

Politecnico di Torino

Corso di Laurea Magistrale in Ingegneria Matematica



**Politecnico
di Torino**

Modellazione Spazio-Temporale Bayesiana delle Precipitazioni Estreme in Piemonte

Relatore:

Prof. Gianluca MASTRANTONIO

Candidato:

Matilde CARNEVALE

Sessione di Laurea Luglio 2025

A.a. 2024/2025

Abstract

Questa tesi si propone di analizzare gli eventi di precipitazioni massime in Piemonte attraverso un approccio modellistico spazio-temporale. Il dataset considerato comprende le precipitazioni massime annuali registrate presso diverse stazioni meteorologiche nel periodo 1928–2020. L’analisi si basa sull’implementazione di due modelli gerarchici bayesiani mediante il pacchetto R-INLA. Il primo modello adotta la famiglia di distribuzioni bGEV (blended Generalized Extreme Value), proposta come alternativa flessibile alla classica GEV per la modellazione dei massimi annuali. Il secondo segue un approccio a soglia, in cui si modellano le eccedenze oltre un livello elevato tramite una distribuzione Generalized Pareto. In entrambi i casi, è stata sviluppata una struttura spazio-temporale in cui il campo latente è modellato tramite l’approccio SPDE (Stochastic Partial Differential Equation) e una dinamica temporale di tipo AR(1). L’obiettivo è fornire una rappresentazione flessibile e coerente della variabilità spazio-temporale delle precipitazioni estreme, utile per comprendere meglio i pattern di rischio idro-meteorologico sul territorio regionale.

Indice

Elenco delle figure	v
1 Background	1
1.1 Fondamenti teorici dell'approccio INLA	1
1.1.1 Introduzione alla Statistica Bayesiana	1
1.1.2 INLA: nuovo approccio all'inferenza bayesiana	2
1.1.3 Latent Gaussian Models	3
1.1.4 Modelli Additivi	4
1.1.5 Approssimazione di Laplace	5
1.1.6 Integrated Nested Laplace Approximation	7
1.1.7 Stochastic Partial Differential Equation	9
1.1.8 Panoramica su R-INLA	11
1.1.9 Costruzione della mesh	13
1.1.10 Variogramma	14
1.2 Fondamenti Teorici Classici della Teoria dei Valori Estremi	15
1.2.1 La Distribuzione GEV	17
1.2.2 Una riformulazione alternativa: la distribuzione bGEV	17
1.2.3 Threshold Models e Distribuzione di Pareto Generalizzata (GPD)	20
2 Specifica del modello spazio-temporale	22
2.1 Dati	23
2.1.1 Statistiche descrittive sui Dataset	24

2.1.2	Preprocessing dei dati	26
2.2	Costruzione della mesh e specifica del campo spazio-temporale . . .	28
2.3	Predittore lineare e costruzione dello stack	31
2.4	Modello dei massimi di blocco con distribuzione bGEV	32
2.5	Modello Soglia con distribuzione di Pareto	34
3	Risultati	39
3.1	Stima dei parametri del Modello bGEV	39
3.2	Stima dei parametri del Modello GP	45
4	Conclusioni	58
	Bibliografia	60

Elenco delle figure

1.1	Esempio grafico di variogramma empirico [3].	15
2.1	Andamento temporale delle stazioni pluviometriche attive nel periodo 1928-2020.	25
2.2	Variogrammi empirici annuali (punti) e media pesata smussata con LOESS (linea) - Dataset10min.	26
2.3	Variogrammi empirici annuali (punti) e media pesata smussata con LOESS (linea) - Dataset5gg.	26
2.4	Distribuzione per anno, Dataset2gg	28
2.5	Mesh triangolare utilizzata per la discretizzazione del campo spaziale tramite SPDE, ottenuta sulla base delle stazioni disponibili.	30
2.6	Schema dell'algoritmo in R-INLA.	38
3.1	Componente spazio-temporale media stimata dal modello bGEV per il dataset con durata 10 minuti, dal 2012 al 2020. Ogni pannello mostra il valore atteso del campo latente per un anno specifico, evidenziando l'evoluzione spaziale dell'effetto.	46
3.2	Deviazione standard a posteriori del campo latente stimato dal modello bGEV per il dataset con durata 10 minuti, dal 2012 al 2020. I valori rappresentano l'incertezza associata alla componente spaziale del modello in ciascun anno.	47

3.3	Componente spazio-temporale media stimata dal modello bGEV per il dataset con durata 1 ora, dal 2012 al 2020. Ogni pannello mostra il valore atteso del campo latente per un anno specifico, evidenziando l'evoluzione spaziale dell'effetto.	48
3.4	Deviazione standard a posteriori del campo latente stimato dal modello bGEV per la durata di 1 ora, dal 2012 al 2020. I valori rappresentano l'incertezza associata alla stima della componente spaziale del modello in ciascun anno.	49
3.5	Componente spazio-temporale media stimata dal modello GP per il dataset con durata 10 minuti, dal 2012 al 2020. Ogni pannello mostra il valore atteso del campo latente per un anno specifico, evidenziando l'evoluzione spaziale dell'effetto.	54
3.6	Componente spazio-temporale media stimata dal modello GP per il dataset con durata 1 ora, dal 2012 al 2020. Ogni pannello mostra il valore atteso del campo latente per un anno specifico, evidenziando la persistenza temporale e l'evoluzione spaziale dell'effetto.	55
3.7	Varianza del campo latente stimato dal modello GP per il dataset con durata 1 ora, dal 2012 al 2020. I valori rappresentano l'incertezza associata alla componente spaziale del modello in ciascun anno.	56
3.8	Varianza del campo latente stimato dal modello GP per il dataset con durata 1 ora, dal 2012 al 2020. I valori rappresentano l'incertezza associata alla componente spaziale del modello in ciascun anno.	57

Capitolo 1

Background

1.1 Fondamenti teorici dell'approccio INLA

In questo capitolo vengono introdotti i principali concetti teorici e computazionali alla base dell'approccio INLA, utilizzato per effettuare inferenza bayesiana in modelli gerarchici latenti. Si introducono inoltre l'approccio SPDE, utile per descrivere fenomeni con struttura spaziale o spazio-temporale, e il pacchetto **R-INLA**, che permette di applicare questi metodi in modo pratico all'interno dell'ambiente R.

1.1.1 Introduzione alla Statistica Bayesiana

La statistica bayesiana rappresenta un'alternativa ai metodi inferenziali tradizionali di tipo frequentista, offrendo una prospettiva probabilistica che include anche l'incertezza sui parametri del modello. Ciò che la distingue maggiormente è la possibilità di esprimere e aggiornare in modo coerente le proprie conoscenze (o incertezze) su tutti gli elementi del modello, parametri inclusi, man mano che si raccolgono nuovi dati. Questo aggiornamento delle informazioni avviene grazie alla regola di Bayes, che costituisce il principio fondamentale dell'approccio bayesiano. Il processo inferenziale bayesiano si basa su tre elementi fondamentali, che insieme permettono di aggiornare le nostre conoscenze sui parametri del modello alla luce dei dati osservati:

- *Distribuzione a priori*: Prima ancora di osservare i dati, si parte da un'idea, una convinzione o un'ipotesi sui possibili valori dei parametri del modello. Questa informazione iniziale viene formalizzata attraverso la distribuzione a priori, che può riflettere esperienze precedenti, risultati di studi simili, oppure essere volutamente non informativa, quando si preferisce lasciare che siano i dati a determinare la forma della distribuzione.
- *Funzione di verosimiglianza*: La verosimiglianza (o **likelihood**) misura quanto i dati osservati siano compatibili con un certo insieme di valori dei parametri. In altre parole, indica quanto "plausibili" appaiono i dati, dato un certo modello, e rappresenta il collegamento chiave tra la teoria statistica e l'evidenza empirica.
- *Distribuzione a posteriori*: Una volta combinata la distribuzione a priori $p(\boldsymbol{\theta})$ con la verosimiglianza $p(\mathbf{y} \mid \boldsymbol{\theta})$ tramite il teorema di Bayes, si ottiene la distribuzione a posteriori (Formula 1.1). Questa rappresenta il risultato finale del processo inferenziale: una sintesi aggiornata delle conoscenze sui parametri dopo aver tenuto conto delle informazioni fornite dai dati.

$$p(\boldsymbol{\theta} \mid \mathbf{y}) = \frac{p(\mathbf{y} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta})}{\int p(\mathbf{y} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta}) d\boldsymbol{\theta}} \propto p(\mathbf{y} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta}) \quad (1.1)$$

Nel denominatore compare un integrale calcolato su tutto lo spazio dei parametri: è la cosiddetta costante di normalizzazione, che garantisce che la distribuzione a posteriori sia una vera densità di probabilità, cioè che l'integrale (o la somma) delle probabilità totali sia pari a uno [1].

1.1.2 INLA: nuovo approccio all'inferenza bayesiana

Uno dei principali ostacoli nell'applicazione pratica dell'inferenza bayesiana è il calcolo della distribuzione a posteriori, che può diventare particolarmente complicato quando il numero di parametri è elevato. Per affrontare questo problema, si ricorre spesso a metodi basati sul campionamento, come gli algoritmi *Markov chain Monte Carlo* (MCMC). Questi strumenti, sebbene molto versatili e diffusi, richiedono notevoli risorse computazionali e possono risultare difficili da gestire.

In questo contesto si inserisce l'approccio dell'*Integrated Nested Laplace Approximation* (INLA), proposto da Rue, Martino e Chopin (2009) [2], che costituisce una valida alternativa deterministica ai metodi di campionamento. Questo metodo si fonda su due idee principali: innanzitutto, invece di stimare l'intera distribuzione a posteriori congiunta dei parametri, INLA si concentra sul calcolo efficiente delle distribuzioni marginali a posteriori, che nella maggior parte dei casi sono sufficienti per l'inferenza; in secondo luogo, l'approccio è specificamente pensato per una classe di modelli noti come *Latent Gaussian Models* (LGM), in cui il campo latente può essere rappresentato come un *Gaussian Markov Random Field* (GMRF). Quest'ultima scelta consente di sfruttare la struttura sparsa della matrice di precisione associata, con notevoli vantaggi in termini di efficienza computazionale [3].

1.1.3 Latent Gaussian Models

Il modello LGM utilizzato da INLA è organizzato secondo una struttura gerarchica a tre livelli [4] presentata nella Formula 1.2,

$$\begin{aligned}
 \mathbf{y} \mid \mathbf{x}, \boldsymbol{\theta}_1 &\sim \prod_{i \in \mathcal{I}} \pi(y_i \mid x_i, \boldsymbol{\theta}_1), && \text{(Likelihood)} \\
 \mathbf{x} \mid \boldsymbol{\theta}_2 &\sim \mathcal{N}(\boldsymbol{\mu}(\boldsymbol{\theta}_2), \mathbf{Q}(\boldsymbol{\theta}_2)^{-1}), && \text{(Latent Gaussian Field)} \\
 \boldsymbol{\theta} &\sim \pi(\boldsymbol{\theta}), && \text{(Prior su } \boldsymbol{\theta})
 \end{aligned} \tag{1.2}$$

dove:

- \mathbf{x} è il Latent Gaussian Field (LGF), cioè l'insieme dei componenti non direttamente osservabili del modello;
- $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ è il vettore degli iperparametri, che controllano aspetti strutturali del modello, come varianze, scale spaziali o autocorrelazioni;
- \mathbf{y} rappresenta i dati osservati e sono assunti mutuamente indipendenti da \mathbf{x} e dagli iperparametri. Questo implica che ciascuna osservazione y_i dipende solo da una componente del campo latente;
- $\boldsymbol{\mu}(\boldsymbol{\theta}_2)$ è la media del campo latente;
- $\mathbf{Q}(\boldsymbol{\theta}_2)$ è la matrice di precisione (l'inverso della matrice di covarianza) di \mathbf{x} .

In pratica, questa struttura gerarchica riflette il modo in cui si costruisce il modello: partendo dagli iperparametri $\boldsymbol{\theta}$, si definisce un campo latente \mathbf{x} che incorpora tutti gli effetti non osservabili e infine si modellano i dati osservati \mathbf{y} in funzione di questo campo latente.

Un'assunzione fondamentale è che \mathbf{x} deve essere rappresentato come un GMRF quando le dimensioni sono elevate. In altre parole, \mathbf{x} è un campo gaussiano in cui le dipendenze tra le variabili sono limitate a quelle "vicine", e questa struttura locale è codificata nella matrice di precisione $\mathbf{Q}(\boldsymbol{\theta}_2)$. Questa scelta non è casuale: usare un GMRF come prior per il campo latente è ciò che permette a INLA di essere così efficiente dal punto di vista computazionale. Infatti, grazie al fatto che \mathbf{Q} è una matrice sparsa, le operazioni numeriche richieste per l'approssimazione delle distribuzioni posteriori possono essere eseguite in modo molto più veloce e stabile, anche quando si lavora con modelli complessi o con grandi quantità di dati.

1.1.4 Modelli Additivi

Nel contesto dell'analisi statistica, uno degli obiettivi più comuni è capire come un insieme di covariate influenzi una certa caratteristica dei dati osservati, ad esempio il valore atteso della variabile risposta. Quando il fenomeno analizzato mostra una dipendenza spaziale o spazio-temporale, è fondamentale che il modello riesca a catturare anche queste correlazioni [5]. Un approccio efficace per affrontare questo tipo di problemi consiste nella costruzione di un *predittore lineare additivo*, espresso su una scala trasformata tramite una funzione di link appropriata (ad esempio, logit per dati binari o la funzione logaritmica per dati di conteggio). Il valore atteso μ_i della variabile risposta è quindi legato al predittore η_i tramite la funzione inversa mostrata nella formula 1.3.

$$\mu_i = g^{-1}(\eta_i) \tag{1.3}$$

La forma generale del predittore lineare [3] è mostrata nella Formula 1.4

$$\eta_i = \alpha + \sum_{k=1}^{n_\beta} \beta_k z_{ki} + \sum_{j=1}^{n_f} f^{(j)}(w_{ji}) \tag{1.4}$$

dove α rappresenta l'intercetta, mentre z_{ki} denotano le covariate fisse (come età, temperatura, altitudine, ecc.), associate ai rispettivi coefficienti β_k .

La differenza principale tra questa formulazione e un modello lineare generalizzato classico risiede nei termini $f^{(j)}$, i quali rappresentano componenti aggiuntive del modello utilizzate per descrivere effetti strutturati o non strutturati che non possono essere adeguatamente rappresentati tramite semplici covariate fisse. Questi termini possono includere, ad esempio, effetti spaziali, temporali, e vengono modellati come realizzazioni di processi gaussiani. Si assume che tali componenti siano a priori indipendenti tra loro e che gli effetti fissi (intercetta e coefficienti β_k) abbiano distribuzioni gaussiane a priori e siano indipendenti dalle componenti di modello [4].

Questa formulazione conduce naturalmente alla definizione di un campo latente gaussiano \mathbf{x} che raccoglie al suo interno tutte le componenti del modello 1.5:

$$\mathbf{x} = (\eta, \alpha, \beta, \mathbf{f}^{(1)}, \mathbf{f}^{(2)}, \dots). \quad (1.5)$$

In questo modo, l'intero campo latente ha una distribuzione congiunta gaussiana e costituisce il secondo livello nella struttura gerarchica dei LGM nella Formula 1.2.

1.1.5 Approssimazione di Laplace

Un concetto fondamentale da ricordare prima di procedere è la definizione della serie di Taylor:

“La serie di Taylor di una funzione in un punto è la rappresentazione della funzione come serie di termini calcolati a partire dalle derivate della funzione stessa nel punto [6].”

Formalmente, la serie di Taylor della funzione $f(x)$ in un intorno di a è mostrata nella Formula 1.6.

$$f(x) = f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \frac{f'''(a)}{3!}(x-a)^3 + \dots \quad (1.6)$$

Questo strumento è alla base dell'approssimazione di Laplace, una tecnica analitica utilizzata per stimare in modo efficiente integrali di funzioni che presentano un

picco marcato attorno a un punto, tipicamente il massimo. L'idea centrale è che, in prossimità di tale massimo, la funzione può essere approssimata da una distribuzione normale centrata in quel punto.

Sia $f(x)$ una funzione due volte derivabile. Per semplificare il calcolo dell'integrale, si considera lo sviluppo di Taylor al secondo ordine del logaritmo di $f(x)$ attorno al punto di massimo x_0 :

$$\log f(x) \approx \log f(x_0) + \frac{\partial \log f(x_0)}{\partial x}(x - x_0) + \frac{\partial^2 \log f(x_0)}{2\partial x^2}(x - x_0)^2. \quad (1.7)$$

Poiché x_0 è un punto di massimo, la derivata prima della funzione logaritmica in x_0 è nulla. Lo sviluppo si riduce quindi ai soli termini costante e quadratico. Inoltre, la seconda derivata è negativa, poiché la funzione presenta concavità rivolta verso il basso nel punto di massimo.

Per ottenere una forma analoga al logaritmo di una densità normale, si riscrive l'approssimazione come:

$$\log f(x) \approx \log f(x_0) - \frac{1}{2\hat{\sigma}^2}(x - x_0)^2, \quad (1.8)$$

dove $\hat{\sigma}^2$ rappresenta la varianza stimata della distribuzione normale che approssima $f(x)$ in un intorno del punto x_0 . Essa è definita come l'inverso dell'opposto della curvatura della funzione logaritmica nel punto di massimo [7]:

$$\hat{\sigma}^2 = - \left(\frac{\partial^2 \log f(x)}{\partial x^2} \Big|_{x=x_0} \right)^{-1} \quad (1.9)$$

Questa relazione evidenzia il legame tra la curvatura della funzione nel punto di massimo e la varianza della distribuzione approssimante: una curvatura più pronunciata (ossia un picco più stretto) corrisponde a una varianza minore, mentre una curvatura più dolce implica una maggiore dispersione.

