

POLITECNICO DI TORINO

Corso di Laurea Magistrale in Ingegneria per l'Ambiente e il territorio.

Indirizzo: Gestione dei Rischi Naturali.



Politecnico di Torino

Tesi di Laurea Magistrale

**“Analisi spaziale della frequenza di eventi estremi
sulla Greater Alpine Region.”**

Relatori:

Prof. Alberto Viglione
Prof. Daniele Ganora

Candidato:

BUCCA Filippo

ANNO ACCADEMICO 2024/2025

ABSTRACT [IT]

La valutazione dei danni causati da eventi idrologici estremi su vaste regioni rappresenta una sfida complessa e di notevole interesse sociale ed economico. Nel presente lavoro di tesi è stato applicato un modello statistico-matematico capace di classificare gli eventi regionali in base a un apposito indice di danno e di simulare numerosi scenari futuri, con implicazioni dirette per la gestione del rischio idrogeologico. In particolare, lo scopo principale del progetto consiste nell'indagare la frequenza di eventi alluvionali che possono causare importanti danni in regioni vaste con il fine di poter applicare i risultati dell'analisi, ad esempio, in ambito assicurativo. Infatti, sulla base delle simulazioni di eventi estremi a scala regionale e dell'esposizione al rischio dei beni, gli istituti assicurativi possono valutare in modo più accurato la distribuzione delle risorse per ridurre il rischio di insolvenza. L'area analizzata è la GAR (Greater Alpine Region) caratterizzata da elevata complessità orografica, idrologica e di particolare interesse climatico, dove la combinazione di ambiente naturale e intervento umano determina un rischio idrogeologico generalmente elevato. I dati idrologici impiegati per questo studio, relativi alle portate fluviali del complesso reticolo idrografico dell'area, provengono da una pluralità di enti e istituzioni di ricerca. In tale contesto, si è sviluppato un approccio metodologico mediante procedure di analisi implementate in linguaggio R e in ambiente GIS. Questo ha permesso di estrarre gli eventi più gravosi del passato dalle serie di dati delle portate giornaliere e di impiegare questi come base di possibili scenari futuri così da costruire uno strumento utile sia per l'analisi deterministica del rischio sia per la valutazione di possibili prossimi impatti. Il progetto di tesi è diviso in tre macro-capitoli: il primo si concentra su concetti teorici dell'idrologia statistica; il secondo capitolo ha lo scopo di individuare, attraverso analisi multivariate, i quindici eventi regionali con indice di danno maggiore che hanno colpito la GAR dal 1899 al 2024; infine nel terzo capitolo è stata effettuata un'analisi probabilistica con simulazioni con metodo Monte Carlo il cui scopo è quello di generare eventi alluvionali con determinati tempi di ritorno, basandosi sulle portate osservate alle stazioni presenti sul territorio della GAR. Tale approccio ha consentito di esplorare scenari ipotetici che forniscono indicazioni utili sulla capacità dei sistemi idrici di sostenere situazioni di stress. Le verifiche finali di confronto tra eventi osservati e simulati hanno fornito riscontri incoraggianti presentando una buona correlazione tra le loro proprietà statistiche. L'applicazione di questo procedimento evidenzia, quindi, come i metodi probabilistici possano integrare in maniera efficace i dati storici, permettendo un'analisi più robusta del rischio idrogeologico e di conseguenza una valutazione più consapevole dei potenziali costi assicurativi legati a eventi

catastrofali. In questa prospettiva, il lavoro costituisce una base solida per sviluppi futuri orientati sia alla ricerca accademica sia al supporto di politiche di adattamento e mitigazione dei rischi naturali nella regione alpina.

ABSTRACT [EN]

The assessment of damage caused by extreme hydrological events over large regions represents a complex challenge of considerable social and economic interest. In this thesis work, a statistical-mathematical model was applied capable of classifying regional events based on a specific damage index and simulating numerous future scenarios, with direct implications for hydrogeological risk management. In particular, the main aim of the project is to investigate the frequency of flood events that can cause significant damage in large regions with the aim of being able to apply the results of the analysis, for example, in the insurance sector. Indeed, based on simulations of extreme events at the regional scale and the risk exposure of assets, insurance institutions can more accurately assess the distribution of resources to reduce the risk of insolvency. The area analyzed is the GAR (Greater Alpine Region) characterized by high orographic, hydrological complexity and special climatic interest, where the combination of natural environment and human intervention determines a generally high hydrogeological risk. The hydrological data employed for this study, related to the river flow rates of the complex hydrographic network of the area, come from a plurality of research bodies and institutions. In this context, a methodological approach was developed using analysis procedures implemented in R language and in a GIS environment. This allowed us to extract the most burdensome events of the past from the daily flow data series and to use these as the basis of possible future scenarios so as to build a useful tool both for the deterministic analysis of risk and for the evaluation of possible next impacts. The thesis project is divided into three macro-chapters: the first focuses on theoretical concepts of statistical hydrology; the second chapter aims to identify, through multivariate analyses, the fifteen regional events with the highest damage index that affected the GAR from 1899 to 2024; finally, in the third chapter, a probabilistic analysis was carried out with simulations using the Monte Carlo method whose aim is to generate flood events with certain return times, based on the flow rates observed at the stations present in the GAR territory. Such an approach has allowed to explore hypothetical scenarios that provide useful guidance on the ability of water systems to sustain stressful situations. The final checks comparing observed and simulated events provided encouraging feedback by presenting a good correlation between their statistical properties. The application

of this procedure highlights, therefore, how probabilistic methods can effectively integrate historical data, allowing a more robust analysis of hydrogeological risk and consequently a more conscious evaluation of the potential insurance costs linked to catastrophe events. From this perspective, the work constitutes a solid basis for future developments oriented both towards academic research and the support of natural risk adaptation and mitigation policies in the Alpine region.

INDICE

1. Introduzione.....	7
2. L'analisi idrologica: inferenza statistica.....	15
2.1 Elaborazioni preliminari dei dati.....	21
2.2 Scelta dei modelli probabilistici.....	23
2.3 Stima dei parametri dei modelli.....	28
2.4 Verifica dei modelli con test di adattamento.....	34
2.5 Stima dell'incertezza.....	45
3. Estrazione e visualizzazione dei maggiori eventi storici della GAR attraverso analisi multivariate.....	54
3.1 Creazione indice di magnitudo degli eventi sulla GAR.....	55
3.2 Estrazione eventi regionali da picchi della serie storica dei D_m	59
3.3 Visualizzazione eventi regionali da picchi della serie storica dei D_m	69
3.4 Ricostruzione storico-documentale dei 15 eventi più gravosi.....	81
4. Analisi probabilistica di eventi simulati sulla GAR.....	98
4.1 Formazione di serie multivariate in idrologia.....	99
4.2 Parametrizzazione della distribuzione normale multivariata.....	103
4.3 Simulazione di dati multivariati normali e back-transformation.....	104
4.4 Verifica del modello multivariato.....	105
4.5 Eventi alluvionali simulati su scala internazionale (GAR).....	113
5. Conclusioni.....	139
6. Bibliografia e sitografia.....	140
7. Allegati.....	143

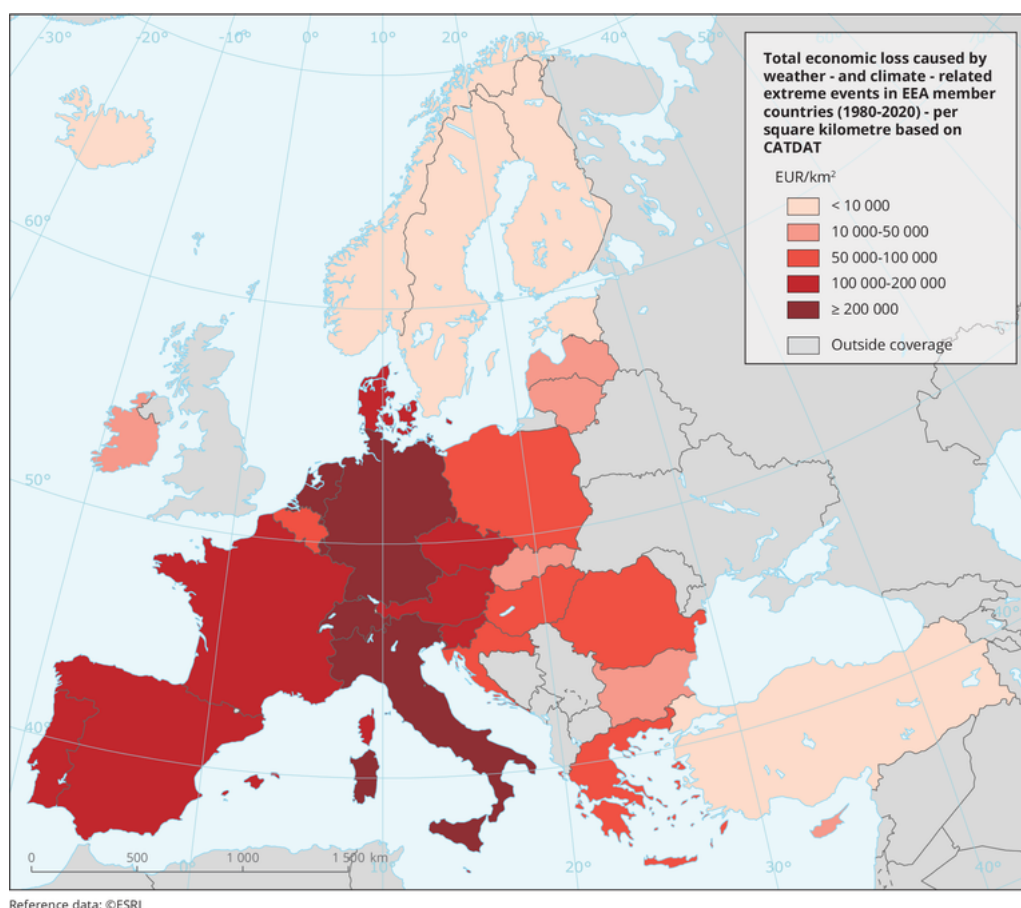
1. Introduzione

Gli eventi idrologici estremi costituiscono uno degli aspetti più critici e impattanti della variabilità climatica degli ultimi decenni. Con il termine “evento estremo”, secondo (Katz & Brown, 1992), si intende un fenomeno meteorologico che si colloca nella parte più alta della distribuzione statistica delle osservazioni, ovvero un accadimento raro, ma con potenziali conseguenze molto gravi. In questa categoria rientrano precipitazioni eccezionalmente intense che possono causare piene fluviali, tempeste violente ma anche siccità prolungate.

Gli eventi alluvionali e le portate estreme rivestono un ruolo centrale nell'analisi idrologica. Esse non solo rappresentano una minaccia diretta per la sicurezza delle popolazioni e la stabilità delle infrastrutture, ma costituiscono anche un campanello d'allarme delle modifiche in atto nel ciclo idrologico terrestre. Le variazioni nella frequenza e intensità degli eventi idrologici estremi sono infatti strettamente connesse ai cambiamenti climatici globali e alla crescente pressione, da parte dell'uomo, sugli ecosistemi fluviali.

Data l'importanza scientifica e sociale di tali fenomeni, si è reso indispensabile lo sviluppo di metodologie sempre più complesse per la loro analisi e, soprattutto, previsione. In questo quadro si colloca la presente tesi, che si propone di investigare la frequenza e distribuzione spaziale degli eventi estremi idrologici nella Greater Alpine Region (GAR), un'area che si è dimostrata essere molto sensibile e vulnerabile nel passato a questa tipologia di eventi.

Il Sesto Rapporto di Valutazione dell'IPCC (AR6, 2021) conferma con elevata sicurezza che il riscaldamento globale, già pari a circa 1,1 °C rispetto all'epoca preindustriale, sta intensificando il ciclo idrologico: ciò si traduce in una frequenza maggiore di precipitazioni eccezionali e di eventi meteorologici severi non solo in Europa ma in molte aree del pianeta. La rivista “*il Grand Continent*” riporta i seguenti dati dell'Agenzia Europea per l'Ambiente sui danni da eventi catastrofici: si stimano danni economici totali legati agli eventi meteorologici estremi tra il 1980 e il 2022 pari a 650 miliardi di euro solo in Europa, di cui oltre 100 miliardi di euro solo negli ultimi cinque anni. In figura 1 viene mostrato il grafico della perdita economica totale causata da eventi severi legati al clima e alle condizioni meteorologiche nei Paesi membri dell'EEA (1980–2020) per chilometro quadrato. L'Italia risulta essere nella categoria peggiore con oltre 200000 euro di danni per chilometro quadrato. Il report EEA (2022) sottolinea come le inondazioni costituiscano l'evento naturale più ricorrente e costoso per il continente, con danni economici complessivi di più di 500 miliardi di euro tra il 1980 e il 2020.



*Figura 1- Perdita economica totale causata da eventi estremi legati al clima e alle condizioni meteorologiche nei Paesi membri dell'EEA (1980–2020) – per chilometro quadrato, basata su CATDAT. Fonte: Briefing no. 21/2021 Title: **Economic losses and fatalities from weather- and climate-related events in Europe**; doi: 10.2800/530599.*

Inoltre, in Europa, numerosi studi hanno documentato una tendenza crescente nella frequenza e severità delle precipitazioni estreme negli ultimi decenni (Van den Besselaar et al., 2013). Proprio in quest'ultimo articolo viene descritto come gli estremi di precipitazione, sia giornalieri (RX1day) sia su più giorni (RX5day), abbiano mostrato una tendenza all'aumento in gran parte dell'Europa, in linea con l'incremento delle temperature e della capacità dell'atmosfera di trattenere vapore acqueo (Allen e Ingram, 2002; Willett et al., 2008). Si osservano comportamenti diversi nei trend di frequenza di eventi estremi tra Europa settentrionale e meridionale. In Europa settentrionale, i cambiamenti nelle precipitazioni estreme seguono andamenti simili a quelli della piovosità totale, con un incremento significativo della frequenza di eventi intensi (Trenberth et al., 2007). In Europa meridionale, invece, le precipitazioni totali mostrano un calo in estate e inverno, ma gli estremi di breve durata tendono comunque a intensificarsi nelle altre stagioni, indicando una crescente variabilità (Klein Tank e Können, 2003; Moberg et al., 2006). In generale, gli studi confermano che eventi con tempi di ritorno pluridecennali stanno diventando più frequenti, trasformandosi in eventi con tempi di ritorno molto più brevi: la

riduzione mediana del tempo di ritorno tra il primo e l'ultimo intervallo ventennale dell'analisi effettuata nell' *International Journal of Climatology* (Van den Besselaar et al., 2013) è risultata pari a circa il 21%, con variazioni comprese tra una diminuzione di circa il 2% e il 58%. Si comprendono quindi le conseguenze dirette dell'aumento della frequenza di eventi con tempi di ritorno pluridecennali sulla gestione del rischio idrologico e climatico (Fowler e Kilsby, 2003; Groisman et al., 2005).

Un dato importante riguarda il fatto che molte simulazioni climatiche concordano nell'indicare che, al crescere della temperatura media globale, aumenta anche la capacità dell'atmosfera di trattenere vapore acqueo (circa +7% per ogni grado di riscaldamento secondo la legge di Clausius-Clapeyron). Questo comporta precipitazioni più abbondanti e concentrate in brevi periodi di tempo, che incrementano la probabilità di piene improvvise e alluvioni lampo (Trenberth, 2011). Inoltre, l'aumento delle temperature ha un impatto diretto anche sullo scioglimento dei ghiacciai e sul regime nivale, con conseguenze particolarmente rilevanti nelle regioni montane (tipologia di territorio che caratterizza quasi totalmente l'area di studio di questo progetto). Lo scioglimento accelerato ovviamente è causa di problematiche serie in quanto modifica la stagionalità delle portate fluviali, incrementando i volumi di deflusso in primavera e riducendo la riserva idrica estiva, con effetti a cascata sulla disponibilità di risorse idriche e sulla frequenza di eventi estremi fluviali.

La regione alpina, quindi, rappresenta un hotspot climatico particolarmente sensibile. Le Alpi hanno registrato un aumento della temperatura media superiore di circa +2 °C rispetto ai livelli preindustriali, quindi quasi il doppio rispetto alla media globale (Beniston et al., 2018). Questo effetto particolare sulla regione alpina implica impatti idrologici particolarmente rilevanti: aumento della frequenza di eventi di piena e modifiche del regime nivale che causano maggiore instabilità idrogeologica. Infatti, secondo il quarto rapporto ISPRA (Istituto Superiore per la Protezione e la Ricerca Ambientale) sul dissesto idrogeologico in Italia del 2024 (Trigila A. et al., 2025), è stato rilevato un aumento del 15% della superficie del territorio italiano a pericolosità elevata per frane dei Piani di Assetto Idrogeologico – PAI, passando dai 55.400 km² del 2021 ai 69.500 km² del 2024, che corrispondono al 23% del territorio nazionale. Inoltre, la quasi totalità dei comuni italiani risulta essere a rischio frana, alluvione, erosione costiera o valanghe.

Gli effetti degli eventi estremi hanno forti ripercussioni soprattutto a livello economico e sociale. Di seguito vengono riportate alcune statistiche che descrivono ciò:

- secondo l'EEA (2022), i disastri naturali legati al clima hanno causato oltre 145000 vittime in Europa tra il 1980 e il 2020;
- le perdite economiche dirette ammontano a circa 500 miliardi di euro, di cui oltre il 40% attribuibili ad eventi come inondazioni e tempeste;
- l'assicuratore MunichRe (2023) riporta che il numero di eventi naturali estremi documentati annualmente è più che triplicato dagli anni '80 ad oggi.

Un caso emblematico è rappresentato dalle alluvioni in Germania e Belgio del luglio 2021, che hanno provocato più di 200 vittime e danni stimati in 30 miliardi di euro. Grazie a questi avvenimenti si evidenzia la necessità di migliorare i sistemi di allerta precoce e la difficoltà di prevedere eventi estremi a scala regionale.



*Figura 2- Immagine dell'evento alluvionale del 14 luglio 2021 che colpì con forza Germania e Belgio.
<https://www.meteoeradar.it/notizie-meteo/>*

Un altro esempio significativo fu l'alluvione del 2002 in Austria e Germania meridionale: record storico nella città di Passau, dove il Danubio ha raggiunto i 12.5 metri, livello mai toccato negli ultimi 500 anni (questo evento verrà analizzato nel dettaglio in seguito).



Figura 3- August-Hochwasser 2002: immagine di Passau durante l'inondazione del Danubio | BR24 | BR.de

Per regioni come la GAR, caratterizzate da elevata densità abitativa, attività economiche ad alta intensità energetica e infrastrutture critiche (dighe, reti di trasporto, centrali idroelettriche), il rischio idrologico rappresenta una sfida strategica. La GAR si estende attraverso otto Paesi europei e comprende una vasta gamma di climi, ecosistemi e morfologie fluviali. È un'area particolarmente vulnerabile agli effetti dei cambiamenti climatici, in quanto racchiude:

- bacini idrografici montani di piccole e medie dimensioni con tempi di corrivazione brevi;
- una forte interazione tra precipitazioni estreme e dinamiche geomorfologiche (frane, debris flow, mud flow, ecc.);
- un'elevata densità abitativa nelle vallate e nei principali centri urbani.

La regione di studio del progetto di tesi, inoltre, svolge un ruolo strategico nella disponibilità idrica dell'Europa centrale, poiché rappresenta la sorgente di alcuni dei principali fiumi europei (Danubio, Reno, Rodano, Po). Ciò rende evidente come eventi estremi in questo territorio possano avere ripercussioni che si estendono ben oltre i confini locali, influenzando la disponibilità idrica e la gestione del rischio in vaste aree dell'Europa centrale.

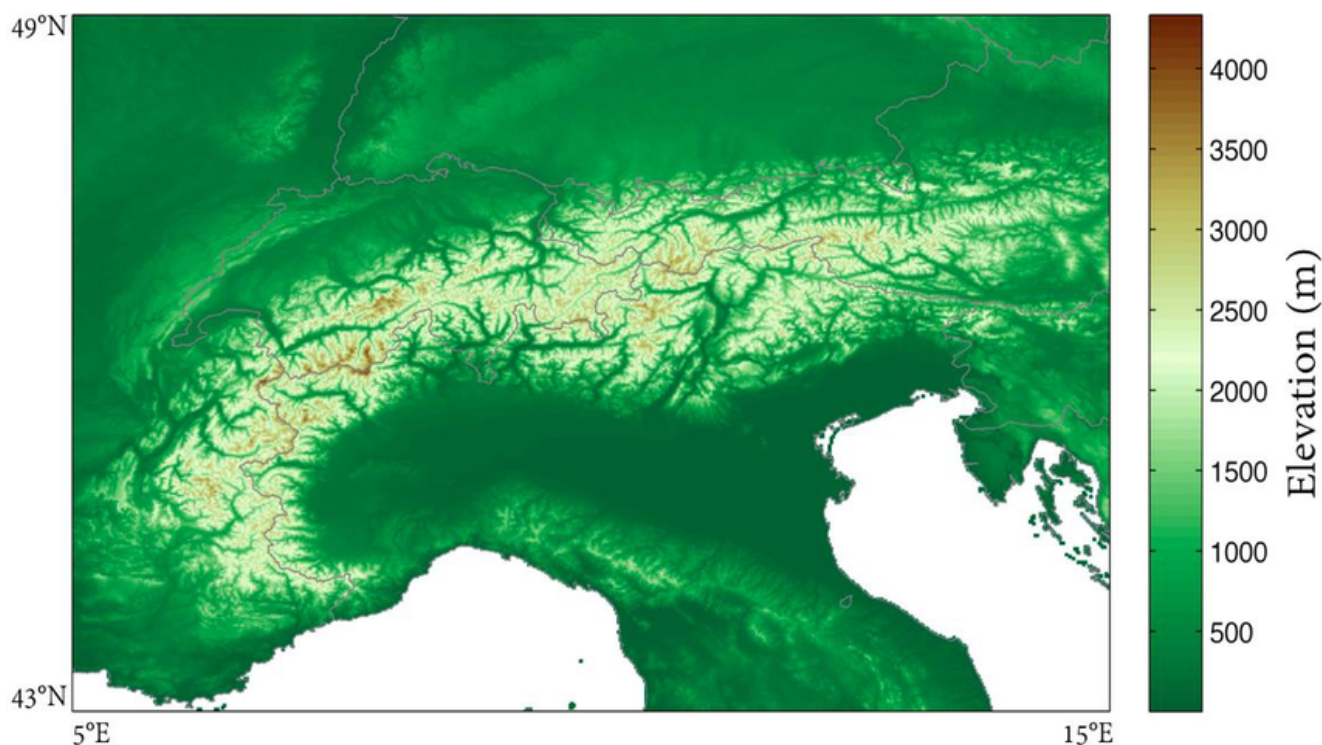


Figura 4- Greater Alpine Region (GAR) localizzata tra 5-15°E e 43-49°N. Development of a Glaciological Spatial Data Infrastructure to assess glaciers response to climatic fluctuations - Scientific Figure on ResearchGate. DOI:[10.13140/RG.2.1.1947.6244](https://doi.org/10.13140/RG.2.1.1947.6244)

La vulnerabilità del sistema idrologico alpino, unita alla crescente pressione esercitata dai cambiamenti climatici, impone dunque un'attenzione particolare all'analisi delle dinamiche degli eventi estremi in questa regione così vasta ed eterogenea.

Una sfida importante per questa tipologia di lavori riguarda la qualità e la continuità dei dati: molte serie storiche presentano lacune, eterogeneità nelle metodologie di misura o interruzioni dovute a modifiche e manutenzioni nelle stazioni di rilevamento. In questo contesto, l'utilizzo di tecniche statistiche robuste diventa indispensabile.

La costruzione di tecniche di analisi regionali come il metodo della piena indice, o il metodo VAPI e la selezione di eventi tramite metodo Peak Over Threshold (POT) consentono di estrarre

informazioni rilevanti anche da serie rumorose o incomplete (metodi che verranno approfonditi nei paragrafi successivi).

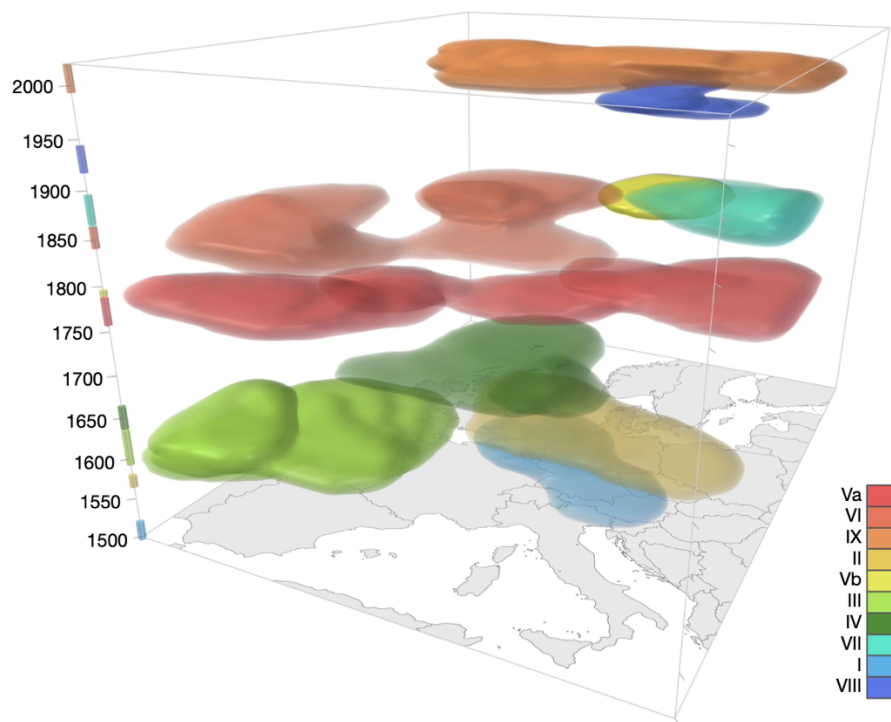


Figura 5- Periodi ricchi di alluvioni in Europa negli ultimi 500 anni. I periodi sono colorati in base al loro rango, con il rosso (periodo Va) che indica il più intenso e il blu (periodo VIII) che indica il meno intenso. Blöschl, G., Kiss, A., Viglione, A. et al. Current European flood-rich period exceptional compared with past 500 years. *Nature* **583**, 560–566 (2020).

Le tecniche di analisi spaziale e di analisi di tipo probabilistico offrono strumenti per caratterizzare la frequenza e la severità degli eventi estremi e, in particolare, l'approccio spaziale assume un ruolo strategico: come mostrato in precedenza non è sufficiente analizzare gli eventi in maniera isolata, ma è necessario valutare le loro caratteristiche in un contesto regionale più ampio. La GAR, per le sue dimensioni e complessità, richiede un'analisi che vada oltre la scala locale e che integri informazioni provenienti da più bacini e stazioni idrometriche. In questo modo, è possibile non solo individuare gli eventi estremi più rilevanti, ma anche valutare la loro estensione spaziale e la simultaneità con cui colpiscono più aree, aspetti cruciali per la gestione delle emergenze e la pianificazione infrastrutturale. In questo progetto di tesi, infatti, si è deciso di procedere nell'analisi dei dati idrologici tramite diverse strategie in cui i protagonisti sono stati:

- l'uso di distribuzioni di probabilità come la Pareto Generalizzata (GP) o la GEV consente di stimare i tempi di ritorno degli eventi estremi;

- l'applicazione di metodi GIS e di analisi spaziale che permettono di mappare la distribuzione geografica degli eventi e di individuare pattern regionali ricorrenti;

Alla luce del contesto descritto, il presente lavoro si propone quindi di:

1. analizzare la frequenza e distribuzione spaziale degli eventi idrologici estremi nella Greater Alpine Region, programmando in linguaggio R;
2. valutare la severità degli eventi attraverso indicatori statistici ovvero principalmente tempi di ritorno e danni stimati tramite un indice di danno creato appositamente;
3. integrare i dati storici con metodi di analisi spaziale per identificare le aree a maggiore vulnerabilità e simulare quindi come gli istituti di assicurazione approcciano il problema dei danni catastrofali;
4. evidenziare l'importanza di tali analisi come strumento di supporto alle strategie di mitigazione e adattamento ai cambiamenti climatici.

In particolare, il progetto di tesi è diviso in tre macro-capitoli:

- il primo si concentrerà sulla teoria di idrologia statistica dei metodi applicati nei due capitoli successivi avendo come riferimento ***“Applied hydrology”*** di Ven Te Chow, David R. Maidment, Larry W. Mays (ISBN 0-07-010810-2) e ***“Applied Statistics for Civil and Environmental Engineers”*** di Nathabandu T. Kottegoda, Renzo Rosso (ISBN: 978-1-405-17917-1);
- il secondo capitolo avrà lo scopo di individuare i quindici eventi regionali con indice di danno maggiore che hanno colpito la GAR dal 1899 al 2024, attraverso analisi multivariate con lo scopo di visualizzarli con mappe interattive di tipologia *leaflet* e collegarli a informazioni e immagini degli stessi eventi del passato;
- nel terzo capitolo verrà effettuata invece un'analisi probabilistica con simulazioni con metodo Monte Carlo il cui scopo è quello di generare eventi alluvionali simulati con determinati tempi di ritorno, basandosi sulle portate osservate dalle stazioni presenti sul territorio della GAR, in modo da poter individuare le aree e i bacini idrografici più vulnerabili a possibili eventi severi futuri.

2. L'analisi idrologica: inferenza statistica.

L'acqua è un composto fondamentale per la vita ed è presente in natura in diversi stati. È in continuo movimento attraverso la successione dei fenomeni di spostamento e circolazione dell'acqua all'interno dell'idrosfera, passando dagli oceani all'atmosfera e viceversa. A livello ingegneristico può essere studiata come risorsa: è possibile, infatti, quantificarla per diversi scopi, tra cui l'uso potabile, irriguo, industriale o per l'idroelettrico. Può anche però essere studiata come fonte di rischio: è possibile quantificare le portate di progetto per la progettazione di opere idrauliche come dighe e traverse o per calcolare i valori delle portate che si raggiungono durante una piena fluviale per poter calcolare l'altezza minima degli argini che dovranno contenerla.

La disciplina che studia il ciclo dell'acqua, gli eventi estremi, la quantificazione delle risorse idriche e come queste sono distribuite nel territorio è l'idrologia operativa.

I processi idrologici sono caratterizzati da una duplice natura: da un lato presentano componenti deterministiche, legate ai meccanismi fisici che ne regolano l'evoluzione, dall'altro lato manifestano una forte componente casuale, dovuta alla complessità e variabilità dei fattori atmosferici e ambientali. In molte circostanze, soprattutto quando la variabilità aleatoria è predominante rispetto a quella deterministica, l'analisi idrologica si concentra sugli aspetti probabilistici, trattando il fenomeno come puramente casuale.

Questo approccio è particolarmente rilevante nello studio degli eventi estremi, come le piene fluviali, che per loro natura non possono essere previsti con precisione puntuale, ma soltanto descritti in termini probabilistici. Allo stesso modo, esso risulta utile per variabili idrologiche aggregate su intervalli temporali estesi, come le precipitazioni annuali o per analisi su lunghi periodi come nel caso di questo progetto di tesi.

L'adozione di metodi statistici in idrologia verte sulla descrizione della loro variabilità attraverso parametri e funzioni matematiche. Questo approccio permette di trarre informazioni significative anche in presenza di serie temporali discontinue, rumorose o influenzate da fattori locali. La statistica, quindi, rappresenta una scienza descrittiva che permette di interpretare i dati idrologici e di stimare la probabilità di accadimento degli eventi estremi.

Uno degli scopi principali dell'idrologia statistica è studiare il dato di natura idrologica (portate, livelli, precipitazioni, ecc.) e legare quest'ultimo ad una probabilità di accadimento per definire il pericolo a cui è sottoposto un target: esso può essere un centro abitato o una singola abitazione.

In questo progetto di tesi si è scelto di studiare le portate dei corsi d'acqua che fluiscono nel territorio della GAR in quanto questa tipologia di dato non dipende dalla geometria degli alvei, a differenza dei livelli idrometrici, e permette quindi di essere indipendente dalle condizioni locali, ipotesi fondamentale per effettuare un'analisi spaziale.

I valori di portate più interessanti ai fini di questo studio sono i picchi degli idrogrammi che spesso però sono dati mancanti o affetti da incertezza: non è possibile conoscerli esattamente ma è possibile associarli a degli intervalli di valori che li caratterizzano.

Si trattano quindi queste tipologie di dati come variabili casuali o aleatorie ovvero come una variabile descritta da una distribuzione di probabilità. La distribuzione specifica la probabilità che un'osservazione della variabile cada all'interno di un intervallo prefissato di valori.

I dati idrologici impiegati per questo studio, relativi alle portate fluviali nella Greater Alpine Region (GAR), provengono da una pluralità di enti e istituzioni di ricerca, tra cui agenzie idrometeorologiche nazionali, università e centri di monitoraggio ambientale. Tale eterogeneità riflette la complessità geografica e amministrativa della regione di studio e consente di disporre di un set di dati ampio e in parte rappresentativo. A proposito di ciò, come si vedrà in seguito, soprattutto fino agli settanta/ottanta si hanno relativamente pochi dati a disposizione. Subito sotto viene riportato l'elenco degli istituti ed enti, divisi per stato, che gestiscono le stazioni idrometriche a cui fa riferimento il progetto:

- **Italia:**

- Arpa Piemonte <https://www.arpa.piemonte.it/>
- Regione Valle d'Aosta <https://cf.regione.vda.it/it/>
- Arpa Lombardia <https://www.arpalombardia.it>
- Arpa Veneto <https://www.arpa.veneto.it/>
- Arpa Friuli-Venezia Giulia <https://www.arpa.fvg.it/>
- Regione Friuli-Venezia Giulia <https://www.regione.fvg.it/>
- Servizio Prevenzioni Rischi Dighe della Provincia autonoma di Trento <https://www.floods.it/public/homepage.php>
- SUDTIROL Open Data Alto Adige <https://data.civis.bz.it/>
- UniTN (Università di Trento). <https://www.unitn.it/it>

- **Francia:** HydroPortail <https://hydro.eaufrance.fr/>
- **Svizzera:** Ufficio federale dell'ambiente UFAM <https://www.bafu.admin.ch/hydro-info>
- **Germania:** Länderübergreifendes Pegelportal <https://www.pegelportal.de/>
- **Austria:** WebGIS eHYD <https://ehyd.gv.at/#>
- **Slovenia:** Slovenian Environment Agency <https://www.arso.gov.si/>
- **GRDC:** Global Runoff Data Centre <https://grdc.bafg.de/>

L'insieme di dati provenienti da tutte le realtà menzionate precedentemente costituisce il dataset di base dell'analisi idrologica del progetto e rappresenta il campione della popolazione di dati ovvero un numero finito di n elementi, pari alla dimensione del campione, estratto dalla popolazione. In figura 6 viene riportata un immagine esplicativa tratta da **"Applied hydrology"** di Ven Te Chow et al. che mostra le differenze tra campione e popolazione (frequenza e probabilità), descritta quest'ultima dalla funzione di densità di probabilità (PDF). Grazie alla statistica si può passare dal campione alla popolazione tramite distribuzioni di probabilità, ovvero modelli statistici.

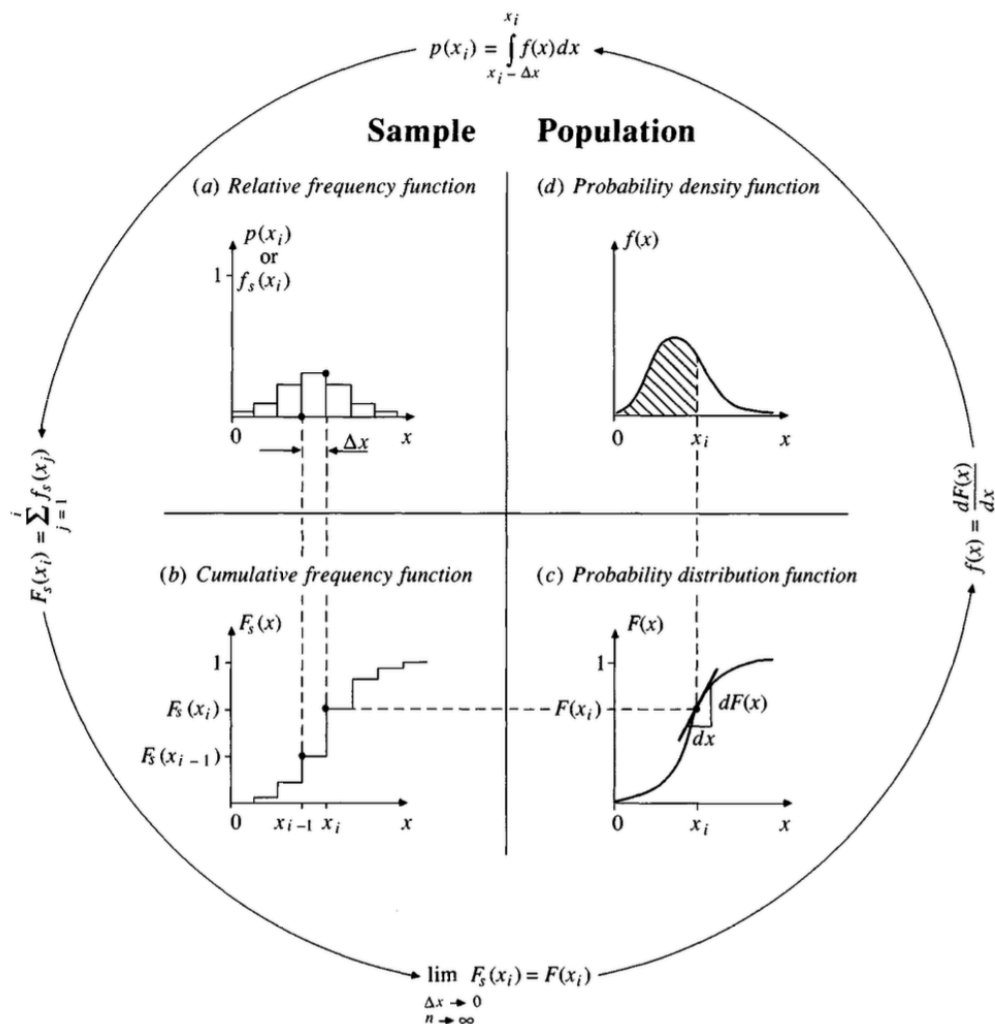


Figura 6- Funzioni di frequenza dai dati campionari e funzioni di probabilità dalla popolazione. **"Applied hydrology"** di Ven Te Chow, David R. Maidment, Larry W. Mays (ISBN 0-07-010810-2)

La procedura statistica che permette di adattare al campione una funzione continua che lo rappresenti (distribuzione statistica) è chiamata inferenza statistica.

Questa procedura si applica quando la funzione di frequenza empirica non è adeguata a rappresentare l'intera popolazione e questo accade quando la dimensione del campione è relativamente piccola rispetto alla popolazione stessa: in idrologia questo è il caso che si presenta nella maggior parte delle analisi in quanto, come già descritto in precedenza, le serie storiche di dati delle grandezze idrologiche spesso sono incomplete o eterogenee.

Di seguito vengono mostrati tre diversi grafici che descrivono proprio questo concetto. Il primo è un grafico di consistenza dei dati che mostra l'effettivo periodo di rilevazione e acquisizione delle portate giornaliere di ogni singola stazione. Le serie storiche sono rappresentate come linee orizzontali, una per ogni stazione. L'asse delle ascisse riporta gli anni, dal 1899 al 2024, mentre l'asse delle ordinate elenca le stazioni considerate. I segmenti neri indicano i periodi per i quali è disponibile una registrazione continua dei dati, mentre gli spazi bianchi corrispondono a lacune, interruzioni o assenza di osservazioni.

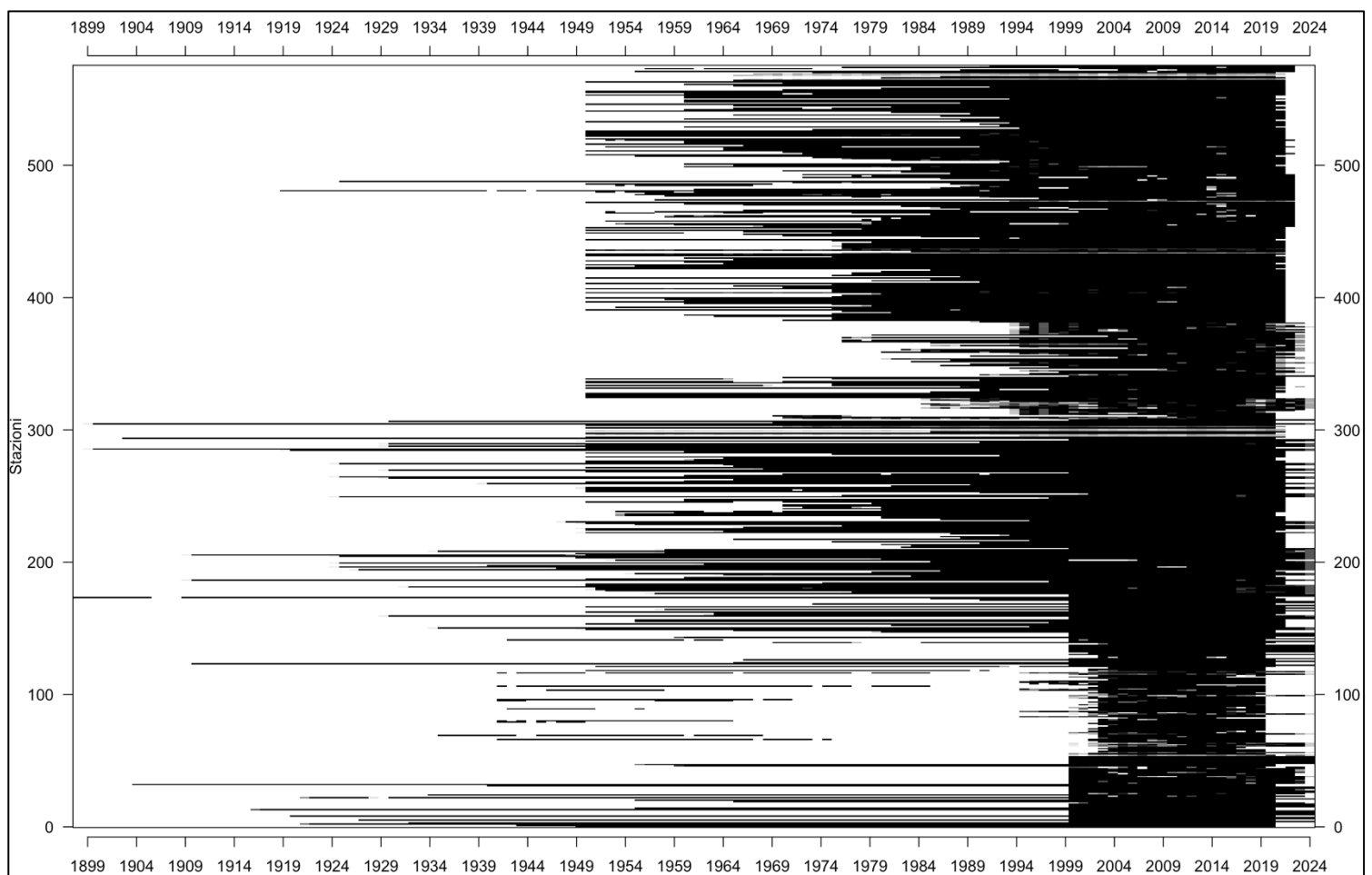


Figura 7- Grafico della consistenza dei dati, sull'asse x gli anni presi in considerazione per l'analisi mentre sull'asse y si riporta il numero delle stazioni.

Si osserva una notevole disomogeneità nella lunghezza e nella completezza delle serie, infatti, soltanto un numero limitato di stazioni presenta registrazioni che si estendono fino ai primi decenni del secolo scorso, mentre la maggior parte delle osservazioni inizia a partire dagli anni '50. Questo fenomeno riflette l'evoluzione storica delle reti di monitoraggio idrologico, che hanno conosciuto un'importante espansione e potenziamento proprio a partire dalla seconda metà del secolo scorso.

Un altro aspetto rilevante è l'aumento progressivo della densità dei dati a partire dagli anni '80: in questo intervallo temporale il numero di stazioni attive cresce sensibilmente, e molte di esse presentano serie quasi continue fino ad oggi.

Queste ultime osservazioni sono ancora più evidenti nella figura sottostante (figura 8). Essa mostra il Boxplot della copertura media delle serie idrologiche suddivise in tre intervalli temporali: 1899–1950, 1951–1979 e 1980–2024.

1. Si osserva come il primo intervallo (1899–1950) sia caratterizzato da valori estremamente bassi e fortemente dispersi: la mediana è prossima allo zero e la presenza di numerosi outlier suggerisce che soltanto poche stazioni abbiano fornito dati consistenti, mentre la maggior parte presenta serie frammentarie o incomplete.
2. Nel periodo 1951–1979 la situazione migliora, con un aumento della copertura media e una distribuzione più ampia, sebbene la mediana rimanga relativamente bassa.
3. Infine, l'intervallo 1980–2024 mostra un deciso salto qualitativo: la copertura media risulta elevata, con valori prossimi a 1 per molte stazioni, e la mediana raggiunge circa 0.9. La distribuzione è più compatta e le stazioni con bassa copertura diventano meno numerose, segnalando un miglioramento sostanziale nella continuità e affidabilità delle osservazioni recenti.

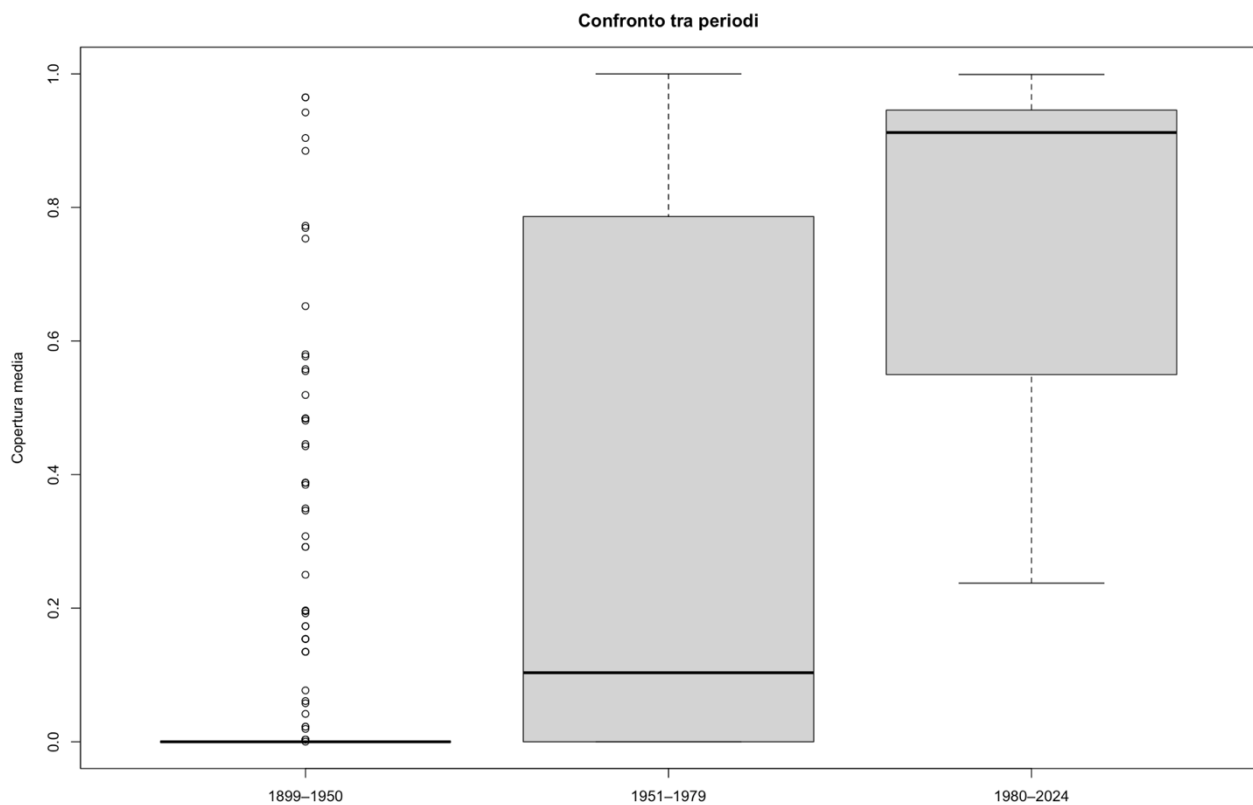


Figura 8- Boxplot del confronto della disponibilità di dati in tre intervalli temporali che mette in evidenza come solo negli ultimi 40/50 anni si ha una buona disponibilità di dati nelle stazioni della GAR.

Infine, con una mappa leaflet è possibile visualizzare la copertura di dati a livello spaziale su tutta la GAR (figura 9). Le stazioni con colori tendenti al rosso sono stazioni con una copertura di dati limitata nel periodo di analisi mentre in verde sono mostrate le stazioni che registrano dati da più tempo e con maggiore continuità.

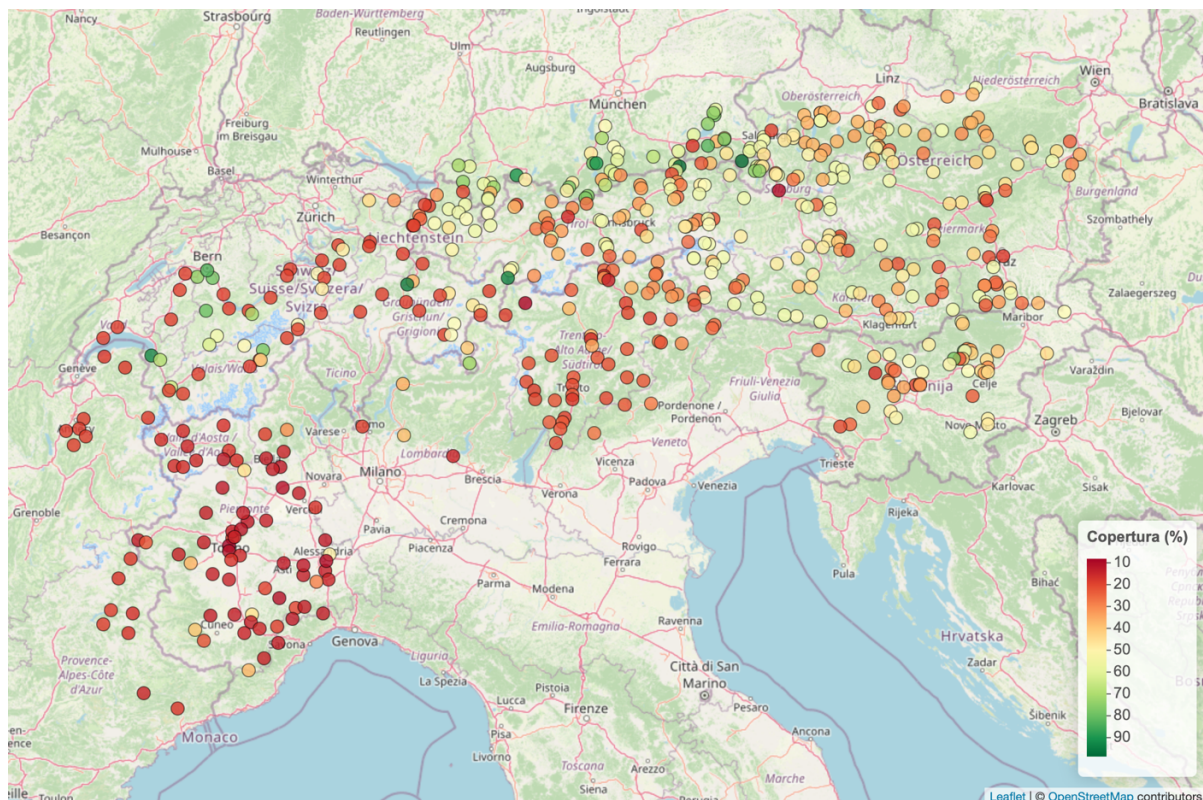


Figura 9- Mappa leaflet che mostra la copertura di dati per ogni stazione (rosso se sono presenti pochi anni di dati rispetto al periodo di analisi, verde se la maggior parte degli anni è coperta).

Di seguito vengono elencati i passaggi della procedura statistica eseguita sui dati delle portate giornaliere dei corsi d'acqua del territorio della GAR:

1. elaborazioni preliminari dei dati;
2. scelta del modello probabilistico;
3. stima dei parametri del modello;
4. verifica del modello con test di adattamento;
5. stima dell'incertezza.

2.1 Elaborazioni preliminari dei dati.

Il primo passaggio della procedura consiste nell'elaborare i dati in modo da evitare di commettere errori grossolani. *"data_GAR_PASS_KGE"* è la large list iniziale che contiene le portate (in metri cubi al secondo) per le 623 stazioni per cui c'erano buoni dati nel periodo 2000-2020. Da questa large list iniziale sono stati infatti rimossi i dati di alcune stazioni, presenti nella lista *"Removal"*, in quanto queste stazioni presentavano problemi di aree a monte. La large list aggiornata è di tipologia *"zoo"*. Un oggetto di questo tipo in R è una struttura dati pensata per gestire serie temporali in modo flessibile, principalmente quindi le liste sono costituite dal valore della portata seguito dalla data di rilevazione del valore.

Al posto della data può anche essere inserito un indice temporale diverso ovvero formato ad esempio da ore, minuti o altri indici. A differenza di altri formati per serie temporali, zoo permette di gestire dati con osservazioni irregolari, cioè non equidistanti nel tempo.

Offre inoltre funzioni per filtrare, aggregare, estrarre sottoperiodi e visualizzare i dati in determinati ordini. Per questo motivo è molto usato in analisi idrologiche e climatologiche.

Come base delle analisi idrologiche rimangono quindi i dati di 575 stazioni che vengono mostrate nella figura successiva in blu (figura 9).

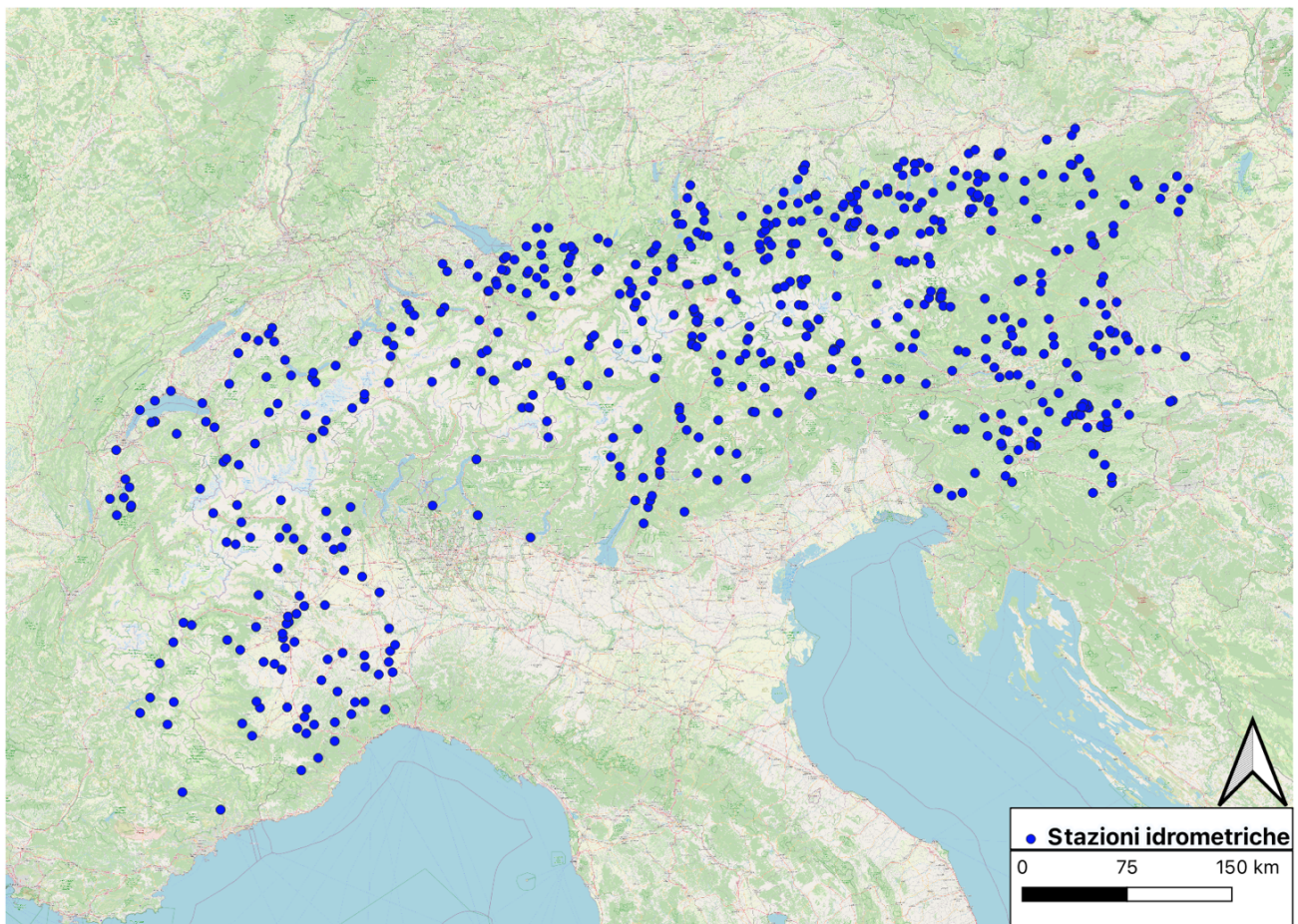


Figura 10- Distribuzione spaziale delle 575 stazioni idrometriche da cui provengono i dati sul territorio della GAR.

La trattazione continua con una seconda operazione di pre-processing utilizzata spesso in idrologia per analisi di estremi, suddivisa in due fasi:

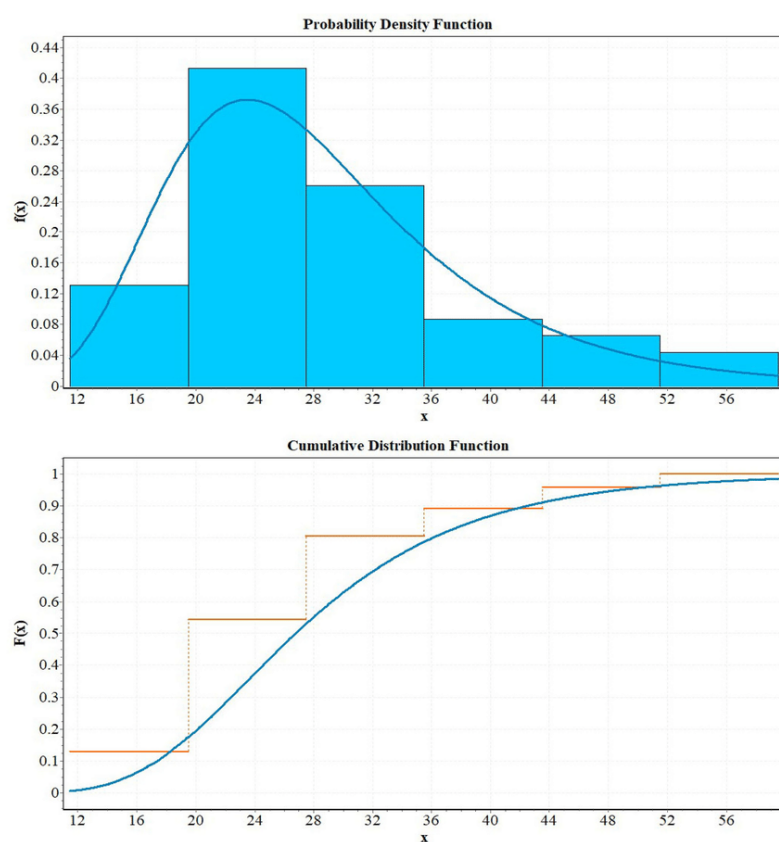
1. identificazione preliminare degli eventi: per rendere ulteriormente robusta l'analisi si è deciso di contare quanti giorni con dati validi ci sono per ogni anno e per ogni bacino in modo da scartare gli anni con meno di 200 giorni di dati validi. Successivamente si sono creati due filtri a media mobile: uno più sensibile su un periodo fisso di 15 giorni e il secondo che controlla i trend su periodi fissi di 100 giorni. L'idea è che quando il filtro a media mobile di 15 giorni supera, come valore, il filtro a media mobile di 100 giorni vuol dire che si è entrati in un evento di piena. Questo permette di riconoscere ed estrapolare dalle serie storiche gli eventi di piena. Quest'ultimi sono salvati in data frame in cui viene segnato la data di inizio, di fine la durata e il valore massimo di portata in metri cubi raggiunto. Quindi per ogni bacino vengono salvati tutti i casi in cui viene riconosciuto un evento;

2. selezione dei picchi significativi: a questo punto si applica il metodo POT (Peak Over Threshold) selezionando un numero limitato di eventi dalla matrice creata precedentemente. Dopo aver ordinato gli eventi in ordine crescente in base al valore di portata “max” raggiunto si selezionano i due più gravosi per ogni anno in cui sono state registrate più di 200 giorni di portate giornaliere valide. Successivamente si riordinano i picchi in ordine cronologico e si salvano gli eventi selezionati nella matrice “POTmax”.

2.2 Scelta dei modelli probabilistici.

Terminata quindi la prima fase di elaborazione dei dati si procede con la scelta del modello probabilistico ovvero del modello matematico che lega un valore della variabile casuale indipendente (x) alla probabilità di accadimento. Si può attuare questa associazione tramite la distribuzione di probabilità cumulata (CDF, $P(x)$) oppure tramite la densità di probabilità (pdf, $p(x)$), legate dalla seguente relazione:

$$P(x) = \int_{-\infty}^{+\infty} p(x)dx \quad (1)$$



I modelli matematici che si vedranno di seguito in dettaglio sono descritti da parametri ovvero numeri calcolati da un campione che ne riassumono le caratteristiche più importanti. I parametri statistici sono quindi le caratteristiche di una popolazione. In particolare, le caratteristiche delle distribuzioni di probabilità sono la moda, il valore della variabile x per cui la pdf è massima (il picco della curva nel grafico superiore), la mediana ovvero il valore della x che divide a metà l'area della pdf (sotto la curva) ed è la x per cui la CDF = 0,5 e infine la media, il parametro di posizione o baricentro della distribuzione.

Figura 11- Rappresentazione grafica delle due funzioni descritte precedentemente. In questo articolo la CDF è $F(x)$ mentre la pdf è $f(x)$. Nabavi, E., Sabour, M., Dezvareh, G.A. et al. Predicting and routing the sub-basin floods and investigating the reservoir dam location in flood risk reduction (case study, Zolachai Dam in West Azerbaijan). Model. Earth Syst. Environ. 8, 3793–3811 (2022). <https://doi.org/10.1007/s40808-021-01330-1>

Per trattare in modo completo come è definita una distribuzione di probabilità bisogna necessariamente citare i momenti centrali di ordine r di quest'ultima. Essi sono i valori che descrivono la forma della distribuzione, misurando come i dati si distribuiscono attorno alla media. L'idea è che non si guarda più ai valori della variabile in sé, ma a quanto si discostano dalla media. Elevando alla potenza r , si pesano questi scostamenti in modo diverso a seconda dell'ordine del momento.

La formula generale che permette di calcolarli è la seguente:

$$M_r = \int (x - \mu)^r \cdot p(x) dx \quad (2)$$

Il momento di ordine 1 è sempre nullo perché la media degli scostamenti dalla media stessa è nulla. Il momento di ordine 2 è la varianza ovvero la distanza quadratica dalla media, utilizzata per misurare la dispersione della distribuzione. È il parametro di scala della distribuzione. Per confrontare distribuzioni di fenomeni molto diversi tra di loro viene spesso impiegato il coefficiente di variazione adimensionale di cui di seguito viene riportata la formula:

$$CV = \frac{\sigma}{\mu} \quad (3)$$

Il momento di ordine 3 (Skewness) descrive invece la simmetria della distribuzione rispetto alla media. Il rapporto tra questo valore e la varianza al cubo viene definito γ ovvero coefficiente di asimmetria: nel caso questo risultasse pari a zero si avrà una distribuzione simmetrica. Nel caso invece risultasse maggiore o minore di zero allora la distribuzione di probabilità sarà rispettivamente ad asimmetria positiva e ad asimmetria negativa.

Il momento di ordine 4 misura invece l'appiattimento o spessore delle code della distribuzione. Il rapporto tra questo valore e la varianza elevata alla quarta permette di calcolare il coefficiente di Kurtosis. Esso aiuta a capire se i dati hanno eventi rari estremi più frequenti di quanto previsto dalla distribuzione normale o gaussiana. Nel caso risultasse un valore di questo coefficiente maggiore di 3 (della gaussiana standard) ci si aspetta maggiori probabilità di valori estremi ovvero valori lontani dalla media. Nel caso invece risultasse un valore minore di 3 allora si avrà una distribuzione più piatta rispetto alla normale, con code leggere e valori estremi meno probabili.

Prima di definire quali sono stati i modelli matematici impiegati in questo progetto di tesi si definiscono altre due caratteristiche dei modelli matematici ovvero la loro flessibilità e la loro

robustezza. La flessibilità di un modello matematico è la sua capacità di adattarsi ai dati ed è descritta dal numero di parametri che lo compongono. I parametri delle distribuzioni coincidono con i loro gradi di libertà. Se si immagina di impiegare un modello con numero di parametri pari alla grandezza del campione allora quest'ultimo sarà riprodotto perfettamente dalla distribuzione.

La robustezza è invece la capacità di ottenere risultati simili quando si utilizzano campioni diversi estratti dalla stessa distribuzione. Per far sì che un modello matematico sia robusto e flessibile allo stesso tempo bisogna trovare il giusto equilibrio tra numero di parametri che descrivono il modello e la tipologia di campione da modellare.

Per l'analisi idrologica sugli estremi di portata fluviale nella GAR, ci si è focalizzati sulle due distribuzioni più impiegate per questa tipologia di analisi: Generalized Extreme Value (GEV) e Generalized Pareto (GP). A differenza di distribuzioni come l'esponenziale che descrive bene l'intensità delle precipitazioni e la Gumbel che descrive invece i massimi di precipitazione, GEV e GP descrivono bene le code estreme dei dati, focus dello studio.

La GEV è la generalizzazione della Gumbel a 3 parametri. Il massimo di N variabili casuali con N molto grande tende alla distribuzione GEV. Avendo un parametro in più della distribuzione Gumbel risulta più flessibile di quest'ultima. La sua funzione di distribuzione cumulativa (CDF) e la corrispondente pdf sono riportate di seguito (equazioni 4 e 5):

$$F(x) = \exp \left[- \left(1 - \frac{\sigma_2}{\sigma_3} (x - \sigma_1) \right)^{\frac{1}{\sigma_3}} \right] \quad (4)$$

$$f(x) = \frac{1}{\sigma_2} e^{-(1-\sigma_3)y} e^{-y} \quad (5)$$

$$\text{Con: } y = \begin{cases} -\theta_3^{-1} \log\{1 - \theta_3(x - \theta_1)/\theta_2\} , & \theta_3 \neq 0 \\ (x - \theta_1)/\theta_2 , & \theta_3 = 0 \end{cases} \quad (6)$$

Dove σ_1 , σ_2 e σ_3 sono i tre parametri che caratterizzano la GEV rispettivamente di posizione, di scala e di forma.

```

### GEV
dGEV <- function(x, xi=1, sigma=1, kappa=-.5){
  y <- -kappa^(-1) * log(1-kappa * (x-xi)/sigma)
  sigma^(-1) * exp(-(1-kappa) * y - exp(-y))
}

pGEV <- function(q, xi=1, sigma=1, kappa=-.5){
  y <- -kappa^(-1) * log(1-kappa * (q-xi)/sigma)
  exp(-exp(-y))
}

qGEV <- function(p, xi=1, sigma=1, kappa=-.5){
  xi + sigma * (1-(-log(p))^kappa)/kappa
}

rGEV<- function(n, xi=1, sigma=1, kappa=-.5){
  p<- runif(n)
  qGEV(p, xi, sigma, kappa)
}

```

Figura 12- Schermata di R studio in cui vengono mostrate le funzioni utilizzate per l'analisi statistica della distribuzione GEV.

Le funzioni rispettivamente permettono di stimare:

- dGEV = densità di probabilità;
- pGEV = probabilità di non superamento;
- qGEV = valori dei quantili;
- rGEV= generare campioni di dati (per la successiva analisi dell'incertezza).

Per la GEV non è possibile costruire una carta probabilistica perché non esiste attualmente una variabile ridotta che elimini i parametri e che sia lineare in x. Viene, in seguito, utilizzata quindi la carta probabilistica di Gumbel. In figura 13 viene mostrata la funzione in R che permette di crearla:

```

pplot <- function(x, distr='Normal', ...) {
  x_i <- sort(x)
  n <- length(x_i)
  i<- 1:n
  F_x_i<- i/(n+1) #weibull plotting position
  if (distr == 'Normal') u_i <- qnorm(F_x_i)
  else if (distr == 'Exp') u_i <- -log(1- F_x_i)
  else if (distr == 'Gumbel') u_i <- -log(-log(F_x_i))
  else u_i <- NA * x_i
  plot(u_i, x_i, ...)
}

```

Figura 13- Funzione in R che permette di plottare la carta probabilistica di Gumbel.

Per quanto riguarda invece la Generalized Pareto, come la GEV, è una distribuzione a tre parametri. Le sue applicazioni includono l'analisi di eventi estremi, la modellazione di grandi sinistri assicurativi, negli studi di affidabilità e, in generale, in qualsiasi situazione in cui si potrebbe impiegare la distribuzione esponenziale ma sia necessaria una maggiore robustezza rispetto ad alternative con code più pesanti o più leggere. (Hosking, J. R. M., & Wallis, J. R. (1987), *Parameter and quantile estimation for the generalized Pareto distribution. Technometrics*, 29 (3), 339–349.).

Le formule che la descrivono vengono riportate di seguito:

$$F(x) = 1 - [1 - \theta_3 \frac{(x-\theta_1)}{\theta_2}]^{1/\theta_3}$$

$$f(x) = \frac{1}{\theta_2} [1 - \theta_3 \frac{(x-\theta_1)}{\theta_2}]^{1/\theta_3 - 1}$$

$$x(F) = \theta_1 + \frac{\theta_2}{\theta_3} [1 - (1 - F)^{\theta_3}]$$

Figura 14- Formule da:

http://www.idrologia.polito.it/didattica/Idrologia/2013/Modulo_1_Statistica/DISPENSA_2_distribuzioni_corr_1.pdf

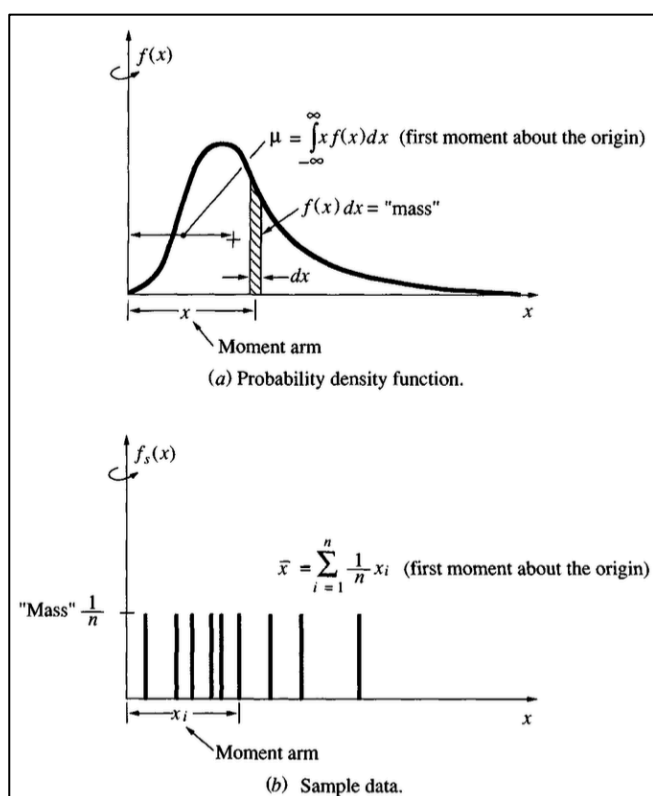
```
#### Generalized Pareto
dGP <- function (x, x0=0, alpha=1, kappa=-.5) { #densità di probabilità
  y <- 1 - kappa*(x - x0)/alpha
  1/alpha * y^(1/kappa - 1)
}
pGP <- function (q, x0=0, alpha=1, kappa=-.5) { #probabilità di non superamento
  y <- 1 - kappa*(q - x0)/alpha
  p <- 1 - y^(1/kappa)
  return(p)
}
qGP <- function (p, x0=0, alpha=1, kappa=-.5) { #quantili
  q <- x0 + alpha*(1 - (1 - p)^kappa)/kappa
  return(q)
}
rGP <- function (n, x0=0, alpha=1, kappa=-.5) { #genera campione di dati
  p <- runif(n, min=1e-10, max=1 - 1e-10)
  x <- qGP(p, x0, alpha, kappa)
  return(x)
}
```

Figura 15- Schermata di R studio in cui vengono mostrate le funzioni utilizzate per l'analisi statistica della distruzione GP.

2.3 Stima dei parametri dei modelli.

Al fine di stimare i parametri delle due distribuzioni introdotte nel paragrafo precedente, e quindi individuare, tra le possibili curve teoriche, quelle che meglio rappresentano l'andamento del campione di portate giornaliere osservate nei molteplici siti della GAR, è possibile adottare differenti metodologie di stima. La letteratura statistica e idrologica propone diversi approcci, ciascuno caratterizzato da specifici punti di forza e limiti applicativi. Tra i più diffusi si annoverano il metodo della massima verosimiglianza, il metodo dei momenti e il metodo dei momenti ponderati lineari (L-moments), particolarmente apprezzati per la loro robustezza in presenza di serie di dati limitate o caratterizzate da valori estremi. L'impiego di tali metodologie consente di adattare in maniera appropriata il modello teorico ai dati osservati, garantendo una stima più affidabile dei parametri e, di conseguenza, una rappresentazione più realistica del fenomeno idrologico in esame.

Il metodo dei momenti, introdotto da Karl Pearson nel 1902, si fonda sull'idea che una buona stima dei parametri di una distribuzione di probabilità debba garantire coerenza tra il modello teorico e i dati osservati. In particolare, secondo questo approccio, i momenti della funzione di densità teorica, calcolati rispetto all'origine, vengono eguagliati ai momenti corrispondenti calcolati sul campione disponibile. In questo modo, i parametri della distribuzione vengono determinati imponendo che le principali caratteristiche statistiche del modello (ad esempio



media, varianza, asimmetria e Kurtosis) coincidano con quelle effettivamente osservate nei dati. In figura 16 viene mostrato un esempio grafico di questo concetto. La media teorica della distribuzione, il momento di primo ordine, viene calcolata tramite l'integrale mostrato nel grafico (a). L'integrale somma i "momenti elementari" ottenuti aumentando il "braccio del momento" per poi ottenere così la media complessiva. Nel grafico (b) invece si ha un campione di dati e la media viene calcolata

Figura 16- Esempio esplicativo del metodo dei momenti. Da: "Applied hydrology" di Ven Te Chow, David R. Maidment, Larry W. Mays (ISBN 0-07-010810-2).

come rapporto tra la sommatoria dei valori della variabile e il numero di osservazioni. La figura mette quindi a confronto il momento teorico (ottenuto integrando la PDF) con il momento campionario (ottenuto mediando i dati osservati). Il metodo dei momenti si basa proprio su questo principio: si eguagliano i momenti teorici ai corrispondenti momenti campionari per poi ricavare, da questa uguaglianza, le stime dei parametri della distribuzione.

Un problema di questo metodo riguarda il fatto che l'efficienza o accuratezza diminuisce all'aumentare dell'ordine r : lo stimatore, ovvero ciò che descrive le proprietà di variabili casuali, della media (ordine uno) infatti è molto più efficiente di quello della varianza (ordine due) e così via. Questa caratteristica rende il metodo dei momenti poco efficiente per campioni asimmetrici e con presenza di outliers soprattutto nelle code. Infatti, i momenti campionari contengono potenze della variabile casuale per cui valori molto grandi della variabile rispetto alla media hanno un peso molto importante nella stima (effetti sulla stima della varianza e skewness).

Proseguendo nella trattazione verranno presentati esempi di questo caso studio che confermano questa caratteristica del metodo.

Per sopperire a queste problematiche i ricercatori J. R. M. Hosking e J. R. Wallis presentano nel 1990 la teoria degli L-momenti. Essa corre parallela a quella dei momenti convenzionali ma il principale vantaggio degli L-momenti rispetto ai momenti convenzionali consiste nel fatto che, essendo funzioni lineari dei dati, essi risentono in misura minore della variabilità campionaria: gli L-momenti risultano infatti più robusti dei momenti convenzionali rispetto alla presenza di valori anomali nei dati e consentono inferenze più affidabili a partire da campioni di dimensioni ridotte sulla distribuzione di probabilità sottostante. (Hosking, J. R. M. (1990). L-Moments: Analysis and Estimation of Distributions Using Linear Combinations of Order Statistics. *Journal of the Royal Statistical Society: Series B (Methodological)*, 52(1), 105–124.)

Gli L-momenti λ_1 e λ_2 , l'L-CV, τ e i rapporti degli L-momenti τ_3 e τ_4 sono le quantità che Hosking and Wallis (Hosking JRM, Wallis JR. *Regional Frequency Analysis: An Approach Based on L-Moments*. Cambridge University Press; 1997.) consigliano di utilizzare per descrivere le distribuzioni di probabilità. Di seguito viene riportata una tabella in cui vengono riassunte e messe a confronto le grandezze protagoniste dei due metodi (figura 17).

	Population moment	Sample moment	Population L-moment	Sample L-moment
Location (mean)	μ	$\bar{x}$	λ_1	ℓ_1
Scale	σ	s	λ_2	ℓ_2
CV	C_v	$\hat{C}_v$	τ	t
Skewness	γ	g	τ_3	t_3
Kurtosis	κ	k	τ_4	t_4

Figura 17- Notazione formale delle grandezze che descrivono i metodi dei momenti e L-momenti. Da: (Hosking JRM, Wallis JR. *Regional Frequency Analysis: An Approach Based on L-Moments*. Cambridge University Press; 1997.)

2.6 Properties of L-moments

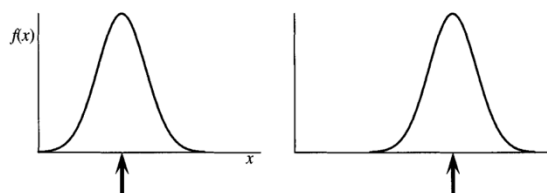


Fig. 2.1. Definition sketch for first L-moment.

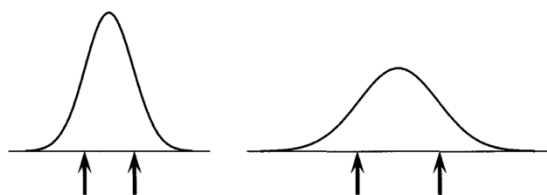


Fig. 2.2. Definition sketch for second L-moment.

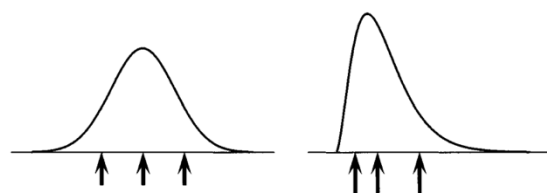


Fig. 2.3. Definition sketch for third L-moment.

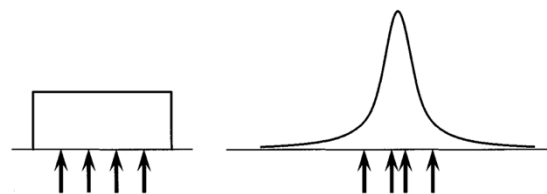


Fig. 2.4. Definition sketch for fourth L-moment.

Figura 18- Rappresentazione grafica delle proprietà degli L-momenti.

Di seguito vengono elencate alcune proprietà di questo metodo:

- **esistenza:** gli L-momenti risultano sempre definiti ogniquale volta la distribuzione considerata ammetta una media finita;
- **unicità:** nel caso in cui la media sia finita, gli L-momenti caratterizzano in modo univoco la distribuzione; non possono dunque esistere due distribuzioni distinte che condividano gli stessi L-momenti.
- **significato statistico:** ciascun L-momento (o rapporto di L-momenti) ha un'interpretazione diretta, analoga a quella dei momenti convenzionali.
- **vincoli algebrici:** alcuni parametri soddisfano condizioni di ammissibilità.
- **invarianza rispetto a trasformazioni lineari;**
- **simmetria:** se la distribuzione di una variabile casuale X è simmetrica rispetto alla media, allora tutti i rapporti degli L-momenti di ordine dispari si annullano.

(Fonti delle proprietà degli L-momenti: Hosking JRM, Wallis JR. *Regional Frequency Analysis: An Approach Based on L-Moments*. Cambridge University Press; 1997. & <http://www.idrologia.polito.it/didattica/Idrologia/2016/blocco1/Dispensa L-Momenti.pdf>).

Il metodo degli L-momenti risulta quindi essere meno sensibile agli outlier e ne consegue che gli stimatori di ordine maggiore di uno risultino essere più efficienti dei momenti convenzionali.

Un metodo particolarmente efficace per confrontare gli L-momenti associati a differenti distribuzioni è l'utilizzo del diagramma dei rapporti degli L-momenti, riportato in Figura 19. In tale rappresentazione, gli L-momenti vengono collocati su un piano cartesiano in cui l'asse delle ascisse corrisponde all'L-skewness e l'asse delle ordinate all'L-kurtosis.

