

POLITECNICO DI TORINO

**Corso di Laurea
in Ingegneria Matematica**

Tesi di Laurea Magistrale

Teoria dei valori estremi per l'analisi spaziale delle precipitazioni estreme in Piemonte



Relatore

Prof. Gianluca Mastrantonio

Co-relatori:

Dott.sa Paola Mazzoglio

Prof. Alberto Viglione

Candidato

Maria Cannistrà

Anno Accademico 2023-2024

Sommario

I valori estremi rappresentano le osservazioni che si discostano in modo significativo dalla media di un insieme di dati e possono avere un impatto rilevante in un determinato fenomeno. Tali eventi estremi possono verificarsi in aree geografiche ampie, per cui la componente spaziale riveste notevole importanza, ad esempio nell'ambito della meteorologia e dell'idrologia. Al fine di ridurre i rischi associati all'avvenimento di eventi estremi, risulta necessario riuscire a prevederne l'intensità, la probabilità di accadimento e il territorio interessato.

Il presente lavoro presenta una revisione approfondita della teoria dei valori estremi. Successivamente presenta i risultati ottenuti dall'analisi su un dataset di precipitazioni massime annuali registrate nella regione Piemonte. In particolare, sono stati confrontati diversi modelli max-stable spaziali, tenendo conto delle coordinate geografiche delle stazioni pluviometriche, su dati dei massimi annui di precipitazione di breve durata (1 ora) e lunga durata (24 ore).

I risultati ottenuti mostrano un comportamento differente tra le precipitazioni di durata 1 e 24 ore, sia per quanto riguarda le zone interessate da eventi particolarmente estremi e sia per quanto riguarda la dipendenza spaziale tra le diverse località sulla base della loro distanza.

Indice

Lista delle figure.....	iii
Lista delle tabelle.....	vi
1. Introduzione.....	1
2. Teoria dei valori estremi.....	3
2.1 Estremi univariati	3
2.1.1 Formulazione del modello asintotico per i massimi	3
2.1.2 Modello asintotico per minimi	7
2.1.3 Generalizzazione al caso di variabili aleatorie dipendenti	8
2.1.4 Caratterizzazione come processi puntuali	9
2.2 Estremi multivariati.....	12
2.2.1 Estremi massimi in ogni componente.....	12
2.2.2 Misure di dipendenza asintotica	16
2.2.3 Variabili di struttura	17
2.2.4 Modello di processo puntuale.....	17
3. Estremi spaziali	19
3.1 Processi max-stable	19
3.1.1 Rappresentazione spettrale	20
3.1.2 Distribuzioni di dimensione finita	22
3.2 Modelli	23
3.2.1 Processo di Smith	23
3.2.2 Processo di Schlather.....	24
3.2.3 Processo di Brown-Resnick.....	25
3.2.4 Processo Extremal-t.....	25
3.3 Dipendenza spaziale	26
3.3.1 Funzione del coefficiente estremo	26
3.3.2 Funzione di probabilità di concomitanza estrema	28
3.4 Simulazione di un processo max-stable	30
4. Inferenza e valutazione di un modello spaziale	34
4.1 Distribuzione marginale GEV arbitraria	34
4.1.1 Stima dei parametri di un modello GEV spaziale	35
4.1.2 Stima dei parametri di un processo max-stable	36
4.2 Selezione del modello	37
4.2.1 TIC.....	37
4.2.2 ANOVA.....	38

4.2.3	Cross-validation.....	38
4.3	Visualizzazione dei dati.....	39
4.3.1	Coefficiente estremo.....	39
4.3.2	Probabilità di concomitanza estrema.....	40
4.3.3	Return levels.....	41
5.	Dataset delle precipitazioni estreme	42
5.1	Descrizione del dataset.....	42
5.2	Caratteristiche del territorio.....	43
5.2.1	Caratteristiche delle stazioni pluviometriche.....	44
5.3	Analisi esplorativa del dataset di precipitazioni	46
5.3.1	Distribuzione temporale e spaziale.....	46
6.	Risultati	49
6.1	Metodologie.....	49
6.2	Precipitazioni massime annuali di durata 1 ora	50
6.2.1	Dataset completo (1928-2020)	50
6.2.2	Anni 1960-2020.....	55
6.2.3	Anni 2000-2020.....	59
6.3	Precipitazioni massime annuali di durata 24 ore	63
6.3.1	Dataset completo (1928-2020)	63
6.3.2	Anni 1960-2020.....	68
6.3.3	Anni 2000-2020.....	71
	Conclusioni	77
A	Risultati supplementari.....	79
B	Codice.....	86
	Bibliografia	96

Lista delle figure

Figura 5.1: Mappa altimetrica del Piemonte e delle aree circostanti. I punti neri rappresentano le stazioni pluviometriche del Piemonte per cui si hanno i valori di precipitazione.	43
Figura 5.2: Distribuzione della quota delle stazioni pluviometriche con risoluzione di 1 km rappresentata da un istogramma (a) e da un boxplot (b).	44
Figura 5.3: Mappa spaziale del valore di <i>openness</i>	45
Figura 5.4: Mappa spaziale (a) e istogramma (b) del valore di precipitazione media annuale (MAP_BIGBANG) per ciascuna stazione pluviometrica.	45
Figura 5.5: Istogramma delle precipitazioni massime annuali di durata 1 ora (a) e 24 ore (b).	47
Figura 5.6: Distribuzione temporale delle precipitazioni massime annuali.	47
Figura 5.7: Istogramma del numero di stazioni pluviometriche che hanno registrato i valori di precipitazioni massime annuali di durata 1 ora (a) e 24 ore (b) rispetto agli anni di osservazione.	48
Figura 6.1: <i>Symbol plot</i> che mostra, per ciascuna stazione pluviometrica, la deviazione della media campionaria locale rispetto alla media campionaria complessiva (a) e la deviazione standard (b) per le precipitazioni di durata 1 ora e l'intero periodo di osservazione (1928-2020).	51
Figura 6.2: Dipendenza spaziale stimata empiricamente dai dati di precipitazione di durata 1 ora per l'intero periodo di osservazione (1928-2020), tramite coefficiente estremo (a, b) e probabilità di concomitanza estrema (c). I punti grigi sono le stime empiriche, mentre i punti neri sono le stime discretizzate con 12000 <i>bins</i> , con ingrandimento sulla versione discretizzata per il coefficiente estremo (b).	52
Figura 6.3: Valutazione della dipendenza spaziale stimata dai modelli max-stable per le precipitazioni di durata 1 ora per l'intero periodo di osservazione (1928-2020), tramite funzione del coefficiente estremo per il modello di Brown-Resnick (a) e per tutti i modelli (b), e funzione di probabilità di concomitanza estrema per ogni modello (c).	53
Figura 6.4: Valutazione locale del modello max-stable Brown-Resnick per i dati di precipitazione di durata 1 ora e per l'intero periodo di osservazione (1928-2020), tramite <i>return level plots</i> (a, b, c, d) per 4 stazioni pluviometriche.	55
Figura 6.5: <i>25-year return levels</i> (a) ed errori standard (b) per l'intero territorio del Piemonte e le aree circostanti per precipitazioni di durata 1 ora e l'intero periodo di osservazione (1928-2020).	55
Figura 6.6: <i>Symbol plot</i> che mostra per ciascuna stazione pluviometrica la deviazione della media campionaria locale rispetto alla media campionaria complessiva (a) e la deviazione standard (b) per le precipitazioni di durata 1 ora e periodo di osservazione 1960-2020.	57
Figura 6.7: Dipendenza spaziale stimata empiricamente dai dati di precipitazione di durata 1 ora per il periodo di osservazione 1960-2020, tramite coefficiente estremo (a, b) e probabilità di concomitanza estrema (c). I punti grigi sono le stime empiriche, mentre i punti neri sono le stime discretizzate con 10000 <i>bins</i> , con ingrandimento sulla versione discretizzata per il coefficiente estremo (b).	57
Figura 6.8: Valutazione della dipendenza spaziale stimata dal modello Extremal-t per le precipitazioni di durata 1 ora per il periodo di osservazione 1960-2020, tramite funzione del coefficiente estremo (a, b), e funzione di probabilità di concomitanza estrema (c). I punti grigi sono le stime empiriche calcolate rispetto alle osservazioni e i punti neri rappresentano la versione discretizzata.	58
Figura 6.9: Valutazione locale del modello max-stable Extremal-t per i dati di precipitazione di durata 1 ora e per il periodo di osservazione 1960-2020, tramite <i>return level plots</i> (a, b, c, d) per 4 stazioni pluviometriche.	58

Figura 6.10: <i>25-year return levels</i> (a) ed errori standard (b) per l'intero territorio del Piemonte e le aree circostanti per precipitazioni di durata 1 ora e il periodo di osservazione 1960-2020.....	59
Figura 6.11: <i>Symbol plot</i> che mostra per ciascuna stazione pluviometrica la deviazione della media campionaria locale rispetto alla media campionaria complessiva (a) e la deviazione standard (b) per le precipitazioni di durata 1 ora e periodo di osservazione 2000-2020.....	60
Figura 6.12: Dipendenza spaziale stimata empiricamente dai dati di precipitazione di durata 1 ora per il periodo di osservazione 2000-2020, tramite coefficiente estremo (a, b) e probabilità di concomitanza estrema (c). I punti grigi sono le stime empiriche, mentre i punti neri sono le stime discretizzate con 400 <i>bins</i> , con ingrandimento sulla versione discretizzata per il coefficiente estremo (b).....	61
Figura 6.13: Dipendenza spaziale stimata dai modelli M1 e M2 per le precipitazioni di durata 1 ora per il periodo di osservazione 2000-2020, tramite funzione del coefficiente estremo (a, b), e funzione di probabilità di concomitanza estrema (c). I punti grigi sono le stime empiriche calcolate rispetto alle osservazioni e i punti neri rappresentano la versione discretizzata.....	62
Figura 6.14: Valutazione locale dei modelli max-stable M1 e M2 selezionati per i dati di precipitazione di durata 1 ora e per il periodo di osservazione 2000-2020, tramite <i>return level plots</i> (a, b, c, d) per M1 e (a', b', c', d') per M2 per 4 stazioni pluviometriche.....	62
Figura 6.15: <i>25-year return levels</i> (a) ed errori standard (b) per l'intero territorio del Piemonte e le aree circostanti per precipitazioni di durata 1 ora e il periodo di osservazione 2000-2020 predetti dal modello M1.....	64
Figura 6.16: <i>25-year return levels</i> (a) ed errori standard (b) per l'intero territorio del Piemonte e le aree circostanti per precipitazioni di durata 1 ora e il periodo di osservazione 2000-2020 predetti dal modello M2.....	64
Figura 6.17: <i>Symbol plot</i> che mostra, per ciascuna stazione pluviometrica, la deviazione della media campionaria locale rispetto alla media campionaria complessiva (a) e la deviazione standard (b) per le precipitazioni di durata 24 ore e l'intero periodo di osservazione (1928-2020).....	65
Figura 6.18: Dipendenza spaziale stimata empiricamente dai dati di precipitazione di durata 24 ore per l'intero periodo di osservazione (1928-2020), tramite coefficiente estremo (a, b) e probabilità di concomitanza estrema (c). I punti grigi sono le stime empiriche, mentre i punti neri sono le stime discretizzate con 11000 <i>bins</i> , con ingrandimento sulla versione discretizzata per il coefficiente estremo (b).....	65
Figura 6.19: Valutazione della dipendenza spaziale stimata dai modelli max-stable per le precipitazioni di durata 24 ore per l'intero periodo di osservazione (1928-2020), tramite funzione del coefficiente estremo per il modello di Brown-Resnick (a) e per tutti i modelli (b), e funzione di probabilità di concomitanza estrema per ogni modello (c).....	67
Figura 6.20: Valutazione locale del modello max-stable di Brown-Resnick per i dati di precipitazione di durata 24 ore e per l'intero periodo di osservazione (1928-2020), tramite <i>return level plots</i> (a, b, c, d) per 4 stazioni pluviometriche.....	68
Figura 6.21: <i>25-year return levels</i> (a) ed errori standard (b) per l'intero territorio del Piemonte e le aree circostanti per precipitazioni di durata 24 ore e l'intero periodo di osservazione (1928-2020).....	68
Figura 6.22: <i>Symbol plot</i> che mostra per ciascuna stazione pluviometrica la deviazione della media campionaria locale rispetto alla media campionaria complessiva (a) e la deviazione standard (b) per le precipitazioni di durata 24 ore e periodo di osservazione 1960-2020.....	69
Figura 6.23: Dipendenza spaziale stimata empiricamente dai dati di precipitazione di durata 24 ore per il periodo di osservazione 1960-2020, tramite coefficiente estremo (a, b) e probabilità di concomitanza estrema (c). I punti grigi sono le stime empiriche, mentre i punti neri sono le stime discretizzate con 10000 <i>bins</i> , con ingrandimento sulla versione discretizzata per il coefficiente estremo (b).....	69

Figura 6.24: Valutazione della dipendenza spaziale stimata dal modello Schlather per le precipitazioni di durata 24 ore per il periodo di osservazione 1960-2020, tramite funzione del coefficiente estremo (a, b), e funzione di probabilità di concomitanza estrema (c). I punti grigi sono le stime empiriche calcolate rispetto alle osservazioni e i punti neri rappresentano la versione discretizzata.	71
Figura 6.25: Valutazione locale del modello max-stable Schlather per i dati di precipitazione di durata 24 ore e per il periodo di osservazione 1960-2020, tramite <i>return level plots</i> (a, b, c, d) per 4 stazioni pluviometriche.....	71
Figura 6.26: <i>25-year return levels</i> (a) ed errori standard (b) per l'intero territorio del Piemonte e le aree circostanti per precipitazioni di durata 24 ore e il periodo di osservazione 1960-2020.....	71
Figura 6.27: <i>Symbol plot</i> che mostra per ciascuna stazione pluviometrica la deviazione della media campionaria locale rispetto alla media campionaria complessiva (a) e la deviazione standard (b) per le precipitazioni di durata 24 ore e periodo di osservazione 2000-2020.....	73
Figura 6.28: Dipendenza spaziale stimata empiricamente dai dati di precipitazione di durata 24 ore per il periodo di osservazione 2000-2020, tramite coefficiente estremo (a, b) e probabilità di concomitanza estrema (c). I punti grigi sono le stime empiriche, mentre i punti neri sono le stime discretizzate con 500 <i>bins</i> , con ingrandimento sulla versione discretizzata per il coefficiente estremo (b).	73
Figura 6.29: Dipendenza spaziale stimata dai modelli M1 e M2 per le precipitazioni di durata 24 ore per il periodo di osservazione 2000-2020, tramite funzione del coefficiente estremo (a, b), e funzione di probabilità di concomitanza estrema (c). I punti grigi sono le stime empiriche calcolate rispetto alle osservazioni e i punti neri rappresentano la versione discretizzata.	75
Figura 6.30: Valutazione locale dei modelli max-stable M1 e M2 selezionato per i dati di precipitazione di durata 24 ore e per il periodo di osservazione 2000-2020, tramite <i>return level plots</i> (a, b, c, d) per M1 e (a', b', c', d') per M2 per 4 stazioni pluviometriche.....	75
Figura 6.31: <i>25-year return levels</i> (a) ed errori standard (b) per l'intero territorio del Piemonte e le aree circostanti per precipitazioni di durata 24 ore e il periodo di osservazione 2000-2020 predetti dal modello M1.	75
Figura 6.32: <i>25-year return levels</i> (a) ed errori standard (b) per l'intero territorio del Piemonte e le aree circostanti per precipitazioni di durata 24 ore e il periodo di osservazione 2000-2020 predetti dal modello M2.	76
Figura A.1: Istogramma del valore di altitudine per le stazioni pluviometriche con registrazioni delle precipitazioni di durata 1 ora nel periodo di osservazione 1960-2020.....	79
Figura A.2: Istogramma del valore di altitudine per le stazioni pluviometriche con registrazioni delle precipitazioni di durata 1 ora nel periodo di osservazione 2000-2020.....	79
Figura A.3: Istogramma del valore di altitudine per le stazioni pluviometriche con registrazioni delle precipitazioni di durata 24 ore nel periodo di osservazione 1960-2020.....	79
Figura A.4: Istogramma del valore di altitudine per le stazioni pluviometriche con registrazioni delle precipitazioni di durata 24 ore nel periodo di osservazione 2000-2020.....	79
Figura A.5: Valutazione della dipendenza spaziale stimata dai modelli max-stable per le precipitazioni di durata 1 ora per il periodo di osservazione 2000-2020, tramite funzione del coefficiente estremo per il modello di Brown-Resnick (a) e per tutti i modelli (b), e funzione di probabilità di concomitanza estrema per ogni modello (c).....	81
Figura A.6: Valutazione della dipendenza spaziale stimata dai modelli max-stable per le precipitazioni di durata 24 ore per il periodo di osservazione 2000-2020, tramite funzione del coefficiente estremo (a) e funzione di probabilità di concomitanza estrema per ogni modello (b).....	83

Lista delle tabelle

Tabella 5.1: Statistiche descrittive (minimo, primo quartile, mediana, media, terzo quartile, massimo e deviazione standard) delle informazioni relative alle stazioni pluviometriche.	44
Tabella 5.2: Statistiche descrittive (minimo, primo quartile, mediana, media, terzo quartile, massimo e deviazione standard) e numeri di valori mancanti dei valori di precipitazioni massime annuali di varie durate.....	47
Tabella 6.1: <i>Trend surfaces</i> per i parametri della distribuzione GEV.....	50
Tabella 6.2: <i>Trend surfaces</i> per il parametro <i>scale</i> della distribuzione GEV.....	50
Tabella 6.3: Valori delle stime delle funzioni del coefficiente estremo e di probabilità di concomitanza estrema in alcune distanze per i modelli max-stable nel caso di precipitazioni di durata 1 ora e l'intero periodo di osservazione (1928-2020).....	53
Tabella 6.4: Statistiche descrittive (minimo, primo quartile, mediana, media, terzo quartile, massimo e deviazione standard) per le precipitazioni di durata 1 ora e periodo di osservazione 1960-2020....	56
Tabella 6.5: Valori delle stime delle funzioni del coefficiente estremo e di probabilità di concomitanza estrema in alcune distanze per il modello Extremal-t selezionato nel caso di precipitazioni di durata 1 ora e il periodo di osservazione 1960-2020.....	58
Tabella 6.6: Statistiche descrittive (minimo, primo quartile, mediana, media, terzo quartile, massimo e deviazione standard) per le precipitazioni di durata 1 ora e periodo di osservazione 2000-2020....	60
Tabella 6.7: Valori delle stime delle funzioni del coefficiente estremo e di probabilità di concomitanza estrema in alcune distanze per i modelli max-stable nel caso di precipitazioni di durata 24 ore e l'intero periodo di osservazione (1928-2020).	67
Tabella 6.8: Statistiche descrittive (minimo, primo quartile, mediana, media, terzo quartile, massimo e deviazione standard) per le precipitazioni di durata 24 ore e periodo di osservazione 1960-2020..	69
Tabella 6.9: Statistiche descrittive (minimo, primo quartile, mediana, media, terzo quartile, massimo e deviazione standard) per le precipitazioni di durata 24 ore e periodo di osservazione 2000-2020..	73
Tabella A.1: Valori di TIC per i modelli GEV spaziali.	80
Tabella A.2: Valori di TIC per i modelli max-stable (prima metodologia).....	80
Tabella A.3: Valori delle stime delle funzioni del coefficiente estremo e di probabilità di concomitanza estrema in alcune distanze per i modelli max-stable nel caso di precipitazioni di durata 1 ora e il periodo di osservazione 2000-2020.....	82
Tabella A.4: Valori delle stime delle funzioni del coefficiente estremo e di probabilità di concomitanza estrema in alcune distanze per i modelli max-stable nel caso di precipitazioni di durata 24 ore e il periodo di osservazione 2000-2020.....	83
Tabella A.5: Errori di previsione 10-fold cross-validation per le precipitazioni di durata 1 ora e periodi di osservazione 1960-2020 e 2000-2020.....	84
Tabella A.6: Valori delle stime delle funzioni del coefficiente estremo e di probabilità di concomitanza estrema in alcune distanze per il modello Brown-Resnick selezionato nel caso di precipitazioni di durata 1 ora e il periodo di osservazione 2000-2020.....	84
Tabella A.7: Errori di previsione 10-fold cross-validation per le precipitazioni di durata 24 ore e periodi di osservazione 1960-2020 e 2000-2020.....	85
Tabella A.8: Valori delle stime delle funzioni del coefficiente estremo e di probabilità di concomitanza estrema in alcune distanze per il modello Extremal-t selezionato nel caso di precipitazioni di durata 24 ore e il periodo di osservazione 2000-2020.....	85

Capitolo 1

Introduzione

Nel campo della statistica e dell'analisi dei dati, è essenziale raccogliere e analizzare dati reali al fine di studiare un determinato fenomeno. Generalmente, si è interessati a studiare la distribuzione di probabilità da cui si suppone provengano i dati o la sua funzione inversa, ovvero la funzione quantile, e il valore atteso della distribuzione come misura della media dei dati ([1], Capitolo 1, p. 1). Questo approccio mira a interpretare il comportamento generale del fenomeno che tali dati descrivono per prendere decisioni basate su di essi. In alcuni contesti, invece, si può essere interessati a calcolare la probabilità di osservare quantità più estreme di quelle registrate, ad esempio per la progettazione di infrastrutture. Tuttavia, la teoria classica delle probabilità non fornisce strumenti adeguati a prevedere tali eventi, poiché l'attenzione non è specificatamente rivolta sulla coda della distribuzione, ovvero la parte della distribuzione dei dati che si estende verso i valori più alti o più bassi.

In una collezione di dati reali, i valori estremi rappresentano quei dati che si discostano significativamente dalla maggioranza dei dati raccolti, ovvero costituiscono la coda della distribuzione. Tali dati possono avere un impatto sostanziale su un fenomeno e possono fornire importanti informazioni su potenziali rischi, opportunità o anomalie che potrebbero non emergere considerando l'intera collezione di dati. La loro analisi consente di stimare la probabilità di accadimento di eventi estremi futuri al fine di mitigare i rischi associati al loro avvenimento. La teoria alla base della trattazione degli eventi estremi si costituisce di concetti probabilistici e statistici avanzati e metodologie specifiche mirate a identificare, caratterizzare e modellare la distribuzione di tali eventi rari, al fine di predirne la probabilità di avvenimento e la loro intensità.

L'analisi dei valori estremi (*Extreme Value Analysis, EVA*) riveste un'importanza crescente nel contesto attuale per molteplici settori, tra cui scienze ambientali, ingegneria e finanza. Ad esempio, particolare interesse viene volto verso eventi meteorologici estremi, come raffiche di vento, temperature o precipitazioni elevate e inondazioni. Tali fenomeni si manifestano su aree geografiche estese e influenzano molti aspetti come agricoltura, tempo di vita di alcuni materiali e funzionamento di alcuni macchinari ([4], Capitolo 1, p. 7), causando danni in termini di vite umane, perdite finanziarie, perdita di infrastrutture necessarie alla società e molto altro. Risulta necessario, dunque, determinare la rarità, l'intensità e la dipendenza spaziale di questi eventi estremi per poter mettere in atto misure di adattamento, mitigazione e allertamento per limitare i danni dovuti a eventi climatici eccezionali futuri e per una migliore gestione delle risorse in risposta a eventi estremi di rilievo. L'analisi dei valori estremi permette infatti di stimare la velocità del vento, il valore di temperatura o di precipitazione e in generale la variabile di interesse sia nelle località osservate sia in nuove località, nei prossimi T anni avendo a disposizione i dati degli eventi estremi per un periodo di tempo molto più limitato di N anni, con N in genere molto minore di T .

Queste informazioni possono essere sfruttate anche in altri settori. Nell'ambito dell'ingegneria strutturale e idraulica, ad esempio, la progettazione di edifici ed infrastrutture, come ponti e dighe, avviene tenendo conto della resistenza a eventi estremi, come raffiche di vento, intensità dei terremoti o alluvioni. Allo stesso modo, le compagnie assicurative devono valutare la probabilità di avvenimento e i rischi associati a eventi catastrofici, come terremoti, uragani e alluvioni, al fine di stabilire le tariffe assicurative ed evitare delle perdite in denaro significative.

Nell'ambito della finanza, invece, vengono analizzati i valori estremi delle serie temporali riguardanti i prezzi speculativi giornalieri o settimanali di attività quali azioni, valute estere o materie prime al fine di valutare il rischio di perdite elevate dovute all'alta variabilità dei prezzi delle attività finanziarie in questione ([5], Capitolo 1, p. 5). Nello specifico, vengono studiati i rendimenti giornalieri o settimanali per analizzare le distribuzioni di perdite/guadagni, tenendo conto di alcuni parametri che misurano il rischio.

Altre applicazioni riguardano la previsione del traffico nelle telecomunicazioni, l'elaborazione dei dati biomedici, scienza alimentare, studi sull'inquinamento e delle concentrazioni di determinate sostanze che superano la soglia critica prevista dalla legge, resistenza alla corrosione di strutture metalliche, resistenza alla rottura di componenti e modellazione dell'affidabilità.

Lo scopo principale di questo lavoro è quello di analizzare un dataset reale di precipitazioni massime annuali registrate nella regione Piemonte attraverso i modelli sviluppati dalla teoria dei valori estremi in un contesto spaziale, con l'obiettivo di modellare i pattern spaziali di questi eventi estremi. A tal fine, la prima parte di questo lavoro è dedicata alla presentazione della teoria classica degli eventi estremi. Nello specifico, nel Capitolo 2, vengono presentati i risultati univariati e multivariati, con riferimento principale a [2]. Nel Capitolo 3, segue una presentazione della teoria degli estremi spaziali, mentre le tecniche di inferenza, selezione e valutazione dei modelli spaziali vengono esposti nel Capitolo 4. L'ultima parte di questo lavoro è dedicata all'applicazione dei modelli sul dataset di precipitazioni. Il Capitolo 5 contiene un'analisi esplorativa del dataset utilizzato, mentre il Capitolo 6 presenta e confronta i risultati ottenuti.

Capitolo 2

Teoria dei valori estremi

Nella prima parte di questo capitolo, viene presentata la teoria classica dei valori estremi nel caso univariato. Essa viene successivamente generalizzata al caso multivariato nella seconda parte. I risultati descritti provengono principalmente da [2].

2.1 Estremi univariati

La teoria dei valori estremi studia il comportamento stocastico dei valori estremi estratti da una collezione di n variabili aleatorie, che si assume siano indipendenti e provenienti dalla stessa distribuzione di probabilità (i.i.d.). I modelli sviluppati applicano un approccio asintotico, valido per $n \rightarrow \infty$, per trovare la distribuzione limite di probabilità dei valori estremi.

Un modo per estrarre i valori estremi è quello di calcolare il massimo o il minimo su ogni blocco di dati (*block maxima/minima*). Ad esempio, è possibile selezionare un evento per ciascuno anno (massimo o minimo annuale). Una generalizzazione consiste nel considerare un numero limitato r di estremi massimi o minimi per ogni blocco (*the r largest/smallest order statistics*). In alternativa, per alcune applicazioni è utile selezionare come estremi i valori che superano una soglia prefissata (*exceedance over/below threshold models*). Questo lavoro si focalizza solo sulla prima caratterizzazione.

2.1.1 Formulazione del modello asintotico per i massimi

Siano X_1, \dots, X_n una sequenza di variabili aleatorie indipendenti provenienti dalla stessa distribuzione di probabilità, con funzione di ripartizione F . Ciascun X_i rappresenta il valore di un processo in esame che viene misurato in un intervallo di tempo regolare, ad esempio le precipitazioni giornaliere. Il valore estremo estratto dal campione di n osservazioni è il valore massimo assunto su tutto il periodo:

$$M_n = \max\{X_1, \dots, X_n\}. \quad (2.1)$$

Ad esempio, se n corrisponde al numero di osservazioni in un anno, ovvero a 365 giorni se il processo viene misurato giornalmente, allora M_n rappresenta il massimo annuale. Se si considerano K anni e per ciascuno anno si hanno n osservazioni, da cui si estrae per ogni anno il valore massimo $M_{n,k}$, con $k = 1, \dots, K$, tale metodo genera dei massimi del blocco (*block maxima*) $M_{n,k}$ per blocchi di dati della stessa lunghezza. Talvolta, invece, le X_i non sono osservabili e i dati a disposizione rappresentano già gli eventi estremi.

La distribuzione teorica di (2.1) è ricavabile in modo esatto dalla teoria, se la distribuzione F è già nota:

$$\mathbb{P}(M_n \leq z) = \mathbb{P}(X_1 \leq z, \dots, X_n \leq z) = \mathbb{P}(X_1 \leq z) \cdot \dots \cdot \mathbb{P}(X_n \leq z) = F^n(z). \quad (2.2)$$

Dal momento che la distribuzione F è ignota, anche la distribuzione (2.2) lo è. Trovare una stima di F per ottenere una stima di F^n può risultare problematico in quanto un piccolo errore nell'approssimazione di F può amplificarsi molto nell'approssimazione di F^n , ottenendo una stima poco accurata della distribuzione di (2.1). Di conseguenza, il problema che si pone è quello di cercare direttamente una stima di (2.2) utilizzando la conoscenza dei valori estremi solamente. Questo

corrisponde a trovare qualche metodo da utilizzare per ottenere un'approssimazione affidabile della distribuzione di (2.1). Una soluzione al problema risulta essere quella di usare un approccio asintotico, considerando la dimensione del campione infinita, e cercare una distribuzione di probabilità G limite per (2.2). In tale situazione, possono sorgere diversi interrogativi sull'esistenza di una o più distribuzioni limite e sulla possibilità che esse dipendano in qualche modo dalla distribuzione di partenza F o se distribuzioni di partenza diverse conducano alla stessa distribuzione limite.

In modo analogo all'approccio utilizzato dal Teorema del Limite Centrale per trovare la distribuzione della media campionaria di una collezione di variabili aleatorie i.i.d., si cerca la distribuzione di probabilità limite per $n \rightarrow \infty$ di

$$M_n^* = \frac{M_n - b_n}{a_n}, \quad (2.3)$$

con $\{a_n > 0\}$ e $\{b_n\}$ sequenze di numeri scelte in modo tale per cui *location* e *scale* di M_n^* siano stabili quando aumenta n . La normalizzazione (2.3) di M_n è necessaria per evitare che la distribuzione limite degeneri in un punto. Infatti, sia z_+ l'estremo superiore della distribuzione, ovvero il più piccolo valore di z tale per cui $F(z) = 1$, allora:

$$\lim_{n \rightarrow \infty} F^n(z) = \begin{cases} 0 & \forall z < z_+ \\ 1 & \text{per } z = z_+ \end{cases}, \quad (2.4)$$

ovvero la distribuzione di M_n degenera nel punto z_+ .

Il problema del limite (estremo), pertanto, è quello di trovare tutte le possibili distribuzioni non degeneri G di (2.3), ed è legato al problema del dominio di attrazione, ovvero caratterizzare la distribuzione F affinché esistano sequenze $\{a_n > 0\}$ e $\{b_n\}$ tali per cui la distribuzione limite G esista ([1], Capitolo 2, p. 46).

Un importante risultato, presentato da Fisher e Tippett nel 1928 e completato e formalizzato da Gnedenko nel 1943, che considera tutte le distribuzioni limite non degeneri è il seguente teorema ([2], Capitolo 3, p. 46).

Teorema 2.1 (Teorema dei Tipi Estremi) Se esistono sequenze di costanti $\{a_n > 0\}$ e $\{b_n\}$ tali che

$$\mathbb{P}\left(\frac{M_n - b_n}{a_n} \leq z\right) \rightarrow G(z) \quad \text{per } n \rightarrow \infty, \quad (2.5)$$

dove G è una funzione di distribuzione non degenera, allora G appartiene a una delle seguenti famiglie:

- I. $G(z) = \exp\left\{-\exp\left[-\left(\frac{z-b}{a}\right)\right]\right\}, \quad -\infty < z < \infty;$
- II. $G(z) = \begin{cases} 0, & z \leq b, \\ \exp\left\{-\left(\frac{z-b}{a}\right)^{-\alpha}\right\}, & z > b; \end{cases}$
- III. $G(z) = \begin{cases} \exp\left\{-\left[-\left(\frac{z-b}{a}\right)^\alpha\right]\right\}, & z < b, \\ 1, & z \geq b, \end{cases}$

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

Un'importante implicazione del precedente teorema è dunque che solo tre famiglie di distribuzioni rappresentano la distribuzione limite degli estremi massimi. Le tre famiglie di distribuzione, note come distribuzioni dei valori estremi, sono rispettivamente:

- I. Gumbel, illimitata sia inferiormente sia superiormente;
- II. Fréchet, limitata inferiormente e illimitata superiormente;
- III. Weibull, illimitata inferiormente e limitata superiormente.

I parametri b , a e α sono rispettivamente *location*, *scale* e *shape* della distribuzione. In particolare, il parametro di *location* indica dove si concentrano i dati e corrisponde all'estremo sinistro per la famiglia Fréchet e all'estremo destro per la famiglia Weibull. Il parametro di *scale*, invece, rappresenta la dispersione della distribuzione e determina la variabilità degli eventi estremi, mentre il parametro di *shape* è relativo alla forma della coda della distribuzione. Infatti, le tre famiglie derivano da un comportamento specifico della coda della distribuzione F delle X_i e, pertanto, rappresentano in modo differente il comportamento dei valori estremi. La funzione di densità della distribuzione Gumbel decade esponenzialmente, mentre quella della distribuzione Fréchet decade polinomialmente. Entrambe le distribuzioni derivano da una distribuzione F con punto estremo superiore $z_+ = \infty$. Al contrario, la distribuzione Weibull deriva da una distribuzione F in cui z_+ assume un valore finito.

La scelta di quale famiglia sia più appropriata nelle applicazioni dipende dai dati che si hanno a disposizione. Tuttavia, esistono molti esempi nelle applicazioni dove una famiglia è stata preferita rispetto a un'altra. In particolare, la distribuzione Gumbel è stata impiegata per descrivere una vasta gamma di fenomeni grazie alla sua semplicità dovuta alla presenza di due parametri anziché tre e al suo dominio d'attrazione per la distribuzione di partenza F che include quelle più tipiche, come esponenziale, gamma, Weibull, normale e lognormale [8]. Le applicazioni più comunemente descritte dalla famiglia Gumbel includono gli estremi idrologici (precipitazioni, inondazioni), velocità estreme del vento, altezze delle onde marine, resistenza elettrica dei materiali, inquinamento atmosferico e problemi geologici ([4], Capitolo 5, p.184). Allo stesso modo, la famiglia Fréchet è stata utilizzata, ad esempio, per modellare le onde del mare e la velocità del vento ([4], Capitolo 5, p. 208), mentre la famiglia Weibull è stata impiegata per analizzare la resistenza alla fatica dei materiali, resistenza alla corrosione, il tempo di vita dei tubi a vuoto e molti altri fenomeni ([4], Capitolo 5, pp. 198-199). Tuttavia, alcuni studi hanno mostrato come la distribuzione Gumbel non sia sempre adatta a descrivere gli estremi idrologici, come alluvioni e precipitazioni estreme, in quanto sottostima le quantità più estreme nel caso di periodi di tempo elevati, comportando un rischio maggiore per le strutture ingegneristiche. Inoltre, è stato studiato come serie storiche di precipitazioni massime annuali estese su 20-50 anni sembrano seguire il comportamento della distribuzione Gumbel, mentre serie estese su molti più anni siano meglio descritte dalla distribuzione Fréchet [8]. Ne consegue che la scelta della famiglia nelle applicazioni sia spesso non banale.

Le difficoltà precedentemente descritte possono essere superate, in quanto è possibile riformulare le tre famiglie di distribuzioni combinandole in un'unica famiglia di distribuzioni, conosciuta con il nome *generalized extreme value (GEV) distribution*, con funzione di ripartizione:

$$G(z) = \exp \left\{ - \left[1 + \xi \left(\frac{z - \mu}{\sigma} \right) \right]^{-1/\xi} \right\}, \quad (2.6)$$

definita sull'insieme $\{z: 1 + \xi(z - \mu)/\sigma > 0\}$, con parametro *location* tale per cui $-\infty < \mu < \infty$, parametro *scale* non negativo $\sigma > 0$ e parametro *shape* tale per cui $-\infty < \xi < \infty$. A partire dalla formulazione (2.6), si ottengono le tre famiglie nel **Teorema 2.1** in base al valore del parametro *shape*: si ottiene la distribuzione Gumbel, con funzione di ripartizione data da I. nel **Teorema 2.1** e parametri b ed a rispettivamente μ e σ , per $\xi = 0$ interpretato come il limite di (2.6) per $\xi \rightarrow 0$, distribuzione Fréchet per $\xi > 0$ e distribuzione Weibull per $\xi < 0$. L'unificazione delle tre famiglie nella distribuzione GEV semplifica l'approccio da utilizzare nelle applicazioni, in quanto evita di dover

scegliere quale delle tre classi è più appropriata per la collezione dei valori estremi in esame o ripetere per tutte e tre la stima dei parametri. Il comportamento della coda della distribuzione deriva direttamente dall'inferenza sul parametro *shape*. È possibile riformulare il **Teorema 2.1** considerando direttamente la distribuzione GEV, come è stato proposto da von Mises nel 1954 e da Jenkinson l'anno successivo.

Teorema 2.2 ([2], Capitolo 3, p. 47) Se esistono sequenze di costanti $\{a_n > 0\}$ e $\{b_n\}$ tali che

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

dove G è una funzione di distribuzione non degenere, allora G appartiene alla famiglia GEV

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

definita sull'insieme $\{z: 1 + \xi(z - \mu)/\sigma > 0\}$, dove $-\infty < \mu < \infty$, $\sigma > 0$ e $-\infty < \xi < \infty$.

Le sequenze di costanti $\{a_n > 0\}$ e $\{b_n\}$ non dipendono dalla distribuzione F delle X_i e sono ignote nella pratica. Questo problema può essere risolto in quanto la normalizzazione (2.3) non modifica la distribuzione limite di M_n . Infatti, assumendo (2.5), si ha, per n sufficientemente grande:

$$\mathbb{P}\left(\frac{M_n - b_n}{a_n} \leq z\right) \approx G(z). \quad (2.7)$$

Allo stesso modo,

$$\mathbb{P}(M_n \leq z) \approx G\left(\frac{z - b_n}{a_n}\right) = G^*(z), \quad (2.8)$$

dove G^* è un altro membro della famiglia GEV. Ne consegue che il **Teorema 2.2**, combinato alle precedenti osservazioni, implica che un'approssimazione della distribuzione limite dei valori massimi estratti da un campione sufficientemente grande di variabili aleatorie i.i.d., come definito in (2.1), appartiene alla famiglia GEV. Nell'implementazione è irrilevante che i parametri di G^* siano diversi da quelli di G , in quanto essi vanno stimati in ogni caso.

2.1.1.1 Proprietà della distribuzione GEV

La funzione di ripartizione della distribuzione GEV è data dall'equazione (2.6). La funzione di densità corrispondente è:

$$g(z) = G'(z) = \exp\left\{-\left[1 + \xi\left(\frac{z - \mu}{\sigma}\right)\right]^{-1/\xi}\right\} \frac{1}{\sigma} \left[1 + \xi\left(\frac{z - \mu}{\sigma}\right)\right]^{-\frac{1}{\xi}-1}, \quad (2.9)$$

definita sull'insieme $\{z: 1 + \xi(z - \mu)/\sigma > 0\}$.

Invertendo (2.6), ovvero ricavando z_p tale che $G(z_p) = \mathbb{P}(M_n \leq z_p) = 1 - p$, invece, si ottengono i quantili:

$$z_p = \begin{cases} \mu - \frac{\sigma}{\xi} \{1 - [-\log(1 - p)]^{-\xi}\} & \text{per } \xi \neq 0, \\ \mu - \sigma \log[-\log(1 - p)] & \text{per } \xi = 0. \end{cases} \quad (2.10)$$

Nelle applicazioni dell'analisi dei valori estremi, essi sono di notevole importanza e sono noti con il nome di *return levels* associati al *return period* $T = 1/p$. Infatti, essi rappresentano il livello che ci si

aspetta venga superato in media ogni T anni. In altre parole, il massimo annuale supera z_p in un particolare anno con probabilità p e questo accade nuovamente dopo $T - 1$ “fallimenti”: $p(1 - p)^{T-1}$, ovvero in media dopo $T = 1/p$ anni.

La distribuzione GEV gode della proprietà di *max-stability*, definita nel seguente modo.

Definizione 2.1 Una distribuzione G è detta **max-stable** se, per ogni $n \geq 2$, esistono costanti $\alpha_n > 0$ e β_n tali che

$$G^n(\alpha_n z + \beta_n) = G(z). \quad (2.11)$$

Ciò equivale a dire che, in seguito a una riscalatura e una traslazione appropriate, l'operazione di prendere il massimo di un campione di variabili aleatorie i.i.d. mantiene la stessa distribuzione.

Si può dimostrare che la condizione di *max-stability* vale solo per la distribuzione GEV, come enunciato dal seguente teorema ([2], Capitolo 3, p. 50).

Teorema 2.3 Una distribuzione è *max-stable* se e solo se è una distribuzione GEV.

Il teorema precedente permette di dimostrare il **Teorema 2.3**. L'idea è quella di considerare il massimo di una collezione di massimi di variabili aleatorie i.i.d. Dal momento che vale la proprietà di *max-stability* e vale il **Teorema 2.3**, la distribuzione limite può essere solo la distribuzione GEV.

2.1.2 Modello asintotico per minimi

È possibile derivare i risultati asintotici ottenuti per il valore massimo di un campione di n variabili aleatorie i.i.d X_1, \dots, X_n anche per il valore minimo:

$$\tilde{M}_n = \min\{X_1, \dots, X_n\}, \quad (2.12)$$

potendo sfruttare la relazione

$$\min_{i=1, \dots, n} \{X_i\} = - \max_{i=1, \dots, n} \{-X_i\}. \quad (2.13)$$

Denotando con $Y_i = -X_i$, la relazione (2.13) corrisponde a dire che piccoli valori di X_i corrispondono a valori grandi di Y_i cambiati di segno. In tal caso, le X_i rappresentano, ad esempio, il tempo di vita di un singolo componente. Di conseguenza, si studia il tempo di vita minimo tra le n componenti di un sistema, in quanto esso corrisponde al tempo di vita dell'intero sistema.

Assumendo che tutte le X_i abbiano la stessa funzione di sopravvivenza $\bar{F} = 1 - F$, la distribuzione di (2.12) è:

$$\begin{aligned} \mathbb{P}(\tilde{M}_n \leq z) &= 1 - \mathbb{P}(\tilde{M}_n > z) = 1 - \mathbb{P}(X_1 > z, \dots, X_n > z) = \\ &= 1 - \mathbb{P}(X_1 > z) \cdot \dots \cdot \mathbb{P}(X_n > z) = 1 - \bar{F}^n(z) = \\ &= 1 - (1 - F(z))^n. \end{aligned} \quad (2.14)$$

Analogamente al risultato ottenuto con i massimi in (2.4), in tal caso si ha:

$$\lim_{n \rightarrow \infty} 1 - (1 - F(z))^n = \begin{cases} 0 & \text{se } F(z) = 0 \\ 1 & \text{se } F(z) \leq 1 \end{cases} \quad (2.15)$$

ottenendo una distribuzione limite degenere.

Dal momento che la funzione F , o equivalentemente \bar{F} , non è nota, trovare una stima per la distribuzione di (2.12) utilizzando la relazione (2.14) non è facile. Tuttavia, grazie alla relazione (2.13) è possibile scrivere:

$$\begin{aligned}\mathbb{P}(\tilde{M}_n \leq z) &= \mathbb{P}(-M_n \leq z) = \mathbb{P}(M_n \geq z) = 1 - \mathbb{P}(M_n \leq z) \approx \\ &\approx 1 - \exp\left\{-\left[1 + \xi\left(\frac{-z - \mu}{\sigma}\right)\right]^{-1/\xi}\right\} = \\ &= 1 - \exp\left\{-\left[1 - \xi\left(\frac{z - \tilde{\mu}}{\sigma}\right)\right]^{-1/\xi}\right\},\end{aligned}\tag{2.16}$$

definito su $\{z: 1 - \xi(z - \tilde{\mu})/\sigma > 0\}$ e con $\tilde{\mu} = -\mu$. La distribuzione in (2.16) prende il nome di distribuzione GEV per minimi. Il risultato ottenuto viene riassunto nel seguente teorema ([2], Capitolo 3, p. 53).

Teorema 2.2.4 *Se esistono sequenze di costanti $\{a_n > 0\}$ e $\{b_n\}$ tali che*

$$\mathbb{P}\left(\frac{\tilde{M}_n - b_n}{a_n} \leq z\right) \rightarrow \tilde{G}(z) \quad \text{per } n \rightarrow \infty,\tag{2.17}$$

per una funzione di distribuzione non degenera \tilde{G} , allora \tilde{G} appartiene alla famiglia GEV per minimi

$$\tilde{G}(z) = 1 - \exp\left\{-\left[1 - \xi\left(\frac{z - \tilde{\mu}}{\sigma}\right)\right]^{-1/\xi}\right\},\tag{2.18}$$

definita sull'insieme $\{z: 1 - \xi(z - \tilde{\mu})/\sigma > 0\}$, dove $-\infty < \tilde{\mu} < \infty$, $\sigma > 0$ e $-\infty < \xi < \infty$.

Nelle applicazioni, è possibile stimare direttamente i parametri della distribuzione GEV per minimi $(\tilde{\mu}, \sigma, \xi)$ a partire dalle realizzazioni z_1, \dots, z_m . In alternativa, è possibile considerare i dati $-z_1, \dots, -z_m$ e stimare i parametri della distribuzione GEV per massimi, che corrispondono agli stessi della distribuzione GEV per minimi, ma considerando $\tilde{\mu} = -\mu$.

2.1.2.1 Min-stability

La distribuzione limite per i minimi gode della proprietà di *min-stability*, definita nel seguente modo ([5], Capitolo 1, p.23).

Definizione 2.2 *Una distribuzione G è min-stable se, per ogni $n \geq 2$, esistono costanti $\alpha_n > 0$ e β_n tali che*

$$1 - (1 - G(\alpha_n z + \beta_n))^n = G(z).\tag{2.19}$$

La proprietà di *min-stability* si può esprimere anche in termini della funzione di sopravvivenza:

$$\bar{G}^n(\alpha_n z + \beta_n) = \bar{G}(z).\tag{2.20}$$

2.1.3 Generalizzazione al caso di variabili aleatorie dipendenti

I risultati e i modelli descritti precedentemente sono validi sotto l'ipotesi che i dati siano indipendenti. Tuttavia, in molte situazioni pratiche questa assunzione non è soddisfatta. La generalizzazione più semplice prevede l'ipotesi che i dati siano stazionari.

Definizione 2.3 *La serie X_1, X_2, \dots è detta stazionaria in senso forse se per ogni m naturale e ogni m -upla (i_1, \dots, i_m) e per ogni h si ha la seguente uguaglianza della funzione di distribuzione congiunta:*

$$\mathbb{P}(X_{i_1} = x_1, \dots, X_{i_m} = x_m) = \mathbb{P}(X_{i_1+h} = x_1, \dots, X_{i_m+h} = x_m). \quad (2.21)$$

Questa assunzione è più realistica dell'indipendenza per molti processi fisici e include la possibilità che le variabili siano mutualmente dipendenti. Tuttavia, esistono molte forme diverse di dipendenza nelle serie stazionarie ed è impossibile sviluppare un'unica caratterizzazione. Un vincolo che può essere imposto è di considerare gli eventi estremi corrispondenti a punti sufficientemente distanti tra loro (nel tempo e/o nello spazio) come se fossero variabili indipendenti. Ad esempio, se si considera una serie sulle precipitazioni, si può supporre che un evento estremo oggi possa influenzare i valori di precipitazione nei giorni successivi o nelle aree adiacenti, ma non gli eventi meteorologici estremi di un periodo maggiore, come qualche mese, o di aree distanti. Sotto questa ipotesi, viene eliminata la dipendenza degli eventi estremi a lungo raggio e si concentra l'attenzione sull'effetto della dipendenza a breve raggio. La distribuzione limite dei massimi di blocco appartiene alla distribuzione GEV, allo stesso modo del caso di variabili aleatorie indipendenti, ma i valori dei parametri della serie stazionaria sono differenti da quelli della serie di variabili indipendenti. Dal momento che i parametri vengono stimati sulla base dei dati, in questo caso la dipendenza può essere ignorata. Tuttavia, l'accuratezza dell'approssimazione della distribuzione GEV ai *block maxima* diminuisce se i livelli di dipendenza nella serie aumentano, dal momento che il numero di campioni non è infinito.