Sostituendo questa approssimazione nella funzione originale ed effettuando l'integrazione, si ottiene la seguente stima dell'integrale, basata su una distribuzione

normale centrata in x_0 :

$$\int f(x)dx = \int \exp[\log f(x)]dx \approx \text{constant} \int \exp\left[-\frac{(x-x_0)^2}{2\hat{\sigma}^2}\right]dx. \quad (1.10)$$

Questa è l'essenza dell'approssimazione di Laplace: sostituire una funzione complessa con una gaussiana centrata nel punto di massimo, sfruttando le proprietà del secondo ordine dello sviluppo di Taylor per semplificare il calcolo integrale.

1.1.6 Integrated Nested Laplace Approximation

A differenza degli algoritmi MCMC, che mirano a campionare dalla distribuzione congiunta a posteriori di tutti i parametri del modello, l'approccio INLA si focalizza sull'approssimazione diretta delle distribuzioni marginali a posteriori. In particolare, l'obiettivo è stimare:

- la marginale a posteriori per ciascun elemento x_i del campo latente:

$$\pi(x_i | \mathbf{y}), \quad i = 1, \dots, n \quad (1.11)$$

- la marginale a posteriori per ciascun iperparametro θ_j :

$$\pi(\theta_j | \mathbf{y}), \quad j = 1, \dots, \dim(\boldsymbol{\theta}) \quad (1.12)$$

Entrambe queste quantità possono essere espresse come integrali che coinvolgono la distribuzione congiunta a posteriori. In particolare:

$$\pi(x_j | \mathbf{y}) = \int \pi(x_j, \boldsymbol{\theta} | \mathbf{y}) d\boldsymbol{\theta} = \int \pi(x_j | \boldsymbol{\theta}, \mathbf{y}) \pi(\boldsymbol{\theta} | \mathbf{y}) d\boldsymbol{\theta} \quad (1.13)$$

$$\pi(\theta_k | \mathbf{y}) = \int \pi(\boldsymbol{\theta} | \mathbf{y}) d\boldsymbol{\theta}_{-k} \quad (1.14)$$

dove $\boldsymbol{\theta}_{-k}$ rappresenta il vettore $\boldsymbol{\theta}$ privato del k -esimo elemento [3].

INLA approssima queste quantità attraverso una formulazione annidata (*nested*), che combina due livelli di approssimazione:

- un'approssimazione analitica per le distribuzioni condizionate $\pi(x_j|\theta, \mathbf{y})$ e $\pi(\theta|\mathbf{y})$;
- un'integrazione numerica sullo spazio degli iperparametri per ottenere le marginali finali.

Usando le regole della probabilità [8] secondo cui:

$$p(b) = \frac{p(a, b)}{p(a|b)}, \quad p(b|c) = \frac{p(a, b|c)}{p(a|b, c)}$$

possiamo riscrivere il termine $\pi(\theta|\mathbf{y})$ come:

$$\pi(\theta|\mathbf{y}) = \frac{\pi(\mathbf{x}, \theta|\mathbf{y})}{\pi(\mathbf{x}|\theta, \mathbf{y})}. \quad (1.15)$$

Espandendo il numeratore e usando un'approssimazione gaussiana al denominatore otteniamo:

$$\pi(\theta|\mathbf{y}) = \frac{\pi(\mathbf{y}|\mathbf{x}, \theta)\pi(\mathbf{x}|\theta)\pi(\theta)}{\tilde{\pi}_G(\mathbf{x}|\theta, \mathbf{y})} \Big|_{x=x^*(\theta)} = \tilde{\pi}(\theta|\mathbf{y}) \quad (1.16)$$

dove $x^*(\theta)$ è il massimo di \mathbf{x} per questa configurazione di θ [9].

Per approssimare il termine $\pi(x_j|\theta, \mathbf{y})$, INLA offre diverse strategie, che si differenziano per complessità computazionale e accuratezza. L'approccio più semplice e veloce consiste nell'usare un'approssimazione gaussiana, ottenuta dalla distribuzione approssimata $\tilde{\pi}_G(\mathbf{x}|\theta, \mathbf{y})$. Tuttavia, questa scelta può risultare poco accurata in presenza di distribuzioni asimmetriche o con code pronunciate.

Una maggiore precisione si ottiene tramite l'approssimazione di Laplace, che migliora la stima della densità condizionata tenendo conto della curvatura della funzione log-posteriori attorno alla moda. Questo metodo, sebbene più accurato, comporta un costo computazionale superiore.

Come compromesso tra velocità ed accuratezza, INLA utilizza di default una versione semplificata dell'approssimazione di Laplace, capace di correggere in modo efficace le principali distorsioni (in particolare sulla media e sull'asimmetria) introdotte dall'approssimazione gaussiana, mantenendo al contempo l'efficienza computazionale [3].

1.1.7 Stochastic Partial Differential Equation

Tra le tante classi di modelli che rientrano nel framework dei modelli gerarchici latenti (LGM), un ruolo di particolare rilievo è occupato dai modelli spaziali e spaziotemporali. In questi casi, l'obiettivo è descrivere e modellare la dipendenza tra osservazioni che si trovano vicine tra loro nello spazio e/o nel tempo. Per affrontare questo tipo di problemi, l'approccio proposto da Lindgren, Rue e Lindström (2011), oggi implementato in INLA, si basa su un'innovativa rappresentazione dei campi gaussiani (GF) con covarianza Matérn attraverso l'uso di equazioni differenziali stocastiche alle derivate parziali (SPDE), risolte numericamente mediante il metodo degli elementi finiti [10].

La chiave di questo approccio consiste nell'osservazione che un campo gaussiano con struttura di dipendenza Matérn può essere visto come soluzione dell'equazione SPDE 1.17:

$$(\kappa^2 - \Delta)^{\alpha/2} x(s) = W(s), \quad s \in \Omega \quad (1.17)$$

dove:

- $x(s)$ è il campo gaussiano spaziale continuo;
- Ω è il dominio spaziale;
- Δ è il Laplaciano (operatore differenziale);
- α controlla la regolarità del campo;
- $W(s)$ è rumore bianco gaussiano spaziale.

La scelta della covarianza di Matérn è particolarmente vantaggiosa, in quanto questa famiglia di funzioni consente di modellare in modo flessibile proprietà essenziali del processo, quali la regolarità, la scala spaziale e la varianza, rendendola adatta a una vasta gamma di fenomeni spaziali. Formalmente, la covarianza Matérn tra due punti s_1 ed s_2 del dominio spaziale dipende dalla distanza euclidea $\|s_1 - s_2\|$ ed è definita come mostrato nella Formula 1.18

$$\text{Cov}(x(s_1), x(s_2)) = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} (\kappa \|s_1 - s_2\|)^\nu K_\nu(\kappa \|s_1 - s_2\|) \quad (1.18)$$

dove:

- σ^2 è la varianza marginale del campo;
- $\nu > 0$ controlla la regolarità (smoothness) del campo: un valore maggiore di ν implica che il campo latente è più regolare e più liscio, con una funzione di covarianza che decresce più lentamente con la distanza; è legato al parametro α della 1.17 tramite la relazione $\nu = \alpha - d/2$ con d dimensione del dominio spaziale;
- $\kappa > 0$ è un parametro di scala inversamente proporzionale al *range spaziale* (ovvero la distanza oltre la quale la correlazione tra punti diventa trascurabile);
- $K_\nu(\cdot)$ è la funzione di Bessel modificata di secondo tipo;
- $\Gamma(\cdot)$ è la funzione Gamma.

Grazie alla loro interpretazione intuitiva, i parametri ν e κ permettono di controllare in modo diretto le proprietà spaziali del processo, facilitando l'analisi e l'interpretazione dei risultati.

Per rendere computazionalmente trattabile il campo continuo $x(s)$, la soluzione dell'equazione SPDE viene approssimata numericamente mediante il metodo degli elementi finiti. Si introduce quindi una discretizzazione spaziale del dominio tramite una mesh triangolare, su cui si costruisce un set di funzioni di base ϕ_k , lineari a tratti, tali che:

$$x(s) \approx \sum_{k=1}^K x_k \phi_k(s) \quad (1.19)$$

dove

- x_k sono coefficienti aleatori associati ai nodi;
- K sono il numero totale di nodi della mesh.

Le funzioni di base sono scelte lineari a tratti su ogni triangolo, il che indica che è uguale ad 1 in corrispondenza del vertice k e 0 altrimenti. Questo implica che ciascun coefficiente x_k interagisce solo con pochi altri nodi adiacenti, producendo

una matrice di precisione sparsa. Questo consente di sfruttare appieno le potenzialità computazionali dei GMRF e rappresenta uno dei principali punti di forza dell'approccio SPDE-INLA.

La mesh ha un ruolo cruciale perché determina la risoluzione spaziale dell'approssimazione del campo, controlla la sparsità della matrice di precisione del GMRF risultante e influisce sulla precisione dell'intero modello.

Nel contesto dei modelli spaziali, il predittore lineare introdotto nella formulazione generale 1.4 assume una forma specifica mostrata nella Formula 1.20.

$$\eta_i = \alpha + \sum_{k=1}^{n_\beta} \beta_k z_{ki} + x(s_i) \quad (1.20)$$

In questa formulazione, l'unico effetto strutturato latente è di natura spaziale ed è rappresentato da un campo gaussiano con covarianza Matérn, approssimato mediante l'approccio SPDE. Una naturale estensione di questo modello si ha nel contesto spazio-temporale, in cui il campo latente è influenzato congiuntamente da spazio e tempo. In questo caso, il predittore lineare assume la forma riportata nella Formula 1.21,

$$\eta_{it} = \alpha + \sum_{k=1}^{n_\beta} \beta_k z_{kit} + w(s_i, t) \quad (1.21)$$

dove $w(s_i, t)$ rappresenta il valore del campo latente spazio-temporale nel punto spaziale s_i e al tempo t .

L'approccio SPDE rappresenta dunque un ponte tra la teoria dei campi gaussiani continui e i GMRF, permettendo di sfruttare la flessibilità della covarianza Matérn insieme alla scalabilità computazionale necessaria per l'analisi di fenomeni spazio-temporali realistici, come la distribuzione delle precipitazioni a livello regionale oggetto di questa tesi.

1.1.8 Panoramica su R-INLA

INLA è stato implementato nel pacchetto R-INLA, che fornisce una sintassi flessibile per la definizione di modelli gerarchici latenti (LGM) e consente di eseguire inferenza bayesiana approssimata in un'ampia gamma di contesti, dai modelli lineari

generalizzati misti ai modelli spaziali e spazio-temporali più complessi [5]. Grazie alla sua versatilità, l'approccio INLA ha trovato applicazione in numerosi ambiti, tra cui finanza (ad esempio, nella stima di modelli di volatilità stocastica), demografia (modelli spaziali per l'analisi del numero di suicidi per area geografica), ecologia e ambiente. Un esempio rilevante è rappresentato dalle analisi spazio-temporali sull'inquinamento atmosferico, come lo studio delle concentrazioni di PM10 in Piemonte durante l'inverno, modellate mediante un campo latente spazio-temporale tramite SPDE [11].

Il pacchetto `R-INLA` consente di specificare i modelli in modo compatto e leggibile, grazie a una sintassi simile a quella di `lm()` e `glm()` in R. In particolare:

- la funzione di verosimiglianza e la *link function* $g(\cdot)$ vengono indicate tra gli argomenti della funzione `inla(...)`;
- il predittore lineare η_i viene specificato utilizzando la notazione a formula di R, in modo analogo a quanto avviene con le funzioni `lm()` e `glm()`. Questo consente di costruire modelli specificando la relazione tra variabili dipendenti e indipendenti in modo compatto e leggibile, ad esempio $\mathbf{y} \sim \mathbf{x}_1 + \mathbf{f}(\mathbf{x}_2)$;
- gli effetti strutturati, come termini spaziali e/o temporali vengono specificati all'interno della formula attraverso la funzione `f(...)`. Ad esempio, `f(id, model = "iid")` definisce un effetto casuale i.i.d., mentre `f(s, model = spde)` definisce un campo spaziale tramite l'approccio SPDE;
- la mesh triangolare necessaria per l'approccio SPDE viene costruita con funzioni dedicate, come `inla.mesh.2d()`.

Un elemento chiave nella modellazione bayesiana con `R-INLA` è la specificazione delle distribuzioni a priori, che può essere gestita in modo flessibile:

- con `control.fixed` per i coefficienti fissi del predittore lineare (β_k);
- con `control.family` per i parametri della distribuzione di risposta;

- con `control.hyper` all'interno di `f(...)` per definire le prior degli iperparametri associati agli effetti strutturati, come la precisione nei modelli "iid", "rw1", "spde", ecc.

Nel caso dell'approccio SPDE, è comune utilizzare prior penalizzanti di tipo PC-priors (Penalized Complexity priors) per i parametri che controllano la varianza e la scala spaziale del campo Matérn, come κ e τ (parametro di precisione ovvero l'inverso della deviazione standard, che regola la variabilità globale del campo). Un esempio è l'uso della prior `pc.cor1` per il parametro di autocorrelazione `rho` nei modelli temporali AR(1), che consente di esprimere in modo esplicito la probabilità a priori che `rho` superi una certa soglia.

1.1.9 Costruzione della mesh

Come è stato detto nei paragrafi precedenti, il ruolo della mesh è cruciale per l'approccio SPDE. Nel pacchetto R-INLA viene costruita con funzioni dedicate, come `inla.mesh.2d()`. Questa funzione permette di controllare in modo preciso la discretizzazione del dominio spaziale grazie ai seguenti argomenti [3]:

- `loc`: specifica le coordinate spaziali dei punti osservati. Queste coordinate vengono utilizzate come vertici iniziali della mesh, assicurando che l'approssimazione del campo sia ben definita nei punti in cui sono presenti dati.
- `boundary`: definisce il contorno del dominio spaziale. Serve a delimitare l'area all'interno della quale la mesh viene costruita, evitando che la discretizzazione si estenda inutilmente oltre la regione di interesse.
- `offset`: controlla la distanza delle estensioni interne ed esterne della mesh rispetto ai dati. Valori adeguati per questo parametro evitano effetti al bordo, garantendo che il campo sia ben approssimato anche nelle aree periferiche.
- `cutoff`: imposta la distanza minima consentita tra due vertici. Questo è utile per ridurre la densità della mesh in aree con molte osservazioni ravvicinate,

evitando la creazione di triangoli troppo piccoli e migliorando l'efficienza computazionale.

- **max.edge**: stabilisce la lunghezza massima dei lati dei triangoli, sia all'interno del dominio sia nelle aree di estensione. Permette di controllare la risoluzione della mesh: valori più piccoli portano a una discretizzazione più fine (quindi più accurata ma computazionalmente più costosa), mentre valori più grandi producono una mesh più grossolana.

1.1.10 Variogramma

Prima di introdurre l'analisi sviluppata in questo lavoro, è utile presentare brevemente uno strumento esplorativo impiegato per indagare la dipendenza spaziale nei dati: il *variogramma empirico*. Questo strumento consente di permettere di riassumere la struttura di dipendenza di un campo gaussiano spaziale $Z(\cdot)$. In particolare, il semivariogramma misura la dissimilarità media tra coppie di osservazioni separate da un dato vettore di distanza \mathbf{h} ed è calcolato come 1.22 ([3]).

$$2\gamma(\mathbf{h}) = \frac{1}{|N(\mathbf{h})|} \sum_{(s_i, s_j) \in N(\mathbf{h})} (Z(s_i) - Z(s_j))^2 \quad (1.22)$$

dove $N(\mathbf{h})$ rappresenta l'insieme delle coppie di punti distanti \mathbf{h} e $|N(\mathbf{h})|$ è il numero di tali coppie.

In generale, il valore del variogramma tende ad aumentare con la distanza: punti vicini nello spazio mostrano valori simili, mentre quelli più distanti tendono a essere meno correlati. Oltre una certa soglia di distanza, detta *range*, la dipendenza spaziale si attenua fino a diventare trascurabile, e il variogramma si stabilizza su un valore chiamato *sill*, che rappresenta approssimativamente la varianza del processo (nel caso di processi stazionari).

Un aspetto particolarmente interessante è il comportamento del variogramma per distanze molto piccole. Se si osserva una discontinuità all'origine, si è in presenza di un *nugget effect* (effetto pepita), che segnala una variabilità residua dovuta a errori di misura o a una componente spaziale non rilevabile alla scala di osservazione [12]. Se invece $\gamma(0) = 0$, il campo è perfettamente regolare.

Un esempio grafico di variogramma empirico è mostrato nella Figura 1.1, dove si può osservare l'andamento tipico della semivarianza rispetto alla distanza tra punti spaziali. In sintesi, la forma del variogramma offre informazioni preziose

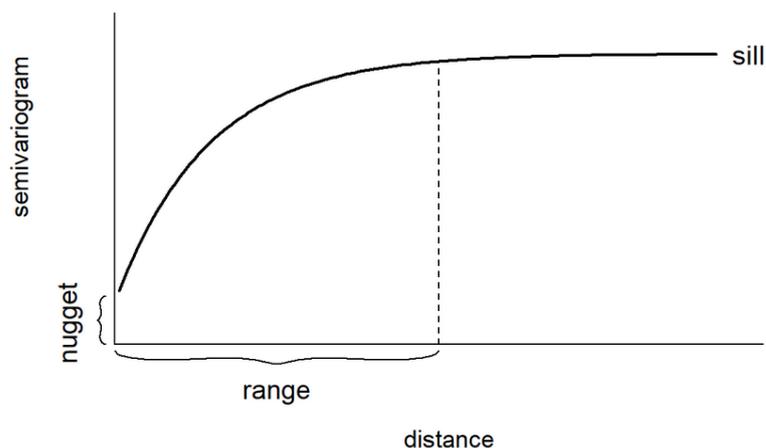


Figura 1.1: Esempio grafico di variogramma empirico [3].

sulla scala di dipendenza, sulla regolarità del campo e sull'eventuale presenza di rumore, rendendolo uno strumento fondamentale nella fase esplorativa di un'analisi geostatistica.

