  di tutte le coppie di stazioni è stata analizzata insieme alla distanza tra rispettivi parametri di bacino nello spazio delle variabili morfoclimatiche. In sostanza, se per due stazioni si hanno curve di crescita molto simili (diverse), si cerca di valutare in che cosa sono simili (diversi) i due bacini.

Per far ciò si è usato un metodo poco noto in ambito idrologico che si basa sul confronto tra le “matrici delle distanze” dove, in questo caso, le distanze sono tra distribuzioni e tra grandezze morfoclimatiche. Le distanze tra tutti i parametri relativi a coppie di bacini vengono raccolte in matrici quadrate  $\Delta$  (matrici delle distanze) il cui elemento generico  $\delta_{hk}$  è una misura della differenza tra i valori assunti dalla variabile considerata riguardo alle entità  $h$  e  $k$  (v.es. *Fabbris*, 1997). Ad esempio, se si considera la variabile area del bacino  $S$ ,  $\delta_{hk}$  sarà data dalla differenza, in valore assoluto, tra le aree dei bacini  $h$  e  $k$ .  $\Delta_S$  sarà una matrice di dimensione  $n \times n$ , simmetrica e con diagonale nulla, contenente la differenza tra le aree di tutte le possibili coppie di bacini. La variabile morfometrica  $S$  verrà quindi utilizzata nella classificazione solo se la relativa matrice delle distanze  $\Delta_S$  risulterà correlata alla matrice  $\Delta_{x(F)}$  delle distanze tra le curve di crescita. Le distanze tra le curve di crescita si possono calcolare con statistiche di omogeneità di coppie di serie. Ad esempio si può utilizzare la misura di eterogeneità di *Hosking & Wallis* (1993) o quella di Anderson-Darling proposta in *Viglione et al.* (2006).

Le matrici delle distanze contengono, per loro natura, valori fortemente correlati tra di loro. Se, ad esempio, si considera ancora  $\Delta_S$ , se l’elemento  $\delta_{12}$  vale 1 e l’elemento  $\delta_{13}$  vale 2,  $\delta_{23}$  potrà solo assumere i valori 1 o 3. Per verificare la presenza di correlazione tra due matrici bisogna quindi ricorrere a tecniche statistiche particolari, che non risentano della mutua dipendenza degli elementi in  $\Delta$ , quali il test di Mantel (*Mantel & Valand*, 1970), sviluppato nell’ambito delle scienze biologico-ambientali. Il test di Mantel è, appunto, una procedura per verificare se la matrice delle distanze  $\Delta_A$  è linearmente indipendente dalla matrice delle distanze  $\Delta_B$ , contenente le distanze calcolate rispetto ad un altro parametro. L’ipotesi nulla (di indipendenza) è testata con una procedura in cui il valore originale della statistica è paragonato alla distribuzione ottenuta da tutte le possibili permutazioni (o un numero elevato di permutazioni casuali) degli oggetti contenuti in una delle due matrici. La statistica usata per la misura della correlazione tra le matrici è il classico coefficiente di correlazione di Pearson:

$$r(\Delta_A \Delta_B) = \frac{1}{N-1} \sum_{i=1}^N \sum_{j=1}^N \left[ \frac{(\delta_{A,ij} - m_A)}{s_A} \right] \left[ \frac{(\delta_{B,ij} - m_B)}{s_B} \right], \quad (6)$$

dove  $N = n(n-1)/2$  è il numero degli elementi della parte triangolare superiore (o inferiore) delle matrici,  $m_A$  è la media degli elementi  $\delta_{A,ij}$  in  $\Delta_A$  e  $s_A$  è la loro deviazione standard.

Un’ulteriore sviluppo del test, il test di Mantel “parziale”, coinvolge invece tre matrici. Per dare una stima della correlazione tra le matrici  $\Delta_A$  e  $\Delta_B$  controllando l’effetto di una terza matrice  $\Delta_C$ , si calcola il coefficiente di correlazione parziale:

$$r(\Delta_A \Delta_B \cdot \Delta_C) = \frac{r(\Delta_A \Delta_B) - r(\Delta_A \Delta_C)r(\Delta_B \Delta_C)}{\sqrt{1 - r(\Delta_A \Delta_C)^2} \sqrt{1 - r(\Delta_B \Delta_C)^2}}, \quad (7)$$

dove  $r(\Delta_A \Delta_B)$  è calcolata con l'Equazione (6). Molti autori hanno suggerito differenti modalità di esecuzione del test. In questo lavoro è stata utilizzata la modalità suggerita da Legendre (2000), basata sulla permutazione dei residui di un "modello nullo". Per l'applicazione del test è stato utilizzato il pacchetto "vegan" (Oksanen et al., 2005) del software R (R Development Core Team, 2006) disponibile all'url <http://microarrays.unife.it/CRAN/>.

### 2.2.2 Selezione della tecnica di raggruppamento

Il risultato del test di Mantel è la selezione di un gruppo di indicatori morfoclimatici significativi nel discriminare la forma delle curve di crescita locali  $\tilde{x}(F)$ . Tali variabili vanno utilizzate nella successiva formazione delle regioni omogenee. La tecnica utilizzata a questo scopo è una metodologia mista, costituita da una prima suddivisione dei bacini con l'algoritmo gerarchico di Ward (1963) e da una successiva rifinitura dei gruppi con un metodo di minimizzazione della dispersione entro i cluster.

L'algoritmo di Ward è di tipo agglomerativo, ovvero parte da una situazione in cui ogni individuo costituisce un cluster. Ad ogni passo dell'analisi si considera l'unione di ogni possibile coppia di cluster e si uniscono quelli la cui fusione determina la minima perdita di informazione, che può essere misurata come la somma delle deviazioni quadratiche di ogni punto dal baricentro del cluster a cui appartiene. L'algoritmo di Ward è stato utilizzato perché è costruito in modo da generare gruppi compatti e con un numero di elementi confrontabile. Come tutti gli algoritmi gerarchici, però, anche quello di Ward ha l'inconveniente di non ammettere riallocazioni di elementi tra i gruppi, per cui non è detto che la configurazione finale sia effettivamente quella ottimale. Per ovviare a questo inconveniente si è utilizzata, a valle dell'algoritmo di Ward, una tecnica di riallocazione degli elementi.

La procedura viene inizializzata con una suddivisione di partenza, nel nostro caso quella in  $k$  cluster ottenuta con Ward. Per questa configurazione può essere calcolata la statistica:

$$W = \sum_{i=1}^k \left( \sum_{j=1}^{n_i} d_{ij}^2 \right), \quad (8)$$

dove  $d_{ij}$  è la distanza euclidea tra l'elemento  $j$ -esimo dell' $i$ -esimo gruppo e il baricentro dell' $i$ -esimo cluster ed  $n_i$  è il numero di elementi dello stesso. A passi successivi si valuta se lo spostamento di un elemento tra due gruppi comporta una diminuzione di  $W$ , nel qual caso viene applicata la correzione. Ciò avviene finché non si giunge ad una configurazione in cui non esistono modifiche in grado di diminuire ulteriormente la dispersione all'interno dei gruppi.