Una distribuzione a due parametri, definita da un parametro di posizione e da uno di scala, si traduce graficamente in un singolo punto all'interno del diagramma.

Una distribuzione a tre parametri, essendo definita da posizione, scala e forma, non è più rappresentata da un singolo punto, bensì da una curva sul diagramma: ciascun punto della linea corrisponde a un diverso valore assunto dal parametro di forma.

(http://www.idrologia.polito.it/didattica/Idrologia/2016/blocco1/Dispensa_L-Momenti.pdf).

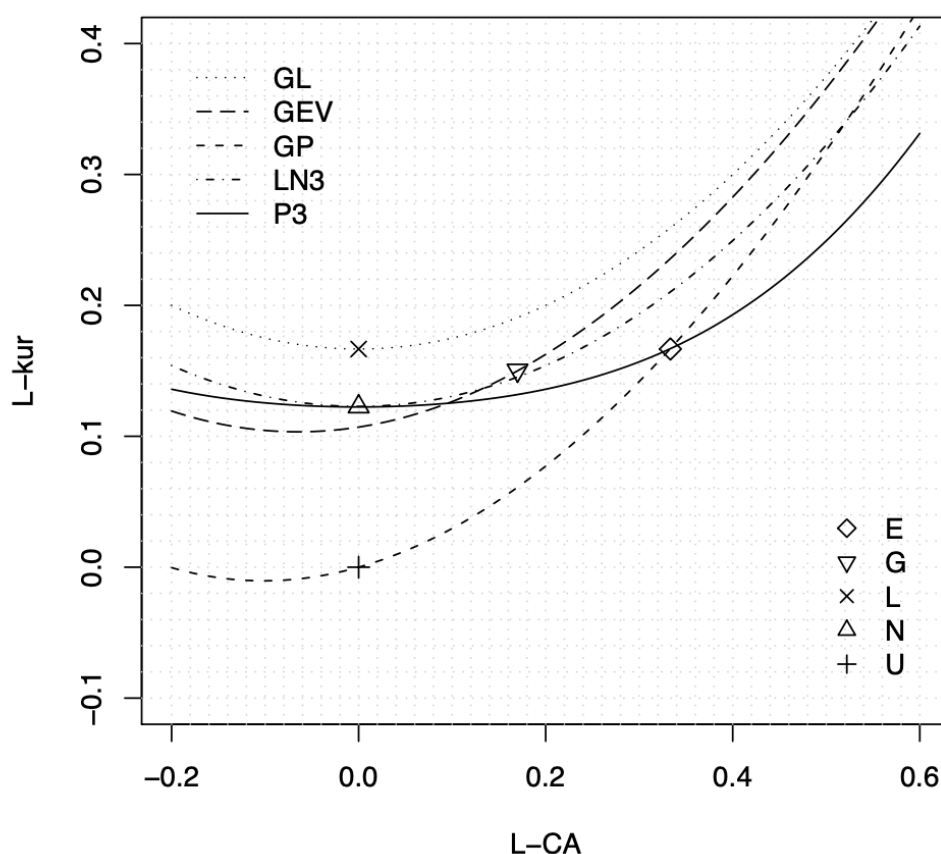


Figura 19-: Diagramma dei rapporti degli L-momenti. Le distribuzioni a due ed a tre parametri sono riportate come punti e come linee rispettivamente. Le distribuzioni a due parametri sono: esponenziale (E), Gumbel (G), log normale (L), normale (N) ed uniforme (U); quelle a tre parametri sono: logistica generalizzata (GL), generalizzata del valore estremo (GEV), Pareto generalizzata (GP), log normale a 3 parametri (LN3) e Pearson tipo III (P3).

Da: http://www.idrologia.polito.it/didattica/Idrologia/2016/blocco1/Dispensa_L-Momenti.pdf

GEV:	$\lambda_1 = \xi + \frac{\alpha}{\kappa} (1 - \Gamma[1 + \kappa])$	$\lambda_2 = \frac{\alpha}{\kappa} (1 - 2^{-\kappa}) \Gamma(1 + \kappa)$
$x = \xi + \frac{\alpha}{\kappa} (1 - [-\ln F]^\kappa)$	$\tau_3 = \left\{ \frac{2(1 - 3^{-\kappa})}{(1 - 2^{-\kappa})} - 3 \right\}$	
	$\tau_4 = \frac{1 - 5(4^{-\kappa}) + 10(3^{-\kappa}) - 6(2^{-\kappa})}{1 - 2^{-\kappa}}$	
Generalized Pareto:	$\lambda_1 = \xi + \frac{\alpha}{1 + \kappa}$	$\lambda_2 = \frac{\alpha}{(1 + \kappa)(2 + \kappa)}$
$x = \xi + \frac{\alpha}{\kappa} \{1 - [1 - F]^\kappa\}$	$\tau_3 = \frac{1 - \kappa}{3 + \kappa}$	$\tau_4 = \frac{(1 - \kappa)(2 - \kappa)}{(3 + \kappa)(4 + \kappa)}$

Figura 20-Valori degli L-momenti e relazioni per l'inverso della funzione di distribuzione cumulata per la GEV e la GP. Da: Stedinger, J. R., Vogel, R. M., & Foufoula-Georgiou, E. (1993). *Frequency analysis of extreme events*. In Maidment, D. R. (Ed.), *Handbook of Hydrology*. McGraw-Hill.

Il metodo della massima verosimiglianza consiste nel determinare i valori dei parametri in modo da massimizzare la funzione di verosimiglianza, che viene riportata di seguito (equazione 7):

$$\mathcal{L}(\theta) = \prod_1^n p(x_i; \theta) \quad (7)$$

La funzione di verosimiglianza, $\mathcal{L}(\theta)$, è costruita a partire da una famiglia parametrica di densità di probabilità, $p(x_i; \theta)$, e da un campione osservato $x_1, x_2, \dots, x_n$. La verosimiglianza è definita come la probabilità congiunta di osservare il campione al variare del parametro θ . Lo scopo della stima è trovare il valore di θ che massimizza questa funzione, ossia rende i dati osservati più probabili sotto il modello scelto. (<https://online.stat.psu.edu/stat415/lesson/1/1.2>).

Si impone quindi la seguente relazione che permette di trovare la massima verosimiglianza:

$$\frac{d\mathcal{L}(\theta)}{d\theta} = 0 \quad (8)$$

Solitamente questo non viene impiegato in idrologia perché, per campioni di dati relativamente brevi, lo stimatore della massima verosimiglianza risulta poco efficiente. In questo studio, invece, dato che si hanno serie storiche importanti di portate giornaliere si è deciso di includerlo nella trattazione. L'approccio è di massimizzare la verosimiglianza rispetto ai dati annuali dei massimi di portate selezionati con il metodo Peak Over Threshold per stimare i parametri di distribuzioni come la GEV o la GP.

In R, grazie alla libreria “*rstan*” è possibile creare una funzione (fare riferimento alla sezione allegati, righe del codice da 203 a 304) che permette di associare una distribuzione Generalized Pareto ai dati con il metodo della massima verosimiglianza. Di seguito viene spiegato il workflow della funzione:

- inizialmente nel blocco functions vengono definiti la densità logaritmica della GP e tre priors ovvero distribuzioni di probabilità che si assegnano a un parametro prima di osservare i dati:
 - $\kappa \sim \text{Normal}(-0.1, 0.122)$. Nello specifico si assegna a κ un prior normale con media -0.1 e deviazione standard 0.122. Significa che prima dei dati si assume che κ sia attorno a -0.1 con variabilità moderata. Questo serve a stabilizzare la stima del parametro forma, κ , nelle serie corte/rumorose (Questi valori sono tratti da Martins, E. S., and J. R. Stedinger (2001), Generalized Maximum Likelihood Pareto-Poisson estimators for partial duration series, *Water Resour. Res.*, 37(10), 2551–2557, doi:[10.1029/2001WR000367](https://doi.org/10.1029/2001WR000367).)
 - x_0 uniforme implicita tra 0 e $\min(y)$ fa sì che la soglia/localizzazione non superi il più piccolo picco e resti maggiore di zero;
 - $\log_alpha$ al posto di alpha per avere la certezza che sia sempre maggiore di zero;
- successivamente vengono forniti i dati ovvero il numero di picchi e i valori osservati;
- nel terzo blocco della funzione vengono quindi calcolati i parametri della GP;
- infine, questi valori vengono salvati in una matrice con numero di righe pari ai siti di rilevazioni dati e con le tre colonne contenenti i tre parametri delle distribuzioni GP per ogni sito.

Si è quindi scelto di calcolare i parametri delle distribuzioni GP anche con il metodo degli L-momenti. In R questo è possibile grazie alle librerie “*lmom*” e “*lmomco*”. La funzione fondamentale per questo passaggio è “*pargpa*” che restituisce i parametri delle GP. Di seguito vengono riportati i risultati per alcune stazioni dei valori dei parametri ottenuti con i due metodi proposti. Le stime dei parametri della distribuzione Generalized Pareto ottenute tramite il metodo della massima verosimiglianza (GML) e quelle derivate dal metodo degli L-momenti risultano sostanzialmente coerenti, mostrando valori tra loro comparabili (come mostrato in figura 21).

	xi	alpha	kappa		x0	alpha	kappa
141	3.9022831	2.1712345	-0.023025728	141	3.871	2.1267828	-0.0582348939
155	7.6867828	3.0087731	0.145483905	155	7.290	3.3236633	0.1242171067
13	11.1047170	4.9316578	-0.013370493	13	11.011	5.0321166	-0.0125496723
95	25.4546420	10.2777169	0.206509961	95	25.300	9.1360804	0.0615524404
84	60.2332267	22.1908588	-0.105157782	84	61.601	20.6414610	-0.1155734945
10	46.3628971	17.2663307	-0.061032779	10	46.689	16.8413481	-0.0698153788
41	91.9952099	24.5485253	-0.246780842	41	91.213	28.6832907	-0.1299409786
80	64.1324406	23.3779933	-0.144230658	80	63.191	25.9063684	-0.0869032923
77	35.2068052	14.0173307	-0.055093049	77	34.800	14.5564044	-0.0465876837
101	83.4378493	38.6790984	-0.081527566	101	82.818	40.2988708	-0.0642298626
21	81.9505840	34.6310797	-0.445841762	21	86.742	33.0176225	-0.3014362819
8	127.0080332	49.9317131	-0.407581809	8	131.071	53.1988968	-0.2489149345
31	257.3567342	45.7062738	-0.194885297	31	256.042	51.9912844	-0.1048610924
7	271.0542995	58.1649800	0.151297907	7	268.508	55.0105521	0.0406925886

Figura 21- Tabelle dei valori dei parametri delle GP di alcuni siti (righe): a sinistra la stima con gli L-momenti e a destra con GML.

Per quanto riguarda la GEV in R, l'analisi è stata svolta con il metodo dei momenti e degli L-momenti. Inizialmente si costruiscono due funzioni “*fparGEV*” e “*skew*” i cui scopi sono rispettivamente stimare i parametri della GEV con il metodo dei momenti e fornire la terza misura necessaria a *fparGEV*, ovvero il valore della skew la cui formula è riportata di seguito:

$$skew = \frac{\sum(x_i - \bar{x})^3}{s^3} \cdot \frac{n}{(n-1) \cdot (n-2)} \quad (9)$$

Successivamente si inizializzano due matrici con numero di righe pari al numero di stazioni e tre colonne con il fine di contenere i parametri delle distribuzioni GEV per ogni sito. Infine, con due cicli for che creano due loop sulle righe delle due matrici si calcolano i parametri: in un caso richiamando la funzione *fparGEV* nel ciclo for, nell'altro caso, ovvero per il metodo degli L-momenti, impiegando la funzione *pelgev*. In figura 22 vengono riportati i risultati delle stime dei parametri per alcuni siti e anche in questo caso esse sono simili e confrontabili.

	xi_GEV_mom	alpha_GEV_mom	kappa_GEV_mom		xi_GEV	alpha_GEV	kappa_GEV
141	5.0365788	1.3994235	-0.1701050829	141	5.0299769	1.2058428	-0.25358123
155	9.2186347	1.6524295	-0.0798593469	155	9.1821186	1.5022502	-0.15236070
13	13.8002691	3.3681059	-0.0978879214	13	13.6597302	2.7222733	-0.24761479
95	30.5113624	4.9408242	-0.1117828393	95	30.4807535	4.9375051	-0.11715215
84	72.7621123	16.8966641	-0.1319834579	84	72.0039361	12.9782157	-0.30520334
10	56.0768391	12.9199408	-0.0874936872	10	55.4188557	9.8217104	-0.27727420
41	106.1371482	24.0925381	-0.1616647097	41	105.4848445	15.6847597	-0.39808195
80	77.4914443	19.3487625	-0.1283140646	80	76.6560398	14.0117996	-0.33032473
77	42.8054091	9.9316473	-0.1338526241	77	42.5475056	7.9437890	-0.27354954
101	106.1623092	31.1474202	-0.0438879255	101	103.8312455	22.2877227	-0.29018925
21	102.1443759	57.0300325	-0.1438642172	21	101.8888557	24.9808106	-0.53783372
8	155.5074447	74.3019496	-0.1506680045	8	155.5089568	35.2014075	-0.51007068
31	284.5795331	42.6779539	-0.1049879343	31	282.1537586	28.2757323	-0.36345588
7	299.6776904	29.4455321	-0.1453479327	7	299.9174770	28.9346885	-0.14897445

Figura 22- Tabelle dei valori dei parametri delle GEV di alcuni siti: a sinistra la stima con i momenti e a destra con L-momenti.

2.4 Verifica dei modelli con test di adattamento.

L'analisi statistica dei dati idrologici e ambientali richiede sempre la formulazione di un insieme di ipotesi preliminari, ad esempio riguardanti la forma della distribuzione di probabilità che meglio rappresenta la variabile in esame e la scelta del metodo di stima dei parametri. (Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer). In particolare, un test statistico è il procedimento che consente, sulla base delle osservazioni disponibili, di decidere se accettare o rifiutare H_0 , ovvero una qualsiasi ipotesi statistica formulata come ipotesi nulla. In questo contesto si definisce livello di significatività del test, indicato con α , la probabilità di commettere un errore di tipo I, ovvero rifiutare H_0 quando essa è in realtà vera. Viceversa, la potenza del test è la probabilità di rifiutare H_0 quando essa è falsa, cioè quando vale H_1 . La potenza è pari a $1-\beta$, dove β è la probabilità di commettere un errore di tipo II, ovvero accettare H_0 quando l'ipotesi alternativa è corretta (Casella & Berger, 2002). È importante notare che, a parità di numerosità campionaria n , la riduzione del livello di significatività α comporta inevitabilmente una diminuzione della potenza del test.

Esiste una vasta gamma di test, ciascuno concepito per valutare specifiche ipotesi statistiche. Nel caso dell'analisi idrologica, particolare interesse rivestono i test di adattamento (goodness-of fit tests), volti a verificare l'ipotesi nulla H_0 che un insieme di osservazioni sia stato estratto da una distribuzione di probabilità prefissata $F(x)$ (Stephens, 1974; Hosking & Wallis, 1997). L'obiettivo è quindi valutare il grado di coerenza tra il campione empirico e il modello

probabilistico teorico adottato, ad esempio una distribuzione GEV o GP per lo studio dei valori estremi.

Un concetto centrale nell'interpretazione dei test è il valore-p (p-value), definito come la probabilità di ottenere, sotto l'ipotesi nulla, un valore della statistica test uguale o più estremo rispetto a quello osservato.

Il valore-p rappresenta quindi un livello di significatività "osservato" che può essere direttamente confrontato con α . In particolare:

- se $p > \alpha$, l'evidenza empirica non è sufficiente a rifiutare H_0 , e la distribuzione ipotizzata viene considerata compatibile con i dati;
- se $p \leq \alpha$, si rifiuta H_0 , concludendo che i dati non sono coerenti con la distribuzione proposta (Wilks, D. S. (2011). *Statistical Methods in the Atmospheric Sciences* (3rd ed.). Academic Press).

Questo sistema permette solo di rigettare il modello ma non permette di confermare l'ipotesi nulla. In questa trattazione l'ipotesi nulla consiste nel fatto che il campione sia stato estratto da una popolazione predefinita (GEV, GP, Gumbel, ecc.).

La procedura generale di un test statistico può essere riassunta in quattro passaggi:

1. definizione della variabile test;
2. determinazione della distribuzione della variabile test quando l'ipotesi nulla è vera;
3. determinazione dei valori della variabile test per cui l'ipotesi nulla deve essere rigettata;
4. decisione sull'ipotesi nulla.

Per determinare la distribuzione della variabile test quando l'ipotesi nulla è vera è stato applicato il Metodo Monte Carlo per cui conoscendo l'andamento ($P(x)$) e la dimensione n del campione è possibile:

- estrarre dalla distribuzione un campione di dimensione pari al campione originario (n) in modo casuale e calcolare quindi la grandezza da confrontare con l'ipotesi nulla;
- successivamente si ripete questa operazione (estraendo ogni volta un campione diverso di dimensione n) un numero molto elevato di volte e si costruisce l'istogramma dei valori della grandezza calcolata. Più simulazioni si eseguono più fitto sarà l'istogramma che tenderà sempre di più alla pdf.

Grazie a questo metodo si è ottenuta la distribuzione della variabile test. Il passaggio successivo è generare un intervallo di confidenza definendo il livello di significatività α . Per questo caso

studio si è scelto di avere tre livelli di significatività che sono i valori tipici usati per le analisi idrologiche:

- $\alpha = 0.01$ che corrisponde al 1% di significatività;
- $\alpha = 0.05$ che corrisponde al 5% di significatività;
- $\alpha = 0.10$ che corrisponde al 10% di significatività.

A questo punto resta solo da valutare se il valore della grandezza calcolata si trovi compreso nell'intervallo di significatività, cioè dove l'ipotesi H_0 non è rigettata, oppure no.

Due dei test statistici più utilizzati sono il test di Pearson (o del χ^2) e il test di Anderson- Darling. Il test di adattamento di Pearson richiede che il dominio della variabile casuale X sia suddiviso in k intervalli mutuamente esclusivi. La costruzione del test si basa quindi sul confronto tra le frequenze osservate e quelle teoriche attese all'interno di ciascun intervallo.

L'applicazione di questo test presenta tuttavia tre limitazioni fondamentali:

- La prima è legata alla soggettività nella definizione degli intervalli di classe: scelte differenti possono condurre a risultati diversi in termini di potenza e significatività del test. Studi metodologici hanno mostrato che la potenza del test risulta massimizzata quando le classi vengono costruite in modo da avere probabilità teoriche equiprobabili. (Anderson, T. W., & Darling, D. A. (1954). A test of goodness of fit. *Journal of the American Statistical Association*, 49(268), 765–769; Anderson, T. W., & Darling, D. A. (1954). A test of goodness of fit. *Journal of the American Statistical Association*, 49(268), 765–769).
- Il secondo, e più rilevante, problema riguarda la stima dei parametri della distribuzione teorica $F(X)$. In contesti applicativi, come l'idrologia statistica, i parametri vengono quasi sempre stimati a partire dallo stesso campione di osservazioni oggetto del test. In questo caso, la distribuzione asintotica della statistica test di Pearson non coincide più esattamente con quella di una variabile χ^2 con $k-1$ gradi di libertà. Se i parametri sono stimati mediante massima verosimiglianza, la distribuzione della statistica test si colloca tra quella di una χ^2 con $k-1$ gradi di libertà e quella di una χ^2 con $k-s-1$ gradi di libertà, dove s è il numero di parametri stimati (Rao, C. R., & Chakravarti, I. M. (1956). Some small sample tests of significance for a multinomial distribution. *Biometrika*, 43(4), 608–611; Kendall, M., & Stuart, A. (1979). *The Advanced Theory of Statistics*, Vol. 2. Macmillan.). Ne deriva un'incertezza non eliminabile nella definizione dei limiti di accettazione del test.

- Quando invece si hanno a disposizione molteplici dati il test va bene ma solo se si è interessati al corpo della distribuzione; infatti, esso non si concentra sulle code: in idrologia la coda destra delle distribuzioni descrive l'andamento degli estremi quindi, focus della trattazione. In più esso rigetta più facilmente modelli matematici con più parametri rispetto ai modelli con meno parametri. Caratteristica del test risulta essere infatti la sua poca potenza (alti errori di tipologia II). (D'Agostino, R. B., & Stephens, M. A. (1986). *Goodness-of-Fit Techniques*. Marcel Dekker.)

La situazione diventa ancor più problematica quando i parametri vengono stimati con metodi differenti dalla massima verosimiglianza, quali ad esempio il metodo dei momenti o degli L-momenti. In tali casi, non è più possibile stabilire un legame asintotico con la distribuzione χ^2 , rendendo l'applicazione del test teoricamente meno giustificata (Stedinger, J. R., Vogel, R. M., & Foufoula-Georgiou, E. (1993). Frequency analysis of extreme events. In D. R. Maidment (Ed.), *Handbook of Hydrology*. McGraw-Hill; Hosking, J. R. M., & Wallis, J. R. (1997). *Regional Frequency Analysis: An Approach Based on L-moments*. Cambridge University Press).

Inoltre, l'aggregazione dei dati in classi riduce la quantità di informazione disponibile, rendendo il test di Pearson generalmente meno potente rispetto ad altri test di adattamento (Stephens, M. A. (1974). EDF statistics for goodness of fit and some comparisons. *Journal of the American Statistical Association*, 69(347), 730–737).

Per queste ragioni, in ambito idrologico si sconsiglia l'uso del test di Pearson quando la dimensione campionaria è inferiore a un centinaio di osservazioni e quando il numero di parametri stimati è maggiore di due, condizioni che si verificano frequentemente nelle applicazioni pratiche (Vogel, R. M., & Fennessey, N. M. (1993). L moment diagrams should replace product moment diagrams. *Water Resources Research*, 29(6), 1745–1752). In tali contesti, risulta preferibile ricorrere a test più potenti e robusti, come il test di Anderson-Darling, presentato nella pagina seguente.

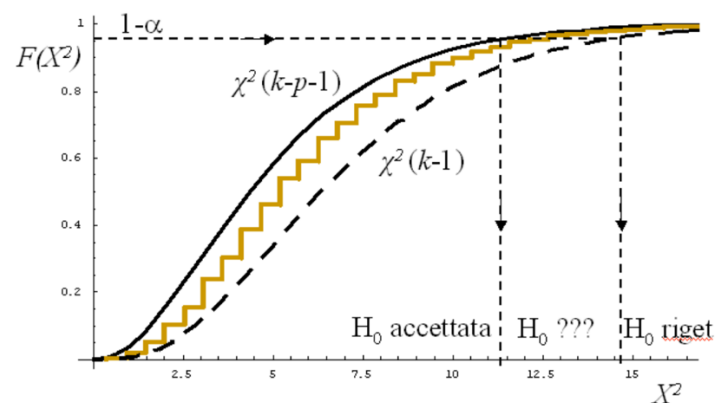


Figura 23- Rappresentazione grafica dei limiti di accettazione del test di Pearson. Da: http://www.idrologia.polito.it/didattica/Idrologia/2016/blocco1/Dispensa_Verifica_Ipotesi_test_2016.pdf

Il test di Anderson-Darling non impiega suddivisioni in classi del campione ma confronta la funzione di probabilità cumulata ipotizzata $P(x)$ con la funzione di frequenza empirica $F(x)$, quindi il test valuta la discrepanza tra campione e popolazione.

La variabile test è A^2 definita come segue:

$$A^2 = n \cdot \int \frac{[P(x)-F(x)]^2}{P(x) \cdot [1-P(x)]} \cdot p(x) dx \quad (10)$$

Dove:

- a numeratore si trova la distanza complessiva tra le due curve al quadrato;
- il denominatore costruito in questo modo serve per attribuire maggiore rilevanza alle code della distribuzione;
- l'integrale si calcola su tutti i valori della variabile x , risulta quindi essere formalmente una media.

Il test si concentra principalmente sulle code della distribuzione, infatti, tanto più piccola è A^2 tanto più vicine sono $F(x)$ e $P(x)$ e tanto è più probabile che H_0 sia vera.

Dalla letteratura (Laio, F. (2004)) viene riportata una formula semplificata di A^2 che è stata utilizzata per l'analisi statistica in questo caso studio.

$$A^2 = -n - \frac{1}{n} \sum_{i=1}^n [(2i-1) \ln[F(x_i, \theta)] + (2n+1-2i) \ln[1-F(x_i, \theta)]] \quad (11)$$

Viene riportata di seguito la corrispondente formula in R:

```
AD <- function (n, P_i) {                                     #variabile test
  -n - (1/n) * sum((seq(1, 2 * n - 1, by = 2)) * log(P_i) +
                  (seq(2 * n - 1, 1, by = -2)) * log(1 - P_i))
}
```

Figura 24- Funzione in R che permette di definire la variabile test di Anderson-Darling

Nella sezione Allegati di questo progetto di tesi viene riportata per intero la funzione "ADtestMC" impiegata per condurre i test sulle distribuzioni di dati, in R, descritte in precedenza. In questo paragrafo viene riportato solamente il workflow della funzione, diviso in sei passaggi, che ne permette di capire il funzionamento:

1. Si definiscono gli input della funzione ossia i dati e le grandezze di cui la funzione ha bisogno:

- `x`: il campione di dati osservati;
- `distr`: la distribuzione teorica da testare (*Normal*, *Exp*, *Gumbel*, *GEV*, *GP*);
- `method`: metodo di stima dei parametri della distribuzione:
 - `"mom"` = metodo dei momenti;
 - `"lmom"` = metodo degli L-momenti;
 - `"GML"` = stima con generalised maximum likelihood (GML).
- `Nsim`: numero di simulazioni Monte Carlo (default = 1000).

2. Successivamente avviene il calcolo dei parametri della distribuzione ipotizzata a seconda della distribuzione e del metodo scelto:

- per la *Normale*: media e deviazione standard, stimati con momenti o L-momenti;
- per la *Esponenziale*: solo il parametro di scala ($\lambda = 1/\text{mean}$);
- per la *Gumbel*: parametri ricavati da media e deviazione standard, oppure da L-momenti;
- per la *GEV*: parametri stimati con il metodo dei momenti (funzione *fparGEV*) oppure con L-momenti (*pelgev*);
- per la *GP* (Generalized Pareto): parametri stimati con L-momenti (*pargpa*) o con GML (*fitGP*).

3. In seguito avviene il calcolo della statistica di Anderson–Darling sui dati reali:

- i dati vengono prima ordinati (`x_i <- sort(x)`);
- si calcolano poi le CDF teoriche della distribuzione ipotizzata, valutata nei punti osservati (`P_x_i`);
- ed infine la statistica di Anderson–Darling A^2 viene calcolata con la formula riportata nella pagina precedente.

4. Si procede quindi con la generazione delle simulazioni Monte Carlo:

- si generano `Nsim` campioni sintetici, della stessa dimensione del campione originale, estratti dalla distribuzione teorica stimata;
- poi per ciascun campione simulato:
 1. si stimano di nuovo i parametri (per tenere conto della variabilità dovuta alla stima);
 2. si calcolano le CDF teoriche corrispondenti;
 3. si calcola la statistica A^2 anche per quel campione.

- si ottiene quindi un vettore di statistiche A_{sim}^2 che rappresenta la distribuzione empirica della statistica test sotto H_0 .

5. Ora avviene il test vero e proprio con il calcolo dei valori critici e del p-value:

- si fissano tre livelli di significatività standard: 10%, 5% e 1% (*signif* <- c (0.1, 0.05, 0.01));
- dalla distribuzione simulata delle statistiche A_{sim}^2 si calcolano i quantili corrispondenti che rappresentano i valori critici oltre i quali si rigetta H_0 ;
- il p-value è stimato come: $p_value = 1 - ecdf(A2_sim)(A2)$ cioè la proporzione di statistiche simulate più estreme rispetto a quella osservata.

6. Gli output della funzione risultano quindi essere sotto forma di vettore con:

- A^2 : valore osservato della statistica di Anderson–Darling;
- lim: valori critici della statistica ai livelli 10%, 5%, 1%;
- pval: p-value stimato.

Questa analisi è stata svolta per ogni stazione della GAR e i risultati dei test sono salvati in file in formato csv che vengono riportati nella sezione Allegati. Qui di seguito vengono riportati solamente due esempi, presi casualmente tra i 575 siti, in modo che sia possibile visualizzare l'output della funzione *ADtestMC*:

Modello	A2	lim,90%	lim,95%	lim,99%	pval
mom_norm	2,847	0,561	0,713	0,933	0,00
mom_lognorm	1,707	0,690	0,738	0,989	0,00
mom_gumbel	1,455	0,914	0,999	1,268	0,01
mom_gev	0,572	0,997	1,160	4,387	0,31
lmom_norm	2,551	0,585	0,642	0,749	0,00
lmom_lognorm	1,615	0,574	0,778	1,004	0,00
lmom_gumbel	0,978	0,682	0,851	1,028	0,02
lmom_gev	0,327	0,537	0,643	1,009	0,41
lmom_gp	0,304	1,185	2,221	4,705	0,66
gml_gp	0,739	1,490	1,591	1,804	0,78

Figura 24- Risultati test Anderson-Darling per il sito 1037.

Modello	A2	lim.90%	lim.95%	lim.99%	pval
mom_norm	2,37	0,63	0,78	1,03	0,00
mom_lognorm	1,00	0,58	0,70	0,98	0,00
mom_gumbel	1,04	0,78	0,88	1,37	0,04
mom_gev	0,79	0,56	0,63	0,73	0,00
lmom_norm	2,46	0,60	0,70	0,93	0,00
lmom_lognorm	1,01	0,53	0,66	0,81	0,00
lmom_gumbel	0,99	0,60	0,71	1,00	0,02
lmom_gev	0,41	0,49	0,56	0,70	0,24
lmom_gp	0,28	2,38	4,66	8,43	0,75
gml_gp	0,77	1,50	1,59	1,88	0,79

Figura 25- Risultati test Anderson-Darling per il sito 3274.

Le tabelle dei risultati sono strutturate in questo modo:

- modello ovvero la distribuzione teorica e metodo di stima dei parametri (mom = momenti, lmom = L-momenti, gml = massima verosimiglianza generalizzata);
- A^2 ovvero valore della statistica di Anderson–Darling calcolata sul campione reale;
- lim.90%, lim.95%, lim.99% che contengono i valori critici della statistica AD simulati con Monte Carlo: se A^2 supera questi limiti, l'ipotesi nulla H_0 , cioè che i dati seguano la distribuzione ipotizzata, viene rifiutata con il corrispondente livello di significatività.
- l'ultima colonna contiene il pval ovvero il p-value empirico ottenuto dalle simulazioni. Un valore piccolo (es. < 0.05) indica che la distribuzione scelta non si adatta bene ai dati.

Da questi due esempi, ma anche per le altre stazioni, si evince che le distribuzioni normale e log-normale, calcolate sia con il metodo dei momenti che L-momenti, non descrivono bene i dati delle portate giornaliere (p-value sempre pari a zero). Per quanto riguarda la Gumbel si può affermare che l'adattamento ai dati è sicuramente migliore delle due distribuzioni precedenti ma comunque non è sufficientemente appropriato. I modelli migliori risultano la GEV (lmom) e la Generalized Pareto (lmom, gml), che mostrano p-value > 0.05 e statistiche A^2 ben al di sotto dei limiti critici.

Per riassumere i risultati dei test su tutti i 575 siti selezionati della GAR vengono proposti tre grafici che descrivono rispettivamente qual è la distribuzione con p-value più alti in ogni sito (figura 26), la proporzione di siti con p-value maggiore di 0.05 (figura 27) e infine un grafico che mette a confronto le distribuzioni impiegate nel test di Anderson-Darling tramite boxpot (figura 28).

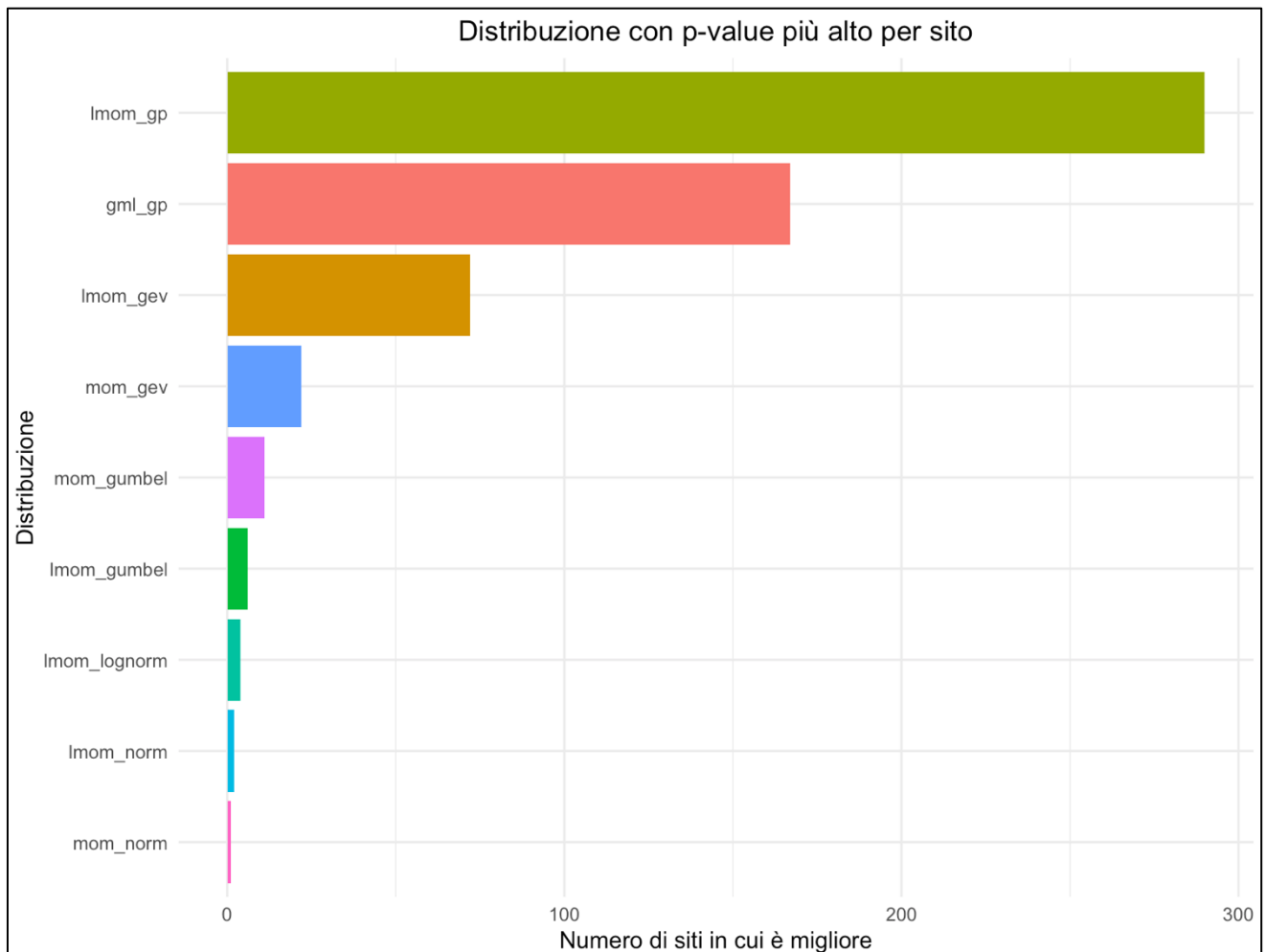


Figura 26- Distribuzioni con p-value più alto per sito: la distribuzione GP calcolata tramite il metodo degli L-momenti risulta quella migliore in quasi 300 dei 575 siti. Le altre due distribuzioni migliori sono la GP calcolata con massima verosimiglianza e la GEV calcolata con L-momenti.

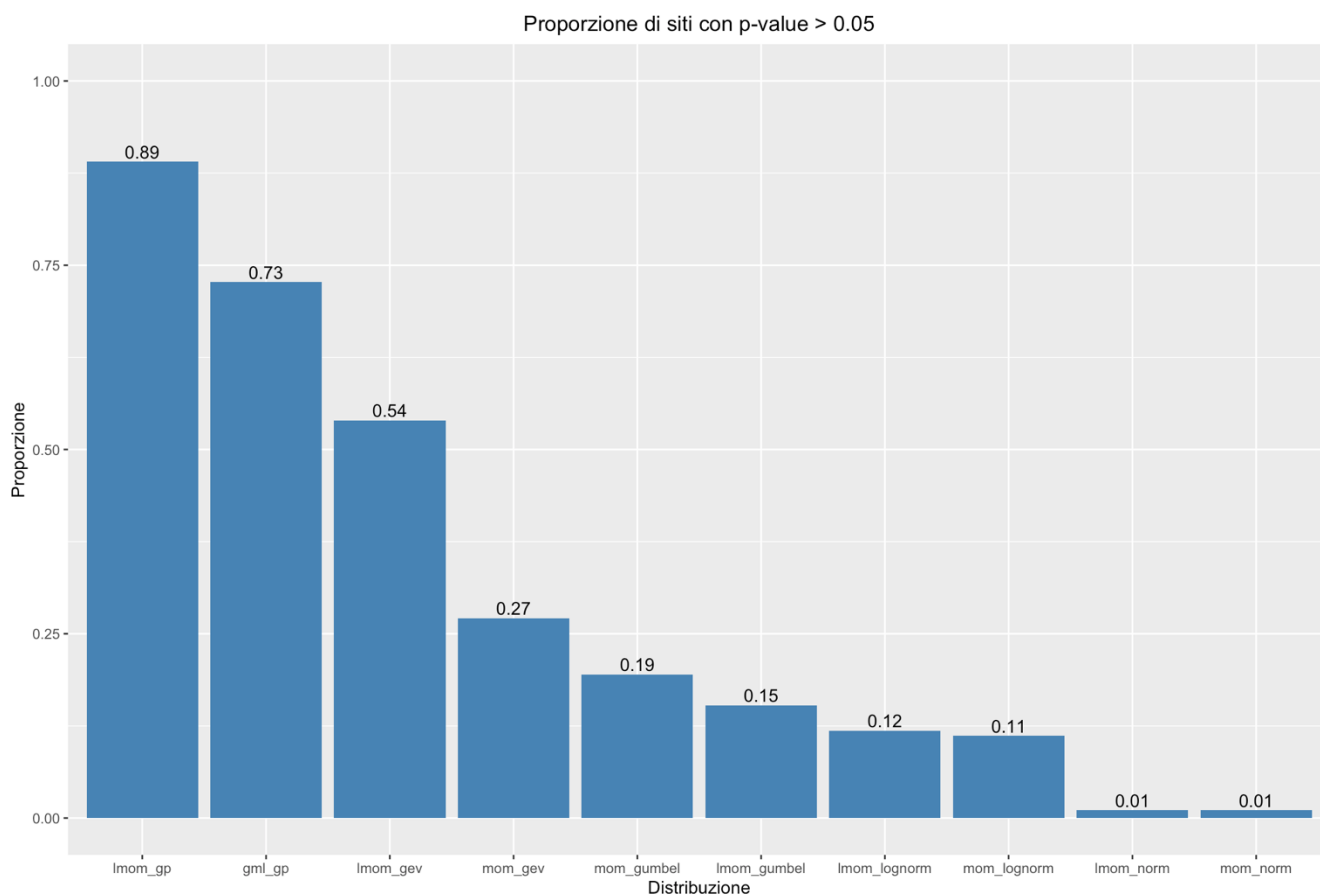


Figura 27- Proporzione di siti con p-value maggiore di 0.05. L'89% dei siti ha p-value maggiore di 0.05 con distribuzione GP calcolata con L-momenti, il 73% dei siti ha p-value maggiore di 0.05 con distribuzione GP calcolata con massima verosimiglianza e il 54% dei siti ha p-value maggiore di 0.05 con distribuzione GEV calcolata con L-momenti.

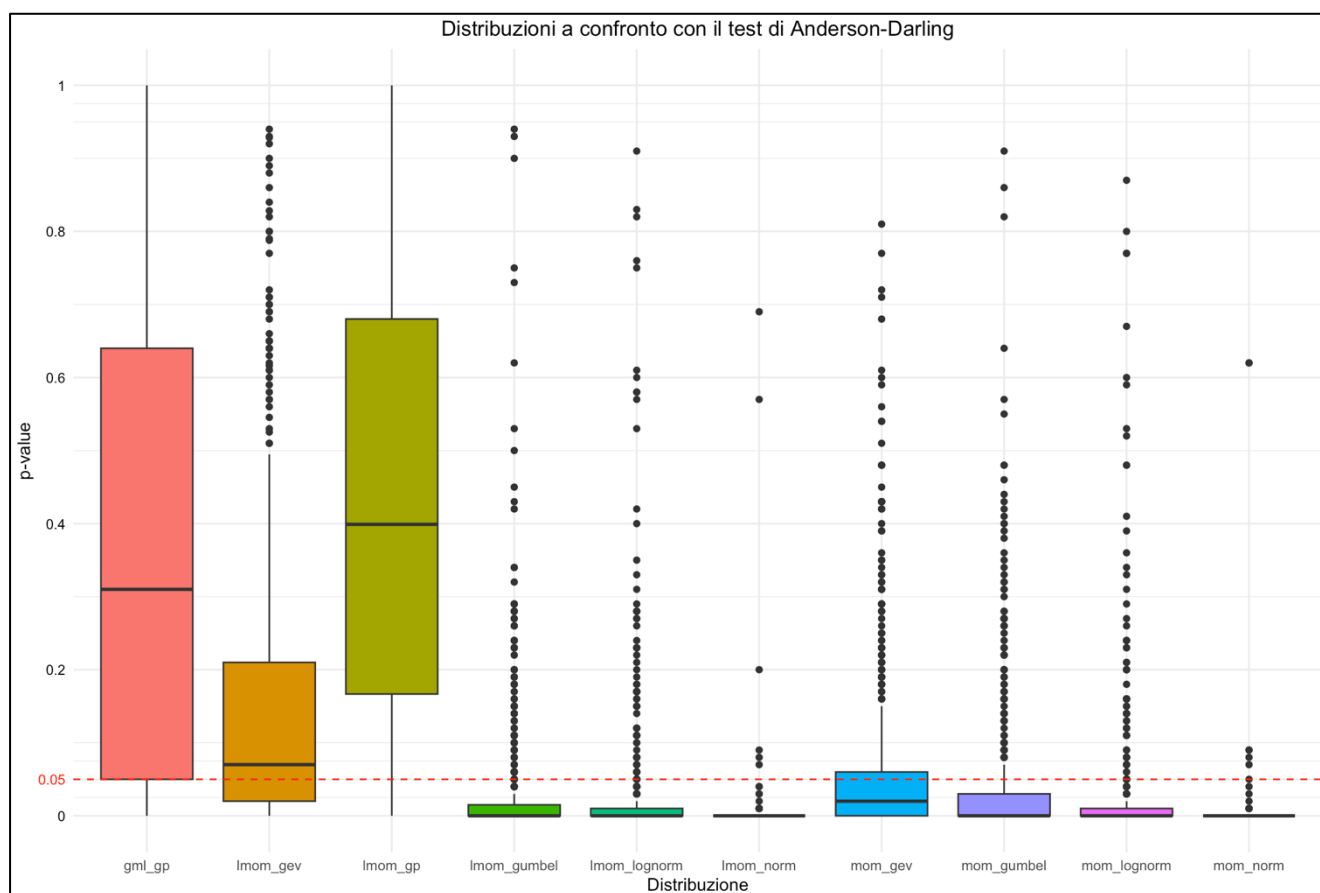


Figura 28-Distribuzioni a confronto con il test di Anderson-Darling. Il Boxplot mostra come solo le prime tre distribuzioni del grafico abbiano la mediana maggiore di 0.05.

2.5 Stima dell'incertezza.

Lo studio dell'incertezza delle distribuzioni idrologiche rappresenta un aspetto decisamente importante dell'analisi di frequenza idrologica, in quanto le decisioni progettuali basate su portate di progetto o tempi di ritorno dipendono fortemente dalla robustezza delle stime probabilistiche.

Le principali fonti di incertezza si possono ricondurre a tre categorie:

- la prima è l'incertezza campionaria, dovuta alla limitata lunghezza delle serie storiche idrologiche, che raramente superano i 50–60 anni (Stedinger et al., 1993). In tali casi, le stime dei parametri delle distribuzioni possono risultare altamente variabili;
- la seconda è l'incertezza legata alla scelta della distribuzione teorica: diversi modelli, come la GEV, la Gumbel o la GP, possono produrre quantili molto diversi soprattutto per tempi di ritorno lunghi (Hosking & Wallis, 1997; Coles, 2001);
- infine, vi è l'incertezza parametrica, connessa al metodo di stima utilizzato. Ad esempio, gli stimatori basati su massima verosimiglianza (GML) sono asintoticamente efficienti ma instabili su campioni piccoli, mentre gli L-momenti offrono maggiore robustezza in presenza di dati scarsi o distribuzioni asimmetriche (Hosking & Wallis, 1997).

Per affrontare tali problematiche, sono stati sviluppati diversi approcci metodologici. In questo studio si è scelto di agire attraverso il metodo Monte Carlo per cui è stato possibile stimare bande di confidenza attraverso campionamenti ripetuti dalle distribuzioni stimate.

Di recente, diversi studi hanno mostrato come la scelta proprio della distribuzione e del metodo di stima possano portare a differenze significative nelle stime di piena. Un esempio è *Serinaldi F. et al. (2009)* che evidenzia come stimatori diversi possono produrre quantili divergenti anche su dataset identici, sottolineando la necessità di confrontare più modelli.

Quindi il problema dell'incertezza non può essere eliminato, ma deve essere quantificato e comunicato, così da fornire strumenti decisionali più solidi nella gestione del rischio idraulico e nella progettazione di opere di difesa.

In questa trattazione per quantificare l'incertezza delle curve di frequenza di piena e dei quantili di progetto si è impiegata la funzione "*FFCuncertMC*". Essa permette di confrontare più modelli matematici, di valutare gli intervalli di confidenza e permette anche di valutare quanto sono affidabili le stime dei quantili estremi, evidenziando l'incertezza dovuta alla stima dei parametri del modello probabilistico scelto. Allo stesso modo della funzione *ADtestMC*, non verrà riportata l'intera funzione (presente nella sezione Allegati) ma solamente il workflow che ne riassume e spiega il funzionamento in cinque passaggi:

1. Gli input e il set-up della funzione: primo input sono le serie x di portate (massimi selezionati con il metodo POT), secondo è il tipo di distribuzione (*distr* = GEV o GP), terzo il metodo di stima (*method* = mom, lmom, GML) e infine quarto il Nsim ovvero il numero di simulazioni (sono state impostate mille simulazioni). Per riassumere i risultati la funzione crea dei grafici con sui due assi:
 - a. *u*: coincide con la variabile ridotta della distribuzione Gumbel (con valori da -2 a 7). $F(x) = \exp(-\exp(-u))$;
 - b. *rp*: tempi di ritorno (2, 5, 10, 20, 50, 100, 200, 500, 1000 anni); inizializza le matrici *q_u_sim* e *q_rp_sim* che conterranno i quantili simulati con metodo Monte Carlo.
2. Successivamente avviene la stima dei parametri sui dati reali a seconda di *distr* e *method*:
 - GEV + mom: *fparGEV(mean, sd, skew)*
 - GEV + L-mom: *pelgev(samlmu(x))*
 - GP + L-mom: *pargpa(lmoms(x))\$para*
 - GP + GML: *optimizing(fitGP, ...)*

Con i parametri stimati calcola:

- *q_u*: quantili su griglia *u*;
- *q_rp*: quantili per i tempi di ritorno.

3. A questo punto si generano le Nsim con il metodo Monte Carlo (per i da 1 a 1000):

- 1) simula un campione della stessa lunghezza *n* della distribuzione stimata (*rGEV* o *rGP*).
- 2) ristima i parametri dal campione simulato (stesso metodo spiegato per la funzione *ADtestMC*);
- 3) ricalcola i quantili su *u* e su *rp* che poi salva in *q_u_sim[i,]*, *q_rp_sim[i,]*.

4. Ora avviene l'aggregazione dell'incertezza:

su ogni colonna (cioè su ciascun u e ciascun T) calcola i percentili (5%, 10%, 25%, 50%, 75%, 90%, 95%), la media tra simulazioni e con *rbind* aggiunge anche il punto-stima (calcolato sui dati reali).

5. Si ottengono infine i seguenti output come lista:

- u e q_u : bande di incertezza della curva in scala Gumbel.
- rp e q_{rp} : bande (percentili) e media dei quantili di progetto per ciascun T .
- q_{rp_sim} : la matrice grezza delle simulazioni sui tempi di ritorno.

Per ogni stazione, selezionata in precedenza, della GAR è stata eseguita la stima dell'incertezza e di seguito ne vengono riportati solamente due esempi scelti anche questa volta in modo casuale tra i 575 siti analizzati: per ognuno dei siti inoltre sono stati salvati in file csv i valori quantili corrispondenti a diversi tempi di ritorno, le medie e le deviazioni standard.

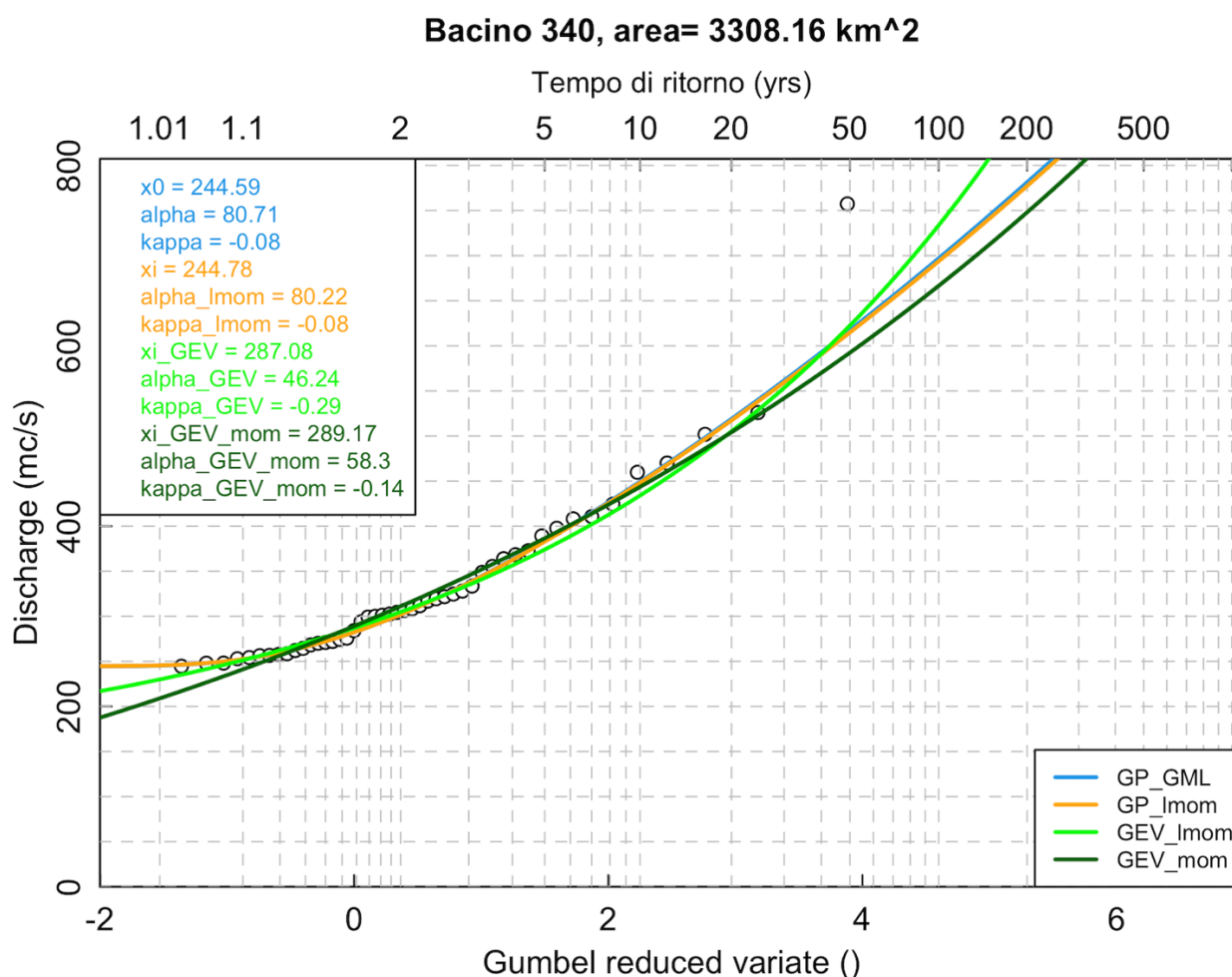


Figura 29- Rappresentazione grafica su carta di Gumbel dell'andamento delle distribuzioni GP e GEV per il sito 340. I pallini bianchi rappresentano i dati osservati.

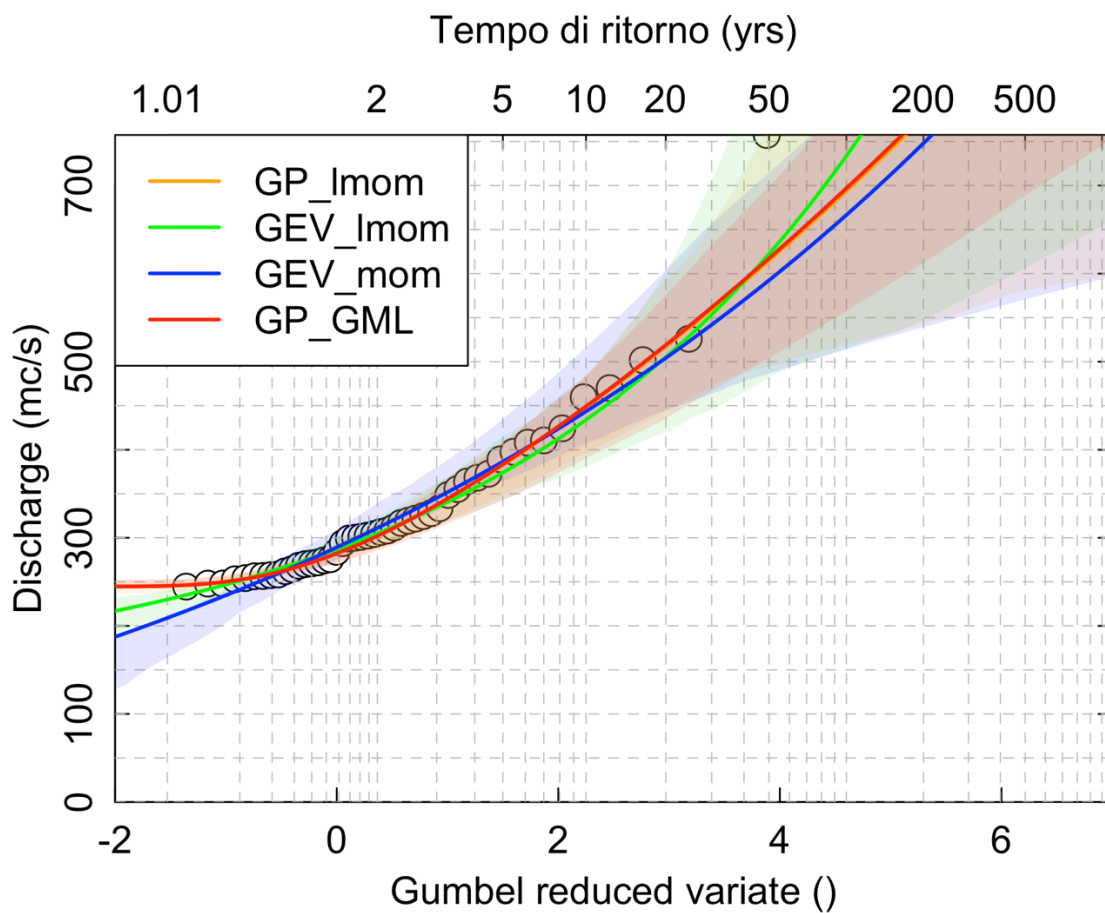


Figura 30- Rappresentazione grafica su carta di Gumbel dei risultati dell'analisi incertezza per il sito 340.

Tabella 1- Dati di quantili delle portate del sito 340 con medie e deviazioni standard.

quantili_340								
	T_2yrs	T_5yrs	T_20yrs	T_50yrs	T_100yrs	T_200yrs	T_500yrs	T_1000yrs
GP_GML	302	384	519	617	697	781	900	997
GP_Imom	302	383	517	615	694	777	895	990
GEV_Imom	305	374	505	622	734	869	1096	1312
GEV_mom	311	387	504	592	667	748	868	970

Media		SD	
	x		x
T_2yrs	306	T_2yrs	12
T_5yrs	381	T_5yrs	22
T_10yrs	440	T_10yrs	33
T_20yrs	503	T_20yrs	51
T_50yrs	598	T_50yrs	94
T_100yrs	680	T_100yrs	149
T_200yrs	774	T_200yrs	233
T_500yrs	924	T_500yrs	416
T_1000yrs	1063	T_1000yrs	640

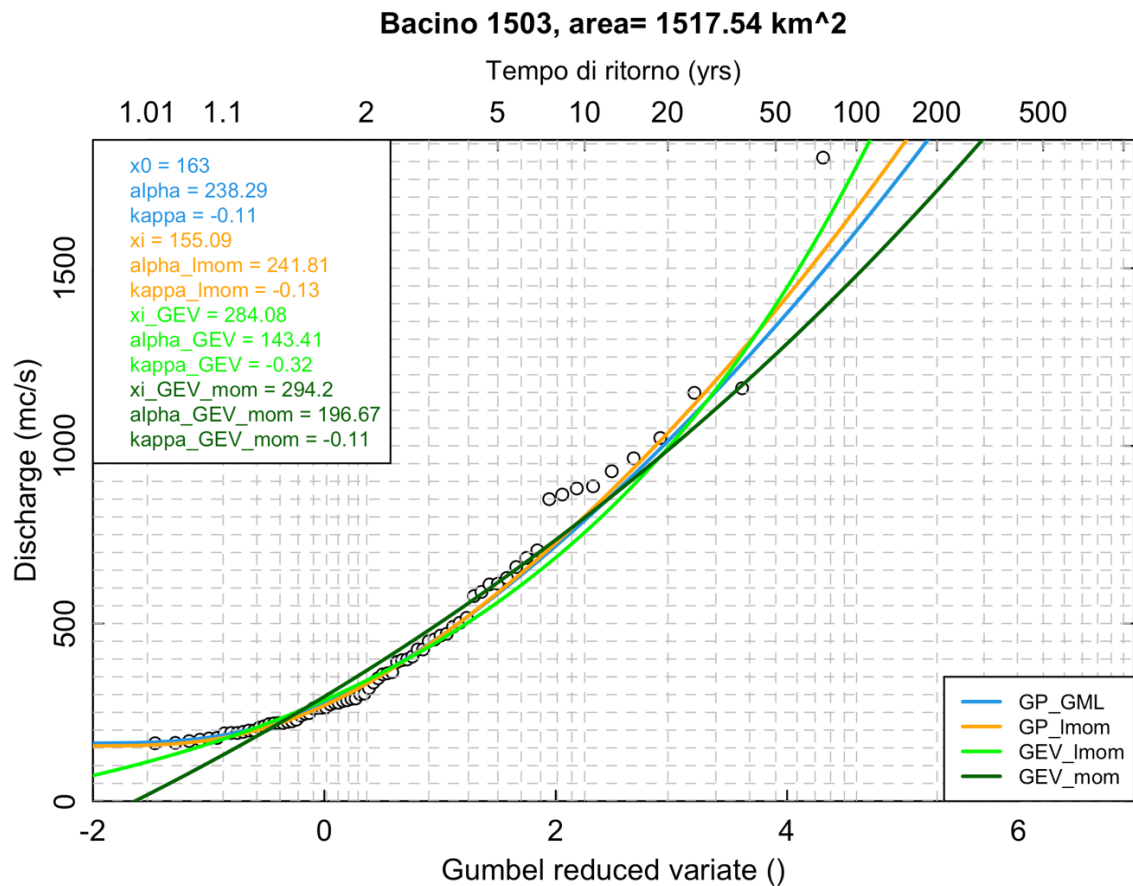


Figura 31- Rappresentazione grafica su carta di Gumbel dell'andamento delle distribuzioni GP e GEV per il sito 1503. I pallini bianchi rappresentano i dati osservati.

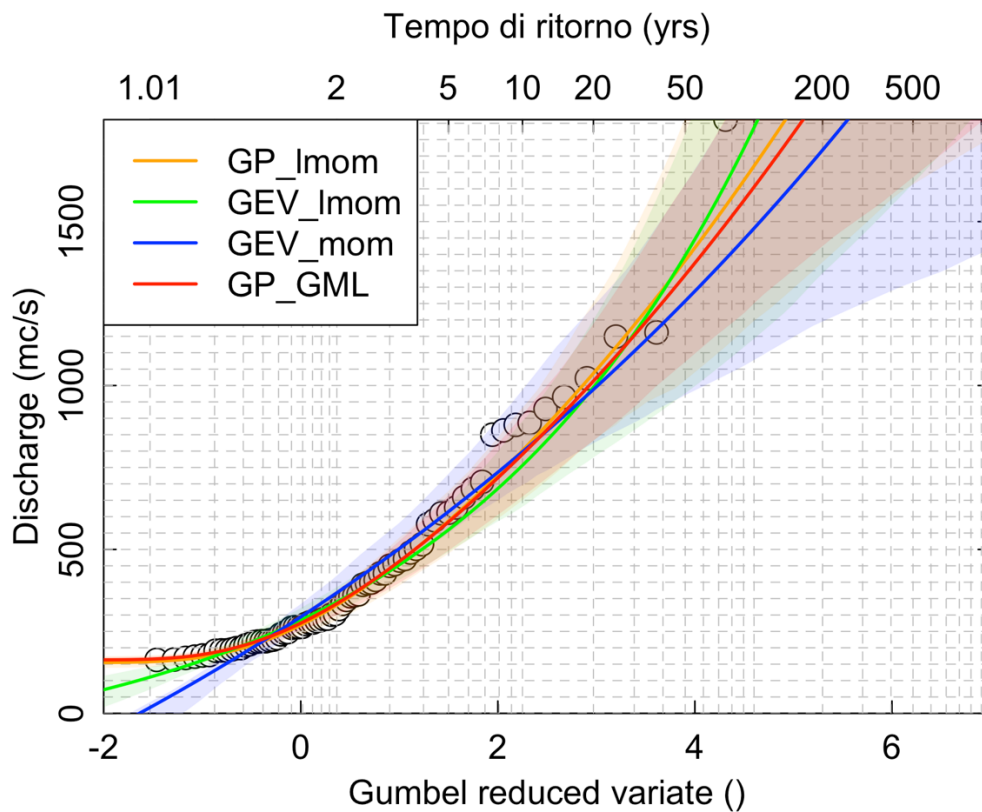


Figura 32- Rappresentazione grafica su carta di Gumbel dei risultati dell'analisi incertezza per il sito 1503.

Tabella 2- Dati di quantili delle portate del sito 1503 con medie e deviazioni standard.

quantili_1503

	T_2yrs	T_5yrs	T_20yrs	T_50yrs	T_100yrs	T_200yrs	T_500yrs	T_1000yrs
GP_GML	335	584	1014	1337	1605	1896	2316	2665
GP_Imom	330	587	1037	1381	1670	1985	2446	2833
GEV_Imom	340	560	995	1396	1787	2272	3102	3913
GEV_mom	368	616	988	1258	1479	1718	2063	2349

Media		SD	
	x		x
T_2yrs	346	T_2yrs	32
T_5yrs	588	T_5yrs	61
T_10yrs	783	T_10yrs	88
T_20yrs	996	T_20yrs	131
T_50yrs	1317	T_50yrs	228
T_100yrs	1597	T_100yrs	346
T_200yrs	1917	T_200yrs	516
T_500yrs	2420	T_500yrs	854
T_1000yrs	2877	T_1000yrs	1225

Per entrambi i casi riportati si possono fare alcune osservazioni sull'andamento delle curve e i valori numerici corrispondenti. Per tempi di ritorno bassi ($T < 20$ anni) tutte le distribuzioni forniscono stime molto simili, con intervalli stretti e buona coerenza con i dati osservati. Ciò significa che l'incertezza è ridotta e la scelta del modello incide poco nella stima.

Per tempi di ritorno maggiori (20–100 anni) iniziano invece a comparire differenze tra le distribuzioni, soprattutto tra GEV (più alta) e GP (più bassa). Si riscontra che l'incertezza cresce progressivamente, come atteso.

Per tempi di ritorno alti (≥ 200 anni): le divergenze diventano marcate, infatti nel caso del sito 340:

- la GEV con L-momenti produce quantili sensibilmente più elevati (ad esempio per $T=1000$ anni si ottiene $1312 \text{ m}^3/\text{s}$);
- la GP, sia L-momenti che GML, fornisce valori più bassi e simili tra loro (rispettivamente 990 e $997 \text{ m}^3/\text{s}$).

Gli intervalli di confidenza diventano molto ampi, a conferma dell'incertezza elevata nella stima degli eventi rari. Ciò si riscontra in modo ancora più evidente per il sito 1503.

Per quanto riguarda la media dei quantili, essa mostra una crescita regolare con il tempo di ritorno mentre la deviazione standard aumenta molto velocemente (per il sito 340 da 12 m³/s a T=2 anni fino a 640 m³/s a T=1000 anni), indicando che l'incertezza cresce in modo esponenziale con il tempo di ritorno.

La GEV con L-momenti fornisce le stime più elevate e quindi più conservative per la gestione del rischio idraulico mentre la GP stimata con GML e L-momenti mostra valori simili e più contenuti, suggerendo maggiore stabilità ma con rischio di sottostima per eventi rari.

Grazie ai seguenti tre grafici è possibile valutare di quanto si discostano le stime con le diverse distribuzioni impiegate. Inoltre, è possibile capire la relazione tra i risultati della GEV e la GP.

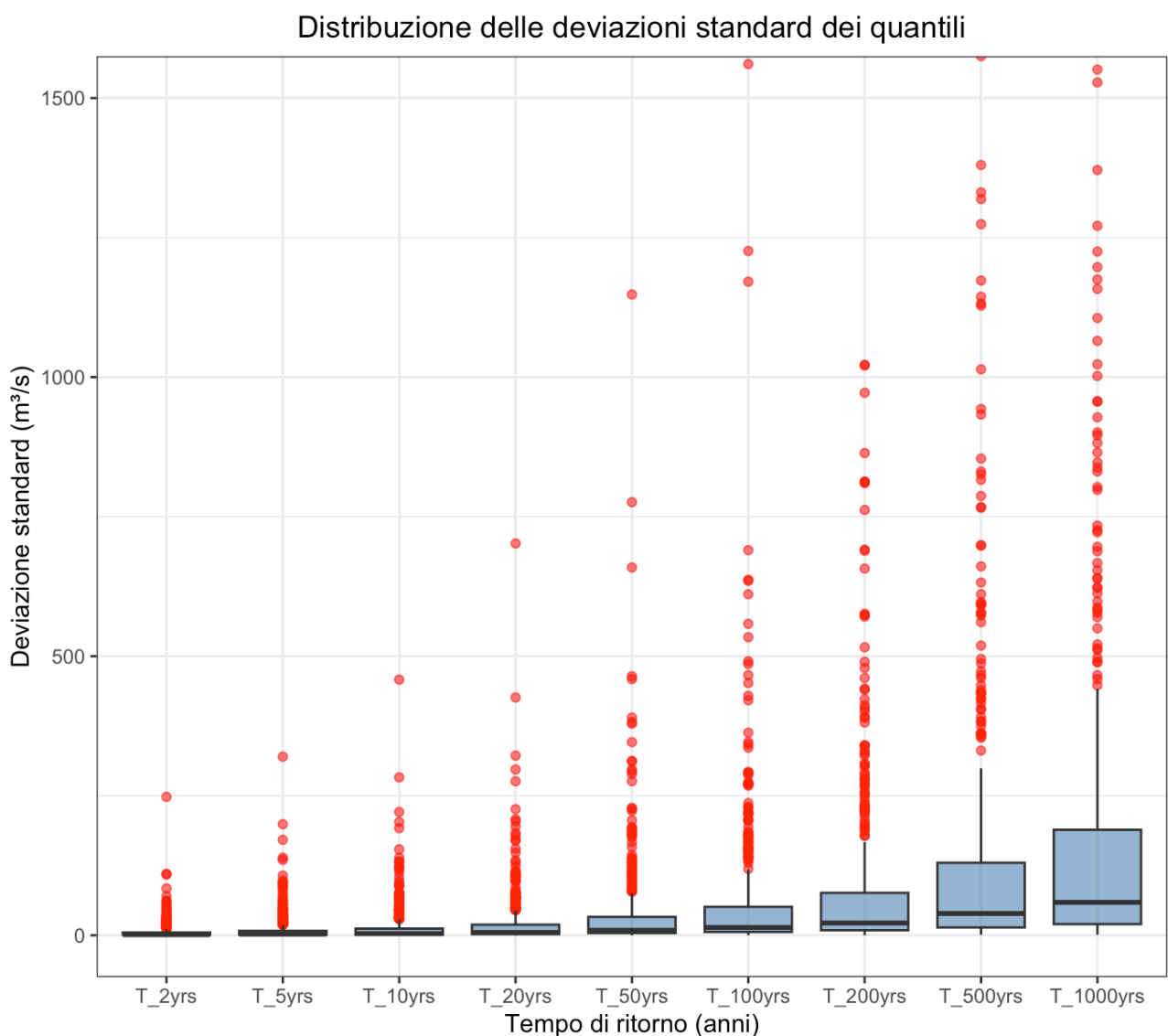
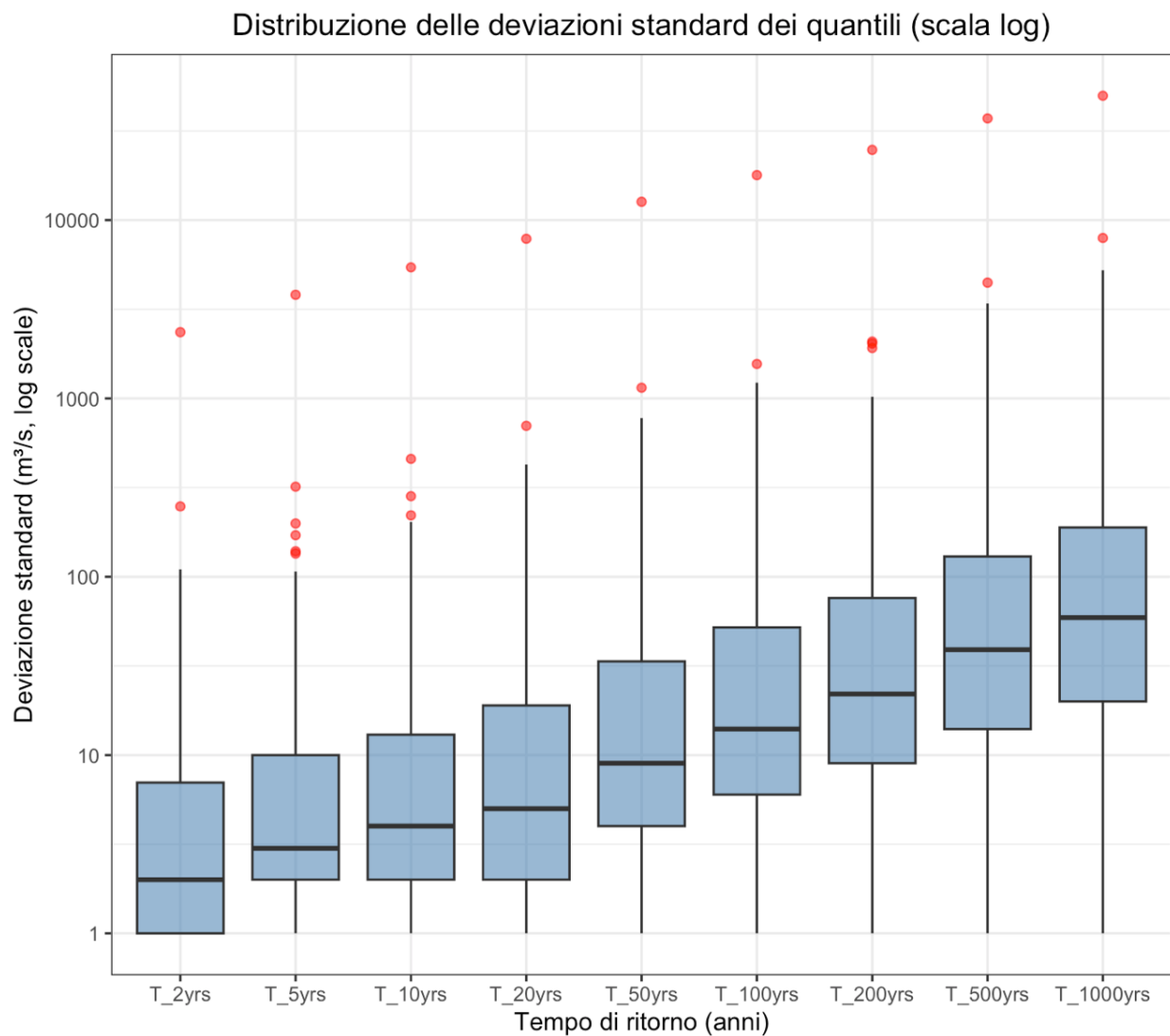


Figura 33-Distribuzione delle SD dei quantili di tutti della GAR: questo grafico mostra per quali tempi di ritorno l'incertezza cresce di più. Come atteso per tempi di ritorno alti l'incertezza aumenta, rimanendo comunque con mediane abbondantemente inferiori ai 250 m³/s.



*Figura 34- Visualizzazione in scala logaritmica della distribuzione delle SD dei quantili.
In questo modo è possibile visualizzare gli outlier*

In figura 33 viene proposto uno zoom della distribuzione delle deviazioni standard su valori minori di 2000 m³/s, in modo che siano più chiare dimensioni dei Boxplot.

Per descrivere del tutto i risultati dell'analisi viene anche riportata la stessa tipologia di grafico ma in scala logaritmica. Quest'ultimo, in figura 34, mostra gli elevati valori di deviazione standard che vengono raggiunti in rari casi. È da sottolineare comunque che la mediana dei Boxplot si assesta su valori inferiori ai 100 m³/s anche per tempi di ritorno pari a 1000 anni, suggerendo che la stima è relativamente robusta.

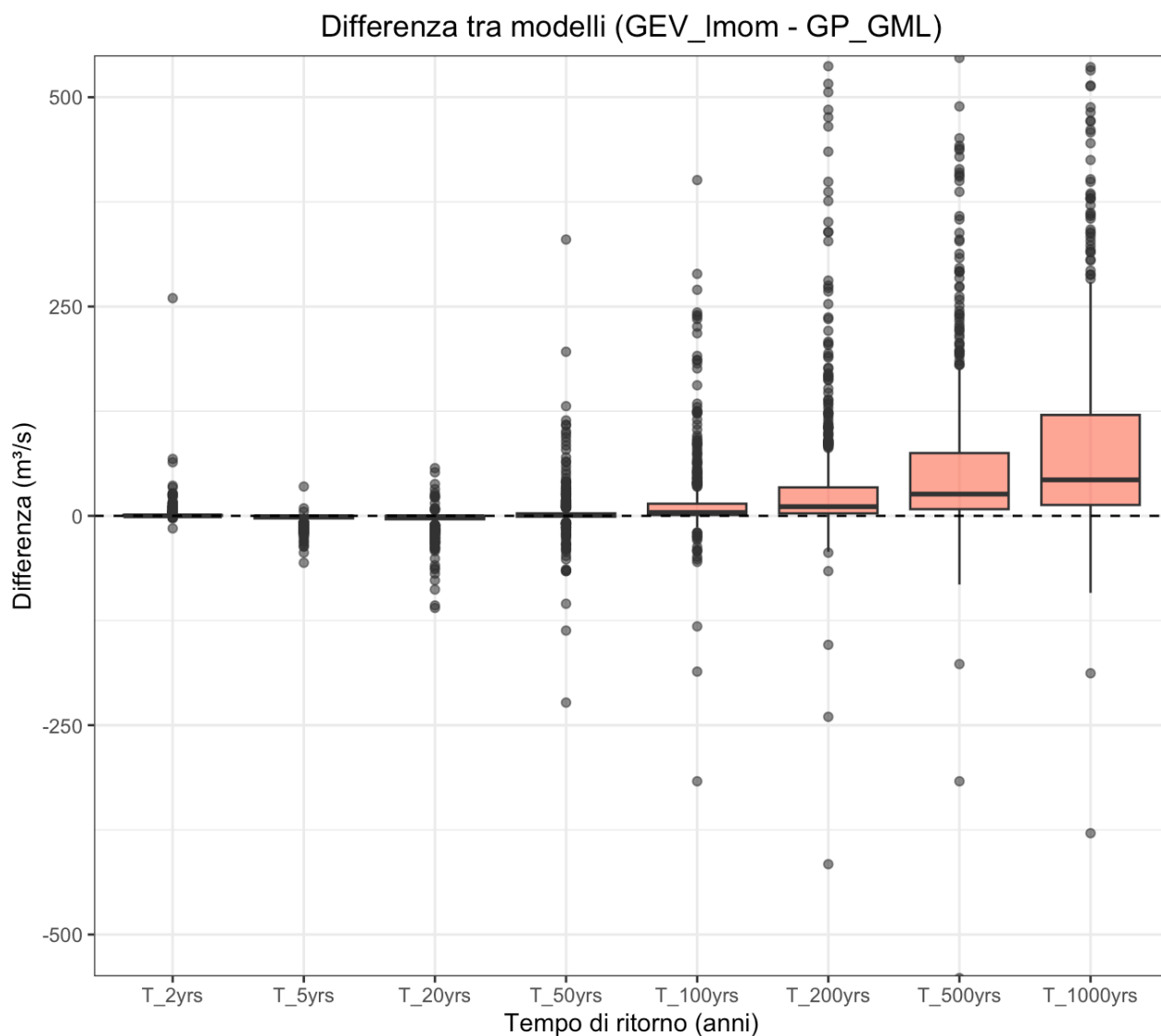


Figura 35- Grafico che mostra la differenza tra i valori dei quantili calcolati con i modelli GEV_lmom e GP_GML: in questo modo si può valutare se, in media, GEV_lmom sovrastima o sottostima rispetto a GP_GML. Se la mediana del Boxplot è sopra lo 0 allora $GEV_lmom > GP_GML$, se sotto invece $GEV_lmom < GP_GM$.

Dal grafico in figura 35 è possibile notare come per tempi di ritorno inferiori a 50 anni la differenza tra le due distribuzioni è nella maggior parte dei casi sostanzialmente trascurabile quindi, la scelta del modello per queste stime non è fonte sostanziale di incertezza.

Diversamente per tempi di ritorno pari a 100, 200, 500 e 1000 anni il grafico mostra il Boxplot gradualmente ampliarsi e la mediana aumentare. In particolare, la differenza tra le stime delle due distribuzioni risulta essere nella maggior parte dei casi positiva evidenziando quindi che, per questi tempi di ritorno, la GEV tende sovrastimare le stime di portata rispetto alla GP.

3. Estrazione e visualizzazione dei maggiori eventi storici della GAR attraverso analisi multivariate.

Una volta conclusa la prima della trattazione e determinato qual è il modello matematico che meglio descrive i dati di portate giornaliere della maggior parte delle stazioni di questo caso studio, ovvero la Generalized Pareto calcolata con il metodo degli L-momenti, si procede con l'estrazione e visualizzazione dei grandi eventi di piena sovraregionali della GAR nel periodo tra il 1899 e il 2024.

Attraverso analisi multivariate, che verranno introdotte nel paragrafo successivo, ci si pone l'obiettivo di creare un indice di danno che descriva a livello spaziale la gravità di un evento: questo permetterà di creare una classifica dei 15 eventi più gravosi. Essi verranno quindi visualizzati su mappe leaflet in R e si confronteranno i risultati con immagini dell'epoca e informazioni dei quotidiani locali.

Come base teorico-metodologica dei due macro-capitoli che seguono è stato utilizzato lo studio Blöschl et al. (2022) dal titolo HOchwasserRisikozonierung Austria 3.0 (HORA 3.0) (Flood risk zoning Austria 3.0), realizzato con la collaborazione del Professore Alberto Viglione nel 2022, relatore della presente tesi.