Quando l'ipotesi di stazionarietà non risulta essere valida, le caratteristiche dei processi cambiano sistematicamente nel tempo e/o nello spazio, come nel caso di effetti stagionali, trend a lungo termine o variazioni spaziali. In questo caso, la non-stazionarietà viene incorporata all'interno dei parametri della distribuzione GEV. Si può considerare, per ognuno dei tre parametri, un effetto lineare, quadratico, che cambia in base al periodo o un termine armonico per descrivere una stagionalità, garantendo la positività del parametro di *scale* σ mediante l'utilizzo della funzione esponenziale. Inoltre, è possibile considerare anche la dipendenza rispetto ad altre variabili. Il modello generale viene descritto da:

$$\theta(t) = h(\mathbf{X}^T \boldsymbol{\beta}), \quad (2.22)$$

dove θ denota il parametro μ , σ o ξ , h è una funzione, $\boldsymbol{\beta}$ è un vettore di parametri da stimare e X contiene le covariate. Nel caso di trend lineari h è la funzione identità, mentre in caso di stagionalità si considera la funzione indicatrice per distinguere i periodi.

2.1.4 Caratterizzazione come processi puntuali

Il comportamento dei valori estremi in un processo può essere descritto in vari modi. Una caratterizzazione, proposta da Pickands nel 1971 e sviluppata da Leadbetter, Lindgren e Rootzén nel 1983, utile per trovare sia la frequenza che l'intensità degli eventi estremi è quella dei processi puntuali. In tal modo, risulta più agevole calcolare il tempo di attesa previsto per l'avvenimento di un evento, essendo a conoscenza di uno di essi, oppure calcolare la probabilità di un certo numero di eventi in un periodo specificato. Un esempio noto dei processi puntuali è il processo di Poisson omogeneo e la sua generalizzazione non-omogenea nel caso in cui gli eventi avvengano casualmente nel tempo e/ nello spazio. Nel caso di eventi estremi descritti da (2.1), vale il seguente risultato ([2], Capitolo 7, pp. 129-130).

Teorema 2.5 Sia X_1, \dots, X_n una serie di variabili aleatorie indipendenti e identicamente distribuite per cui esistono sequenze di costanti $\{a_n > 0\}$ e $\{b_n\}$ tali che

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

dove

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

e siano z_- e z_+ gli estremi inferiore e superiore di G rispettivamente. Allora, la sequenza di processi puntuali

$$N_n = \left\{ \left(\frac{i}{n+1}, \frac{X_i - b_n}{a_n} \right) : i = 1, \dots, n \right\} \quad (2.23)$$

converge nelle regioni della forma $(0,1) \times [u, \infty)$, per ogni $u > z_-$, a un processo di Poisson, con misura di intensità su $A = [v_1, v_2] \times [z, z_+)$ data da

$$\Lambda(A) = (v_2 - v_1) \left[1 + \xi \left(\frac{z - \mu}{\sigma} \right) \right]^{-1/\xi}. \quad (2.24)$$

Secondo il precedente teorema, viene definito un processo puntuale N_n tale per cui, sotto le ipotesi del **Teorema 2.2**, esso converge in distribuzione a un processo di Poisson non-omogeneo. La prima ordinata del processo puntuale (2.23) rappresenta l'asse temporale e/o spaziale che viene mappato nell'intervallo $(0,1)$, mentre la seconda ordinata rappresenta le variabili aleatorie normalizzate in modo tale da stabilizzare il comportamento degli estremi nel limite per $n \rightarrow \infty$. Una traccia della dimostrazione è la seguente.

DIM. Consideriamo una regione della forma $A = (0,1) \times [u, \infty)$ per qualche valore grande di u . Ognuno degli n punti di N_n ha una probabilità p di cadere nella regione A , dove

$$p = \mathbb{P} \left(\frac{X_i - b_n}{a_n} > u \right) \approx \frac{1}{n} \left[1 + \xi \left(\frac{u - \mu}{\sigma} \right) \right]^{-1/\xi}.$$

Dal momento che le variabili X_i sono mutuamente indipendenti, $N_n(A)$ segue una distribuzione binomiale di parametri n e p :

$$N_n(A) \sim \text{Bin}(n, p).$$

La convergenza di $N_n(A)$, per $n \rightarrow \infty$, a un processo di Poisson segue dalla classica teoria delle probabilità e dei processi stocastici. La misura di intensità si ricava da:

$$\Lambda(A) = \int_0^1 \lambda(s) ds = \int_0^1 np ds = \int_0^1 n \frac{1}{n} \left[1 + \xi \left(\frac{u - \mu}{\sigma} \right) \right]^{-1/\xi} ds = \left[1 + \xi \left(\frac{u - \mu}{\sigma} \right) \right]^{-1/\xi}.$$

Dal momento che il processo è omogeneo rispetto al tempo/spazio, segue che, per ogni regione della forma $A = [v_1, v_2] \times [u, \infty)$, con $[v_1, v_2] \subset [0,1]$, la distribuzione limite di $N_n(A)$ è una $Poiss(\Lambda(A))$, dove

$$\Lambda(A) = (v_2 - v_1) \left[1 + \xi \left(\frac{z - \mu}{\sigma} \right) \right]^{-1/\xi}. \quad \square$$

Il dominio del processo di Poisson non è legato agli estremi della distribuzione degli X_i , bensì a quelli della distribuzione limite G . In particolare, la normalizzazione applicata ai punti in N_n tende a concentrare i valori non estremi vicino al termine inferiore dell'intervallo $(0,1) \times (z_-, z_+)$. Di conseguenza, la convergenza nel **Teorema 2.5** è valida solo nelle regioni limitate da z_- . Sulla base dei risultati ottenuti, dunque, è possibile considerare il processo di Poisson come un'approssimazione ragionevole per il comportamento dei campioni nel caso in cui questi ultimi siano finiti ma numerosi (n sufficientemente grande). Tuttavia, poiché la convergenza è ristretta agli insiemi che escludono

l'estremità inferiore, si ottiene una buona approssimazione solo in insiemi della forma $(0,1) \times (u, \infty)$, dove u è sufficientemente grande. Ne consegue che il **Teorema 2.5** fornisce una rappresentazione solo per il comportamento dei valori estremi degli X_i , e nelle applicazioni è necessario determinare il valore di u abbastanza estremo da rendere il limite un'approssimazione ragionevole.

Una proprietà fondamentale dei processi di Poisson riguarda l'indipendenza dei punti. Dunque, l'occorrenza di un punto in un determinato luogo $x \in A$ non influenza in alcun modo, né incoraggiando né impedendo, l'avvenimento di un altro punto nelle vicinanze di x o in qualunque altra parte dell'area. Inoltre, la presenza di una misura di intensità non costante permette alcune particolarità, ad esempio consente la possibilità di avere un numero diverso di punti in differenti sotto-regioni dello spazio A .

La caratterizzazione dei valori estremi come processi puntuali si collega al modello dei *block maxima* considerando l'evento estremo

$$\left\{ \frac{M_n - b_n}{a_n} \leq z \right\}$$

equivalente all'evento

$$N_n(A_z) = 0, \quad \text{con } A_z = (0,1) \times (z, \infty).$$

Di conseguenza,

$$\begin{aligned} \mathbb{P}\left(\frac{M_n - b_n}{a_n} \leq z\right) &= \mathbb{P}(N_n(A_z) = 0) \rightarrow \mathbb{P}(N(A_z) = 0) = \exp(-\Lambda(A_z)) = \\ &= \exp\left\{-\left[1 + \xi\left(\frac{z - \mu}{\sigma}\right)\right]^{-1/\xi}\right\}, \end{aligned}$$

perciò la distribuzione limite dei *block maxima* normalizzati è la distribuzione GEV.

Analogamente al modello dei *block maxima*, il problema della mancanza di informazioni riguardanti le sequenze di costanti $\{a_n > 0\}$ e $\{b_n\}$ viene risolto assorbendo i coefficienti all'interno dei parametri di *location* e *scale* del modello. Il **Teorema 2.5** viene dunque riformulato come segue ([2], Capitolo 7, p. 133).

Teorema 2.6 Sia X_1, \dots, X_n una serie di variabili aleatorie indipendenti e identicamente distribuite, e sia

$$N_n = \left\{ \left(\frac{i}{n+1}, X_i \right) : i = 1, \dots, n \right\}. \quad (2.25)$$

Allora, per u sufficientemente elevato, nelle regioni della forma $(0,1) \times [u, \infty)$, N_n è approssimativamente un processo di Poisson, con misura di intensità su $A = [v_1, v_2] \times (z, \infty)$ data da

$$\Lambda(A) = (v_2 - v_1) \left[1 + \xi \left(\frac{z - \mu}{\sigma} \right) \right]^{-1/\xi}.$$

Per convenzione, i limiti dei valori estremi vengono espressi in termini di distribuzioni approssimate di massimi annuali, piuttosto che utilizzare un'altra scala temporale. Ne consegue che, se i dati sono stati osservati per n_y anni, la misura d'intensità del processo di Poisson diventa:

$$\Lambda(A) = n_y (v_2 - v_1) \left[1 + \xi \left(\frac{z - \mu}{\sigma} \right) \right]^{-1/\xi}. \quad (2.26)$$

2.2 Estremi multivariati

In alcune situazioni gli eventi estremi si riferiscono a due o più processi per i quali risulta importante tenere in considerazione la combinazione tra di essi e le interrelazioni che intercorrono. In altre situazioni, invece, viene investigato il comportamento di diverse caratteristiche dello stesso processo oppure viene osservato un processo in diversi luoghi ([7], Capitolo 3, pp. 144-145). In tutte queste situazioni, risulta necessario considerare gli estremi multivariati, piuttosto che analizzarli separatamente sotto l'ipotesi di indipendenza. Alcuni casi di valori estremi multivariati che sono stati investigati riguardano l'ambito ambientale, ad esempio l'analisi spaziale di precipitazioni in diverse località di una regione o lo studio di come una combinazione di livello del mare e altezza d'onda possa provocare il collasso di una diga marittima; oppure l'ambito finanziario, con esempi riguardanti l'analisi dei rendimenti estremi congiunti di diverse valute o un'indagine sulla dipendenza tra mercati azionari internazionali in periodi di alta volatilità ([1], Capitolo 8, p. 251); oppure la progettazione strutturale, come la quantificazione del rischio di fallimento dovuta a una combinazione di più processi ([7], Capitolo 3, pp. 146-147).

Una prima questione che insorge nel caso di estremi multivariati riguarda la definizione e l'identificazione di tali eventi rari, in quanto bisogna stabilire se sia necessario che una sola coordinata raggiunga un valore eccezionale o se sia richiesta l'estremità in tutte le dimensioni contemporaneamente. Un tipico approccio consiste nell'estendere il concetto di estremo massimo, ad esempio, considerando il massimo per ciascuna componente separatamente (*componentwise maxima*). Di conseguenza, vanno ridefiniti e interpretati tutti i concetti della teoria dei valori estremi univariati tradizionali. Inoltre, si generano difficoltà dovute alla dimensionalità sia per la validazione del modello che per i calcoli. Infine, un'altra importante problematica riguarda la gestione della dipendenza degli estremi e il sistema per stimare le diverse strutture di dipendenza che possono intercorrere tra le variabili ([1], Capitolo 8, pp. 251-252).

Nell'analisi degli estremi multivariati, vengono presi in considerazione due aspetti principali: le distribuzioni marginali e la struttura di dipendenza tra le variabili. Dapprima vengono standardizzate le distribuzioni marginali su una scala comune mediante l'uso delle tecniche univariate, per poi passare ad affrontare la questione della dipendenza con un approccio innovativo, dal momento che non è possibile rappresentare la vasta gamma di strutture di dipendenza limite con una semplice famiglia parametrica ([1], Capitolo 8, p. 252).

2.2.1 Estremi massimi in ogni componente

Sia $\{(X_{i,1}, \dots, X_{i,d}), i = 1, \dots, n\}$ una sequenza di realizzazioni indipendenti di un vettore aleatorio X di d dimensioni con funzione di distribuzione congiunta $F(x_1, \dots, x_d)$. Il vettore di estremi massimi multivariati (*componentwise maxima*) è definito estraendo per ciascuna componente il valore massimo:

$$\mathbf{M}_n = (M_{n,1}, \dots, M_{n,d}) = \left(\max_{i=1, \dots, n} X_{i,1}, \dots, \max_{i=1, \dots, n} X_{i,d} \right). \quad (2.27)$$

Dal momento che il valore massimo viene estratto indipendentemente per ciascuna componente, non è detto che esso corrisponda alla stessa realizzazione i ovunque. Di conseguenza, il vettore \mathbf{M}_n non corrisponde necessariamente a un vettore osservato della serie originale. Ciascuna componente $M_{n,j}$, con $j \in \{1, \dots, d\}$, considerata separatamente dalle altre, corrisponde all'estremo massimo di una sequenza di variabili aleatorie indipendenti appartenenti alla stessa distribuzione marginale F_j . Pertanto, è possibile applicare i modelli dei valori estremi univariati per ciascuna componente ed ottenere una distribuzione limite.

La funzione di distribuzione del vettore \mathbf{M}_n è:

$$\begin{aligned}
\mathbb{P}(\mathbf{M}_n \leq \mathbf{x}) &= \mathbb{P}(M_{n,1} \leq x_1, \dots, M_{n,d} \leq x_d) = \\
&= \mathbb{P}(X_{1,1} \leq x_1, \dots, X_{n,1} \leq x_1, \dots, X_{1,d} \leq x_d, \dots, X_{n,d} \leq x_d) = \\
&= \mathbb{P}(X_{1,1} \leq x_1, \dots, X_{1,d} \leq x_d) \cdots \mathbb{P}(X_{n,1} \leq x_1, \dots, X_{n,d} \leq x_d) = \\
&= \mathbb{P}(\mathbf{X}_1 \leq \mathbf{x}) \cdots \mathbb{P}(\mathbf{X}_n \leq \mathbf{x}) = F^n(\mathbf{x}) = F^n(x_1, \dots, x_d),
\end{aligned} \tag{2.28}$$

con $\mathbf{x} \in \mathbb{R}^d$. Analogamente al caso univariato, è necessario normalizzare \mathbf{M}_n in modo tale da ottenere una distribuzione limite non banale quando la dimensione del campione n tende ad infinito. Ciò equivale a trovare sequenze di vettori $\{\mathbf{a}_n > \mathbf{0}\}$, in cui ciascuna componente è positiva, e $\{\mathbf{b}_n\}$ in modo tale che il limite per $n \rightarrow \infty$ di $\mathbf{a}_n^{-1}(\mathbf{M}_n - \mathbf{b}_n)$ sia non degenere e sia descritto da una funzione di distribuzione G con margini non degeneri, ovvero:

$$\begin{aligned}
\mathbb{P}\left(\frac{\mathbf{M}_n - \mathbf{b}_n}{\mathbf{a}_n} \leq \mathbf{x}\right) &= \mathbb{P}\left(\frac{M_{n,1} - b_{n,1}}{a_{n,1}} \leq x_1, \dots, \frac{M_{n,d} - b_{n,d}}{a_{n,d}} \leq x_d\right) = \\
&= F^n(\mathbf{a}_n \mathbf{x} + \mathbf{b}_n) \rightarrow G(\mathbf{x}), \quad n \rightarrow \infty.
\end{aligned} \tag{2.29}$$

Se vale (2.29), G prende il nome di distribuzione dei valori estremi multivariata e F è nel dominio di attrazione di G . La questione diventa, sotto l'ipotesi che valga (2.29), quella di trovare una caratterizzazione della distribuzione congiunta limite G ([1], Capitolo 8, p. 254).

La teoria sviluppata per identificare e analizzare gli estremi multivariati include diverse caratterizzazioni e modelli che non dipendono dalla dimensione dei vettori. Pertanto, è possibile esprimere i concetti restringendo la dimensionalità al caso bivariato, ovvero a vettori di due dimensioni, al fine di semplificare la notazione. In tal caso, il vettore aleatorio viene denotato con (X, Y) .

Una semplificazione delle rappresentazioni, che permette di studiare meglio la struttura di dipendenza, si ottiene assumendo che ciascuna componente segua la stessa distribuzione marginale. La scelta esatta della distribuzione marginale non è fondamentale, a patto che la distribuzione scelta sia in grado di catturare le caratteristiche essenziali delle variabili in questione. Un'opzione particolarmente utile è quella di assumere che X_i e Y_i seguano una distribuzione di Fréchet standardizzata, con funzione di ripartizione:

$$F(z) = \exp\left(-\frac{1}{z}\right), \quad z > 0. \tag{2.30}$$

Tale distribuzione si ottiene come un caso speciale della distribuzione GEV con parametri $\mu = 0$, $\sigma = 1$ e $\xi = 1$. Inoltre, se si considera il caso univariato con M_n data da (2.1) e distribuzione di partenza F come in (2.30), si ottiene che la distribuzione limite G è anch'essa una distribuzione di Fréchet standardizzata:

$$\mathbb{P}\left(\frac{M_n}{n} \leq z\right) = F^n(nz) = \left[\exp\left(-\frac{1}{nz}\right)\right]^n \rightarrow \exp\left(-\frac{1}{z}\right), \quad n \rightarrow \infty, \quad z > 0. \tag{2.31}$$

Di conseguenza, per ottenere lo stesso risultato nel caso multivariato si considera il vettore riscaldato:

$$\mathbf{M}_n^* = (M_{x,n}^*, M_{y,n}^*) = \left(\max_{i=1, \dots, n} \frac{X_i}{n}, \max_{i=1, \dots, n} \frac{Y_i}{n}\right) \tag{2.32}$$

Con questa normalizzazione, tutti i valori, ad eccezione di quelli insolitamente elevati, tendono a 0 quando la dimensione del campione n va all'infinito. In questo modo solo le quantità estreme assumono valori non nulli. Sotto queste ipotesi, vale il seguente risultato ([2], Capitolo 8, p. 144).

Teorema 2.7 Sia M_n^* definito come (2.32), dove (X_i, Y_i) sono vettori indipendenti con distribuzioni marginali Fréchet standardizzate. Allora se

$$\mathbb{P}(M_{x,n}^* \leq x, M_{y,n}^* \leq y) \stackrel{d}{\rightarrow} G(x, y), \quad (2.33)$$

dove G è una funzione di distribuzione non-degenere, G ha la forma

$$G(x, y) = \exp[-V(x, y)], \quad x > 0, y > 0, \quad (2.34)$$

dove

$$V(x, y) = 2 \int_0^1 \max\left(\frac{w}{x}, \frac{1-w}{y}\right) dH(w) \quad (2.35)$$

e H è una funzione di distribuzione su $[0,1]$ che soddisfa il vincolo di media

$$\int_0^1 w dH(w) = \frac{1}{2}. \quad (2.36)$$

La classe di distribuzioni G è in corrispondenza biunivoca con l'insieme delle funzioni di distribuzione H , che soddisfano (2.36), ma che possono essere non differenziabili. Nel caso in cui H è differenziabile e ha densità h , (2.35) diventa

$$V(x, y) = 2 \int_0^1 \max\left(\frac{w}{x}, \frac{1-w}{y}\right) h(w) dw. \quad (2.37)$$

Le funzioni H tali per cui vale (2.35) e che soddisfano la proprietà (2.36) sono molteplici e non è possibile ottenere una distribuzione limite G descritta da un numero finito di parametri. Una possibilità è quella di adottare tecniche non parametriche per ottenere delle stime basate sui dati senza assumere forme specifiche per le distribuzioni. Tuttavia, bisogna garantire che le stime non parametriche soddisfino vincoli del tipo (2.36), e questo comporta delle complicazioni. Un'alternativa è usare sottofamiglie parametriche di distribuzioni per H , che permettono di approssimare una vasta gamma di sottofamiglie per la distribuzione G . Per creare questi modelli si considera una famiglia di distribuzioni H tale per cui la media è pari a 0.5 per ogni valore del parametro. Successivamente, sostituendo la famiglia trovata in (2.35) e (2.34) si ottengono le distribuzioni finali. Alcune classi di distribuzioni comunemente utilizzate per descrivere gli estremi bivariati sono la famiglia logistica, bilogistica e Dirichlet.

La funzione V è detta misura esponenziale e soddisfa

$$\begin{aligned} V(x, \infty) &= \frac{1}{x}, \quad x > 0, \\ V(\infty, y) &= \frac{1}{y}, \quad y > 0, \end{aligned} \quad (2.38)$$

ovvero garantisce che le distribuzioni marginali siano Fréchet unitarie. Inoltre, V è omogenea di ordine -1 , ovvero soddisfa la seguente proprietà:

$$V(a^{-1}x, a^{-1}y) = aV(x, y), \quad \forall a > 0. \quad (2.39)$$

Utilizzando (2.39) in (2.34) si ottiene:

$$G^n(x, y) = G(n^{-1}x, n^{-1}y), \quad n = 2, 3, \dots, \quad (2.40)$$

ovvero, se (X, Y) ha distribuzione G , allora anche \mathbf{M}_n ha distribuzione G , a parte per il fattore n^{-1} . La proprietà (2.40) rappresenta la versione multivariata della proprietà di *max-stability*.

Generalizzando le distribuzioni marginali che appartengono alla famiglia GEV per i risultati della teoria dei valori estremi univariati con parametri (μ_x, σ_x, ξ_x) e (μ_y, σ_y, ξ_y) rispettivamente per X e Y , è possibile ottenere, nel caso bivariato, una distribuzione limite G della forma:

$$G(x, y) = \exp[-V(\tilde{x}, \tilde{y})], \quad (2.41)$$

con V che soddisfa (2.35) per qualche scelta di H e con

$$\tilde{x} = \left[1 + \xi_x \left(\frac{x - \mu_x}{\sigma_x}\right)\right]^{1/\xi_x} \quad \text{e} \quad \tilde{y} = \left[1 + \xi_y \left(\frac{y - \mu_y}{\sigma_y}\right)\right]^{1/\xi_y}, \quad (2.42)$$

a condizione che

$$\left[1 + \xi_x \left(\frac{x - \mu_x}{\sigma_x}\right)\right] > 0 \quad \text{e} \quad \left[1 + \xi_y \left(\frac{y - \mu_y}{\sigma_y}\right)\right] > 0.$$

La funzione di densità del modello si ricava da (2.34) ed è:

$$g(x, y) = \left[\frac{\partial V(x, y)}{\partial x} \frac{\partial V(x, y)}{\partial y} - \frac{\partial^2 V(x, y)}{\partial x \partial y} \right] \exp[-V(x, y)], \quad x > 0, y > 0. \quad (2.43)$$

La funzione V definita in (2.35) si generalizza in più dimensioni con la rappresentazione spettrale dovuta a Pickands nel 1981 data da ([11], [12]):

$$V(\mathbf{x}) = \int_{S_d} \max_{j=1, \dots, d} \frac{w_j}{x_j} dH(\mathbf{w}), \quad (2.44)$$

dove H è una misura finita, detta misura spettrale, definita sul simpleso d -dimensionale

$$S_d = \left\{ \mathbf{w} \in [0, 1]^d : \sum_{j=1}^d w_j = 1 \right\}, \quad (2.45)$$

e tale che

$$\int_{S_d} w_j dH(\mathbf{w}) = 1, \quad \forall j = 1, \dots, d. \quad (2.46)$$

La condizione (2.38) che garantisce che le marginali siano Fréchet unitari si generalizza di conseguenza con:

$$V(\infty, \dots, \infty, x_j, \infty, \dots, \infty) = \frac{1}{x_j}, \quad x_j > 0, \quad \forall j = 1, \dots, d. \quad (2.47)$$

È possibile considerare assunzioni diverse riguardo la distribuzione marginale per ciascuna componente del vettore di estremi massimi multivariati (2.27). Nella letteratura, sono state esplorate, ad esempio, distribuzioni come esponenziale, Weibull, Gumbel e uniforme. La scelta della distribuzione marginale non cambia sostanzialmente i risultati, ma comporta una caratterizzazione differente di (2.35). Queste distribuzioni alternative offrono prospettive diverse per analizzare i dati e possono essere più adatte a certe applicazioni, come ad esempio la distribuzione esponenziale utilizzata tipicamente per

analizzare un vettore di estremi minimi multivariato. Infatti, alcune proprietà o caratterizzazioni sono più facilmente osservabili con una distribuzione specifica. Inoltre, diverse scelte di distribuzione marginale possono talvolta portare all'uso di metodi statistici diversi ([1], Capitolo 8, p. 271-273).

2.2.2 Misure di dipendenza asintotica

Il caso speciale di completa dipendenza asintotica tra le componenti X_j , con $j = 1, \dots, d$, in (2.29) si verifica quando la distribuzione limite G si esprime come:

$$G(\mathbf{x}) = \min_{j=1, \dots, d} G_j(x_j), \quad \mathbf{x} \in \mathbb{R}^d. \quad (2.48)$$

e la funzione esponenziale si può scrivere come:

$$V(\mathbf{x}) = \frac{1}{\min(x_1, \dots, x_d)}, \quad \mathbf{x} \in \mathbb{R}^d. \quad (2.49)$$

Al contrario, il caso speciale di totale indipendenza asintotica tra le componenti X_j , con $j = 1, \dots, d$, in (2.29) si verifica quando la distribuzione limite G si fattorizza rispetto alle distribuzioni marginali:

$$G(\mathbf{x}) = \prod_{j=1}^d G_j(x_j), \quad \mathbf{x} \in \mathbb{R}^d, \quad (2.50)$$

e la funzione esponenziale si può scrivere come:

$$V(\mathbf{x}) = \frac{1}{x_1} + \dots + \frac{1}{x_d}, \quad \mathbf{x} \in \mathbb{R}^d. \quad (2.51)$$

Nello specifico, si ottiene un'espressione semplice per la distribuzione limite G nei seguenti casi particolari di distribuzioni marginali appartenenti alle tre classi Gumbel, Fréchet e Weibull [7]:

I. Gumbel: $G_j(x_j) = \exp[-\exp(-x_j)]$, allora $G(\mathbf{0}) = \exp(-d)$;

II. Fréchet: $G_j(x_j) = \exp[-x_j^{-\alpha_j}]$, $\alpha_j > 0$, allora $G(\mathbf{1}) = \exp(-d)$;

III. Weibull $G_j(x_j) = \exp[-(-x_j)^{\alpha_j}]$, $\alpha_j > 0$, allora $G(-\mathbf{1}) = \exp(-d)$.

Un modo per misurare la forza di dipendenza degli estremi è quello di raggruppare le caratteristiche essenziali della struttura di dipendenza in alcuni coefficienti accuratamente definiti che forniscono una rappresentazione approssimativa ma accurata dell'intera struttura di dipendenza. Nel caso di un vettore (X, Y) , in cui le distribuzioni marginali sono F_X e F_Y , si definisce un primo coefficiente come:

$$\chi = \lim_{u \rightarrow 1} \mathbb{P}(F_Y(Y) > u | F_X(X) > u), \quad (2.52)$$

Questa misura limite rappresenta la tendenza di una variabile ad essere grande condizionalmente all'essere grande dell'altra variabile. Essa assume valori tra 0 e 1, e si ottiene il caso di indipendenza asintotica tra X e Y quando $\chi = 0$. Il valore di χ aumenta nel caso di una dipendenza degli estremi maggiori. Tuttavia, la misura χ è in grado di discriminare la dipendenza estrema nel caso di distribuzioni asintoticamente dipendenti. Nel caso di distribuzioni asintoticamente indipendenti, invece, si può definire il coefficiente:

$$\bar{\chi} = \lim_{u \rightarrow 1} \bar{\chi}(u), \quad (2.53)$$

dove, per $0 < u < 1$:

$$\begin{aligned}\bar{\chi}(u) &= \frac{2\log[\mathbb{P}(F_X(X) > u)]}{\log[\mathbb{P}(F_X(X) > u, F_Y(Y) > u)]} - 1 = \\ &= \frac{2\log[1 - u]}{\log(\mathbb{P}(F_X(X) > u, F_Y(Y) > u))} - 1.\end{aligned}\tag{2.54}$$

In tal caso, $\bar{\chi} \in [-1, 1]$ e si ottiene il caso di dipendenza asintotica tra X e Y quando $\bar{\chi} = 1$, mentre per variabili indipendenti $\bar{\chi} = 0$. Per variabili asintoticamente indipendenti, il valore di $\bar{\chi}$ aumenta nel caso di una maggiore dipendenza ai livelli estremi.

La coppia $(\chi, \bar{\chi})$ va dunque interpretata nel seguente modo:

- se $\bar{\chi} = 1$ le variabili sono asintoticamente dipendenti e il valore di χ misura la forza della dipendenza estrema,
- se $\bar{\chi} < 1$ allora $\chi = 0$, le variabili sono asintoticamente indipendenti e il valore di $\bar{\chi}$ è una misura più adatta della forza della dipendenza estrema.

2.2.3 Variabili di struttura

Nel caso in cui si voglia studiare il comportamento estremo di una qualche combinazione delle variabili, è possibile definire la variabile di struttura (*structure variable*):

$$Z = \Phi(\mathbf{M}_n).\tag{2.55}$$

Alcune possibilità includono il massimo, il minimo o la somma delle componenti di \mathbf{M}_n . Denotando con g la funzione di densità di \mathbf{M}_n , la funzione di distribuzione di Z si ottiene da:

$$\mathbb{P}(Z \leq z) = \int_{A_z} g(x, y) dx dy,\tag{2.56}$$

dove $A_z = \{(x, y): \Phi(x, y) \leq z\}$.

Nel caso particolare in cui Φ rappresenti il massimo, è possibile evitare l'integrazione e la distribuzione di Z si trova facilmente come:

$$\mathbb{P}(Z \leq z) = G(z, z),\tag{2.57}$$

con G distribuzione congiunta di \mathbf{M}_n .

2.2.4 Modello di processo puntuale

La caratterizzazione come processo puntuale può essere generalizzata al caso bivariato dal seguente risultato ([2], Capitolo 8, p. 156).

Teorema 2.8 *Sia $(x_1, y_1), (x_2, y_2), \dots$ una sequenza di osservazioni bivariate indipendenti da una distribuzione con margini standard Fréchet che soddisfa la convergenza per componentwise maxima*

$$\mathbb{P}(M_{x,n}^* \leq x, M_{y,n}^* \leq y) \rightarrow G(x, y).$$

Sia $\{N_n\}$ una sequenza di processi puntuali definiti da

$$N_n = \left\{ \left(\frac{x_1}{n}, \frac{y_1}{n} \right), \dots, \left(\frac{x_n}{n}, \frac{y_n}{n} \right) \right\}.\tag{2.58}$$

Allora,

$$N_n \xrightarrow{d} N, \quad (2.59)$$

su regioni limitate dall'origine $(0,0)$, dove N è un processo di Poisson non-omogeneo su $(0, \infty) \times (0, \infty)$. Inoltre, denotando

$$r = x + y \quad e \quad w = \frac{x}{x + y}, \quad (2.60)$$

la funzione di intensità di N è

$$\lambda(r, w) = 2 \frac{dH(w)}{r^2}, \quad (2.61)$$

dove H è legata a G attraverso (2.34) e (2.35).

La trasformazione $(x, y) \rightarrow (r, w)$ definita in (2.60) equivale a passare da coordinate cartesiane a coordinate pseudo-polari, tali per cui r misura la distanza dall'origine, con dominio $(0, \infty)$, e w misura l'angolo in una scala $[0, 1]$. In altre parole, w misura quanto è grande x rispetto alla somma $x + y$. Nello specifico, $w = 0$ corrisponde al caso di x molto piccolo rispetto a y fino a ottenere l'asse $x = 0$, mentre $w = 1$ corrisponde al caso di x molto grande rispetto a y fino a ottenere l'asse $y = 0$, e $w \approx 1/2$ indica che x e y hanno valori simili. Ne consegue che, guardando l'equazione (2.61), la frequenza di avere entrambi gli eventi estremi diminuisce quadraticamente con l'aumentare della distanza tra di essi, mentre H determina l'ampiezza angolare dei punti di N , la quale è invece indipendente dalla distanza radiale. Se H è una funzione differenziabile con densità h , allora $h(\cdot)$ indica la frequenza relativa degli eventi rispetto alla distanza angolare. Quando gli estremi sono quasi indipendenti, grandi valori di x/n si verificano insieme a piccoli valori di y/n e viceversa. In questo scenario, $h(w)$ è elevata vicino a $w = 0$ e $w = 1$, mentre è bassa altrove. Invece, se esiste una forte dipendenza tra i due eventi estremi, facendo sì che i valori x/n e y/n siano simili, allora $h(w)$ sarà elevata vicino a $w = 1/2$. La necessità di mantenere una maggiore generalità in (2.61), senza presumere che H abbia sempre una densità h , permette di considerare due casi limite speciali: quando G rappresenta variabili indipendenti, la funzione di misura H ha atomi con massa $1/2$ a $w = 0$ e $w = 1$; mentre, quando G rappresenta variabili perfettamente dipendenti, H consiste in un atomo con massa unitaria a $w = 1/2$.

Considerando il modello dei *componentwise block maxima*, la caratterizzazione come processi puntuali corrisponde a:

$$\mathbb{P}(M_{x,n}^* \leq x, M_{y,n}^* \leq y) = \mathbb{P}(N_n(A) = 0), \quad (2.62)$$

dove N_n è il processo definito da (2.58) e $A = \{(0, \infty) \times (0, \infty)\} \setminus \{(0, x) \times (0, y)\}$. Ne consegue che:

$$\mathbb{P}(M_{x,n}^* \leq x, M_{y,n}^* \leq y) \rightarrow \mathbb{P}(N(A) = 0) = \exp[-\Lambda(A)], \quad (2.63)$$

dove

$$\begin{aligned} \Lambda(A) &= \int_A 2 \frac{dr}{r^2} dH(w) = \int_{w=0}^1 \int_{r=\min\{x/w, y/(1-w)\}}^{\infty} 2 \frac{dr}{r^2} dH(w) = \\ &= 2 \int_{w=0}^1 \max\left(\frac{w}{x}, \frac{1-w}{y}\right) dH(w). \end{aligned} \quad (2.64)$$

Capitolo 3

Estremi spaziali

Numerosi eventi estremi, come ondate di calore, forti precipitazioni e neviccate, maree e tempeste di vento, sono manifestazioni rare e intense di rischi naturali che, per loro stessa natura, si estendono nello spazio, coinvolgendo aree geografiche ampie e variabili. L'analisi degli estremi in un contesto spaziale può avere molteplici scopi. Ad esempio, permette di stimare i cambiamenti negli eventi estremi, come l'aumento delle temperature massime giornaliere, integrando i dati provenienti da vari siti e considerando la loro dipendenza reciproca. Inoltre, aiuta a collegare eventi rari a cause specifiche e a valutare il rischio complessivo di grandi eventi, come uragani o inondazioni, utili per le compagnie di assicurazione e la pianificazione della sicurezza pubblica. Infine, è fondamentale per stimare il rischio in località cruciali, come gli impianti nucleari e industriali oppure le zone densamente urbanizzate, utilizzando dati provenienti da diverse aree per ridurre l'incertezza delle stime, soprattutto quando i dati locali sono limitati [20].

Nei contesti pratici, tali fenomeni estremi sono osservati in un numero finito di località e per uno specifico periodo di tempo. I siti e i tempi delle osservazioni possono essere deterministici, ad esempio dati presi giornalmente da una griglia di punti di una regione, oppure casuali, come nel caso di grandi incendi boschivi. Si potrebbe pensare di utilizzare i modelli multivariati per descrivere gli estremi spaziali. Tuttavia, quando il numero di siti è elevato, tali modelli diventano difficili da gestire e quando si vuole aggregare i dati su aree più grandi o prevedere il valore che essi possono assumere in nuove località, non è possibile attenersi semplicemente ai risultati degli estremi multivariati. In alcuni settori, come l'idrologia, risulta necessario poter rispondere ai precedenti quesiti, e, di conseguenza, estendere i modelli multivariati.

Nel Capitolo 2, è stata introdotta e presentata la teoria dei valori estremi in una dimensione finita, ovvero estremi di variabili o vettori aleatori. Nel presente capitolo, invece, si esplorano gli estremi in una dimensione infinita, ovvero estremi di processi stocastici, e vengono definiti modelli statistici in grado di catturare la dipendenza spaziale del processo stocastico e fornire una base affidabile per l'estrapolazione in ambienti ad alta dimensionalità. Nel contesto dei processi continui (nello spazio, nel tempo o in entrambi), il concetto di massimo viene esteso al massimo “punto per punto” (*pointwise maxima*).

3.1 Processi max-stable

Sia S un sottoinsieme compatto di \mathbb{R}^d , con $d \geq 1$, e sia $\mathcal{C}(S)$ lo spazio delle funzioni continue (casuali) definite su S . Una naturale estensione della teoria dei valori estremi ai processi stocastici è rappresentata dai processi max-stable, definiti nel seguente modo in [13].

Definizione 3.1 Sia X_1, X_2, \dots , una sequenza di copie indipendenti di un processo stocastico $\{X(s): s \in S\}$. Se per ogni $n \geq 1$ esistono funzioni di normalizzazione continue $\{a_n(s) > 0, s \in S\}$ e $\{b_n(s) \in \mathbb{R}, s \in S\}$ tali che per ogni $s \in S$

$$\frac{\max_{i=1, \dots, n} X_i(s) - b_n(s)}{a_n(s)} \xrightarrow{d} Z(s), \quad (3.1)$$

allora $\{Z(s): s \in S\}$ è detto *max-stable*.

Quando si parla di processi stocastici con traiettorie (*sample paths*) continue definiti su un insieme compatto, dire che due processi hanno la stessa distribuzione implica che, per ogni insieme finito di punti all'interno del dominio considerato, le distribuzioni congiunte delle variabili casuali in questi punti siano identiche.

L'utilizzo dei processi max-stable per descrivere gli estremi spaziali si basa su argomenti asintotici, come nel caso univariato, in quanto si manifestano come i massimi puntuali presi su un numero infinito di processi stocastici (appropriatamente riscaldati), a condizione che il limite esista. Ciò equivale a dire che, considerando un numero elevato di realizzazioni indipendenti di un processo stocastico spaziale, i massimi di queste realizzazioni in ogni sito, dopo una corretta normalizzazione e a patto che il limite esista, convergeranno in distribuzione a un processo max-stable. Il seguente risultato è stato dimostrato da [16].

Teorema 3.1 Sia X_1, X_2, \dots , una sequenza di copie indipendenti di un processo stocastico $\{X(s): s \in S\}$ con traiettorie continue. Se esistono funzioni di normalizzazione continue $\{c_n(s) > 0, s \in S, n \geq 1\}$ e $\{d_n(s) \in \mathbb{R}, s \in S, n \geq 1\}$ tali che il processo limite $\{Z(s): s \in S\}$ definito da

$$\frac{\max_{i=1, \dots, n} X_i(s) - d_n(s)}{c_n(s)} \rightarrow Z(s), \quad s \in S, \quad n \rightarrow \infty, \quad (3.2)$$

è non degenera, allora il processo $\{Z(s): s \in S\}$ è un processo max-stable.

La convergenza in (3.2) è una convergenza debole nello spazio delle funzioni continue in S . In modo analogo al caso univariato, in cui l'attenzione si focalizza sulla distribuzione limite dei *block maxima*, ovvero la distribuzione GEV, si assume che il processo stocastico limite $\{Z(s): s \in S\}$ sia una buona approssimazione per modellare il processo dei massimi parziali $\{\max_{i=1, \dots, n} X_i(s): s \in S\}$, per n sufficientemente grande. Inoltre, la sua distribuzione marginale unidimensionale è la distribuzione GEV, per coerenza con i risultati univariati, mentre la distribuzione marginale K -dimensionale appartiene alla classe delle distribuzioni dei valori estremi multivariate. Per scopi teorici, è più conveniente assumere che i margini siano Fréchet unitari:

$$\mathbb{P}(Z(s) \leq z) = \exp\left(-\frac{1}{z}\right), \quad z > 0, \quad s \in S, \quad (3.3)$$

al fine di restringere l'attenzione sulla struttura di dipendenza spaziale. Il processo stocastico $\{Z(s): s \in S\}$ con margini Fréchet unitari viene detto processo max-stable semplice.

3.1.1 Rappresentazione spettrale

Al fine di comprendere meglio la struttura e le proprietà dei processi max-stable, facilitando la loro analisi, modellazione e simulazione, sono state introdotte alcune caratterizzazioni di tali processi. La rappresentazione spettrale, proposta da [16] e ripresa da [17], è una delle più utilizzate in quanto esprime un processo max-stable come combinazioni di componenti più semplici.

Teorema 3.2 ([13]) *Ogni processo max-stable semplice non degenero $\{Z(s): s \in S\}$ definito su un insieme compatto $S \subset \mathbb{R}^d$, $d \geq 1$, con traiettorie continue soddisfa:*

$$Z(s) \stackrel{d}{=} \max_{i \geq 1} \zeta_i f_i(s), \quad s \in S, \quad (3.4)$$

dove $\{(\zeta_i, f_i): i \geq 1\}$ sono punti di un processo di Poisson su $(0, \infty) \times \mathcal{C}$ con intensità $\zeta^{-2} d\zeta v(df)$ per qualche misura finita locale v definita nello spazio \mathcal{C} delle funzioni continue non negative su S tali che

$$\int f(s) v(df) = 1 \quad s \in S. \quad (3.5)$$

La caratterizzazione spettrale (3.4) non è unica, in quanto misure v differenti possono portare allo stesso processo max-stable semplice $\{Z(s): s \in S\}$. Inoltre, il vincolo di considerare solo funzioni non negative è imposto per comodità. Infatti, nella caratterizzazione spettrale, dove si prendono i massimi puntuali di un numero infinito di funzioni, è possibile utilizzare anche funzioni reali. L'unica condizione necessaria è che la misura sia positiva: $v\{f(s) > 0\} > 0$ per ogni $s \in S$.

I margini del processo max-stable $\{Z(s): s \in S\}$ sono ancora Fréchet unitari. Infatti:

$$\begin{aligned} \mathbb{P}(Z(s) \leq z) &= \mathbb{P}((\zeta, f) \notin (\zeta, f) \in (0, \infty) \times S: \zeta f(s) > z) = \\ &= \exp \left[- \int_{\mathfrak{X}} \int_{z/f(s)}^{\infty} \zeta^{-2} d\zeta v(df) \right] = \\ &= \exp \left[- \int_{\mathfrak{X}} z^{-1} f(s) v(df) \right] = \exp \left(-\frac{1}{z} \right). \end{aligned} \quad (3.6)$$

È possibile estendere i risultati precedenti ai processi con distribuzioni marginali appartenenti ad altre distribuzioni dei valori estremi, oppure derivare le proprietà dei processi $\{Z^{1/\alpha}(s), s \in S\}$, $\{\log Z(s): s \in S\}$, $\{-Z^{-1/\alpha}(s), s \in S\}$ da quelle del processo max-stable $\{Z(s): s \in S\}$ considerato finora, o introdurre i processi min-stable [16].

La proprietà di *max-stability* di $Z(\cdot)$ deriva dal fatto che la sovrapposizione di n processi di Poisson indipendenti e identici è un processo di Poisson con la sua intensità moltiplicata per n . Più precisamente, per ogni sottoinsieme di punti $\{s_1, \dots, s_k\} \subset S$:

$$\left\{ \max_{i=1, \dots, n} X_i(s_1), \dots, \max_{i=1, \dots, n} X_i(s_k) \right\} \sim n\{Z(s_1), \dots, Z(s_k)\}, \quad k \in \mathbb{N}, \quad (3.7)$$

ovvero hanno la stessa distribuzione.

[17] ha fornito un'interpretazione basata sui fenomeni temporaleschi per questa rappresentazione spettrale. Ha supposto che dei temporali si abbattano sulla regione S e che la misura v descriva come sono distribuite. Ogni ζ_i rappresenta l'intensità delle precipitazioni mentre la funzione f_i ne descrive la forma. Dunque, il prodotto $\zeta_i f_i(s)$ rappresenta la quantità di pioggia che cade in una specifica posizione $s \in S$ durante l'evento. Il processo max-stable appare come la pioggia massima osservata rispetto a un numero infinito di eventi $\{(\zeta_i, f_i(s): s \in S): i \geq 1\}$ indipendenti.

Una seconda rappresentazione è stata proposta da [18], dove le funzioni deterministiche f_i vengono sostituite da processi casuali stazionari.

Teorema 3.3 *Sia Y_1, Y_2, \dots una sequenza di copie indipendenti di un processo stocastico non-negativo $\{Y(s): s \in S\}$ stazionario e tali che $\mathbb{E}[Y(s)] = 1$ per ogni $s \in S$. Allora:*

$$Z(s) \stackrel{d}{=} \max_{i \geq 1} \zeta_i Y_i(s), \quad s \in S, \quad (3.8)$$

è un processo max-stable semplice stazionario, dove $\{\zeta_i: i \geq 1\}$ sono punti di un processo di Poisson su $(0, \infty)$ con intensità $d\Lambda(\zeta) = \zeta^{-2} d\zeta$.

La caratterizzazione (3.8) può essere vista come un caso speciale di (3.4), quando ν è una misura di probabilità. Con questa rappresentazione spettrale, i punti $\{\zeta_i: i \geq 1\}$ rappresentano il raggio, in quanto ζ_i determina quanto è grande e rilevante il contributo specifico i al massimo puntuale $\{Z(s): s \in S\}$. Invece, i processi stocastici $\{Y_i(s): s \in S\}$ rappresentano l'angolo, in quanto $Y_i(s)$ rappresenta la forma e la distribuzione spaziale, proprio come l'angolo determina la direzione del punto nelle coordinate polari. Questa rappresentazione è potente perché separa l'informazione sulla magnitudine (raggio) dall'informazione sulla struttura spaziale (angolo), facilitando l'analisi e la comprensione del comportamento estremo in termini di dipendenza spaziale (e temporale) nei processi max-stable. Inoltre, la rappresentazione spettrale può facilitare la simulazione e l'analisi dei processi max-stable, fornendo un framework teorico solido per lo sviluppo di metodi computazionali e algoritmi.

Riprendendo l'interpretazione di [17], gli eventi $\zeta_i Y_i(s)$ rappresentano le precipitazioni spaziali giornaliere. L'estensione spaziale è guidata da $\{Y(s): s \in S\}$, ovvero la struttura di dipendenza è la stessa per ogni evento, mentre l'intensità ζ_i è diversa. Quando $\{Y(s): s \in S\}$ rappresenta una forte dipendenza spaziale, è probabile che la tempesta influenzi l'intera regione di studio S . A differenza della rappresentazione (3.4), la forma della tempesta non è più deterministica.

La stazionarietà del processo max-stable nella formulazione del **Teorema 3.3** comporta che è possibile descrivere le distribuzioni congiunte di $\{Z(s): s \in S\}$ per coppie di punti $s_i, s_j \in S$ considerando i punti θ ed $h = s_i - s_j$ senza perdita di generalità.

È possibile esprimere la caratterizzazione (3.8) anche come:

$$Z(s) \stackrel{d}{=} \max_{i \geq 1} \varphi_i(s), \quad s \in S, \quad (3.9)$$

dove $\Phi = \{\varphi_i(s) = \zeta_i Y_i(s), s \in S, i \geq 1\}$ è un processo di Poisson sullo spazio delle funzioni continue non negative \mathcal{C}^+ , con misura di intensità

$$\Lambda(A) = \int_0^\infty \mathbb{P}(\zeta Y \in A) \zeta^{-2} d\zeta, \quad (3.10)$$

per tutti gli insiemi di Borel $A \subset \mathcal{C}^+$. Le funzioni $\{\varphi_i, i \geq 1\}$ vengono dette funzioni spettrali [11]. Il processo Φ può essere decomposto in due sottoprocessi puntuali:

$$\begin{aligned} \Phi^+ &= \{\varphi \in \Phi: \exists s \in S, \varphi(s) = Z(s)\}, \\ \Phi^- &= \{\varphi \in \Phi: \varphi(s) < Z(s), s \in S\}, \end{aligned} \quad (3.11)$$

i cui atomi, rispettivamente, vengono denotati da Dombry, Eyi-Minko e Ribatet nel 2013 con funzioni estreme e sub-estreme [13].