1.2 Fondamenti Teorici Classici della Teoria dei Valori Estremi

La teoria dei valori estremi si distingue all'interno della statistica per il suo focus sullo studio dell'insolito, ovvero degli eventi rari e estremi, piuttosto che sul comportamento medio o tipico, come descritto in [13]. Il suo punto di partenza è l'analisi del massimo di una sequenza di variabili casuali indipendenti e identicamente distribuite (i.i.d.):

$$M_n = \max\{X_1, \dots, X_n\}$$

dove $X_1, \dots, X_n \sim F$. In contesti applicativi, ciascuna X_i rappresenta spesso una misura rilevata in modo regolare nel tempo e M_n corrisponde al valore massimo osservato in un intervallo temporale, come un anno.

La distribuzione esatta del massimo può essere espressa per ogni n :

$$P\{M_n \leq z\} = P\{X_1 \leq z, \dots, X_n \leq z\} = P\{X_1 \leq z\} \cdots P\{X_n \leq z\} = \{F(z)\}^n$$

Per evitare che la distribuzione del massimo degeneri su un solo valore, si introduce una normalizzazione tramite due sequenze di costanti $a_n > 0$ e b_n , definendo una variabile trasformata:

$$M_n^* = \frac{M_n - b_n}{a_n}.$$

Questa trasformazione consente di "stabilizzare" la distribuzione di M_n^* al crescere di n , permettendo lo studio del suo comportamento limite. Il risultato fondamentale, noto come teorema dei valori estremi, stabilisce che, se la distribuzione limite di M_n^* esiste, essa deve appartenere a una delle seguenti tre famiglie: 1.2.1.

Teorema 1.2.1. *Se esistono due sequenze di costanti $\{a_n > 0\}$ e $\{b_n\}$ tali che*

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\frac{M_n - b_n}{a_n} \leq z \right) = G(z),$$

dove G è una funzione di ripartizione non degenera, allora G è necessariamente una delle seguenti:

- *Tipo I (Gumbel):* $G(z) = \exp(-e^{-z})$, $z \in \mathbb{R}$
- *Tipo II (Fréchet):* $G(z) = \begin{cases} 0, & z \leq 0 \\ \exp(-z^{-\alpha}), & z > 0 \end{cases}$, $\alpha > 0$
- *Tipo III (Weibull):* $G(z) = \begin{cases} \exp(-(-z)^\alpha), & z \leq 0 \\ 1, & z > 0 \end{cases}$, $\alpha > 0$

per i parametri $a > 0$, b e, nel caso delle famiglie II e III, $\alpha > 0$.

Queste tre famiglie costituiscono le distribuzioni limite dei massimi e rappresentano l'intero insieme dei possibili comportamenti asintotici del massimo normalizzato M_n^* , indipendentemente dalla forma della distribuzione di partenza F .

1.2.1 La Distribuzione GEV

Le tre famiglie classiche di distribuzioni dei valori estremi (Gumbel, Fréchet e Weibull) possono essere ricondotte a un'unica formulazione unificata: la distribuzione dei valori estremi generalizzata (*Generalized Extreme Value*, GEV), definita dalla funzione di ripartizione mostrata nella Formula 1.23,

$$G(z) = \exp \left\{ - \left[1 + \xi \left(\frac{z - \mu}{\sigma} \right) \right]^{-1/\xi} \right\}, \quad (1.23)$$

valida per tutti i valori z tali che $1 + \xi(z - \mu)/\sigma > 0$, con parametri $\mu \in \mathbb{R}$ (location), $\sigma > 0$ (scale) e $\xi \in \mathbb{R}$ (shape).

Questa parametrizzazione consente di rappresentare le tre distribuzioni limite in modo coerente:

- per $\xi > 0$ si ottiene il caso Fréchet (tipo II);
- per $\xi < 0$ si ottiene il caso Weibull (tipo III)
- mentre il caso Gumbel (tipo I) corrisponde al limite per $\xi \rightarrow \infty$.

L'introduzione della famiglia GEV rappresenta un vantaggio significativo dal punto di vista pratico: consente di modellare l'intero spettro dei possibili comportamenti di coda attraverso un unico modello, demandando ai dati il compito di indicare quale tra le tre tipologie di estremi è più appropriata tramite la stima del parametro di *shape* ξ . Inoltre, l'inferenza su ξ fornisce anche una misura dell'incertezza riguardo al tipo di distribuzione più adatto: un'alta variabilità stimata di ξ suggerisce ambiguità nella classificazione, mentre una stima stabile ne rafforza l'identificazione.

1.2.2 Una riformulazione alternativa: la distribuzione bGEV

Come visto, la distribuzione GEV fornisce una rappresentazione unificata dei tre tipi di distribuzioni limite per i massimi: Gumbel, Fréchet e Weibull. Tuttavia, la sua parametrizzazione classica in termini di *location* μ , *scale* σ e *shape* ξ presenta alcune difficoltà pratiche, soprattutto quando si desidera modellare la dipendenza

dei parametri da covariate (come nei modelli regressivi o spazio-temporali). In particolare:

- la media della GEV è definita solo per $\xi < 1$,
- la varianza solo per $\xi < 0.5$

compromettendo la possibilità di assegnare interpretazioni standard ai parametri e rendendo complessa la definizione di distribuzioni a priori sensate.

Per superare queste limitazioni, è stata proposta una parametrizzazione alternativa che si basa su:

- un quantile di riferimento q_α , che funge da parametro di posizione;
- una misura di dispersione s_β , definita come differenza tra due quantili simmetrici (es. tra il 75esimo e il 25esimo percentile);
- il parametro di coda ξ , che resta invariato.

La funzione di ripartizione della GEV in questa parametrizzazione diventa:

$$F(y \mid q_\alpha, s_\beta, \xi) = \exp \left\{ - \left[\left(\frac{y - q_\alpha}{s_\beta (\ell_{1-\beta/2, \xi} - \ell_{\beta/2, \xi})^{-1}} + \ell_{\alpha, \xi} \right)^+ \right]^{-1/\xi} \right\}$$

Dove:

$$\ell_{a, \xi} = (-\log a)^{-\xi}, \quad (x)^+ = \max(x, 0)$$

e il caso $\xi = 0$, corrisponde alla forma limite di Gumbel. Questa formulazione consente una rappresentazione più interpretabile della posizione e della dispersione della distribuzione, rendendola più adatta all'uso in modelli con covariate o componenti spaziali.

Su queste basi, è stata sviluppata la distribuzione bGEV (*blended Generalized Extreme Value*), che rappresenta una combinazione pesata tra una distribuzione di tipo Fréchet (GEV con $\xi > 0$) e una distribuzione di tipo Gumbel ($\xi = 0$). La funzione di distribuzione cumulativa è definita come:

$$H(y \mid \theta) = F(y \mid q_{\alpha_F}, s_{\beta_F}, \xi)^{p(y)} \cdot G(y \mid q_{\alpha_G}, s_{\beta_G})^{1-p(y)}$$

dove:

- $F(\cdot)$ è una distribuzione Fréchet parametrizzata tramite quantili
- $G(\cdot)$ è una distribuzione Gumbel, anch'essa parametrizzata tramite quantili,
- $p(y)$ è una funzione peso che varia in $[0, 1]$ e controlla la transizione tra le due distribuzioni.

Il vettore completo dei parametri è:

$$\theta = (q_{\alpha_F}, s_{\beta_F}, \xi, q_{\alpha_G}, s_{\beta_G}, sp_1, sp_2, a, b)$$

La funzione peso è definita come la distribuzione cumulativa di una Beta trasformata:

$$p(y | sp_1, sp_2, a, b) = \Pr\left(Y \leq \frac{y - a}{b - a} \mid Y \sim \text{Beta}(sp_1, sp_2)\right)$$

dove:

- $sp_1 > 1, sp_2 > 1$ sono i parametri di forma della distribuzione Beta;
- a e b sono due quantili della distribuzione Fréchet, detti *quantili mixing*, cioè la regione in cui le due distribuzioni vengono fuse. All'interno dell'intervallo $[a, b]$, la funzione peso $p(x)$ passa gradualmente da 1 a 0, permettendo una transizione regolare tra Fréchet e Gumbel. Al di fuori dell'intervallo, il contributo proviene interamente da una delle due componenti.

L'implementazione di default in R-INLA prevede $sp_1 = sp_2 = 5$, generando una funzione peso simmetrica centrata nell'intervallo *mixing*. In sintesi, la distribuzione bGEV consente una modellazione flessibile dei dati estremi. Grazie a questa formulazione, è possibile evitare problemi di discontinuità e interpretabilità nei modelli con covariate o effetti spaziali, mantenendo al contempo la capacità di rappresentare comportamenti di coda estremi.

Per un approfondimento teorico e applicativo si rimanda a [14].

1.2.3 Threshold Models e Distribuzione di Pareto Generalizzata (GPD)

L'approccio classico della teoria dei valori estremi prevede l'analisi dei massimi a blocchi, ossia la selezione di un solo valore massimo da ciascun sottoinsieme temporale della serie (ad esempio, il massimo annuale). Tuttavia, questa strategia può risultare inefficiente, soprattutto quando si dispone di serie storiche dense e dettagliate, come rilevazioni giornaliere o orarie. In questi casi, un'alternativa più efficace è rappresentata dai *Threshold Models*, che permettono di sfruttare tutte le osservazioni che eccedono un certo valore elevato u . Questo consente di modellare direttamente i valori estremi senza dover ricorrere a una suddivisione in blocchi. Sia quindi X_1, X_2, \dots una sequenza di variabili casuali i.i.d. con funzione di distribuzione F . Consideriamo gli eventi estremi come quei valori di X_i che superano una soglia fissata u . Una strategia naturale per modellare tali eccessi è quella di considerare la distribuzione condizionata:

$$\mathbb{P}(X > u + y \mid X > u) = \frac{1 - F(u + y)}{1 - F(u)}, \quad y > 0.$$

Se F fosse nota, la distribuzione degli eccessi risulterebbe nota. Tuttavia, poiché ciò non accade nella pratica, si ricorre a un'approssimazione teorica valida per soglie sufficientemente elevate. Questo approccio è del tutto analogo all'uso della distribuzione GEV per approssimare i massimi di blocchi. Il risultato teorico alla base dei modelli a soglia, tratto da [13], è formalizzato nel teorema 1.2.2.

Teorema 1.2.2. *Sia X_1, X_2, \dots una successione di variabili casuali indipendenti con funzione di distribuzione comune F , e indichiamo*

$$M_n = \max\{X_1, \dots, X_n\}$$

Supponiamo che F soddisfi le condizioni del Teorema 1.2.1, cioè che esista una funzione limite G tale che per n grande:

$$\mathbb{P}(M_n \leq z) \approx G(z).$$

Quindi, per u sufficientemente grande, la distribuzione condizionata dell'eccesso $Y = X - u$ dato che $X > u$ è approssimativamente

$$H(y) = 1 - \left(1 + \frac{\xi y}{\tilde{\sigma}}\right)^{-1/\xi}, \quad (1.24)$$

definita per $y > 0$ tali che $1 + \xi y/\tilde{\sigma} > 0$, dove

$$\tilde{\sigma} = \sigma + \xi(u - \mu).$$

La distribuzione 1.24 definisce la *Generalized Pareto Distribution* (GPD), una famiglia che si dimostra essere la controparte naturale della GEV per il caso dei superamenti di soglia. Il legame tra le due distribuzioni è diretto: se i massimi a blocchi seguono una GEV, allora gli eccessi sopra soglia seguono approssimativamente una GPD con gli stessi parametri di forma. In particolare, il parametro ξ di forma è condiviso tra le due famiglie, e controlla la coda della distribuzione. A seconda del segno di ξ , la GPD assume diverse forme:

- Se $\xi < 0$, la distribuzione presenta un supporto limitato superiormente a $u - \tilde{\sigma}/\xi$;
- Se $\xi > 0$, la distribuzione ha coda pesante e supporto illimitato superiormente;
- Se $\xi = 0$, si ottiene per limite una distribuzione esponenziale:

$$H(y) = 1 - \exp\left(-\frac{y}{\tilde{\sigma}}\right), \quad y > 0. \quad (1.25)$$

La GPD fornisce così un quadro teorico robusto per modellare le code della distribuzione dei dati, sfruttando in modo più efficiente le informazioni sugli eventi estremi rispetto al solo utilizzo dei massimi per blocco. Il parametro ξ , come nella GEV, determina la natura e il rischio associato agli eventi rari, ed è cruciale per l'analisi delle code e per l'estrapolazione oltre i valori osservati.

Capitolo 2

Specifica del modello spazio-temporale

In questo capitolo vengono presentati nel dettaglio i modelli statistici sviluppati per l'analisi delle precipitazioni estreme nella regione Piemonte. L'obiettivo dell'analisi è modellare gli estremi pluviometrici tenendo conto della loro dipendenza spaziale e temporale, attraverso l'approccio INLA, implementato nel software R.

Il capitolo è strutturato come segue: si comincia con una descrizione del dataset utilizzato, includendo un'analisi esplorativa delle precipitazioni giornaliere, con particolare attenzione alla stagionalità e alla struttura spaziale dei dati. Segue una sezione dedicata al preprocessing, in cui vengono illustrate le operazioni di aggregazione temporale, controllo di qualità e selezione delle stazioni meteorologiche.

Nella parte centrale del capitolo vengono poi introdotti e descritti i due modelli principali sviluppati:

- un secondo modello unico, basato sulla famiglia BGEV, che consente di modellare direttamente i massimi di blocco, secondo l'approccio proposto da [14];
- un secondo modello a due stadi, che combina una distribuzione Gamma per l'intera serie e un modello Generalized Pareto (GP) per le eccedenze oltre una soglia elevata. L'impostazione metodologica segue le linee guida di [15].

Entrambi i modelli includono componenti spaziali e temporali, covariate esplicative e vengono stimati tramite l'approccio SPDE per il campo latente, in combinazione con la struttura AR(1) nel dominio temporale. La struttura del modello spazio-temporale è ispirata a Blangiardo et al. (2013) [5], che fornisce una guida completa alla costruzione e all'interpretazione di modelli di questo tipo, sia dal punto di vista teorico che implementativo.

2.1 Dati

L'analisi è stata condotta su un dataset contenente informazioni relative a massime precipitazioni annue raccolte presso una rete di stazioni meteorologiche localizzate in Piemonte nel periodo dal 1928 al 2020. Il fenomeno è stato registrato per differenti durate temporali:

- 10, 20, 30 minuti
- 1, 3, 6, 12, 24 ore
- 2, 3, 4, 5 giorni

Quindi, per ciascun anno nel periodo di osservazione e per ogni stazione meteorologica inclusa nella rete, è stato selezionato il valore massimo di precipitazione. In questo modo, ogni riga del dataset rappresenta un'osservazione spazialmente localizzata (una stazione) e temporalmente indicizzata (un anno), associata a un evento di pioggia estrema. Ogni finestra temporale è stata studiata separatamente, ottenendo un totale di 12 dataset. Ci si riferirà a questi dataset come Dataset10min, Dataset20min ecc.

Oltre all'informazione sul valore di precipitazione, il dataset include un insieme di covariate esplicative di tipo spaziale, morfologico e climatico. In particolare, ogni osservazione è caratterizzata da:

- Informazioni spaziali, rappresentate dalle coordinate X e Y espresse in m con sistema di riferimento EPSG:32632 ovvero WGS84 UTM 32N;
- Informazioni temporali, sotto forma del minuto, giorno, ora ed anno solare in cui è stato registrato l'evento di precipitazione considerato;

- Variabili orografiche, come la quota del pluviometro (ottenuta dal database ufficiale di ARPA Piemonte), espressa in m (variabile `quota`);
- Indici morfologici, tra cui l'indice di apertura del terreno (variabile
- Indici morfologici, tra cui l'indice di apertura del terreno (variabile `openness`), calcolato a partire da un DEM con risoluzione spaziale di 500 m e utilizzando un raggio di 5 km. L'`openness` quantifica il grado di apertura di un punto rispetto all'area circostante, ovvero se il punto si trovi in posizione dominante (rilievo) o incassata (depressione). Valori elevati indicano aree sporgenti, come crinali o dorsali, mentre valori bassi caratterizzano zone concave o avvallate, come valli o conche;
- Informazioni climatiche, rappresentate dalla precipitazione media annua registrata in corrispondenza del pluviometro, estratta dal dataset BIGBANG4.0 di ISPRA, espressa in mm (variabile `map`).

2.1.1 Statistiche descrittive sui Dataset

Le informazioni disponibili derivano da 498 stazioni pluviografiche, attive in periodi differenti e con una copertura temporale non uniforme, come mostrato in Figura 2.1. Nel 1928, ad esempio, i dati relativi ai picchi di precipitazione sono disponibili soltanto per 3 stazioni. Negli anni successivi, il numero di stazioni attive cresce inizialmente, poi subisce una flessione, per poi tornare ad aumentare fino a raggiungere una certa stabilità attorno al 2000.

A scopo esplorativo, si è analizzata la struttura di dipendenza spaziale della precipitazione, attraverso la costruzione di un variogramma campionario medio. Questo strumento aiuta a capire come varia la somiglianza tra i valori misurati al crescere della distanza tra i punti di osservazione. L'approccio adottato è ispirato allo studio [16].