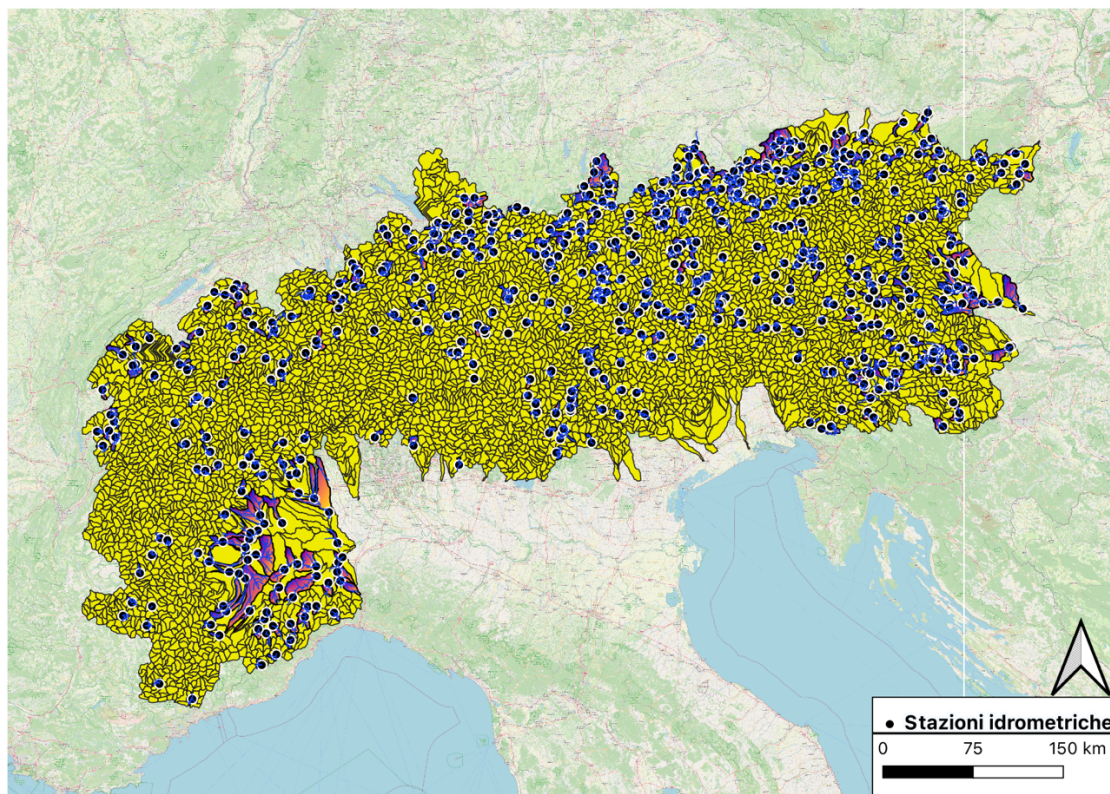


Figura 36- Totalità dei bacini idrografici della GAR (giallo), I bacini colorati sono quelli corrispondenti alle stazioni selezionate per questo caso studio.

3.1 Creazione dell'indice di magnitudo degli eventi sulla GAR.

Prima di iniziare a trattare gli eventi alluvionali ed in particolare come questi sono stati estratti da serie storiche di portate giornaliere delle stazioni della GAR, bisogna definire nello specifico cosa si intende per evento alluvionale estremo.

Nel presente lavoro, con eventi di piena rilevanti (*major flood events*) si intendono quegli eventi che sono caratterizzati da episodi di piena che superano di molto le condizioni di deflusso normali e che generano impatti idrologici e/o sociali significativi. In particolare, essi sono definiti come gli eventi di piena regionali caratterizzati da un picco dell'indice di danno medio giornaliero normalizzato (D_m), la cui definizione verrà trattata in seguito. Esso è calcolato su base multivariata e spazialmente pesata. Grazie a un filtro passa-basso, definito su un periodo di cento giorni, è stato possibile superare la variabilità di fondo dei dati ed identificare quindi i picchi delle distribuzioni di dati (e di conseguenza i rispettivi indici di danno che li descrivono): questo consente di concentrare l'analisi sui valori significativi, riducendo rumore e distorsioni statistiche. Definito, quindi, il target dello studio, si procede con l'analisi dell'algoritmo che ha portato all'estrazione degli eventi sopra descritti.

Inizialmente si vuole assegnare una probabilità di non superamento a tutti i valori delle portate giornaliere, raccolti nella matrice *dailyQ* dello script presente nella sezione Allegati. Per farlo bisogna però definire i limiti inferiore e superiore dei valori di portata da modellare: eventi al di sotto della soglia inferiore (*low*) sono considerati non estremi e non modellabili dalla GP in quanto in termini idrologici piccole portate giornaliere o minime precipitazioni non generano rischi significativi.

Per quanto riguarda il limite superiore (*upp*), in particolare nei casi in cui il parametro kappa è maggiore di zero, la GP definisce un massimo teorico oltre il quale la distribuzione non può andare. Questo avviene perché la GP stessa con parametro di forma (k) maggiore di zero tende a produrre valori estremi molto grandi (illimitati teoricamente). È utile introdurre un limite superiore perché in natura esiste chiaramente un limite fisico per portate o precipitazioni, imposto da dimensione del bacino, capacità del fiume o vincoli geomorfologici. Inoltre, dati anche i risultati dell'analisi dell'incertezza svolta in precedenza, valori oltre il limite superiore non possono essere stimati con affidabilità dal modello.

Non definire *low* e *upp* può portare quindi a:

- stime distorte dei parametri della GP;
- probabilità di non superamento maggiori di 1 o minori di 0;
- previsioni di eventi “impossibili” (ad esempio portate negative o infinite).

I valori delle soglie dipendono dai parametri della distribuzione:

- il valore di soglia inferiore coincide con il parametro di posizione della distribuzione GP, diverso per ogni sito, stimato con il metodo degli L-momenti;
- il valore di soglia superiore invece, nei casi in cui $k > 0$, risulta essere pari a:

$$upp = xi + \frac{\alpha}{k} \quad (12)$$

Con: xi , parametro di posizione della GP, α parametro di scala e $kappa$ parametro di forma. Si crea successivamente un filtro logico che permetta di estrapolare solo i valori interni all'intervallo modellabile e di questi valori si calcolano le probabilità di non superamento grazie alla funzione p_{GP} descritta nel paragrafo 2.2. Si assegna probabilità 0 a tutte le portate sotto la soglia minima modellabile e 1 a quelle sopra la soglia massima.

Si trasforma poi la probabilità in tempo di ritorno medio per ciascun giorno e ciascun sito grazie alla seguente relazione:

$$T = \frac{1}{ev_x_yr \cdot (1 - P_u)} \quad (13)$$

Per passare dal concetto di tempo di ritorno (matrice t_n) ai danni (matrice d_n) si è deciso di utilizzare una funzione di pseudo-danno che trasforma il logaritmo naturale del tempo di ritorno in un valore di pseudo-danno normalizzato tra 0 e 1 con andamento lineare (funzione a rampa, mostrata graficamente in figura 37).

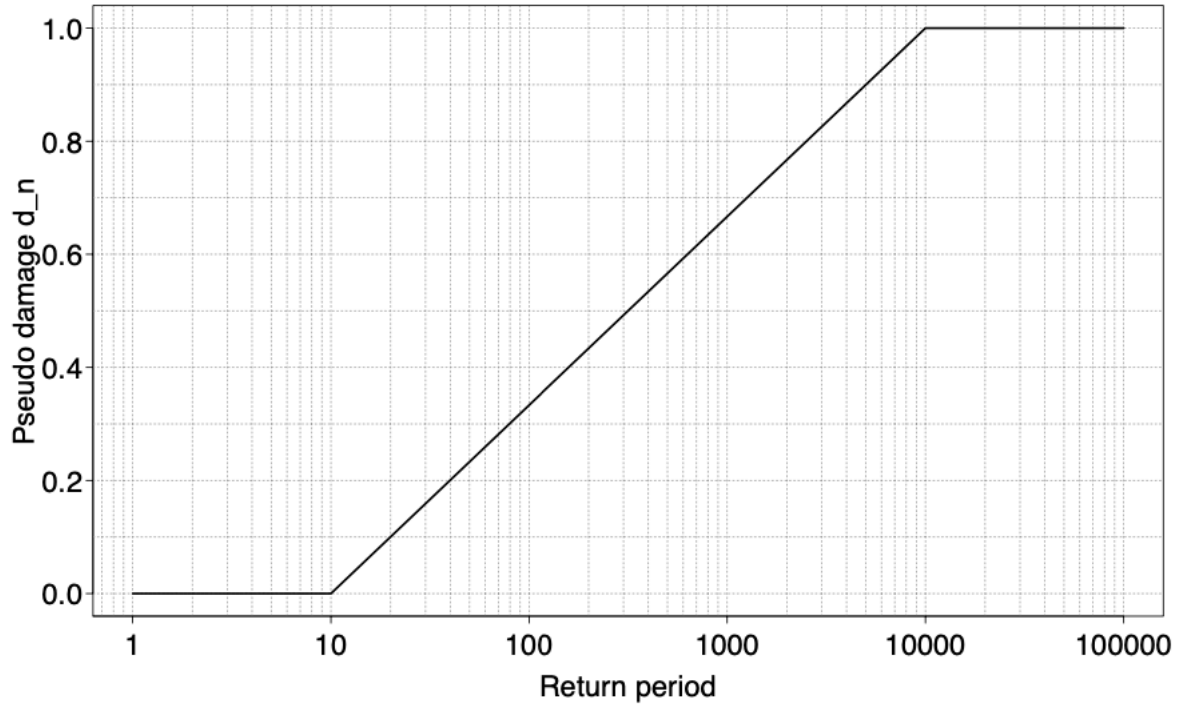


Figura 37- Funzione “ramp” di pseudo-danno. Dallo studio Blöschl et al. (2022)..

Agli eventi frequenti, ovvero con tempo di ritorno tra 1 e 10 anni, è assegnato valore del danno 0 (nessun danno). Per gli eventi estremi, con tempo di ritorno pari o superiore a 10000 anni viene assegnato invece valore 1 (danno massimo). Per quanto riguarda gli eventi intermedi: *ramp* assegna danni crescenti per eventi sempre più rari, in modo lineare come mostrato in figura 37.

Si applica questa funzione a tutte le serie di t_n , ottenendo la matrice d_n con lo pseudo-danno giornaliero per ciascun sito.

Successivamente, per ciascun giorno i viene determinato un danno medio giornaliero (D_m) a livello spaziale, calcolato come media su tutti gli n siti con dati disponibili grazie alla seguente relazione:

$$D_m = \frac{1}{n} \cdot \sum_{j=1}^n g(j) \cdot window[d_n(j, i), f_b] \quad (14)$$

La funzione *window()* ha lo scopo di individuare, per ciascun giorno i , i valori massimi di $d_n(j, i)$ all'interno di una finestra temporale mobile di ampiezza $f_b = 15$ giorni. In pratica, per ogni giorno considerato vengono inclusi anche i 7 giorni precedenti e i 7 giorni successivi, così da cogliere la persistenza e l'intensità degli eventi di piena su scala plurigiornaliera. Contestualmente, $g(j)$ rappresenta un coefficiente di ponderazione legato alle caratteristiche

del sito, ovvero la componente spaziale del peso, che in questa applicazione è stato assunto pari al logaritmo naturale dell'area del bacino imbrifero, ipotizzando quindi che bacini più estesi siano associati a potenziali danni maggiori.

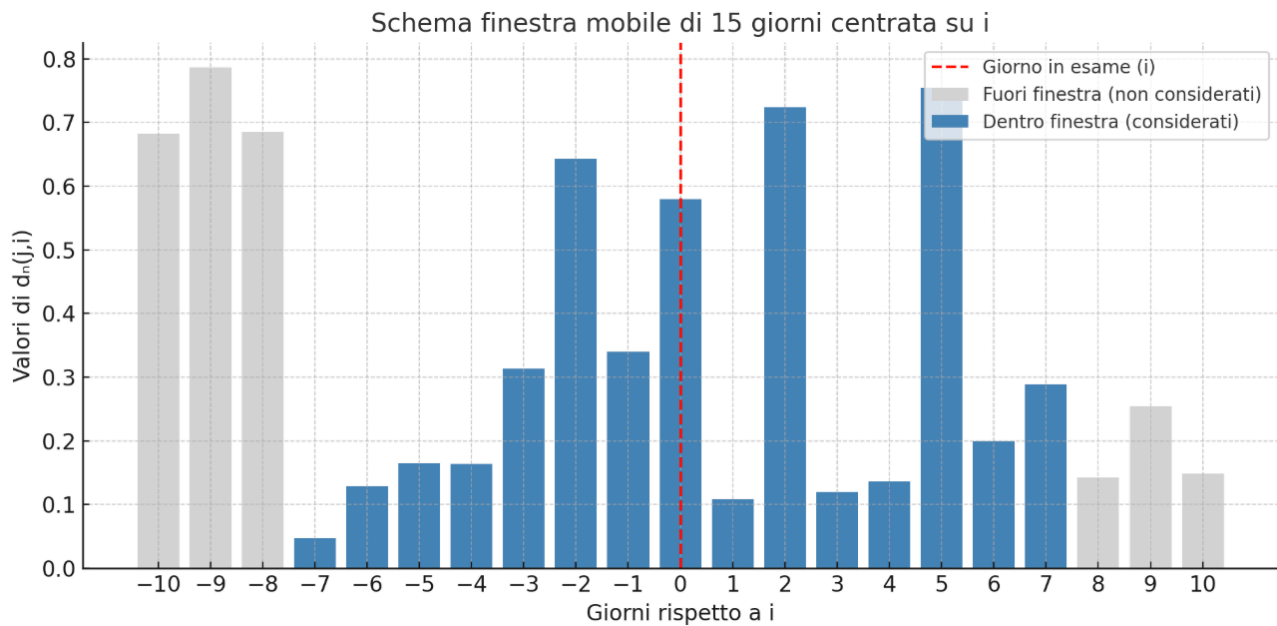


Figura 38- Esempio esplicativo della finestra mobile impiegata in questa analisi, i valori di d_n sono randomici in questo caso.

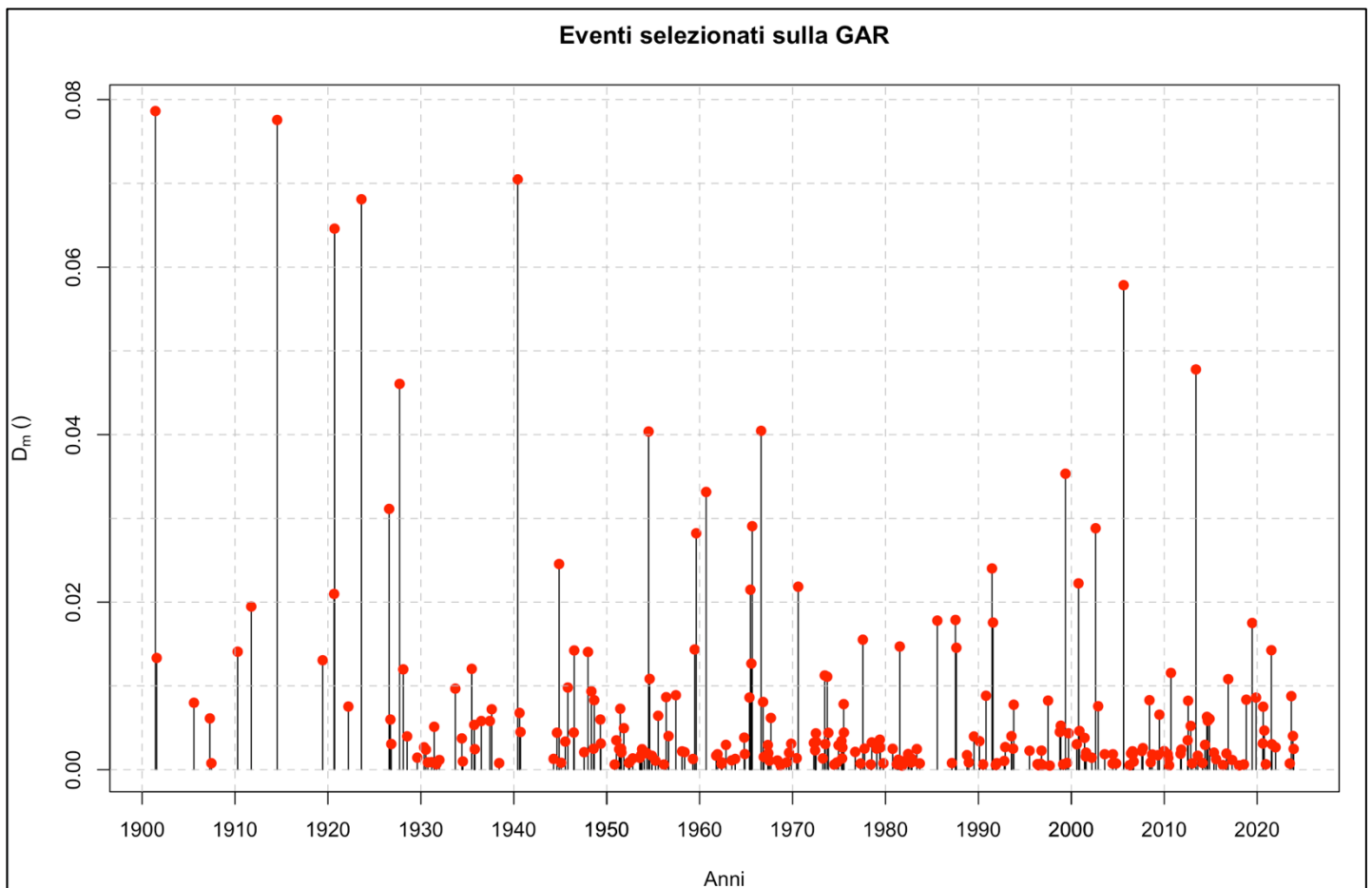


Figura 39- Plot di tutti gli eventi con danni intermedi dal 1899 al 2024.

Dall'analisi emergono picchi di D_m particolarmente elevati nei primi decenni del Novecento. Tuttavia, questo risultato va interpretato con cautela. Infatti, nello stesso intervallo temporale il grafico di consistenza dei dati mostra che il numero di stazioni idrometriche attive era molto ridotto. Ciò può comportare due possibili effetti:

- 1) il primo è la sovrastima dei valori di D_m come effetto del fatto che con poche stazioni un singolo evento estremo localizzato può influenzare in maniera sproporzionata la media spaziale pesata, facendo apparire l'evento più rilevante a livello regionale di quanto non fosse realmente;
- 2) il secondo riguarda la distribuzione spaziale delle stazioni che all'epoca era probabilmente non omogenea per cui gli eventi dei primi del Novecento non sono da considerarsi completamente affidabili.

Nei periodi successivi, con l'aumento del numero di stazioni attive e quindi con una copertura spaziale più estesa ed omogenea, i valori di D_m risultano più affidabili come indicatore del danno medio giornaliero.

3.2 Estrazione eventi regionali da picchi della serie storica dei D_m .

Per approfondire le caratteristiche degli eventi più rilevanti individuati nella serie storica (figura 39), è stata predisposta una procedura che associa ad ogni evento i valori massimi di portata giornaliera, tempo di ritorno e pseudo-danno registrati presso ciascun sito idrometrico.

L'individuazione dei massimi locali del danno medio giornaliero (D_m) è stata effettuata confrontando la serie temporale originale con una sua versione filtrata tramite un filtro passa-basso (media mobile di cento giorni). In questo modo, i punti in cui D_m supera la curva smussata ottenuta dal filtraggio rappresentano i picchi significativi, cioè i potenziali eventi di piena da considerare. Nelle seguenti pagine vengono mostrati prima la tabella risultante dei quindici eventi con valore massimo di danno medio giornaliero (D_m) più elevato, considerati come rappresentativi delle piene di maggiore rilevanza a scala regionale, con le date dell'evento (*start & end*), la data del picco (*tmaxDM*) e il valore dell'indice di magnitudo (*maxDm*) (figura 40) e successivamente i grafici degli andamenti di D_m negli anni degli eventi regionali massimi.


```
> maxeventi
```

	start	end	tmaxDm	maxDm
1	1901-06-08	1901-06-22	1901-06-08	0.06762829
8	1914-07-16	1914-08-13	1914-07-16	0.03184111
11	1920-09-16	1920-09-30	1920-09-17	0.03933363
14	1923-08-09	1923-08-23	1923-08-09	0.03217453
15	1926-08-02	1926-08-16	1926-08-02	0.03322907
18	1927-09-18	1927-10-02	1927-09-18	0.03845562
40	1940-05-23	1940-06-09	1940-06-03	0.06834886
73	1954-06-25	1954-07-17	1954-07-04	0.03845259
88	1959-08-06	1959-08-20	1959-08-20	0.02645097
106	1965-08-26	1965-09-10	1965-08-28	0.02612505
108	1966-08-11	1966-08-26	1966-08-14	0.03347145
188	1999-05-05	1999-05-29	1999-05-16	0.03591847
197	2002-08-05	2002-08-19	2002-08-13	0.02629243
205	2005-08-15	2005-08-30	2005-08-17	0.05121943
225	2013-05-26	2013-06-09	2013-05-29	0.04694963

Figura 40- Visualizzazione della console di RStudio in cui viene mostrato il contenuto del data.frame contenente i 15 eventi massimi.

Nelle seguenti pagine vengono mostrati i grafici specifici per ognuno dei quindici eventi in cui viene mostrato l'andamento dell'indice di magnitudo nell'anno e nel mese dell'evento selezionato. Gli eventi sono identificati e riportati in base all'anno in cui essi sono avvenuti.

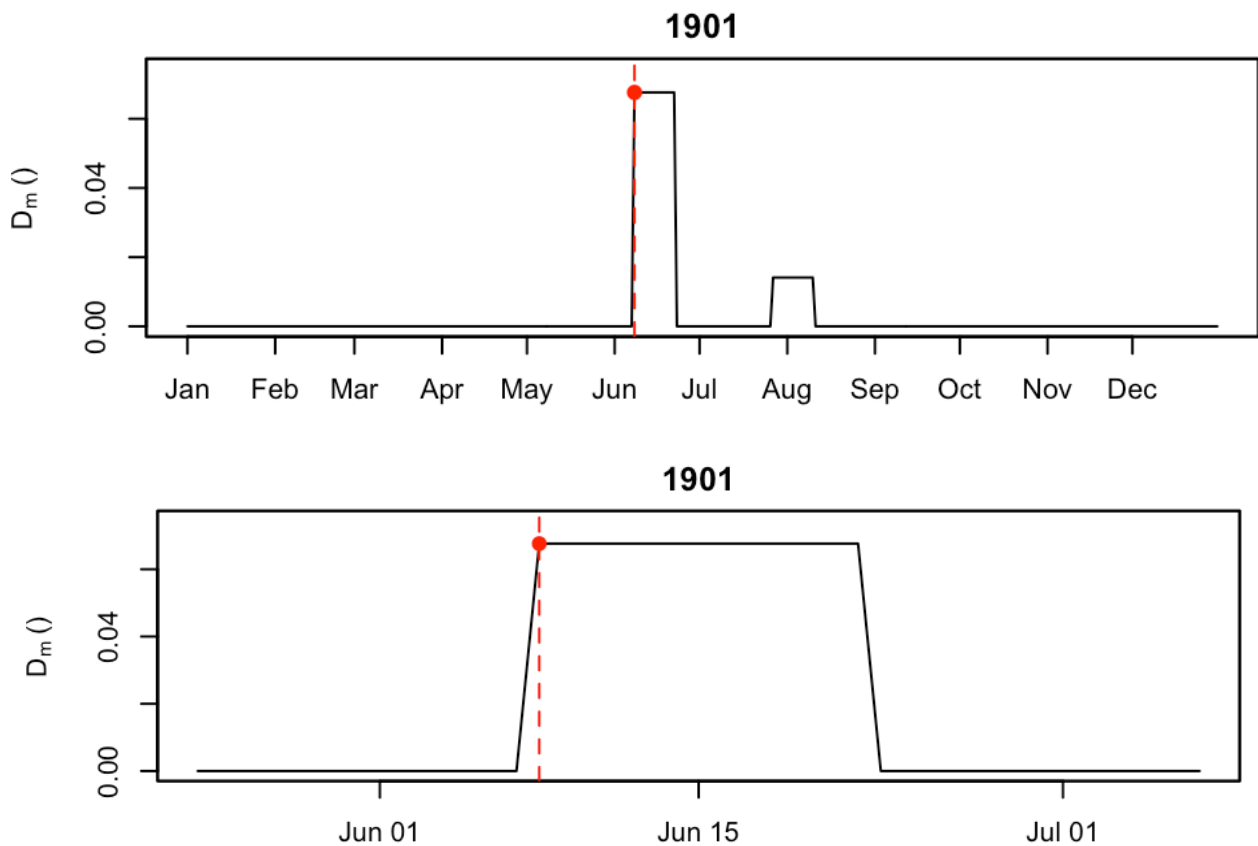


Figura 41- Andamento indice di magnitudo per l'evento del 1901.

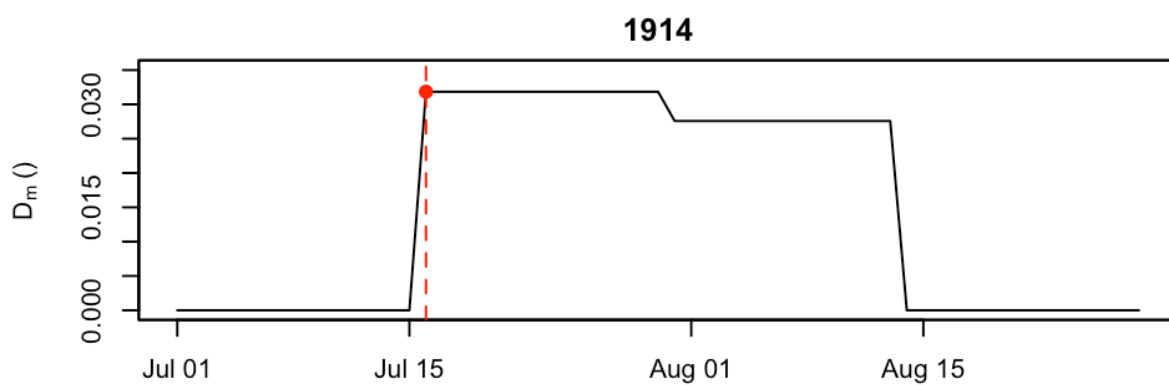
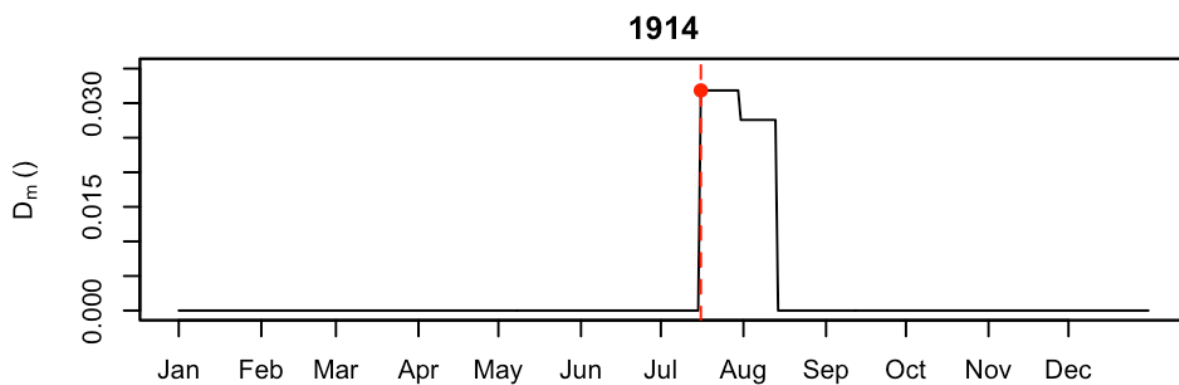


Figura 42- Andamento indice di magnitudo per l'evento del 1914.

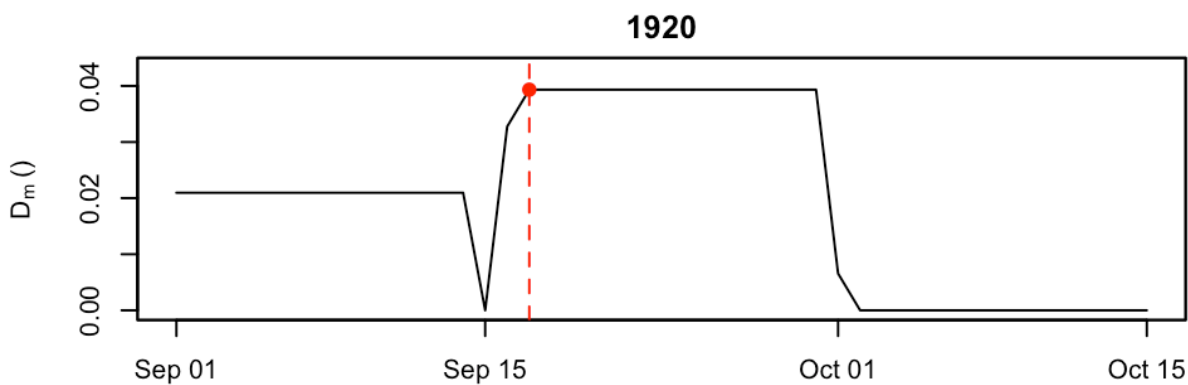
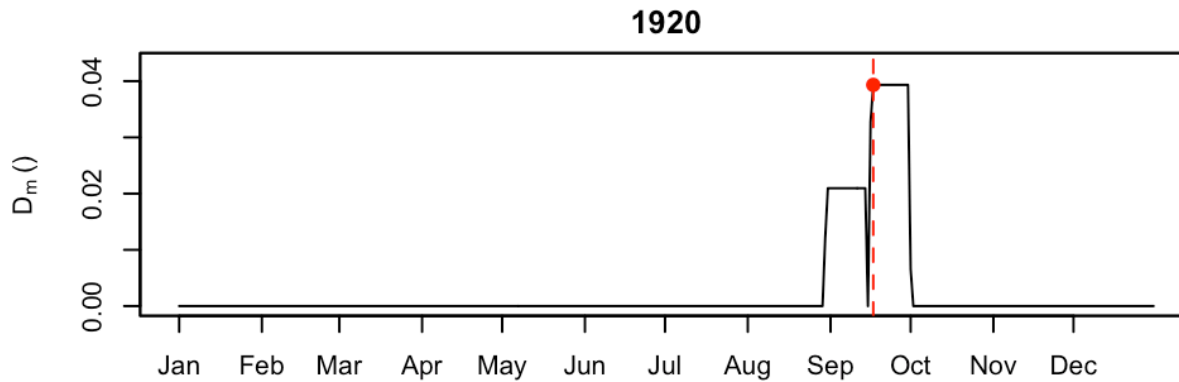


Figura 43- Andamento indice di magnitudo per l'evento del 1920.

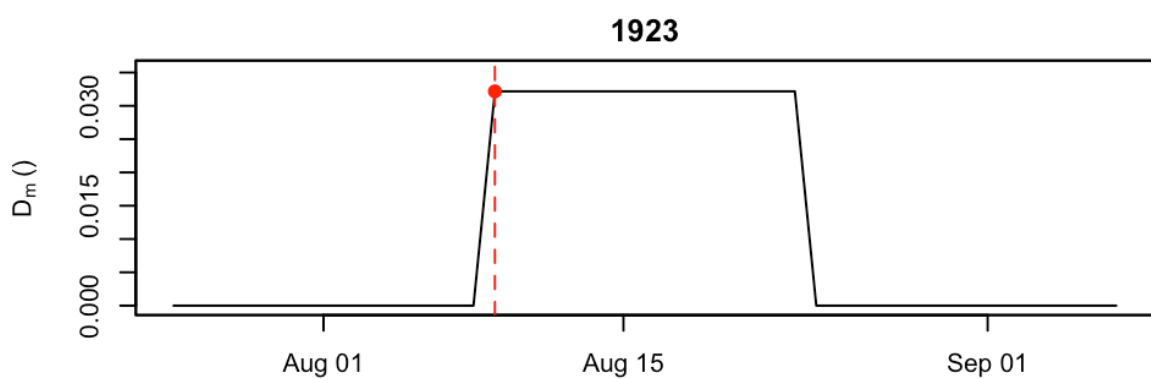
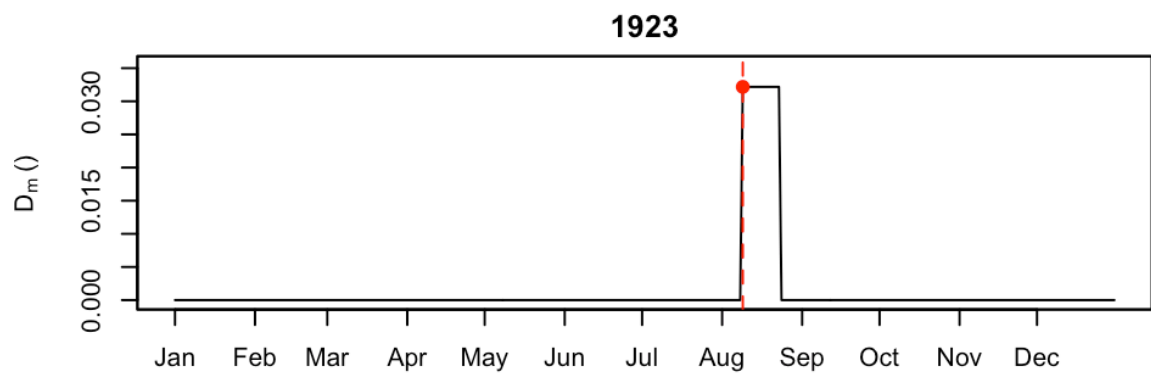


Figura 44- Andamento indice di magnitudo per l'evento del 1923.

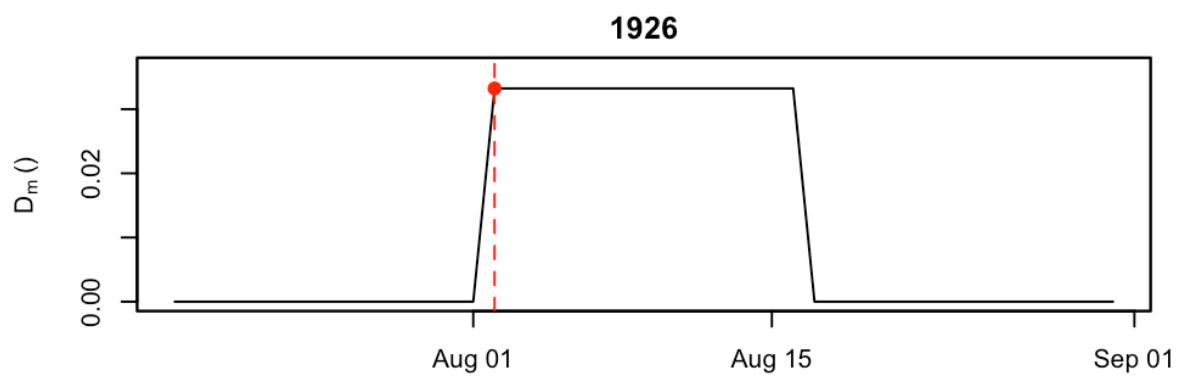
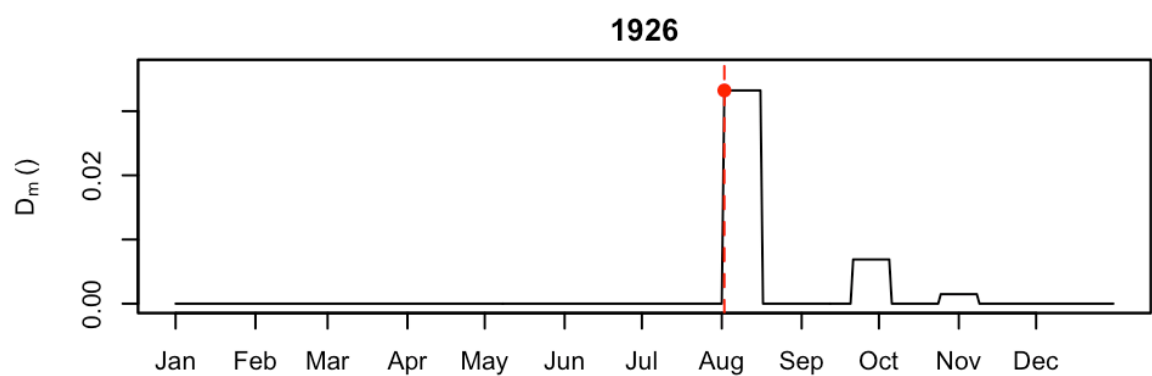


Figura 45- Andamento indice di magnitudo per l'evento del 1926.

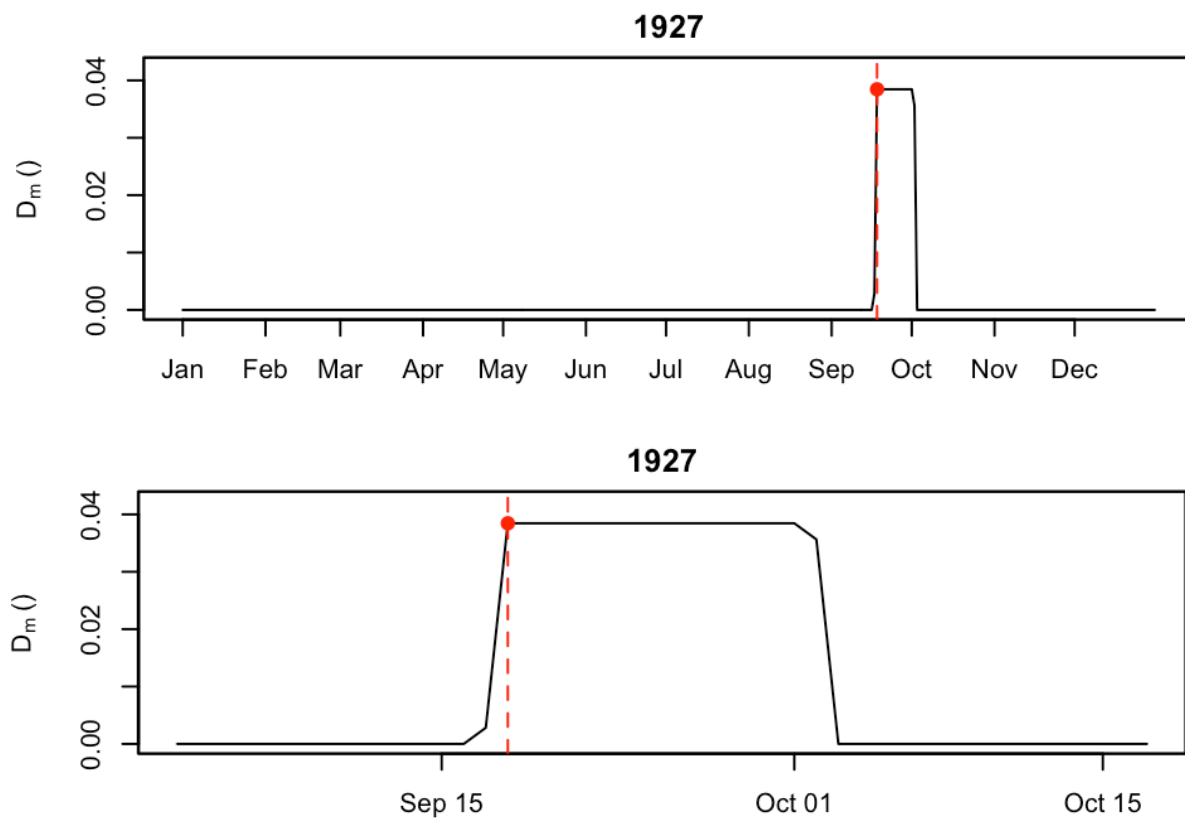


Figura 46- Andamento indice di magnitudo per l'evento del 1927.

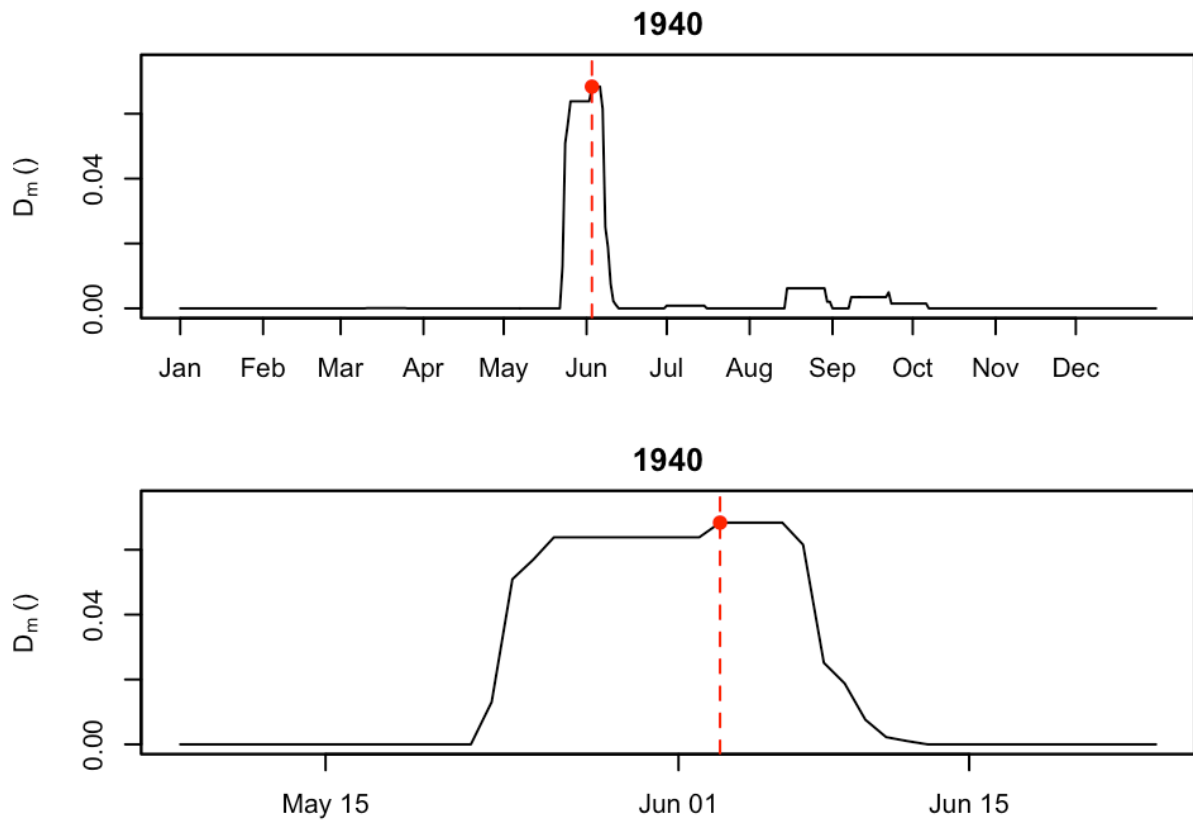


Figura 47- Andamento indice di magnitudo per l'evento del 1940.

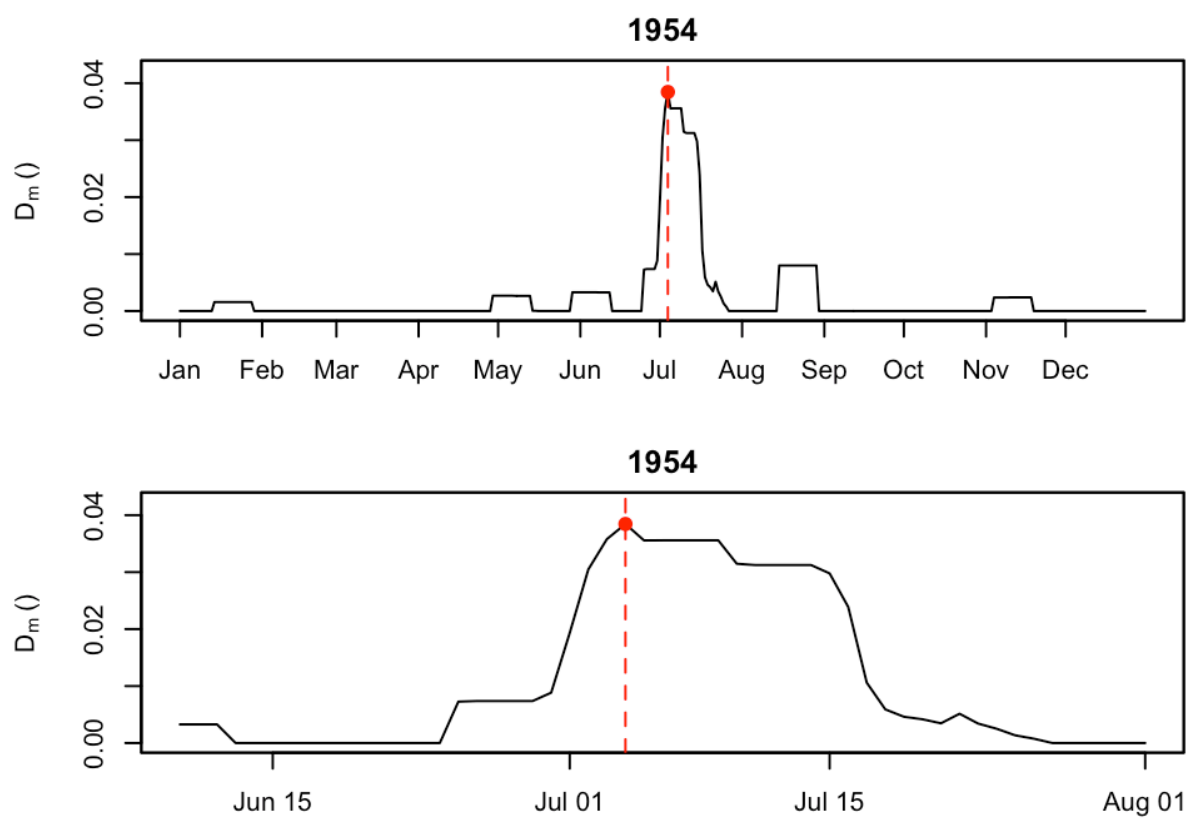


Figura 48- Andamento indice di magnitudo per l'evento del 1954.

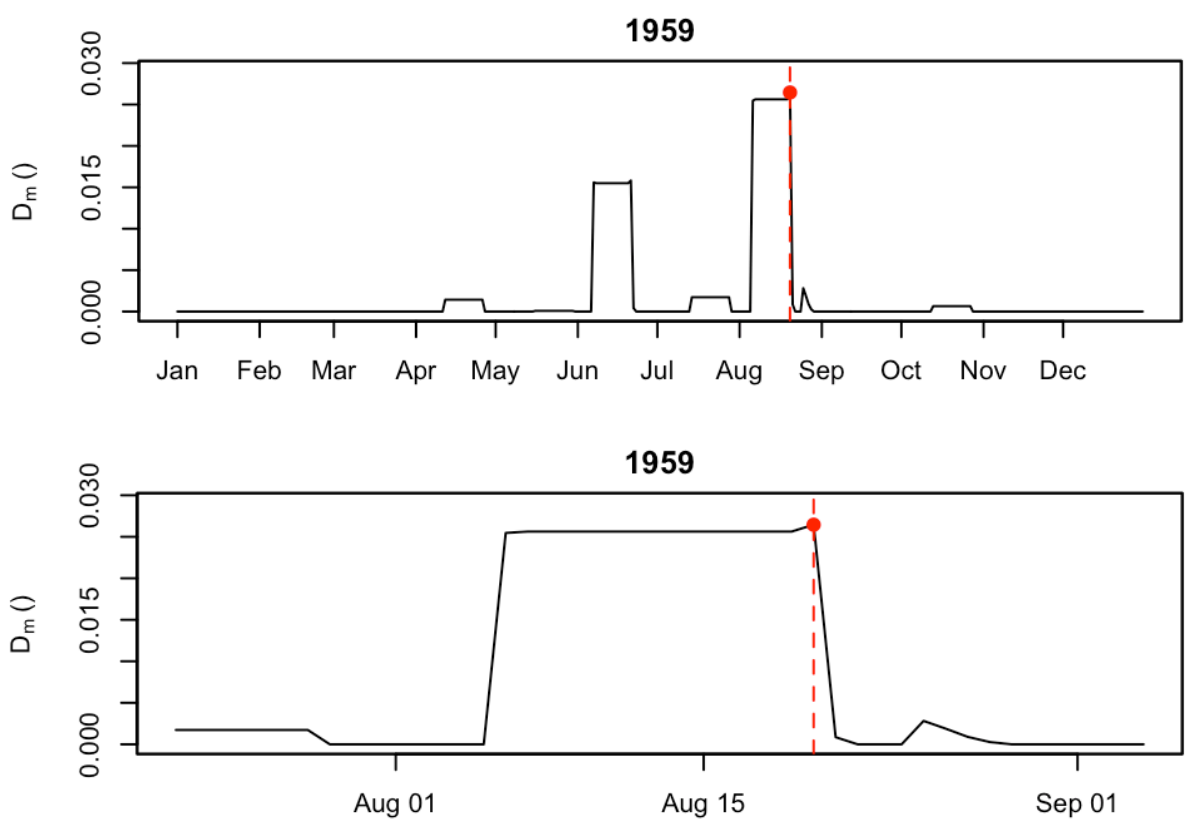


Figura 49- Andamento indice di magnitudo per l'evento del 1959.

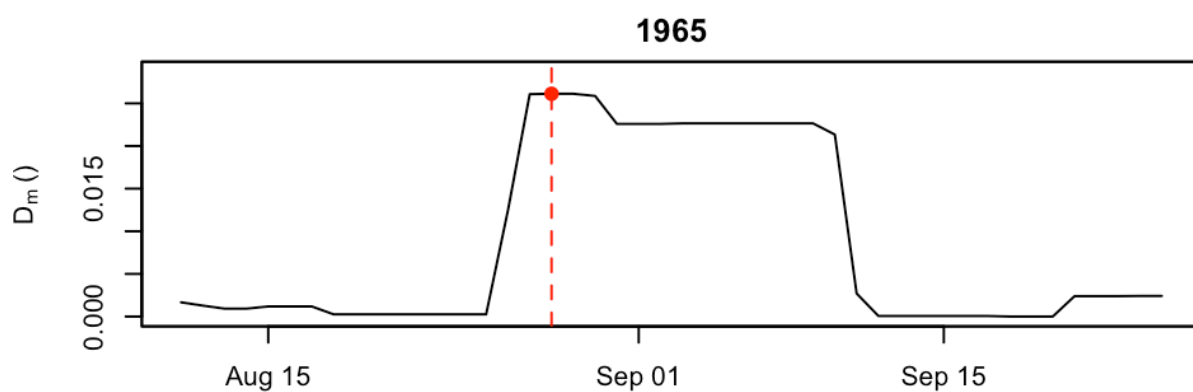
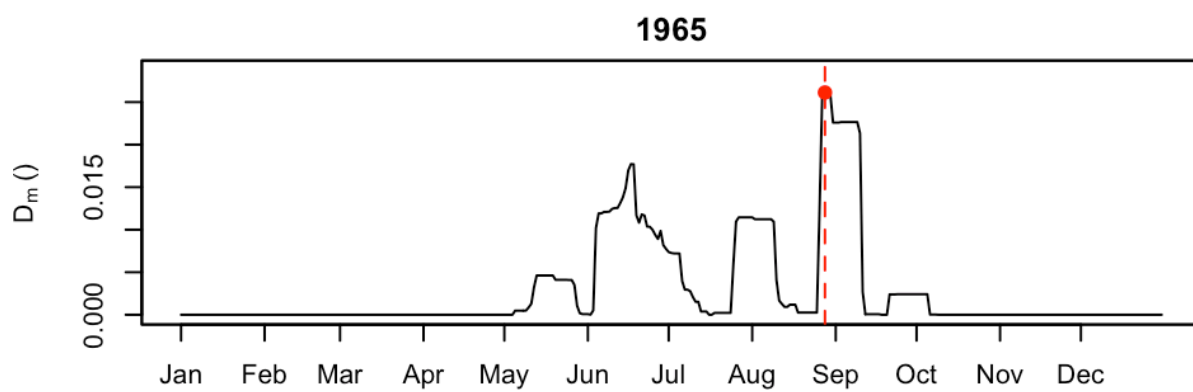


Figura 50- Andamento indice di magnitudo per l'evento del 1965.

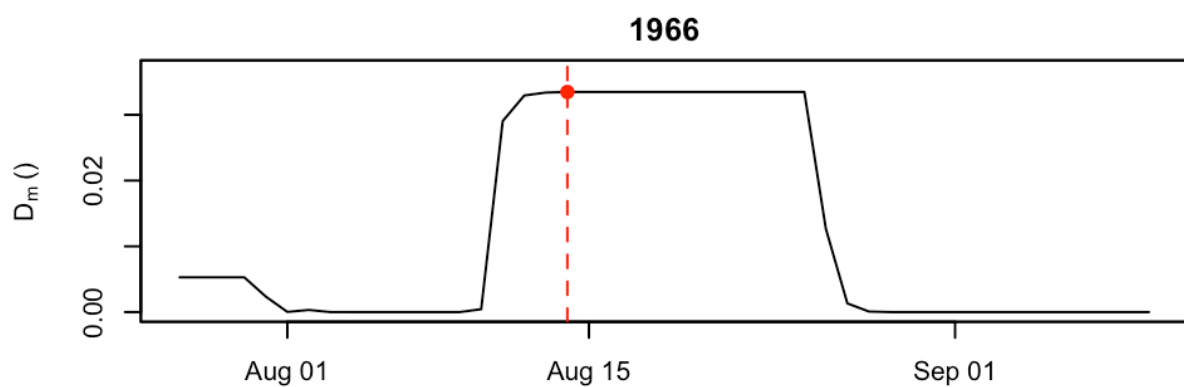
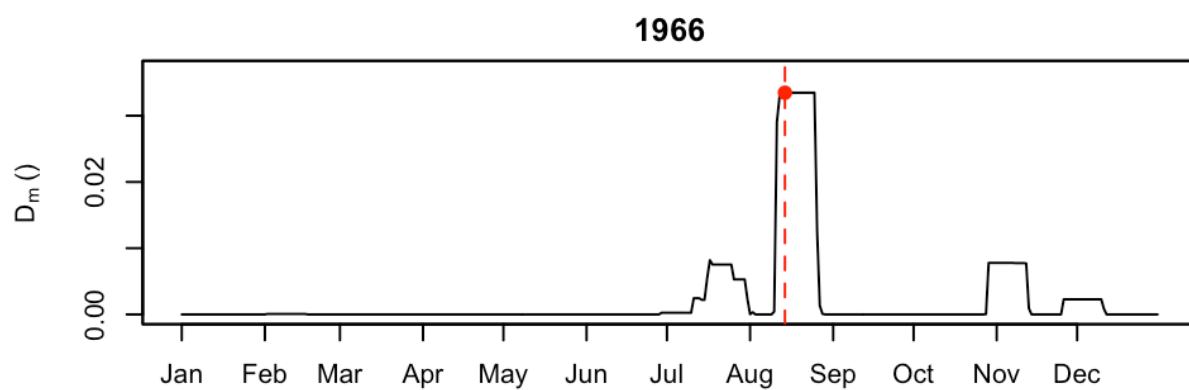


Figura 51- Andamento indice di magnitudo per l'evento del 1966.

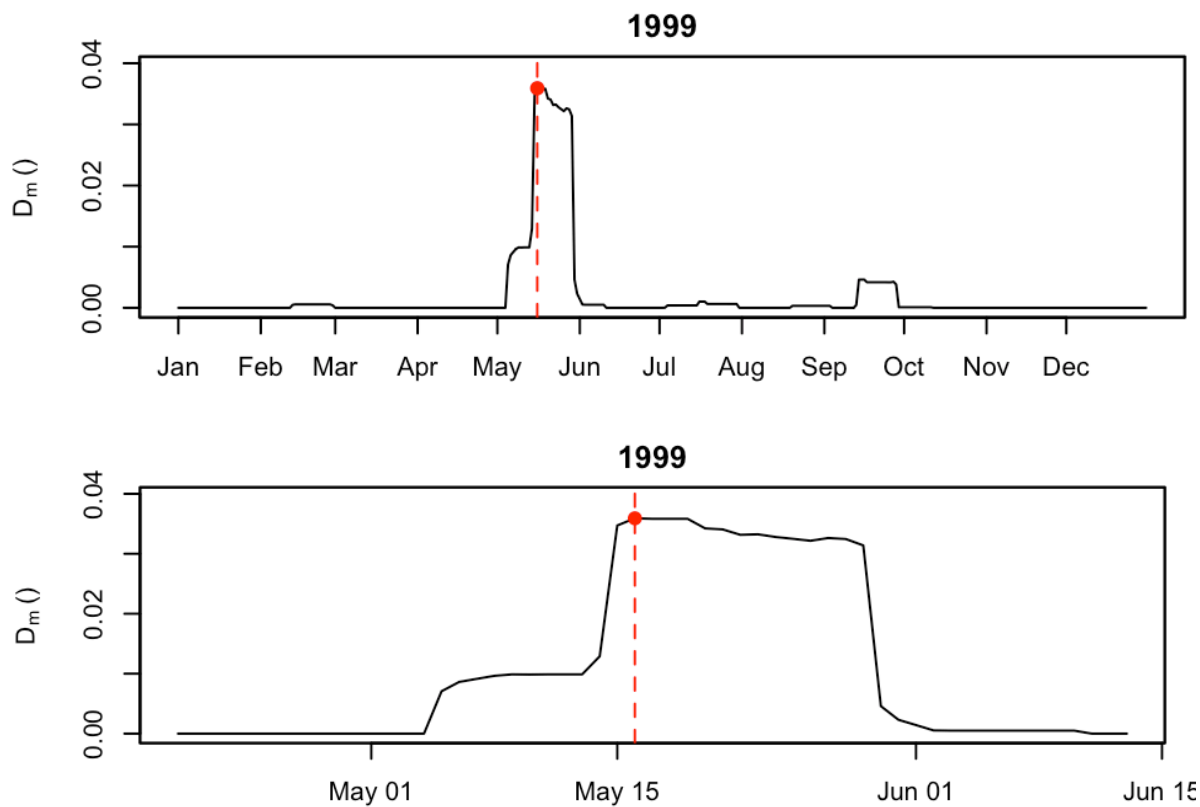


Figura 52- Andamento indice di magnitudo per l'evento del 1999.

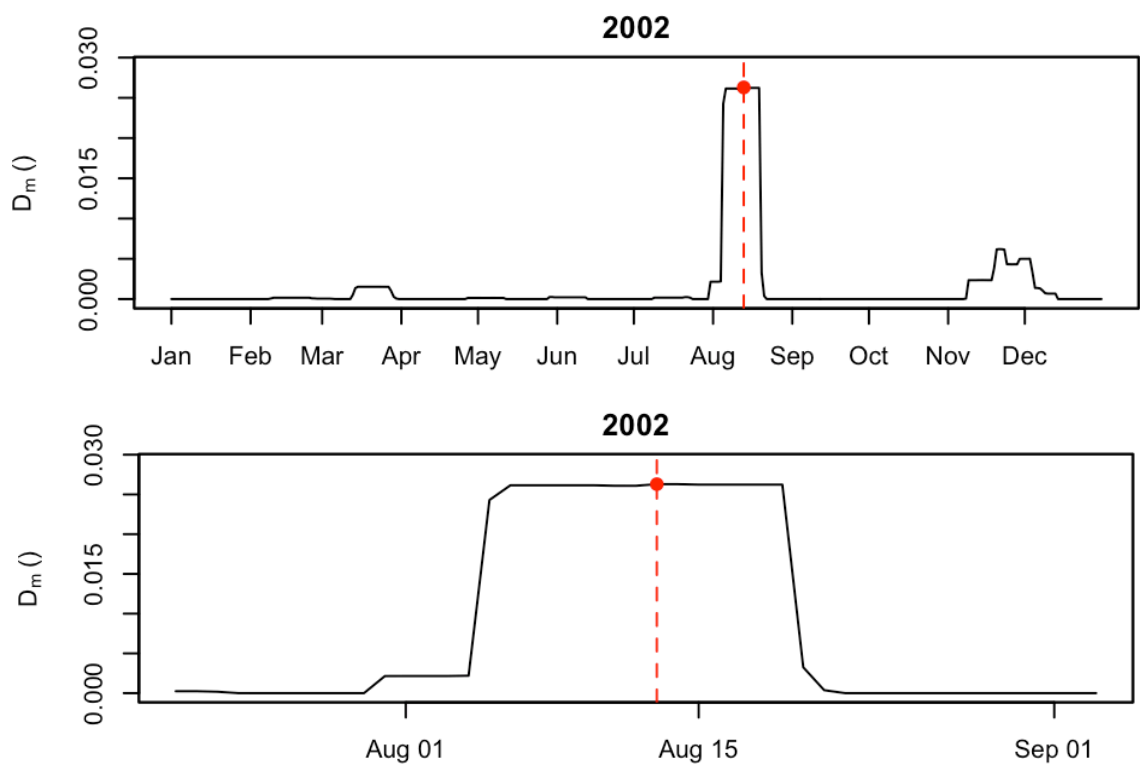


Figura 53- Andamento indice di magnitudo per l'evento del 2002.

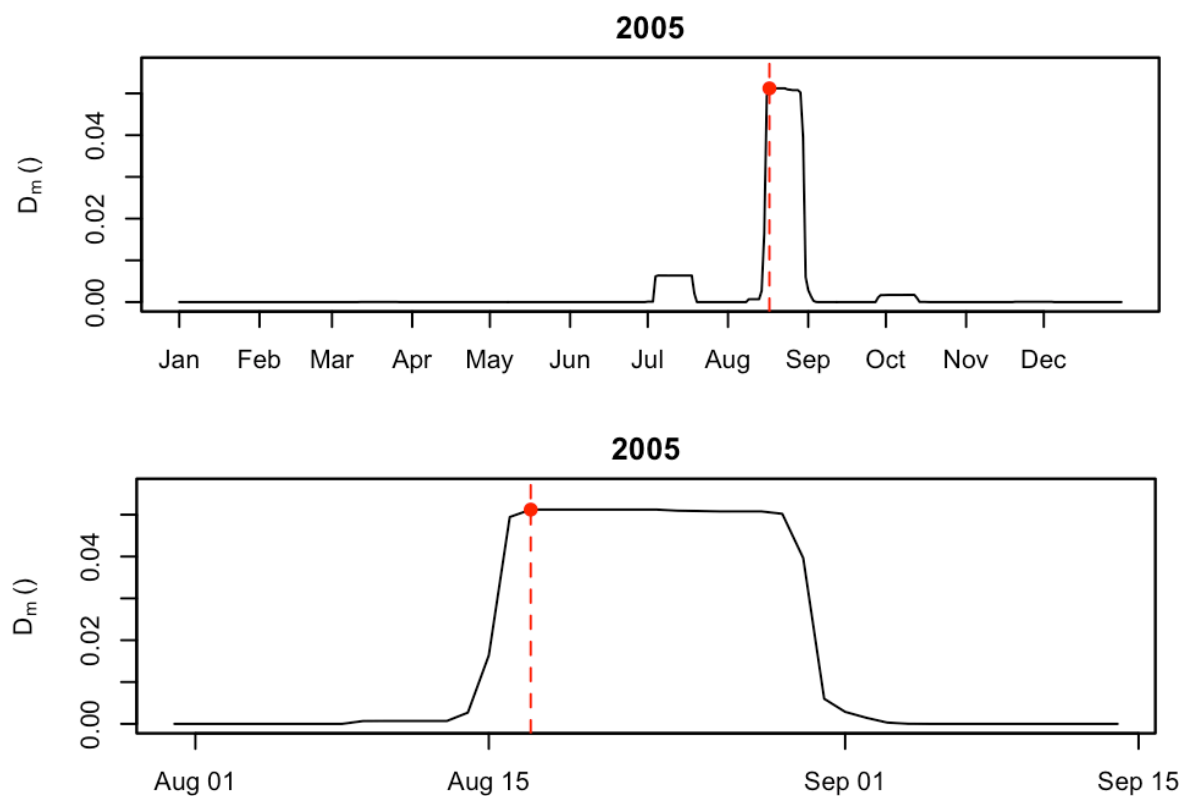


Figura 54- Andamento indice di magnitudo per l'evento del 2005.

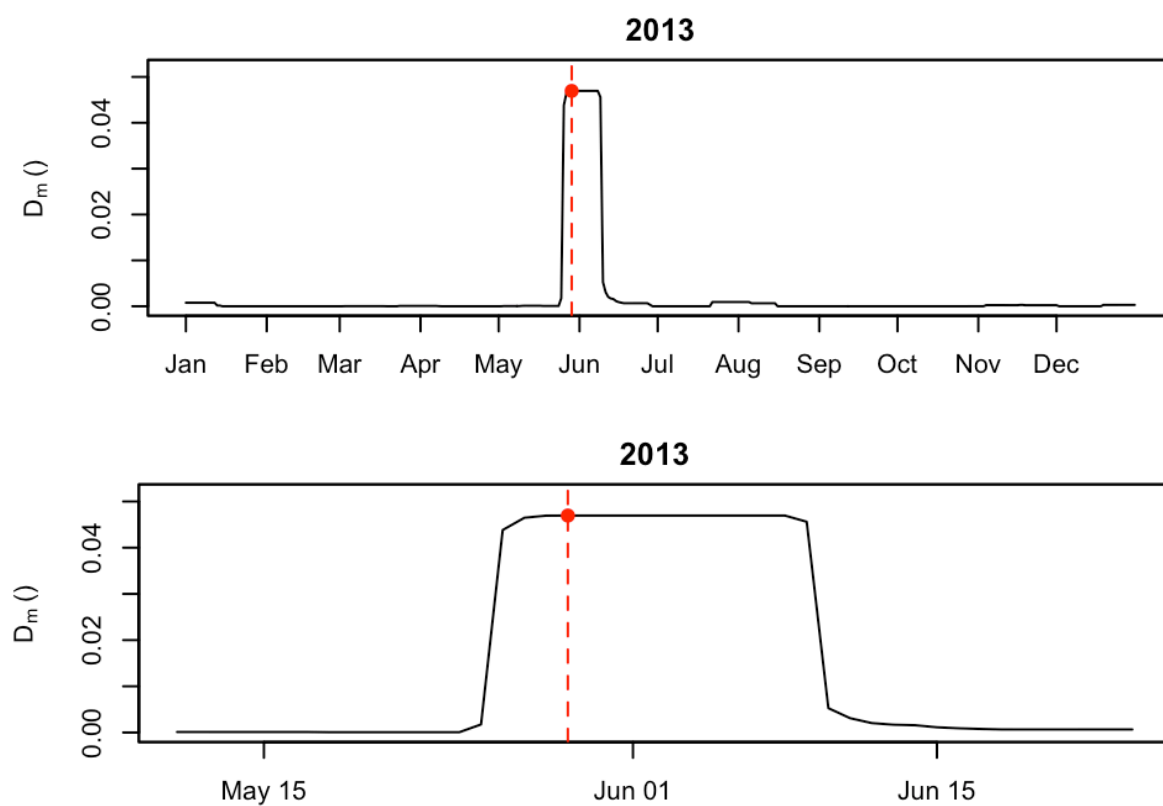


Figura 55- Andamento indice di magnitudo per l'evento del 2013.

Distribuzione stagionale degli eventi estremi (maxeventi)

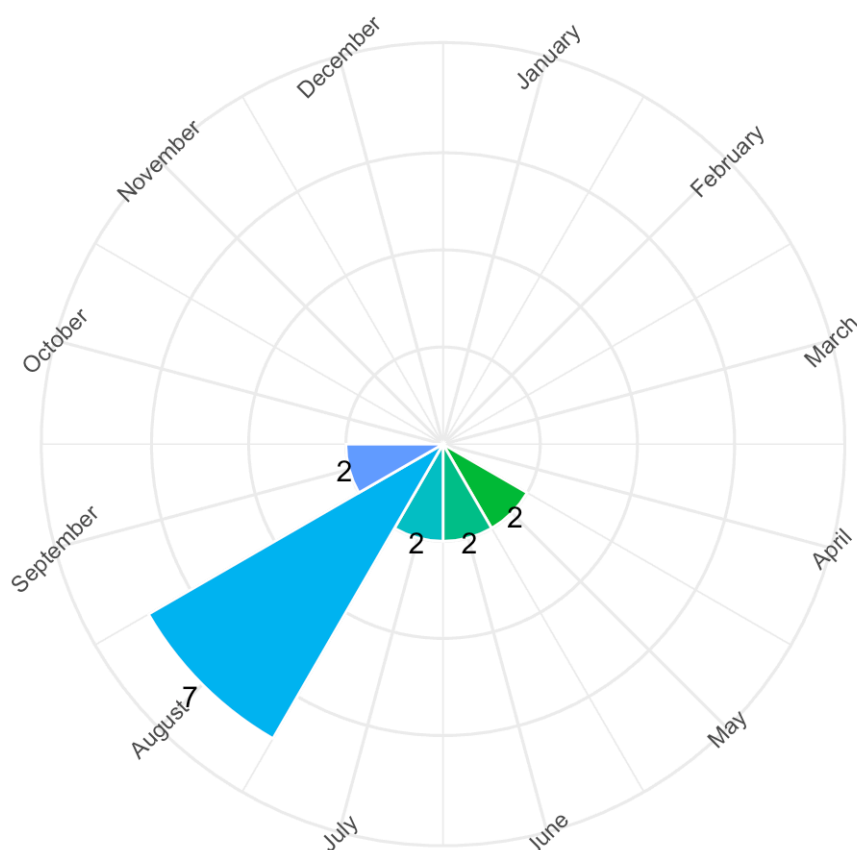


Figura 56- Distribuzione stagionale dei 15 eventi estremi mostrati in precedenza.

Questo andamento stagionale suggerisce che gli eventi di piena con i valori massimi di danno medio regionale (D_m) sono legati alla stagionalità delle condizioni idrologiche e meteorologiche della GAR. In particolare, primavera inoltrata ed estate, ovvero il periodo tra maggio e agosto mostrato in figura 56, corrispondono al periodo in cui possono coesistere tre condizioni:

- precipitazioni intense convettive come temporali estivi, spesso molto localizzate ma con elevata intensità;
- fenomeni di fusione nivale residua in bacini montani ad alta quota (nelle prime settimane di maggio e giugno), che possono aumentare la base di deflusso;
- suoli già saturi a seguito delle piogge primaverili, che riducono la capacità di assorbimento e amplificano il deflusso superficiale.

L'assenza di eventi selezionati nei mesi autunnali e invernali suggerisce che, nel contesto della GAR, le piene di questa stagione, pur rilevanti in altre regioni, non hanno generato valori di D_m comparabili a quelli estivi. Ciò è probabilmente legato al fatto che le precipitazioni in questo periodo risultano generalmente più distribuite nel tempo, che in molte aree assumono carattere nevoso, e che le condizioni idrologiche dei bacini (maggiore capacità di ritenzione, assenza della fusione nivale) non favoriscono la formazione di piene estreme su scala regionale.

3.3 Visualizzazione eventi regionali da picchi della serie storica dei D_m .

Per la rappresentazione spaziale degli eventi selezionati, è stato sviluppato uno script che integra i risultati dell'analisi idrologica con strumenti di visualizzazione interattiva.

L'obiettivo è quello di costruire mappe dinamiche in cui, per ciascun evento regionale individuato tramite la serie storica di D_m , vengano riportate le principali informazioni idrologiche e di impatto a livello di singolo bacino/stazione. Per ciascun evento, inoltre, vengono estratti e organizzati i valori massimi registrati nei siti di monitoraggio durante il periodo di durata dell'evento:

- la portata massima (Q_{max});
- il tempo di ritorno associato (RP);
- il valore del danno normalizzato.

Questi tre insiemi di informazioni vengono combinati in un unico data frame, che associa a ciascuna stazione sia le grandezze idrologiche che quelle legate al danno (*df_eventi*).

Successivamente, per la rappresentazione cartografica, vengono letti gli shapefile contenenti la rete idrografica (grazie alla gentile concessione dei file di Anna Basso) e le coordinate delle stazioni, che vengono riportate in un sistema di riferimento geografico coerente (WGS84).

Per ogni evento selezionato, viene generata una mappa interattiva con la libreria Leaflet, nella quale: i fiumi vengono rappresentati come linee blu semitrasparenti; le stazioni di misura vengono mostrate come cerchi colorati, la cui tonalità varia in funzione del tempo di ritorno (RP), secondo una scala cromatica dal giallo al rosso; al clic su ciascun punto, un popup riporta in modo sintetico le informazioni associate: il valore di Q_{max} ; il tempo di ritorno stimato; il danno corrispondente.

Nel caso in cui la stazione selezionata in quell'anno non fosse attiva il popup riporta la dicitura "Dato non disponibile" e il pallino raffigurante la stazione rimane trasparente con contorno grigio.

Per facilitare l'interpretazione, in alto a destra della mappa è presente un riquadro informativo che riporta il valore massimo di D_m relativo all'evento e la sua data di occorrenza, mentre in basso a destra viene aggiunta una legenda con la scala cromatica dei tempi di ritorno.

Infine, ogni mappa viene salvata in formato HTML interattivo, così da poter essere esplorata dinamicamente anche al di fuori dell'ambiente R. Per questioni pratiche di seguito vengono riportate le schermate dei file HTML creati, una per evento, con selezionate le stazioni con le condizioni più gravose (figure da 57 a 71).

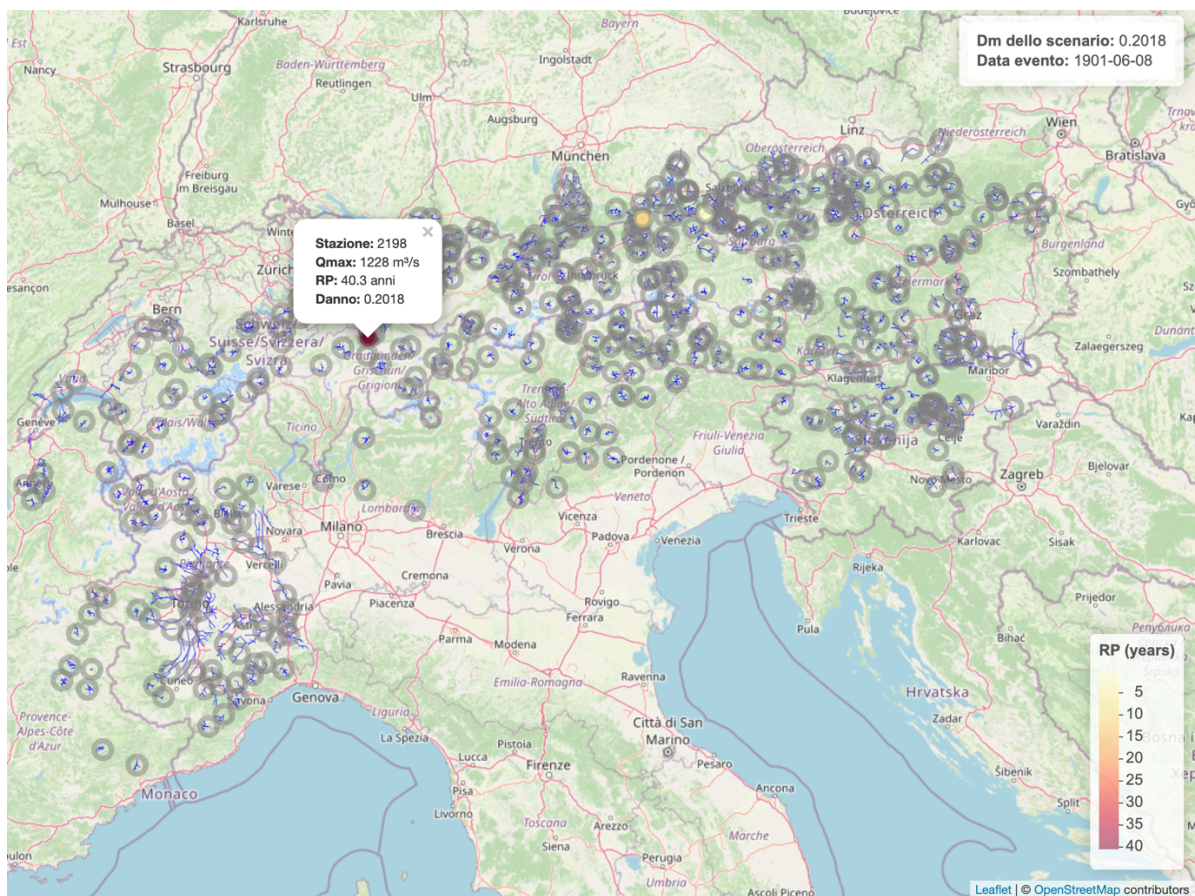


Figura 57- Visualizzazione dell'evento del 1901.

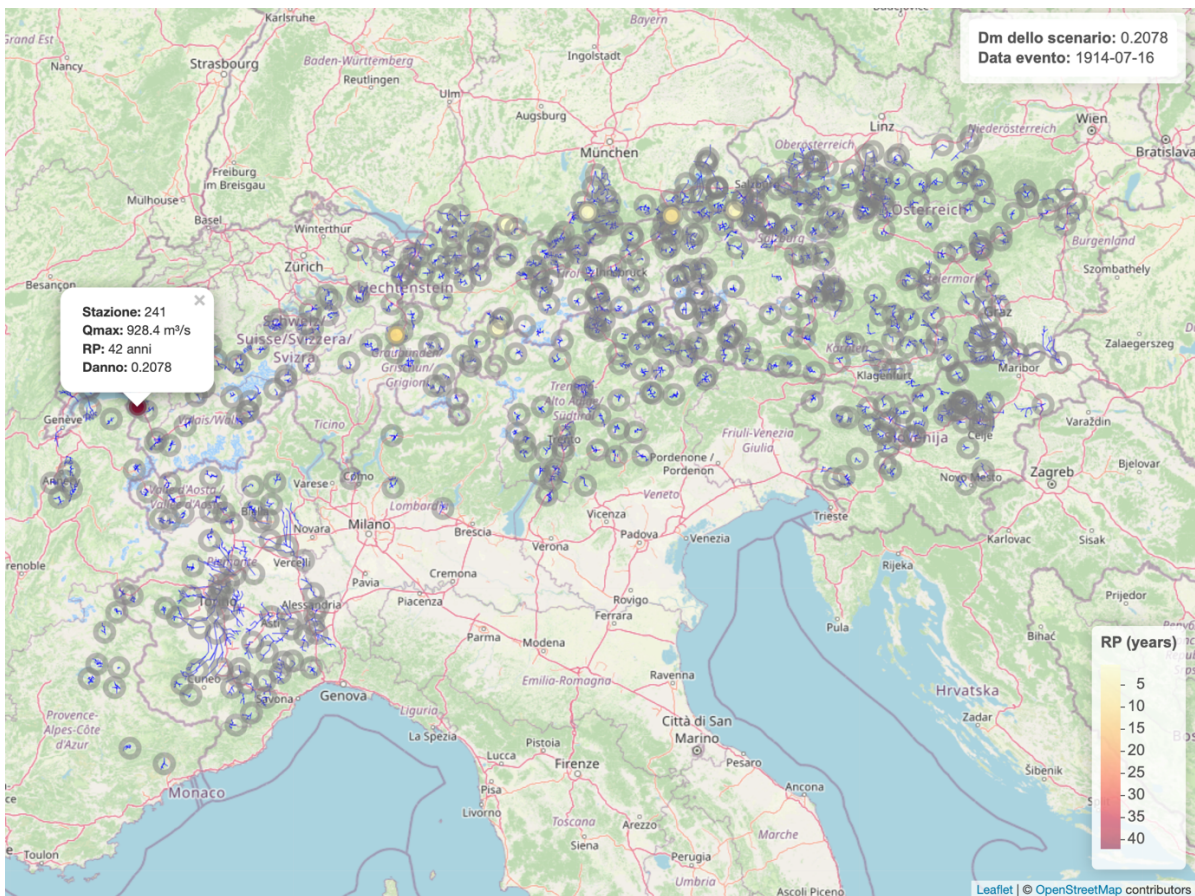


Figura 58- Visualizzazione dell'evento del 1914.

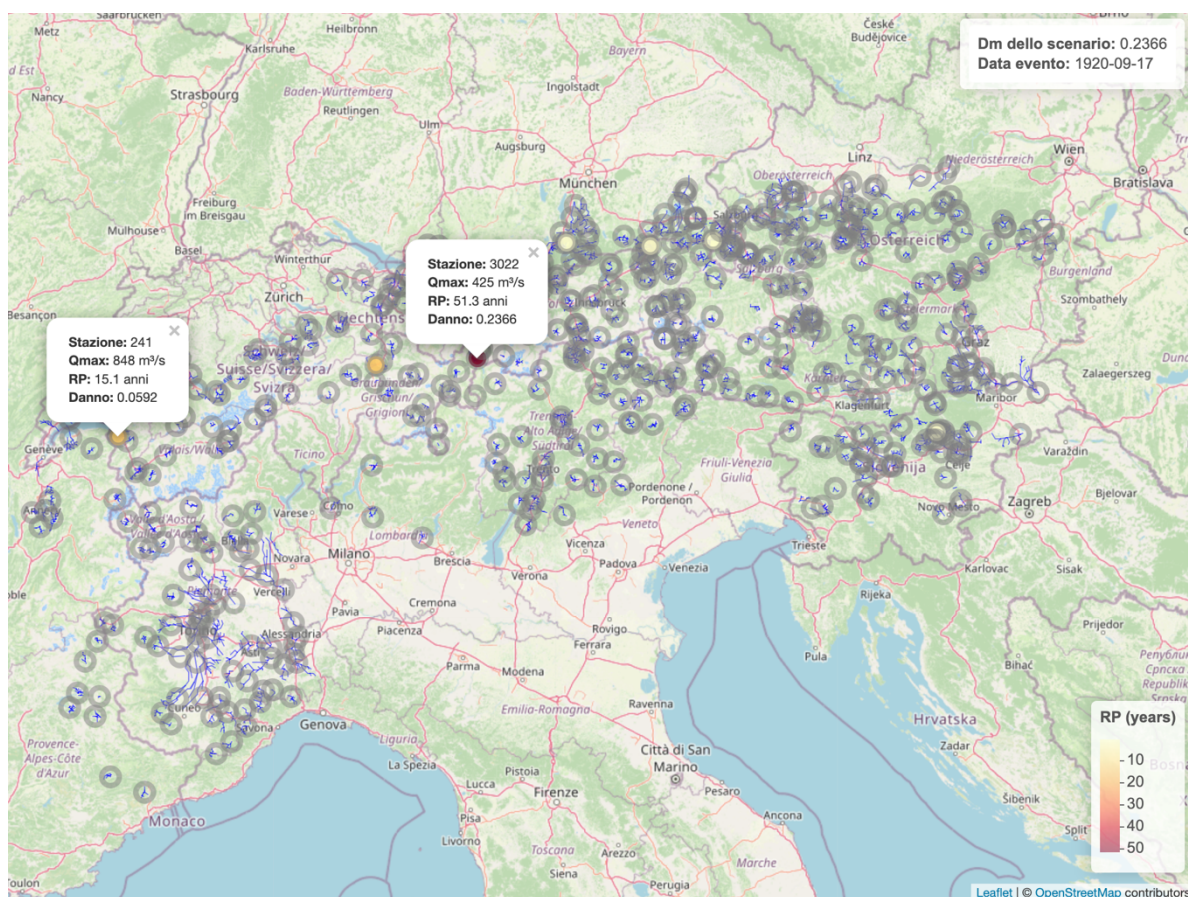


Figura 59- Visualizzazione dell'evento del 1920.