Tale metodo comporta la riallocazione dei punti finché tutti gli individui di ogni gruppo sono più vicini (come distanza euclidea) al centro del proprio cluster che al centro degli altri. Nel caso in cui le variabili di classificazione siano due, ciò comporta che i gruppi possano essere delimitati da "poligoni di Thiessen", rendendo molto semplice la definizione del criterio di assegnazione di una nuova entità ad un cluster (vedi Paragrafo 3.4).

### 2.2.3 Identificazione del numero di gruppi tramite test di omogeneità

Le tecniche di cluster analysis consentono di raggruppare i dati, ma il numero di gruppi che si formano va scelto in maniera indipendente: nel nostro caso si è cercato di suddividere i bacini nel minor numero possibile di gruppi, verificando che le regioni che si vengono a formare siano statisticamente omogenee. In pratica si parte dalla condizione in cui tutti i bacini sono raggruppati in un unico cluster e si esegue un test per valutare l'omogeneità del macrogruppo; successivamente si suddividono i bacini in due gruppi, in tre gruppi e così via, utilizzando la metodologia descritta nel Paragrafo 2.2.2; ci si arresta quando tutte le regioni passano il test.

Nel presente lavoro è stato utilizzato il test di eterogeneità di Hosking & Wallis (1993), in ragione della scarsa asimmetria dei deflussi annui (v.es. Viglione et al., 2006). Il test di Hosking & Wallis (1993) valuta il grado di eterogeneità  $H$  di un gruppo di stazioni e viene utilizzato per stabilire se essi possono essere ragionevolmente trattati come appartenenti ad una regione omogenea. Più specificamente, la misura di eterogeneità paragona la variabilità degli  $L$ -momenti campionari dei siti con quella che ci si aspetterebbe per

una regione omogenea. *Hosking & Wallis* (1993) suggeriscono di considerare la regione come “accettabilmente omogenea” se  $H < 1$ , come “possibilmente eterogenea” se  $1 \leq H < 2$ , e come “certamente eterogenea” se  $H \geq 2$ .

#### 2.2.4 Selezione di un modello per la curva di crescita

Una volta determinate le regioni omogenee, i dati di ognuna di esse vengono raggruppati e utilizzati per la stima delle curve di crescita regionali. In pratica si costruisce, per ogni raggruppamento, un campione contenente tutti i valori di deflusso delle stazioni appartenenti alla regione (divisi per i corrispondenti valori indice), ossia si applica il cosiddetto metodo “station-year” (v.es. *Hosking & Wallis*, 1997). Successivamente si valuta quale distribuzione di probabilità descrive meglio il campione. A tal fine si impiegano test atti a provare l’ipotesi che una data variabile casuale sia distribuita secondo un’assegnata distribuzione di probabilità, noti come test di adattamento. Assegnato un campione di dati  $x_i$  ( $i = 1, \dots, m$ ) estratto da una distribuzione  $F_R(x)$ , lo scopo del test è provare l’ipotesi statistica  $H_0 : F_R(x) = F(x, \theta)$ , dove  $F(x, \theta)$  è la distribuzione ipotetica e  $\theta$  è un vettore di parametri stimati dal campione  $x_i$ .

Per determinare se la distribuzione di probabilità scelta si adatta bene agli  $m$  dati di cui si dispone in un’assegnata regione, si è utilizzato un test di adattamento basato su una misura dello scostamento medio quadratico tra la distribuzione ipotetica  $F(x, \theta)$  e la funzione di frequenza cumulata  $F_m(x)$ , definita come:

$$\begin{cases} F_m(x) = 0, & x < x_{(1)} \\ F_m(x) = \frac{i}{m}, & x_{(i)} \leq x < x_{(i+1)} \\ F_m(x) = 1, & x_{(m)} \leq x \end{cases}, \quad (9)$$

dove con  $x_{(i)}$  si è indicato l’ $i$ -esimo elemento del campione di dati ordinato in senso crescente. In generale questi tipi di test possono essere ricondotti alla formulazione di una statistica test:

$$Q^2 = m \int_x [F_m(x) - F(x, \theta)]^2 \Psi(x) dF(x), \quad (10)$$

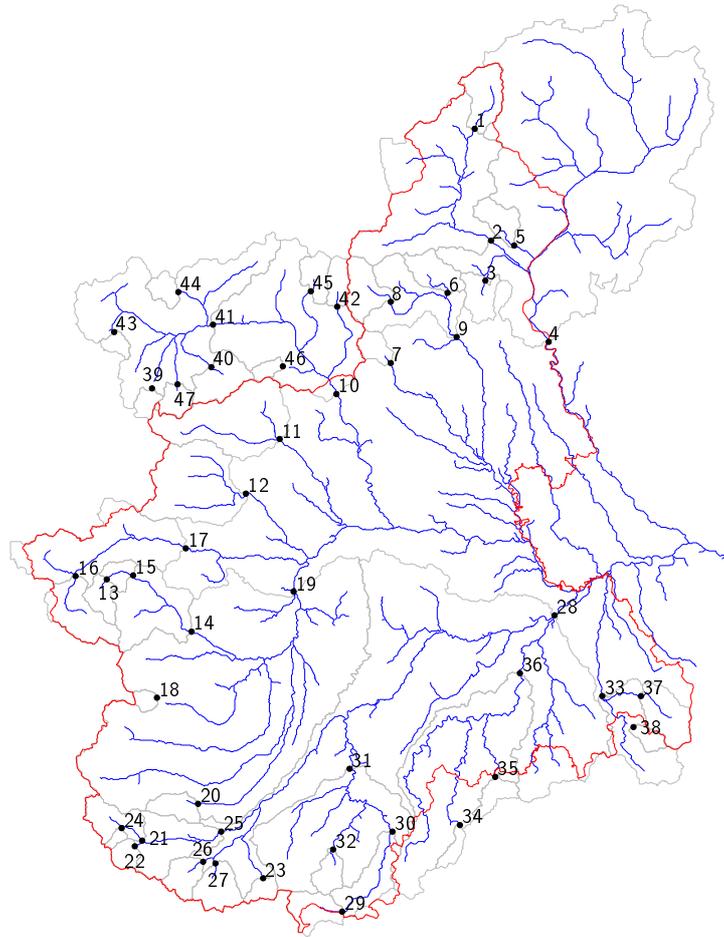
dove  $\Psi(x)$  è una funzione che può valere 1, nel qual caso si ha la statistica di Cramer-von Mises, oppure  $\Psi(x) = [F(x, \theta)(1 - F(x, \theta))]^{-1}$ , che definisce la statistica di Anderson-Darling (*Laio*, 2004). In questo lavoro è stata utilizzata quest’ultima formulazione che, nella pratica, viene calcolata come:

$$A^2 = -m - \frac{1}{m} \sum_{i=1}^m \{ (2i - 1) \ln[F(x_{(i)}, \theta)] + (2m + 1 - 2i) \ln[1 - F(x_{(i)}, \theta)] \}. \quad (11)$$

