Le nuove frontiere dell'Idrologia
Potenza, 4 dicembre 2006
Università degli Studi della Basilicata
Facoltà di Ingegneria

Verso il superamento (?) dell'analisi statistica regionale in idrologia.

Attilio Castellarin – Università di Bologna
Francesco Laio – Politecnico di Torino

Analisi regionale di frequenza degli estremi idrologici: possibili aree di indagine

Approccio classico

[es. Hosking and Wallis, 1997]

A. Identificazione delle regioni omogenee

B. Scelta della distribuzione di frequenza

C. Stima dei parametri regionali

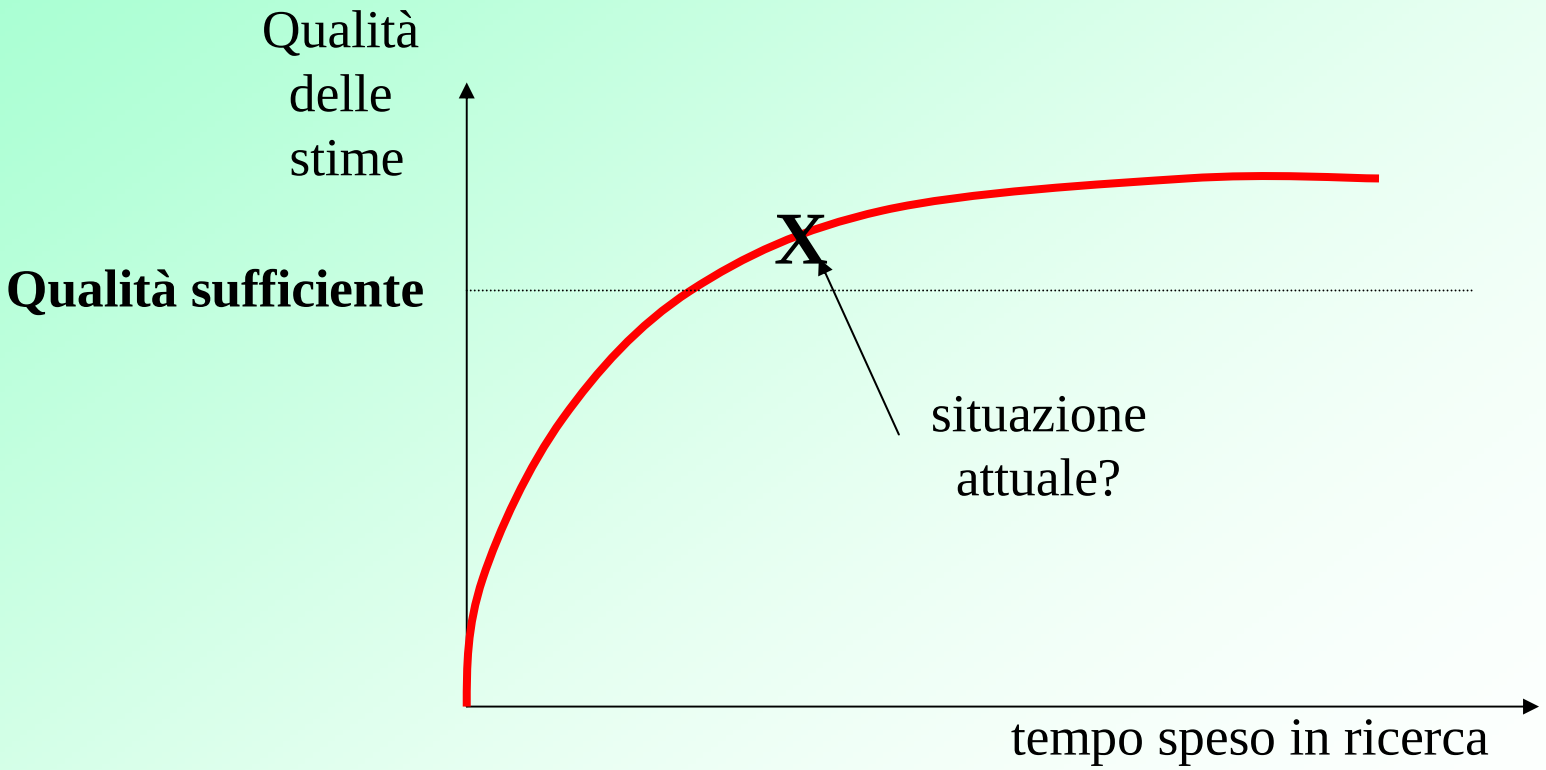
D. Verifica del modello regionale (validazione del modello e incertezza delle stime)

- Metodo di regionalizzazione
- Descrittori geomorfoclimatici
- Test di omogeneità
- Scelta della distribuzione
- Metodo di stima dei parametri regionali
- Stima della grandezza indice in siti non strumentati

Ha senso continuare a fare ricerca sui metodi statistici di analisi regionale degli estremi?

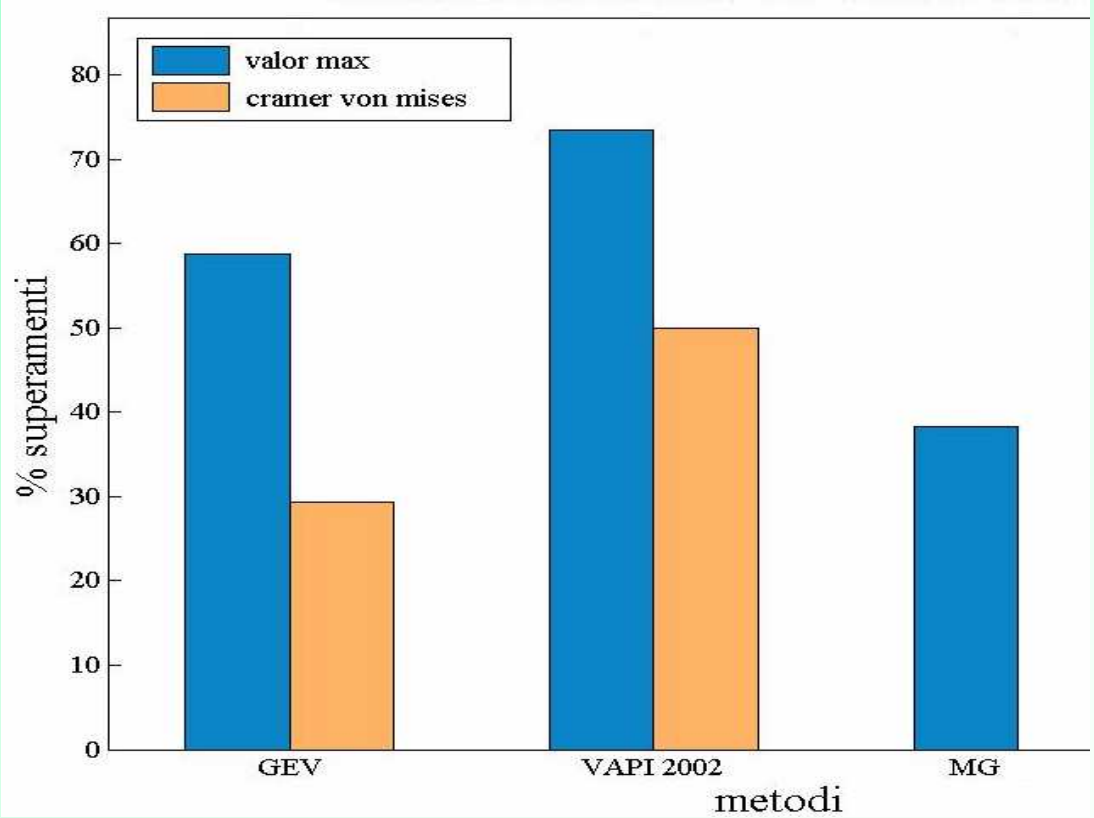
Non ha senso se:

a) i metodi esistenti producono già risultati soddisfacenti



Ha senso continuare a fare ricerca sui metodi statistici di analisi regionale degli estremi?

In realtà quando si applicano i metodi esistenti per la stima di Q_T in siti non strumentati spesso si ottengono risultati insoddisfacenti

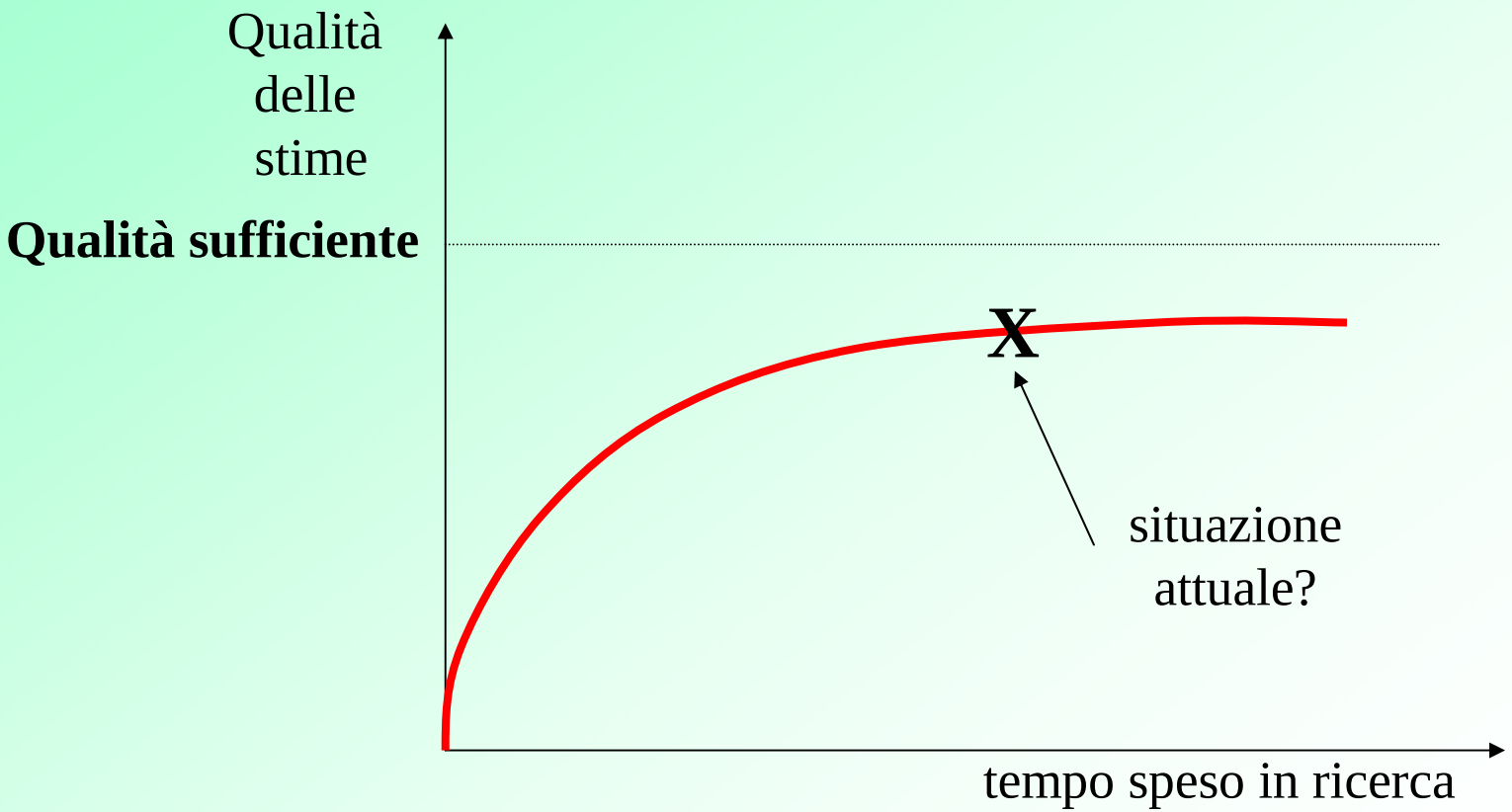


Esempio di test di adattamento applicati a stime regionali di Q_T in Piemonte

Ha senso continuare a fare ricerca sui metodi statistici di analisi regionale degli estremi?