In particolare, per ogni anno disponibile nel dataset, è stato calcolato il variogramma empirico utilizzando i valori medi annui di precipitazione e le coordinate spaziali delle stazioni. Poiché il numero di stazioni cambia da un anno all'altro, e quindi anche l'affidabilità del variogramma associato, si è scelto di dare più peso

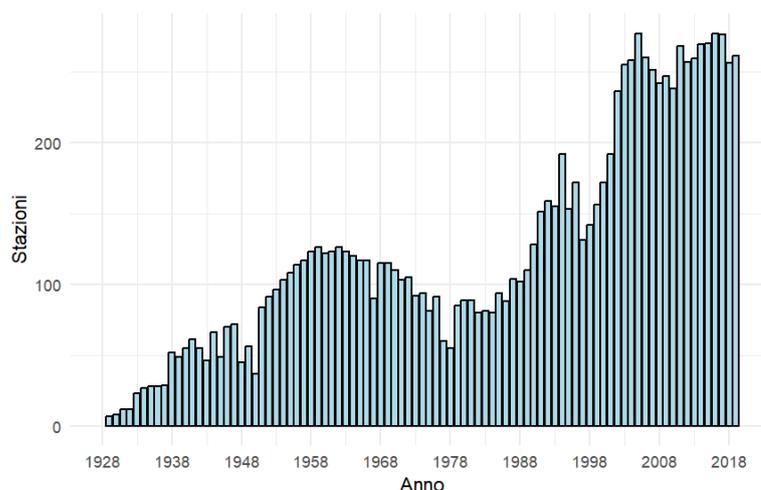


Figura 2.1: Andamento temporale delle stazioni pluviometriche attive nel periodo 1928-2020.

agli anni con un numero maggiore di stazioni attive. I variogrammi così ottenuti sono stati interpolati su una griglia comune di distanze, per renderli confrontabili, e poi mediati in modo pesato. Per rendere più chiaro l'andamento generale, la curva finale è stata smussata con un'interpolazione loess, una tecnica flessibile che permette di evidenziare le tendenze principali senza imporre una forma predefinita. Il risultato è un variogramma medio che offre una panoramica complessiva sulla struttura di dipendenza spaziale delle precipitazioni nel periodo analizzato. In particolare, osservando la forma della curva è possibile farsi un'idea della presenza di discontinuità a distanza zero (nugget), della distanza oltre la quale i valori non sono più correlati (range), e della varianza stabilizzata oltre tale soglia (sill).

Per evitare di appesantire la trattazione, vengono mostrati solo i grafici relativi ai Dataset20min (2.2) e Dataset5gg (2.3) a titolo d'esempio.

Nei grafici vengono mostrati, con punti cerchiati in blu, le semivarianze empiriche calcolate per ogni anno, in funzione della distanza tra le stazioni (in km). La curva rossa è la media pesata interpolata dei variogrammi annuali.

Il grafico 2.2 evidenzia una chiara crescita iniziale della semivarianza, suggerendo un range spaziale (cioè la distanza oltre la quale i dati non sono più correlati) intorno ai 80–100 km. Questo range aumenta a circa 150 km nel grafico 2.3.

Questo tipo di rappresentazione, pur essendo di natura esplorativa, si è rivelata

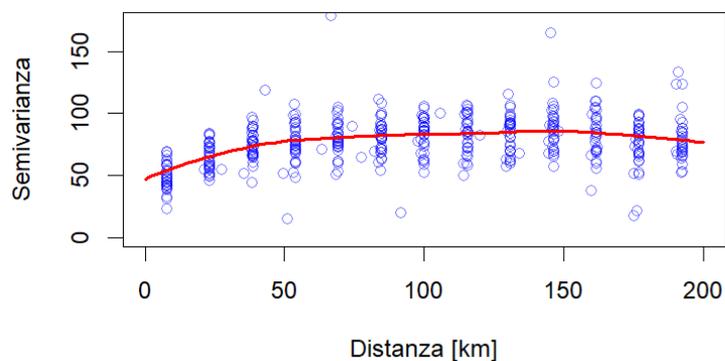


Figura 2.2: Variogrammi empirici annuali (punti) e media pesata smussata con LOESS (linea) - Dataset10min.

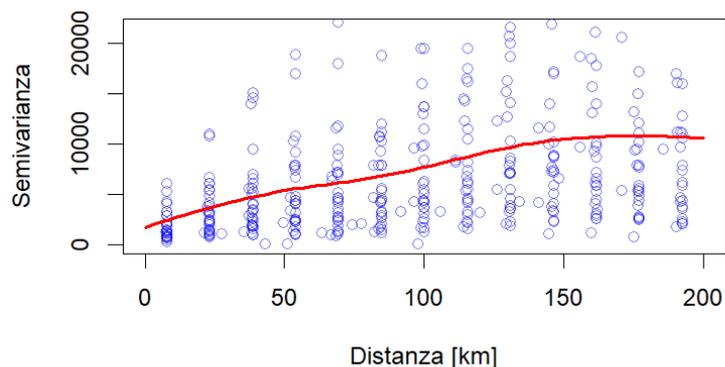


Figura 2.3: Variogrammi empirici annuali (punti) e media pesata smussata con LOESS (linea) - Dataset5gg.

utile per farsi un'idea della scala di correlazione spaziale delle precipitazioni estreme nei diversi dataset.

2.1.2 Preprocessing dei dati

Per assicurare un corretto trattamento delle covariate nel modello, si è proceduto a una fase preliminare di preprocessing dei dati.

Per prima cosa, al fine di ottenere anche le coordinate geografiche in gradi decimali

(latitudine e longitudine), necessarie per alcune rappresentazioni, è stata effettuata una trasformazione di sistema di riferimento tramite il pacchetto `sf` di R. Nello specifico:

- le coordinate X e Y originarie sono state convertite in un oggetto `sf` con CRS EPSG:32632;
- successivamente, è stata effettuata la trasformazione nel sistema geografico EPSG:4326;
- sono quindi state estratte le coordinate latitudinali e longitudinali, che sono state aggiunte al dataset anagrafico.

Inoltre, per agevolare l'interpretazione spaziale e la costruzione della mesh, sono state create nuove variabili che esprimono le coordinate originarie in chilometri. Le variabili così ottenute, sono state trasformate standardizzate sottraendo la media e dividendo per la deviazione standard. Questa operazione consente di ottenere coordinate centrate e scalate, rendendo più stabile la stima dei parametri spaziali. Faremo riferimento a queste variabili con `X_km_std` e `Y_km_std`.

Anche le variabili continue sono state standardizzate tramite z-score, ossia trasformate in variabili con media nulla e deviazione standard unitaria.

Successivamente, per consentire la costruzione del campo spazio-temporale nel contesto dell'approccio SPDE, è stato necessario assegnare un indice temporale intero a ciascun anno del periodo di osservazione. Operativamente, la variabile temporale è stata definita tramite l'assegnazione `time = rep(1:n_anni, each = n_stazioni)`, in modo da associare a ciascuna stazione, per ogni anno, un indice temporale coerente con l'ordine cronologico delle osservazioni. Questa numerazione sequenziale del tempo è fondamentale per strutturare il campo latente come un processo AR(1) nel dominio temporale, permettendo di modellare in modo coerente l'autocorrelazione tra gli anni.

Tutte le operazioni descritte sono state eseguite su ciascuno dei 12 dataset, indipendentemente dalla finestra temporale considerata.

Per ogni dataset è stata successivamente condotta un'analisi preliminare volta a individuare un intervallo temporale che massimizzasse la quantità di informazioni

disponibili, evitando al contempo la presenza di lacune temporali. Ad esempio, nei dataset con durata compresa tra i 2 a 5 giorni è presente un'interruzione tra il 2001 e il 2003. Per questo motivo si è deciso di svolgere lo studio dal 2004 al 2020.

Un esempio relativo al Dataset2gg è riportato in Figura 2.4. Gli intervalli finali

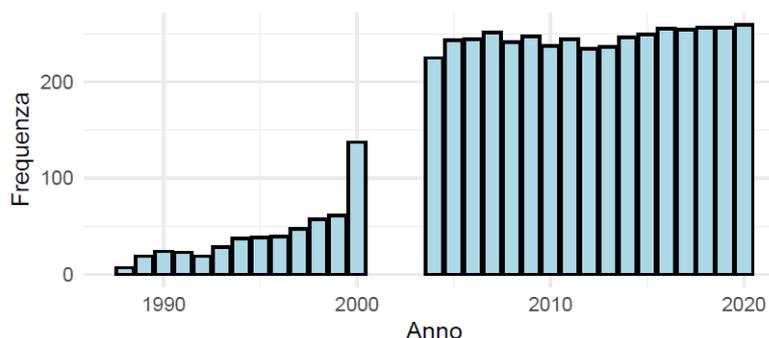


Figura 2.4: Distribuzione per anno, Dataset2gg

per ogni dataset sono specificati nella Tabella 2.1

Dataset	Intervallo
Durata minuti e ore	2000 - 2020
Durata giorni	2004 - 2020

Tabella 2.1: Intervalli temporali analizzati per ogni dataset

2.2 Costruzione della mesh e specifica del campo spazio-temporale

Per modellare la componente spaziale latente del fenomeno, è stata costruita una mesh triangolare utilizzando le coordinate delle stazioni meteorologiche espresse in chilometri e standardizzate.

La mesh spaziale è stata costruita utilizzando la funzione `fm_mesh_2d_inla()` del pacchetto `fmesher`, con la seguente configurazione:

```
mesh_st <- fmesher::fm_mesh_2d_inla(
  loc = anagrafica[, c("X_km_std", "Y_km_std")],
  loc.domain = pborders[, c("UTM_X_std", "UTM_Y_std")],
```

```
offset = c(0.1, 0.3),
max.edge = c(0.7, 2),
min.angle = c(21, 21),
cutoff = 0.1
)
```

La scelta dei parametri è stata guidata dall'obiettivo di ottenere una discretizzazione spaziale sufficientemente estesa da coprire l'intero dominio osservato, mantenendo al contempo un numero contenuto di triangoli per limitare i costi computazionali. Tale scelta è stata inoltre adattata alla nuova scala delle coordinate, standardizzate, affinché i parametri della mesh risultassero coerenti con l'unità priva di dimensione e mantenessero un'adeguata risoluzione spaziale nella nuova scala. In particolare:

- L'*offset* è stato impostato a 0.1 verso l'interno e 0.3 verso l'esterno del dominio osservato (in unità standardizzate), così da estendere opportunamente l'area di discretizzazione del campo latente e attenuare gli effetti di bordo.
- La dimensione massima dei triangoli è stata fissata a 0.7 nelle aree più dense e a 2 nelle zone periferiche, consentendo di rappresentare in modo adeguato la variabilità spaziale pur mantenendo sotto controllo il numero complessivo di nodi.
- È stato applicato un *cutoff* pari a 0.1 per evitare la formazione di triangoli eccessivamente piccoli nelle aree in cui le stazioni risultano molto ravvicinate nella scala standardizzata.
- Come nella versione precedente della mesh, è stato mantenuto un angolo minimo di 21° nei triangoli generati, al fine di garantire una buona qualità geometrica della discretizzazione.

La mesh risultante presenta 394 vertici e 741 triangoli, e consente di rappresentare in modo adeguato la variabilità spaziale del campo latente, mantenendo il costo computazionale su livelli compatibili con l'inferenza SPDE. La mesh costruita è mostrata in Figura 2.5.

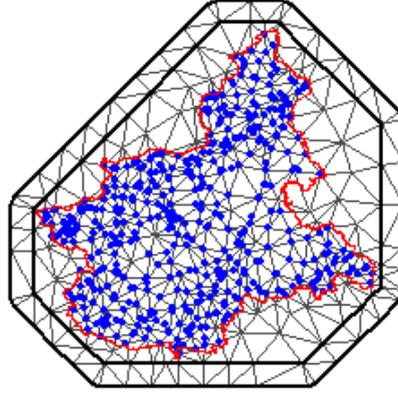


Figura 2.5: Mesh triangolare utilizzata per la discretizzazione del campo spaziale tramite SPDE, ottenuta sulla base delle stazioni disponibili.

Una volta costruita la mesh triangolare sul dominio spaziale, il campo spaziale latente è stato modellato come un processo gaussiano continuo, tramite l'approccio SPDE implementato nella funzione `inla.spde2.pcmatern()` del pacchetto R-INLA. Questo approccio consente di rappresentare un campo gaussiano con covarianza di tipo Matérn. Poiché l'obiettivo dell'analisi è spiegare la variabilità dei dati nello spazio e nel tempo, il campo spaziale è stato replicato per ciascun istante temporale e la dinamica nel tempo è stata modellata tramite una struttura autoregressiva di ordine 1 (AR(1)). Più precisamente, il modello assume che il campo al tempo t , denotato con $x_t(s)$, dipenda da quello al tempo precedente secondo la relazione 2.1:

$$x_t(s) = \rho x_{t-1}(s) + \epsilon_t(s) \quad (2.1)$$

dove $\rho \in (-1,1)$ è il parametro di autocorrelazione temporale e $\epsilon_t(s)$ è un campo spaziale indipendente, distribuito come un GMRF con la stessa struttura SPDE.

Nel codice, la variabile `field` rappresenta l'effetto spazio-temporale, mentre `field.group` indica il gruppo temporale, cioè il tempo a cui ogni osservazione è associata. Questo consente a INLA di replicare il campo spaziale per ogni istante temporale e di collegare le repliche attraverso una struttura autoregressiva. La prior specificata su ρ è una Penalized Complexity (PC) prior, costruita per penalizzare la complessità rispetto a un modello base in cui non c'è correlazione temporale

($\rho = 0$). Il parametro della prior viene specificato tramite l'argomento `param = c(0.9, 0.5)`, che equivale a imporre la seguente condizione:

$$P(|\rho| > 0.9) = 0.5. \quad (2.2)$$

In altre parole, si assegna una probabilità del 50% al fatto che l'autocorrelazione temporale tra due istanti successivi superi (in valore assoluto) la soglia di 0.9, indicando una moderata propensione verso una forte dipendenza temporale, ma lasciando comunque margine di adattamento ai dati. Questa scelta riflette l'ipotesi che il campo spaziale latente evolva lentamente nel tempo, mantenendo una struttura simile da un anno all'altro, senza però forzare un legame troppo rigido.

Nel codice, tale specifica viene implementata come segue:

```
f(field, model = spde, group = field.group, control.group = list(
  model = "ar1",
  hyper = list(
    rho = list(prior = "pc.cor1", param = c(0.9, 0.5))
  )
))
```

2.3 Predittore lineare e costruzione dello stack

Nel modello formulato, il predittore lineare assume la seguente forma:

$$\eta_i = \beta_0 + \beta_1 \cdot \text{map}_i + \beta_2 \cdot \text{openness}_i + \beta_3 \cdot \text{quota}_i + \beta_4 \cdot X_km_std_i + \beta_5 \cdot Y_km_std_i + f(s_i, t_i) \quad (2.3)$$

dove

- β_0 è l'intercetta;
- β_1, \dots, β_5 sono i coefficienti associati alle covariate standardizzate (precipitazione media annua, openness, quota e coordinate);

- $f(s_i, t_i)$ rappresenta il campo latente spazio-temporale, modellato come un processo SPDE in spazio con dipendenza AR(1) nel tempo.

Successivamente, è stato creato lo stack dei dati tramite la funzione `inla.stack`. Lo stack consente di combinare le informazioni sul campo latente e sulle covariate fisse in un'unica struttura coerente per l'inferenza INLA. Il comando `inla.stack()` specifica:

- la variabile di risposta,
- la lista delle matrici degli effetti (per gli effetti latenti e per le covariate fisse),
- la lista degli effetti da includere nel modello.

Questo approccio consente di costruire e gestire in modo flessibile modelli gerarchici complessi, come nel caso del presente studio, in cui coesistono componenti spaziali, temporali e covariate fisse. La struttura dello stack garantisce che INLA possa eseguire l'inferenza in modo efficiente, preservando la corretta associazione tra i dati, il campo latente e le componenti deterministiche del predittore.

2.4 Modello dei massimi di blocco con distribuzione bGEV

Dopo aver definito la struttura del campo latente e costruito lo stack dei dati, il modello è stato stimato mediante il pacchetto R-INLA. Per modellare i valori di precipitazione è stata utilizzata la famiglia `bgev` (*blended Generalized Extreme Value*), che consente di rappresentare distribuzioni dei massimi di blocco in un contesto completamente bayesiano, parametrizzate tramite quantili.

Siano $y_{i,t}$ le precipitazioni massime annuali osservate nella stazione $i = 1, \dots, N$ durante l'anno $t = 1, \dots, T$, per un totale di $n = N \times T$ osservazioni. Il modello assume:

$$y_{i,t} \sim \text{bGEV}(q_{\alpha,it}, s_{\beta,it}, \xi) \quad (2.4)$$

dove:

- $q_{\alpha,it}$ rappresenta il quantile $\alpha = 0.5$ (ossia la mediana), funzione delle covariate e del campo latente;
- $s_{\beta,it}$ misura la dispersione, calcolata tramite una differenza interquantile centrata sulla mediana, con $\beta = 0.25$;
- ξ è il parametro di forma, comune a tutte le osservazioni;
- **bGEV** è la distribuzione *blended Generalized Extreme Value*, combinazione tra la distribuzione di Fréchet e quella di Gumbel.

La struttura del modello permette di specificare un predittore lineare per ciascuna delle tre componenti principali della distribuzione: *location*, *spread* e *tail*.

Per la componente *location* è stata utilizzata la seguente formulazione:

$$q_{\alpha,it} = \beta_0 + \sum_{k=1}^K \beta_k x_{kit} + u(s_i, t) \quad (2.5)$$

dove β_0 è l'intercetta, x_{kit} rappresentano le covariate ambientali e geografiche (`quota`, `openness`, `map`, `X_km`, `Y_km`) con coefficienti $\beta_k \sim \mathcal{N}(0, 100)$, mentre $u(s_i, t)$ è il campo spazio-temporale.

