

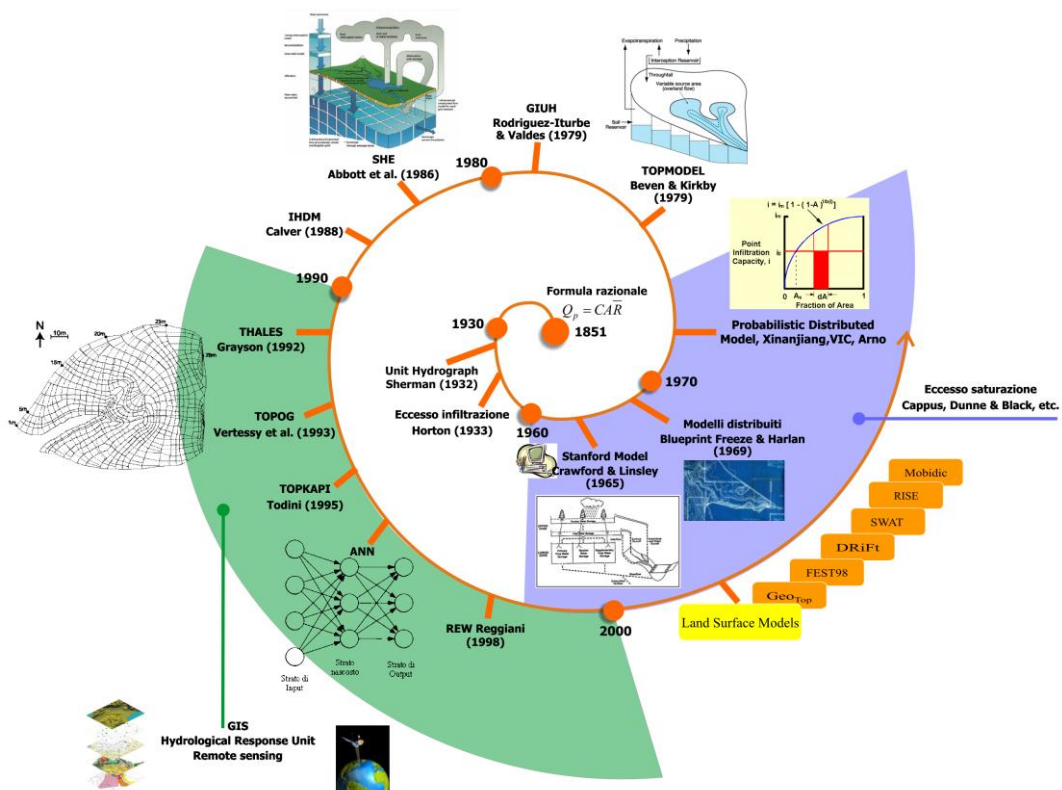
# Università della Calabria

## FACOLTA' DI INGEGNERIA

### Corso di Laurea in Ingegneria Civile

CORSO DI IDROLOGIA N.O.

Prof. Pasquale Versace



## SCHEDA DIDATTICA N°12

### ANALISI STATISTICA DELLE VARIABILI CASUALI IDROLOGICHE

## ANALISI STATISTICA DELLE VARIABILI CASUALI IDROLOGICHE

### PREMESSA

Per la progettazione di opere idrauliche, sia che provvedano alla difesa idraulica del territorio e delle attività che in esso si svolgono, sia che interessino l'utilizzazione delle acque, è necessario identificare le caratteristiche idrologiche dell'area interessata. In molti casi è sufficiente sviluppare un'analisi statistica considerando le variabili idrologiche come variabili casuali caratterizzate da una specifica distribuzione di probabilità.

Le variabili casuali di interesse possono essere diverse a seconda del caso di studio, ma in ogni caso devono essere correttamente individuate e ben definite. Quando si affrontano problemi di uso delle risorse idriche le variabili casuali da considerare possono essere:

- portata media annua;
- pioggia annua;
- portata media nel periodo secco;
- pioggia nel periodo secco;
- massimo annuale del numero di giorni consecutivi non piovosi;
- minimo annuale del deflusso medio in  $k$  giorni consecutivi ( $k = 1, \dots, m$ ) con  $m$  che in Calabria può essere posto pari a 100-150.

Quando, invece, si affrontano problemi di difesa dalle piene o dalle inondazioni o dagli allagamenti le variabili casuali che interessano sono:

- massimo annuale della portata al colmo;
- massimo annuale della portata media in  $k$  ore consecutive ( $k$  varia tra 0.5 e 1.2);
- massimo annuale della portata media giornaliera;
- massimo annuale delle piogge di  $t$  ore ( $t$  varia tra 0.5 e 24 ore);
- massimo annuale delle piogge di  $k$  giorni consecutivi ( $k$  varia tra 1 e 5 o 10).

Come si vede da questi esempi nell'identificazione della variabile idrologica di interesse, da considerare come variabile casuale, è importante stabilire con chiarezza il passo temporale di misura (1 ore,  $t$  ore, 1 giorno,  $k$  giorni, etc.), il valore da considerare (massimo, minimo, media, valore totale), l'orizzonte temporale di riferimento (anno, periodo secco, mese, etc.).

L'applicazione delle metodologie statistiche presenta naturalmente alcune limitazioni e può comportare incertezze ed errori anche rilevanti. Le serie storiche, infatti, dovrebbero rispettare le condizioni di omogeneità e stazionarietà, che possono però venir meno per effetto di rilevanti

interventi antropici. I dati disponibili, inoltre, non sempre sono rilevati con tecniche adeguate e sono, quindi, affetti da notevole incertezza.

L'analisi dei valori misurati della grandezza idrologica di interesse consente, comunque, di inquadrare, nell'ambito delle approssimazioni accettabili, i fenomeni idrologici e di provvedere, quindi, con la progettazione, alla soluzione dei diversi problemi tecnici, definendo quantitativamente anche il rischio di insuccesso.

La prima fase dell'analisi idrologica individua le grandezze idrologiche di interesse e le stazioni nelle quali sono misurate. Si procede quindi alla raccolta dei dati, alla loro analisi statistica ed alla valutazione dei valori che le variabili indagate potranno assumere nel futuro.

La stima dei valori di una qualsiasi variabile idrologica consiste essenzialmente nella identificazione della relazione  $x = x(T)$ , che lega la variabile al periodo di ritorno  $T$ , o, che è lo stesso (vedi pag.4), nella determinazione della funzione di probabilità cumulata  $F_X(x)$ .

La metodologia consueta adottata nelle analisi statistiche è applicata attraverso le seguenti fasi:

- *formulazione dell'ipotesi di lavoro* riguardo la composizione della popolazione della  $X$ , individuando quale fra le leggi del calcolo delle probabilità meglio si adatta a descrivere la suddetta popolazione;
- *precisazione dell'ipotesi di lavoro* avanzata, calcolando i parametri, in funzione degli  $N$  valori osservati della  $X$ . In questo modo si trovano le migliori stime dei parametri che caratterizzano la distribuzione utilizzata traendo dal campione la massima informazione sulla popolazione da cui esso proviene;
- *verifica dell'ipotesi di lavoro*, con test statistici atti a verificare che gli scarti che si riscontrano fra la composizione della popolazione e quella del campione siano o meno significativi.

Le analisi statistiche dei fenomeni idrologici si basano essenzialmente sull'uso delle distribuzioni di probabilità, quindi, è utile un breve richiamo delle principali definizioni di statistica e delle distribuzioni di probabilità più comunemente adoperate nello studio di tali grandezze.