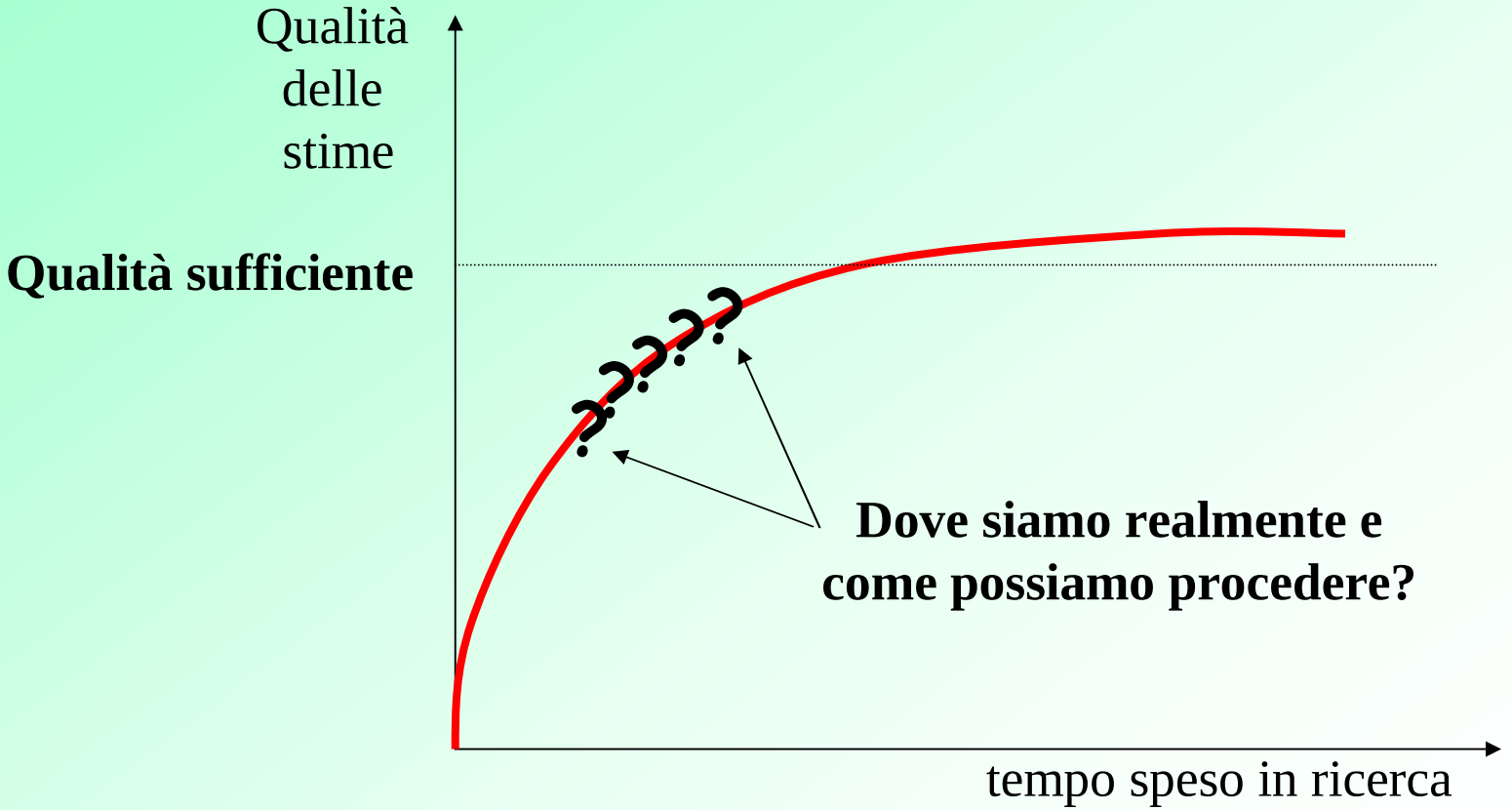
Non ha senso neppure se:

b) i metodi non sono pienamente soddisfacenti, ma si è raggiunto un “asintoto” nella qualità della rappresentazione



Ha senso continuare a fare ricerca sui metodi statistici di analisi regionale degli estremi?

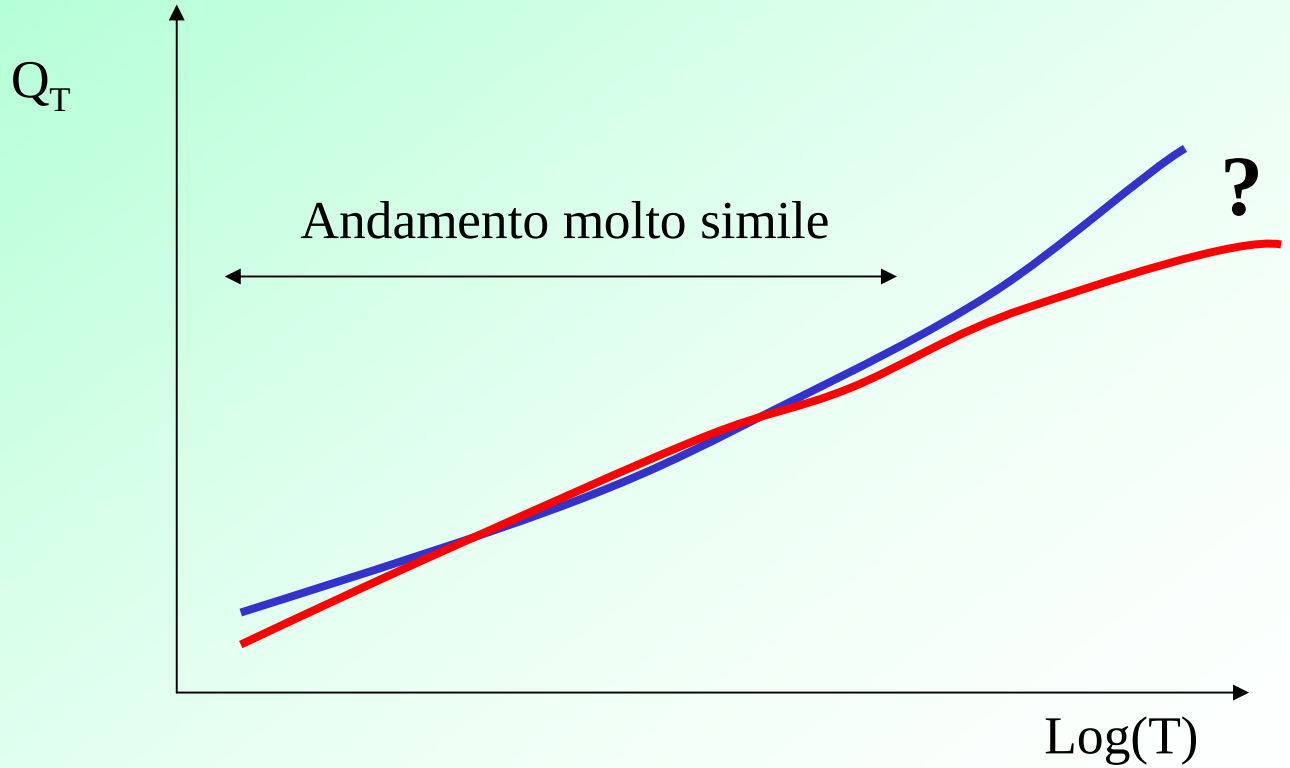
Dunque le stime sono spesso insoddisfacenti ai fini delle applicazioni pratiche. Inoltre, ci sono questioni aperte molto rilevanti che fanno pensare che non si sia raggiunto un asintoto.



Cosa può produrre un salto di qualità?

1. Scelta della distribuzione di probabilità?

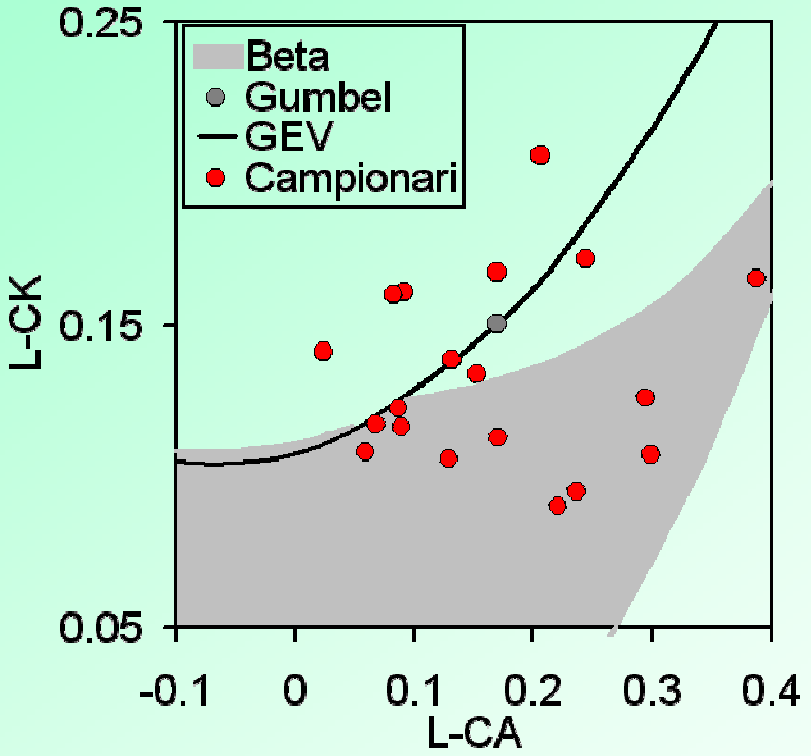
Probabilmente no. Sembra un ambito “chiuso” in cui non si hanno elementi concreti per stabilire se sia migliore una distribuzione rispetto all’altra per T alti, mentre per T bassi le distribuzioni sono molto simili se hanno (L-)momenti uguali



Cosa può produrre un salto di qualità?

Influenza della distribuzione di probabilità:

22 serie idrometriche statunitensi dei massimi annuali, $n \geq 50$ anni (PMF da US-NRC, reg. guide 1.59, 1977)



Rappresentazione del regime di frequenza di Q/PMF:

Distribuzione GEV
distribuzione di riferimento
(Met. L-mom.: $\lambda 1$, L-CV, L-CA)

Distribuzione EV1
(Met. L-mom.: λ_1, L-CV)

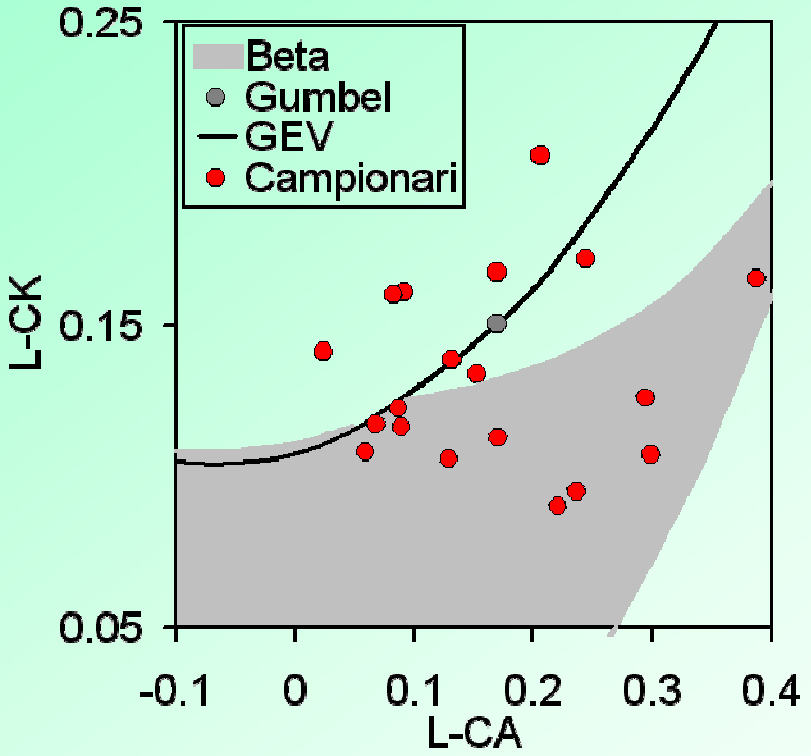
Distribuzione Beta
(Met. L-mom.: λ_1, L-CV)

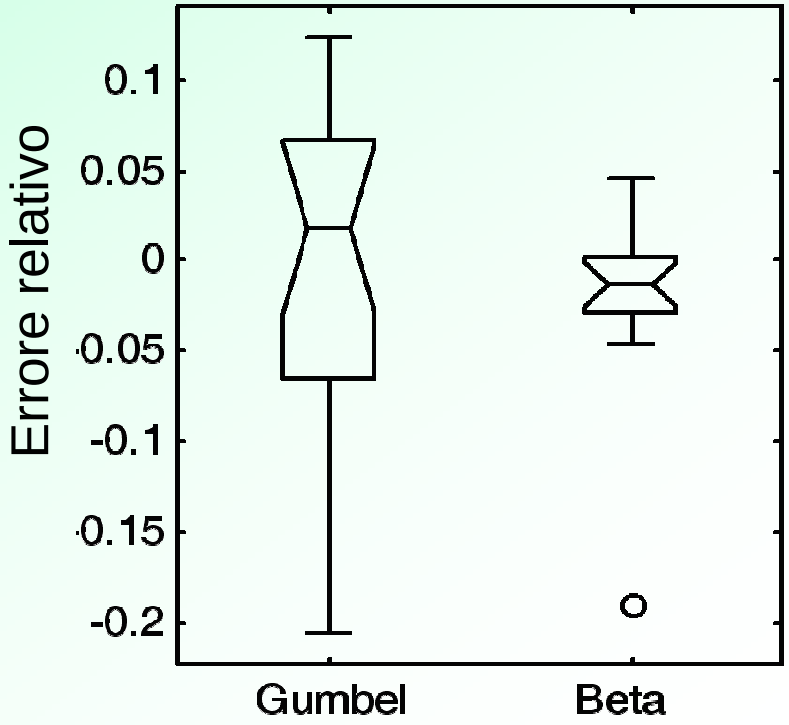
Cosa può produrre un salto di qualità?