La statistica  $A^2$  ottenuta dai dati deve essere paragonata alla popolazione delle  $A^2$  che si avrebbero per campioni effettivamente estratti dalla distribuzione ipotetica  $F(x, \theta)$ , sempre con parametri stimati dal campione. Tale popolazione può essere ricavata con una procedura di tipo Monte-Carlo: si genera un grande numero di campioni di lunghezza  $m$  dalla  $F(x, \theta)$ ; per ognuno di essi si stimano i parametri  $\hat{\theta}$  di  $F$  e si calcola la statistica  $A_0^2$  di Anderson-Darling con l’Equazione (11) (con  $\hat{\theta}$  al posto di  $\theta$ ). L’insieme degli  $A_0^2$  così calcolati permette di determinare la funzione di frequenza  $G(A_0^2)$  della statistica test sotto l’ipotesi  $H_0$ ; ad esempio, se si vuole verificare l’adattamento della distribuzione ai dati originali con significatività del 5%, si rigetta l’ipotesi  $H_0$  se l’ $A^2$  calcolato con i dati originali risulta maggiore del quantile 0.95 di  $G(A_0^2)$ .

### 3 APPLICAZIONE

I metodi descritti nei Paragrafi precedenti sono stati applicati per l’analisi di frequenza del deflusso annuo nelle regioni Piemonte e Valle d’Aosta. L’eterogeneità che contraddistingue questo territorio rende particolarmente complicata, e nello stesso tempo interessante, la ricerca di modelli regionali di stima delle variabili idrologiche. In questo ambito spaziale relativamente limitato coesistono infatti conformazioni orografiche e situazioni climatiche estremamente differenti: in poche centinaia di chilometri si va dal clima appenninico-mediterraneo del sud-est collinare piemontese a quello alpino-continentale della montagnosa Valle d’Aosta, passando per tutte le condizioni intermedie.



**Figura 1:** Bacini idrografici del Servizio Idrografico utilizzati nello studio (la numerazione corrisponde al codice dei bacini come indicato in Tabella 1).

### 3.1 Dati utilizzati

Per la stima dei modelli regionali si sono considerati 47 bacini idrografici Piemontesi e Valdostani sottesi da stazioni idrometriche del SIMN (Servizio Idrografico e Mareografico Nazionale) con superfici comprese tra 20 e 8000 km<sup>2</sup> (Figura 1).

I dati medi di afflusso e deflusso sono stati reperiti nella Pubblicazione n. 17 del Servizio Idrografico, che riporta i dati caratteristici dei corsi d'acqua italiani aggiornati all'anno 1970. Le serie storiche dei deflussi sono state integrate fino all'anno 1986, dove possibile, con i dati degli Annali Idrologici del Servizio Idrografico.

Su questi bacini sono state condotte in precedenza un'analisi morfometrica (Viglione, 2003) ed uno studio di stima dell'evapotraspirazione media mensile (Viglione, 2004), i risultati dei quali sono stati utilizzati in questo lavoro.

Le variabili morfoclimatiche dei bacini prese in considerazione per l'analisi regionale comprendono (i valori sono riportati in Tabella 1 per i 47 bacini):

- Afflusso medio annuo  $A_m$  [mm];
- Superficie  $S$  [km<sup>2</sup>];
- Altitudine media  $H_m$  [m s.m.m.];
- Pendenza media  $P_m$  [%]: angolo alla base del triangolo rettangolo che ha per base la radice quadrata dell'area del bacino e per altezza il doppio dell'altitudine mediana (relativa alla sezione di chiusura)

**Tabella 1:** Parametri morfoclimatici considerati nello studio per i bacini idrografici (num è il numero di anni delle serie storiche dei deflussi annui).

