



L'analisi statistica e probabilistica delle precipitazioni piovose estreme

Luigi Fanizzi. Ecoacque® (info@ecoacque.it)

L'elaborazione e l'analisi dei dati pluviometrici, vengono effettuate al fine di ricercare la relazione fra l'altezza h_p , delle precipitazioni piovose, e la loro durata t . Le relazioni $h_p = h_p(t)$, sono generalmente espresse nella forma monomia $h_p = a \times t^n$, dove le costanti a ed n sono determinate col metodo dei minimi quadrati (G. C. Frega, 1984) ed opportunamente ragguagliate, per tener conto dell'estensione della superficie del bacino scolante, secondo le equazioni di U. Puppini (L. Fanizzi, 2010). Le curve che si ottengono sono dette **curve segnalatrici di possibilità pluviometrica (CSPP)**; l'analisi viene eseguita considerando il cosiddetto tempo di ritorno idrologico (T_r), cioè quel periodo medio di tempo, espresso in anni, in cui un valore, d'intensità assegnata, viene uguagliato, o superato, almeno una volta. Stime ragionevolmente affidabili, per la summenzionata analisi, sono possibili solo per valori di T_r non troppo elevati: orientativamente $T_r \leq 2 \times N+1$ (con N pari al numero di anni, definiti dal campione di osservazioni; mod. A. Guadagni, 2003).

Brevi note teoriche

Scopo dell'analisi statistico-probabilistica, dei dati relativi alle massime precipitazioni piovose intense (p. di scroscio, di breve durata e massima intensità), è quello di far corrispondere, ad ogni valore di una variabile casuale o stocastica (che può assumere, cioè, valori diversi, in dipendenza dell'aleatorietà dei fenomeni meteorologici), la probabilità che si verifichi un evento maggiore od uguale a quel valore, almeno una volta, ossia di individuare, per ogni evento, il suo tempo di ritorno T_r . L'analisi probabilistica è necessaria in quanto, mentre per i dati, rilevati in passato, si può definire la frequenza (ossia il numero di volte in cui un evento si è presentato, in una serie di manifestazioni), per i dati futuri, occorre introdurre il concetto di **probabilità**, ossia il rapporto tra il numero di casi favorevoli, al verificarsi di un certo evento, ed il numero dei casi possibili. Poiché, quindi, non è lecito identificare frequenza con probabilità, è necessario estendere, artificialmente, il campo delle osservazioni, in studio, individuando una **distribuzione di probabilità** che si adatti alla serie di osservazioni note. L'analisi probabilistica consente di valutare eventi con T_r superiore al numero di anni definito dalla consistenza del campione dei dati, con un'attendibilità che si riduce, però, all'aumentare dello stesso T_r . Nell'elaborazione probabilistica i dati di precipitazione piovosa si considerano **variabili casuali**, cioè governate dalla Legge del Caso e si suppone che la serie dei valori osservati nel passato, costituisca un campione estratto dalla popolazione di tale variabile casuale.

Distribuzione di probabilità

Com'è noto, si definisce funzione di densità di probabilità $p(x)$, quella funzione che moltiplicata per l'ampiezza infinitesima dx , rappresenta la probabilità che si verifichi un valore nell'intervallo $[x, x + dx]$. Tale

funzione è legata alla funzione di probabilità $P(x \leq X)$, denominata anche **probabilità cumulata di non superamento**, secondo la relazione:

$$P(x \leq X) = \int_{-\infty}^x p(x) dx \text{ (integrale da } -\infty \text{ ad } X)$$

Poiché la probabilità che la variabile x_{tesima} (in **posizione d'ordine i**), assuma un qualsiasi valore compreso tra $-\infty$ e $+\infty$ è uguale ad **1** (evento certo), si avrà che:

$$\int_{-\infty}^{+\infty} p(x) dx = 1 \text{ (integrale da } -\infty \text{ a } +\infty)$$