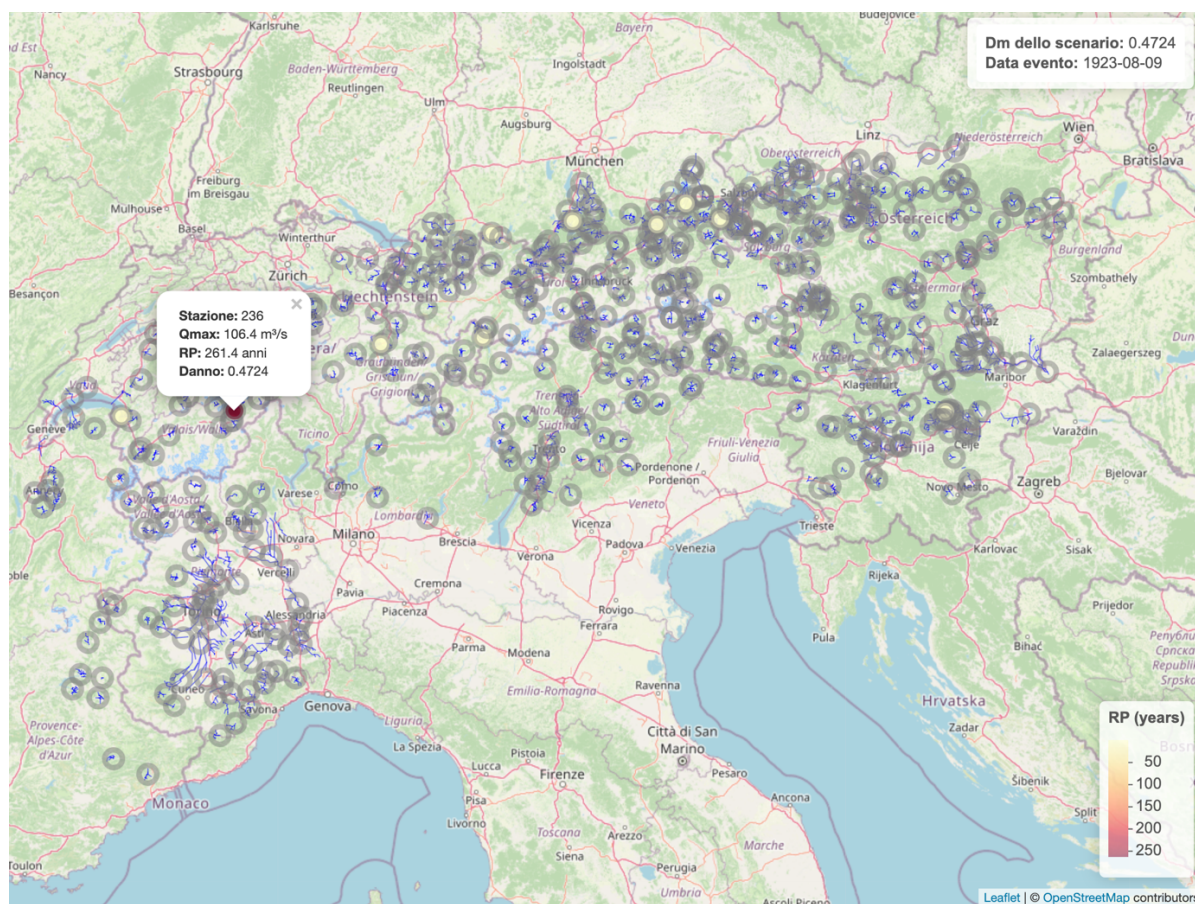


Figura 60- Visualizzazione dell'evento del 1923.

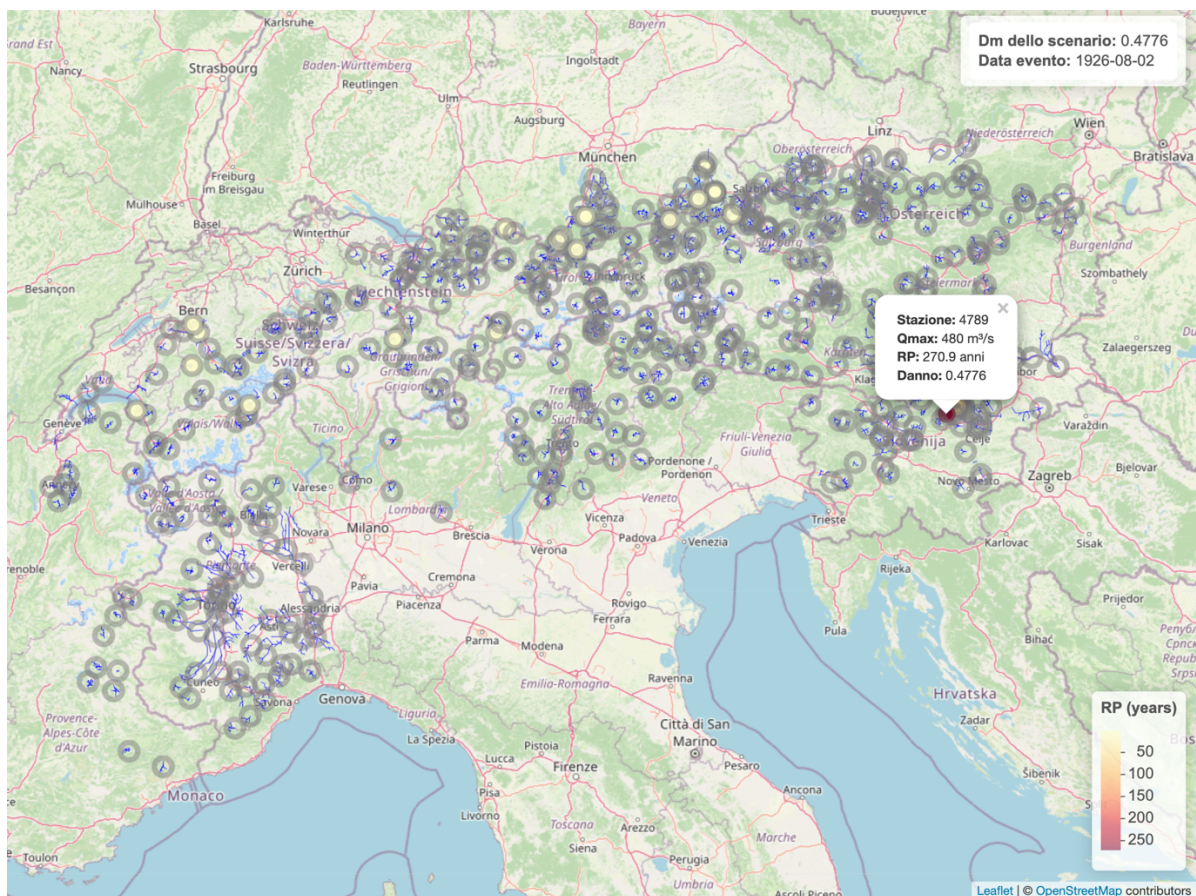


Figura 61- Visualizzazione dell'evento del 1926.

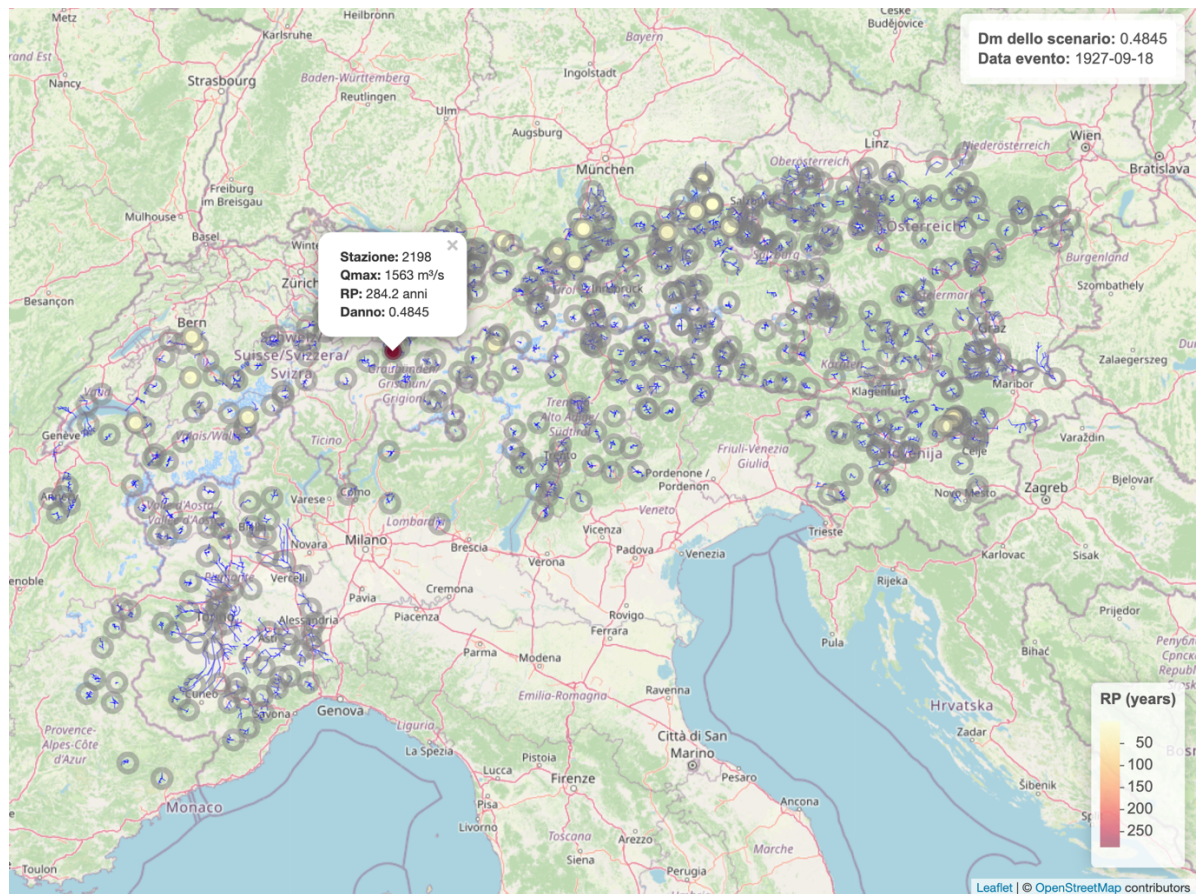


Figura 62- Visualizzazione dell'evento del 1927.

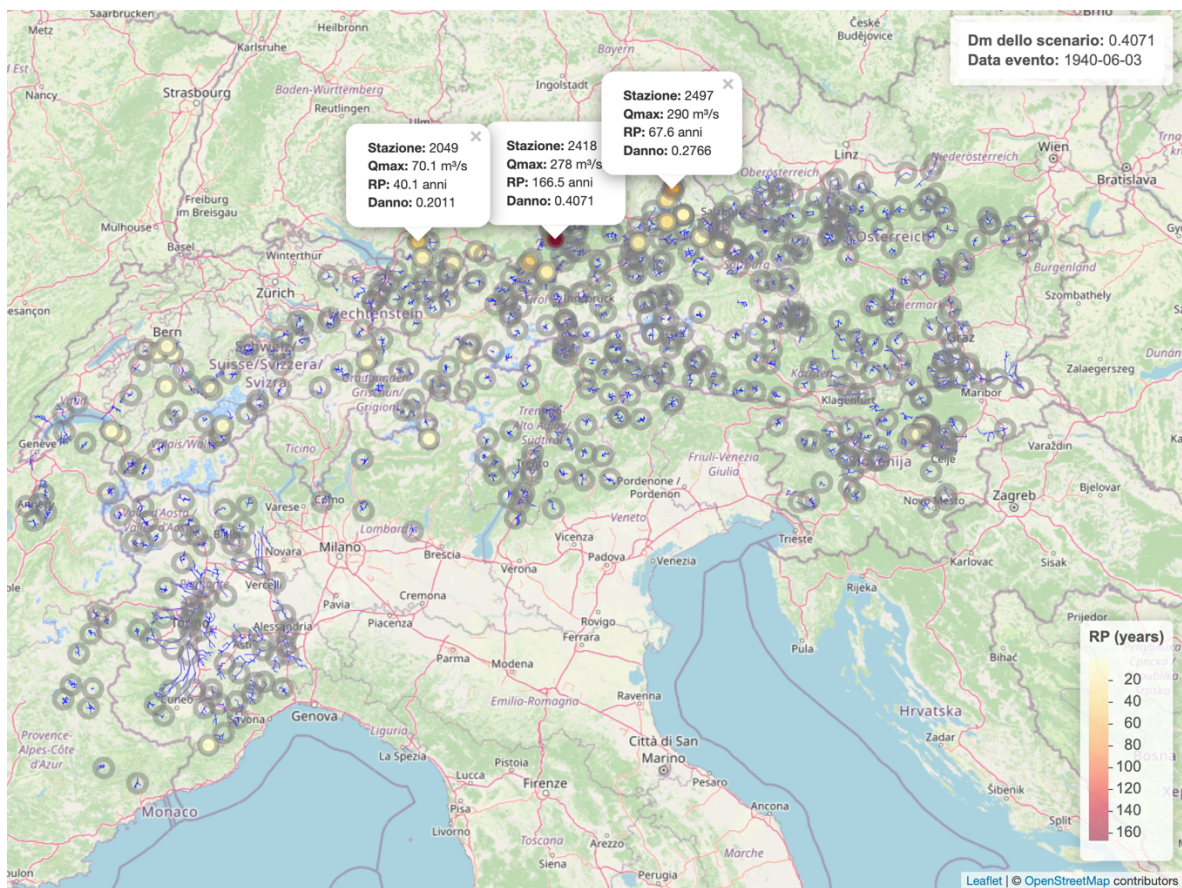


Figura 63- Visualizzazione dell'evento del 1940.

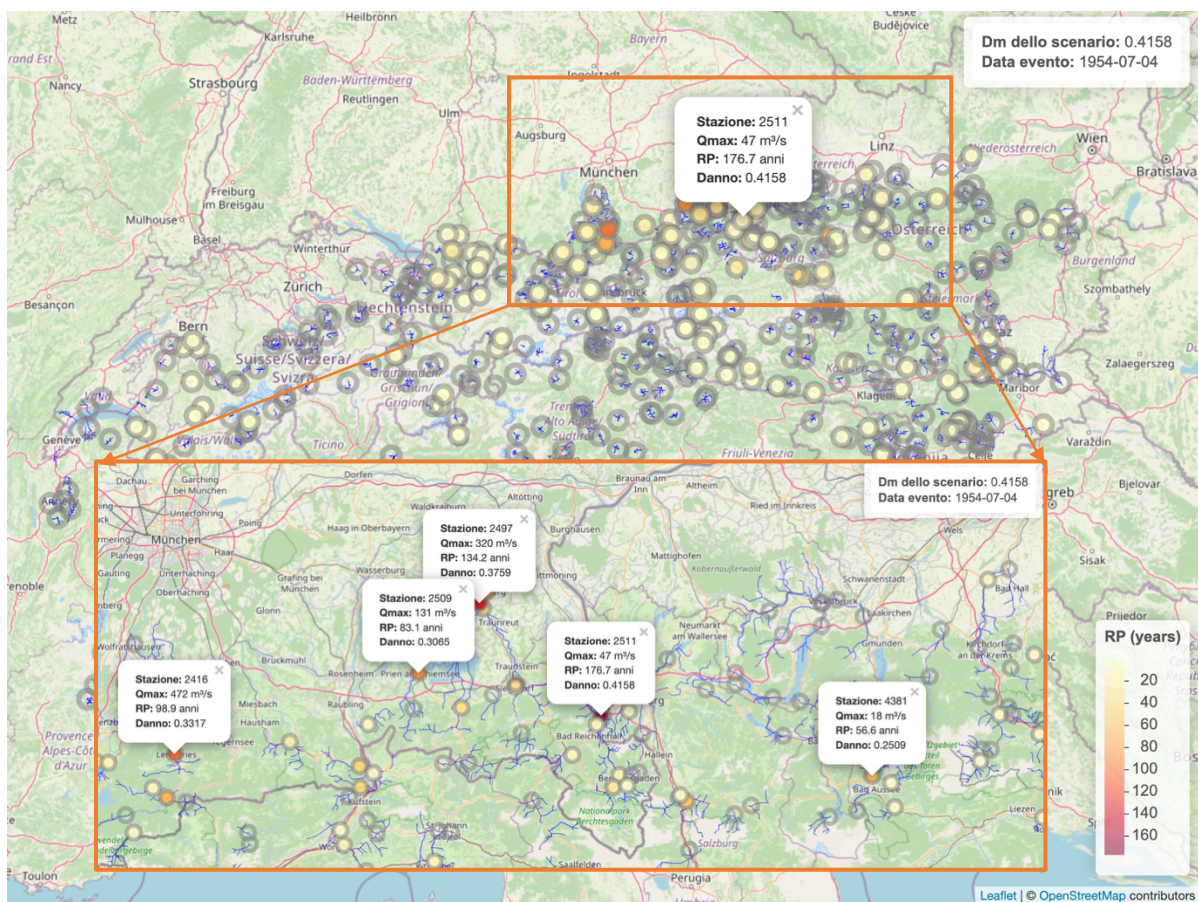


Figura 64- Visualizzazione dell'evento del 1954, con zoom sull'area di Salisburgo.

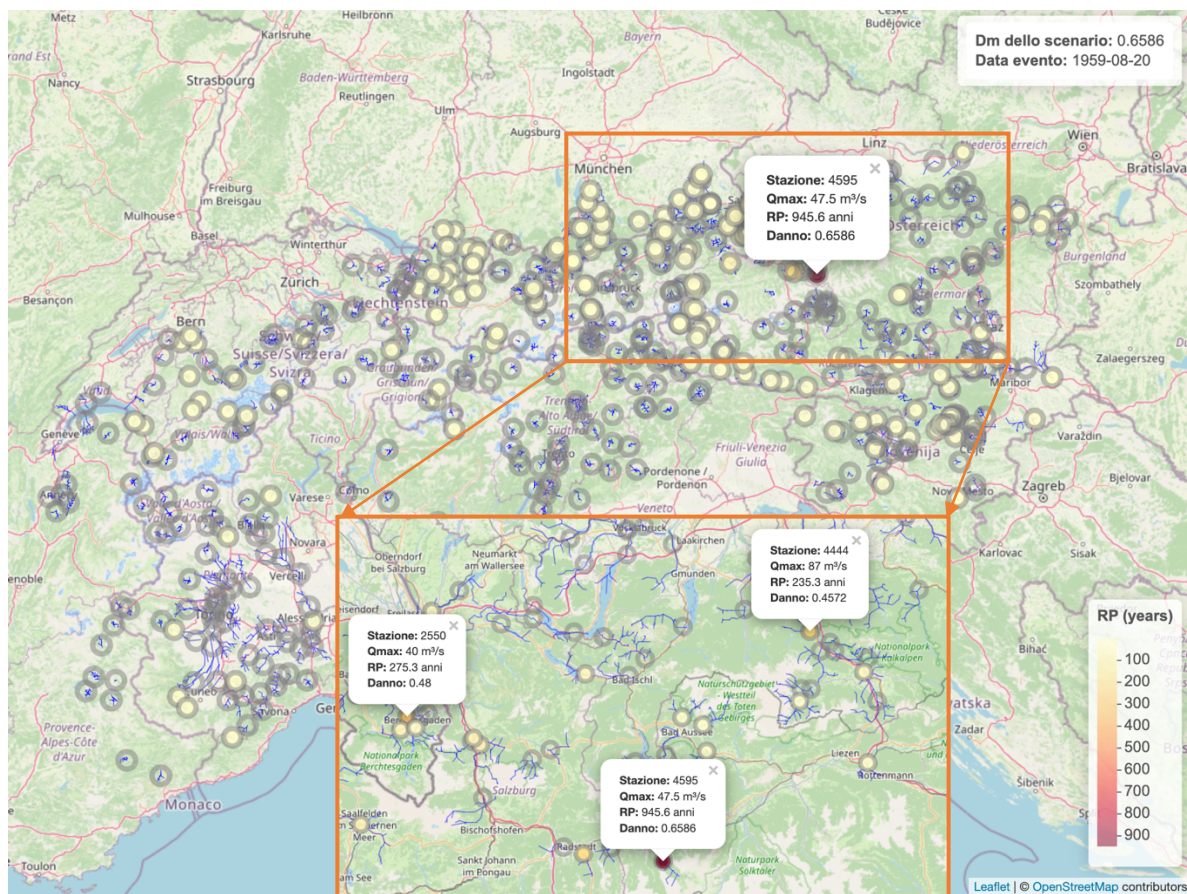


Figura 65- Visualizzazione dell'evento del 1959, con zoom sull'area di Salisburgo.

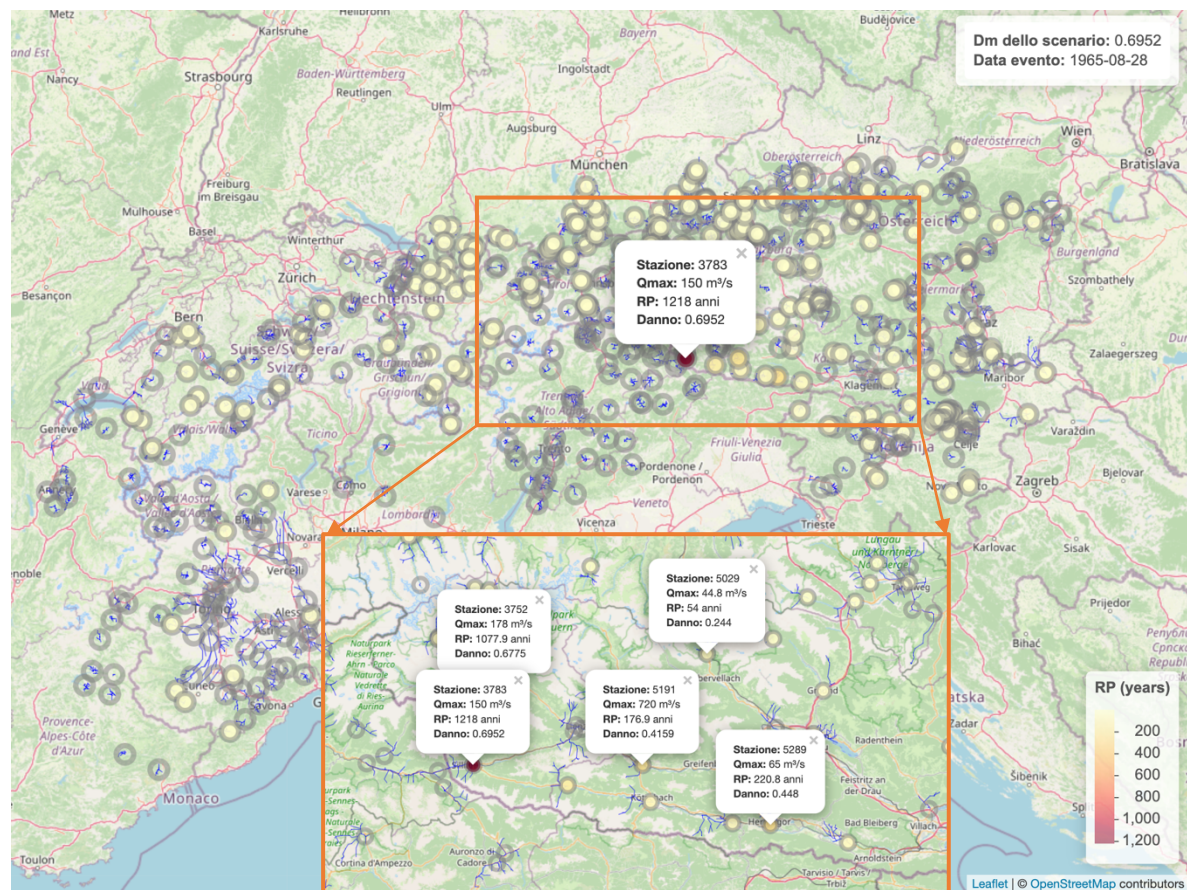


Figura 66- Visualizzazione dell'evento del 1965.

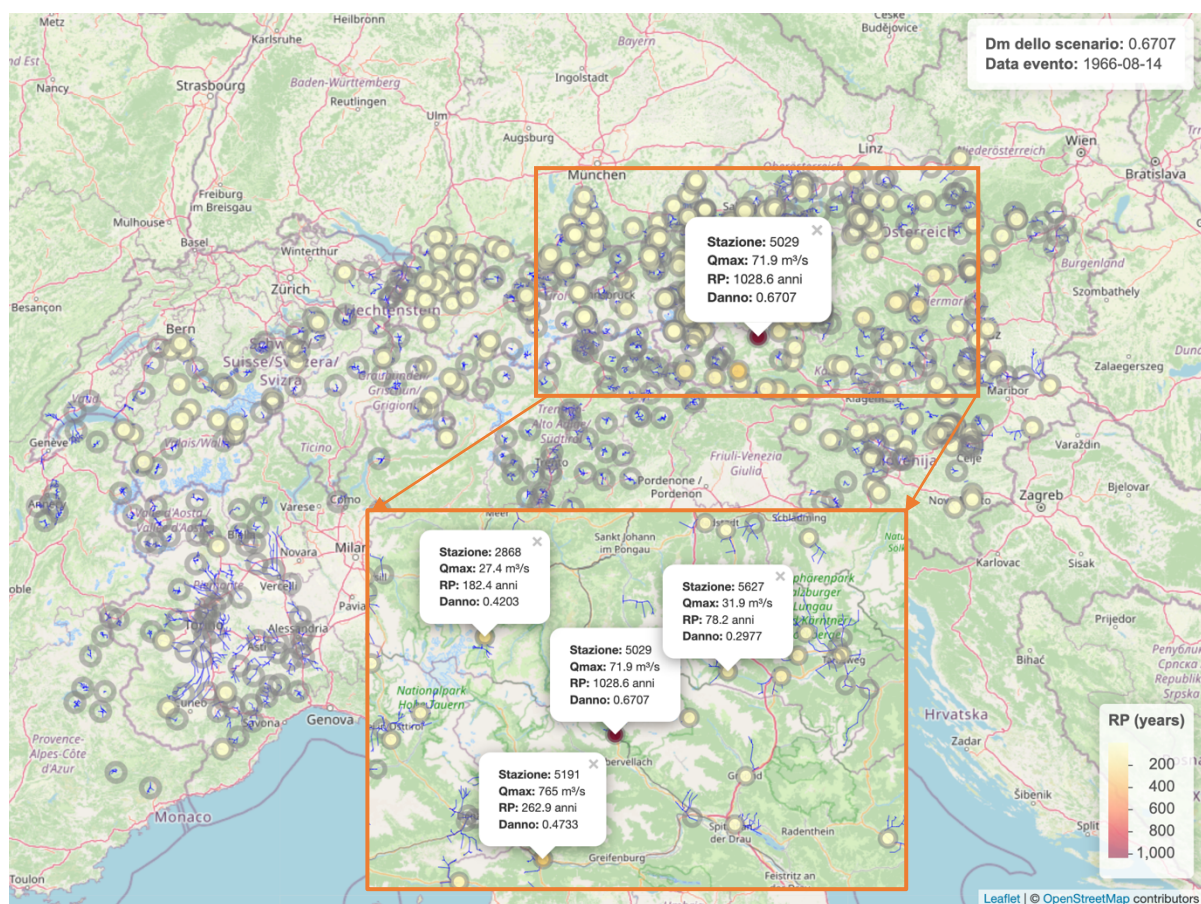


Figura 67- Visualizzazione dell'evento del 1966, con zoom sull'area di Obervellach.

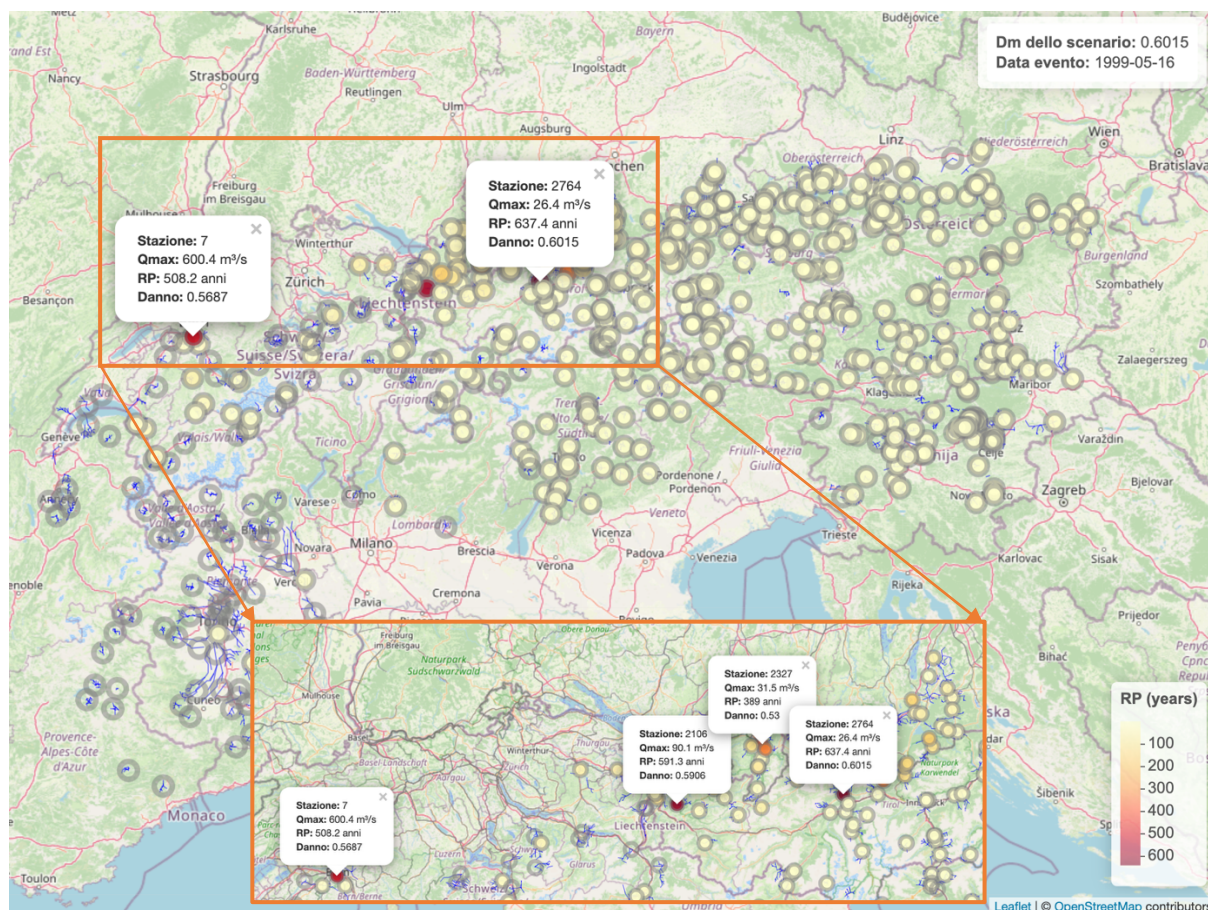


Figura 68- Visualizzazione dell'evento del 1999, con zoom sull'area di Berna e Innsbruck.

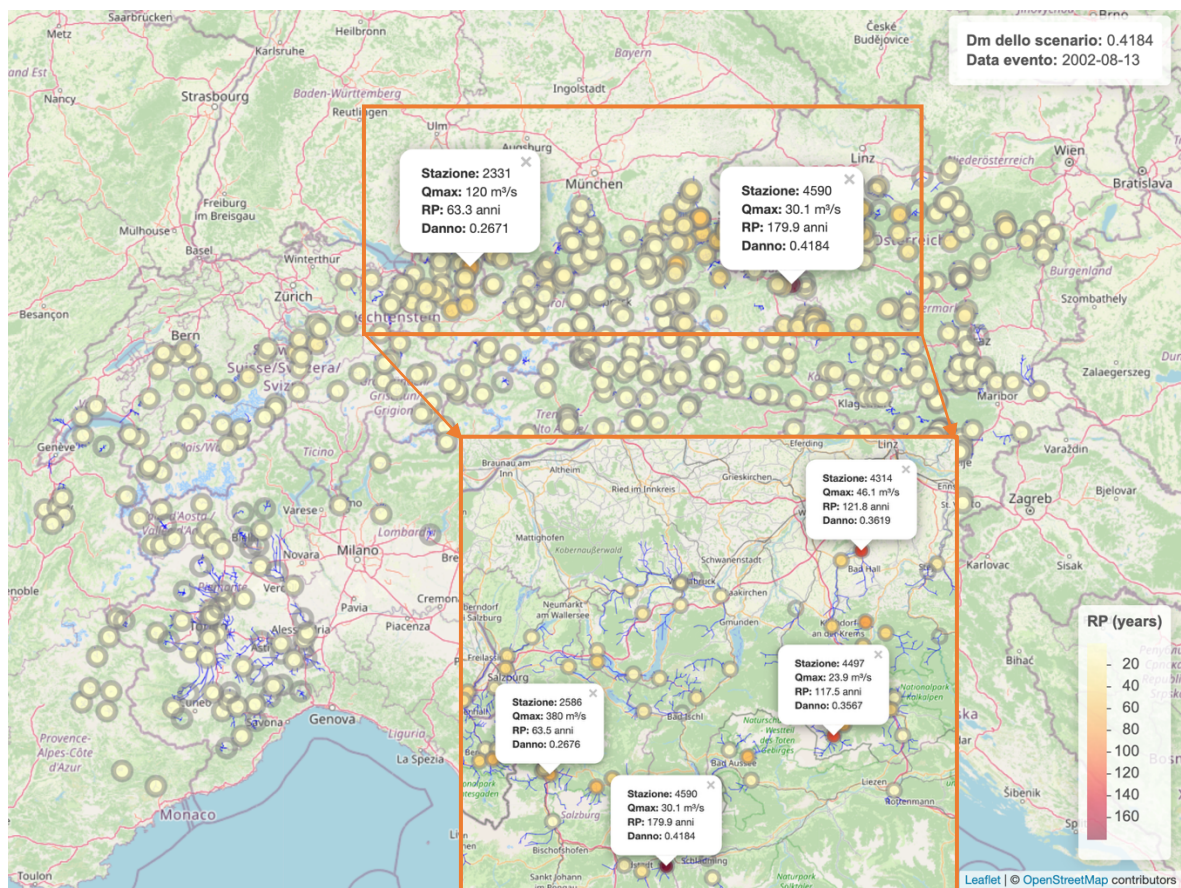


Figura 69- Visualizzazione dell'evento del 2002, con zoom sull'area di Sonthofen e Salisburgo.

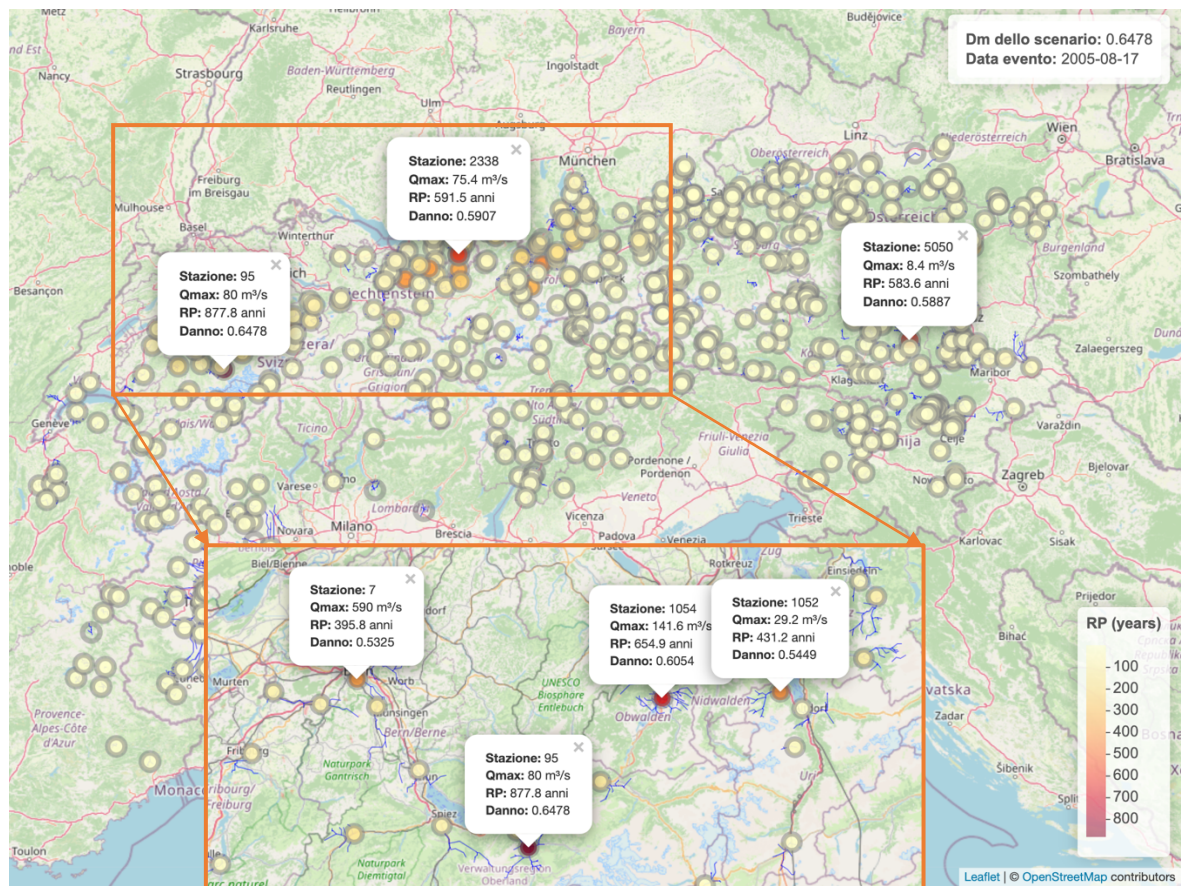


Figura 70- Visualizzazione dell'evento del 2005, con zoom sull'area di Berna.

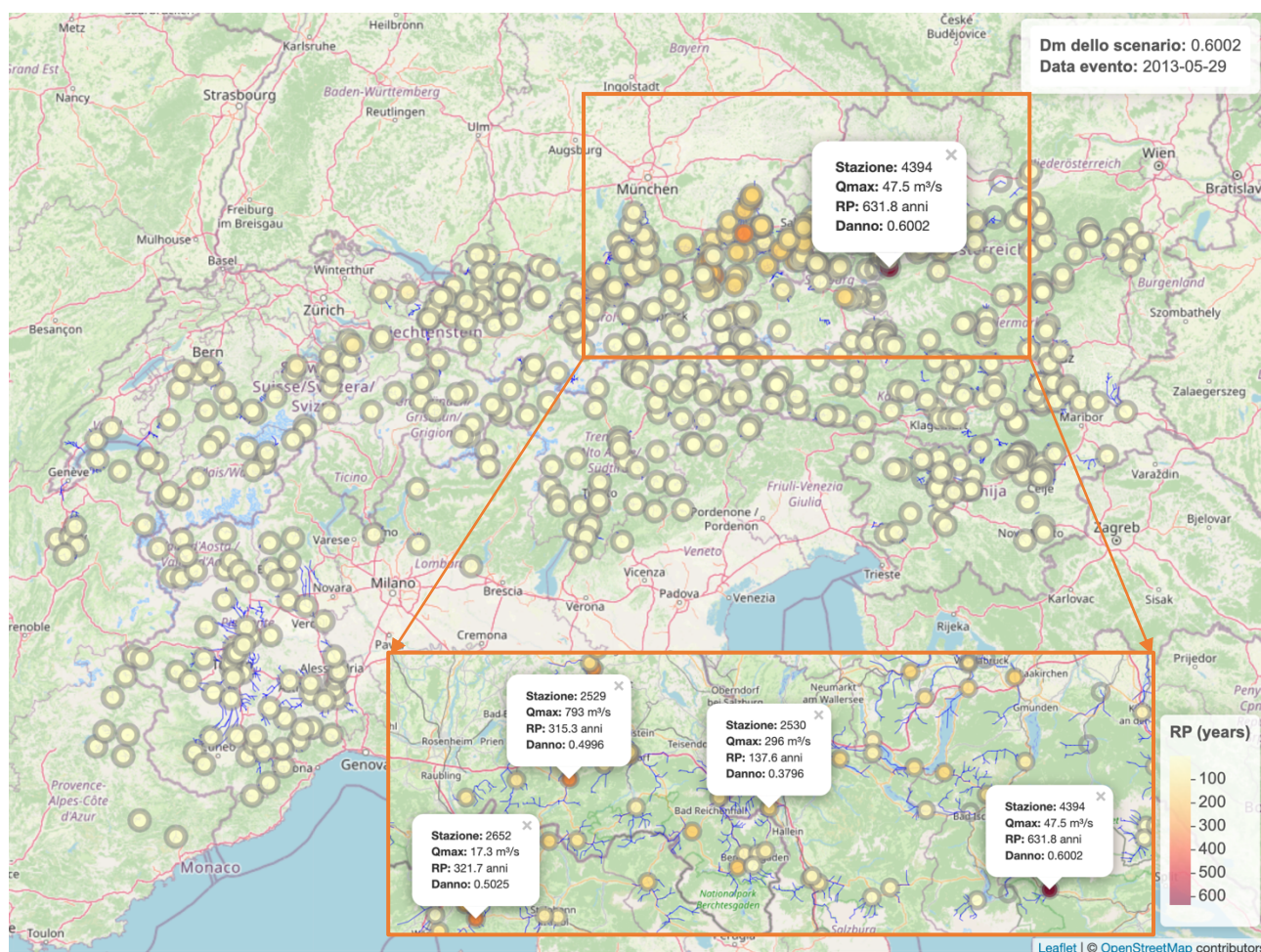


Figura 71- Visualizzazione dell'evento del 2013, con zoom sull'area di Kufstein.

Le mappe delle figure appena riportate, realizzate con la libreria Leaflet di R, per i 15 eventi di piena regionale più rilevanti (*maxeventi*) mostrano in modo chiaro l'evoluzione sia della rete di osservazione sia delle caratteristiche idrologiche degli eventi stessi. Nelle prime mappe, relative agli eventi dei primi decenni del Novecento, il numero di stazioni disponibili è molto limitato: i pallini colorati sono pochi e la copertura spaziale decisamente frammentata. Con l'avanzare del tempo, invece, si osserva un progressivo aumento della densità di stazioni monitorate, che porta le mappe più recenti a presentare un quadro molto più ricco e dettagliato della distribuzione spaziale delle piene. Un altro aspetto evidente è l'aumento dei tempi di ritorno massimi con il passare degli anni. Mentre gli eventi storici mostrano valori relativamente contenuti, gli eventi più recenti raggiungono tempi di ritorno decisamente più elevati: questo è un fenomeno che può essere letto anche come un segnale dei cambiamenti climatici in atto, che hanno reso sicuramente più intensi gli eventi estremi su scala regionale. Dal punto di vista spaziale, emerge con chiarezza come una delle aree più colpite sia quella austriaca intorno a Salisburgo, dove si concentrano ben 5 dei 15 eventi selezionati.

Al contrario, nella zona del Piemonte e del Nord-Ovest italiano la rappresentazione è affidabile solo a partire dal 1999, a causa della mancanza di dati nei periodi precedenti.

Queste osservazioni portano a concludere che gli eventi dei primi del Novecento non siano, come anticipato nei paragrafi precedenti, completamente attendibili: la scarsità di stazioni disponibili non consente di avere una rappresentazione robusta del danno medio (D_m), e i valori stimati risultano comunque caratterizzati da tempi di ritorno relativamente bassi rispetto a quelli osservati, per eventi simili, in epoca recente nelle stesse aree.

Per tali eventi sono stati inoltre calcolati:

- la media geometrica dei tempi di ritorno per ciascun sito, utile a descrivere la severità media delle piene in termini probabilistici;
- la media dei valori di pseudo-danno, che fornisce una misura sintetica dell'impatto medio stimato per ogni bacino;
- il numero di occorrenze per le quali il tempo di ritorno ha superato i due anni ($nevRPgr2$) e il numero di eventi con pseudo-danno maggiore di zero ($nevDgr0$), come indicatori della frequenza e della diffusione spaziale degli eventi significativi.

I grafici di questi dati calcolati vengono riportati qui sotto come mappe statiche create grazie alla libreria *ggplot2* di R (figure da 72 a 75).

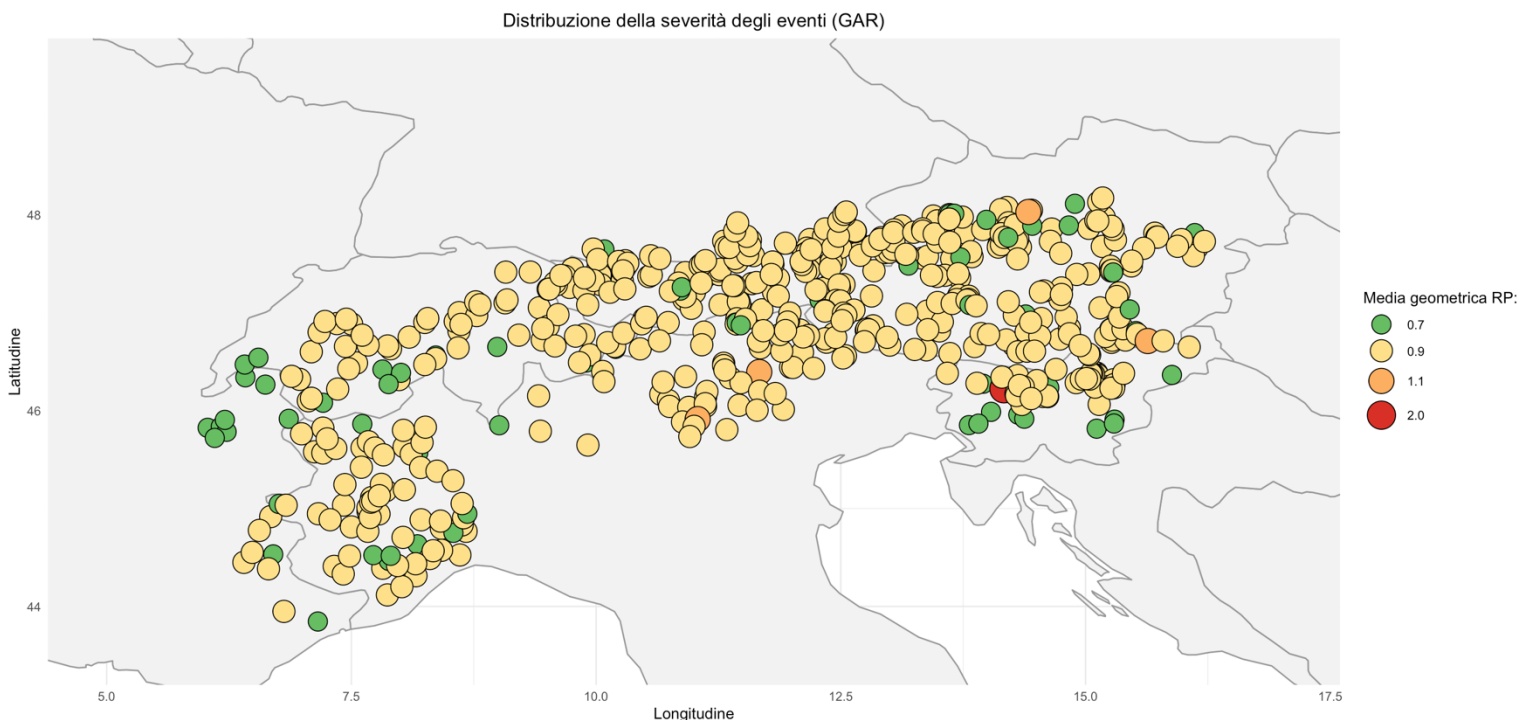


Figura 72- Media geometrica dei tempi di ritorno per ciascun sito.

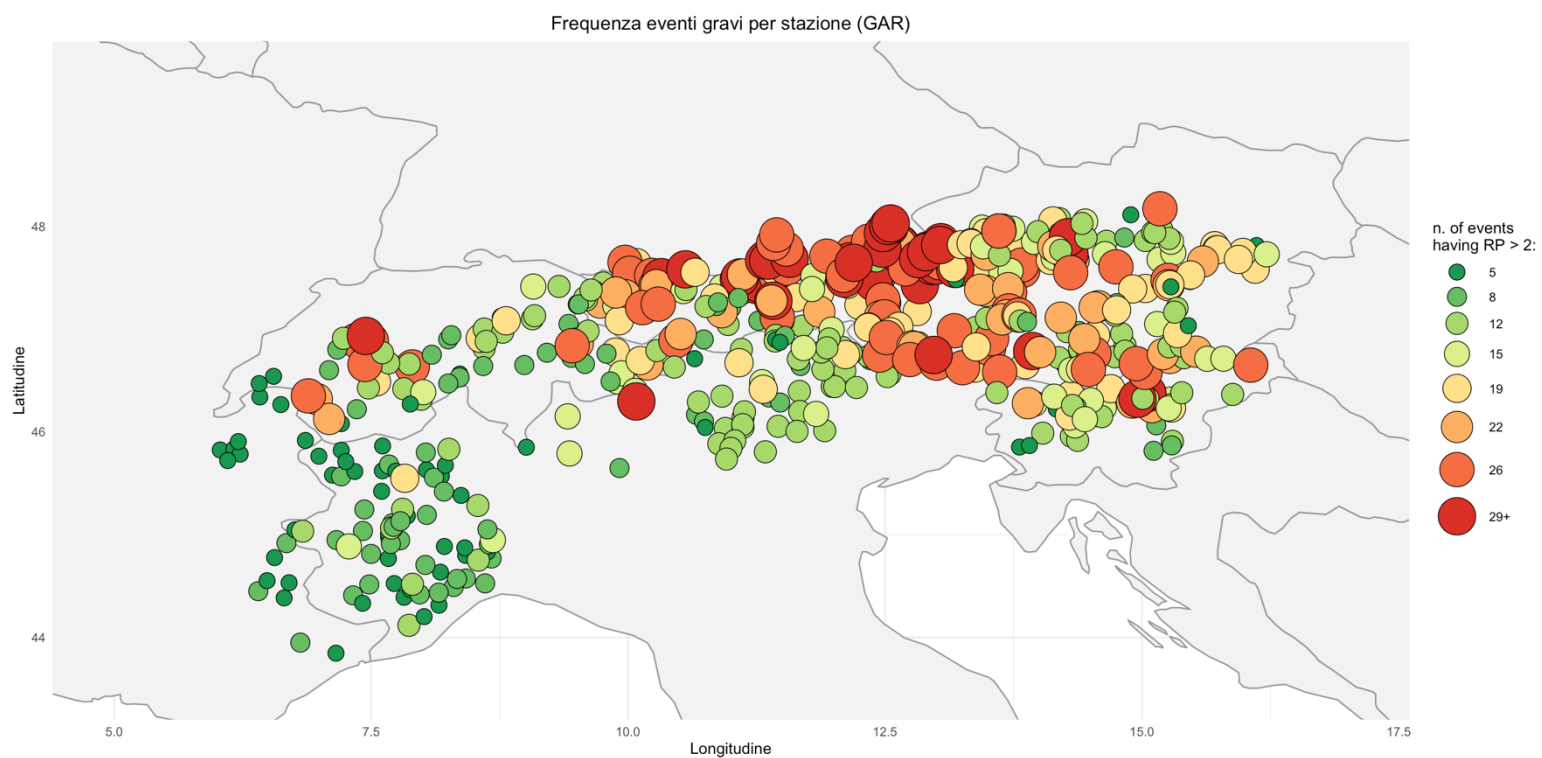


Figura 73 Frequenza eventi gravi per ogni stazione: numero di volte in cui si sono riscontrati eventi con tempo di ritorno maggiore di due anni.

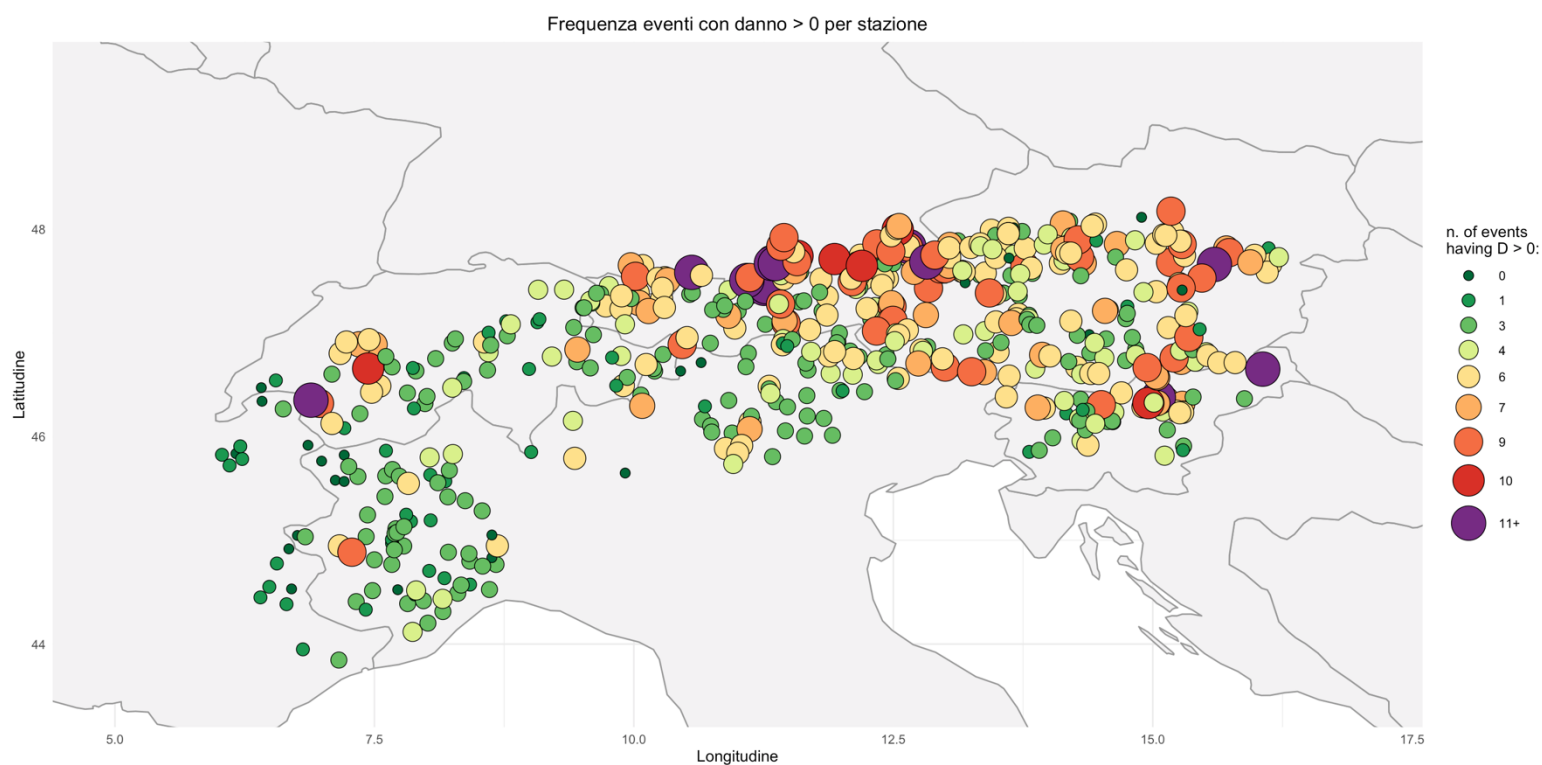


Figura 74- Frequenza eventi gravi per ogni stazione: numero di volte in cui D_m è risultato essere maggiore di zero.

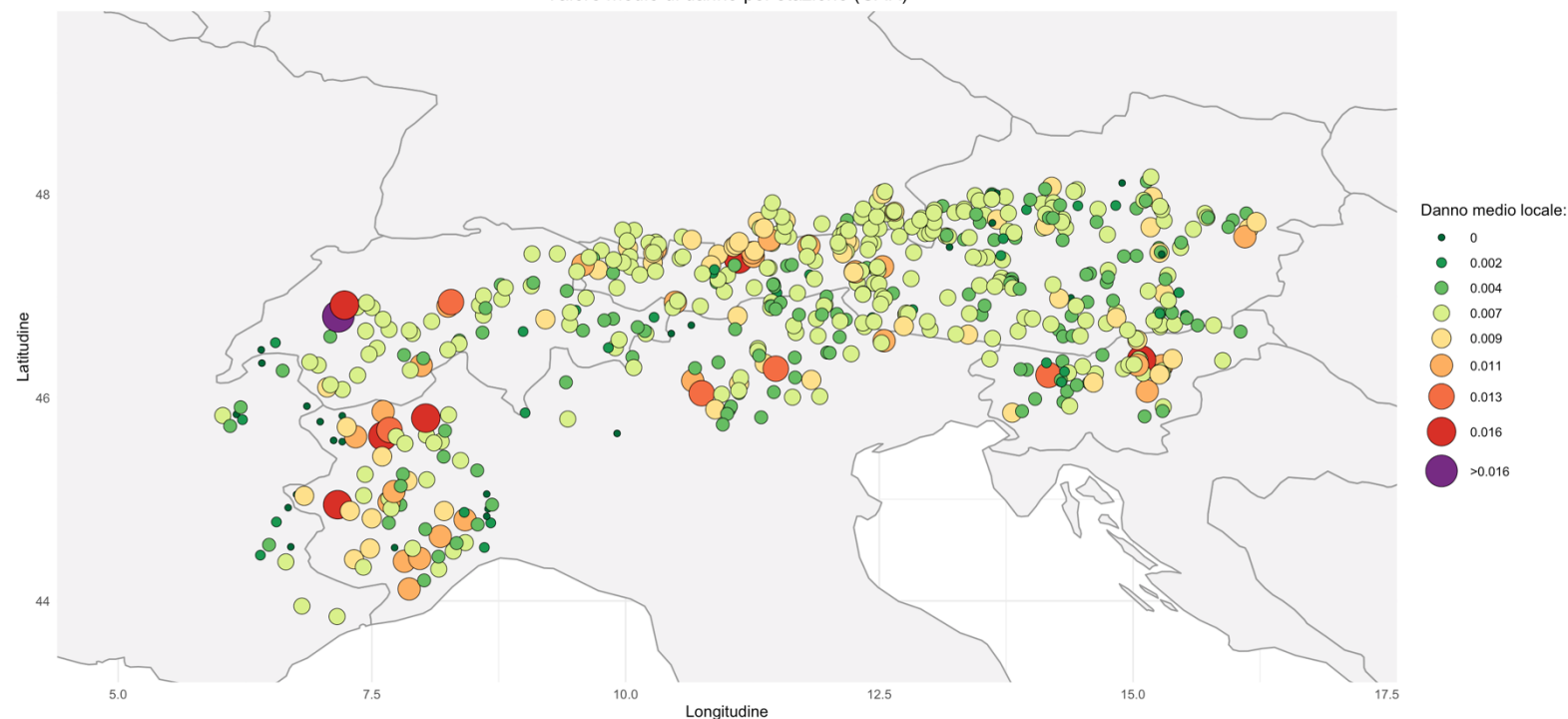


Figura 75- Valore medio di danno per stazione.

La mappa in figura 72 mostra il Periodo di Ritorno (RP) locale di eventi idrologici estremi, con colore che varia dal verde (RP più basso, eventi più frequenti) al rosso (RP più alto, eventi rari). La distribuzione spaziale evidenzia un certo grado di eterogeneità: le aree verdi sono caratterizzate da una maggiore frequenza di eventi con periodi di ritorno tendenzialmente bassi (eventi poco dannosi) mentre, in Slovenia ad esempio sono più frequenti eventi con media geometrica dei tempi di ritorno più alta (in grado quindi di causare più danni). Tale caratteristica può chiaramente essere dovuta all'influenza congiunta di orografia, uso del suolo e caratteristiche dei bacini.

I grafici 73 e 74 mostrano rispettivamente i siti in cui si sono registrati più volte eventi con tempo di ritorno pari 2 anni e si è registrato un danno maggiore di zero. I risultati indicano che, come visto in precedenza, è confermato il fatto che l'area più esposta a eventi in grado di causare danni è l'area di Salisburgo in Austria: i colori più accessi si registrano anche in Friuli-Venezia Giulia e in Slovenia. Un caso interessante è l'area di Berna (Svizzera), che appare relativamente isolata tra stazioni con colori prevalentemente verdi ma presenta comunque tonalità rosse/viola.

La figura 75 mette in luce, invece, un elemento non emerso nelle mappe precedenti: nel Nord-Ovest italiano, in particolare in Valle d'Aosta, si osservano eventi con indici di danno superiori alla media. Considerata la brevità delle serie storiche di portata giornaliera disponibili per questa zona rispetto ad altre aree della GAR, è plausibile che in passato non siano stati registrati eventi di pari intensità.

3.4 Ricostruzione storico-documentale dei 15 eventi più gravosi.

In questo paragrafo si riportano i risultati della ricerca documentale condotta per i 15 eventi di portata massima selezionati: sono state raccolte e valutate prove fotografiche, articoli di stampa e resoconti locali, dove disponibili, per verificare e contestualizzare i picchi misurati nelle serie giornaliere.

Evento di giugno 1901 a Domat/Chur (Svizzera): essendo l'evento più remoto è stato più complicato trovare informazioni storiche ma di seguito vengono riportati i dati del <https://www.sturmarchiv.ch> che descrivono in parte l'evento.

Fatti in breve	
Tipo di evento	Alluvione
Stato di verifica	QC1
ESWD	Non segnalato
Posizione	Ticino (TI), San Bernardino (GR), Hinterrein (GR)
Tempo / Durata	24 ore
Data	14-15 giugno 1901
Magnitudine / Dimensione	>100mm di pioggia in 24 ore
Danno	-
Decessi	-
Lesioni	-
Segnala la fonte	cronache, archivio generale
Osservazioni	-

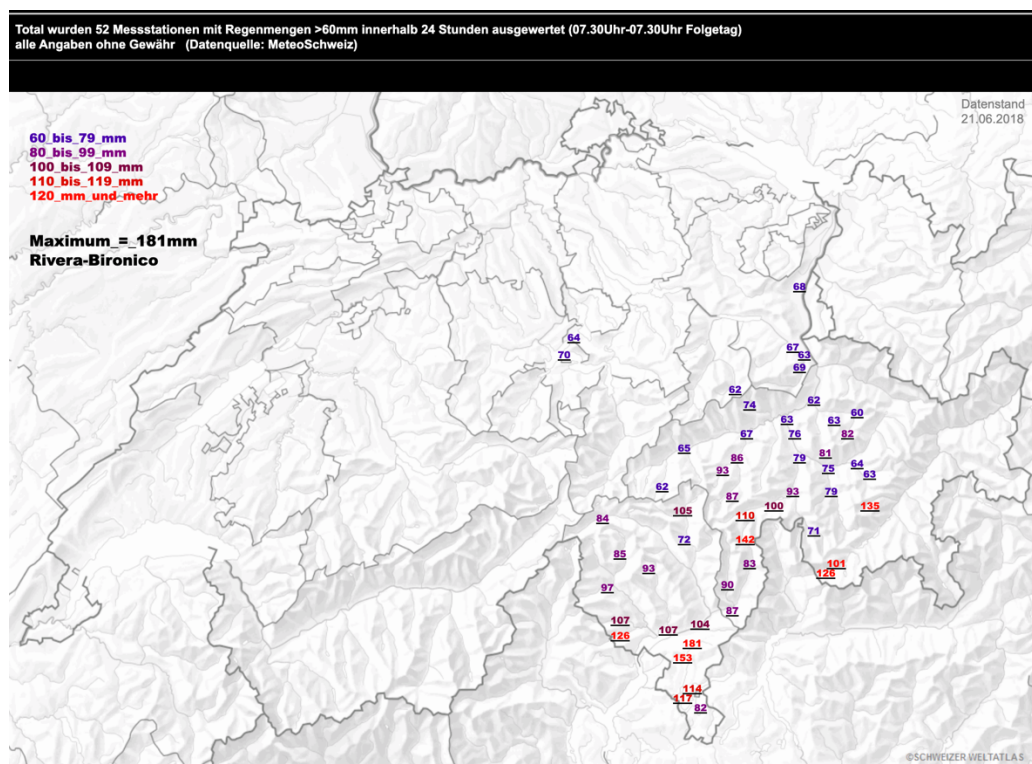


Figura 76- Immagini con i mm di pioggia caduti nell'area di Chur.

Witterung im Juni

Auf die Ende Mai eingetretene kühle Periode folgte vom 2. an rasches Steigen der Temperatur bei trockenem, meist heiterem Wetter. Am Abend des 6. brachten ausgedehnte Gewitterregen [...] wieder Abkühlung. Dann folgte [...] vom 8. an [...] wieder heiteres trockenes Wetter mit steigender Temperatur. Vom 13. an stellte sich über Südwesteuropa neuerdings Hochdruck ein, dessen Regime sich bis zum Schluss des Monats geltend machte. Die Witterung war dabei [...] etwas veränderlich, zeitweise trübe und regnerisch. (p. 170)

Figura 77- Breve descrizione del meteo pre e post evento da: <https://www.meteosuisse.admin.ch/>

Evento di luglio 1914 stazione di Chessel (Svizzera): situata nella pianura del Rodano tra Vouvry–Port-Valais–Noville, all'estremità opposta a Ginevra del Lago Lemano. L'evento avviene durante l'estate 1914 che è segnata da episodi temporaleschi intensi nel bacino alto del Rodano e, più a sud-est, da grandi piene sull'Isère. Nella letteratura classica sul Rodano, luglio 1914 è citato tra le piene più temibili nell'alto corso fino a Saint-Maurice (poco a monte del delta dell'Haut-Lac. Cronache meteo locali segnalano temporali e allagamenti nel bacino lemanico già a fine giugno 1914 (precisamente a Vevey il 27/06): ciò indica una fase instabile prolungata prima di metà luglio. Il quotidiano M. Pardé (1925, 1928) (*Pardé Maurice. Le régime du Rhône. In: Revue de géographie alpine, tome 13, n°3, 1925. pp. 459-547*) cita esplicitamente luglio 1914 come episodio di piena importante nel sistema del Rodano e come crue «exceptionnelle» in area grenoblese (utile per il quadro sinottico). Documenti cantonali ricordano che nel 1914 iniziano gli scavi del Grand Canal nella piana del Rodano con funzione di bonifica e difesa idraulica del territorio. (<https://www.faovd.ch/alphabet-des-communes/225/Chessel>)



Figura 78- Mappa che mostra l'area dell'evento del 1914 che colpì Chessel e Saint-Maurice. Viene evidenziata la distanza tra i due paesi.

LE RÉGIME DU RHONE

Par MAURICE PARDÉ.

J'avais fait part à Maurice Pardé, le jour de la soutenance de ses thèses, de mon intention de donner à la Revue de Géographie Alpine un compte rendu de son beau travail sur le régime du Rhône, d'en célébrer les mérites, et surtout de signaler tout ce qu'il apporte de nouveau. C'est alors qu'avec sa spontanéité habituelle il m'offrit, par reconnaissance pour la Revue qui a accueilli ses premiers travaux, de résumer lui-même pour cette Revue les principaux résultats de son travail. Je ne pouvais qu'accepter avec gratitude, et je le remercie cordialement. Je suis persuadé d'ailleurs que la lecture du vigoureux article qui suit ne pourra qu'inspirer le désir de connaître le livre lui-même, et je me permets d'assurer tous ceux qui auraient cette pensée qu'ils ne seront pas déçus¹.

Raoul BLANCHARD.

¹ Maurice Pardé, *Le Régime du Rhône*. Etude hydrologique, Université de Lyon, Institut des Etudes Rhodaniennes, 1925, 2 vol. in-8°, XIV+887 et 440 p., 117 fig. Voir également : *Le calcul des débits du Rhône et de ses affluents*. Grenoble, Bibliothèque de l'Institut de Géographie alpine, 1925, in-8°, 168 p.

Mais c'est dans les Alpes, à savoir sur leur versant piémontais et encore plus sur eur versant méridional, au Sud de la Suisse, puis un peu au Nord de la frontière oro- graphique, dans les vallées voisines longitudinales du Rhône, du Rhin et de l'Inn supérieurs, que ces crues méditerranéennes d'été paraissent les plus fréquentes ; de sorte qu'elles concourent fortement et assez curieusement au caractère continental que contribue à donner ici au climat l'abondance estivale moyenne des chutes d'eau. Nombre de maxima qui, dans les tableaux récapitulatifs des Annuaire hydrographiques de la Suisse, figurent au passif de juillet ou d'août pour les rivières septentrionales, dans leurs secteurs non éloignés des pays italiens ou tessinois, doivent avoir eu pour causes de grandes averses amenées par le vent classique de type föhn à composante Sud-Est, et qui ont pu s'étendre au Nord de la ligne de partage montagneuse, grâce aux échancrures des cols vers des affluents à couloirs orientés S-N. Ce fut le cas très certainement pour les hautes eaux des 23-24 juillet 1914, qui ont sévi à la fois sur le Rhin et le Rhône alpestres, le Tessin, la haute Durance, le Drac et l'Isère et les rivières cévenoles.

Figura 79– In questo passaggio del giornale dell'epoca "Regime du Rhone" si parla dell'evento in questione.

Come ulteriore conferma il servizio meteorologico NFO LIMAT francese riporta negli archivi un riferimento all'evento in questione nel seguente modo:



"12 au 13 juillet: Dans la nuit un gros orage frappe les hauteurs de Vevey, à Blonay et à Saint-Légier, avec des dégâts d'eaux."

(<https://forums.infoclimat.fr/f/topic/23756-orages-en-r%C3%A9gion-l%C3%A9manique>)

Evento di settembre del 1920 a Martina (confine tra Austria e Svizzera):

A fine settembre 1920 si registrò una grande ondata di piena che interessò varie regioni alpine. Tra il 19 e il 24 settembre 1920 risultano colpiti fra gli altri l'Engadina, il Münstertal e il basso Plessurtal (Martina si trova proprio all'interno di quest'area). (Da: <https://www.gra-nat.ch>)
Martina è il villaggio al confine proprio sul passo con Nauders (Austria), posto sulle sorgenti del fiume Inn.

Sempre da <https://sturmarchiv.ch> si trovano informazioni su questo evento che vengono riportate di seguito:

September. Witterung sehr trüb, mit häufigen und außerordentlich reichlichen Niederschlägen. Besonders der Gewitterregen vom 19. wies eine ungewöhnlich große Niederschlagsmenge auf (Chur 28,1 mm). Und noch gewaltiger waren die Regengüsse vom 21. und 22. Sie bewirkten ein mächtiges Anschwellen der Flüsse im ganzen Lande. Als dann am 23. morgens die Niederschläge neuerdings einsetzten, war man auf sehr Schlimmes gefaßt. Aus einzelnen Gegenden trafen auch Meldungen über erfolgten Hochwasserschaden und über Ausbrüche von Rufen ein. Zum Glück hörte es am 23. bald zu regnen auf, und wir blieben vor einer größeren Hochwasserkatastrophe verschont.

In Zillis ging eine Rufe durch das Dorf herunter, am der Post vorbei, und ergoß sich über die darunter liegenden Wiesen.

Bei Andeer vor Bärenburg wurde durch einen Rüfenausbruch die Straße überschüttet.

In Zernez unterhalb der Bahnbrücke durchbrach der Inn die Bewehrung auf einer Strecke von ca. 50 Metern. Seit 1888 hat man den Inn nie so hoch gesehen.

Im Münstertal bedrohte der Muranzabach mit seinen Geschiebmassen einige Häuser in Sta. Maria, und etwas unterhalb von dessen Einmündung in den Rambach wurde von letzterem eine Brücke weggerissen.

Der September brachte uns eine reiche Steinobsternte.

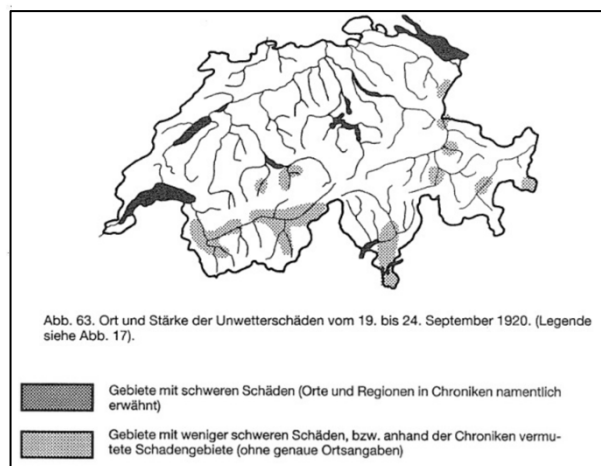


Figura 80- Testimonianza dell'evento di settembre el 1920 da: <https://www.e-periodica.ch/Naturchronik> 1919 und 1920 mit Witterungsnotizen für Chur.

Evento del 1923 ad agosto a Brig (Svizzera):

L'evento viene citato nell'articolo (Yihua Zhong et.al, Historical flood reconstruction in a torrential alpine catchment and its implication for flood hazard assesment, Journal of Hydrology, Volume 629, 2024, 130547, ISSN 0022-1694, <https://doi.org/10.1016/j.jhydrol.2023.130547>). Non sono state trovate ulteriori informazioni per l'area di Brig.

Evento del 1926 ad agosto a Nazarje (Slovenia):

L'evento di agosto 1926 è riconosciuto come una delle inondazioni più significative tra le due guerre, insieme a quelle del 1923 e del 1933. La causa fu una pioggia eccezionale in poche ore che ha dato origine a massicce inondazioni in vaste aree del territorio sloveno. Le zone colpite principali includevano Celje, la valle della Savinja, con espansione anche a nord-ovest, estendendosi fino al Prekmurje. Il quotidiano <https://www.slovenskenovice.si> riporta in un articolo recente alcune immagini dell'evento alluvionale dell'epoca, riportate di seguito: "600 chilometri quadrati di Slovenia e dove, secondo i calcoli, con una media di un decilitro di pioggia per metro quadrato alla fine di settembre di 92 anni fa, caddero 60 miliardi di litri d'acqua, che dovettero defluire entro 24 ore", secondo un rapporto dell'ing. Anton Šivic in "Inondazioni nella regione di Lubiana e silvicoltura".



Figura 81-Immagini dell'evento dell'agosto del 1926 nelle zone di Nazarje.

Evento del 1927 ad agosto a Chur (Svizzera):



Vengono riportate tre immagini dall'archivio digitale il cui link è riportato di seguito. L'immagine a colori rappresenta ciò che resta di un edificio dell'epoca che subì gravi danni dall'evento alluvionale. Le altre due immagini vennero scattate sul posto subito dopo l'evento e mostrano edifici rasi al suolo e

gravi danni al territorio. (<https://ueberschwemmungsgedaechtnis.hochwasserrisiko.ch>). Di

seguito viene anche mostrata la mappa dell'evento da: <https://sturmarchiv.ch>. Si segnala che durante l'evento annegarono 12 persone, con danni che risultarono pari a 7,5 milioni di franchi.

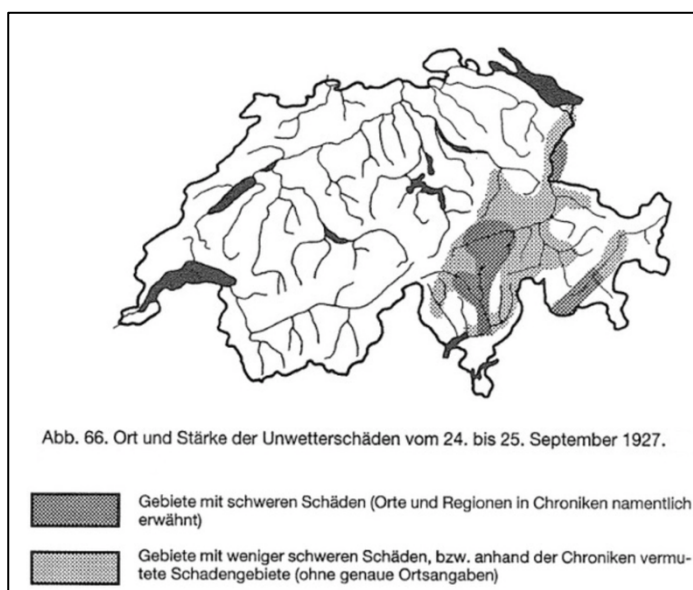


Figura 82- Immagini dell'evento del 1927

Evento del 1940 a giugno area del Kochelsee (Alta Baviera, Germania):

Il bacino del Kochelsee (alimentato dall'Isar superiore e dai torrenti di montagna circostanti) fu colpito da piene improvvise e trasporto detritico. Le cronache idrologiche bavaresi segnalano danni diffusi a strade e ponti, e livelli eccezionali registrati lungo il Loisach e nell'area tra Kochel e Benediktbeuern.

Riguardo questo evento non sono disponibili rapporti tecnici che lo documentano: la situazione è comprensibile in quanto avvenne in piena Seconda Guerra Mondiale e molte testate giornalistiche dell'epoca non hanno digitalizzato gli articoli che lo riguardano.

Evento del 1954 a giugno nell'area di Salisburgo (Austria/Baviera):

L'alluvione del Danubio del 1954 fu un evento eccezionalmente violento, provocato dalle precipitazioni nelle Prealpi bavaresi e austriache tra il 7 e l'11 luglio 1954.

Le precipitazioni caddero ininterrottamente per circa 70/90 ore su terreni già saturi da un precedente periodo piovoso iniziato il 27 giugno. In totale, circa 150000 ettari di terreno furono allagati, di cui 57888 ettari nella Bassa Baviera, 51286 ettari nell'Alta Baviera e 38535 ettari nell'Alto Palatinato. I fiumi Danubio, Inn e Ilz furono i protagonisti delle inondazioni. Queste ultime hanno causato complessivamente 12 vittime e oltre 9000 persone hanno dovuto essere evacuate. Complessivamente il disastro ha causato danni economici per quasi 60 milioni di euro. Informazioni da: (<https://undine.bafg.de/donau>).



Figura 83- Immagini di Palau (Germania) sommersa dall'esondazione del Danubio.

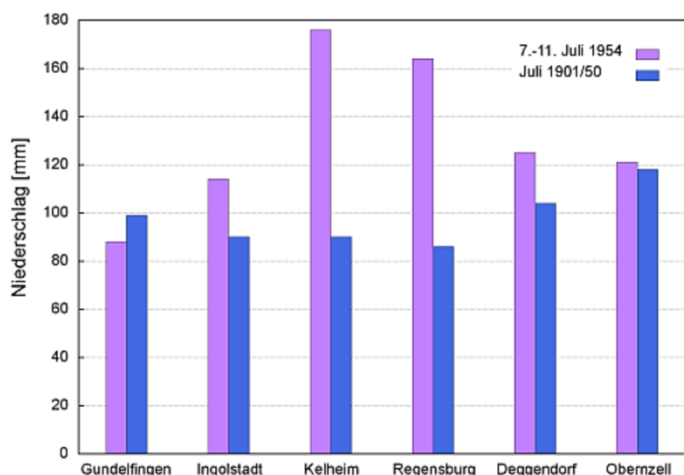


Figura 84- Precipitazioni (totale di quattro giorni dal 7 all'11 luglio 1954) nei punti di misurazione nella valle del Danubio rispetto alla media delle precipitazioni totali di luglio della serie temporale dal 1901 al 1950.

In particolare, il bacino della Salzach fu particolarmente colpito. A Salisburgo città il fiume superò gli argini in diversi tratti, con allagamenti estesi, danni a edifici, strade e ferrovie. Le conseguenze furono molteplici: centinaia di case danneggiate, ponti distrutti ed evacuazioni in larga scala. L'evento del 1954 è rimasto a lungo come riferimento nelle statistiche idrologiche e nella pianificazione delle opere di difesa della regione del bacino di Salzach.