3.1.2 Distribuzioni di dimensione finita

Nelle applicazioni, il processo stocastico $\{Z(s): s \in S\}$ non è osservato nell'intera regione S , ma in un numero finito di località $s_1, \dots, s_k \in S, k \geq 1$. Ne consegue che il vettore aleatorio $\{Z(s_1), \dots, Z(s_k)\}$ è necessariamente max-stable e la sua distribuzione è facilmente derivabile dalla caratterizzazione

spettrale (3.4) o (3.8) e dalla misura di conteggio del processo di Poisson associato. Siano $z_1, \dots, z_k > 0$ e sia

$$\begin{aligned} A &= \{(\zeta, f) \in (0, \infty) \times \mathcal{C}^+(S): \zeta f(s_j) > z_j \text{ per qualche } j = 1, \dots, k\} = \\ &= \left\{ (\zeta, f) \in (0, \infty) \times \mathcal{C}^+(S): \zeta > \min_{j=1, \dots, k} \frac{z_j}{f(s_j)} \right\}. \end{aligned} \quad (3.12)$$

Allora:

$$\begin{aligned} \mathbb{P}(Z(s_1) \leq z_1, \dots, Z(s_k) \leq z_k) &= \mathbb{P}(\text{nessun atomo di } \{(\zeta_i, Y_i): i \in \mathbb{N}\} \in A) = \\ &= \exp \left[- \int \int_0^\infty 1_{\{(\zeta, f) \in A\}} \zeta^{-2} d\zeta dP(f) \right] = \\ &= \exp \left[- \int \max_{j=1, \dots, k} \frac{f(s_j)}{z_j} dP(f) \right] = \\ &= \exp \left[- \mathbb{E} \left(\max_{j=1, \dots, k} \frac{Y(s_j)}{z_j} \right) \right] \\ &= \exp[-V(z_1, \dots, z_k)], \end{aligned} \quad (3.13)$$

dove dP denota la misura di probabilità del processo stocastico Y e V è la misura esponenziale introdotta in 2.2.1. Quando $k = 1$, si ottiene la distribuzione Fréchet standard in (3.13).

Si ottiene lo stesso processo max-stable per diverse scelte per Y in (3.8). Infatti, sia

$$\tilde{Z}(s) = \max_{i \geq 1} \zeta_i W_i Y_i(s), \quad s \in S, \quad (3.14)$$

dove $\{(\zeta_i, Y_i): i \in \mathbb{N}\}$ sono come in (3.8) e W_i sono copie indipendenti di una variabile casuale non negativa W tale che $\mathbb{E}(W) = 1$, con W indipendente dai punti $\{(\zeta_i, Y_i): i \in \mathbb{N}\}$. Da (3.13) si ha:

$$\begin{aligned} \mathbb{P}(\tilde{Z}(s_1) \leq z_1, \dots, \tilde{Z}(s_k) \leq z_k) &= \exp \left[- \mathbb{E} \left(\max_{j=1, \dots, k} \frac{WY(s_j)}{z_j} \right) \right] = \\ &= \exp \left[- \mathbb{E} \left(\max_{j=1, \dots, k} \frac{Y(s_j)}{z_j} \right) \right] = \\ &= \mathbb{P}(Z(s_1) \leq z_1, \dots, Z(s_k) \leq z_k), \end{aligned} \quad (3.15)$$

ovvero Z e \tilde{Z} hanno la stessa distribuzione.

3.2 Modelli

La rappresentazione spettrale (3.8) è molto flessibile e può essere adattata per costruire vari modelli specifici. È possibile scegliere diverse funzioni di peso e distribuzioni di $\{Y(s): s \in S\}$ per catturare la dipendenza spaziale desiderata.

3.2.1 Processo di Smith

Un primo modello è stato proposto da [17], da cui prende il nome di processo di Smith o anche processo dei valori estremi Gaussiano. Esso è ottenuto considerando

$$f_i(s) = \varphi(s - U_i; 0, \Sigma), \quad s \in S, \quad (3.16)$$

dove $(\zeta_i, U_i): i \geq 1$ sono punti di un processo di Poisson su $(0, \infty) \times \mathbb{R}^d$ con misura di intensità $\zeta^{-2} d\zeta du$, mentre $\varphi(\cdot, 0, \Sigma)$ è la funzione di densità d-dimensionale di un vettore aleatorio Gaussiano con media zero e matrice di covarianza Σ , che determina la forma degli eventi estremi e controlla la dipendenza tra le componenti $Z(s_i)$. Nel caso bivariato, la funzione di distribuzione di questo modello max-stable è data da:

$$\mathbb{P}(Z(s_1) \leq z_1, Z(s_2) \leq z_2) = \exp \left[-\frac{1}{z_1} \Phi \left(\frac{a}{2} + \frac{1}{a} \log \frac{z_2}{z_1} \right) - \frac{1}{z_2} \Phi \left(\frac{a}{2} + \frac{1}{a} \log \frac{z_1}{z_2} \right) \right], \quad (3.17)$$

dove Φ è la funzione di distribuzione cumulativa di una normale standard e $a^2 = h^T \Sigma^{-1} h$ è la distanza di Mahalanobis, con $h = s_1 - s_2$. Tuttavia, il processo di Smith è poco utilizzato nelle applicazioni in quanto non è flessibile dal momento che la forma delle densità Gaussiane multivariate è troppo restrittiva e irrealistica.

3.2.2 Processo di Schlather

[18] ha proposto un nuovo modello per i processi max-stable, da cui prende il nome di processo di Schlather o anche processo Gaussiano estremo. Esso è ottenuto considerando

$$Y_i(s) = \sqrt{2\pi} \max\{0, \varepsilon_i(s)\}, \quad s \in S, \quad (3.18)$$

dove ε_i sono copie indipendenti di un processo Gaussiano standard con funzione di correlazione ρ . Il fattore $\sqrt{2\pi}$ è necessario affinché il valore atteso di Y_i sia uguale a 1 per qualsiasi $s \in S$. Nel caso bivariato, la funzione di distribuzione di questo modello max-stable è data da:

$$\mathbb{P}(Z(s_1) \leq z_1, Z(s_2) \leq z_2) = \exp \left[-\frac{1}{2} \left(\frac{1}{z_1} + \frac{1}{z_2} \right) \left(1 + \sqrt{1 - \frac{2[1 + \rho(h)]z_1 z_2}{(z_1 + z_2)^2}} \right) \right], \quad (3.19)$$

dove $h = s_1 - s_2$.

La funzione di correlazione ρ può assumere valori in $[-1, 1]$, dove $\rho = 1$ corrisponde a variabili completamente dipendenti e $\rho = -1$ corrisponde a variabili indipendenti. In quest'ultimo caso, infatti, (3.19) diventa semplicemente:

$$\mathbb{P}(Z(s_1) \leq z_1, Z(s_2) \leq z_2) = \exp \left[-\left(\frac{1}{z_1} + \frac{1}{z_2} \right) \right] = \mathbb{P}(Z(s_1) \leq z_1) \cdot \mathbb{P}(Z(s_2) \leq z_2). \quad (3.20)$$

Una tipica scelta per la funzione di correlazione è quella di considerare la famiglia *powered exponential*:

$$\rho(h) = \exp \left[-\left(\frac{h}{\lambda} \right)^\kappa \right], \quad \lambda > 0, \quad 0 < \kappa \leq 2, \quad (3.21)$$

dove λ è il parametro di *range*, che controlla la distanza a cui la correlazione diminuisce significativamente, mentre κ è il parametro di *smooth*, che determina la "morbidezza" della correlazione. In particolare, per $\kappa = 1$ si ottiene la funzione di correlazione esponenziale semplice. Inoltre, è possibile aggiungere un *sill* c_1 e un effetto *nugget* v_2 alla funzione di correlazione tali per cui $v_2 = 1 - c_1$, ottenendo

$$\rho_*(h) = \begin{cases} \nu_2 + c_1, & h = 0, \\ c_1 \rho(h), & h > 0. \end{cases} \quad (3.22)$$

Tale funzione di correlazione ρ è isotropa, ovvero è identica per qualsiasi direzione del vettore $s_1 - s_2$, e dipende solo dalla distanza h tra le due località, ma è possibile renderla anisotropa con delle trasformazioni.

Altre tipiche scelte per la funzione di correlazione sono Whittle-Matérn, Cauchy e Bessel.

3.2.3 Processo di Brown-Resnick

Nel 1977 Brown e Resnick hanno proposto un modello per i processi max-stable, da cui il nome di processo di Brown-Resnick, che si ottiene considerando

$$Y_i(s) = \exp\{\varepsilon_i(s) - \gamma(s)\}, \quad s \in S, \quad (3.23)$$

dove ε_i sono copie indipendenti di un processo Gaussiano centrato con incrementi $\varepsilon_i(s+h) - \varepsilon_i(s)$ stazionari e semi-variogramma $\gamma(h) = \text{Var}(\varepsilon_i(s+h) - \varepsilon_i(s))/2$, che può essere illimitato. Tipicamente, il processo $\{\varepsilon_i(s), s \in S\}$ è un moto Browniano frazionario, e il semi-variogramma è:

$$\gamma(h) = \left(\frac{h}{\lambda}\right)^\kappa, \quad \lambda > 0, \quad 0 < \kappa \leq 2, \quad (3.24)$$

dove λ è il parametro di *range* e κ è il parametro di *smooth*. Nel caso particolare di $\kappa = 2$ si ottiene il processo di Smith. Ne consegue che il processo di Brown-Resnick ne è una generalizzazione che supera le limitazioni del processo più semplice. Nel caso bivariato, la funzione di distribuzione di questo modello max-stable è data da (3.17), ma con $a^2 = \text{Var}(Y(s_1 - s_2))$. Anche se il processo Y non è stazionario, il processo max-stable ad esso associato lo è.

3.2.4 Processo Extremal-t

Un altro modello tipicamente utilizzato nelle applicazioni è il processo Extremal-t. Esso è stato introdotto sulla base della rappresentazione spettrale da [19] e si ottiene considerando

$$Y_i(s) = \sqrt{\pi} 2^{-\frac{\nu-2}{2}} \Gamma\left(\frac{\nu+1}{2}\right)^{-1} \max\{0, \varepsilon_i(s)\}^\nu, \quad s \in S, \nu \geq 1, \quad (3.25)$$

dove ε_i sono copie indipendenti di un processo Gaussiano standard con funzione di correlazione ρ e Γ è la funzione Gamma. La funzione di correlazione ρ è come nel processo di Schlather. Inoltre, quando $\nu = 1$, si ottiene come caso particolare il processo di Schlather (3.18), mentre, quando valgono determinate condizioni e $\nu \rightarrow \infty$ si ottiene il processo di Brown-Resnick. Nel caso bivariato, la funzione di distribuzione del modello max-stable extremal-t è data da [14]:

$$\mathbb{P}(Z(s_1) \leq z_1, Z(s_2) \leq z_2) = \exp \left[-\frac{1}{z_1} T_{\nu+1} \left\{ -\frac{\rho(h)}{b} + \frac{1}{b} \left(\frac{z_2}{z_1}\right)^{1/\nu} \right\} - \frac{1}{z_2} T_{\nu+1} \left\{ -\frac{\rho(h)}{b} + \frac{1}{b} \left(\frac{z_1}{z_2}\right)^{1/\nu} \right\} \right], \quad (3.26)$$

dove T_ν è la funzione di distribuzione cumulativa di una variabile aleatoria di Student con ν gradi di libertà e $b^2 = [1 - \rho(h)^2]/(\nu + 1)$.

3.3 Dipendenza spaziale

La dipendenza spaziale nei processi max-stable è fondamentale per modellare eventi estremi che si verificano su aree geografiche estese. A tal fine, sono state proposte diverse misure per descrivere e misurare la dipendenza spaziale per coppie di località. Ogni modello max-stable presentato in 3.2 cattura la struttura di dipendenza tra massimi registrati in diverse località in modo diverso, permettendo stime più accurate del rischio. Ne consegue che le diverse misure di dipendenza spaziale assumono una forma specifica per ciascun modello. Le misure tipicamente utilizzate nelle applicazioni sono la funzione del coefficiente estremo e la funzione di probabilità di concomitanza estrema.

3.3.1 Funzione del coefficiente estremo

Un modo per misurare la dipendenza spaziale del processo stocastico $\{Z(s): s \in S\}$ è attraverso il coefficiente estremo (*extremal coefficient*). Esso è definito tramite la funzione esponenziale quando (3.13) viene valutata per lo stesso valore di z in k località, ovvero:

$$\begin{aligned} \mathbb{P}(Z(s_1) \leq z, \dots, Z(s_k) \leq z) &= \exp[-V(z, \dots, z)] = \exp\left[-\frac{V(1, \dots, 1)}{z}\right] = \\ &= \exp\left(-\frac{\theta_k}{z}\right), \end{aligned} \quad (3.27)$$

dove θ_k è il coefficiente estremo k -variato. Esso assume valori in $[1, k]$, con estremi inferiore e superiore corrispondenti ai casi di completa dipendenza e indipendenza rispettivamente, e può essere interpretato come il numero effettivo di stazioni indipendenti. Anche se non caratterizza completamente la struttura di dipendenza, il coefficiente estremo è indipendente dal valore di z .

Partendo dalla definizione del coefficiente estremo, viene definita, da Schlather e Tawn nel 2003, come misura della dipendenza spaziale di un processo max-stable, la funzione del coefficiente estremo per una coppia di località $s_1, s_2 \in S$ a distanza $h = s_1 - s_2$:

$$\theta(h) = -z \log[\mathbb{P}(Z(s_1) \leq z, Z(s_2) \leq z)] = \mathbb{E}[\max\{Y(s_1), Y(s_2)\}] \in [1, 2] \quad (3.28)$$

Se si considerano solo i due casi di completa dipendenza e indipendenza si ha:

$$\begin{aligned} \mathbb{P}(Z(s+h) \leq z | Z(s) \leq z) &= \mathbb{P}(Z(s+h) \leq z)^{\theta(h)-1} = \\ &= \begin{cases} 1, & \theta(h) = 1, \\ \mathbb{P}(Z(s+h) \leq z), & \theta(h) = 2. \end{cases} \end{aligned} \quad (3.29)$$

La funzione $\theta(\cdot)$ non è differenziabile nell'origine, a meno che $\theta = 1$, ed è tale per cui $2 - \theta(h)$ è una funzione semi-definita positiva, ovvero una funzione di correlazione. Inoltre, se $d \geq 2$ e il processo $Z(\cdot)$ è isotropo, allora $\theta(\cdot)$ ha al massimo un salto nell'origine ed è continua altrove. Ne consegue che la struttura di dipendenza spaziale può essere caratterizzata da una funzione di correlazione. Tuttavia, una funzione di correlazione valida non sempre porta a una funzione del coefficiente estremo valida. In particolare, si ottiene una funzione del coefficiente estremo valida se $1 - 2(\theta(h) - 1)^2$ è definita positiva. Inoltre, valgono le seguenti disuguaglianze [21]:

1. $\theta(h+l) \leq \theta(h)\theta(l)$,
 2. $\theta(h+l)^\tau \leq \theta(h)^\tau + \theta(l)^\tau - 1, \quad \forall 0 \leq \tau \leq 1$,
 3. $\theta(h+l)^\tau \geq \theta(h)^\tau + \theta(l)^\tau - 1, \quad \forall \tau \leq 0$.
- (3.30)

3.3.1.1 *F-madogramma e funzione del coefficiente estremo*

Un altro modo per esprimere la funzione del coefficiente estremo è mediante il F-madogramma, definito da [21] per un processo max-stable stazionario $\{Z(s): s \in S\}$ con margini Fréchet unitari come:

$$v_F(h) = \frac{1}{2} \mathbb{E}[|F(Z(s+h)) - F(Z(s))|], \quad (3.31)$$

dove F rappresenta la funzione di distribuzione cumulativa di $\{Z(s): s \in S\}$ tale per cui si assume $F(Z(s)) \sim U(0,1)$. Di conseguenza, $\mathbb{E}[F(Z(s))] = 1/2$. Considerando la relazione $|a - b| = 2\max(a, b) - a - b$, (3.31) si può riscrivere come:

$$\begin{aligned} v_F(h) &= \mathbb{E}[\max\{F(Z(s+h)), F(Z(s))\}] - \mathbb{E}[F(Z(s))] = \\ &= \mathbb{E}[F(\max\{Z(s+h), Z(s)\})] - \frac{1}{2}. \end{aligned} \quad (3.32)$$

Da (3.27) si ha:

$$F(\max\{Z(s+h), Z(s)\}) = \mathbb{P}(Z(s+h) \leq z, Z(s) \leq z) = \exp\left(-\frac{\theta(h)}{z}\right). \quad (3.33)$$

Di conseguenza,

$$\begin{aligned} \mathbb{E}[F(\max\{Z(s+h), Z(s)\})] &= \int_0^\infty \exp\left(-\frac{1}{z}\right) \frac{\theta(h)}{z^2} \exp\left(-\frac{\theta(h)}{z}\right) dz = \\ &= \frac{\theta(h)}{\theta(h) + 1}. \end{aligned} \quad (3.34)$$

Combinando (3.32) e (3.34) si ottiene la funzione del coefficiente estremo espressa in termini del F-madogramma:

$$\theta(h) = \frac{1 + 2v_F(h)}{1 - 2v_F(h)}. \quad (3.35)$$

3.3.1.2 *Funzioni del coefficiente estremo per i modelli max-stable*

È possibile ottenere le funzioni del coefficiente estremo per ciascun processo max-stable presentato in 3.2. In particolare, per il processo di Smith si ha:

$$\theta(h) = 2\Phi\left(\frac{\alpha(h)}{2}\right), \quad (3.36)$$

ottenendo completa dipendenza $\theta(h) = 1$ quando $h \rightarrow 0$ e indipendenza $\theta(h) = 2$ quando $h \rightarrow \infty$.

Per il processo di Schlather, invece, la funzione del coefficiente estremo è:

$$\theta(h) = 1 + \sqrt{\frac{1 - \rho(h)}{2}}. \quad (3.37)$$

Nel caso particolare in cui il processo è definito in \mathbb{R}^2 e le Y_i sono stazionarie e isotropiche, si ottiene che $\theta(h) < 1.838$. Di conseguenza, questo modello non può tener conto degli estremi che diventano indipendenti quando la distanza h aumenta indefinitamente.

Per il processo di Brown-Resnick, la funzione del coefficiente estremo è data da:

$$\theta(h) = 2\Phi\left(\sqrt{\frac{\gamma(h)}{2}}\right), \quad (3.38)$$

e si ottiene completa dipendenza quando $\gamma(h) \rightarrow 0$, mentre se il semi-variogramma non è limitato si ha $\gamma(h) \propto h^\alpha$, si ottiene indipendenza quando $h \rightarrow \infty$.

Infine, il processo extremal-t ha funzione del coefficiente estremo:

$$\theta(h) = 2T_{\nu+1}\left(\sqrt{\frac{\nu+1}{1-\rho(h)^2}} - \sqrt{\frac{1-\rho(h)^2}{\nu+1}}\rho(h)\right), \quad (3.39)$$

il cui estremo superiore, quando $h \rightarrow \infty$, è:

$$\theta(h) = 2T_{\nu+1}(\sqrt{\nu+1}), \quad (3.40)$$

per cui si ottiene indipendenza.

3.3.2 Funzione di probabilità di concomitanza estrema

A differenza delle metodologie già sviluppate che si concentrano sull'intensità degli estremi, [22] ha proposto un'altra misura di dipendenza spaziale che tiene conto dei tempi di occorrenza degli estremi in diverse località. Nello specifico, la probabilità di concomitanza estrema (*extremal concurrence probability*) è definita come la probabilità che gli estremi si verifichino simultaneamente in diverse posizioni spaziali e si ottiene in maniera asintotica dalla probabilità di concomitanza campionaria (*sample concurrence probability*). Quest'ultima si definisce come:

$$\begin{aligned} p_n(s_1, \dots, s_k) &= \mathbb{P}(\text{per qualche } i_0 \in \{1, \dots, n\}: Z_n(s_j) = X_{i_0}(s_j), \forall j = 1, \dots, k) = \\ &= n \mathbb{E} \left[F(X(s_1), \dots, X(s_k))^{n-1} \right], \end{aligned} \quad (3.41)$$

dove $Z_n(s) = \max_{i=1, \dots, n} X_i(s)$, $s \in S$ e F è la distribuzione cumulativa multivariata. Ne consegue che la probabilità di concomitanza campionaria non dipende dalla distribuzione marginale di X , ma solo dalla struttura di dipendenza e dalla dimensione del campione n . Inoltre, la definizione (3.41) equivale a dire che solo un'osservazione $\{X_{i_0}(s), s \in S\}$ genera il massimo puntuale in ogni località s_1, \dots, s_k , ovvero un singolo evento estremo che colpisce tutti i siti è responsabile dei massimi storici. Di conseguenza, le probabilità di concomitanza possono aiutare a comprendere e quantificare l'estensione areale degli eventi estremi.

Considerando la rappresentazione (3.9), la probabilità di concomitanza estrema si definisce come:

$$p(s_1, \dots, s_k) = \mathbb{P}(\text{per qualche } l \geq 1: Z(s_j) = \varphi_l(s_j), j = 1, \dots, k), \quad (3.42)$$

ovvero i valori del processo stocastico $\{Z(s): s \in S\}$ nelle località s_1, \dots, s_k sono ottenuti dalla stessa funzione spettrale φ_l . Come previsto, $p(s_1, \dots, s_k) \in [0, 1]$, dove gli estremi inferiore e superiore corrispondono rispettivamente a indipendenza e completa dipendenza del processo stocastico nelle k località.

È possibile esprimere tale probabilità in funzione di un *hitting scenario*, che rappresenta come diversi eventi contribuiscono al massimo componente per componente di un processo stocastico. Nello

specifico, si definisce *hitting scenario estremo* π come una partizione di $\{1, \dots, k\}$ tale che due indici $j_1, j_2 \in \{1, \dots, k\}$ appartengono alla stessa componente di π se e solo se

$$\arg \max_{i \geq 1} \varphi_i(s_{j_1}) = \arg \max_{i \geq 1} \varphi_i(s_{j_2}), \quad \varphi_i \in \Phi. \quad (3.43)$$

Si può dimostrare che esso si ottiene in modo asintotico dal *hitting scenario campionario*, definito come una partizione di $\{1, \dots, k\}$:

$$\pi_n = \{C_i: C_i \neq \emptyset\}, \quad \text{dove } C_i = \{j: Z_n(s_j) = X_i(s_j)\}, i = 1, \dots, n, \quad (3.44)$$

ovvero suddivide le località in base a quando esse contribuiscono al massimo. Sia nel caso di concomitanza campionario sia nel caso di concomitanza estrema, gli estremi sono concorrenti se e solo se π_n o π corrispondono all'intero insieme $\{1, \dots, k\}$. Vale il seguente risultato asintotico per la probabilità di concomitanza [22].

Teorema 3.4 *Se il processo stocastico X è nel dominio di attrazione di qualche processo max-stabile Z , la probabilità di concomitanza campionario converge alla sua controparte estrema*

$$p_n(s_1, \dots, s_k) = \mathbb{P}(\pi_n = \{1, \dots, k\}) \rightarrow \mathbb{P}(\pi = \{1, \dots, k\}) = p(s_1, \dots, s_k), \quad n \rightarrow \infty. \quad (3.45)$$

La probabilità di concomitanza estrema può essere espressa come:

$$p(s_1, \dots, s_k) = \mathbb{E}_Y \left[\left(V(Y(s_1), \dots, Y(s_k)) \right)^{-1} \right]. \quad (3.46)$$

Quando si considera una coppia di località (s_1, s_2) , si definisce la funzione di probabilità di concomitanza estrema come la misura di Kendall τ di dipendenza del vettore max-stabile $\{Z(s_1), Z(s_2)\}$:

$$p(h) = \mathbb{E}[\text{sign}\{Z(s_1) - Z_*(s_1)\} \text{sign}\{Z(s_2) - Z_*(s_2)\}], \quad (3.47)$$

dove $\{Z_*(s): s \in S\}$ è una copia indipendente di $\{Z(s): s \in S\}$. Si può dimostrare che esiste una relazione che lega la funzione probabilità di concomitanza estrema al coefficiente estremo:

$$p(s_1 - s_2) = 2 - \theta_\xi(s_1 - s_2), \quad (3.48)$$

dove θ_ξ è definito come in (3.28), ma con $z = 1$. La funzione in (3.47), dunque, indica quanto è probabile che coppie di punti nello spazio a distanza h siano coinvolti in eventi estremi contemporaneamente, fornendo informazioni sull'estensione areale dell'evento estremo d'interesse che si verifica in una posizione $s \in S$. Essa è semi-definita positiva e non è differenziabile nell'origine, a meno che $p(h) = 1, \forall h \in S$. Se $d \geq 1$ e il processo $Z(\cdot)$ è isotropo, allora la funzione di probabilità di concomitanza estrema ha al massimo un salto nell'origine ed è continua altrove. Inoltre, valgono le seguenti disuguaglianze per ogni $h, l \in S$:

1. $2 - p(h + l) \leq (2 - p(h))(2 - p(l))$,
 2. $(2 - p(h + l))^\alpha \leq (2 - p(h))^\alpha + (2 - p(l))^\alpha - 1, \quad 0 \leq \alpha \leq 1$,
 3. $(2 - p(h + l))^\alpha \geq (2 - p(h))^\alpha + (2 - p(l))^\alpha - 1, \quad \alpha < 0$.
- (3.49)

Invece di considerare coppie di punti, è possibile considerare una cella di concomitanza per una località $q \in S$, definita come un insieme casuale di siti che hanno una relazione di concomitanza con s :

$$C(s) = \{q \in S: \text{estremi sono concomitanti in } s \text{ e } q\}. \quad (3.50)$$

Il valore atteso del volume d-dimensionale di una cella di concomitanza vale:

$$\mathbb{E}[|C(s)|] = \int_{q \in S} p(s, q) dq, \quad (3.51)$$

e misura quanto rapidamente diminuisce la dipendenza negli estremi allontanandosi da s .

3.3.2.1 Funzioni di probabilità di concomitanza estrema per i modelli max-stable

Non è sempre possibile ottenere delle formule esplicite per la funzione di probabilità di concomitanza estrema espressa come (3.46), ma è possibile stimarla campionando copie indipendenti di Y e calcolando la media campionaria. Nel caso dei modelli max-stable presentati in 3.2, la funzione di concomitanza estrema per coppie di punti assume specifiche forme.

Per il processo di Brown-Resnick, in particolare, la funzione di probabilità di concomitanza estrema bivariata è:

$$p(h) = \mathbb{E} \left[\left(\Phi(Z) + \exp(\gamma(h) - \sqrt{2\gamma(h)}Z) \Phi(\sqrt{2\gamma(h)} - Z) \right)^{-1} \right], \quad (3.52)$$

dove $Z \sim N(0,1)$ con funzione di distribuzione cumulativa Φ . Nel punto di riferimento o la probabilità è chiaramente 1, mentre si ottiene indipendenza $p(h) \rightarrow 0$ per $\|h\| \rightarrow \infty$ quando il semi-variogramma è non limitato, ovvero $\gamma(h) \rightarrow \infty$ per $\|h\| \rightarrow \infty$.

Per il processo extremal-t e il caso particolare del processo di Schlather, invece, la funzione di probabilità di concomitanza estrema bivariata è:

$$p(h) = \mathbb{E} \left[\left(T_{\nu+1}(T) + (\rho(h) + \sigma(h)T)^{-\nu} T_{\nu+1} \left(-\frac{\rho(h)}{\sigma(h)} + \frac{1}{\sigma(h)(\rho(h) + \sigma(h)T)} \right) \right)^{-1} \mathbf{1}_{\{\rho(h) + \sigma(h)T > 0\}} \right], \quad (3.53)$$

dove $\sigma(h) = \sqrt{[1 - \rho(h)^2]/(1 + \nu)}$ e T è una variabile aleatoria di Student con $\nu + 1$ gradi di libertà e funzione di distribuzione cumulativa $T_{\nu+1}$.

3.4 Simulazione di un processo max-stable

In vari contesi applicativi, si è interessati a valutare alcune caratteristiche più avanzate del processo stocastico che descrive gli eventi estremi oppure si vuole fare previsione del fenomeno in nuove località. In tali casi, si ricorre alla simulazione stocastica. La simulazione esatta dei processi max-stable può sembrare impraticabile, poiché, secondo la **Definizione 3.1**, richiederebbe di determinare i massimi in ogni punto per un numero infinito di funzioni casuali, mentre a livello pratico il numero di realizzazioni è necessariamente finito. È possibile aggirare il problema e ottenere simulazioni esatte del processo max-stable $\{Z(s): s \in S\}$ su una finestra di campionamento finita $B \subset \mathbb{R}^d$, sotto alcune condizioni, considerando una delle rappresentazioni spettrali presentate in 3.1.1 [18]. Inoltre, se si considera la decomposizione delle funzioni spettrali Φ introdotte nella rappresentazione (3.9) in funzioni estreme e sub-estreme, la simulazione del processo max-stable Z in una nuova località s corrisponde a simulare in modo esatto dalla funzione estrema $\varphi \in \Phi^+$ corrispondente [24].

Ne consegue che, considerando ad esempio la caratterizzazione (3.8), per simulare il processo max-stable $\{Z(s): s \in S\}$, bisogna simulare i punti ζ_i del processo di Poisson e copie indipendenti del processo stocastico non negativo $\{Y(s): s \in S\}$, che solitamente è un processo Gaussiano, a parte una trasformazione. Dal momento che la misura d'intensità $d\Lambda(\zeta) = \zeta^{-2}d\zeta$ diverge per ogni insieme della forma $(0, a)$, con $a > 0$, ne consegue che 0 è un punto di accumulazione, ovvero infiniti punti ζ_i sono vicini a 0 e dunque trascurabili per ottenere una buona approssimazione del processo $\{Z(s): s \in S\}$ [11]. Risulta, dunque, più efficiente supporre che i punti $\zeta_i: i \geq 1$ siano decrescenti, ovvero $\{\zeta_i \geq \zeta_{i+1}, \forall i \geq 1\}$, in modo tale da approssimare $\{Z(s): s \in S\}$ con il massimo sui primi N punti più significativi, con N numero finito dato da un qualche criterio di arresto [23]. Per quanto riguarda la simulazione dei punti ζ_i , è possibile sfruttare il fatto che $\{\zeta_i^{-1}, i \geq 1\}$ è un processo di Poisson standard su $(0, \infty)$, ovvero $\zeta_1^{-1}, \zeta_2^{-1} - \zeta_1^{-1}, \zeta_3^{-1} - \zeta_2^{-1}, \dots$ sono indipendenti e distribuiti come un'esponenziale standard. Di conseguenza, si ottiene un'uguaglianza in distribuzione tra il processo $\{Z(s): s \in S\}$ e il processo

$$\max_{i \geq 1} \frac{Y_i(s)}{\sum_{j=1}^i E_j}, \quad s \in S, \quad (3.54)$$

dove $E_j \sim \text{Exp}(1)$. Per quanto riguarda il processo $\{Y(s): s \in S\}$, si suppone che esso sia limitato uniformemente da una costante C positiva, ovvero $\mathbb{P}(Y(s) \leq C) = 1$ con $s \in S$, e che abbia supporto in una palla $b(o, r)$ per qualche $r \in (0, \infty)$. Sotto queste ipotesi vale il seguente risultato dimostrato in [18] che permette di ottenere delle simulazioni esatte del processo max-stable $\{Z(s): s \in S\}$.

Teorema 3.5 *Siano Y, ζ_i e Z come nel Teorema 3.3. Assumiamo che Y sia limitato uniformemente da $C \in (0, \infty)$ e abbia supporto in una palla $b(o, r)$ per qualche $r \in (0, \infty)$. Siano B un insieme compatto, Y_i copie i.i.d. di Y , U_i i.i.d. con distribuzione uniforme su $B_r = \cup_{s \in B} b(s, r)$, E_i variabili i.i.d. con distribuzione esponenziale standard e tutti indipendenti tra loro. Allora, su B , il processo Z ha quasi sicuramente la stessa distribuzione di*

$$|B_r| \sup \left\{ \frac{Y_i(s - U_i)}{\sum_{j=1}^i E_j}, i = 1, \dots, N, \text{ e } N \text{ tale che } \frac{C}{\sum_{j=1}^N E_j} \leq \max_{1 \leq i \leq N} \frac{Y_i(s - U_i)}{\sum_{j=1}^i E_j} \right\}. \quad (3.55)$$

Pertanto, il processo Z può essere simulato esattamente in tempo finito mediante l'**Algoritmo 3.1** [14].

L'efficienza dell'**Algoritmo 3.1** è misurata in base al numero previsto di funzioni spettrali che devono essere simulate prima che sia soddisfatto il criterio di arresto, e, in particolare, il numero di funzioni sub-estreme considerate dovrebbe essere minimo. Minimizzare il numero di funzioni spettrali necessarie per soddisfare il criterio di arresto è fondamentale per ridurre il tempo computazionale richiesto per ottenere risultati accurati. Nello specifico, si può dimostrare che il numero atteso di funzioni considerate per l'algoritmo precedente è [23]:

$$\mathbb{E}[N] = C \mathbb{E}[(\inf_{x \in \mathbb{X}} Z(s))^{-1}] < \infty. \quad (3.56)$$

Se le ipotesi del **Teorema 3.5** non sono soddisfatte, ovvero il supporto di Y non è incluso in una palla $b(o, r)$ oppure non è limitato uniformemente da una costante C , è possibile utilizzare delle approssimazioni di r e C e ottenere delle simulazioni accurate, anche se non esatte. Ad esempio, si può considerare un limite pseudo uniforme $C^* < \infty$ sufficientemente grande tale che $\mathbb{P}(Y(s) > C^*) = \varepsilon$, per $\varepsilon > 0$ sufficientemente piccolo. La scelta di C^* è un compromesso tra l'accuratezza della simulazione e il tempo computazionale, in quanto un valore troppo grande comporta un numero di iterazioni maggiori per soddisfare il criterio di arresto. [18] ha proposto $C^* = 3$ per il processo di Schlather e r tale che $\varphi(r) = 0,001$ per la densità normale standard φ per il processo di Smith. Per valutare la qualità del

Algoritmo 3.1: Simulazione di un processo max-stable semplice.

Input: limite superiore $C > 0$

Output: una realizzazione per processo max-stable

```

1   $i = 1, flag = true, W = 0, Z \equiv 0;$ 
2  while  $flag$  is true do
3     $W = W + E, E \sim Exp(1);$ 
4     $\zeta_i = 1/W;$ 
5     $Y_i \sim Y;$ 
6     $Z = \max(Z, \zeta_i Y_i);$ 
7    if  $\zeta_i C < \inf(Z)$  then
8       $flag = false;$ 
9    else
10      $flag = true;$ 
11   end if
12    $i = i + 1;$ 
13 end while
14 Return  $Z.$ 

```

procedimento approssimativo dell'**Algoritmo 3.1** quando vengono considerate delle approssimazioni di r e C , si misura la probabilità di ottenere realizzazioni incorrette. In particolare, si considera una realizzazione prodotta dall'algoritmo incorretta se esiste almeno una funzione estremale che non viene presa in considerazione dall'algoritmo e, per ogni $C^* > 0$, la probabilità di ottenere realizzazioni incorrette è [23]:

$$\begin{aligned}
& \mathbb{P}\left(\exists(\zeta_i, Y_i) \in \Phi^+ : \zeta_i C^* < \inf_{s \in S} \max_{j=1, \dots, i-1} \zeta_j Y_j(s)\right) \leq \\
& \leq \sqrt{\mathbb{E}\left(\sup_{s \in S} \left(\frac{Y(s)}{Z(s)}\right)^2\right)} \cdot \sqrt{\mathbb{P}(\sup_{s \in S} Y(s) > C^*)}.
\end{aligned} \tag{3.57}$$

Tuttavia, quando il processo $\{Y(s) : s \in S\}$ non è stazionario, ma il processo max-stable risultante lo è, come nel caso del processo di Brown-Resnick, l'**Algoritmo 3.1** fallisce in quanto si ottengono delle realizzazioni non stazionarie anche per valori elevati di C . Un'alternativa semplice ed efficiente consiste nell'utilizzare traslazioni casuali per mitigare l'impatto della non stazionarietà di Y , ovvero ottenere campioni da

$$\tilde{Z}(s) = \max_{i \geq 1} \zeta_i Y_i(s - U_i), \quad s \in S, \tag{3.58}$$

dove U_i sono indipendenti e appartenenti a una distribuzione F definita almeno su S . Il processo \tilde{Z} ha la stessa distribuzione di Z . Infatti,

$$\begin{aligned}
\mathbb{P}(\tilde{Z}(s_1) \leq z_1, \dots, \tilde{Z}(s_k) \leq z_k) &= \exp\left[-\int \mathbb{E}\left(\max_{j=1, \dots, k} \frac{Y(s_j - u)}{z_j}\right) dF(u)\right] = \\
&= \mathbb{P}(Z(s_1 - u) \leq z_1, \dots, Z(s_k - u) \leq z_k) = \\
&= \mathbb{P}(Z(s_1) \leq z_1, \dots, Z(s_k) \leq z_k),
\end{aligned} \tag{3.59}$$

dal momento che Z è stazionario [14].

Sono stati proposti altri algoritmi derivanti da rappresentazioni differenti, in particolar modo per risolvere il problema della non-stazionarietà del processo $\{Y(s): s \in S\}$ nel modello di Brown-Resnick. La scelta di quale caratterizzazione considerare per la simulazione di un processo di Brown-Resnick dipende sia dal semi-variogramma corrispondente sia dal dominio di simulazione [23].

Un altro aspetto che è stato preso in considerazione riguarda la simulazione del processo max-stable $\{Z(s): s \in S\}$ condizionatamente ai valori che esso assume in un numero limitato di località, ovvero

$$Z(s) | \{Z(s_1) = z_1, \dots, Z(s_k) = z_k\}, \quad s_1, \dots, s_k \in S. \quad (3.60)$$

Dombry, Éyi-Minko e Ribatet nel 2013 hanno proposto un algoritmo a 3 step che permette di ottenere campioni da (3.60) che consiste nel sorteggiare un *hitting scenario*, solitamente usando un campionatore di Gibbs, utilizzando le osservazioni del processo max-stable. Successivamente, vengono simulate le funzioni estreme dal *hitting scenario* e si definisce un processo prendendo il massimo rispetto alle funzioni estreme. Infine, indipendentemente dagli step precedenti, vengono simulate le funzioni sub-estreme e si definisce un processo max-stable “assottigliato” dove le realizzazioni sono filtrate in modo da essere inferiori ai valori osservati. Il processo condizionato (3.60) si ottiene prendendo il massimo tra i processi definiti nelle ultime due fasi ([11], [14]).

Capitolo 4

Inferenza e valutazione di un modello spaziale

Nel Capitolo 3 sono stati presentati i modelli max-stable per l'analisi degli estremi spaziali e le misure di dipendenza spaziale. Tuttavia, per applicare efficacemente questi modelli a dataset reali, è essenziale adottare una metodologia robusta e coerente che permetta di effettuare inferenze accurate sui parametri del modello. Questo capitolo è dedicato alla descrizione dettagliata della metodologia impiegata per l'inferenza dei parametri nei modelli max-stable e alla discussione delle tecniche pratiche per l'applicazione di questi modelli ai dati reali. Verranno analizzate anche le tecniche per la scelta del modello più appropriato e per la valutazione della sua adeguatezza, nonché gli strumenti grafici utili per visualizzare e interpretare i risultati.

4.1 Distribuzione marginale GEV arbitraria

I modelli presentati nel Capitolo 3 sono stati sviluppati sotto l'ipotesi che la distribuzione marginale fosse Fréchet unitaria. Tuttavia, nelle situazioni concrete questa assunzione non è realistica, in quanto supporre che i parametri dei modelli di valori estremi siano costanti su un'intera area geografica non riflette la realtà. Di conseguenza, è necessario tenere conto della distribuzione marginale nello spazio, ovvero permettere ai parametri generalizzati dei valori estremi μ , σ e ξ di variare nello spazio, rendendo il modello più accurato e rappresentativo della realtà. In tal caso, l'equazione (2.22) descrive le *trend surfaces* per i parametri della distribuzione marginale GEV, in cui possibili covariate sono longitudine, latitudine e altitudine dei punti dell'area geografica.

Attraverso delle trasformazioni marginali è possibile passare da una distribuzione Fréchet unitaria a una distribuzione GEV arbitraria e viceversa per un singolo punto dello spazio. Nello specifico, sia Y una variabile aleatoria con distribuzione GEV con parametri μ , σ e ξ , e sia Z una variabile aleatoria con distribuzione Fréchet unitaria. Allora valgono le seguenti trasformazioni:

$$Z = \begin{cases} \left[1 + \xi \frac{Y - \mu}{\sigma} \right]_+^{1/\xi}, & \xi \neq 0, \\ \exp\left(\frac{Y - \mu}{\sigma}\right), & \xi = 0, \end{cases} \quad (4.1)$$

$$Y = \begin{cases} \mu + \sigma \frac{Z_+^\xi - 1}{\xi}, & \xi \neq 0, \\ \mu + \sigma \log Z_+, & \xi = 0. \end{cases} \quad (4.2)$$

È possibile trasformare i dati a una distribuzione Fréchet unitaria anche utilizzando la funzione di distribuzione cumulativa empirica, ottenuta assegnando un rango a ciascun elemento e dividendolo per il numero totale di elementi più uno, tramite la seguente equazione:

$$Z = -\frac{1}{\log(F_n(y))}. \quad (4.3)$$

Per quanto riguarda i modelli max-stable, da un punto di vista statistico, esistono due strategie per tenere conto dei parametri della distribuzione marginale GEV ([11], [22]).

- La prima strategia consiste nello stimare i parametri marginali rispetto a delle covariate rilevanti, ignorando la dipendenza spaziale. Successivamente, utilizzare le *trend surfaces* stimate per trasformare i dati a margini Fréchet unitari tramite la trasformazione (4.1). In alternativa, i dati vengono trasformati in modo empirico tramite la trasformazione (4.3). Infine, fittare un processo max-stable semplice ai dati trasformati per stimare la dipendenza spaziale ([11], [20], [26], [27]).
- La seconda strategia consiste nel fittare un processo max-stable direttamente ai dati, stimando contemporaneamente i parametri marginali e la dipendenza spaziale ([11], [12], [14], [25], [28]).

Il primo approccio ha diversi svantaggi. In primo luogo, poiché i parametri GEV vengono stimati separatamente dalla struttura di dipendenza spaziale, non si tiene conto della possibile interazione tra la stima dei parametri marginali e la struttura di dipendenza. In secondo luogo, gli errori standard delle stime dei parametri di dipendenza spaziale, che descrivono come le variabili estreme sono correlate nello spazio, sono basati sull'ipotesi che i dati trasformati siano margini di Fréchet unitari esatti, ovvero senza errori. Tuttavia, poiché questa trasformazione è basata su stime dei parametri GEV che hanno un'incertezza associata (errori standard), non si tiene conto di questa incertezza aggiuntiva quando si fitta la struttura di dipendenza spaziale. Di conseguenza, gli errori standard relativi ai parametri di dipendenza spaziale risultano sottostimati in quanto si ignora l'incertezza derivante dalla stima preliminare dei parametri GEV. In altre parole, gli errori standard dei parametri di dipendenza spaziale non riflettono pienamente l'incertezza complessiva del modello, in quanto non si tiene conto dell'incertezza introdotta nella prima fase della trasformazione dei dati, portando a una valutazione troppo ottimistica della precisione delle stime della struttura di dipendenza [42].

Il secondo approccio, invece, sfrutta tutte le informazioni disponibili nei dati contemporaneamente, potenzialmente migliorando l'accuratezza delle stime in quanto evita la propagazione degli errori tra la stima dei margini e la stima della dipendenza spaziale. Tuttavia, la funzione di verosimiglianza diventa più complessa e, di conseguenza, più difficile da ottimizzare, portando a stime meno precise e a superfici di verosimiglianza irregolari. Inoltre, se le *trend surfaces* sono scarse, cioè se i dati non forniscono una buona rappresentazione delle variazioni spaziali, la stima simultanea può introdurre bias, portando a stime di dipendenza spaziale distorte [11].

Da questo punto in avanti, si considererà esclusivamente la seconda strategia, ovvero la stima simultanea dei parametri GEV e della struttura di dipendenza spaziale. Spesso, il modello max-stable viene fittato utilizzando delle stime iniziali dei parametri. In tal caso, le stime iniziali dei parametri marginali corrispondono alle stime ottenute nella prima fase della prima strategia.

4.1.1 Stima dei parametri di un modello GEV spaziale

Un modello GEV spaziale è definito tramite le *trend surfaces* per i parametri μ , σ e ξ :

$$\mu = X_{\mu}\beta_{\mu}, \quad \sigma = X_{\sigma}\beta_{\sigma}, \quad \xi = X_{\xi}\beta_{\xi}, \quad (4.4)$$

dove X_{θ} è la matrice delle covariate, ovvero le coordinate spaziali per ogni località, e β_{θ} è il vettore dei parametri che deve essere stimato. Esistono diverse metodologie per l'inferenza dei parametri, tra cui uno dei più utilizzati è il metodo della massima verosimiglianza (MLE). La funzione di log-verosimiglianza per un modello GEV spaziale è definita come:

$$l(\boldsymbol{\beta}) = \sum_{i=1}^n \sum_{j=1}^m \log f(y_{i,j}; \boldsymbol{\theta}_i(\boldsymbol{\beta}_{\theta_i})), \quad (4.5)$$

dove n è il numero di località, m è il numero di osservazioni e $\boldsymbol{\theta}_i$ è il vettore dei parametri della distribuzione GEV per la località i . Con questa formulazione, si suppone che le stazioni siano reciprocamente indipendenti, in quanto si somma la funzione di log-verosimiglianza derivante dalla densità GEV univariata (2.9) per ciascuna località. Per ogni località $i = 1, \dots, n$, la funzione di log-verosimiglianza univariata è ([2], Capitolo 3, p. 55):

$$\begin{cases} \sum_{j=1}^m -\log(\sigma_i) - \left(1 + \xi_i \frac{y_{i,j} - \mu_i}{\sigma_i}\right)^{-1/\xi_i} - \left(1 + \frac{1}{\xi_i}\right) \log\left(1 + \xi_i \frac{y_{i,j} - \mu_i}{\sigma_i}\right), & \xi_i \neq 0, \\ \sum_{j=1}^m -\log(\sigma_i) - \frac{y_{i,j} - \mu_i}{\sigma_i} - \exp\left(-\frac{y_{i,j} - \mu_i}{\sigma_i}\right), & \xi_i = 0. \end{cases} \quad (4.6)$$

La funzione di log-verosimiglianza ottenuta combinando (4.5) e (4.6), viene ottimizzata per ottenere le stime dei parametri utilizzando, ad esempio, un algoritmo di tipo Newton. Di conseguenza, stime affidabili dell'errore standard per ciascun parametro $\hat{\beta}_k$ di (4.4) vengono ottenute con le loro stime a “sandwich”:

$$\text{std. err}(\hat{\beta}_k) = \sqrt{[\hat{\mathbf{H}}^{-1} \hat{\mathbf{K}} \hat{\mathbf{H}}^{-1}]_{kk}}, \quad (4.7)$$

dove $\hat{\mathbf{H}}$ è la matrice hessiana della funzione di verosimiglianza e $\hat{\mathbf{K}}$ è la matrice di varianza empirica, ovvero una stima della varianza dei gradienti per ciascun parametro, e $\hat{\mathbf{V}} = \hat{\mathbf{H}}^{-1} \hat{\mathbf{K}} \hat{\mathbf{H}}^{-1}$ denota la matrice di varianza e covarianza dei parametri $\hat{\boldsymbol{\beta}}$.

4.1.2 Stima dei parametri di un processo max-stable

Quando si vuole fittare un processo max-stable a dei dati reali, poiché il numero di località è finito, l'analisi rientra nella categoria dei modelli multivariati. Tuttavia, i metodi di verosimiglianza standard per questi modelli sono impraticabili per due motivi. In primo luogo, molti processi max-stable non hanno una distribuzione in forma chiusa in più di due dimensioni. In secondo luogo, anche quando tale distribuzione esiste in forma chiusa, la funzione di verosimiglianza risulta estremamente complessa, poiché è definita con una somma su tutte le partizioni delle località. Il numero di queste partizioni corrisponde al numero di Bell [29], che produce un'esplosione combinatoria anche per un numero moderato di località. Di conseguenza, risulta computazionalmente infattibile ottenere una stima dei parametri mediante il metodo di massima verosimiglianza.