cod	nome	num	$D_m$	$A_m$	$S$	$H_m$	$P_m$	$L_{LDP}$	$PLDP$	$S_{2000}$	$EST$	$NORD$	$R_c$	$X_{bar}$	$Y_{bar}$	$I_T$	$I_B$
			[mm]	[mm]	[km <sup>2</sup> ]	[m s.m.m.]	[%]	[km]	[%]	[%]	[%]	[deg]	[deg]				
1	Toce a Cadarese	15	1571	1457	190	2137	22.10	31.6	18.20	66.0	-0.29	-0.96	0.52	8.397	46.375	1.52	0.65
2	Toce a Candoglia	32	1382	1519	1540	1674	7.70	82.4	10.20	36.4	0.63	-0.78	0.31	8.225	46.149	1.31	0.63
3	Niguglia a Omegna	10	1353	1901	122	637	4.80	16.0	8.30	0.0	0.33	0.94	0.41	8.384	45.821	1.28	0.51
4	Ticino a Miorina	65	1395	1695	6692	1286	2.60	168.1	7.80	20.2	0.00	-1.00	0.30	8.652	46.169	1.32	0.56
5	SBernardino a Santino	15	1730	2113	119	1251	17.90	22.6	26.30	2.4	0.51	-0.86	0.53	8.456	46.035	1.92	0.45
6	Mastallone a PonteFolle	32	1600	1936	147	1319	12.60	23.8	22.60	6.3	0.53	-0.85	0.49	8.206	45.888	1.69	0.50
7	Cervo a Passobrevé	13	1461	1803	75	1490	20.20	14.4	22.90	13.5	0.70	-0.71	0.62	7.978	45.679	1.49	0.54
8	Sesia a Campertogno	26	1275	1427	170	2112	19.80	21.8	26.10	57.4	0.90	-0.44	0.49	7.936	45.838	1.10	0.77
9	Sesia a PonteAranco	24	1428	1735	703	1491	8.30	62.2	16.20	21.9	0.75	-0.66	0.47	8.091	45.833	1.04	0.68
10	DoraBaltea a Tavagnasco	61	918	949	3311	2090	6.60	110.9	10.80	58.1	0.85	-0.52	0.39	7.395	45.728	0.50	1.04
11	Orco a PonteCanavese	44	1034	1263	615	1924	12.10	47.9	18.60	46.7	0.93	-0.36	0.43	7.425	45.470	0.98	0.79
12	SturaLanzo a Lanzo	49	1090	1296	577	1773	10.60	40.3	21.00	37.4	0.99	-0.17	0.54	7.287	45.290	0.90	0.77
13	Chisone a SoucheresBasses	10	819	966	92	2222	15.40	17.0	17.60	73.3	0.25	0.97	0.49	6.938	44.974	0.48	1.06
14	Chisone a SMartino	34	694	1058	581	1730	11.20	56.6	13.70	36.9	0.87	-0.49	0.49	7.084	44.963	0.48	0.96
15	Chisone a Fenestrelle	23	654	910	157	2144	15.90	26.6	15.30	64.5	0.88	0.47	0.41	6.965	45.001	0.37	1.12
16	DoraRiparia a Oulx	30	663	851	254	2165	13.20	34.9	16.80	63.9	-0.09	1.00	0.46	6.851	44.932	0.24	1.20
17	DoraRiparia a SAntonino	27	591	841	993	1867	9.90	78.0	11.80	46.3	0.99	0.16	0.24	6.912	45.070	0.16	1.20
18	Po a Crissolo	39	1254	1271	38	2261	28.80	8.4	29.30	73.7	0.97	0.25	0.74	7.115	44.693	1.03	0.83
19	Po a Moncalieri	44	507	952	5032	924	0.80	114.0	5.40	14.5	0.61	0.79	0.39	7.398	44.736	0.18	1.10
20	Grana a Monterosso	41	811	1135	103	1565	15.00	19.0	19.50	20.5	0.99	0.13	0.54	7.240	44.403	0.67	0.93
21	SturaDemonte a Pianche	23	925	1112	180	2074	17.30	26.8	16.10	61.8	0.83	-0.55	0.47	7.007	44.356	0.76	0.96
22	RioBagni a BagniVinadio	20	1241	1398	62	2138	19.10	9.7	26.00	66.6	0.72	0.70	0.73	7.053	44.267	1.25	0.77
23	Vermenagna a Limone	16	1128	1364	57	1677	17.60	10.7	23.10	20.9	0.06	1.00	0.56	7.576	44.178	1.09	0.79
24	RioPiz a Pietrapozio	23	1272	1273	21	2194	38.70	8.3	24.70	71.5	0.42	0.91	0.55	7.018	44.311	1.06	0.84
25	SturaDemonte a Gaiola	18	1011	1219	560	1814	10.20	55.3	12.10	43.1	1.00	0.09	0.41	7.137	44.316	0.85	0.88
26	GessoValletta a SLorenzo	10	1384	1392	110	2105	22.30	17.1	19.50	61.2	0.79	0.61	0.59	7.277	44.212	1.28	0.77
27	GessoEntracque a Entracque	12	1404	1468	157	1894	17.00	16.9	23.80	44.4	-0.10	0.99	0.61	7.407	44.182	1.31	0.73
28	Tanaro a Montecastello	58	501	997	8024	651	0.80	209.8	6.40	6.0	0.75	0.66	0.28	8.064	44.548	0.18	1.07
29	Tanaro a PonteNava	34	1030	1281	148	1576	11.30	19.5	23.80	17.6	1.00	-0.04	0.50	7.771	44.124	0.90	0.84
30	Tanaro a Nucetto	31	902	1233	376	1222	7.20	55.4	16.10	7.8	0.60	0.80	0.28	7.901	44.179	0.65	0.87
31	Tanaro a Farigliano	36	776	1120	1516	938	2.70	93.2	12.30	5.2	0.17	0.98	0.54	7.852	44.298	0.47	0.96
32	Corsaglia a Molline	29	1068	1366	89	1513	17.80	18.8	20.60	17.0	0.17	0.99	0.58	7.828	44.226	1.19	0.79
33	Scrivia a Serravalle	25	827	1389	616	688	3.50	51.9	8.10	0.0	-0.80	0.60	0.51	9.040	44.628	0.77	0.76
34	BormidaMallare a Ferrania	22	965	1228	50	602	5.90	18.0	9.60	0.0	0.26	0.97	0.39	8.300	44.297	0.38	0.87
35	Erro a Sassello	16	882	1200	83	605	5.30	17.6	6.40	0.0	0.00	1.00	0.36	8.458	44.447	0.30	0.88
36	Bormida a Cassine	12	510	971	1542	481	1.80	131.1	5.80	0.0	0.50	0.87	0.35	8.322	44.500	0.04	1.09
37	Borbera a Baracche	18	779	1220	202	867	6.40	25.3	13.40	0.0	-0.80	0.61	0.57	9.112	44.668	0.57	0.86
38	Vobbia a Vobbietta	13	835	1461	57	727	9.70	14.9	14.90	0.0	-0.81	0.58	0.55	9.046	44.605	0.88	0.72
39	DoraRhemes a Pelaud	6	1453	1041	54	2743	24.00	12.6	19.60	97.8	0.36	0.93	0.57	7.091	45.514	1.09	0.97
40	GrandEyvia a Cretaz	9	1109	940	179	2593	17.10	15.3	25.80	86.4	-0.60	0.80	0.53	7.377	45.584	0.75	1.06
41	DoraBaltea a Aosta	17	898	952	1824	2267	8.40	55.7	15.40	67.9	0.99	0.11	0.29	7.177	45.718	0.60	1.04
42	Lys a Gressoney	23	1357	1191	91	2625	24.90	16.4	22.50	84.2	-0.03	-1.00	0.59	7.830	45.855	1.23	0.83
43	Rutor a Promise	31	1648	1314	46	2512	31.10	10.8	27.30	89.7	-0.19	0.98	0.53	6.970	45.672	1.26	0.75
44	Artanavaz a StOyen	17	1023	1283	71	2229	22.30	11.9	22.10	71.8	0.98	-0.19	0.59	7.151	45.828	1.15	0.76
45	Evancon a Champoluc	30	977	1048	105	2631	20.20	15.1	21.20	88.6	-0.31	-0.95	0.54	7.742	45.872	0.96	0.94
46	Ayasse a Champorcher	24	1258	1179	41	2352	30.40	12.4	17.90	82.4	0.98	0.19	0.56	7.559	45.613	0.95	0.84
47	Savara a EauRousse	15	1079	987	84	2723	22.90	11.5	28.30	95.8	0.10	1.00	0.60	7.206	45.523	1.11	1.02

del bacino. In sostanza tale pendenza è calcolata rispetto ad un bacino di forma quadrata equivalente a quello reale, e non tiene conto della sua effettiva forma, che può essere più o meno allungata. Si è scelto di calcolare  $P_m$  in questo modo, piuttosto che con i classici strumenti di analisi del DEM (Digital Elevation Model), in modo che il suo valore non fosse influenzato dalla risoluzione del DEM stesso;

- Lunghezza del Longest Drainage Path  $L_{LDP}$  [km]: lunghezza del percorso tra la sezione di chiusura ed il punto più lontano da essa, sul bordo del bacino, seguendo le direzioni di drenaggio. Esso coincide sostanzialmente con l'asta principale;
- Pendenza media del Longest Drainage Path  $P_{LDP}$  [%]: valore medio delle pendenze associate ad ogni pixel del Longest Drainage Path;
- Percentuale d'area a quota superiore ai 2000 m s.m.m.  $S_{2000}$  [%];
- Easting  $EST$ : seno dell'angolo formato dal vettore di orientamento con il nord. Il vettore di orientamento è quel segmento che unisce il baricentro del bacino con la sezione di chiusura. Il parametro  $EST$  ha valore massimo uguale a 1 se il bacino è orientato verso est e valore minimo pari a -1 se è orientato verso ovest;
- Northing  $NORD$ : coseno dell'angolo formato dal vettore di orientamento con il nord che ha valore massimo uguale a 1 se il bacino è orientato verso nord e valore minimo pari a -1 se è orientato verso sud;
- Rapporto di circolarità  $R_c$ : rapporto tra l'area del bacino e l'area del cerchio avente lo stesso perimetro del bacino;
- Coordinate del baricentro del bacino  $X_{bar}$  e  $Y_{bar}$  [deg] espresse come longitudine e latitudine (nel sistema di riferimento di Greenwich) in gradi esadecimali;
- Indice di Thornthwaite  $I_T$ : indice di umidità globale che, nella forma più semplice, si presenta come un indice di bilancio idrico a scala annua:

$$I_T = \frac{A_m - ET_p}{ET_p}, \quad (12)$$

dove  $A_m$  è la precipitazione media annua ed  $ET_p$  l'evapotraspirazione potenziale media annua relative al bacino. La seconda grandezza è stata stimata tramite la formulazione di Hargreaves (v.es. *Viglione, 2004*);