Le prior sui coefficienti delle covariate sono distribuzioni normali $\mathcal{N}(0, 100)$ e vengono specificate con l'argomento `control.fixed`.

La componente *spread* è stata modellata come:

$$\log(s_{\beta,it}) = \gamma_0 + \gamma_1 \cdot \text{map}_{it} + \gamma_2 \cdot \text{quota}_{it}. \quad (2.6)$$

I coefficienti γ_1 e γ_2 seguono prior normali con varianza ampia ($\mathcal{N}(0, 300)$), mentre la varianza residua ha una prior `log-Gamma` con parametri $(1, 0.01)$. Questa parte del modello viene specificata usando la funzione `inla.mdata()`, che consente di associare predittori diversi a ciascuna componente della distribuzione.

Il parametro di forma ξ , che determina il comportamento della coda della distribuzione, è trattato come costante. Per esso si utilizza una prior `Penalized Complexity`, costruita per penalizzare deviazioni dalla distribuzione Gumbel (che corrisponde a

$\xi = 0$). La prior assume la forma:

$$\mathbb{P}(\xi > x) \approx \exp(-\lambda x) \quad \xi \in (0,0.5), \quad \lambda = 7 \quad (2.7)$$

Internamente, la trasformazione logit garantisce che ξ rimanga entro il supporto desiderato.

La funzione di ripartizione della distribuzione blended GEV è mostrata nella Formula 2.8,

$$H(y) = F(y|q_\alpha^{(F)}, s_\beta^{(F)}, \xi)^{p(y)} \cdot G(y|q_\alpha^{(G)}, s_\beta^{(G)})^{1-p(y)} \quad (2.8)$$

dove F è una distribuzione GEV di tipo Fréchet ($\xi > 0$), G una distribuzione di Gumbel ($\xi = 0$), entrambe parametrizzate tramite quantili. Il peso $p(y)$ che regola il miscela tra le due componenti della distribuzione bGEV è calcolato come la probabilità, sotto una distribuzione Beta(5,5), che una variabile aleatoria T sia inferiore a una trasformazione lineare di y , normalizzata rispetto a due soglie di riferimento:

$$p(y) = \mathbb{P}\left(T \leq \frac{y-a}{b-a} \mid T \sim \text{Beta}(5, 5)\right) \quad (2.9)$$

I parametri a e b corrispondono, rispettivamente, al 5° e al 20° percentile della componente Fréchet, e definiscono l'intervallo in cui avviene la transizione dalla distribuzione Gumbel a quella Fréchet. Questa scelta permette di concentrare la miscela nella parte inferiore della distribuzione, riservando alla Fréchet (più adatta a modellare code pesanti) il controllo della coda destra.

La stima del modello avviene con la chiamata alla funzione `inla()` che estituisce le stime a posteriori di tutte le componenti del predittore lineare e dei parametri della distribuzione.

2.5 Modello Soglia con distribuzione di Pareto

Il secondo modello implementato si basa su un approccio a soglia per l'analisi delle precipitazioni estreme, e combina l'uso di una distribuzione Gamma per l'intera variabile risposta con una distribuzione Generalizzata di Pareto (GP) per modellare

le eccedenze oltre una soglia elevata. Questo approccio è spunto del lavoro [17]. Come già evidenziato, ciascun dataset contiene i massimi annuali di precipitazione osservati per una determinata durata. Tuttavia, la distribuzione di questi valori è eterogenea e include anche eventi non necessariamente estremi. Ad esempio, nel caso della durata di 10 minuti, i valori variano da circa 2 mm a oltre 40 mm. Per cogliere l'intera struttura dei dati, è stato inizialmente stimato un modello con famiglia **Gamma**, scelta adatta a variabili continue positive come le precipitazioni. In questo modello, la precipitazione massima y_i è assunta distribuita come:

$$y_i \sim \text{Gamma}(\mu_i, \theta), \quad \text{con} \quad \log(\mu_i) = \eta_i$$

dove il predittore lineare $\eta_{i,t}$ include intercetta, covariate fisse (**quota**, **openness**, **map**, coordinate spaziali) e un campo latente spazio-temporale $u(s_i, t)$, modellato tramite SPDE con struttura AR(1) nel tempo. Nel modello sono state adottate distribuzioni a priori pensate per garantire un buon equilibrio tra flessibilità e controllo della complessità. In particolare, si è fatto uso sia di priori debolmente informative, che lasciano ai dati il compito di guidare la stima, sia di priori di tipo Penalised Complexity (PC), utili a evitare overfitting penalizzando esplicitamente deviazioni da modelli più semplici.

Per gli effetti fissi, inclusa l'intercetta, si è utilizzata una prior normale con media zero e varianza elevata:

$$\beta_k \sim \mathcal{N}(0, 10^4)$$

Si tratta di una scelta classica e poco informativa, che riflette l'assenza di conoscenza a priori forte sui coefficienti.

Il campo spaziale latente, costruito tramite l'approccio SPDE, è stato regolato tramite prior PC sui parametri della funzione di covarianza Matérn. Queste prior modellano l'incertezza sul range spaziale r e sulla deviazione standard σ tramite probabilità esplicite:

$$\mathbb{P}(r < r_0) = p_r, \quad \mathbb{P}(\sigma > \sigma_0) = p_\sigma$$

Nel caso specifico, sono stati utilizzati i valori $r_0 = 5$, $p_r = 0.5$, $\sigma_0 = 5$, $p_\sigma = 0.01$, che riflettono una moderata incertezza sul range spaziale ma una preferenza a priori per deviazioni standard contenute.

Infine, il parametro di dispersione della distribuzione Gamma, indicato con θ , è stato modellato con una prior di tipo log-

$$\theta \sim \log - \text{Gamma}(2,2)$$

Tale distribuzione è appropriata per parametri di scala strettamente positivi.

A partire da questo modello, si è poi definito un secondo stadio per la modellazione delle eccedenze. La soglia è stata fissata empiricamente come il quantile al 95% delle precipitazioni stimate:

$$soglia = \hat{Q}_{0.95}(\hat{Y})$$

dove \hat{Y} sono i valori stimati dal modello Gamma. In questo modo, si seleziona il 5% delle osservazioni considerate estreme nel contesto del dominio spaziale e temporale analizzato.

Le eccedenze sono state calcolate come:

$$e_i = \frac{y_i - soglia}{\hat{\mu}_i}$$

dove μ_i è il valore atteso stimato dal modello Gamma ed ha lo scopo di standardizzare l'eccedenza rispetto al livello atteso nel punto specifico, in modo da correggere per eterogeneità spaziale. Le eccedenze così definite sono modellate con un secondo modello INLA, questa volta con famiglia `gp`, che assume che e_i sia distribuita secondo una distribuzione Generalizzata di Pareto.

Il predittore lineare mantiene la stessa struttura del modello Gamma, includendo le stesse covariate e lo stesso campo latente. L'obiettivo è stimare, condizionatamente al superamento della soglia, l'intensità dell'eccedenza attesa, e quindi la componente realmente estrema del rischio. I risultati ottenuti sono utili per la mappatura del rischio localizzato di precipitazioni intense.

Per offrire una visione d'insieme delle operazioni implementate nel modello, si propone di seguito un diagramma nella Figura 2.6 che sintetizza le principali fasi dell'algoritmo.

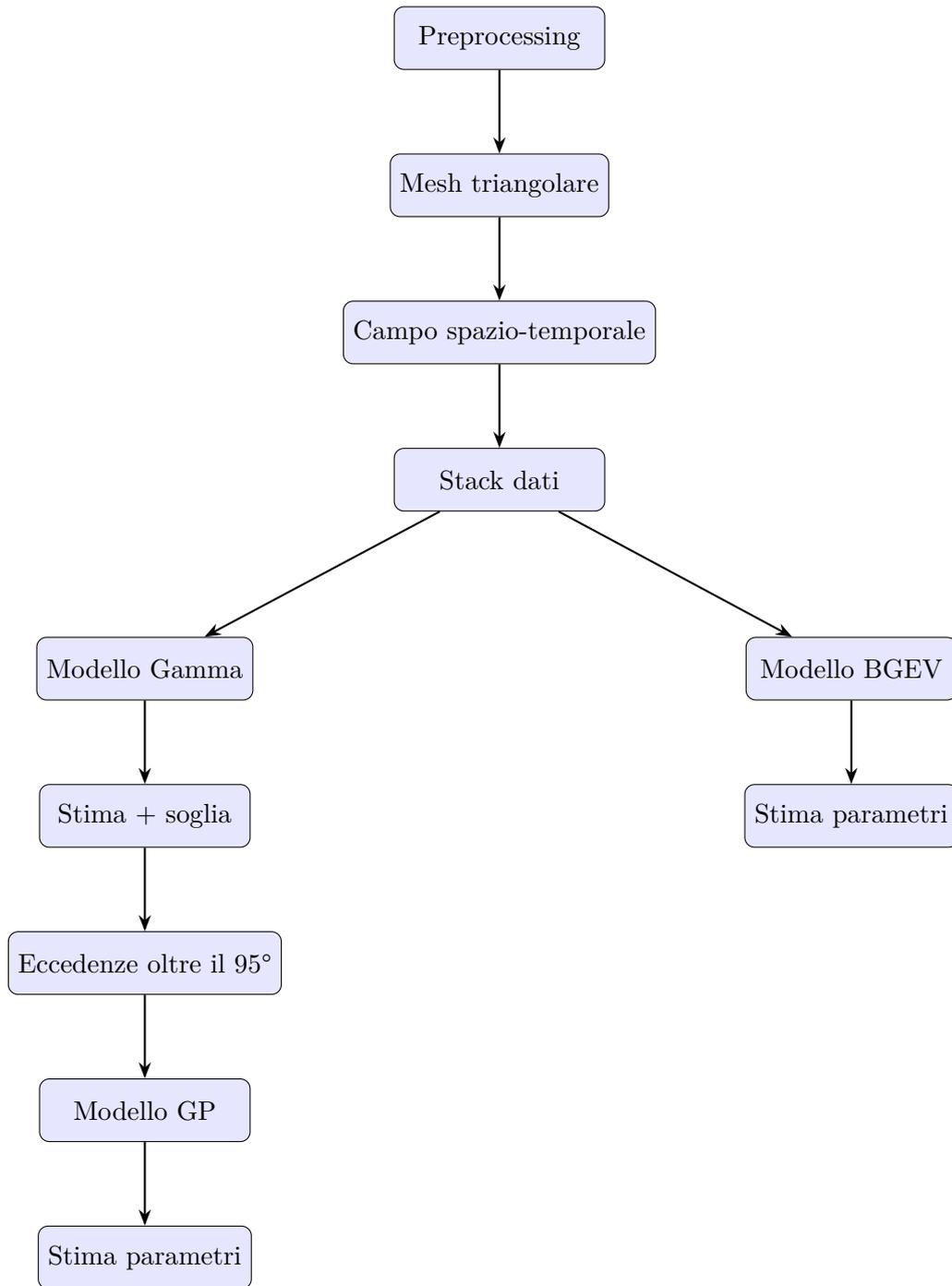


Figura 2.6: Schema dell'algoritmo in R-INLA.

Capitolo 3

Risultati

In questo capitolo vengono presentati i risultati dei due modelli spazio-temporale presentati nel capitolo 2. Verranno descritti gli effetti fissi, la struttura spaziale latente e i principali iperparametri del modello.

3.1 Stima dei parametri del Modello bGEV

Per motivi legati alla capacità computazionale, il modello bGEV è stato applicato unicamente ai dataset relativi alle durate di 10, 20, 30 minuti e 1 ora. Per ciascuna di queste durate è stato considerato un sottoinsieme temporale di 9 anni, dal 2012 al 2020 inclusi. L'analisi degli effetti fissi stimati, mostrati nella Tabella 3.1, evidenzia il ruolo significativo delle covariate nel descrivere l'intensità delle precipitazioni estreme. In particolare, vengono riportate le stime a posteriori degli effetti fissi associati alle covariate inserite nel modello per ciascuna durata considerata. Le covariate utilizzate sono comuni a tutti i modelli e includono l'intercetta, la piovosità media annua (`map`), l'indice di apertura (`openness`), la `quota` e le coordinate spaziali standardizzate. Si osserva innanzitutto un incremento netto e progressivo dell'intercetta al crescere della durata, coerente con l'attesa che eventi di precipitazione più lunghi siano mediamente più intensi. Tutti gli effetti fissi, ad eccezione delle coordinate, risultano significativamente diversi da zero, come indicato dagli intervalli di credibilità al 95% che non includono lo zero.

In particolare, l'effetto di `map` ha un effetto positivo con coefficienti che aumentano progressivamente con la durata dell'evento. Questo comportamento è atteso: aree con precipitazione media più elevata tendono a essere più soggette a eventi estremi di maggiore intensità, e tale effetto si amplifica quando si considerano durate più lunghe, passando da circa 2.1 (per 10 minuti) fino a oltre 7.0 (per 1 ora). Anche l'indice di `openness` mostra un effetto positivo e significativo in tutti i modelli. Questo risultato indica che, a parità delle altre condizioni, le aree in rilievo rispetto al terreno circostante, come crinali, dorsali o colline, tendono a essere associate a intensità maggiori degli eventi di precipitazione estrema. Dal punto di vista fisico, ciò può essere spiegato con la dinamica dell'innalzamento forzato di masse d'aria umida: quando un flusso d'aria incontra un ostacolo orografico, è costretto a salire. Durante questa risalita, l'aria si raffredda progressivamente e il vapore acqueo che contiene condensa in goccioline, dando origine a precipitazioni. Pertanto, la maggiore esposizione topografica può favorire la formazione o il rafforzamento di fenomeni convettivi localizzati. Al contrario, la quota ha un effetto negativo significativo in tutti i modelli. Questo risultato, indica che le precipitazioni estreme di breve durata tendono a concentrarsi maggiormente a quote medio-basse, mentre alle altitudini più elevate l'intensità degli eventi estremi brevi tende a ridursi. Nel complesso, i risultati confermano la rilevanza delle covariate selezionate nel descrivere la variabilità spaziale dell'intensità delle precipitazioni estreme, e mostrano come la sensibilità del modello a tali covariate aumenti con la durata dell'evento. La Tabella 3.2 riassume le stime a posteriori degli iperparametri associati alla componente spaziale latente dei modelli. In particolare, si riportano il range spaziale e la deviazione standard del campo spaziale (`Stdev`). Il parametro `Stdev` misura l'intensità della variabilità spaziale residua catturata dalla componente spaziale del modello, dopo aver tenuto conto delle covariate e non va confuso con l'errore standard riportato accanto, che riflette l'incertezza sulla stima a posteriori del parametro stesso. La deviazione standard del campo spaziale aumenta progressivamente all'aumentare della durata degli eventi passando da circa 2.1 per la durata di 10 minuti fino a circa 4.8 per la durata di un'ora. Questo suggerisce che le componenti spaziali non spiegate dal modello diventano più rilevanti man mano che si considerano eventi di durata maggiore. In altre parole, la struttura spaziale

Tabella 3.1: Effetti fissi stimati dal modello bGEV per diverse durate di precipitazione. I valori riportati sono le medie a posteriori con la rispettiva deviazione standard.

Durata	Parametro	Media	Dev. std	2.5% quantile	97.5% quantile
10 minuti	Intercept	13.983	0.313	13.369	14.597
	map	2.126	0.227	1.681	2.570
	openness	1.327	0.181	0.972	1.683
	quota	-1.522	0.173	-1.861	-1.184
	X_km_std	-0.259	0.297	-0.840	0.323
	Y_km_std	0.026	0.281	-0.525	0.577
20 minuti	Intercept	19.816	0.556	18.727	20.905
	map	3.530	0.368	2.808	4.251
	openness	1.897	0.282	1.344	2.450
	quota	-2.123	0.269	-2.650	-1.596
	X_km_std	-0.372	0.506	-1.364	0.621
	Y_km_std	-0.332	0.482	-1.277	0.614
30 minuti	Intercept	22.886	1.107	20.7718	25.055
	map	4.816	0.492	3.851	5.781
	openness	2.078	0.372	1.349	2.807
	quota	-2.222	0.352	-2.911	-1.533
	X_km_std	-0.063	0.882	-1.791	1.665
	Y_km_std	-0.780	0.849	-2.445	0.885
1 ora	Intercept	28.496	1.477	25.601	31.391
	map	7.060	0.630	5.826	8.295
	openness	2.271	0.467	1.357	3.186
	quota	-2.541	0.445	-3.413	-1.670
	X_km_std	0.355	1.136	-1.871	2.581
	Y_km_std	-1.216	1.104	-3.380	0.948

latente cattura un ruolo crescente nel descrivere la distribuzione delle precipitazioni estreme più prolungate, evidenziando la presenza di pattern spaziali persistenti che non sono completamente attribuibili alle sole covariate.

Il range, espresso in unità della mesh (1 unità corrisponde a circa 54 km), indica la distanza entro la quale due punti mantengono una forte correlazione spaziale nel campo latente. Anche in questo caso si osserva un incremento del range con

Tabella 3.2: Stime a posteriori dei parametri spaziali del campo latente per ciascuna durata, ottenute dal modello bGEV.