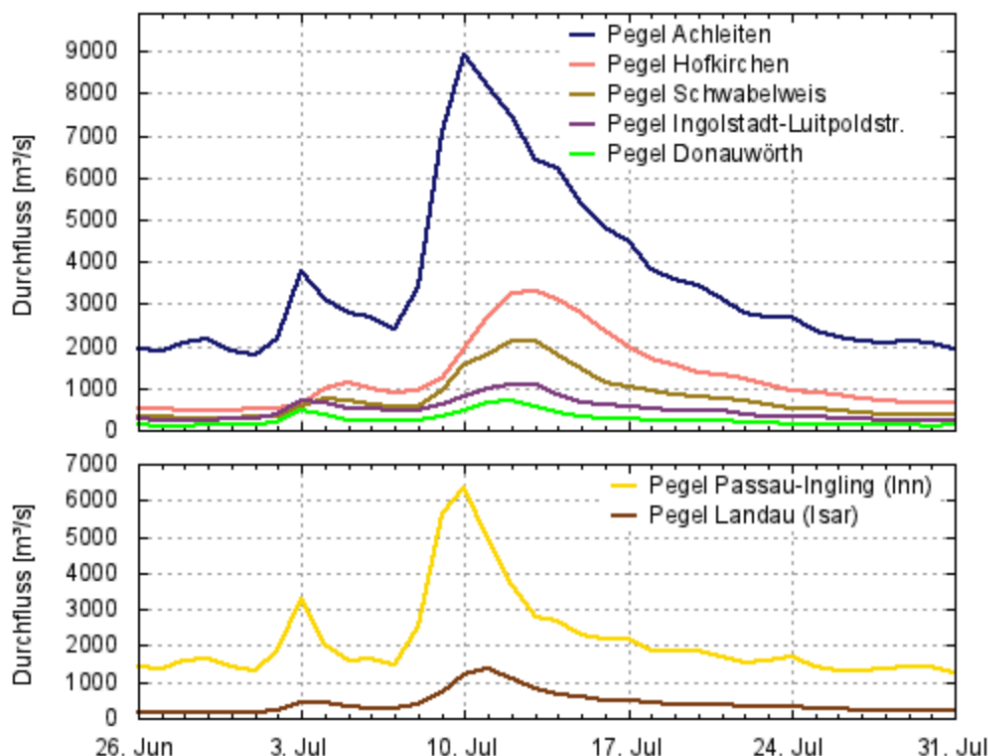


Figura 85- Portate (valori medi giornalieri: 26.6.-31.7.1954) presso misuratori selezionati nel bacino del Danubio. Da: (<https://undine.bafg.de/donau>)

Evento dell'agosto 1959 nelle zone di Shladming (Austria): nubifragi e 30 ore di pioggia causarono una catastrofe nella Baviera sud-orientale e in Austria, persino peggiore dell'alluvione del 1954. Nella contea di Salisburgo, le inondazioni furono quelle di maggiore entità. Infatti, lì il fiume Salzach raggiunse la piena più alta dal 1920. L'acqua di quest'ultimo strappò via una parte del ponte autostradale che collegava Salisburgo a Vienna (mostrato in figura 86). Di conseguenza il ponte crollò e con ciò, il collegamento tra queste due città fu interrotto. Si stima che i danni ammontarono a circa 35 milioni di scellini austriaci.

(<https://www.alamy.com/aug-15-1959-floine-in-austria-cloud-bursts-and-30-hours-of-rain-caused-image69359592.html>)

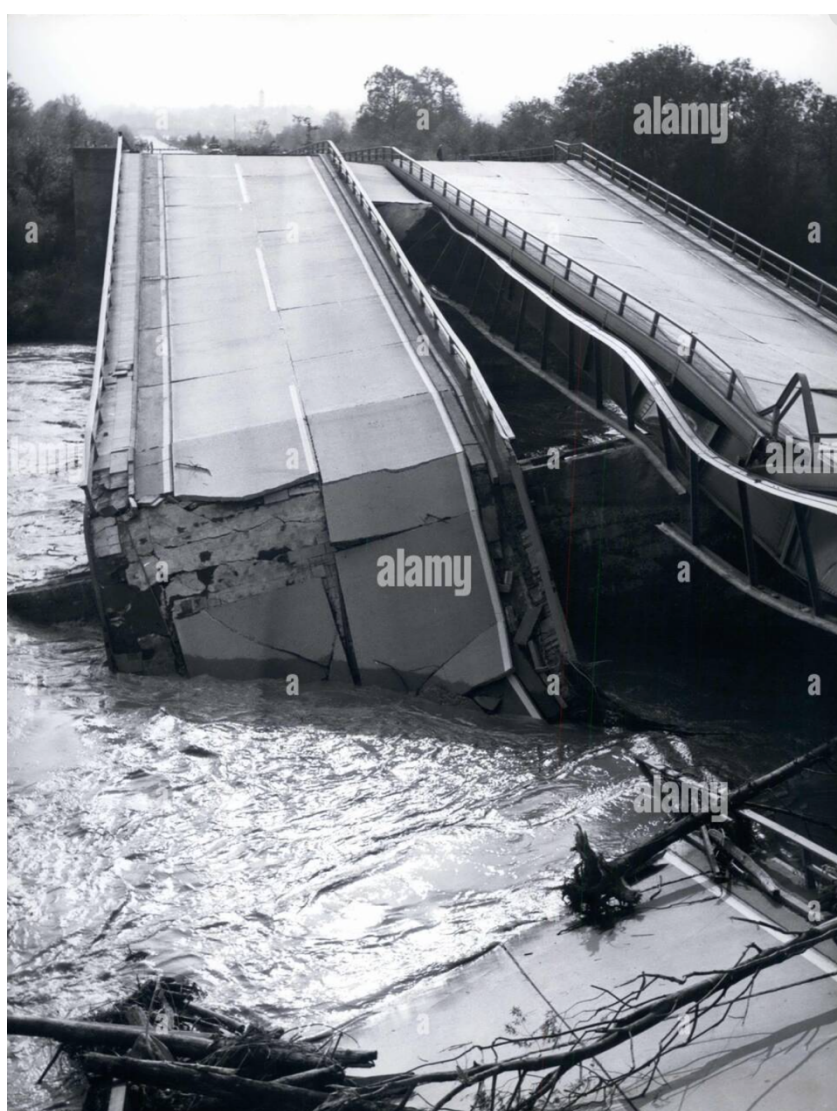


Figura 86- Immagine dell'evento del 1959 in Austria.

Evento dell'agosto 1965 nelle zone di Sillian e Dolach (Austria al confine con l'Italia vicino ad Auronzo di Cadore): in particolare, nell'area della Drava (Drautal), i comuni di Sillian e Dölsach furono colpiti in pieno dalla furia delle acque. Situazioni simili si trovarono anche a Bezirk, Lienz e nell'Osttirol.



Figura 87- Collezione Vigili del Fuoco volontari città di Lienz 1965. Da: <https://www.tiroler-photoarchiv.eu>



Figura 88- Immagini dell'evento del 1965. Da:
<https://tirol.kommunalarchive.at/index.php/at-gema-dolsach-b-1-1-334>

In due giorni, 400 milioni di metri cubi d'acqua si riversarono sul Tirolo Orientale con effetti devastanti. Dopo il disastro, a Vienna si prese seriamente in considerazione l'ipotesi di trasferire tutti gli abitanti della valle del Deferegggen. (<https://tirol.orf.at/v2/news/stories>) Furono costruite barriere e fu istituito il fondo di soccorso. Ci furono anche vittime: nella frazione di St. Veit, Gassen, una gigantesca frana di fango distrusse una casa bifamiliare, uccidendo tragicamente sei persone. (<https://www.meinbezirk.at/osttirol>).

Evento dell'agosto 1966 nelle zone di Obervellach (Austria/Tirolo):

Nell'estate 1966 la valle della Drau e l'Osttirol subirono danni da piene ripetute: dopo l'evento del 1965 si ripeté una fase di grandi precipitazioni che portò nell'agosto 1966 ad esondazioni con danni gravi a Sillian e altri centri della valle. L'evento del 1966 è ricordato come una delle più rilevanti crisi idriche locali del Dopoguerra e ha poi motivato importanti interventi di difesa fluviale nella Drau (misure di arginatura e consolidamento tra gli anni '70 e '80).



Figura 89- Immagini relative al grave evento che colpì il Tirolo, soprattutto a Sillian (<https://tirol.kommunalarchive.at/>).

Evento del maggio 1999 nell'alta Austria al confine con la Germania (Fussen) con picchi a Innerlaterns e Nassereith:

L'evento di maggio 1999 (Pfingsthochwasser) fu un'ondata di piena sulla Baviera, Vorarlberg e Tirolo causata da forti piogge convettive sommate agli apporti della fusione nivale. Molte stazioni alpine registrarono innalzamenti eccezionali dei livelli idrici.

Il temporale "Quartus" si è abbattuto sulle Alpi settentrionali causando l'espansione dell'Ammersee, dell'Amper ed allagando vaste aree. Quest'ultime si sono rapidamente estese, coprendo parti dell'Alta Algovia sul fiume Iller. L'alluvione è poi proseguita a valle, raggiungendo alcune delle aree più abitate della Baviera. (<http://news.bbc.co.uk/>). L'evento fu molto esteso ci furono anche disagi a Thun (Svizzera), mostrati nella figura i seguito.



Figura 90- Immagine dell'alluvione del 1999 a Thun.



Figura 91- Immagini dell'epoca a Schnepfegg.

Evento dell'agosto 2002 nell'alta Austria e nella zona di Salisburgo:

Le inondazioni nell'Europa centrale dell'agosto 2002 hanno causato disastri in Germania, Repubblica Ceca e Austria. Gravi inondazioni si sono verificate nella Germania orientale e settentrionale, in particolare lungo il fiume Elba in Baviera, e in Austria lungo il Danubio. Le inondazioni sono state causate da giorni di piogge intense che hanno portato a settimane di soccorsi, almeno 45 morti e danni per un totale di circa 15 miliardi di euro.

L'alluvione del 2002 ha superato per gravità l'alluvione del 1954, il più grave del XX secolo, nella regione dell'Elba e in tutta l'Austria, nonché in alcune parti della Baviera.

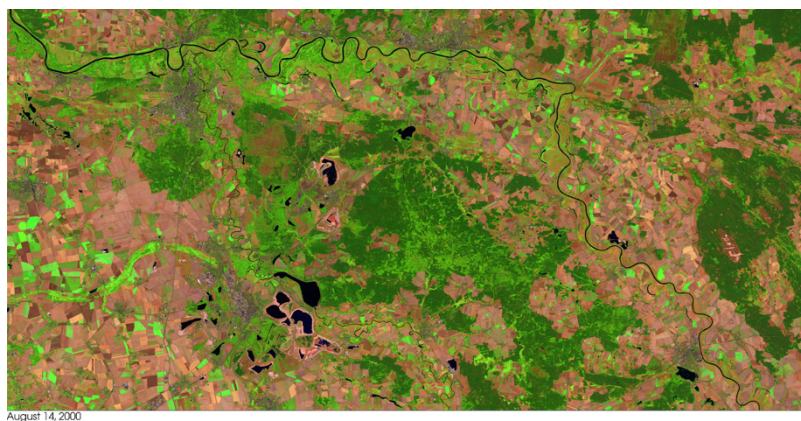


Figura 92- Immagine satellitare dell'Elba tra Torgau e Aken il 14 agosto 2000 (prima) e il 20 agosto 2002 (durante l'evento) Da: http://visibleearth.nasa.gov/view_detail.php?id=3584

Oltre 110 tra fiumi e torrenti esondarono in Austria. Ci furono anche molteplici segnalazioni di distruzione di infrastrutture (ponti, strade, ferrovie). Inoltre, ci furono danni economici enormi, solo in Austria stimati a tre miliardi di euro e numerose vittime.



Figura 93- Golling, Kuchl e Hallein furono dichiarate zone disastrose. Nella foto l'insediamento di Golling con il Salzach sommerse in gran parte la cittadina.

Evento dell'agosto 2005 che interessò Svizzera, Austria, Baviera e Repubblica Ceca:

Tra il 19 e il 23 agosto 2005 precipitazioni eccezionali hanno colpito l'arco alpino settentrionale: in Svizzera si registrarono in poche ore livelli e deflussi mai osservati in molte stazioni (Interlaken, Sarner, Berna), mentre in Baviera e Tirolo si verificarono piene e colate che interessarono anche Oberstdorf.



Figura 94- Svizzera, 24 agosto 2005. Da: <https://www.cbsnews.com/pictures/switzerland-floods>

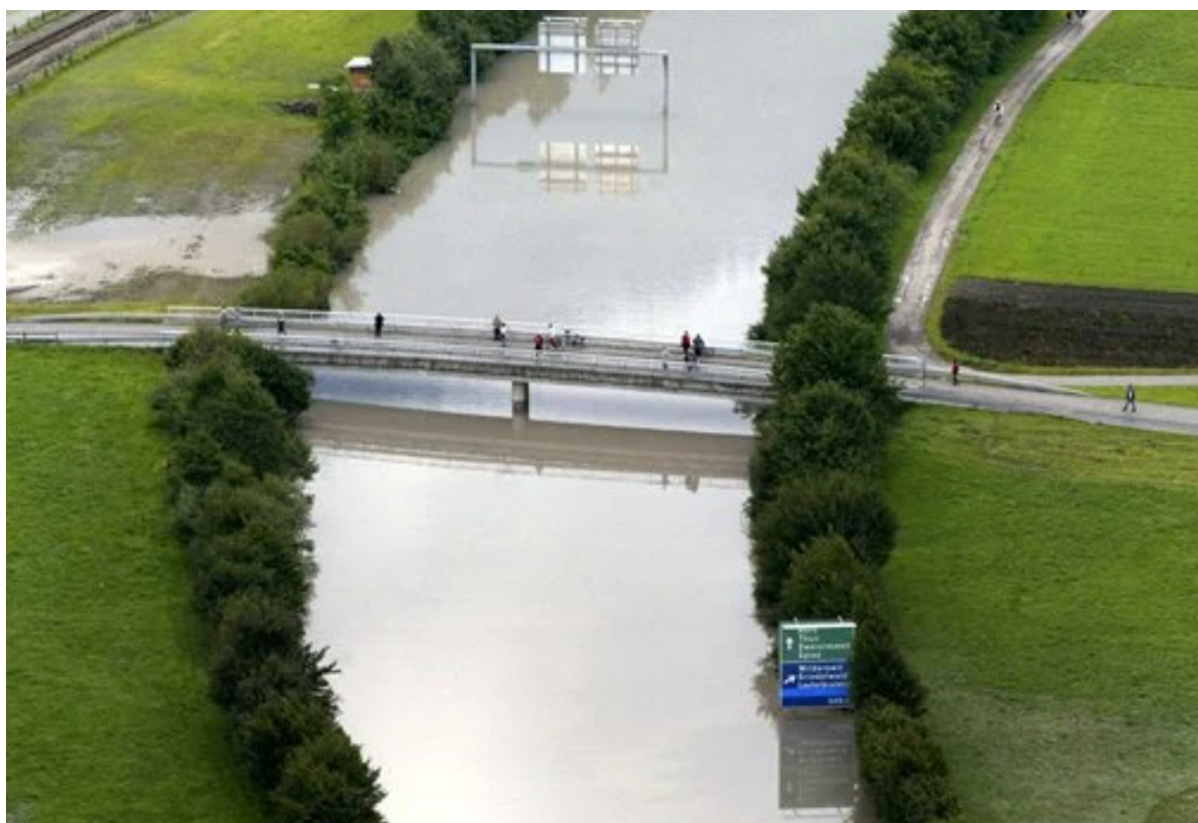


Figura 95- Immagine dell'alluvione in Svizzera: autostrada A8 fuori Interlaken chiusa al traffico a causa di un'alluvione, il 23 agosto 2005. Da: <https://www.cbsnews.com/pictures/switzerland-floods>



Figura 96- Immagini della vista aerea che mostra il quartiere Mattequartier, a Berna, in Svizzera, completamente circondato dall'acqua dell'alluvione, il 23 agosto 2005. Da: <https://www.cbsnews.com/pictures/switzerland-floods>

Evento di giugno 2013 che interessò Austria, Baviera:

Alla fine di maggio 2013 l'aria umida bloccata contro le Alpi, i rilievi della Baviera e della Boemia, produsse piogge molto intense e persistenti per diversi giorni consecutivi (dal 30 maggio al 3 giugno). In alcune aree alpine e prealpine si superarono i 300 mm di pioggia in pochi giorni, con suoli già saturi da settimane di precipitazioni primaverili.

La Svizzera orientale, il Tirolo e la zona di Salisburgo furono le prime ad essere colpite. In Austria, tra il 2 e il 5 giugno, i livelli dei fiumi Enns, Salzach, Inn e soprattutto Danubio salirono rapidamente. A Linz il Danubio raggiunse quasi i 10 m di altezza, il secondo valore più alto mai misurato. In Baviera ci fu il caso più drammatico: nella città di Passau il 3 giugno 2013 il Danubio raggiunse 12,89 m, il livello più alto dal 1501. La città storica costruita all'incrocio tra Danubio, Inn e Ilz, rimase completamente allagata. A Salisburgo e in alta Austria, si verificarono anche frane e colate detritiche oltre alle inondazioni fluviali.

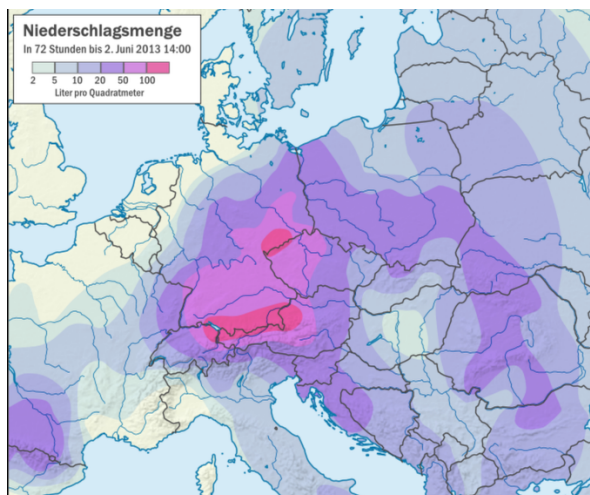


Figura 97-Le tre immagini rappresentano rispettivamente: mappa delle precipitazioni durante l'evento; Pianura alluvionale del basso Elba; Passau, 3 giugno 2013.

Gli impatti furono molteplici:

- le vittime furono molteplici in Europa centrale e si segnalano evacuazioni di decine di migliaia di persone nelle regioni della Baviera e in Austria;
- i danni stimati in Austria furono pari a circa tre miliardi di euro, con forti impatti sui trasporti, energia e turismo. In Germania invece oltre otto miliardi di euro di danni. Furono anche mobilitati in Austria oltre 20000 uomini del Bundesheer e della protezione civile per rinforzare dighe, montare barriere mobili e assistere la popolazione.



Figura 98- Fermata dell'autobus allagata a Worms (Germania), 3 giugno 2013.

4. Analisi probabilistica di eventi simulati sulla GAR.

Una volta mostrati gli eventi più gravosi del passato secondo l'indice di magnitudo D_m , la trattazione procede con il seguente capitolo dove si descrive la procedura probabilistica adottata per la generazione e l'analisi statistica di eventi di piena simulati su scala spaziale. L'obiettivo principale è costruire una larga popolazione di eventi idrologici, ottenuta mediante simulazioni create con il metodo Monte Carlo, che riproduca sia le caratteristiche marginali dei dati osservati in ciascuna stazione sia la struttura di dipendenza spaziale fra le stazioni stesse. La disponibilità di una tale catalogazione consente di stimare in modo robusto probabilità di superamento e periodi di ritorno regionali, nonché di confrontare scenari e valutare il rischio idraulico con l'obiettivo di poter prevedere, per quanto possibile, quali aree della GAR saranno più vulnerabili ad eventi alluvionali estremi futuri.

Per raggiungere questo scopo sono stati prima individuati i massimi spaziali osservati e le corrispondenti portate massime alle diverse stazioni. Successivamente è stata eseguita una serie di simulazioni Monte Carlo finalizzate a generare un elevato numero (10000) di eventi spaziali partendo dalle caratteristiche univariate empiriche e dalla matrice di correlazione stimata.

È necessario passare da un'analisi univariata a una multivariata per poter descrivere gli eventi simulati quindi, per poter applicare agevolmente una dipendenza multivariata parametrica, le serie univariate sono state trasformate in modo da approssimare una distribuzione normale (nell'analisi si è assunto come distribuzione marginale di riferimento la Burr III che verrà descritta meglio in seguito). Tale trasformazione garantisce le proprietà di distribuzione richieste per l'impiego della distribuzione normale multivariata come strumento di modellizzazione della dipendenza.

Il procedimento seguito, riportato nel dettaglio nella trattazione Blöschl et al. (2022), può essere sintetizzato nei passaggi principali che verranno poi illustrati più approfonditamente nei sottoparagrafi seguenti:

- trasformazione delle serie univariate alla scala normale, partendo dalle stime delle marginali (Burr III);
- parametrizzazione della distribuzione normale multivariata coerente con la matrice di correlazione osservata fra siti;
- generazione, tramite Monte Carlo, di n valori normalmente distribuiti per ciascuna delle m serie mantenendo media, varianza e struttura di correlazione;

- back-transformation dei valori generati alla scala originale delle portate;
- determinazione delle probabilità di superamento e dei corrispondenti periodi di ritorno a scala locale e spaziale.

Nei paragrafi successivi verranno descritti nel dettaglio i criteri di stima delle distribuzioni marginali, la costruzione e la validazione della matrice di correlazione, le scelte numeriche relative al campionamento Monte Carlo e le procedure di verifica della bontà della simulazione (confronto fra statistiche osservate e simulate).

Infine, il capitolo conclude con l'impiego della popolazione simulata per la stima dei periodi di ritorno e per un'analisi critica dei limiti e delle incertezze associate al metodo.

4.1 Formazione di serie multivariate in idrologia.

Nell'ambito dell'idrologia, i fenomeni estremi, come le alluvioni fluviali, sono spesso il risultato di interazioni complesse tra diverse variabili come ad esempio precipitazioni, portata fluviale, volume degli afflussi, durata delle piogge, saturazione del suolo. Per comprenderne meglio la natura, è indispensabile adottare un approccio basato su serie multivariate, cioè sequenze temporali che descrivono congiuntamente più variabili idrologiche correlate.

Le serie univariate analizzano una sola dimensione del fenomeno, ad esempio i volumi di pioggia o le portate dei fiumi, ma questi dati isolati possono risultare fuorvianti se usati per descrivere eventi complessi. Le serie multivariate, invece, permettono di:

- catturare la correlazione tra variabili interdipendenti (es. picco di piena e durata dell'evento);
- studiare l'occorrenza congiunta di condizioni estreme;
- analizzare in modo più realistico il rischio idrogeologico.

Per poter applicare agevolmente un modello parametrico di dipendenza multivariata, è fondamentale individuare trasformazioni monotone che riportino le distribuzioni marginali su scala normale. In particolare, per la modellazione delle marginali è stata assunta la distribuzione Burr III, scelta per la sua elevata flessibilità e la capacità di rappresentare code pesanti tipiche delle portate estreme; le trasformazioni utilizzate consentono di "gaussianizzare" le marginali e di rendere quindi applicabile la distribuzione normale multivariata per la costruzione e la simulazione degli eventi simulati. Per la trasformazione in scala normale si applica un taglio superiore: i quantili simulati a livello locale vengono limitati

al valore massimo di piena locale stimato con la Pareto sui dati POT, evitando così simulazioni che eccedano i massimi fisicamente plausibili.

Di seguito viene riportata la formula della densità di probabilità della Burr III:

$$F(x) = \left(1 + \frac{1}{\gamma_1} \cdot \left(\frac{x}{\beta}\right)^{-\frac{1}{\gamma_2}}\right)^{-\gamma_1 \gamma_2} \quad (15)$$

```
dBrIII <- function (x, beta, gamma1, gamma2) {
  d <- 1/beta * (1 + 1/gamma1 * (x/beta)^(-1/gamma2))^(-gamma1*gamma2 - 1) * (x/beta)^(-1/gamma2 - 1)
  return(d)
}
pBrIII <- function (x, beta, gamma1, gamma2) {
  p <- (1 + 1/gamma1 * (x/beta)^(-1/gamma2))^(-gamma1*gamma2)
  return(p)
}
qBrIII <- function (p, beta, gamma1, gamma2) {
  q <- beta * (gamma1 * (p^(-1/(gamma1*gamma2)) - 1))^(-gamma2)
  return(q)
}
rBrIII <- function (n, beta, gamma1, gamma2) {
  p <- runif(n, min=0.000000001, max=0.999999999)
  x <- qBrIII(p, beta, gamma1, gamma2)
  return(x)
}
```

Figura 99- Funzioni che permettono di integrare la distribuzione Burr III nella trattazione.

Nella sezione Allegati viene riportato il codice (righe da 1512 a 1570) che ha lo scopo di stimare, per ciascuna stazione (colonna della matrice *events*), i parametri della legge marginale Burr III, quindi di combinare la CDF stimata della Burr III con una CDF di tipo Pareto/GPD (stimata in precedenza, parametri contenuti in *paramPareto_lmom*) e infine trasformare ogni osservazione in uno score normale $z=\Phi^{-1}(p)$ dove p è la probabilità cumulata risultante dalla scelta tra le due CDF. Questo produce una matrice *trnsevents* di valori sulla scala normale, pronta per utilizzare una dipendenza multivariata.

Grazie alla funzione, mostrata in figura 100, è possibile la scelta tra le due CDF: se per un dato x la CDF Pareto p_1 è maggiore di quella Burr p_2 , si usa p_1 . Usare il massimo permette di adottare la probabilità cumulativa che assegna a x il valore di probabilità più elevato, e conseguentemente il punteggio normale $z=\Phi^{-1}(p)$ sarà maggiore (più verso la coda destra) rispetto al caso in cui si usasse p_2 . Questo comportamento realizza una vera e propria "correzione della coda": la coda Pareto, stimata sui POT, prevale quando assegna una probabilità più alta agli x grandi, impedendo che la marginale Burr sottovaluti la rarità di grandi eventi (o viceversa, a seconda delle forme). Si sta scegliendo, per ogni x , la valutazione di probabilità che rende la trasformazione verso la scala normale più coerente con il modello di code GP dove questo è dominante.

```
pGPBrIIIImax <- function (x, x0, alpha, kappa, beta, gamma1, gamma2) {
  p1 <- 1 - (1 - kappa*(x - x0)/alpha)^(1/kappa) #CDF collegata al modello Pareto
  p2 <- (1 + 1/gamma1 * (x/beta)^(-1/gamma2))^(gamma1*gamma2) #CDF della Burr III
  apply(cbind(p1, p2), 1, max) #la probabilità cumulativa massima fra le due stime
}
```

Figura 100- Funzione che permette di combinare le due CDF (GP e Burr III).

Ora di seguito vengono mostrati tre grafici che mettono a confronto i parametri risultanti delle due distribuzioni protagoniste di questi paragrafi: GP e BurrIII. Essendo entrambe a tre parametri si mettono a confronto rispettivamente: x_0 e β (figura 101), α e γ_1 (figura 102), κ e γ_2 (figura 103).

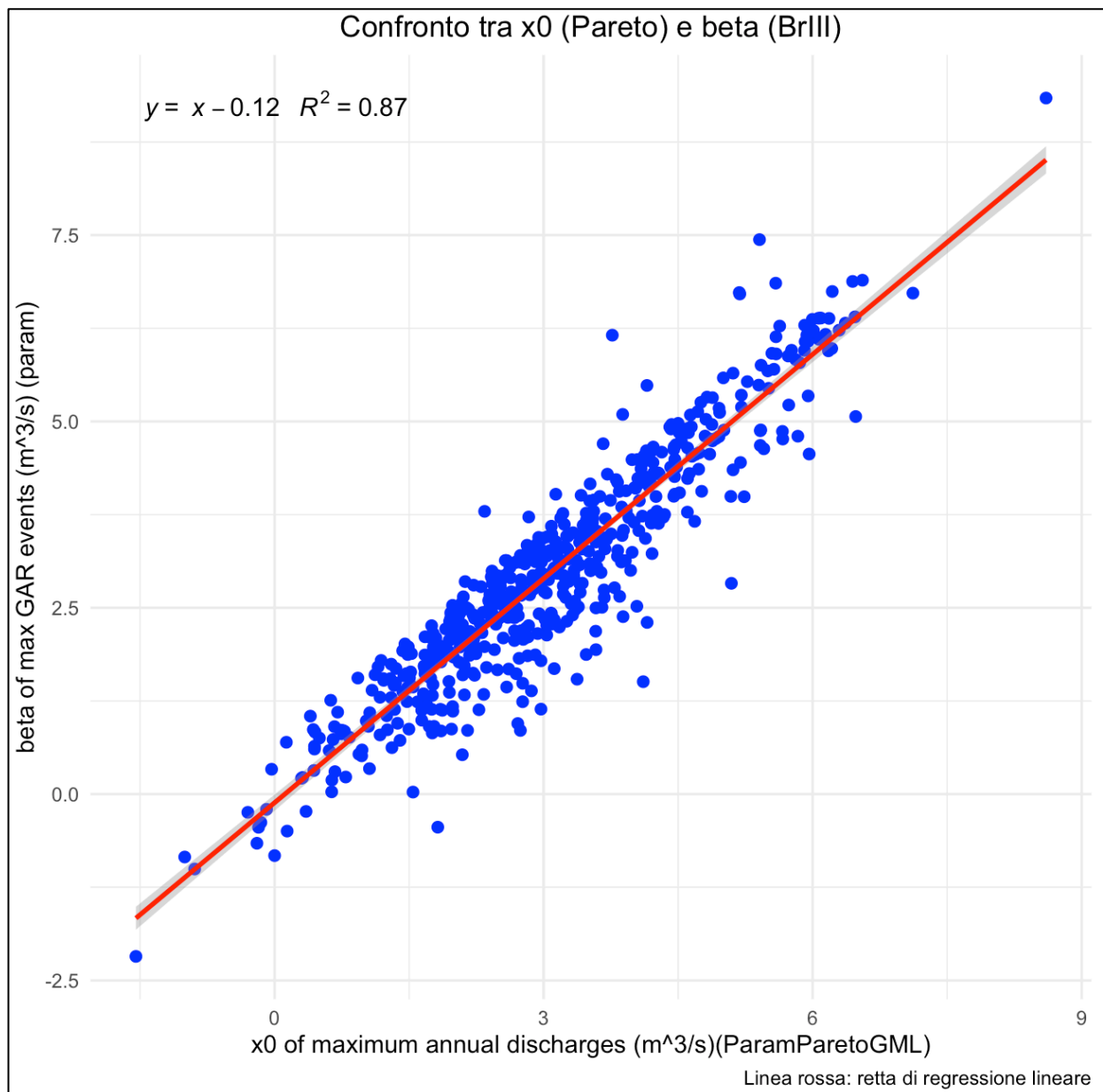


Figura 101- Confronto tra x_0 (GP) e β (BurrIII).

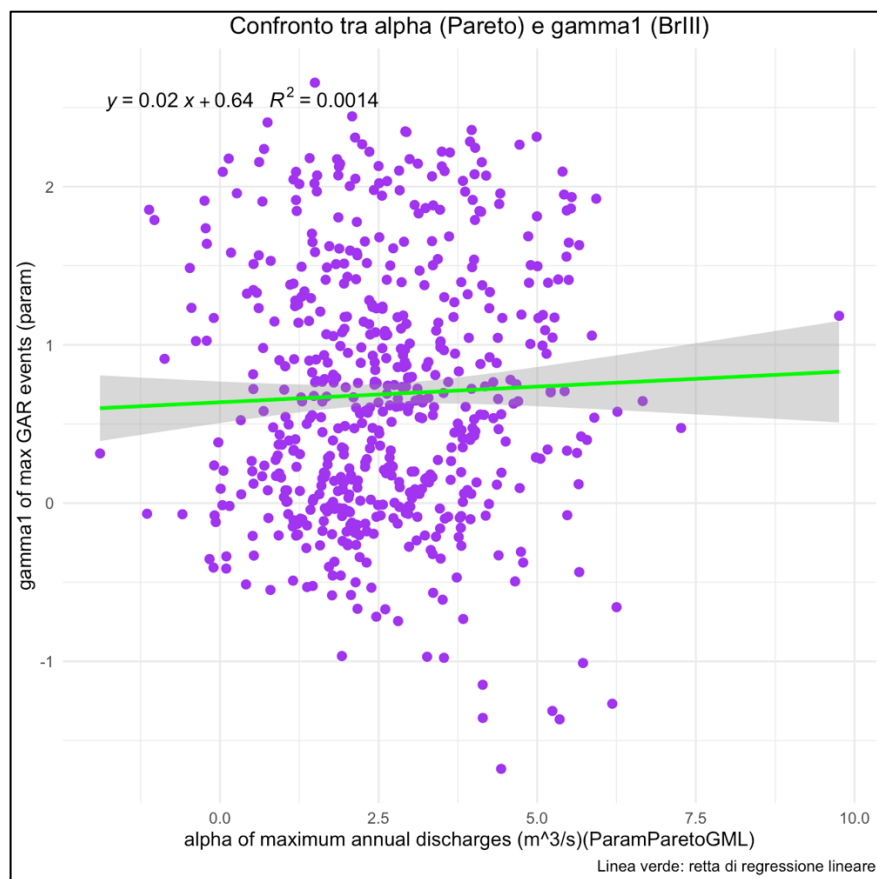


Figura 102- Confronto tra alpha (GP) e gamma1 (BurrIII).

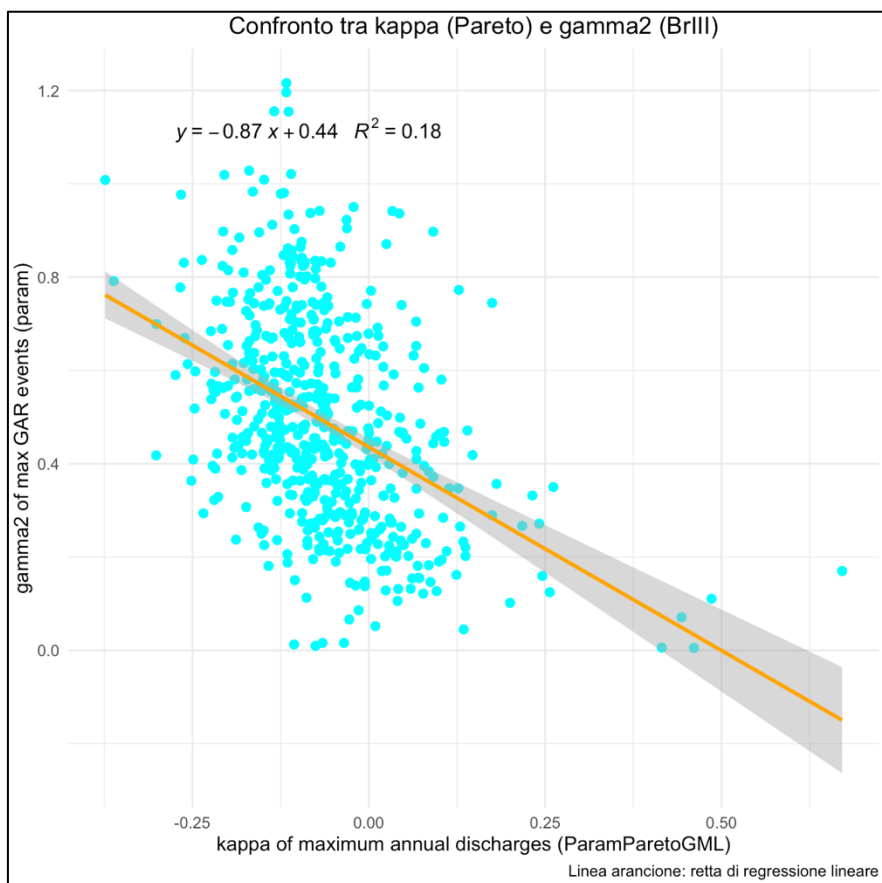


Figura 103- Confronto tra kappa (GP) e gamma2 (BurrIII).

Risulta evidente come ci sia una buona correlazione tra la prima coppia di parametri (parametri di posizione) mentre non ci sia affatto per le altre due coppie (parametri di scala e forma).

4.2 Parametrizzazione della distribuzione normale multivariata.

L'obiettivo principale di questo paragrafo è costruire una matrice di correlazione valida per la distribuzione normale multivariata a partire dai normal scores (*trnsevents*), risolvendo i problemi numerici che emergono dalla stima empirica (valori mancanti, valori non validi rumore, ecc.).

La correlazione è stata calcolata tramite la tecnica “correlazioni per coppia” (*pairwise.complete.obs*) al fine di sfruttare il maggior numero possibile di osservazioni non nulle: operazione che però impone correlazione zero dove l'informazione manca. Dopo aver effettuato questo calcolo si riscontra comunque che la matrice empirica risultante non sia definita positiva perché risultano molti autovalori negativi, condizione che impedisce l'impiego diretto in procedimenti che richiedono una matrice di covarianza valida.

Per rendere la matrice utilizzabile si è applicata la procedura numerica standard basata sull'algoritmo di Higham (2002): la funzione *nearPD()* (pacchetto *Matrix* in R) è stata usata per ricercare la matrice più prossima a quella ottenuta in precedenza che sia definita positiva. L'operazione di correzione ha modificato alcuni elementi, incluse le entrate diagonali: di conseguenza la matrice corretta è stata successivamente riscalata in modo da ottenere una matrice di correlazione propriamente detta (si è convertita la matrice di covarianza ottenuta inizialmente in una matrice di correlazione). A causa di imprecisioni numeriche e della presenza di autovalori molto piccoli, si sono osservate discrepanze aritmetiche minime (non simmetria numerica dell'ordine di 10^{-10}), risolte mediante un arrotondamento controllato. Il risultato della matrice di correlazione è proposto nella figura sottostante (figura 104).

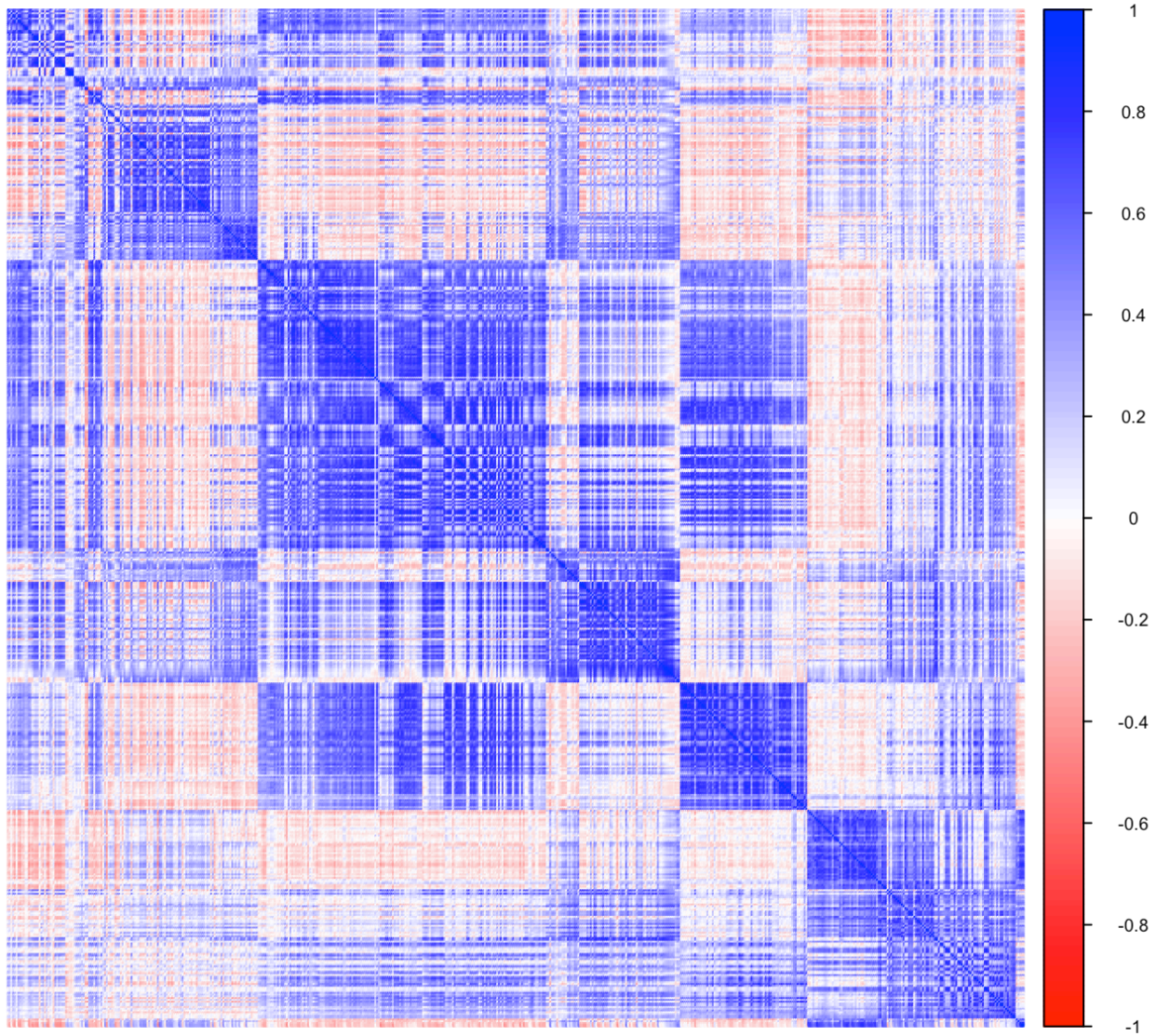


Figura 104- Grafico della correlazione per la distribuzione normale multivariata.

4.3 Simulazione di dati multivariati normali e back-transformation.

Se un vettore di dati y è composto da p variabili aleatorie congiuntamente distribuite, caratterizzate da un vettore dei valori attesi μ e da una matrice di varianza-covarianza Σ , allora la funzione di densità multivariata dei componenti del vettore è data da (Blöschl et al. (2022)):

$$f(y) = \frac{1}{(2\pi)^{p/2} \cdot \sqrt{|\Sigma|}} \exp \left[-\frac{1}{2} (y - \mu) \cdot \Sigma^{-1} \cdot (y - \mu)^t \right] \quad (16)$$

Nell'ambito delle simulazioni Monte Carlo, l'utilizzo della distribuzione multinormale risulta particolarmente conveniente poiché richiede soltanto due elementi fondamentali: il vettore dei valori attesi μ , che descrive la posizione media delle variabili considerate, e la matrice di

varianza-covarianza Σ , che ne rappresenta la struttura di dipendenza congiunta. Tali parametri permettono di generare campioni che rispettano sia le caratteristiche marginali delle distribuzioni che le correlazioni osservate tra le variabili, garantendo così una rappresentazione coerente degli eventi simulati. In particolare, per generare la popolazione simulata di eventi del caso studio si è proceduto nel seguente modo.

A partire dalla matrice di correlazione stimata sui *normal scores*, *Strnsevents*, si è simulata una variabile aleatoria multivariata normale con media nulla e covarianza *Strnsevents* mediante la funzione *mvrnorm* della libreria *MASS* in R, per un totale di $N=ev_yr \times NYRsim$ realizzazioni, con $NYRsim = 10000$: si è quindi ottenuto un totale di 20000 simulazioni avendo impostato l'analisi su due eventi all'anno. Le probabilità marginali per ciascuna componente sono state ottenute col legame normale inverso $p=\Phi(z)$ e convertite in valori sulla scala originale delle portate tramite i quantili inversi delle due leggi marginali considerate: la Generalized Pareto stimata sui dati POT e la Burr III stimata per la marginale. Per ogni probabilità p si è scelto il valore:

$$x(p) = \min (F_{Pareto}^{-1}(p), F_{BurrIII}^{-1}(p)) \quad (17)$$

ovvero il più piccolo fra i due quantili a pari probabilità, in modo da garantire coerenza con la strategia adottata nella fase di trasformazione diretta (dove si era impiegata la corrispondente regola sulle CDF). La matrice *simevents* così ottenuta rappresenta la popolazione simulata di portate su cui sono poi calcolate probabilità di superamento e tempi di ritorno regionali.

4.4 Verifica del modello multivariato.

Per valutare la coerenza tra i dati osservati e quelli generati dal modello, sono state confrontate le principali statistiche descrittive delle serie di portata: media, deviazione standard e 90° percentile. A partire dalle simulazioni Monte Carlo, le stime di queste grandezze sono state calcolate su blocchi equivalenti alla lunghezza delle osservazioni, ottenendo distribuzioni empiriche e relativi intervalli di confidenza al 95%, un esempio viene riportato nella figura sottostante (figura 105).

	site	mean_obs	mean_lo	mean_hi	sd_obs	sd_lo	sd_hi	q90_obs	q90_lo	q90_hi
141	1	3.5196417	3.2601313	3.9082450	3.0678528	2.5990619	3.5568702	6.8444	6.3021392	8.1078394
155	2	5.3833911	5.0352211	6.0644945	4.2283306	3.7700393	4.4605490	10.6541	10.1576088	11.9767796
13	3	8.9774635	8.3830291	9.8375933	6.9976254	6.2827358	8.4111067	19.5264	15.3232861	20.8699227
95	4	20.0364286	18.8021506	21.7149516	13.5537709	11.8461946	14.3303216	37.6388	34.7512462	40.6756144
84	5	53.4588000	49.8798141	57.6965152	39.0425971	33.6352488	44.8142759	101.1054	92.2556505	109.5948258
10	6	36.5699867	34.6478471	40.4450071	28.0983337	25.7446996	32.7281217	71.5670	65.6526665	79.7637832
41	7	86.5117333	82.6387822	92.7144673	53.6760340	44.8079414	60.8202681	132.6202	137.7492604	163.1820510
80	8	60.2162833	58.4620044	66.0631107	43.4355595	38.4838279	49.9108384	117.6983	105.1303536	126.0150453
77	9	30.9042961	29.2676944	32.9770610	22.0683107	19.7437527	24.9351452	57.0772	51.8730186	63.6631025
101	10	69.3197833	62.3347810	76.5863884	59.4536589	53.9648195	71.3079441	154.5333	122.5198502	175.5143324
21	11	100.7225167	80.9679923	102.1257274	101.4180905	55.4482477	124.8799090	197.7083	128.6742855	188.0536519
8	12	145.5133833	122.7249947	155.7042630	133.8343392	83.2589765	175.9095054	263.0073	210.6401442	283.1135160
31	13	224.8925167	217.5196591	240.4947012	119.9398056	107.6414965	134.7452258	385.5612	354.9017108	400.4971501
7	14	225.9562924	218.3353018	237.9459438	105.3941961	94.2563306	108.2671534	349.0000	334.5966636	372.0596603

Figura 105- Esempio per i primi 14 siti della GAR: si riportano rispettivamente per ogni sito media dei dati osservati, media del quantile 2.5% dei dati simulati (mean_lo), media del quantile 97.5% dei dati simulati (mean_hi). Allo stesso modo per le deviazioni standard e 90° percentile.

Tali risultati sono stati poi confrontati con i valori osservati mediante rappresentazioni grafiche, che consentono di verificare visivamente l'aderenza del modello alle osservazioni (figure che verranno mostrate nella pagina seguente).

Successivamente, è stato calcolato il bias relativo tra simulato e osservato per ciascun sito e per ciascun indice, in modo da quantificare eventuali sovrastime o sottostime sistematiche. Di seguito viene riportata l'equazione che permette di calcolarli:

$$bias = \frac{\text{valore simulato mediano} - \text{valore osservato}}{\text{valore osservato}} \quad (18)$$

Vengono quindi creati tre grafici:

- bias relativo della media;
- bias relativo della deviazione standard;
- bias relativo del 90° percentile.

Infine, per valutare la capacità del modello di preservare la struttura di dipendenza spaziale, sono state confrontate le correlazioni di Pearson e di Spearman tra le serie osservate e simulate, considerando tutte le coppie di stazioni (ciò verrà approfondito in seguito).

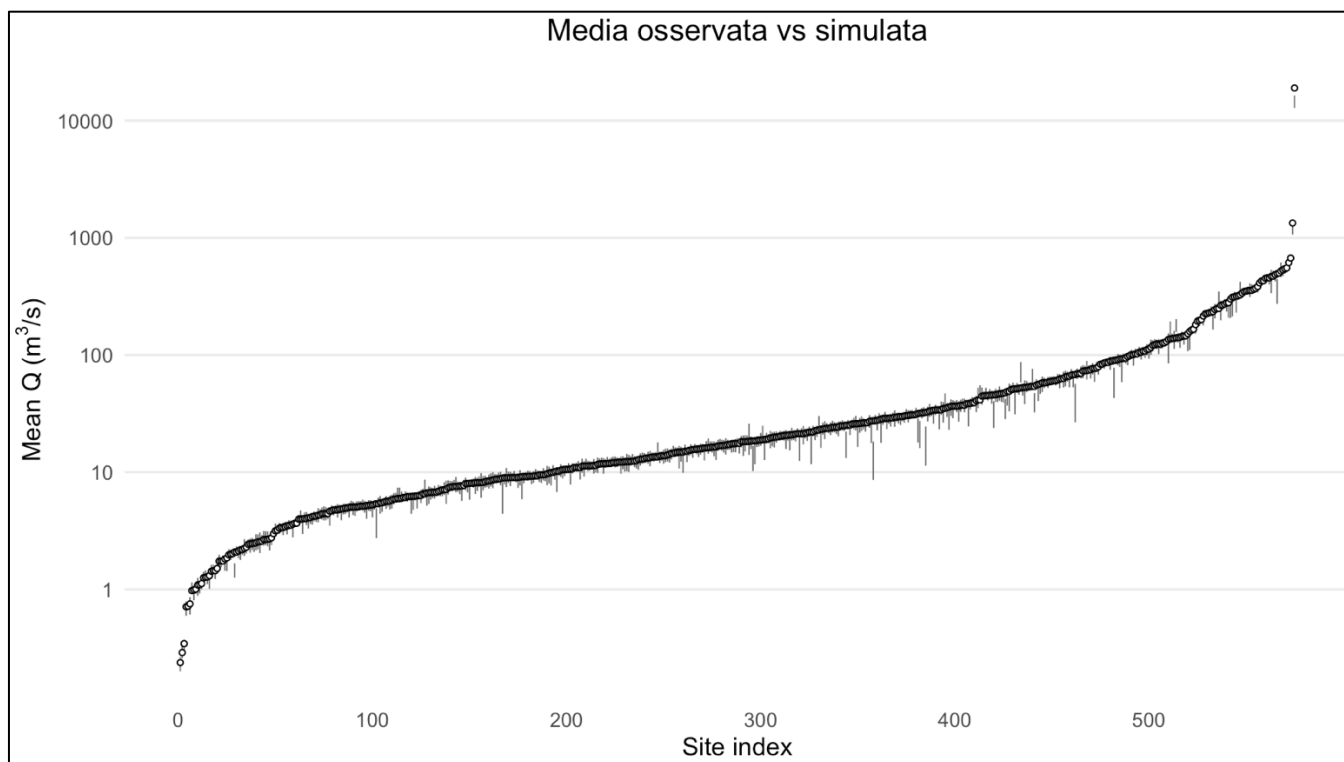


Figura 106- Grafico della media osservata confrontata con la media simulata (con intervalli di confidenza).

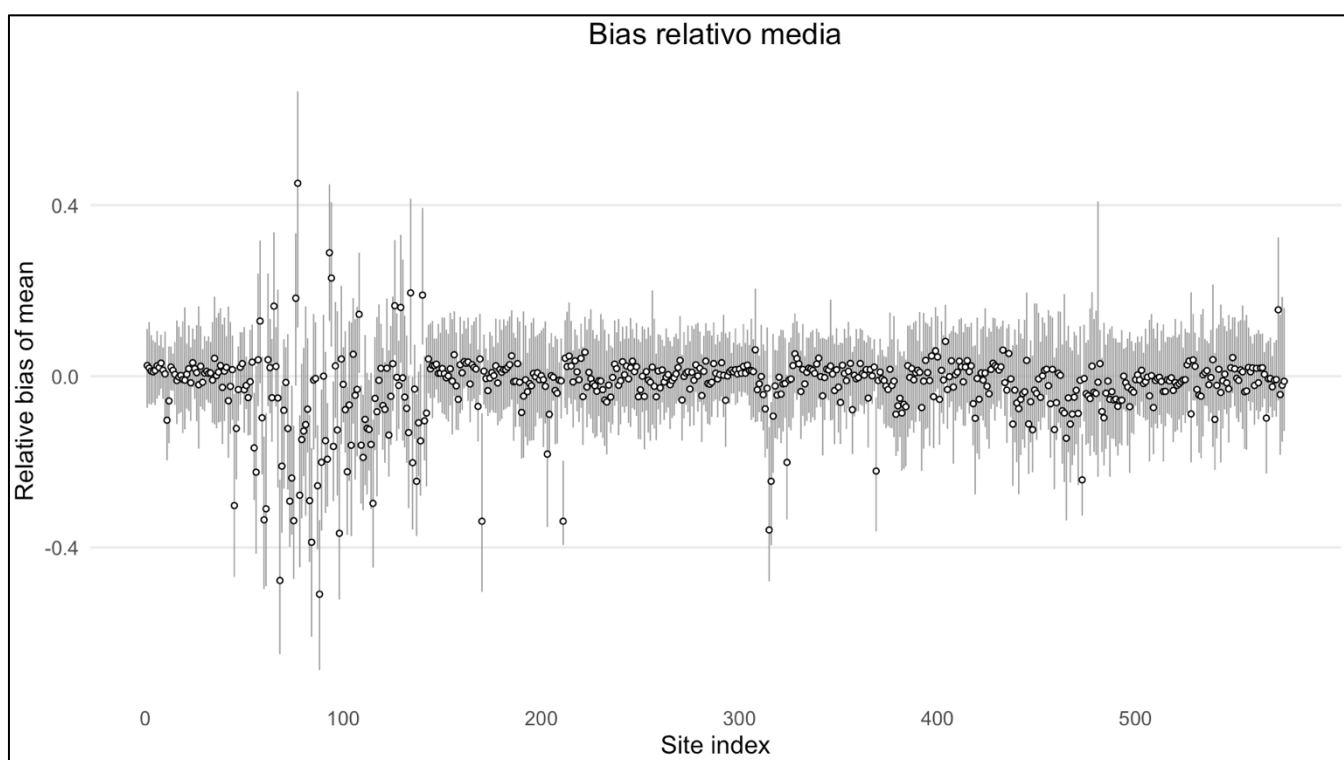


Figura 107- Grafico del bias relativo alla media.

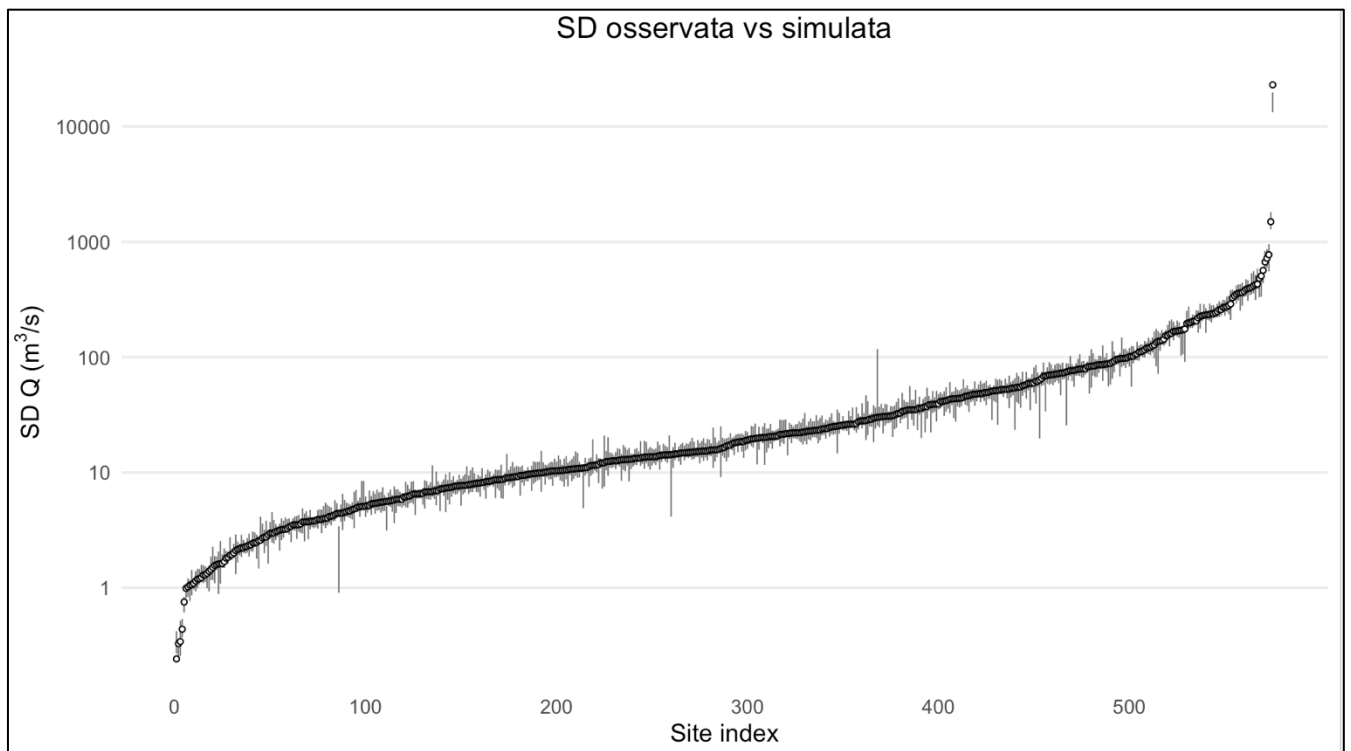


Figura 108- Grafico della SD osservata confrontata con la SD simulata (con intervalli di confidenza).

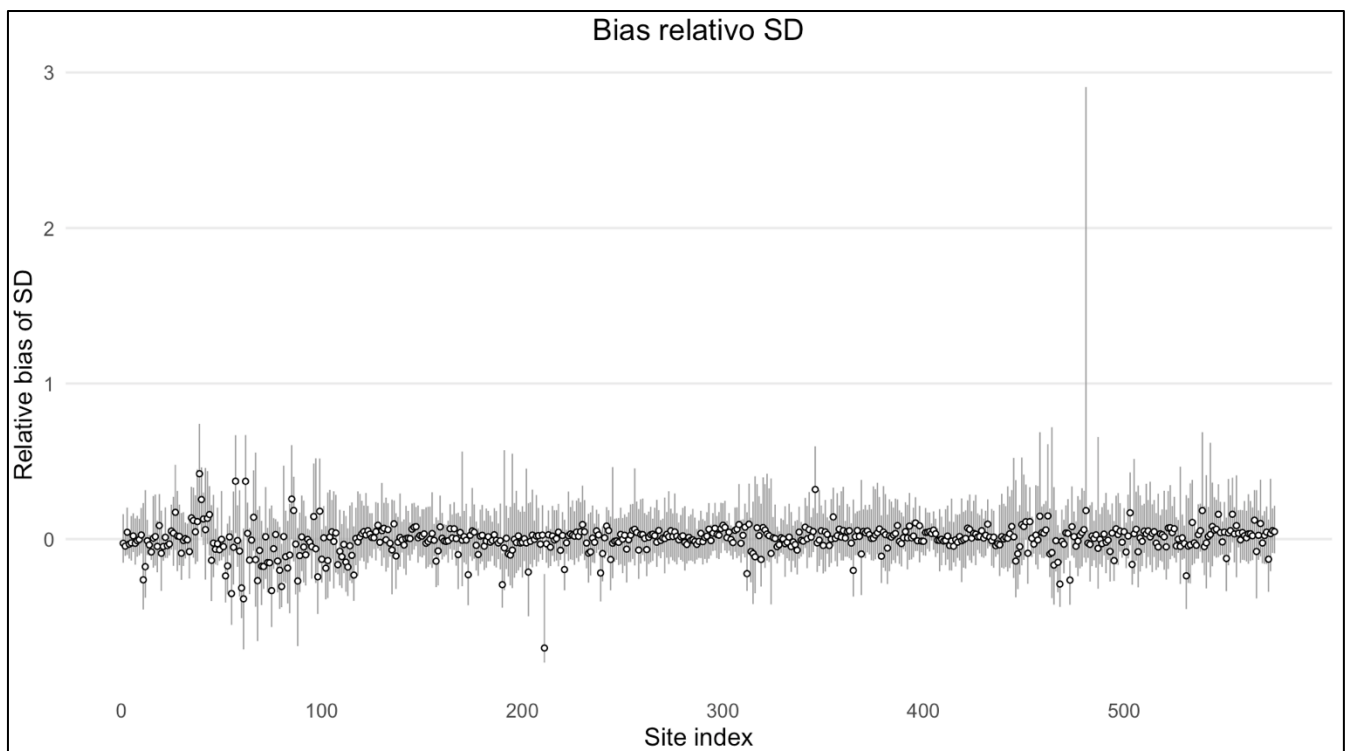


Figura 109- Grafico del bias relativo alla SD.

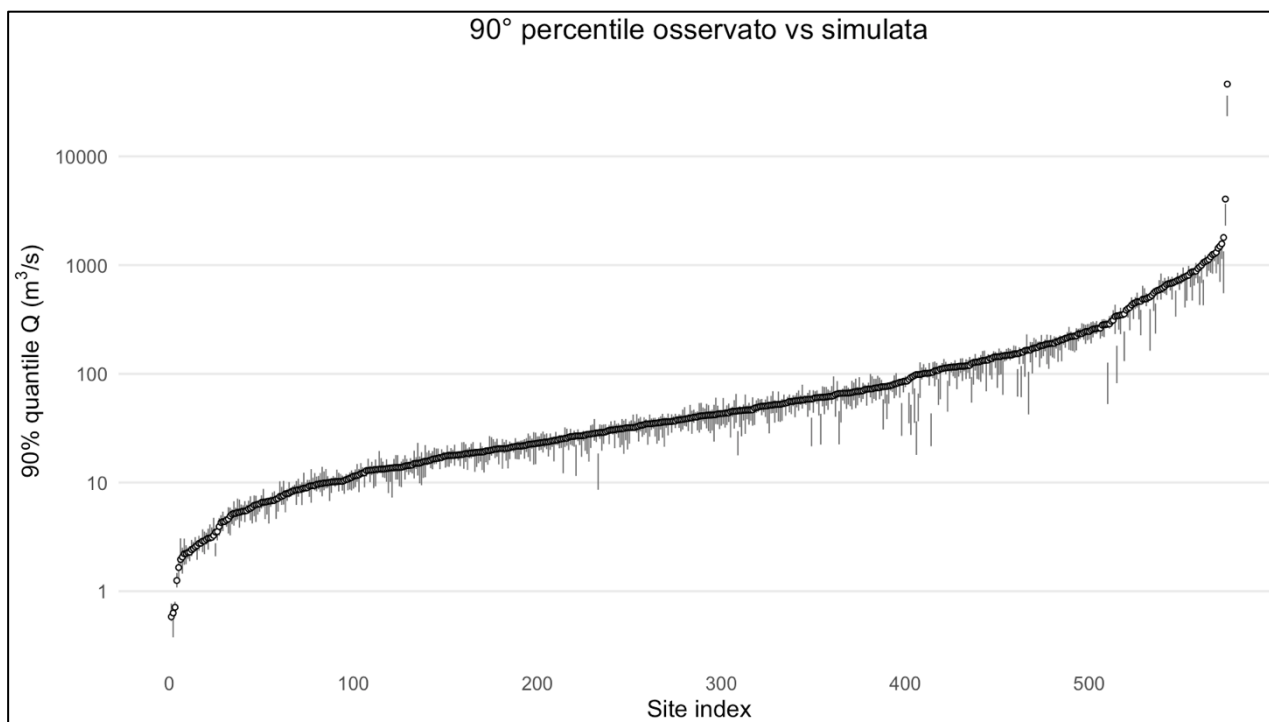


Figura 110 - Grafico del 90° percentile osservato confrontato con il 90° percentile simulato (con intervalli di confidenza).

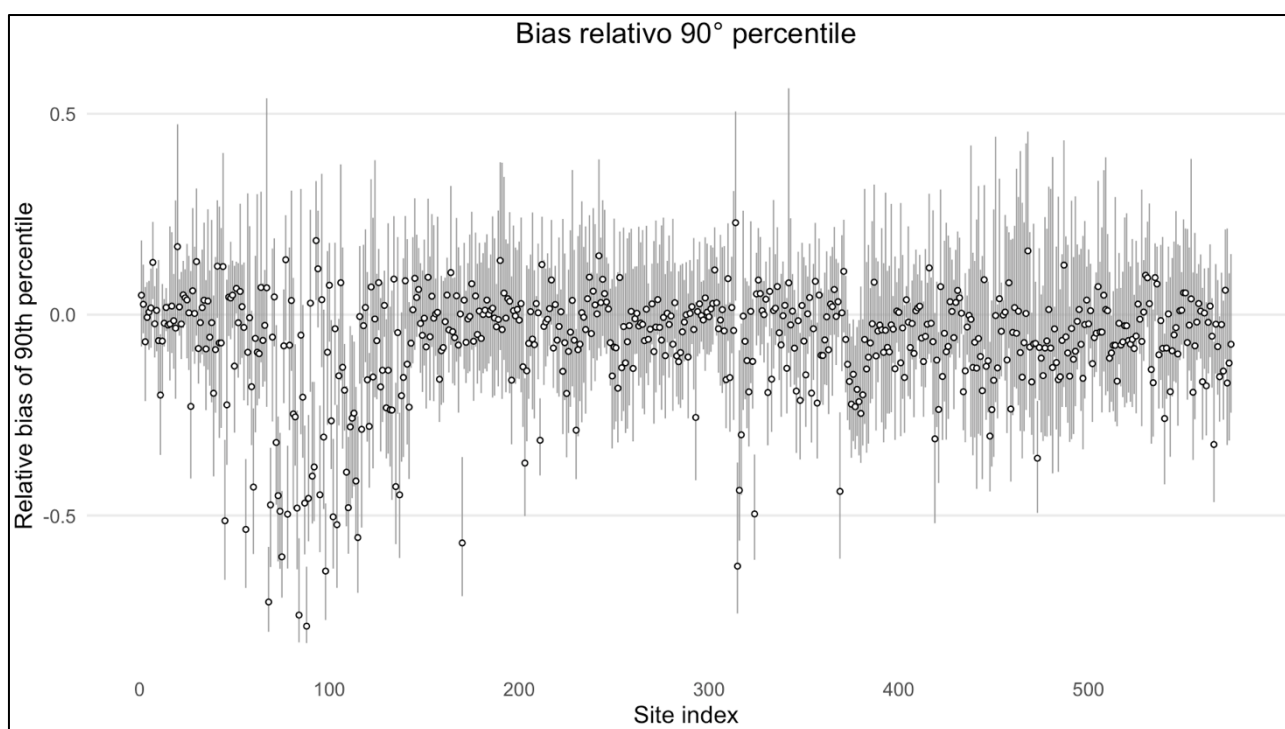


Figura 111- Grafico del bias relativo al 90° percentile.

Dai grafici riportati (figure 106-111) è possibile notare come la quasi totalità delle stime simulate abbia un bias compreso tra -0.5 e 0.5 per quanto riguarda la media e il 90° percentile mentre per la SD esiste un caso anomalo che esce da questo intervallo di valori. Dato che si tratta comunque di un caso su 575 si ritiene il modello robusto.

Come anticipato in precedenza, per verificare che la procedura di simulazione preservi la struttura di dipendenza spaziale tra le stazioni, sono stati confrontati i coefficienti di correlazione calcolati sulle serie osservate e su quelle simulate. Per ciascuna coppia di stazioni si sono ottenuti i coefficienti di Pearson (misura della dipendenza lineare) e di Spearman (misura della dipendenza monotona basata sui ranghi): dalle matrici di correlazione si è estratta la metà inferiore non ridondante e sono stati tracciati degli scatter-plot che riportano r_{obs} sull'asse delle ascisse e r_{sim} su quello delle ordinate. La retta tratteggiata in rosso è utilizzata come riferimento: punti vicini a quest'ultima indicano una buona riproduzione della correlazione. Per un giudizio quantitativo sono stati riportati il coefficiente di correlazione tra i vettori r_{obs} e r_{sim} , il valore RMSE e il bias medio. La contemporanea analisi di Pearson e Spearman consente di distinguere fra errori relativi all'intensità lineare della dipendenza ed errori relativi all'ordine (ranking) delle associazioni tra stazioni.

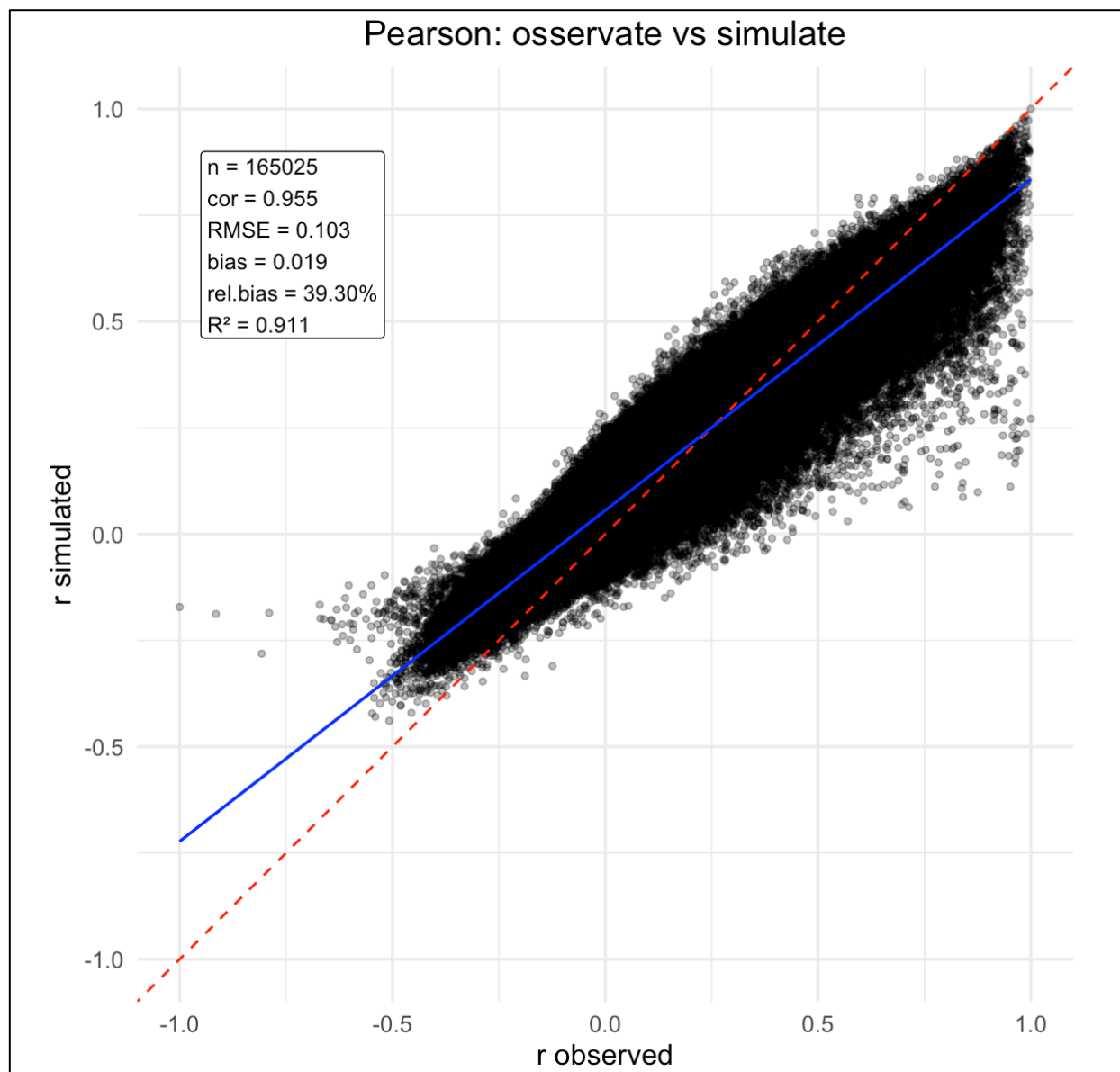


Figura 112- Grafico della correlazione dei dati osservati e simulati con Pearson.

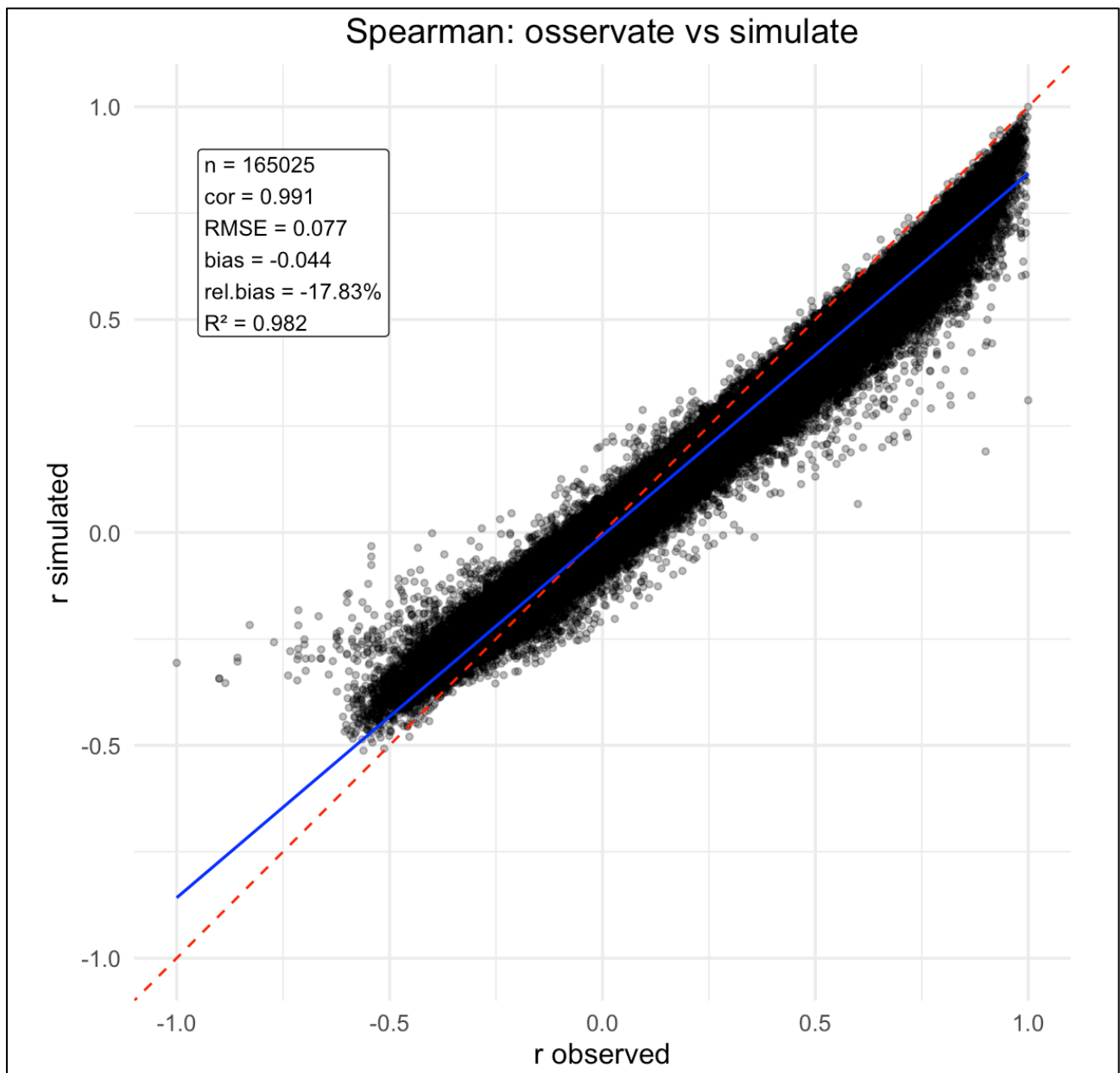


Figura 113- Grafico della correlazione dei dati osservati e simulati con Spearman.

Nei due grafici vengono riportati anche numero di coppie analizzate (n), correlazione tra vettori obs e sim e il coefficiente R^2 della regressione lineare sim- obs.

Globalmente, i risultati indicano che la procedura proposta è appropriata per la riproduzione della struttura di probabilità multivariata. A supporto di questa affermazione vi è la consistente corrispondenza osservata sia per le medie che per i quantili al 90% e per i coefficienti di correlazione di Pearson e Spearman.

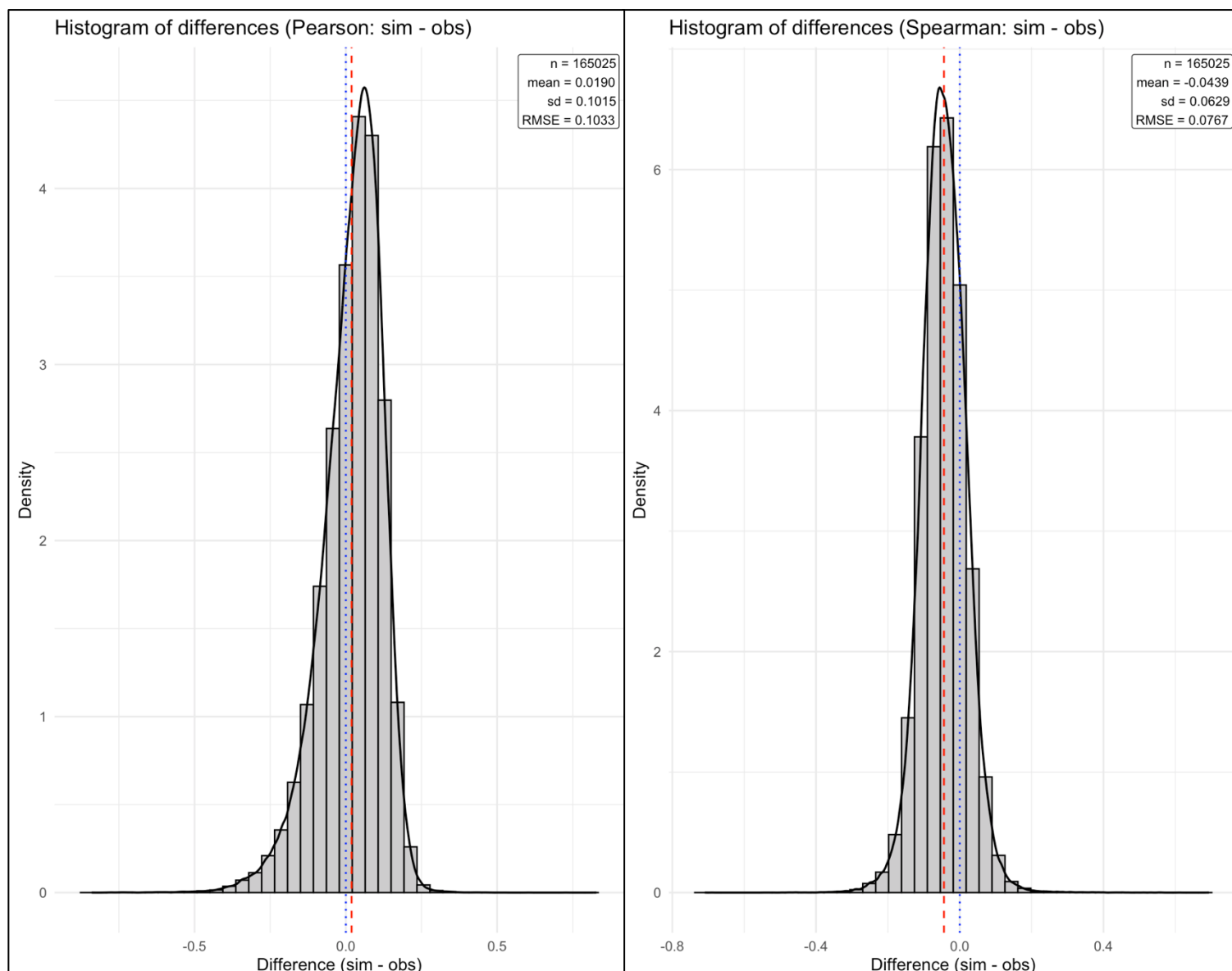


Figura 114- Istogramma delle differenze (sim – obs) per Pearson & Spearman.

L'istogramma delle differenze (sim - obs) rivela la distribuzione degli errori di correlazione:

- la media delle differenze (in rosso) vicina a zero indica assenza di bias sistematico nelle correlazioni simulate;
- la larghezza (SD) relativamente piccola indica che le correlazioni simulate sono costantemente vicine a quelle osservate;
- leggere asimmetrie segnalano che il modello tende rispettivamente a sovra-stimare per la Pearson e sotto-stimare per Spearman;

Le linee verticali in rosso rappresentano la media delle differenze, in blu la linea tratteggiata rappresenta lo zero (nessun errore). Si evince dai grafici in figura che la media delle differenze è in entrambi i casi prossima a zero a conferma dell'idoneità della procedura proposta.

4.5 Eventi alluvionali simulati su scala internazionale (GAR).

Una volta validato il modello multiparametrico si procede generando gli eventi alluvionali simulati sulla GAR con una procedura molto simile a quella impiegata per estrarre gli eventi reali dalle serie storiche di portate giornaliere riportata nel capitolo 3.2.

Per ogni evento simulato e per ciascun sito si calcola la probabilità cumulata marginale $P=F(x)$ utilizzando la funzione CDF Pareto stimata sui dati POT. Poi sulla base di P si ottengono i periodi di ritorno locali mediante $T=1/(\lambda(1-P))$, dove λ è il numero medio di eventi per anno considerati. I periodi di ritorno vengono trasformati, tramite la funzione *ramp* (identica a quella impiegata per i dati osservati) applicata a $\log(T)$, in un indice di danno locale D . Successivamente si costruisce per ciascun evento un indice regionale aggregato D_m come media pesata dei danni locali elevati alla potenza p e pesati per il logaritmo naturale dell'area. Infine, ordinando gli indici aggregati e associandoli ai ranghi empirici, si ricavano i corrispondenti periodi di ritorno a scala regionale.

I quantili e i tempi di ritorno così ottenuti sono impiegati per caratterizzare la frequenza e la gravità degli eventi simulati su ampia scala.