#### **ALCUNE DEFINIZIONI DI BASE**

Indicando con  $X$  una variabile casuale è possibile individuare una funzione  $F_X(x)$  che, per ciascun valore reale  $x$  che la  $X$  può assumere, misuri la probabilità che si verifichino valori di  $X$  minori o al più uguali a  $x$ :

$$F_X(x) = P(X \leq x)$$

La funzione  $F_X(x)$  è detta *funzione di probabilità cumulata* ed in generale è funzione, oltre che del valore della variabile casuale in esame, di altri parametri  $\beta_i$  da stimare.

$$F_X(x) = f(x, \beta_i)$$

Il frattile  $x_p$  indica il valore della variabile casuale  $X$  che ha probabilità  $P$  di non essere superato:

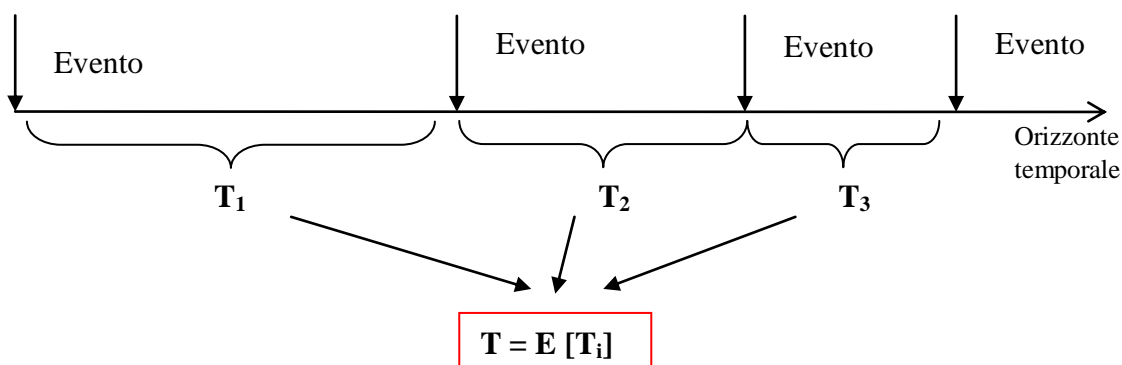
$$x_p : F_X(x_p) = P(X \leq x_p) = P$$

Si definisce **popolazione** l'insieme di tutti i valori che la variabile casuale  $X$  può assumere in determinate condizioni ambientali, ovvero in condizioni che sono rimaste immutate rispetto al passato e che resteranno tali anche in futuro, almeno per i tempi che interessano la sede applicativa.

Si definisce **campione di dimensione  $N$** , la serie statistica costituita da  $N$  valori  $x_1, x_2, \dots, x_i, \dots, x_N$  assunti dalla  $X$  in una determinata stazione di misura o estratti in maniera casuale dalla popolazione.

Si definisce **periodo di ritorno  $T$**  associato ad un dato valore  $x$  di una variabile casuale  $X$ , il numero medio di anni che bisogna attendere perché si verifichi nuovamente un determinato evento.

Il termine evento può generare equivoci e deve essere pertanto correttamente definito. In generale lo scopo dell'analisi idrologica è l'identificazione del cosiddetto valore di progetto (pioggia di progetto, portata di progetto) in base alla quale si dimensiona l'opera. L'evento si ha quando il valore di progetto si dimostra inadeguato.



In particolare, nel caso di problemi connessi all'uso delle risorse idriche, l'evento è costituito dal fatto che la variabile casuale  $X$  assuma valori inferiori al valore di progetto  $x$ , che cioè le disponibilità idriche risultino inferiori a quelle previste

**Evento**       $X < x$

Nel caso di problemi di difesa dalle piene l'evento è, invece, rappresentato dal superamento del valore di progetto  $x$ , e che quindi si verifichino piogge o portate maggiori di quelle previste.

**Evento**  $X > x$

L'approccio statistico consente di associare a ciascun valore del periodo di ritorno  $T$  un valore  $x$  di  $X$ . E' necessario però conoscere il legame tra periodo di ritorno e probabilità cumulata.

Possiamo immaginare che ogni anno sia una prova indipendente ed ipotizzare che la probabilità  $P$  che si verifichi l'evento sia ogni anno la stessa.

Nei due casi considerati si ha rispettivamente:

<b>Evento</b>	<b>P[Evento]</b>
$X < x$	$P = P[X < x] = F_X(x)$
$X > x$	$P = P[X > x] = 1 - F_X(x)$

Il modello probabilistico che descrive la probabilità di avere per la prima volta l'evento considerato (il primo successo) all' $n$ -esima prova è la distribuzione geometrica:

$$P_N(n) = (1-P)^{n-1}P$$

con  $N$  = variabile casuale = numero di prove

Si definisce periodo di ritorno e si indica con  $T$ , il valore atteso del numero di anni (ovvero il numero medio di prove necessarie) per avere il primo successo. Come è noto nel caso di distribuzione geometrica la media è pari a  $\frac{1}{P}$ , pertanto a seconda di come si definisce l'evento il valore di  $T$  assume valori diversi.

<b>Evento</b>	<b>Tempo di ritorno</b>
$X < x$	$T = E[N] = \frac{1}{P} = \frac{1}{F_X(x)}$
$X > x$	$T = E[N] = \frac{1}{P} = \frac{1}{1 - F_X(x)}$

### Esempio

Si consideri il valore  $x$  della variabile casuale  $X$  tale che  $F_X(x) = 0.2$ .

Se l'evento che intendiamo considerare è del tipo  $X > x$  (piene fluviali), allora la probabilità che ogni anno si osservi tale occorrenza è

$$P = 1 - F_X(x) = 0.8$$

ed il periodo di ritorno associato ad  $x$  risulta:

$$T = E[N] = \frac{1}{P} = \frac{1}{1 - F_X(x)} = \frac{1}{0.8} = 1.25 \text{ anni}$$

Viceversa, se l'evento è rappresentato da  $X < x$  (risorse idriche), la probabilità  $P$  di suddetto evento risulta:

$$P = F_X(x) = 0.2$$

Il corrispondente tempo di ritorno sarà:

$$T = E[N] = \frac{1}{P} = \frac{1}{F_X(x)} = \frac{1}{0.2} = 5 \text{ anni}$$

□

La scelta del periodo di ritorno da assumere come base di progetto avviene in base a considerazioni di carattere tecnico-economico, tenendo presenti le caratteristiche dell'opera e il confronto, almeno qualitativo, tra costi e benefici connessi. Infatti, fissando il periodo di ritorno si accetta il rischio che mediamente una volta ogni  $T$  anni possano verificarsi eventi più gravosi di quelli presi in considerazione. E' evidente che tenuto conto dei franchi di sicurezza che si assumono nella progettazione delle opere, si possono prendere a riferimento periodi di ritorno relativamente bassi per opere il cui mancato funzionamento comporti danni comunque accettabili e di solo carattere economico e periodi di ritorno molto più alti per opere per le quali il verificarsi di eventi più gravosi di quello previsto potrebbe comportare danni enormi e perdite umane.

L'uso del periodo di ritorno è molto diffuso in idrologia perché apparentemente intuitivo. Si utilizza quindi il frattile  $x_T$  cioè il valore che ha periodo di ritorno  $T$  in luogo del frattile  $x_P$  prima definito. Si è detto che il termine periodo di ritorno è solo apparentemente intuitivo perché molto spesso si crede che il valore  $x_T$  si verifichi ogni  $T$  anni. Nella realtà la probabilità  $P^*$  di avere in  $T$  anni almeno 1 superamento di  $x_T$  è pari a:

$$P^* = 1 - P_T(0)$$

dove  $P_T(0)$  è la probabilità di avere 0 superamento in  $T$  prove.

Utilizzando la distribuzione binomiale

$$P_T(0) = \binom{T}{0} P^0 (1-p)^T = (1-P)^T$$

avendo indicato con  $P$  la probabilità di avere un superamento in una singola prova. Poiché

$$P = \frac{1}{T}$$

allora

$$P_T(0) = \left(1 - \frac{1}{T}\right)^T$$

che per  $T \rightarrow \infty$  tende a  $\frac{1}{e}$  e quindi  $P^*$  tende a  $1 - \frac{1}{e} = 0.63$

Più in generale la probabilità di avere almeno un superamento in  $N$  anni del valore di progetto  $x_T$ , caratterizzato da un periodo di ritorno  $T$ , è pari a:

$$H_N(x_T) = 1 - \left(1 - \frac{1}{T}\right)^N$$

che si ottiene facilmente sostituendo nella (\*)  $N$  al posto di  $T$ .

La tabella riporta alcuni valori di  $H_N(x_T)$  al variare di  $T$  e di  $N$ .

$T \backslash N$	20	50	100	500
20	0.642	0.923	0.994	~1.000
50	0.332	0.636	0.867	~1.000
100	0.182	0.395	0.634	0.993
1000	0.020	0.049	0.095	0.394

## MODELLI PROBABILISTICI

Le distribuzioni di probabilità di variabili continue più comunemente usate nell'analisi delle variabili idrologiche possono essere divise in tre "gruppi", quelle del tipo Normale, le distribuzioni relative ai processi di conteggio di occorrenze casuali e le distribuzioni dei valori estremi.

### LE DISTRIBUZIONI DEL TIPO NORMALE

Le grandezze idrologiche possono essere considerate come il risultato di un processo di tipo additivo, cioè come somma di un gran numero di singoli contributi che si sono verificati in un determinato intervallo di tempo, e quindi distribuite asintoticamente, per il teorema limite centrale, secondo la legge *Normale*. Tali grandezze, possono essere valutate anche come il risultato di un processo di tipo moltiplicativo, cioè come il prodotto degli effetti di un numero  $v$  di fattori, e quindi distribuite secondo la legge *Rad- $v$ -Normale*, cioè normale delle radici  $v$ -esime (Stidd, 1970), o asintoticamente, per  $v$  sufficientemente grande, secondo la legge *Log-Normale*, ancora per il teorema limite centrale.

