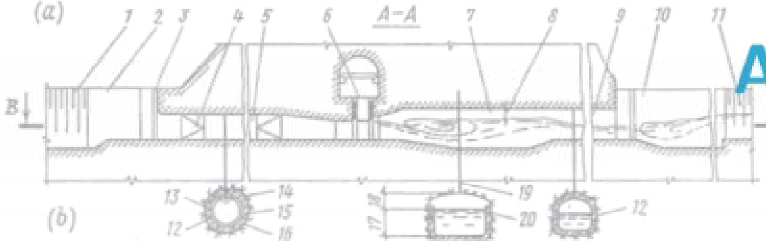


Un accenno ai metodi di regressione lineare

Il problema: Determinazione della miglior retta in grado di descrivere la relazione lineare che intercorre tra due variabili a partire dalle loro misura sperimentale … (ammesso che una relazione lineare esista …)

Consideriamo cioè: $N \; coppie \; di \; misure \; (x_i, y_i) \; con \; i = 1, \ldots, N$

Supponiamo che tra *x* e *y* sussista una relazione lineare

Assumiamo inoltre, per semplicità, che l'errore sperimentale sulla variabile *x* sia trascurabile (*y* variabile "risposta" e *x* variabile "esplicativa")

$$y_i = a x_i + b \quad + \epsilon_i$$

Metodo dei minimi quadrati (1) o metodo di Gauss

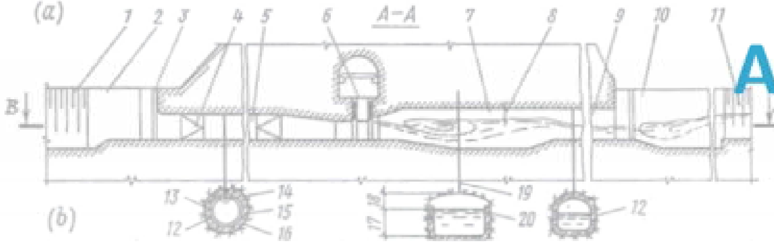
Definiamo quindi: $\epsilon_i = y_i - y_i^t$ $\quad con \quad$ $y_i^t = ax_i + b$

Per cui si ha: $\epsilon_i = y_i - ax_i - b$

Sia inoltre $\sigma_i^2 \ \ l'errore \ \ sperimentale \ \ sull'i-esimo \ \ valore \ \ di \ \ y$

$$\frac{\epsilon_i^2}{\sigma_i^2}$$

Scarto quadratico in rapporto all'errore di misura



Metodo dei minimi quadrati (2)

$$z = \sum_{i=1}^{N} \frac{\epsilon_i^2}{\sigma_i^2} = \sum_{i=1}^{N} \frac{(y_i - a x_i - b)^2}{\sigma_i^2}$$

Il problema si riduce quindi a trovare i valori di *a* e *b* in grado di rendere minimo *z*.

$$z = z(a, b)$$

$$\frac{\partial z}{\partial a} = -2 \sum_{i=1}^{N} \frac{x_i (y_i - a x_i - b)}{\sigma_i^2} = 0$$

$$\frac{\partial z}{\partial b} = -2 \sum_{i=1}^{N} \frac{(y_i - a x_i - b)}{\sigma_i^2} = 0$$

Metodo dei minimi quadrati (3)

Per cui, risolvendo il sistema di 2 equazioni in 2 incognite dopo aver posto:

$$S_{xx} = \sum_{i=1} \frac{x_i^2}{\sigma_i^2} \qquad S_{xy} = \sum_{i=1} \frac{x_i y_i}{\sigma_i^2}$$

$$S_x = \sum_{i=1} \frac{x_i}{\sigma_i^2} \qquad S_y = \sum_{i=1} \frac{y_i}{\sigma_i^2} \qquad S_0 = \sum_{i=1} \frac{1}{\sigma_i^2}$$

Metodo dei minimi quadrati (Soluzione)

Si ricava: $a = \frac{S_{xy} \cdot S_0 - S_x S_y}{S_{xx} \cdot S_0 - S_x^2}$ $\qquad b = \frac{S_y \cdot S_{xx} - S_x S_{xy}}{S_{xx} \cdot S_0 - S_x^2}$