Influenza della distribuzione di probabilità:

22 serie idrometriche statunitensi dei massimi annuali, $n \geq 50$ anni (PMF da US-NRC, reg. guide 1.59, 1977)

Confronto tra le stime dei quantili con T=50 anni:





Cosa può produrre un salto di qualità?

2. Scelta del metodo di stima dei parametri della distribuzione di probabilità regionale?
Probabilmente no. Piccole variazioni nell'incertezza di stima al variare del metodo. Asintoto raggiunto con il metodo degli L-momenti?

3. Scelta del metodo per verificare l'omogeneità delle regioni (test di omogeneità)?
Probabilmente no. L'omogeneità statistica non è così importante per ottenere buoni risultati. Inoltre diversi metodi di letteratura producono risultati simili $\Rightarrow$ asintoto raggiunto?

Cosa può produrre un salto di qualità?

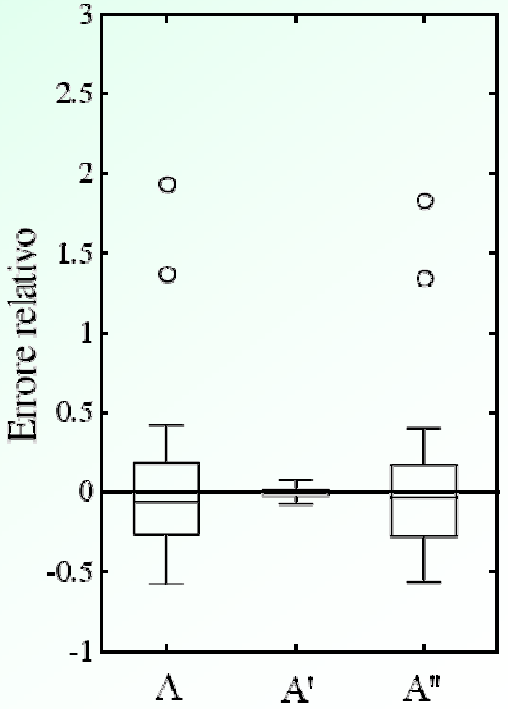
4. Stima della grandezza indice in siti non strumentati?
Probabilmente sì. Problema ancora aperto, la qualità delle stime è usualmente abbastanza scarsa ⇒ si può probabilmente migliorare, in particolare nella scelta dei descrittori o nell'impiego di metodi di derivazione concettuale.

Quantili regionali: esempio di cross-validazione
30 bacini emiliano-romagnoli e marchigiani

A) Metodo della piena indice

A') Componente di A) legata al fattore di crescita (distribuzione regionale)

A'') Componente di A) legata alla piena indice



Cosa può produrre un salto di qualità?

5. Scelta del metodo di regionalizzazione (piena indice, regioni gerarchiche, area di influenza, ….)?
Forse. In realtà l'impressione è che metodi diversi portino a variazioni marginali (vedi dopo) $\Rightarrow$ asintoto raggiunto?

6. Utilizzo di descrittori geomorfoclimatici nella fase di formazione delle regioni
Sì. E' un ambito finora piuttosto trascurato ma essenziale per migliorare la qualità delle stime in sezioni non strumentate

Metodi di regionalizzazione (premessa)

Evoluzione del concetto di regione omogenea
[v. ad es. Ouarda et al., *J. Hydrol.* 2001].

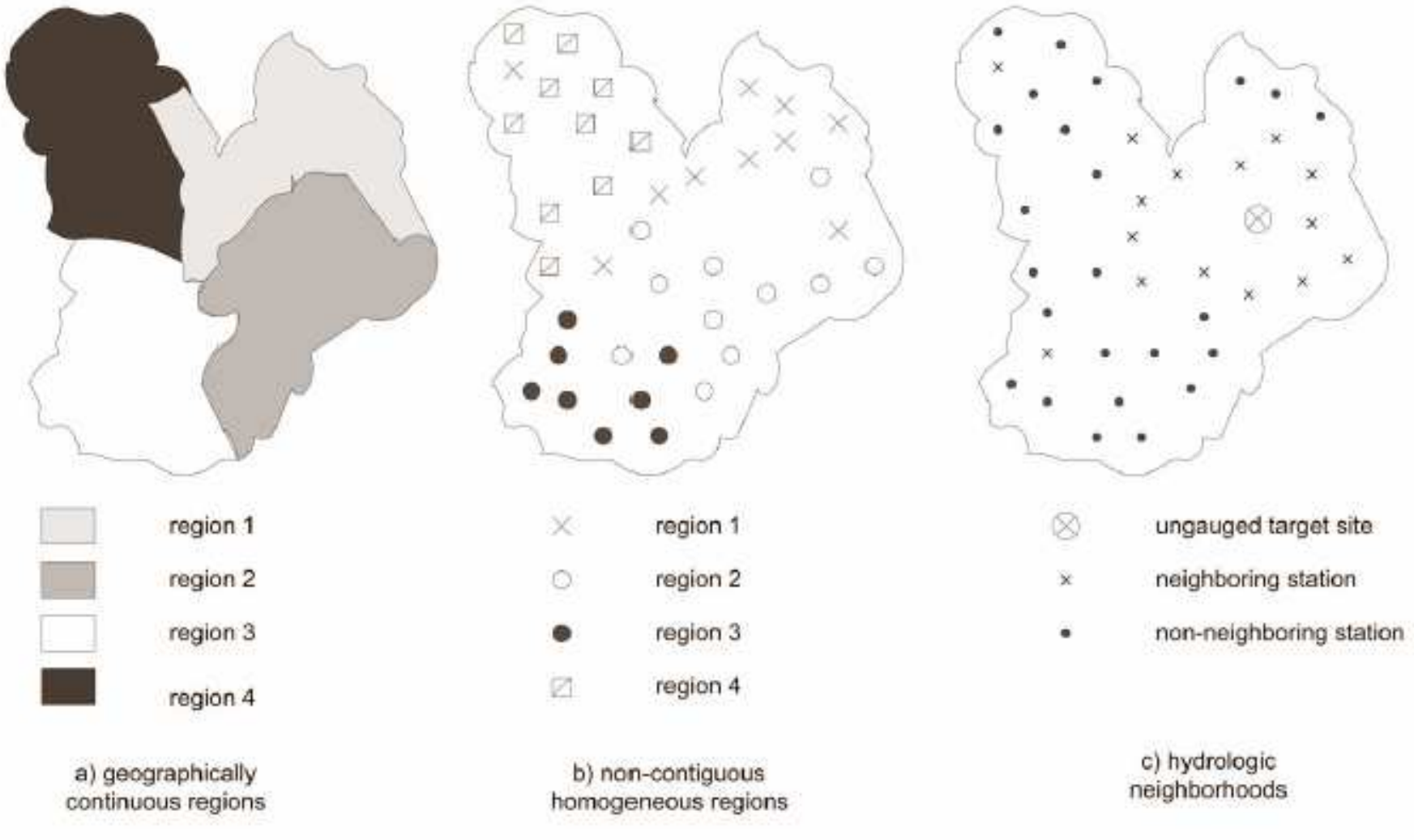


Fig. 1. Approaches for the delineation of homogeneous regions.

Metodi di regionalizzazione (premessa)

Il criterio dell'affinità idrologica (vicinaggio idrologico) è un concetto consolidato nell'analisi regionale di frequenza, v. ad es.:

Burn, WRR [1990] (Region of Influence approach): applicazione all'analisi di frequenza degli estremi idrometrici

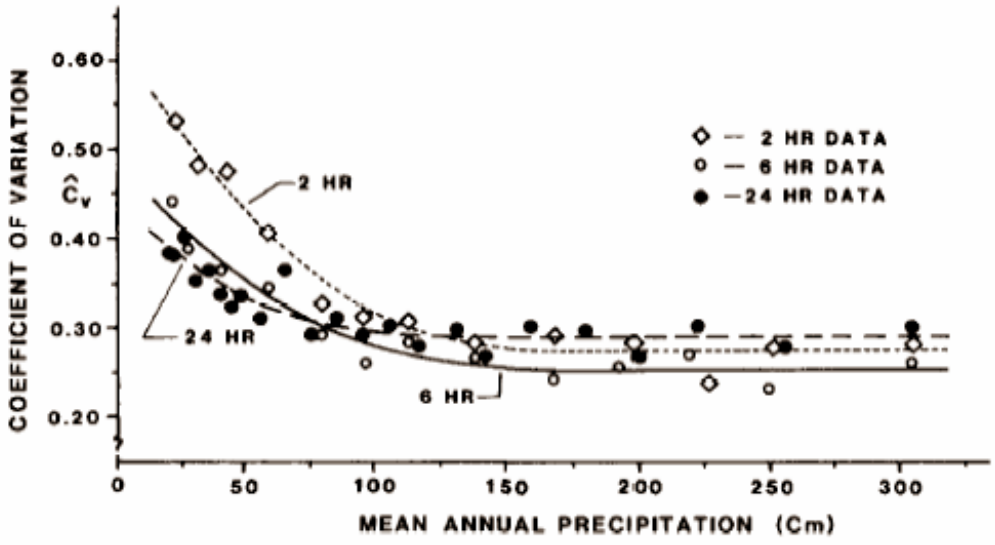
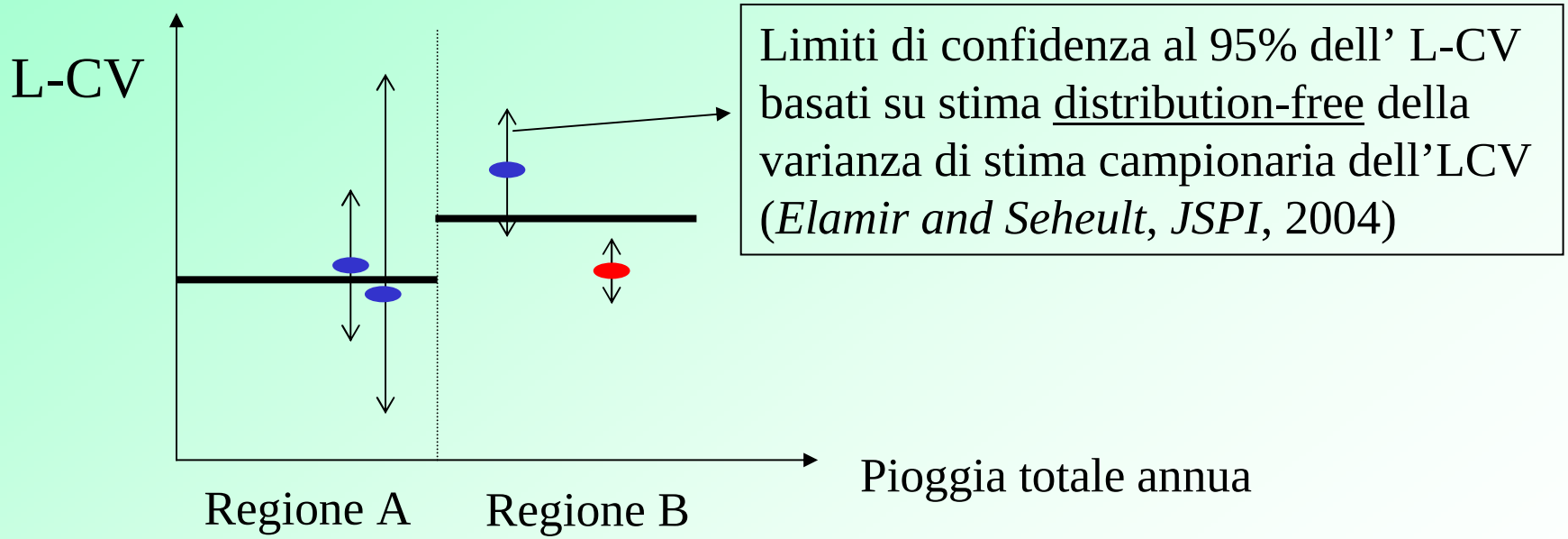


Fig. 5*a*. Regional generalized extreme value solutions for coefficient of variation as a function of mean annual precipitation.