#### Distribuzione Normale

Una variabile  $X$  si dice distribuita secondo la legge Normale (o distribuita normalmente) se la legge matematica, con cui la funzione distribuzione di probabilità cumulata (Cumulative Distribution Function, CDF) varia in funzione di  $x$ , assume la forma:

$$F_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^x e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

e la funzione distribuzione di probabilità (Probability Density Function, PDF) assume l'espressione:

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

in cui le grandezze  $\mu$  e  $\sigma$  vengono definite parametri della distribuzione e ne rappresentano rispettivamente la media e lo scarto quadratico medio.

Sostituendo alla variabile  $X$  la variabile ausiliaria  $U$  definita come

$$U = \frac{X - \mu}{\sigma}$$

chiamata *variabile ridotta standardizzata*, che ha media zero e scarto quadratico medio uno, la CDF diventa:



$$F_U(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^u e^{-\frac{u^2}{2}} du.$$

L'andamento di tale funzione è nota ed esistono tabelle che associano ad ogni valore di  $U$  il corrispondente valore di  $F_U(u)$  (Su EXCEL basta richiamare la funzione DISTRIB.NORM.ST. a cui bisogna fornire in ingresso il valore della variabile standardizzata  $u$ ).

### Distribuzione Log-Normale

Si dice che la variabile casuale  $X$  è distribuita secondo la legge log-normale se la variabile trasformata

$$Y = \log X$$

è distribuita secondo la legge normale.

La CDF e la PDF diventano in questo caso:

$$F_Y(y) = \frac{1}{\sqrt{2\pi}\sigma_y} \int_{-\infty}^y e^{-\frac{(y-\mu_y)^2}{2\sigma_y^2}} dy$$

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma_y} e^{-\frac{(y-\mu_y)^2}{2\sigma_y^2}}$$

in cui  $\mu_Y = \mu (\log X)$  e  $\sigma_Y = \sigma (\log X)$  sono rispettivamente la media e lo scarto della variabile trasformata  $Y$ .

### Distribuzione Box e Cox

La distribuzione proposta da Box e Cox per una generica variabile casuale  $X$  considera che una trasformata  $Y$  di  $X$  tra quelle definite secondo le seguenti espressioni:

$$Y = X^v \quad 0 < v \leq 1$$

$$Y = \log X \quad v = 0$$

sia distribuita secondo la legge normale.

Il valore di  $v$  da considerare è quello per cui la corrispondente serie trasformata risulta meglio interpretata dalla distribuzione Normale.

A tale scopo, per diversi valori di  $v$  si calcola il coefficiente di asimmetria  $G$  delle serie trasformate,

$$G = N^{\frac{1}{2}} \frac{\sum_{i=1}^N (y_i - \bar{y})^3}{\left[ \sum_{i=1}^N (y_i - \bar{y})^2 \right]^{\frac{3}{2}}}$$

adottando il valore di  $\nu$  per il quale lo stimatore  $G$  più si avvicina al valore zero.

Si ricorda, infatti, che la distribuzione Normale è caratterizzata da un coefficiente di asimmetria  $\gamma_1=0$ .

Per  $\nu = 0$ ,  $X$  risulta distribuita secondo la legge lognormale. Per  $\nu = 1$ ,  $X$  risulta distribuita secondo la legge Normale.

Altre distribuzioni comunemente usate sono la radnormale ( $\nu = 1/2$ ) e la radcubicnormale ( $\nu = 1/3$ ).

E' da considerare tuttavia che  $G$  è uno stimatore distorto di  $\gamma_1$

$$E[G] \neq \gamma_1$$

ed è affetto da elevata variabilità.

Pertanto è preferibile valutare  $\nu$  in base ad un numero più elevato di serie storiche con analisi di tipo regionale (si veda esercitazione **Pioggie annue** ).

Inoltre, per l'incertezza nella stima si assumono come valori di  $\nu$  plausibile solo i seguenti: 0, 1/3, 1/2, 1.

## **LE DISTRIBUZIONI RELATIVE AI PROCESSI DI CONTEGGIO DI OCCORRENZE CASUALI**

### **Distribuzione di Poisson**

La v.c. di Poisson può essere considerata come un caso particolare della v.c. binomiale ed è utilizzata per rappresentare una vasta gamma di fenomeni casuali, in particolare quelli connessi ad un qualche "conteggio" di eventi in un prefissato periodo di tempo (ad esempio il numero di incidenti stradali per giorno ad un dato incrocio, il numero di chiamate telefoniche per ora in un dato centralino, etc.).

Ipotizziamo di essere interessati a conoscere il numero totale di veicoli che arrivano in una certa località in un fissato periodo di tempo di  $t$  secondi. Assumiamo di suddividere il periodo temporale  $t$  in  $N$  intervalli (ognuno dei quali rappresenta una prova) e di conoscere, inoltre,  $P$  la probabilità che arrivi un veicolo in ciascuno di questi intervalli. Allora la funzione di probabilità della variabile casuale  $X$  numero totale di veicoli (=numero di successi) è data dalla distribuzione binomiale:

$$P_N(X = x) = \binom{N}{x} P^x (1-P)^{N-x}$$

Facciamo ora tendere  $N$  ad infinito e contemporaneamente  $P$  a zero, in modo tale che il prodotto  $NP$  si mantenga sempre pari a  $v$ , che rappresenta il numero medio di successi nel periodo  $t$ . La funzione di probabilità della variabile discreta  $X$  tende allora ad assumere la forma asintotica:

$$P(X = x) = \frac{v^x e^{-v}}{x!}$$

che prende il nome di distribuzione di Poisson.

Abbiamo visto che la distribuzione di Poisson può essere ricavata dalla binomiale quando  $N$  cresce indefinitamente e  $P$  decresce verso lo zero di modo che risulti sempre  $NP=v$ . Poiché  $P$  rappresenta la probabilità di avere un successo, la variabile casuale di Poisson può essere assimilata ad una binomiale nella quale la probabilità che si verifichi un successo è prossima allo zero: per tale motivo la variabile casuale di Poisson si presta bene ad interpretare i cosiddetti *eventi rari*.

La distribuzione di Poisson può essere ricondotta anche ai processi stocastici.

Spesso nelle applicazioni idrologiche si incontrano delle variabili casuali la cui distribuzione di probabilità presenta la caratteristica di essere dipendente dal tempo. Un processo stocastico è in generale una funzione casuale del tempo.

Nell'esempio appena descritto ci siamo interessati ad un processo stocastico  $X(t)$ , il cui valore (casuale) ad ogni istante  $t$  è il numero di arrivi occorsi da  $t=0$ .

In tal caso possiamo riscrivere la distribuzione di Poisson considerando  $v=\lambda t$ , in cui  $\lambda$  è la probabilità di avere un successo nell'intervallo unitario di tempo.

$$P(X = x) = \frac{(\lambda t)^x e^{-\lambda t}}{x!}$$

La relazione precedente vale però solo nel caso di processo stocastico Poissoniano e cioè di un processo che ha le seguenti caratteristiche:

**Stazionarietà** - la probabilità di avere un successo nel singolo intervallo  $(t, t+h)$  non dipende dal tempo  $t$ ;

**Non molteplicità**- la probabilità di avere più di un evento nel singolo intervallo è nulla;

**Indipendenza**- il numero di successi in ogni intervallo di tempo è indipendente dal numero di successi degli intervalli precedenti (gli intervalli sono disgiunti ossia non si sovrappongono).

### Distribuzione esponenziale

Se consideriamo un processo Poissoniano, di parametro  $\lambda$ , il periodo di tempo  $T$  intercorrente tra due successi consecutivi (ovvero il tempo a cui si osserva il primo successo) è una variabile casuale che si distribuisce con legge esponenziale:

$$\begin{aligned} f_T(t) &= \lambda e^{-\lambda t} \\ F_T(t) &= 1 - e^{-\lambda t} \end{aligned} \quad (**)$$

Caratterizzata da  $E[T] = 1/\lambda$  e  $Var[T] = 1/\lambda^2$

L'espressione (\*\*) può essere facilmente ricavata a partire dall'espressione della distribuzione di Poisson. Infatti, basta considerare che la probabilità che la v.c.  $T$  superi un certo valore  $t$  è pari alla probabilità (calcolata con Poisson) che nessun evento si verifichi nell'intervallo di tempo di lunghezza  $t$ , ossia:

$$\begin{aligned} P(T > t) &= 1 - F_T(t) \\ P(T > t) &= P(X = 0) = \frac{(\lambda t)^0 e^{-\lambda t}}{0!} = e^{-\lambda t} \\ 1 - F_T(t) &= e^{-\lambda t} \end{aligned}$$

da cui si ottiene  $F_T(t) = 1 - e^{-\lambda t}$

### Distribuzione Gamma

Se si considera un processo Poissoniano un'altra variabile casuale di interesse è l'intervallo di tempo  $X_k$ , necessario per avere il  $k$ -esimo successo.