E con $\sigma_i^2 = \sigma^2 = costante$

$$a = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\overline{x^2} - \bar{x}^2}$$

$$b = \frac{\overline{x^2} - \bar{x} \cdot \overline{xy}}{\overline{x^2} - \bar{x}^2}$$

Ancora sulla regressione lineare semplice

La miglior retta così determinata passa per il punto di coordinate:

$$x^* = \frac{\sum_{i=1}^{N} \frac{x_i}{\sigma_i^2}}{\sum_{i=1}^{N} \frac{1}{\sigma_i^2}} \qquad y^* = \frac{\sum_{i=1}^{N} \frac{y_i}{\sigma_i^2}}{\sum_{i=1}^{N} \frac{1}{\sigma_i^2}}$$

Baricentro dei dati sperimentali

Errore sui parametri

$$\sigma^2(a) = \frac{S_0}{S_0 S_{xx} - S_x^2} \qquad \sigma^2(b) = \frac{S_{xx}}{S_0 S_{xx} - S_x^2}$$

La distribuzione di Gumbel

Deriva da una distribuzione degli x_i del campione di tipo esponenziale (ma anche altre distribuzioni a coda grassa sono buone candidate: Weibull, Gamma, Lognormale...)

Convergono ad un'esponenziale per valori elevati di X

$$F_{X_{(k)}}(x) = \sum_{j=k}^{n} \binom{n}{j} [F_X(x)]^j \, [1 - F_X(x)]^{n-j}$$