Per essere $P(x \leq X)$ la probabilità cumulata di non superamento e $P(x > X)$ la probabilità di superamento, si può scrivere:

$$P(x \leq X) = 1 - \left(\frac{1}{T_r}\right) = \frac{(T_r - 1)}{T_r}$$

In quanto T_r è notoriamente definito come il **numero medio di anni per cui un determinato evento è eguagliato o superato almeno una volta** (E. Usai, 2008):

$$T_r = \frac{1}{(1 - P(x \leq X))} = \frac{1}{P(x > X)}$$

Statisticamente, cioè, il tempo di ritorno di un determinato evento è formalmente definibile come l'**inverso della probabilità di accadimento** dell'evento stesso. Nel presente contributo, ogni **campione** si considera costituito dai massimi valori annuali di **precipitazione piovosa intensa effettiva** (> 1 mm/giorno), di varia durata e può essere assimilato ad un sotto campione particolare di tutti i possibili valori verificatisi nel periodo di osservazione. La distribuzione di probabilità, da adattare al campione, e qui utilizzata, è la **distribuzione Normale** o di **Gauss**. Una variabile x , pertanto, si dirà distribuita secondo la **Legge Normale** (o distribuita *normalmente*) se la sua **funzione di densità di probabilità (PDF)** e la sua **funzione di distribuzione cumulata (CDF)**, hanno rispettivamente la forma (G. Anglani Frega, 1982):

$$p(x) = \frac{1}{\sqrt{2 \cdot \pi \cdot \sigma^2}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \text{e} \quad P(x) = \int_{-\infty}^x \frac{1}{\sqrt{2 \cdot \pi \cdot \sigma^2}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

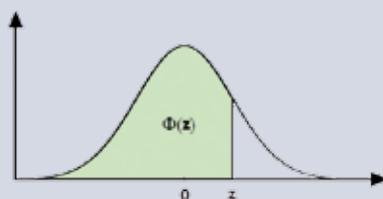
in cui le grandezze μ e σ sono i **parametri caratteristici della distribuzione** e ne rappresentano rispettivamente la **media aritmetica** e lo **scarto quadratico medio**. Sostituendo alla variabile x , la **variabile ausiliaria z** , definita come:

$$z = \frac{x - \mu}{\sigma}$$

chiamata **variabile ridotta standardizzata**, che ha **media aritmetica** pari a "0" e **scarto quadratico medio** pari a "1", la **CDF** (cd funzione di ripartizione frequenze cumulate), diventa:

$$P(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz$$

L'andamento della **funzione normale $P(z)$** , è noto (f. di **Gauss-Laplace**) ed esistono Tabelle che ne associano il valore ad ogni valore di z (vedi **Tabella 1**; C. Fallico, 1992).



z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
4.0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Per valori di $z < 0$, data la simmetria della distribuzione Gaussiana, vale la relazione: $\Phi(-z) = 1 - \Phi(+z)$

Tabella 1 - Funzione di ripartizione della distribuzione normale standard $\Phi(z)$ per valori di $z \geq 0$.



Accettabilità della distribuzione normale

La distribuzione Normale viene adattata al campione stesso, attraverso la determinazione dei parametri caratteristici. Si assume, perciò, come vera l'ipotesi statistica che la variabile casuale, cioè il valore dei massimi annuali, sia distribuita secondo un'assegnata funzione di probabilità (campione gaussiano). È necessario, dunque, verificare l'accettabilità di tale ipotesi e, di conseguenza, valutare l'adattamento della distribuzione al campione. Tale verifica può essere effettuata con l'utilizzo dei test statistici, cioè di procedimenti che consentono di decidere, sulla base di osservazioni, di cui si dispone, se accettare o meno una generica ipotesi statistica H_0 (esempio: campione gaussiano). Nel caso in cui l'ipotesi statistica, si basi sulla verifica di una determinata distribuzione, il Test è detto non parametrico (ossia non dipendente dalla distribuzione) è, cioè, un Test di verifica dell'ipotesi statistica (verifica della forma della distribuzione campionaria). Con l'utilizzo del Test s'introduce una variabile casuale ordinale continua con distribuzione di probabilità nota e si verifica che il valore del parametro teorico, desunto dallo studio del campione, sia inferiore ad un determinato valore critico che dipende dal livello di significatività prescelto. Il livello di significatività "a" indica la probabilità di rigettare l'ipotesi statistica H_0 quando, invece, questa sia verificata; si può così individuare anche una regione di accettazione alla quale corrisponde la probabilità $(1 - \alpha)$, cioè la probabilità che il parametro stimato cada all'interno della regione prefissata; in tal caso, l'ipotesi H_0 , viene accettata. Usualmente, in idrologia applicata, il valore del livello di significatività a, che si presceglie è pari al valore 0,05; questo significa che la probabilità massima, con cui si accetta di rischiare, di compiere un errore, è del 5%. La verifica dell'ipotesi che la variabile casuale segua effettivamente una funzione di probabilità assegnata viene qua effettuata, con il Test non parametrico di Kolmogorov-Smirnov. Tale Test è basato sull'esame dello spostamento fra la funzione di distribuzione osservata $F(x)$ e la funzione di distribuzione teorica ipotizzata $P(x)$. La grandezza adottata, come misura dello spostamento, è la massima differenza D , in valore assoluto, tra le due funzioni $F(x)$ e $P(x)$. Per quanto summenzionato, la distribuzione della cosiddetta statistica D , è indipendente dalla funzione di distribuzione teorica cumulata (CDF), ipotizzata come vera, ed il suo unico parametro è rappresentato dalla dimensione del campione N .