Scheafer, WRR [1990]: applicazione all'analisi di frequenza degli estremi pluviometrici

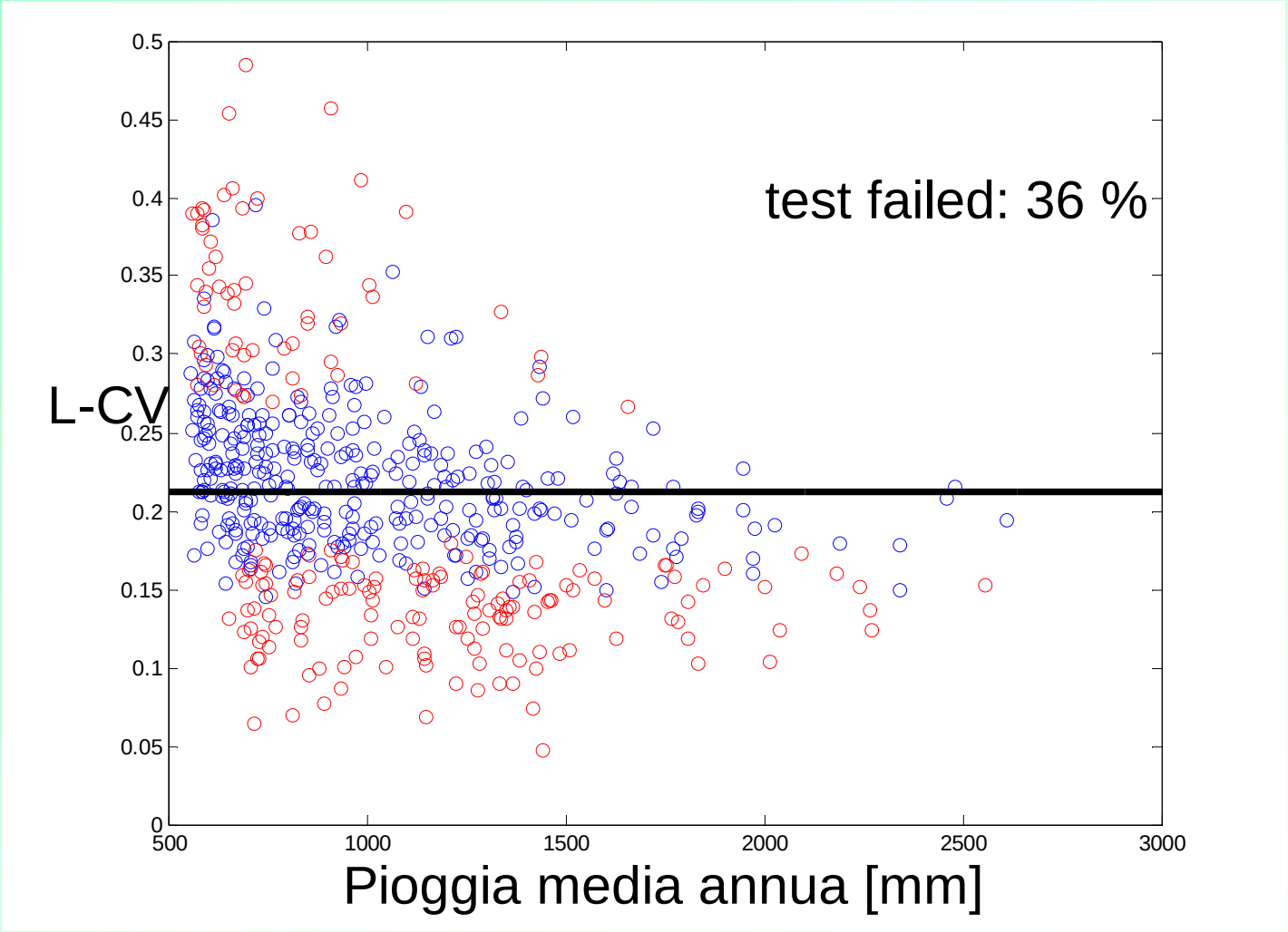
Scelta del metodo di regionalizzazione

Ad esempio si può ipotizzare di scegliere un **unico descrittore** (pioggia totale annua, PTA) per formare le regioni.
Si rappresentano L-CV o L-CA (con la loro incertezza) vs. PTA



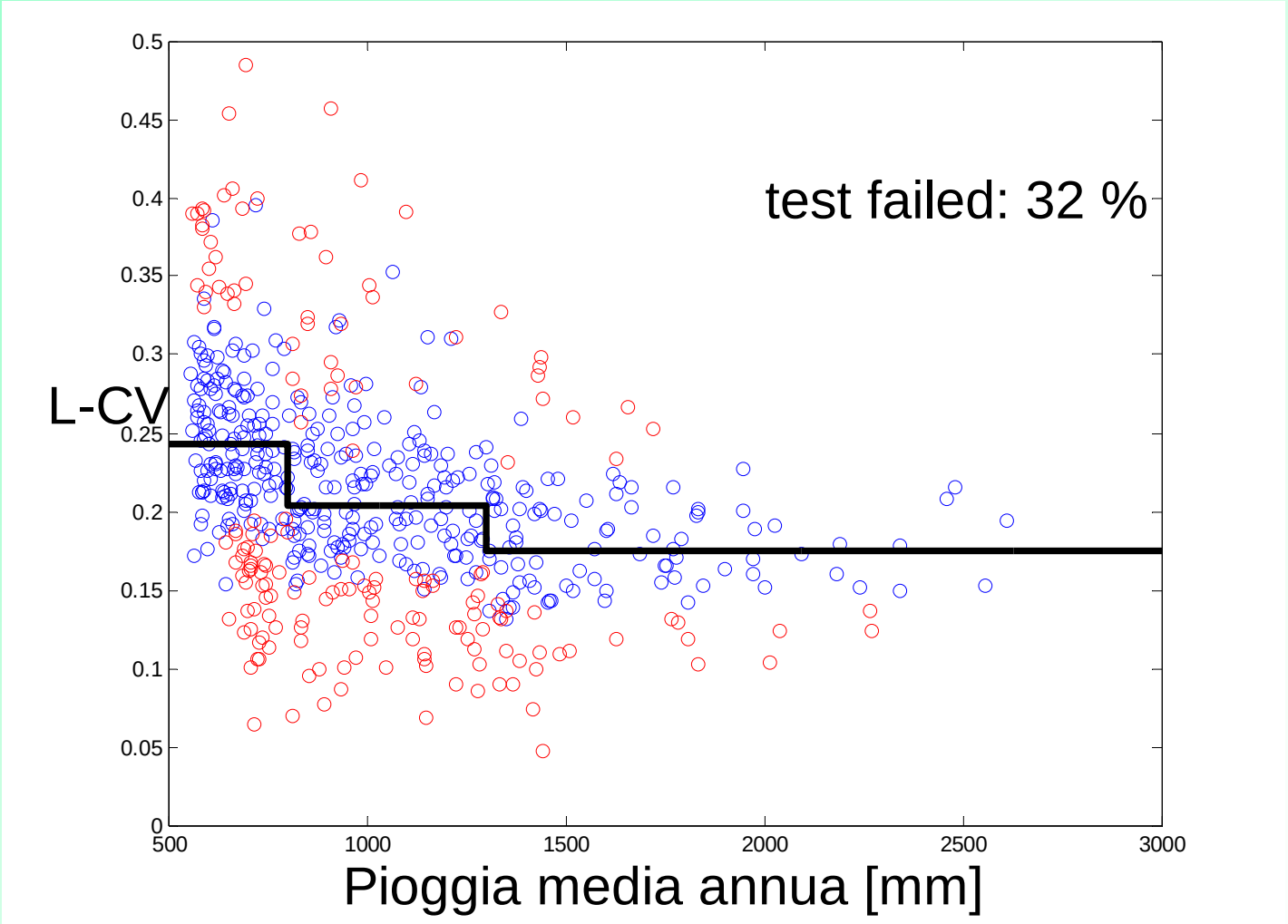
Si valuta quindi la percentuale di casi in cui i risultati dell'analisi regionale sono consistenti con i dati (punti blu: OK; punti rossi: NO)

Scelta del metodo di regionalizzazione



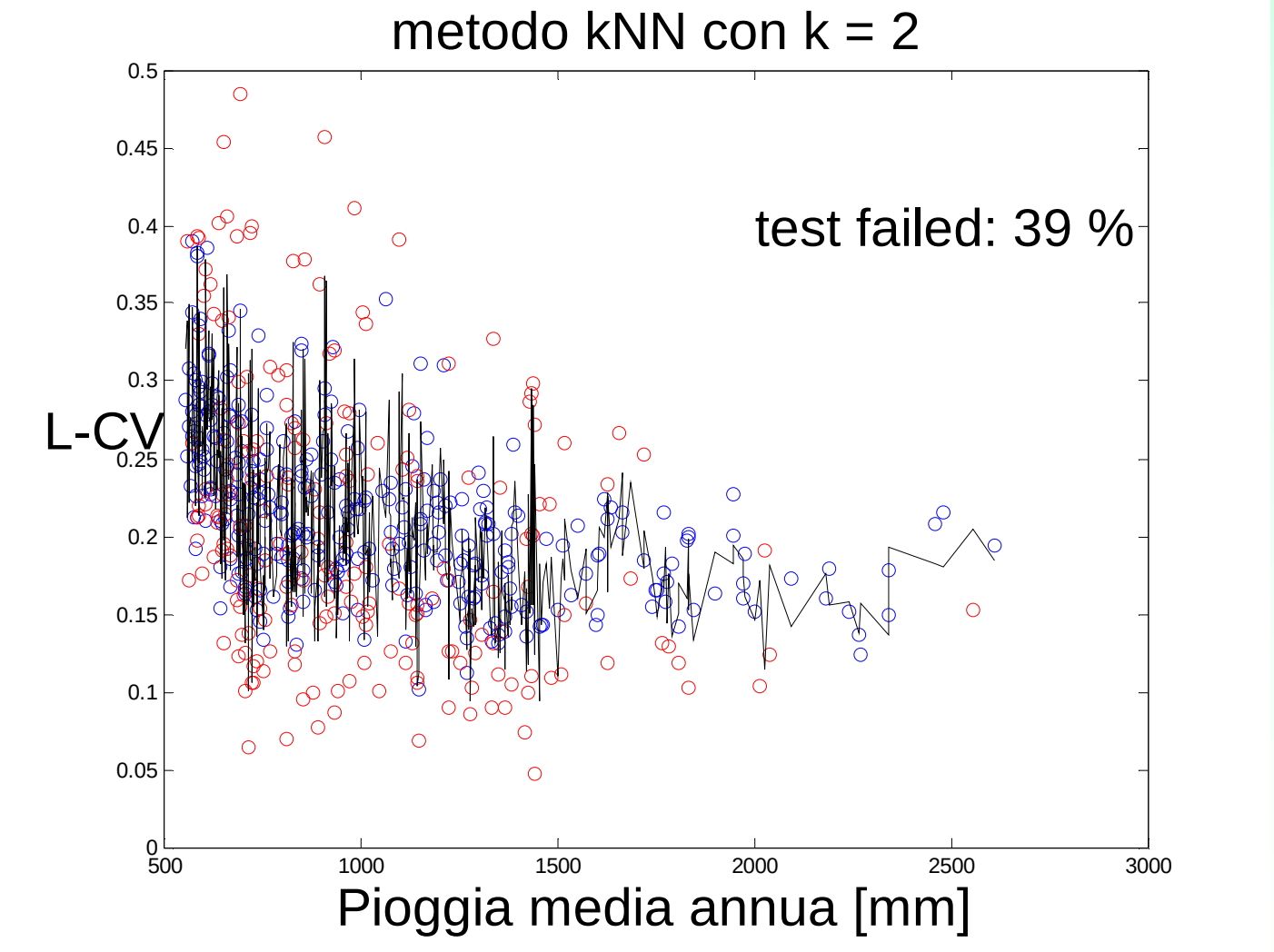
1. **Regione unica** $\Rightarrow$ non spiegato il 36% degli L-CV campionari