Beta

$$F_{X_{MAX}}(x) = [F_X(x)]^n = (1 - e^{-\lambda x})^n$$

$$F_{X_{MAX}}(x) = [F_X(x)]^n = e^{-e^{-\frac{(x-u)}{\alpha}}}$$

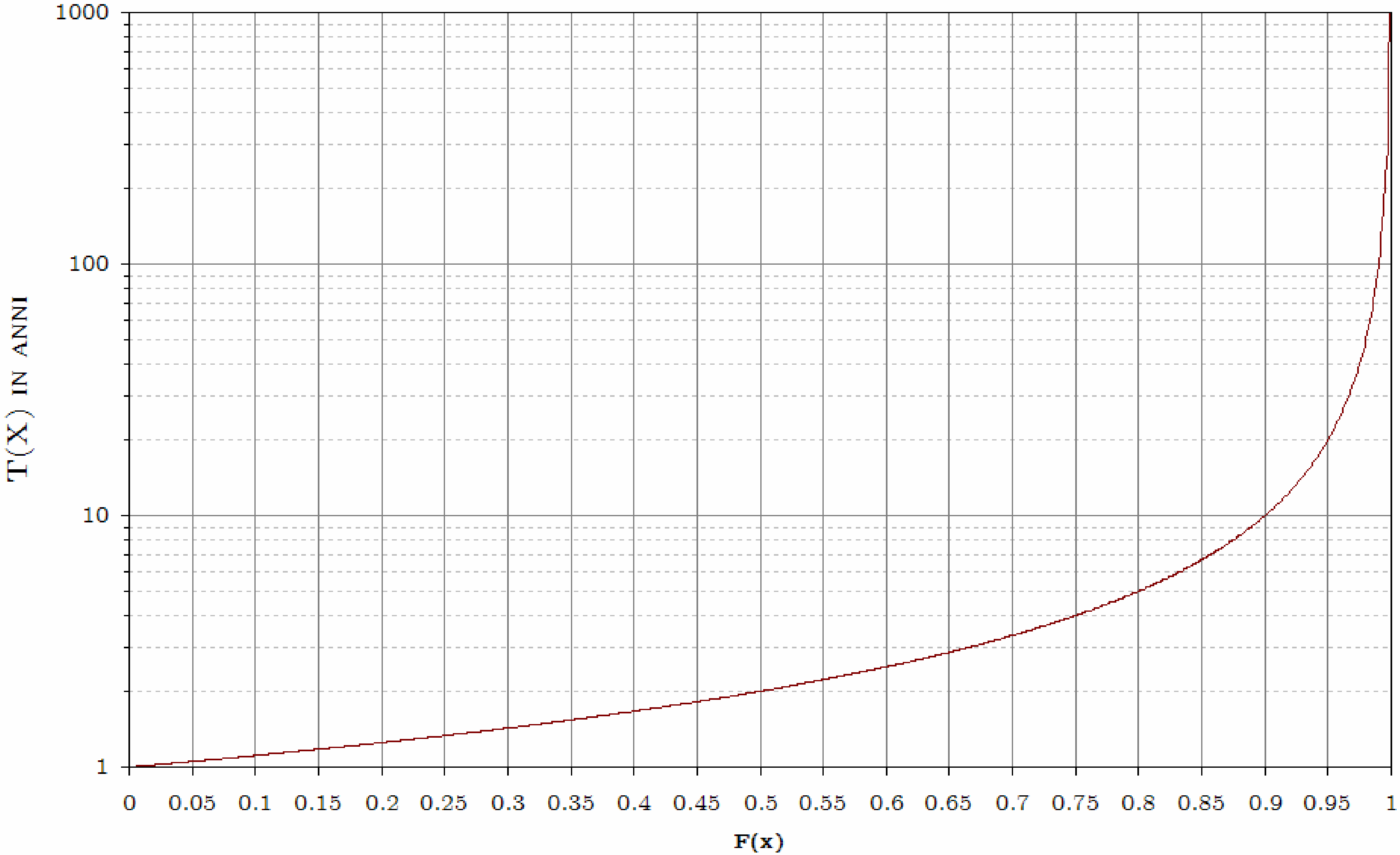
Ancora sulla distribuzione di Gumbel

$$f_{X_{MAX}}(x) = \frac{1}{\alpha}\exp\left[-\frac{(x-u)}{\alpha} - e^{-\frac{(x-u)}{\alpha}}\right] \qquad con\ -\infty < x < +\infty$$

$$u = \mu_{X_{MAX}} - \frac{n_e\sqrt{6}}{\pi}\sigma_{X_{MAX}}$$

$$\frac{\sqrt{6}}{\pi}\cdot\sigma_{X_{MAX}}$$

$$u = \mu_{X_{MAX}} - 0.5772\cdot\alpha$$


1000
100
10
1
T(X) IN ANNI
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
0.45
0.5
0.55
0.6
0.65
0.7
0.75
0.8
0.85
0.9
0.95
1
F(x)