Durata	Parametro	Media	Dev. std	2.5% quantile	97.5% quantile
10 minuti	Range	1.006	0.175	0.711	1.399
	Stdev	2.117	0.179	1.788	2.493
20 minuti	Range	1.080	0.188	0.763	1.499
	Stdev	3.213	0.271	2.722	3.789
30 minuti	Range	1.609	0.499	0.863	2.808
	Stdev	3.731	0.543	2.761	4.895
1 ora	Range	1.668	0.258	1.210	2.225
	Stdev	4.824	0.367	4.144	5.589

la durata dell'evento, da circa ~ 54 km per la durata di 10 minuti, fino a circa ~ 90 km per la durata di un'ora. Questo significa che la correlazione spaziale si estende su distanze più ampie per eventi più lunghi, coerentemente con il fatto che le precipitazioni più persistenti tendono a coinvolgere aree geografiche più estese, mentre gli eventi più brevi sono più localizzati. Nel complesso, i risultati mostrano che la componente spaziale latente cattura fenomeni non osservabili che variano sia in intensità che in scala spaziale a seconda della durata dell'evento, e che questi effetti diventano più marcati man mano che si considerano eventi con maggiore durata.

Per quanto riguarda la dimensione temporale, la Tabella 3.3 riporta le stime del parametro ρ di autocorrelazione temporale (`GroupRho`) del campo latente, che modella la dipendenza tra gli anni successivi. I valori stimati di ρ sono compresi tra circa 0.34 e 0.65, con un generale incremento al crescere della durata. Un valore più elevato di ρ indica una maggiore persistenza temporale dei pattern spaziali, ovvero una struttura spaziale più stabile da un anno all'altro. Nel complesso, l'analisi dei parametri spaziali conferma che la struttura spazio-temporale latente cattura efficacemente le dipendenze presenti nei dati, e che tali dipendenze si intensificano con l'aumentare della durata dell'evento estremo.

Un altro aspetto importante nella modellazione delle precipitazioni massime riguarda il comportamento della coda superiore della distribuzione, che rappresenta

Tabella 3.3: Stime del parametro di autocorrelazione temporale (**GroupRho**) del campo spazio-temporale per ciascuna durata di precipitazione, ottenute dal modello bGEV.

Durata	Media	Dev. std	2.5% quantile	97.5% quantile
10 minuti	0.317	0.133	0.046	0.565
20 minuti	0.440	0.142	0.139	0.691
30 minuti	0.692	0.123	0.403	0.879
1 ora	0.651	0.053	0.535	0.742

Tabella 3.4: Stime a posteriori del parametro **tail** per ciascuna durata di precipitazione, ottenute dal modello bGEV.

Durata pioggia	Media	Dev. Std	0.025 quantile	0.975 quantile
10 min	0.014	0.011	0.002	0.043
20 min	0.033	0.018	0.012	0.082
30 min	0.044	0.022	0.015	0.098
1 h	0.122	0.015	0.094	0.121

la probabilità di osservare eventi di intensità eccezionalmente elevata. Nel modello bGEV, questo comportamento è controllato dal parametro **tail**, il cui valore stimato fornisce indicazioni sulla pesantezza della coda. In tutti i modelli stimati, il parametro di forma risulta positivo, in accordo con la struttura della distribuzione BGEV adottata. Come si osserva in Tabella 3.4, il valore medio di **tail** mostra un incremento progressivo al crescere della durata dell'evento piovoso: da circa 0.014 per la durata di 10 minuti, fino a 0.122 per la durata di un'ora. Questo andamento suggerisce che, al crescere della durata, non solo aumenta l'intensità media degli eventi estremi (come evidenziato dagli effetti fissi), ma aumenta anche la probabilità di osservare accumuli eccezionalmente elevati. In altre parole, gli eventi estremi tendono a diventare più gravi e più rari su scale temporali più lunghe. Per ultimo, la Tabella 3.5 riporta gli effetti fissi stimati per la componente di dispersione (**spread**) della distribuzione BGEV, che descrive la variabilità residua attorno alla media condizionata del processo. In altre parole, questi parametri indicano come la variabilità delle precipitazioni estreme dipenda da alcune caratteristiche orografiche,

Tabella 3.5: Effetti fissi associati alla componente di dispersione (**spread**) per ciascuna durata, ottenute dal modello bGEV. Vengono riportate le stime posteriori medie con deviazione standard e intervallo di credibilità al 95%.

Durata	Parametro	Media	Dev. std.	2.5%	97.5%
10 minuti	openness	0.002	0.022	-0.042	0.045
	quota	-0.086	0.026	-0.137	-0.035
20 minuti	openness	0.017	0.025	-0.026	0.061
	quota	-0.098	0.025	-0.148	-0.048
30 minuti	openness	0.036	0.022	-0.007	0.078
	quota	-0.104	0.025	-0.152	-0.056
1 ora	openness	0.101	0.022	0.058	0.144
	quota	-0.115	0.026	-0.166	-0.064

come l'openness e la quota.

Per quanto riguarda la quota, si osserva un effetto negativo e statisticamente significativo in tutti i modelli, abbastanza stabile per le diverse durate temporali. Questo risultato suggerisce che, a parità delle altre condizioni, la variabilità delle precipitazioni estreme tende a diminuire con l'aumentare dell'altitudine. In altre parole, alle quote più elevate gli eventi estremi mostrano una dispersione minore, ovvero una maggiore omogeneità dell'intensità.

L'effetto dell'openness, invece, mostra un pattern diverso. Per le durate brevi (10, 20 e 30 minuti) l'effetto stimato è positivo ma non statisticamente significativo, in quanto gli intervalli di credibilità al 95% includono lo zero. Tuttavia, per la durata di 1 ora, l'effetto diventa chiaramente positivo e significativo, suggerendo che, su scale temporali più lunghe, le aree più esposte (es. crinali, versanti dominanti) non solo tendono a registrare eventi più intensi, ma anche caratterizzati da una maggiore variabilità. Questo comportamento potrebbe riflettere un'influenza più marcata delle condizioni locali nella modulazione degli estremi, che rende il fenomeno più irregolare e differenziato nello spazio.

Per facilitare l'interpretazione dei risultati, si riportano i grafici relativi alla componente spazio-temporale media e alla varianza stimata del campo latente, riferiti ai dataset delle precipitazioni massime di durata 10 minuti (Figura 3.1 e

Figura 3.2) e 1 ora (Figura 3.3 e Figura 3.4) registrate in Piemonte. Per entrambe le durate considerate, emerge un effetto spaziale mediamente negativo nelle aree settentrionali, prossime al confine con la Svizzera, e nelle zone meridionali adiacenti all'Appennino Ligure. Questo pattern si mantiene sostanzialmente stabile nel tempo, coerentemente con la struttura autoregressiva temporale (AR(1)) imposta nel modello, la cui dipendenza è governata dal parametro ρ . Al contrario, nelle aree centrali della regione e in prossimità della Valle d'Aosta si osserva un effetto spaziale positivo, indicativo di livelli attesi più elevati degli eventi estremi, a parità di covariate. Questo suggerisce una maggiore intensità media delle precipitazioni di breve durata in tali zone.

3.2 Stima dei parametri del Modello GP

I modelli basati sulla distribuzione Generalized Pareto (GP), utilizzati per analizzare le eccedenze oltre una soglia prefissata, sono stati stimati per tutti i dataset disponibili. La Tabella 3.6 riporta le stime a posteriori degli effetti fissi associati alle covariate incluse nel modello GP, per cinque durate rappresentative delle precipitazioni estreme: 10 minuti, 3 ore, 24 ore, 3 giorni e 5 giorni. Per ciascun parametro (intercetta e coefficienti delle covariate) sono riportati la media e la deviazione standard della distribuzione marginale a posteriori, insieme al 25° e al 95° percentile. Anche in questo modello, emerge una chiara significatività dell'effetto della covariata `map`, che risulta positivamente associata al parametro di scala per tutte le durate, con valori medi crescenti da circa 0.15 a 0.48. Le variabili `openness` e `quota` mostrano un'influenza generalmente non significativa per la maggior parte delle durate, con un effetto significativo solo nei modelli a durata più lunga (5 giorni), dove il 2.5° percentile della distribuzione marginale supera lo zero. In particolare, `openness` tende ad avere un effetto negativo e questo potrebbe indicare che zone più "chiuse", come fondovalle o bacini montani, siano più soggette ad accumuli estremi prolungati. Infine, la coordinata spaziale `Y_km_std` mostra un effetto negativo e significativo per durate da 3 ore in su, suggerendo che le aree più settentrionali del Piemonte tendano a presentare una minore intensità degli eventi estremi oltre soglia.

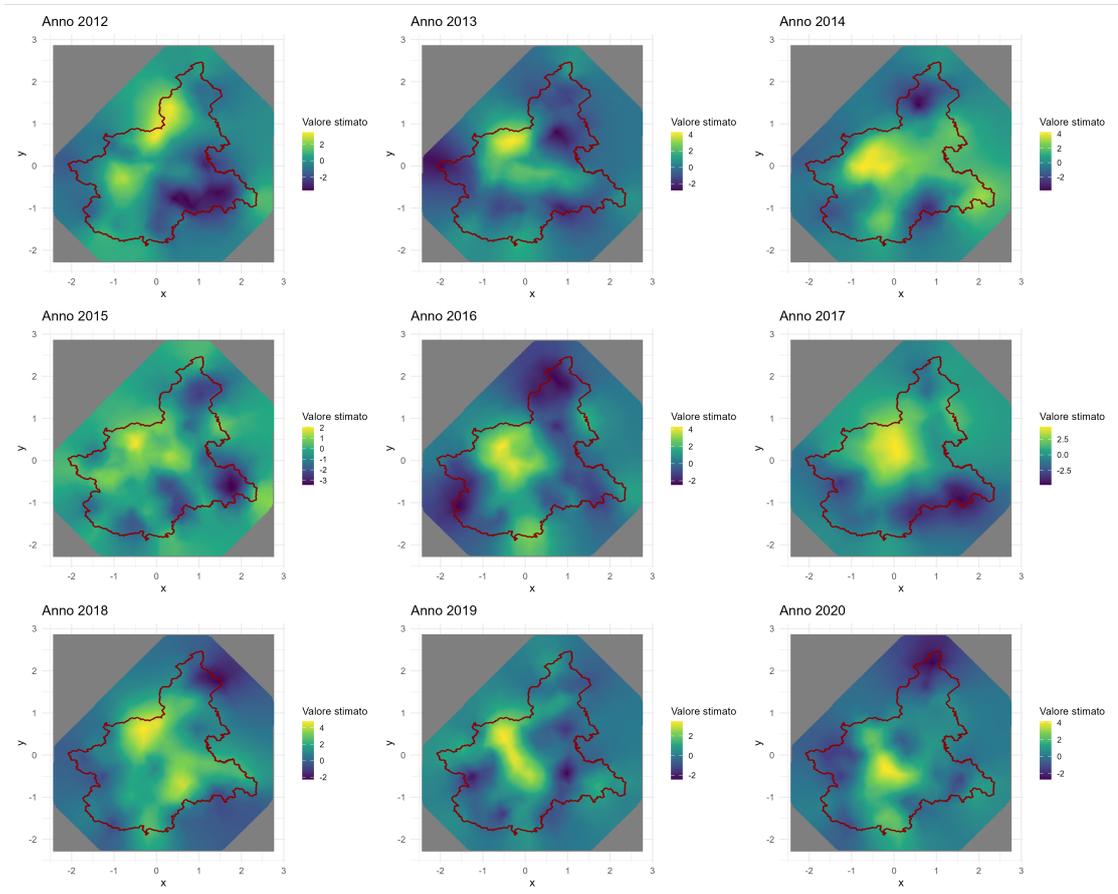


Figura 3.1: Componente spazio-temporale media stimata dal modello bGEV per il dataset con durata 10 minuti, dal 2012 al 2020. Ogni pannello mostra il valore atteso del campo latente per un anno specifico, evidenziando l'evoluzione spaziale dell'effetto.

Anche in questo caso si riportano le stime a posteriori degli iperparametri associati alla componente spaziale latente dei modelli (Tabella 3.7). In particolare, vengono mostrati il range spaziale e la deviazione standard del campo spaziale. Il range aumenta all'aumentare della durata, passando da circa 1.7 (circa 91km) per eventi

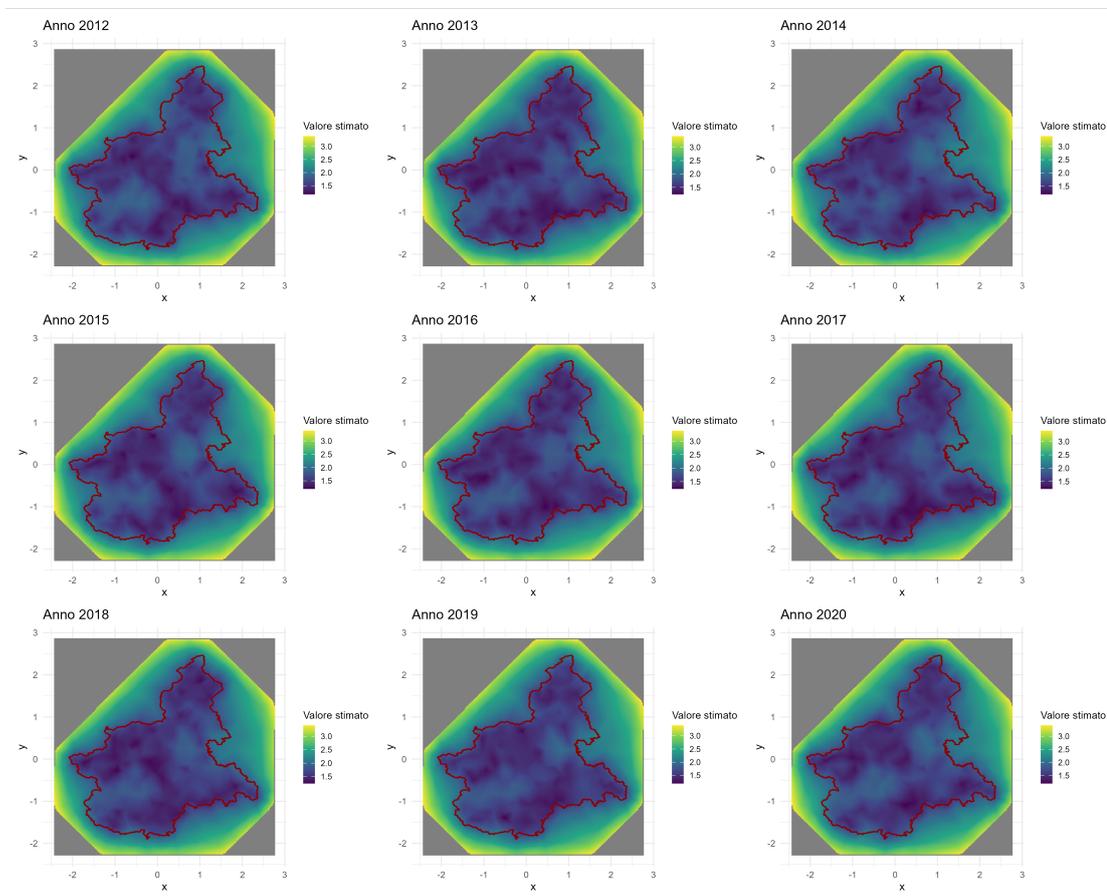


Figura 3.2: Deviazione standard a posteriori del campo latente stimato dal modello bGEV per il dataset con durata 10 minuti, dal 2012 al 2020. I valori rappresentano l'incertezza associata alla componente spaziale del modello in ciascun anno.

di 10 minuti fino a 2.4 (circa 128km) per eventi della durata di 5 giorni. Anche la deviazione standard del campo spaziale aumenta con la durata fino a raggiungere un picco a 24 ore, indicando una maggiore variabilità spaziale residua nei valori di intensità per eventi di durata media. Successivamente, si osserva una lieve

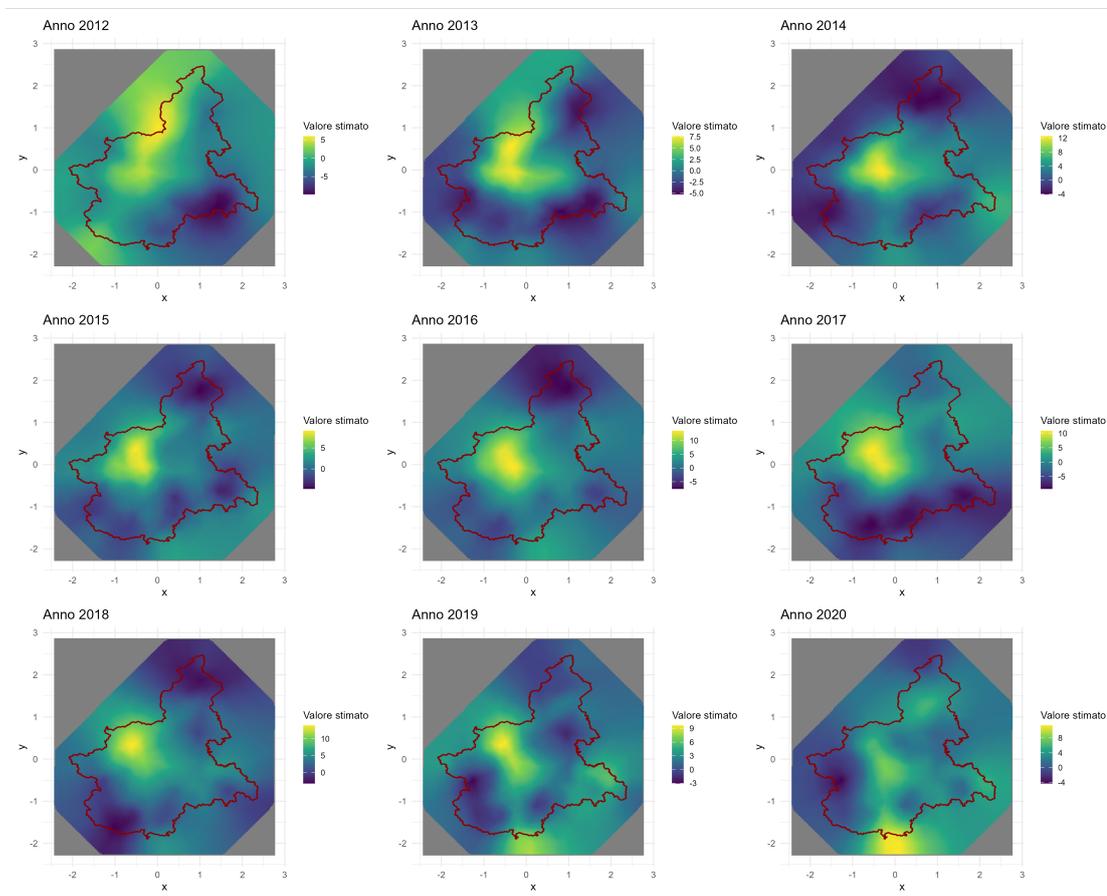


Figura 3.3: Componente spazio-temporale media stimata dal modello bGEV per il dataset con durata 1 ora, dal 2012 al 2020. Ogni pannello mostra il valore atteso del campo latente per un anno specifico, evidenziando l'evoluzione spaziale dell'effetto.