Scelta del metodo di regionalizzazione



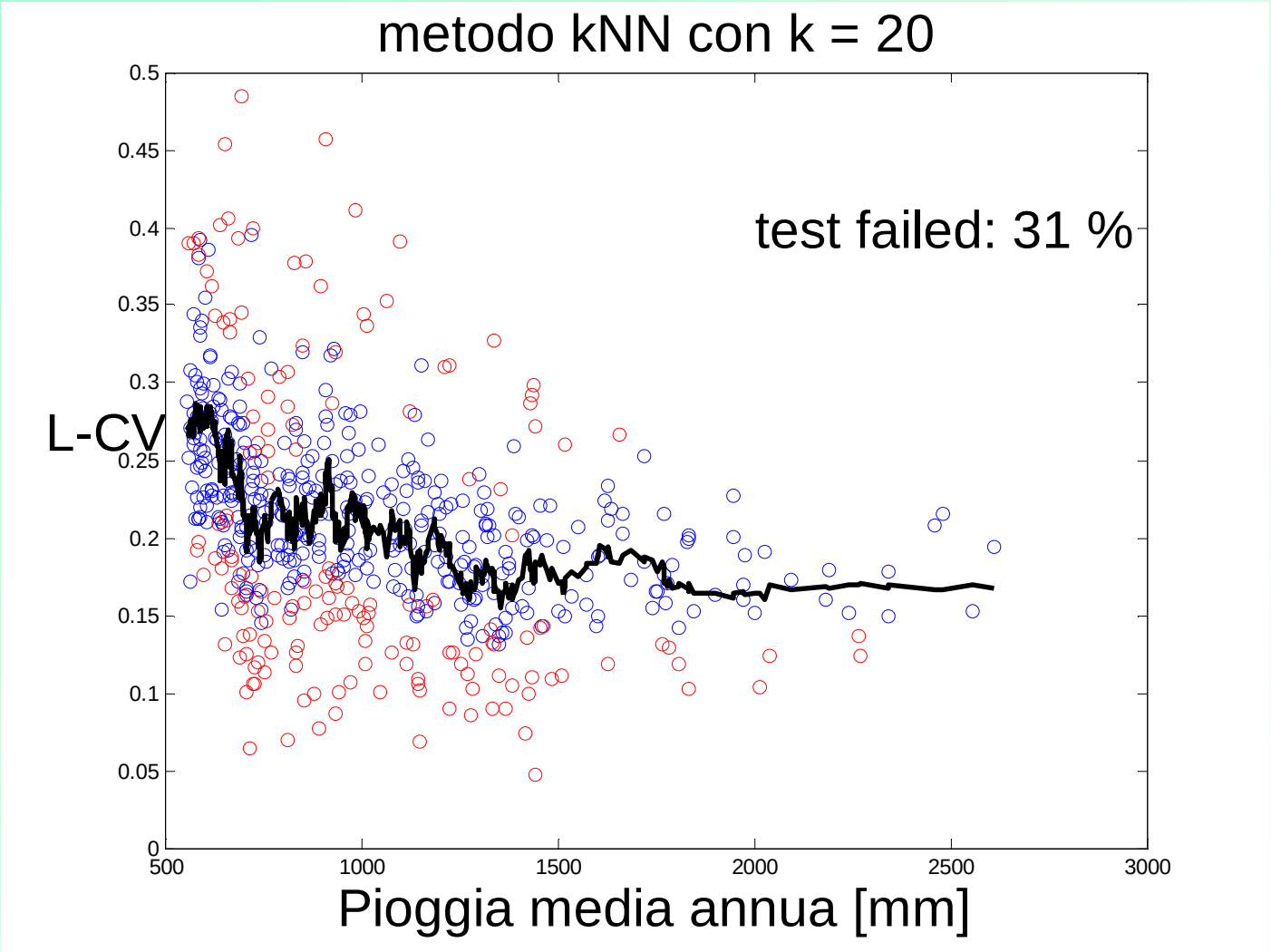
2. **Regioni fisse** $\Rightarrow$ non spiegato il 32% degli L-CV campionari

Scelta del metodo di regionalizzazione



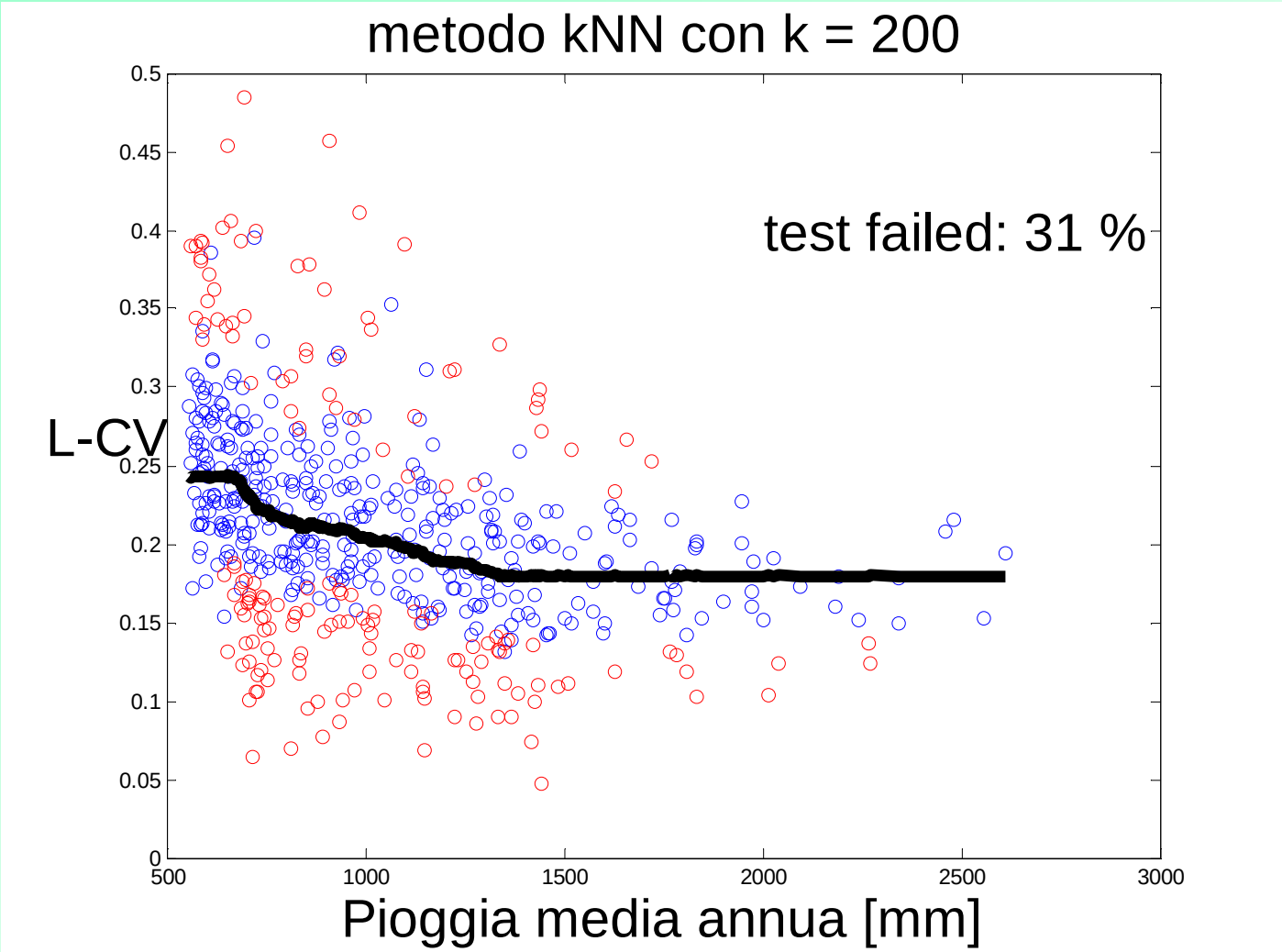
3. **Region of influence (regioni piccole:** 2 siti) ⇒ non spiegato il 39% degli L-CV campionari

Scelta del metodo di regionalizzazione



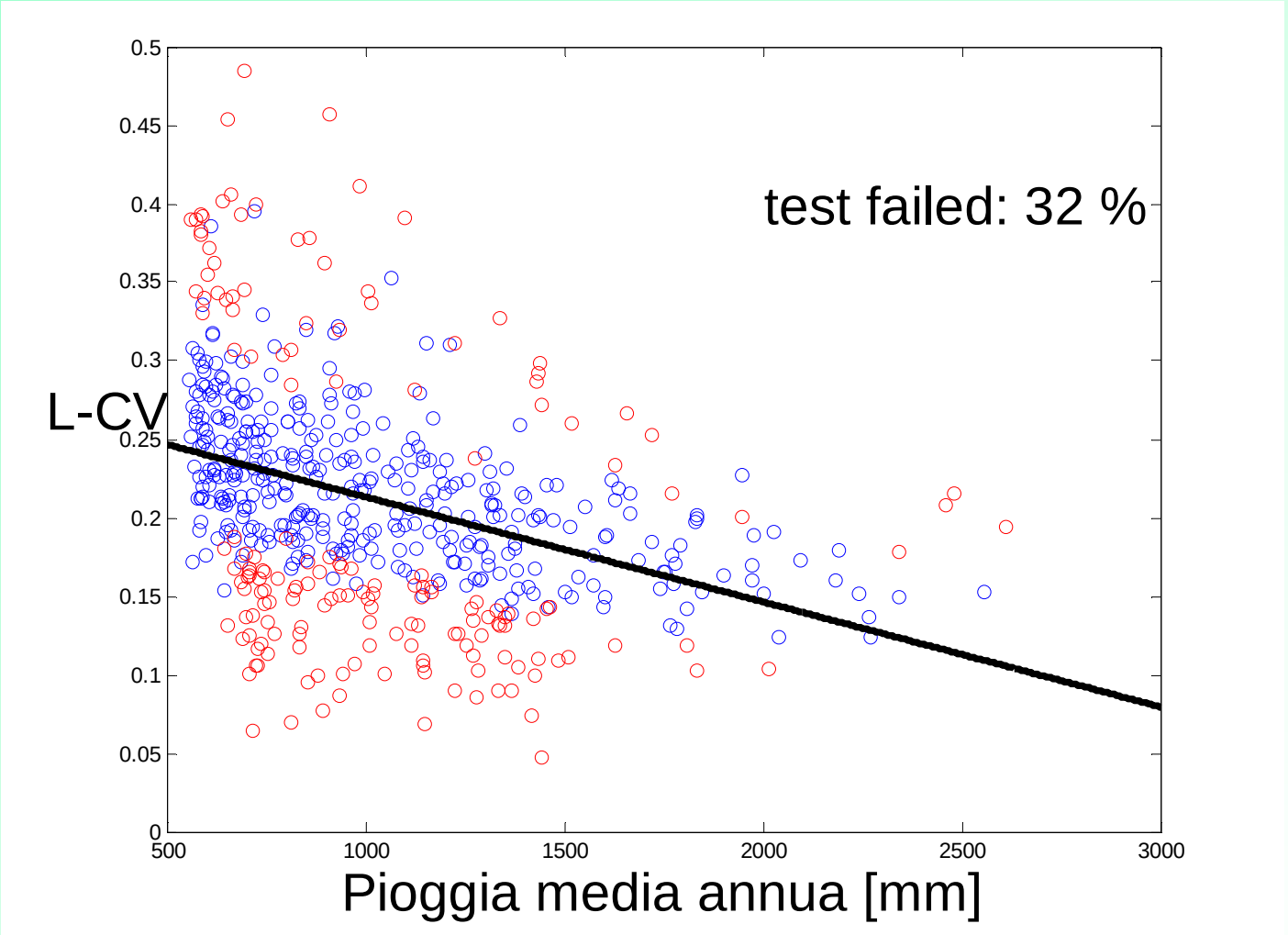
4. **Region of influence (regioni medie**: 20 siti) ⇒ non spiegato il 31% degli L-CV campionari