Un metodo alternativo consiste nel considerare la funzione di verosimiglianza composita (MCLE). Esso permette di stimare i parametri marginali GEV congiuntamente ai parametri di dipendenza in un quadro unificato, producendo risultati flessibili e affidabili a un costo computazionale moderato. Dal momento che sono note le distribuzioni bivariate dei processi max-stable, è possibile definire una funzione di log-verosimiglianza composita a coppie (*pairwise composite log-likelihood*). Essa è definita come la somma delle funzioni di log-verosimiglianza relative alle funzioni di densità bivariate per ogni coppia di località [25]:

$$l_P(\boldsymbol{\psi}) = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \sum_{k=1}^m w_{ij} \log f(y_{i,m}, y_{j,m}; \boldsymbol{\psi}), \quad (4.8)$$

dove n è il numero di località, $n(n-1)/2$ è il numero totale di coppie, m è il numero di osservazioni, w_{ij} il peso per la coppia $(\mathbf{y}_i, \mathbf{y}_j)$ e $\boldsymbol{\psi}$ il vettore dei parametri che deve essere stimato. Se tutti i pesi $w_{ij} > 0$, tali che $\sum_{ij} w_{ij} = 1$, tutte le coppie di località sono incluse nel calcolo della funzione di log-verosimiglianza a coppie (4.8). Tuttavia, i requisiti computazionali e l'efficienza statistica possono essere migliorati, ad esempio, considerando solo coppie di osservazioni vicine. Dal momento che le funzioni di densità bivariate sono note per processi max-stable semplici, è possibile includere l'informazione dei parametri marginali GEV considerando la trasformazione $(Y_i, Y_j) = g(Z_i, Z_j)$, con g^{-1} definita da (4.1). La funzione di densità bivariata risultante è data da:

$$f_{Y_i, Y_j}(y_i, y_j) = f_{Z_i, Z_j}(g^{-1}(y_i, y_j)) |J(y_i, y_j)|, \quad (4.9)$$

dove $f_{Z_i, Z_j}(z_i, z_j)$ è la funzione di densità bivariata di un processo max-stable semplice e J è la Jacobiana della trasformazione:

$$|J(y_i, y_j)| = \frac{1}{\sigma_i} \left(1 + \xi_i \frac{y_i - \mu_i}{\sigma_i}\right)_+^{1/\xi_i - 1} \cdot \frac{1}{\sigma_j} \left(1 + \xi_j \frac{y_j - \mu_j}{\sigma_j}\right)_+^{1/\xi_j - 1}. \quad (4.10)$$

La funzione di log-verosimiglianza ottenuta combinando (4.8), (4.9) e (4.10), viene ottimizzata per ottenere le stime dei parametri utilizzando, ad esempio, un algoritmo di tipo Newton. Poiché la verosimiglianza a coppie è un'approssimazione della “verosimiglianza completa”, gli errori standard non possono essere calcolati direttamente dall'inverso della matrice di informazione di Fisher. Invece, è necessario utilizzare una stima a sandwich per tenere conto di un'errata specificazione del modello, analogamente a (4.7) [42].

4.2 Selezione del modello

La scelta del modello più appropriato è un altro aspetto fondamentale nell'analisi degli estremi spaziali, in quanto consente di identificare il modello che meglio rappresenta il processo generativo dei dati tra una serie di modelli candidati. Un'accurata selezione del modello aiuta a migliorare le previsioni, a comprendere meglio le relazioni tra le variabili e a evitare problemi di sovra-adattamento o sotto-adattamento. La selezione del modello può avvenire adattando un criterio di informazione ([12], [14], [20], [25], [26], [27]), un test del rapporto di verosimiglianza per i modelli annidati ([14], [25]) o mediante la validazione del modello su un set di dati diverso da quello di training ([12], [26], [27], [28]).

4.2.1 TIC

Il criterio di informazione di Takeuchi (TIC) ha lo scopo di identificare il modello che fornisce il miglior compromesso tra bontà di adattamento e complessità del modello. Esso si basa sulla distanza di Kullback-Leibler, che misura la differenza di informazione tra una distribuzione di probabilità vera e una distribuzione di probabilità stimata da un modello. Il TIC penalizza i modelli più complessi per la loro maggiore flessibilità, che può portare a un sovra-adattamento ai dati di training e a una scarsa generalizzazione su nuovi dati. Nel caso di funzione di verosimiglianza composita, si seleziona il modello che minimizza ([14], [25]):

$$TIC(f_\psi) = -2l_p(\hat{\boldsymbol{\psi}}) + 2tr\left(\hat{\mathbf{K}}(\hat{\boldsymbol{\psi}})\hat{\mathbf{H}}(\hat{\boldsymbol{\psi}})^{-1}\right), \quad (4.11)$$

dove $\hat{\boldsymbol{\psi}}$ è lo stimatore di massima verosimiglianza a coppie, $\hat{\mathbf{K}}$ e $\hat{\mathbf{H}}$ sono le stime delle matrici di varianza empirica e hessiana, mentre il secondo termine corrisponde al termine di penalizzazione della log-verosimiglianza composita. In caso di una specificazione errata del modello, il TIC corrisponde all'AIC.

4.2.2 ANOVA

In caso di modelli annidati, è possibile utilizzare il test del rapporto di verosimiglianza per verificare se i dati supportano un'ipotesi specifica su un parametro del modello. Sia $\boldsymbol{\psi} = (\boldsymbol{\psi}', \boldsymbol{\psi}'')$. Nello specifico, si definisce un'ipotesi nulla H_0 corrispondente al modello più semplice, ad esempio, $\boldsymbol{\psi}' = \boldsymbol{\psi}_o (= 0)$, e un'ipotesi alternativa H_1 che rappresenta il modello con cui si vuole confrontarlo, ad esempio $\boldsymbol{\psi}' \neq \boldsymbol{\psi}_o$. La statistica del test è la funzione di devianza:

$$D(\boldsymbol{\psi}_o) = 2[l(\hat{\boldsymbol{\psi}}) - l((\boldsymbol{\psi}_o, \hat{\boldsymbol{\psi}}''))], \quad (4.12)$$

che, sotto l'ipotesi nulla, ha distribuzione asintotica $\sum_{i=1}^r \lambda_i X_i$, dove r è la dimensione del vettore dei parametri $\boldsymbol{\psi}$, X_i sono variabili aleatorie i.i.d. provenienti da una χ_1^2 , e λ_i sono gli autovalori di $[H^{-1}K]_{\boldsymbol{\psi}'}$, ovvero la sottomatrice ristretta al parametro $\boldsymbol{\psi}'$ ([14], [25]). Il *p-value* associato al test determina se accettare o rifiutare (tipicamente, quando $p < 0.05$) l'ipotesi nulla.

4.2.3 Cross-validation

La cross-validation è una tecnica non parametrica che valuta la performance del modello suddividendo il dataset in sottoinsiemi di addestramento e test. Essa può essere utilizzata per selezionare il modello più adatto evitando il sovra-adattamento e garantendo una buona capacità di generalizzazione su dati non visti. Questa tecnica, infatti, fornisce una stima robusta della capacità predittiva del modello. Per ogni modello candidato e addestrato sull'insieme di training, si calcola la previsione di una quantità di interesse (ad esempio il valore estremo) nelle località dell'insieme di test e si misura l'errore rispetto alle quantità reali osservate. Nel *k-fold cross-validation*, si ripete il procedimento per k sottoinsiemi di addestramento e i risultati delle k iterazioni vengono mediati per ottenere una stima complessiva delle prestazioni del modello. Alcune misure tipiche dell'errore, calcolate in ogni iterazione, sono lo scarto quadratico medio (RMSE), definito come:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_{i,p} - y_{i,o})^2}, \quad (4.13)$$

e l'errore assoluto medio (MAE), definito come:

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_{i,p} - y_{i,o}|, \quad (4.14)$$

dove n è il numero di osservazioni, $\hat{y}_{i,p}$ sono i valori predetti dal modello e $y_{i,o}$ sono i valori osservati. Per entrambe le misure, il modello selezionato tra i vari candidati è quello con errore medio più basso. Tra RMSE e MAE è preferibile utilizzare il primo, in quanto esso penalizza maggiormente gli errori grandi rispetto a quelli piccoli, mentre il secondo tratta tutti gli errori allo stesso modo.

Nel caso in cui la quantità predetta è il valore estremo nelle nuove località dell'insieme di test, è necessario applicare una tecnica di simulazione del processo max-stable. Gli algoritmi descritti in 3.4 considerano un processo max-stable semplice. Nei contesti pratici, in cui il processo ha una distribuzione marginale GEV arbitraria, è possibile ottenere delle simulazioni del processo max-stable semplice e successivamente trasformarle in realizzazioni del processo con distribuzione marginale GEV mediante le trasformazioni marginali (4.2).

4.3 Visualizzazione dei dati

Gli strumenti grafici di visualizzazione dei dati giocano un ruolo importante nell'analisi esplorativa dei dati, nella valutazione in termini di bontà di adattamento dei modelli max-stable e nella visualizzazione di quantità d'interesse stimate dai modelli. Ad esempio, tramite istogrammi è possibile visualizzare la distribuzione dei valori estremi e identificare la frequenza con cui i valori delle variabili di interesse, come l'altitudine delle diverse località, si presentano. Invece, tramite boxplot è possibile visualizzare la distribuzione dei dati attraverso quartili, mediana e valori anomali.

Una rappresentazione molto utile è quella delle mappe spaziali, che mostrano la variazione di un determinato valore, come l'altitudine, i valori di precipitazione o i parametri marginali della distribuzione GEV, su un'area geografica, permettendo di visualizzare immediatamente le differenze spaziali e identificare pattern geografici di interesse.

Un grafico adatto a identificare trend spaziali rispetto ai valori estremi che vengono assunti in diverse località è il *symbol plot*. Esso mostra la deviazione della media campionaria dei valori estremi di ciascuna località rispetto alla media campionaria complessiva, ovvero la media campionaria calcolata su tutte le località. In particolare, per ciascuna località viene disegnato un cerchio il cui raggio è proporzionale allo scostamento del valore in quella località rispetto al valore medio areale. Il colore del cerchio mostra se la deviazione è positiva (rosso) o se è negativa (blu). In modo analogo, è possibile mostrare la deviazione standard dei valori estremi per ciascuna località. Tramite questi grafici è possibile individuare località caratterizzate da valori estremi elevati, che si discostano significativamente dal valore medio, e visualizzare facilmente come tali località sono distribuite spazialmente. In tal modo, è possibile rilevare eventuali pattern geografici e identificare le aree con anomalie estreme. Inoltre, questi grafici forniscono informazioni visive che possono supportare decisioni e analisi approfondite riguardo alla struttura spaziale dei parametri dei modelli max-stable.

Dopo aver fittato e selezionato un modello max-stable, risulta necessario valutare se esso rappresenta in modo affidabile la realtà dei fenomeni estremi che si stanno analizzando, catturando correttamente la distribuzione dei valori estremi osservati nei dati e la struttura di dipendenza spaziale. Una volta validato un modello max-stable, è possibile utilizzarlo per fare previsioni su eventi estremi futuri. Tracciando su un grafico le misure di dipendenza spaziale presentate in 3.3, è possibile visualizzare la struttura di dipendenza spaziale stimata dai dati e dai modelli max-stable. Di conseguenza, è possibile verificare facilmente se la struttura di dipendenza dei modelli sia stata catturata correttamente o se sia sottostimata o sovrastimata. Al fine di analizzare se il modello rispecchia correttamente le caratteristiche degli estremi in ogni singola località, è possibile visualizzare il *return level plot* per alcune località prese singolarmente. Infine, è possibile rappresentare i valori predetti dai modelli dei *return levels* per uno specifico *return period* (2.10) in una mappa spaziale, al fine di prevedere le aree con anomalie estreme predette dal modello.

4.3.1 Coefficiente estremo

La funzione del coefficiente estremo per un modello max-stable assume la forma presentata in 3.3.1.2, dove vengono utilizzate le stime dei parametri $\hat{\psi}$ ottenute massimizzando la log-verosimiglianza a coppie. Per verificare se la struttura di dipendenza misurata dalla funzione del coefficiente estremo sia stata catturata correttamente dal modello, è possibile confrontarla con la stima del coefficiente estremo ottenuta direttamente considerando coppie di dati. Data la relazione (3.35) tra la funzione del coefficiente estremo e il F-madogramma, uno stimatore per la funzione del coefficiente estremo, calcolato direttamente sulle osservazioni, è data da [42]:

$$\hat{\theta}(h) = \frac{1 + 2\hat{\nu}_F(h)}{1 - 2\hat{\nu}_F(h)} \quad (4.15)$$

dove lo stimatore del F-madogramma per ciascuna coppia di località (s_i, s_j) a distanza $h = s_j - s_i$ è dato da:

$$\hat{\nu}_F(h) = \frac{1}{2n} \sum_{k=1}^n |\hat{F}(Z_k(s_j)) - \hat{F}(Z_k(s_i))|, \quad (4.16)$$

dove n è il numero di osservazioni e \hat{F} è ottenuto normalizzando il rango delle osservazioni. È possibile visualizzare anche la versione discretizzata dello stimatore della funzione del coefficiente estremo calcolato sulle osservazioni, ottenuta discretizzando il valore h in un numero fissato di intervalli e calcolando la media dello stimatore del F-madogramma (4.16) per ciascun intervallo.

4.3.2 Probabilità di concomitanza estrema

La funzione di probabilità di concomitanza estrema per un modello max-stable assume la forma presentata in 3.3.2.1, dove vengono utilizzate le stime dei parametri $\hat{\psi}$ ottenute massimizzando la log-verosimiglianza a coppie. Per verificare se la struttura di dipendenza misurata dalla funzione di concomitanza estrema sia stata catturata correttamente dal modello, è possibile confrontarla con la stima della probabilità di concomitanza estrema ottenuta direttamente considerando coppie di dati. [22] ha proposto tre stimatori di tale misura, mostrando che lo stimatore di Kendall è il più accurato. Esso è definito per la coppia di località (s_i, s_j) a distanza $h = s_j - s_i$ come:

$$\hat{p}_2(s_i, s_j) = \frac{2}{n(n-1)} \sum_{1 \leq k < l \leq n} \text{sign}\{Z_k(s_i) - Z_l(s_i)\} \text{sign}\{Z_k(s_j) - Z_l(s_j)\}, \quad (4.17)$$

dove n è il numero di osservazioni e $n(n-1)/2$ è il numero di coppie che vengono valutate nella sommatoria. Tale quantità, dunque, misura quanto frequentemente gli eventi estremi nella coppia di località (s_i, s_j) avvengono simultaneamente in modo concordante. Lo stimatore (4.17) è non distorto e asintoticamente normale. L'errore standard asintotico può essere stimato accuratamente mediante il metodo Jackknife, che consiste nel ricalcolare più volte \hat{p}_2 lasciando fuori dal campione una osservazione alla volta [30], ovvero tramite:

$$\text{std. err}(\hat{p}_2(s_i, s_j)) = \frac{1}{n} \sum_{l=1}^n \left(\hat{p}_{2-l}(s_i, s_j) - \hat{p}_2(s_i, s_j) \right)^2, \quad (4.18)$$

dove

$$\begin{aligned} \hat{p}_{2-l}(s_i, s_j) &= \frac{1}{(n-1)(n-2)} \sum_{\substack{1 \leq m < r \leq n, \\ m, r \neq l}} \text{sign}(Z_m(s_i) - Z_r(s_i)) \\ &\quad \times \text{sign}(Z_m(s_j) - Z_r(s_j)) = \\ &= \frac{n}{(n-2)} \hat{p}_2(s_i, s_j) - \frac{2}{(n-1)(n-2)} \sum_{m=1}^n \text{sign}(Z_m(s_i) - Z_l(s_i)) \\ &\quad \times \text{sign}(Z_m(s_j) - Z_l(s_j)). \end{aligned} \quad (4.19)$$

È possibile visualizzare anche la versione discretizzata dello stimatore della funzione di probabilità di concomitanza estrema calcolato sulle osservazioni, ottenuta discretizzando il valore della distanza h in un numero fissato di intervalli e calcolando la media della stima della probabilità di concomitanza estrema (4.17) per ciascun intervallo.

4.3.3 Return levels

Per quanto riguarda l'analisi marginale del modello max-stable, è possibile confrontare i *return levels* stimati dal modello con quelli osservati per una singola località per valori crescenti di *return periods*. Tale grafico viene chiamato *return level plot*. Nello specifico, i valori stimati del modello calcolano (2.10) utilizzando le stime dei parametri delle *trend surfaces* (4.4) per la singola località, mentre i valori osservati sono semplicemente i dati originali mostrati in ordine crescente. Inoltre, viene considerato l'intervallo di confidenza di significatività 5% calcolato rispetto a n realizzazioni della distribuzione GEV marginale. Tali realizzazioni vengono ottenute mediante la seguente formula:

$$\hat{z}_T = \begin{cases} \hat{\mu} + \frac{\hat{\sigma}}{\hat{\xi}} \{ [Exp(n)]^{-\hat{\xi}} - 1 \} & \text{per } \hat{\xi} \neq 0, \\ \hat{\mu} - \hat{\sigma} \log[Exp(n)] & \text{per } \hat{\xi} = 0, \end{cases} \quad (4.20)$$

dove $Exp(n)$ indica n realizzazioni dalla distribuzione esponenziale standard. Dunque, il *return level plot* mostra come la probabilità di superare un certo livello aumenta all'aumentare del periodo di ritorno in una specifica località.

Nei contesti pratici, spesso, un aspetto importante legato ai valori estremi riguarda l'estrapolazione di valori maggiori rispetto a quelli osservati nel periodo disponibile e in nuove località, ovvero la previsione di dove sia più probabile che si verifichino eventi più estremi di quelli osservati in un'area geografica [15]. Di conseguenza, è possibile mostrare una mappa spaziale dei *return levels* associati a uno specifico *return period*. Tali quantità sono determinate dalle distribuzioni marginali di un processo max-stable e ottenute mediante (2.10) per ogni località. I parametri delle distribuzioni marginali GEV vengono stimati attraverso le *trend surfaces* (4.4), dove i parametri β sono le stime del modello max-stable $\hat{\beta}$ (una parte del vettore stimato $\hat{\psi}$) con errori standard dati da (4.7). L'errore standard del valore predetto dei parametri della distribuzione marginale GEV nella nuova località k è:

$$std. err(\hat{\theta}_k) = \sqrt{[\mathbf{X}_\theta \hat{\mathbf{V}}_\theta \mathbf{X}_\theta^T]_{kk}}, \quad (4.21)$$

dove θ rappresenta il parametro di *location*, *scale* o *shape*, \mathbf{X}_θ è la matrice delle covariate relative al parametro θ e $\hat{\mathbf{V}}_\theta$ è la matrice di varianza e covarianza relativa al parametro θ . Di conseguenza, anche la stima del *return level* per la località i ha un errore standard, che può essere calcolato mediante il metodo Delta ([2], Capitolo 3, p. 56):

$$std. err(\hat{z}_{T_i}) \approx \sqrt{[\nabla z_T^T(\hat{\theta}) \hat{\mathbf{V}}_{\hat{\theta}} \nabla z_T(\hat{\theta})]_{ii}}, \quad (4.22)$$

dove $\hat{\mathbf{V}}_{\hat{\theta}} = \mathbf{X}_\theta \hat{\mathbf{V}}_\theta \mathbf{X}_\theta^T$ è la matrice di varianza e covarianza relativa alla stima dei parametri $\hat{\theta}$, mentre il gradiente di z_T calcolato in $\hat{\theta}$, denotando $y_T = -\log(1 - 1/T)$, è pari a:

$$\begin{aligned} \nabla z_T(\hat{\theta}) &= \left[1, \frac{y_T^{-\hat{\xi}} - 1}{\hat{\xi}}, -\frac{\hat{\sigma} y_T^{-\hat{\xi}} \log(y_T)}{\hat{\xi}}, -\frac{\hat{\sigma} [y_T^{-\hat{\xi}} - 1]}{\hat{\xi}^2} \right] = \\ &= \left[1, \frac{\hat{z}_T - \hat{\mu}}{\hat{\sigma}}, -\frac{\hat{\sigma} y_T^{-\hat{\xi}} \log(y_T)}{\hat{\xi}}, -\frac{\hat{z}_T - \hat{\mu}}{\hat{\xi}} \right]. \end{aligned} \quad (4.23)$$

Capitolo 5

Dataset delle precipitazioni estreme

Nei capitoli precedenti è stata esposta la teoria dei valori estremi e dei processi spaziali estremi, oltre alle metodologie per condurre inferenze sui modelli da addestrare su dati reali, per selezionare il modello più rappresentativo dei dati e valutarne l'accuratezza. Questo capitolo, invece, si concentra sulla descrizione dettagliata dei dati reali utilizzati per le analisi svolte nel presente lavoro. In particolare, si focalizza sulla descrizione del dataset utilizzato, relativo alle precipitazioni massime annuali del Piemonte, fornendo una panoramica completa delle caratteristiche del dataset, delle stazioni di monitoraggio e degli aspetti spaziali e temporali delle precipitazioni.

5.1 Descrizione del dataset

Il dataset utilizzato comprende dati sulle precipitazioni massime annuali di varia durata registrate in un periodo di riferimento che va dal 1928 al 2020 e in una rete di 498 stazioni pluviometriche ubicate nel territorio italiano della regione Piemonte. Le variabili registrate includono:

- un codice identificativo univoco della stazione pluviometrica (ID);
- l'anno relativo all'evento estremo registrato;
- i valori di precipitazione massima annuale, espressa in mm, per diverse durate temporali che variano da brevi periodi (10, 20 e 30 minuti, 1 e 3 ore) a periodi più prolungati, ovvero alcune ore (6 e 12 ore) o giorni interi (24 ore, 1, 2, 3, 4 e 5 giorni);
- data di registrazione dell'evento estremo, sotto forma di anno, mese e giorno;
- ora e minuto di registrazione dell'evento estremo, solo per le precipitazioni di durata fino a 24 ore dal momento che i dati relativi a precipitazioni di durata giornaliera si riferiscono allo stesso periodo di riferimento (9 del mattino in passato, mezzanotte attualmente).

Il dataset analizzato è stato estratto dall'*Improved Italian - Rainfall Extreme Dataset (I²-RED)* [31] che contiene le precipitazioni massime annuali di durata da 1 a 24 ore in Italia, ora esteso al fine di includere anche le durate sub-orarie e plurigiornaliere.

Ad ogni ID della stazione pluviometrica corrispondono le seguenti informazioni:

- coordinate geografiche della stazione di monitoraggio, ovvero longitudine, latitudine e altitudine, espresse in metri;
- *openness*, ovvero la visibilità rispetto alla posizione circostante in un'area di raggio 5 km dalla posizione attuale (*OPENNESS_5km*) tenendo conto del rilievo del terreno, valutata usando come input un Modello Digitale di Elevazione (DEM) a 500 metri di risoluzione;
- precipitazione media annuale (*MAP_BIGBANG*) registrata in corrispondenza del pluviometro ed espressa in mm, che è stata estratta dal dataset *BIGBANG4.0* di ISPRA [32];
- nome della stazione pluviometrica.

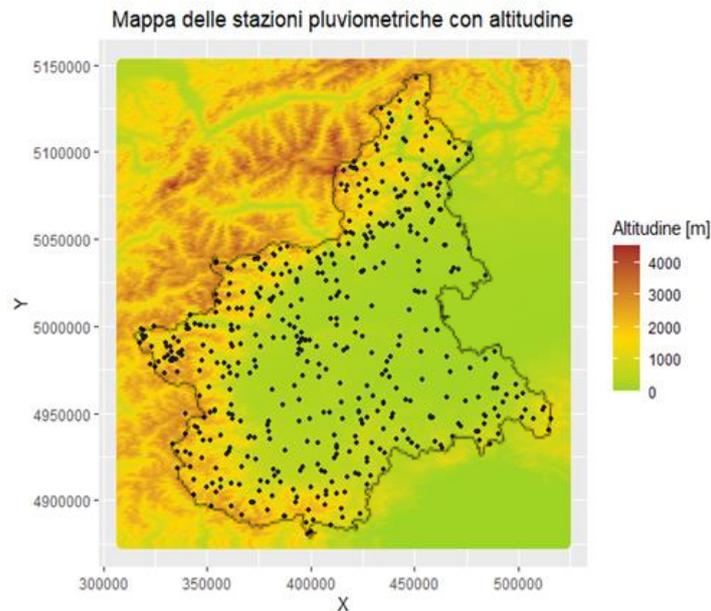


Figura 5.1: Mappa altimetrica del Piemonte e delle aree circostanti. I punti neri rappresentano le stazioni pluviometriche del Piemonte per cui si hanno i valori di precipitazione.

In particolare, longitudine (X) e latitudine (Y) sono espresse rispetto al sistema di riferimento EPSG:32632, ovvero WGS84 UTM 32N. La longitudine esprime la distanza orizzontale dal meridiano centrale, mentre la latitudine esprime la distanza verticale dall'equatore. Per quanto riguarda l'altitudine, sono forniti due valori: la quota del pluviometro ottenuta dal database ufficiale di ARPA Piemonte [33] (QUOTA_UFFICIALE) e la quota del pluviometro estratta da un modello digitale del terreno a bassa risoluzione (1 km), più rappresentativa della quota esatta del punto quando si analizzano gli estremi spazialmente (QUOTA_1km). Infatti, tale valore tiene conto dell'effetto di caratteristiche topografiche locali, come variazioni locali di altitudine, pendenze ed esposizioni, che possono influenzare il verificarsi di eventi di precipitazioni estreme. Di conseguenza, fornisce una misura più accurata delle condizioni topografiche che influenzano le precipitazioni ed aiuta ad ottenere modelli più affidabili nel caso di analisi regionali ([36], [37]).

Infine, è stato utilizzato un DEM per l'intera area del Piemonte e le zone circostanti per fini di visualizzazione, ottenuto dal *Shuttle Radar Topography Mission* (SRTM) [34] e con un ricampionamento ad 1 km di risoluzione al fine di ottenere la quota media dell'intorno del pluviometro e per uniformità con i valori di QUOTA_1km.

5.2 Caratteristiche del territorio

Considerando l'intera area rappresentata nella Figura 5.1, parte del territorio è a carattere montuoso, con rilievi più elevati lungo la catena delle Alpi a nord e a ovest e in parte degli Appennini a sud, mentre l'area pianeggiante si estende principalmente al centro-est. Tale geografia influenza e regola il clima del Piemonte. Inoltre, la regione si trova al crocevia di diverse masse d'aria: aria continentale proveniente dalla Pianura Padana, aria umida proveniente dal Mediterraneo e correnti atlantiche nord-occidentali. L'interazione di queste masse d'aria con i rilievi montuosi genera frequenti circolazioni locali e dà vita a una varietà di microclimi [35]. In particolare, nel caso di catene montuose elevate, tale interazione dà origine a precipitazioni di carattere orografico, intense e di lunga durata. Al contrario, ad altitudini minori si possono verificare precipitazioni intense e di breve durata di carattere convettivo. Di conseguenza, dal momento che la regione è caratterizzata da una variabilità significativa del valore di

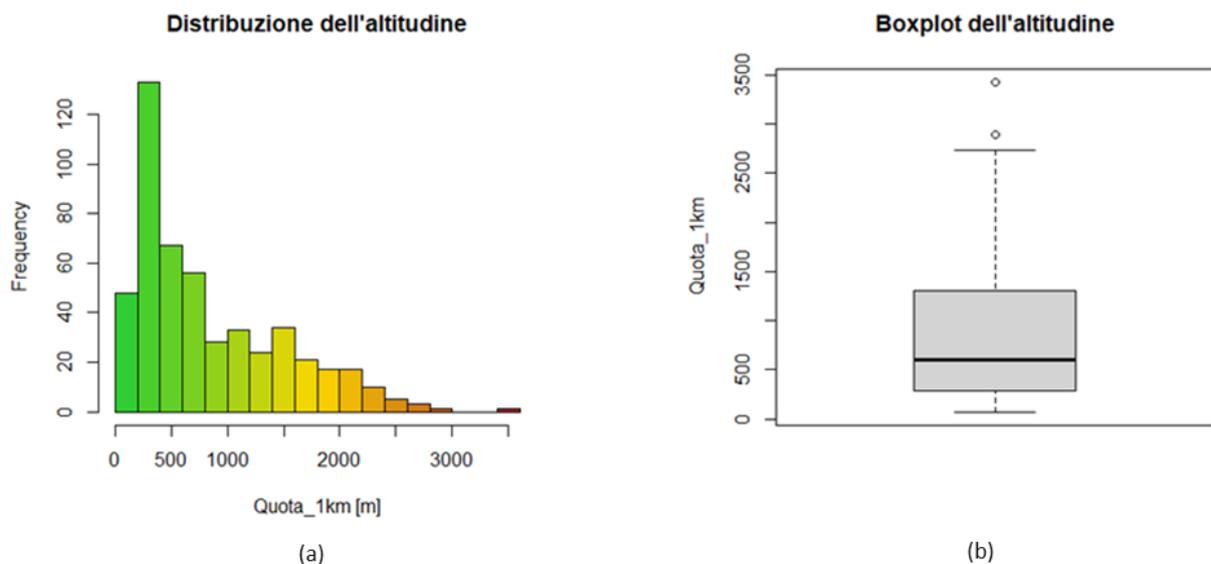


Figura 5.2: Distribuzione della quota delle stazioni pluviometriche con risoluzione di 1 km rappresentata da un istogramma (a) e da un boxplot (b).

altitudine e le tipologie di precipitazione dipendono da tale quantità, risulta necessario considerare la quota come covariata nei modelli max-stable. Inoltre, verranno analizzate le precipitazioni intense sia di breve durata sia persistenti.

5.2.1 Caratteristiche delle stazioni pluviometriche

Data la morfologia del Piemonte, le stazioni pluviometriche sono distribuite principalmente ad altitudini inferiori ai 1500 metri, come risulta evidente dalla Figura 5.2, mentre le stazioni poste ad un'altitudine superiore a circa 2700 metri sono in numero molto limitato. Infatti, in zone ad alta quota potrebbero non essere presenti corrente elettrica e una connessione internet stabile, entrambe necessarie per il funzionamento dei pluviometri e la trasmissione dei dati. Inoltre, la presenza di neve rappresenta un ulteriore problema per le stazioni pluviometriche, per cui oltre una certa quota in genere si installano solo nivometri.

I valori di altitudine ufficiale e di quella ottenute con una risoluzione di 1 km sono simili, come mostrato dalle statistiche descrittive presentate nella Tabella 5.1. Di conseguenza, è possibile considerare la quota con risoluzione di 1 km nelle successive analisi, per le motivazioni precedentemente descritte.

	X	Y	$QUOTA_{UFFICIALE}$	$QUOTA_{1km}$	$OPENNESS_{5km}$	$MAP_{BIGBANG}$
MIN	317833	4880823	77.0	72.48	1.183	660.8
$1^{\circ} QUANTILE$	371738	4946656	296.2	287.01	1.372	844.6
$MEDIANA$	411524	4990081	620.0	601.56	1.495	1062.4
$MEDIA$	408682	4995259	816.0	832.46	1.459	1130.3
$3^{\circ} QUANTILE$	444561	5039272	1224.5	1289.48	1.554	1320.7
MAX	515945	5142603	3325.0	3421.44	1.665	2237.0
$DEV. STD$	45747.6	61169.19	635.6237	655.5692	0.1085394	356.1877

Tabella 5.1: Statistiche descrittive (minimo, primo quartile, mediana, media, terzo quartile, massimo e deviazione standard) delle informazioni relative alle stazioni pluviometriche.

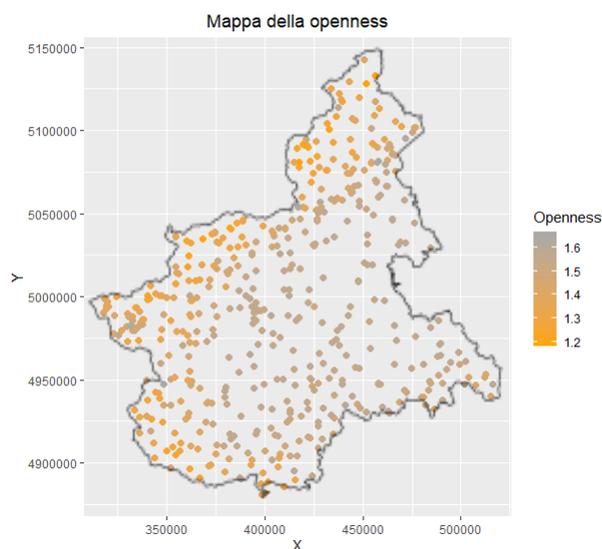


Figura 5.3: Mappa spaziale del valore di *openness*.

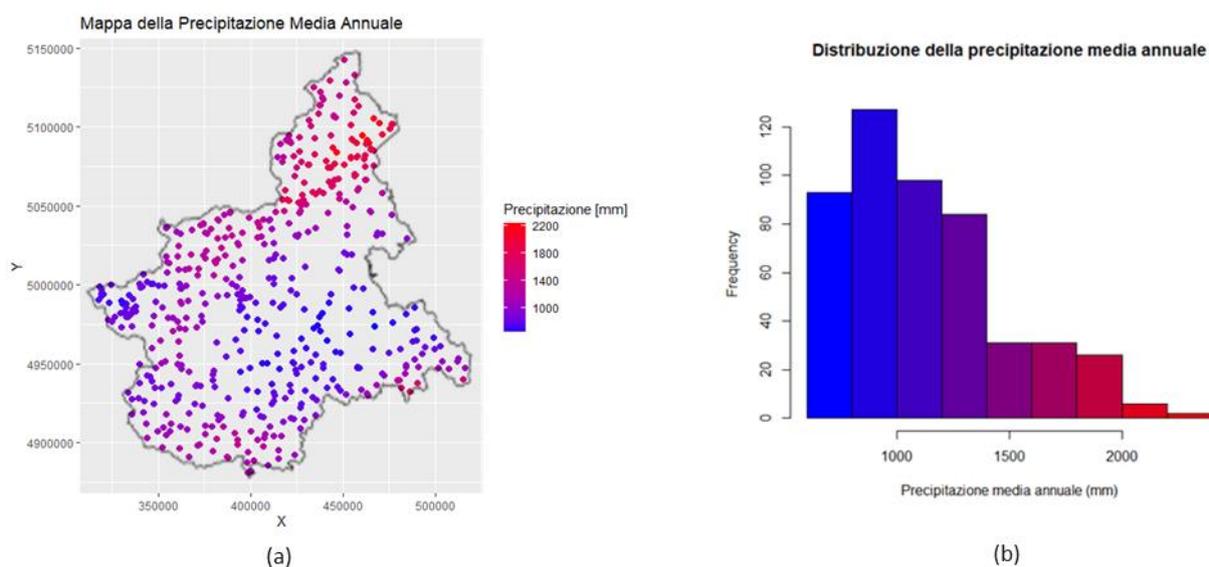


Figura 5.4: Mappa spaziale (a) e istogramma (b) del valore di precipitazione media annuale (MAP_BIGBANG) per ciascuna stazione pluviometrica.

Per quanto riguarda i valori di *openness*, essi sono strettamente legati alla morfologia del territorio, ovvero alla conformazione e alla struttura del paesaggio. Valori elevati sono tipici di zone relativamente piatte, come pianure e altipiani, e indicano che il paesaggio è molto aperto, con pochi ostacoli che bloccano la visuale. Valori bassi, invece, sono tipici di catene montuose e colline con molte variazioni altimetriche. Ciò è conforme a quanto mostrato dalla Figura 5.3. Il range di valori varia da un valore minimo di 1.183, diffuso nelle zone montuose, a un valore massimo di 1.665, caratteristico delle zone ad altitudine inferiore ai 1000 metri. L'*openness* è un indicatore topografico che potrebbe influenzare il microclima locale, ma non ha un impatto diretto e uniforme sulle precipitazioni estreme. Di conseguenza, la sua inclusione in un modello max-stable aumenterebbe la complessità del modello inutilmente senza migliorarne la precisione.

Un'informazione utile per comprendere preliminarmente la distribuzione spaziale delle precipitazioni è rappresentata dal valore medio di precipitazione annuale. Osservando le statistiche descrittive relative a tale quantità nella Tabella 5.1, si può notare che c'è un'alta variabilità, indicando un

comportamento diverso nelle varie stazioni pluviometriche. In particolare, dalla Figura 5.4 si nota che valori più alti di precipitazione media annuale sono localizzati a nord-est della regione, corrispondenti a valori di altitudine intermedi (circa 1500 metri), con picchi di 2200 mm. Dunque, solo poche località sono interessate da valori molto elevati di precipitazione media annuale, come mostrato anche dall'istogramma. A quote più elevate, la precipitazione media annuale varia intorno ai 1400-1800 mm. Tuttavia, la maggior parte del territorio, corrispondente alla zona pianeggiante o l'alta Val di Susa (al confine con la Francia), è caratterizzato da valori medi di precipitazione annuale inferiori ai 1000 mm. I valori osservati rispecchiano la distribuzione spaziale delle precipitazioni registrata in [35]. Tuttavia, la media annuale delle precipitazioni rappresenta le condizioni generali della regione e non necessariamente le caratteristiche estreme degli eventi. Inoltre, la distinzione tra le zone in cui è stata registrata una precipitazione media annuale elevata e zone con una precipitazione media annuale bassa può essere fatta considerando altre variabili incluse nei modelli max-stable, come longitudine, latitudine e altitudine. Di conseguenza, tale variabile non verrà inclusa nei modelli max-stable.

5.3 Analisi esplorativa del dataset di precipitazioni

Prima di procedere all'analisi spaziale delle precipitazioni estreme del Piemonte, risulta necessario comprendere meglio la distribuzione e le caratteristiche delle precipitazioni massime annuali registrate. La Tabella 5.2 presenta le statistiche descrittive per le precipitazioni di diversa durata, fornendo una sintesi numerica dei dati ed evidenziando le tendenze centrali e la variabilità delle precipitazioni massime annuali. In particolare, si nota che, in generale, per le precipitazioni di durata maggiore si registrano quantità più elevate, come previsto, ma anche deviazioni standard maggiori. Inoltre, i valori di mediana e media sono abbastanza vicini nel caso delle precipitazioni di breve durata, ma tendono a distanziarsi maggiormente per gli eventi più prolungati.

Ci sono valori mancanti per le precipitazioni di ciascuna durata. Tuttavia, per le precipitazioni di durata da 1 a 24 ore sono stati registrati più valori e, di conseguenza, il numero di valori mancanti è limitato. In particolare, il range di valori delle precipitazioni massime annuali di durata 1 ora può essere considerato rappresentativo degli eventi estremi di breve durata, quasi sempre prodotti da fenomeni convettivi. Al contrario, le precipitazioni massime annuali di durata 24 ore possono essere considerate rappresentative degli eventi estremi di lunga durata, in genere associati a fenomeni di tipo stratiforme. Di conseguenza, le successive analisi si concentreranno sulle precipitazioni estreme di queste due durate.

Osservando nel dettaglio i valori delle statistiche descrittive relative alle precipitazioni di durata 1 e 24 ore nella Tabella 5.2, si può notare che la mediana, la media e il terzo quartile assumono dei valori significativamente più bassi del valore massimo registrato. Di conseguenza, ci sono poche osservazioni molto estreme. Ciò è evidente anche nella Figura 5.5, dove si nota che la maggior parte delle osservazioni relative alle precipitazioni massime annuali di durata 1 ora appartiene al range di valori 10-40 mm, mentre poche osservazioni rispetto al totale superano 70 mm. In modo analogo, si nota che la maggior parte delle registrazioni relative alle precipitazioni massime annuali di durata 24 ore appartiene al range di valori 50-150 mm, mentre poche osservazioni rispetto al totale superano 250 mm.

5.3.1 Distribuzione temporale e spaziale

Nonostante le precipitazioni massime annuali siano state registrate dal 1928 al 2020, poche osservazioni rispetto al totale riguardano il periodo iniziale tra il 1928 e il 1960 circa, mentre la maggior parte delle osservazioni riguardano il periodo finale tra il 2000 e il 2020, come si può notare dalla Figura 5.6.

	<i>MIN</i>	<i>1° QUARTILE</i>	<i>MEDIANA</i>	<i>MEDIA</i>	<i>3° QUARTILE</i>	<i>MAX</i>	<i>DEV. STD</i>	<i>NA</i>
<i>10 MIN</i>	2.20	10.00	13.80	14.45	18.00	41.60	5.869	6021
<i>20 MIN</i>	3.60	14.10	19.60	20.83	26.10	69.20	9.016	6026
<i>30 MIN</i>	4.50	16.50	23.00	24.73	31.10	81.90	11.161	6025
<i>1 H</i>	4.0	18.0	25.40	28.10	35.2	135.6	14.014	27
<i>3 H</i>	7.40	26.90	36.30	41.08	49.90	277.00	21.077	31
<i>6 H</i>	9.80	35.75	47.50	54.23	64.60	406.30	28.266	33
<i>12 H</i>	10.80	47.60	64.00	74.13	89.00	538.10	40.017	31
<i>24 H</i>	10.8	63.4	87.40	101.60	124.5	633.0	55.634	31
<i>1 GG</i>	9.40	56.80	78.00	92.75	113.60	619.00	54.338	6919
<i>2 GG</i>	18.4	78.2	110.80	130.20	161.0	644.8	74.689	6919
<i>3 GG</i>	23.4	90.8	126.20	150.80	190.6	697.4	86.085	6919
<i>4 GG</i>	26.4	99.2	137.60	164.50	206.8	773.0	93.358	6919
<i>5 GG</i>	26.4	104.4	145.6	175.2	219.6	799.4	100.533	6919

Tabella 5.2: Statistiche descrittive (minimo, primo quartile, mediana, media, terzo quartile, massimo e deviazione standard) e numeri di valori mancanti dei valori di precipitazioni massime annuali di varie durate.

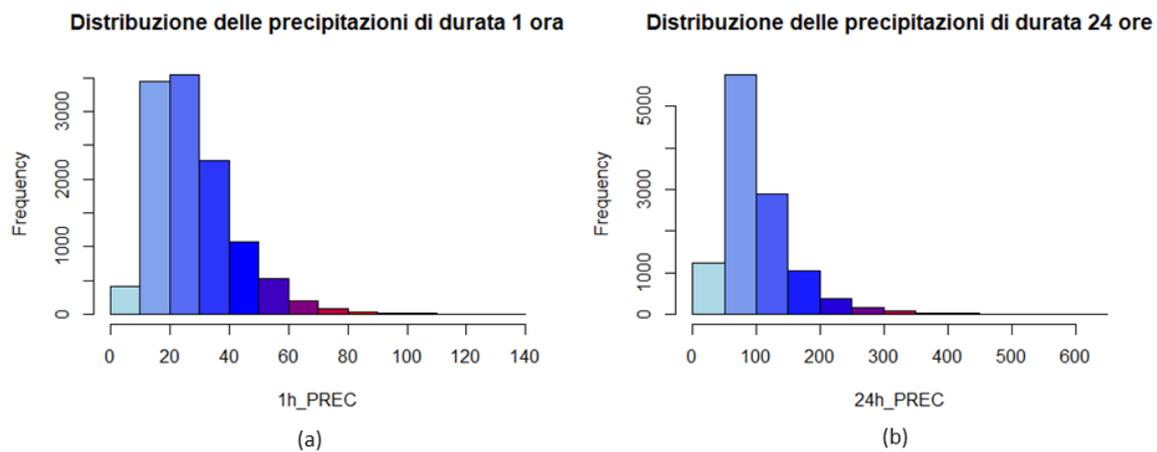


Figura 5.5: Istogramma delle precipitazioni massime annuali di durata 1 ora (a) e 24 ore (b).

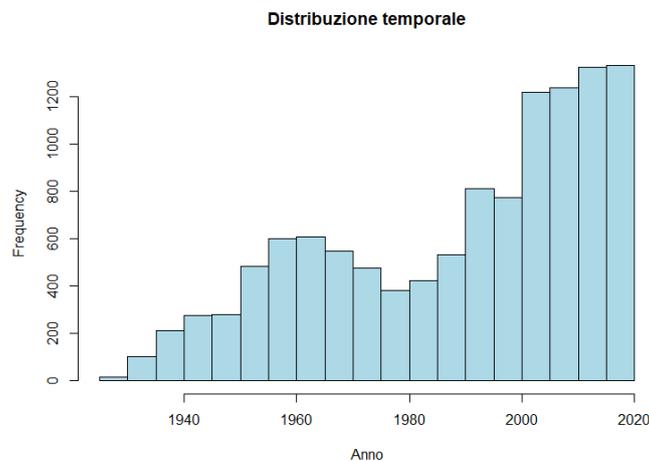


Figura 5.6: Distribuzione temporale delle precipitazioni massime annuali.

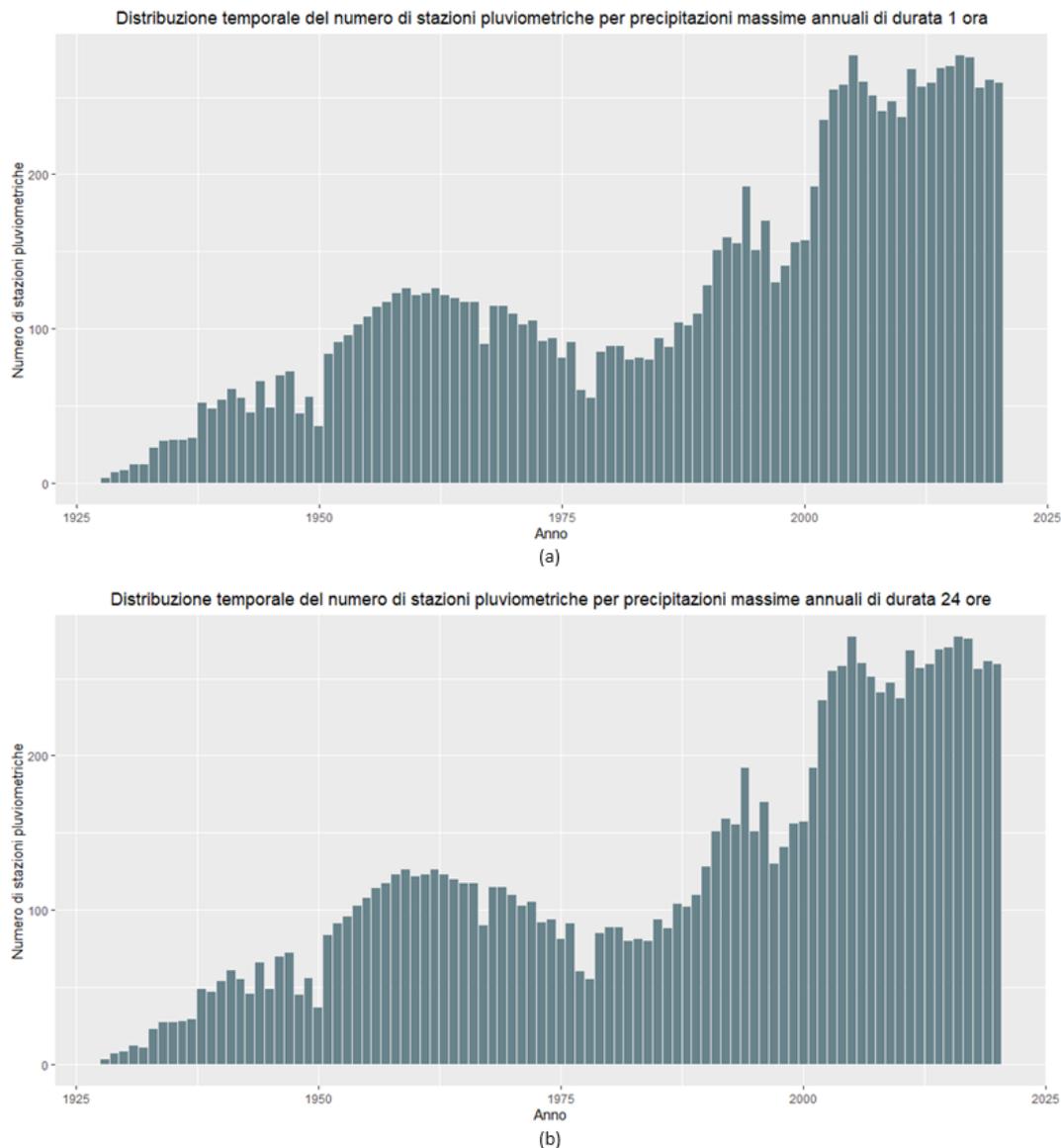


Figura 5.7: Istogramma del numero di stazioni pluviometriche che hanno registrato i valori di precipitazioni massime annuali di durata 1 ora (a) e 24 ore (b) rispetto agli anni di osservazione.

Considerando contemporaneamente la distribuzione temporale e spaziale degli eventi estremi registrati, si può notare dalla Figura 5.7 che le osservazioni registrate nel periodo iniziale (1928-1960 circa) derivano da poche stazioni pluviometriche per le precipitazioni massime annuali sia di durata 1 ora sia di durata 24 ore. Dunque, nel primo periodo si hanno poche osservazioni che interessano una copertura spaziale limitata. Al contrario, se si considera il periodo finale (2000-2020), le osservazioni registrate derivano dalla (quasi) totalità delle stazioni pluviometriche del Piemonte considerate per quasi ogni anno. Dunque, l'ultimo periodo di osservazione risulta essere ricco e con un'ampia copertura spaziale. Di conseguenza, in seguito a una prima analisi che considera l'intero periodo di osservazione o il periodo compreso tra il 1960 e il 2020, è possibile concentrarsi sugli ultimi 21 anni. Inoltre, per quattro stazioni pluviometriche è stato registrato un solo valore di precipitazione, e di conseguenza non si ha a disposizione una serie temporale per tali località. I modelli max-stable spaziali possono essere applicati anche in presenza di valori mancanti. Tuttavia, è necessario che siano presenti almeno due osservazioni per ciascuna stazione pluviometrica, ovvero che siano stati registrati i valori delle precipitazioni massime per almeno due anni. Di conseguenza, per ogni periodo di osservazione considerato, verranno escluse le stazioni pluviometriche con solo un valore registrato per tale periodo.