Questa variabile casuale è distribuita secondo la distribuzione Gamma, la cui funzione densità di probabilità (PDF) è:

$$f_{X_k}(x) = \frac{\lambda (\lambda x)^{k-1} e^{-\lambda x}}{\Gamma(k)}$$

dove  $\Gamma(k)$  indica la funzione matematica Gamma, definita in generale come:

$$\Gamma(k) = \int_0^{\infty} t^{k-1} e^{-t} dt$$

e che risulta pari a  $(k-1)!$  se  $k$  è un numero intero.

Ponendo  $\alpha = k$  e  $\beta = 1/\lambda$  si ottiene l'espressione della distribuzione Gamma così come è riportata nelle successive dispense.

$$f_{X_k}(x) = \frac{(x/\beta)^{\alpha-1} e^{-x/\beta}}{\beta \Gamma(\alpha)}$$

I momenti di tale distribuzione espressi in funzione dei parametri  $\alpha$  e  $\beta$  valgono:

<b>E[X]</b>	$\alpha\beta$
<b>Var[X]</b>	$\alpha\beta^2$

Pertanto utilizzando il metodo dei momenti i parametri della distribuzione  $\alpha$  e  $\beta$  si possono esprimere in funzione della media  $\mu$  e dello scarto quadratico medio  $\sigma$ , attraverso le relazioni:

$$\alpha = \frac{\mu^2}{\sigma^2},$$

$$\beta = \frac{\sigma^2}{\mu}.$$

## **LE DISTRIBUZIONI DEI VALORI ESTREMI**

L'analisi statistica degli estremi idrologici si può condurre secondo due diversi approcci non alternativi. Il primo consiste nel considerare solo i massimi (o minimi) valori in un assegnato intervallo di tempo, generalmente un anno, utilizzando cioè la serie dei massimi (minimi) annuali e quindi un valore per ogni anno. Il secondo approccio, invece, tiene conto di tutti i valori che eccedono (o sono inferiori) una soglia arbitrariamente prefissata, in genere abbastanza elevata (bassa), e si basa quindi su un numero annuo di eventi che può cambiare da anno in anno. In un caso si estrae dal campione di dati idrologici solo la serie dei massimi (minimi) annuali, nell'altro quella dei valori superiori (inferiori) ad una soglia.

### **Il processo dei massimi annuali**

All'interno dei modelli probabilistici, che schematizzano il processo dei massimi annuali, particolare sviluppo hanno quelli derivanti da alcuni classici risultati asintotici del calcolo delle probabilità.

Se si considerano i massimi annuali  $X$  di una grandezza idrologica (piogge giornaliere, portate al colmo, etc.) come i massimi di una serie di  $N$  variabili casuali  $Y_i$  ( $i=1, \dots, N$ ), indipendenti e identicamente distribuite, è possibile, per  $N$  tendente ad infinito, individuare la distribuzione asintotica di  $X$  prescindere dalla distribuzione delle  $Y$ , ma soltanto in base all'andamento della sua coda superiore (ovvero di come la funzione di probabilità cumulata delle  $Y$  tende ad uno). In particolare le distribuzioni asintotiche possono essere di tre tipi:

- distribuzione EV1 o di Gumbel;
- distribuzione EV2 o di Frechet;
- distribuzione EV3 o di Weibull.

### **Distribuzione EV1 o di Gumbel**

La funzione di ripartizione (CDF) della distribuzione di Gumbel, o distribuzione del massimo valore del primo tipo, EV1 (Extreme Value Type-1), ha la seguente espressione:

$$F_X(x) = e^{-e^{-\alpha(x-\varepsilon)}} \quad \text{con } \alpha > 0;$$

mentre la funzione di densità di probabilità (PDF) assume la forma:

$$f_x(x) = \alpha e^{-\alpha(x-\varepsilon)} e^{-e^{-\alpha(x-\varepsilon)}}.$$

I parametri della funzione si possono stimare utilizzando sia il metodo dei momenti, con il quale le stime di  $\varepsilon$  ed  $\alpha$  si ottengono in funzione della media  $\bar{x}$  e dello scarto quadratico medio  $s_x$  degli  $n$  dati del campione a disposizione attraverso le relazioni:

$$\varepsilon = \bar{x} - \frac{0,577}{\alpha}$$

$$\alpha = \sqrt{\frac{1,645}{s_x^2}},$$

sia quello della massima verosimiglianza che permette di stimare i parametri  $\alpha$  ed  $\varepsilon$  risolvendo in maniera iterativa le espressioni:

$$\frac{1}{\hat{\alpha}} = \bar{x} - \frac{\sum_{i=1}^n x_i e^{-\hat{\alpha}x_i}}{\sum_{i=1}^n e^{-\hat{\alpha}x_i}}$$

$$e^{-\hat{\alpha}\hat{\varepsilon}} = \frac{1}{n} \sum_{i=1}^n e^{-\hat{\alpha}x_i}.$$

Operativamente si procede considerando un valore di primo tentativo  $\hat{\alpha} = \alpha_1$  e lo si sostituisce nel termine a destra della prima espressione, ottenendo, quindi, come soluzione un nuovo valore  $\hat{\alpha} = \alpha_2$ . Se  $|\alpha_1 - \alpha_2|$  non risulta minore della tolleranza fissata (es.  $\Delta=0.001$ ), allora si procede al calcolo di un nuovo valore  $\alpha_3$  che sia compreso tra  $\alpha_1$  e  $\alpha_2$ .

$$\text{se } (\alpha_1 < \alpha_2) \quad \alpha_3 = \alpha_1 + \frac{\alpha_2 - \alpha_1}{3}$$

$$\text{se } (\alpha_1 > \alpha_2) \quad \alpha_3 = \alpha_2 + \frac{\alpha_1 - \alpha_2}{3}$$

Una volta stimato  $\alpha$  è possibile ricavare la corrispondente stima di  $\varepsilon$  dalla seconda espressione.

Tra le caratteristiche di questa distribuzione si segnalano il valore costante del coefficiente di asimmetria teorico  $\gamma_1$  pari a 1.1396 e la condizione di stabilità rispetto al massimo, per cui se il massimo annuale è distribuito secondo la EV1 con parametri  $\alpha$  e  $\varepsilon$ , il massimo in  $N$  anni sarà sempre distribuito secondo la EV1, con parametri:

$$\alpha_N = \alpha$$

$$\varepsilon_N = \varepsilon + \frac{\ln N}{\alpha}.$$

Il modello probabilistico EV1 è tra i più usati nello studio dei massimi idrologici, ma l'esperienza fatta in diverse analisi idrologiche, ha evidenziato come tale modello non descriva sempre in modo soddisfacente la distribuzione di probabilità di tali variabili. Infatti, spesso la il tipico andamento rettilineo della funzione di ripartizione EV1, non riesce ad interpretare la presenza di uno o più valori che si discostano molto dagli altri (outliers).

### **Distribuzione EV2 o di Frechet**

In questo caso la CDF è limitata inferiormente ed ha la seguente espressione:

$$F_x(x) = \exp \left[ - \left( \frac{v - \beta}{x - \beta} \right)^{\frac{1}{K}} \right] \quad \text{con } K > 0, x \geq \beta, v > \beta \geq 0$$

Si noti che se la v.c.  $X$  segue la distribuzione EV2 con parametri  $K, \beta, v$ , allora la variabile

$$Y = \ln(X - \beta)$$

segue la distribuzione EV1 con parametri

$$\alpha = 1/K$$

$$\varepsilon = \ln(v - \beta)$$

### **Distribuzione EV3 o di Weibull**

La CDF della distribuzione EV3 ha la seguente espressione:

$$F_x(x) = \exp \left[ - \left( \frac{\omega - x}{\omega - u} \right)^{1/K} \right] \quad \text{con } K > 0, x \leq \omega, u < \omega$$

Questa distribuzione presenta un limite superiore ( $\omega$ ) per la variabile  $X$  e pertanto non è di rilevante interesse pratico nel settore dell'analisi di frequenza delle piene.