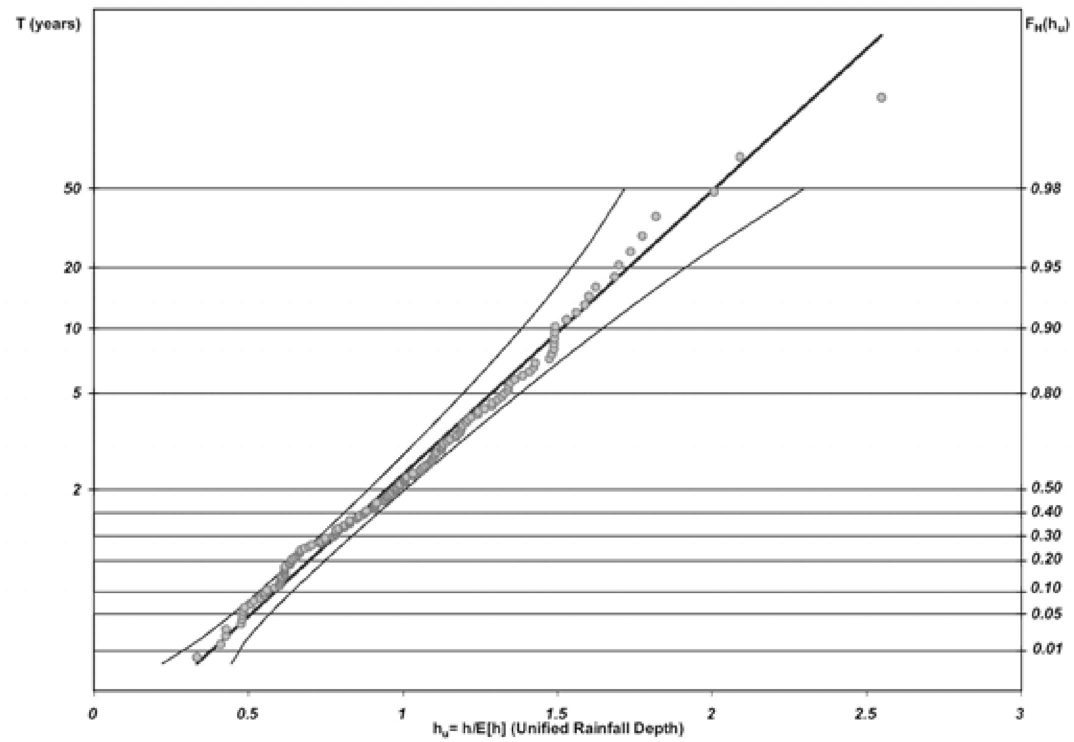
Carta di Gumbel

La distribuzione di Gumbel

$$F(x) = \exp\{-\exp[-(x - u)/\alpha\ ]\ \}$$

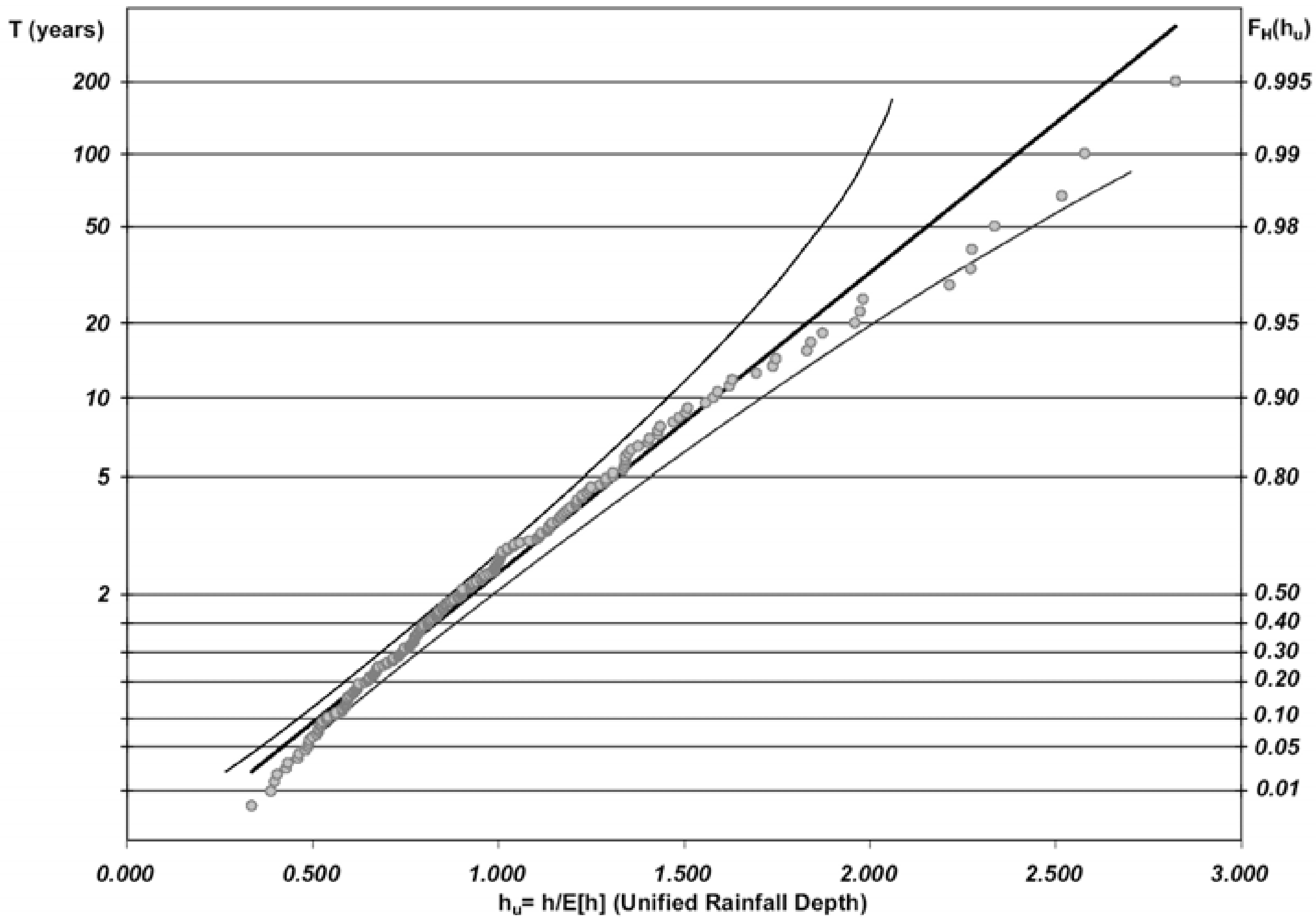
con la trasformazione lineare
$y = (\ x - u\ )/\alpha$
si riduce all' espressione

$$F(x) = \exp\{-\exp(-y)\ \}$$



La carta probabilistica di Gumbel è una carta che ha sui due assi rispettivamente le variabili x e y .
Sulla carta probabilistica ogni distribuzione di Gumbel è rappresentata da una retta.

