

ALMA MATER STUDIORUM - UNIVERSITÀ DI BOLOGNA

SCUOLA DI INGEGNERIA E ARCHITETTURA

DICAM

*CORSO DI LAUREA MAGISTRALE IN
INGEGNERIA PER L'AMBIENTE E IL TERRITORIO*

TESI DI LAUREA MAGISTRALE

in

Modellistica Idrologica M

**Tecniche regionali di stima della portata di progetto: impatto della
correlazione spaziale tra le serie di osservazioni**

Candidato

Armando Maria Famulari

Relatore

Prof. Ing. Attilio Castellarin

Correlatori

Dott. Ing. Alessio Pugliese
Dr. Stacey A. Archfield, Ph.D.

Anno Accademico 2012-2013

III Sessione

«Hai appreso anche tu quel segreto del fiume:
Che il tempo non esiste?»

- Herman Hesse, *Siddharta*

Indice

Introduzione	1
Capitolo 1 – Cenni preliminari sull'Analisi regionale delle piene.....	3
1.1 Dall'analisi locale a quella regionale.....	3
1.2 Articolazione di un'analisi regionale di frequenza delle piene	4
1.3 Valutazione regionale delle portate di piena in Italia: progetto VAPI	7
Capitolo 2 – Cenni di Geostatistica.....	8
2.1 Definizioni	8
2.2 Dall'approccio deterministico a quello probabilistico	8
2.3 Modelli di base	10
2.4 Variogrammi sperimentali e teorici	14
2.5 Stima locale di una variabile.....	19
2.6 Ordinary Kriging	22
2.7 Topological Kriging.....	24
Capitolo 3 – Metodo di regressione ai Minimi Quadrati Generalizzati	27
Capitolo 4 – Area di studio	31
4.1 Dati geomorfologici	33
4.2 Dati idrometrici	34
Capitolo 5 – Applicazione di GLS e Top-kriging per la stima della piena di progetto in bacini non strumentati appartenenti all'area di studio	37
5.1 Applicazione del metodo GLS	37
5.2 Applicazione Top-kriging	38
5.3 Procedimento seguito per l'indagine.....	39
5.4 Risultati	45
5.5 Discussione.....	52
Conclusioni.....	54

Introduzione

Il principale argomento della presente trattazione riguarda la stima della portata di progetto in bacini idrografici in cui non siano disponibili osservazioni idrometriche. Questo problema è oggetto del progetto di ricerca internazionale denominato PUB (*Prediction in Ungauged Basin*) promosso da parte della International Association of Hydrological Sciences e svoltosi durante il decennio appena conclusosi (2003-2012).

Nell'ambito generale appena richiamato, il presente elaborato si concentra in particolare su due tipi di procedure per la stima di indici idrometrici (ad. es. portata di piena di progetto, ovvero la portata al colmo di piena associata ad una ben precisa frequenza di accadimento) in bacini non strumentati: la Regressione ai Minimi Quadrati Generalizzati (GLS) e il Topological Kriging (in seguito abbreviato con Top-kriging o TK). La prima prende in considerazione una serie di fattori geomorfologici e climatici caratterizzanti i bacini di studio e, sulla base di alcune assunzioni di tipo statistico, ne estrae i pesi per un modello di regressione lineare delle osservazioni (Griffis, Stedinger, 2007). Il secondo è un metodo di tipo geostatistico che tiene conto del supporto areale della variabile regionalizzata (c.d. l'area del bacino imbrifero), della collocazione nello spazio di tale variabile e della struttura annidata dei bacini oggetto di studio (Skøien et al., 2006). Le tecniche citate sono state confrontate tra loro per la stima delle portate di progetto in bacini non strumentati e, date le caratteristiche di complementarità degli approcci, si è cercato di capire se fosse possibile accoppiarle per migliorare l'accuratezza di stima della grandezza di interesse (piena di progetto) in bacini non strumentati. Le prestazioni delle due tecniche, applicate separatamente l'una dall'altra o accoppiate sono state quantificate in base ad alcuni indici statistici, come l'indice di efficienza di Nash-Sutcliffe (Nash, Sutcliffe, 1970) e l'errore medio relativo. Il lavoro di Tesi è stato effettuato su un'area di studio comprendente 61 bacini degli Stati Uniti sud-orientali, localizzati fra gli stati dell'Alabama, della Florida e della Georgia e i cui dati sono stati messi a disposizione dall'USGS (United States Geological Survey).

Il lavoro di Tesi è suddiviso in cinque capitoli.

Nel *primo capitolo* si prendono in considerazione alcuni cenni preliminari sul tema dell'Analisi Regionale di frequenza delle piene.

Il *secondo capitolo* è dedicato a una sintetica panoramica sulla Geostatistica, i cui concetti saranno richiamati nelle parti successive per la trattazione degli argomenti principali, non ultimo il Top-kriging.

Il *terzo capitolo* accenna brevemente ai metodi regressivi di stima ai Minimi Quadrati Generalizzati.

Il *quarto capitolo* descrive sinteticamente l'area di studio, sui cui dati è stata applicata la procedura descritta in questo elaborato.

Il *quinto e ultimo capitolo* illustra in dettaglio l'indagine che è stata effettuata sui residui dei quantili, ne illustra i risultati e fornisce alcune note conclusive in merito.

Capitolo 1 – Cenni preliminari sull'Analisi regionale delle piene

1.1 Dall'analisi locale a quella regionale

Nei problemi di difesa idraulica del territorio si pone spesso la necessità di valutare il deflusso relativo ad una determinata sezione fluviale ed associato ad un assegnato livello di rischio. Tale problema può essere risolto tramite metodi di analisi di frequenza *locale* delle piene, qualora si disponga di un campione sufficientemente numeroso di osservazioni relative alla sezione d'interesse; questi metodi consistono principalmente nell'identificazione di una distribuzione di probabilità che meglio rappresenti le osservazioni disponibili, considerandole come variabili aleatorie. A esse si attribuisce una certa probabilità di verificarsi, che viene spesso quantificata, in ambito idrologico, con il concetto di *tempo di ritorno*.

Per tempo di ritorno si intende la frequenza con cui una certa portata x viene mediamente uguagliata o superata e con cui i manufatti dimensionati in base ad essa vengono a trovarsi in condizioni di crisi (v. Brath, 1995) ed è espresso tramite la relazione:

$$T_R(x) = \frac{1}{1 - F_X(x)} = \frac{1}{F_X'(x)} \quad (1.1)$$

In cui $F_X(x)$ è la *frequenza di non superamento* della variabile, ovvero la probabilità che la variabile X non superi il valore x , mentre $F_X'(x)$, per contro, è la *frequenza di superamento*. Il *quantile*, di conseguenza, è la portata associata a un certo tempo di ritorno $x(T_R)$ ed è ottenuto generalmente mediante un'analisi di frequenza dei valori al colmo di piena annuali. L'obiettivo dell'analisi di frequenza delle piene, dunque, è quello di ottenere una stima ragionevolmente affidabile del quantile $x(T_R)$ per tempi di ritorno di interesse tecnico, che possono variare da qualche decina a qualche migliaio di anni, a seconda delle strutture a cui sono associati e al margine di sicurezza richiesto.

La stima della portata massima x ad assegnato tempo di ritorno T_R effettuata su base *puntuale*, ossia utilizzando i dati di una singola stazione, risulta tuttavia fortemente incerta, se non addirittura impossibile se:

1. si prendono in considerazione tempi di ritorno molto più elevati della dimensione del campione utilizzato: una stima di questo genere, infatti, si può ritenere sufficientemente attendibile finché il tempo di ritorno rimane inferiore alla dimensione del campione disponibile;
2. si prendono in considerazione sezioni non strumentate, e quindi prive di osservazioni.

Il ricorso a tecniche di *analisi regionale di frequenza* permette di risolvere questi problemi e di ottenere una stima accurata di una variabile idrologica caratterizzata da un elevato tempo di ritorno, poiché sono basate sul ricorso a realizzazioni di variabili casuali *simili* a quella da cui provengono le osservazioni della stazione considerata. Ciò consente di aumentare, in un certo senso, il numero di dati a propria disposizione, e di allargare la stima anche a sezioni prive di misurazioni. La letteratura scientifica riporta svariati esempi di applicazione di stime regionali di quantili al colmo di piena. La maggior parte di queste tecniche si basano sulla preventiva individuazione di *regioni omogenee* per le quali si ipotizza che la distribuzione in frequenza dei colmi di piena possa ritenersi la stessa sulla medesima regione. Solo dopo aver identificato tali regioni si può procedere alla stima dell'informazione idrometrica in sezioni non strumentate avendo a disposizione informazioni raccolte in quelle sezioni dotate di osservazioni.

1.2 Articolazione di un'analisi regionale di frequenza delle piene

Perché si possano individuare e raggruppare zone idrologicamente omogenee nell'area di indagine, è necessario che queste abbiano la medesima distribuzione di frequenza dei picchi di piena massimi annuali, a meno di un fattore di scala. Le caratteristiche climatiche e geomorfologiche dei bacini sono discriminanti fondamentali sulla scelta della migliore distribuzione di probabilità da adottare per una determinata regione omogenea.

La delimitazione può essere condotta da un punto vista soggettivo o oggettivo. Fra i *criteri soggettivi* si menziona ad esempio quello geografico-amministrativo, mentre fra i *criteri oggettivi* rientrano quelli basati sulla *cluster analysis*: si associa a ciascun sito una serie di parametri descrittivi delle caratteristiche geomorfoclimatiche, per poi accoppiare o separare i bacini in base alla *similitudine* di tali parametri. Poiché un

criterio non esclude l'altro, l'associazione di un bacino a un'area omogenea piuttosto che ad un'altra è sottoposta a test statistici. La letteratura scientifica nel corso degli anni ha classificato l'individuazione delle regioni omogenee secondo tre tipologie: geograficamente contigue, geograficamente non contigue, o per vicinanza idrologica (v. fig. 1.1). Bacini privi di osservazioni vengono assegnati all'una o all'altra regione omogenea in base ad uno di questi criteri.

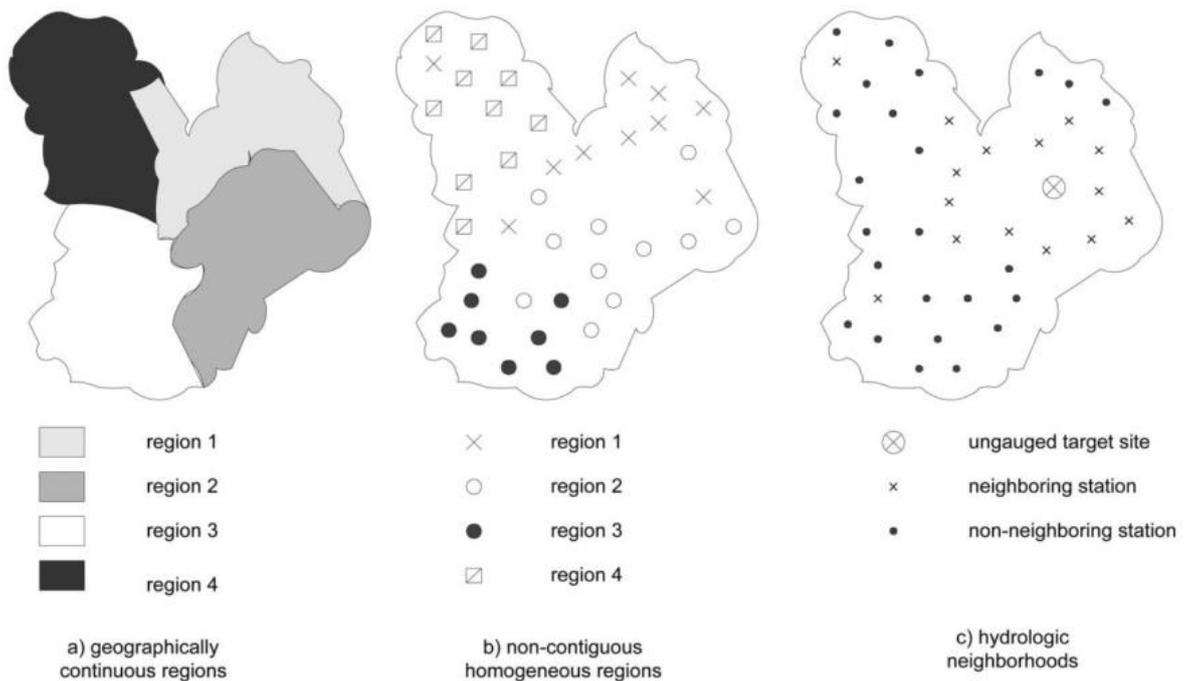


Figura 1.1. Evoluzione temporale del concetto di regione (o raggruppamento di bacini) idrologicamente omogenea (Ouarda et al., 2001)

Una volta scelta la suddivisione più appropriata, si passa all'identificazione di una relazione in grado di esprimere le portate al colmo relative ai corsi d'acqua ricadenti in ciascuna regione omogenea. Questa fase può essere affrontata facendo uso di diversi metodi:

1. Una prima metodologia prevede l'applicazione dei metodi di analisi puntuale di frequenza delle piene alle osservazioni di massimi annuali relative a ciascuna stazione ricadente nella regione. Questi metodi stimano il legame fra portata al colmo e tempo di ritorno, $x = x(T_R)$, con tecniche di inferenza statistica che mettono in relazione le stime di portata con un opportuno insieme di

caratteristiche riguardanti il bacino. Comunemente, con questo approccio si ricorre a metodi di tipo multiregressivo;

2. Un secondo approccio si basa sull'identificazione di una distribuzione di probabilità nella quale inquadrare l'informazione idrologica. Attraverso tests di adattamento, si calcolano alcune statistiche di sintesi dai dati e si verifica la consistenza dei valori con quelli attesi. Questo tipo di approccio, già utilizzato per l'analisi locale, trova validità anche in quella regionale, seppur con alcuni accorgimenti. A valle della scelta del tipo di distribuzione, si opera la stima dei parametri a essa relativi.
3. Un terzo metodo che può essere impiegato al problema della regionalizzazione è il cosiddetto *metodo della portata indice*, che si basa sull'identificazione di un fattore di scala fra le diverse distribuzioni di probabilità che regolano le portate al colmo.

1.2.1 Analisi regionale basata sulla portata indice

L'analisi regionale consente di correlare spazialmente la variabile idrologica, slegandola così dal contesto temporale. La metodologia più diffusamente utilizzata per eseguire questo tipo di analisi è quella detta della *portata indice* (Dalrymple, 1960). Per operare la stima del quantile in base al metodo della portata indice, è necessario riferirsi a una regione *idrologicamente omogenea* e definire un'opportuna funzione di probabilità di non superamento $F_X(x)$ di una variabile casuale, strutturata in questo modo:

$$X' = \frac{X}{Q_{ind}} \quad (1.2)$$

In cui X è il massimo valore annuo della portata al colmo, adimensionalizzato rispetto a un'apposita grandezza Q_{ind} , detta *portata indice*. Tale grandezza è locale, diversa per ciascuna sezione (o sito) in esame. In idrologia, la funzione $F_X(x)$, la sua inversa $x'(F)$ e l'equivalente relazione $x(T_R)$ sono indicate con il termine *curva di crescita*.

Nella pratica, il metodo viene applicato adimensionalizzando le serie storiche dei massimi annuali delle portate al colmo relative a ciascuna regione, dividendo ciascun elemento per la media della relativa serie (che rappresenta la stima della portata indice Q_{ind}). Accorpendo successivamente le serie adimensionalizzate in un unico

campione, se ne identifica la distribuzione e si ricercano una serie di equazioni regionali di tipo multiregressivo che mettano in relazione la portata indice con alcune caratteristiche geomorfologiche e climatiche del bacino (v. Brath, 1995).

In letteratura vengono proposti diversi approcci per individuare le relazioni fra la portata indice e i parametri geomorfoclimatici del bacino idrografico.

1.3 Valutazione regionale delle portate di piena in Italia: progetto VAPI

La grande varietà di tecniche di regionalizzazione disponibili, unitamente ai molteplici risultati che è possibile ottenere a seconda del metodo e dell'approccio adottati, ha spinto negli ultimi decenni verso l'esigenza di definire procedure standard di riferimento a livello nazionale. È in quest'ottica che il Gruppo Nazionale per la Difesa dalle Catastrofi Idrogeologiche (GNDCI) del CNR ha attivato il Progetto speciale Valutazione delle Piene (VAPI), il cui scopo è di definire una metodologia statistica comune, attraverso la quale sia possibile la valutazione e la quantificazione del rischio idrometeorologico.

Attraverso il lavoro congiunto di organi pubblici e privati, il progetto prevede un'analisi di tipo regionale basata sul metodo della *portata indice* (v. paragrafo 1.2.1) e l'utilizzo integrato di tutte le informazioni di natura idrologica disponibili. La procedura adotta un particolare modello probabilistico a doppia componente denominato TCEV (*Two Component Extreme Values Distribution*). Per maggiori dettagli al riguardo si rimanda a Franchini e Galeati (1996).

I principali vantaggi dell'adozione di questo metodo sono una migliore comprensione dei risultati relativi a regioni diverse del territorio nazionale e la possibilità di confrontare agevolmente il rischio idrogeologico di aree differenti.