Sono state nuovamente create le seguenti matrici, questa volta con la dicitura “_sim”:

- P_u_sim = per ogni stazione “j” il codice R trasforma le portate simulate in probabilità di non superamento usando la funzione *pGP()*. Quindi, P_u_sim è la matrice contenente le probabilità cumulative marginali per ogni evento simulato e ogni sito;
- T_n_sim = contiene il periodo di ritorno (in anni) per ogni valore di probabilità simulato grazie alla seguente relazione:

$$T = \frac{1}{ev_x_yr \cdot (1 - P)}$$

con a denominatore il prodotto che coincide con la probabilità di superamento per anno, il cui inverso è il periodo di ritorno misurato in anni;

- D_n_sim = è la matrice dei danni stimati per ogni evento simulato e per ogni sito. Essa è ottenuta grazie alla funzione *ramp* che è la relazione rischio-danno che trasforma la rarità di un evento (T) in gravità di danno atteso su quel sito;
- *aree* = viene creato un vettore contenente tutte le aree dei bacini della GAR con il fine di utilizzare quest'ultime come pesi nell'analisi spaziale. L'uso di $\log(aree)$ come pesi dà più peso a siti con area maggiore, ma in modo attenuato rispetto ad usare l'area diretta;

- D_m_{sim} = contiene il danno aggregato spazialmente per ciascun evento simulato: serve a confrontare eventi a livello regionale. Viene impiegato per caratterizzare la frequenza e la gravità degli eventi simulati.

Si è deciso inizialmente di fare un esperimento di visualizzazione degli eventi di piena simulati, su eventi random ovvero senza una caratteristica particolare e vengono proposti di seguito alcuni esempi (il ciclo for creato permette di plottare 15 eventi casuali ma ne vengono riportati solo alcuni di questi per brevità):

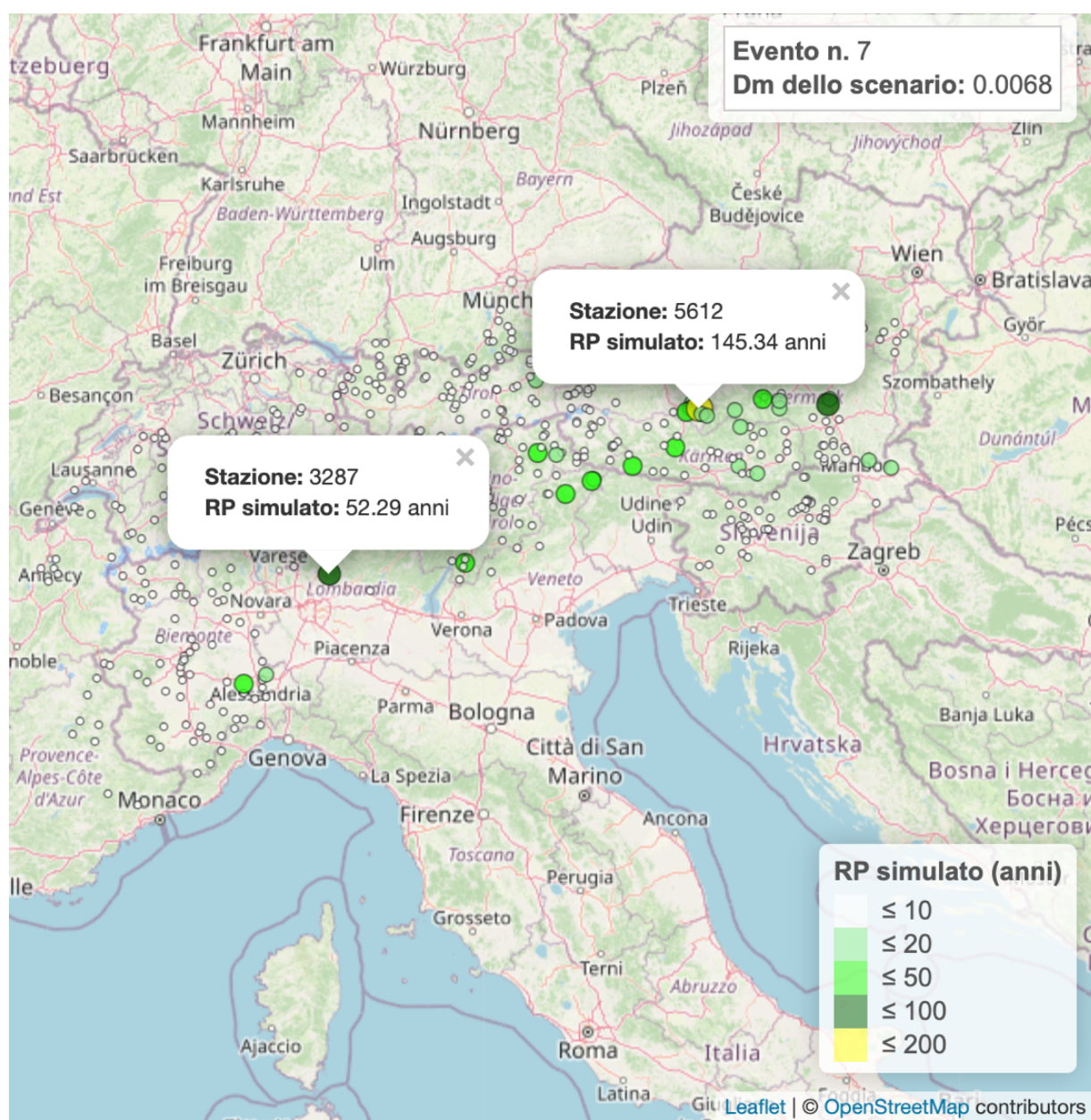


Figura 115- Mappa leaflet di un evento simulato sulla GAR.



Figura 116-Mappa leaflet che mostra un evento simulato in cui tutte le stazioni presentano tempo di ritorno minore di 10 anni. L'indice di danno risultante è pari a 0.

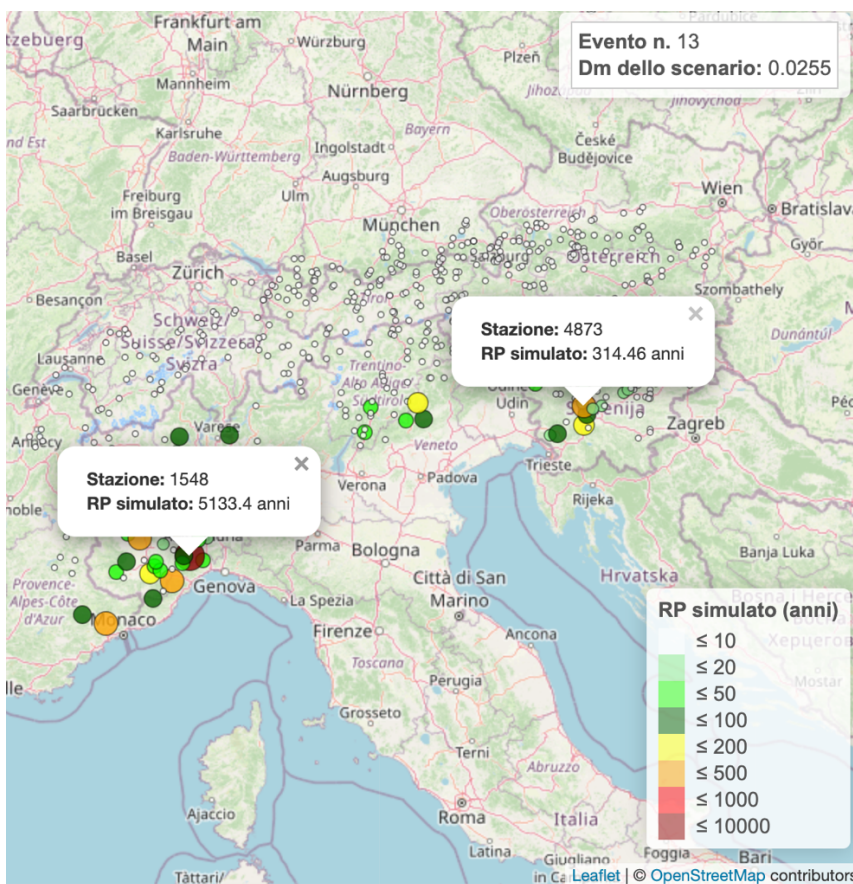


Figura 117- Mappa leaflet che mostra un evento simulato in cui la Liguria e il Piemonte sono coinvolti da un evento estremo con tempo di ritorno decisamente elevato. La Slovenia è simultaneamente coinvolta con tempi di ritorno dell'ordine dei trecento anni.

Per quanto riguarda la distribuzione spaziale ogni evento simulato mostra un pattern diverso delle portate alle stazioni. Alcuni eventi appaiono più localizzati (con RP elevati solo in poche aree), mentre altri hanno un'estensione più ampia, con alte ricorrenze in più bacini contemporaneamente. Questa variabilità riflette la natura stocastica del modello multivariato, che non riproduce un singolo evento storico ma un ventaglio plausibile di scenari futuri.

Da notare che le classi di tempo di ritorno (RP) sono state create in modo che la legenda consentisse di distinguere rapidamente le zone colpite da eventi rari e potenzialmente catastrofici. Le stazioni, infatti, sono classificate in base al RP simulato, con una determinata codifica del colore e dimensione crescente del marker. I colori vanno dal verde (eventi più frequenti, ≤ 20 anni) al rosso/viola (eventi estremi, > 10000 anni). In più ogni mappa, in alto a destra, riporta il valore D_m che rappresenta una misura aggregata dell'intensità dell'evento a scala regionale. Gli eventi con D_m più elevato corrispondono a scenari con più stazioni simultaneamente esposte a piene rare, mentre quelli con lo stesso indice basso sono caratterizzati da eventi più localizzati o meno intensi.

Successivamente si è deciso di plottare gli eventi simulati più gravosi, ovvero con indice di magnitudo aggregato più alto: si tratta di eventi catastrofici che fortunatamente hanno tempi di ritorno dell'ordine di 20000 anni. Vengono proposte le mappe leaflet che li descrivono subito di seguito:

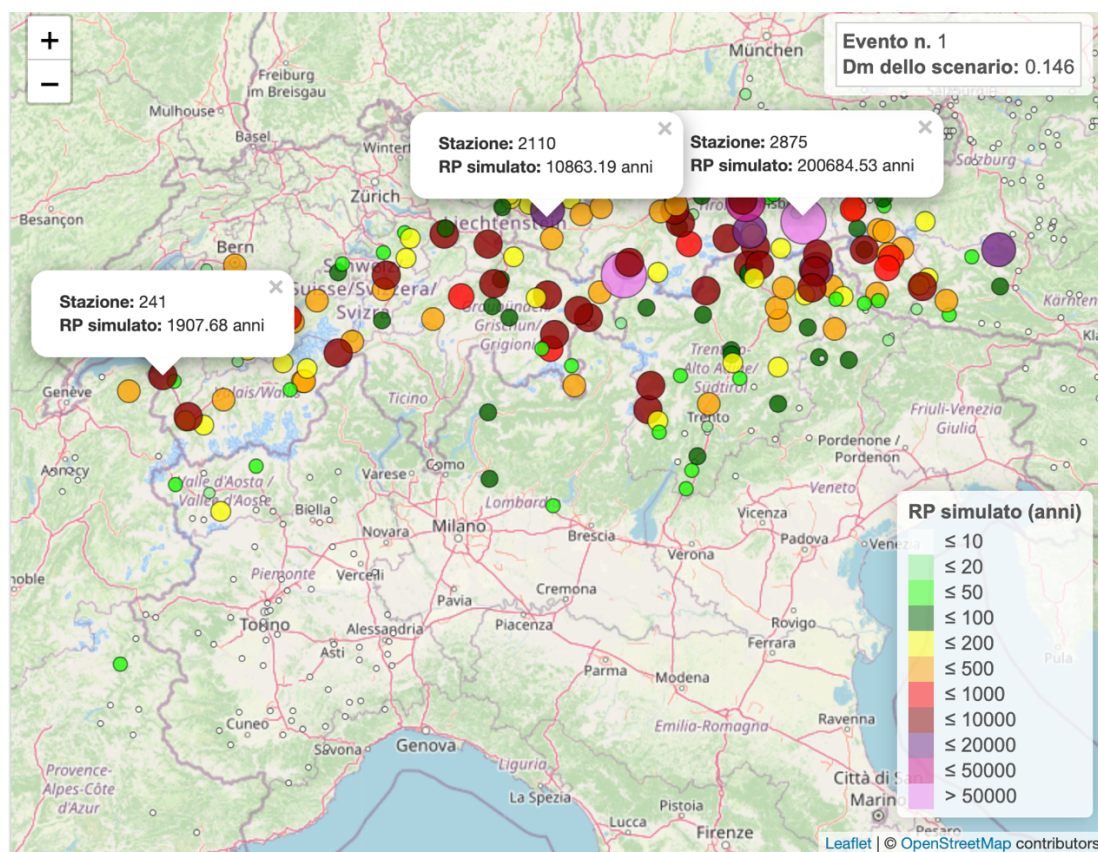


Figura 118- Mappa leaflet di uno dei 15 eventi simulati più gravosi.

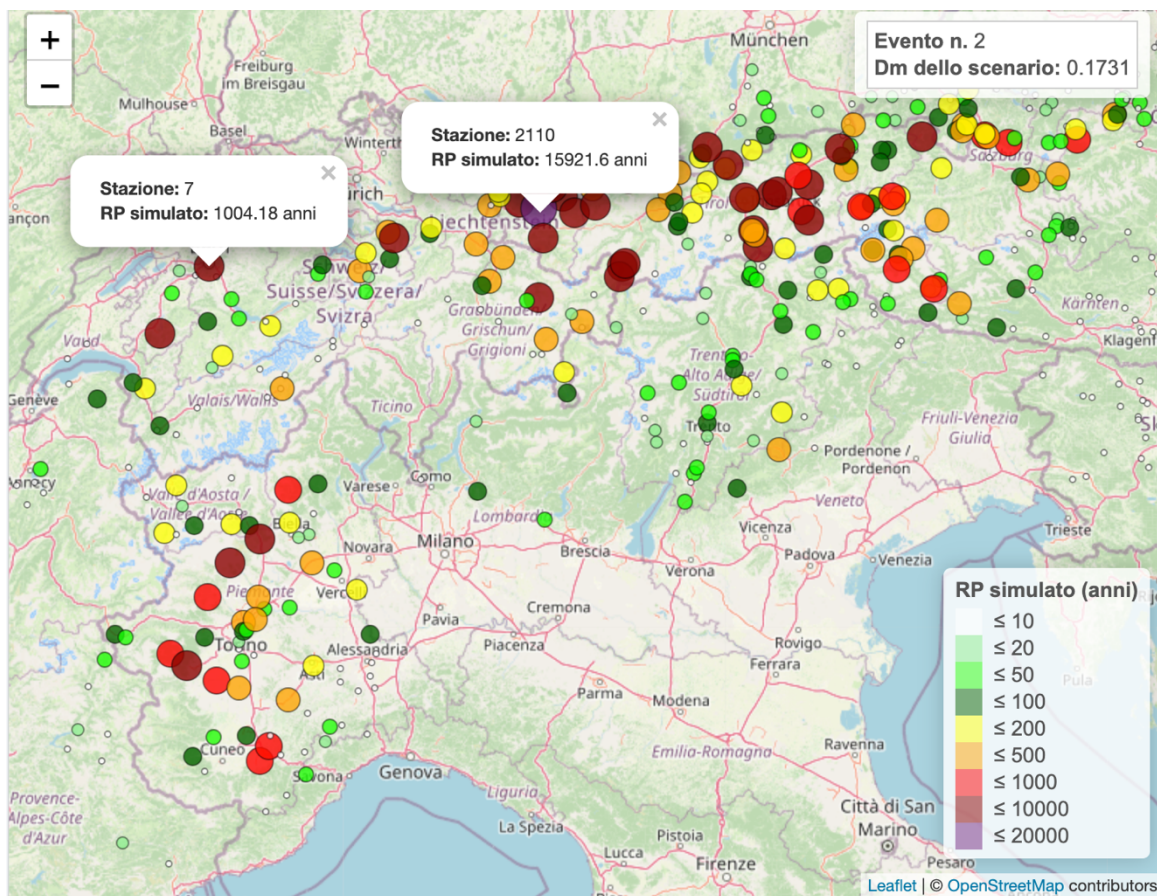


Figura 119- Mappa leaflet di uno dei 15 eventi simulati più gravosi.

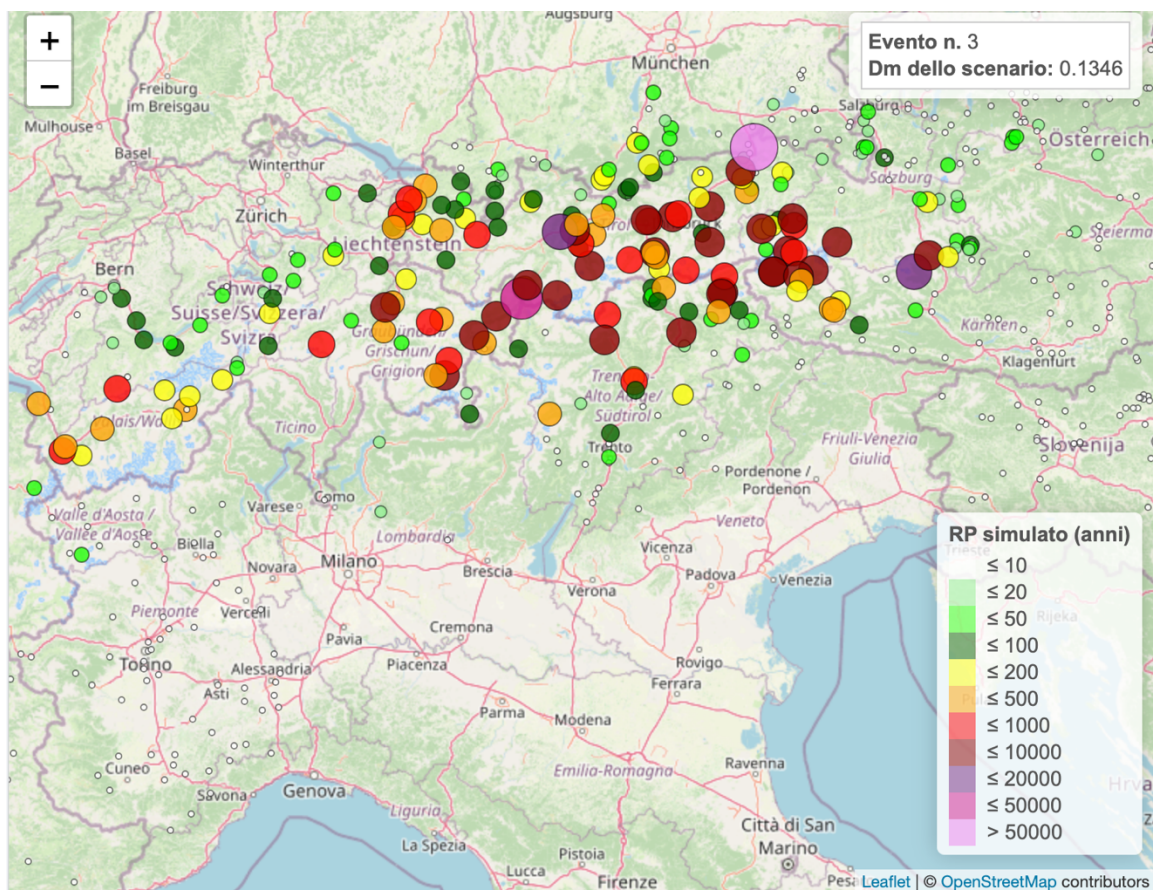


Figura 120- Mappa leaflet di uno dei 15 eventi simulati più gravosi.

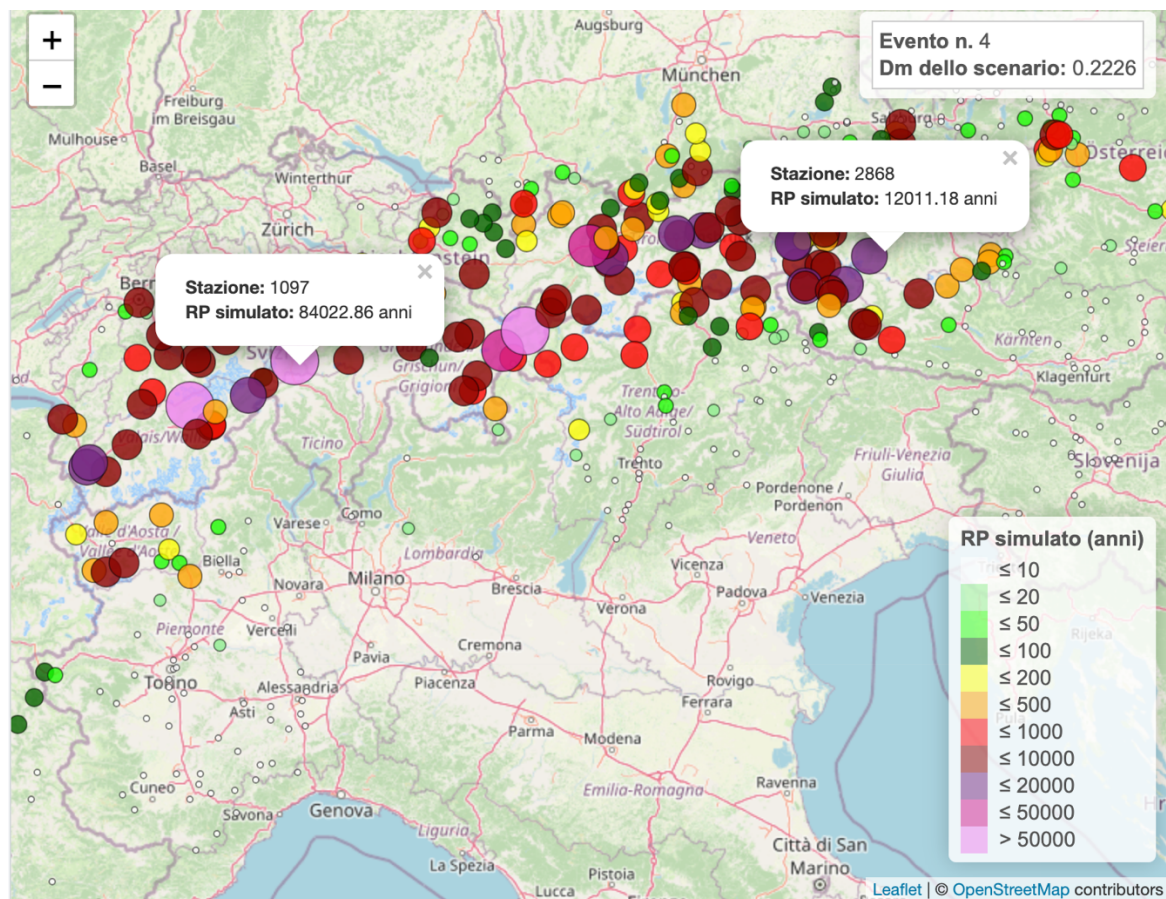


Figura 121- Mappa leaflet di uno dei 15 eventi simulati più gravosi.

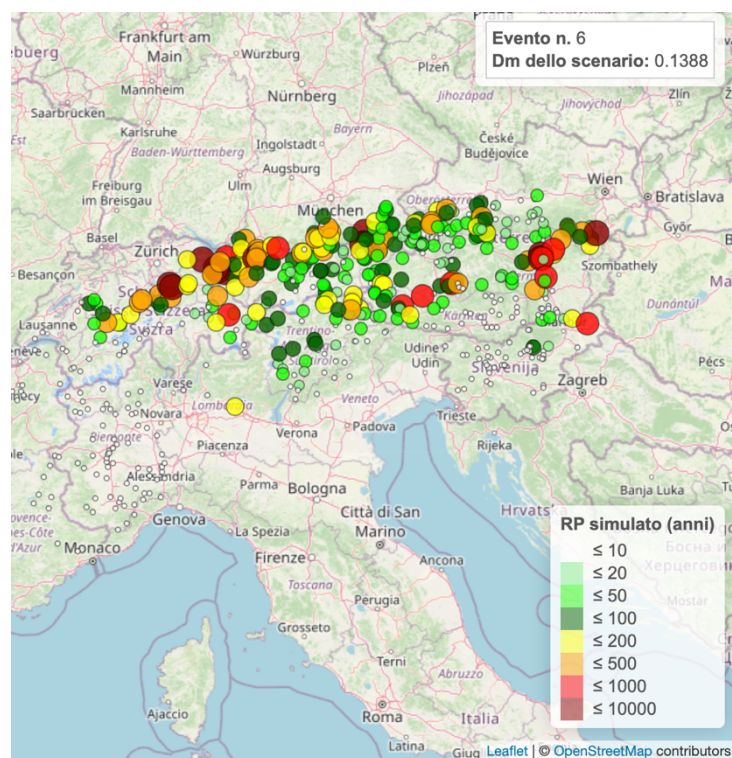
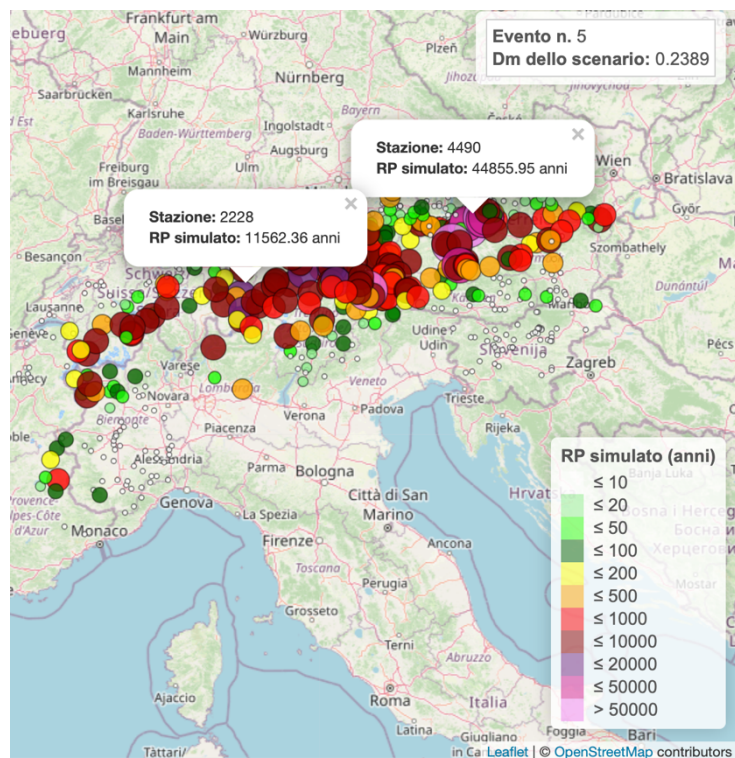


Figura 122- Mappe leaflet di due dei 15 eventi simulati più gravosi.

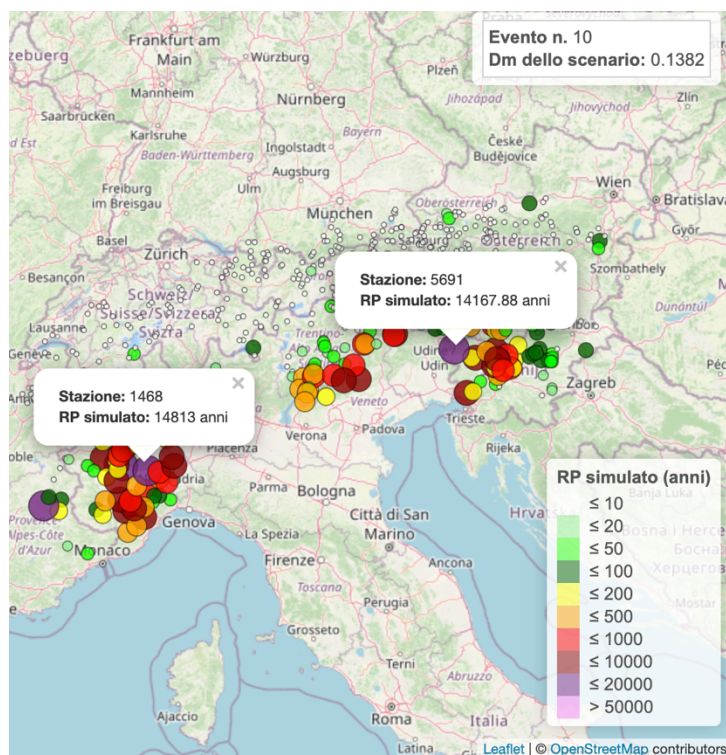
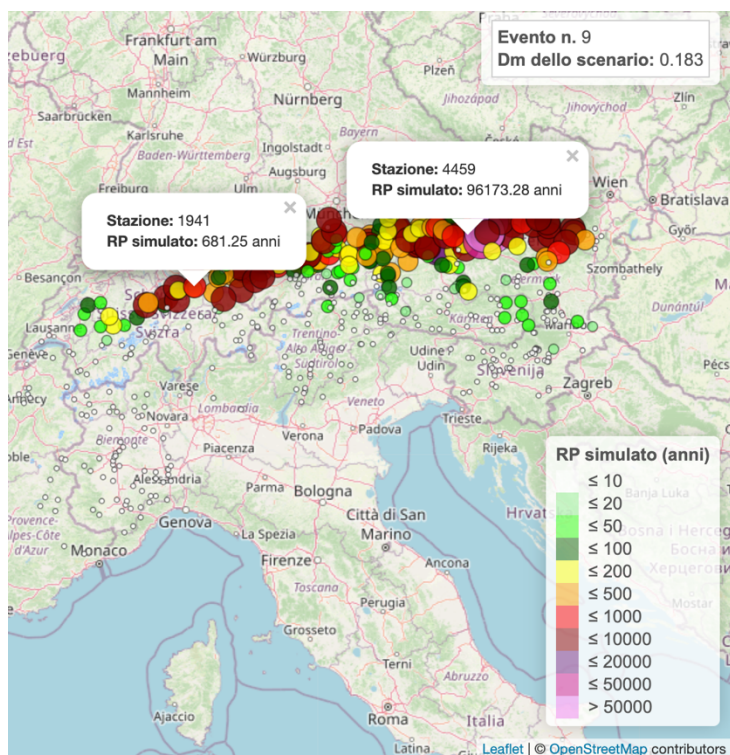
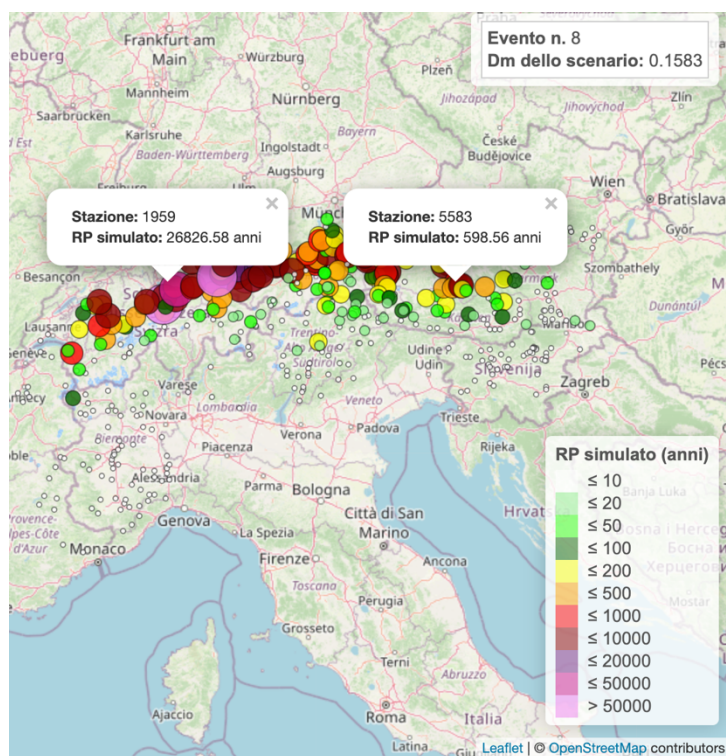
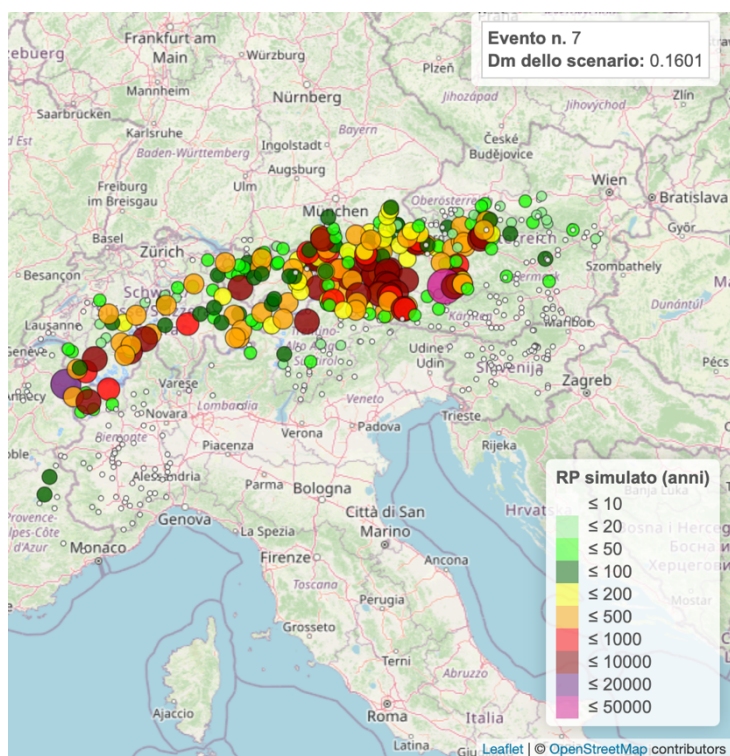


Figura 123- Mappa leaflet di quattro dei 15 eventi simulati più gravi.

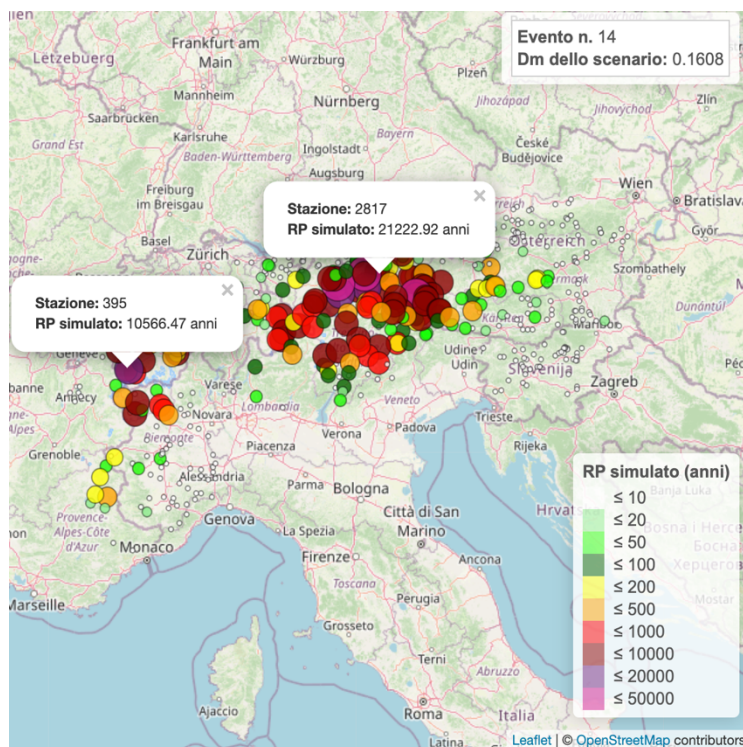
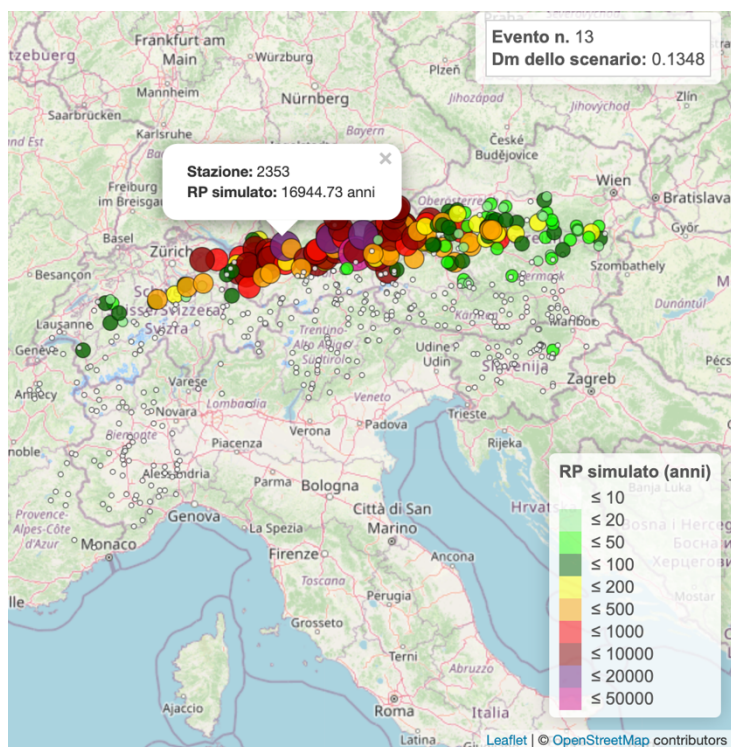
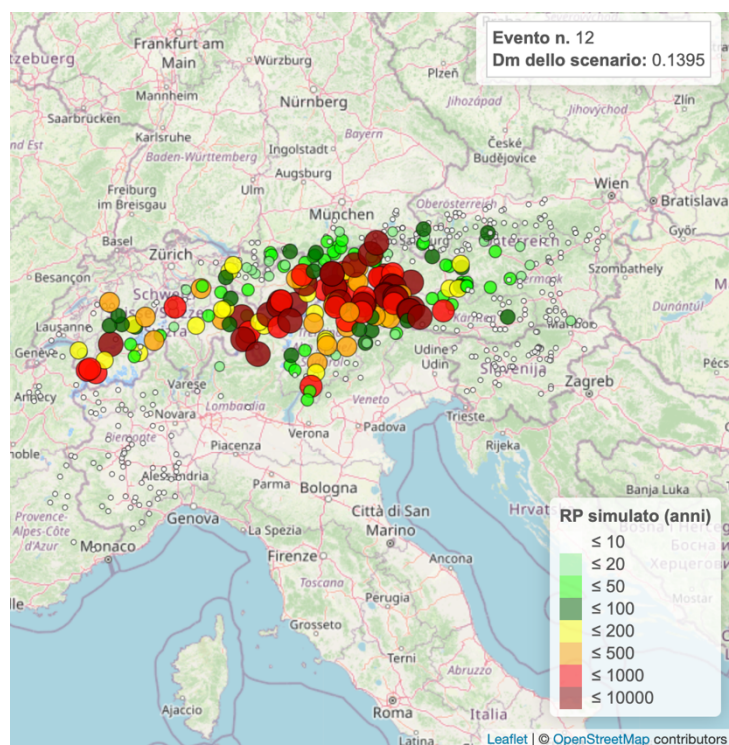
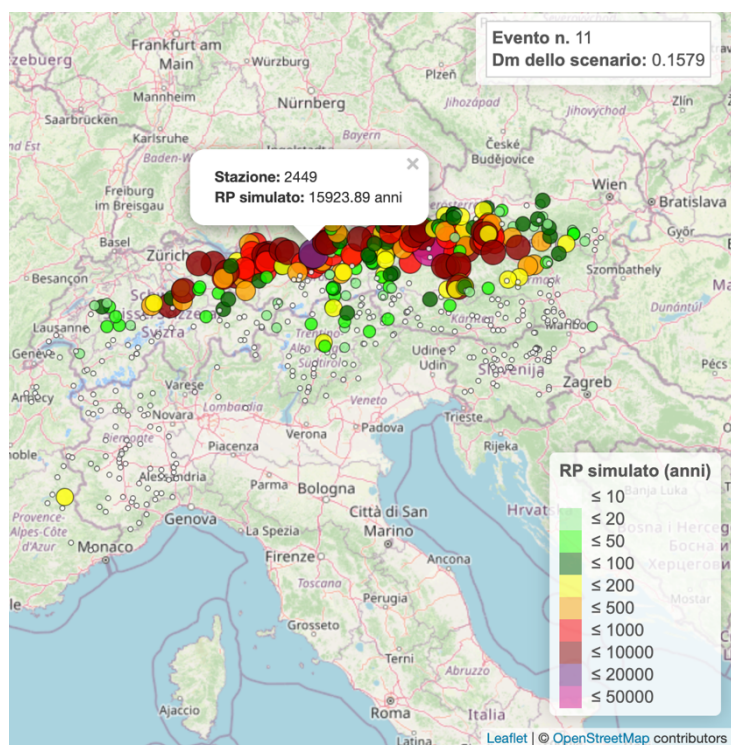


Figura 124- Mappa leaflet di quattro dei 15 eventi simulati più gravi.

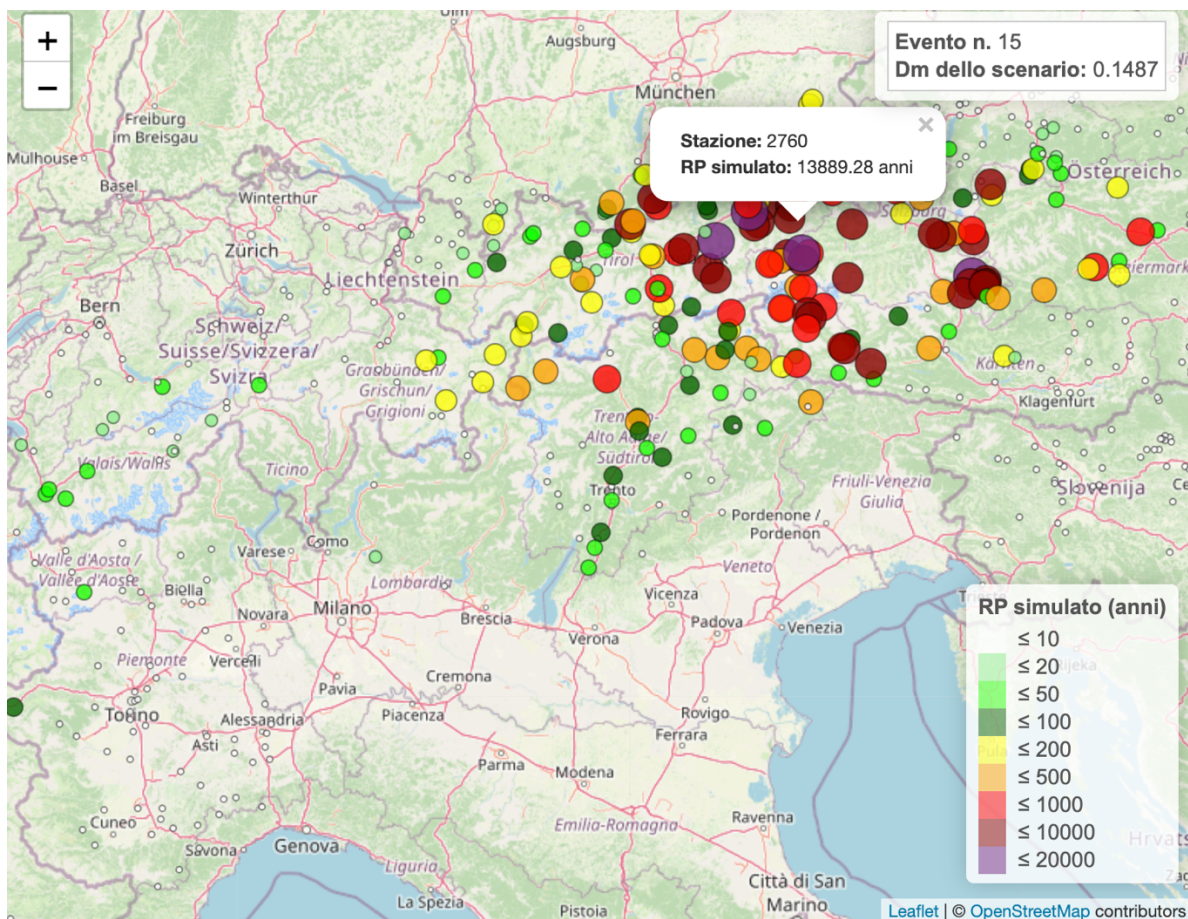


Figura 125- Mappa leaflet di uno dei 15 eventi simulati più gravosi.

Gli eventi con i valori massimi di D_m rappresentano scenari estremamente severi, con tempi di ritorno regionali che possono raggiungere anche i 200000 anni. A differenza degli eventi più comuni, questi scenari estremi non sono caratterizzati da picchi isolati in singole stazioni, ma da un coinvolgimento simultaneo di un gran numero di bacini, con portate eccezionali su gran parte del territorio. Le mappe, infatti, mostrano un'elevata concentrazione di stazioni classificate nelle classi di tempo di ritorno più alte (oltre 10000 anni). I colori rosso scuro e viola, associati a tempi di ritorno catastrofici, si distribuiscono in maniera diffusa, indicando che l'intero sistema idrografico nazionale sarebbe sottoposto a condizioni estreme nello stesso evento.

La generazione di eventi con $T > 100.000$ anni riflette la coda pesante delle distribuzioni di piena (modellata con Generalized Pareto e Burr III), che permette di quantificare il rischio di eventi nazionali catastrofici rarissimi non presenti nelle osservazioni storiche, rilevanti in un'ottica di pianificazione strategica e valutazione delle perdite massime possibili.

Continuando l'analisi sugli eventi simulati si è deciso di estrarre 15 eventi con tempo di ritorno medio di 100 anni. Questi eventi rivestono particolare importanza per due motivi principali:

- a differenza degli eventi estremamente rari (con tempi di ritorno di migliaia di anni), gli eventi con $T_m \approx 100$ anni hanno una probabilità di occorrenza concreta nel corso della vita utile delle infrastrutture idrauliche. Essi rappresentano quindi scenari di piena plausibili e coerenti con la scala temporale della pianificazione idraulica e della protezione civile.
- valutazione dell'impatto socio-economico: infatti pur non raggiungendo l'estensione catastrofica degli scenari estremi, questi eventi possono comunque coinvolgere aree molto ampie e determinare danni significativi a infrastrutture, insediamenti e attività economiche.

Essendo un tempo di ritorno medio su tutta la GAR, localmente possono verificarsi fenomeni con tempi di ritorno molto maggiori, come mostrato nella figura riportata sotto:

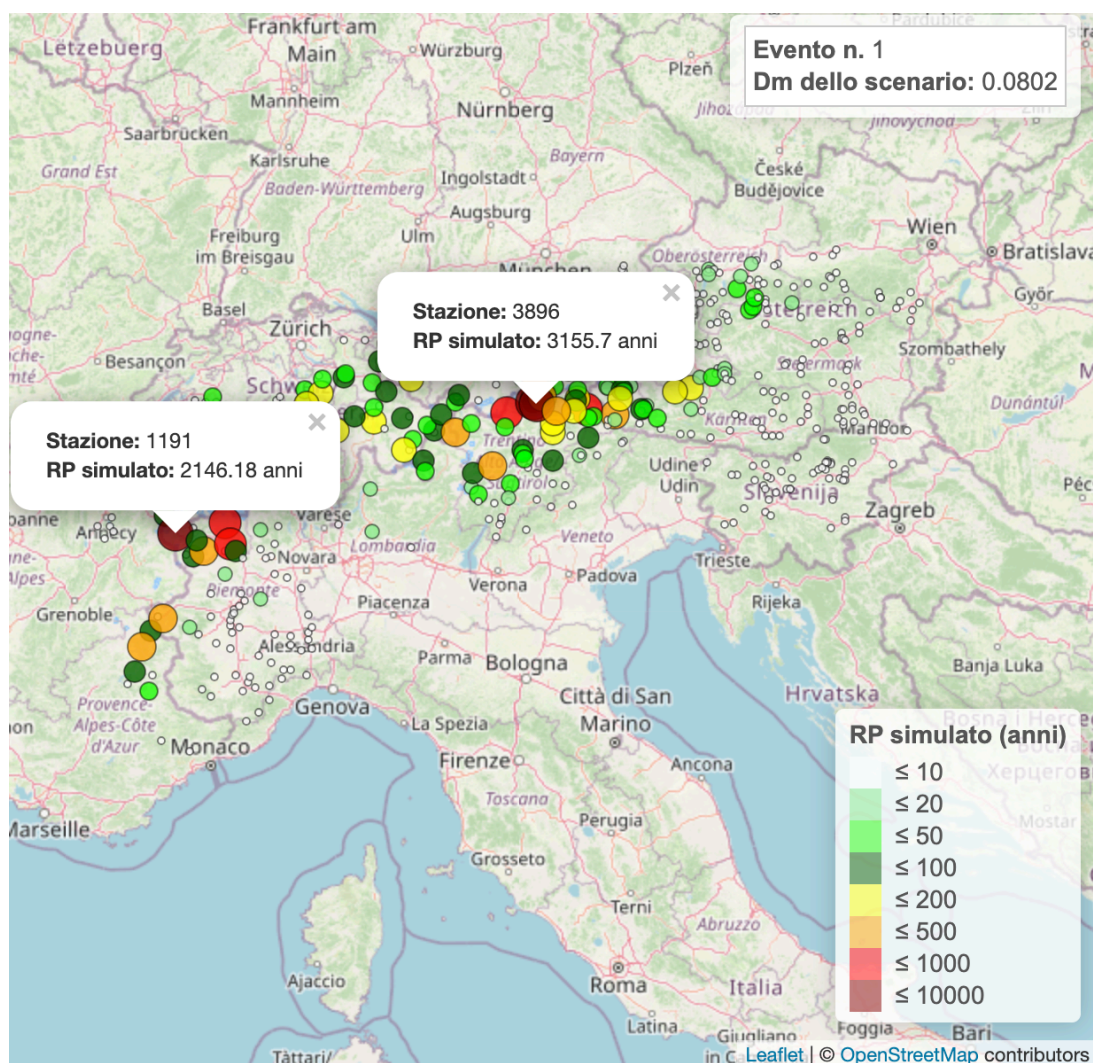


Figura 126-Evento con T_m medio pari a 100 anni.

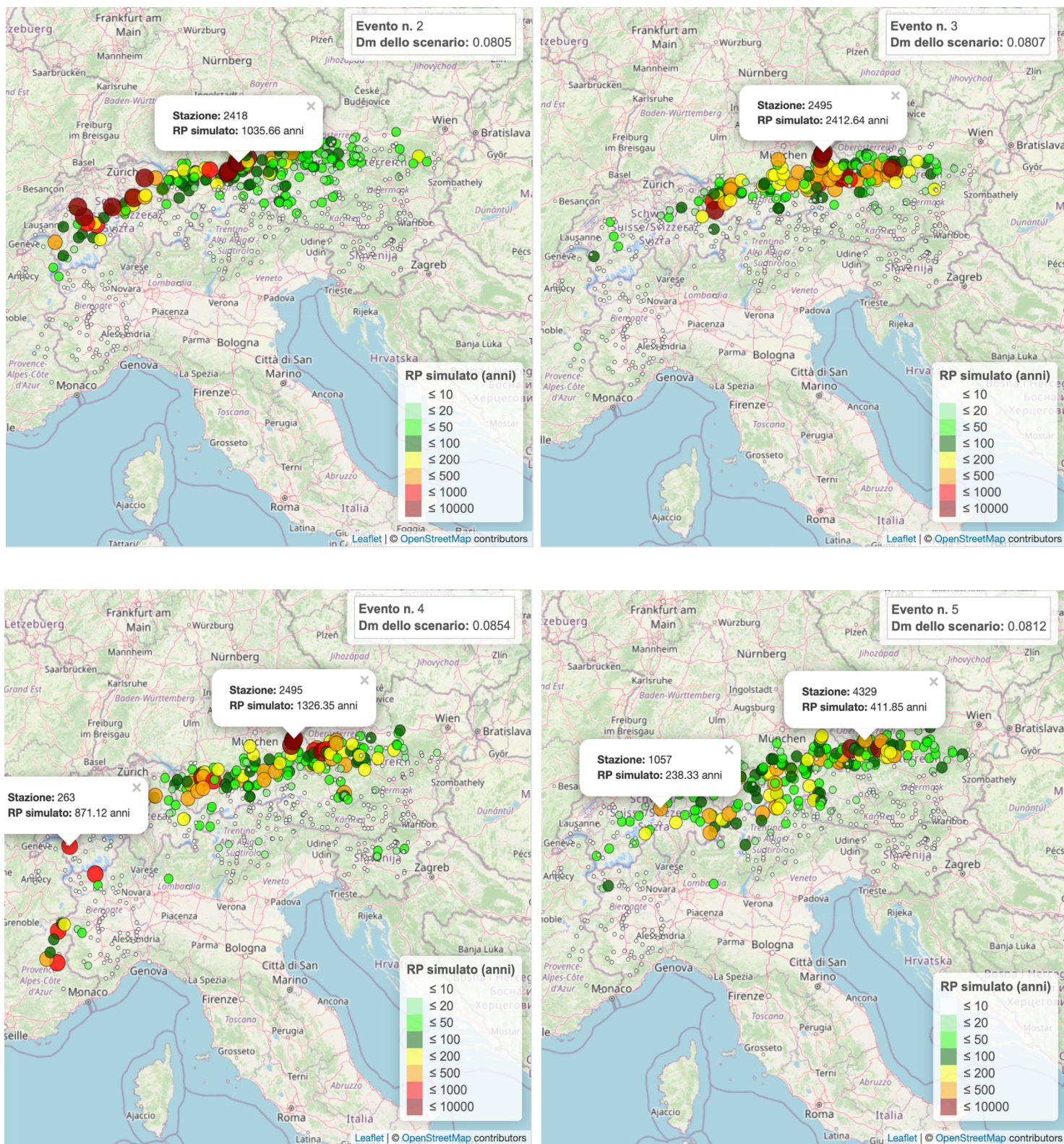


Figura 127- Eventi con T_m medio pari a 100 anni.

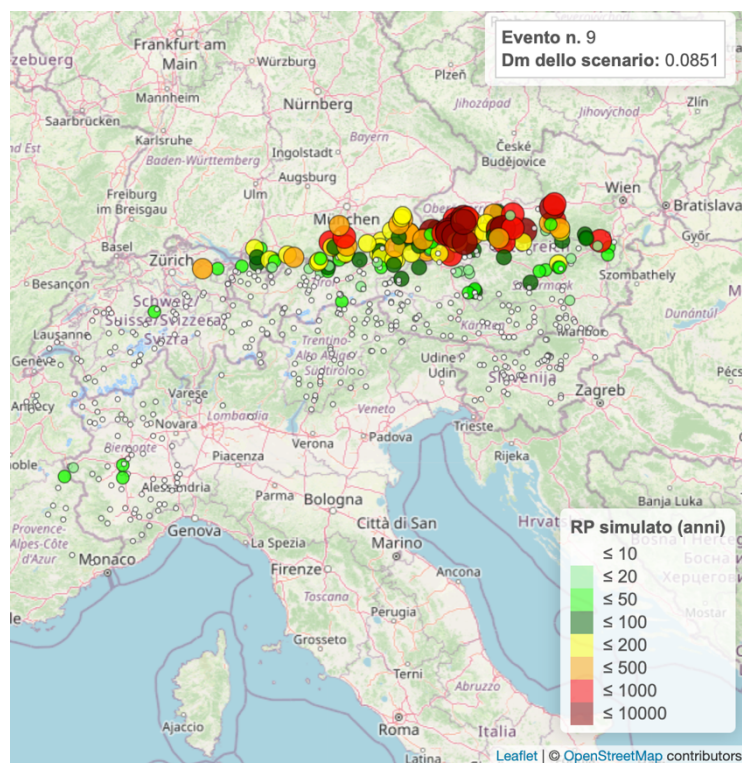
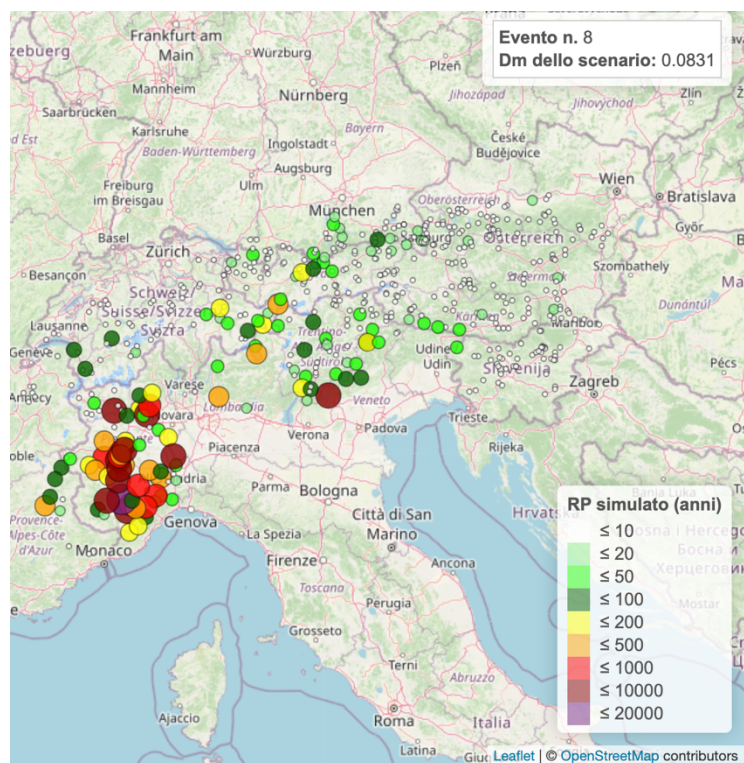
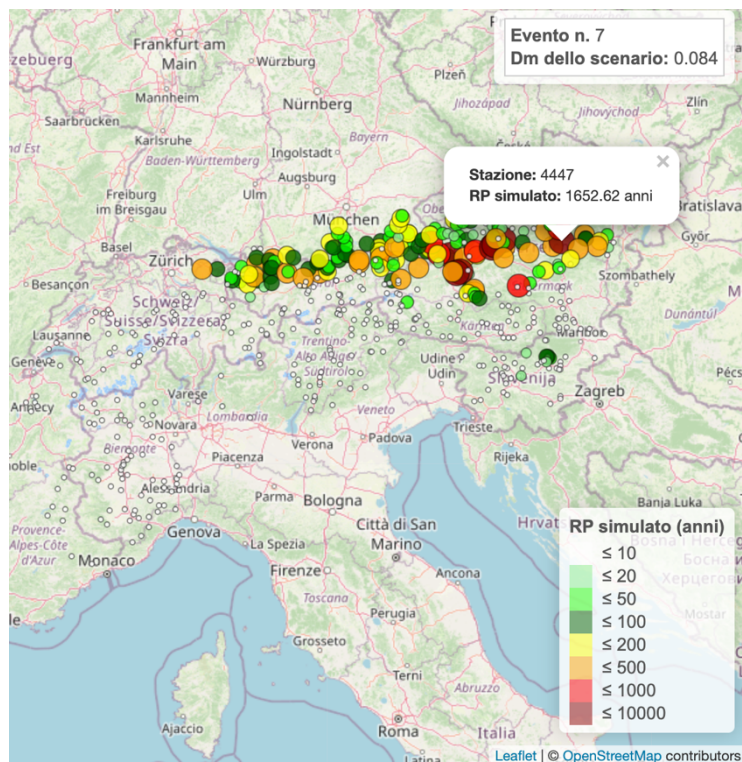
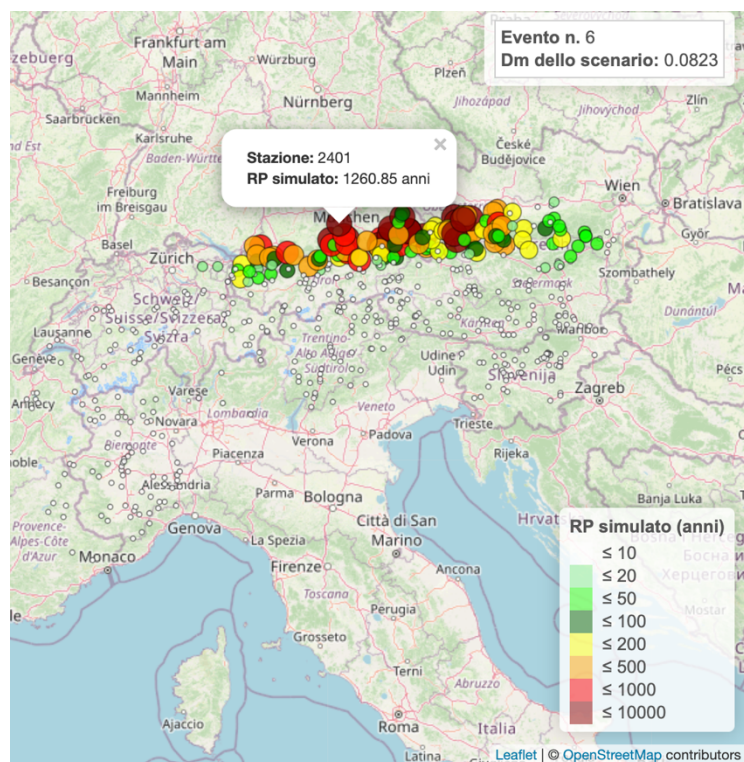


Figura 128- Eventi con T_m medio pari a 100 anni.

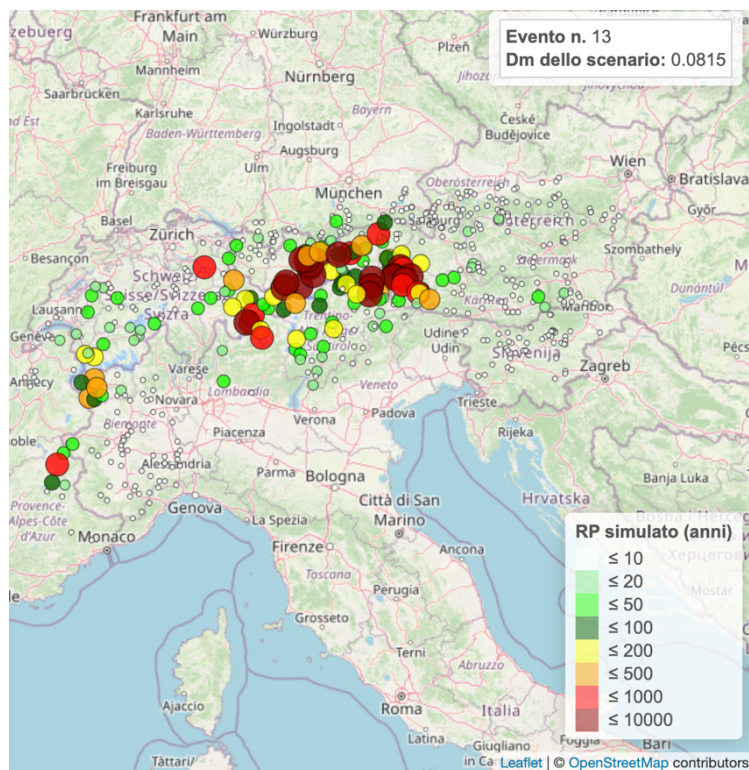
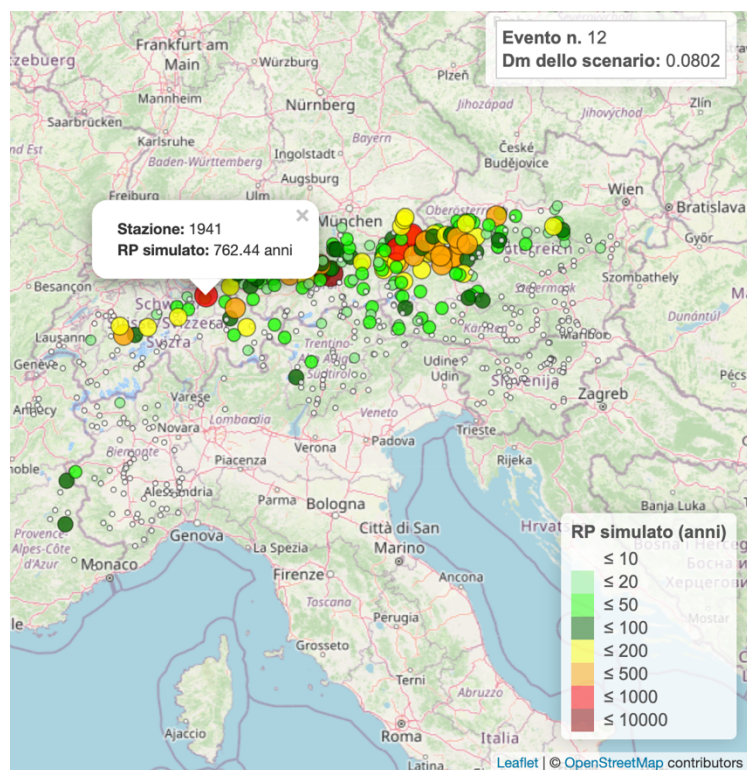
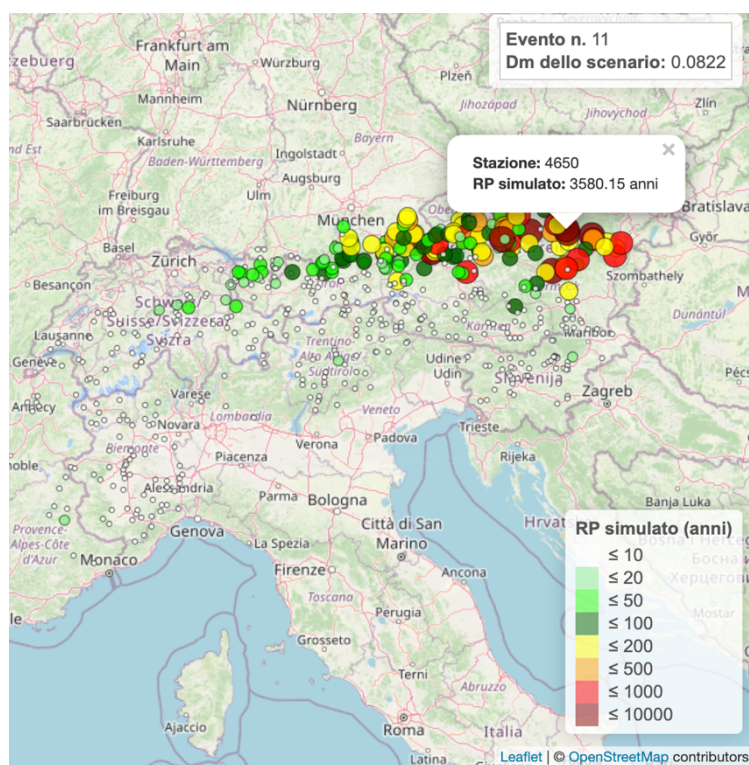
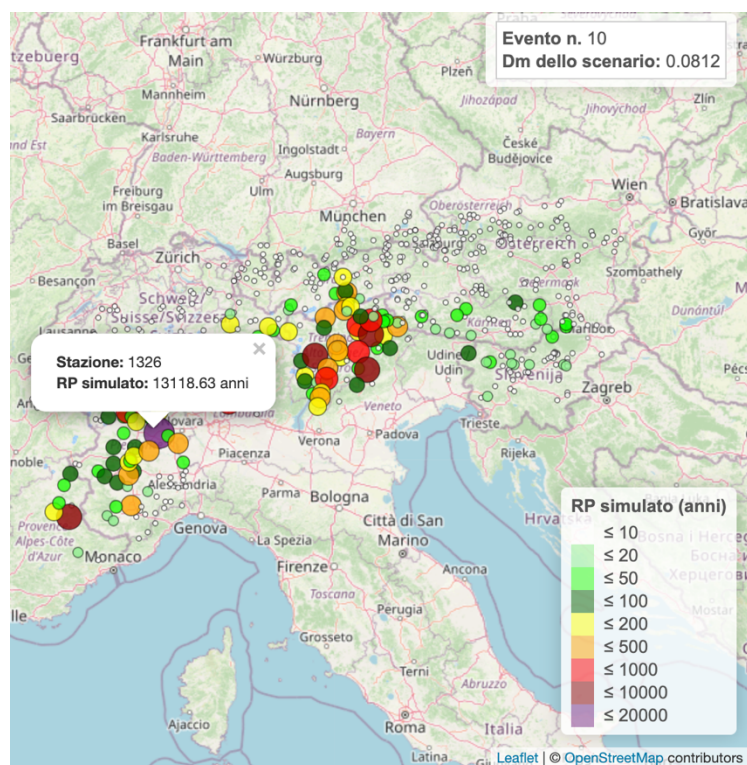


Figura 129- Eventi con T_m medio pari a 100 anni.

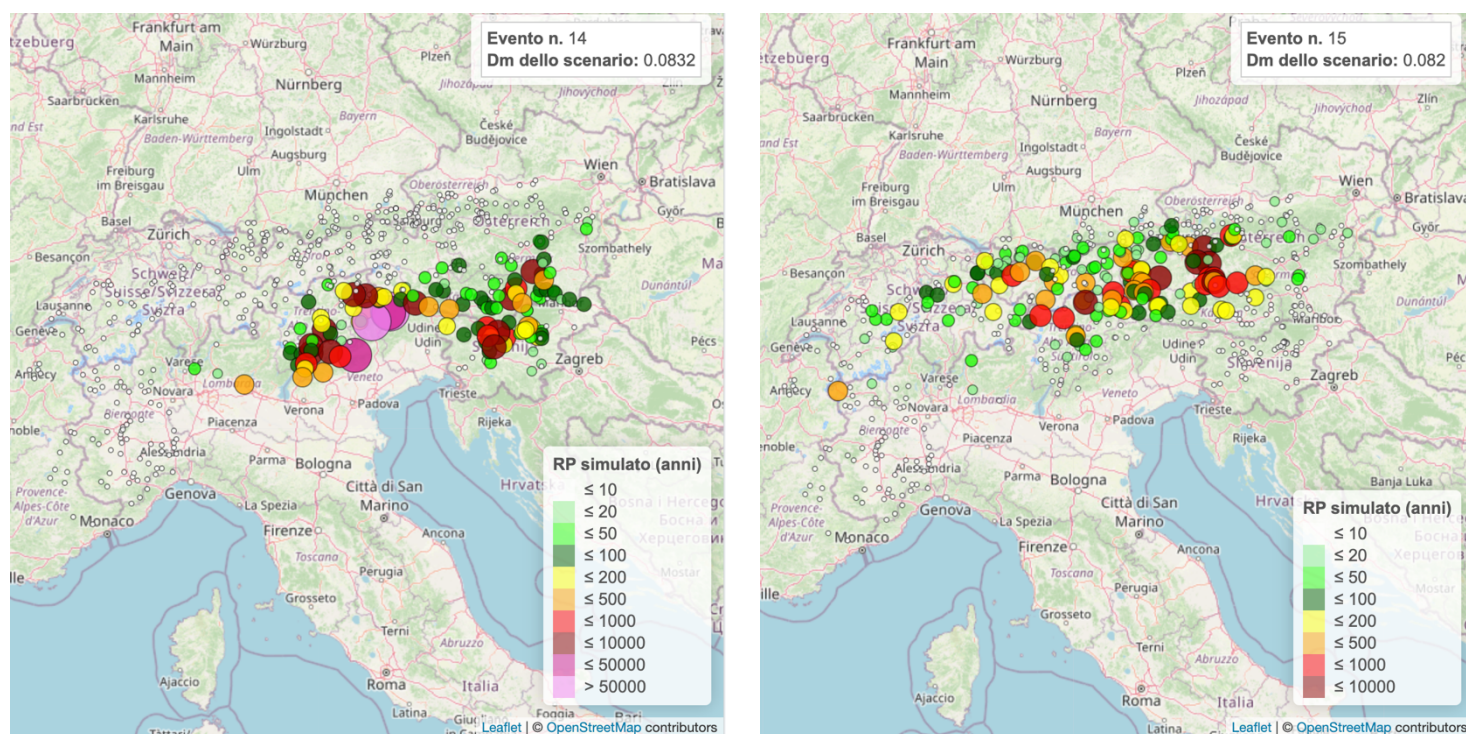


Figura 130- Eventi con T_m medio pari a 100 anni.

Dall'analisi delle mappe mostrate sopra si osserva innanzitutto che il valore dell'indice di severità D_m risulta praticamente identico per tutti gli eventi estratti. Questo è coerente con il fatto che gli eventi sono stati selezionati condizionandoli ad avere un tempo di ritorno medio nazionale di circa cento anni: il criterio di estrazione fissa implicitamente il livello medio di severità e ciò comporta una ridotta variabilità nei valori di D_m .

Un secondo aspetto rilevante riguarda l'estensione spaziale degli eventi: circa 10 dei 15 scenari simulati mostrano un coinvolgimento idrologico che interessa quasi la totalità del territorio austriaco, evidenziando come un evento con tempo di ritorno medio centenario non sia necessariamente confinato a scala locale, ma possa assumere una dimensione sincronizzata a livello nazionale. Da notare è che anche a parità di D_m gli eventi mostrano configurazioni diverse in termini di intensità locale e distribuzione dei tempi di ritorno, indicando che la stessa severità media può realizzarsi attraverso combinazioni idrologiche molto differenti.

Per completezza, nella pagina seguente, vengono riportati anche eventi con tempo medio di ritorno di dieci anni: anche per questa tipologia di eventi ci si aspetta che localmente ci possano essere situazioni ben più gravose del tempo medio di ritorno.

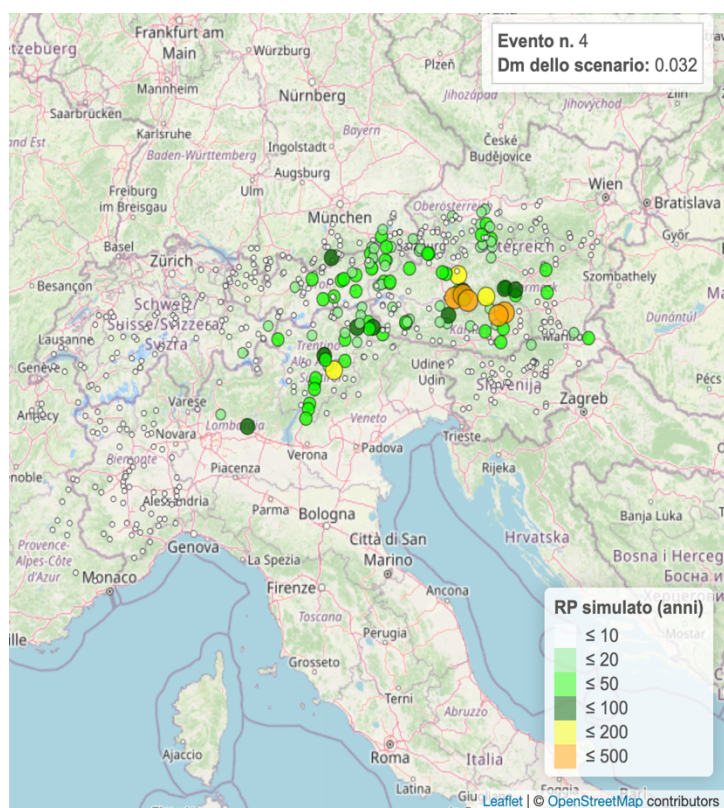
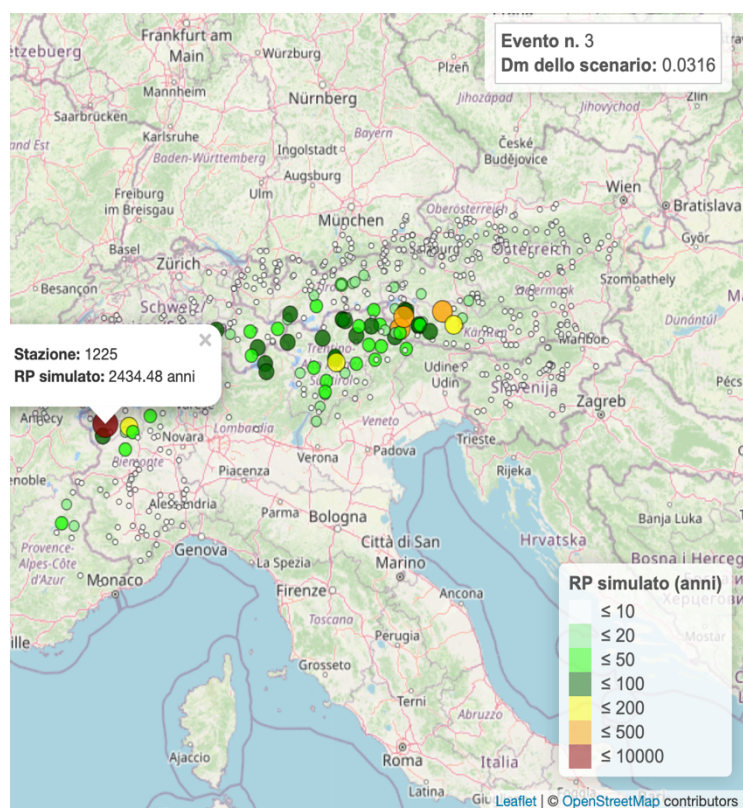
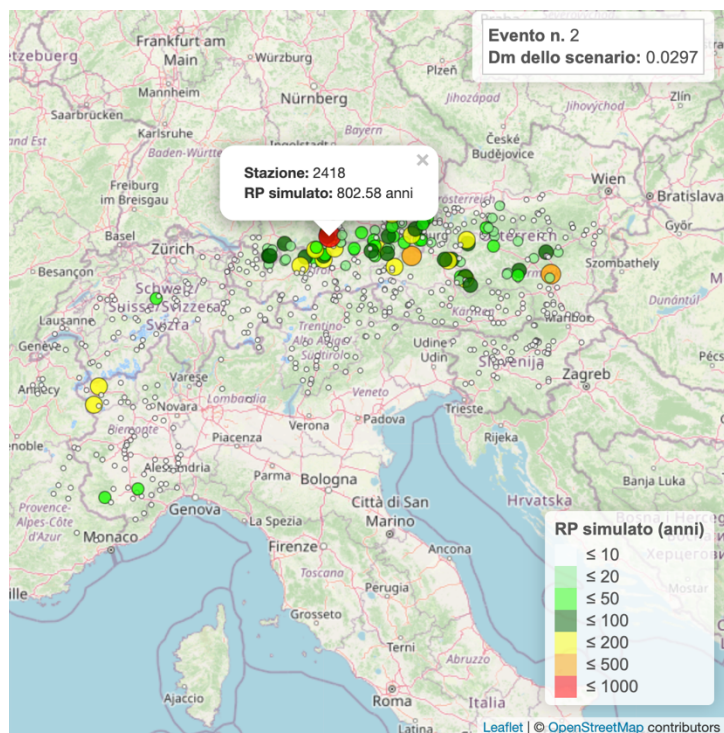
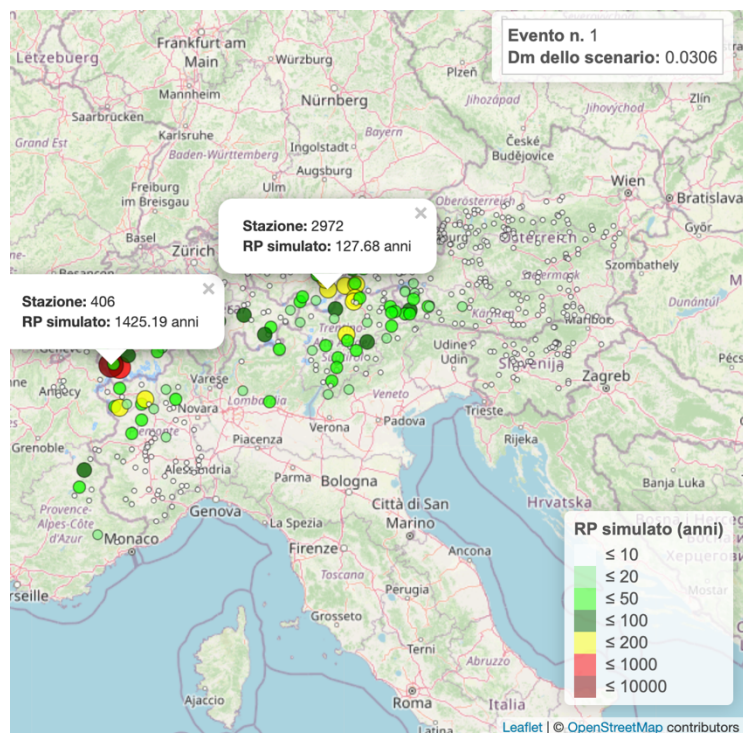


Figura 131- Eventi con T_m medio pari a 10 anni.