Capitolo 6

Risultati

In seguito ad un'analisi esplorativa del dataset fornita nel Capitolo 5, questo capitolo sarà dedicato ai risultati ottenuti applicando alle precipitazioni massime annuali di durata 1 e 24 ore i modelli max-stable spaziali presentati nel Capitolo 3 e considerando le tecniche di inferenza, selezione e validazione del modello presentate nel Capitolo 4. Nello specifico, verrà dapprima presentata un'analisi empirica della distribuzione spaziale degli eventi estremi e della stima della dipendenza spaziale, e successivamente verranno confrontate due metodologie per la selezione del modello max-stable. Inoltre, in questo lavoro non verrà considerata alcuna dipendenza temporale per i parametri della distribuzione GEV marginale, ma solo una dipendenza spaziale attraverso l'uso di longitudine, latitudine e/o altitudine come covariate. Infine, verrà valutato il modello attraverso i grafici presentati nel paragrafo 4.3 e verrà predetto il *25-year return level* per l'intero territorio del Piemonte e le aree circostanti.

6.1 Metodologie

La prima metodologia consiste nel selezionare dapprima tramite TIC (o ANOVA quando possibile) le *trend surfaces* per i parametri della distribuzione GEV e successivamente selezionare tramite TIC il modello max-stable, addestrato tenendo conto delle *trend surfaces* precedentemente scelte. Nello specifico, sono state considerate le *trend surfaces* riportate nella Tabella 6.1, dove X rappresenta la longitudine, Y la latitudine e Z l'altitudine. I modelli max-stable considerati includono i modelli Schlather, Brown-Resnick ed Extremal-t. In particolare, è stata considerata la funzione di correlazione *powered exponential* per entrambi i modelli Schlather ed Extremal-t e valore di *nugget* nullo.

La seconda metodologia consiste nel selezionare tramite k-fold cross-validation direttamente un modello max-stable addestrato considerando diverse combinazioni per le *trend surfaces*. Dunque, l'inferenza sui parametri della distribuzione GEV marginale e della dipendenza spaziale avviene in contemporanea. I modelli max-stable considerati sono gli stessi della prima metodologia, ovvero Schlather, Brown-Resnick ed Extremal-t. Per quanto riguarda i parametri della distribuzione marginale, al fine di ridurre il numero di combinazioni e, di conseguenza, il tempo computazionale della procedura, sono state considerate diverse *trend surfaces* solo per il parametro *scale*, tenendo fissate quelle relative ai parametri di *location* e *shape*. Nello specifico, le covariate relative al parametro di *location* sono longitudine, latitudine e altitudine, mentre il parametro di *shape* è costante. Le *trend surfaces* considerate per il parametro di *scale* sono riportate nella Tabella 6.2.

Ciascun modello è stato addestrato considerando i valori di precipitazione registrati dal 90% delle stazioni pluviometriche, selezionate casualmente, ed è stato valutato l'errore di predizione sui valori di precipitazione relativi al restante 10% delle stazioni, tramite RMSE e MAE. Il procedimento è stato ripetuto considerando 10 blocchi di stazioni pluviometriche, ciascuno equilibrato in termini di dispersione delle località e valori di altitudine. Infine, viene selezionato il modello con RMSE medio minore.

La prima metodologia verrà utilizzata per analizzare il dataset di precipitazioni massime annuali di durata 1 e 24 ore per l'intero periodo di osservazione (1928-2020) e per il periodo 2000-2020. La seconda metodologia, invece, verrà utilizzata per il periodo di osservazione 1960-2020, al fine di ridurre

	<i>Location</i> $\mu(x)$	<i>Scale</i> $\sigma(x)$	<i>Shape</i> $\xi(x)$
M_0	$\beta_{1,\mu}$	$\beta_{1,\sigma}$	$\beta_{1,\xi}$
M_1	$\beta_{1,\mu} + \beta_{2,\mu}X + \beta_{3,\mu}Y$	$\beta_{1,\sigma}$	$\beta_{1,\xi}$
M_{1a}	$\beta_{1,\mu} + \beta_{2,\mu}X + \beta_{3,\mu}Y + \beta_{4,\mu}Z$	$\beta_{1,\sigma}$	$\beta_{1,\xi}$
M_{2x}	$\beta_{1,\mu} + \beta_{2,\mu}X + \beta_{3,\mu}Y$	$\beta_{1,\sigma} + \beta_{2,\sigma}X$	$\beta_{1,\xi}$
M_{2xa}	$\beta_{1,\mu} + \beta_{2,\mu}X + \beta_{3,\mu}Y + \beta_{4,\mu}Z$	$\beta_{1,\sigma} + \beta_{2,\sigma}X$	$\beta_{1,\xi}$
M_{2xa2}	$\beta_{1,\mu} + \beta_{2,\mu}X + \beta_{3,\mu}Y + \beta_{4,\mu}Z$	$\beta_{1,\sigma} + \beta_{2,\sigma}X + \beta_{3,\sigma}Z$	$\beta_{1,\xi}$
M_{2y}	$\beta_{1,\mu} + \beta_{2,\mu}X + \beta_{3,\mu}Y$	$\beta_{1,\sigma} + \beta_{2,\sigma}Y$	$\beta_{1,\xi}$
M_{2ya}	$\beta_{1,\mu} + \beta_{2,\mu}X + \beta_{3,\mu}Y + \beta_{4,\mu}Z$	$\beta_{1,\sigma} + \beta_{2,\sigma}Y$	$\beta_{1,\xi}$
M_{2ya2}	$\beta_{1,\mu} + \beta_{2,\mu}X + \beta_{3,\mu}Y + \beta_{4,\mu}Z$	$\beta_{1,\sigma} + \beta_{2,\sigma}Y + \beta_{3,\sigma}Z$	$\beta_{1,\xi}$
M_2	$\beta_{1,\mu} + \beta_{2,\mu}X + \beta_{3,\mu}Y$	$\beta_{1,\sigma} + \beta_{2,\sigma}X + \beta_{3,\sigma}Y$	$\beta_{1,\xi}$
M_{2a}	$\beta_{1,\mu} + \beta_{2,\mu}X + \beta_{3,\mu}Y + \beta_{4,\mu}Z$	$\beta_{1,\sigma} + \beta_{2,\sigma}X + \beta_{3,\sigma}Y$	$\beta_{1,\xi}$
M_{2a2}	$\beta_{1,\mu} + \beta_{2,\mu}X + \beta_{3,\mu}Y + \beta_{4,\mu}Z$	$\beta_{1,\sigma} + \beta_{2,\sigma}X + \beta_{3,\sigma}Y + \beta_{4,\sigma}Z$	$\beta_{1,\xi}$
M_3	$\beta_{1,\mu} + \beta_{2,\mu}X + \beta_{3,\mu}Y$	$\beta_{1,\sigma} + \beta_{2,\sigma}X + \beta_{3,\sigma}Y$	$\beta_{1,\xi} + \beta_{2,\xi}X + \beta_{3,\xi}Y$

Tabella 6.1: *Trend surfaces* per i parametri della distribuzione GEV.

<i>Scale</i> $\sigma(x)$
$\beta_{1,\sigma} + \beta_{2,\sigma}X$
$\beta_{1,\sigma} + \beta_{2,\sigma}Y$
$\beta_{1,\sigma} + \beta_{2,\sigma}X + \beta_{3,\sigma}Y$
$\beta_{1,\sigma} + \beta_{2,\sigma}X + \beta_{3,\sigma}Z$
$\beta_{1,\sigma} + \beta_{2,\sigma}Y + \beta_{3,\sigma}Z$
$\beta_{1,\sigma} + \beta_{2,\sigma}X + \beta_{3,\sigma}Y + \beta_{4,\sigma}Z$

Tabella 6.2: *Trend surfaces* per il parametro *scale* della distribuzione GEV.

i tempi computazionali e dal momento che il numero di stazioni pluviometriche considerate nel periodo temporale antecedente il 1960 è molto limitato, e per il periodo 2000-2020. La scelta di questi periodi è stata, dunque, influenzata dalla disponibilità dei dati delle stazioni pluviometriche. È stato considerato l'intero periodo di osservazione per ottenere una visione complessiva delle precipitazioni estreme nel Piemonte. Tuttavia, il numero ridotto di stazioni pluviometriche disponibili durante i primi decenni di questo periodo può influire sulla rappresentatività spaziale dei dati. Il periodo di osservazione 2000-2020 è stato incluso per analizzare le precipitazioni estreme nel Piemonte, considerando i dati più recenti e provenienti da un numero maggiore di stazioni pluviometriche.

6.2 Precipitazioni massime annuali di durata 1 ora

6.2.1 Dataset completo (1928-2020)

Escludendo le stazioni meteorologiche per cui non è stato registrato alcun valore oppure è stato registrato solamente un valore nell'intero periodo di osservazione, l'analisi è stata condotta sulle rimanenti 494 località. Dal momento che solo quattro stazioni meteorologiche sono state escluse, la distribuzione dell'altitudine complessiva non è variata significativamente rispetto alla Figura 5.2. In modo analogo, le statistiche descrittive sono pressoché inalterate rispetto a quelle nella Tabella 5.2.

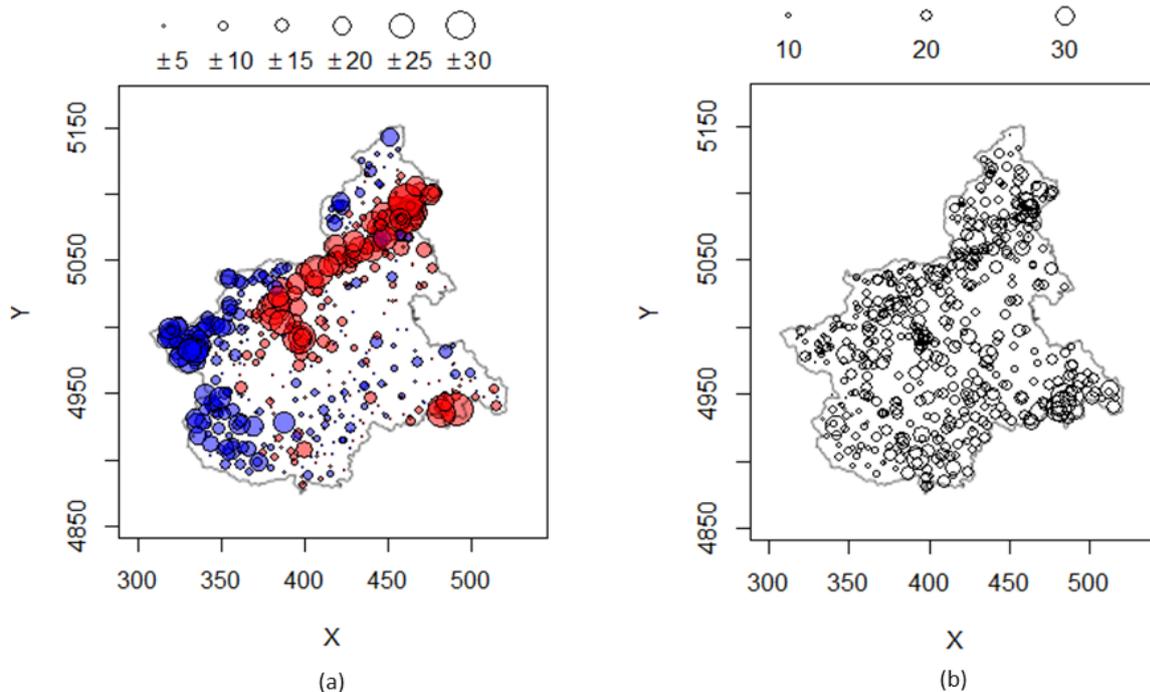


Figura 6.1: *Symbol plot* che mostra, per ciascuna stazione pluviometrica, la deviazione della media campionaria locale rispetto alla media campionaria complessiva (a) e la deviazione standard (b) per le precipitazioni di durata 1 ora e l'intero periodo di osservazione (1928-2020).

Il pannello a sinistra della Figura 6.1 mostra la deviazione della media campionaria di ciascuna stazione pluviometrica rispetto al valore di media complessiva pari a 28.1 mm, mentre il pannello a destra mostra la deviazione standard di ciascuna località. Tenendo conto della mappa altimetrica nella Figura 5.1, si può osservare che le zone caratterizzate da valori medi significativamente inferiori rispetto alla media complessiva, rappresentati da cerchi blu più ampi, corrispondono alle aree alpine ad altitudini maggiori, situate ai confini nord e ovest del Piemonte. Inoltre, in queste località le deviazioni standard dei valori di precipitazione sono basse. Di conseguenza, il valore medio calcolato è rappresentativo del valore di precipitazioni di breve durata registrate in queste zone. Dunque, ci si aspetta che tali aree non siano particolarmente interessate a fenomeni di precipitazione intensi e di breve durata, dal momento che presentano generalmente quantitativi minimi di precipitazione per l'intero periodo di osservazione (1928-2020). Al contrario, le zone caratterizzate da valori medi significativamente maggiori rispetto alla media complessiva, rappresentati da cerchi rossi più ampi, corrispondono alle località montuose con altitudine inferiore, poste a nord del Piemonte, e agli Appennini, a sud-est della regione. Tali località sono, tuttavia, caratterizzate da deviazioni standard maggiori. Di conseguenza, queste zone sono soggette a precipitazioni di breve durata abbondanti alternate ad anni in cui le precipitazioni sono più moderate. Dunque, ci si può aspettare che tali aree siano particolarmente soggette a fenomeni estremamente intensi di breve durata. Infine, le zone pianeggianti sono caratterizzate da valori medi simili alla media complessiva e con deviazioni standard moderate, rappresentati da cerchi di misura inferiore. Di conseguenza, queste zone sono soggette a precipitazioni di breve durata di intensità perlopiù pari a circa 28 mm. Dunque, ci si aspetta che in tali aree si registrino valori simili di precipitazioni di durata 1 ora. L'analisi della Figura 6.1 suggerisce dunque l'utilizzo delle coordinate geografiche delle stazioni pluviometriche come covariate dei parametri della distribuzione GEV e in particolar modo del parametro *location*, dal momento che è possibile distinguere aree interessate da eventi particolarmente estremi rispetto ad aree dove sono state registrate quantità minori di precipitazione, e del parametro *scale*, al fine di modellare variazioni nella variabilità spaziale degli eventi estremi.

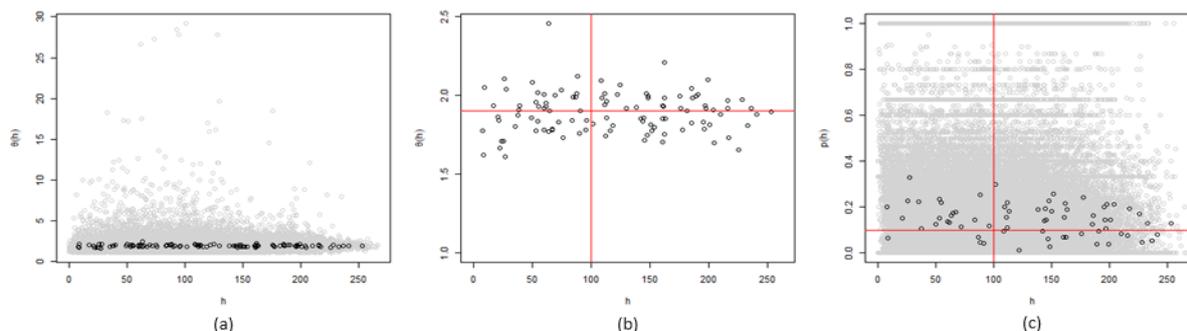


Figura 6.2: Dipendenza spaziale stimata empiricamente dai dati di precipitazione di durata 1 ora per l'intero periodo di osservazione (1928-2020), tramite coefficiente estremo (a, b) e probabilità di concomitanza estrema (c). I punti grigi sono le stime empiriche, mentre i punti neri sono le stime discretizzate con 12000 *bins*, con ingrandimento sulla versione discretizzata per il coefficiente estremo (b).

La stima empirica della dipendenza spaziale tra le stazioni pluviometriche per i dati di precipitazione di durata 1 ora è molto bassa, prossima all'indipendenza, già a basse distanze tra due stazioni pluviometriche, come mostrato dalla Figura 6.2. Nello specifico, i primi due pannelli mostrano la stima del coefficiente estremo calcolata per ciascuna coppia di località. Osservando nel dettaglio la versione discretizzata nel pannello (b), si può notare che i punti stanno nel range di valori tra 1.7 circa a poco sopra 2.0 quasi per ogni distanza h , con valore intermedio pari a circa 1.9 (linea rossa orizzontale) già a 100 km di distanza tra due stazioni pluviometriche. Solo a distanze inferiori ai 50 km sono presenti valori leggermente più bassi ma superiori a 1.5. Il pannello (c) mostra, invece, la stima della probabilità di concomitanza estrema calcolata per ciascuna coppia di località. Osservando i punti neri, si può notare che essi stanno nel range di valori tra 0.0 e poco sotto 0.4 circa per quasi ogni distanza. Ad esempio, a 100 km di distanza tra due stazioni pluviometriche la stima della probabilità di concomitanza estrema è di poco superiore a 0.1. Questo risultato ottenuto è conforme alle aspettative, considerando che gli eventi convettivi sono altamente concentrati nello spazio; pertanto, è prevedibile che la dipendenza spaziale diminuisca rapidamente a breve distanza.

Sulla base delle considerazioni effettuate, sono stati addestrati diversi modelli GEV spaziali al fine di selezionare le *trend surfaces* per i parametri della distribuzione. I risultati sono presentati nell'Appendice A nella Tabella A.1. Il modello M_{2xa2} ha registrato il TIC più basso tra tutti i modelli, pari a 88900.20. Il test ANOVA tra tale modello e il secondo con TIC più basso, ovvero il modello M_{2xa} con TIC pari a 88997.46, ha restituito un *p-value* inferiore a $2.2e-16$. Di conseguenza, le *trend surfaces* selezionate comprendono longitudine, latitudine e altitudine come covariate per il parametro *location*, longitudine e altitudine per il parametro *scale* e nessuna per il parametro *shape*. Le stime dei coefficienti e relativi errori standard sono presentati in (A.1). In particolare, si può notare che i coefficienti relativi all'altitudine sono di segno opposto rispetto alle altre covariate e maggiori in valore assoluto, mentre il parametro di *shape* è positivo, seppur piccolo, suggerendo una distribuzione di tipo Fréchet.

In base ai risultati ottenuti sui modelli max-stable successivamente addestrati, il modello Brown-Resnick performa meglio rispetto agli altri due modelli. Infatti, la prima riga nella Tabella A.2 mostra un valore di TIC significativamente inferiore per il modello Brown-Resnick e la Figura 6.3 mostra che tale modello cattura la dipendenza spaziale con una crescita (diminuzione) graduale della funzione del coefficiente estremo (funzione di probabilità di concomitanza estrema) in base ai valori di distanza tra coppie di stazioni pluviometriche. In particolare, nel caso della funzione del coefficiente estremo (b), si nota per il modello di Brown-Resnick una crescita repentina a bassissime distanze fino ad assumere un valore pari a 1.6 circa, seguita da una crescita graduale fino a circa 1.9 a 50 km di distanza, per poi

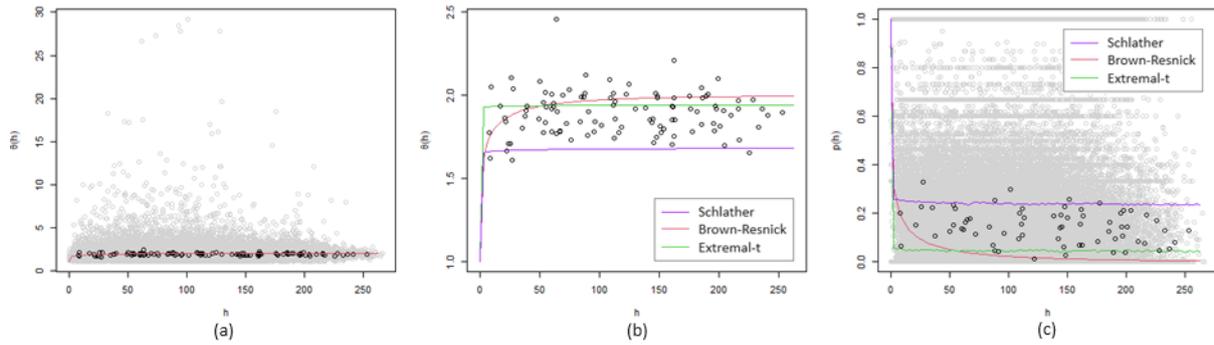


Figura 6.3: Valutazione della dipendenza spaziale stimata dai modelli max-stable per le precipitazioni di durata 1 ora per l'intero periodo di osservazione (1928-2020), tramite funzione del coefficiente estremo per il modello di Brown-Resnick (a) e per tutti i modelli (b), e funzione di probabilità di concomitanza estrema per ogni modello (c).

<i>Funzione</i>	<i>Distanza</i>	<i>Schlather</i>	<i>Brown-Resnick</i>	<i>Extremal-t</i>
<i>Coefficiente estremo</i>	<i>h = 5 km</i>	1.658865	1.694459	1.929736
	<i>h = 10 km</i>	1.663539	1.778198	1.932331
	<i>h = 20 km</i>	1.667939	1.854829	1.934684
	<i>h = 30 km</i>	1.670385	1.893632	1.935954
	<i>h = 50 km</i>	1.673333	1.933983	1.937447
	<i>h = 100 km</i>	1.677093	1.971619	1.939289
<i>Probabilità di concomitanza estrema</i>	<i>h = 5 km</i>	0.6034875	0.24325325	0.06279817
	<i>h = 10 km</i>	0.2569516	0.17299354	0.05220661
	<i>h = 20 km</i>	0.2509686	0.11197724	0.05316410
	<i>h = 30 km</i>	0.2459039	0.08217762	0.04511245
	<i>h = 50 km</i>	0.2462662	0.05166118	0.04643710
	<i>h = 100 km</i>	0.2392663	0.02311458	0.04487048

Tabella 6.3: Valori delle stime delle funzioni del coefficiente estremo e di probabilità di concomitanza estrema in alcune distanze per i modelli max-stable nel caso di precipitazioni di durata 1 ora e l'intero periodo di osservazione (1928-2020).

continuare a crescere per distanze maggiori sino a un massimo di 2.0. Al contrario, la funzione del coefficiente estremo stimata dagli altri due modelli assume quasi subito un valore pressoché costante. Nello specifico, il modello Extremal-t assume sin da subito una bassa dipendenza spaziale tra le stazioni pluviometriche, dal momento che i valori della funzione del coefficiente estremo stimata si aggirano intorno a 1.9. Invece, il modello Schlather stima maggiore dipendenza spaziale di quella stimata dai dati, dal momento che la funzione del coefficiente estremo assume valori intorno a 1.65. A differenza della stima della funzione del coefficiente estremo che assume perlopiù gli stessi valori di quelli stimati empiricamente, la funzione di probabilità di concomitanza estrema (c) stimata dal modello Brown-Resnick assume valori leggermente inferiori alle stime effettuate sui dati, ovvero assume una minore dipendenza, prossima all'indipendenza, già a basse distanze. Analogamente alla stima della funzione del coefficiente estremo, la funzione di probabilità di concomitanza estrema stimata dagli altri due modelli assume quasi subito un valore pressoché costante, intorno a 0.25 per il modello Extremal-t e intorno a 0.05 per il modello Schlather. I valori stimati dai tre modelli delle due funzioni per alcuni valori di distanza h sono riportati nella Tabella 6.3.

I parametri e gli errori standard stimati dal modello Brown-Resnick per le *trend surfaces* e per i valori di *range* λ e *smooth* κ sono riportati di seguito.

$$\begin{aligned}
\mu(x) &= -68.20392_{(1.57264)} + 0.06924_{(0.00039)}lon(x) + \\
&\quad + 0.01303_{(0.00033)}lat(x) - 3.07528_{(0.02410)}alt(x), \\
\sigma(x) &= -2.91832_{(0.11531)} + 0.03352_{(0.00025)}lon(x) - 1.57287_{(0.01853)}alt(x), \\
\xi(x) &= 0.10310_{(0.00923)}, \\
\lambda &= 1.1599_{(0.0097)}, \\
\kappa &= 0.5077_{(0.0011)}.
\end{aligned} \tag{6.1}$$

Si può notare che le stime dei coefficienti relativi al parametro di *scale* sono simili ai valori stimati in (A.1), mentre le stime dei coefficienti dei parametri di *location* e *shape* hanno subito maggiori variazioni. In particolare, il coefficiente $\beta_{1,\mu}$ è aumentato significativamente e la stima sembra più accurata in quanto l'errore standard è diminuito. In modo analogo, il coefficiente relativo al parametro di *shape* è aumentato leggermente, convalidando l'ipotesi di una distribuzione marginale di tipo Fréchet. Gli errori standard sono diminuiti in generale per ciascun coefficiente. Infine, la bassa dipendenza spaziale notata nella Figura 6.3 si riscontra nei parametri di *range* e *smooth*.

L'analisi locale del modello max-stable Brown-Resnick selezionato è stata effettuata per quattro stazioni pluviometriche casuali, mostrate nel pannello a sinistra della Figura 6.4. Le stazioni selezionate si trovano in zone pianeggianti (b), collinari (a, d) e montuose con altitudine intermedia (c). I quattro *return level plots* mostrano che il modello è in parte coerente con i dati osservati in alcune località, come nel caso mostrato dal pannello (b). Tuttavia, in alcuni siti il modello sovrastima la frequenza degli eventi estremi, come nei casi mostrati dai pannelli (a, d), mentre in altre stazioni il modello sottostima la frequenza degli eventi estremi, come nel caso mostrato dal pannello (c).

Infine, è stato predetto dal modello Brown-Resnick il *25-year return level* in tutta la regione e le zone circostanti. I valori predetti e i relativi errori standard sono mostrati in Figura 6.5. In particolare, si può notare che, tenendo conto della mappa altimetrica nella Figura 5.1, nelle zone alpine ad elevata altitudine, i valori predetti di precipitazione non superano in media 30 mm e gli errori standard corrispondenti sono bassi (circa 0.10). Nelle zone montuose ad altitudini intermedie, come gli Appennini e le Alpi a quota intermedia, i valori predetti di precipitazione non superano in media il range di valori tra 30 e 60 mm. Gli errori standard in tali aree sono in genere relativamente bassi e inferiori a 0.15, tranne a nord-est dove sono circa 0.30. Infine, nelle zone collinari e pianeggianti, i valori predetti di precipitazione non superano in media 90 mm e gli errori standard sono compresi tra 0.15 e 0.25 circa. Di conseguenza, dal momento che ovunque gli errori standard non sono molto elevati, l'incertezza dei valori predetti dei *return levels* è bassa. In particolare, le aree ad altitudini elevate rispecchiano quanto osservato dai dati registrati nell'intero periodo di osservazione e mostrato nella Figura 6.1. Tuttavia, i risultati ottenuti assumono che le aree principalmente interessate da eventi estremi di breve durata siano quelle pianeggianti e collinari ad est del Piemonte, mentre le aree montuose a quote intermedie siano soggette a precipitazioni di circa 50 mm.

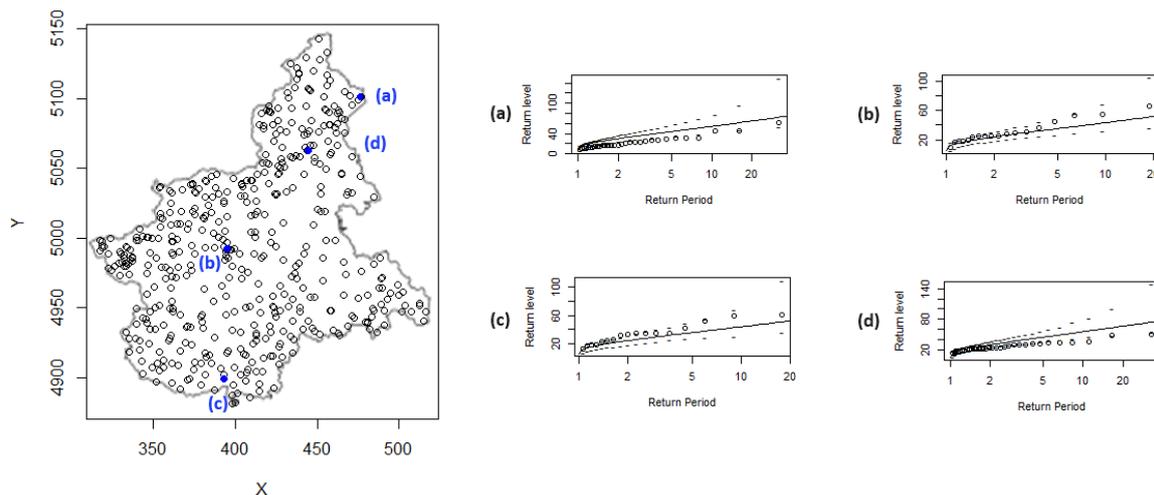


Figura 6.4: Valutazione locale del modello max-stable Brown-Resnick per i dati di precipitazione di durata 1 ora e per l'intero periodo di osservazione (1928-2020), tramite *return level plots* (a, b, c, d) per 4 stazioni pluviometriche.

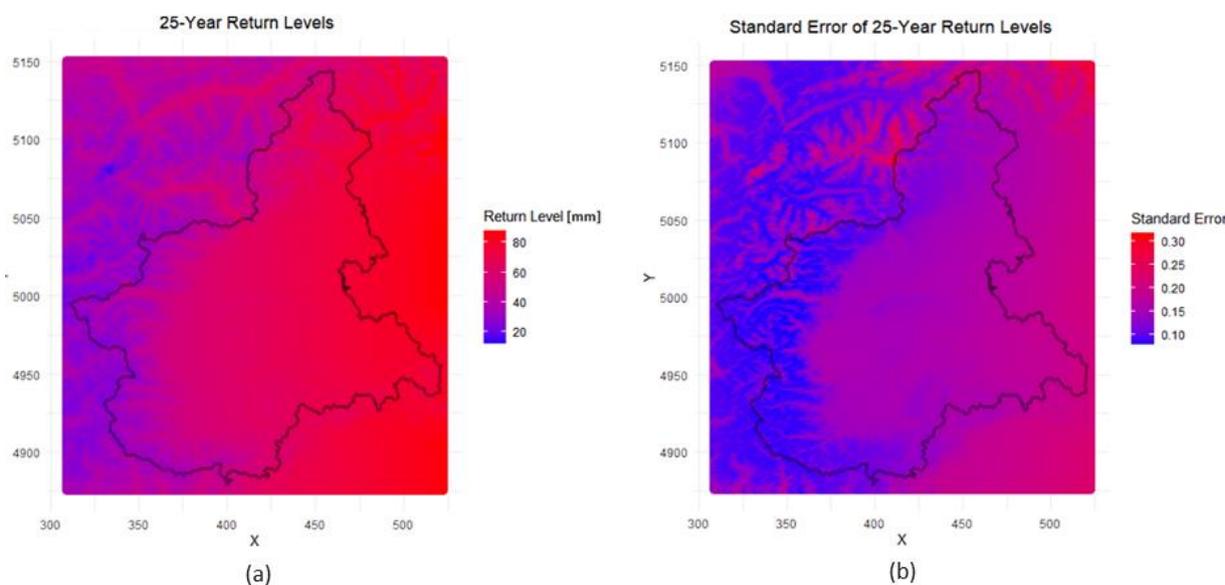


Figura 6.5: 25-year return levels (a) ed errori standard (b) per l'intero territorio del Piemonte e le aree circostanti per precipitazioni di durata 1 ora e l'intero periodo di osservazione (1928-2020).

6.2.2 Anni 1960-2020

Escludendo le stazioni meteorologiche per cui non è stato registrato alcun valore oppure è stato registrato solamente un valore nell'intero periodo di osservazione, l'analisi è stata condotta sulle rimanenti 487 località. La distribuzione dell'altitudine complessiva non è variata significativamente rispetto al dataset completo. Infatti, dalla Figura A.1 si nota che la maggior parte delle stazioni pluviometriche ha un'altitudine inferiore ai 1000 metri e pochissime stazioni si trovano ad una quota superiore ai 2500 metri. Le statistiche descrittive per i valori di precipitazione di durata 1 ora nel periodo temporale considerato (1960-2020), indipendentemente dalla stazione pluviometrica, sono presentate nella Tabella 6.4 e si discostano leggermente dai valori calcolati sull'intero periodo temporale.

Il pannello a sinistra della Figura 6.6 mostra la deviazione della media campionaria di ciascuna stazione pluviometrica rispetto al valore di media complessiva pari a 28.26 mm, mentre il pannello a

destra mostra la deviazione standard di ciascuna località. I valori osservati sono simili a quelli registrati per l'intero periodo (1928-2020), dunque valgono le stesse considerazioni fatte con la Figura 6.1.

La stima empirica della dipendenza spaziale tra le stazioni pluviometriche per i dati di precipitazione di durata 1 ora è molto bassa, prossima all'indipendenza, già a basse distanze tra due stazioni pluviometriche, come mostrato dalla Figura 6.7. Nello specifico, il pannello (a) mostra la stima del coefficiente estremo calcolata per ciascuna coppia di località. Osservando nel dettaglio la versione discretizzata nel pannello (b), si può notare che i punti stanno nel range di valori tra 1.7 circa a 2.2 circa per ogni distanza h , con valore intermedio pari a circa 1.9 (linea rossa orizzontale) già a 50 km di distanza tra due stazioni pluviometriche. Il pannello (c) mostra, invece, la stima della probabilità di concomitanza estrema calcolata per ciascuna coppia di località. Osservando i punti neri, si può notare che essi stanno nel range di valori tra 0.0 e 0.3 circa per quasi ogni distanza. Ad esempio, a 50 km di distanza tra due stazioni pluviometriche la stima della probabilità di concomitanza estrema è circa 0.15. In generale, la dipendenza spaziale stimata è simile a quella osservata sull'intero dataset nella Figura 6.2.

I risultati ottenuti con cross-validation sono presentati nella Tabella A.5 e mostrano che, in generale, il modello Schlather ha prestazioni inferiori, il modello Brown-Resnick si comporta meglio e il modello Extremal-t offre le migliori performance, per qualsiasi combinazione di covariate. In particolare, il modello con errore medio di predizione minore, considerando RMSE, è il modello Extremal-t e con covariate per il parametro *scale* date da latitudine e altitudine. Le stime dei parametri per le *trend surfaces* e per i valori di *range* λ , *smooth* κ e gradi di libertà (DoF) ν , e relativi errori standard del modello selezionato, addestrato sull'intero dataset di precipitazioni di durata 1 ora nel periodo 1960-2020, sono riportati in (6.2). Il valore di TIC del modello, invece, è pari a 14458237.

$$\begin{aligned}
 \mu(x) &= -89.14393_{(2.32732)} + 0.04412_{(0.00051)}lon(x) + \\
 &\quad + 0.01941_{(0.00049)}lat(x) - 3.73869_{(0.03683)}alt(x), \\
 \sigma(x) &= -45.51668_{(0.98092)} + 0.01147_{(0.00020)}lat(x) - 2.55753_{(0.02302)}alt(x), \\
 \xi(x) &= 0.09426_{(0.00146)}, \\
 \lambda &= 0.03472_{(0.04937)}, \\
 \kappa &= 0.12909_{(0.02704)}, \\
 \nu &= 5.40371_{(0.13339)}.
 \end{aligned} \tag{6.2}$$

Dai valori in (6.2), si può notare che il parametro di *location* aumenta in base ai valori di longitudine e latitudine ma diminuisce con un peso maggiore in base al valore di altitudine. Il parametro di *scale* aumenta in base al valore di latitudine e diminuisce con peso maggiore rispetto al valore di altitudine, pur mantenendo dei valori positivi in ogni località. Infine, il parametro di *shape* è positivo, seppur piccolo, suggerendo una distribuzione marginale di tipo Fréchet. I valori di *range* e *smooth* sono bassi, suggerendo una dipendenza spaziale tra le stazioni pluviometriche molto bassa. Gli errori standard sono bassi per (quasi) tutti i coefficienti, indicando una bassa incertezza delle stime.

<i>Min</i>	<i>1° quartile</i>	<i>Mediana</i>	<i>Media</i>	<i>3° quartile</i>	<i>Max</i>	<i>Deviazione standard</i>
4.00	17.80	25.50	28.26	35.50	135.60	14.28604

Tabella 6.4: Statistiche descrittive (minimo, primo quartile, mediana, media, terzo quartile, massimo e deviazione standard) per le precipitazioni di durata 1 ora e periodo di osservazione 1960-2020.

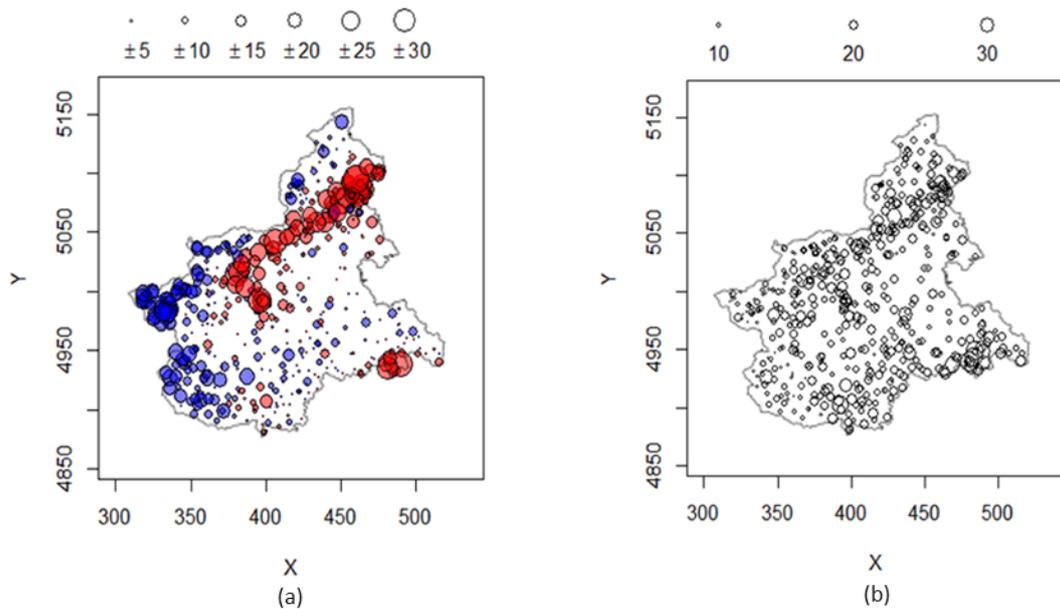


Figura 6.6: *Symbol plot* che mostra per ciascuna stazione pluviometrica la deviazione della media campionaria locale risetto alla media campionaria complessiva (a) e la deviazione standard (b) per le precipitazioni di durata 1 ora e periodo di osservazione 1960-2020.

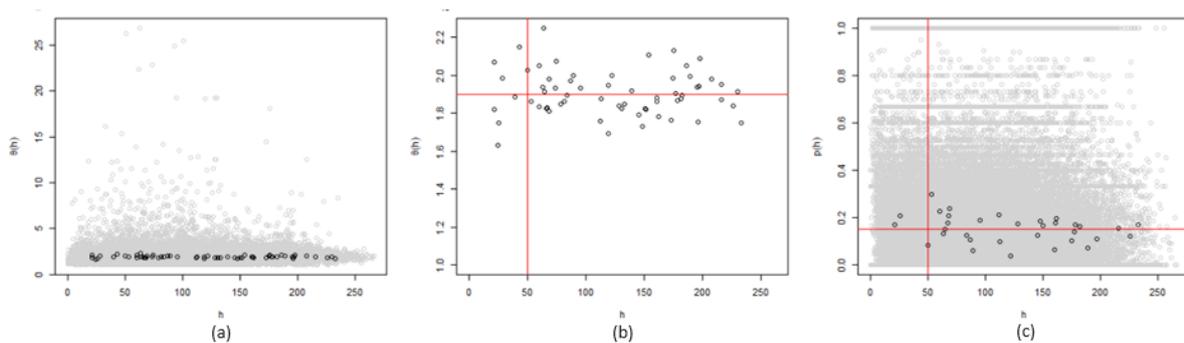


Figura 6.7: Dipendenza spaziale stimata empiricamente dai dati di precipitazione di durata 1 ora per il periodo di osservazione 1960-2020, tramite coefficiente estremo (a, b) e probabilità di concomitanza estrema (c). I punti grigi sono le stime empiriche, mentre i punti neri sono le stime discretizzate con 10000 *bins*, con ingrandimento sulla versione discretizzata per il coefficiente estremo (b).

La Figura 6.8 mostra la funzione del coefficiente estremo e la funzione di probabilità di concomitanza estrema stimate dal modello Extremal-t. In particolare, nel caso della funzione del coefficiente estremo si nota una crescita repentina a bassissime distanze tra coppie di stazioni pluviometriche fino ad assumere un valore quasi costante intorno a 1.9, valore osservato anche nelle stime empiriche. Analogamente, la funzione di probabilità di concomitanza estrema stimata dal modello decresce repentinamente a bassissime distanze fino ad assumere un valore quasi costante prossima a 0. Tuttavia, la funzione stimata assume valori inferiori di quelli osservati, assumendo una maggiore indipendenza tra le stazioni pluviometriche. Dunque, la dipendenza spaziale del modello viene catturata in modo incorretto specialmente a basse distanze, in quanto viene assunta una dipendenza molto bassa, prossima all'indipendenza, anche tra località relativamente vicine. I valori stimati dal modello delle due funzioni per alcuni valori di distanza h sono riportati nella Tabella 6.5.

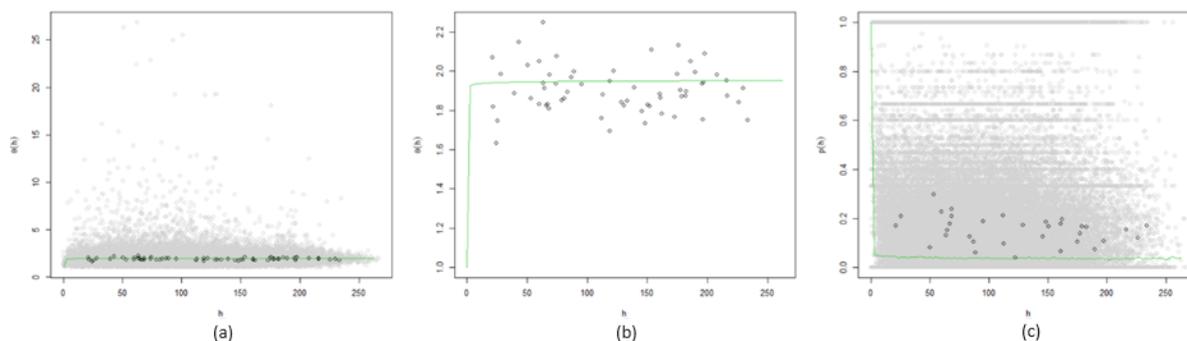


Figura 6.8: Valutazione della dipendenza spaziale stimata dal modello Extremal-t per le precipitazioni di durata 1 ora per il periodo di osservazione 1960-2020, tramite funzione del coefficiente estremo (a, b), e funzione di probabilità di concomitanza estrema (c). I punti grigi sono le stime empiriche calcolate rispetto alle osservazioni e i punti neri rappresentano la versione discretizzata.

Funzione	Distanza	Extremal-t
Coefficiente estremo	$h = 5 \text{ km}$	1.930422
	$h = 20 \text{ km}$	1.940078
	$h = 50 \text{ km}$	1.944924
	$h = 100 \text{ km}$	1.947883
Probabilità di concomitanza estrema	$h = 5 \text{ km}$	0.05710053
	$h = 20 \text{ km}$	0.04191922
	$h = 50 \text{ km}$	0.03942080
	$h = 100 \text{ km}$	0.03859814

Tabella 6.5: Valori delle stime delle funzioni del coefficiente estremo e di probabilità di concomitanza estrema in alcune distanze per il modello Extremal-t selezionato nel caso di precipitazioni di durata 1 ora e il periodo di osservazione 1960-2020.

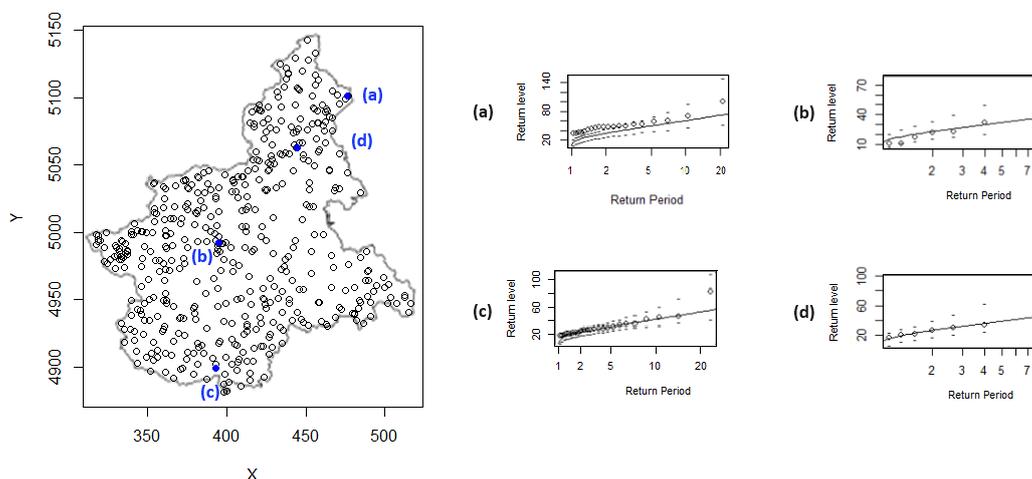


Figura 6.9: Valutazione locale del modello max-stable Extremal-t per i dati di precipitazione di durata 1 ora e per il periodo di osservazione 1960-2020, tramite *return level plots* (a, b, c, d) per 4 stazioni pluviometriche.

L'analisi locale del modello max-stable Extremal-t selezionato è stata effettuata per le stesse quattro stazioni pluviometriche considerate nella Figura 6.4 e riportate per comodità nel pannello a sinistra della Figura 6.9. I quattro *return level plots* mostrano che il modello è in parte coerente con i dati osservati in alcune località, come nei casi mostrati dai pannelli (b, c, d). Tuttavia, in alcuni siti il modello sottostima la frequenza degli eventi estremi, come nel caso mostrato dal pannello (a).

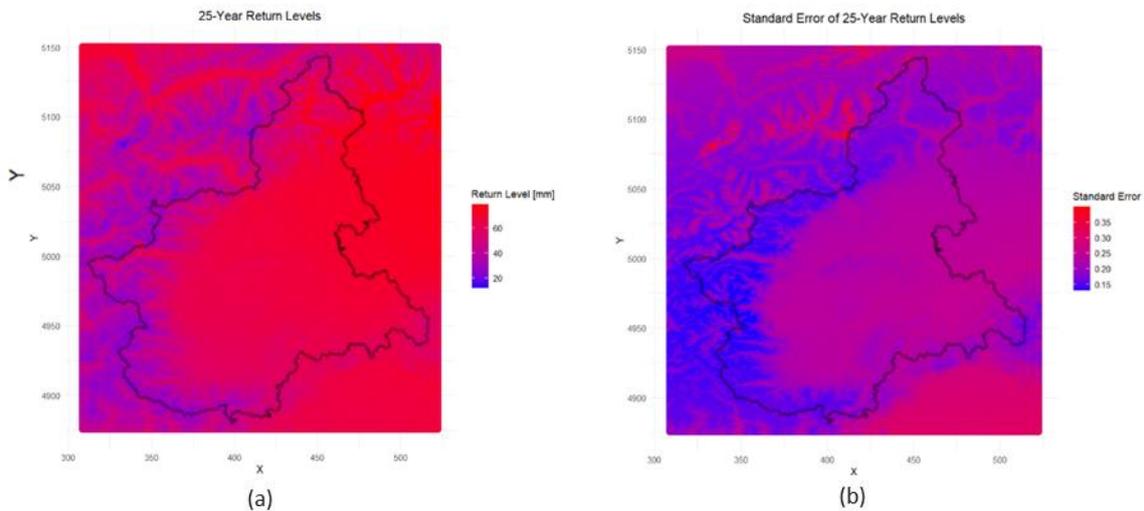


Figura 6.10: *25-year return levels* (a) ed errori standard (b) per l'intero territorio del Piemonte e le aree circostanti per precipitazioni di durata 1 ora e il periodo di osservazione 1960-2020.