- Indice di Budyko  $I_B$ : indice di aridità radiazionale che si esprime come:

$$I_B = \frac{R_n}{\lambda A_m} \quad (13)$$

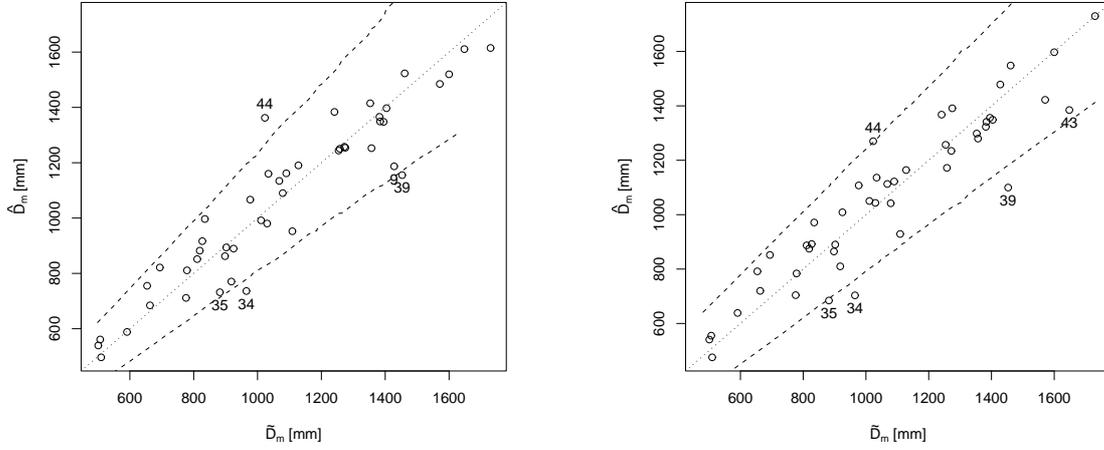
dove  $R_n$  è la radiazione netta media annua e  $\lambda$  è il calore latente di vaporizzazione. I valori assunti da  $I_B$  sono inferiori all'unità in regioni umide e superiori in regioni aride. La radiazione netta è stata stimata nel lavoro di *Viglione (2004)*.

### 3.2 Regionalizzazione del deflusso indice

Per la stima di  $D_m$  si sono considerate tutte le possibili regressioni lineari tra i deflussi medi campionari  $\tilde{D}_m$  dei 47 bacini e le variabili morfoclimatiche descritte nel Paragrafo 3.1. La regressione risultata migliore in base ai criteri descritti nel Paragrafo 2.1 è la seguente:

$$\ln(\hat{D}_m) = 7.86 + 2.91 \cdot 10^{-4} \cdot H_m + 7.22 \cdot 10^{-2} \cdot NORD - 1.70 \cdot I_B, \quad (14)$$

caratterizzata da un coefficiente di determinazione  $R_{adj}^2 = 0.900$  e da  $RMSE_{cv} = 110.5$  mm (riferito alla variabile non trasformata  $D_m$ ). Delle variabili utilizzate per la scrittura dell'Equazione (14) l'indice di Budyko  $I_B$  non è di facile stima. La determinazione della radiazione solare netta  $R_n$  sul bacino (si veda l'Equazione (13)) richiede la conoscenza delle distribuzioni spaziali di temperatura, umidità relativa



**Figura 2:** Risultato delle cross validazioni per il modello di Equazione (14) a sinistra e di Equazione (15) a destra; si sono indicati i siti caratterizzati dagli errori maggiori: Sesia a Ponte Aranco (9), Bormida di Mallare a Ferrania (34), Erro a Sassello (35), Dora di Rhemes a Pelaud (39), Rutor a Promise (43) e Artavanaz a St.Oyen (44).

e nuvolosità media (o stime di esse). Per questo motivo si è deciso di considerare anche modelli i cui parametri fossero di più semplice determinazione. Il migliore tra questi è risultato essere:

$$\hat{D}_m^{1/3} = -22.7 + 4.37 \cdot \ln(A_m) + 10^{-3} \cdot H_m, \quad (15)$$

caratterizzato da un coefficiente di determinazione  $R_{adj}^2 = 0.883$  e da  $RMSE_{cv} = 115.8$  mm (riferito alla variabile non trasformata  $D_m$ ). I grafici di Figura 2 riportano il risultato delle cross-validazioni: il deflusso stimato in ogni bacino è stato ottenuto escludendo il bacino stesso dalla taratura delle Regressioni (14) e (15); le linee tratteggiate rappresentano l'intervallo di previsione del 95% (v.es. *Montgomery et al.*, 2001), ovvero la fascia entro la quale dovrebbe ricadere il 95% dei valori stimati con il modello se le ipotesi fatte fossero corrette.

Una relazione analoga alla (15) è stata ottenuta nella regionalizzazione del deflusso annuo in Basilicata (*Claps et al.*, 1998).

### 3.3 Regionalizzazione della curva di crescita

Si sono considerate le 38 stazioni idrometriche la cui serie storica supera i 10 anni di osservazioni. Per valutare la distanza tra le curve di crescita campionarie  $\hat{x}(F)$  si è utilizzata la differenza tra gli  $L$ -CV campionari delle serie, il che equivale ad utilizzare la statistica test di *Hosking & Wallis* (1993). In questo caso l'utilizzo delle matrici delle distanze non è indispensabile, essendo possibile anche procedere con un'analisi regressiva tra gli  $L$ -CV campionari e le variabili morfoclimatiche (cosa che è stata fatta ottenendo gli stessi risultati). È stata però utilizzata la metodologia descritta nel Paragrafo 2.2.1 per dimostrarne la validità e per la sua maggiore generalità: nel caso in cui risultasse conveniente utilizzare un altro test di omogeneità (*Viglione et al.*, 2006), ad esempio nell'analisi regionale delle piene, il ricorso all'analisi regressiva non sarebbe possibile.