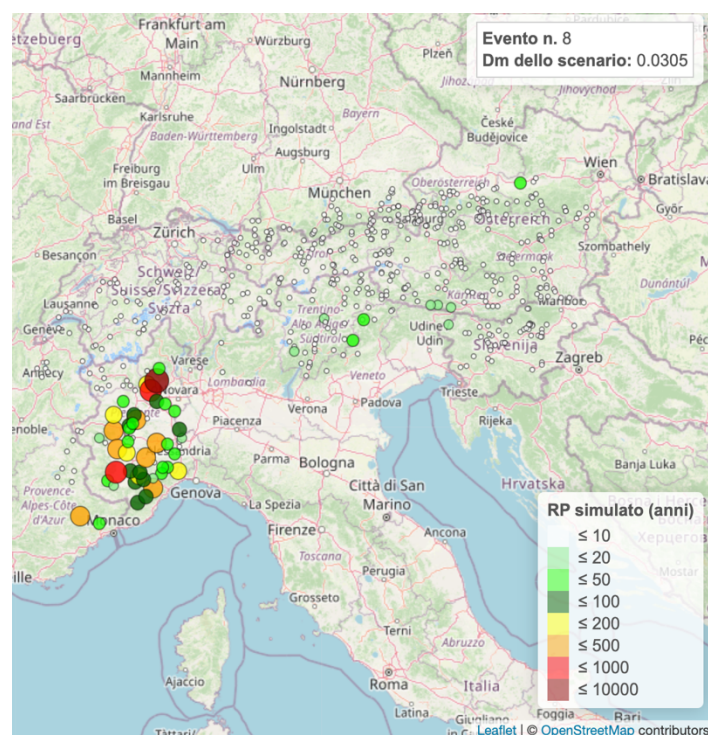
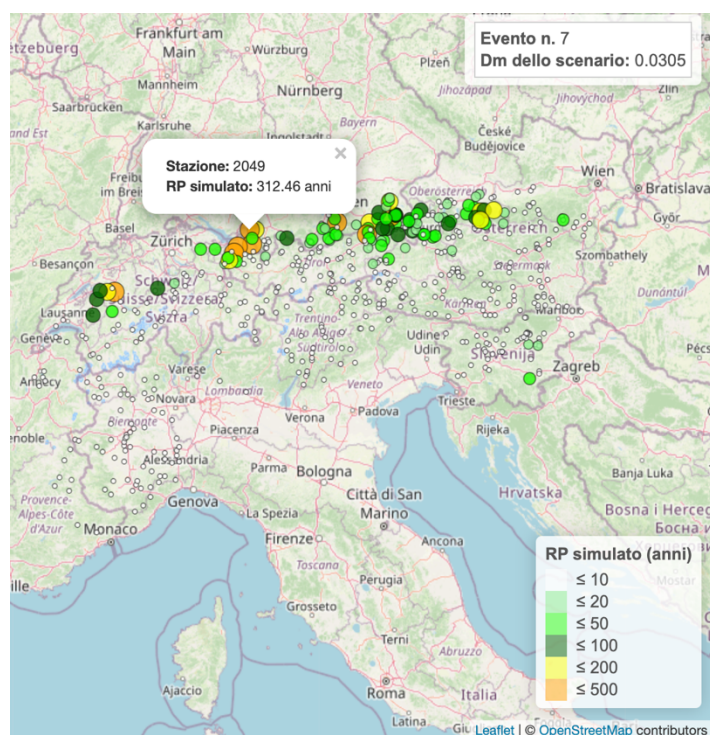
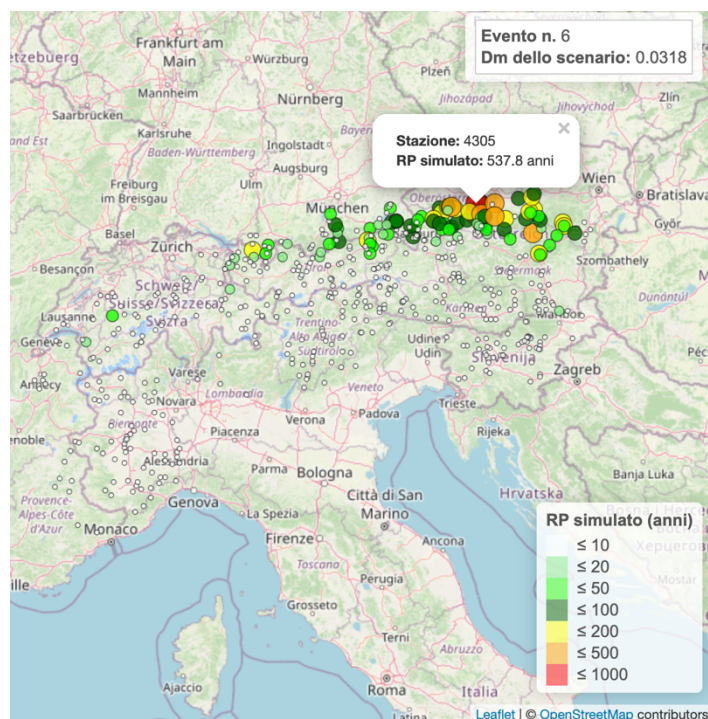
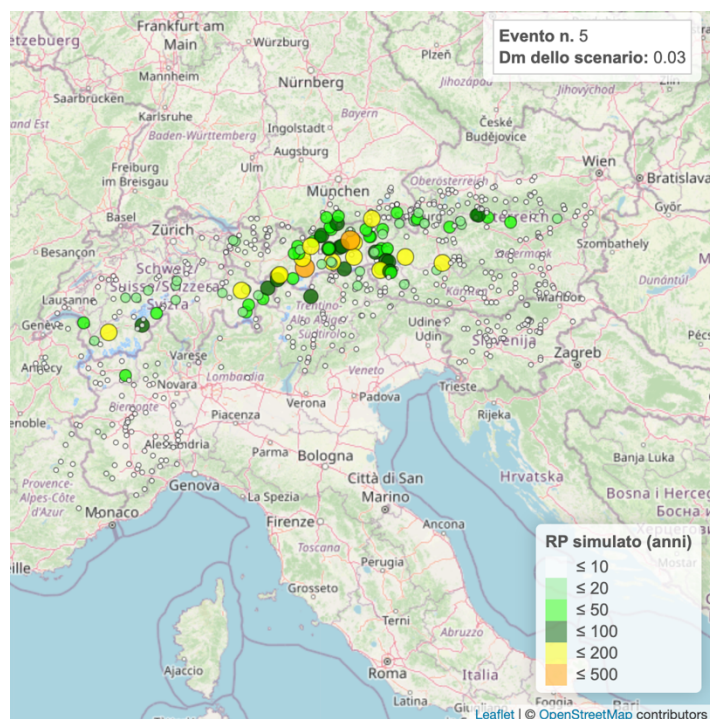


Figura 132- Eventi con T_m medio pari a 10 anni.

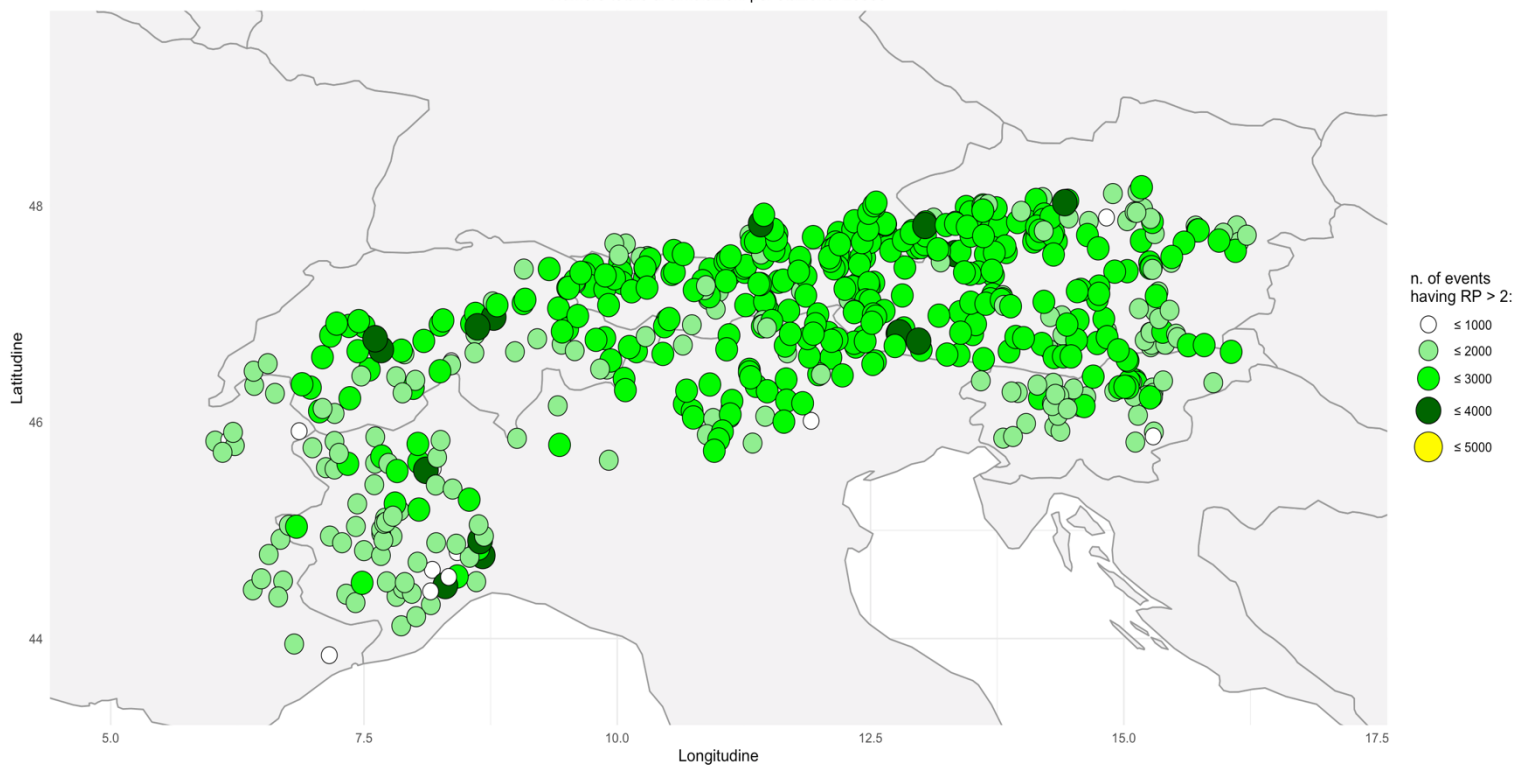
Questo tipo di mappatura permette di confrontare la varietà degli scenari simulati: dalla piena molto localizzata alla piena diffusa e catastrofica. È uno strumento utile soprattutto per l'analisi di rischio a scala regionale perché non guarda solo al singolo sito ma al quadro complessivo delle piene nella regione di analisi.

Allo stesso modo dell'analisi svolta sugli eventi osservati sono state inoltre svolte alcune analisi sulla distribuzione spaziale degli eventi simulati, tra cui:

- frequenza eventi simulati con $RP > 2$ per stazione: si mostra il numero di simulazioni in cui il tempo di ritorno locale supera 2 anni per poterlo confrontare con il numero totale di simulazioni svolte. Se molte stazioni mostrano un numero elevato di eventi simulati che causano danni significa che gli scenari simulati tendono a produrre eventi diffusi anziché punti isolati. Se invece i conteggi sono concentrati in poche aree, i fenomeni sono più localizzati. I marker nelle mappe sono colorati e scalati in funzione di questo conteggio: le stazioni con marker più grandi e di colore più scuro sono quelle che, nelle simulazioni, risultano più frequentemente colpite da eventi non banalmente ricorrenti. Questi sono quindi i luoghi più spesso raggiunti da eventi significativi.
- frequenza eventi simulati con danno > 0 per stazione: si contano quante simulazioni producono un danno locale positivo rispetto al numero totale di simulazioni svolte. Ciò permette di evidenziare la vulnerabilità spaziale delle stazioni: quelle con molti casi con $D > 0$ sono coloro che più spesso subiscono impatti (anche se piccoli). Mettere a confronto questo plot con il precedente aiuta a distinguere luoghi dove si hanno spesso picchi non-dannosi (ovvero con $RP > 2$ ma $D \approx 0$) da luoghi dove anche eventi meno rari provocano danni.
- media geometrica locale degli RP: si plotta la media geometrica ($\exp(\text{media dei } \log T)$) dei periodi di ritorno simulati per ogni stazione perché è una misura robusta della rarità tipica osservata al sito. Infatti, valori elevati di RP_{geom} indicano stazioni dove, in media, gli eventi sono più rari, cioè occorrono meno frequentemente eventi intensi. Valori bassi, invece, indicano stazioni più esposte a eventi frequenti. Si impiega la media geometrica e la si preferisce alla media aritmetica per dati asimmetrici, come i tempi di ritorno, perché riduce l'influenza di outlier.
- danno medio locale: si plotta la media aritmetica del danno locale D_m per ciascuna stazione. Ciò permette di evidenziare le aree con maggiore severità attesa: stazioni con D_m più alto sono quelle dove si aspettano impatti maggiori in media. È interessante confrontare quest'ultima con la mappa RP_{geom} così da poter distinguere se alte medie di danno derivano da eventi rari molto grandi o da eventi più frequenti di intensità moderata.

In generale, le mappe descritte consentono di individuare le stazioni più frequentemente esposte e le aree con maggiore severità media. Nelle figure seguenti vengono mostrati i risultati.

Frequenza eventi con $RP > 2$ simulati per stazione (GAR)
 Numero totale di simulazioni per stazione: 20000



Frequenza eventi simulati con danno > 0 per stazione (GAR)
 Numero totale di simulazioni per stazione: 20000

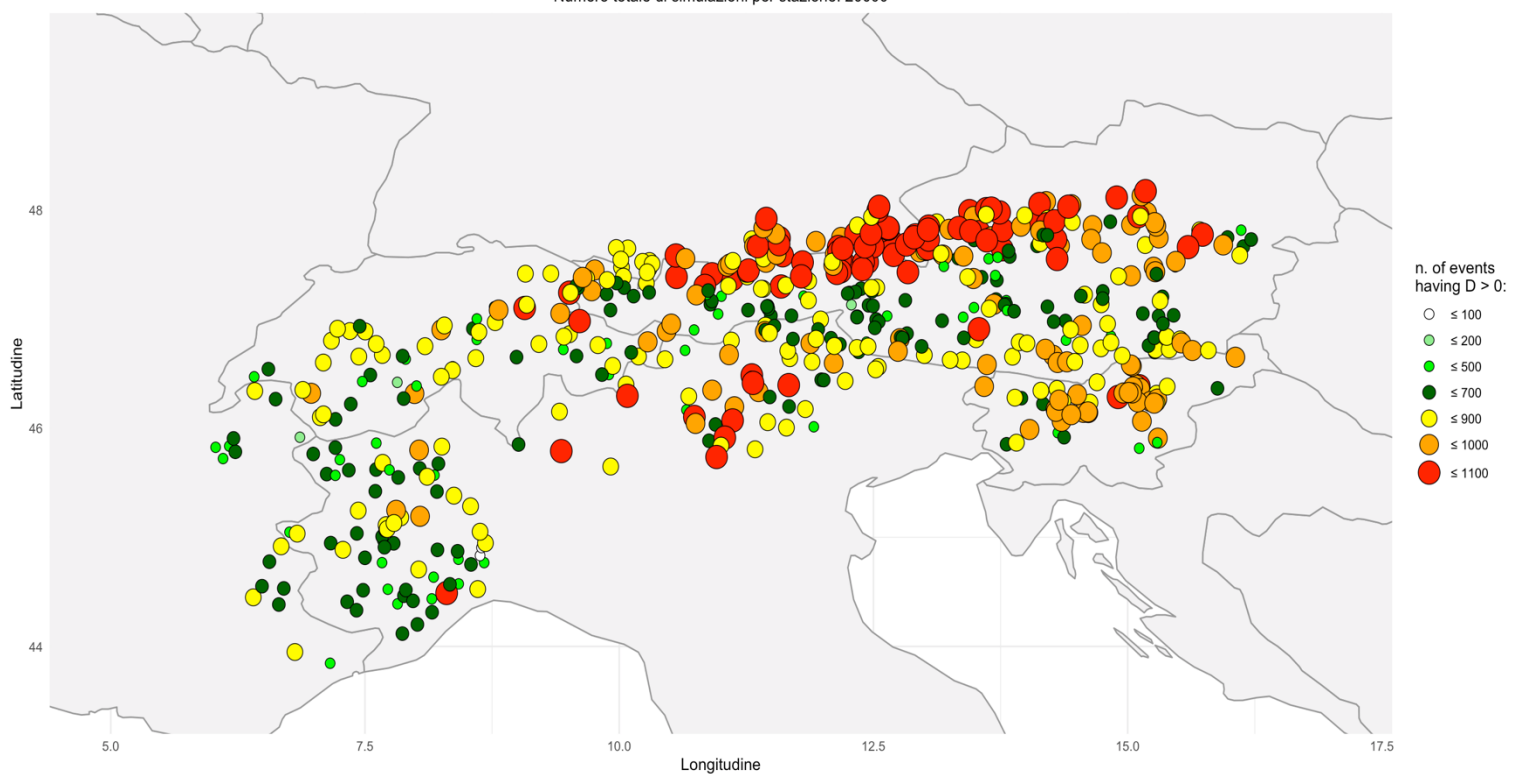


Figura 133- Grafici che mostrano rispettivamente la frequenza di eventi con $RP > 2$ per stazione e la frequenza di eventi con $D_m > 0$ per stazione.

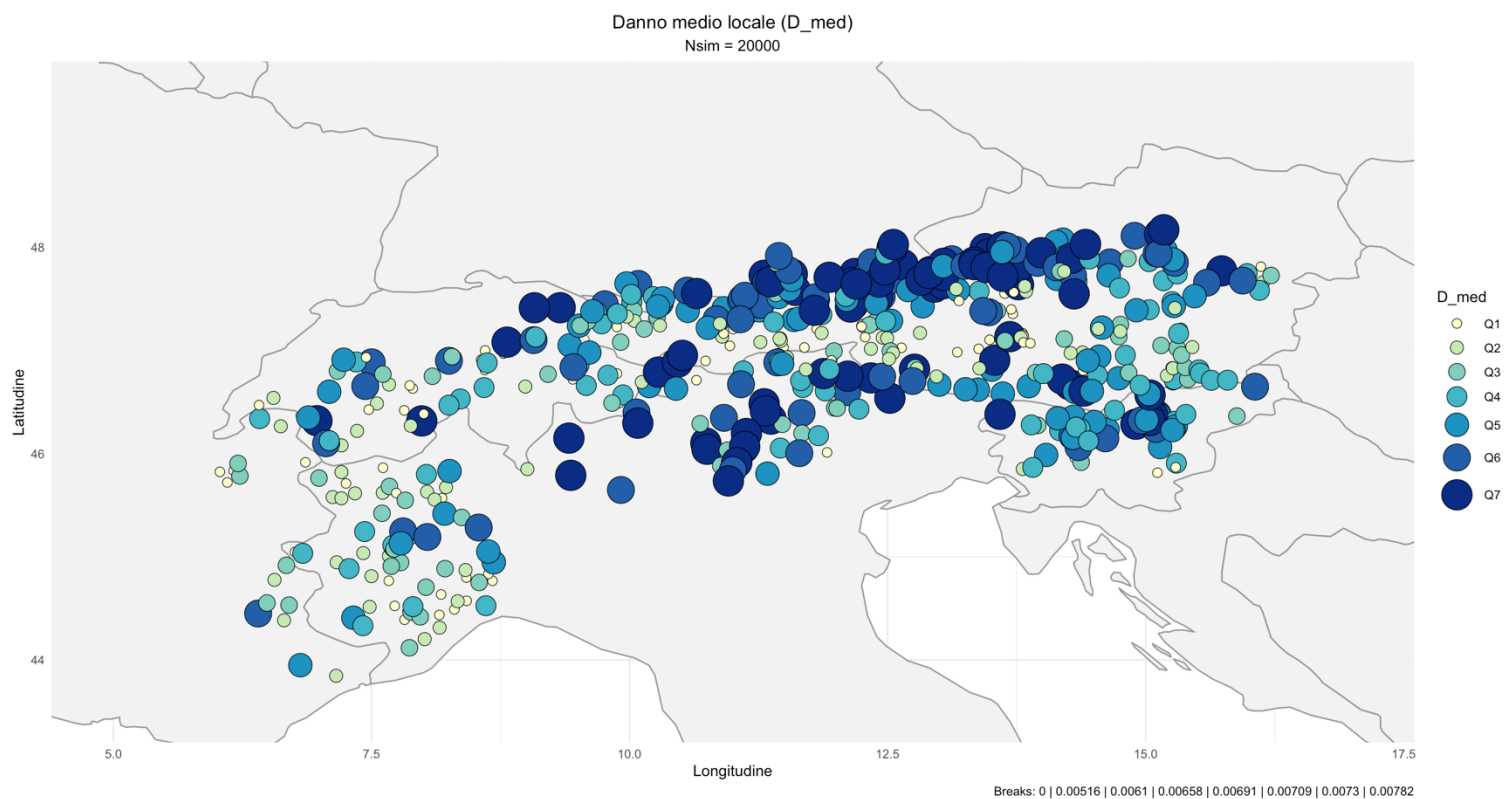
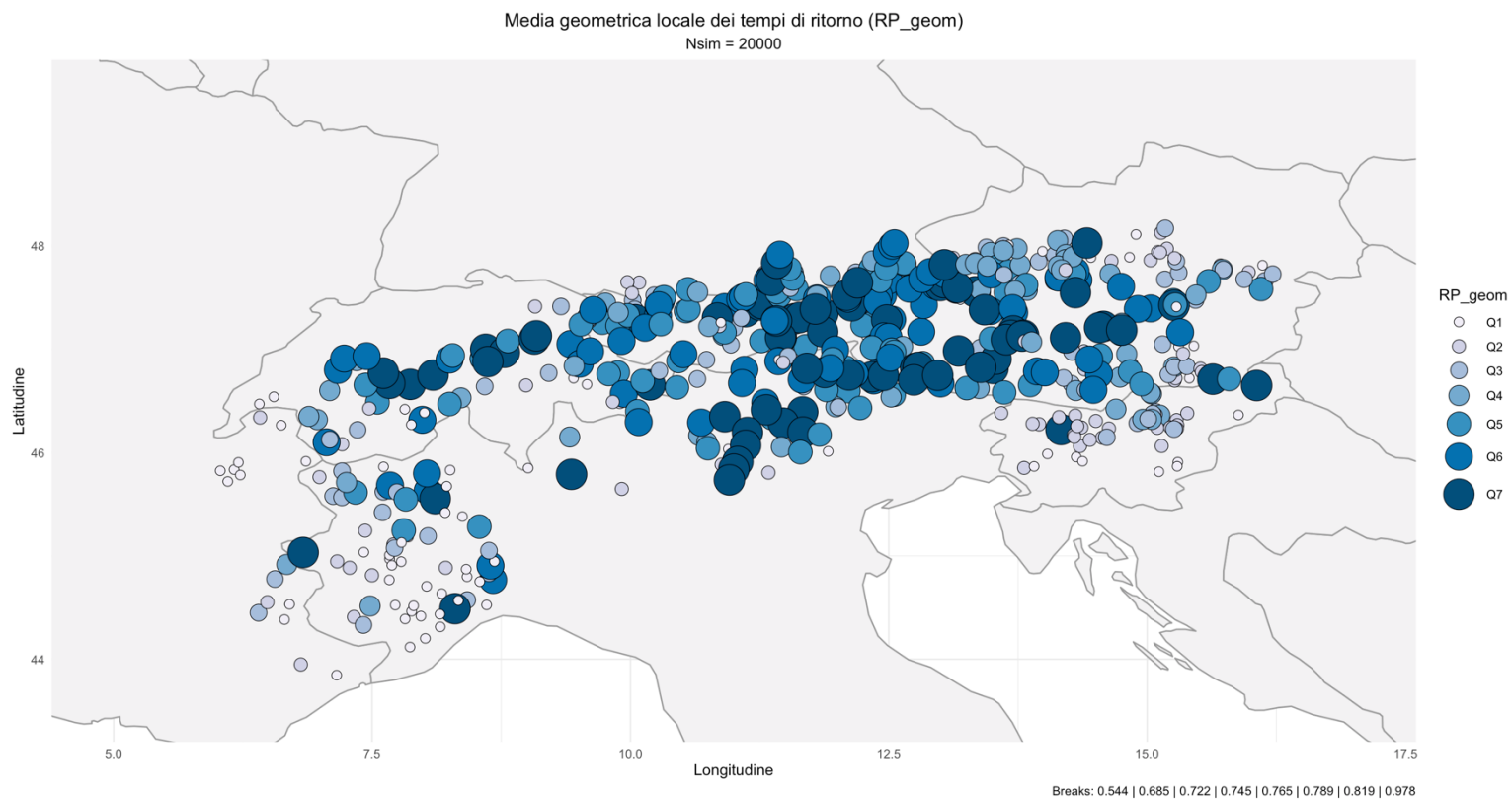


Figura 134- Grafici che mostrano rispettivamente le medie degli RP e D_m divise in 7 classi.

Si possono fare alcune considerazioni sui risultati mostrati in queste mappe:

- le classi più elevate di D_m (i marker più grandi e i colori più scuri) sono concentrate lungo la catena alpina, specialmente nella fascia montana settentrionale ovvero Austria centrale/orientale e in alcune porzioni delle Alpi occidentali. Le zone di pianura e le stazioni periferiche presentano in genere D_m più basso (con colori chiari). L'aggregazione spaziale suggerisce che i fattori orografici e la rete idrografica alpina abbiano un'influenza importante sulla distribuzione degli eventi nella GAR. Infatti, è da notare proprio la presenza di gruppi ampi di marker scuri. Questo implica che eventi simulati hanno spesso carattere sincrono su più bacini della stessa zona montana;
- le classi di RP_{geom} più alte (cioè i quantili superiori) risultano prevalentemente ubicate lungo la stessa fascia alpina dove D_m risulta maggiore. Da notare il fatto che la distribuzione dei valori di questa variabile è relativamente compressa: i valori di RP_{geom} sono tutti piuttosto vicini tra loro;
- la mappa dei conteggi $D > 0$ evidenzia molte stazioni con un numero elevato di simulazioni che generano danno (i marker rossi/arancione), soprattutto lungo la fascia alpina. Alcune aree hanno conteggi intermedi (giallo) o bassi (verde), indicando eterogeneità locale nella vulnerabilità o nell'esposizione. Questo aspetto deve però ovviamente essere integrato con informazioni puntuali da rilievi in campo per poter avere un'analisi del rischio affidabile.

In sintesi, le simulazioni evidenziano una concentrazione degli effetti più severi lungo la catena alpina in particolare l'area a nord delle Alpi, corrispondente all'Austria e alle zone di confine, con alcuni effetti importanti anche nelle parti occidentali (Svizzera/Vorarlberg). Un fattore molto importante è che le mappe mostrano coerenza tra di loro: le stazioni con D_m elevato tendono a corrispondere a zone con RP_{geom} tendenzialmente più alte e con maggiore numero di simulazioni che producono danno positivo.

Di seguito vengono, infine, mostrati i diagrammi a dispersione finali che permettono di tirare le fila del confronto eventi osservati vs eventi simulati.

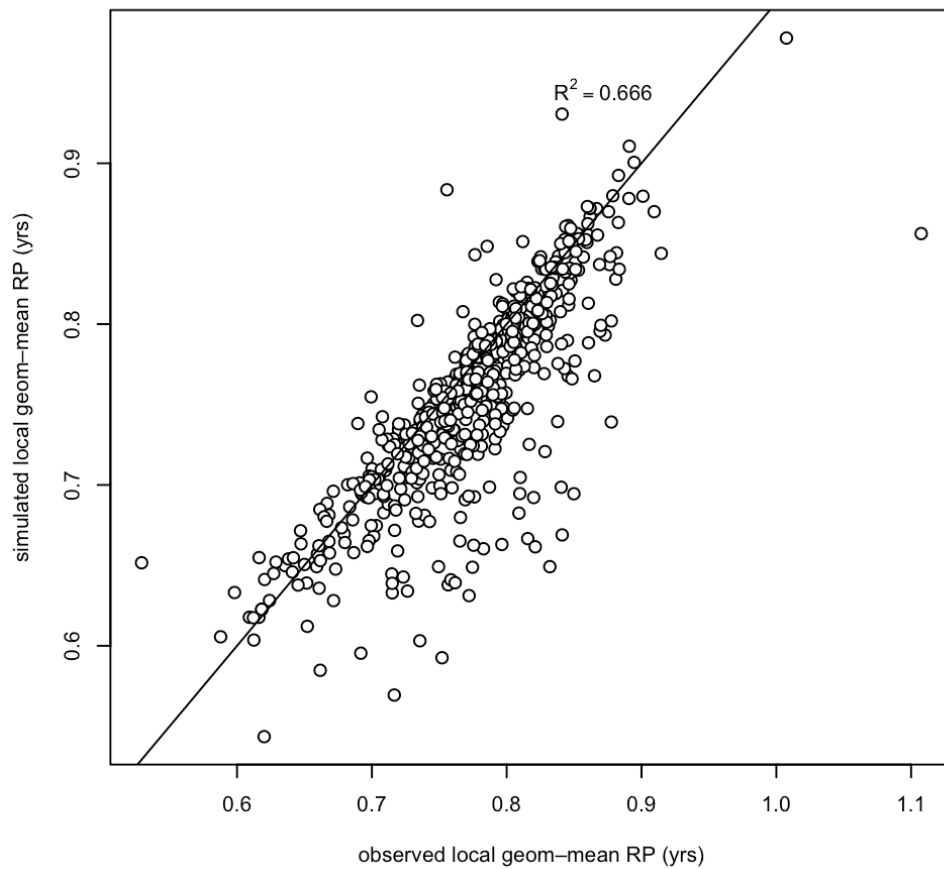


Figura 135- Confronto tra media geometrica degli RP osservati e simulati.

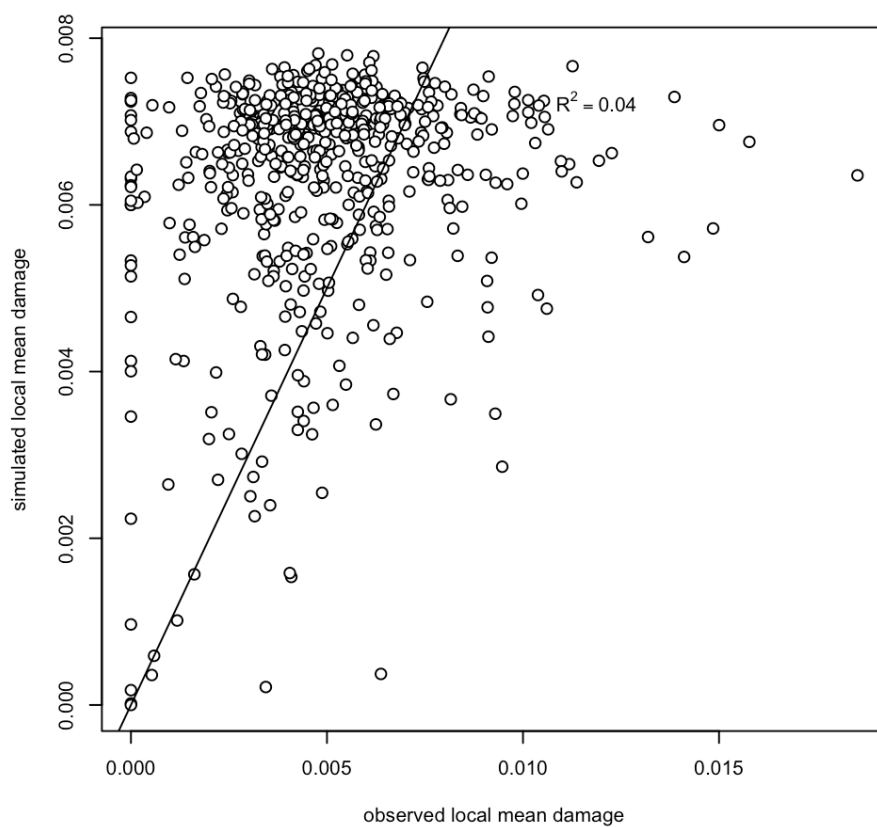


Figura 136- Confronto tra danno medio locale degli eventi osservati e simulati.

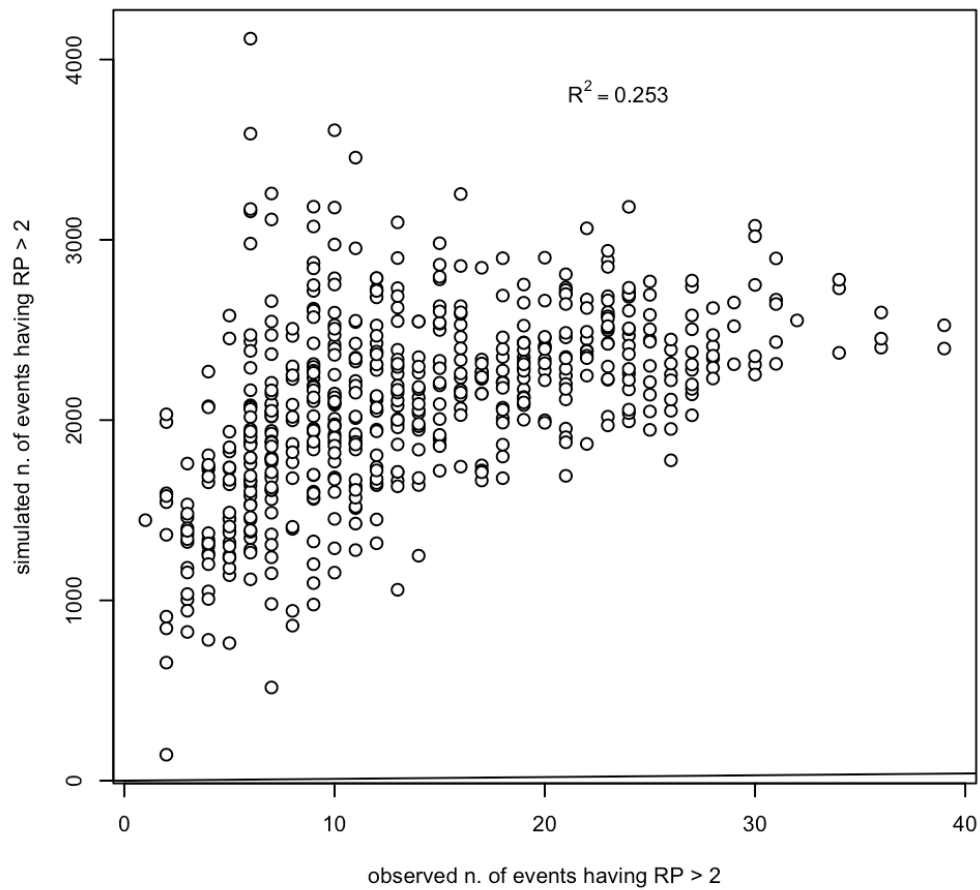


Figura 137- Confronto tra numero di eventi osservati e simulati con $RP > 2$.

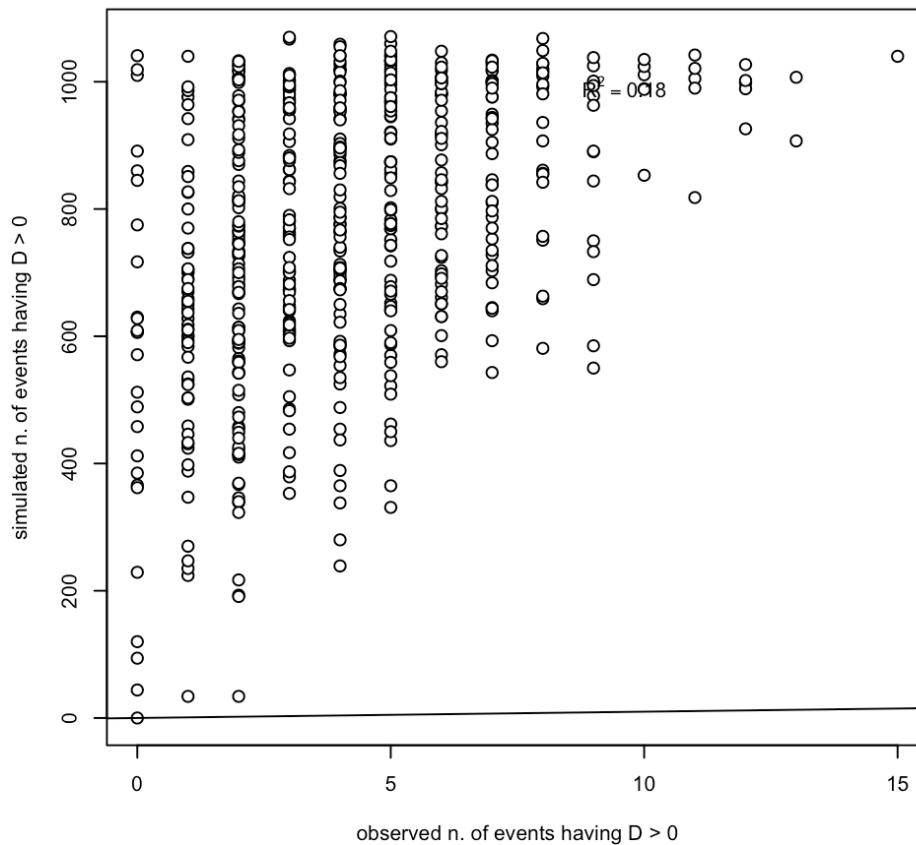


Figura 138- Confronto tra numero di eventi simulati e osservati con $D > 0$.

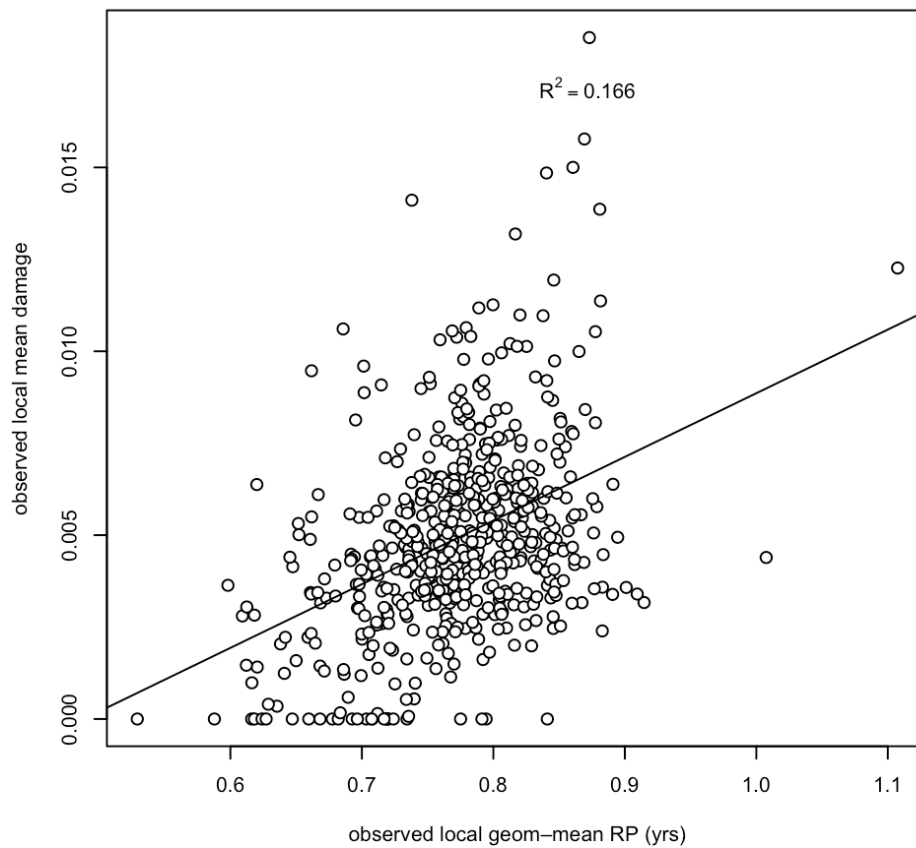


Figura 139- Relazione tra media geometria degli RP osservati e il danno medio locale osservato.

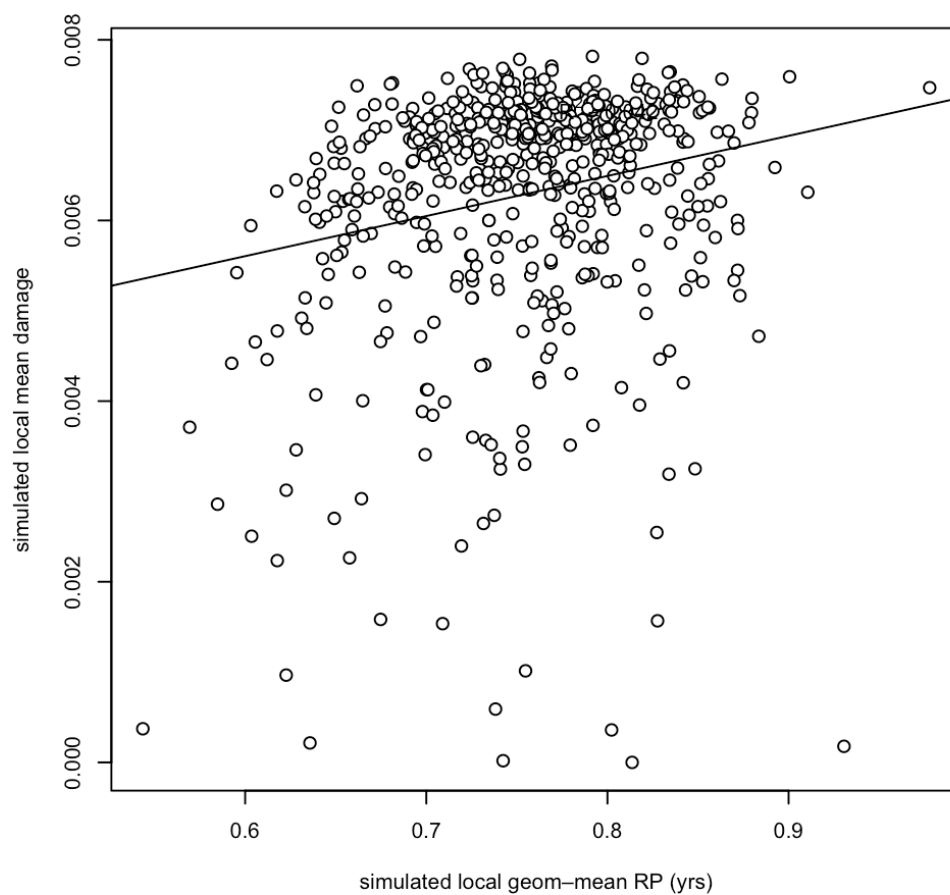


Figura 140- Relazione tra media geometrica degli RP simulati e danno medio locale simulato.

I grafici di confronto tra gli eventi osservati e simulati forniscono un quadro complessivamente positivo della capacità del modello di riprodurre la struttura probabilistica multivariata degli eventi di piena, ma evidenziano anche alcuni importanti limiti. In primo luogo, il confronto tra tempi di ritorno geometrici locali mostra una buona coerenza, con un coefficiente di correlazione ($R^2 \approx 0.67$) che indica come gran parte della variabilità osservata venga catturata dal modello. Questo risultato conferma che la procedura di simulazione è in grado di replicare adeguatamente le statistiche di base delle distribuzioni di frequenza.

Per quanto riguarda i danni medi locali, invece, le prestazioni risultano decisamente meno soddisfacenti. I valori simulati si distribuiscono in un intervallo ristretto e mostrano una correlazione molto debole con quelli osservati ($R^2 \approx 0.04$), segnalando una sottostima della variabilità e degli estremi stessi. Ciò suggerisce che, mentre il modello riesce a fornire stime plausibili dell'ordine di grandezza dei danni, non è in grado di rappresentare in modo accurato la complessità e l'eterogeneità spaziale riscontrabile negli eventi reali.

Anche il confronto tra il conteggio di eventi estremi simulati e osservati evidenzia differenze importanti: le simulazioni producono generalmente più eventi con tempo di ritorno superiore a due anni e molti più eventi con danno positivo rispetto a quanto osservato. Tuttavia, tale discrepanza va interpretata alla luce del diverso orizzonte temporale che si ha a disposizione. È infatti importante sottolineare il fatto che i dati osservati coprono alcune decadi, mentre le simulazioni Monte Carlo consentono di esplorare un ventaglio molto più ampio di possibili realizzazioni. In questo senso, il maggiore numero di eventi simulati non va letto come una sovrastima del rischio ma come una conseguenza naturale del diverso campionamento.

Infine, il confronto tra tempi di ritorno e danni medi rivela un aspetto interessante. Nei dati osservati la correlazione è presente ma debole, segno che oltre alla magnitudo idrologica giocano un ruolo importante fattori locali e socio-economici. Nelle simulazioni, invece, la relazione appare più marcata con un chiaro incremento dei danni al crescere del tempo di ritorno. Ciò indica che il modello riesce a riprodurre il legame teorico tra severità dell'evento e conseguenze ma al prezzo di una semplificazione che non cattura a pieno la variabilità reale.

Per approfondire ulteriormente e concludere il confronto tra eventi osservati e simulati, vengono proposti i grafici di probabilità riportati in figura 141 e 142, nei quali i pallini bianchi rappresentano i valori osservati dell'indice D_m , mentre la curva rossa descrive la distribuzione degli 20000 eventi simulati. Sull'asse delle ascisse è riportata la probabilità di non superamento (in basso) con la corrispondente scala dei tempi di ritorno (in alto), mentre sull'asse y è rappresentato l'indice di magnitudo degli eventi.

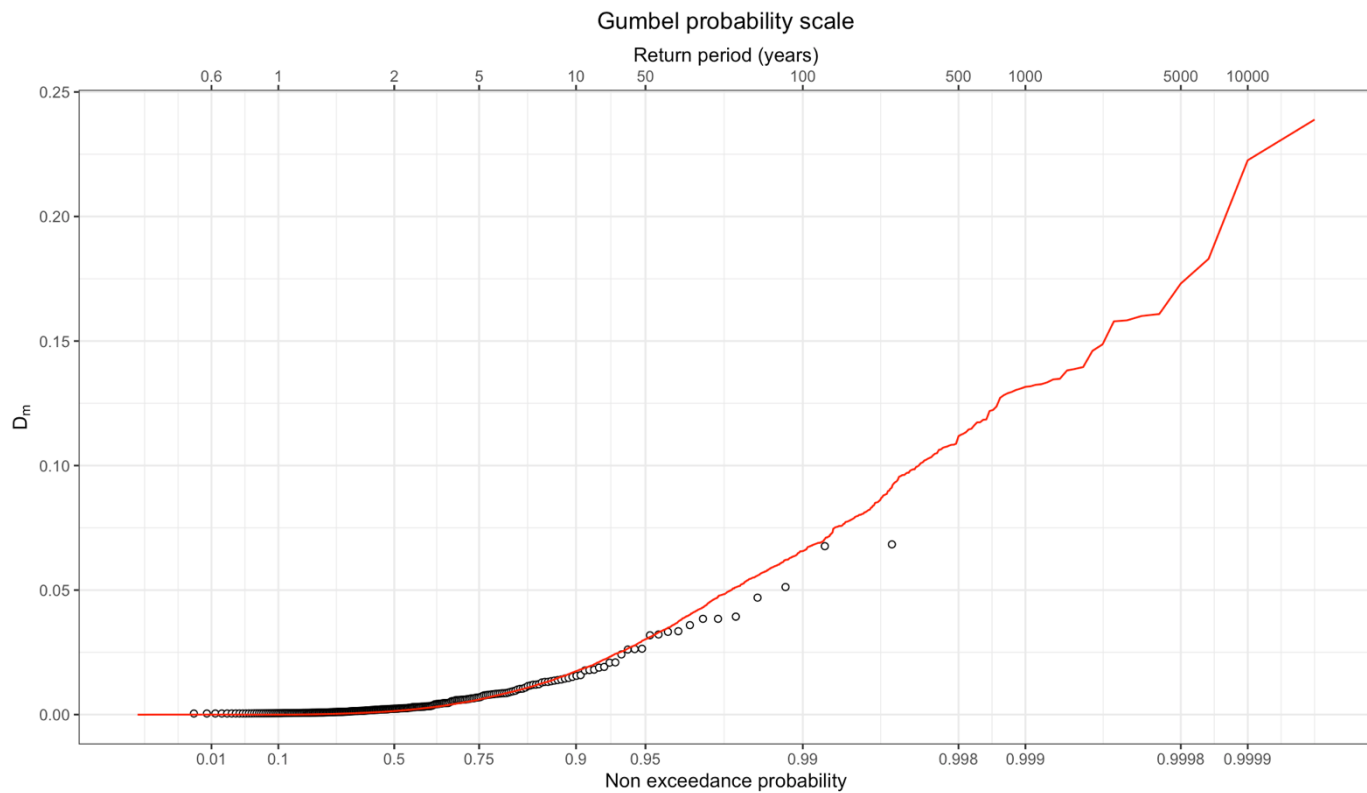


Figura 141- Grafico del confronto tra gli indici di magnitudo degli eventi osservati (pallini bianchi) e degli eventi simulati (curva rossa).

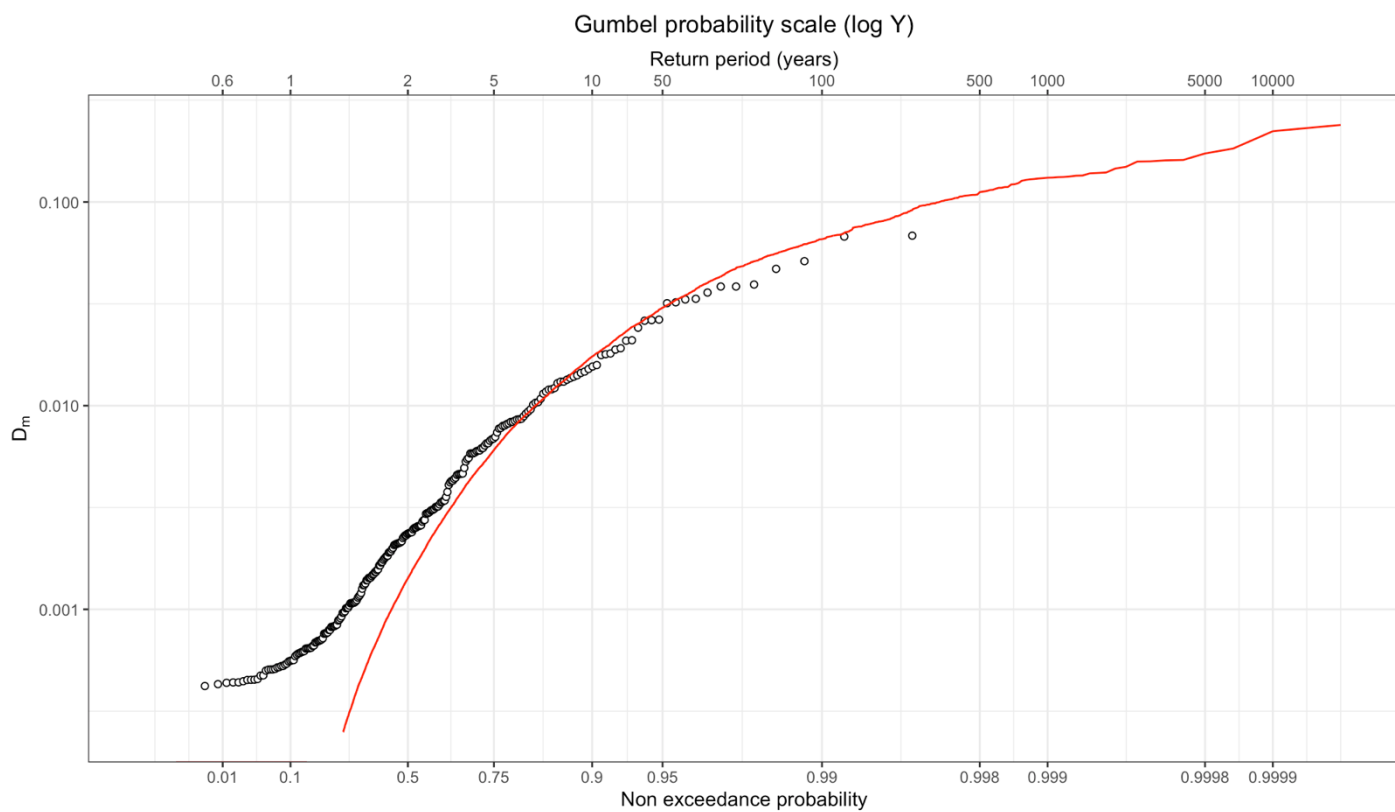


Figura 142- Stesso grafico della figura 141 ma proposto in scala logaritmica.

Dall'analisi risulta esserci una buona coerenza tra le osservazioni e la distribuzione simulata, in particolare per i valori di probabilità intermedie (tra 0.1 e 0.95), dove la curva rossa si sovrappone molto bene ai punti. Si osserva invece che per le code della distribuzione, cioè gli eventi più rari e con tempi di ritorno elevati, i valori dell'indice D_m sono leggermente più elevati rispetto agli osservati (comportamento coerente con i grafici mostrati in precedenza).

Allo stesso tempo in figura 142 tramite l'uso della scala logaritmica si mette meglio in evidenza la divergenza per le basse probabilità, confermando che le simulazioni riescono a riprodurre bene il comportamento medio ma tendono ad amplificare la frequenza e l'intensità degli eventi estremi.

Nel complesso, questi risultati confermano la solidità del metodo per la caratterizzazione statistica e probabilistica del rischio di piena su larga scala, evidenziando al contempo la necessità di integrare ulteriori componenti, ad esempio relativi alla vulnerabilità locale e alla distribuzione degli elementi esposti, per una rappresentazione più realistica delle perdite.

5. Conclusioni.

Il lavoro di ricerca presentato in questo progetto di tesi ha avuto come obiettivo principale l'analisi spaziale della frequenza di eventi estremi nella Greater Alpine Region (GAR), con particolare attenzione agli eventi di piena e alle loro conseguenze sul territorio. La GAR rappresenta un'area ad elevata complessità idrologica e climatica, dove la combinazione tra topografia, variabilità meteorologica e numerosa presenza di insediamenti antropici determina un rischio idrogeologico generalmente elevato. In tale contesto, si è voluto sviluppare un approccio metodologico che fosse in grado di collegare i dati storici delle portate giornaliere, delle stazioni presenti in diversi stati europei, alle dinamiche di danno osservate, così da costruire uno strumento utile sia per l'analisi deterministica del rischio sia per la valutazione di possibili impatti futuri.

Il percorso metodologico è stato sviluppato seguendo una struttura modulare. Nel Capitolo 2, vengono riportati i fondamenti teorici di idrologia statistica impiegati come base del lavoro svolto: è stato mostrato come sono state eseguite le elaborazioni preliminari sui dati, la scelta dei modelli probabilistici appropriati, la stima robusta dei parametri, la verifica dell'adattamento con test statistici e la valutazione dell'incertezza associata alle stime. Questi passaggi hanno garantito una solida base statistica alle fasi successive, permettendo di caratterizzare le code delle distribuzioni delle piene fluviali in ciascuna stazione. Nel Capitolo 3, è stato definito un indice di magnitudo regionale (D_m), sono poi stati estratti e visualizzati gli eventi regionali dai picchi della serie storica dei D_m e sono state condotte ricostruzioni storico-documentali per i quindici eventi più gravosi ovvero quelli che risultavano avere valori di D_m più elevati. Questo capitolo ha fornito l'insieme di eventi che ha guidato la calibrazione e la verifica del modello simulativo.

Infine, nel Capitolo 4, "*Analisi probabilistica di eventi simulati sulla GAR*", si è introdotta la formazione delle serie multivariate, si è parametrizzata la distribuzione normale multivariata derivante dall'unione della distribuzione GP e BurrIII e successivamente si sono simulate popolazioni di dati multivariate riportate poi alla scala originale mediante back-transformation. Si è quindi verificata la qualità del modello tramite confronti statistici e grafici, fino alla produzione delle simulazioni a scala internazionale.

Quindi in primo luogo sono stati identificati e caratterizzati gli eventi estremi che hanno causato danni significativi nel passato e, in secondo luogo, tali informazioni sono state utilizzate per costruire un modello di simulazione basato sulla distribuzione multivariata normale e su

tecniche Monte Carlo. Tale approccio ha consentito non solo di ricreare eventi già osservati, ma anche di esplorare scenari ipotetici, inclusi quelli caratterizzati da tempi di ritorno estremamente elevati (fino a 300.000 anni), che pur non essendo realistici in termini di frequenza forniscono indicazioni utili sulla capacità massima di stress del sistema.

L'analisi ha anche mostrato come eventi con tempi di ritorno dell'ordine dei cento anni, ovvero quelli che realisticamente possono accadere e tendono a divenire più frequenti in un contesto di cambiamento climatico, rivestono un ruolo cruciale per la stima dei danni economici. Essi rappresentano infatti il compromesso tra eventi relativamente probabili e allo stesso tempo sufficientemente distruttivi da generare perdite consistenti e, ad esempio, aumenti significativi dei costi delle polizze che coprono i clienti degli istituti di assicurazioni dai danni economici derivanti da eventi catastrofici.

Le verifiche finali di confronto tra eventi osservati e simulati hanno fornito risultati incoraggianti ma anche evidenziato alcuni limiti. La media geometrica locale dei tempi di ritorno è stata riprodotta con buona accuratezza ($R^2 \approx 0.67$), confermando la validità del modello nella rappresentazione della struttura marginale e spaziale della rarità degli eventi. Inoltre, nelle simulazioni la relazione tra tempo di ritorno e danno medio locale appare più regolare e lineare rispetto agli eventi osservati, indicando che il modello cattura correttamente la dipendenza concettuale tra magnitudo e conseguenze. Nonostante ciò, le simulazioni tendono a sottostimare la variabilità locale e i valori estremi dei danni, con una correlazione relativamente bassa rispetto ai dati osservativi ($R^2 \approx 0.04$). È da contare il fatto che i conteggi assoluti degli eventi risultano molto più elevati nelle simulazioni rispetto alle osservazioni: ciò è spiegabile sia con la maggiore numerosità del campione simulato (20000 eventi) sia con l'inevitabile semplificazione dei processi reali introdotta dal modello probabilistico.

Complessivamente, si può concludere che l'approccio sviluppato si è dimostrato idoneo a ricostruire la struttura multivariata della frequenza delle piene nella GAR e a produrre scenari futuri plausibili. La sua applicazione evidenzia come i metodi probabilistici possano integrare in maniera efficace i dati storici, permettendo sia un'analisi più robusta del rischio idrogeologico sia una valutazione più consapevole dei potenziali costi assicurativi legati a eventi catastrofici. Rimane tuttavia fondamentale affiancare al modello idrologico probabilistico informazioni più dettagliate su vulnerabilità ed esposizione locali, così da tradurre le simulazioni in stime di danno più affidabili e direttamente utilizzabili in ambito gestionale e decisionale. In questa prospettiva, il lavoro costituisce una base solida per sviluppi futuri orientati sia alla ricerca accademica sia al supporto di politiche di adattamento e mitigazione dei rischi naturali nella regione alpina.

6. BIBLIOGRAFIA

Beniston, M., Farinotti, D., Stoffel, M., Andreassen, L. M., Coppola, E., Eckert, N., ... & Vincent, C. (2018). The European mountain cryosphere: A review of its current state, trends, and future challenges. *The Cryosphere*, 12(2), 759–794. <https://doi.org/10.5194/tc-12-759-2018>

EEA – European Environment Agency. (2022). *Climate change adaptation in the agriculture sector in Europe*. Luxembourg: Publications Office of the European Union. <https://www.eea.europa.eu>

IPCC. (2021). *Climate Change 2021: The Physical Science Basis*. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press. <https://www.ipcc.ch/ar6/>

Katz, R. W., & Brown, B. G. (1992). Extreme events in a changing climate: Variability is more important than averages. *Climatic Change*, 21(3), 289–302. <https://doi.org/10.1007/BF00139728>

Merz, B., Aerts, J. C., Arnbjerg-Nielsen, K., Baldi, M., Becker, A., Bichet, A., ... & Weiler, M. (2021). Impact of climate change on flood risk: A review. *Earth's Future*, 9(1), e2020EF001882. <https://doi.org/10.1029/2020EF001882>

MunichRe. (2023). *NatCatSERVICE: Natural catastrophe statistics online*. Munich Reinsurance Company. <https://www.munichre.com>

Trenberth, K. E. (2011). Changes in precipitation with climate change. *Climate Research*, 47(1–2), 123–138. <https://doi.org/10.3354/cr00953>

Van den Besselaar, E. J. M., Klein Tank, A. M. G., & Buishand, T. A. (2013). Trends in European precipitation extremes over 1951–2010. *International Journal of Climatology*, 33(12), 2682–2689. <https://doi.org/10.1002/joc.3619>

Trigila A., Lastoria B., Iadanza C., Bussettini M., Mariani S., D'Ascola F., Salmeri A., Cassese M.L., Pesarino V., Di Paola G., Romeo S., Rischia I., Dessì B., Spizzichino D., Licata V., Gallozzi P.L. (2025) Dissesto idrogeologico in Italia: pericolosità e indicatori di rischio - Edizione 2024. ISPRA, Rapporti 415/2025

Hosking JRM, Wallis JR. Regional Frequency Analysis: An Approach Based on L-Moments. Cambridge University Press; 1997.

“Applied hydrology” di Ven Te Chow, David R. Maidment, Larry W. Mays (ISBN 0-07-010810-2)

Hosking, J. R. M. (1990). L-Moments: Analysis and Estimation of Distributions Using Linear Combinations of Order Statistics. *Journal of the Royal Statistical Society: Series B (Methodological)*, 52(1), 105–124.

Stedinger, J. R., Vogel, R. M., & Foufoula-Georgiou, E. (1993). Frequency analysis of extreme events. In Maidment, D. R. (Ed.), *Handbook of Hydrology*. McGraw-Hill

Laio, F. (2004), Cramer–von Mises and Anderson-Darling goodness of fit tests for extreme value distributions with unknown parameters, *Water Resour. Res.*, 40, W09308, doi:10.1029/2004WR003204.

Serinaldi, F. (2009). *Assessing the applicability of fractional order statistics for flood frequency analysis*. *Journal of Hydrology*, 379(3–4), 312–323.

Kuczera, G. (1999), Comprehensive at-site flood frequency analysis using Monte Carlo Bayesian inference, *Water Resour. Res.*, 35(5), 1551–1557, doi:[10.1029/1999WR900012](https://doi.org/10.1029/1999WR900012).

Blöschl, G., J. Waser, A. Buttinger-Kreuzhuber, D. Cornel, J. Eisl, M. Hofer, M. Hollaus, Z. Horváth, J. Komma, A. Konev, J. Parajka, N. Pfeifer, A. Reithofer, J.L. Salinas, P. Valent, A. Viglione, M.H. Wimmer and H. Stiefelmeyer (2022) HOchwasserRisikozonierung Austria 3.0 (HORA 3.0) (Flood risk zoning Austria 3.0), *Österreichische Wasser- und Abfallwirtschaft*, 74, 212–223, doi:10.1007/s00506-022-00848-7.

FONTI E SITOGRAFIA DELLE FIGURE

Figura 1: Briefing no. 21/2021

Title: *Economic losses and fatalities from weather- and climate-related events in Europe*

HTML - TH-AM-21-018-EN-Q - ISBN: 978-92-9480-419-8 - ISSN: 2467-3196 -

doi:10.2800/7654

PDF - TH-AM-21-018-EN-N - ISBN: 978-92-9480-418-1 - ISSN: 2467-3196 -

doi:10.2800/530599

Figura 2: <https://www.meteoeradar.it/notizie-meteo/>

Figura 3: <https://www.br.de/nachricht/hochwasser-2002-bayern-100.html>

Figura 4: Development of a Glaciological Spatial Data Infrastructure to assess glaciers response to climatic fluctuations - Scientific Figure on ResearchGate. DOI:[10.13140/RG.2.1.1947.6244](https://doi.org/10.13140/RG.2.1.1947.6244).

Figura 5: Blöschl, G., Kiss, A., Viglione, A. *et al.* Current European flood-rich period exceptional compared with past 500 years. *Nature* **583**, 560–566 (2020).

<https://doi.org/10.1038/s41586-020-2478-3>

Figura 6: “Applied hydrology” di Ven Te Chow, David R. Maidment, Larry W. Mays (ISBN 0-07-010810-2)

Figura 11: Nabavi, E., Sabour, M., Dezvareh, G.A. *et al.* Predicting and routing the sub-basin floods and investigating the reservoir dam location in flood risk reduction (case study, Zolachai Dam in West Azerbaijan). *Model. Earth Syst. Environ.* 8, 3793–3811 (2022).

<https://doi.org/10.1007/s40808-021-01330-1>

Figura 14:

http://www.idrologia.polito.it/didattica/Idrologia/2013/Modulo_1_Statistica/DISPENSA_2_distribuzioni_corr_1.pdf

Figura 16: *“Applied hydrology” di Ven Te Chow, David R. Maidment, Larry W. Mays (ISBN 0-07-010810-2)*

Figura 17: Hosking JRM, Wallis JR. Regional Frequency Analysis: An Approach Based on L-Moments. Cambridge University Press; 1997.

Figura 18: Hosking JRM, Wallis JR. Regional Frequency Analysis: An Approach Based on L-Moments. Cambridge University Press; 1997.

Figura 19: <http://www.idrologia.polito.it/didattica/Idrologia/2016/blocco1/Dispensa L-Momenti.pdf>.

Figura 20: Stedinger, J. R., Vogel, R. M., & Foufoula-Georgiou, E. (1993). Frequency analysis of extreme events. In Maidment, D. R. (Ed.), Handbook of Hydrology. McGraw-Hill

Figura 23:
<http://www.idrologia.polito.it/didattica/Idrologia/2016/blocco1/Dispensa Verifica Ipotesi t est 2016.pdf>

7. ALLEGATI

Di seguito viene riportato lo script creato e impiegato in questo progetto di tesi da Filippo Bucca, seguendo lo studio Blöschl et al. (2022):

```
#-----PROGETTO TESI-----#
#-----elaborazione preliminare dei dati-----#
library(zoo)
#carico dati
setwd("/Users/fil_bucca/Desktop/POLITO /5° ANNO/TESI_MAGISTRALE/File_tesi")
data_orig<-load("/Users/fil_bucca/Desktop/POLITO /5°
ANNO/TESI_MAGISTRALE/File_tesi/data_GAR_PASS_KGE.RData")
data_removal <-load("/Users/fil_bucca/Desktop/POLITO /5°
ANNO/TESI_MAGISTRALE/File_tesi/Removal.RData")
data_aree <-load("/Users/fil_bucca/Desktop/POLITO /5°
ANNO/TESI_MAGISTRALE/File_tesi/area_cats_GAR.RData")
backup_data_GAR_PASS_KGE <- data_GAR_PASS_KGE

# Eseguo il filtro: rimuovo da qobs le serie che sono in Removal
ids_to_remove <- as.character(Removal)
data_GAR_PASS_KGE$qobs <- data_GAR_PASS_KGE$qobs[!names(data_GAR_PASS_KGE$qobs)
%in% ids_to_remove]

# Verifica
names(data_GAR_PASS_KGE$qobs)

#seleziono le aree corrispondenti a qobs
nomi_qobs<- names(data_GAR_PASS_KGE$qobs)
```

```

area_cats_GAR_filtrata <- area_cats_GAR[nomi_qobs]
# Rimuove elementi NULL (nel caso in cui alcuni nomi in qobs non siano presenti in
area_cats_GAR_filtrata)
area_cats_GAR_filtrata <- area_cats_GAR_filtrata[!sapply(area_cats_GAR_filtrata, is.null)]

# Estraggo la lista qobs dalla lista principale
qobs_list <- data_GAR_PASS_KGE$qobs

#-----funzione che estrae i due massimi per ogni bacino-----#
extract_q_stats <- function(flow_series) {
  # Estraggo i valori numerici (portate) dal zoo object
  flow_values <- coredata(flow_series)
  sorted_values <- sort(flow_values, decreasing = TRUE, na.last = NA)

  # Estraggo massimo e secondo massimo
  q_max <- if (length(sorted_values) >= 1) sorted_values[1] else NA
  q_max_sec <- if (length(sorted_values) >= 2) sorted_values[2] else NA

  # Calcolo media tra i due
  q_max_mean <- mean(c(q_max, q_max_sec), na.rm = TRUE)

  return(c(q_max = q_max, q_max_sec = q_max_sec, q_max_mean = q_max_mean))
}

# Applico la funzione a ciascun bacino
result_matrix <- t(sapply(qobs_list, extract_q_stats))

# Aggiungo i nomi dei bacini come rownames
if (!is.null(names(qobs_list))) {
  rownames(result_matrix) <- names(qobs_list)
}
# Salvo la matrice in un file CSV
write.csv(result_matrix, file = "qobs_summary.csv", row.names = TRUE)

# Unisco la lista qobs in un'unica matrice zoo (colonne = bacini)
dailyQ <- do.call(merge, qobs_list)

# Verifica struttura
str(dailyQ)

#####
#####
# Fase 1: calcolo osservazioni per anno, definisce eventi grazie al confronto tra due filtri a
media mobile
obsQxanno <- aggregate(dailyQ, by=as.numeric(substr(time(dailyQ), 1, 4)),
  FUN=function(x){sum(!is.na(x))})

ev_x_yr = 2 # eventi per anno

POTmax <- vector('list', ncol(dailyQ))

```

```

names(POTmax) <- colnames(dailyQ)

for(j in 1:ncol(dailyQ)) { # su tutti i bacini
  print(j)
  nyrj <- sum(obsQxanno[,j] > 200) #somma di giorni con dati significativi in un anno deve
  essere maggiore di 200.

  filtro1 <- zoo(filter(dailyQ[,j], rep(1,15)/15), time(dailyQ))
  filtro1[is.na(filtro1)] <- median(filtro1, na.rm=TRUE)

  filtro2 <- zoo(filter(dailyQ[,j], rep(1,100)/100), time(dailyQ))
  filtro2[is.na(filtro2)] <- median(filtro2, na.rm=TRUE)

  ev <- rep(0, length(dailyQ[,j]))
  eventi <- data.frame(start=character(0), end=character(0),
                      tmax=character(0), max=numeric(0),
                      stringsAsFactors=FALSE)
  count = 0
  for (i in 2:length(dailyQ[,j])) {
    if (!is.na(dailyQ[i,j]) & (filtro1[i] > filtro2[i])) {
      if (ev[i-1] == 0) {
        count <- count+1
        eventi[count,] <- NA
        eventi[count,'start'] <- as.character(time(dailyQ[i,j]))
        eventi[count,'end'] <- as.character(time(dailyQ[i,j]))
        eventi[count,'tmax'] <- as.character(time(dailyQ[i,j]))
        eventi[count,'max'] <- dailyQ[i,j]
      } else {
        eventi[count,'end'] <- as.character(time(dailyQ[i,j]))
        if (dailyQ[i,j] > eventi[count,'max']) {
          eventi[count,'tmax'] <- as.character(time(dailyQ[i,j]))
          eventi[count,'max'] <- dailyQ[i,j]
        }
      }
    }
    ev[i] <- count
  }
  POTmax[[j]] <- eventi
}

# Fase 2: selezione dei picchi significativi
POT <- POTmax
POTmax <- vector('list', length(POT))
names(POTmax) <- names(POT)

for (j in 1:length(POT)) {
  nyrj <- sum(obsQxanno[,j] > 200)
  dummy <- POT[[j]]
  dummy <- dummy[order(-dummy$max), c('tmax','max')][1:(ev_x_yr*nyrj), ]
  dummy <- dummy[order(dummy['tmax'], )

```

```

rownames(dummy) <- 1:nrow(dummy)
POTmax[[j]] <- dummy
}
save(POTmax, file = "POT_peaks_qobs_575sites.RData")
load("/Users/fil_bucca/Desktop/POLITO /5°
ANNO/TESI_MAGISTRALE/File_tesi/POT_peaks_qobs_575sites.RData")

#-----scelta del modello probabilistico-----#
pplot <- function(x, distr='Normal', ...) {
  x_i <- sort(x)
  n <- length(x_i)
  i<- 1:n
  F_x_i<- i/(n+1) #weibull plotting position
  if (distr == 'Normal') u_i <- qnorm(F_x_i)
  else if (distr == 'Exp') u_i <- -log(1- F_x_i)
  else if (distr == 'Gumbel') u_i <- -log(-log(F_x_i))
  else u_i <- NA * x_i
  plot(u_i, x_i, ...)
}

##funzioni Gumbel
dgumb <- function(x, theta1 =0, theta2 =1) { #densità probablità
  u <- (x-theta1)/theta2
  d <- exp(-u - exp(-u))/theta2
  return(d)
}

pgumb <- function(q, theta1 =0, theta2 =1) { #probabilità di non superamento
  u <- (q-theta1)/theta2
  p <- exp(-exp(-u))
  return(p)
}

qgumb <- function(p, theta1 =0, theta2 =1) { #quantili
  u <- -log(-log(p))
  q <- theta1 + theta2*u
  return(q)
}

rgumb <- function(n, theta1 =0, theta2 =1) { #genero campione da una gumbel
  p <- runif(n)
  x <- qgumb(p, theta1, theta2)
  return(x)
}

#### Generalized Pareto
dGP <- function (x, x0=0, alpha=1, kappa=-.5) { #densità di probabilità
  y <- 1 - kappa*(x - x0)/alpha
  1/alpha * y^(1/kappa - 1)
}

```



```

pGP <- function (q, x0=0, alpha=1, kappa=-.5) { #probabilità di non superamento
  y <- 1 - kappa*(q - x0)/alpha
  p<- 1 - y^(1/kappa)
  return(p)
}

qGP <- function (p, x0=0, alpha=1, kappa=-.5) { #quantili
  q<- x0 + alpha*(1 - (1 - p)^kappa)/kappa
  return(q)
}

rGP <- function (n, x0=0, alpha=1, kappa=-.5) { #genera campione di dati
  p <- runif(n, min=1e-10, max=1 - 1e-10)
  x<- qGP(p, x0, alpha, kappa)
  return(x)
}

### GEV
dGEV <- function(x, xi=1, sigma=1, kappa=-.5){
  y <- -kappa^(-1) * log(1-kappa * (x-xi)/sigma)
  sigma^(-1) *exp(-(1-kappa) * y - exp(-y))
}

pGEV <- function(q, xi=1, sigma=1, kappa=-.5){
  y <- -kappa^(-1) * log(1-kappa * (q-xi)/sigma)
  exp(-exp(-y))
}

qGEV <- function(p, xi=1, sigma=1, kappa=-.5){
  xi + sigma *(1-(-log(p))^kappa)/kappa
}

rGEV<- function(n, xi=1, sigma=1, kappa=-.5){
  p<- runif(n)
  qGEV(p, xi, sigma, kappa)
}

# Generalised maximum likelihood
library('rstan')
options(mc.cores = parallel::detectCores())
rstan_options(auto_write = TRUE)
code01 <- 'functions {
  real gp_lpdf (real y, real x0, real alpha, real kappa) {
    real z;
    z = 1 - kappa*(y - x0)/alpha;
    return -log(alpha) + (1/kappa - 1)*log(z);
  }
}
data {
  int<lower=1> m; // number of floods exceeding xi
  real y[m];    // piene osservate

```

```

    real hpar[2]; // (mean, sd) of kappa
  }
  transformed data {
    real miny;
    miny = min(y);
  }
  parameters {
    real<lower=0, upper=miny> x0; // soglia inferiore
    real log_alpha;           // log parametro di scala (in log per garantire positività)
    real kappa;               // parametro di forma
  }
  transformed parameters {
    real alpha;
    alpha = exp(log_alpha);
  }
  model {                      //costruzione del modello di verosimiglianza
    // priors on component parameters
    // x0 ~ uniform
    // log_alpha ~ uniform
    kappa ~ normal(hpar[1], hpar[2]);
    for(i in 1:m) {
      y[i] ~ gp(x0, alpha, kappa);
    }
  }
}'

fitGP <- stan_model(model_code=code01, model_name='fitGP') #compilazione del modello GP
in stan

paramParetoGML <- array(NA, dim=c(length(POTmax), 3), dimnames=list(names(POTmax),
c('x0','alpha','kappa')))) #creo tabella vuota con i parametri xo alfa e kappa che compilo con il
ciclo for seguente

for (j in 1:length(POTmax)) {
  print(j)
  sample <- sort(as.numeric(POTmax[[j]]['max']))
  paramParetoGML[j,] <- optimizing(fitGP, data=list(m=length(sample), y=sample, hpar=c(-
0.1,0.122)), # Martins and Stedinger (2000, WRR) #con optimizing stimo i parametri per ogni
bacino
                        init=list(x0=min(sample), log_alpha=log(sd(sample)), kappa=-
0.1))$par[c('x0','alpha','kappa')])
}

#-----calcolo i parametri della GP con gli Lmom-----#
library(lmomco)
library(lmom)
paramPareto_lmom <- array(NA, dim=c(length(POTmax),3),dimnames=list(names(POTmax),
c('xi','alpha','kappa'))))
for (l in 1:length(POTmax)) {
  print(l)

```

```

q_sample <- sort(as.numeric(POTmax[[l]][, 'max']))
# Calcolo L-momenti
lmoments <- lmoms(q_sample)
# Stima parametri Generalized Pareto (GPA nel pacchetto)
params_gpa <- pargpa(lmoments)
param_GP_lmom <- params_gpa[["para"]]
paramPareto_lmom[l,]<- param_GP_lmom
}

# -----GEV - metodo dei momenti e l-momenti-----#
fparGEV <- function (mom) {
  # mom = vettore con c(mu, sigma, gamma)
  t3 <- seq(-0.3333, 1, by=.001)
  CA <- sign(t3) * (-gamma(1 + 3*t3) + 3*gamma(1 + t3)*gamma(1 + 2*t3) - 2*(gamma(1 +
t3))^3)/(gamma(1 + 2*t3) - (gamma(1 + t3))^2)^(3/2)
  T3 <- approx(CA, t3, mom[3])$y
  T2 <- abs(T3) * mom[2]/(gamma(1 + 2*T3) - (gamma(1 + T3))^2)^0.5
  T1 <- mom[1] - T2/T3 * (1 - gamma(1 + T3))
  return(c(T1, T2, T3))
}
skew <- function (x) {
  x_i <- sort(x)
  n <- length(x_i)
  m <- mean(x_i)
  s <- sd(x_i)
  sk <- sum((x_i - m)^3)*n/((n - 1)*(n - 2) * s^3)
  return(sk)
}

param_GEV_mom <- array(NA, dim=c(length(POTmax),3),dimnames=list(names(POTmax),
c('xi_GEV_mom','alpha_GEV_mom','kappa_GEV_mom')))
param_GEV_lmom <- array(NA, dim=c(length(POTmax),3),dimnames=list(names(POTmax),
c('xi_GEV','alpha_GEV','kappa_GEV')))

#GEV metodo dei momenti
for (f in 1:length(POTmax)) {
  print(f)
  q_sample <- sort(as.numeric(POTmax[[f]][, 'max']))
  par_m_GEV <- fparGEV(c(mean(q_sample), sd(q_sample), skew(q_sample)))
  param_GEV_mom[f,]<- par_m_GEV
}
#GEV metodo L-mom
for (g in 1:length(POTmax)) {
  print(g)
  q_sample <- sort(as.numeric(POTmax[[g]][, 'max']))
  par_lm_GEV <- pelgev(samlmu(q_sample))
  param_GEV_lmom[g,]<- par_lm_GEV
}

#-----Creo grafici su carta di Gumbel per ogni bacino-----#

```

```

Tref <- c(2, 5, 10, 20, 50, 100, 200, 500, 1000)

# Cartella output
base_dir<- "gumbel_plots"
if (!dir.exists(base_dir)){
  dir.create("gumbel_plots", showWarnings = FALSE)}

#-----Verifica dei modelli con test di adattamento-----#

AD <- function (n, P_i) {          #variabile test
  -n - (1/n) * sum((seq(1, 2 * n - 1, by = 2)) * log(P_i) +
    (seq(2 * n - 1, 1, by = -2)) * log(1 - P_i))
}

#----- Funzione per "clippare" le CDF tra (eps, 1-eps)-----#
.clip01 <- function(p, eps = 1e-12) pmin(pmax(p, eps), 1 - eps)

#----- Funzione per stimare i parametri GP via L-momenti-----#
safe_pargpa <- function(x) {
  lm <- tryCatch(lmoms(x), error = function(e) NULL)
  if (is.null(lm)) return(NA)
  par <- tryCatch(pargpa(lm)[["para"]], error = function(e) NA)
  if (any(is.na(par))) return(NA)
  return(par)
}

ADtestMC <- function (x, distr='Normal', method='mom', Nsim=1000) {
  x_i <- sort(x)
  n <- length(x_i)
  A2_sim <- rep(NA, Nsim)
  if (method == 'mom') {
    if (distr == 'Normal') {
      par <- c(mean(x_i), sd(x_i))
      P_x_i <- .clip01(pnorm(x_i, mean=par[1], sd=par[2]))
      A2 <- AD(n, P_x_i)
      for (i in 1:Nsim) {
        sim_i <- sort(rnorm(n, mean=par[1], sd=par[2]))
        par_sim <- c(mean(sim_i), sd(sim_i))
        P_sim_i <- .clip01(pnorm(sim_i, mean=par_sim[1], sd=par_sim[2]))
        A2_sim[i] <- AD(n, P_sim_i)
      }
    } else if (distr == 'Exp') {
      par <- mean(x_i)
      P_x_i <- .clip01(pexp(x_i, rate=1/par[1]))
      A2 <- AD(n, P_x_i)
      for (i in 1:Nsim) {
        sim_i <- sort(rexp(n, rate=1/par[1]))
        par_sim <- mean(sim_i)
        P_sim_i <- .clip01(pexp(sim_i, rate=1/par_sim[1]))
        A2_sim[i] <- AD(n, P_sim_i)
      }
    }
  }
}