### **Distribuzione GEV**

Le leggi di distribuzione asintotiche viste in precedenza possono essere considerate come casi particolari della legge generale dei valori estremi GEV (Generalized Extreme Value), proposta da Jenkinson nel 1955.

La relativa CDF assume l'espressione:



$$F_x(x) = \exp \left[ - \left( 1 + K \frac{x - \varepsilon}{a} \right)^{\frac{1}{K}} \right]; \text{ per } K \neq 0, a > 0$$

$$F_x(x) = \exp[-\exp[-\alpha(x - \varepsilon)]]; \text{ per } K=0, \text{ con } \alpha = \frac{1}{a} > 0.$$

La GEV, quindi, ammette la EV1 come funzione limite per  $K$  uguale a zero, coincide con la EV2 per  $K < 0$  (basta porre  $a = \frac{\nu - \beta}{K}$  e  $\varepsilon = \nu$ ) e per  $K > 0$  con la EV3 (basta porre  $a = \frac{u - \omega}{K}$  e  $\varepsilon = u$ ).

### Il processo dei minimi annuali

Analogamente a quanto visto per i massimi sono state ricavate le distribuzioni asintotiche per i valori minimi  $Z$  di una serie di  $N$  variabili casuali  $Y_i$  ( $i=1, \dots, N$ ), indipendenti e identicamente distribuite, per  $N$  tendente ad infinito.

#### EV1 minimi

La distribuzione asintotica di tipo 1 per  $Z$  è caratterizzata da:

$$F_Z(z) = 1 - e^{-e^{\alpha(z - \varepsilon)}}$$

$$f_Z(z) = \alpha \exp[\alpha(z - \varepsilon) - e^{\alpha(z - \varepsilon)}]$$

#### EV2 minimi

La distribuzione asintotica **EV2** per i minimi può essere ricavata, ma è priva di interesse pratico.

#### EV3 minimi (o di Weibull)

La CDF della distribuzione **EV3** minimi ha la seguente espressione:

$$F_Z(z) = 1 - \exp \left[ - \left( \frac{z - \varepsilon}{u - \varepsilon} \right)^k \right] \quad \text{con } z \geq \varepsilon$$

Se una variabile casuale  $Z$  è distribuita secondo la EV3 minimi con parametri  $\varepsilon$ ,  $u$  e  $k$  allora la variabile casuale  $X = \ln(Z - \varepsilon)$  segue la EV1 minimi, con parametri  $\varepsilon_l = \ln(u - \varepsilon)$  ed  $\alpha = k$ .

### Distribuzione TCEV

Una migliore interpretazione probabilistica di serie caratterizzate dalla presenza di outliers, si ha con il modello a doppia componente denominato TCEV (acronimo di Two Component Extreme Value), che si rappresenta con una funzione di probabilità cumulata del tipo:

$$F_x(x) = \exp\{-\Lambda_1 \exp(-x/\theta_1) - \Lambda_2 \exp(-x/\theta_2)\} \quad x \geq 0 \quad (1)$$

Questa distribuzione, in cui si possono distinguere formalmente una componente *base* (pedice 1), relativa agli eventi normali e più frequenti, ed una componente *straordinaria* (pedice 2), relativa ad eventi più gravosi e rari, permette di interpretare fisicamente il processo dei massimi annuali tramite due popolazioni distinte.

I quattro parametri del modello TCEV espresso nella forma (1) hanno un chiaro significato fisico dal momento che  $\Lambda_1$  e  $\Lambda_2$  esprimono il numero medio annuo di eventi superiori ad una soglia delle due componenti, e  $\theta_1$  e  $\theta_2$  esprimono il valore medio di tali eventi.

L'espressione (1) si può mettere nella forma:

$$F_x(x) = \exp\{-\exp[-\alpha_1(x - \varepsilon_1)] - \exp[-\alpha_2(x - \varepsilon_2)]\} \quad x \geq 0 \quad (2)$$

equivalendo formalmente al prodotto di due funzioni di distribuzione cumulata di Gumbel, avendo posto  $\varepsilon_i = \theta_i \ln \Lambda_i$  e  $\alpha_i = 1/\theta_i$  con  $i=1,2$ .

La funzione di probabilità cumulata (1) è esprimibile ancora in altra forma effettuando la trasformazione di variabili  $\theta_* = \theta_2/\theta_1$  e  $\Lambda_* = \Lambda_2/\Lambda_1^{1/\theta_*}$ . In questo caso, in modo del tutto equivalente, la (1) si può scrivere:

$$F_x(x) = \exp\{-\Lambda_1 \exp(-x/\theta_1) - \Lambda_* \Lambda_1^{1/\theta_*} \exp[-x/(\theta_* \theta_1)]\} \quad x \geq 0 \quad (3)$$

e i quattro parametri che caratterizzano il modello diventano  $\Lambda_*$ ,  $\theta_*$ ,  $\Lambda_1$  e  $\theta_1$ .

Per la determinazione di  $x_T$  occorre avere in definitiva una stima dei quattro parametri  $\Lambda_1$ ,  $\Lambda_2$ ,  $\theta_1$  e  $\theta_2$  o equivalentemente dei quattro parametri  $\Lambda_*$ ,  $\theta_*$ ,  $\Lambda_1$  e  $\theta_1$ , con i quali si può ricostruire integralmente la funzione di probabilità cumulata (1) o (3).

La stima dei quattro parametri si può ottenere ricorrendo al metodo dei momenti o al metodo della massima verosimiglianza, vincolando con quest'ultimo metodo i parametri da stimare alla conoscenza di quelli già noti da indagini a livello regionale.

### Stima regionale dei parametri del modello TCEV

I parametri della TCEV sono 4 ed è quindi elevata l'incertezza della stima ottenuta con le serie storiche disponibili la cui dimensione campionaria è in genere inferiore a 80.

Per ridurre l'incertezza si utilizzano tecniche di analisi regionale che consentono di stimare almeno alcuni dei parametri sulla base di tutte le serie storiche ricadenti all'interno di vaste aree indicate come zone e sottozone omogenee.

Al 1° livello di regionalizzazione per i due parametri di forma del modello,  $\theta_*$  e  $\Lambda_*$ , si può assumere un valore costante all'interno di ampie zone omogenee. La stima dei valori che tali parametri assumono nella singola zona omogenea risulta pertanto molto affidabile, perché si può ottenere utilizzando tutti i dati delle serie ricadenti all'interno di essa.

Al 2° livello di regionalizzazione, oltre ai valori costanti dei parametri  $\theta_*$  e  $\Lambda_*$  nelle zone omogenee, all'interno di queste è possibile identificare sottozone omogenee, entro cui si può ritenere costante anche il parametro di scala  $\Lambda_1$ . Anche in questo caso, utilizzando per la stima di  $\Lambda_1$  tutti i dati delle serie ricadenti all'interno della singola sottozona, risulta essere accresciuta l'affidabilità della stima di questo parametro. In totale quindi per questo livello di analisi sono tre i parametri di cui si può assumere a priori un valore regionale.

Al 3° livello di regionalizzazione, oltre ai tre parametri  $\theta_*$ ,  $\Lambda_*$  e  $\Lambda_1$  di cui si può assumere un valore regionale, identificato al livello precedente, si persegue in modo regionale anche la stima del quarto parametro che sia  $\theta_1$  o  $\mu$  in relazione all'approccio che si intende adottare (vedi paragrafo successivo).

### Il metodo del valore indice

Al secondo ed al terzo livello di regionalizzazione la determinazione di  $x_T$  può essere effettuata attraverso due metodologie alternative.

La prima consiste, come abbiamo già applicato per i modelli probabilistici presentati in precedenza, nella stima dei quattro parametri  $\Lambda_1$ ,  $\Lambda_2$ ,  $\theta_1$  e  $\theta_2$ , o equivalentemente dei quattro parametri  $\Lambda_*$ ,  $\theta_*$ ,  $\Lambda_1$  e  $\theta_1$ , con i quali si può ricostruire integralmente la funzione di probabilità cumulata (1) o (3). Tale procedura nel prosieguo della scheda verrà indicata come approccio  $F_X(x)$ .

La stima di  $x_T$ , nel modello TCEV, può essere alternativamente effettuata con il **metodo del valore indice**. Con tale metodo si analizza in luogo di  $X$  una variabile adimensionale  $X/X_I$  dove  $X_I$  è un valore caratteristico della distribuzione di  $X$  ed assume il nome di valore indice.

Nelle applicazioni quasi sempre si utilizza come valore indice la media  $\mu$  e si analizza la variabile  $X'=X/\mu$  che viene indicata come fattore di crescita.