Capitolo 2 – Cenni di Geostatistica

2.1 Definizioni

La Geostatistica studia i fenomeni naturali che si sviluppano su base spaziale a partire dalle informazioni derivanti da un loro campionamento. In particolare, studia la variabilità spaziale dei parametri che descrivono tali fenomeni, estraendone le regole in un quadro modellistico di riferimento e usandole per effettuare le operazioni volte a dare soluzione a specifiche problematiche riguardanti la caratterizzazione e la stima dei fenomeni stessi (v. Raspa e Bruno, 1994a).

Se si considera un fenomeno spaziale caratterizzato da una variabile z , localizzato nel generico punto x di coordinate (x_u, x_v) , la funzione $z(x)$ lega la grandezza indagata ai punti dello spazio. È possibile pertanto dare le seguenti definizioni:

- La **variabile regionalizzata** (VR) è la funzione matematica $z(x)$ che descrive la variabilità nello spazio della grandezza d'interesse e che si presenta strutturata spazialmente;
- Il **campo** è il dominio di $z(x)$, all'interno del quale la VR può assumere determinati valori e in cui se ne studia la variabilità;
- Il **supporto** è l'entità geometrica su cui viene definita o misurata la variabile z e di cui sono note dimensioni e forma. Se le dimensioni sono particolarmente piccole, il supporto può anche considerarsi puntuale. Inoltre, al variare della forma e dimensione del supporto, si ottiene una variabile regionalizzata diversa dalla prima, anche se analoga.

2.2 Dall'approccio deterministico a quello probabilistico

Osservando come varia una grandezza nello spazio, si può capire se la sua variabilità incoraggi o meno verso uno studio matematico diretto, ovvero se essa sia strutturata o no. In base a questa distinzione, la variabile regionalizzata viene studiata con un approccio *deterministico* (in ogni punto dello spazio la grandezza assume un certo valore) o *probabilistico* (in vicinanza dei punti considerati, vi sarà una certa probabilità che la grandezza assuma un certo valore). Quest'ultimo

approccio tiene in considerazione entrambi gli aspetti della variabilità della VR ed è quello di maggior interesse ai fini del nostro studio.

Da un punto di vista probabilistico, dunque, si passa da una variabile regionalizzata VR in un punto x_0 del campo S ad una *Variabile Aleatoria* (VA) continua $Z(x_0)$. Essa assume valori numerici appartenenti a un certo intervallo, definito da una legge di densità di probabilità $f_0(Z)$. In queste condizioni:

- In x_0 , il valore puntuale $z(x)$ costituisce una *realizzazione* della variabile aleatoria $Z(x_0)$;
- In ogni altro punto x di S si può definire una variabile aleatoria $Z(x)$;
- L'insieme di tutte le variabili aleatorie in S è detto *Funzione Aleatoria* (FA).

La Funzione Aleatoria è caratterizzata dall'insieme di tutte le funzioni di distribuzione multivariabili che si possono definire in S , per qualsiasi numero intero k di punti e per qualsiasi loro configurazione:

$$F_{Z_1, z_2, \dots, z_k}(x_1, x_2, \dots, x_k) = \text{prob}\{Z(x_1) < z_1, Z(x_2) < z_2, \dots, Z(x_k) < z_k\} \quad (2.1)$$

L'insieme di queste funzioni, data la natura spaziale del fenomeno da modellizzare, costituisce la legge spaziale della FA $Z(x)$. Se le variabili $Z(x_1), Z(x_2), \dots, Z(x_k)$ sono indipendenti a due a due, la legge spaziale di $Z(x)$ è costituita dall'insieme di tutte le funzioni di distribuzione monovariabile:

$$F_{Z(x)} = \text{prob}\{Z(x) < z\}, \quad \forall x \in S \quad (2.2)$$

A questo proposito si ricorda che l'interpretazione probabilistica dei fenomeni spaziali è soltanto un intermediario di calcolo e non una rappresentazione della realtà fisica. Il ricorso a un modello cosiddetto topo-probabilista, infatti, ci consente di adottare strumenti di calcolo in problemi per i quali non bastano soluzioni basate su modelli empirici o deterministici.

Avendo a disposizione un modello, se ne stimeranno i parametri a partire dai dati di campionatura del fenomeno mentre, per quanto riguarda le stime (o *predizioni*), le più diffuse sono quelle lineari.

2.3 Modelli di base

La grande varietà di modelli probabilistici ci pone di fronte alla questione relativa al modello da scegliere. Poiché la maggior parte dei problemi è di natura *locale*, cioè interessa una piccola porzione del campo comprendente il supporto, si possono fare alcune considerazioni:

- Definiamo tale porzione del campo con il termine *vicinaggio* e la utilizziamo per le nostre stime;
- Tendendo a fare stime su tutto il campo, senza limitarsi a un solo problema locale, tale vicinaggio rimane pressoché identico per forma e dimensione e con esso si percorre tutto il campo: viene quindi definito *vicinaggio mobile*.

Poiché il ricorso a un vicinaggio mobile è compatibile con la natura locale dei nostri problemi, è lecito poterlo adottare in sostituzione di una molteplicità di realizzazioni della FA, in un modello che non interessi tutto il campo. Per far sì che ciò si verifichi, e poiché si dispone di un'unica realizzazione del modello, è necessario sollecitare condizioni di *stazionarietà* per la FA.

2.3.1. Stazionarietà *strictu sensu* dei modelli

Una FA si definisce stazionaria in presenza di stazionarietà *strictu sensu*, ossia della proprietà di non variare la propria legge di distribuzione. In altri termini, per qualunque numero intero $k > 0$ ed un insieme di k punti nel campo S localizzati in $\{x_1, x_2, \dots, x_k\}$, la variabile aleatoria $\{Z(x_1), Z(x_2), \dots, Z(x_k)\}$ e $\{Z(x_1 + h), Z(x_2 + h), \dots, Z(x_k + h)\}$ con h vettore qualsiasi, mantiene la stessa legge di distribuzione k -variabile. La proprietà della stazionarietà è riferita ai primi due momenti della FA, sui quali si effettuano la maggior parte delle operazioni geostatistiche, e al variogramma.

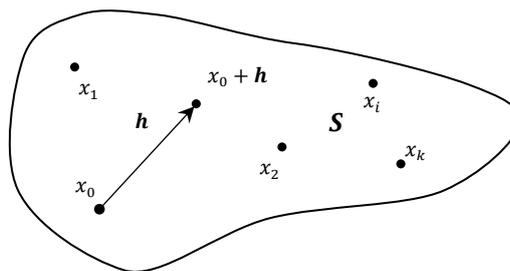


Figura 2.1. Schematizzazione del campo S e dei punti x_i in cui localizzare la VA

Momento primo

Sia S il campo di indagine, e si definisca in ogni suo punto $x \in S$ una VA $Z(x)$. Il suo momento primo, se esiste, è generalmente funzione di x :

$$E[Z(x)] = m(x) \quad (2.3)$$

Dove, lo si ricorda, per valore atteso di una variabile casuale continua Z con funzione di densità di probabilità $f(z)$ si intende per definizione:

$$E[Z] = \int_{-\infty}^{+\infty} zf(z)dz \quad (2.4)$$

Momento secondo

Si considerino i punti x_1 e x_2 del campo S e le variabili aleatorie a essi associate $Z(x_1)$ e $Z(x_2)$. La loro *covarianza* è data da:

$$\begin{aligned} \text{Cov}(x_1, x_2) &= E\{[Z(x_1) - m(x_1)] \cdot [Z(x_2) - m(x_2)]\} \\ &= E[Z(x_1) \cdot Z(x_2)] - m(x_1) \cdot m(x_2) \end{aligned} \quad (2.5)$$

Se x_1 e x_2 coincidono, allora si ottiene la *varianza*:

$$\text{Var}(x) = \text{Cov}(x, x) = E[Z^2(x)] - E[Z(x)]^2 \quad (2.6)$$

Variogramma

Il variogramma è la funzione di uso più comune nella Geostatistica ed è impiegato principalmente per caratterizzare la variabilità spaziale di un fenomeno regionalizzato.

Se si considera la FA $Z(x)$ definita in due punti x_0 e $x_0 + h$ del campo S , in cui sia definita la FA e distanti h , la loro differenza è detta *incremento* o *accrescimento*:

$$[Z(x_0) - Z(x_0 + \mathbf{h})]$$

ed è a sua volta una variabile regionalizzata. Si definisce il variogramma come la semi-varianza dell'incremento ed è funzione di h e di x_0 :

$$\gamma(x_0, \mathbf{h}) = \frac{1}{2} \text{Var}[Z(x_0 + \mathbf{h}) - Z(x_0)] \quad (2.7)$$

2.3.2. Modelli stazionari di ordine 2

Un modello di Funzione Aleatoria si definisce *stazionario di ordine 2* in presenza delle seguenti condizioni:

- Il momento primo, ovvero la *media*, esiste ed è invariante rispetto alla posizione x :

$$E[Z(x)] = m \quad (2.8)$$

- Il momento secondo, ovvero la covarianza, esiste e non dipende dalla posizione x dei punti, ma soltanto dalla distanza h :

$$\text{Cov}(x_1, x_2) = C(x_1, x_1 + \mathbf{h}) = C(\mathbf{h}) \quad (2.9)$$

Poiché per punti coincidenti la covarianza decade nella varianza, facendo tendere a 0 la loro distanza, si ottiene:

$$\lim_{|\mathbf{h}| \rightarrow 0} C(x_1, x_1 + \mathbf{h}) \approx C(x_1, x_1) = \text{Var}(x_1) = C(0) \quad (2.10)$$

Quando si verificano queste due condizioni, dunque, la funzione $C(h)$ è detta *funzione covarianza* ed esprime la correlazione fra le variabili $Z(x)$ e $Z(x + h)$ in funzione della sola distanza h fra i rispettivi punti d'appoggio.

Per poter giungere a una formulazione più compatta della funzione covarianza, è utile richiamare la definizione di variogramma per un modello stazionario di ordine 2. Sviluppiamo perciò la varianza secondo le equazioni (2.6) e (2.7) e teniamo presente che il momento primo è invariante per traslazione, $E[Z(x + h)] = E[Z(x)]$:

$$\begin{aligned} \gamma(x, \mathbf{h}) &= \frac{1}{2} \text{Var}[Z(x + \mathbf{h}) - Z(x)] \\ &= \frac{1}{2} E\{[Z(x + \mathbf{h}) - Z(x)]^2\} \end{aligned} \quad (2.11)$$

Sviluppando il quadrato del binomio fra parentesi quadre e ricordando le definizioni di varianza e covarianza, si ottiene:

$$\begin{aligned}\gamma(x, \mathbf{h}) &= \frac{1}{2} E[Z(x + \mathbf{h})^2 + Z(x)^2 - 2 \cdot Z(x)Z(x + \mathbf{h})] \\ &= \frac{1}{2} \{ \text{Var}[Z(x + \mathbf{h})] + \text{Var}[Z(x)] - 2\text{Cov}[Z(x + \mathbf{h}), Z(x)] \}\end{aligned}\quad (2.12)$$

Ricordando che il momento secondo, nel caso di modello stazionario di ordine 2, è invariante per traslazione, $\text{Var}[Z(x + \mathbf{h})] = \text{Var}[Z(x)]$:

$$\gamma(x, \mathbf{h}) = \frac{1}{2} \{ 2\text{Var}[Z(x)] - 2\text{Cov}[Z(x)] \} \quad (2.13)$$

Ricordando le equazioni (2.9) e (2.10), si ottiene la seguente relazione:

$$\gamma(\mathbf{h}) = C(0) - C(\mathbf{h}) \quad (2.14)$$

Questa relazione mostra che la funzione variogramma è legata alla funzione covarianza nel caso di modelli stazionari, e consente di affermare che anch'essa è invariante per traslazione. Si può pertanto dedurre che la correlazione fra le variabili $Z(x)$ e $Z(x + \mathbf{h})$ diminuisce all'aumentare della distanza fra i loro punti d'appoggio, e quindi la funzione $C(\mathbf{h})$ è decrescente all'aumentare di \mathbf{h} fino ad annullarsi: le due variabili a questo punto perdono la loro correlazione e la misura di \mathbf{h} alla quale essi diventano indipendenti è detta *range*. Viceversa, la funzione $\gamma(\mathbf{h})$ è crescente fino a raggiungere un valore di soglia detto *sill*, corrispondente alla varianza $C(0)$.

2.3.3. Modelli non stazionari

Si menzionano ora, per completezza di trattazione, i modelli di FA non stazionari. Le FA *non stazionarie* soddisfano anche una sola delle seguenti condizioni:

- La media $E[Z(x)] = m(x)$ non rimane costante nel campo;
- La funzione covarianza non esiste, o comunque non è invariante per traslazione.

I modelli non stazionari, benché molto ricorrenti nella pratica della Geostatistica, vengono ricondotti a situazioni di stazionarietà lavorando, più comunemente, sulla *deriva* o sull'*incremento*.

Modelli con deriva

Questo tipo di modello si basa su una variazione sistematica, più o meno accentuata, della variabile, che può manifestarsi a scala regionale o inglobare una deriva (*dérive* in francese, o *trend* in inglese). Se un fenomeno manifesta con evidenza, alla corretta scala di osservazione, un progressivo aumento o diminuzione, potrebbe essere indicato simulare questo comportamento attraverso un modello con media variabile, ovvero $E[Z(x)] = m(x)$, applicato a vicinaggi mobili all'interno dell'area di indagine.

Nel modello con deriva, la FA viene considerata costituita da:

- Una componente deterministica, ovvero la *deriva* $m(x)$;
- Una componente aleatoria, formata da un *residuo* $Y(x) = Z(x) - m(x)$.

La variabile risulterà ovviamente:

$$Z(x) = Y(x) + m(x) \quad (2.15)$$

La deriva viene generalmente modellizzata con una funzione polinomiale:

$$m(x) = \sum_{i=1}^k a_i f^i(x) \quad (2.16)$$

Nella quale i coefficienti a_i dovranno essere stimati, e $f^i(x)$ sono monomi di primo grado, generalmente funzione delle coordinate spaziali dei punti.

Il residuo, essendo a media nulla (quindi costante), può essere a sua volta modellizzato con una FA stazionaria o intrinseca, a seconda che ammetta covarianza o solo variogramma.

2.4 Variogrammi sperimentali e teorici

La funzione variogramma viene stimata a partire dai dati provenienti dal campionamento del fenomeno oggetto di studio o da altri dati puntuali di natura indiretta. Una volta scelta una direzione principale, il variogramma sperimentale è calcolato sugli n punti del campo di cui si abbiano informazioni, tramite uno stimatore di questo tipo:

$$\gamma^*(\mathbf{h}) = \frac{1}{2n} \sum_{i=1}^n [Z(x + \mathbf{h}) - Z(x)]^2 \quad (2.19)$$

A partire da un certo tipo di supporto, a prescindere dalla sua regolarità e con opportune tecniche di calcolo, è possibile stimare per diverse distanze h il variogramma sperimentale $\gamma^*(h)$, il cui andamento permette di riconoscere il modello di FA da applicare. Nello studiarne l'andamento, si osserva la funzione variogramma nei pressi dell'origine degli assi e lo stile di dipendenza in funzione di h .

Comportamento nell'origine

Nell'origine del variogramma, ovvero per $h \rightarrow 0$, è possibile individuare tre forme di comportamento:

- *Parabolico*: denota un'elevata regolarità e continuità nell'origine, permettendo di derivare due volte la funzione;
- *Lineare*: è caratterizzato da continuità, ma non derivabilità;
- *Discontinuo*: la funzione non è continua, né derivabile. Nei pressi dell'origine, per h che tende a zero, $\gamma(h)$ non tende a zero a sua volta. Questa discontinuità è detta *effetto pepita* o *nugget effect*.

Andamento in funzione di h

All'aumentare di h , il variogramma aumenta di valore e può evolvere sostanzialmente in due modi:

1. Raggiungere un valore di soglia, o *sill* (FA stazionaria);
2. Aumentare indefinitamente (FA intrinseca).

In entrambi i casi, piuttosto che cercare di descrivere direttamente il comportamento dei variogrammi sperimentali, risulta molto più agevole e pratico ricorrere ad apposite funzioni analitiche comunemente assunte per descriverne il comportamento. La funzione variogramma ha le seguenti proprietà:

- Assume sempre valori positivi: $\gamma(h) \geq 0$;
- Per $h = 0$, si ha sempre $\gamma(h) = 0$;
- È una funzione pari: $\gamma(-h) = \gamma(h)$;

- Quando la FA è stazionaria, $\gamma(h) = C(0) + C(h)$;
- $\gamma(h)$ cresce all'infinito più rapidamente di $|h^2|$: $\lim_{|h| \rightarrow 0} \frac{\gamma(h)}{|h^2|} = 0$;
- Deve essere tale da poter dare luogo a combinazioni lineari autorizzate.

Una volta scelta la funzione matematica più adatta a descrivere il comportamento del variogramma sperimentale, l'aggiustamento viene effettuato tramite un'operazione di *fitting* e la stima dei parametri che costituiscono la funzione variogramma. Il risultato è un variogramma teorico, una funzione che si è in grado di studiare analiticamente, continua e derivabile in ogni punto del proprio dominio.

2.4.1 Modelli di variogramma teorico

Modello pepitico

$$\gamma(h) = c[1 - \delta(r)], \quad r = |h| \geq 0 \quad \begin{cases} \delta(r) = 1, & \text{se } r = 0 \\ \delta(r) = 0, & \text{per qualsiasi altro valore di } r \end{cases}$$

$\delta(r)$ rappresenta la funzione impulso di Dirac, c è il valore soglia del variogramma. Questo modello esprime una discontinuità nell'origine.