Infine, analogamente alla prima metodologia, è stato predetto dal modello Extremal-t il *25-year return level* in tutta la regione e le zone circostanti. I valori predetti e i relativi errori standard sono mostrati in Figura 6.10. In particolare, si può notare che nelle zone alpine ad elevata altitudine, i valori predetti di precipitazione non superano in media 30 mm e gli errori standard corrispondenti sono bassi (circa 0.15, tranne a nord-ovest dove sono circa 0.40). Nelle zone montuose ad altitudini intermedie, come gli Appennini e le Alpi a quota intermedia, i valori predetti di precipitazione non superano in media il range di valori tra 30 e 40 mm. Gli errori standard in tali aree sono in genere relativamente bassi e inferiori a 0.20. Infine, nelle zone collinari e pianeggianti, i valori predetti di precipitazione non superano in media 70 mm e a nord-est 80 mm, e gli errori standard sono compresi tra 0.20 e 0.30 circa. Di conseguenza, dal momento che ovunque gli errori standard non sono molto elevati, l'incertezza dei valori predetti dei *25-year return levels* è bassa. In particolare, le aree ad altitudini elevate rispecchiano quanto osservato dai dati registrati nell'intero periodo di osservazione e mostrato nella Figura 6.6. Tuttavia, i risultati ottenuti assumono che le aree principalmente interessate da eventi estremi di breve durata siano quelle pianeggianti e collinari a nord-est del Piemonte, ovvero una zona più estesa di quella osservata nella Figura 6.6. Invece, in base ai risultati ottenuti, si prevede che le aree montuose a quote intermedie siano soggette a precipitazioni di circa 40 mm, leggermente superiori rispetto alla media osservata.

6.2.3 Anni 2000-2020

Escludendo le stazioni meteorologiche per cui non è stato registrato alcun valore oppure è stato registrato solamente un valore nell'intero periodo di osservazione, l'analisi è stata condotta sulle rimanenti 317 località. La distribuzione dell'altitudine complessiva non è variata significativamente rispetto al dataset completo. Infatti, dalla Figura A.2 si nota che la maggior parte delle stazioni pluviometriche ha un'altitudine inferiore a 800 metri e pochissime stazioni si trovano ad una quota superiore ai 2500 metri. Le statistiche descrittive per i valori di precipitazione di durata 1 ora nel periodo temporale considerato (2000-2020), indipendentemente dalla stazione pluviometrica, sono presentate nella Tabella 6.6. In particolare, il range di valori assunti dalle precipitazioni negli ultimi 21 anni di osservazione si è leggermente ridotto, ovvero il valore minimo è superiore e il valore massimo è leggermente inferiore rispetto ai valori calcolati sull'intero periodo temporale (1928-2020). Di conseguenza, sono aumentati anche le statistiche descrittive (primo quartile, mediana, media e terzo quartile), mentre la deviazione standard è aumentata leggermente.

<i>Min</i>	<i>1° quartile</i>	<i>Mediana</i>	<i>Media</i>	<i>3° quartile</i>	<i>Max</i>	<i>Deviazione standard</i>
6.40	20.00	28.00	30.79	38.60	135.10	14.99021

Tabella 6.6: Statistiche descrittive (minimo, primo quartile, mediana, media, terzo quartile, massimo e deviazione standard) per le precipitazioni di durata 1 ora e periodo di osservazione 2000-2020.

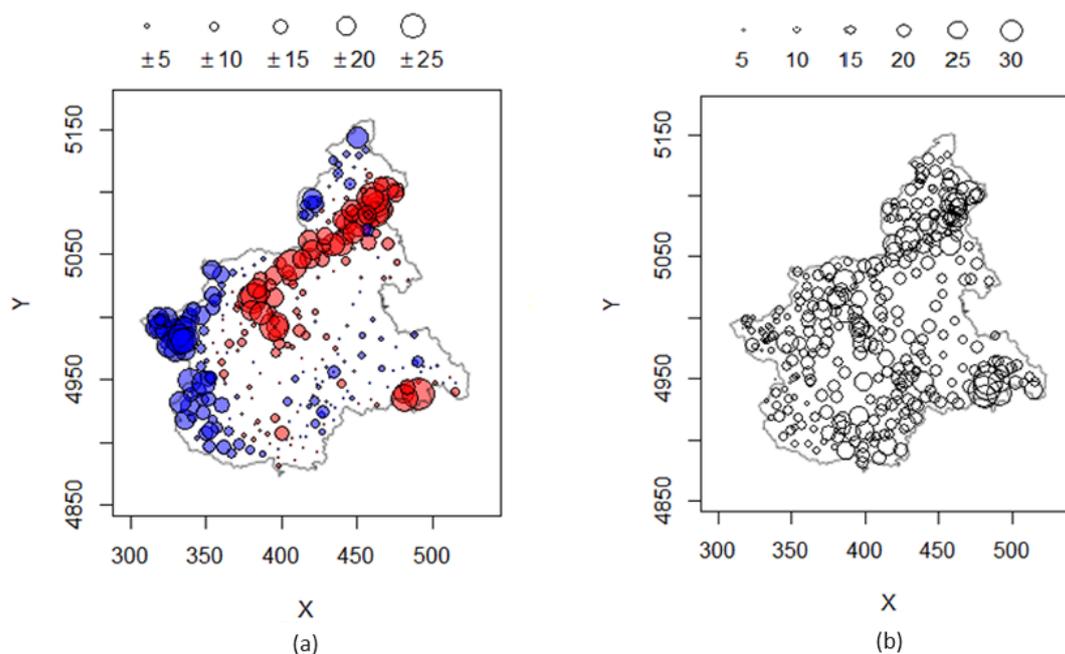


Figura 6.11: *Symbol plot* che mostra per ciascuna stazione pluviometrica la deviazione della media campionaria locale risetto alla media campionaria complessiva (a) e la deviazione standard (b) per le precipitazioni di durata 1 ora e periodo di osservazione 2000-2020.

Il pannello a sinistra della Figura 6.11 mostra la deviazione della media campionaria di ciascuna stazione pluviometrica rispetto al valore di media complessiva pari a 30.79 mm, mentre il pannello a destra mostra la deviazione standard di ciascuna località. I valori osservati sono simili a quelli registrati per l'intero periodo (1928-2020), dunque valgono le stesse considerazioni fatte con la Figura 6.1.

La stima empirica della dipendenza spaziale tra le stazioni pluviometriche per i dati di precipitazione di durata 1 ora è molto bassa, prossima all'indipendenza, già a basse distanze tra due stazioni pluviometriche, come mostrato dalla Figura 6.12. Nello specifico, osservando nel dettaglio il pannello (b), si può notare che i punti relativi al coefficiente estremo stanno nel range di valori tra 1.7 e 1.9 circa (linea rossa orizzontale) per distanze h tra due località inferiori ai 50 km, e tra 1.8 e 2.0 per distanze maggiori. Osservando invece le stime della probabilità di concomitanza estrema nel pannello (c), si può notare che i punti neri stanno nel range di valori tra 0.1 e 0.2 circa per distanze minori di 50 km e si stabilizzano intorno a 0.1 a distanze maggiori. Rispetto alla Figura 6.2, i punti della versione discretizzata sono meno dispersi e si può notare un andamento evidente della crescita e decrescita rispettivamente delle stime delle funzioni di coefficiente estremo e di probabilità di concomitanza estrema.

Successivamente all'analisi esplorativa precedentemente esposta, sono state applicate entrambe le metodologie precedentemente presentate. Con la prima metodologia è stato selezionato il modello (M1) Brown-Resnick con *trend surfaces* costruite considerando longitudine, latitudine e altitudine come covariate per il parametro *location*, longitudine e altitudine per il parametro *scale* e nessuna per il parametro *shape*. Con la seconda metodologia è stato selezionato il modello (M2) Brown-Resnick con

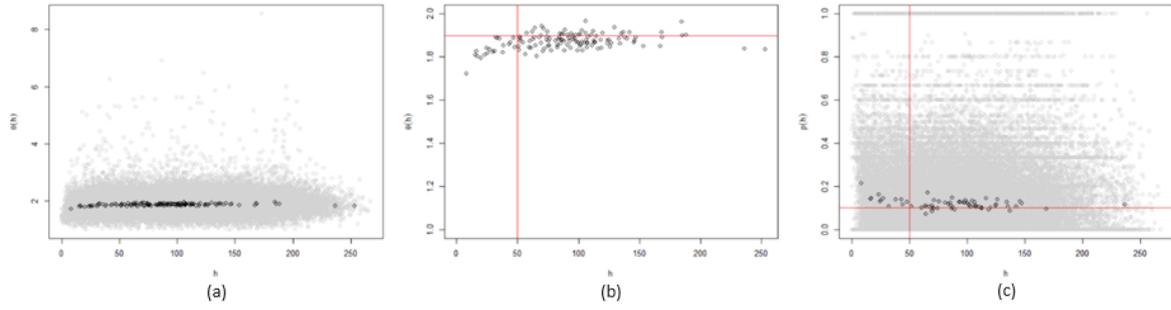


Figura 6.12: Dipendenza spaziale stimata empiricamente dai dati di precipitazione di durata 1 ora per il periodo di osservazione 2000-2020, tramite coefficiente estremo (a, b) e probabilità di concomitanza estrema (c). I punti grigi sono le stime empiriche, mentre i punti neri sono le stime discretizzate con 400 *bins*, con ingrandimento sulla versione discretizzata per il coefficiente estremo (b).

stesse *trend surfaces* per i parametri *location* e *shape*, mentre le covariate considerate per il parametro di *scale* sono latitudine e altitudine. Il valore di TIC del modello M1 è pari a 10412180 e del modello M2 è pari a 10422778. Maggiori dettagli sui risultati di entrambe le metodologie sono esposti nell'Appendice A.

I parametri e gli errori standard stimati dal modello M1 per le *trend surfaces* e per i valori di *range* λ e *smooth* κ sono riportati in (6.3). Nello specifico, le stime dei parametri di *range* e *smooth* sono leggermente superiori rispetto ai valori ottenuti sull'intero dataset in (6.1), suggerendo maggiore dipendenza spaziale tra le stazioni pluviometriche vicine e maggiore indipendenza tra le stazioni lontane.

$$\begin{aligned}
\mu(x) &= -81.16980_{(17.63916)} + 0.05092_{(0.00334)}lon(x) + \\
&\quad + 0.01766_{(0.00361)}lat(x) - 4.51461_{(0.22976)}alt(x), \\
\sigma(x) &= 1.33257_{(1.12624)} + 0.02509_{(0.00250)}lon(x) - 1.89896_{(0.18934)}alt(x), \\
\xi(x) &= 0.09067_{(0.00962)}, \\
\lambda &= 1.3304_{(0.06734)}, \\
\kappa &= 0.6023_{(0.00561)}.
\end{aligned} \tag{6.3}$$

Invece, i parametri e gli errori standard stimati dal modello M2 per le *trend surfaces* e per i valori di *range* λ e *smooth* κ sono riportati in (6.4). Nello specifico, le stime dei parametri di *location* e *shape* sono abbastanza simili a quelle ottenute in (6.3). Inoltre, per entrambi i modelli l'altitudine ha un peso maggiore ed è di segno opposto rispetto alle altre covariate, mentre le stime dei parametri di *range* e *smooth* sono molto simili.

$$\begin{aligned}
\mu(x) &= -96.36977_{(16.49754)} + 0.03289_{(0.00301)}lon(x) + \\
&\quad + 0.02222_{(0.003398)}lat(x) - 4.74677_{(0.22724)}alt(x), \\
\sigma(x) &= -32.72847_{(6.24939)} + 0.00900_{(0.00124)}lat(x) - 2.57733_{(0.14678)}alt(x), \\
\xi(x) &= 0.08696_{(0.00985)}, \\
\lambda &= 1.3125_{(0.06337)}, \\
\kappa &= 0.6145_{(0.00631)}.
\end{aligned} \tag{6.4}$$

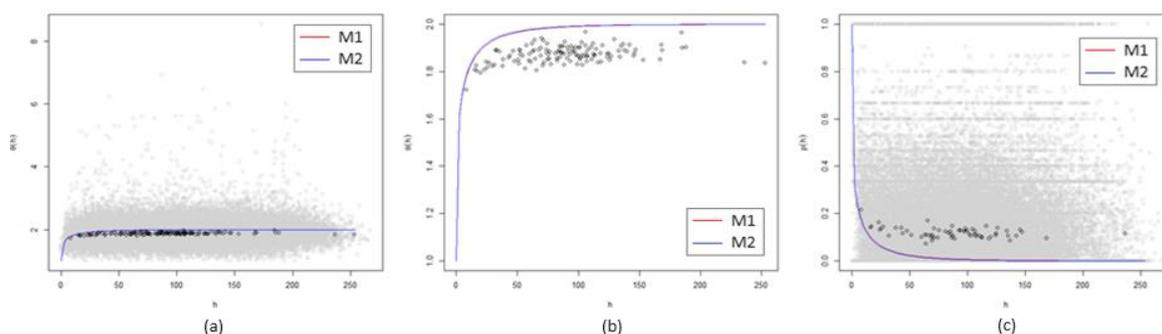


Figura 6.13: Dipendenza spaziale stimata dai modelli M1 e M2 per le precipitazioni di durata 1 ora per il periodo di osservazione 2000-2020, tramite funzione del coefficiente estremo (a, b), e funzione di probabilità di concomitanza estrema (c). I punti grigi sono le stime empiriche calcolate rispetto alle osservazioni e i punti neri rappresentano la versione discretizzata.

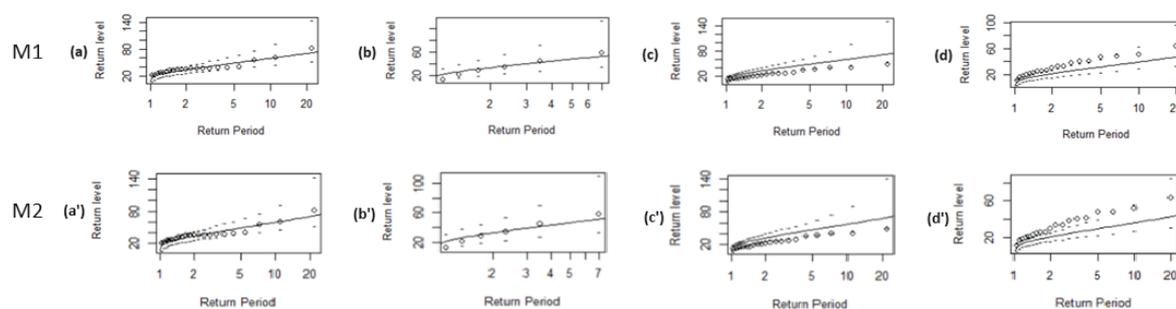


Figura 6.14: Valutazione locale dei modelli max-stable M1 e M2 selezionati per i dati di precipitazione di durata 1 ora e per il periodo di osservazione 2000-2020, tramite *return level plots* (a, b, c, d) per M1 e (a', b', c', d') per M2 per 4 stazioni pluviometriche.

La Figura 6.13 mostra le funzioni del coefficiente estremo e di probabilità di concomitanza estrema stimate dai modelli M1 e M2. In particolare, si nota che le due funzioni sono quasi sovrapposte per entrambi i modelli, come previsto dalle stime dei coefficienti, e assumono maggiore indipendenza rispetto a quella stimata empiricamente dai dati. L'andamento delle due funzioni è repentino a bassissime distanze fino a raggiungere il valore di 1.7 circa per la funzione del coefficiente estremo e di 0.2 circa per la funzione di concomitanza estrema, per poi proseguire in maniera graduale sino a divenire prossimo all'indipendenza a circa 100 km di distanza tra due stazioni pluviometriche. I valori ottenuti dalle due funzioni per alcuni valori di distanza sono riportati nella Tabella A.3 per il modello M1 e nella Tabella A.6 per il modello M2.

L'analisi locale di entrambi i modelli, mostrata nella Figura 6.14, è stata effettuata per le stesse quattro stazioni pluviometriche considerate nella Figura 6.4 e nella Figura 6.9. Gli otto *return level plots* mostrano che sia il modello M1 sia il modello M2 sono in parte coerenti con i dati osservati in alcune località, come nei casi mostrati dai pannelli (a, b, a', b'). Tuttavia, in alcuni siti i modelli sovrastima la frequenza degli eventi estremi, come nei casi mostrati dai pannelli (c, c'), mentre in altre stazioni i modelli sottostimano la frequenza degli eventi estremi, come nei casi mostrati dai pannelli (d, d').

Infine, è stato predetto dal modello M1 il *25-year return level* in tutta la regione e le zone circostanti. I valori predetti e i relativi errori standard sono mostrati in Figura 6.15. In particolare, si può notare che nelle zone alpine ad elevata altitudine, i valori predetti di precipitazione non superano in media 20 mm e gli errori standard corrispondenti sono generalmente bassi (circa 1.0), tranne a nord, esternamente al Piemonte (circa 3.0). Nelle zone montuose ad altitudini intermedie, come gli Appennini e le Alpi a quota intermedia, i valori predetti di precipitazione non superano in media il range di valori tra 30 e 40 mm.

Gli errori standard in tali aree sono in genere relativamente bassi e inferiori a 1.5. Infine, nelle zone collinari e pianeggianti al centro-sud, i valori predetti di precipitazione non superano in media 60 mm, mentre a nord e ad est i valori predetti di precipitazione possono arrivare fino a poco sopra 80 mm, con zone particolarmente interessate ad eventi così estremi esteriormente al Piemonte. In generale, gli errori standard dei valori predetti di precipitazione nelle zone collinari e pianeggianti sono intorno a 1.5 all'interno della regione e 2.0 esternamente. In generale, si notano errori standard maggiori di un ordine di grandezza rispetto a quanto precedentemente osservato nei risultati relativi all'intero dataset. Tuttavia, i valori sono abbastanza inferiori rispetto ai valori predetti di precipitazione. Di conseguenza, l'incertezza dei valori predetti dei *25-year return levels* è relativamente bassa. In particolare, le aree ad altitudini elevate rispecchiano quanto osservato dai dati registrati nell'intero periodo di osservazione e mostrato nella Figura 6.11. In modo analogo, i risultati ottenuti assumono che le aree principalmente interessate da eventi estremi di breve durata siano nelle zone collinari a nord-est del Piemonte. Tuttavia, si assume che le restanti aree pianeggianti e collinari e le aree montuose a quote intermedie siano soggette a precipitazioni di circa 40-70 mm, più estreme di quelle osservate nella Figura 6.11.

Analogamente alla prima metodologia, è stato predetto dal modello M2 il *25-year return level* in tutta la regione e le zone circostanti. I valori predetti e i relativi errori standard sono mostrati in Figura 6.16. In particolare, si può notare che nelle zone alpine ad elevata altitudine, i valori predetti di precipitazione non superano in media 30 mm e gli errori standard corrispondenti sono bassi (inferiori a 1.0, tranne a nord-ovest dove sono circa 2.5). Nelle zone montuose ad altitudini intermedie, come gli Appennini e le Alpi a nord-est, i valori predetti di precipitazione non superano in media il range di valori tra 30 e 50 mm. Gli errori standard in tali aree sono in genere relativamente bassi e inferiori a 1.25. Infine, nelle zone collinari e pianeggianti a sud, i valori predetti di precipitazione non superano in media 60 mm e a nord-est 80 mm, e gli errori standard sono 1.5 circa e 2.0 a sud esternamente al Piemonte. Per quanto riguarda gli errori standard, valgono le stesse considerazioni fatte con il modello M1. Dunque, le aree ad altitudini elevate rispecchiano quanto osservato dai dati registrati nell'intero periodo di osservazione e mostrato nella Figura 6.11. Tuttavia, i risultati ottenuti assumono che le aree principalmente interessate da eventi estremi di breve durata siano quelle pianeggianti e collinari a nord-est del Piemonte, ovvero una zona più estesa di quella osservata nella Figura 6.11. Invece, in base ai risultati ottenuti, si prevede che le aree montuose a quote intermedie siano soggette a precipitazioni di circa 40 mm, leggermente superiori rispetto alla media osservata.

In definitiva, i modelli M1 e M2 sono molto simili sia per quanto riguarda la dipendenza spaziale catturata sia per quanto riguarda il comportamento nelle singole stazioni pluviometriche. Inoltre, anche le mappe spaziali dei *25-year return levels* sono abbastanza simili e, per il modello M2 il range di valori relativi agli errori standard sono leggermente inferiori rispetto a M1.

6.3 Precipitazioni massime annuali di durata 24 ore

6.3.1 Dataset completo (1928-2020)

Escludendo le stazioni meteorologiche per cui non è stato registrato alcun valore oppure è stato registrato solamente un valore nell'intero periodo di osservazione, l'analisi è stata condotta sulle rimanenti 493 località. Dal momento che solo cinque stazioni meteorologiche sono state escluse, la distribuzione dell'altitudine complessiva non è variata significativamente rispetto alla Figura 5.2. In modo analogo, le statistiche descrittive sono pressoché inalterate rispetto a quelle nella Tabella 5.2. Dal momento che il range di valori e, di conseguenza, il valore medio e la deviazione standard relativi ai dati di precipitazione di durata 24 ore sono notevolmente maggiori rispetto ai dati di precipitazione

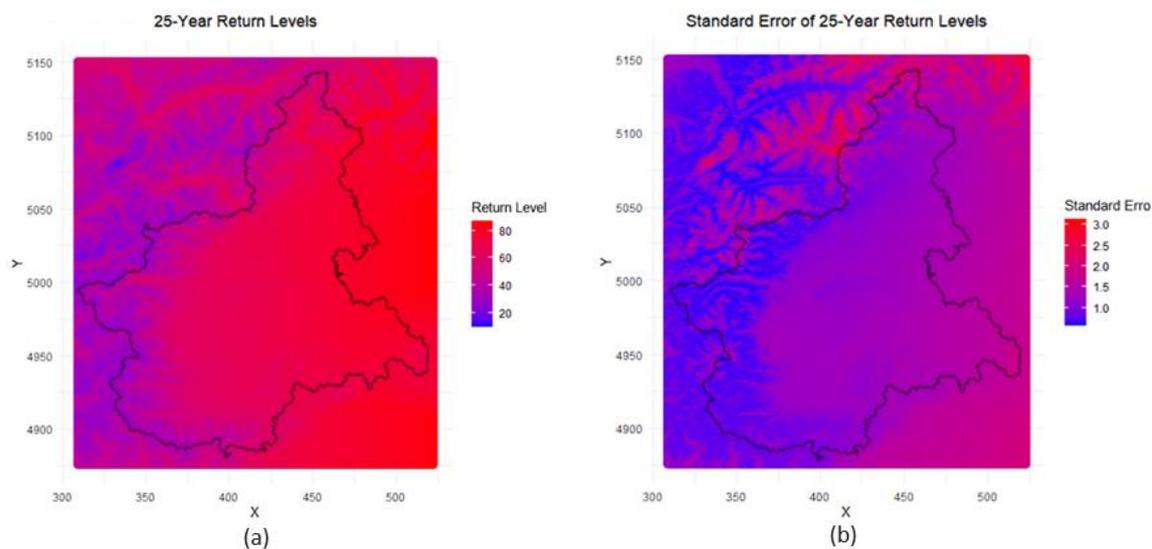


Figura 6.15: 25-year return levels (a) ed errori standard (b) per l'intero territorio del Piemonte e le aree circostanti per precipitazioni di durata 1 ora e il periodo di osservazione 2000-2020 predetti dal modello M1.

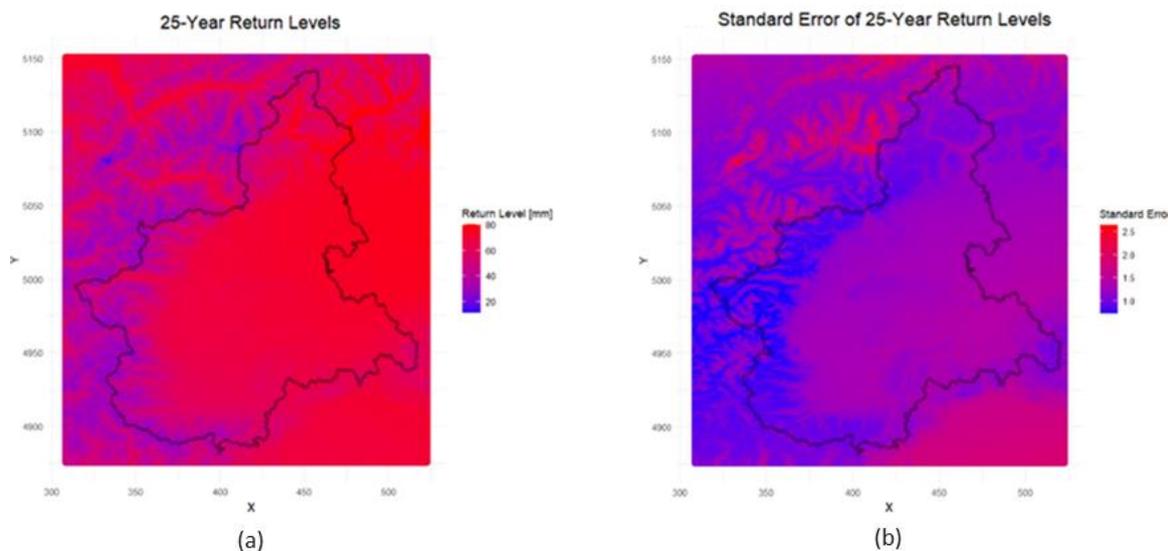


Figura 6.16: 25-year return levels (a) ed errori standard (b) per l'intero territorio del Piemonte e le aree circostanti per precipitazioni di durata 1 ora e il periodo di osservazione 2000-2020 predetti dal modello M2.

di durata 1 ora, ci si aspettano valori maggiori dei parametri della distribuzione GEV.

Il pannello a sinistra della Figura 6.17 mostra la deviazione della media campionaria di ciascuna stazione pluviometrica rispetto al valore di media complessiva pari a 101.6 mm, mentre il pannello a destra mostra la deviazione standard di ciascuna località. Si può osservare che le zone caratterizzate da valori medi significativamente inferiori rispetto alla media complessiva, rappresentati da cerchi blu più ampi, corrispondono alle aree alpine a ovest (Val di Susa) e in parte le zone pianeggianti e collinari al centro del Piemonte. Inoltre, in queste località le deviazioni standard dei valori di precipitazione sono relativamente basse. Di conseguenza, il valore medio calcolato è rappresentativo del valore di precipitazioni di durata prolungata registrate in queste zone. Dunque, ci si aspetta che tali aree non siano particolarmente interessate a fenomeni di precipitazione intensi e prolungati, dal momento che presentano generalmente quantitativi minimi di precipitazione per l'intero periodo di osservazione (1928-2020). Al contrario, le zone caratterizzate da valori medi significativamente maggiori rispetto alla

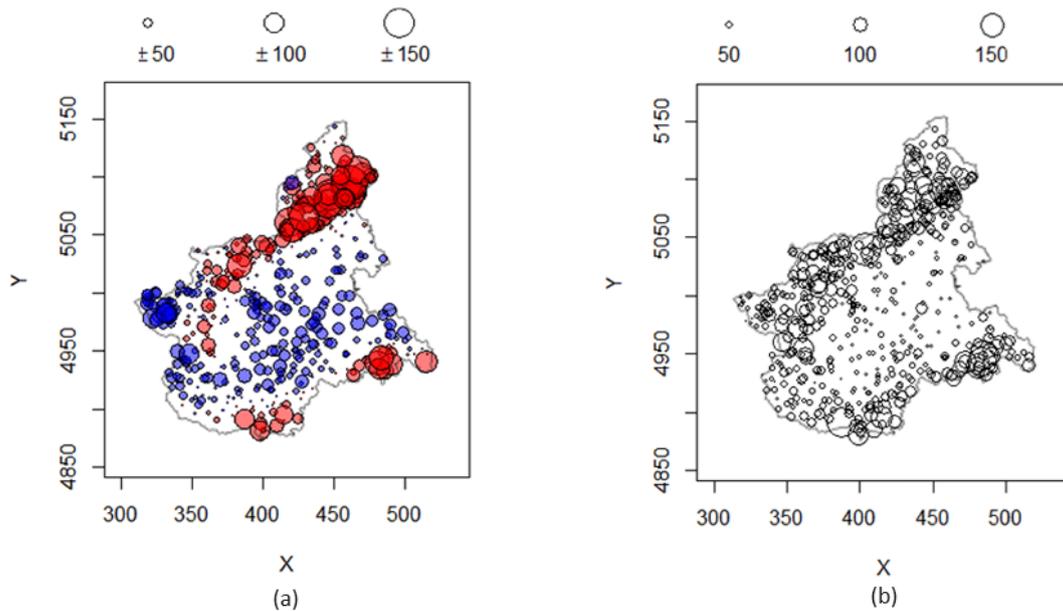


Figura 6.17: *Symbol plot* che mostra, per ciascuna stazione pluviometrica, la deviazione della media campionaria locale rispetto alla media campionaria complessiva (a) e la deviazione standard (b) per le precipitazioni di durata 24 ore e l'intero periodo di osservazione (1928-2020).

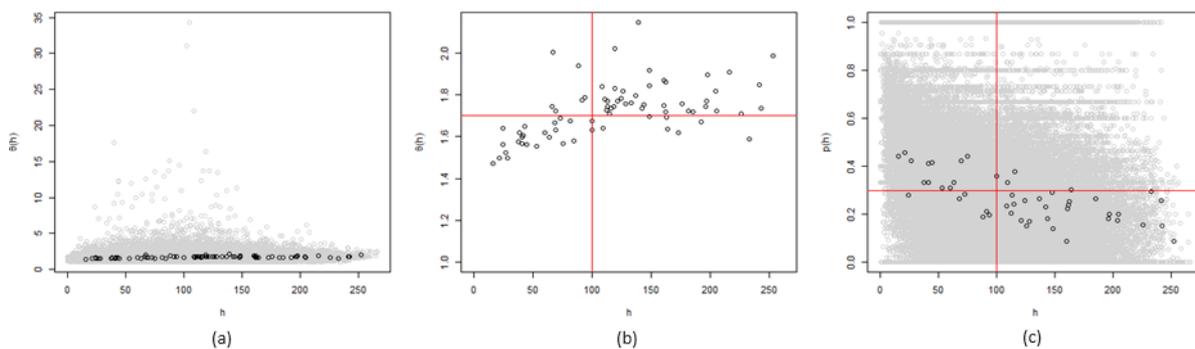


Figura 6.18: Dipendenza spaziale stimata empiricamente dai dati di precipitazione di durata 24 ore per l'intero periodo di osservazione (1928-2020), tramite coefficiente estremo (a, b) e probabilità di concomitanza estrema (c). I punti grigi sono le stime empiriche, mentre i punti neri sono le stime discretizzate con 11000 *bins*, con ingrandimento sulla versione discretizzata per il coefficiente estremo (b).

media complessiva, rappresentati da cerchi rossi più ampi, corrispondono alle località montuose a nord e a sud del Piemonte, coprendo un'area più estesa di quella osservata con i dati di precipitazione di durata 1 ora. Tali località sono, tuttavia, caratterizzate da deviazioni standard elevate. Di conseguenza, queste zone sono soggette a precipitazioni di lunga durata abbondanti alternate ad anni in cui le precipitazioni sono più moderate. Dunque, ci si può aspettare che tali aree siano particolarmente soggette a fenomeni estremamente intensi e prolungati. Infine, le zone caratterizzate da valori simili alla media complessiva sono in numero limitato e sparpagliate in diverse aree della regione. Per tali zone, le deviazioni standard sono relativamente basse. Dunque, queste località sono soggette a precipitazioni di durata giornaliera di intensità perlopiù pari a circa 100 mm e ci si aspetta che in tali aree si registrino valori simili di precipitazioni di durata 24 ore. L'analisi della Figura 6.17 suggerisce dunque l'utilizzo delle coordinate geografiche delle stazioni pluviometriche come covariate dei parametri della distribuzione GEV e in particolar modo del parametro *location*, dal momento che è possibile distinguere aree interessate da eventi particolarmente estremi rispetto ad aree dove sono state registrate quantità

minori di precipitazione, e del parametro *scale*, al fine di modellare variazioni nella variabilità spaziale degli eventi estremi.

La stima empirica della dipendenza spaziale tra le stazioni pluviometriche per i dati di precipitazione di durata 24 ore è mostrata nella Figura 6.18. In particolare, si evince che la funzione del coefficiente estremo (probabilità di concomitanza estrema) stimata per le precipitazioni di lunga durata ha un andamento crescente (decescente) rispetto ai valori di distanza tra le stazioni pluviometriche, partendo da un valore di circa 1.45 (0.5 rispettivamente) e raggiungendo circa 1.7 (0.3 rispettivamente) a distanza di 100 km tra due località. Dunque, i dati di precipitazione di lunga durata mostrano una dipendenza spaziale più forte rispetto alle precipitazioni di breve durata precedentemente analizzate. Di conseguenza, ci si aspetta che le stime dei parametri di dipendenza siano di valore maggiore rispetto ai valori stimati per i dati di precipitazione di durata 1 ora.

In base ai risultati ottenuti per la selezione delle *trend surfaces* mostrati nella Tabella A.1, sono state utilizzate longitudine, latitudine e altitudine come covariate sia per il parametro *location* sia per il parametro *scale* e nessuna per il parametro *shape*. I valori stimati dal modello GEV spaziali sono presentati in (A.3). Il modello max-stable selezionato è quello di Brown-Resnick, come suggerito dal valore di TIC nella terza riga della Tabella A.2. La dipendenza spaziale stimata dal modello Brown-Resnick è molto simile a quella stimata dal modello Extremal-t, come mostrato dalla Figura 6.19. In particolare, si nota una diminuzione graduale della dipendenza spaziale tra le stazioni pluviometriche all'aumentare della distanza tra di esse. Tuttavia, entrambi i modelli assumono minore dipendenza rispetto a quella stimata empiricamente dai dati per qualsiasi distanza tra le località. Ciò risulta evidente anche dai valori delle funzioni del coefficiente estremo e di probabilità di concomitanza estrema riportati nella Tabella 6.7. Ad esempio, per due stazioni pluviometriche poste a distanza di 100 km viene stimato un valore per il coefficiente estremo pari a circa 1.87 e per la probabilità di concomitanza estrema pari a circa 0.1. Il modello Schlather, invece, non cattura correttamente la dipendenza spaziale in quanto assume un valore costante per qualsiasi distanza.

I parametri e gli errori standard stimati dal modello Brown-Resnick per le *trend surfaces* e per i valori di *range* λ e *smooth* κ sono riportati in (6.5). Come previsto, i coefficienti dei parametri della distribuzione GEV marginale sono in generale maggiori, in valore assoluto, rispetto alle stime relative ai dati di precipitazione di durata 1 ora. In particolare, si nota un peso notevolmente maggiore per il valore di altitudine sia nel parametro di *location* sia nel parametro di *scale*. Il parametro di *shape* evidenzia una distribuzione marginale di tipo Fréchet, anche in tale caso. Inoltre, anche le stime dei parametri relativi alla dipendenza spaziale, ovvero *range* e *smooth*, assumono valori maggiori rispetto al caso di precipitazioni di breve durata.

$$\begin{aligned}
\mu(x) &= -754.4482_{(6.4127)} + 0.1619_{(0.0025)}lon(x) + 0.1516_{(0.0012)}lat(x) + \\
&\quad + 11.6241_{(0.1557)}alt(x), \\
\sigma(x) &= -225.86439_{(6.52059)} + 0.11780_{(0.00188)}lon(x) + 0.04034_{(0.00123)}lat(x) + \\
&\quad + 11.50369_{(0.14542)}alt(x), \\
\xi(x) &= 0.1688_{(0.0024)}, \\
\lambda &= 6.3817_{(0.1479)}, \\
\kappa &= 0.5656_{(0.0048)}.
\end{aligned} \tag{6.5}$$

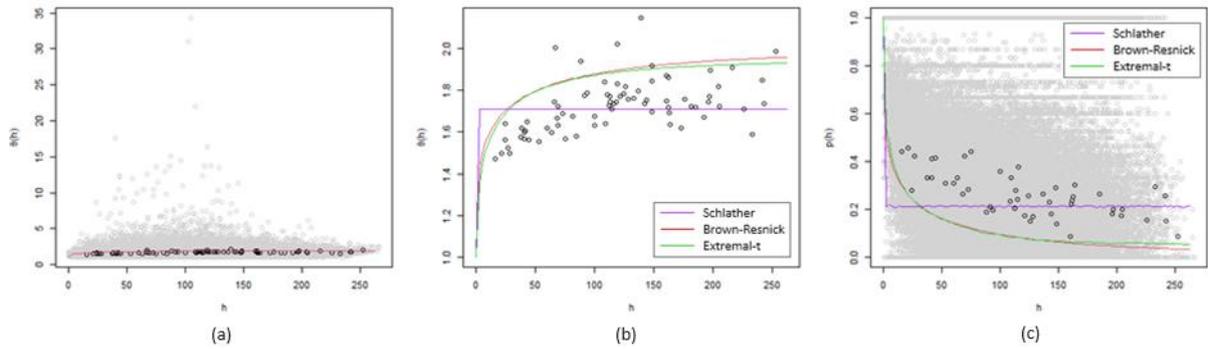


Figura 6.19: Valutazione della dipendenza spaziale stimata dai modelli max-stable per le precipitazioni di durata 24 ore per l'intero periodo di osservazione (1928-2020), tramite funzione del coefficiente estremo per il modello di Brown-Resnick (a) e per tutti i modelli (b), e funzione di probabilità di concomitanza estrema per ogni modello (c).

Funzione	Distanza	Schlather	Brown-Resnick	Extremal-t
Coefficiente estremo	$h = 20 \text{ km}$	1.707107	1.671313	1.653116
	$h = 50 \text{ km}$	1.707107	1.794381	1.792018
	$h = 100 \text{ km}$	1.707107	1.876395	1.870789
	$h = 200 \text{ km}$	1.707107	1.938972	1.917888
Probabilità di concomitanza estrema	$h = 20 \text{ km}$	0.5901797	0.26884037	0.30561805
	$h = 50 \text{ km}$	0.2159504	0.16600830	0.16219125
	$h = 100 \text{ km}$	0.2162411	0.09930096	0.09552852
	$h = 200 \text{ km}$	0.2164289	0.04773698	0.06154074

Tabella 6.7: Valori delle stime delle funzioni del coefficiente estremo e di probabilità di concomitanza estrema in alcune distanze per i modelli max-stable nel caso di precipitazioni di durata 24 ore e l'intero periodo di osservazione (1928-2020).

L'analisi locale del modello max-stable di Brown-Resnick selezionato è stata effettuata per le stesse quattro stazioni pluviometriche casuali considerate in precedenza e riportate per comodità nel pannello a sinistra della Figura 6.20. I quattro *return level plots* mostrano che il modello è in parte coerente con i dati osservati in alcune località, come nel caso mostrato dal pannello (b). Tuttavia, in alcuni siti il modello sottostima leggermente la frequenza degli eventi estremi, come nei casi mostrati dai pannelli (a, c, d).

Infine, in Figura 6.21 sono mostrati il *25-year return level* predetto dal modello Brown-Resnick in tutta la regione e le zone circostanti e i relativi errori standard. In particolare, si può notare che nelle zone alpine ad elevata altitudine, i valori predetti di precipitazione sono molto estremi, con valori di circa 500 mm a nord e 400 mm a ovest, e gli errori standard corrispondenti sono maggiori (circa 4 a nord e 3 a ovest). Di conseguenza, queste aree sono particolarmente interessate da fenomeni di precipitazione intensa e prolungata. Questo fenomeno era stato osservato nella Figura 6.17 specialmente nella zona settentrionale del Piemonte, ma per un'area più estesa. Nella catena montuosa degli Appennini, a sud della regione, invece, i valori predetti di precipitazione sono di circa 350 mm ed errori standard di circa 2.5, in linea con quanto osservato nella Figura 6.17. Le aree in cui sono stati predetti valori di precipitazione in media non superiori a 100 mm sono situate nelle zone collinari e montuose ad altitudini minori ad ovest del territorio considerato, con errori standard inferiori a circa 1.5. Tali aree, dunque, non sono particolarmente interessate da precipitazioni estreme e prolungate, rispecchiando quanto osservato nella Figura 6.17. Infine, le zone in cui sono stati predetti valori di precipitazione nel range tra 200 e 300 mm corrispondono alle aree pianeggianti, collinari e montuose ad altitudini minori

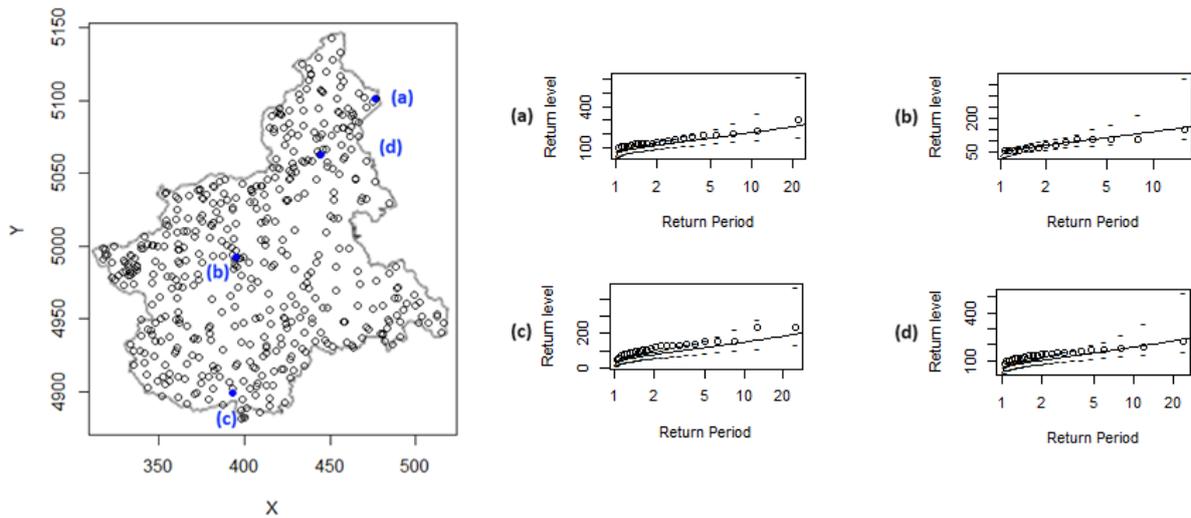


Figura 6.20: Valutazione locale del modello max-stable di Brown-Resnick per i dati di precipitazione di durata 24 ore e per l'intero periodo di osservazione (1928-2020), tramite *return level plots* (a, b, c, d) per 4 stazioni pluviometriche.

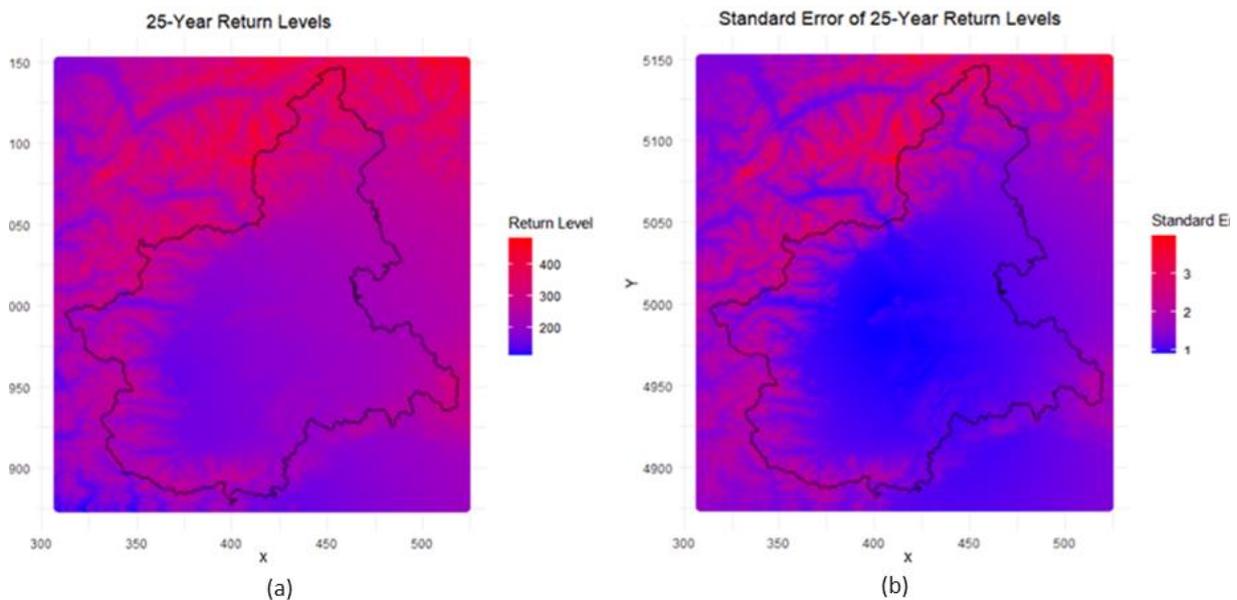


Figura 6.21: 25-year return levels (a) ed errori standard (b) per l'intero territorio del Piemonte e le aree circostanti per precipitazioni di durata 24 ore e l'intero periodo di osservazione (1928-2020).

a est della regione, con errori standard inferiori a 1.5. Per tali aree, i valori predetti sono in linea con quanto osservato nella Figura 6.17 nel caso delle aree pianeggianti e collinari. Tuttavia, nelle aree montuose a nord della regione, i valori predetti sono inferiori rispetto a quanto osservato dai dati, come era stato indicato anche dalla Figura 6.20.

6.3.2 Anni 1960-2020

Per il periodo di osservazione 1960-2020, l'analisi è stata condotta su 487 località. La distribuzione dell'altitudine complessiva non è variata significativamente rispetto al dataset completo, come mostrato dalla Figura A.3. Le statistiche descrittive per i valori di precipitazione di durata 24 ore nel periodo temporale considerato, indipendentemente dalla stazione pluviometrica, sono presentate nella Tabella 6.8 e si discostano leggermente dai valori calcolati sull'intero periodo temporale.

<i>Min</i>	<i>1° quartile</i>	<i>Mediana</i>	<i>Media</i>	<i>3° quartile</i>	<i>Max</i>	<i>Deviazione standard</i>
10.8	63.4	88.2	102.5	125.6	633.0	57.02535

Tabella 6.8: Statistiche descrittive (minimo, primo quartile, mediana, media, terzo quartile, massimo e deviazione standard) per le precipitazioni di durata 24 ore e periodo di osservazione 1960-2020.

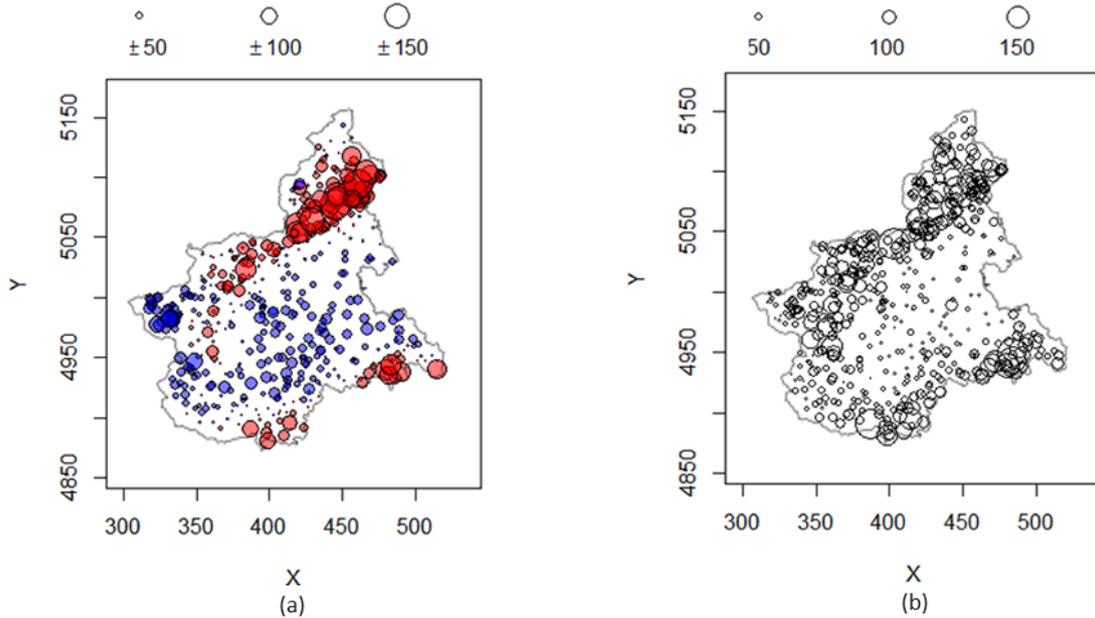


Figura 6.22: *Symbol plot* che mostra per ciascuna stazione pluviometrica la deviazione della media campionaria locale rispetto alla media campionaria complessiva (a) e la deviazione standard (b) per le precipitazioni di durata 24 ore e periodo di osservazione 1960-2020.