Scelta del metodo di regionalizzazione



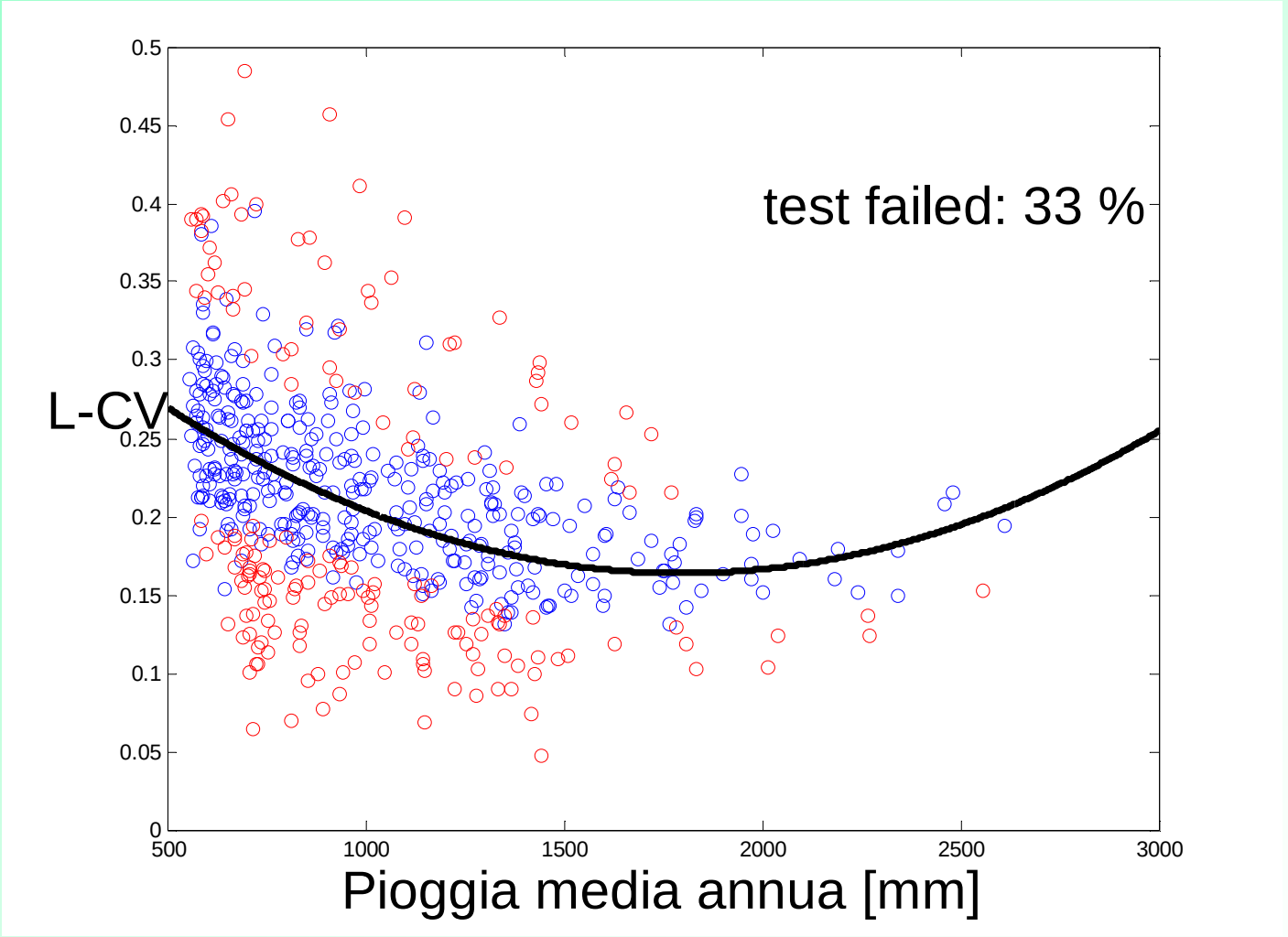
5. **Region of influence (regioni grandi**: 200 siti) ⇒ non spiegato il 31% degli L-CV campionari

Scelta del metodo di regionalizzazione



6. **Variabilità continua (lineare)** $\Rightarrow$ non spiegato il 32% degli L-CV campionari

Scelta del metodo di regionalizzazione



7. **Variabilità continua (quadratica)** ⇒ non spiegato il 32% degli L-CV campionari

Scelta del metodo di regionalizzazione

Si hanno **variazioni poco significative al variare del metodo**, e si rimane su valori non accettabili (% failed dovrebbe essere 5%!!)

metodo	**% test fallito**
regione unica	36
regioni fisse	32
ROI reg. piccole	39
ROI reg. medie	31
ROI reg. grandi	31
variab. continua (lin.)	32
variab. continua (quadr.)	33

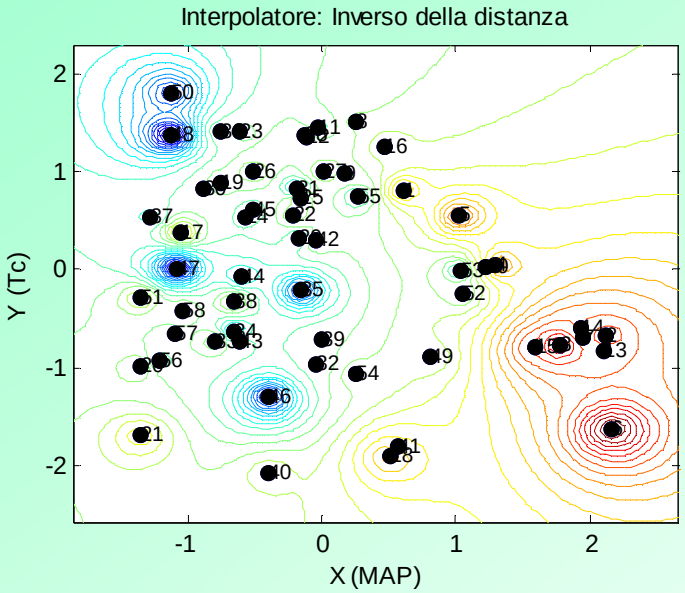
Cosa può produrre il salto di qualità?

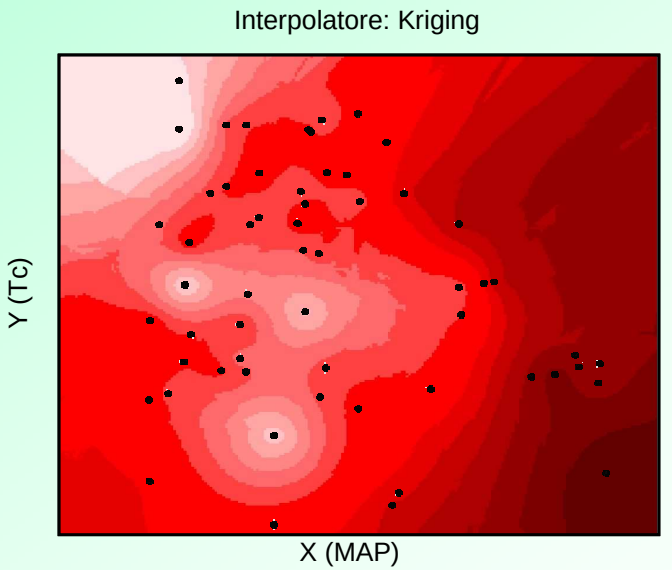
- investire nella ricerca di descrittori geomorfoclimatici e sul loro legame con i descrittoti statistici (ad esempio gli L-momenti)
- tecniche di interpolazione spaziale

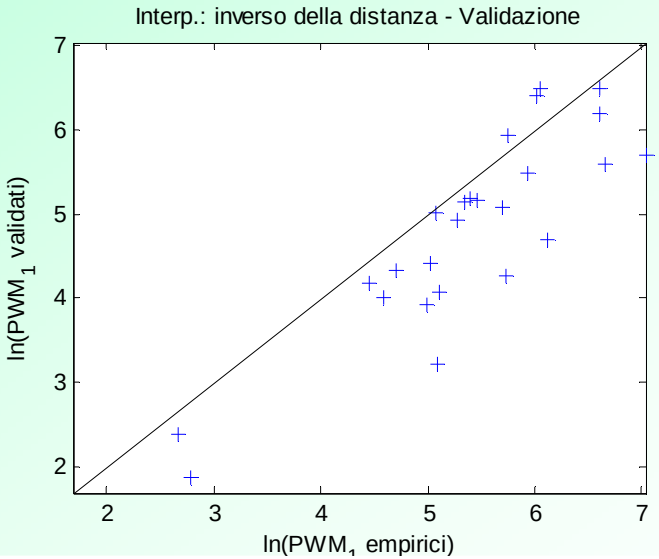
Impiego di tecniche di interpolazione spaziale

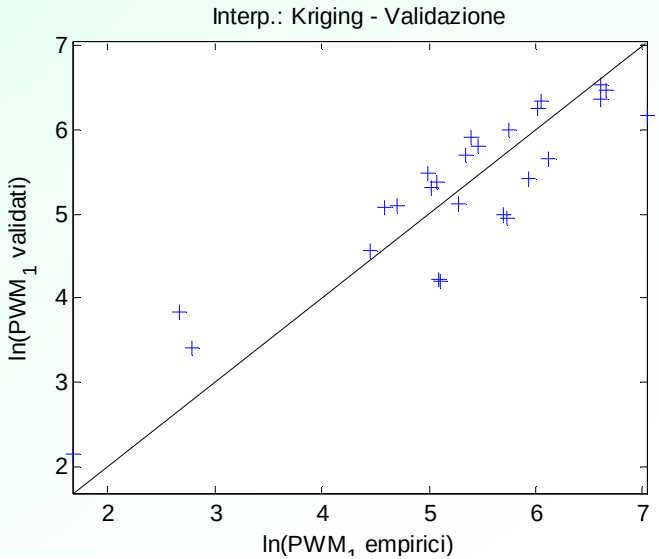
Stima indiretta della piena indice:

Cross-validazione per due diversi interpolatori spaziali.









Ci sono tematiche ancora da esplorare?

Approccio classico

[es. Hosking and Wallis, 1997]

A. Identificazione delle regioni omogenee

B. Scelta della distribuzione di frequenza

C. Stima dei parametri regionali

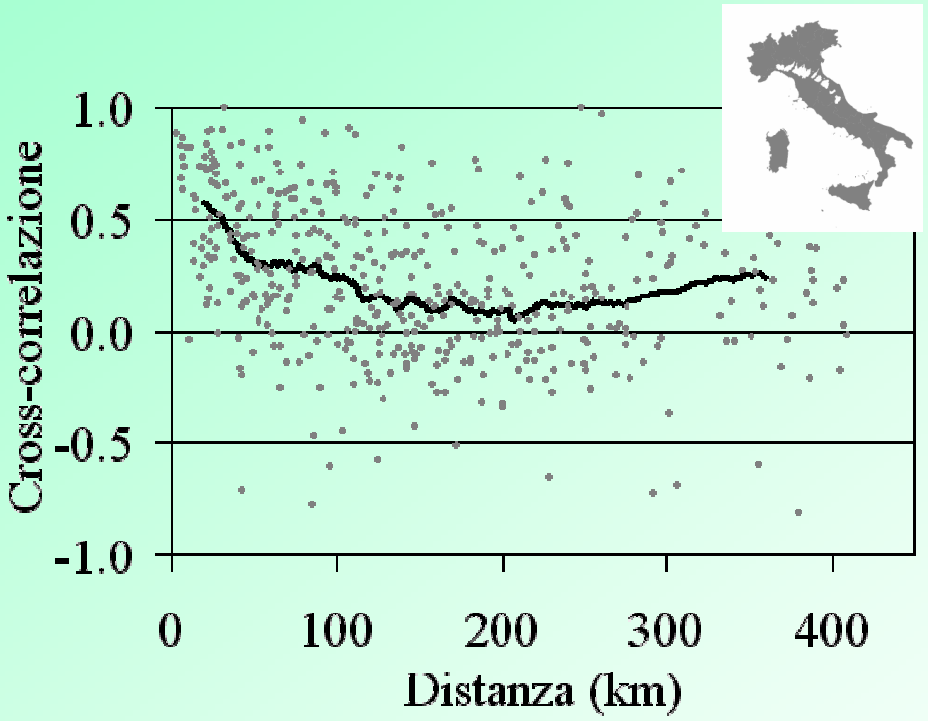
D. Verifica del modello regionale (validazione del modello e incertezza delle stime)