Il test di Kolmogorov-Smirnov

I valori del campione osservato, vengono ordinati in funzione dell'ampiezza, mettendo al primo posto il valore con ampiezza minore (ordinamento crescente). Si costruisce, quindi, la funzione di distribuzione osservata $F_o(x)$, mediante la relazione (M. Greppi, 2005):

$$F_o(x(i)) = \frac{i}{N}$$

ove:

i = indicatore della posizione nel campione ordinato;

N = numerosità del campione.

Con tutti i punti, così determinati, si ricava una curva continua. Dopo aver costruito la funzione di distribuzione osservata, si mette a confronto con quella teorica ipotizzata (per differenza, in valore assoluto), costruendo la seguente variabile D , che verrà impiegata come statistica del Test (Im. Chakravarti et Al., 1967):

$$D = \max_{i=1}^N \{ |F_o(x_i) - F_G(x_i)| \}$$

Praticamente D , indica di quanto le due curve discostano nei punti osservati. A questo punto, per il test si fissa il livello d'incertezza ossia di significatività α e da opportune tabelle, specifiche per il test di Kolmogorov-Smirnov (K-S), si ricava il valore d_{α} , da usare nel test:

$$P_r \{ (K-S) \geq d_{\alpha} \} = \alpha$$

Infine si confronta la statistica del test con d_{α} e l'ipotesi fatta sulla funzione di distribuzione viene accettata se:

$$P_r \{ 0 < D < d_{\alpha} \} = 1 - \alpha$$

Il vantaggio di impiegare questo test è nella sua semplicità e, soprattutto, dal fatto che l'esito del test non è condizionato dalla dimensione del campione.

Esempio di analisi

Per $Tr = 5$ anni (p.to 8.3.5. di cui al DPCM 4 marzo 1996), sia dato il campione, costituito da $N = 11$ osservazioni (AA 1990 ÷ 2001: Stazione pluviometrica di Bari Osservatorio - $L_T 41^\circ 07' 05''$, 7 ; $L_G 16^\circ 52' 21''$, 40), delle massime altezze annuali di precipitazione intensa di durata pari ad 1 ora:

$$h_{ii} = x_i = (23.4, 21.8, 23.4, 17.2, 16, 12.6, 20.6, 24.4, 19.4, 24, 29)$$

Poiché si sta facendo l'ipotesi di distribuzione gaussiana, si possono utilizzare dei valori stimati per descrivere la funzione di distribuzione teorica $F_G(x)$. I principali parametri che descrivono le caratteristiche della funzione di probabilità, del campione, possono raggrupparsi in due categorie (L. Di Comite et Al., 1998):

- 1) Parametri che valutano la tendenza centrale, dell'insieme di dati, come la media aritmetica (momento di primo ordine o di ordine 1, rispetto all'origine $\mu = 0$):

$$\mu = \frac{1}{N} \sum_{i=1}^N x_i = 21.07$$

- 2) Parametri che misurano la dispersione che, in genere, è valutata rispetto alla media aritmetica, come la varianza (momento di secondo ordine o di ordine 2, rispetto a μ):

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2 = 13.32$$

I valori della funzione di distribuzione teorica ipotizzata, si possono ricavare dalle tabelle per la gaussiana $N(21.07, 13.32)$.