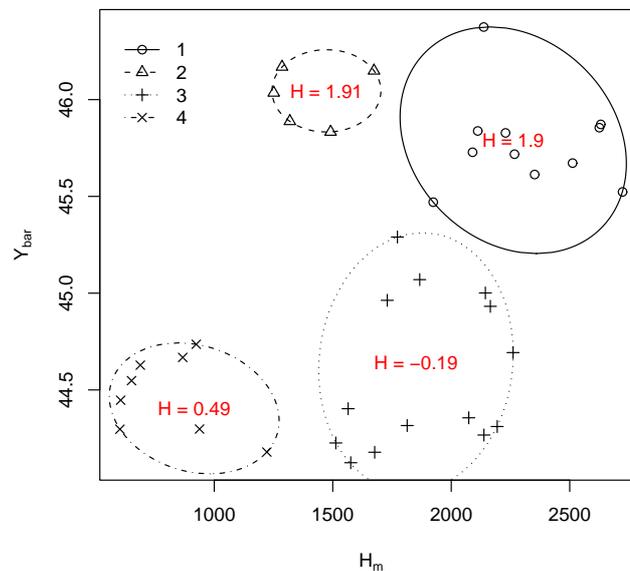
Eseguendo il test di Mantel semplice e parziale tra  $\Delta_{L-CV}$  e le matrici delle distanze delle grandezze morfoclimatiche per tutte le combinazioni possibili, si è ottenuto che la migliore tra le correlazioni significative è quella tra  $\Delta_{L-CV}$  e le matrici  $\Delta_{H_m}$  e  $\Delta_{Y_{bar}}$ . I coefficienti di correlazione tra la matrice delle distanze degli  $L$ -CV ed ognuna delle matrici dei parametri morfoclimatici condizionata all'altra (Equazione 7) valgono:

$$r(\Delta_{L-CV} \Delta_{H_m} \Delta_{Y_{bar}}) = 0.431$$

$$r(\Delta_{L-CV} \Delta_{Y_{bar}} \Delta_{H_m}) = 0.295$$

In entrambi i casi il test di Mantel parziale con livello di significatività dell'1% consente di rigettare l'ipotesi di indipendenza tra le matrici. Per questo motivo la quota media  $H_m$  e la latitudine del baricentro  $Y_{bar}$  dei bacini sono state scelte come variabili di classificazione per suddividere in gruppi i bacini del SIMN.

Applicando la procedura mista di cluster analysis descritta nel Paragrafo 2.2.2 ed arrestando la procedura al superamento del test di omogeneità di Hosking & Wallis (1993) in tutti i gruppi, si sono individuati i 4 cluster rappresentati in Figura 3. Per due delle regioni individuate i valori della statistica  $H$  sono di poco inferiori a 2, per cui tali regioni dovrebbero ritenersi "possibilmente eterogenee", ma si è deciso di accettare questo raggruppamento per evitare di dover utilizzare gruppi troppo piccoli.



**Figura 3:** Suddivisione dei bacini in 4 regioni omogenee sul piano  $H_m - Y_{bar}$ ; per ogni gruppo è stato indicato il valore della statistica di omogeneità di Hosking e Wallis.

La scelta della distribuzione di probabilità caratteristica di ogni gruppo è stata effettuata considerando distribuzioni a 3 parametri (v.es. *Basson et al.*, 1994) per la stima dei quali il numero elevato di osservazioni disponibili per ogni regione determina adeguate condizioni di robustezza. La tecnica di "model

**Tabella 2:** Stima della probabilità  $G(A^2)$ , associata alla statistica di Anderson-Darling di Equazione (11); valori superiori a 0.9 comportano il rigetto della distribuzione ipotetica.

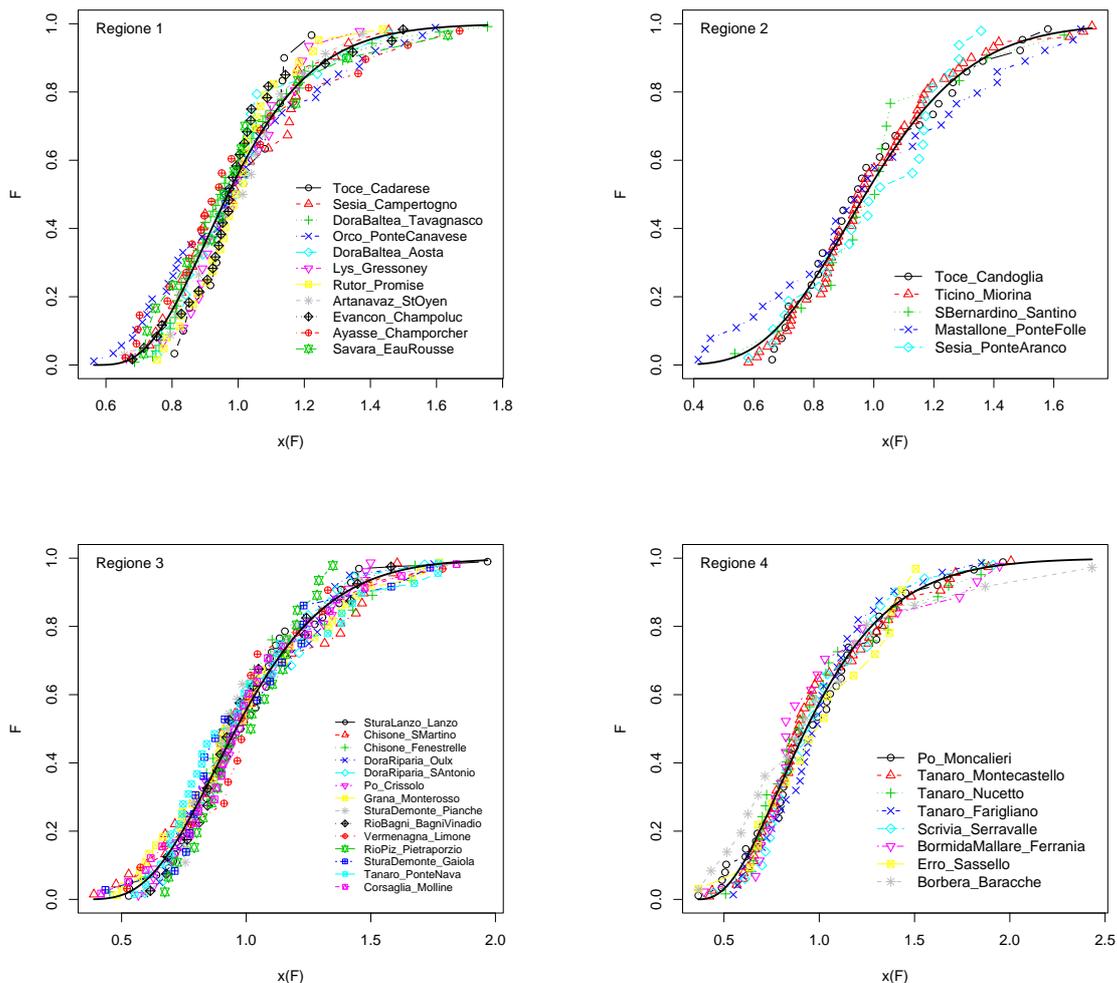
	REGIONE			
	1	2	3	4
GP	1.00	1.00	1.00	1.00
GEV	0.172	0.390	0.917	0.849
GL	0.474	0.848	0.999	0.986
LN3	0.300	0.425	0.924	0.867
P3	0.589	0.388	0.863	0.816

selection" descritta nel Paragrafo 2.2.4 è stata applicata sui quattro raggruppamenti definiti in precedenza, considerando le seguenti distribuzioni di probabilità:

- distribuzione di Pareto generalizzata (GP);

- distribuzione generalizzata dei valori estremi (GEV);
- distribuzione logistica generalizzata (GL);
- distribuzione log-normale a 3 parametri (LN3);
- distribuzione di Pearson tipo III (P3);

delle quali solo le ultime due sono comunemente proposte per la descrizione dei deflussi annui (v.es. *Vogel & Wilson*, 1996). I parametri delle distribuzioni sono stati stimati con il metodo degli  $L$ -momenti (*Hosking & Wallis*, 1997) utilizzando il campione ottenuto in ogni regione raggruppando. In Tabella 2 sono riportati i valori della probabilità  $\hat{G}(A^2)$ , associata alla statistica di Anderson-Darling di Equazione (11): tanto più vicino a zero è il valore ottenuto, tanto migliore può ritenersi l'adattamento della distribuzione al campione. Dal momento che tutte le distribuzioni prese in considerazione sono a 3 parametri, il valore di  $\hat{G}(A^2)$  può essere effettivamente utilizzato per confronto. Poiché si è scelto di utilizzare un solo tipo di distribuzione per tutto il territorio considerato, si è scelta la distribuzione di Pearson (o Gamma a 3 parametri) che, nel caso del terzo gruppo di bacini, è l'unica che non viene rigettata dal test con significatività 10%.



**Figura 4:** Curve di crescita campionarie e curva di crescita regionale per le regioni omogenee individuate.

La distribuzione di probabilità di Pearson tipo III è definita come:

$$f(x) = \frac{(x - \xi)^{\alpha-1} e^{-(x-\xi)/\beta}}{\beta^\alpha \Gamma(\alpha)}, \quad F(x) = G\left(\alpha, \frac{x - \xi}{\beta}\right) / \Gamma(\alpha), \quad (16)$$

dove  $f(x)$  è la funzione densità di probabilità,  $F(x)$  è la distribuzione cumulata,  $\xi$  è il parametro di posizione,  $\beta$  quello di scala,  $\alpha$  quello di forma,  $\Gamma$  è la funzione gamma e  $G$  la funzione gamma incompleta (v.es. Kottegoda & Rosso, 1998).

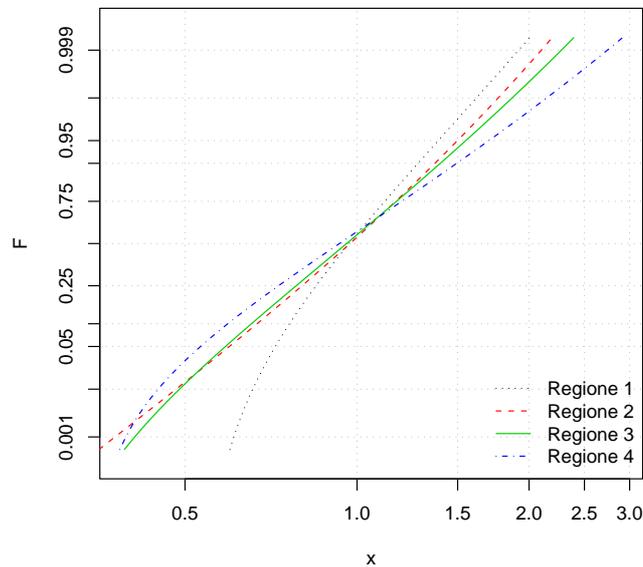
La curva di crescita è definita come:

$$x(F) = \xi + \beta \cdot \Theta(F, \alpha), \quad (17)$$

dove  $\Theta(F, \alpha)$  è l'inversa della funzione gamma generalizzata. Per le quattro regioni individuate si sono ottenuti i seguenti valori dei parametri:

$$\begin{array}{lll} \alpha_1 = 4.764 & \beta_1 = 9.534 \cdot 10^{-2} & \xi_1 = 0.5458 \\ \alpha_2 = 10.60 & \beta_2 = 8.508 \cdot 10^{-2} & \xi_2 = 9.843 \cdot 10^{-2} \\ \alpha_3 = 5.817 & \beta_3 = 0.1237 & \xi_3 = 0.2801 \\ \alpha_4 = 3.107 & \beta_4 = 0.2093 & \xi_4 = 0.3496 \end{array} \quad (18)$$

In Figura 4 si sono rappresentate le curve di crescita regionali sovrapposte a quelle campionarie relative alle singole stazioni. Osservando i grafici relativi alle regioni 1 e 2 si comprende perché la misura di eterogeneità di Hosking e Wallis si avvicina al valore di non accettazione. In Figura 5 le quattro curve regionali sono state



**Figura 5:** Rappresentazione delle curve di crescita regionali in carta probabilistica lognormale.

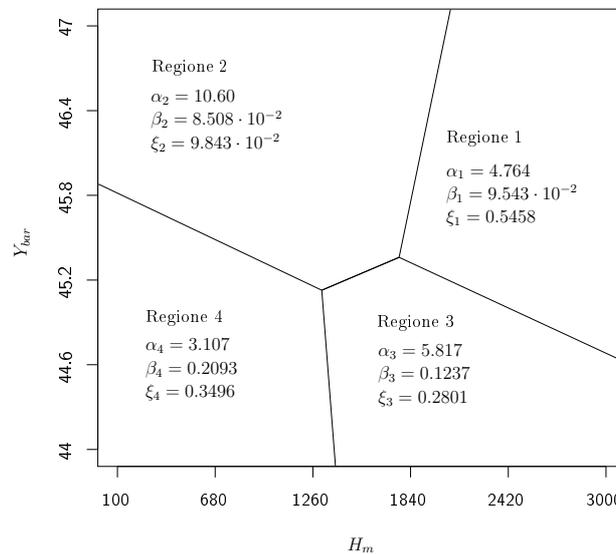
rappresentate tutte insieme in carta probabilistica lognormale in modo da farne risaltare le differenze. Si noti come le curve di crescita siano significativamente asimmetriche (non-normali), presentando i seguenti coefficienti di asimmetria:

$$\gamma_1 = 0.974, \quad \gamma_2 = 0.624, \quad \gamma_3 = 0.803, \quad \gamma_4 = 1.106.$$

### 3.4 Utilizzo del modello regionale

Volendo esemplificare la procedura, se si vuole stimare il deflusso annuo  $D(T)$  che non viene raggiunto, in media, una volta ogni  $T$  anni ( $T = 1/F$ ) si procede come segue:

- se nel sito di interesse è disponibile una serie storica con numerosità sufficiente (almeno  $n > 2T$ ),  $D(T)$  può essere stimato con un'analisi di frequenza locale senza ricorrere alla tecnica di regionalizzazione;
- se sono disponibili solo pochi dati, ad esempio 5-10,  $D_m$  viene stimato come media aritmetica dei deflussi annui misurati, mentre per  $x(T)$  si determina quale tra le quattro curve di crescita individuate nel Paragrafo 3.3 sia da utilizzare. La scelta è fatta in base all'assegnazione del bacino idrografico sotteso dalla sezione considerata ad una delle quattro regioni omogenee, entrando nel diagramma rappresentato in Figura 6 con i valori delle variabili quota media  $H_m$  e latitudine del baricentro del bacino  $Y_{bar}$ . Le linee di separazione nel diagramma di Figura 6 sono state ricavate con il criterio



**Figura 6:** Suddivisione del piano  $H_m - Y_{bar}$  nelle 4 regioni omogenee; per ognuna di esse sono stati indicati i parametri della curva di crescita regionale (distribuzione Tipo III di Pearson).

della minima distanza tra il baricentro delle regioni omogenee (rappresentate in Figura 3) e i punti dello spazio delle variabili morfoclimatiche;

- per una sezione sprovvista di misure,  $D_m$  viene stimato tramite uno dei modelli regressivi proposti (Equazioni (14) e (15)) mentre per  $x(T)$  si procede come nel caso precedente. Anche in questo caso i parametri necessari all'applicazione del modello regionale sono pochi e di facile determinazione: se si utilizza il modello regressivo più semplice (Equazione (15)) i parametri richiesti sono solamente l'afflusso medio annuo  $A_m$ , la quota media  $H_m$  e la latitudine del baricentro del bacino  $Y_{bar}$ .

Una volta stimati il deflusso indice e la curva di crescita, il deflusso annuo corrispondente al tempo di ritorno  $T$  sarà ottenuto come  $D(T) = D_m \cdot x(T)$ .

## CONCLUSIONI

In questo lavoro si è utilizzato il metodo della grandezza indice per l'analisi di frequenza regionale del deflusso annuo nella regione idrografica che comprende il Piemonte e la Valle d'Aosta. Si è esaminata la

dipendenza del deflusso medio annuo da numerose variabili morfoclimatiche pervenendo ad una regressione multipla con grandezze facilmente determinabili. Nella definizione delle curve di crescita regionali, è stata proposta una tecnica basata sul confronto tra matrici delle distanze, che consente di determinare oggettivamente i parametri morfoclimatici più adatti ad indirizzare il raggruppamento dei siti in regioni statisticamente omogenee. Per la suddivisione effettiva è stata applicata una metodologia mista di cluster analysis (algoritmo gerarchico e ottimizzazione numerica) seguita da una verifica di omogeneità. I risultati dei test e la qualità degli adattamenti ottenuti confermano la validità della metodologia proposta.

## Ringraziamenti

Questo lavoro è stato sviluppato nell'ambito del progetto C.I.P.E. 2004 (Comitato Interministeriale per la Programmazione Economica) della Regione Piemonte e col supporto di fondi MIUR (PRIN 2003 e 2005).

## BIBLIOGRAFIA

- Basson M., Allen R., Pegram G., van Rooyen J., Probabilistic management of water resource and hydropower systems. Water Resources Publications (1994).
- Brath A., Camorani G., Castellarin A., *Una tecnica di stima regionale della curva di durata delle portate in bacini non strumentati*. In *XXIX Convegno di Idraulica e Costruzioni Idrauliche*, volume 2, pp. 391–398, Trento. Università di Trento (2004).
- Claps P., Mancino L., *Impiego di classificazioni climatiche quantitative nell'analisi regionale del deflusso annuo*. In *XXVIII Convegno di Idraulica e Costruzioni Idrauliche*, pp. 169–178, Potenza (2002). 16-19 settembre 2002.
- Claps P., Fiorentino M., Silvagni G., *Studio per la valorizzazione e la salvaguardia delle risorse idriche in basilicata. valutazione delle risorse idriche e possibilità di regolazione dei deflussi*. Regione Basilicata (1998).
- Dalrymple T., Flood frequency analyses, volume 1543-A di *Water Supply Paper*. U.S. Geological Survey, Reston, Va. (1960).
- Fabbris L., *Statistica multivariata: analisi esplorativa dei dati*. Serie di Matematica. McGraw-Hill Libri Italia srl, Milano (1997).
- Ferraresi M., Todini E., Franchini M., *Un metodo per la regionalizzazione dei deflussi medi*. In *XXI Convegno di Idraulica*, L'Aquila (1988).
- Hosking J., Wallis J., *Some statistics useful in regional frequency analysis*. Water Resources Research (1993), 29 (2), pp. 271–281.
- Hosking J., Wallis J., *Regional Frequency Analysis: An Approach Based on L-Moments*. Cambridge University Press (1997).
- Kottogoda N., Rosso R., *Statistics, probability, and reliability for civil and environmental engineers*. McGraw-Hill, International Edition (1998).
- Laio F., *Cramer-von mises and anderson-darling goodness of fit tests for extreme value distributions with unknown parameters*. Water Resources Research (2004), 40, pp. W09308, doi:10.1029/2004WR003204.
- Legendre P., *Comparison of permutation methods for the partial correlation and partial mantel tests*. Journal of Statistical Computation and Simulation (2000), 67, pp. 37–73.
- Mantel N., Valand R., *A technique of nonparametric multivariate analysis*. Biometrics (1970), 26, pp. 547–558.
- Montgomery D., Peck E., Vining G., *Introduction to linear regression analysis*. Wiley, New York (2001).
- Oksanen J., Kindt R., O'Hara R. B., *vegan: Community Ecology Package* (2005). R package version 1.6-10.
- R Development Core Team, *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria (2006). ISBN 3-900051-07-0.
- Robson A., Reed D., *Statistical procedures for flood frequency estimation*. In *Flood Estimation Handbook*, volume 3. Institute of Hydrology Crowmarsh Gifford, Wallingford, Oxfordshire (1999).
- Viglione A., *Determinazione automatica di parametri morfometrici dei bacini idrografici*. Dipartimento di Idraulica, Trasporti e Infrastrutture Civili, Politecnico di Torino (2003). Working Paper 2003-01, disponibile su <http://www2.polito.it/strutture/ditic/Ricerca/>

Working\_papers.

Viglione A., *Stima dell'evapotraspirazione media mensile sul territorio piemontese*. Dipartimento di Idraulica, Trasporti e Infrastrutture Civili, Politecnico di Torino (2004). Working Paper 2004-01 disponibile su [http://www2.polito.it/strutture/ditic/Ricerca/Working\\_papers](http://www2.polito.it/strutture/ditic/Ricerca/Working_papers).

Viglione A., Laio F., Claps P., *A comparison of homogeneity tests for regional frequency analysis* (2006). In revisione.

Vogel R., Wilson I., *Probability distribution of annual maximum, mean, and minimum streamflows in the united states*. Journal of Hydrologic Engineering (1996), 1 (2), pp. 69–76.

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