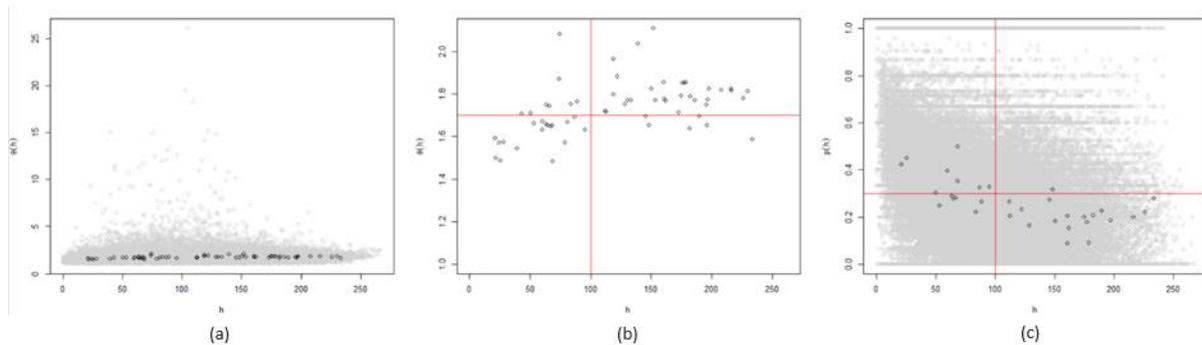


Figura 6.23: Dipendenza spaziale stimata empiricamente dai dati di precipitazione di durata 24 ore per il periodo di osservazione 1960-2020, tramite coefficiente estremo (a, b) e probabilità di concomitanza estrema (c). I punti grigi sono le stime empiriche, mentre i punti neri sono le stime discretizzate con 10000 *bins*, con ingrandimento sulla versione discretizzata per il coefficiente estremo (b).

Il pannello a sinistra della Figura 6.22 mostra la deviazione della media campionaria di ciascuna stazione pluviometrica rispetto al valore di media complessiva pari a 102.5 mm, mentre il pannello a destra mostra la deviazione standard di ciascuna località. I valori osservati sono simili a quelli registrati per l'intero periodo (1928-2020), dunque valgono le stesse considerazioni fatte con la Figura 6.17.

La stima empirica della dipendenza spaziale tra le stazioni pluviometriche per i dati di precipitazione di durata 24 ore e per periodo di osservazione 1960-2020, mostrata nella Figura 6.23, è simile alla stima nel caso dell'intero periodo di osservazione (1928-2020), mostrata nella Figura 6.18. In particolare, nel caso di stazione pluviometriche situate a distanza inferiore a 100 km, si stima un coefficiente estremo

nel range di valori tra 1.5 circa a 1.7, continuando ad aumentare gradualmente fino a circa 1.8 per 250 km di distanza. Invece, la probabilità di concomitanza estrema assume valori tra 0.4 circa a 25 km di distanza a 0.3 a 100 km.

I risultati ottenuti con cross-validation sono presentati nella Tabella A.7 e mostrano valori di errori medi di previsione più elevati rispetto ai valori ottenuti con i dati di precipitazione di durata 1 ora. Ne consegue che i modelli considerati non rispecchiano del tutto i dati osservati. In particolare, il modello con errore medio minore, considerando RMSE, è il modello Schlather e con covariate per il parametro di *scale* date da longitudine e latitudine. Le stime dei parametri per le *trend surfaces* e per i valori di *range* λ e *smooth* κ , e relativi errori standard del modello selezionato, addestrato sull'intero dataset di precipitazioni di durata 24 ore nel periodo 1960-2020, sono riportati in (6.6). Il valore di TIC del modello, invece, è pari a 19298308.

$$\begin{aligned}
 \mu(x) &= -721.8412_{(7.6479)} + 0.1530_{(0.0031)}lon(x) + 0.1472_{(0.0015)}lat(x) + \\
 &\quad + 2.6205_{(0.1062)}alt(x), \\
 \sigma(x) &= -260.89801_{(7.69869)} + 0.06868_{(0.00213)}lon(x) + \\
 &\quad + 0.05353_{(0.00152)}lat(x), \\
 \xi(x) &= 0.2138_{(0.0044)}, \\
 \lambda &= 0.000205_{(0.004162)}, \\
 \kappa &= 0.167067_{(0.402141)}.
 \end{aligned} \tag{6.6}$$

Dai valori in (6.6), si può notare che i coefficienti delle covariate assumono valori molto bassi, compreso quello relativo all'altitudine, e i parametri legati alla dipendenza spaziale sono estremamente bassi. Ne consegue che tale modello ha assunto che le varie stazioni pluviometriche siano distribuite marginalmente in modo simile e che siano quasi indipendenti tra loro, senza tenere conto della distanza. Inoltre, nel caso dei parametri di *range* e *smooth*, gli errori standard sono maggiori o dello stesso ordine di grandezza delle quantità stimate, indicando un'alta incertezza dei valori ottenuti.

La Figura 6.24 mostra la funzione del coefficiente estremo e la funzione di probabilità di concomitanza estrema stimate dal modello Schlather. In particolare, si nota che entrambe le funzioni assumono quasi subito un valore pressoché costante, pari a 1.71 circa per la funzione del coefficiente estremo e 0.21 circa per la funzione di probabilità di concomitanza estrema. Ne consegue che, come previsto osservando le stime in (6.6), tale modello non cattura correttamente la dipendenza spaziale tra le stazioni pluviometriche.

L'analisi locale del modello max-stable Schlather selezionato è stata effettuata per le stesse quattro stazioni pluviometriche considerate precedentemente. I quattro *return level plots* mostrano che in alcuni siti il modello sottostima la frequenza degli eventi estremi, come nei casi mostrati dai pannelli (a, c), mentre in altri siti sovrastima la frequenza degli eventi estremi, come nei casi mostrati dai pannelli (b, d). Dunque, in generale, il modello non cattura correttamente le caratteristiche locali delle diverse stazioni pluviometriche.

La Figura 6.26 mostra il *25-year return level* e relativi errori standard predetti dal modello Schlather. In tal caso, si osserva un andamento crescente dei valori predetti da sud-ovest verso nord-est che non tiene minimamente conto dei valori di altitudine. Ne consegue che, anche se gli errori standard sono relativamente bassi rispetto al range di valori predetti, il risultato ottenuto non è coerente con quanto osservato.

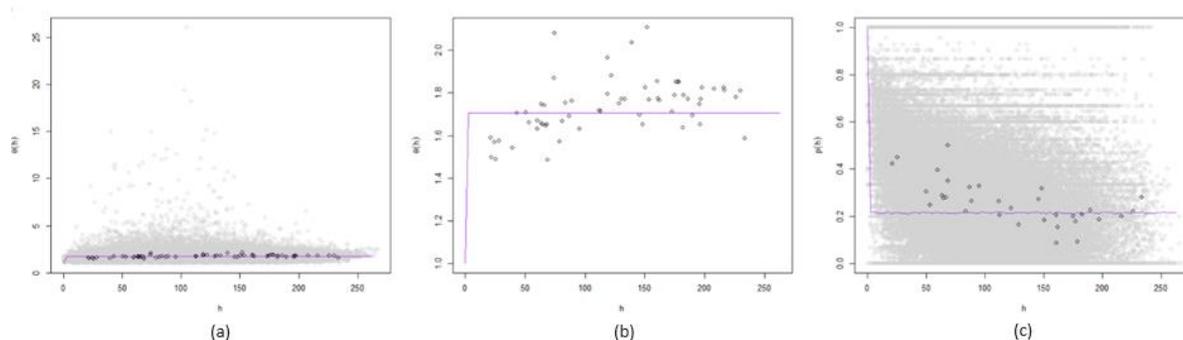


Figura 6.24: Valutazione della dipendenza spaziale stimata dal modello Schlather per le precipitazioni di durata 24 ore per il periodo di osservazione 1960-2020, tramite funzione del coefficiente estremo (a, b), e funzione di probabilità di concomitanza estrema (c). I punti grigi sono le stime empiriche calcolate rispetto alle osservazioni e i punti neri rappresentano la versione discretizzata.

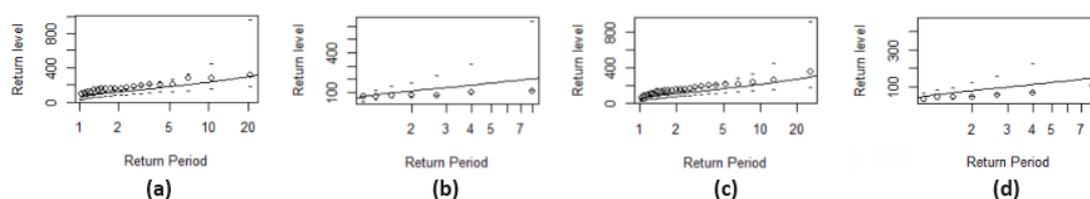


Figura 6.25: Valutazione locale del modello max-stable Schlather per i dati di precipitazione di durata 24 ore e per il periodo di osservazione 1960-2020, tramite *return level plots* (a, b, c, d) per 4 stazioni pluviometriche.

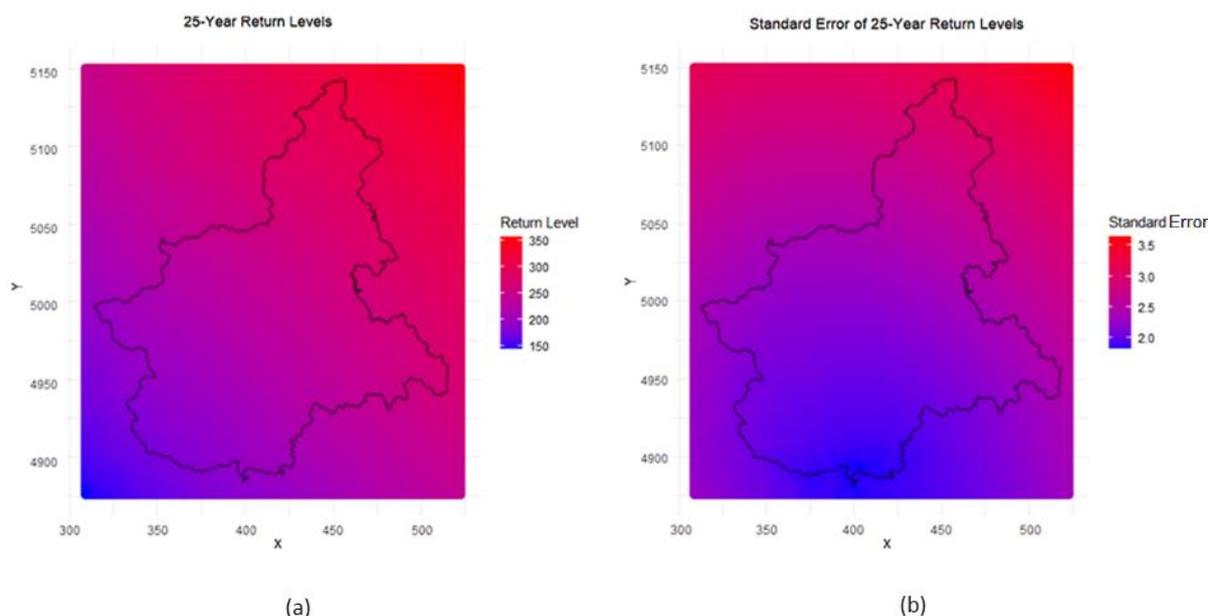


Figura 6.26: 25-year return levels (a) ed errori standard (b) per l'intero territorio del Piemonte e le aree circostanti per precipitazioni di durata 24 ore e il periodo di osservazione 1960-2020.

6.3.3 Anni 2000-2020

Per il periodo di osservazione 2000-2020, l'analisi è stata condotta su 317 località. La distribuzione dell'altitudine complessiva non è variata significativamente rispetto al dataset completo. Infatti, dalla Figura A.4 si nota che la maggior parte delle stazioni pluviometriche ha un'altitudine inferiore a 800 metri e pochissime stazioni si trovano ad una quota superiore ai 2500 metri. Le statistiche descrittive per i valori di precipitazione di durata 24 ore nel periodo temporale considerato, indipendentemente

dalla stazione pluviometrica, sono presentate nella Tabella 6.9. In particolare, il range di valori assunti dalle precipitazioni negli ultimi 21 anni di osservazione si è leggermente ridotto, ovvero il valore minimo è superiore rispetto ai valori calcolati sull'intero periodo temporale (1928-2020). Di conseguenza, sono aumentate anche le statistiche descrittive, compresa la deviazione standard.

Il pannello a sinistra della Figura 6.27 mostra la deviazione della media campionaria di ciascuna stazione pluviometrica rispetto al valore di media complessiva pari a 105.57 mm, mentre il pannello a destra mostra la deviazione standard di ciascuna località. Rispetto a quanto osservato considerando l'intero periodo di osservazione (1928-2020), si nota che le stazioni pluviometriche a sud-ovest del Piemonte assumono dei valori simili alla media complessiva. Invece, per le altre zone valgono le stesse considerazioni fatte con la Figura 6.17.

La stima empirica della dipendenza spaziale tra le stazioni pluviometriche per i dati di precipitazione di durata 24 ore è mostrata nella Figura 6.28. Nello specifico, si può notare lo stesso andamento riscontrato per periodi di osservazione maggiori ma le stime sono meno disperse.

Successivamente all'analisi esplorativa precedentemente esposta, sono state applicate entrambe le metodologie precedentemente presentate. Con la prima metodologia è stato selezionato il modello (M1) Extremal-t con *trend surfaces* costruite considerando longitudine, latitudine e altitudine come covariate per i parametri *location* e *scale* e nessuna per il parametro *shape*. Con la seconda metodologia è stato selezionato il modello (M2) Extremal-t con stesse *trend surfaces* per i parametri *location* e *shape*, mentre l'unica covariata considerata per il parametro *scale* è la longitudine. Il valore di TIC del modello M1 è pari a 13765201 e del modello M2 è pari a 13813230. Maggiori dettagli sui risultati di entrambe le metodologie sono esposti nell'Appendice A.

I parametri e gli errori standard stimati dal modello M1 per le *trend surfaces* e per i valori di *range* λ , *smooth* κ e gradi di libertà ν sono riportati in (6.7). Nello specifico, le stime dei parametri di *range* e *smooth* sono superiori rispetto ai valori ottenuti sull'intero dataset in (6.5), suggerendo una correlazione tra le stazioni pluviometriche anche a maggiori distanze.

$$\begin{aligned}
 \mu(x) &= -729.3713_{(62.6074)} + 0.1662_{(0.0264)}lon(x) + 0.1462_{(0.0115)}lat(x) + \\
 &\quad + 11.3885_{(1.3742)}alt(x), \\
 \sigma(x) &= -175.1616_{(72.6973)} + 0.1123_{(0.0203)}lon(x) + 0.0307_{(0.0137)}lat(x) + \\
 &\quad + 11.4967_{(1.3623)}alt(x), \\
 \xi(x) &= 0.1854_{(0.0240)}, \\
 \lambda &= 92.3838_{(16.0897)}, \\
 \kappa &= 0.9672_{(0.0568)}, \\
 \nu &= 5.2481_{(0.56340)}.
 \end{aligned} \tag{6.7}$$

Invece, i parametri e gli errori standard stimati dal modello M2 per le *trend surfaces* e per i valori di *range* λ e *smooth* κ sono riportati in (6.8). Nello specifico, tale modello dà un peso notevolmente ridotto all'altitudine e stima valori leggermente superiori per i parametri legati alla dipendenza spaziale rispetto al modello M1. Gli errori standard sono in generale bassi per tutti i coefficienti. Tuttavia, gli errori relativi ai coefficienti $\beta_{4,\mu}$ e $\beta_{1,\sigma}$ sono maggiori delle quantità stimate, indicando un'alta incertezza dei valori ottenuti.

<i>Min</i>	<i>1° quartile</i>	<i>Mediana</i>	<i>Media</i>	<i>3° quartile</i>	<i>Max</i>	<i>Deviazione standard</i>
22.85	65.40	90.60	105.57	128.50	633.00	60.37155

Tabella 6.9: Statistiche descrittive (minimo, primo quartile, mediana, media, terzo quartile, massimo e deviazione standard) per le precipitazioni di durata 24 ore e periodo di osservazione 2000-2020.

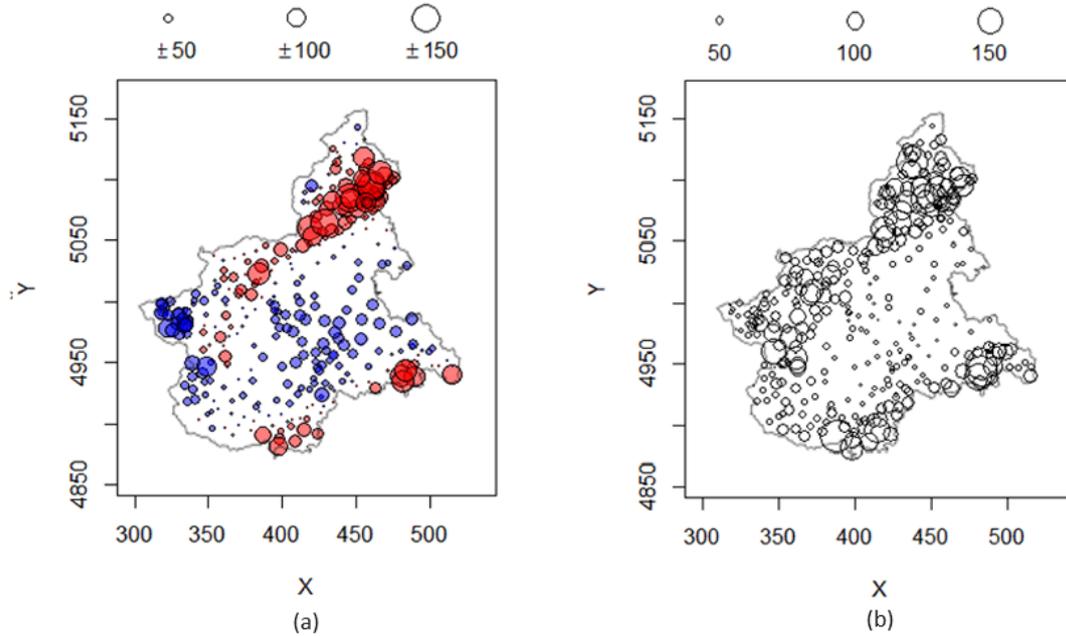


Figura 6.27: *Symbol plot* che mostra per ciascuna stazione pluviometrica la deviazione della media campionaria locale rispetto alla media campionaria complessiva (a) e la deviazione standard (b) per le precipitazioni di durata 24 ore e periodo di osservazione 2000-2020.

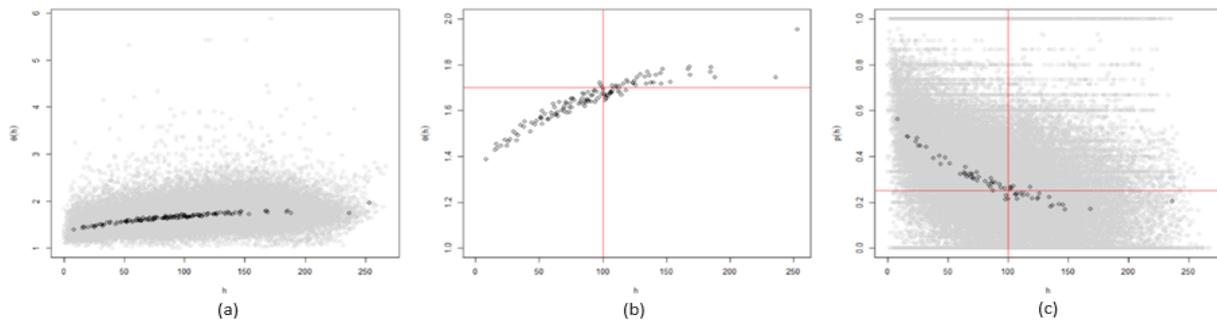


Figura 6.28: Dipendenza spaziale stimata empiricamente dai dati di precipitazione di durata 24 ore per il periodo di osservazione 2000-2020, tramite coefficiente estremo (a, b) e probabilità di concomitanza estrema (c). I punti grigi sono le stime empiriche, mentre i punti neri sono le stime discretizzate con 500 *bins*, con ingrandimento sulla versione discretizzata per il coefficiente estremo (b).

$$\mu(x) = -623.7991_{(41.1653)} + 0.1480_{(0.0253)}lon(x) + 0.1284_{(0.0079)}lat(x) + 0.2095_{(0.5785)}alt(x),$$

$$\sigma(x) = 1.16596_{(8.02525)} + 0.07946_{(0.01897)}lon(x),$$

$$\xi(x) = 0.1997_{(0.0261)}, \tag{6.8}$$

$$\lambda = 92.746_{(14.947)},$$

$$\kappa = 1.026_{(0.050)},$$

$$v = 5.750_{(0.660)}.$$

La Figura 6.29 mostra le funzioni del coefficiente estremo e di probabilità di concomitanza estrema stimate dai modelli M1 e M2. In particolare, si nota che le due funzioni sono abbastanza simili per entrambi i modelli e assumono maggiore indipendenza rispetto a quella stimata empiricamente dai dati, nonostante l'andamento graduale. Tuttavia, il modello M1 è leggermente più accurato. I valori ottenuti dalle due funzioni per alcuni valori di distanza sono riportati nella Tabella A.4 per il modello M1 e nella Tabella A.8 per il modello M2.

L'analisi locale di entrambi i modelli, mostrata nella Figura 6.30, è stata effettuata per le stesse quattro stazioni pluviometriche considerate precedentemente. Gli otto *return level plots* mostrano che sia il modello M1 sia il modello M2 sono in parte coerente con i dati osservati in alcune località, come nei casi mostrati dai pannelli (a, b, a', b'). Tuttavia, in alcuni siti i modelli sovrastima la frequenza degli eventi estremi, come nei casi mostrati dai pannelli (c, c'), mentre in altre stazioni i modelli sottostimano la frequenza degli eventi estremi, come nei casi mostrati dai pannelli (d, d').

La Figura 6.31 mostra il *25-year return level* e relativi errori standard predetti dal modello M1. In particolare, si può notare che nelle zone alpine ad elevata altitudine, i valori predetti di precipitazione, i valori predetti di precipitazione sono molto estremi, con valori di circa 500 mm a nord e 400 mm a ovest, e gli errori standard corrispondenti sono maggiori (circa 40 a nord e 30 a ovest). Di conseguenza, queste aree sono particolarmente interessate da fenomeni di precipitazione intensa e prolungata. Questo fenomeno era stato osservato nella Figura 6.27 specialmente nella zona settentrionale del Piemonte, ma per un'area più estesa. Nella catena montuosa degli Appennini, a sud della regione, invece, i valori predetti di precipitazione sono di circa 350 mm ed errori standard di circa 20-25, in linea con quanto osservato nella Figura 6.27. Le aree in cui sono stati predetti valori di precipitazione in media non superiori a 100 mm sono situate nelle zone collinari e montuose ad altitudini minori ad ovest del territorio considerato, con errori standard inferiori a circa 15-10. Tali aree, dunque, non sono particolarmente interessate da precipitazioni estreme e prolungate, rispecchiando quanto osservato nella Figura 6.27. Infine, le zone in cui sono stati predetti valori di precipitazione nel range tra 200 e 250 mm corrispondono alle aree pianeggianti, collinari e montuose ad altitudini minori a est della regione, con errori standard inferiori a 10. Per tali aree, i valori predetti sono in linea con quanto osservato nella Figura 6.27 nel caso delle aree pianeggianti e collinari. Tuttavia, nelle aree montuose a nord della regione, i valori predetti sono inferiori rispetto a quanto osservato dai dati, come era stato indicato anche dalla Figura 6.30.

La Figura 6.32, invece, mostra il *25-year return level* e i relativi errori standard predetti dal modello M2. In tal caso, si osserva un andamento crescente dei valori predetti da sud-ovest verso nord-est che non tiene minimamente conto dei valori di altitudine. Ne consegue che il risultato ottenuto non è coerente con quanto osservato.

In definitiva, i modelli M1 e M2 sono simili sia per quanto riguarda la dipendenza spaziale catturata sia per quanto riguarda il comportamento nelle singole stazioni pluviometriche. Tuttavia, le mappe spaziali dei *25-year return levels* sono differenti, ottenendo stime più accurate con il modello più complesso M1.

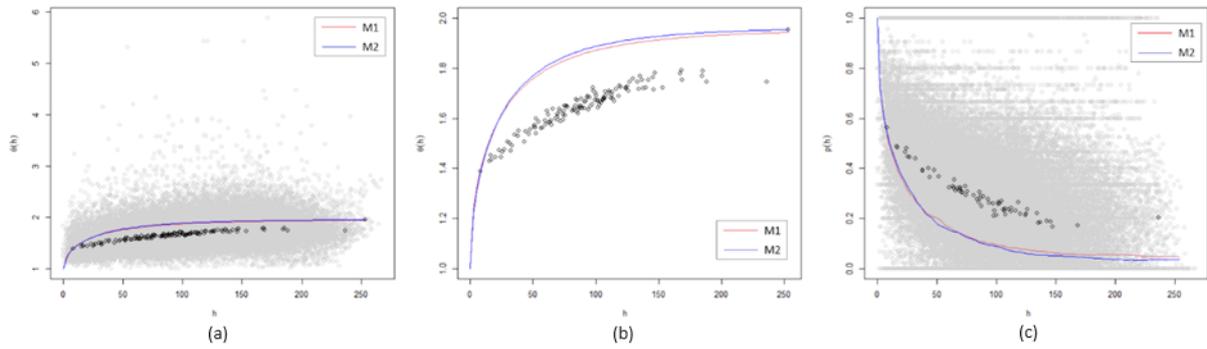


Figura 6.29: Dipendenza spaziale stimata dai modelli M1 e M2 per le precipitazioni di durata 24 ore per il periodo di osservazione 2000-2020, tramite funzione del coefficiente estremo (a, b), e funzione di probabilità di concomitanza estrema (c). I punti grigi sono le stime empiriche calcolate rispetto alle osservazioni e i punti neri rappresentano la versione discretizzata.

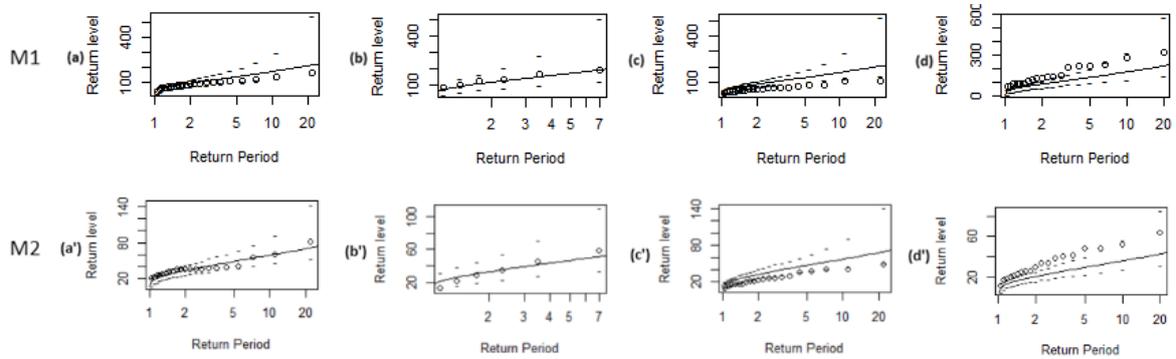


Figura 6.30: Valutazione locale dei modelli max-stable M1 e M2 selezionato per i dati di precipitazione di durata 24 ore e per il periodo di osservazione 2000-2020, tramite *return level plots* (a, b, c, d) per M1 e (a', b', c', d') per M2 per 4 stazioni pluviometriche.

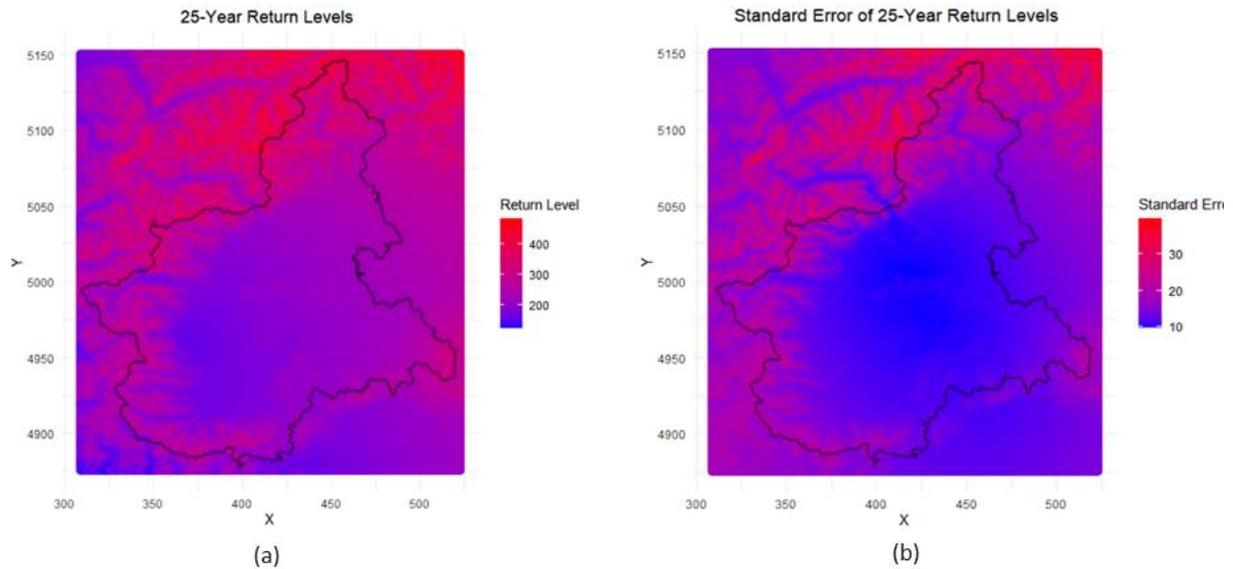


Figura 6.31: 25-year return levels (a) ed errori standard (b) per l'intero territorio del Piemonte e le aree circostanti per precipitazioni di durata 24 ore e il periodo di osservazione 2000-2020 predetti dal modello M1.

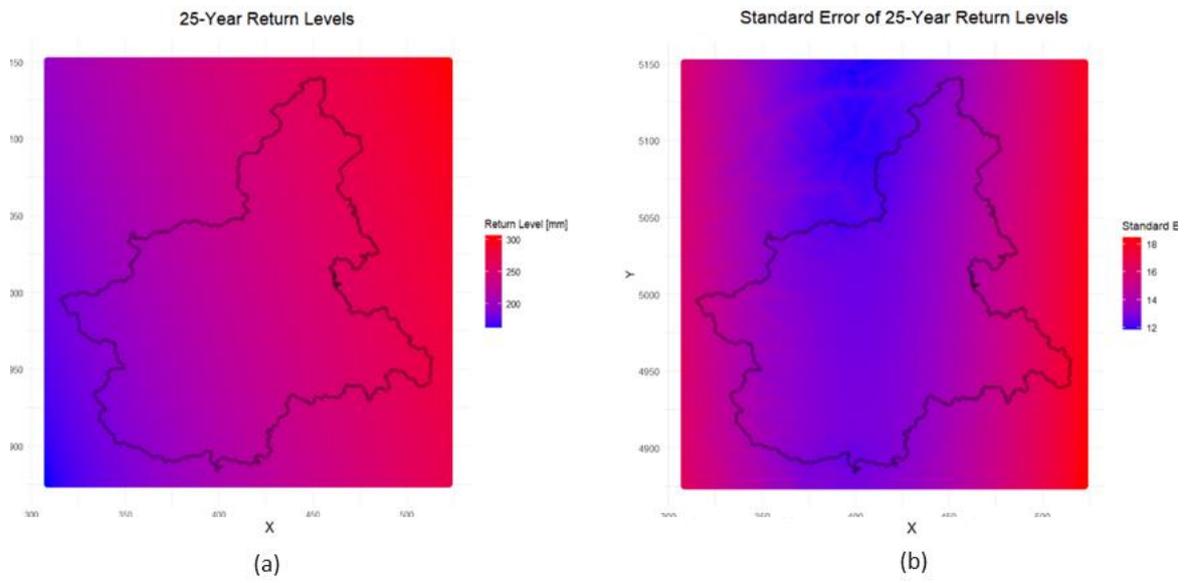


Figura 6.32: *25-year return levels* (a) ed errori standard (b) per l'intero territorio del Piemonte e le aree circostanti per precipitazioni di durata 24 ore e il periodo di osservazione 2000-2020 predetti dal modello M2.

Conclusioni

Lo scopo principale di questo lavoro è stato quello di usufruire delle proprietà statistiche relative alla teoria dei valori estremi spaziali per caratterizzare il comportamento delle precipitazioni massime annuali registrate nella regione Piemonte, considerando la loro intrinseca dipendenza spaziale. Nello specifico, sono stati confrontati diversi modelli max-stable spaziali con diverse configurazioni per i parametri della distribuzione marginale GEV, per descrivere dati relativi alle precipitazioni estreme di breve durata (1 ora) e persistenti (24 ore) registrate in un periodo di osservazione di 93 anni. Inoltre, sono state confrontate due metodologie per la selezione del modello che rappresenti meglio i dati osservati.

Dai risultati ottenuti si evince un comportamento differente tra le precipitazioni di durata 1 ora e quelle di durata 24 ore, associate a diversi tipi di fenomeno (convettivo le prime, stratiforme le seconde). In particolare, le precipitazioni di breve durata (1 ora) si verificano con intensità maggiore nelle aree montuose ad altitudine intermedia sino a raggiungere valori più estremi di 80 mm. In particolar modo, il fenomeno caratterizza principalmente la zona a nord-est della regione e in parte a sud-est, lungo la catena montuosa degli Appennini. Tuttavia, queste zone sono caratterizzate da un'alta variabilità nell'intensità delle precipitazioni. Le zone a quote più elevate, invece, non sono interessate da questi fenomeni estremi, in quanto le quantità di precipitazioni osservate si aggirano intorno ai 10 mm, senza differenze significative nell'intero periodo di osservazione. Infine, le aree pianeggianti e collinari della regione sono caratterizzate da valori intermedi di precipitazioni di durata 1 ora. Al contrario, le precipitazioni di lunga durata (24 ore) si verificano con intensità maggiore nelle zone montuose con altitudini elevate sino a raggiungere quantità più estreme di 400 mm. In particolare, i territori interessati sono le catene montuose delle Alpi e, in maniera leggermente ridotta, degli Appennini. Il fenomeno si riduce nelle zone collinari e pianeggianti della regione, arrivando a quantità di circa 100 mm. Di conseguenza, il comportamento osservato con l'analisi delle precipitazioni di durata 24 ore rispecchia il carattere orografico delle precipitazioni.

Una caratteristica che evidenzia il differente comportamento tra le precipitazioni di differente durata è la dipendenza spaziale tra le stazioni pluviometriche stimata empiricamente. Infatti, nel caso di precipitazioni di breve durata, la dipendenza spaziale stimata è minima già a basse distanze tra due stazioni pluviometriche, ed è prossima all'indipendenza a distanze moderate. Al contrario, nel caso di precipitazioni persistenti, viene stimata una significativa dipendenza tra stazioni pluviometriche vicine, una discreta dipendenza a distanze moderate e si raggiunge indipendenza solo a grandi distanze. La maggior parte dei modelli max-stable selezionati, sia nel caso delle precipitazioni di durata 1 ora sia nel caso delle precipitazioni di durata 24 ore e per i diversi periodi di osservazione considerati, stimano una minore dipendenza spaziale di quella stimata empiricamente. Inoltre, il modello Schlather e a volte anche il modello Extremal-t non catturano correttamente l'andamento della dipendenza spaziale rispetto ai valori di distanza tra le località, assumendo un valore quasi costante per qualsiasi distanza. Per quanto riguarda le distribuzioni marginali, i diversi modelli riescono a catturare il comportamento locale per alcune località, mentre sovrastimano o sottostimano la frequenza degli eventi estremi in altre località.

La selezione del modello max-stable attraverso la seconda metodologia è computazionalmente più costosa e non sempre comporta previsioni più accurate. In particolare, nel caso delle precipitazioni di durata 1 ora e per qualsiasi periodo di osservazione considerato, si ottengono mappe per i *25-year return*

levels simili con qualsiasi modello selezionato. Inoltre, nel caso del periodo di osservazione 2000-2020, i modelli selezionati con le due metodologie sono paragonabili e assumono la dipendenza spaziale e la distribuzione marginale in modo simile. Nel caso delle precipitazioni di durata 24 ore, invece, i modelli selezionati con la seconda metodologia considerano meno covariate, escludendo l'altitudine per le *trend surfaces* per parametro *scale* della distribuzione GEV. Di conseguenza, i *25-year return levels* predetti seguono un gradiente da nord-est a sud-ovest, ovvero le precipitazioni predette sono più abbondanti nelle zone situate a nord-est e diminuiscono gradualmente spostandosi verso sud-ovest, senza tenere conto dei diversi valori di altitudine dei territori interessati. Tuttavia, ciò non è coerente con il comportamento orografico osservato. Al contrario, i modelli selezionati con la prima metodologia considerano le tre coordinate geografiche come covariate sia per il parametro di *location* sia per il parametro di *scale*, aumentando la complessità del modello. Tuttavia, i valori predetti dei *return levels* rispecchiano il comportamento orografico osservato. Nel caso del periodo di osservazione 2000-2020, infine, i due modelli selezionati con le due metodologie catturano la struttura di dipendenza spaziale in modo simile.

Appendice A

Risultati supplementari

Questa appendice contiene risultati supplementari e ulteriori dettagli a supporto delle analisi e delle discussioni presentate nel Capitolo 6. Le informazioni fornite in questa sezione sono state incluse per migliorare la comprensione riguardo ai metodi utilizzati e ai dati ottenuti, offrendo un quadro più completo del lavoro svolto.

Le seguenti figure mostrano la distribuzione del valore di altitudine delle stazioni pluviometriche considerate per i dati di precipitazione di durata 1 e 24 ore e i diversi periodi di osservazione.

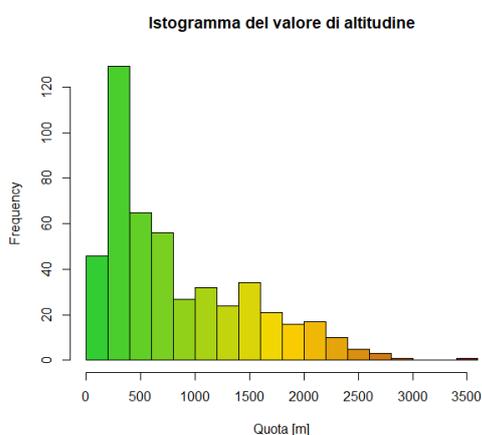


Figura A.1: Istogramma del valore di altitudine per le stazioni pluviometriche con registrazioni delle precipitazioni di durata 1 ora nel periodo di osservazione 1960-2020.

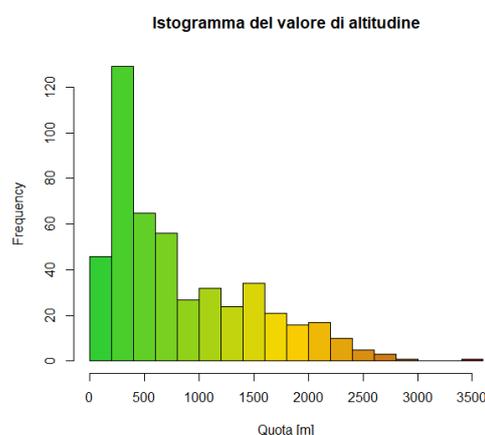


Figura A.3: Istogramma del valore di altitudine per le stazioni pluviometriche con registrazioni delle precipitazioni di durata 24 ore nel periodo di osservazione 1960-2020.

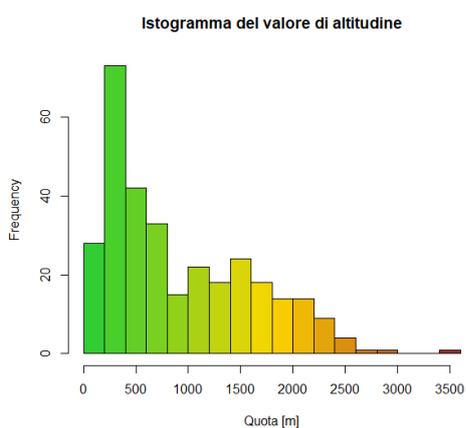


Figura A.2: Istogramma del valore di altitudine per le stazioni pluviometriche con registrazioni delle precipitazioni di durata 1 ora nel periodo di osservazione 2000-2020.

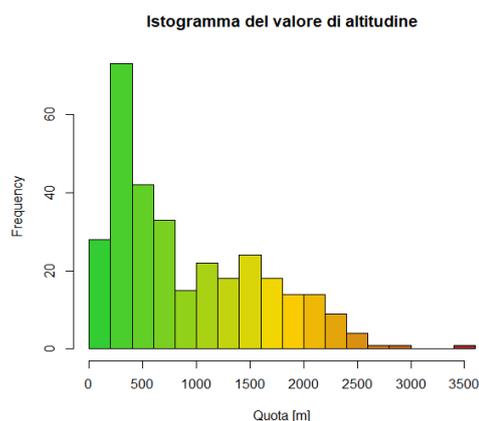


Figura A.4: Istogramma del valore di altitudine per le stazioni pluviometriche con registrazioni delle precipitazioni di durata 24 ore nel periodo di osservazione 2000-2020.

La tabella seguente mostra i risultati ottenuti per la selezione delle *trend surfaces* della distribuzione GEV per il dataset di precipitazioni di durata 1 e 24 ore e per i periodi di osservazione 1928-2020 e 2000-2020.

Modello	TIC			
	1 h		24 ore	
	1928-2020	2000-2020	1928-2020	2000-2020
M_0	91038.15	42215.56	120362.3	55395.02
M_1	89700.93	41722.15	119524.8	55017.35
M_{1a}	89549.31	41473.52	119471.7	55174.45
M_{2x}	89181.72	41569.59	119344.9	54977.14
M_{2xa}	88997.46	41363.01	119330.9	55004.30
M_{2xa2}	88900.20	41287.39	118848.1	54824.29
M_{2y}	89946.20	41783.09	119491.6	55139.8
M_{2ya}	89598.57	41548.84	119239.7	55044.76
M_{2ya2}	89227.29	41381.66	119387.5	55027.35
M_2	89309.05	41588.16	119215.9	55000.71
M_{2a}	89222.37	41470.51	119267.3	55014.36
M_{2a2}	89064.33	41366.79	118741.2	54812.00
M_3	89100.86	41511.39	119148.4	54977.05

Tabella A.1: Valori di TIC per i modelli GEV spaziali.

La tabella seguente mostra i risultati ottenuti per la selezione del modello max-stable con la prima metodologia per il dataset di precipitazioni di durata 1 e 24 ore e per i periodi di osservazione 1928-2020 (dataset completo) e 2000-2020. In particolare, si nota che nel caso di precipitazioni di durata 24 ore e periodo di osservazione 1928-2020 il modello Schlather ha riscontrato alcuni problemi dal momento che non ha potuto stimare gli errori standard.

Dataset	Schlather	Brown-Resnick	Extremal-t
1h/1928-2020	15704510	15527621	15558499
1h/2000-2020	10512114	10412180	10427603
24h/1928-2020	NA	20631586	20632780
24h/2000-2020	13800820	13766846	13765201

Tabella A.2: Valori di TIC per i modelli max-stable (prima metodologia).

Di seguito le stime dei coefficienti e relativi errori standard come pedici per i parametri della distribuzione GEV corrispondenti ai modelli GEV spaziali selezionati sulla base dei risultati presentati nella Tabella A.1 e corrispondenti modelli max-stable considerati.

Precipitazioni di durata 1 ora per anni 1928-2020: modello M_{2xa2}

$$\begin{aligned}
 \mu(x) &= -159.77201_{(4.19104)} + 0.06816_{(0.00094)} \text{lon}(x) + \\
 &\quad + 0.03133_{(0.00085)} \text{lat}(x) - 2.93968_{(0.05417)} \text{alt}(x), \\
 \sigma(x) &= -2.33500_{(0.27772)} + 0.03064_{(0.00063)} \text{lon}(x) - 1.12454_{(0.04000)} \text{alt}(x), \\
 \xi(x) &= 0.07468_{(0.00257)}.
 \end{aligned} \tag{A.1}$$

Nello specifico, si può notare che il parametro di *location* aumenta in base ai valori di longitudine e latitudine ma diminuisce con un peso maggiore in base al valore di altitudine. Per quanto riguarda il parametro di *scale*, esso aumenta in base al valore di longitudine e diminuisce con peso maggiore rispetto al valore di altitudine, pur mantenendo dei valori positivi in ogni località. Infine, il parametro di *shape* è positivo, seppur piccolo, suggerendo una distribuzione di tipo Fréchet.

Precipitazioni di durata 1 ora per anni 2000-2020: modello M_{2xa2}

Il test ANOVA tra il modello M_{2xa2} e il modello M_{2xa} ha restituito un p -value pari a $5.698e-12$. Di seguito le *trend surfaces* relative al modello M_{2xa2} .

$$\begin{aligned}\mu(x) &= -164.86229_{(22.47325)} + 0.03999_{(0.00427)}lon(x) + \\ &\quad + 0.03543_{(0.00462)}lat(x) - 4.99886_{(0.29524)}alt(x), \\ \sigma(x) &= 1.31996_{(1.39704)} + 0.02451_{(0.00315)}lon(x) - 1.52633_{(0.22949)}alt(x), \\ \xi(x) &= 0.07493_{(0.01195)}.\end{aligned}\tag{A.2}$$

In particolare, si nota che il parametro di *location* aumenta in base ai valori di longitudine e latitudine ma diminuisce con un peso maggiore in base al valore di altitudine. Per quanto riguarda il parametro di *scale*, esso aumenta in base al valore di longitudine e diminuisce con peso leggermente maggiore rispetto al valore di altitudine, pur mantenendo dei valori positivi in ogni località. Infine, il parametro di *shape* è positivo, seppur piccolo, suggerendo una distribuzione di tipo Fréchet, e molto simile al valore ottenuto in (A.1) nell'intero dataset.

Il modello Brown-Resnick performa meglio rispetto ai modelli Schlather ed Extremal-t. Infatti, la seconda riga nella Tabella A.2 mostra un valore di TIC significativamente inferiore per il modello Brown-Resnick. Invece, la Figura A.5 mostra che tale modello assume quasi indipendenza tra le stazioni pluviometriche già a distanze moderate, ma l'andamento è graduale. Al contrario, la funzione del coefficiente estremo stimata dagli altri due modelli assume quasi subito un valore pressoché costante. Nello specifico, il modello Extremal-t assume sin da subito una bassa dipendenza spaziale tra le stazioni pluviometriche, dal momento che i valori della funzione del coefficiente estremo stimata si aggirano intorno a 1.93, come evidenziato anche dalla Tabella A.3. Tale valore è inferiore alle stime del modello di Brown-Resnick e più vicino alle stime empiriche già a distanze minori di 50 km. Invece, il modello di Schlather stima maggiore dipendenza spaziale da quella stimata dai dati, dal momento che la funzione del coefficiente estremo assume valori intorno a quasi 1.7. In modo analogo, la funzione di probabilità di concomitanza estrema (c) stimata dal modello Extremal-t raggiunge quasi subito il valore quasi costante di 0.05 circa, mentre per il modello Schlather di 0.25 circa.