Modello sferico

$$\gamma(h) = \begin{cases} c \left(\frac{3}{2} \cdot \frac{r}{a} - \frac{1}{2} \cdot \frac{r^3}{a^3} \right), & \text{per } r = |h| \leq a \\ c, & \text{per } r = |h| \geq a \end{cases}$$

dove r e c rappresentano rispettivamente il *range* e il *sill* e sono parametri del modello. Il comportamento nell'origine è lineare, potendo trascurare il termine cubico nei pressi dell'origine.

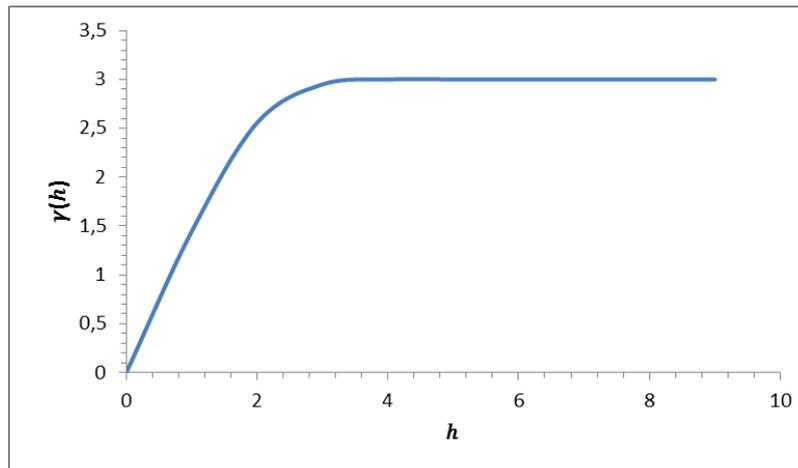


Figura 2.2. Esempio di semivariogramma teorico di tipo sferico

Modello esponenziale

$$\gamma(h) = c \left(1 - e^{-\frac{h}{r}} \right), \quad \text{per } |h| \geq 0$$

Il *sill*, rappresentato dal parametro c , viene in questo caso raggiunto asintoticamente: il *range* risulta quindi infinito. Per poter avere una misura della distanza entro cui si manifesta la correlazione e poter operare un confronto con altri modelli, si introduce un *range pratico* a' , definito come la distanza alla quale si raggiunge il 95% del *sill*, pari a circa $3r$.

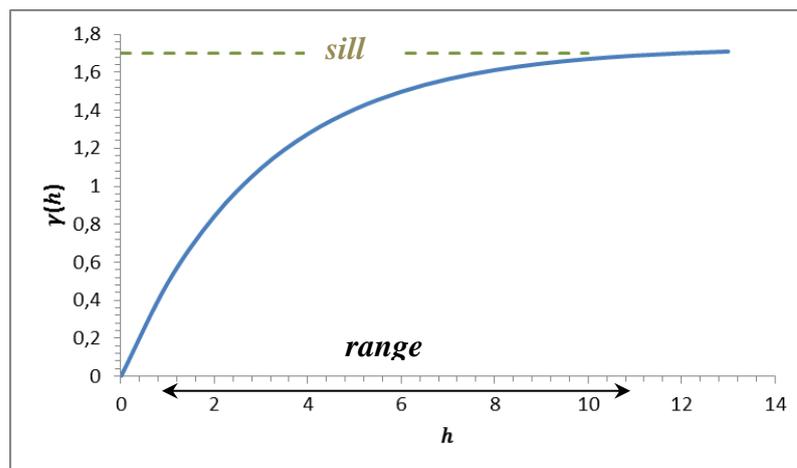


Figura 2.3. Esempio di andamento di un variogramma teorico di tipo esponenziale, con significato di *sill* e *range*

Modello gaussiano

$$\gamma(h) = c \left(1 - e^{-\frac{h^2}{r^2}} \right), \quad \text{per } |h| \geq 0$$

Il parametro c rappresenta anche qui il *sill*, ancora una volta raggiunto asintoticamente con *range* infinito. Per il modello gaussiano viene altresì introdotto un *range pratico* a' , pari alla distanza alla quale si raggiunge il 95% del *sill*, pari ad $a\sqrt{3}$.

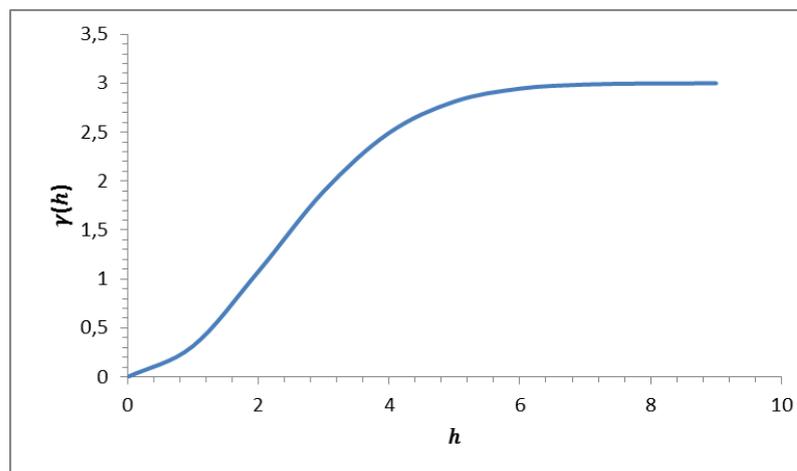


Figura 2.4. Esempio di andamento di un variogramma teorico di tipo gaussiano

Modelli potenza

Questi modelli corrispondono a FA con illimitate capacità di dispersione e che quindi non ammettono covarianza (non stazionarie). I modelli potenza si presentano in questa forma:

$$\gamma(h) = ch^\beta$$

In cui $c > 0$ e al variare di β si può ottenere un diverso tipo di andamento: ad esempio, se $\beta = 0$, il modello è lineare.

Ovviamente quelle appena illustrate sono funzioni analitiche dal comportamento ideale e continuo, con le quali si cerca di modellizzare il variogramma sperimentale. Come si è detto, capita spesso che la funzione sperimentale presenti una discontinuità nell'origine (effetto pepita), per la quale si

rivela necessario adattare il modello con un parametro aggiuntivo, con cui si trasla verso l'alto il variogramma:

$$g(h) = nug + \gamma(h) \quad (2.20)$$

In cui si somma al modello $\gamma(h)$ (uno di quelli proposti in precedenza) la costante positiva nug .

Può accadere, inoltre, che in un variogramma sperimentale si possano riconoscere e distinguere più variogrammi elementari, ciascuno con diversi valori di *range*. Queste strutture elementari inscatolate l'una dentro l'altra si definiscono strutture annidate e possono essere studiate separatamente come variogrammi teorici separati, sfruttando il principio della sovrapposizione degli effetti. Si dimostra infatti che se la variabile $Z(x)$ è composta da m variabili indipendenti e stazionarie $Z_i(x)$, con $i = 1, \dots, m$ e dotate ciascuna di variogramma $\gamma_i(h)$, anche il variogramma della variabile composta $\gamma(h)$ può essere espresso come la sommatoria dei variogrammi elementari:

$$\gamma(h) = \sum_{i=1}^m \gamma_i(h)$$

2.5 Stima locale di una variabile

La *stima* è un'operazione che permette di attribuire un valore ad una variabile in un punto in cui essa non è nota. Attraverso un apposito stimatore, che definisce le modalità con cui si calcola il valore della variabile, questa operazione è possibile a partire dalle caratteristiche locali della variabile in esame, e non quelle globali: pertanto, essa è definita *stima locale*. Gli stimatori più comuni e adatti a questo tipo di operazione sono quelli lineari, i quali permettono di stimare la variabile di interesse in un punto x_0 del campo mediante una combinazione lineare dei valori noti nei dintorni di quel punto, all'interno di una porzione ristretta del campo chiamato *vicinaggio di stima*.

Consideriamo il punto x_0 in cui non sia noto il valore della variabile $Z(x_0)$, intesa come una realizzazione della FA stazionaria $Z(x)$. In una serie di n punti nelle

vicinanze di x_0 , invece, ipotizziamo di conoscere il valore della VA, come anche le funzioni covarianza $C(h)$ e variogramma $\gamma(h)$. Lo stimatore lineare è definito come:

$$Z^*(x_0) = \sum_{i=1}^n \lambda_i Z(x_i) \quad (2.21)$$

Dove i coefficienti λ_i sono i *ponderatori* della stima, in quanto lo stimatore lineare non è altro che la media ponderata delle variabili note all'interno del vicinaggio di stima.

Per caratterizzare questa tecnica da un punto di vista geostatistico, la stima è associata ad un *errore di stima* $\varepsilon(x_0)$ e una *varianza di stima* $\sigma_s^2(x_0)$.

2.5.1 Errore di stima e correttezza della stima

L'errore di stima non è altro che la differenza fra il valore vero ed il valore stimato. Anch'esso, è importante ricordarlo, è una vera e propria Variabile Aleatoria:

$$\varepsilon(x_0) = Z(x_0) - Z^*(x_0) \quad (2.22)$$

Lo stimatore deve avere la fondamentale proprietà di essere *corretto (unbiased)*: questa condizione è verificata se l'errore è a media nulla:

$$E[\varepsilon(x_0)] = 0 \quad (2.23)$$

Osservando la precedente equazione e scomponendola nei suoi addendi come in (2.23),

$$E[Z(x_0) - Z^*(x_0)] = 0 \quad (2.24)$$

Ricordiamo di essere in condizioni di stazionarietà, per cui $E[Z(x)] = m = \text{cost}$, $\forall s \in S$ e sfruttiamo la linearità dell'operatore valore atteso e raccogliamo il termine contenente $Z(x)$:

$$E \left[Z(x_0) - \sum_{i=1}^n \lambda_i Z(x_i) \right] = 0 \quad (2.25)$$

$$= m \left[1 - \sum_{i=1}^n \lambda_i \right] = 0$$

Portando al seguente risultato, importante condizione sui ponderatori e conseguenza della media nulla dell'errore di stima:

$$\sum_{i=1}^n \lambda_i = 1 \quad (2.26)$$

2.5.2 Varianza di stima e accuratezza della stima

Gli errori di stima sono caratterizzati da una legge di densità di probabilità con media nulla (stimatore corretto) e una certa dispersione che indica la frequenza con cui si verificano errori elevati. Non disponendo tuttavia di questa legge, si caratterizza la qualità della stima tramite la sua varianza, detta appunto *varianza di stima*, che dipende strettamente dalla funzione variogramma della VA e dalla covarianza.

Per la definizione di varianza, la varianza di stima risulta:

$$\sigma_s^2 = E \left[\left(Z(x_0) - \sum_{i=1}^n \lambda_i Z(x_i) \right)^2 \right] \quad (2.27)$$

Sviluppando il quadrato del binomio, si ottiene:

$$\sigma_s^2 = E \left[Z(x_0)^2 - 2 \cdot Z(x_0) \sum_{i=1}^n \lambda_i Z(x_i) + \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j Z(x_i) Z(x_j) \right]$$

Il quale, per le proprietà di linearità del valore atteso e le eq. (2.6) e (2.10), diventa:

$$\begin{aligned} \sigma_s^2 &= C(0) + \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j C(x_i, x_j) - 2 \sum_{i=1}^n \lambda_i C(x_i, x_0) \\ &= C(0) + \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j [C(0) - \gamma(h_{i,j})] - 2 \sum_{i=1}^n \lambda_i [C(0) - \gamma(h_{i,0})] \end{aligned}$$

E che, per la (2.26), ci consente di ottenere:

$$\sigma_s^2 = 2 \sum_{i=1}^n \lambda_i \gamma_{i,0} - \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \gamma_{i,j} \quad (2.28)$$

Questo consente di notare che la varianza di stima non dipende dalla covarianza $C(0)$, ma soltanto dal variogramma, per cui non c'è bisogno dell'ipotesi di stazionarietà introdotta in precedenza.

Per quanto riguarda il calcolo dei ponderatori λ_i dell'equazione (2.21) per poter effettuare la stima della variabile regionalizzata, si fa riferimento a tecniche individuabili secondo due differenti approcci:

- **Tecniche deterministiche**, che considerano l'aspetto puramente geometrico, come il *metodo dei poligoni di Thiessen*;
- **Tecniche geostatistiche**, note come *Kriging*, basate sulla regressione, che consentono di trovare i ponderatori minimizzando la varianza di stima.

Benché esistano diverse tipologie di Kriging, i paragrafi seguenti trattano l'*Ordinary Kriging* ed il *Topological Kriging*.

2.6 Ordinary Kriging

Dato un set di ponderatori, è possibile calcolare, con l'ausilio del variogramma, la precisione della stima corrispondente. Il problema che ci si pone, in particolare, è quello di trovare quei coefficienti che diano luogo alla stima con la maggior precisione: in altri termini, quella che presenti una varianza di stima minima. In termini matematici, il metodo di ottimizzazione è quello di Lagrange, che minimizza la varianza di stima eguagliando a zero le sue derivate parziali rispetto ai ponderatori λ_i e, ricordando l'equazione (2.26), pone:

$$\frac{\partial \sigma_s^2}{\partial \lambda_i} = 0 \quad \forall i = 1, \dots, n \quad \sum_{i=1}^n \lambda_i = 1$$

Ricordando l'espressione della varianza di stima, si ottiene:

$$\frac{\partial \sigma_s^2}{\partial \lambda_i} = \frac{\partial}{\partial \lambda_i} \left[2 \sum_{i=1}^n \lambda_i \gamma_{i,0} - \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \gamma_{i,j} - 2\mu \left(1 - \sum_{i=1}^n \lambda_i \right) \right] = 0 \quad \forall i = 1, \dots, n$$

Dove $\mu \in \mathbb{R}$ è il parametro lagrangiano incognito che descrive il vincolo della funzione. Svolgendo la derivata, si ottiene:

$$\frac{\partial \sigma_s^2}{\partial \lambda_i} = 2\gamma_{i,0} - 2 \sum_{j=1}^n \lambda_j \gamma_{i,j} - 2\mu = 0 \quad \forall i = 1, \dots, n \quad (2.29)$$

Semplificando e mettendo a sistema con la (2.26), si ottiene:

$$\begin{cases} \sum_{j=1}^n \lambda_j \gamma_{i,j} + \mu = \gamma_{i,0} \\ \sum_{i=1}^n \lambda_i = 1 \end{cases} \quad \forall i = 1, \dots, n \quad (2.30)$$

Si può esprimere questo sistema in forma esplicita:

$$\begin{aligned} \lambda_1 \gamma_{1,1} + \lambda_2 \gamma_{1,2} + \dots + \lambda_n \gamma_{1,n} + \mu &= \gamma_{1,0} \\ \lambda_1 \gamma_{2,1} + \lambda_2 \gamma_{2,2} + \dots + \lambda_n \gamma_{2,n} + \mu &= \gamma_{2,0} \\ &\vdots \\ \lambda_1 \gamma_{n,1} + \lambda_2 \gamma_{n,2} + \dots + \lambda_n \gamma_{n,n} + \mu &= \gamma_{n,0} \\ \lambda_1 + \lambda_2 + \dots + \lambda_n &= 1 \end{aligned} \quad (2.31)$$

Mentre in forma matriciale, lo stesso sistema corrisponde a:

$$\begin{bmatrix} \gamma_{1,1} & \gamma_{1,2} & \dots & \gamma_{1,2} & \gamma_{1,2} & 1 \\ \gamma_{2,1} & \gamma_{2,2} & \dots & \gamma_{2,2} & \gamma_{2,2} & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \gamma_{n,1} & \gamma_{n,2} & \dots & \gamma_{n,n} & \gamma_{n,2} & 1 \\ 1 & 1 & \dots & 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \\ \mu \end{bmatrix} = \begin{bmatrix} \gamma_{1,0} \\ \gamma_{2,0} \\ \vdots \\ \gamma_{n,0} \\ 1 \end{bmatrix} \quad (2.32)$$

Questo sistema suggerisce che la matrice dei coefficienti e il termine noto dipendono esclusivamente dalla funzione variogramma, calcolabile in tutti i punti del campo – almeno teoricamente. Il sistema ammette un'unica soluzione se le posizioni i-esime sono distinte, risolvendo in definitiva il sistema lineare qui sopra esposto e di seguito riportato in forma compatta:

$$\lambda = \Gamma^{-1}g_0$$

2.7 Topological Kriging

Il Topological kriging (o Top-kriging) è un metodo di stima di tipo geostatistico sviluppato da Skøien et al. (2006) e basato sul Block-kriging, il quale definisce la variabile regionalizzata su supporto *non puntuale*, ma pari all'estensione areale del supporto stesso (nelle applicazioni idrologiche il bacino imbrifero sotteso dalla sezione fluviale di interesse). Per tenere conto della relazione fra la variabile regionalizzata e il supporto, si media la variabile $z(x)$ sull'area A , o più precisamente la si standardizza rispetto all'area (si rimuove cioè la dipendenza dall'area della variabile stessa, il ché coincide il più delle volte con una media sull'area):

$$\bar{z}(A) = \frac{1}{A} \int_A w(x)z(x)dx \quad (2.33)$$

dove \bar{z} è la variabile regolarizzata e $w(x)$ una funzione di peso, tipicamente assunta uguale a 1 per l'ipotesi che la variabile sia costante su tutta l'area. Il sistema di krigaggio da risolvere resta lo stesso se si prende in considerazione un supporto non puntuale, anche se il variogramma utilizzato andrà regolarizzato anch'esso (Skøien et al. 2011):

$$\begin{aligned} \gamma_{12} &= \frac{1}{2} \times Var(z(A_1) - z(A_2)) = \\ &= \frac{1}{A_1 A_2} \int_{A_1} \int_{A_2} \gamma_p(|\mathbf{x}_1 - \mathbf{x}_2|) d\mathbf{x}_1 d\mathbf{x}_2 + \\ &\quad - \frac{1}{2} \left[\frac{1}{A_1^2} \int_{A_1} \int_{A_1} \gamma_p(|\mathbf{x}_1 - \mathbf{x}_2|) d\mathbf{x}_1 d\mathbf{x}_2 + \frac{1}{A_2^2} \int_{A_2} \int_{A_2} \gamma_p(|\mathbf{x}_1 - \mathbf{x}_2|) d\mathbf{x}_1 d\mathbf{x}_2 \right] \end{aligned} \quad (2.34)$$

in cui \mathbf{x}_1 e \mathbf{x}_2 sono i vettori posizione relativi a ciascun'area utilizzata nell'integrazione, mentre γ_p è il variogramma puntuale. Questa espressione mette in evidenza l'effetto regolarizzante, producendo una variabile più bassa rispetto a quella puntuale: maggiore è l'area, più evidente è questo effetto.