In generale seguendo tale approccio, la stima di  $x_T$  si ottiene con due passi distinti:

- stima del fattore di crescita  $x'_T$ , relativo al periodo di ritorno  $T$ ;
- stima del valore indice,  $\mu$ .

In definitiva la stima di  $x_T$  si ottiene con il prodotto  $x_T = x'_T \cdot \mu$ .

La stima del fattore di crescita, riferita al periodo di ritorno imposto dal problema in esame, è ovviamente una stima probabilistica. La distribuzione di probabilità (curva di crescita) di tale variabile interpretata con la legge probabilistica TCEV assume espressione:

$$F_{x'}(x') = \exp\left[-\Lambda_1 \exp(-\eta x') - \Lambda_* \Lambda_1^{1/\theta_*} \exp(-\eta x'/\theta_*)\right] \quad (4)$$

dove:

$$\eta = \frac{\mu}{\theta_1} = \ln \Lambda_1 + \gamma_\varepsilon - \sum_{j=1}^{\infty} \frac{(-1)^{j-1} \Lambda_*^j}{j!} \Gamma(j/\theta_*) \quad (5)$$

con  $\gamma_\varepsilon = 0.57722$  (costante di Eulero) e sommatoria che può essere limitata da 1 a 20 con buona approssimazione per scopi pratici.

E' importante sottolineare che la curva di crescita (4), come si può notare facilmente dalle espressioni riportate, dipende dai soli parametri  $\Lambda_*$ ,  $\theta_*$  e  $\Lambda_1$ . In definitiva, utilizzando l'approccio del valore indice, è possibile ottenere una stima di  $x_T$  dalla conoscenza dei parametri  $\theta_*$ ,  $\Lambda_*$ ,  $\Lambda_1$  (mediante i quali si stima la curva di crescita) e di  $\mu$  (portata indice). L'equivalenza tra l'approccio del valore indice, e quello indicato come approccio  $F_X(x)$  è evidente dalla relazione (5) che lega tra loro i parametri di posizione  $\theta_1$  e  $\mu$ . Per confronto con la procedura  $F_X(x)$  descritta in precedenza, chiameremo questa procedura approccio  $F_{x'}(x')$ .

E' opportuno sottolineare che mentre per il secondo livello di regionalizzazione il valore indice può essere considerato pari alla media aritmetica della serie campionaria della variabile idrologica considerata, al terzo livello anche questo parametro sarà stimato considerando relazioni empiriche derivate su base regionale, e di cui si riportano nell'ultimo paragrafo della scheda, le espressioni ottenute per la Calabria.

### Scelta della procedura da adottare

La scelta della procedura da utilizzare, e quindi del livello di regionalizzazione al quale fare riferimento, dipende sostanzialmente dalla dimensione campionaria.

- La stima puntuale di tutti e 4 i parametri (*livello 0 di regionalizzazione*) presenta una elevata incertezza per le dimensioni usuali delle serie campionarie ed è quindi poco utilizzata.
- Quando si dispone di almeno 40÷50 anni di osservazione, si può adottare il *1° livello di regionalizzazione*, utilizzando le stime regionali dei parametri di forma  $\theta_*$  e  $\Lambda_*$ . Per i rimanenti due parametri del modello si utilizzano i dati della serie campionaria, effettuando una stima puntuale di  $\Lambda_1$  e  $\theta_1$  vincolata ai parametri  $\theta_*$  e  $\Lambda_*$  con il metodo della massima verosimiglianza (approccio  $F_X(x)$ ).

1° LIVELLO ) 2 parametri → stima regionale ( $\theta_*$  e  $\Lambda_*$ )    2 parametri → stima puntuale ( $\Lambda_1$  e  $\theta_1$ )

- Se si dispone di almeno 20÷30 anni di osservazione, adottando al *2° livello di regionalizzazione* i valori regionali stimati per i parametri  $\theta_*$ ,  $\Lambda_*$  e  $\Lambda_1$ , si può effettuare sulla base dei dati campionari una stima vincolata di  $\theta_1$  ai parametri  $\theta_*$ ,  $\Lambda_*$ ,  $\Lambda_1$  con il metodo della massima verosimiglianza (approccio  $F_X(x)$ ). Se invece si utilizza la curva di crescita, al valore regionale dei parametri  $\theta_*$ ,  $\Lambda_*$  e  $\Lambda_1$  si può affiancare la media aritmetica della serie campionaria,  $\bar{X}$ , come stima del valore indice  $\mu$ . Quest'ultimo metodo è più semplice e di immediata applicabilità.

2° LIVELLO ) 3 parametri → stima regionale ( $\theta_*$ ,  $\Lambda_*$  e  $\Lambda_1$ )    1 parametro → stima puntuale ( $\theta_1$ )

OPPURE

2° LIVELLO ) 3 parametri → stima regionale ( $\theta_*$ ,  $\Lambda_*$  e  $\Lambda_1$ )    1 parametro → stima puntuale ( $\bar{X}$ )

- Nel caso in cui le osservazioni campionarie manchino completamente o siano scarse per qualità e dimensione, al *3° livello di analisi regionale* si preferisce adottare la curva di crescita e affiancare ai valori regionali di  $\theta_*$ ,  $\Lambda_*$  e  $\Lambda_1$  la stima di  $\bar{X}$  ottenuta dalle relazioni empiriche identificate per la singola area omogenea (tabelle 3-4).

3° LIVELLO ) 4 parametri → stima regionale ( $\theta_*$ ,  $\Lambda_*$ ,  $\Lambda_1$  e  $\bar{X}$ )

Equivalentemente è possibile, ottenuta la stima di  $\bar{X}$  dalle relazioni empiriche, risalire al parametro  $\theta_1$  considerando la relazione (5) in cui  $\eta$  dipende dai soli parametri regionali  $\Lambda_*$ ,  $\theta_*$  e  $\Lambda_1$ .

L'analisi statistica di una singola serie campionaria prevede, in definitiva, la scelta di uno solo dei livelli di regionalizzazione proposti, in funzione della dimensione della serie. E' comunque consigliabile confrontare i risultati ottenibili dai diversi livelli di regionalizzazione. Nell'utilizzo del 3° livello di analisi regionale, infine, è buona norma confrontare tra loro i risultati ottenuti per la stima di  $\bar{X}$  da più formulazioni empiriche tra quelle proposte.

### I risultati ottenuti per la regione Calabria

Il modello TCEV è stato impiegato nel progetto VAPI (VALutazione delle Piene in Italia) sviluppato dal Consiglio Nazionale delle Ricerche. In generale poiché le stazioni idrometriche sono molto meno numerose di quelle pluviometriche, l'analisi delle piene in Calabria, come d'altronde nelle altre regioni italiane, è stata preceduta dall'analisi delle piogge estreme. In particolare l'analisi dei massimi annuali di pioggia giornaliera fornisce utili indicazioni circa l'identificazione delle sottozone idrometriche che si ammette coincidano in prima approssimazione con le sottozone pluviometriche. L'interesse per le piogge orarie deriva invece dal fatto che la piena di progetto è legata al valore medio dell'intensità di pioggia  $I_{tc}$  relativa ad una durata critica  $t_c$  scelta a caratterizzare la risposta del bacino idrografico, che per i bacini della Calabria è pari a qualche ora. La classica espressione che lega le medie dei massimi annuali delle altezze di pioggia  $\bar{X}_t$  alla durata  $t$  è  $\bar{X}_t = a t^n$  (curva di probabilità pluviometrica) con parametri  $a$  ed  $n$  che possono variare da sito a sito in dipendenza della quota sul mare o di altri parametri geografici.

Al 1° e al 2° livello di regionalizzazione i valori dei parametri di forma  $\theta_*$ ,  $\Lambda_*$ , del parametro di scala  $\Lambda_1$  e di  $\eta$  del modello TCEV per i massimi annuali di pioggia giornaliera e di portata al colmo assumono i valori riportati nella tabella 1.

	<b>Piogge giornaliere</b>	<b>Portate</b>
	<b>1° LIVELLO</b>	
$\theta_*$	2.154	2.654
$\Lambda_*$	0.418	0.350
	<b>2° LIVELLO</b>	
$\Lambda_1$ <i>TIRRENO</i>	48.914	10.147
$\Lambda_1$ <i>CENTRO</i>	22.878	5.519
$\Lambda_1$ <i>IONIO</i>	10.987	3.047

$\eta_{TIRRENO}$	5.183	3.651
$\eta_{CENTRO}$	4.423	3.042
$\eta_{IONIO}$	3.690	2.448

Tab. 1- Valori regionali dei parametri TCEV relativi al 1° e al 2° livello di analisi regionale.