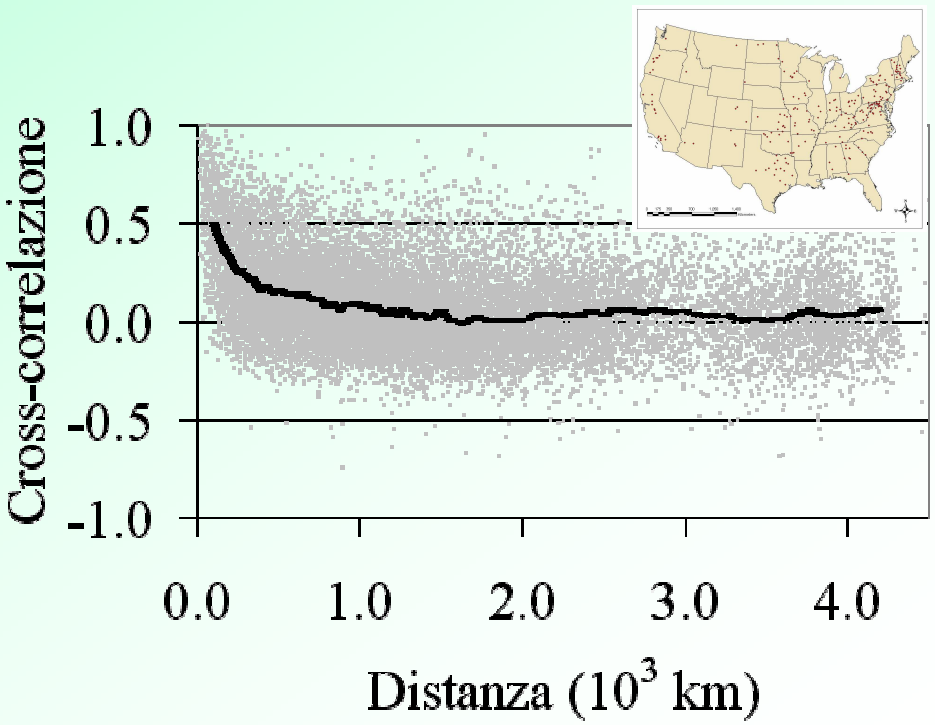
- Metodo di regionalizzazione
- Descrittori geomorfoclimatici
- Test di omogeneità
- Scelta della distribuzione
- Metodo di stima dei parametri regionali
- Stima della grandezza indice in siti non strumentati
- **Effetti della cross-correlazione**

Ci sono tematiche ancora da esplorare?

La cross-correlazione tra le serie annuali degli estremi idrologici è spesso non trascurabile

Massimi annuali di portata al colmo di piena





Considerazioni conclusive

- Ambiti di indagine più promettenti:
  - stima della grandezza indice per siti privi di osservazioni
  - ricerca di descrittori geomorfoclimatici sui quali basare il vicinaggio idrologico
- In questi ambiti le schematizzazioni concettuali (ad es. derivazione analitica della distribuzione di probabilità) possono produrre risultati significativi in termini di:
  - modelli di stima della grandezza indice
  - identificazione dei descrittori più adatti
  - messa a punto di modelli interpretativi con riferimento alle relazioni tra le (L)statistiche ed i descrittori, ma anche a modelli di cross-correlazione

Ci sono tematiche ancora da esplorare?

Probabilmente manca ancora una comprensione a tutto campo degli effetti della cross-correlazione sulle tecniche di regionalizzazione.

Il problema è generalmente “aggirato” in base ai risultati ottenuti in alcuni classici studi:

> *la cross-correlazione non produce distorsione degli stimatori regionali, ne altera semplicemente la varianza di stima* [v. ad esempio Hosking and Wallis, 1988]

-Varianza di stima $\Rightarrow$ Incertezza delle stime regionali

- Altri effetti?

Cross-correlazione – Test di omogeneità

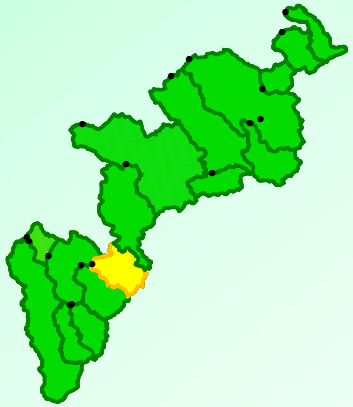
I test vengono normalmente applicati in maniera acritica, non considerando gli effetti della cross-correlazione.

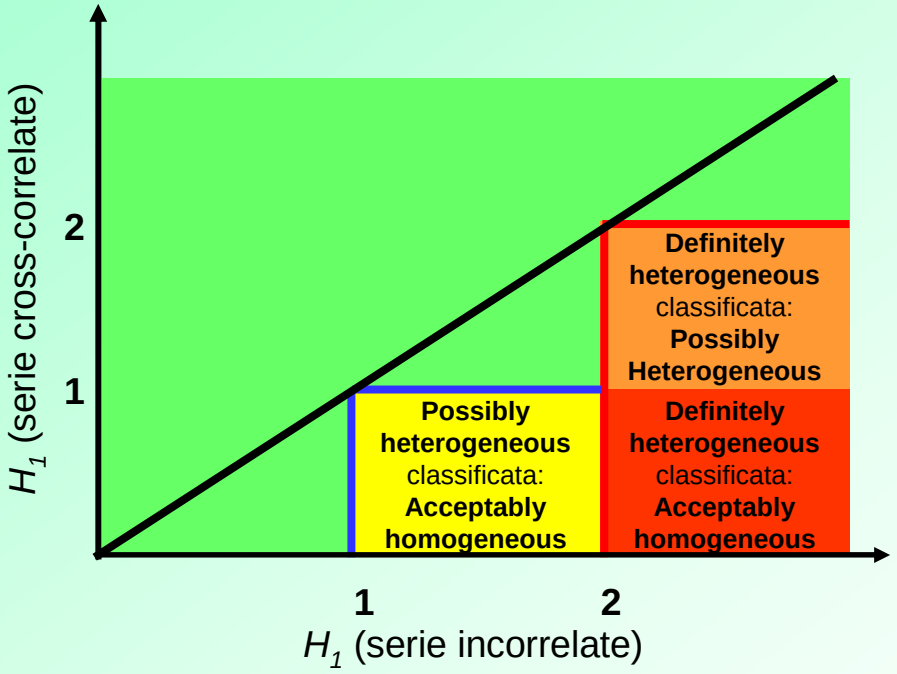
Esempio:
Misura di eterogeneità H_1, basata sulla variabilità campionaria degli L-Cv [Hosking and Wallis, WRR, 1993]:

"Acceptably homogeneous" se $H_1 < 1$
"Possibly heterogeneous" se $1 \leq H_1 < 2$
"Definitely heterogeneous" se $H_1 \geq 2$

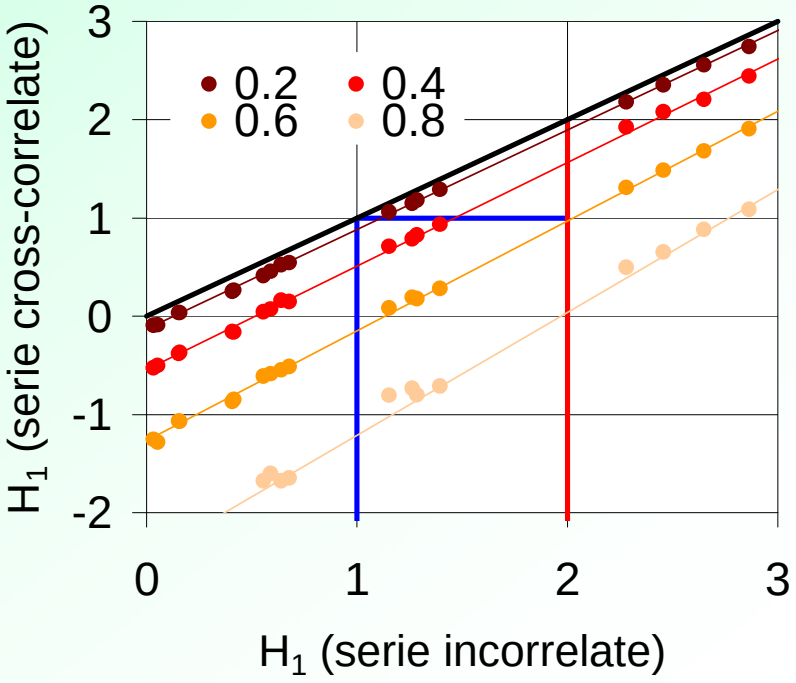
Cross-correlazione – Test di omogeneità

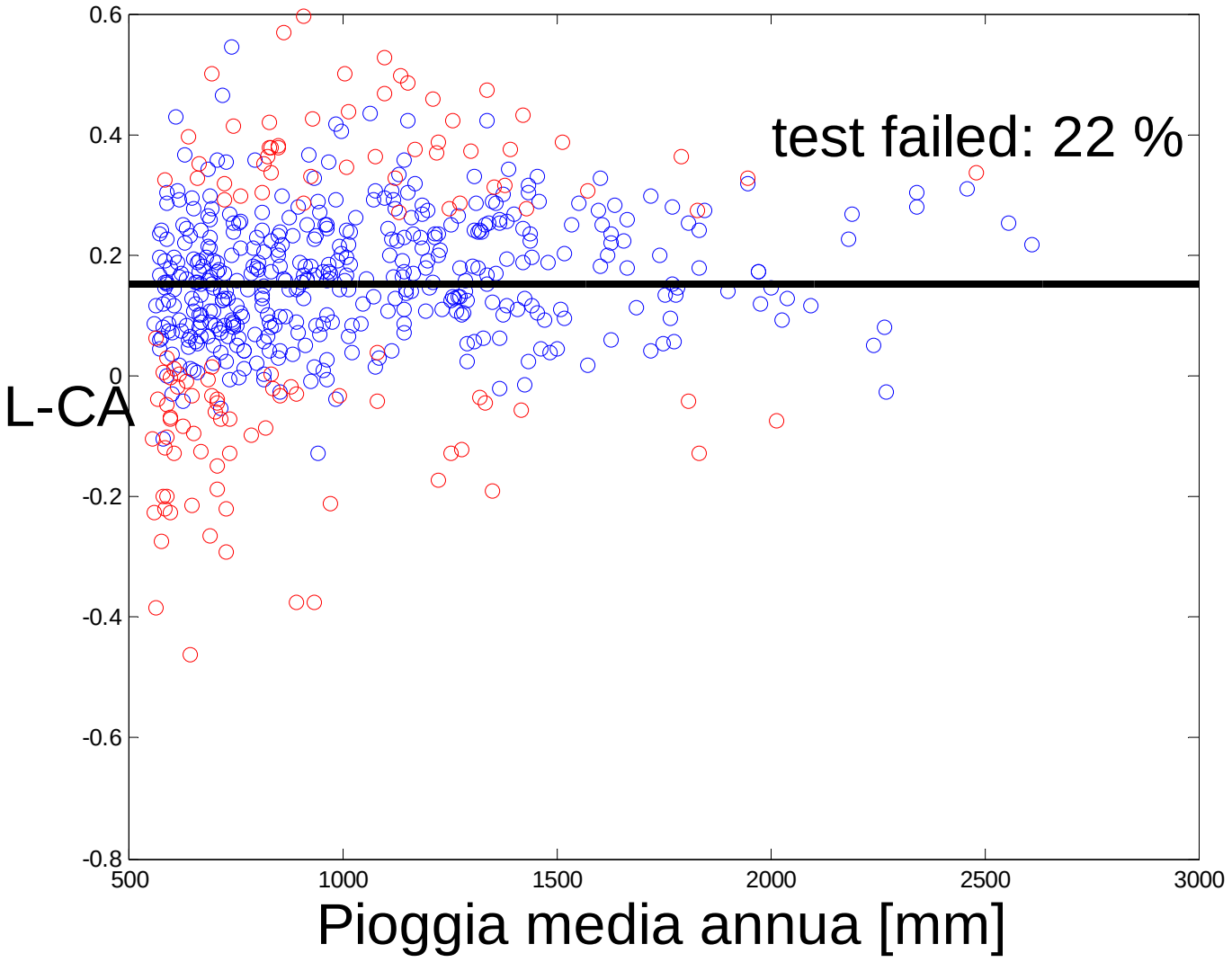
Simulazioni Monte Carlo:
Regioni sintetiche (*N* serie contemporanee di lunghezza *n* anni, EV1con *CV*=0.4) con diverso grado di cross-correlazione. Un solo sito è caratterizzato da una distribuzione EV1(*CV**=0.1,0.2,...0.7).