Poiché si ha a disposizione la Tabella della funzione gaussiana standard o di ripartizione della variabile casuale normale standardizzata (Tabella 1), si potrà porre, per il campione ordinato:

$$F_G(x) = \Phi \left(\frac{x - \mu}{\sqrt{\sigma^2}} \right) = \Phi(z)$$

$$F_G(12.6) = \Phi \left(\frac{12.6 - 21.07}{\sqrt{13.32}} \right) = \Phi(-2.32) = 1 - 0.9898 = 0.0102$$

$$F_G(16.0) = \Phi \left(\frac{16 - 21.07}{\sqrt{13.32}} \right) = \Phi(-1.39) = 0.0823$$

$$F_G(17.2) = \Phi \left(\frac{17.2 - 21.07}{\sqrt{13.32}} \right) = \Phi(-1.06) = 0.1446$$

$$F_G(19.4) = \Phi \left(\frac{19.4 - 21.07}{\sqrt{13.32}} \right) = \Phi(-0.46) = 0.3228$$

$$F_G(20.6) = \Phi \left(\frac{20.6 - 21.07}{\sqrt{13.32}} \right) = \Phi(-0.13) = 0.4483$$

$$F_G(21.8) = \Phi \left(\frac{21.8 - 21.07}{\sqrt{13.32}} \right) = \Phi(0.20) = 0.5793$$

$$F_G(23.4) = \Phi \left(\frac{23.4 - 21.07}{\sqrt{13.32}} \right) = \Phi(0.64) = 0.7389$$

$$F_G(23.4) = \Phi \left(\frac{23.4 - 21.07}{\sqrt{13.32}} \right) = \Phi(0.64) = 0.7389$$

$$F_G(24) = \Phi \left(\frac{24 - 21.07}{\sqrt{13.32}} \right) = \Phi(0.80) = 0.7881$$

$$F_G(24.4) = \Phi \left(\frac{24.4 - 21.07}{\sqrt{13.32}} \right) = \Phi(0.91) = 0.8186$$

$$F_G(29) = \Phi \left(\frac{29 - 21.07}{\sqrt{13.32}} \right) = \Phi(2.17) = 0.985$$

Mentre, i valori della **funzione di distribuzione osservata**, per $N = 11$, valgono:

$$F_o(12.6) = \frac{1}{11} = 0,0909 \quad F_o(16.0) = \frac{2}{11} = 0,1818$$

$$F_o(17.2) = \frac{3}{11} = 0,2727 \quad F_o(19.4) = \frac{4}{11} = 0,3636$$

$$F_o(20.6) = \frac{5}{11} = 0,4545 \quad F_o(21.8) = \frac{6}{11} = 0,5455$$

$$F_o(23.4) = \frac{7}{11} = 0,6364 \quad F_o(23.4) = \frac{8}{11} = 0,7273$$

$$F_o(24.0) = \frac{9}{11} = 0,8182 \quad F_o(24.4) = \frac{10}{11} = 0,9091$$

$$F_o(29.0) = \frac{11}{11} = 1.0000$$

Il punto di massima distanza, dalla **curva teorica ipotizzata** (vedi **Figura 1**), è:

$$D(17.2) = |F_o(17.2) - F_c(17.2)| = \text{ASS}(0.2727 - 0.1446) = \text{ASS}(0.1281).$$

Se si pone un livello d'incertezza pari ad $\alpha = 0.05$, la Tabella (**K-S**, Tab. 2; U. Maione et Al., 2003), per la **curva gaussiana** fornisce, per $N = 11$, un valore $d_\alpha = 0.3910$, poiché la statistica del test **D** è minore di questo valore, l'ipotesi statistica H_0 fatta (**campione gaussiano**), può essere accettata.