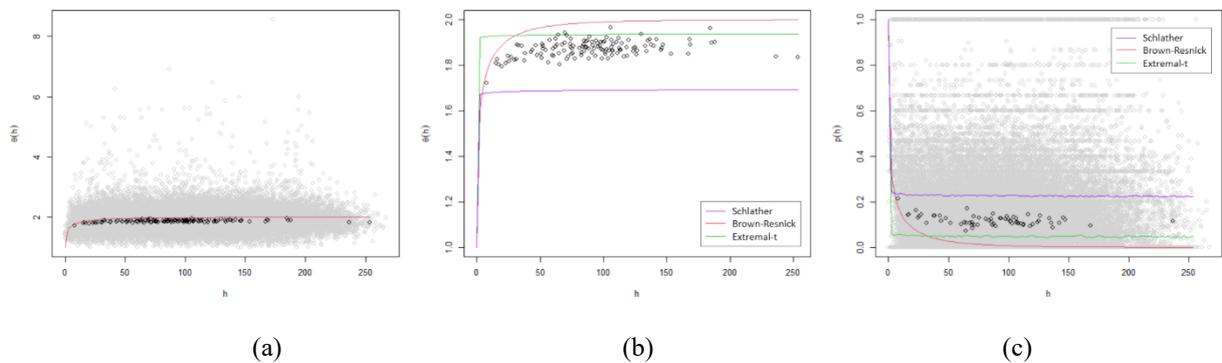


Figura A.5: Valutazione della dipendenza spaziale stimata dai modelli max-stable per le precipitazioni di durata 1 ora per il periodo di osservazione 2000-2020, tramite funzione del coefficiente estremo per il modello di Brown-Resnick (a) e per tutti i modelli (b), e funzione di probabilità di concomitanza estrema per ogni modello (c).

Funzione	Distanza	Schlather	Brown-Resnick (M1)	Extremal-t
Coefficiente estremo	$h = 5 \text{ km}$	1.669516	1.707902	1.924893
	$h = 10 \text{ km}$	1.673851	1.805742	1.927756
	$h = 20 \text{ km}$	1.677863	1.890270	1.930367
	$h = 30 \text{ km}$	1.680061	1.929258	1.931783
	$h = 50 \text{ km}$	1.682675	1.964940	1.933454
	$h = 100 \text{ km}$	1.685949	1.990593	1.935528
Probabilità di concomitanza estrema	$h = 5 \text{ km}$	0.5991332	0.235644522	0.05829929
	$h = 10 \text{ km}$	0.2456796	0.154731249	0.05524629
	$h = 20 \text{ km}$	0.2407364	0.087731522	0.05140194
	$h = 30 \text{ km}$	0.2374761	0.057542789	0.05259848
	$h = 50 \text{ km}$	0.2386768	0.029855158	0.04952644
	$h = 100 \text{ km}$	0.2328017	0.008783205	0.04651237

Tabella A.3: Valori delle stime delle funzioni del coefficiente estremo e di probabilità di concomitanza estrema in alcune distanze per i modelli max-stable nel caso di precipitazioni di durata 1 ora e il periodo di osservazione 2000-2020.

Precipitazioni di durata 24 ore per anni 1928-2020: modello M_{2a2}

Il test ANOVA tra il modello M_{2a2} e il modello M_{2xa2} ha restituito un p -value pari a $1.069e-15$. Di seguito le *trend surfaces* relative al modello M_{2a2} .

$$\begin{aligned} \mu(x) &= -969.3611_{(17.3071)} + 0.2168_{(0.0062)}lon(x) + 0.1896_{(0.0034)}lat(x) + \\ &\quad + 15.1463_{(0.4801)}alt(x), \\ \sigma(x) &= -246.71580_{(15.30740)} + 0.14260_{(0.00459)}lon(x) + 0.04231_{(0.00295)}lat(x) + \\ &\quad + 12.81375_{(0.40655)}alt(x), \\ \xi(x) &= 0.2129_{(0.0065)} \end{aligned} \quad (A.3)$$

In particolare, si nota che i parametri di *location* e *scale* aumentano in base ai valori di longitudine, latitudine e, con un peso maggiore, altitudine, mentre il parametro di *shape* è positivo, suggerendo una distribuzione di tipo Fréchet. Gli errori standard sono bassi per tutti i coefficienti, indicando che le stime siano abbastanza accurate, ad eccezione di $\beta_{1,\mu}$ e $\beta_{1,\sigma}$. I coefficienti sono, in generale, maggiori rispetto a quelli ottenuti per i dati di precipitazione di durata 1 ora, come era stato previsto.

Precipitazioni di durata 24 ore per anni 2000-2020: modello M_{2a2}

Il test ANOVA tra il modello M_{2a2} e il modello M_{2xa2} ha restituito un p -value pari a a . Di seguito le *trend surfaces* relative al modello M_{2a2} .

$$\begin{aligned} \mu(x) &= -803.6833_{(77.6305)} + 0.2070_{(0.0337)}lon(x) + 0.1577_{(0.0144)}lat(x) + \\ &\quad + 12.8010_{(1.8683)}alt(x), \\ \sigma(x) &= -110.42783_{(91.22317)} + 0.13812_{(0.02638)}lon(x) + 0.01555_{(0.01715)}lat(x) + \\ &\quad + 12.37809_{(1.93480)}alt(x), \\ \xi(x) &= 0.1668_{(0.0295)}. \end{aligned} \quad (A.4)$$

In particolare, si nota che tutti i coefficienti legati alle covariate spaziali sono positivi e quelli relativi all'altitudine assumono valori maggiori. Invece, il parametro di *shape* è positivo, suggerendo una distribuzione di tipo Fréchet.

Il modello Extremal-t performa meglio rispetto ai modelli Schlather e Brown-Resnick. Infatti, la quarta riga nella Tabella A.2 mostra un valore di TIC inferiore per il modello Extremal-t. Invece, la Figura A.5 mostra che tale modello e il modello Brown-Resnick stimano le funzioni del coefficiente estremo e di probabilità di concomitanza estrema in modo simile. In particolare, assumono minore dipendenza spaziale tra le stazioni pluviometriche rispetto a quanto stimato dai dati, ma l'andamento è graduale. Al contrario, le funzioni stimate dal modello Schlather assumono quasi subito un valore pressoché costante. Alcuni valori assunti dalle funzioni stimate dai tre modelli sono riportati nella Tabella A.4.

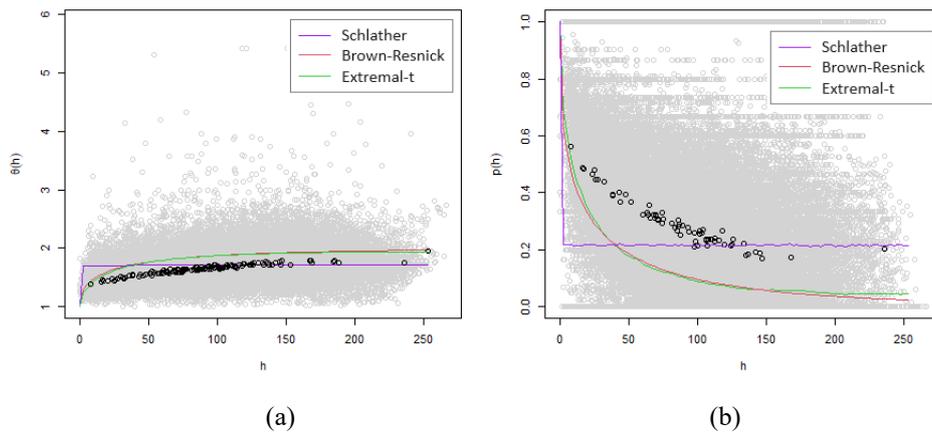


Figura A.6: Valutazione della dipendenza spaziale stimata dai modelli max-stable per le precipitazioni di durata 24 ore per il periodo di osservazione 2000-2020, tramite funzione del coefficiente estremo (a) e funzione di probabilità di concomitanza estrema per ogni modello (b).

Funzione	Distanza	Schlather	Brown-Resnick (M1)	Extremal-t
Coefficiente estremo	$h = 20 \text{ km}$	1.705622	1.598026	1.568903
	$h = 50 \text{ km}$	1.706553	1.762470	1.759591
	$h = 100 \text{ km}$	1.706873	1.874293	1.872141
	$h = 200 \text{ km}$	1.707019	1.952863	1.932916
Probabilità di concomitanza estrema	$h = 20 \text{ km}$	0.5868022	0.3341635	0.36392978
	$h = 50 \text{ km}$	0.2167791	0.1930500	0.18004711
	$h = 100 \text{ km}$	0.2128081	0.1014145	0.10153259
	$h = 200 \text{ km}$	0.2159866	0.03821082	0.04996804

Tabella A.4: Valori delle stime delle funzioni del coefficiente estremo e di probabilità di concomitanza estrema in alcune distanze per i modelli max-stable nel caso di precipitazioni di durata 24 ore e il periodo di osservazione 2000-2020.

Di seguito maggiori dettagli relativi ai modelli selezionati tramite 10-fold cross validation per le precipitazioni di durata 1 e 24 ore e periodi di osservazione 1960-2020 e 2000-2020.

La tabella seguente mostra gli errori medi calcolati per ogni modello valutato tramite cross validation per il dataset di precipitazioni di durata 1 ora, dove nella prima colonna viene mostrato il modello max-stable e le covariate per il parametro *scale*.

Modello	1960-2020		2000-2020	
	RMSE	MAE	RMSE	MAE
<i>Sch, X</i>	7.788472	6.298044	8.564102	7.115681
<i>Sch, Y</i>	7.905314	6.393577	8.568325	7.196251
<i>Sch, X+Y</i>	7.704037	6.173963	8.563234	7.142614
<i>Sch, X+Z</i>	7.945916	6.370607	8.780223	7.382990
<i>Sch, Y+Z</i>	7.869394	6.358574	8.525076	7.166590
<i>Sch, X+Y+Z</i>	7.879193	6.369337	8.322460	6.930974
<i>B-R, X</i>	7.620172	5.984354	7.736427	6.357157
<i>B-R, Y</i>	7.603401	6.046929	7.931915	6.459553
<i>B-R, X+Y</i>	7.610947	5.980164	7.766695	6.380114
<i>B-R, X+Z</i>	7.680064	6.008330	7.773940	6.364410
<i>B-R, Y+Z</i>	7.575717	5.957256	7.709493	6.358759
<i>B-R, X+Y+Z</i>	7.648598	5.992192	7.777553	6.365384
<i>E-t, X</i>	7.574452	5.948413	7.915057	6.515320
<i>E-t, Y</i>	7.598339	6.047494	7.942410	6.557885
<i>E-t, X+Y</i>	7.584104	5.951239	7.912188	6.485702
<i>E-t, X+Z</i>	7.682183	6.028227	7.824941	6.481160
<i>E-t, Y+Z</i>	7.559653	5.957144	7.869552	6.543185
<i>E-t, X+Y+Z</i>	7.651170	5.998113	7.858017	6.487413

Tabella A.5: Errori di previsione 10-fold cross-validation per le precipitazioni di durata 1 ora e periodi di osservazione 1960-2020 e 2000-2020.

In generale, per gli anni 2000-2020 il modello Schlather ha prestazioni inferiori, mentre il modello Extremal-t si comporta meglio, ma il modello Brown-Resnick offre le migliori performance, per qualsiasi combinazione di covariate. Dunque, rispetto ai risultati ottenuti per il periodo di osservazione 1960-2020, le performance dei modelli Extremal-t e Brown-Resnick risultano invertite.

Funzione	Distanza	Brown-Resnick (M_2)
Coefficiente estremo	$h = 5 \text{ km}$	1.713813
	$h = 20 \text{ km}$	1.897516
	$h = 50 \text{ km}$	1.969530
	$h = 100 \text{ km}$	1.992584
Probabilità di concomitanza estrema	$h = 5 \text{ km}$	0.439703552
	$h = 20 \text{ km}$	0.238218182
	$h = 50 \text{ km}$	0.088559200
	$h = 100 \text{ km}$	0.006411538

Tabella A.6: Valori delle stime delle funzioni del coefficiente estremo e di probabilità di concomitanza estrema in alcune distanze per il modello Brown-Resnick selezionato nel caso di precipitazioni di durata 1 ora e il periodo di osservazione 2000-2020.

La tabella seguente mostra gli errori medi calcolati per ogni modello valutato tramite cross validation per il dataset di precipitazioni di durata 24 ore, dove nella prima colonna viene mostrato il modello max-stable e le covariate per il parametro *scale*.

<i>Modello</i>	<i>1960-2020</i>		<i>2000-2020</i>	
	RMSE	MAE	RMSE	MAE
<i>Sch, X</i>	35.04108	28.57438	34.89358	29.70161
<i>Sch, Y</i>	34.43203	28.14852	35.54260	30.34836
<i>Sch, X+Y</i>	33.85026	27.61741	34.86268	29.79547
<i>Sch, X+Z</i>	34.77163	27.73192	35.30307	29.46540
<i>Sch, Y+Z</i>	35.50940	27.86904	36.81397	30.06796
<i>Sch, X+Y+Z</i>	34.41440	27.76966	35.50086	29.43960
<i>B-R, X</i>	34.76633	27.81593	34.20424	29.00896
<i>B-R, Y</i>	34.41959	27.44104	34.57115	29.30693
<i>B-R, X+Y</i>	34.20786	27.45950	34.26754	29.06952
<i>B-R, X+Z</i>	34.30336	27.35219	34.53975	28.70959
<i>B-R, Y+Z</i>	35.01781	27.31351	35.95191	29.49791
<i>B-R, X+Y+Z</i>	34.08440	27.14503	34.52266	28.71098
<i>E-t, X</i>	34.71075	27.78179	34.18294	28.99303
<i>E-t, Y</i>	34.46438	27.52027	34.58129	29.30919
<i>E-t, X+Y</i>	34.21548	27.45447	34.23499	29.03878
<i>E-t, X+Z</i>	34.28464	27.31314	34.61996	28.78235
<i>E-t, Y+Z</i>	34.97308	27.30904	36.31916	29.63825
<i>E-t, X+Y+Z</i>	34.10713	27.18626	34.84920	28.90101

Tabella A.7: Errori di previsione 10-fold cross-validation per le precipitazioni di durata 24 ore e periodi di osservazione 1960-2020 e 2000-2020.

<i>Funzione</i>	<i>Distanza</i>	<i>Extremal-t (M2)</i>
<i>Coefficiente estremo</i>	<i>h = 20 km</i>	1.567673
	<i>h = 50 km</i>	1.772093
	<i>h = 100 km</i>	1.889747
	<i>h = 200 km</i>	1.948692
<i>Probabilità di concomitanza estrema</i>	<i>h = 20 km</i>	0.3877881
	<i>h = 50 km</i>	0.1980232
	<i>h = 100 km</i>	0.0835941
	<i>h = 200 km</i>	0.0378160

Tabella A.8: Valori delle stime delle funzioni del coefficiente estremo e di probabilità di concomitanza estrema in alcune distanze per il modello Extremal-t selezionato nel caso di precipitazioni di durata 24 ore e il periodo di osservazione 2000-2020.

Appendice B

Codice

Le analisi presentate in questo lavoro sono state effettuate usando il pacchetto R *SpatialExtremes* [41], mentre le simulazioni dei processi max-stable attraverso il pacchetto R *mev* [43]. Di seguito viene riportato il codice utilizzato, da cui provengono i risultati e le figure precedentemente presentati. Si fa riferimento a tre file: *anagrafica* contiene le informazioni relative alle stazioni pluviometriche, *database* contiene i dati di precipitazione e *punti* contiene il DEM del territorio del Piemonte e delle aree circostanti.

```
# Librerie
library(readxl)
library(ggplot2)
library(data.table)
library(dplyr)
library(SpatialExtremes)
library(caret)
library(mev)

##### ANALISI ESPLORATIVA DEL DATASET
# PUNTI
grafico = ggplot(punti, aes(x=x, y=y, color=punti[, "z"])) +
  geom_point(size=3) + scale_color_gradient2(low="limegreen", mid="gold",
  high="brown", midpoint = 1650, name = "Altitudine") + labs(title="Mappa con
  altitudine")
grafico + geom_point(data = anagrafica, aes(x = X, y = Y), shape=1,
  color="black", size = 1)
hist(punti$z, main = "Distribuzione dell'altitudine", xlab = "Altitudine
  [m]", breaks = 12, col = colorRampPalette(c("limegreen", "gold",
  "brown"))(10))
boxplot(punti$z)

colnames(punti) = c("X", "Y", "QUOTA_1km")
punti = data.matrix(punti/1000)

# ANAGRAFICA
dim(anagrafica)
head(anagrafica)
length(unique(anagrafica$ID)) # verifico ID univoci
summary(anagrafica)
sd(anagrafica$X)
sd(anagrafica$Y)
sd(anagrafica$QUOTA_UFFICIALE)
sd(anagrafica$QUOTA_1km)
sd(anagrafica$OPENNESS_5km)
sd(anagrafica$MAP_BIGBANG)

ggplot(anagrafica, aes(x = X, y = Y, color = QUOTA_1km)) +
```

```

geom_point(size = 2) + scale_color_gradient2(low = "limegreen", mid="gold",
high = "brown", midpoint = 1650, name = "Quota [m]") +
ggtitle("Distribuzione delle stazioni pluviometriche con quota") +
theme(plot.title = element_text(hjust = 0.5))
plot(anagrafica[,2:7])
plot(anagrafica$QUOTA_1km, anagrafica$OPENNESS_5km)
par(mfrow=c(2,2))
hist(anagrafica$QUOTA_UFFICIALE, main = "Distribuzione della Quota
Ufficiale", xlab = "Quota Ufficiale [m]", breaks = 12,
col = colorRampPalette(c("limegreen", "gold", "brown"))(18))
hist(anagrafica$QUOTA_1km, main = "Distribuzione dell'altitudine",
xlab = "Quota_1km [m]", breaks = 12, col = colorRampPalette(c("limegreen",
"gold", "brown"))(18))
boxplot(anagrafica$QUOTA_1km, main = "Boxplot dell'altitudine",
ylab = "Quota_1km")
par(mfrow=c(1,2))
hist(anagrafica$OPENNESS_5km, main = "Distribuzione dell'openness",
xlab = "Openness")
boxplot(anagrafica$OPENNESS_5km, main="Boxplot dell'openness",
ylab = "Openness")
hist(anagrafica$MAP_BIGBANG, main = "Distribuzione della precipitazione media
annuale", xlab = "Precipitazione media annuale (mm)",
col = colorRampPalette(c("blue", "red"))(9))
boxplot(anagrafica$MAP_BIGBANG, main = "Boxplot della Precipitazione Media
Annuale", ylab = "Precipitazione (mm)")
par(mfrow=c(1,1))
ggplot(data = anagrafica, aes(x = X, y = Y, color = OPENNESS_5km)) +
geom_point(size=2) + scale_color_gradient(low = "orange", high =
"darkgray", name = "Openness") + labs(title = "Mappa della openness") +
theme(plot.title = element_text(hjust = 0.5))
ggplot(data = anagrafica, aes(x = X, y = Y, color = MAP_BIGBANG)) +
geom_point(size=2) + scale_color_gradient(low = "blue", high = "red",
name = "Precipitazione [mm]") + labs(title = "Mappa della Precipitazione
Media Annuale") + theme(plot.title = element_text(hjust = 0.5))

# DATABASE
dim(database)
summary(database)
sd(unlist(database[, "10min_PREC"]), na.rm = TRUE)
sd(unlist(database[, "20min_PREC"]), na.rm = TRUE)
sd(unlist(database[, "30min_PREC"]), na.rm = TRUE)
sd(unlist(database[, "1h_PREC"]), na.rm = TRUE)
sd(unlist(database[, "3h_PREC"]), na.rm = TRUE)
sd(unlist(database[, "6h_PREC"]), na.rm = TRUE)
sd(unlist(database[, "12h_PREC"]), na.rm = TRUE)
sd(unlist(database[, "24h_PREC"]), na.rm = TRUE)
sd(unlist(database[, "1gg_PREC"]), na.rm = TRUE)
sd(unlist(database[, "2gg_PREC"]), na.rm = TRUE)
sd(unlist(database[, "3gg_PREC"]), na.rm = TRUE)
sd(unlist(database[, "4gg_PREC"]), na.rm = TRUE)
sd(unlist(database[, "5gg_PREC"]), na.rm = TRUE)

hist(database$anno, main = "Distribuzione temporale", xlab = "Anno",

```

```

col='lightblue')
boxplot(database$anno, main = "Boxplot degli anni", ylab='Anno',
col='lightblue')
colPREC = grep("PREC", colnames(database))
hist(unlist(database[,colPREC[4]]), xlab="1h_PREC", main= "Distribuzione
delle precipitazioni di durata 1 ora", col= colorRampPalette(c('lightblue',
'blue', 'red'))(9))
hist(unlist(database[,colPREC[8]]), xlab="24h_PREC", main= "Distribuzione
delle precipitazioni di durata 24 ore", col= colorRampPalette(c('lightblue',
'blue', 'red'))(8))
hist(unlist(database[,colPREC[4]+1]), xlab="1h_YY", main= "Distribuzione
temporale delle precipitazioni di durata 1 ora", col='lightblue')
hist(unlist(database[,colPREC[8]+1]), xlab="24h_YY", main= "Distribuzione
temporale delle precipitazioni di durata 24 ore", col='lightblue')

df = dati_uniti[, c("ID", "1h_PREC", "1h_YY")]
colnames(df) = c("ID", "PREC_1h", "Anno_1h")
df = df[complete.cases(df$PREC_1h), ] # Elimino valori mancanti
df = as.data.table(df)
conteggi_per_anno <- df %>% group_by(Anno_1h) %>%
summarise(NumeroIDUnivoci = n_distinct(ID))
ggplot(conteggi_per_anno, aes(x = Anno_1h, y = NumeroIDUnivoci)) +
geom_bar(stat = "identity", fill = "lightblue4") +
labs(title = "Distribuzione temporale del numero di stazioni pluviometriche
per precipitazioni massime annuali di durata 1 ora", x = "Anno",
y = "Numero di stazioni pluviometriche") +
theme(plot.title = element_text(hjust = 0.5))
df = dati_uniti[, c("ID", "24h_PREC", "24h_YY")]
colnames(df) = c("ID", "PREC_24h", "Anno_24h")
df = df[complete.cases(df$PREC_24h), ] # Elimino valori mancanti
df = as.data.table(df)
conteggi_per_anno <- df %>% group_by(Anno_24h) %>%
summarise(NumeroIDUnivoci = n_distinct(ID))
ggplot(conteggi_per_anno, aes(x = Anno_24h, y = NumeroIDUnivoci)) +
geom_bar(stat = "identity", fill = "lightblue4") +
labs(title = "Distribuzione temporale del numero di stazioni pluviometriche
per precipitazioni massime annuali di durata 24 ore", x = "Anno",
y = "Numero di stazioni pluviometriche") +
theme(plot.title = element_text(hjust = 0.5))

##### Funzione per mostrare symbol plot con deviazione standard (simile a
##### symbolplot del pacchetto SpatialExtremes
symbolplot_sd <- function(data, coord, plot.legend = TRUE, scale = 1){
add <- FALSE
values <- apply(data, 2, sd, na.rm = TRUE) # deviazioni standard

if (plot.legend)
layout(matrix(2:1, 2), heights = c(0.15, 1)) # per grafico legenda sopra

mar <- par("mar")
radius <- values
norm.factor <- scale * min(dist(coord)) / max(radius)

```

```

symbols(coord, circles = radius * norm.factor, add = add, inches = FALSE)

if (plot.legend){
  radius.legend <- pretty(radius, 4)
  radius.legend <- radius.legend[radius.legend > 0]
  n.circles <- length(radius.legend)

  xlim <- par("usr")[1:2]
  ylim <- par("usr")[3:4]
  centers <- cbind(seq(xlim[1], xlim[2], length = n.circles), mean(ylim))

  par(mar = c(0, 5, 0, 5))
  plot(-10^6, xlim = xlim, ylim = ylim, axes = FALSE, bty = "n",
       xlab = "", ylab = "")
  symbols(centers, circles = radius.legend * norm.factor, inches = FALSE,
         bty = "n", xaxt = "n", yaxt = "n", xlab = "", ylab = "",
         add = TRUE)

  for (j in 1:n.circles)
    text(centers[j,1], mean(ylim) - 1.1 * max(radius.legend) * norm.factor,
        bquote(.(radius.legend[j])), pos = 1, offset = 0.75)

  par(mar = mar)
}
}

##### Analisi dataset 1h (analogo per 24 h)
# Elimino le stazioni in cui è presente un solo valore (problemi in seguito)
conteggio <- table(database[,1])
valori_unici <- names(conteggio[conteggio == 1])
database = database[!(database[, 1] %in% valori_unici), , drop = FALSE]
rm(conteggio, valori_unici) # Rimuovo dalla memoria i dati che non servono

df = database[, c("ID", "1h_PREC", "1h_YY")] # 1h Precipitazioni
colnames(df) = c("ID", "PREC_1h", "Anno_1h")
summary(df) # 27 valori mancanti
df = df[complete.cases(df[, "PREC_1h"]), ] # Elimino valori mancanti
df = as.data.table(df)
dataset_1h = dcast(df, Anno_1h ~ ID, value.var = "PREC_1h")

## Estraggo il dataset
# Se periodo 2000-2020
# data_0020 <- dataset_1h[dataset_1h$Anno_1h >= 2000 & dataset_1h$Anno_1h <=
#                               2020, ]
# da qui in poi analogo
dataset_completo <- dataset_1h %>% select_if(~!all(is.na(.)))
dataset_completo = dataset_completo %>%
  select(names(which(!colSums(is.na(dataset_completo)) ==
    (nrow(dataset_completo)-1))))
dataset_completo = dataset_completo[, -1]
coord = anagrafica[, c("X", "Y", "QUOTA_1km")]

```

```

rownames(coord) = anagrafica[, "ID"]
coord = coord[intersect(rownames(coord), colnames(dataset_completo)),
              drop = FALSE]
dataset_completo = data.matrix(dataset_completo)
coord = data.matrix(coord)
coord = coord/1000 # in km

data = df %>% filter(ID %in% colnames(dataset_completo))
data = merge(anagrafica[, c("ID", "X", "Y", "QUOTA_1km")], data, by = "ID")
ggplot(data, aes(x = X, y = Y, color = data[, "QUOTA_1km"])) +
  geom_point(size = 3) + scale_color_gradient2(low = "limegreen", mid =
  "gold", high = "brown", midpoint = 1650, name = "Quota [m]") +
  labs(title = "Mappa con altitudine")
hist(coord[, "QUOTA_1km"]*1000, main = "Istogramma del valore di altitudine",
      xlab = "Quota [m]", breaks = 12,
      col = colorRampPalette(c("limegreen", "gold", "brown"))(18))

summary(data)
sd(data$PREC_1h)

## Analisi esplorativa
symbolplot(dataset_completo, coord, which = "mean", plot.legend = TRUE,
           scale = 500)
symbolplot_sd(dataset_completo, coord, plot.legend = TRUE, scale = 400)
# Dipendenza spaziale stimata empiricamente
par(mfrow = c(1,3))
fmadogram(dataset_completo, coord, which = "ext", col = "lightgray")
fmadogram(dataset_completo, coord, which = "ext", n.bins = 12000, add = TRUE)
fmadogram(dataset_completo, coord, which = "ext", n.bins = 12000)
abline(v=100, col="red")
abline(a=1.9, b=0, col="red")
concprob(dataset_completo, coord, ylim = c(0,1), col = "lightgray")
concprob(dataset_completo, coord, add = TRUE, n.bins = 12000,
         ylim = c(0, 1))
abline(v=100, col="red")
abline(a=0.1, b=0, col="red")
par(mfrow=c(1,1))

## METODOLOGIA 1
# Selezione trend surface
loc.form <- scale.form <- shape.form <- y ~ 1
M0 = fitspatgev(dataset_completo, coord, loc.form, scale.form, shape.form)

loc.form <- y ~ X + Y
scale.form <- shape.form <- y ~ 1
M1 = fitspatgev(dataset_completo, coord, loc.form, scale.form, shape.form)

loc.form <- y ~ X + Y + QUOTA_1km
scale.form <- shape.form <- y ~ 1
M1_a = fitspatgev(dataset_completo, coord, loc.form, scale.form, shape.form)

loc.form <- y ~ X + Y
scale.form <- y ~ X

```

```

shape.form <- y ~ 1
M2_x = fitspatgev(dataset_completo, coord, loc.form, scale.form, shape.form)

loc.form <- y ~ X + Y + QUOTA_1km
scale.form <- y ~ X
shape.form <- y ~ 1
M2_xa = fitspatgev(dataset_completo, coord, loc.form, scale.form, shape.form)

loc.form <- y ~ X + Y + QUOTA_1km
scale.form <- y ~ X + QUOTA_1km
shape.form <- y ~ 1
M2_xa2 = fitspatgev(dataset_completo, coord, loc.form, scale.form,
shape.form)

loc.form <- y ~ X + Y
scale.form <- y ~ Y
shape.form <- y ~ 1
M2_y = fitspatgev(dataset_completo, coord, loc.form, scale.form, shape.form)

loc.form <- y ~ X + Y + QUOTA_1km
scale.form <- y ~ Y
shape.form <- y ~ 1
M2_ya = fitspatgev(dataset_completo, coord, loc.form, scale.form, shape.form)

loc.form <- y ~ X + Y + QUOTA_1km
scale.form <- y ~ Y + QUOTA_1km
shape.form <- y ~ 1
M2_ya2 = fitspatgev(dataset_completo, coord, loc.form, scale.form,
shape.form)

loc.form <- scale.form <- y ~ X + Y
shape.form <- y ~ 1
M2 = fitspatgev(dataset_completo, coord, loc.form, scale.form, shape.form)

loc.form <- y ~ X + Y + QUOTA_1km
scale.form <- y ~ X + Y
shape.form <- y ~ 1
M2_a = fitspatgev(dataset_completo, coord, loc.form, scale.form, shape.form)

loc.form <- y ~ X + Y + QUOTA_1km
scale.form <- y ~ X + Y + QUOTA_1km
shape.form <- y ~ 1
M2_a2 = fitspatgev(dataset_completo, coord, loc.form, scale.form, shape.form)

loc.form <- scale.form <- shape.form <- y ~ X + Y
M3 = fitspatgev(dataset_completo, coord, loc.form, scale.form, shape.form)

TIC(M0, M1, M1_a, M2_x, M2_xa, M2_xa2, M2_y, M2_ya, M2_ya2, M2, M2_a, M2_a2,
M3) # The best model is M2_xa2
anova(M2_xa2, M2_xa) # M2_xa2

M2_xa2
loc.form <- y ~ X + Y + QUOTA_1km

```

```

scale.form <- y ~ X + QUOTA_1km
shape.form <- y ~ 1

# Selezione modello max-stable
start.trend <- as.list(M2_xa2$fitted)
start <- c(list(range = 150, smooth = 0.2), start.trend)
schlat <- fitmaxstab(dataset_completo, coord, "powexp", loc.form, scale.form,
                    shape.form, nugget = 0, method = "nlm",
                    start = start, typsize = unlist(start))
start <- c(list(range = 13, smooth = 0.24), start.trend)
brown <- fitmaxstab(dataset_completo, coord, "brown", loc.form, scale.form,
                   shape.form, start = start, method = "nlm",
                   typsize = unlist(start))
start <- c(list(range = 150, smooth = 0.2, DoF = 1), start.trend)
extt <- fitmaxstab(dataset_completo, coord, "tpowexp", loc.form, scale.form,
                  shape.form, nugget = 0, method = "nlm",
                  start = start, typsize = unlist(start))

TIC(schlat, brown, extt) # Brown-Resnick

# Valutazione modello
# Dipendenza spaziale
par(mfrow=c(1,3))
fmadogram(dataset_completo, coord, which = "ext", col="lightgray")
fmadogram(dataset_completo, coord, brown, which = "ext", add=TRUE,
n.bins=12000)
fmadogram(dataset_completo, coord, brown, which = "ext", n.bins=12000)
fmadogram(dataset_completo, coord, extt, which = "ext", n.bins=12000,
col=c('black','limegreen'), add=TRUE)
fmadogram(dataset_completo, coord, schlat, which = "ext", n.bins=12000,
col=c('black','purple'), add=TRUE)
concprob(dataset_completo, coord, ylim=c(0,1), col="lightgray")
concprob(dataset_completo, coord, brown, ylim=c(0,1), add=TRUE,
n.bins=12000)
concprob(dataset_completo, coord, extt, ylim=c(0,1), add=TRUE, n.bins=12000,
col=c('black','limegreen'))
concprob(dataset_completo, coord, schlat, ylim=c(0,1), add=TRUE,
n.bins=12000, col=c('black','purple'))

brown$ext.coeff(c(5,10,20,30,50,100))
brown$conc.prob(c(5,10,20,30,50,100))
extt$ext.coeff(c(5,10,20,30,50,100))
extt$conc.prob(c(1,5,10,20,30,50,100))
schlat$ext.coeff(c(5,10,20,30,50,100))
schlat$conc.prob(c(5,10,20,30,50,100))

brown # Parametri stimati
set.seed(42)
sites <- sample(1:nrow(coord), 4)
plot(coord, type = "n")
points(coord[-sites,])
points(coord[sites,], pch = c("1", "2", "3", "4"), col = "blue")
plot.maxstab(brown) # Valutazione locale

```

```

par(mfrow=c(1,1))

# Get the prediction for the GEV parameters
predictions <- predict(brown, punti[1:20000,], ret.per = 25)
predictions2 <- predict(brown, punti[20001:40000,], ret.per=25)
predictions3 <- predict(brown, punti[40001:59340,], ret.per=25)
predictions = rbind(predictions, predictions2, predictions3)
rm(predictions2, predictions3)

# 25-year return level
return_data <- data.frame(X = predictions[, "X"], Y = predictions[, "Y"],
ReturnLevel = predictions[,"Q25"], stderror = predictions[,"Q25std.err"])
ggplot(return_data, aes(x = X, y = Y, color = ReturnLevel)) +
  geom_point(size = 3) +
  scale_color_gradient(low = "blue", high = "red", name = "Return Level") +
  labs(title = "25-Year Return Levels") +
  theme_minimal()
ggplot(return_data, aes(x = X, y = Y, color = stderror)) +
  geom_point(size = 3) +
  scale_color_gradient(low = "blue", high = "red", name = "Standard Error")
+ labs(title = "Standard Error of 25-Year Return Levels") +
  theme_minimal()

rm(dataset_completo)

## METODOLOGIA 2 (CROSS-VALITADION)
# Anni 1960-2020
loc.form = y ~ X + Y + QUOTA_1km # Altitudine importante
scale.form = c(y ~ X, y ~ Y, y ~ X + Y, y ~ X + QUOTA_1km,
              y ~ Y + QUOTA_1km, y ~ X + Y + QUOTA_1km)
shape.form = y ~ 1 # per non complicare troppo il modello
modello = list("powexp", "brown", "tpowexp") # Schlather, Brown-Resnick,
Extremal-t
parametri = expand.grid(scale.form, modello)

# k = 10 fold Cross-Validation: partizione stazioni pluviometriche
set.seed(42)
test = createDataPartition(coord[,"X"], p=0.1, list = FALSE, times = 10)
# Verifico distribuzione delle partizioni (per tutte e 10)
ggplot(data.frame(coord[test[,1],]), aes(x = X, y = Y, color =
  coord[test[,1], "QUOTA_1km"]*1000)) + geom_point(size = 3) +
  scale_color_gradient2(low = "limegreen", mid="gold", high = "brown",
  midpoint = 1650, name = "Quota [m]") +
  labs(title = "Mappa con altitudine per il test set 1")
hist(coord[test[,1],"QUOTA_1km"]*1000, xlab = "Quota [m]", breaks = 12,
  main = "Istogramma del valore di altitudine per il test set 1",
  col = colorRampPalette(c("limegreen", "gold", "brown"))(18))
# i 10 folds sono equilibrati: località sparse e con diverse altitudini

errors_prec = matrix(0, nrow = nrow(parametri), ncol = 3)
colnames(errors_prec) = c("RMSE", "Rsquared", "MAE")
# semi-variogramma Brown-Resnick
semiv = function(x,range, smooth) {(x/range)^smooth}

```

```

for(i in 1:nrow(parametri))
{
  for(k in 1:10)
  {
    coord_train = coord[-test[,k],]
    coord_test = coord[test[,k],]
    data_train = data_6020[ , intersect(rownames(coord_train),
                                         colnames(data_6020)), drop = FALSE]
    data_test = data_6020[, intersect(rownames(coord_test),
                                      colnames(data_6020)), drop = FALSE]

    M = fitspatgev(data_train, coord_train, loc.form = loc.form,
                  scale.form = parametri[[i,1]], shape.form = shape.form)
    if(parametri[[i,2]] == "powexp"){
      start = c(list(range = 150, smooth = 0.2))
    }else if(parametri[[i,2]] == "brown"){
      start = c(list(range = 13, smooth = 0.24))
    }else{
      start = c(list(range = 150, smooth = 0.2, DoF = 1))
    }
    start = c(start, as.list(M$fitted))
    if(parametri[[i,2]] == "brown"){
      maxstab = fitmaxstab(data_train, coord_train,
                          cov.mod = parametri[[i,2]], loc.form = loc.form,
                          scale.form = parametri[[i,1]],
                          shape.form = shape.form, method = "nlm",
                          start = start, tysize = unlist(start))
    }else{
      maxstab = fitmaxstab(data_train, coord_train,
                          cov.mod = parametri[[i,2]], loc.form = loc.form,
                          scale.form = parametri[[i,1]],
                          shape.form = shape.form, method = "nlm",
                          start = start, tysize = unlist(start),
                          nugget = 0)
    }

    param_pred = predict(maxstab, coord_test, ret.per = c(10, 25, 50))
    new_pred = matrix(0, nrow = nrow(data_test), ncol = nrow(coord_test))
    for(j in 1:nrow(coord_test))
    {
      set.seed(42)
      # new_pred[,j] = rgev(nrow(data_test), param_pred[j,"loc"],
      #                    param_pred[j,"scale"], param_pred[j,"shape"])
      if(parametri[[i,2]] == "powexp"){
        new_pred[,j] = frech2gev(rmev(n= nrow(data_test), model="schlather",
                                     sigma=matrix(schlat$cov.fun(sqrt(sum(
                                       coord_test[j,]-c(0,0,0))))),
                                     coord=t(coord_test[1,]),
                                     range=schlat$param["range"],
                                     smooth=schlat$param["smooth"],
                                     nugget=schlat$param["nugget"]),
                                loc=param_pred_s[1,"loc"], scale=param_pred_s[1,"scale"],
                                shape=param_pred_s[1,"shape"])
      }
    }
  }
}

```

```

}else if(parametri[[i,2]] == "brown"){
  new_pred[,j] = frech2gev(rmev(n=10, model="br",
    vario=semiv,coord=t(coord_test[1,]),
    range=brown$param["range"],
    smooth=brown$param["smooth"]),
    loc=param_pred[1,"loc"], scale=param_pred[1,"scale"],
    shape=param_pred[1,"shape"])
}else{
  new_pred[,j] = frech2gev(rmev(n=10, model="xstud",
    param= extt$param["DoF"],
    sigma=matrix(maxstab$cov.fun(sqrt(sum(
      coord_test[j,]-c(0,0,0))),
      ncol=1,nrow=1)),
    coord=t(coord_test[1,]),
    range=extt$param["range"],
    smooth=extt$param["smooth"],
    DoF=extt$param["DoF"]),
    loc=param_pred[1,"loc"],
    scale=param_pred[1,"scale"],
    shape=param_pred[1,"shape"])
}
}
errors_prec[i,] = errors_prec[i,] +
  postResample(apply(new_pred, 2, mean),
    apply(data_test, 2, mean, na.rm = TRUE))
}
}
# circa 36 ore
errors_prec = errors_prec/10 # media degli errori della previsione delle
# precipitazioni per ogni modello sui 10 folds
save(erroros_prec, file = "crosslh6020.Rdata")
#load("crosslh6020.Rdata")
errors_prec

# modello selezionato
scale.form = parametri[[which.min(errors_prec[,1],1)]]
# da qui si fitta modello max-stable
# valutazione e plot analoghi a prima

```

Bibliografia

- [1] Beirlant, J., Goegebeur, Y., Segers, J. e Teugels, J. (2004) *Statistics of Extremes: Theory and Applications*. Wiley.
- [2] Coles, S. G. (2001) *An Introduction to Statistical Modeling of Extreme Values*. Londra: Springer.
- [3] Claps P., Laio F., Allamano P., Gli Estremi degli Estremi: Eventi eccezionali e piogge di progetto, Atti dei Convegni Lincei n.270: ‘Cosa non funziona nella difesa dal rischio idro-geologico nel nostro paese? Analisi e rimedi’, Accademia dei Lincei, Roma, 23 marzo 2012, ACL, VOL 270, ISBN: 978-88-218-1065-7; ISSN: 0391-805X, 2013.
- [4] Castillo, E. (1988) *Extreme Value Theory in engineering*. Boston: Academic Press.
- [5] Reiss, R.-D. e Thomas, M. (2007) *Statistical Analysis of Extreme Values with Applications to Insurance, Finance, Hydrology and Other Fields*. Boston, Berlino: Birkhäuser Basel.
- [6] Cont, R. (2001) ‘Empirical properties of asset returns: stylized facts and statistical issues’, *Quantitative Finance*, 1(2), pp. 223-236.
- [7] Kotz, S. e Nadarajah, S. (2000) *Extreme Value Distributions: Theory and Applications*. Londra: Imperial College Press.
- [8] Koutsoyiannis, D. (2004) ‘Statistics of extremes and estimation of extreme rainfall: I. Theoretical investigation / Statistiques de valeurs extrêmes et estimation de précipitations extrêmes: I. Recherche théorique’, *Hydrological Sciences Journal*, 49(4), p. 590. doi: 10.1623/hysj.49.4.575.54430.
- [9] Gilleland, E. e Katz, R. W. (2016) ‘extRemes 2.0: An Extreme Value Analysis Package in R’, *Journal of Statistical Software*, 72(8), pp. 1-39. doi: 10.18637/jss.v072.i08
- [10] De Haan, L. e Resnick, S. I. (1977) ‘Limit theory for multivariate sample extremes’, *Zeit. Wahrscheinl.-theorie* 40, pp. 317-337. doi: 10.1007/BF00533086
- [11] Ribatet, M. (2017) ‘Modelling spatial extremes using max-stable processes’, Book chapter in *Nonlinear and stochastic climate dynamics*. Editors: Franzke, C. and O’Kane, T.
- [12] Davison, A., Padoan, S. e Ribatet, M. (2012) ‘Statistical modelling of spatial extremes’, *Statistical Science*, 27(2), pp. 161-186. doi: 10.1214/11-STS376
- [13] Ribatet, M., Dombry, C. e Oesting, M. (2015) ‘Spatial extremes and max-stable processes’, Book chapter in *Extreme value modeling and risk analysis: Methods and applications*. Editors: Dey, D.K. and Yan, J.
- [14] Ribatet, M. (2013) ‘Spatial extremes: Max-stable processes at work’, *Journal de la Société Française de Statistique*, 154(2), pp. 156-177.
- [15] Padoan, S. A. (2011), *Max-Stable Processes and Their Applications*. Job Market Paper. Univerisutà di Padova.
- [16] de Haan, L. (1984) ‘A spectral representation for max-stable processes’, *The Annals of Probability*, 12(4), pp. 1194–1204. doi: 10.1214/aop/1176993148
- [17] Smith, R. L. (1990) ‘Max-stable processes and spatial extreme’, Unpublished manuscript.
- [18] Schlather, M. (2002) ‘Models for stationary max-stable random fields’, *Extremes*, 5(1), pp. 33–44. doi: 10.1023/A:1020977924878
- [19] Opitz, T. (2012) ‘Extremal-t process: Elliptical domain of attraction and a spectral representation’, *Journal of Multivariate Analysis*, 122, pp. 409-413. doi: 10.1016/j.jmva.2013.08.008
- [20] Davison, A. C., Huser, R. e Thibaud, E. (2019) ‘Spatial extremes’, In *Handbook of Environmental and Ecological Statistics*. Editors A. E. Gelfand, M. Fuentes, J. A. Hoeting and R. L. Smith, CRC Press.

- [21] Cooley, D., Naveau, P. e Poncet, P. (2006) ‘Variograms for spatial max-stable random fields’. In: Bertail, P., Soulier, P., Doukhan, P. (eds) *Dependence in Probability and Statistics*. Lecture Notes in Statistics, vol 187. Springer, New York, NY. doi: 10.1007/0-387-36062-X_17
- [22] Dombry, C., Ribatet, M. e Stoev, S. (2018) ‘Probabilities of concurrent extremes’, *Journal of the American Statistical Association (Theory & Methods)*, 113(524), pp. 1565-1582.
- [23] Oesting, M., Ribatet, M. e Dombry, C. (2015) ‘Simulation of max-stable processes’, Book chapter in *Extreme value modeling and risk analysis: Methods and applications*. Editors: Dey, D.K. and Yan, J.
- [24] Dombry, C., Engelke, S. e Oesting, M. (2016). ‘Exact Simulation of Max-Stable Processes’, *Biometrika*, 103 (2), pp. 303–317. doi: 10.1093/biomet/asw008.
- [25] Padoan, S. A., Ribatet, M. e Sisson, S. A. (2010) ‘Likelihood-based inference for max-stable processes’, *Journal of the American Statistical Association*, 105 (489), pp. 263–277. doi: 10.1198/jasa.2009.tm08577
- [26] Nicolet, G., Eckert, N., Morin, S. e Blanchet, J. (2017), ‘A multi-criteria leave-two-out cross-validation procedure for max-stable process selection’, *Spatial Statistics*, 22(1), pp. 107-128. doi: 10.1016/j.spasta.2017.09.004
- [27] Blanchet, J. e Davison, A. C. (2011) ‘Spatial modeling of extreme snow depth’, *The Annals of Applied Statistics*, 5(3), pp. 1699–1725. doi: 10.1214/11-AOAS464
- [28] Gaume, J., Eckert, N., Chambon, G., Naaim, M. e Bel, L. (2013) ‘Mapping extreme snowfalls in the French Alps using max-stable processes’, *Water Resources Research*, 49, pp. 1079-1098. doi: 10.1002/wrcr.20083
- [29] Wikipedia, Numeri di Bell. Disponibile su: https://it.wikipedia.org/wiki/Numeri_di_Bell
- [30] Wikipedia, Metodo Jackknife. Disponibile su: https://it.wikipedia.org/wiki/Metodo_jackknife (Accesso: 12 giugno 2024).
- [31] Mazzoglio, P., Butera, I. e Claps, P. (2020) ‘I²-RED: A Massive Update and Quality Control of the Italian Annual Extreme Rainfall Dataset’, *Water*, 12(12), 3308. doi: 10.3390/w12123308
- [32] Istituto Superiore per la Protezione e la Ricerca Ambientale (ISPRA). Disponibile su: <https://www.isprambiente.gov.it/it>
- [33] Agenzia Regionale per la Protezione Ambientale (ARPA) Piemonte. Disponibile su: <https://www.arpa.piemonte.it/>
- [34] Farr, T. G., Rosen, P. A., Caro, E., Crippen, R., Duren, R. Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., Seal, D., Shaffer, S., Shimada, J., Umland, J., Werner, M., Oskin, M., Burbank, D. e Alsdorf, D. (2007) ‘The Shuttle Radar Topography Mission’, *Reviews of Geophysics*, 45(2), pp. 1-33. doi: 10.1029/2005RG000183
- [35] ARPA Piemonte, Clima ed indicatori climatici del Piemonte. Disponibile su: http://www.idrologia.polito.it/web2/open-data/cd_Dati_Regione_Piemonte/50_anni_dati_meteo_Piemonte/clima_ed_indicatori.htm (Accesso: 15 giugno 2024)
- [36] Mazzoglio, P., Butera, I. e Claps, P. (2023) ‘A local regression approach to analyze the orographic effect on the spatial variability of sub-daily rainfall annual maxima’, *Geomatics, Natural Hazards and Risk*, 14(1). doi: 10.1080/19475705.2023.2205000
- [37] Daly, C., Gibson, W. P., Taylor, G. H., Johnson, G. L. e Pasteris, P. (2002) ‘A knowledge-based approach to the statistical mapping of climate’, *Climate Research, Clim Res*, 22, pp. 99-113, doi:10.3354/cr022099
- [38] Dutang, C. (2023). CRAN Task View: Extreme Value Analysis. Version 2023-11-04. URL <https://CRAN.R-project.org/view=ExtremeValue>.
- [39] Belzile, L., Dutang, C., Northrop, P. e Opitz, T. (2023) ‘A modeler’s guide to extreme value software’, *Extremes*, 26, pp. 595-638. doi:10.1007/s10687-023-00475-9.

- [40] Gilleland, E., Ribatet, M. e Stephenson, A. G. (2013) ‘A comparative software review for extreme value analysis’, *Extremes*, 16(1), pp. 103-119. doi: 10.1007/s10687-012-0155-0
- [41] Ribatet, M. (2022). SpatialExtremes: Modelling Spatial Extremes. R package version 2.1-0. Disponibile su: <http://spatialextremes.r-forge.r-project.org/>
- [42] Ribatet, M. (2009) ‘A User’s Guide to the SpatialExtremes Package’. In SpatialExtremes (2.1-0). Retrieved from <https://CRAN.R-project.org/package=SpatialExtremes>
- [43] Belzile, L. R. (2023). mev: Modelling of Extreme Values. R package version 1.16. Disponibile su: <https://lbelzile.github.io/mev/>.