Poiché può essere molto complicato risolvere analiticamente l'espressione (3.4), si ricorre ad un'opportuna discretizzazione del campo, che permette di trasformare l'integrale in una sommatoria.

Skøien, inoltre, propone un modo per regolarizzare l'effetto *nugget* fra due supporti di diversa estensione:

$$C_{0p}(A_1, A_2) = \frac{1}{2} \left(\frac{C_{0p}}{A_1} + \frac{C_{0p}}{A_2} - \frac{2C_{0p} \text{Mis}(A_1 \cap A_2)}{A_1 A_2} \right) \quad (2.35)$$

dove C_{0p} è l'effetto pepita, mentre $\text{Mis}(A_1 \cap A_2)$ corrisponde a:

- $\min\{A_1, A_2\}$ se le aree sono in sovrapposizione;
- 0 se le aree sono disgiunte.

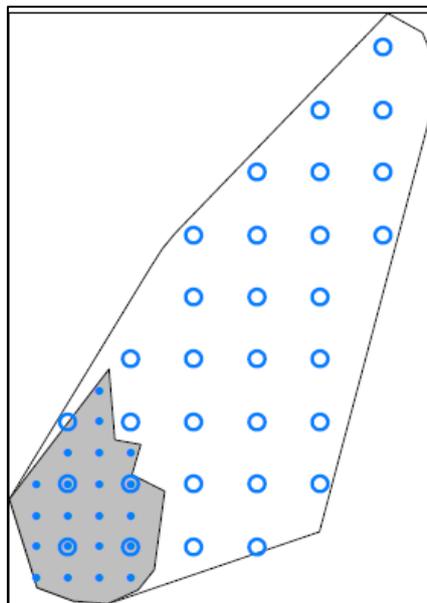


Figura 2.5. Esempio di discretizzazione in due diverse aree a differenti risoluzioni (immagine tratta da Skøien et al., 2006)

Nel determinare i ponderatori, viene fatto notare dall'autore come le dimensioni delle aree usate come supporto influenzino il calcolo dei pesi, così come l'eventuale struttura annidata dei bacini oggetto di studio. Si è mostrato, ad esempio, che un ponderatore relativo a un sottobacino nel bacino su cui si opera la stima risulta maggiore rispetto ad uno calcolato per un bacino di pari estensione e nelle vicinanze.

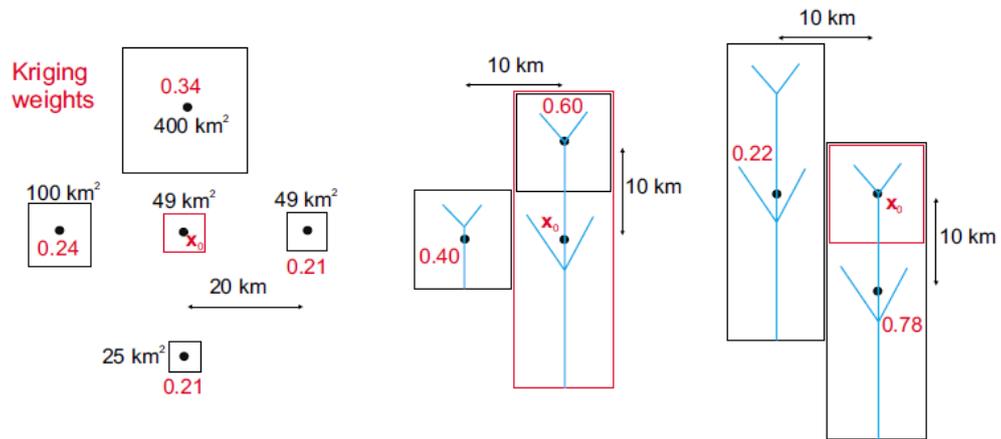


Figura 2.6. Esempio degli effetti sui ponderatori delle dimensioni dei bacini (a sinistra) e della loro struttura annidata, a parità di distanza (Skøien et al., 2006)

Capitolo 3 – Metodo di regressione ai Minimi Quadrati Generalizzati

Nell'ambito dei modelli di stima regionale delle portate al colmo di piena, la letteratura scientifica ha in passato proposto metodi basati sull'uso dei minimi quadrati generalizzati, meglio conosciuti in letteratura anglosassone come *Generalized Least Squares* (GLS). Questa tecnica ha avuto particolare successo e diffusione, specialmente nelle applicazioni pratiche, per la sua robustezza analitica e teorica, oltre che per lo sviluppo di software dedicati. Negli Stati Uniti il GLS è il metodo di riferimento per la stima delle piene in bacini non strumentati (v. Tasker e Stedinger, 1989; Griffis e Stedinger, 2007).

Tali metodi mettono in relazione le caratteristiche geomorfologiche e climatiche del bacino con la portata d'interesse e in base a questa relazione si stima la portata, o più di frequente, quantili di portata di piena al colmo relativi ad un bacino per il quale non sono note le osservazioni di portata. Per una serie di n quantili Y , ottenuti da altrettante stazioni di misura, l'equazione di regressione si esprime tipicamente nella forma:

$$Y = \sum_{i=1}^M \beta_i x_i + \varepsilon$$

Dove Y è il vettore delle portate in forma logaritmica relative alle sezioni strumentate, x_i i vettori che contengono gli M valori, sempre in forma logaritmica, delle caratteristiche geomorfologiche e climatiche dei bacini oggetto di studio, β_i sono i coefficienti stimati dalla regressione ed ε rappresenta l'errore del modello, spesso chiamato vettore dei residui.

Quando i residui del modello di regressione sono indipendenti e hanno la stessa varianza (sono *omoschedastici*), può essere applicata una regressione ai minimi quadrati ordinari (OLS) per stimare i coefficienti del modello. Tuttavia, è improbabile che corsi d'acqua appartenenti alla stessa area geografica abbiano eventi di piena completamente indipendenti l'uno dall'altro: di conseguenza, l'ipotesi di avere residui indipendenti rischia di non essere valida (Stedinger et al., 1993). Inoltre, poiché per ciascun bacino sono disponibili dati in quantità diverse, la precisione di stima del quantile differirebbe fra una sezione strumentata e l'altra, e quindi verrebbe meno

l'assunzione di avere una stessa varianza per tutti i residui. Per evitare che cadano le ipotesi di indipendenza e stessa varianza, si ricorre alla regressione ai minimi quadrati generalizzati (GLS), che tiene conto delle differenti varianze e la correlazione spaziale fra le portate attribuendo un peso ad ogni quantile di piena, che dipende dalla lunghezza delle osservazioni stesse, da una stima della cross-correlazione fra le piene e della loro varianza attraverso una regressione OLS (Stedinger, Tasker, 1985).

Se la vera grandezza y_i può essere espressa come una funzione lineare delle caratteristiche del bacino x_{ik} in forma logaritmica e dell'errore del modello δ_i , avremo:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + \delta_i \quad (3.1)$$

assumendo che gli errori δ_i abbiano distribuzione normale, cioè con media nulla e varianza costante σ_δ^2 . Va ricordato, tuttavia, che la differenza fra la vera grandezza e l'osservazione dipende da un errore casuale η_i , che influenza l'osservazione \hat{y}_i :

$$\hat{y}_i = y_i + \eta_i \quad (3.2)$$

Risulta perciò:

$$\hat{y}_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + \delta_i + \eta_i \quad (3.3)$$

Poiché $y_i = \log(Q_i)$, riportando la precedente equazione in forma logaritmica ed esprimendola successivamente in funzione dell'argomento, si ottiene un vettore un vettore così costituito:

$$\hat{Q}_i = b_0 R_{i1}^{b_1} R_{i2}^{b_2} \dots R_{ik}^{b_k} 10^{\delta_i + \eta_i} \quad (3.4)$$

in cui b_j è una stima del j -esimo coefficiente beta ($j = 0, \dots, k$) e R_{ij} sono i valori osservati delle k caratteristiche del bacino ($R_{ij} = 10^{x_{ij}}$ per $j = 1, \dots, k$). L'equazione (3.3) può essere riscritta in forma matriciale:

$$\hat{Y} = X\beta + \varepsilon \quad (3.5)$$

dove X è una matrice $N \times (k + 1)$ che contiene le k caratteristiche del bacino, con l'aggiunta di un'ulteriore colonna di 1, β un vettore dei $(k + 1)$ parametri della

regressione, ed $\varepsilon = \delta + \eta$ contiene i $(k + 1)$ errori casuali, per i quali $E[\varepsilon] = 0$ (valore atteso dell'errore casuale) e $E[\varepsilon\varepsilon^T] = \Lambda$ (valore atteso della correlazione).

A causa della correlazione fra i residui, l'analisi tradizionale ai minimi quadrati ordinari (OLS) non è considerata appropriata, pertanto viene impiegato il metodo dei minimi quadrati generalizzati (GLS), che mette in relazione i quantili stimati alle caratteristiche geomorfoclimatiche del bacino e descrive gli errori. β è stimato come:

$$\mathbf{b} = (\mathbf{X}^T \mathbf{\Lambda}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{\Lambda}^{-1} \hat{\mathbf{Y}} \quad (3.6)$$

Le stime risulterebbero non deviate ed esatte, ovvero saremmo in presenza di un modello B.L.U.E. (*Best Linear Unbiased Estimator*) se fosse nota con precisione Λ . Poiché quest'ultima non è nota, è necessario stimarla tramite:

$$\hat{\mathbf{\Lambda}}(\sigma_\delta^2) = \sigma_\delta^2 \mathbf{I}_N + \hat{\mathbf{\Sigma}} \quad (3.7)$$

di cui \mathbf{I}_N è una matrice identità di ordine N e σ_δ^2 è la varianza di stima, che descrive la precisione con cui il modello $\mathbf{X}\beta$ predice i valori di $\hat{\mathbf{Y}}$. L'obiettivo che ci si pone nel ricorrere a questo tipo di modello è, evidentemente, di ottenere stime con il minimo valore possibile di σ_δ^2 .

$\hat{\mathbf{\Sigma}}$ la matrice delle covarianze di stima, le cui componenti lungo la diagonale e fuori di essa sono stimate a partire da Tasker e Stedinger (1989), che propongono la seguente formula:

$$\begin{aligned} \Sigma_{ii} &= \left[1 + K_i \gamma_i + \frac{1}{2} K_i^2 (1 + 0,75 \gamma_i^2) \right] \frac{\sigma_i^2}{n_i} && \text{per } i = j \\ \Sigma_{ij} &= \left[1 + \frac{1}{2} K_i \gamma_i + \frac{1}{2} K_j \gamma_j + \frac{1}{2} K_i K_j (\rho_{ij} + 0,75 \gamma_i \gamma_j) \right] \rho_{ij} \frac{m_{ij} \sigma_i \sigma_j}{n_i n_j} && \text{per } i \neq j \end{aligned} \quad (3.8)$$

Dove m_{ij} è la lunghezza simultanea del numero dei campioni fra i siti i e j , ρ_{ij} il coefficiente di cross-correlazione delle portate fra i bacini i e j , σ_i e σ_j le deviazioni standard delle osservazioni relative ai rispettivi siti, γ_i e γ_j i coefficienti di asimmetria campionari riferiti ai due bacini, e K_i e K_j sono coefficienti che esprimono il fattore di frequenza della distribuzione.

Per evitare la correlazione fra i residui e i quantili, gli autori raccomandano che:

- i. ρ_{ij} sia stimato in base alla distanza fra i siti i e j ;

- ii. Le deviazioni standard σ_i e σ_j siano ottenute mediante una regressione a parte sull'area d'interesse;
- iii. Sia utilizzato un coefficiente di asimmetria regionale G_i al posto di γ_i .

Uno stimatore della matrice di covarianza alternativo per evitare la correlazione è proposto da Griffis e Stedinger (2007).

Capitolo 4 – Area di studio

L'area di studio considerata nella presente trattazione si trova negli Stati Uniti sud-orientali e riguarda uno studio sulla regionalizzazione delle piene pubblicato in un rapporto dell'USGS (*United States Geological Survey*, v. Gotvald et al., 2009), il quale ha lo scopo di stimare la portata di piena di progetto in bacini rurali non strumentati. Delle 943 stazioni oggetto di studio da parte dell'USGS, ne sono state prese in considerazione 61, situate negli stati dell'Alabama (3), della Florida (1) e della Georgia (57) e rappresentate in fig. 4.2.

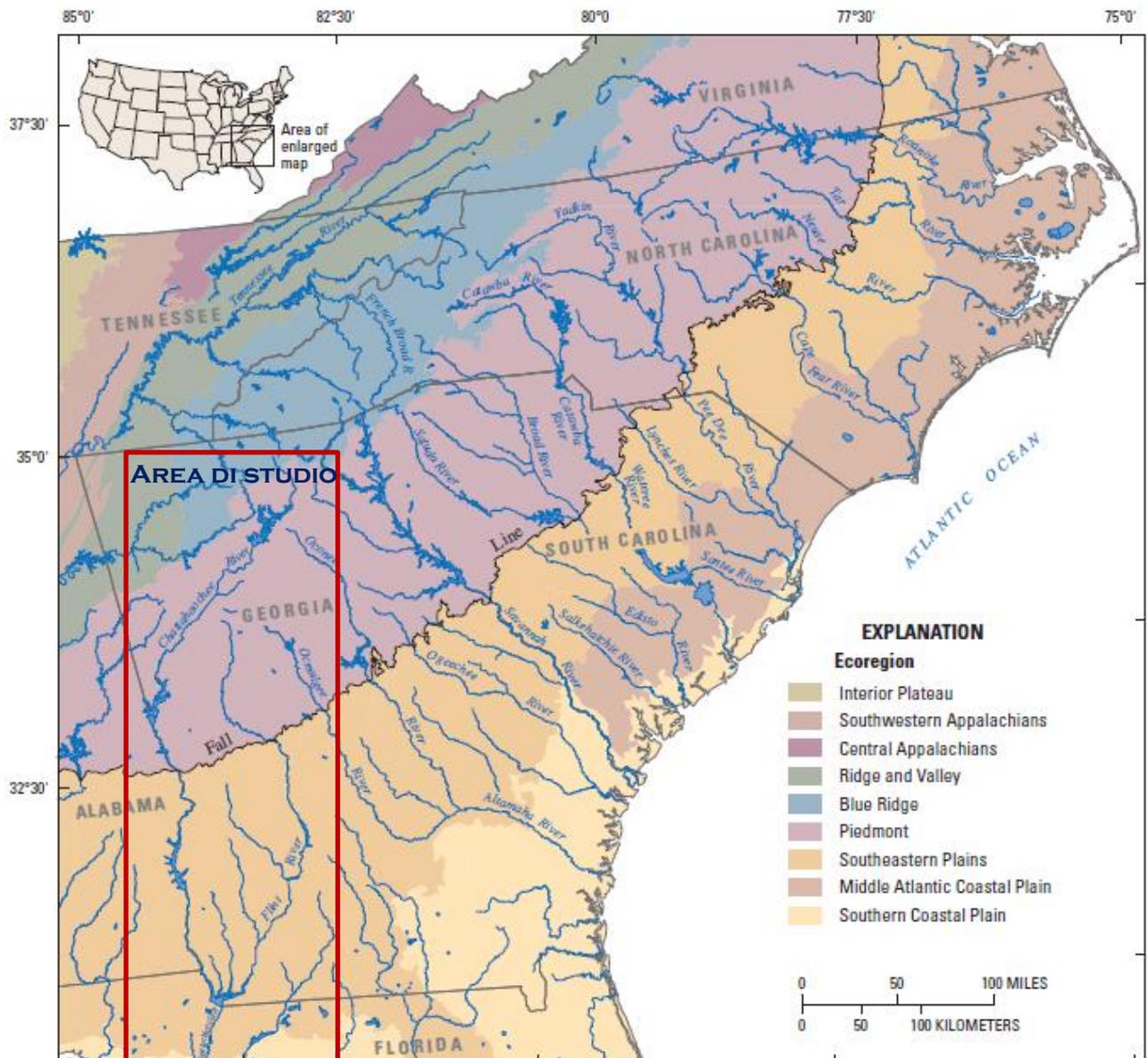


Figura 4.1. Inquadramento geografico dell'area di studio (Gotvald et al., 2009)

L'area in questione, è caratterizzata da una precipitazione media annua compresa fra i 1000 e i 1500 mm, associata principalmente al movimento di fronti caldi e freddi fra i mesi di novembre e aprile, con temporali estivi isolati fra maggio e ottobre, e una temperatura media annua compresa fra i 12 e i 20°C. Tempeste tropicali ed uragani, che si verificano occasionalmente sulle coste dell'Atlantico e del Golfo, possono provocare forti aumenti di piovosità sui bacini di studio. L'area è suddivisa in eco-regioni classificate dall'USEPA (*U.S. Environmental Protection Agency*). Le eco-regioni raggruppano aree che presentano ecosistemi con caratteristiche generali di similarità per tipologia, qualità e quantità di risorse ambientali, e costituiscono una base spaziale per la ricerca, la valutazione, la gestione e il monitoraggio degli ecosistemi e le loro componenti. Esse sono individuate a partire dalle caratteristiche spaziali del territorio e dall'insieme dei fenomeni biotici e abiotici, che comprendono la geologia, la morfologia, la biologia, la flora e la fauna, il clima, i suoli e il loro uso e l'idrologia. Le eco-regioni comprese nella zona oggetto di studio sono principalmente tre, come illustrato in figura 4.1: l'area delle pianure sud-orientali (*Southeastern Plains*) si estende in direzione nord fino alla linea di demarcazione denominata *Fall Line*, che segna il confine con una vasta area pedemontana (*Piedmont*); a nord si incontrano altre due eco-regioni, denominate *Blue Ridge* e *Ridge and Valley*, marginali rispetto alla nostra area di studio.