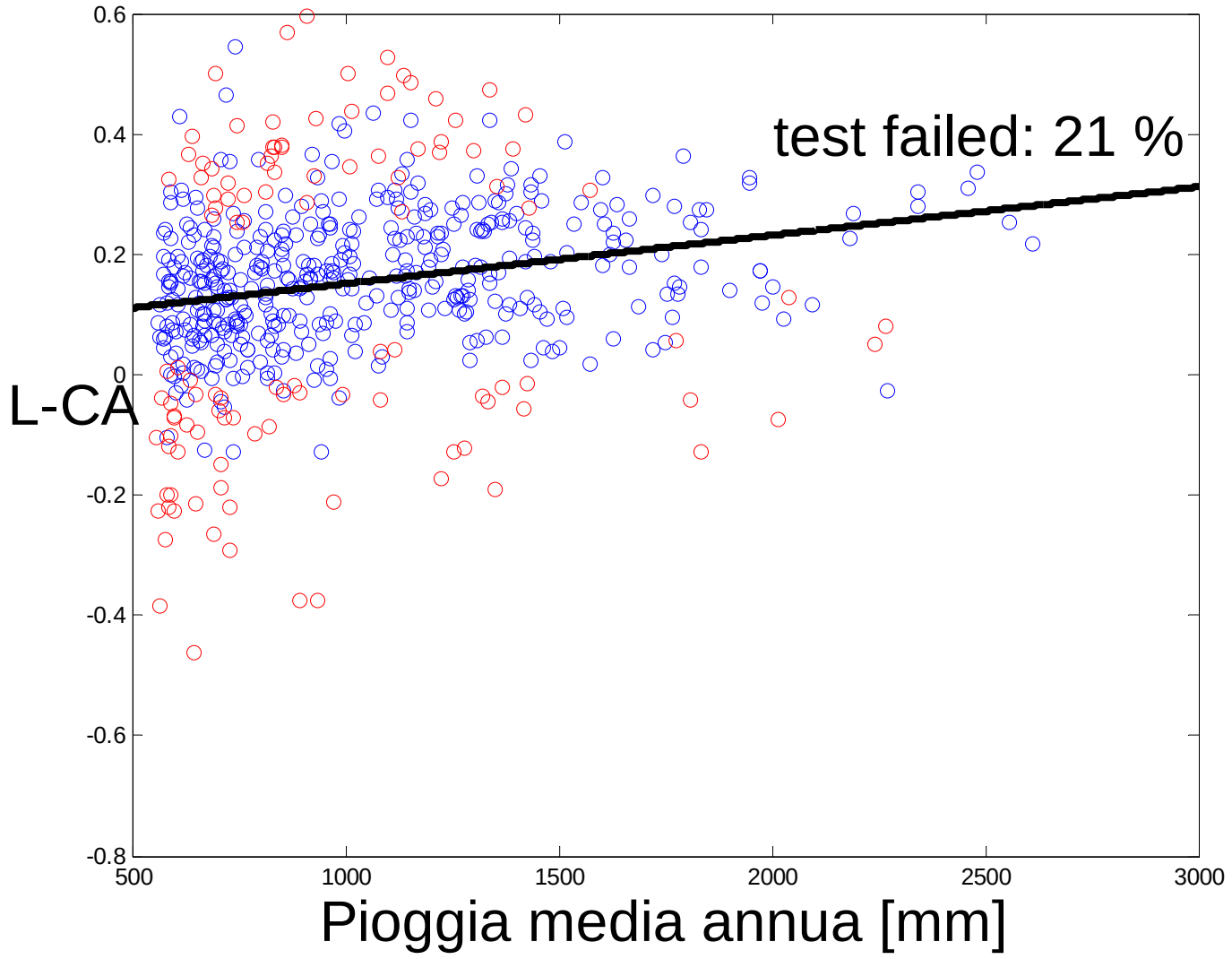




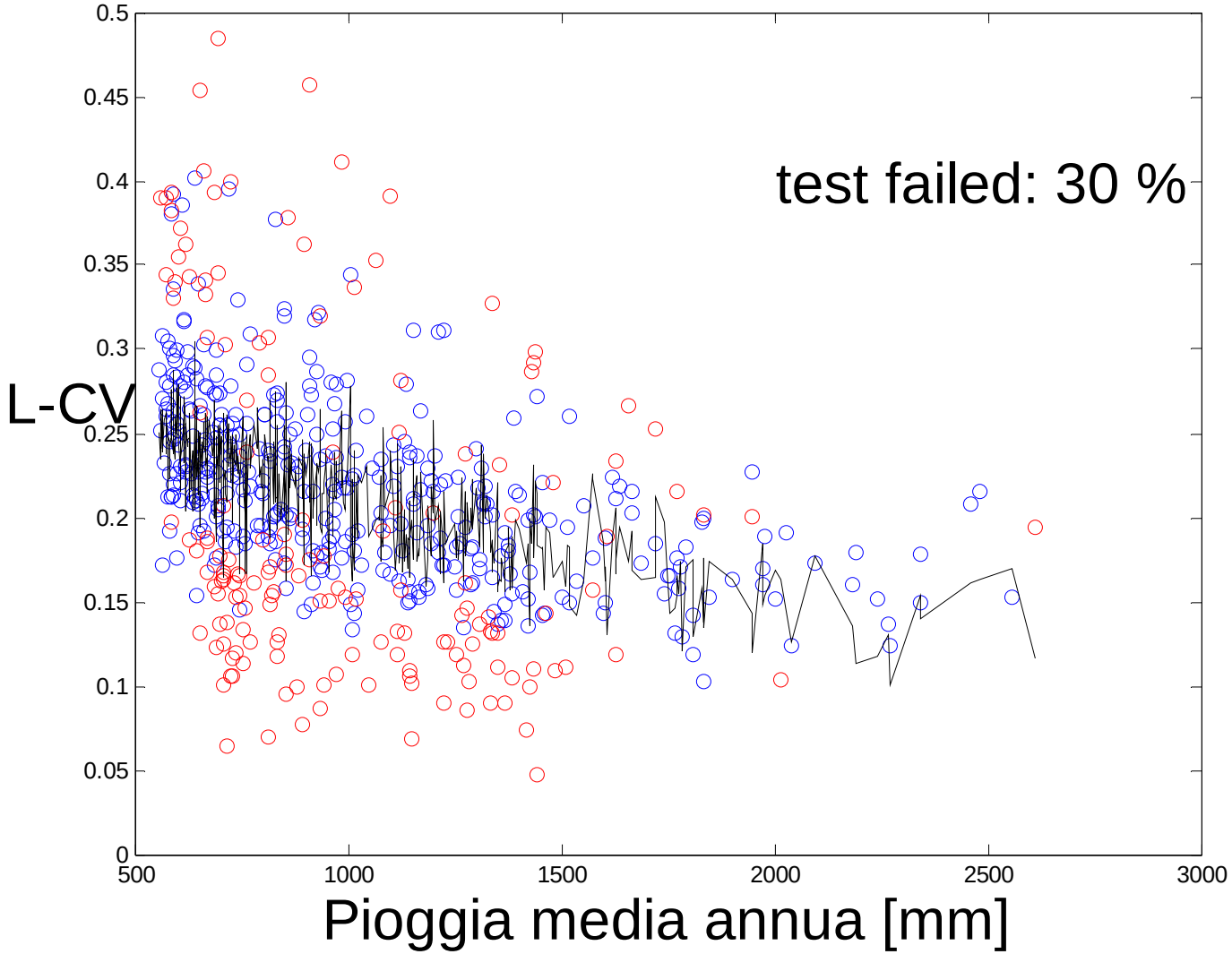
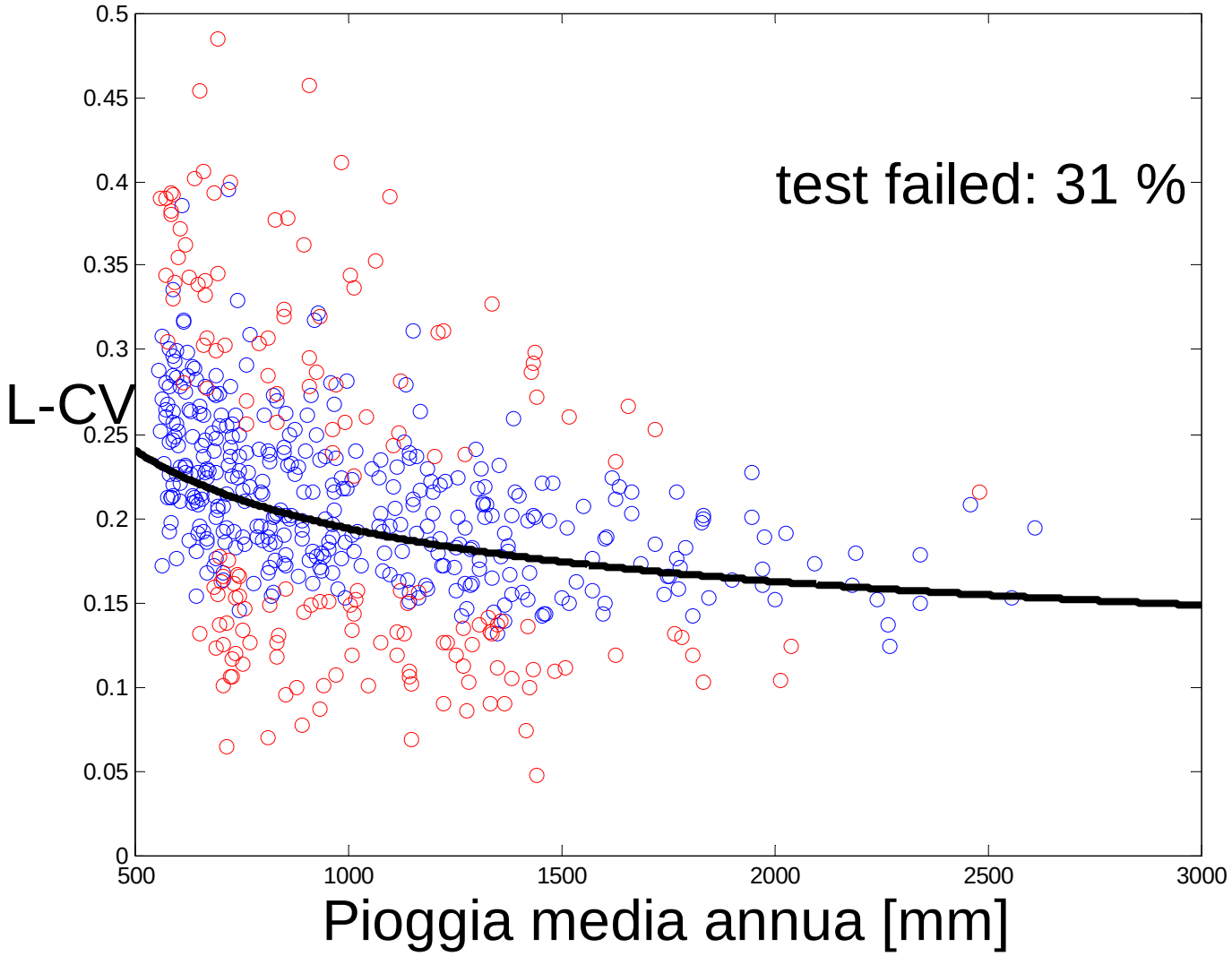
N =30 siti (*n* =10,25,50 anni)




test failed: 22 %
L-CA
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
500
1000
1500
2000
2500
3000
Pioggia media annua [mm]


test failed: 21 %
L-CA
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
500
1000
1500
2000
2500
3000
Pioggia media annua [mm]

L-CV funzione lineare di PTA, log(A) e Hm



Modello compound poisson con $\lambda=PTA/\alpha$, $\alpha=const=50$ mm.
Si ottiene L-CV=log(2)/(0.5772+log(PTA/50))