Un esempio: distribuzione delle precipitazioni estreme a Chiavari (Genova)



Un po' di bibliografia ...

$con \quad -\infty < x < +\infty$

- Ugo Moisello (1999), *Idrologia Tecnica*, La Goliardica Pavese
- A.A.V.V.(1999), Sistemi di Fognatura: Manuale di progettazione, a cura del Centro Studi Idraulica Urbana, Hoepli Editore

Relazioni Altezza-Durata e Intensità-durata

costante

$$h(d,T) = a(T)\cdot d^{b}$$

Relazione altezza-durata per assegnato T

$$i(d,T) = a(T)\cdot d^{b-1}$$

Relazione intensità media-durata per assegnato T

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

Rischio di fallanza su un periodo di N anni

Alcuni semplici passaggi per la costruzione delle curve di possibilità pluviometrica a partire da una legge probabilistica di gumbel per gli estremi annuali di precipitazione

Supponiamo di voler valutare la massima altezza totale di pioggia h che si verifica con assegnato tempo di ritorno T ed in funzione della durata d.
Come abbiamo già avuto occasione di osservare, una schematizzazione semplice della legge di dipendenza di h da T e d può essere:

$$h(d,T) = a(T) \cdot d^b$$

Supponiamo inoltre che la distribuzione di probabilità dei massimi annuali di precipitazione (per durata d assegnata) sia di tipo EV1 (Gumbel):

$$F_X(x) = e^{-e^{-\frac{(x-u)}{\alpha}}}$$

Alcuni semplici passaggi per la costruzione delle curve di possibilità pluviometrica (2)

$$F_X(x) = e^{-e^{-\frac{(x-u)}{\alpha}}}$$

Data quindi la variabile ridotta y: $y = \frac{x-u}{\alpha}$

$$h = c\left\{1 - k \ln\left[\ln\left(\frac{T}{T-1}\right)\right]\right\} \cdot d^b$$

Che per assegnato T possiamo anche scrivere come:

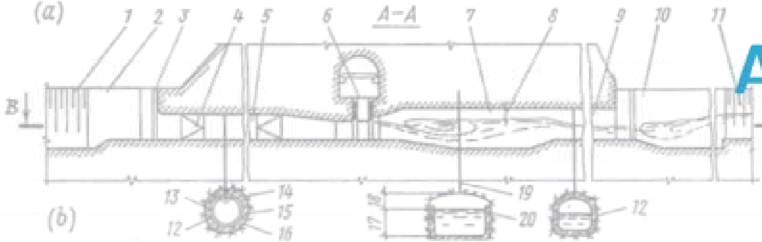
$$y_T = -\ln\left[\ln\left(\frac{T}{T-1}\right)\right]$$

$$k = \frac{0.78 \cdot c_{v\ unico}}{1 - 0.45 \cdot c_{v\ unico}}$$

Sotto l'ipotesi di Gumbel si ricava quindi con un po' di algebra:

$$h(d,T) = c\left\{1 - k \ln\left[\ln\left(\frac{T}{T-1}\right)\right]\right\} \cdot d^b$$

$$c_{v\ unico} = \sqrt{\frac{1}{D}\sum_{d=1} c_{v_d}^2}$$



Alcuni semplici passaggi per la costruzione delle curve di possibilità pluviometrica (3)

Se andiamo ora a calcolare il tempo di ritorno T corrispondente alla media delle altezze μ_h (per una qualsiasi durata) si ha:

$$\frac{T-1}{T} = \exp\{-\exp[-\alpha(\mu_h - \mu_h + 0.45\sigma_h)]\} = 0.57$$

$$\downarrow \qquad \downarrow \qquad \downarrow$$

$$T = \frac{1}{(1-0.57)} = 2.32$$

E sostituendo nell'espressione di a(T) per tutte le curve si ha:

$$a = c \cdot \{1 - k \ln[\ln \left(\frac{2.32}{2.32-1}\right)]\} =$$

$$= c \cdot (1 + 0.577k)$$

Alcuni semplici passaggi per la costruzione delle curve di possibilità pluviometrica (4)