In base ai valori riportati nella tabella 1, le espressioni ottenute per le leggi di crescita nelle diverse sottozone omogenee (2° livello di regionalizzazione,  $\theta_*$ ,  $\Lambda_*$  e  $\Lambda_1$  costanti) sono riportate nella tabella 2. In Tabella 3 sono riportati, per ogni sottozona, i valori dei fattori di crescita per alcuni periodi di ritorno, per quanto riguarda gli estremi pluviometrici ed idrometrici.

<b>Piogge giornaliere</b>		
$F_{X'}(x') = \exp[-48.914(177.96)^{-x'} - 2.542(11.068)^{-x'}]$	TIRRENO	
$F_{X'}(x') = \exp[-22.878(83.341)^{-x'} - 1.786(7.794)^{-x'}]$	CENTRO	
$F_{X'}(x') = \exp[-10.987(39.986)^{-x'} - 1.271(5.549)^{-x'}]$	IONIO	
<b>Portate al colmo</b>		
$F_{X'}(x') = \exp[-10.147(38.458)^{-x'} - 0.837(3.956)^{-x'}]$	TIRRENO	
$F_{X'}(x') = \exp[-5.519(20.918)^{-x'} - 0.665(3.145)^{-x'}]$	CENTRO	
$F_{X'}(x') = \exp[-3.047(11.550)^{-x'} - 0.532(2.514)^{-x'}]$	IONIO	

Tab. 2- Espressioni delle leggi di crescita di portate al colmo e piogge giornaliere della regione Calabria.

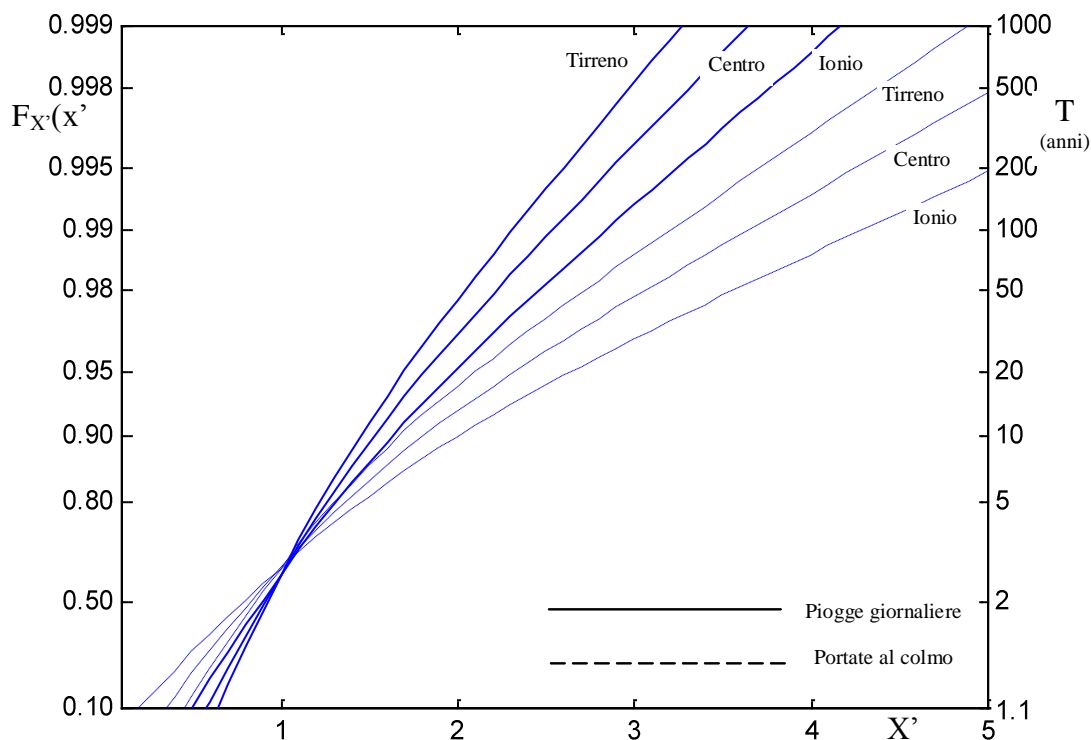


Fig.1- Leggi di crescita di portate al colmo e piogge giornaliere relative al 2° livello di analisi regionale

<b>T (anni)</b>	<b>Piogge</b>			<b>Portate</b>		
	<b>Tirreno</b>	<b>Centro</b>	<b>Ionio</b>	<b>Tirreno</b>	<b>Centro</b>	<b>Ionio</b>
5	1.22	1.26	1.31	1.31	1.37	1.46
10	1.45	1.52	1.63	1.67	1.81	2.01
20	1.69	1.81	1.97	2.09	2.32	2.64
50	2.04	2.21	2.46	2.71	3.07	3.57
100	2.31	2.54	2.84	3.21	3.67	4.31
200	2.59	2.87	3.24	3.7	4.27	5.06
500	2.97	3.31	3.77	4.36	5.07	6.05
1000	3.26	3.65	4.17	4.86	5.67	6.8

Tab. 3- Valori dei fattori di crescita per diversi periodi di ritorno

Le espressioni empiriche identificate al 3° livello per la stima dei valori medi delle piogge giornaliere ed orarie sono le seguenti:

$$\text{piogge giornaliere} \quad \log \bar{X} = a y + b$$

$$\text{piogge orarie} \quad \bar{X}_t = c t^{(d + a y)/\log 24}$$

dove  $y$  è la quota della stazione rispetto al livello del mare.

I parametri da impiegare per le diverse aree omogenee sono riportate nella tabella seguente:

<b>3° LIVELLO - Piogge giornaliere ed orarie</b>				
Area omogenea	$a$	$b$	$c$	$d$
T1	0.00014	1.907	27.79	0.521
T2	0.00021	1.683	23.75	0.365
T3	0.00022	1.769	26.61	0.402
T4	0.00028	1.736	16.73	0.367
C1	0.00049	1.690	21.73	0.411
C2	0.00021	1.683	23.75	0.365
C3	0.00016	1.951	31.02	0.517
C4	0.00032	1.840	33.22	0.377
C5	0.00036	1.815	34.99	0.329
I1	0.00026	1.778	24.37	0.449
I2	0.00025	1.922	30.97	0.489
I3	0.00043	1.953	39.58	0.414
I4	0.00027	1.817	34.13	0.342

Tab. 4- Valori dei parametri al 3° livello delle piogge estreme per singola area omogenea.

Per le portate al colmo nel 3° livello di analisi regionale si sono invece ottenute le formulazioni empiriche riportate nella tabella 5.



<b>3° LIVELLO – Portate</b>	
$\bar{X} = 1.578A^{0.839}$	$A = [Km^2]$
$\bar{X} = C^* A \bar{I}_{tr} / 3.6$	$C^* = 0.158; A = [Km^2]; \bar{I}_{tr} = [mm/h]$

Tab. 5- Relazioni empiriche per la stima della portata indice al 3° livello delle portate al colmo.

Nella tabella 5  $\bar{I}_{tr}$  è la media dei massimi annuali dell'intensità di pioggia di durata pari al tempo di ritardo del bacino, dato dalla distanza temporale tra il baricentro di un pluviogramma efficace e quello dell'idrogramma superficiale ad esso corrispondente o, in altri termini, dal primo momento dell'idrogramma unitario istantaneo rispetto all'origine nell'ipotesi che il sistema bacino sia lineare e stazionario. La stima dell'intensità  $\bar{I}_{tr}$  si ottiene dalla curva di probabilità pluviometrica ricavata a partire dai valori medi registrati nei singoli pluviografi.

Nella seconda relazione di tabella 4,  $C^*$  è il coefficiente di deflusso di piena. Il valore  $C^* = 0.158$  può essere considerato come un valore medio per l'intero territorio regionale.  $C^*$  può oscillare tra valori minimi pari a  $0.07 \div 0.08$  nei bacini carbonatici del Pollino e valori massimi pari a 0.4 nei bacini argillosi come l'Esaro di Crotona.

### STIMA DEI PARAMETRI

La funzione di probabilità cumulata e la funzione densità di probabilità della distribuzione TCEV sono:

$$F_X(x) = \exp\left\{-\Lambda_1 \exp(-x/\theta_1) - \Lambda_* \Lambda_1^{1/\theta_*} \exp[-x/(\theta_* \theta_1)]\right\} \quad x \geq 0$$

$$f_X(x) = \frac{dF_X(x)}{dx} = F_X(x) \left( \frac{\Lambda_1}{\theta_1} e^{-x/\theta_1} + \frac{\Lambda_* \Lambda_1^{1/\theta_*}}{\theta_* \theta_1} e^{-x/(\theta_* \theta_1)} \right) = F_X(x) \Psi_*(x)$$

in cui  $\Lambda_*$ ,  $\Lambda_1$  e  $\theta_1$  sono i quattro parametri che caratterizzano il modello probabilistico.