L'eco-regione *Southeastern Plains* è costituita da territori pianeggianti e irregolari che comprendono un mosaico di terreni coltivabili, pascoli ed aree boschive. Caratterizzata da quote e rilievi maggiori rispetto all'adiacente eco-regione delle pianure costiere sud-orientali (*Southeastern Coastal Plain*), presenta tuttavia una situazione geologica più recente rispetto alla zona pedemontana a nord. I corsi d'acqua di quest'area presentano tendenzialmente basse pendenze e alvei sabbiosi. A nord, l'eco-regione *Piedmont* costituisce una zona di transizione fra le zone montuose degli Appalachi a nord-ovest e l'area pressoché pianeggiante della pianura costiera a sud-est. Caratterizzata da formazioni di rocce metamorfiche ed ignee, quest'area presenta pianure irregolari alternate a zone collinari. Un tempo destinati per lo più ad uso agricolo, i suoli di questa regione sono regrediti a foreste di pini e alberi da legno duro, con un recente aumento della conversione in aree urbane e suburbane.

4.1 Dati geomorfologici

I dati messi a disposizione dall'USGS per i 61 bacini d'interesse presentano informazioni di natura morfologica, climatica, idrologica e di uso del suolo relative a ciascuna stazione, in base alle quali è possibile effettuare le stime sui quantili.

I confini degli spartiacque sono stati individuati con due differenti modalità, entrambe basate sulla tecnologia dei sistemi informativi geografici (GIS): una parte è stata generata da modelli digitali del terreno (DEM) provenienti dal *National Elevation Dataset* (NED), con risoluzione orizzontale variabile fra i 10 e i 30 metri – le stazioni in esame rientrano in questa prima casistica; un'altra proviene da DEM elaborati a partire da rilievi LIDAR (*Light Detection and Ranging*) con risoluzione di 3 metri.

In tabella 4.1 sono elencati alcuni dei 22 parametri utilizzati nello studio, per i quali si rimanda al rapporto dell'USGS (Gotvald et al., 2009). Essi sono l'area e il perimetro del bacino, la quota media, la lunghezza dell'asta fluviale, la frazione impermeabile di bacino, la densità di drenaggio e la precipitazione media annua. Per ciascuna variabile, la tabella 4.2 riporta il primo quartile, la mediana, il terzo quartile e i valori minimo e massimo.

Tabella 4.1. Descrizione delle variabili geomorfologiche prese in considerazione nello studio

Parametro	Significato
A [km^2]	Area del bacino
L [km]	Lunghezza dell'asta fluviale
P [km]	Perimetro del bacino
H_M [m]	Quota media del bacino
FI [%]	Frazione di bacino impermeabile
MAP [mm]	Precipitazione media annua
D_D [km/km^2]	Densità di drenaggio

Tabella 4.2. Parametri geomorfologici e climatici considerati per l'area di studio

	Min	Max	Media	I quartile	Valore mediano	III quartile
A [km^2]	1,9	13752,1	792,7	121,7	328,9	727,7
L [km]	2,9	410,5	54,8	22,6	43,1	65,2
P [km]	5,7	1071,6	159,9	73,2	113,7	205,5
H_M [m]	47,3	758,2	214,3	129,5	174,1	243,0
FI [%]	0,1	7,9	1,0	0,4	0,6	1,1
MAP [mm]	1079,5	1899,2	1342,0	1265,0	1300,2	1397,0
D_D [km/km^2]	0,7	5,8	1,3	1,2	1,3	1,4

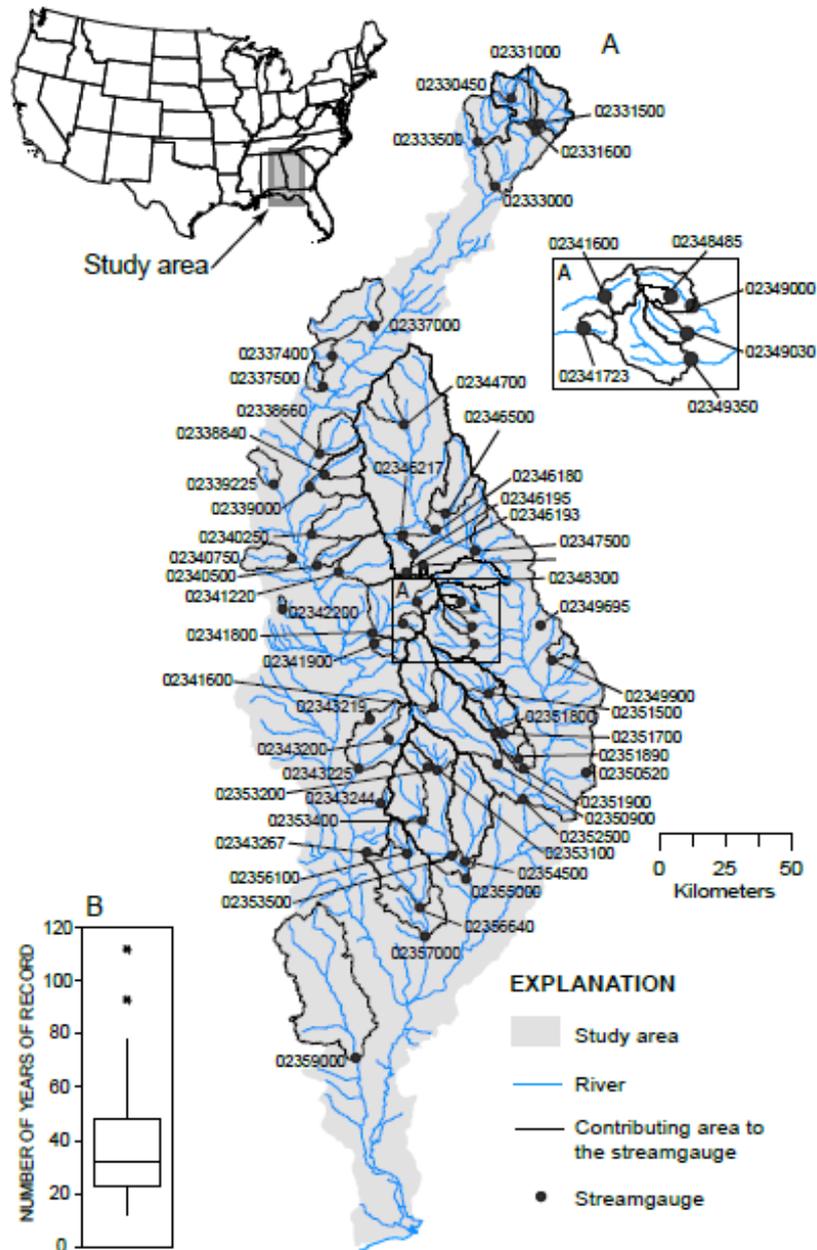


Figura 4.2. I 61 bacini dell'area di studio con relativo codice identificativo, sezioni di chiusura, reticolo idrografico (in blu) e spartiacque (confini in nero). Nel riquadro B è mostrato l'intervallo di lunghezza delle registrazioni relative alle stazioni di misura. (Archfield et al., 2013)

4.2 Dati idrometrici

Le informazioni idrometriche relative all'area di studio sono costituite da una serie di stime di quantili associati a tempi di ritorno di 10, 50, 100 e 500 anni e provengono da un'analisi di frequenza delle portate di picco annuali relative a osservazioni di almeno 10 anni e il cui regime non è significativamente influenzato dalla presenza di dighe, casse di espansione, maree o aree urbanizzate.

I quantili empirici sono stati calcolati in base alle linee guida fornite dal “Bulletin 17-B” rilasciato dalla *Hydrology Subcommittee of the Interagency Advisory Committee on Water Data* (1982), come menzionato in Gotvald et al. (2009). La stima è avvenuta adottando una distribuzione di Pearson di Tipo III e adattandola alle serie temporali relative alle portate di picco annuali registrate in ciascuna stazione tramite il calcolo della *media*, della *deviazione standard* e del *coefficiente di asimmetria (skewness)* delle osservazioni in forma logaritmica, i quali descrivono, rispettivamente, il punto medio, la pendenza e la curvatura della concavità/convessità della curva di frequenza. La probabilità di non superamento è espressa in Gotvald et al. (2009) in termini di percentuale, anziché in anni come di consueto: di conseguenza, una portata a cui sia associata una probabilità di non superamento dell’1% corrisponde alla cosiddetta “portata centennale”, in quanto è riferita a un evento che può essere eguagliato o superato con una probabilità di 1 a 100.

Le stime sulla probabilità di superamento in termini di percentuale P sono eseguite inserendo i tre momenti sopra menzionati nella seguente equazione:

$$\log Q_P = X + KS \quad (4.1)$$

In cui Q_P è il quantile associato alla probabilità in percentuale P , X corrisponde alla media delle osservazioni di portata al colmo di piena annuale in forma logaritmica, K è un fattore ottenuto tramite il coefficiente di asimmetria e della probabilità di non superamento assegnata, mentre S è la deviazione standard dei picchi di piena annuali, sempre in forma logaritmica, che costituisce una misura del grado di variazione dei valori annuali rispetto al valor medio.

Come nota riguardante il rapporto USGS, gli autori hanno spiegato la preferenza di considerare il tempo di ritorno in termini di percentuale, piuttosto che in anni, per evitare confusione da parte del pubblico, che potrebbe interpretare questo concetto come un vero e proprio intervallo temporale in cui si verificano piene di una certa importanza. Le piene, viene ricordato, sono eventi casuali da considerare in termini probabilistici, e il fraintendimento ha a che fare, tipicamente, con il numero di volte in cui un certo evento di piena si verifica durante il periodo a cui fa riferimento il tempo di ritorno. Il fatto che l’evento si verifichi con una data frequenza attesa, infatti, non significa che non possa accadere più volte durante quello stesso arco di tempo o, viceversa, che non possa verificarsi affatto.

Nella presente trattazione, comunque, i tempi di ritorno sono espressi come d'usuale, ossia in termini di anni. Un quadro sulle portate al colmo per i tempi di ritorno d'interesse è presentato nella seguente tabella, nella quale si riportano, come per le variabili geomorfologiche di cui al paragrafo precedente, i valori minimo, medio e massimo, il primo quartile, la mediana e il terzo quartile.

Tabella 4.1. Dati sui quantili di partenza in base ai tempi di ritorno

	Q₁₀ [m ³ /s]	Q₅₀ [m ³ /s]	Q₁₀₀ [m ³ /s]	Q₅₀₀ [m ³ /s]
Min	3,5	7,5	9,9	13,8
Max	1673,5	2489,1	2860,0	3822,8
Media	243,0	397,9	475,7	690,9
I quartile	57,5	96,8	118,1	165,1
Valore mediano	168,8	303,0	368,1	535,2
III quartile	264,5	481,4	603,1	923,1

Capitolo 5 – Applicazione di GLS e Top-kriging per la stima della piena di progetto in bacini non strumentati appartenenti all'area di studio

5.1 Applicazione del metodo GLS

Come già accennato nel cap. 3, il metodo di regressione ai Minimi Quadrati Generalizzati consente di stimare i quantili di portata di piena al colmo attraverso un sistema lineare che tiene conto di una serie di parametri dipendenti dalle caratteristiche geomorfologiche del bacino di riferimento.

L'indagine condotta da Gotvald et al. (2009) sull'area di studio e pubblicata nel rapporto USGS ha previsto l'applicazione del metodo GLS per l'analisi di regressione regionale della stima delle portate di piena al colmo in bacini privi di strumentazioni idrometriche, fra i quali vi sono le stazioni di nostro interesse. Le equazioni sviluppate in seguito a questo studio mettono in relazione le stime sui quantili associati ai tempi di ritorno considerati con le caratteristiche geomorfoclimatiche dei bacini.

In principio, sono state condotte alcune analisi preliminari basate sulla regressione ai Minimi Quadrati Ordinari (OLS) con lo scopo di:

- determinare quali fossero le variabili geomorfoclimatiche che meglio interpretassero relazioni log-lineari;
- suddividere l'area di studio in 5 regioni idrologicamente omogenee.

Tali analisi preliminari hanno dimostrato che i descrittori che interpretano meglio la natura dei dati di portata sono l'area drenante del bacino (*DA*) e la frazione di appartenenza di un bacino ad una determinata regione omogenea (*PCT*); si è dimostrato che l'introduzione di ulteriori variabili aumenterebbe la complessità del modello, ma non apporterebbe benefici significativi, in termini prestazionali, tali da giustificare una loro implementazione nell'analisi regionale. La stima dei quantili è stata, pertanto, effettuata ricorrendo al metodo dei Minimi Quadrati Generalizzati (GLS) il cui risultato finale consiste in una serie di equazioni di regressione che esprimono il quantile associato a un determinato valore del tempo di ritorno in funzione dell'area drenante appartenente al bacino d'interesse, qui riportate in forma logaritmica:

$$\log Q_{TR} = a_1(PCT_1) + a_2(PCT_2) + \dots + a_n(PCT_n) + b_0 \log DA$$

In cui $PCT_1, PCT_2, \dots, PCT_n$ sono le frazioni di bacino che, in percentuale, ricadono nell’area omogenea designata $1, 2, \dots, n$; $a_1, a_2, \dots, a_n, b_0$ sono i parametri della regressione e DA è l’estensione dell’area drenante. In questo modo si sono ottenuti i quantili empirici relativi a ciascuno dei bacini appartenenti all’area di studio, per ognuno dei tempi di ritorno assegnati.

5.2 Applicazione Top-kriging

Recenti studi (Castiglioni et al., 2011, Chokmani & Ouarda, 2004, Archfield et al., 2013) hanno dimostrato che le tecniche di stima geostatistiche, innovative nelle applicazioni idrologiche, possono avere un ruolo fondamentale nei problemi di stima di variabili idrologiche (dai quantili di piena ad indici di magra) in bacini non strumentati (PUB: *Prediction in Ungauged Basins*).

Tra tutte le tecniche utilizzate, l’interpolazione dei dati idrometrici mediante il metodo Top-kriging permette di considerare il supporto non puntuale della variabile idrologica, il quale è pari all’area drenante del bacino, mediante l’attribuzione di pesi differenti in dipendenza dall’estensione di quest’ultima, dalla posizione relativa dei bacini e della loro eventuale struttura annidata (v. Par. 2.7).

Applicando questa tecnica di tipo geostatistico Archfield et al. (2013) hanno ottenuto stime di valori di portata al colmo di piena attraverso un algoritmo di cross-validazione di tipo *jack-knife*, meglio noto come *Leave-One-Out-Cross-Validation* (LOOCV) (v. Par. 5.3.4). Il calcolo è stato condotto su 4 differenti serie di quantili di piena al colmo con assegnato tempo di ritorno, rispettivamente di 10, 50, 100, 500 anni. I risultati, riportati in tabella 5.2, mostrano che l’applicazione del Top-kriging ottiene incoraggianti performances su tutti i quantili considerati, seppur con un evidente calo di capacità predittiva all’aumentare del tempo di ritorno.

Nello stesso studio, la stima delle variabili idrologiche con questa tecnica costituisce solo una fase preliminare di un’indagine più approfondita, il cui scopo è stato quello di confrontare le prestazioni di differenti tecniche di kriging e della loro combinazione applicata ai residui delle rispettive stime.