Da cui si ottiene che se:

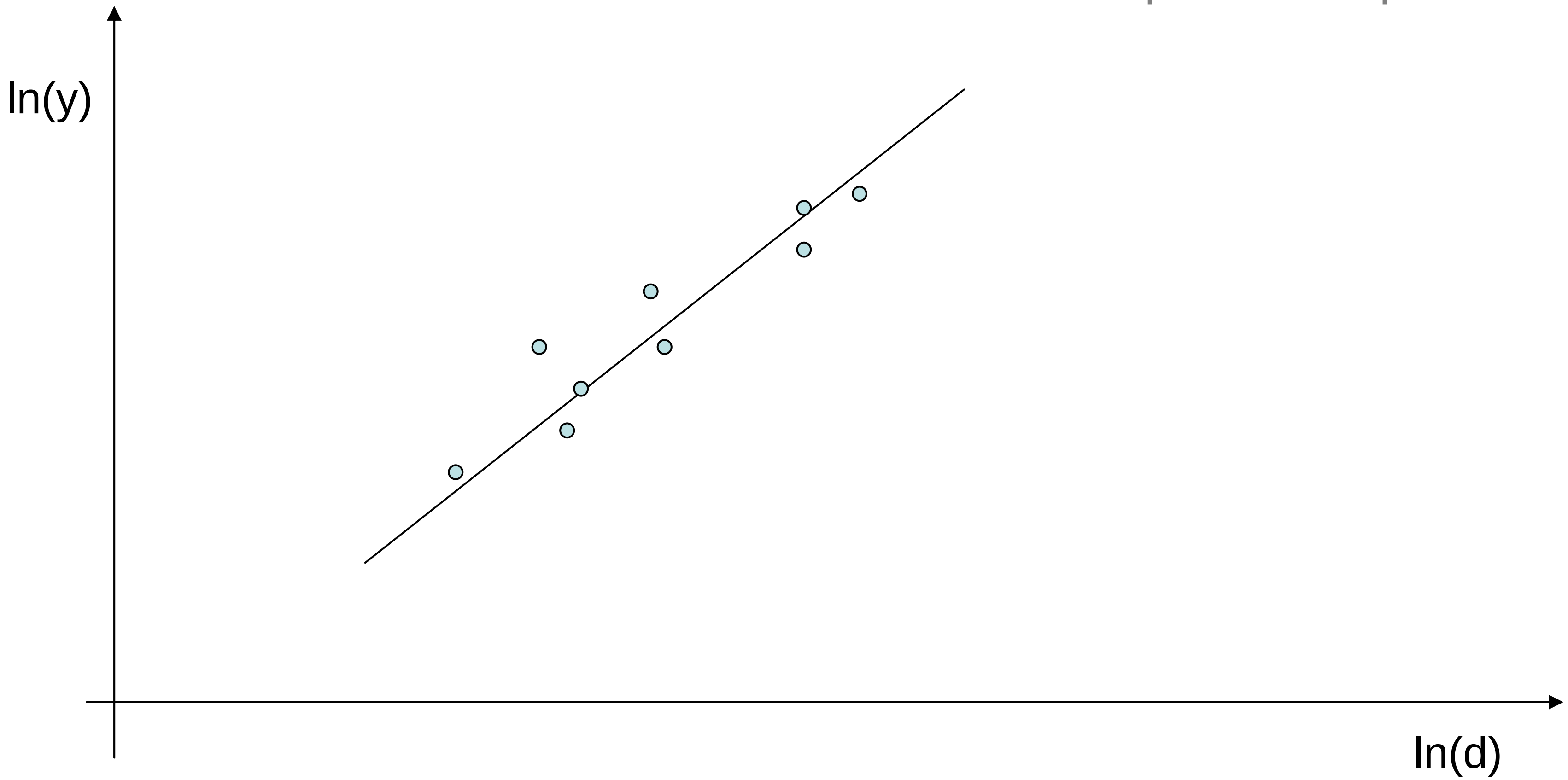
$$a = c \cdot (1 + 0.577k) \longrightarrow \mu_h = c \cdot (1 + 0.577k) \cdot d^b$$