diminuzione per le durate di 3 e 5 giorni. Per quanto riguarda la dimensione temporale, la Tabella 3.8 riporta le stime del parametro ρ di autocorrelazione temporale (`GroupRho`) del campo latente, che modella la dipendenza tra gli anni successivi. Per la durata di 1 ora, l'analisi dei parametri spaziali conferma che la

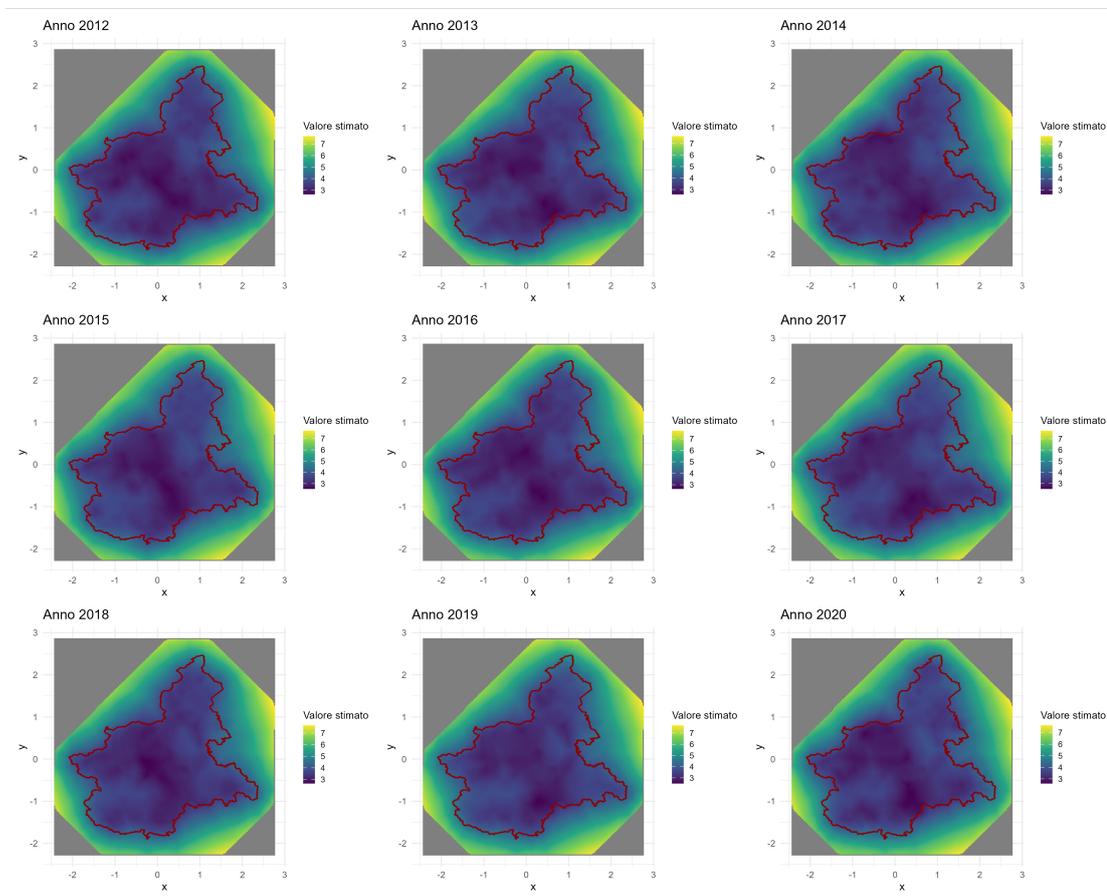


Figura 3.4: Deviazione standard a posteriori del campo latente stimato dal modello bGEV per la durata di 1 ora, dal 2012 al 2020. I valori rappresentano l'incertezza associata alla stima della componente spaziale del modello in ciascun anno.

struttura spazio-temporale latente cattura efficacemente le dipendenze presenti nei dati, e che tali dipendenze si intensificano con l'aumentare della durata dell'evento estremo. Per le altre durate invece il parametro ρ non risulta statisticamente significativo.

Tabella 3.6: Effetti fissi stimati dal modello GP per diverse durate di precipitazione. I valori riportati sono le medie a posteriori con la rispettiva deviazione standard.

Durata	Parametro	Media	Dev. std	2.5% quantile	97.5% quantile
10 minuti	Intercept	0.246	0.099	0.049	0.443
	MAP	0.151	0.068	0.018	0.286
	Openness	0.090	0.066	-0.040	0.221
	Quota	-0.023	0.066	-0.151	0.107
	X_km_std	-0.113	0.087	-0.275	0.068
	Y_km_std	0.002	0.082	-0.164	0.160
3 ore	Intercept	0.602	0.125	0.352	0.844
	MAP	0.342	0.086	0.173	0.512
	Openness	-0.042	0.077	-0.194	0.110
	Quota	-0.075	0.073	-0.218	0.067
	X_km_std	0.062	0.104	-0.143	0.267
	Y_km_std	-0.316	0.093	-0.500	-0.135
24 ore	Intercept	0.662	0.206	0.243	1.058
	MAP	0.492	0.120	0.257	0.727
	Openness	-0.161	0.093	-0.345	0.020
	Quota	-0.001	0.086	-0.169	0.169
	X_km_std	0.054	0.145	-0.231	0.342
	Y_km_std	-0.448	0.128	-0.699	-0.196
3 giorni	Intercept	0.290	0.213	-0.147	0.696
	MAP	0.349	0.121	0.114	0.589
	Openness	-0.146	0.085	-0.314	0.021
	Quota	0.057	0.079	-0.097	0.213
	X_km_std	-0.015	0.132	-0.279	0.242
	Y_km_std	-0.279	0.128	-0.534	-0.029
5 giorni	Intercept	0.138	0.225	-0.321	0.572
	MAP	0.477	0.125	0.236	0.725
	Openness	-0.182	0.088	-0.355	-0.010
	Quota	0.172	0.083	0.011	0.337
	X_km_std	-0.015	0.152	-0.012	0.278
	Y_km_std	-0.415	0.144	-0.414	-0.132

Infine, la Tabella 3.9 presenta le stime a posteriori del parametro `tail` della distribuzione GP, che governa il comportamento della coda superiore della distribuzione.

Tabella 3.7: Stime a posteriori dei parametri spaziali del campo latente, ottenute dal modello GP.

Durata	Parametro	Media	Dev. std	2.5% quantile	97.5% quantile
10 minuti	Range	1.703	0.884	0.601	3.976
	Stdev	0.292	0.084	0.157	0.483
3 ore	Range	1.227	0.526	0.510	2.541
	Stdev	0.605	0.104	0.423	0.831
24 ore	Range	2.194	0.798	1.040	4.139
	Stdev	0.720	0.111	0.524	0.961
3 giorni	Range	2.305	1.116	0.813	5.094
	Stdev	0.487	0.118	0.296	0.756
5 giorni	Range	2.406	1.675	0.579	6.814
	Stdev	0.558	0.151	0.317	0.908

Tabella 3.8: Stima del parametro di autocorrelazione temporale (GroupRho) del campo spazio-temporale per ciascuna durata di precipitazione, ottenute dal modello GP.

Durata	Media	Dev. std	2.5% quantile	97.5% quantile
10 minuti	0.392	0.404	-0.558	0.920
20 minuti	0.768	0.306	-0.174	0.996
30 minuti	0.414	0.597	-0.860	0.999
1 ora	0.939	0.090	0.685	0.998
3 ore	0.180	0.213	-0.256	0.568
6 ore	-0.025	0.235	-0.478	0.431
12 ore	0.011	0.219	-0.409	0.438
24 ore	0.072	0.244	-0.418	0.522
2 giorni	0.218	0.302	-0.407	0.738
3 giorni	0.036	0.357	-0.619	0.710
4 giorni	-0.133	0.351	-0.738	0.573
5 giorni	0.079	0.436	-0.722	0.837

In tutti i modelli stimati, il parametro di forma risulta positivo. Si osserva un progressivo aumento del parametro fino alla durata delle 24 ore, per poi diminuire di nuovo per le durate giornaliere.

Tabella 3.9: Stime a posteriori del parametro `tail` per ciascuna durata di precipitazione, ottenute dal modello GP.

Durata pioggia	Media	Dev. Std	0.025 quantile	0.975 quantile
10 min	0.208	0.044	0.128	0.300
20 min	0.308	0.037	0.234	0.378
30 minuti	0.343	0.041	0.258	0.419
1 ora	0.363	0.036	0.287	0.429
3 ore	0.314	0.061	0.186	0.421
6 ore	0.338	0.053	0.225	0.429
12 ore	0.408	0.047	0.298	0.477
24 ore	0.475	0.023	0.414	0.498
2 giorni	0.262	0.048	0.168	0.354
3 giorni	0.247	0.049	0.152	0.343
4 giorni	0.269	0.048	0.174	0.361
5 giorni	0.261	0.056	0.152	0.368

Per visualizzare e confrontare le dipendenze spazio-temporali catturate dal modello, si riportano i campi spaziali stimati per ciascun anno nel periodo 2012–2020, sia per il dataset a risoluzione temporale di 10 minuti (Figura 3.5) sia per quello a risoluzione di 1 ora (Figura 3.6). Questo confronto è particolarmente utile per evidenziare come la struttura di dipendenza vari al variare della risoluzione temporale dei dati: nel caso a 10 minuti, il modello rileva principalmente una componente spaziale, mentre per la durata di 1 ora emerge in modo più evidente anche una componente temporale, caratterizzata da una dipendenza positiva tra gli anni, come indicato dal parametro ρ dell'AR(1). Nella Figura 3.5 si osserva il comportamento del campo spaziale per nove anni consecutivi nel caso dei dati a 10 minuti. Sebbene l'intervallo di valori stimati non sia costante tra gli anni, emerge una variabilità significativa da un anno all'altro, senza una struttura temporale coerente. Tuttavia, a partire dal 2018 si osserva una certa stabilità nella configurazione spaziale del campo: i valori più elevati (rappresentati in giallo nelle mappe) si concentrano con maggiore frequenza nella zona di Torino e dintorni. Questo può essere interpretato come un segnale di intensificazione degli eventi estremi in quell'area negli ultimi anni. In particolare, l'anno 2020 si distingue per una maggiore intensità del campo

spaziale latente: i valori stimati risultano più estremi rispetto agli anni precedenti, con un contrasto più marcato tra le diverse zone del territorio. Questo indica che nel 2020 la distribuzione spaziale degli effetti latenti è stata più disomogenea, con alcune aree che presentano valori significativamente più alti (soprattutto attorno a Torino) e altre con valori molto più bassi. Tale evidenza suggerisce un aumento della variabilità spaziale degli eventi estremi proprio in quell'anno.

Per quanto riguarda invece il dataset a risoluzione oraria, rappresentato in Figura 3.6, il comportamento del campo spazio-temporale risulta notevolmente più coerente nel tempo. In particolare, si osserva una configurazione spaziale relativamente stabile, con valori negativi (blu) predominanti nei primi anni del periodo considerato (2012–2015), seguiti da un progressivo incremento del campo, che diventa più positivo negli anni successivi. Questo comportamento riflette la presenza di una componente temporale significativa, coerente con il valore stimato positivo del parametro ρ del processo AR(1), che suggerisce una persistenza nel tempo del campo latente. La presenza di dipendenza temporale ben definita in questo caso può essere legata alla maggiore stabilità statistica della serie oraria, che essendo meno rumorosa rispetto a quella a 10 minuti, consente al modello di catturare più efficacemente l'evoluzione temporale della componente spaziale latente.

A differenza dei grafici precedenti, nelle Figure 3.8 e 3.7 l'obiettivo non è osservare l'effetto medio del campo, ma valutarne l'incertezza stimata nello spazio e nel tempo. In questo caso è possibile fare un discorso abbastanza analogo. In particolare, nel complesso, si osserva una struttura della varianza piuttosto stabile nel corso degli anni, con valori generalmente più elevati lungo i bordi del dominio spaziale e valori più bassi al centro, in corrispondenza della zona più densamente osservata. Questa distribuzione è coerente con le aspettative: le regioni periferiche della mesh triangolare presentano una maggiore incertezza, dovuta alla scarsità o assenza di dati osservati in quelle aree, mentre la varianza si riduce nelle zone centrali dove la densità di stazioni è maggiore e questo permette stime più precise. Confrontando i diversi anni, non si evidenziano variazioni sostanziali della varianza tra un anno e l'altro; questo suggerisce che il livello di incertezza associato alla componente spaziale del modello è rimasto relativamente costante nel tempo. In sintesi, l'analisi della varianza del campo evidenzia che il modello è stato in grado di fornire stime

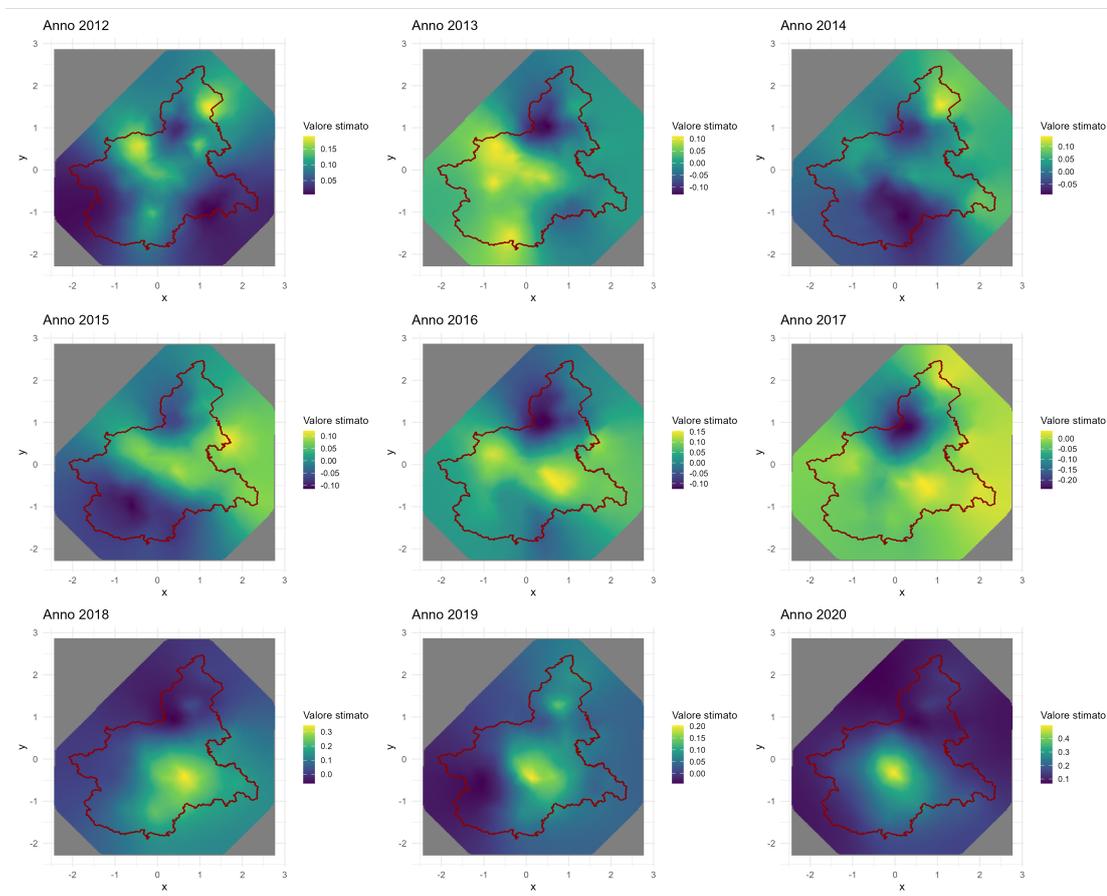


Figura 3.5: Componente spazio-temporale media stimata dal modello GP per il dataset con durata 10 minuti, dal 2012 al 2020. Ogni pannello mostra il valore atteso del campo latente per un anno specifico, evidenziando l'evoluzione spaziale dell'effetto.

stabili nelle aree centrali del dominio, con incertezza più elevata verso le regioni meno coperte dalle osservazioni.