5.3 Procedimento seguito per l’indagine

I residui, o gli scarti, ottenuti dall’applicazione del modello regionale RFFA-GLS per la stima di quantili di piena al colmo in bacini non strumentati, possono essere ritenuti, teoricamente, una variabile regionalizzata a tutti gli effetti (v. Par. 2.1), data la loro distribuzione spaziale. Seguendo l’analisi condotta da Archfield et al. (2013) il Top-kriging può essere applicato per modellare i residui ottenuti dalle stime GLS, in modo tale che la correzione operata dal modello geostatistico possa, in qualche maniera, migliorare la capacità predittiva dello stimatore regressivo. La presente dissertazione prosegue, quindi, in direzione di un confronto tra diverse tecniche regionali di stima applicate alla stessa area di studio utilizzata negli studi sopracitati, con lo scopo di verificare se i punti di forza del procedimento possano essere confermati accoppiando tecniche idrologiche come il GLS con tecniche tipicamente geostatistiche. La struttura dell’analisi si articola in tre fasi procedurali, partendo dal dataset relativo ai 61 bacini appartenenti all’area di studio considerata da Gotvald et al. (2009).

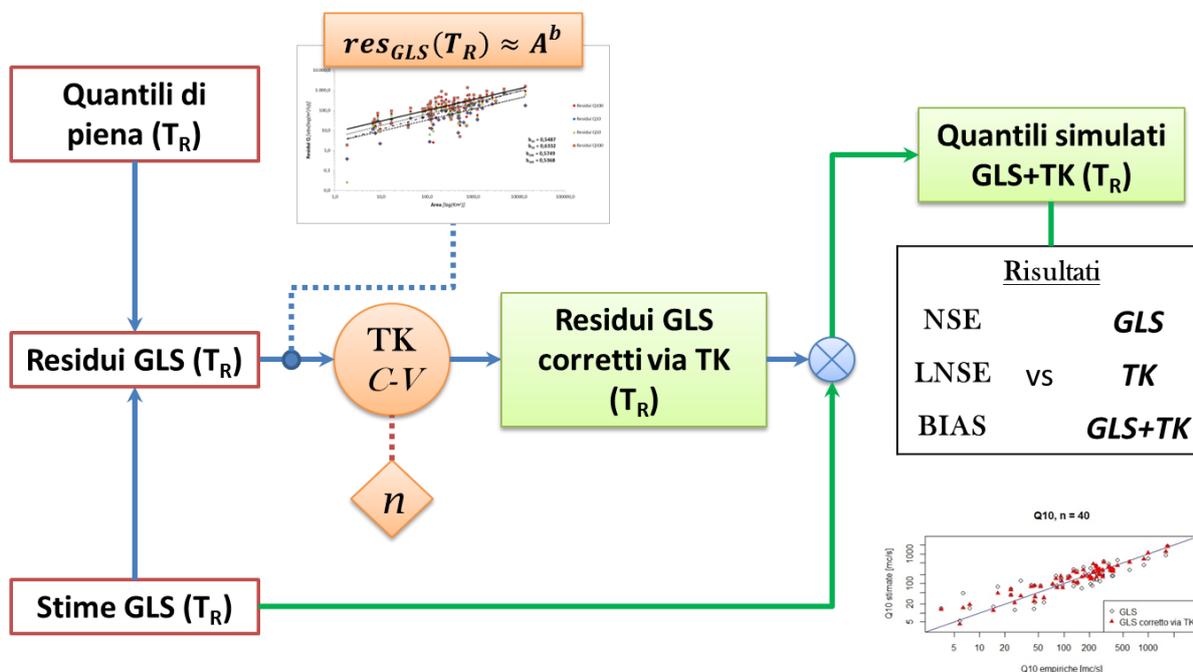


Figura 5.1. Schema della struttura dell’indagine

5.3.1. Modellazione residui GLS

L’indagine qui presentata parte dall’analisi dei *residui* ottenuti dall’applicazione del GLS, ovvero la differenza fra la variabile empirica e quella simulata, ottenuta mediante il modello:

$$res_{GLS}(T_R) = Q_{emp}(T_R) - Q_{GLS}(T_R) \quad (5.1)$$

In cui $Q_{emp}(T_R)$ sono i quantili empirici di partenza associati a tempi di ritorno $T_R = 10, 50, 100, 500$ anni, ciascuno relativo a un bacino ricadente nella regione d’interesse, e $Q_{GLS}(T_R)$ le stime ottenute dal modello multiregressivo ai Minimi Quadrati Generalizzati in cross-validazione (v. par. 5.1). I dati, ottenuti per gentile concessione dall’USGS per scopi didattici e/o di ricerca, sono riferiti spazialmente all’area di studio ed importati mediante un software GIS (*Geographical Information System*).

5.3.2. Regionalizzazione dei residui mediante Top-kriging

Come già accennato, i residui delle stime GLS possono essere considerati, data la loro natura spaziale, una variabile regionalizzata, perciò il passo successivo consiste nell’interpolarli tramite il Top-kriging il quale, lavorando su un supporto areale, induce un effetto regolarizzante sulla variabile regionalizzata e sul variogramma, tanto più evidente quanto più grande è l’estensione del supporto considerato (v. cap. 2).

Questo passaggio è stato eseguito con l’ausilio del software di calcolo numerico **R** (*The R Project for Statistical Computing*, <http://www.r-project.org/>), programma freeware e open-source compatibile con molti sistemi operativi, in grado di fornire una varietà di strumenti di calcolo di tipo statistico e grafico, oltre a mettere a disposizione una vasta gamma di pacchetti aggiuntivi per applicazioni specifiche. Il pacchetto utilizzato per l’interpolazione è denominato *rtop* ed è stato messo a punto da J. O. Skøien ed altri autori (2011). Con questo pacchetto è stato possibile importare in ambiente **R** le informazioni contenute nei file GIS di partenza, le attraverso i quali il programma esegue il calcolo dei variogrammi sperimentale e teorico, ed opera la stima dei residui regionalizzati direttamente in cross-validazione:

questo lo rende un valido strumento di cui avvalersi per eseguire l’interpolazione in tempi brevi, data la complessità delle operazioni.

Archfield et al. (2013) hanno dimostrato che l’interpolazione risulta avere prestazioni migliori adottando valori di portata specifica, piuttosto che utilizzare direttamente i quantili, adottando una scelta analoga per il trattamento dei residui. Il valore di un certo quantile di portata conserva una dipendenza evidente dall’area di drenaggio A , ma lo stesso si può affermare per i residui, i quali sono stati normalizzati rispetto alle rispettive aree. Per ciascuno dei quattro valori del tempo di ritorno, si è ottenuta la seguente relazione di scala:

$$res_{GLS}(T_R) \approx A^b \quad (5.2)$$

Applicando una trasformazione logaritmica alla relazione (5.2), mediante regressione lineare, si sono eseguite diverse prove preliminari calcolando diversi valori di tentativo di b al fine di trovare quello che offrì una migliore modellazione dei residui rispetto alle aree, scegliendo infine un esponente medio pari a 0.57, valore omnicomprensivo per una legge di scala unica adottabile per tutte le serie a diverso tempo di ritorno considerate. I “residui specifici” così trovati sono pertanto resi in questa forma:

$$res'_{GLS}(T_R) = \frac{res_{GLS}(T_R)}{A^{0.57}} \quad (5.3)$$

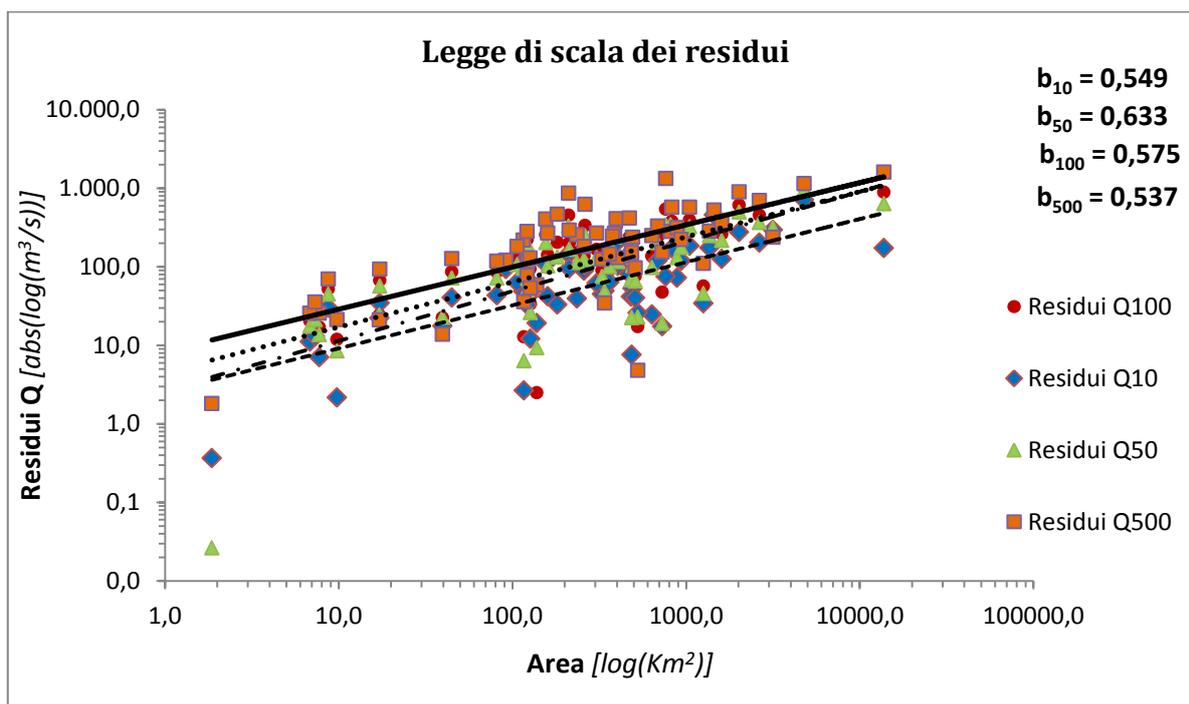


Figura 5.2. Leggi di scala empirica tra i residui dei quantili di piena stimati in cross-validazione attraverso la tecnica GLS (Gotvald et al., 2009) e l’area del bacino corrispondente

Oltre alla ricerca di un valore di b che offrisse una normalizzazione adeguata, i calcoli sono stati reiterati in funzione di valori di vicinaggio diversi rispetto all’impostazione predefinita (10 bacini). Il pacchetto `rtop` consente infatti l’impostazione e la modifica di una varietà di parametri che regolano l’interpolazione, il calcolo dei variogrammi e la risoluzione di discretizzazione del supporto areale, ma anche del vicinaggio, identificato d’ora in avanti con n . Analisi preliminari hanno dimostrato che esiste una certa sensibilità dei risultati al variare del vicinaggio: per questo motivo, nell’indagine si è ritenuto opportuno adottare, a turno, vicinaggi di 10 e 40 bacini, al fine di verificare le prestazioni del modello all’aumentare del vicinaggio. A valle dell’interpolazione, dunque, si sono ottenute due serie di residui in base ai valori di n scelti, per ciascuno dei tempi di ritorno assegnati e in cross-validazione:

$n = 10$	$n = 40$
$res'_{TK}(T_R, 10)$	$res'_{TK}(T_R, 40)$

5.3.3. Accoppiamento GLS-TK

Una volta interpolati con il metodo Top-kriging i residui dei quantili di piena stimati con il metodo GLS, per tempi di ritorno pari a 10, 50, 100 e 500 anni e con diversi valori di vicinaggio, si è passati alla seconda fase del procedimento. Ai residui regionalizzati via TK è stata dapprima ripristinata la dipendenza dall’area, semplicemente invertendo la formula (5.3):

$$res_{TK}(T_R, n) = res'_{TK}(T_R, n) \cdot A^{0.57} \quad (5.4)$$

Questi residui “de-normalizzati”, poi, sono stati sommati ai quantili empirici iniziali ottenuti con il metodo GLS, al fine di individuare e quantificare le differenze fra i quantili modellati e quelli empirici. La somma fornisce sostanzialmente due nuove serie di valori di portata simulata, uno per ciascun valore di vicinaggio scelto per l’interpolazione mediante il Top-kriging, sempre al variare del tempo di ritorno:

$$Q_{sim}(T_R) = Q_{GLS}(T_R) + res_{TK}(T_R, n) \quad (5.5)$$

Si è quindi operata, sostanzialmente, una modulazione dei quantili empirici stimati tramite la regressione con i residui della medesima stima, questi ultimi regionalizzati con una tecnica di tipo geostatistico.

5.3.4. Procedura di Cross-Validazione

Sia la stima dei quantili con il metodo GLS, che l’interpolazione Top-kriging dei residui sono state effettuate in cross-validazione. Questo significa che durante l’analisi è stata operata una validazione del modello per appurare l’accuratezza del procedimento e simulare l’assenza di osservazioni su ciascun bacino (obiettivo dell’analisi). Nel nostro caso, la procedura di cross-validazione, detta *leave-one-out*, o anche *jack-knife*, consiste nella scelta di un sottoinsieme dei bacini a disposizione (“lasciandone fuori uno”), i cui dati vengono utilizzati per effettuare le stime relative alle stazioni escluse dal calcolo. Più dettagliatamente, si può riassumere la procedura di stima dei residui applicata all’area di studio, nei seguenti passi:

- I variogrammi sperimentale e teorico vengono stimati utilizzando i residui ottenuti per tutte le 61 stazioni dell’area di studio;
- Dati 61 valori residui in ciascuna stazione, si esclude la stazione i -esima;
- Impiegando i residui delle 60 stazioni rimaste, si stimano i residui relativi al bacino i -esimo escluso, considerato incognito;
- Per il bacino i -esimo è possibile confrontare residui empirici e stimati dal modello;
- La procedura viene iterata, a turno, per ciascuna stazione dell’area di studio.

Al termine dell’ultima iterazione, i quantili e i residui simulati così ottenuti possono essere confrontati con quelli empirici, attraverso una serie di indici statistici di prestazione, per i quali si rimanda al paragrafo 5.3.5.

Una tale procedura risulta facilmente applicabile all’analisi in oggetto poiché il pacchetto `rtop` in ambiente **R** presenta routine che automatizzano il processo di cross-validazione.

5.3.5. Indici di prestazione

La valutazione dell’efficienza del procedimento è stata eseguita, passo dopo passo, con l’ausilio di due principali indici prestazionali: l’efficienza di Nash-Sutcliffe e l’Errore Medio Relativo.

L’indice di efficienza di Nash e Sutcliffe (1970) viene calcolato a partire dagli n valori simulati P da un certo modello, e i valori osservati O oggetto della simulazione, relativi alla medesima serie temporale, normalizzati con la varianza delle osservazioni:

$$NSE = 1 - \frac{1}{n-1} \frac{\sum_{i=1}^n (O_i - P_i)^2}{\text{var}(O_i)}$$

I valori di questo indice sono compresi nell’intervallo fra 1 e $-\infty$, stando a significare che quanto più è accurato il modello, tanto più NSE sarà vicino a 1; al contrario, avvicinandosi a 0, le prestazioni del modello diventano sempre meno significative; valori negativi indicano infine un peggioramento della capacità predittiva dei valori simulati rispetto a quelli osservati, rendendo di fatto inutile il modello preso in esame.

È stato fatto notare da alcuni autori che l’indice NSE, dipendendo dal quadrato delle differenze fra stime ed osservazioni, presenta una certa sensibilità a valori estremi (v. Legates, McCabe, 1999), in particolare tende a sovrastimare le prestazioni del modello quando si registrano valori di picco nelle portate e, viceversa, a sottostimarle in condizioni di magra (Krause et al., 2005). Per ovviare a questo inconveniente, l’indice NSE viene calcolato in forma logaritmica per attenuare gli effetti dei picchi da un lato, e mantenere più o meno sullo stesso livello i valori più bassi del quantile:

$$\text{LNSE} = 1 - \frac{1}{n-1} \frac{\sum_{i=1}^n (\log O_i - \log P_i)^2}{\text{var}(\log O_i)}$$

L’Errore Medio Relativo, MRE o BIAS, è invece la media del rapporto fra la differenza tra le osservazioni e le predizioni e le osservazioni stesse, relative allo intervallo di tempo:

$$\text{BIAS} = \frac{1}{n} \sum_{i=1}^n \frac{O_i - P_i}{O_i}$$

5.4 Risultati

L’indagine fin qui descritta è stata articolata a partire dai dati sui quantili di portata al colmo relativi a tempi di ritorno di 10, 50, 100 e 500 anni relativi a un campione di 61 bacini rurali degli Stati Uniti sud-occidentali ricadenti prevalentemente nello stato della Georgia e caratterizzati da scarsa o nulla regimentazione dei deflussi idrici, per i quali si è potuto disporre di almeno dieci anni di osservazioni delle portate. Insieme alle portate empiriche, i dati a disposizione comprendono, per ciascuno dei quattro tempi di ritorno, le portate stimate in cross-validazione attraverso il metodo del GLS. I residui di queste ultime stime sono state interpolati mediante Top-kriging, sempre in cross-validazione, per simulare le condizioni di bacini non strumentati. Questo approccio di correzione GLS corretto via Top-kriging è stato applicato per differenti vicinaggi e i risultati rappresentati in grafici di dispersione e resi attraverso gli indici di prestazione menzionati nel paragrafo 5.3.5.

L’indice NSE è stato calcolato inizialmente confrontando i residui stimati ottenuti applicando il Top-kriging ai residui empirici GLS, per ognuno dei quattro tempi di

ritorno e per diversi valori di vicinaggio: lo scopo è stato quello di appurare il grado di efficienza del metodo TK rispetto al metodo GLS, al variare del numero di bacini limitrofi utilizzati per la stima. In seguito, dopo aver sommato le stime così ottenute ai residui GLS di partenza, lo stesso indice è stato ottenuto da questi nuovi valori rispetto ai valori empirici dei quantili di portata al colmo di piena ottenuti in base alle osservazioni di portata disponibili. Quest’ultimo risultato è stato infine confrontato con le prestazioni ottenute dalla sola regressione GLS e con l’ausilio, in aggiunta, del BIAS.