### 1° Livello di regionalizzazione

Per i due parametri di forma del modello,  $\theta_*$  e  $\Lambda_*$ , si può assumere un valore costante all'interno di ampie zone omogenee.

I parametri  $\theta_1$  e  $\Lambda_1$  possono essere invece desunti dalla singola serie applicando il metodo della massima verosimiglianza.

Si definisce la funzione di verosimiglianza come:

$$\ln L = \sum_{i=1}^n \ln f_X(x_i) = \sum_{i=1}^n \ln F_X(x_i) + \sum_{i=1}^n \ln \Psi_*(x_i)$$

con  $n$  numero di dati della serie. Le formule risolutive per la stima dei parametri sono le seguenti (da non imparare a memoria!):

$$\frac{\partial \ln L}{\partial \theta_1} = 0 \Leftrightarrow \theta_1 = \frac{\sum_{i=1}^n x_i e^{-x_i/\theta_1} + \frac{\Lambda_* \Lambda_1^{1/\theta_*}}{\theta_*^2 \Lambda_1} \sum_{i=1}^n x_i e^{-x_i/\theta_* \theta_1}}{\sum_{i=1}^n x_i e^{-x_i/\theta_1} + \frac{\Lambda_* \Lambda_1^{1/\theta_*}}{\theta_* \Lambda_1} \sum_{i=1}^n x_i e^{-x_i/\theta_* \theta_1} + \sum_{i=1}^n \frac{e^{-x_i/\theta_1}}{\Psi_*(x_i)} + \frac{\Lambda_* \Lambda_1^{1/\theta_*}}{\theta_* \Lambda_1} \sum_{i=1}^n \frac{e^{-x_i/\theta_* \theta_1}}{\Psi_*(x_i)}}$$

$$\frac{\partial \ln L}{\partial \Lambda_1} = 0 \Leftrightarrow \Lambda_1 = \frac{\frac{\Lambda_1}{\theta_1} \left[ \sum_{i=1}^n \frac{e^{-x_i/\theta_1}}{\Psi_*(x_i)} + \frac{\Lambda_* \Lambda_1^{1/\theta_*}}{\theta_*^2 \Lambda_1} \sum_{i=1}^n \frac{e^{-x_i/\theta_* \theta_1}}{\Psi_*(x_i)} \right]}{\sum_{i=1}^n e^{-x_i/\theta_1} + \frac{\Lambda_* \Lambda_1^{1/\theta_*}}{\theta_* \Lambda_1} \sum_{i=1}^n e^{-x_i/\theta_* \theta_1}}$$

Noti  $\theta_*$  e  $\Lambda_*$ , lo schema iterativo combina queste due ultime espressioni ipotizzando 2 valori di primo tentativo per  $\theta_1$  e  $\Lambda_1$  nella prima delle due, ricavando  $\theta_1$  e sostituendolo nella seconda insieme con il  $\Lambda_1$  di 1° tentativo. Si ottiene così il valore  $\Lambda_1$  di 2° tentativo e si ritorna alla prima espressione ripetendo il calcolo con i 2 valori di secondo tentativo.

$$\theta_1^{i+1} = f_a(\theta_1^i, \Lambda_1^i)$$

$$\Lambda_1^{i+1} = f_b(\theta_1^{i+1}, \Lambda_1^i)$$

Il criterio di terminazione del calcolo si basa sul confronto tra due stime successive di  $\theta_1$  e  $\Lambda_1$ .

$$\text{Errore} = \left| \theta_1^{i+1} - \theta_1^i \right| + \left| \Lambda_1^{i+1} - \Lambda_1^i \right| \leq \text{tolleranza}$$

Come valori di primo tentativo per  $\theta_1$  e  $\Lambda_1$  si possono adottare i parametri  $\alpha$  ed  $\varepsilon$  della legge di Gumbel stimati dalla serie di dati in esame e quindi porre  $\theta_1 = 1/\alpha$  e  $\Lambda_1 = e^{\alpha\varepsilon}$

## 2° livello di regionalizzazione

Oltre ai valori costanti dei parametri  $\theta_*$  e  $\Lambda_*$  nelle zone omogenee, all'interno di queste è possibile identificare sottozone omogenee, entro cui si può ritenere costante anche il parametro di scala  $\Lambda_1$ . Dalla singola serie viene stimato, quindi, solo il parametro  $\theta_1$  con il metodo della massima verosimiglianza utilizzando l'espressione fornita precedentemente per il livello 1.

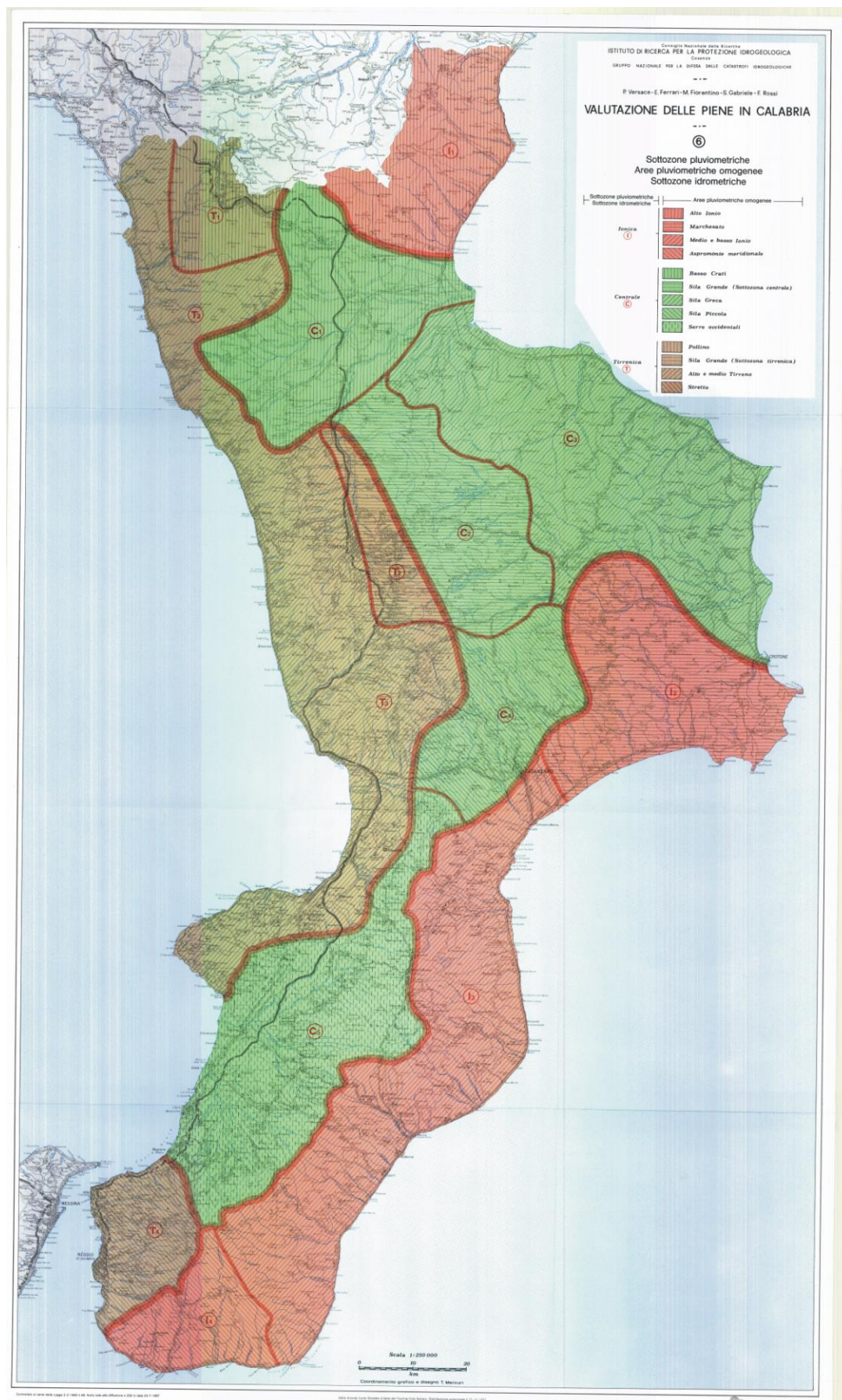


Fig. 2 - Sottozone e aree omogenee individuate nel progetto VAPI Calabria