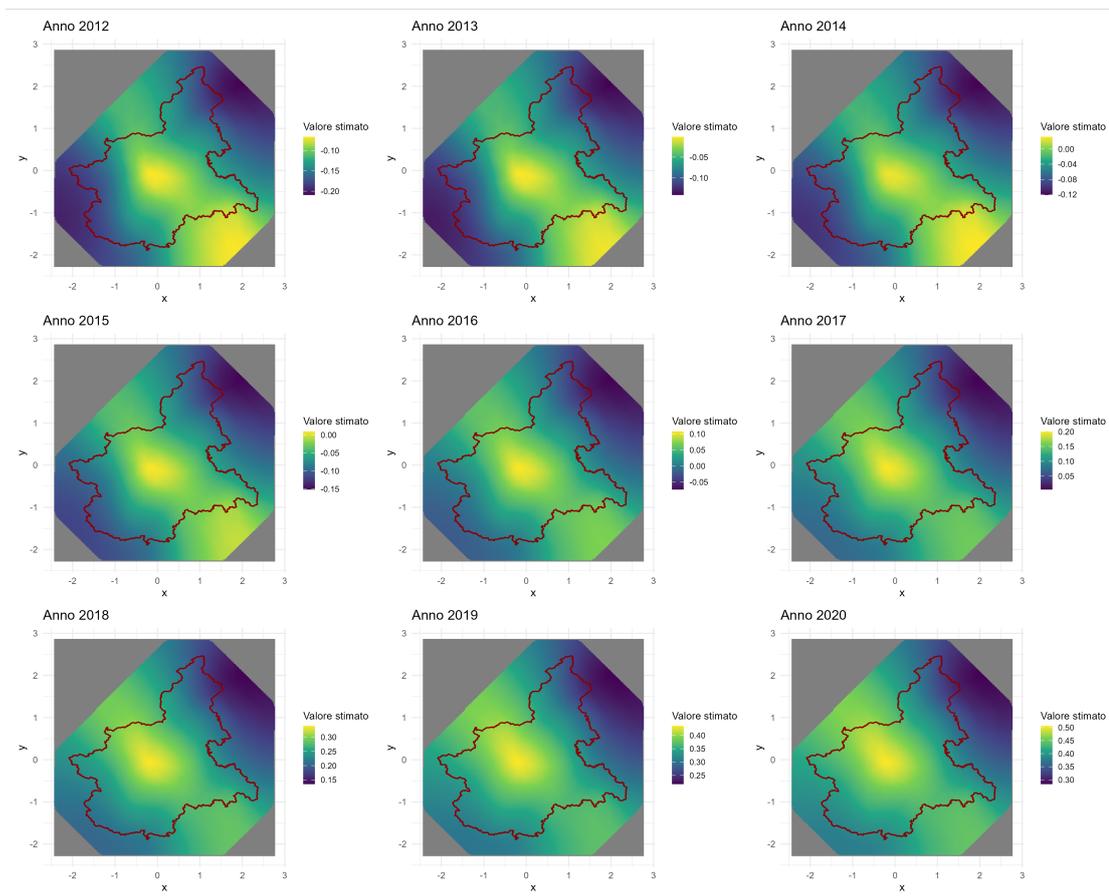


Figura 3.6: Componente spazio-temporale media stimata dal modello GP per il dataset con durata 1 ora, dal 2012 al 2020. Ogni pannello mostra il valore atteso del campo latente per un anno specifico, evidenziando la persistenza temporale e l'evoluzione spaziale dell'effetto.

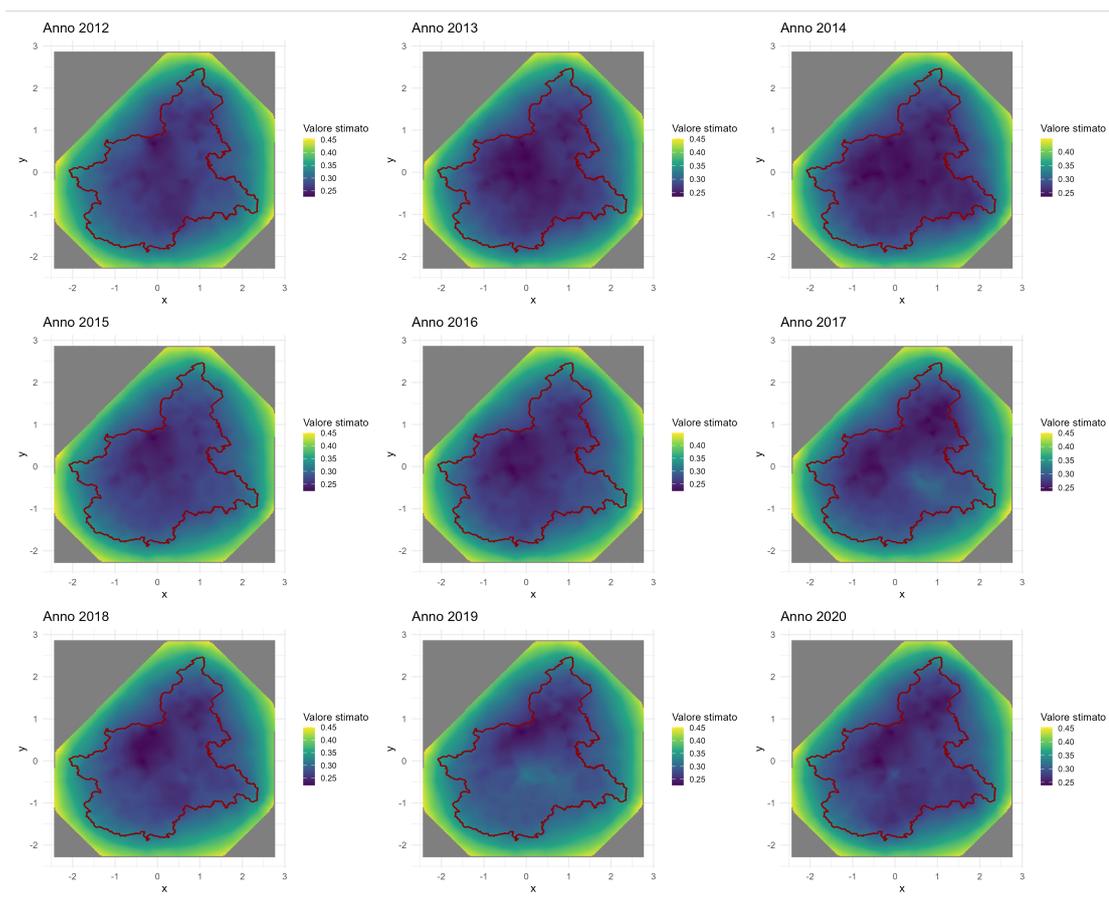


Figura 3.7: Varianza del campo latente stimato dal modello GP per il dataset con durata 1 ora, dal 2012 al 2020. I valori rappresentano l'incertezza associata alla componente spaziale del modello in ciascun anno.

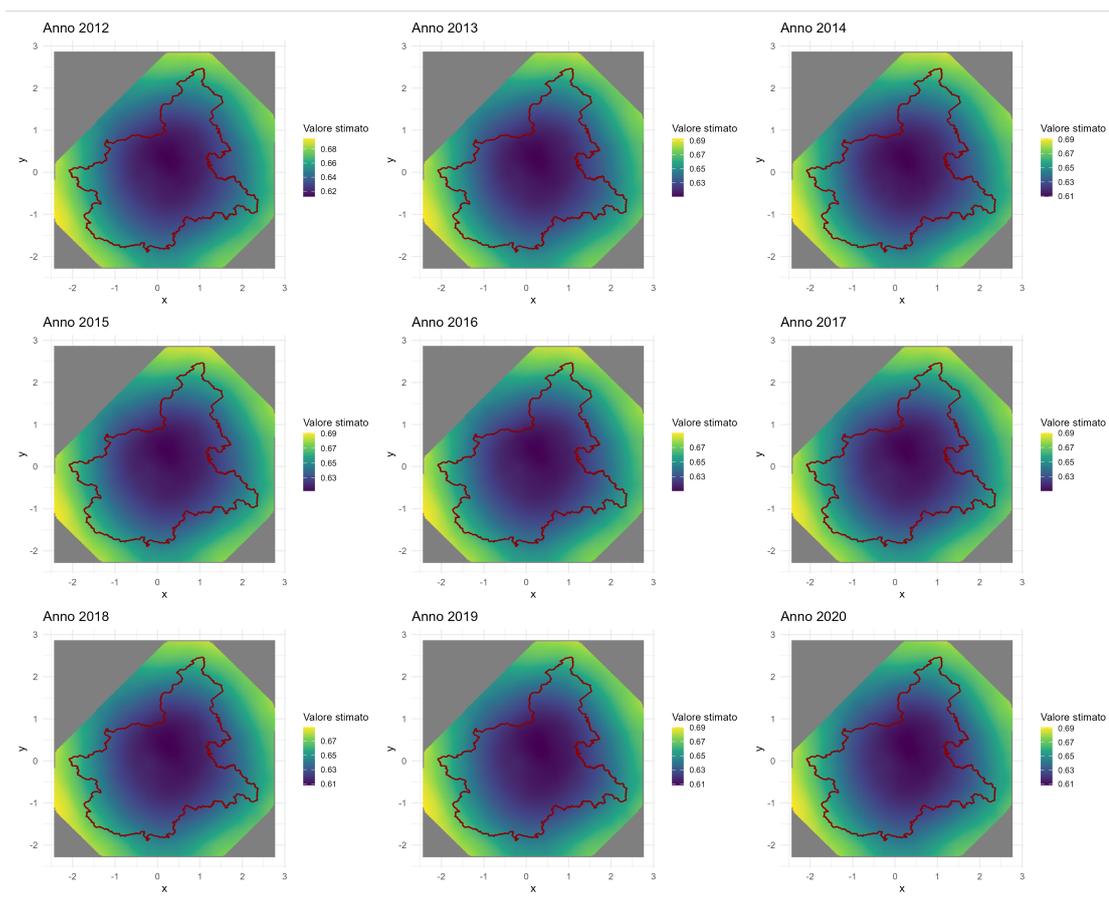


Figura 3.8: Varianza del campo latente stimato dal modello GP per il dataset con durata 1 ora, dal 2012 al 2020. I valori rappresentano l'incertezza associata alla componente spaziale del modello in ciascun anno.

Capitolo 4

Conclusioni

In questo lavoro di tesi è stato sviluppato un modello spazio-temporale per l'analisi delle precipitazioni estreme in Piemonte, utilizzando l'approccio bayesiano integrato implementato nel pacchetto R-INLA. L'analisi ha riguardato diverse scale temporali, dai 10 minuti ai 5 giorni, e si è basata su un ampio database di massimi annuali, con l'obiettivo di modellare la variabilità spaziale e temporale degli eventi estremi. Dal punto di vista metodologico, la tesi ha mostrato come l'approccio SPDE (Stochastic Partial Differential Equation) consenta una rappresentazione efficiente del campo spaziale latente tramite un processo gaussiano a media nulla e covarianza Matérn, discretizzato su una mesh triangolare. Inoltre, la formulazione autoregressiva di ordine 1 (AR(1)) ha permesso di incorporare in modo naturale la dipendenza temporale nella componente spaziale latente, migliorando la coerenza nella modellazione della dinamica spazio-temporale.

Sono stati stimati due modelli distinti. Il primo, basato sulla famiglia *blended Generalized Extreme Value* (bGEV), è stato utilizzato per descrivere l'intera distribuzione dei massimi annuali, fornendo una visione complessiva del comportamento delle precipitazioni estreme. Il secondo, con famiglia *Generalized Pareto* (GP), è stato invece dedicato allo studio delle eccedenze sopra soglia, con l'obiettivo di caratterizzare la coda della distribuzione e valutare il rischio associato agli eventi più intensi.

Il lavoro apre la strada a diversi sviluppi futuri. Potrebbe essere interessante

valutare la capacità predittiva del modello su dati di test o su eventi estremi recenti, o anche estenderlo per includere ulteriori covariate. Inoltre, altri approfondimenti potrebbero riguardare la sensibilità del modello alla scelta delle prior distribuite sugli iperparametri, oppure l'applicazione del modello bGEV anche alle durate più lunghe, al fine di confrontare il comportamento degli eventi estremi su scale temporali differenti.

Nel complesso, il lavoro ha evidenziato l'efficacia dell'approccio bayesiano con INLA nel modellare componenti spazio-temporali complesse, non catturate dalle sole covariate, fornendo stime interpretabili e geograficamente dettagliate.

Bibliografia

- [1] Javier Prieto Tejedor. *Bayesian Inference*. Rijeka: IntechOpen, nov. 2017. ISBN: 978-953-51-3578-4. DOI: 10.5772/66264. URL: <https://doi.org/10.5772/66264> (cit. a p. 2).
- [2] Håvard Rue, Sara Martino e Nicolas Chopin. «Approximate Bayesian Inference for Latent Gaussian models by using Integrated Nested Laplace Approximations». In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 71.2 (apr. 2009), pp. 319–392. ISSN: 1369-7412. DOI: 10.1111/j.1467-9868.2008.00700.x. eprint: https://academic.oup.com/jrsssb/article-pdf/71/2/319/49686253/jrsssb_71_2_319.pdf. URL: <https://doi.org/10.1111/j.1467-9868.2008.00700.x> (cit. a p. 3).
- [3] Paula Moraga. *Geospatial Health Data: Modeling and Visualization with R-INLA and Shiny*. Chapman & Hall/CRC Biostatistics Series. Chapman & Hall/CRC, 2019. ISBN: 9780367357955 (cit. alle pp. 3, 4, 7, 8, 13–15).
- [4] Håvard Rue, Andrea Riebler, Sigrunn Sørbye, Janine Illian, Daniel Simpson e Finn Lindgren. «Bayesian computing with INLA: A review». In: *Annual Review of Statistics and Its Application* 4 (apr. 2016). DOI: 10.1146/annurev-statistics-060116-054045 (cit. alle pp. 3, 5).
- [5] Marta Blangiardo, Michela Cameletti, Gianluca Baio e Håvard Rue. «Spatial and spatio-temporal models with R-INLA». In: *Spatial and Spatio-temporal Epidemiology* 4 (2013), pp. 33–49. ISSN: 1877-5845. DOI: <https://doi.org/10.1016/j.sste.2012.12.001>. URL: <https://www.sciencedirect.com/science/article/pii/S1877584512000846> (cit. alle pp. 4, 12, 23).

- [6] Wikipedia. *Serie di Taylor* — *Wikipedia, L'enciclopedia libera*. [Online; in data 2-giugno-2025]. 2025. URL: [%5Curl%7B//it.wikipedia.org/w/index.php?title=Serie_di_Taylor&oldid=144734540%7D](https://it.wikipedia.org/w/index.php?title=Serie_di_Taylor&oldid=144734540) (cit. a p. 5).
- [7] Andrew Gelman, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari e Donald B. Rubin. *Bayesian Data Analysis*. English. 3rd ed. United States: CreateSpace, 2013. ISBN: 9781439840955 (cit. a p. 6).
- [8] Precision Analytics. *A Gentle INLA Tutorial*. <https://www.precision-analytics.ca/articles/a-gentle-inla-tutorial/>. Accessed: 2025-06-02. 2023 (cit. a p. 8).
- [9] Virgilio Gómez-Rubio. *Bayesian Inference with INLA*. Boca Raton, FL: Chapman & Hall/CRC Press, 2020 (cit. a p. 8).
- [10] Elias T. Krainski, Virgilio Gómez-Rubio, Haakon Bakka, Andrea Lenzi, Daniela Castro-Camilo, Dan Simpson, Finn Lindgren e Håvard Rue. *Advanced Spatial Modeling with Stochastic Partial Differential Equations Using R and INLA*. Boca Raton, FL: Chapman e Hall/CRC, 2019 (cit. a p. 9).
- [11] Ro'fah Nur Rachmawati, Anik Djuraidah, Anwar Fitrianto e I Made Sumer-tajaya. «Spatio-temporal Models Using R-INLA with Generalized Extreme Value Distribution in Hierarchical Bayes Regression». In: *International Journal of Scientific Research in Science, Engineering and Technology IJSRSET* 4 (2018). ISSN: 2394-4099 (cit. a p. 12).
- [12] Romano Nunzio. *Applicazioni di tecniche geostatistiche per la caratterizzazione idraulica dei suoli*. <https://www.academia.edu/25390555/>. Rivista di Ingegneria Agraria, Dipartimento di Ingegneria Agraria e Agronomia del Territorio, Università degli Studi di Napoli "Federico II". 2000 (cit. a p. 14).
- [13] Stuart Coles. «Classical Extreme Value Theory and Models». In: *An Introduction to Statistical Modeling of Extreme Values*. London: Springer London, 2001, pp. 45–73. ISBN: 978-1-4471-3675-0. DOI: 10.1007/978-1-4471-3675-0_3. URL: https://doi.org/10.1007/978-1-4471-3675-0_3 (cit. alle pp. 15, 20).

- [14] Daniela Castro-Camilo. *Blended GEV: a tutorial using R-INLA*. Ver. 1.2. Available from <https://www.r-inla.org/doc/examples/blended-gev.pdf>. 2023 (cit. alle pp. 19, 22).
- [15] Thomas Opitz, Raphaël Huser, Haakon Bakka e Havard Rue. «INLA goes extreme: Bayesian tail regression for the estimation of high spatio-temporal quantiles». In: *Extremes* 21.3 (2018), pp. 441–462. DOI: 10.1007/s10687-018-0324-x. URL: <https://hal.inrae.fr/hal-02626414> (cit. a p. 22).
- [16] Elisa Bartolini, Paola Allamano, Francesco Laio e Pierluigi Claps. *Analisi spaziale delle precipitazioni medie ed intense su Piemonte e Valle d’Aosta*. Technical report, Dipartimento di Idraulica, Trasporti ed Infrastrutture Civili, Politecnico di Torino. 2011 (cit. a p. 24).
- [17] Thomas Opitz, Raphaël Huser, Haakon Bakka e Håvard Rue. «INLA goes extreme: Bayesian tail regression for the estimation of high spatio-temporal quantiles». In: *Extremes* 21 (set. 2018). DOI: 10.1007/s10687-018-0324-x (cit. a p. 35).