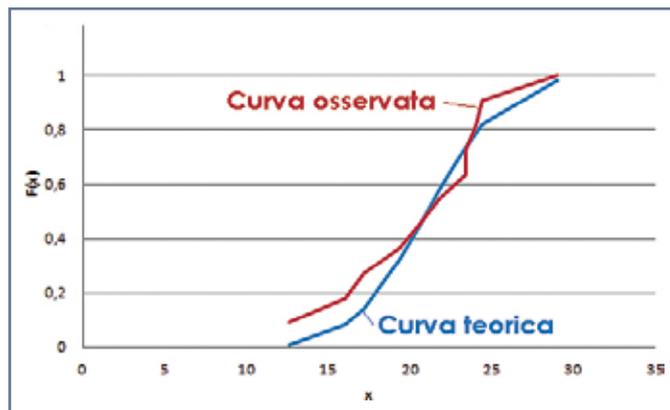


Figura 1 - Elaborazione grafica del Test di Kolgomorov-Smirnov.

$$D(12.6) = \text{ASS}(0.0807) \quad D(16.0) = \text{ASS}(0.0995)$$

$$D(17.2) = \text{ASS}(0.1281) \quad D(19.4) = \text{ASS}(0.0408)$$

$$D(20.6) = \text{ASS}(0.0062) \quad D(21.8) = \text{ASS}(0.0338)$$

$$D(23.4) = \text{ASS}(0.1025) \quad D(23.4) = \text{ASS}(0.0116)$$

$$D(24.0) = \text{ASS}(0.04301) \quad D(24.4) = \text{ASS}(0.0905)$$

$$D(29.0) = \text{ASS}(0.0150)$$

Dimensione campionaria N	Livello di significatività α				
	.20	.15	.10	.05	.01
1	.684	.925	.950	.975	.995
2	.684	.726	.776	.842	.929
3	.565	.597	.642	.708	.828
4	.494	.525	.564	.624	.733
5	.446	.474	.510	.565	.669
6	.410	.436	.470	.521	.618
7	.381	.405	.438	.486	.577
8	.358	.381	.411	.457	.543
9	.339	.360	.388	.432	.514
10	.322	.342	.368	.410	.490
11	.307	.326	.352	.391	.468
12	.295	.313	.338	.375	.450
13	.284	.302	.325	.361	.433
14	.274	.292	.314	.349	.418
15	.266	.283	.304	.338	.404
16	.258	.274	.295	.328	.392
17	.250	.266	.286	.318	.381
18	.244	.259	.278	.309	.371
19	.237	.252	.272	.301	.363
20	.231	.246	.264	.294	.356
25	.210	.220	.240	.270	.320
30	.190	.200	.220	.240	.290
35	.180	.190	.210	.230	.270
>35	1.07N	1.14N	1.22N	1.36N	1.63N

Tabella 2 - Valori di d_α per il Test di Kolgomorov-Smirnov.

Bibliografia

- (1) Im. Chakravarti, R.G. Laha, and J. Roy, (1967): "Handbook of Methods of Applied Statistics", Volume I, Ed. John Wiley and Sons, Mishawaka, Indiana (USA).
- (2) L. Fanizzi (2010): "Analisi delle piogge estreme", L'Ambiente, n. 1, Ed. ICSA, Milano (ITA).
- (3) A. Guadagni (2003): "Prontuario dell'ingegnere", Ed. U. Hoepli, Milano.
- (4) E. Usai (2008): "Manuale di idrologia per la progettazione", Ed. U. Hoepli, Milano.
- (5) G. Anglani Frega (1982): "Elementi di statistica idrologica", Ed. Liguori, Napoli.
- (6) M. Greppi (2005): "Idrologia", Ed. U. Hoepli, Milano.
- (7) U. Maione, U. Moisello (2003): "Elementi di statistica per l'idrologia", Ed. La Goliardica Pavese, Pavia.
- (8) L. Di Comite, P. Iaquina (1998): "Esercitazioni di statistica", Ed. Cacucci, Bari.
- (9) C. Fallico (1992): "Applicazioni di statistica idrologica", Ed. Bios, Cosenza.
- (10) G. C. Frega (1984): "Lezioni di acquedotti e fognature", Ed. Liguori, Napoli.