5.4.1 Top-kriging applicato ai residui GLS

Si riportano di seguito i grafici di dispersione risultanti dall’interpolazione via Top-kriging dei residui GLS (che sono sia positivi che negativi) in cross-validazione, seguiti dalle tabelle riassuntive che riportano gli indici di Nash-Sutcliffe e BIAS, per ciascuno dei quattro quantili e dei due valori di vicinaggio (parametro $n = 10$ e $n = 40$ bacini rispettivamente). Nelle figure da 5.1 a 5.3, sono riportati in ordinata i residui dei quantili ottenuti dalla regressione ai Minimi Quadrati Generalizzati, e in ascissa quelli ottenuti modellando i residui via Top-kriging, scalati con le rispettive aree come spiegato nel paragrafo 5.3.2. Il modello risulta tanto più performante quanto più i punti si trovano allineati alla bisettrice degli assi.

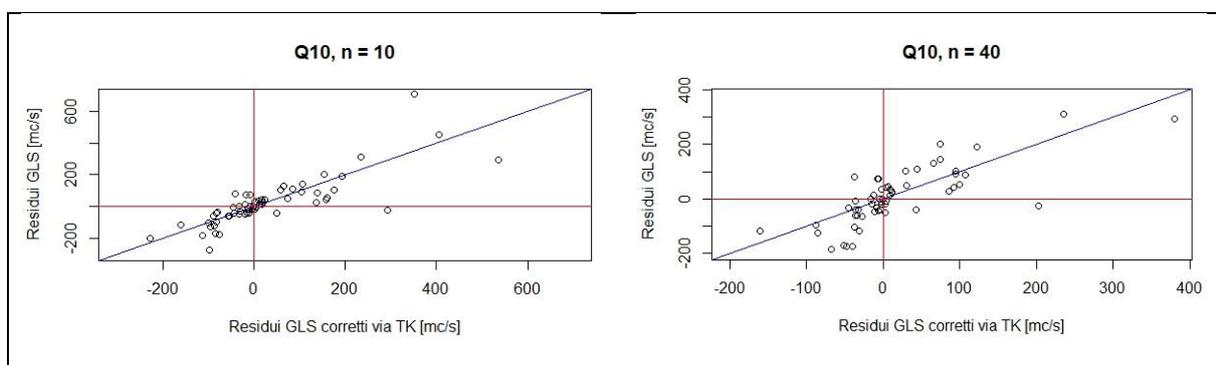


Figure 5.1a e b. Diagrammi di dispersione fra residui empirici GLS e residui GLS corretti via TK. $T_R = 10$ anni

Capitolo 5 – Applicazione di GLS e Top-kriging per la stima della piena di progetto in bacini non strumentati appartenenti all’area di studio

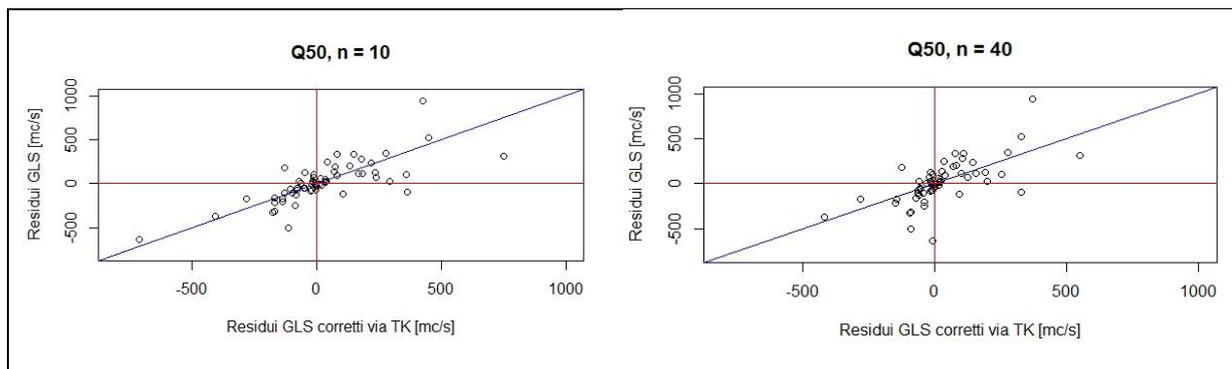


Figure 5.2a e b. Diagrammi di dispersione fra residui empirici GLS e residui GLS corretti via TK. $T_R = 50$ anni

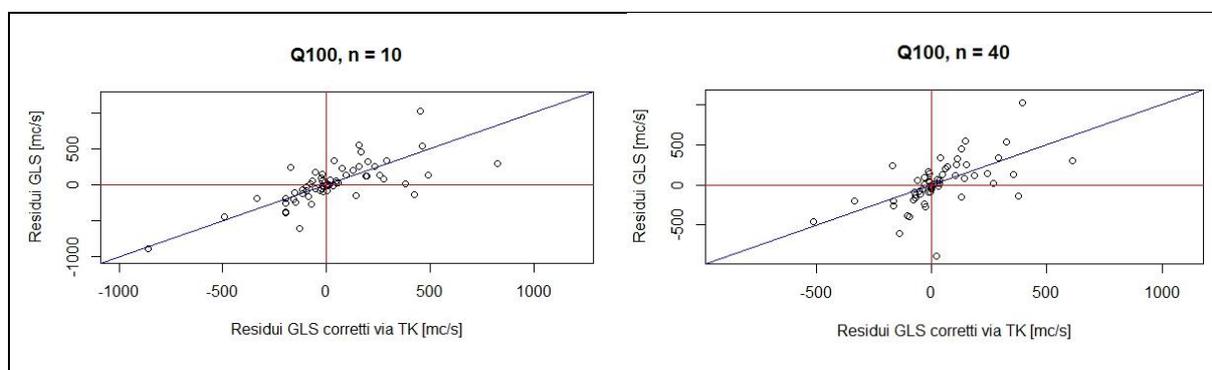


Figure 5.3a e b. Diagrammi di dispersione fra residui empirici GLS e residui GLS corretti via TK. $T_R = 100$ anni

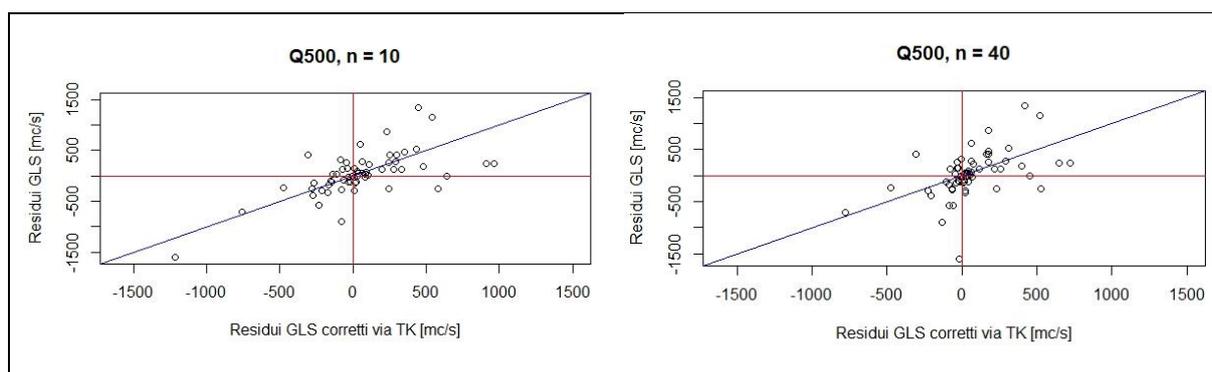


Figure 5.4a e b. Diagrammi di dispersione fra residui empirici GLS e residui GLS corretti via TK. $T_R = 500$ anni

Tabella 5.1. Indici NSE e BIAS relativi all’interpolazione dei residui GLS corretti via TK

Quantili	<i>n = 10</i>		<i>n = 40</i>	
	NSE	BIAS	NSE	BIAS
Q_{10}	0,5881	0,0940	0,6457	0,5296
Q_{50}	0,5432	1,3922	0,4301	-0,6258
Q_{100}	0,4986	-0,1184	0,3424	0,2775
Q_{500}	0,3977	-1,7569	0,2232	-0,9045

Come si può notare dalla tabella 5.1, l’interpolazione dei residui empirici GLS mediante l’applicazione del modello TK mostra risultati simili al variare del vicinaggio, confrontabili tra loro per entrambi gli indici di prestazione e, com’era prevedibile, evidenziando un andamento decrescente all’aumentare del tempo di ritorno considerato.

5.4.2 Quantili empirici GLS corretti via TK

Nelle figure seguenti si riportano i diagrammi scatterplot fra i quantili simulati $Q_{sim}(T_R)$ attraverso la combinazione delle due tecniche, risultati dalla somma fra le portate stimate tramite GLS ed i residui ottenuti attraverso il TK (eq. 5.4), e le portate empiriche di partenza $Q^{emp}(T_R)$ (eq. 5.1), per ciascuno dei tempi di ritorno e delle scelte operate sui vicinaggi. Nei grafici, i cerchi rappresentano i quantili GLS in funzione delle portate empiriche, mentre con i triangoli rossi sono identificati i valori simulati (v. eq. 5.4).

Capitolo 5 – Applicazione di GLS e Top-kriging per la stima della piena di progetto in bacini non strumentati appartenenti all’area di studio

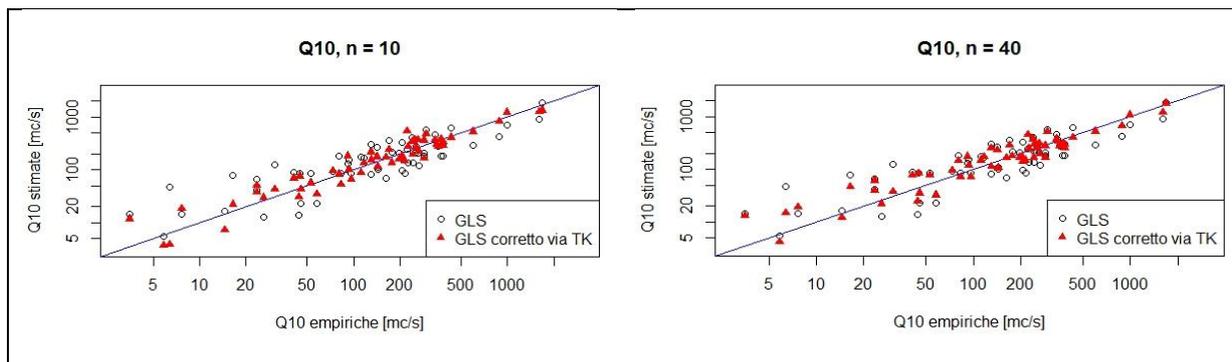


Figure 5.5a e b. Diagrammi di dispersione fra quantili simulati con il metodo GLS e GLS via TK ed empirici. $T_R = 10$ anni

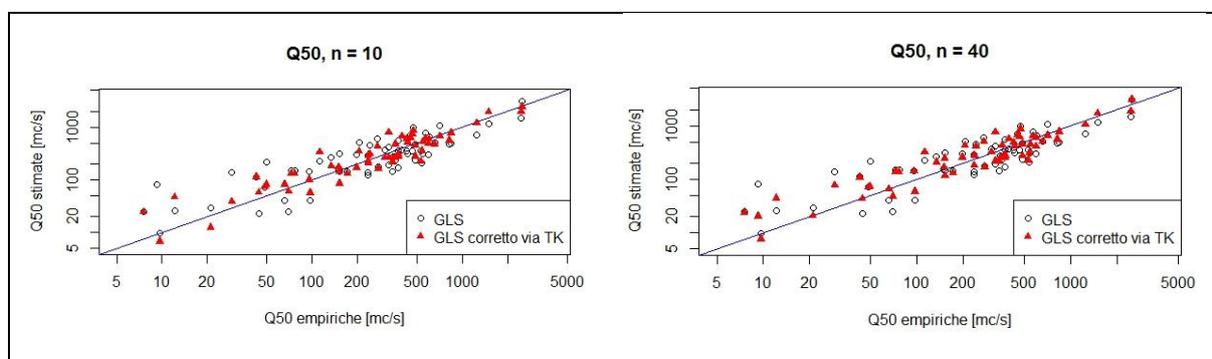


Figure 5.6a e b. Diagrammi di dispersione fra quantili simulati con il metodo GLS e GLS via TK ed empirici. $T_R = 50$ anni

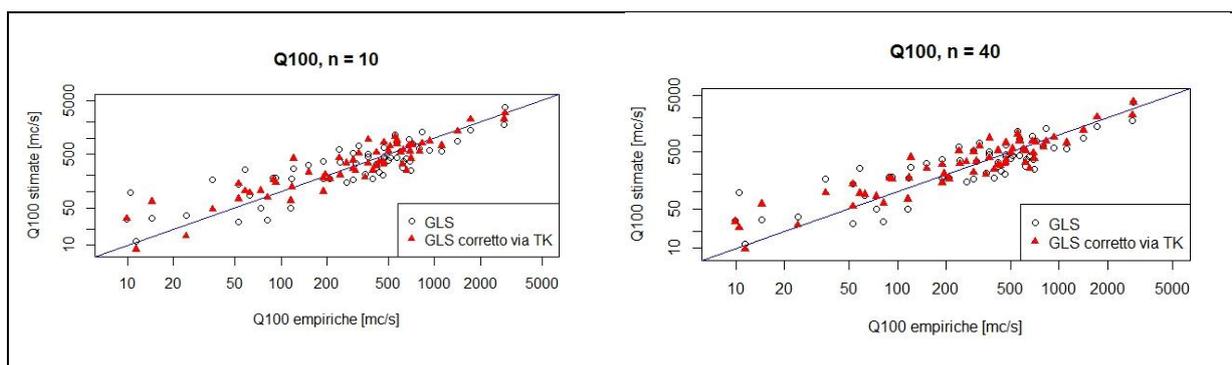


Figure 5.7a e b. Diagrammi di dispersione fra quantili simulati con il metodo GLS e GLS via TK ed empirici. $T_R = 100$ anni

Capitolo 5 – Applicazione di GLS e Top-kriging per la stima della piena di progetto in bacini non strumentati appartenenti all’area di studio

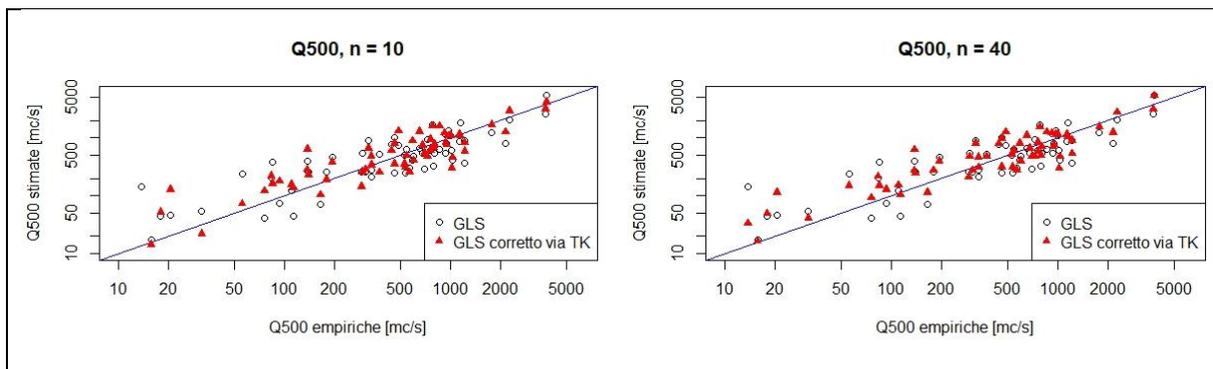


Figure 5.8a e b. Diagrammi di dispersione fra quantili simulati con il metodo GLS e GLS via TK ed empirici. $T_R = 500$ anni

Nella seconda serie di grafici, si riportano in ascissa gli scarti relativi in valore assoluto tra i quantili stimati con il metodo proposto e i valori empirici, e in ordinata gli scarti relativi in valore assoluto tra i quantili stimati con il GLS e i valori empirici. Una tale rappresentazione mette in evidenza la distribuzione dell’errore rispetto ai due metodi di stima. La linea bisettrice indica l’equivalenza in termini di errore tra le due metodologie, mentre un valore che cade nella parte inferiore o superiore del grafico rispetto alla diagonale indicare, rispettivamente, una maggiore o minore propensione all’errore del metodo GLS tal quale, o dell’approccio accoppiato GLS+TK. Per ciascuna delle figure qui sotto rappresentate, quindi, è possibile apprezzare quanto ci si discosti dal valore nullo, e quanti valori indichino una maggiore imprecisione fra i due metodi considerati.