Ponendo quindi: $$y_d = \frac{\mu_{h_d}}{(1+0.577k)} \longrightarrow y_d = c \cdot d^b$$

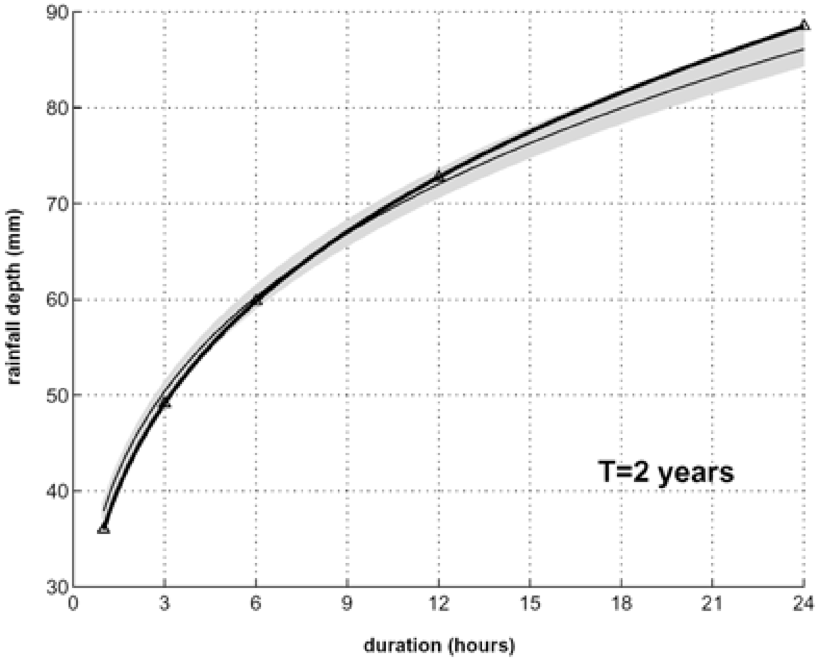
E linearizzando: $$\ln(y_d) = \ln(c) + b \cdot \ln(d)$$

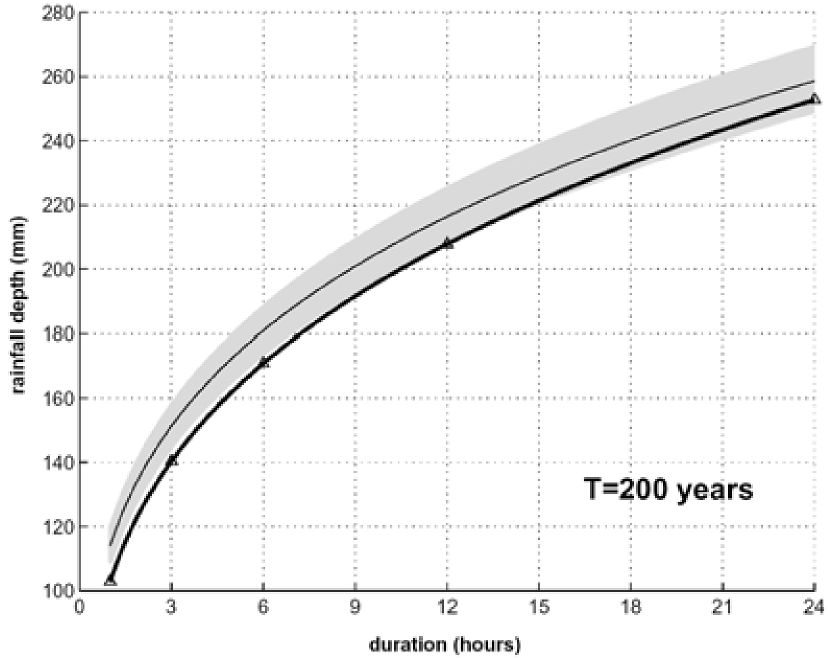
Per cui c e b possono essere stimati, tramite una regressione lineare, dal diagramma sperimentale ln(y) in funzione di ln(d)

Alcuni semplici passaggi per la costruzione delle curve di possibilità pluviometrica (5)



Un esempio: le cpp di Chiavari





Ietogramma Chicago