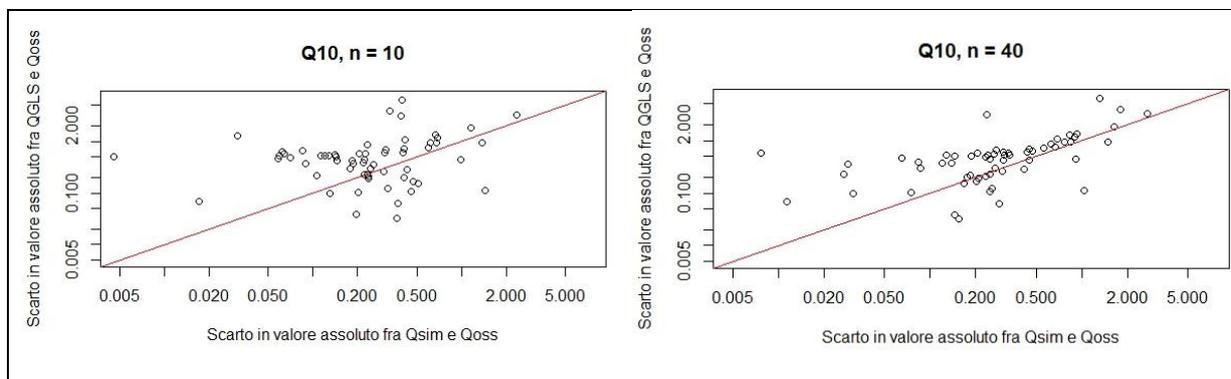


Figure 5.10a e b. Confronto dell’errore in valore assoluto fra quantili simulati con il metodo GLS e GLS via TK ed empirici. $T_R = 10$ anni

Capitolo 5 – Applicazione di GLS e Top-kriging per la stima della piena di progetto in bacini non strumentati appartenenti all’area di studio

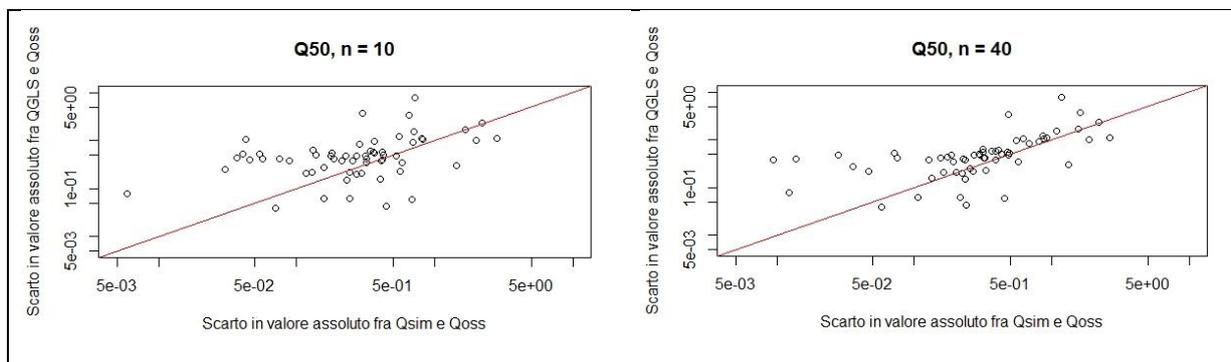


Figure 5.9a e b. Confronto dell'errore in valore assoluto fra quantili simulati con il metodo GLS e GLS via TK ed empirici. $T_R = 50$ anni

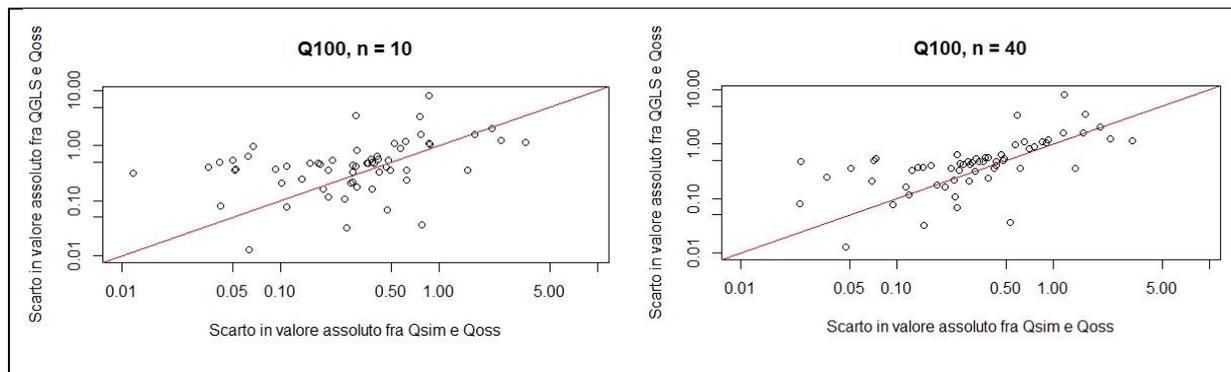


Figure 5.11a e b. Confronto dell'errore in valore assoluto fra quantili simulati con il metodo GLS e GLS via TK ed empirici. $T_R = 100$ anni

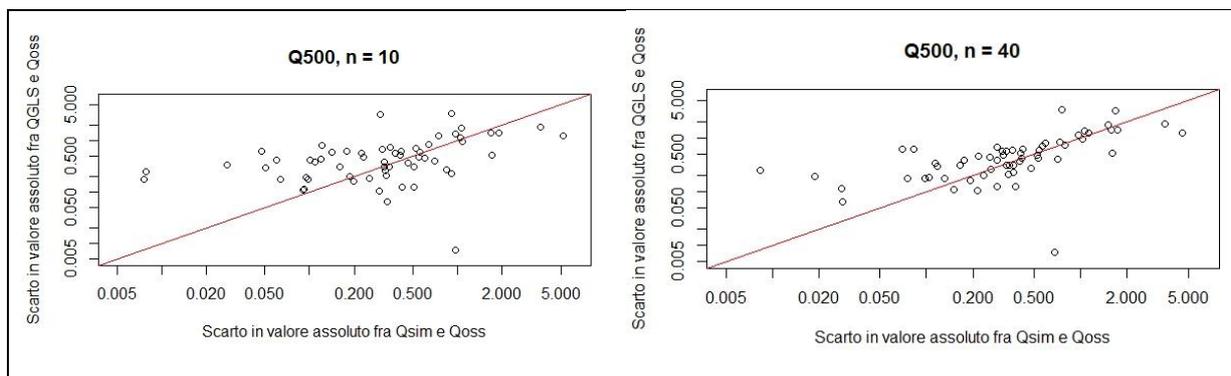


Figure 5.12a e b. Confronto dell'errore in valore assoluto fra quantili simulati con il metodo GLS e GLS via TK ed empirici. $T_R = 500$ anni

Infine, la tabella 5.2 riassume le prestazioni di questo procedimento, mettendo a confronto gli indici NSE, LNSE e BIAS ottenuti dai quantili GLS corretti via TK, rispetto alle portate GLS tal quali di partenza e ai quantili stimati direttamente tramite Top-kriging (Archfield et al., 2013). Si ricorda che gli indici NSE e LNSE denotano prestazioni migliori se più vicini a 1, mentre per gli indici BIAS migliori prestazioni si hanno quando i valori si avvicinano maggiormente a 0.

Tabella 5.2a, b e c. Prestazioni in termini di indici NSE, LNSE ed Errore Medio Relativo del metodo GLS+TK rispetto al solo GLS e al TK puro.

Quantili	Prestazioni GLS			Prestazioni TK puro		
	NSE	LNSE	BIAS	NSE	LNSE	BIAS
Q ₁₀	0,7666	0,7312	-0,3681	0,9330	0,9020	0,1350
Q ₅₀	0,7566	0,7297	-0,3890	0,8982	0,8766	0,1674
Q ₁₀₀	0,7394	0,7266	-0,3930	0,8774	0,8628	0,1765
Q ₅₀₀	0,6713	0,7158	-0,3941	0,8262	0,8207	0,2113

Quantili	Prestazioni GLS+TK (n = 10)			Prestazioni GLS+TK (n = 40)		
	NSE	LNSE	BIAS	NSE	LNSE	BIAS
Q ₁₀	0,9054	0,9096	-0,1448	0,9186	0,8816	-0,2315
Q ₅₀	0,8894	0,8582	-0,1969	0,8620	0,8571	-0,2681
Q ₁₀₀	0,8698	0,8164	-0,2206	0,8293	0,8434	-0,2889
Q ₅₀₀	0,8028	0,7813*	-0,3020	0,7457	0,7967	-0,3769

5.5 Discussione

Come si evince dai risultati riportati in tabella 5.2, sia gli indici NSE che gli errori medi relativi risultanti dall’accoppiamento delle stime GLS con i residui modellati via Top-kriging presentano notevoli miglioramenti rispetto all’applicazione del solo metodo di regressione. Scegliendo un vicinaggio standard di 10 bacini, l’indice NSE può superare lo 0.9, come nel caso della portata decennale, anche se rimane al di sopra dello 0.8 per tutti gli altri quantili e, rispetto alle stime operate con il metodo GLS,

comporta un miglioramento di circa il 12%. Questo margine tocca il 14% estendendo il vicinaggio a 40 bacini, caso in cui l'indice di efficienza si mantiene maggiore o uguale a 0.75, laddove le stime della regressione GLS si mantengono in un intervallo di NSE compreso fra 0.77 e 0.67.

L'indice BIAS mostra parimenti un notevole miglioramento e anche in questo caso, i risultati più evidenti si riscontrano nei quantili associati a un tempo di ritorno di 10 anni, rimanendo comunque più basso in valore assoluto degli errori riscontrati con la sola regressione, anche nell'introduzione di un vicinaggio maggiore e con l'aumento dei tempi di ritorno.

In tabella 5.2, l'indice LNSE relativo al quantile di portata Q500 calcolato con il metodo GLS+TK è stato ottenuto eliminando dal vettore dei residui un elemento (il n° 46) che altrimenti avrebbe determinato una portata negativa e impedito il calcolo dell'indice di prestazione.

Conclusioni

L'indagine condotta nel presente lavoro di Tesi ha avuto l'obiettivo di analizzare e mettere a confronto l'efficacia di due tecniche per la stima di variabili idrologiche, come la portata al colmo di piena, in bacini idrografici non strumentati: i Minimi Quadrati Generalizzati (GLS) e il Topological kriging (Top-kriging o TK). Tali tecniche sono state applicate a differenti serie di quantili di piena al colmo con tempo di ritorno 10, 50, 100 e 500 anni relativi a un'area di studio comprendente 61 bacini degli Stati Uniti sud-orientali.

Il metodo GLS, basato su una regressione lineare ai minimi quadrati, mette in relazione, per ciascun bacino imbrifero, l'informazione idrologica con informazioni geomorfoclimatiche tenendo conto della mutua correlazione spaziale tra i quantili di piena. Il Top-kriging, invece, è una tecnica di stima di tipo geostatistico che tiene conto del supporto areale della variabile idrologica, coincidente con la superficie dei bacini, della rispettiva posizione geografica e della loro eventuale struttura annidata. L'applicazione ha riguardato in particolare l'accoppiamento di questi due metodi, operato attraverso l'interpolazione mediante il Top-kriging dei residui delle stime GLS, in cross-validazione e per due diversi vicinaggi. La stima è stata eseguita, inoltre, tenendo conto dell'esistenza di una legge di scala fra residui GLS e area del bacino. Successivamente, sono stati calcolati i quantili empirici sommando le stime GLS con i residui cross-validati ottenuti dall'applicazione del TK ai residui GLS. Le prestazioni di questa procedura sono state valutate e confrontate con quelle dei due metodi applicati separatamente per la stessa area di studio e lo stesso dataset. La comparazione quantitativa sulle performance si è valutata attraverso il calcolo di 3 indici statistici: efficienza di Nash-Sutcliffe (NSE), la stessa efficienza applicando una trasformazione logaritmica alle serie confrontate (Log-NSE o LNSE) e l'errore medio relativo.

La modellazione dei residui GLS mediante Top-kriging risulta performante nonostante sia evidente la presenza di un trend decrescente dell'efficienza NSE, LNSE all'aumentare del tempo di ritorno per entrambe le scelte di vicinaggio (10 e 40 bacini). Tuttavia, l'accoppiamento delle due tecniche presenta notevoli miglioramenti sia in termini di NSE che di BIAS se si considera il metodo GLS applicato indipendentemente, e ottenendo risultati del tutto paragonabili a quelli conseguiti con la stima dei quantili direttamente via Top-kriging.

L'indagine, in conclusione, ha evidenziato che l'accoppiamento di tecniche di regressione come il GLS con metodi geostatistici innovativi, come ad es. il Top-kriging, può effettivamente incrementare la capacità predittiva di un modello di stima regionale confermando, altresì, che le applicazioni geostatistiche in campo idrologico risultano un approccio affidabile nei problemi di stima di quantili di piena al colmo in bacini non strumentati.

Bibliografia

- Archfield, S. A., Pugliese, A., Castellarin, A., Skøien, J. O., & Kiang, J. E. (2013). Topological and canonical kriging for design flood prediction in ungauged catchments: an improvement over a traditional regional regression approach? *Hydrology and Earth System Sciences*, *17*(4), 1575–1588. doi:10.5194/hess-17-1575-2013
- Brath, A., & Franchini, M. (1998). La valutazione regionale del rischio di piena con il metodo della portata indice. *CNR-GNDCI*.
- Gotvald, A. J., Feaster, T. D., & Weaver, J. C. (2009). *Magnitude and Frequency of Rural Floods in the Southeastern United States* (Vol. 1).
- Griffis, V. W., & Stedinger, J. R. (2007). The use of GLS regression in regional hydrologic analyses. *Journal of Hydrology*, *344*(1-2), 82–95. doi:10.1016/j.jhydrol.2007.06.023
- Krause, P., & Boyle, D. P. (2005). Advances in Geosciences Comparison of different efficiency criteria for hydrological model assessment, 89–97.
- Legates, D., & McCabe, G. (1999). Evaluating the use of “goodness of fit” measures in hydrologic and hydroclimatic model validation. *Water Resources Research*, *35*(1), 233–241
- Martins, E. S., & Stedinger, J. R. (2002). Cross correlations among estimators of shape. *Water Resources Research*, *38*(11), 34–1–34–7. doi:10.1029/2002WR001589
- Ouarda, T. B. M. J., Girard, C., Cavadias, G. S., & Bobée, B. (2001). Regional flood frequency estimation with canonical correlation analysis. *Journal of Hydrology*, *254*(1-4), 157–173. doi:10.1016/S0022-1694(01)00488-7
- Pugliese, A., Castellarin, A., Montanari, A., & Archfield, S. (2012). *TECNICHE DI INTERPOLAZIONE GEOSTATISTICA PER LA STIMA DELLA PIENA DI PROGETTO IN BACINI NON STRUMENTATI*.
- Raspa, G, Bruno, R. (n.d.). Dispense di geostatistica applicata (pp. 1–96). Retrieved from <http://w3.uniroma1.it/geostatistica/Geostatistica/>

Skøien, J., Merz, R., & Blöschl, G. (2006). Top-kriging--geostatistics on stream networks. *Hydrology & Earth System ...*, 277–287.

Skøien, J. O., & Laaha, G. (2013). Rtop-an R package for interpolation of data with a variable spatial support-examples from river networks. *EGU General ...*, (1993), 1–14

Ringraziamenti

Per primo non posso che ringraziare il prof. Castellarin, che in questi mesi ha sempre trovato il tempo di seguire il mio lavoro, nonostante i numerosi impegni. Se oggi adoro questa materia e in generale le discipline che riguardano le risorse idriche, lo devo a persone come lui e il prof. Montanari (di cui mi mancheranno le videolezioni). Subito dopo, il mio più sincero grazie va ad Alessio Pugliese, con cui ho scritto questa Tesi e senza il quale non ce l'avrei mai fatta. Il suo supporto, i suoi consigli e la sua infinita pazienza sono stati fondamentali, soprattutto durante le ultime battute, quando il tempo stringeva e la stanchezza era tanta. Gli auguro ogni bene.

Special thanks vanno infine all'altra mia correlatrice, Stacey Archfield, che non ho ancora avuto il piacere di conoscere personalmente.

Lavorare con queste persone è stato faticoso, entusiasmante, ma soprattutto istruttivo.

Mi accingo ancora una volta a scrivere la parola *Fine* a un pezzo di vita, e voltare pagina su un nuovo capitolo già iniziato. Questi due anni e mezzo a Bologna hanno cambiato radicalmente la mia vita e il mio modo di vedere e pensare molte cose. Da quando mi sono lasciato Ancona alle spalle mi sembra sia passata un'eternità. Ho incontrato molte persone, lasciatene indietro molte altre, fatto tante scoperte e considerazioni, sono cresciuto. Per questo ringrazio Bologna, che sarà per sempre la mia città, quella in cui sono rinato, insieme a Napoli, dove invece sono nato.

Il primo grazie va alla mia famiglia, che ha vissuto in prima linea e condiviso con me insuccessi e soddisfazioni: i miei genitori, mio fratello e le mie dolci sorelline, e l'onnipresente zia Ester. Non posso non ringraziare i miei compagni di corso e amici Simone 'gneddu' Pinna Nossai, Annalisa 'capo' Gionni, Elisa 'arbitro' Grandi, Rob e Viviana, che mi hanno accompagnato durante questo cammino.

Un grazie va ai miei ex coinquilini e ai miei amici più stretti.

Dulcis in fundo, ringrazio la mia amata Anna, senza il cui sostegno non sarei qui a scrivere.

Un pensiero a chi continua a vigilare sul mio sentiero. Non Vi dimentico mai.