```



```

}
} else if (distr == 'Gumbel') {
  par <- c(mean(x_i) - 0.5772*sd(x_i)*sqrt(6)/pi, sd(x_i)*sqrt(6)/pi)
  P_x_i <- .clip01(pgumb(x_i, theta1=par[1], theta2=par[2]))
  A2 <- AD(n, P_x_i)
  for (i in 1:Nsim) {
    sim_i <- sort(rgumb(n, theta1=par[1], theta2=par[2]))
    par_sim <- c(mean(sim_i) - 0.5772*sd(sim_i)*sqrt(6)/pi, sd(sim_i)*sqrt(6)/pi)
    P_sim_i <- .clip01(pgumb(sim_i, theta1=par_sim[1], theta2=par_sim[2]))
    A2_sim[i] <- AD(n, P_sim_i)
  }
} else if (distr == 'GEV') {
  par <- fparGEV(c(mean(x_i), sd(x_i), skew(x_i)))
  P_x_i <- .clip01(pGEV(x_i, xi=par[1], sigma=par[2], kappa=par[3]))
  A2 <- AD(n, P_x_i)
  for (i in 1:Nsim) {
    sim_i <- sort(rGEV(n, xi=par[1], sigma=par[2], kappa=par[3]))
    par_sim <- fparGEV(c(mean(sim_i), sd(sim_i), skew(sim_i)))
    P_sim_i <- .clip01(pGEV(sim_i, xi=par_sim[1], sigma=par_sim[2], kappa=par_sim[3]))
    A2_sim[i] <- AD(n, P_sim_i)
  }
}
} else if (method == 'lmom') {
  if (distr == 'Normal') {
    par <- c(samlmu(x_i)[1], sqrt(pi)*samlmu(x_i)[2])
    P_x_i <- .clip01(pnorm(x_i, mean=par[1], sd=par[2]))
    A2 <- AD(n, P_x_i)
    for (i in 1:Nsim) {
      sim_i <- sort(rnorm(n, mean=par[1], sd=par[2]))
      par_sim <- c(samlmu(sim_i)[1], sqrt(pi)*samlmu(sim_i)[2])
      P_sim_i <- .clip01(pnorm(sim_i, mean=par_sim[1], sd=par_sim[2]))
      A2_sim[i] <- AD(n, P_sim_i)
    }
  } else if (distr == 'Exp') {
    cat("\nUse method == \"mom\" that is the same!\n")
  } else if (distr == 'Gumbel') {
    par <- c(samlmu(x_i)[1] - 0.5772*samlmu(x_i)[2]/log(2), samlmu(x_i)[2]/log(2))
    P_x_i <- .clip01(pgumb(x_i, theta1=par[1], theta2=par[2]))
    A2 <- AD(n, P_x_i)
    for (i in 1:Nsim) {
      sim_i <- sort(rgumb(n, theta1=par[1], theta2=par[2]))
      par_sim <- c(samlmu(sim_i)[1] - 0.5772*samlmu(sim_i)[2]/log(2),
samlmu(sim_i)[2]/log(2))
      P_sim_i <- .clip01(pgumb(sim_i, theta1=par_sim[1], theta2=par_sim[2]))
      A2_sim[i] <- AD(n, P_sim_i)
    }
  } else if (distr == 'GEV') {
    par <- pelgev(samlmu(x_i))
    P_x_i <- .clip01(pGEV(x_i, xi=par[1], sigma=par[2], kappa=par[3]))
    A2 <- AD(n, P_x_i)
  }
}

```

```

for (i in 1:Nsim) {
  sim_i <- sort(rGEV(n, xi=par[1], sigma=par[2], kappa=par[3]))
  par_sim <- pelgev(samlmu(sim_i))
  P_sim_i <- .clip01(pGEV(sim_i, xi=par_sim[1], sigma=par_sim[2], kappa=par_sim[3]))
  A2_sim[i] <- AD(n, P_sim_i)
}
} else if (distr == 'GP') {
  par <- safe_pargpa(x_i) #pargpa(lmoms(x_i))["para"]
  if (any(is.na(par))) return(NA)
  P_x_i <- .clip01(pGP(x_i, x0= par[1], alpha= par[2], kappa= par[3]))
  A2 <- AD(n, P_x_i)
  for (i in 1:Nsim) {
    sim_i <- sort(rGP(n, x0=par[1], alpha =par[2], kappa =par[3]))
    par_sim <- safe_pargpa(sim_i)#pargpa(lmoms(sim_i))["para"]
    if (any(is.na(par_sim))) next
    P_sim_i <- .clip01(pGP(sim_i, par_sim[1], par_sim[2], par_sim[3]))
    A2_sim[i] <- AD(n, P_sim_i)
  }
}
} else if (method == 'GML') {
  if (distr == 'GP') {
    par <- optimizing(fitGP, data=list(m=length(x_i), y=x_i, hpar=c(-
0.1,0.122)),init=list(x0=min(x_i), log_alpha=log(sd(x_i)), kappa=-
0.1))$par[c('x0','alpha','kappa')]
    P_x_i <- .clip01(pGP(x_i, x0= par[1], alpha= par[2], kappa= par[3]))
    A2 <- AD(n, P_x_i)
    for (i in 1:Nsim) {
      sim_i <- sort(rGP(n, x0 = par[1], alpha= par[2], kappa=par[3]))
      par_sim <- optimizing(fitGP, data=list(m=length(sim_i), y=sim_i, hpar=c(-
0.1,0.122)),init=list(x0=min(sim_i), log_alpha=log(sd(sim_i)), kappa=-
0.1))$par[c('x0','alpha','kappa')]
      P_sim_i <- .clip01(pGP(sim_i, par_sim[1], par_sim[2], par_sim[3]))
      A2_sim[i] <- AD(n, P_sim_i)
    }
  }
}
signif <- c(0.1, 0.05, 0.01)
A2_signif <- quantile(A2_sim, prob=1 - signif, na.rm=T)
p_value <- 1 - ecdf(A2_sim)(A2)
output <- c(A2=A2, lim=A2_signif, pval=p_value)
return(output)
}

```

```

#-----Funzione per l'analisi dell'incertezza-----#
FFCuncertMC <- function (x, distr='GEV', method='mom', Nsim=1000) {
  u <- seq(-2, 7, by=.1) # gumbel reduced variate
  rp <- c(2, 5, 10, 20, 50, 100, 200, 500, 1000) # return periods
  x_i <- sort(x)
  n <- length(x_i)
  q_u_sim <- matrix(NA, ncol=length(u), nrow=Nsim)

```

```

colnames(q_u_sim) <- sprintf('%.1f', u)
q_rp_sim <- matrix(NA, ncol=length(rp), nrow=Nsim)
colnames(q_rp_sim) <- paste0('T_', rp, 'yrs')
if (method == 'mom') {
  if (distr == 'GEV') {
    par <- fparGEV(c(mean(x_i), sd(x_i), skew(x_i)))
    q_u <- qGEV(exp(-exp(-u)), xi=par[1], sigma=par[2], kappa=par[3])
    q_rp <- qGEV(1 - 1/rp, xi=par[1], sigma=par[2], kappa=par[3])
    for (i in 1:Nsim) {
      sim_i <- sort(rGEV(n, xi=par[1], sigma=par[2], kappa=par[3]))
      par_sim <- fparGEV(c(mean(sim_i), sd(sim_i), skew(sim_i)))
      q_u_sim[i,] <- qGEV(exp(-exp(-u)), xi=par_sim[1], sigma=par_sim[2], kappa=par_sim[3])
      q_rp_sim[i,] <- qGEV(1 - 1/rp, xi=par_sim[1], sigma=par_sim[2], kappa=par_sim[3])
    }
  }
} else if (method == 'lmom') {
  if (distr == 'GEV') {
    par <- pelgev(samlmu(x_i))
    q_u <- qGEV(exp(-exp(-u)), xi=par[1], sigma=par[2], kappa=par[3])
    q_rp <- qGEV(1 - 1/rp, xi=par[1], sigma=par[2], kappa=par[3])
    for (i in 1:Nsim) {
      sim_i <- sort(rGEV(n, xi=par[1], sigma=par[2], kappa=par[3]))
      par_sim <- pelgev(samlmu(sim_i))
      q_u_sim[i,] <- qGEV(exp(-exp(-u)), xi=par_sim[1], sigma=par_sim[2], kappa=par_sim[3])
      q_rp_sim[i,] <- qGEV(1 - 1/rp, xi=par_sim[1], sigma=par_sim[2], kappa=par_sim[3])
    }
  }
} else if (distr == 'GP') {
  par <- pargpa(lmoms(x_i))["para"]
  q_u <- qGP(exp(-exp(-u)), x0= par[1], alpha= par[2], kappa= par[3])
  q_rp <- qGP(1 - 1/rp, x0= par[1], alpha=par[2],kappa= par[3])
  for (i in 1:Nsim) {
    sim_i <- sort(rGP(n, x0= par[1], alpha= par[2], kappa= par[3]))
    par_sim <- pargpa(lmoms(sim_i))["para"]
    q_u_sim[i,] <- qGP(exp(-exp(-u)), par_sim[1], par_sim[2], par_sim[3])
    q_rp_sim[i,] <- qGP(1 - 1/rp,par_sim[1], par_sim[2], par_sim[3])
  }
}
} else if (method == 'GML') {
  if (distr == 'GP') {
    par <- optimizing(fitGP, data=list(m=length(x_i), y=x_i, hpar=c(-
0.1,0.122)),init=list(x0=min(x_i), log_alpha=log(sd(x_i)), kappa=-
0.1))$par[c('x0','alpha','kappa')]
    q_u <- qGP(exp(-exp(-u)), x0= par[1], alpha= par[2], kappa= par[3])
    q_rp <- qGP(1 - 1/rp, x0= par[1], alpha=par[2],kappa= par[3])
    for (i in 1:Nsim) {
      sim_i <- sort(rGP(n, x0= par[1], alpha= par[2], kappa= par[3]))
      par_sim <- optimizing(fitGP, data=list(m=length(sim_i), y=sim_i, hpar=c(-
0.1,0.122)),init=list(x0=min(sim_i), log_alpha=log(sd(sim_i)), kappa=-
0.1))$par[c('x0','alpha','kappa')]
      q_u_sim[i,] <- qGP(exp(-exp(-u)), par_sim[1], par_sim[2], par_sim[3])
    }
  }
}

```

```

    q_rp_sim[i.] <- qGP(1 - 1/rp,par_sim[1], par_sim[2], par_sim[3])
  }
}
}
q_u_sim_out <- apply(q_u_sim, 2, quantile, prob=c(.05,.1,.25,.5,.75,.9,.95), na.rm=T)
q_u_sim_out <- rbind(q_u_sim_out, mean=apply(q_u_sim, 2, mean), point_est=q_u, na.rm=T)
q_rp_sim_out <- apply(q_rp_sim, 2, quantile, prob=c(.05,.1,.25,.5,.75,.9,.95), na.rm=T)
q_rp_sim_out <- rbind(q_rp_sim_out, mean=apply(q_rp_sim, 2, mean), point_est=q_rp,
na.rm=T)
output <- list(u=u, q_u=q_u_sim_out, rp=rp, q_rp=q_rp_sim_out, q_rp_sim=q_rp_sim)
return(output)
}

```

```

#-Ciclo su ogni sito per creazioni prima delle cartelle e poi dei grafici/csv per ogni bacino-#
for (j in 1:nrow(paramParetoGML)) {
  print(j)

```

```

  # Nome del bacino (usato come nome cartella)
  site_name <- rownames(paramParetoGML)[j]
  site_dir <- file.path(base_dir, site_name)
  if (!dir.exists(site_dir)){
    dir.create(file.path(site_dir),showWarnings = FALSE)}

```

```

  # Parametri stimati GP GML
  params <- paramParetoGML[j, ]
  x0 <- params["x0"]
  alpha <- params["alpha"]
  kappa <- params["kappa"]
  site_name <- rownames(paramParetoGML)[j]

```

```

  # Parametri stimati GP lmom
  params <- paramPareto_lmom[j, ]
  xi <- params["xi"]
  alpha_lmom <- params["alpha"]
  kappa_lmom <- params["kappa"]
  site_name <- rownames(paramPareto_lmom)[j]

```

```

  # Parametri stimati GEV_lmom
  params <- param_GEV_lmom[j, ]
  xi_GEV <- params["xi_GEV"]
  alpha_GEV <- params["alpha_GEV"]
  kappa_GEV <- params["kappa_GEV"]
  site_name <- rownames(param_GEV_lmom)[j]

```

```

  # Parametri stimati GEV_mom
  params <- param_GEV_mom[j, ]
  mean_GEV <- params["xi_GEV_mom"]
  sd_GEV <- params["alpha_GEV_mom"]
  skew_GEV <- params["kappa_GEV_mom"]
  site_name <- rownames(param_GEV_mom)[j]

```



```

# Estraggo eventi osservati per il sito
sample_obs <- sort(as.numeric(POTmax[[j]][, 'max']))
n_obs <- length(sample_obs)

# Inizio grafico
png(filename = file.path(site_dir, paste0("Grafico quantili_", site_name, ".png")),
     width = 2000,    # in pixel
     height = 1600,   # in pixel
     res = 300        # dots per inch
)

# Setup plot
op <- par(mar=c(3,3,5,1)+0.03, mgp=c(1.4,0.3,0), tcl=.3, xaxs='i', yaxs='i', las=0, cex=1)
pplot(sample_obs, distr='Gumbel', pch=21, bg='white', cex=0.9,
      xlab='Gumbel reduced variate ()', ylab='Discharge (mc/s)',
      xlim=c(-2, 7), ylim=c(0, max(sample_obs)+50))
axis(3, at=-log(-log(1 - 1/c(1.01,1.1,2,5,10,20,50,100,200,500,1000))),
     labels=c(1.01,1.1,2,5,10,20,50,100,200,500,1000))
mtext("Tempo di ritorno (yrs)", side = 3, line = 1.5, cex = 0.9, font.main=1)
axis(4, labels=F)
title(paste0("Bacino ", site_name, ", area = ", round(area_cats_GAR_filtrata[[j]]/1000000,2), "
km^2"), cex.main=1, font.main=2, line=3)
abline(h=seq(0, 5000, by=50), lty=2, col='grey', lwd=.7)
abline(v=-log(-log(1 - 1/c(1.01,1.1,1.2,1.3,1.4,1.5,1.6,1.7,1.8,1.9,
      2,3,4,5,6,7,8,9,10,20,30,40,50,60,70,80,90,
      100,200,300,400,500,600,700,800,900,1000))),
      lty=2, col='grey', lwd=.7)

curve(qGP(pgumb(x), x0, alpha, kappa), add=T, col=4, lwd=2)
curve(qGP(pgumb(x), xi, alpha_lmom, kappa_lmom), add=T, col="orange", lwd=2)
curve(qGEV(pgumb(x), xi_GEV, alpha_GEV, kappa_GEV), add=T, col='green', lwd=2)
curve(qGEV(pgumb(x), mean_GEV, sd_GEV, skew_GEV), add=T, col='darkgreen', lwd=2)

# Legenda con parametri
#legend("topleft", legend = sprintf("x0 = %.2f\nalpha = %.2f\nkappa = %.3f\nxi =
%.2f\nalpha_lmom = %.2f\nkappa_lmom = %.3f", x0, alpha, kappa, xi, alpha_lmom,
kappa_lmom), bty = 'n', cex = 0.9)
legend("topleft",
      legend = c(
        paste("x0 =", round(x0, 2)),
        paste("alpha =", round(alpha, 2)),
        paste("kappa =", round(kappa, 2)),
        paste("xi =", round(xi, 2)),
        paste("alpha_lmom =", round(alpha_lmom, 2)),
        paste("kappa_lmom =", round(kappa_lmom, 2)),
        paste("xi_GEV =", round(xi_GEV, 2)),
        paste("alpha_GEV =", round(alpha_GEV, 2)),
        paste("kappa_GEV =", round(kappa_GEV, 2)),
        paste("xi_GEV_mom =", round(mean_GEV, 2)),

```

```

paste("alpha_GEV_mom =", round(sd_GEV, 2)),
paste("kappa_GEV_mom =", round(skew_GEV, 2))
),
text.col = c(4,4,4,
'orange','orange','orange','green','green','green','darkgreen','darkgreen','darkgreen'),
bty = "white",
cex = 0.7)

legend("bottomright", legend=c('GP_GML','GP_lmom','GEV_lmom','GEV_mom'), lty=c(1,1,1,1),
lwd=2, col=c(4,'orange','green','darkgreen'), bty='white', cex=0.7)
dev.off()

#costruisco la tabella dei quantili con i tempi di ritorno per ogni distribuzione analizzata
tabellaquantili <- rbind(round(qGP(1 - 1/Tref, x0, alpha, kappa)),
round(qGP(1 - 1/Tref, xi, alpha_lmom, kappa_lmom)),
round(qGEV(1 - 1/Tref, xi_GEV, alpha_GEV, kappa_GEV)),
round(qGEV(1 - 1/Tref, mean_GEV, sd_GEV, skew_GEV)))
rownames(tabellaquantili) <- c('GP_GML','GP_lmom','GEV_lmom','GEV_mom')
colnames(tabellaquantili) <- paste0("T_", Tref, 'yrs')

#salvo csv con tabella quantili
csv_file <- file.path(site_dir, paste0("quantili_", site_name, ".csv"))
write.csv(tabellaquantili, file = csv_file, row.names = TRUE)

#lista di risultati dei test statistici
risultati <- list(
mom_norm = ADtestMC(sample_obs),
mom_lognorm = ADtestMC(log(sample_obs)),
mom_gumbel = ADtestMC(sample_obs, distr='Gumbel'),
mom_gev = ADtestMC(sample_obs, distr='GEV'),
lmom_norm = ADtestMC(sample_obs, method='lmom'),
lmom_lognorm = ADtestMC(log(sample_obs), method='lmom'),
lmom_gumbel = ADtestMC(sample_obs, method='lmom', distr='Gumbel'),
lmom_gev = ADtestMC(sample_obs, method='lmom', distr='GEV'),
lmom_gp = ADtestMC(sample_obs, method='lmom', distr='GP'),
gml_gp = ADtestMC(sample_obs, method='GML', distr='GP'))

# Trasforma la lista in un data frame
risultati_df <- do.call(rbind, lapply(risultati, function(x) as.data.frame(t(x))))
risultati_df <- cbind(Modello=names(risultati), risultati_df)

csv_file_test_stat<- file.path(site_dir, paste0("risultati_ADtest_", site_name, ".csv"))
write.csv(risultati_df, file = csv_file_test_stat, row.names = FALSE)

#Grafico dell'incertezza per ogni bacino
outGEVlm <- FFCuncertMC(sample_obs, distr='GEV', method='lmom')
outGEVmom<-FFCuncertMC(sample_obs, distr='GEV', method='mom')
outGPIm<- FFCuncertMC(sample_obs, distr='GP', method='lmom')
outGP_GML<-FFCuncertMC(sample_obs, distr='GP', method='GML')

```

```

#salvo il grafico dell'incertezza
png(filename = file.path(site_dir, paste0("Incertezza quantili_", site_name, ".png")),
     width = 2000,      # in pixel
     height = 1600,    # in pixel
     res = 300         # dots per inch
)
#plot dell'incertezza con le diverse distribuzioni
op <- par(mar=c(3,3,3,1)+0.03, mgp=c(1.4,0.3,0), tcl=.3, xaxs='i', yaxs='i', las=0, cex=1.3)
pplot(sample_obs, distr='Gumbel', pch=21, bg='white', cex=1.5,
      xlab='Gumbel reduced variate ()', ylab='Discharge (mc/s)',
      xlim=c(-2, 7), ylim=c(0, max(sample_obs)))
axis(3, at=-log(-log(1 - 1/c(1.01,1.1,2,5,10,20,50,100,200,500,1000))),
     labels=c(1.01,1.1,2,5,10,20,50,100,200,500,1000))
axis(4, labels=F)
title('Tempo di ritorno (yrs)', cex.main=1, font.main=1, line=2)
abline(h=seq(0, 5000, by=50), lty=2, col='grey', lwd=.7)
abline(v=-log(-log(1 - 1/c(1.01,1.1,1.2,1.3,1.4,1.5,1.6,1.7,1.8,1.9,
      2,3,4,5,6,7,8,9,10,20,30,40,50,60,70,80,90,
      100,200,300,400,500,600,700,800,900,1000))),
      lty=2, col='grey', lwd=.7)
polygon(c(outGEVmom$u, rev(outGEVmom$u)), c(outGEVmom$q_u['5%'],
rev(outGEVmom$q_u['95%'],])), col='#0000FF22', border=NA)
polygon(c(outGPIm$u, rev(outGPIm$u)), c(outGPIm$q_u['5%'], rev(outGPIm$q_u['95%'],])),
col = "#FFA50022", border=NA)
polygon(c(outGEVlm$u, rev(outGEVlm$u)), c(outGEVlm$q_u['5%'],
rev(outGEVlm$q_u['95%'],])), col = "#32CD3222", border=NA)
polygon(c(outGP_GML$u, rev(outGP_GML$u)), c(outGP_GML$q_u['5%'],
rev(outGP_GML$q_u['95%'],])), col = "#FF000022", border=NA)
curve(qGP(pgumb(x), xi, alpha_lmom, kappa_lmom), add=T, col="orange", lwd=2)
curve(qGEV(pgumb(x), xi_GEV, alpha_GEV, kappa_GEV), add=T, col='green', lwd=2)
curve(qGEV(pgumb(x), mean_GEV, sd_GEV, skew_GEV), add=T, col='blue', lwd=2)
curve(qGP(pgumb(x), x0, alpha, kappa), add=T, col="red", lwd=2)
legend('topleft', legend=c('GP_lmom','GEV_lmom','GEV_mom','GP_GML'), lty=c(1,1,1,1),
lwd=2, col=c('orange','green','blue','red'), bty='white', cex=1)

dev.off()

### salvo in un csv i valori medi e le sd per tutti i tempi di ritorno
mean_T <- round(apply(rbind(outGPIm$q_rp_sim, outGP_GML$q_rp_sim,
outGEVmom$q_rp_sim, outGEVlm$q_rp_sim), 2, mean))
sd_T <-round(apply(rbind(outGPIm$q_rp_sim, outGP_GML$q_rp_sim, outGEVmom$q_rp_sim,
outGEVlm$q_rp_sim), 2, sd))
csv_file_incertezza_sd <- file.path(site_dir, paste0("SD quantili_", site_name, ".csv"))
write.csv(sd_T, file = csv_file_incertezza_sd, row.names = TRUE)
csv_file_incertezza_mean <- file.path(site_dir, paste0("Media quantili_", site_name, ".csv"))
write.csv(mean_T, file = csv_file_incertezza_mean, row.names = TRUE)
}

#-----Grafici riassuntivi dei test statistici-----#

```

```

library(tidyverse)

# cartella principale
main_folder <- "gumbel_plots"

# cerco tutti i file che si chiamano risultati_ADtest_numero.csv
all_files <- list.files(main_folder,
                        pattern="^risultati_ADtest_\\d+\\.csv$",
                        recursive=TRUE, full.names=TRUE)

# leggo e combino tutti i file
all_results <- map_dfr(all_files, function(f) {
  read.csv(f) %>%
    mutate(site = gsub("[^0-9]", "", basename(f))) # estrae numero sito dal nome file
})

#----- mediana p-value e proporzione di test superati-----#
summary_pvals <- all_results %>%
  group_by(Modello) %>%
  summarise(
    median_pval = median(pval, na.rm=TRUE),
    prop_passed = mean(pval > 0.05, na.rm=TRUE),
    n = n()
  ) %>%
  arrange(desc(median_pval))

print(summary_pvals)

# Barplot proporzioni di test superati
ggplot(summary_pvals, aes(x=reorder(Modello, -prop_passed), y=prop_passed)) +
  geom_col(fill="steelblue") +
  geom_text(aes(label=sprintf("%.2f", prop_passed)), vjust=-0.3) +
  labs(
    title = "Proporzione di siti con p-value > 0.05",
    x = "Distribuzione",
    y = "Proporzione"
  ) +
  ylim(0,1) +
  theme(plot.title = element_text(hjust = 0.5)) # <-- titolo centrato

library(dplyr)

#quale distribuzione "vince" più spesso?
winners <- all_results %>%
  group_by(site) %>%
  slice_max(order_by = pval, n = 1, with_ties = FALSE) %>%
  ungroup() %>%
  count(Modello)

```



```

ggplot(winners, aes(x = reorder(Modello, n), y = n, fill = Modello)) +
  geom_col() +
  coord_flip() +
  theme_minimal() +
  labs(
    title = "Distribuzione con p-value più alto per sito",
    x = "Distribuzione",
    y = "Numero di siti in cui è migliore"
  ) +
  theme(legend.position = "none", plot.title = element_text(hjust = 0.5))

#-----Distribuzioni a confronto con il test di Anderson-Darling-----#
library(ggplot2)

ggplot(all_results, aes(x = Modello, y = pval, fill = Modello)) +
  geom_boxplot() +
  geom_hline(yintercept = 0.05, color = "red", linetype = "dashed") + # linea orizzontale
  theme_minimal() +
  labs(
    title = "Distribuzioni a confronto con il test di Anderson-Darling",
    x = "Distribuzione",
    y = "p-value"
  ) +
  theme(
    legend.position = "none",
    plot.title = element_text(hjust = 0.5), # titolo centrato
    axis.text.y = element_text(color = c("red", rep("black", length(unique(all_results$pval))-1)))
  ) +
  scale_y_continuous(
    breaks = c(0.05, pretty(all_results$pval, 5)),
    labels = function(y) ifelse(y == 0.05, "0.05", as.character(y))
  )

#-----grafici riassuntivi analisi dell'incertezza-----#

library(tidyverse)

# prendi tutte le cartelle dei siti
site_dirs <- list.dirs(main_folder, recursive = FALSE)

all_quantili <- list()

for (site in site_dirs) {
  site_id <- basename(site)

  file_quantili <- file.path(site, paste0("quantili_", site_id, ".csv"))
  file_media <- file.path(site, paste0("Media quantili_", site_id, ".csv"))
  file_sd <- file.path(site, paste0("SD quantili_", site_id, ".csv"))

```

```

if (file.exists(file_quantili)) {
  q <- read.csv(file_quantili) %>% mutate(site_id = site_id, type="quantili")
  m <- read.csv(file_media) %>% mutate(site_id = site_id, type="media")
  s <- read.csv(file_sd) %>% mutate(site_id = site_id, type="sd")

  all_quantili[[site_id]] <- bind_rows(q, m, s)
}
}

df_all <- bind_rows(all_quantili)
#-----#
# Leggi i file con le SD
files_sd <- list.files("gumbel_plots", pattern="^SD quantili_.*\\.csv$",
                      recursive=TRUE, full.names=TRUE)

df_sd <- map_dfr(files_sd, ~{
  site_id <- gsub(".*quantili_(\\d+)\\.csv", "\\1", .x)
  df <- read.csv(.x, check.names=FALSE)
  df$site_id <- site_id
  df
})

# Tieni solo colonne numeriche + site_id
df_sd_clean <- df_sd %>%
  rename(T = ...1, SD = x) %>%
  mutate(T = factor(T, levels=c("T_2yrs", "T_5yrs", "T_10yrs", "T_20yrs",
                                "T_50yrs", "T_100yrs", "T_200yrs", "T_500yrs", "T_1000yrs")))

ggplot(df_sd_clean, aes(x=T, y=SD)) +
  geom_boxplot(fill="steelblue", alpha=0.6, outlier.colour="red") +
  theme_bw() +
  labs(
    title="Distribuzione delle deviazioni standard dei quantili",
    x="Tempo di ritorno (anni)",
    y="Deviazione standard (m³/s)"
  ) +
  theme(plot.title = element_text(hjust=0.5)) +
  coord_cartesian(ylim=c(0,1500))

ggplot(df_sd_clean, aes(x=T, y=SD)) +
  geom_boxplot(fill="steelblue", alpha=0.6, outlier.colour="red") +
  theme_bw() +
  labs(
    title="Distribuzione delle deviazioni standard dei quantili (scala log)",
    x="Tempo di ritorno (anni)",
    y="Deviazione standard (m³/s, log scale)"
  ) +
  theme(plot.title = element_text(hjust=0.5)) +
  scale_y_log10()
# -----Leggo tutti i file quantili-----#

```

```

files_quant <- list.files("gumbel_plots", pattern="^quantili.*\\.csv$",
  recursive=TRUE, full.names=TRUE)

df_quant <- map_dfr(files_quant, ~{
  site_id <- gsub(".*quantili_(\\d+)\\.csv", "\\1", .x)
  df <- read.csv(.x, check.names=FALSE)
  df$site_id <- site_id
  df
})

# Porto in long format
df_quant_long <- df_quant %>%
  pivot_longer(
    cols = starts_with("T_"),
    names_to="T",
    values_to="Q"
  )

# Differenza tra modelli (es. GEV_lmom - GP_GML)
df_diff <- df_quant_long %>%
  select(site_id, ...1, T, Q) %>%
  pivot_wider(names_from=...1, values_from=Q) %>%
  mutate(
    diff = GEV_lmom - GP_GML,
    T = factor(T, levels=c("T_2yrs", "T_5yrs", "T_10yrs", "T_20yrs",
      "T_50yrs", "T_100yrs", "T_200yrs", "T_500yrs", "T_1000yrs"))
  )

# --- Grafico 1: con zoom
p1 <- ggplot(df_diff, aes(x=T, y=diff)) +
  geom_boxplot(fill="tomato", alpha=0.6) +
  geom_hline(yintercept=0, linetype="dashed", color="black") +
  theme_bw() +
  labs(
    title="Differenza tra modelli (GEV_lmom - GP_GML)",
    x="Tempo di ritorno (anni)",
    y="Differenza (m³/s)"
  ) +
  theme(plot.title = element_text(hjust=0.5)) +
  coord_cartesian(ylim=c(-500,500)) # zoom sui valori centrali

# --- Grafico 2: scala logaritmica sull'asse y
p2 <- ggplot(df_diff, aes(x=T, y=abs(diff))) + # uso abs() per il log
  geom_boxplot(fill="tomato", alpha=0.6) +
  theme_bw() +
  labs(
    title="Differenza tra modelli (|GEV_lmom - GP_GML|) - scala log",
    x="Tempo di ritorno (anni)",
    y="Differenza assoluta (m³/s, log scale)"
  ) +

```

```

theme(plot.title = element_text(hjust=0.5)) +
scale_y_log10()
print(p1)
print(p2)

#-----GRAFICO DI CONSISTENZA DEI DATI-----#
library(dplyr)
library(stats)
#Estraggo anni dalle date
years<- as.numeric(format(time(dailyQ), "%Y"))

#Matrice logica dei dati disponibili
presenza_logica <- !is.na(dailyQ)

#Calcolo il numero di giorni con dati per ogni stazione e anno
giorni_disponibili <- apply(presenza_logica, 2, function(x) {
  tapply(x, years, sum, na.rm = TRUE)
})

#Calcolo giorni totali per ogni anno (utile in presenza di anni bisestili, ecc.)
giorni_per_anno <- table(years)[rownames(giorni_disponibili)]

#Normalizza: percentuale di giorni con dati (0-1)
percentuale_dati <- sweep(giorni_disponibili, 1, giorni_per_anno, FUN = "/")

#Traspo per plottare: righe = stazioni, colonne = anni
matrice_plot <- t(percentuale_dati)

#Ordino le stazioni per data di inizio osservazione
inizio_dati <- apply(matrice_plot, 1, function(x) min(which(!is.na(x)), na.rm = TRUE))
matrice_plot <- matrice_plot[order(inizio_dati), ]

anni <- as.numeric(colnames(matrice_plot))

# Grafico
n_stazioni <- nrow(matrice_plot)
par(mar = c(4, 5, 3, 5)) # margini: basso, sinistra, alto, destra
image(t(matrice_plot),
      col = gray.colors(100, start = 1, end = 0),
      axes = FALSE, xlab = "", ylab = "")

# Etichette asse X (anni, ogni 10)
ticks_x <- seq(min(anni), max(anni), by = 5)
pos_x <- match(ticks_x, anni) / length(anni)

axis(1, at = pos_x, labels = ticks_x, line = 0.5)
axis(3, at = pos_x, labels = ticks_x, line = 0.5)

# Etichette asse Y (stazioni)
ticks_y <- pretty(1:n_stazioni)

```



```

pos_y <- ticks_y / n_stazioni

axis(2, at = pos_y, labels = ticks_y, las = 1)
axis(4, at = pos_y, labels = ticks_y, las = 1)
box()
mtext("Stazioni", side = 2, line = 2.5)

#confronto la percentuale di copertura in due periodi
periodi <- list("1899-1950" = which(format(time(dailyQ), "%Y") %in% 1899:1950), "1951-1979" = which(format(time(dailyQ), "%Y") %in% 1951:1979), "1980-2024" = which(format(time(dailyQ), "%Y") %in% 1980:2024))

copertura <- sapply(periodi, function(idx) {
  colMeans(!is.na(dailyQ[idx, ]))
})

boxplot(copertura, ylab = "Copertura media", main = "Confronto tra periodi")

#-----mappa geografica della consistenza-----#
# Calcolo percentuale di giorni con dati per ogni stazione
percentuale_copertura <- colMeans(!is.na(dailyQ)) *100 # valori tra 0 e 100

metadata <- read.csv("/Users/fil_bucca/Desktop/POLITO /5°
ANNO/TESI_MAGISTRALE/File_tesi/Shape_tesi/metadata.csv", stringsAsFactors = FALSE)
colnames(metadata) <- c("lat", "lon", "ID")

metadata_sf <- st_as_sf(metadata, coords = c("lat", "lon"), crs = 3035)
metadata_geo <- st_transform(metadata_sf, crs = 4326)
coords <- st_coordinates(metadata_geo)

metadata$lon <- coords[, 1]
metadata$lat <- coords[, 2]
df_mappa <- data.frame(ID = colnames(dailyQ), copertura = percentuale_copertura)
df_mappa <- merge(df_mappa, metadata, by = "ID")

library(sf)
library(ggplot2)
library(rnaturalearth)
library(rnaturalearthdata)
library(leaflet)
library(ggspatial)
library(htmlwidgets)
#confini_europa<- st_read("/Users/fil_bucca/Desktop/POLITO /5°
ANNO/TESI_MAGISTRALE/File_tesi/Europe/Europe_merged.shp")
#confini_europa <- st_transform(confini_europa, crs = 4326)

#palette di colori (dal rosso al verde)
pal <- colorNumeric(palette = "RdYlGn", domain = df_mappa$copertura)

leaflet(df_mappa) %>%

```

```

addProviderTiles(providers$OpenStreetMap) %>% # Sfondo mappa
addCircleMarkers(
  lng = ~lon,
  lat = ~lat,
  radius = 6,
  fillColor = ~pal(copertura),
  color = "black",
  weight = 1,
  fillOpacity = 0.8,
  popup = ~paste0(
    "<b>Stazione:</b> ", ID, "<br>",
    "<b>Copertura:</b> ", round(copertura, 1), "%"
  )
) %>%
addLegend(
  position = "bottomright",
  pal = pal,
  values = ~copertura,
  title = "Copertura (%)",
  opacity = 1
)

#-----CREO INDICE DI MAGNITUDO DEGLI EVENTI SULLA GAR-----
#(HORAKumulGaussian20240122b)

p_u <- dailyQ*NA #matrice zoo con portate giornaliere per ogni sito (colonne = siti, righe =
date)
for(j in 1:ncol(p_u)) { # loop su ogni sito
  print(j)
  dummy <- as.numeric(dailyQ[,j]) #Estrae la serie temporale del sito j come vettore numerico.

  low=paramPareto_lmom[j,'xi']; upp=+Inf; #Definisce i limiti: low= soglia inferiore (minimo
valore che la GP può modellare), upp= soglia superiore, calcolata solo se kappa > 0
  if(paramPareto_lmom[j,'kappa'] > 0) upp=paramPareto_lmom[j,'xi'] +
paramPareto_lmom[j,'alpha']/paramPareto_lmom[j,'kappa']

  quali <- (!is.na(dummy)) & (dummy > low) & (dummy < upp) #Crea un filtro logico per
selezionare i valori validi:interni all'intervallo modellabile [xi, upp]

  p_u[quali,j] <- pGP(dummy[quali], x0=paramPareto_lmom[j,'xi'],
alpha=paramPareto_lmom[j,'alpha'], kappa=paramPareto_lmom[j,'kappa'])
  p_u[(dummy <= low),j] <- 0 #probabilità 0 ai valori sotto la soglia → eventi troppo piccoli
  p_u[(dummy >= upp),j] <- 1 #probabilità 1 ai valori oltre il massimo → eventi troppo grandi
  per essere modellati
}

t_n <- 1/(ev_x_yr*(1 - p_u)) # in yrs, tempo medio tra due eventi con portata maggiore

#pseudo-damage → tempo di ritorno in danno:

```

```

ramp <- function(lT, minlT=2, maxlT=1e4) { #Trasforma il logaritmo naturale del tempo di
ritorno in un valore di pseudo-danno normalizzato tra 0 e 1 con andamento lineare (funzione
a rampa)
  # lT = ln of return period
  lT0 <- lT; lT0[is.na(lT)] <- 0 #eventi frequenti 0 (nessun danno)
  out <- rep(0, length(lT0))
  out[lT0 > minlT] <- (lT0[lT0 > minlT] - minlT)/(maxlT - minlT) #eventi intermedi: ramp
assegna danni crescenti per eventi sempre più rari.
  out[lT0 > maxlT] <- 1 #eventi estremi 1 (danno massimo)
  out[is.na(lT)] <- NA
  return(out)
}
parramp <- c(log(10), log(10000)) # soglia minima: eventi con T < 10 anni → danno nullo;
soglia massima: eventi con T > 10.000 anni → danno massimo
d_n <- t_n*NA

for(j in 1:ncol(d_n)) {
  print(j)
  d_n[,j] <- ramp(log(as.numeric(t_n[,j])), minlT=parramp[1], maxlT=parramp[2])
}

# window()
wd_n <- d_n*NA
for(j in 1:ncol(wd_n)) {
  print(j)
  wd_n[,j] <- rollapply(d_n[,j], width=15, FUN=max, fill=NA)
}

# weighted mean
D_m <- rep(NA, nrow(wd_n))
p=1 # generalized mean parameter (=1 for arithmetic mean, ~0 tends to geometric mean)
log_weights <- log(unlist(area_cats_GAR_filtrata))

for (ii in 1:nrow(wd_n)) {
  D_m[ii] <- weighted.mean(as.numeric(wd_n[ii,])^p, w=log_weights, na.rm=TRUE)^(1/p)
}

#zoo del D-m con il plot
D_m <- zoo(D_m, time(wd_n))
plot(D_m, xlab= 'Anni', main= 'D_m negli anni sulla GAR') #plot di tutti gli eventi con danni
intermedi
abline(h=seq(0, 0.1, by=0.01), lty=2, col='grey', lwd=.7)
years_tick <- seq(1910, 2020, by = 10)
axis(1, at = as.Date(paste0(years_tick, "-01-01")), labels = years_tick)

filtro <- zoo(stats::filter(D_m, rep(1,100)/100), time(D_m)) #filtro passa-basso media mobile
di 100 giorni

```

```

filtro[is.na(filtro)] <- median(filtro, na.rm=TRUE) #sostituisce eventuali NA nel filtro con la
mediana
ev <- rep(0, length(D_m)) #vettore per marcare l'appartenenza di ogni giorno a un evento (0 =
no evento)
eventi <- data.frame(start=character(0), end=character(0),
                     tmaxDm=character(0), maxDm=numeric(0),
                     stringsAsFactors=FALSE)
count=0 #contatore degli eventi trovati
for (i in 2:length(D_m)) { #Ciclo per rilevare eventi
  if (is.na(D_m[i]) & (D_m[i] > filtro[i])) { if (ev[i-1] == 0) {# Nuovo evento
    count <- count+1
    eventi[count,] <- NA
    eventi[count,'start'] <- as.character(time(D_m[i]))
    eventi[count,'end'] <- as.character(time(D_m[i]))
    eventi[count,'tmaxDm'] <- as.character(time(D_m[i]))
    eventi[count,'maxDm'] <- D_m[i]
  } else { # continuazione evento precedente
    eventi[count,'end'] <- as.character(time(D_m[i]))
    if (D_m[i] > eventi[count,'maxDm']) {
      eventi[count,'tmaxDm'] <- as.character(time(D_m[i]))
      eventi[count,'maxDm'] <- D_m[i]
      # c'è un evento se > filtro
    }
  }
  ev[i] <- count # marca il giorno i come parte dell'evento "count"
}
}
# 1 o più eventi per anno, scegli solo i principali eventi
nyr <- length(D_m)/365.25
dummy <- eventi[order(-eventi$maxDm),][1:(ev_x_yr*nyr),]
dummy <- dummy[order(dummy[, 'tmaxDm']),] #ordina cronologicamente gli eventi
selezionati
rownames(dummy) <- 1:nrow(dummy)
eventi <- dummy #aggiorna eventi con solo i più rilevanti
eventi$tmaxDm <- as.Date(eventi$tmaxDm) #date per il plot

# Plot eventi come SERIE STORICA
plot(eventi$tmaxDm, eventi$maxDm, type = "n", xlab = "Anni", ylab =
expression(D[m]~"()"), main= 'Eventi selezionati sulla GAR')
segments(eventi$tmaxDm, 0, eventi$tmaxDm, eventi$maxDm, col = "black")
points(eventi$tmaxDm, eventi$maxDm, pch = 19, col = "red")
abline(h=seq(0, 0.1, by=0.01), lty=2, col='grey', lwd=.7)
abline(v=as.numeric(as.Date(paste0(seq(1900, 2020, by=10), "-01-01"))),
      lty=2, col='grey', lwd=.7)
axis(1, at = as.Date(paste0(years_tick, "-01-01")), labels = years_tick)

summary(eventi$maxDm)

#-----estrazione eventi regionali da picchi nella serie D_m-----#
events <- matrix(NA, nrow=nrow(eventi), ncol=ncol(dailyQ))

```



```

colnames(events) <- colnames(dailyQ)
RPln3_events <- events
d_n_events <- events
for (ii in 1:nrow(eventi)) {
  dummy <- window(dailyQ, start=as.Date(eventi$start[ii]), end=as.Date(eventi$end[ii]))
  events[ii,] <- apply(dummy, 2, max)
  dummy <- window(t_n, start=as.Date(eventi$start[ii]), end=as.Date(eventi$end[ii]))
  RPln3_events[ii,] <- apply(dummy, 2, max)
  dummy <- window(d_n, start=as.Date(eventi$start[ii]), end=as.Date(eventi$end[ii]))
  d_n_events[ii,] <- apply(dummy, 2, max)
}

maxeventi <- eventi[rank(-eventi$maxDm) <= 15,]
maxeventi
max_RPln3_events <- RPln3_events[rank(-eventi$maxDm) <= 15,]
meanRPln3_events <- exp(apply(log(RPln3_events), 2, mean, na.rm=TRUE))
meand_n_events <- apply(d_n_events, 2, mean, na.rm=TRUE)
nevRPgr2 <- apply(RPln3_events > 2, 2, sum, na.rm=TRUE)
nevDgr0 <- apply(d_n_events > 0.00000001, 2, sum, na.rm=TRUE)

#----- Plot dei 15 eventi principali (maxeventi)--> grafico del mese dell'evento -----#
par(mfrow=c(5,3), mar=c(4,4,2,1)) # 5x3 pannelli per i 15 eventi

for (i in 1:nrow(maxeventi)) {
  # finestra temporale attorno all'evento
  start_ev <- as.Date(maxeventi$start[i])
  end_ev <- as.Date(maxeventi$end[i])

  # estrai la serie D_m nella finestra dell'evento
  Dm_event <- window(D_m, start=start_ev-15, end=end_ev+15) # margine ±15 giorni

  # plot
  plot(Dm_event, type="l", lwd=1,
       xlab="", ylab=expression(D[m]~"()"),
       main=format(maxeventi$tmaxDm[i], "%Y"),
       ylim=c(0, max(Dm_event, na.rm=TRUE)*1.1))

  # evidenza il massimo
  abline(v=maxeventi$tmaxDm[i], col="red", lty=2)
  points(maxeventi$tmaxDm[i], maxeventi$maxDm[i], pch=19, col="red")
}
par(mfrow=c(5,3), mar=c(4,4,2,1)) # layout 5x3

#grafico annuale
for (i in 1:nrow(maxeventi)) {
  year_ev <- format(maxeventi$tmaxDm[i], "%Y")

  # estrai tutto l'anno dell'evento
  start_ev <- as.Date(paste0(year_ev, "-01-01"))
  end_ev <- as.Date(paste0(year_ev, "-12-31"))
}

```

```

Dm_event <- window(D_m, start=start_ev, end=end_ev)

# plot singolo
plot(Dm_event, type="l", lwd=1.2,
     xlab="", ylab=expression(D[m]~"()"),
     main=paste("Evento", i, "-", year_ev),
     ylim=c(0, max(Dm_event, na.rm=TRUE)*1.1),
     xaxt="n")

axis(1, at=seq(start_ev, end_ev, by="month"),
     labels=format(seq(start_ev, end_ev, by="month"), "%b"))

abline(v=maxeventi$tmaxDm[i], col="red", lty=2, lwd=1)
points(maxeventi$tmaxDm[i], maxeventi$maxDm[i],
       pch=19, col="red", cex=1.3)

# aspetta che premi Invio prima di passare al prossimo evento
readline(prompt="Premi [Invio] per vedere il prossimo evento...")
}

#-----mappa leaflet dei 15 eventi selezionati-----#
library(tidyr)

# Estraggo gli eventi peggiori (righe da usare)
righe_eventi <- rank(-eventi$maxDm) <= 15

# Creo dataframe per Qmax
qmax_df <- as.data.frame(events[righe_eventi, ])
qmax_df$evento_id <- 1:15
qmax_long <- pivot_longer(qmax_df, cols = -evento_id, names_to = "ID", values_to = "q_max")

# Tempo di ritorno (RP)
rp_df <- as.data.frame(RPIn3_events[righe_eventi, ])
rp_df$evento_id <- 1:nrow(rp_df)
rp_long <- pivot_longer(rp_df, cols = -evento_id, names_to = "ID", values_to = "RP")

# Danno
dn_df <- as.data.frame(d_n_events[righe_eventi, ])
dn_df$evento_id <- 1:nrow(dn_df)
dn_long <- pivot_longer(dn_df, cols = -evento_id, names_to = "ID", values_to = "danno")

# Combina i tre dataframe
df_eventi <- qmax_long %>%
  left_join(rp_long, by = c("evento_id", "ID")) %>%
  left_join(dn_long, by = c("evento_id", "ID")) %>%
  left_join(df_mappa, by = "ID")

# Seleziono l'evento da mappare (es. evento 1-->ciclo for)-----#
library(htmlwidgets)

```

```

fiumi <- st_read("/Users/fil_bucca/Desktop/POLITO /5°
ANNO/TESI_MAGISTRALE/File_tesi/Shape_tesi/fiumiselez.shp")
fiumi <- st_transform(fiumi, crs = 4326)

for (evento_selezionato in 1:15) {
  df_ev <- df_eventi %>% filter(evento_id == evento_selezionato)
  duplicati <- unique(df_ev$ID[duplicated(df_ev$ID)])
  df_ev <- df_ev[!duplicated(df_ev$ID), ]
  Dm <- max(na.omit(df_ev$danno))
  data_evento <- as.character(maxeventi$tmaxDm[evento_selezionato])
  if (all(is.na(df_ev$RP))) {
    pal <- NULL
  } else {
    pal <- colorNumeric("YlOrRd", domain = range(df_ev$RP, na.rm = TRUE), na.color =
"transparent")
  }
  # popup
  df_ev <- df_ev %>%
  mutate(
    qmax_txt = ifelse(is.na(q_max), "Dato non disponibile", paste0(round(q_max, 1), " m³/s")),
    RP_txt = ifelse(is.na(RP), "Dato non disponibile", paste0(round(RP, 1), " anni")),
    danno_txt = ifelse(is.na(danno), "Dato non disponibile", round(danno, 4))
  )
  mappa <- leaflet(df_ev) %>%
  addProviderTiles(providers$OpenStreetMap) %>%
  addPolylines(data = fiumi, color = "blue", weight = 1, opacity = 0.5, group = "Fiumi") %>%
  addCircleMarkers(
    lng = ~lon, lat = ~lat,
    fillColor = if (!is.null(pal)) pal(df_ev$RP) else "grey",
    fillOpacity = 0.9,
    color = "grey",
    radius = 8,
    popup = ~paste0(
      "<b>Stazione:</b> ", ID, "<br>",
      "<b>Qmax:</b> ", qmax_txt, "<br>",
      "<b>RP:</b> ", RP_txt, "<br>",
      "<b>Danno:</b> ", danno_txt
    ),
    popupOptions = popupOptions(autoClose = FALSE, closeOnClick = FALSE)
  ) %>%
  addControl(
    html = paste0("<div style='background-color: white; padding: 8px; border-radius: 4px;'>
      <b>Dm dello scenario:</b> ", round(Dm, 4), "<br>
      <b>Data evento:</b> ", data_evento, "</div>"),
    position = "topright")
  if (!is.null(pal)) {
    mappa <- mappa %>% addLegend("bottomright", pal = pal, values = df_ev$RP, title = "RP
(years)")
  }
}

```

```

  saveWidget(mappa, file = paste0("mappa_evento_", evento_selezionato, ".html"),
selfcontained = TRUE)
}

#-----distribuzione media dei max eventi nell'anno-----#
library(lubridate)

# Estraggo il mese dell'evento
maxeventi <- maxeventi %>%
  mutate(mese = month(tmaxDm, label = TRUE, abbr = FALSE)) # es. "maggio", "giugno"...

# Aggiungo mese numerico (1=gennaio, 12=Dicembre)
maxeventi <- maxeventi %>%
  mutate(
    mese = month(tmaxDm, label = TRUE, abbr = FALSE),
    mese_num = month(tmaxDm)
  )
#Conto eventi per mese e completa i mesi mancanti con 0
eventi_mese <- maxeventi %>%
  count(mese_num) %>%
  complete(mese_num = 1:12, fill = list(n = 0)) %>%
  mutate(mese = factor(month.name[mese_num], levels = month.name))

# Grafico circolare con 12 spicchi
ggplot(eventi_mese, aes(x = mese_num, y = n, fill = mese)) +
  geom_bar(stat = "identity", width = 1, color = "white") +
  geom_text(
    aes(label = ifelse(n > 0, n, "")), # mostra solo se > 0
    position = position_stack(vjust = 1.05),
    size = 4, color = "black", buffer=T,
  ) +
  scale_x_continuous(
    breaks = 1:12,
    labels = month.name,
    expand = c(0,0)
  ) +
  coord_polar(start = 0) +
  labs(
    title = "Distribuzione stagionale degli eventi estremi (maxeventi)",
    x = NULL,
    y = NULL
  ) +
  theme_minimal() +
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1, size = 9),
    axis.text.y = element_blank(),
    plot.title = element_text(hjust = 0.5, face = "bold"),
    legend.position = "none"
  )
#-----Crea tabella per le mappe successive-----#

```



```

df_map <- data.frame(
  ID = names(meanRPln3_events),
  RP_geom = meanRPln3_events,
  D_med = meand_n_events,
  RP_gt2 = nevRPgr2,
  D_gt0 = nevDgr0,
  mean_dam_loc = meand_n_events
) %>%
  left_join(df_mappa, by = "ID")

#-----media geometrica dei RP-----#
df_map <- df_map %>%
  mutate(
    classe_RP_geom = cut(
      RP_geom,
      breaks = c(0, 0.5, 0.7, 0.9, 1.1, 2, 3, Inf),
      labels = c("0.5", "0.7", "0.9", "1.1", "2.0", "3.0", "5+"),
      right = TRUE,
      include.lowest = TRUE
    )
  )
palette_classi_RP_geom <- c(
  "0.5" = "#1a9850",
  "0.7" = "#66bd63",
  "0.9" = "#fee08b",
  "1.1" = "#fdae61",
  "2.0" = "#d73027",
  "3.0" = "#d73027",
  "5.0+" = "#d73027"
)
ggplot(df_map, aes(x = lon, y = lat)) +
  borders("world", xlim = c(5, 20), ylim = c(44, 50), fill = "gray95", colour = "gray60") +
  #geom_sf(data = fiumi, color = "blue", size = 0.3, alpha = 0.6) +
  geom_point(aes(fill = classe_RP_geom, size = classe_RP_geom), shape = 21, color = "black",
stroke = 0.5) +
  scale_fill_manual(values = palette_classi_RP_geom, name = "Media geometrica RP:.",) +
  scale_size_manual(
    name = "Media geometrica RP:",
    values = c(
      "0.5" = 5,
      "0.7" = 6,
      "0.9" = 7,
      "1.1" = 8,
      "2.0" = 9,
      "3.0" = 10,
      "5.0+" = 11)) +
  coord_fixed(
    xlim = c(5, 17), # longitudine (E)
    ylim = c(43.5, 49.5)) + # latitudine (N))
  theme_minimal() +

```

```

labs(title = "Distribuzione della severità degli eventi (GAR)",
      x = "Longitudine", y = "Latitudine")+
theme(plot.title = element_text(hjust = 0.5))

#-----mappa del numero di eventi severi (RP>2)-----#
df_map <- df_map %>%
mutate(
  classe_RP = cut(
    RP_gt2,
    breaks = c(0, 5, 8, 12, 15, 19, 22, 26, Inf),
    labels = c("5", "8", "12", "15", "19", "22", "26", "29+"),
    right = TRUE,
    include.lowest = TRUE
  )
)
palette_classi <- c(
  "5" = "#1a9850",
  "8" = "#66bd63",
  "12" = "#a6d96a",
  "15" = "#d9ef8b",
  "19" = "#fee08b",
  "22" = "#fdae61",
  "26" = "#f46d43",
  "29+" = "#d73027"
)

ggplot(df_map, aes(x = lon, y = lat)) +
  borders("world", xlim = c(5, 20), ylim = c(44, 50), fill = "gray95", colour = "gray60") +
  geom_point(aes(fill = classe_RP, size = classe_RP), shape = 21, color = "black", stroke = 0.4) +
  scale_fill_manual(values = palette_classi, name = "n. of events\nhaving RP > 2:") +
  scale_size_manual(
    values = c(5, 6, 7, 8, 9, 10, 11, 12),
    name = "n. of events\nhaving RP > 2:") +
  coord_fixed(
    xlim = c(5, 17),
    ylim = c(43.5, 49.5)) +
  theme_minimal() +
  labs(
    title = "Frequenza eventi gravi per stazione (GAR)",
    x = "Longitudine", y = "Latitudine") +
  theme(legend.position = "right", plot.title = element_text(hjust = 0.5))
#-----mappa con conteggio eventi con D>0-----#
df_map <- df_map %>%
mutate(
  classe_D = cut(
    D_gt0,
    breaks = c(-1, 0, 1, 3, 4, 6, 7, 9, 10, Inf),
    labels = c("0", "1", "3", "4", "6", "7", "9", "10", "11+"),
    include.lowest = TRUE,
    right = TRUE
  )
)

```

```

)
)

palette_danno <- c(
  "0" = "#006837",
  "1" = "#1a9850",
  "3" = "#66bd63",
  "4" = "#d9ef8b",
  "6" = "#fee08b",
  "7" = "#fdae61",
  "9" = "#f46d43",
  "10" = "#d73027",
  "11+" = "#762a83")

ggplot(df_map, aes(x = lon, y = lat)) +
  borders("world", xlim = c(5, 20), ylim = c(44, 50), fill = "gray95", colour = "gray60") +
  geom_point(aes(fill = classe_D, size = classe_D), shape = 21, color = "black", stroke = 0.4) +
  scale_fill_manual(values = palette_danno, name = "n. of events\nhaving D > 0:") +
  scale_size_manual(
    values = c(3, 4, 5, 6, 7, 8, 9, 10, 11),
    name = "n. of events\nhaving D > 0:") +
  coord_fixed(
    xlim = c(5, 17),
    ylim = c(43.5, 49.5)) +
  theme_minimal() +
  labs(title = "Frequenza eventi con danno > 0 per stazione", x = "Longitudine", y =
"Latitudine") +
  theme(legend.position = "right", plot.title = element_text(hjust = 0.5))

#-----mappa con media del danno locale-----#
df_map <- df_map %>%
  mutate(
    classe_meand = cut(
      mean_dam_loc,
      breaks = c(-0.001, 0, 0.002, 0.004, 0.007, 0.009, 0.011, 0.013, 0.016, Inf),
      labels = c("0", "0.002", "0.004", "0.007", "0.009", "0.011", "0.013", "0.016", ">0.016"),
      include.lowest = TRUE,
      right = TRUE
    )
  )

palette_meand <- c(
  "0" = "#006837",
  "0.002" = "#1a9850",
  "0.004" = "#66bd63",
  "0.007" = "#d9ef8b",
  "0.009" = "#fee08b",
  "0.011" = "#fdae61",
  "0.013" = "#f46d43",

```

```

"0.016" = "#d73027",
">0.016" = "#762a83"
)

ggplot(df_map, aes(x = lon, y = lat)) +
  borders("world", xlim = c(5, 20), ylim = c(44, 50), fill = "gray95", colour = "gray60") +
  geom_point(aes(fill = classe_meand, size = classe_meand), shape = 21, color = "black", stroke
= 0.3) +
  scale_fill_manual(
    name = "Danno medio locale:",
    values = palette_meand,
    drop = FALSE) +
  scale_size_manual(
    name = "Danno medio locale:",
    values = c(
      "0" = 2,
      "0.002" = 3,
      "0.004" = 4,
      "0.007" = 5,
      "0.009" = 6,
      "0.011" = 7,
      "0.013" = 8,
      "0.016" = 9,
      ">0.016" = 10
    ))+
  coord_fixed(xlim = c(5, 17), ylim = c(43.5, 49.5)) +
  theme_minimal() +
  labs(title = "Valore medio di danno per stazione (GAR)", x = "Longitudine", y = "Latitudine")+
  theme(plot.title = element_text(hjust = 0.5))
#-----#
#parte complessa sulla simulazione di eventi estremi sulla GAR tramite Montecarlo#

# Trasformazione
dBrIII <- function (x, beta, gamma1, gamma2) {
  d <- 1/beta * (1 + 1/gamma1 * (x/beta)^(-1/gamma2))^(-gamma1*gamma2 - 1) *
(x/beta)^(-1/gamma2 - 1)
  return(d)
}
pBrIII <- function (x, beta, gamma1, gamma2) {
  p <- (1 + 1/gamma1 * (x/beta)^(-1/gamma2))^(-gamma1*gamma2)
  return(p)
}
qBrIII <- function (p, beta, gamma1, gamma2) {
  q <- beta * (gamma1 * (p^(-1/(gamma1*gamma2)) - 1))^(-gamma2)
  return(q)
}
rBrIII <- function (n, beta, gamma1, gamma2) {
  p <- runif(n, min=0.0000000001, max=0.9999999999)
  x <- qBrIII(p, beta, gamma1, gamma2)
  return(x)
}

```



```

}

library('rstan') #(HORAkumulGaussian20240122b)
options(mc.cores = parallel::detectCores())
rstan_options(auto_write = TRUE)
# truncated normal priors for shape parameters
code03 <- 'functions {
    real BrIII_lpdf (real x, real beta, real gamma1, real gamma2) {
        return -log(beta) + (-gamma1*gamma2 - 1)*log(1 + 1/gamma1 * (x/beta)^(-
1/gamma2)) + (-1/gamma2 - 1)*log(x/beta);
    }
}
data {
    int<lower=0> N;
    real<lower=0> x[N];
    real hpar[4]; // (mean_gamma1, sd_gamma1, mean_gamma2, sd_gamma2)
}
parameters {
    real log_beta;
    real<lower=0> gamma1;
    real<lower=0> gamma2;
}
transformed parameters {
    real beta;
    beta = exp(log_beta);
}
model {
    // priors on component parameters
    // log_beta ~ uniform
    // gamma1 ~ truncated normal
    gamma1 ~ normal(hpar[1], hpar[2]);
    // gamma2 ~ truncated normal
    gamma2 ~ normal(hpar[3], hpar[4]);
    for(i in 1:N) {
        x[i] ~ BrIII(beta, gamma1, gamma2);
    }
}'

BrIIIfit <- stan_model(model_code=code03, model_name='BrIIIfit')
param <- matrix(NA, nrow=ncol(events), ncol=3)
colnames(param) <- c('beta','gamma1','gamma2')

for (j in 1:ncol(events)) {
    print(j)
    dummy <- sort(events[,j]) # to avoid NAs
    dummy <- dummy[dummy > 0.000000001] # to avoid 0s
    param[j,] <- optimizing(BrIIIfit, data=list(N=length(dummy), x=dummy, hpar=c(0,10,
0.2,0.3)), # weak priors
        init=list(log_beta=log(mean(dummy)), gamma1=0.2,
gamma2=0.2))$par[c('beta','gamma1','gamma2')]
}

```

```

pGPBrIII_max <- function (x, x0, alpha, kappa, beta, gamma1, gamma2) {
  p1 <- 1 - (1 - kappa*(x - x0)/alpha)^(1/kappa) #CDF collegata al modello Pareto
  p2 <- (1 + 1/gamma1 * (x/beta)^(-1/gamma2))^(-gamma1*gamma2) #CDF della Burr III
  apply(cbind(p1, p2), 1, max) #la probabilità cumulativa massima fra le due stime: scelta
  conservativa per far sì che la probabilità assegnata ad un valore x sia la maggiore tra le due
  CDF
}
trnsevents <- events*NA
for (j in 1:ncol(events)) {
  dummy <- events[,j]
  dummy[dummy < 0.000000001] <- NA # evito gli 0
  ppp <- pGPBrIII_max(dummy, x0=paramParetoGML[j,'x0'], alpha=paramParetoGML[j,'alpha'],
  kappa=paramParetoGML[j,'kappa'],
  beta=param[j,'beta'], gamma1=param[j,'gamma1'],
  gamma2=param[j,'gamma2'])
  trnsevents[,j] <- qnorm(ppp)
}

# Creo un dataframe combinato per il plotting xo_beta
df_comp <- data.frame(
  log_x0 = log(paramParetoGML[, "x0"]),
  log_beta = log(param[, "beta"]))

# Plot con retta di regressione
library(ggpubr)
ggplot(df_comp, aes(x = log_x0, y = log_beta)) +
  geom_point(color = "blue", size = 2) + # punti
  geom_smooth(method = "lm", color = "red", se = TRUE) + # retta di regressione
  stat_regline_equation( # aggiunge l'equazione della retta
    aes(label = paste(..eq.label.., ..rr.label.., sep = "~~~")),
    label.x = min(df_comp$log_x0) + 0.1, # posizione
    label.y = max(df_comp$log_beta) - 0.1,
    size = 4) +
  theme_minimal() +
  labs(
    title = "Confronto tra x0 (Pareto) e beta (BrIII)",
    x = "x0 of maximum annual discharges (m^3/s)(ParamParetoGML)",
    y = "beta of max GAR events (m^3/s) (param)",
    caption = "Linea rossa: retta di regressione lineare")+
  theme(plot.title = element_text(hjust = 0.5))

# Creo un dataframe combinato per il plotting alpha_gamma1
df_comp1 <- data.frame(
  log_alpha = log(paramParetoGML[, "alpha"]),
  log_gamma1 = log(param[, "gamma1"]))

# Plot con retta di regressione
ggplot(df_comp1, aes(x = log_alpha, y = log_gamma1)) +
  geom_point(color = "purple", size = 2) + # punti

```

```

geom_smooth(method = "lm", color = "green", se = TRUE) + # retta di regressione
stat_regline_equation( # aggiunge l'equazione della retta
  aes(label = paste(..eq.label.., ..rr.label.., sep = "~~~")),
  label.x = min(df_comp1$log_alpha) + 0.1, # posizione
  label.y = max(df_comp1$log_gamma1) - 0.1,
  size = 4) +
theme_minimal() +
labs(
  title = "Confronto tra alpha (Pareto) e gamma1 (BrIII)",
  x = "alpha of maximum annual discharges (m^3/s)(ParamParetoGML)",
  y = "gamma1 of max GAR events (param)",
  caption = "Linea verde: retta di regressione lineare")+
theme(plot.title = element_text(hjust = 0.5))

```

```

# Creo un dataframe combinato per il plotting k_gamma2
df_comp2 <- data.frame(
  kappa = paramParetoGML[, "kappa"],
  gamma2 = param[, "gamma2"])

```

```

# Plot con retta di regressione
ggplot(df_comp2, aes(x = kappa, y = gamma2)) +
  geom_point(color = "cyan", size = 2) + # punti
  geom_smooth(method = "lm", color = "orange", se = TRUE) + # retta di regressione
  stat_regline_equation( # aggiunge l'equazione della retta
    aes(label = paste(..eq.label.., ..rr.label.., sep = "~~~")),
    label.x = min(df_comp2$kappa) + 0.1, # posizione
    label.y = max(df_comp2$gamma2) - 0.1,
    size = 4) +
  theme_minimal() +
  labs(
    title = "Confronto tra kappa (Pareto) e gamma2 (BrIII)",
    x = "kappa of maximum annual discharges (ParamParetoGML)",
    y = "gamma2 of max GAR events (param)",
    caption = "Linea arancione: retta di regressione lineare")+
  theme(plot.title = element_text(hjust = 0.5))

```

#-----Parametrizzazione della distribuzione normale multivariata-----#

```

m_trnsevents <- apply(trnsevents, 2, mean, na.rm=TRUE)*0
S_trnsevents <- cor(trnsevents, use='pairwise.complete.obs')
S_trnsevents[is.na(S_trnsevents)] <- 0
det(S_trnsevents)

```

```

require(matrixcalc)
is.positive.definite(S_trnsevents) #esce FALSE

```

```

is.positive.semi.definite(S_trnsevents) #esce FALSE

```

```

eigen(S_trnsevents)$values # molti valori negativi

```

```

library(Matrix)
S_trnsevents_R <- nearPD(S_trnsevents) #metodo per migliorare la stima della matrice di
covarianza. In questo caso viene utilizzata la funzione nearPD del pacchetto R Matrix, che
implementa l'algoritmo di Higham (2002).
det(S_trnsevents_R$mat) # 0
require(matrixcalc)
is.positive.definite(as.matrix(S_trnsevents_R$mat))#ora esce TRUE
is.positive.semi.definite(as.matrix(S_trnsevents_R$mat))#ora esce TRUE

round(quantile(S_trnsevents - as.matrix(S_trnsevents_R$mat)), 5) #calcola la differenza tra la
matrice originale e la matrice corretta
# 0% 25% 50% 75% 100%
#-2.48143 -0.00900 0.00041 0.00972 0.40643
mean(diag(S_trnsevents)) #esce 1

mean(diag(S_trnsevents_R$mat))# uguale a 1.144969, nearPD ha alterato le diagonali.
#Poiché la matrice non è una matrice di correlazione, la riscaliamo affinché lo
diventi."Scaling" qui significa: convertire la matrice di covarianza in una matrice di
correlazione

invsigg <- diag(1/sqrt(diag(S_trnsevents_R$mat)))
S_trnsevents2 <- invsigg %*% as.matrix(S_trnsevents_R$mat) %*% t(invsigg)
is.symmetric.matrix(S_trnsevents2) #esce FALSE

is.symmetric.matrix(round(S_trnsevents2, 10))#esce TRUE arrotondando.

S_trnsevents2 <- round(S_trnsevents2, 10)
det(S_trnsevents2) # è 0

is.positive.definite(as.matrix(S_trnsevents2)) #esce TRUE

is.positive.semi.definite(as.matrix(S_trnsevents2)) #esce TRUE

eigen(S_trnsevents2)$values
round(quantile(S_trnsevents - as.matrix(S_trnsevents2)), 5)
#0% 25% 50% 75% 100%
#-0.67332 -0.00400 0.02345 0.05453 0.67465
mean(diag(S_trnsevents2)) # è 1

#-----GRAFICI MATRICE DI CORRELAZIONE-----#
library(corrplot)
library(gridExtra)
library(grid)
# Plot con corrplot
corrplot(S_trnsevents2,
         method = "color", # usa colori
         col = colorRampPalette(c("red", "white", "blue"))(200), # rosso = negativo, blu = positivo
         is.corr = TRUE, # indica che è una matrice di correlazione
         tl.pos = "n", # nasconde le etichette delle variabili (troppo affollato)
         cl.cex = 0.6, # dimensione della legenda

```



```

diag = TRUE,                # mostra la diagonale (1.0)
addgrid.col = NA)          # rimuove la griglia

library(reshape2)
n <- nrow(S_trnsevents2)
rownames(S_trnsevents2) <- 1:n
colnames(S_trnsevents2) <- 1:n
df_corr <- melt(S_trnsevents2)

# Plot
ggplot(df_corr, aes(x = Var2, y = Var1, fill = value)) +
  geom_tile() +
  scale_fill_gradient2(
    low = "red", mid = "white", high = "blue",
    midpoint = 0, limit = c(-1, 1), name = "Correlazione"
  ) +
  scale_x_continuous(breaks = seq(0, n, by = 100)) +
  scale_y_continuous(breaks = seq(0, n, by = 100)) +
  coord_fixed() +
  theme_minimal(base_size = 10) +
  theme(
    axis.title = element_blank(),
    axis.text = element_text(size = 6),
    panel.grid = element_blank())

#-----generazione dei dati simulati e backtransformation-----#

NYRsim=10000
library(MASS)
simtrnsevents <- mvrnorm(n=ev_x_yr*NYRsim, mu=m_trnsevents, Sigma=S_trnsevents2)
qGPBrIIImin <- function (p, x0, alpha, kappa, beta, gamma1, gamma2) {
  q1 <- x0 + alpha*(1 - (1 - p)^kappa)/kappa
  q2 <- beta * (gamma1 * (p^(-1/(gamma1*gamma2)) - 1))^(-gamma2)
  apply(cbind(q1, q2), 1, min)
}
simevents <- simtrnsevents*NA
for (j in 1:ncol(simtrnsevents)) {
  print(j)
  dummy <- pnorm(simtrnsevents[,j])
  simevents[,j] <- qGPBrIIImin(dummy, x0=paramParetoGML[j,'x0'],
alpha=paramParetoGML[j,'alpha'], kappa=paramParetoGML[j,'kappa'],
                             beta=param[j,'beta'], gamma1=param[j,'gamma1'],
gamma2=param[j,'gamma2'])
}

#----- grafici confornto sim e obs-----#

#library(scales)

```

```

clamp_p <- function(p, eps = 1e-12) pmax(pmin(p, 1 - eps), eps)

# CDF e survival della Burr III
pmBurrIII <- function(x, beta, gamma1, gamma2) {
  (1 + 1/gamma1 * (x/beta)^(-1/gamma2))^(-gamma1*gamma2)
}

# CDF Pareto
pParetoGML <- function(x, x0, alpha, kappa) {
  # ensure x >= x0
  p <- rep(NA_real_, length(x))
  idx <- x >= x0
  p[idx] <- 1 - (1 - kappa*(x[idx] - x0)/alpha)^(1/kappa)
  p[!idx] <- 0
  p
}

# Empirical survival (1-CDF) helper
emp_surv <- function(vec, probs = NULL) {
  vec <- vec[!is.na(vec)]
  n <- length(vec)
  if (is.null(probs)) probs <- seq(1, n)/(n+1)
  q <- quantile(vec, probs = probs, names = FALSE, type=8)
  data.frame(x = q, surv = 1 - probs)
}

plot_hist_density <- function(j, observed, simulated, out_file = NULL, sample_size = 10000) {
  obs <- observed[, j]
  obs <- obs[!is.na(obs)]
  sim <- simulated[, j]
  sim <- sim[!is.na(sim)]
  # campiona per velocità se troppo grande
  if(length(sim) > sample_size) sim <- sample(sim, sample_size)

  df <- rbind(
    data.frame(value = obs, type = "Observed"),
    data.frame(value = sim, type = "Simulated")
  )

  p <- ggplot(df, aes(x = value, fill = type)) +
    geom_histogram(aes(y = ..density..), position = "identity", alpha = 0.5, bins = 60) +
    geom_density(aes(color = type), size = 0.8) +
    scale_x_continuous(trans = 'log10', labels = scales::comma) + # spesso utile scala log per
portate
    labs(title = paste("Station", j, "- Observed vs Simulated"),
         x = "Portata (scala log10)", y = "Densità") +
    theme_minimal()

  if(!is.null(out_file)) ggsave(out_file, plot = p, width = 8, height = 5)
  return(p)
}

```

```

}

plot_qq <- function(j, observed, simulated, out_file = NULL, probs = seq(0.001, 0.999,
length.out=1000)) {
  obs <- sort(observed[!is.na(observed[,j]), j])
  sim <- sort(simulated[!is.na(simulated[,j]), j])
  # scegli quantili comuni (interpolazione se necessario)
  p <- probs
  qobs <- quantile(obs, p, names = FALSE, type=8)
  qsim <- quantile(sim, p, names = FALSE, type=8)
  df <- data.frame(obs = qobs, sim = qsim)
  p1 <- ggplot(df, aes(x = obs, y = sim)) +
    geom_point(alpha = 0.4, size = 0.8) +
    geom_abline(slope = 1, intercept = 0, color = "red", linetype = "dashed") +
    scale_x_continuous(trans = 'log10', labels = scales::comma) +
    scale_y_continuous(trans = 'log10', labels = scales::comma) +
    labs(title = paste("QQ-plot Station", j), x = "Quantili osservati", y = "Quantili simulati") +
    theme_minimal()
  if(!is.null(out_file)) ggsave(out_file, plot = p1, width = 6, height = 6)
  return(p1)
}

plot_tail_compare <- function(j, observed, simulated, pareto_params, burr_params, out_file =
NULL, npoints = 200) {
  obs <- observed[!is.na(observed[,j]), j]
  sim <- simulated[!is.na(simulated[,j]), j]
  # Empirical survival
  emp_obs <- emp_surv(obs)
  emp_sim <- emp_surv(sim)

  # Model survival curves from fitted parametri
  x_min <- min(c(min(obs, na.rm=TRUE), min(sim, na.rm=TRUE)), na.rm = TRUE)
  x_max <- quantile(c(obs, sim), 0.999, na.rm = TRUE) * 1.2
  xs <- exp(seq(log(x_min + 1e-9), log(x_max), length.out = npoints))

  par <- pareto_params[j, ] # assumes columns x0, alpha, kappa
  burr <- burr_params[j, ] # assumes columns beta, gamma1, gamma2

  S_pareto <- 1 - pParetoGML(xs, par['x0'], par['alpha'], par['kappa'])
  S_burr <- 1 - pmBurrIII(xs, burr['beta'], burr['gamma1'], burr['gamma2'])

  df_plot <- data.frame(x = c(emp_obs$x, emp_sim$x, xs, xs),
    surv = c(emp_obs$surv, emp_sim$surv, S_pareto, S_burr),
    source = c(rep("Empirical Obs", nrow(emp_obs)),
      rep("Empirical Sim", nrow(emp_sim)),
      rep("Pareto model", length(xs)),
      rep("BurrIII model", length(xs))))

  p <- ggplot(df_plot, aes(x = x, y = surv, color = source, linetype = source)) +
    geom_line() +

```

```

scale_x_log10(labels = scales::comma) + scale_y_log10() +
labs(title = paste("Tail comparison - Station", j),
     x = "Portata (log scale)", y = "Survival function (log scale)") +
theme_minimal()

if(!is.null(out_file)) ggsave(out_file, plot = p, width = 7, height = 5)
return(p)
}

plot_return_level <- function(j, observed, simulated, nyears_obs, NYRsim, out_file = NULL) {
  # calcola i massimi annuali osservati e simulati e i loro livelli di ritorno
  obs <- observed[!is.na(observed[,j]), j]
  sim <- simulated[, j]
  probs <- seq(0.001, 0.999, length.out = 1000)
  qobs <- quantile(obs, probs, names=FALSE, type=8)
  # trattandosi di eventi per anno: return period T = 1/(1-p)
  Tobs <- 1/(1 - probs)

  qsim <- quantile(sim, probs, names=FALSE, type=8)
  Tsim <- 1/(1 - probs) * (NYRsim / 1) # adattamento se necessario

  df <- data.frame(T = c(Tobs, Tsim), Q = c(qobs, qsim),
                  src = rep(c("Observed", "Simulated"), each = length(probs)))
  p <- ggplot(df, aes(x = T, y = Q, color = src)) +
    geom_line() + scale_x_log10() + scale_y_log10() +
    labs(title = paste("Return level plot - Station", j),
         x = "Return period (years, log scale)", y = "Return level (flow)") +
    theme_minimal()
  if(!is.null(out_file)) ggsave(out_file, plot = p, width = 7, height = 5)
  return(p)
}

outdir <- "diagnostic_plots"
if(!dir.exists(outdir)) dir.create(outdir)

stations_to_check <- c(141,155,7,8,10,13) # scelta random.

for (j in stations_to_check) {
  cat("Processing station", j, "\n")
  p1 <- plot_hist_density(j, observed = events, simulated = simevents,
                        out_file = file.path(outdir, paste0("hist_station_", j, ".png")))
  p2 <- plot_qq(j, events, simevents, out_file = file.path(outdir, paste0("qq_station_", j, ".png")))
  p3 <- plot_tail_compare(j, events, simevents,
                        pareto_params = paramParetoGML,
                        burr_params = param,
                        out_file = file.path(outdir, paste0("tail_station_", j, ".png")))
  p4 <- plot_return_level(j, events, simevents, nyears_obs = NA, NYRsim = NYRsim,
                        out_file = file.path(outdir, paste0("return_station_", j, ".png")))
}

```



```

grid.arrange(p1, p2, p3, p4, ncol = 2)
}

#-----Verifica del modello multivariato-----#
m_events <- apply(events, 2, mean, na.rm=TRUE)
sd_events <- apply(events, 2, sd, na.rm=TRUE)
q90_events <- apply(events, 2, quantile, prob=0.9, na.rm=TRUE)
m_simevents <- matrix(NA, nrow=floor(ev_x_yr*NYRsim/nrow(events)),
ncol=length(m_events))
sd_simevents <- m_simevents
q90_simevents <- m_simevents

for (ii in 1:(floor(ev_x_yr*NYRsim/nrow(events)))) {
  dummy <- simevents[(nrow(events)*(ii - 1) + 1):(nrow(events)*ii),]
  m_simevents[ii,] <- apply(dummy, 2, mean, na.rm=TRUE)
  sd_simevents[ii,] <- apply(dummy, 2, sd, na.rm=TRUE)
  q90_simevents[ii,] <- apply(dummy, 2, quantile, prob=0.9, na.rm=TRUE)
}
q025_m_simevents <- apply(m_simevents, 2, quantile, prob=.025)
q500_m_simevents <- apply(m_simevents, 2, quantile, prob=.50)
q975_m_simevents <- apply(m_simevents, 2, quantile, prob=.975)

q025_sd_simevents <- apply(sd_simevents, 2, quantile, prob=.025)
q500_sd_simevents <- apply(sd_simevents, 2, quantile, prob=.50)
q975_sd_simevents <- apply(sd_simevents, 2, quantile, prob=.975)

q025_q90_simevents <- apply(q90_simevents, 2, quantile, prob=.025)
q500_q90_simevents <- apply(q90_simevents, 2, quantile, prob=.50)
q975_q90_simevents <- apply(q90_simevents, 2, quantile, prob=.975)

n_sites <- length(m_events)
df_compare <- data.frame(
  site = 1:n_sites,
  mean_obs = m_events,
  mean_lo = q025_m_simevents,
  mean_hi = q975_m_simevents,
  sd_obs = sd_events,
  sd_lo = q025_sd_simevents,
  sd_hi = q975_sd_simevents,
  q90_obs = q90_events,
  q90_lo = q025_q90_simevents,
  q90_hi = q975_q90_simevents
)
library(ggplot2)
library(dplyr)
library(gridExtra)

make_sorted_plot <- function(obs, lo, hi, y_lab, title) {

```

```

# 1) creo un df temporaneo ordinato
df_tmp <- df_compare %>%
  dplyr::select(site, !!sym(obs), !!sym(lo), !!sym(hi)) %>%
  arrange(!!sym(obs)) %>%
  mutate(idx = row_number())

# 2) plot
ggplot(df_tmp, aes(x = idx)) +
  # error bars per CI 95%
  geom_linerange(aes(ymin = !!sym(lo), ymax = !!sym(hi)),
    color = "grey40", linewidth = 0.3) +
  # punti dei valori osservati
  geom_point(aes(y = !!sym(obs)),
    shape = 21, fill = "white", color = "black", size = 1) +
  scale_x_continuous(name = "Site index", breaks = seq(0, nrow(df_tmp), by = 100)) +
  scale_y_log10(name = y_lab) +
  labs(title = title) +
  theme_minimal(base_size = 12) +
  theme(
    panel.grid.major.x = element_blank(),
    panel.grid.minor = element_blank(),
    plot.title = element_text(hjust = 0.5))
}

p_mean <- make_sorted_plot(
  obs = "mean_obs", lo = "mean_lo", hi = "mean_hi",
  y_lab = expression("Mean Q (m"3"/s)"),
  title = "Media osservata vs simulata"
)

p_sd <- make_sorted_plot(
  obs = "sd_obs", lo = "sd_lo", hi = "sd_hi",
  y_lab = expression("SD Q (m"3"/s)"),
  title = "SD osservata vs simulata"
)

p_q90 <- make_sorted_plot(
  obs = "q90_obs", lo = "q90_lo", hi = "q90_hi",
  y_lab = expression("90% quantile Q (m"3"/s)"),
  title = "90° percentile osservato vs simulata"
)

plots <- list(
  "Media Q" = p_mean,
  "Deviazione SD" = p_sd,
  "90° Percentile" = p_q90)

for (nm in names(plots)) {
  cat("\n--- Adesso il grafico:", nm, "---\n")
  print(plots[[nm]])
  readline(prompt = "Premi <Invio> per passare al prossimo grafico")
}

```

```

}

df_bias <- data.frame(
  site = df_compare$site,

  bias_mean  = (q500_m_simevents - m_events) / m_events,
  bias_mean_lo = (q025_m_simevents - m_events) / m_events,
  bias_mean_hi = (q975_m_simevents - m_events) / m_events,

  bias_sd  = (q500_sd_simevents - sd_events) / sd_events,
  bias_sd_lo = (q025_sd_simevents - sd_events) / sd_events,
  bias_sd_hi = (q975_sd_simevents - sd_events) / sd_events,

  bias_q90  = (q500_q90_simevents - q90_events) / q90_events,
  bias_q90_lo = (q025_q90_simevents - q90_events) / q90_events,
  bias_q90_hi = (q975_q90_simevents - q90_events) / q90_events
)

```

```

make_bias_plot <- function(df, obs, lo, hi, y_lab, title) {
  ggplot(df, aes(x = site)) +
    # barre di confidenza
    geom_linerange(aes(ymin = .data[[lo]], ymax = .data[[hi]]),
      color = "grey60", size = 0.3) +
    # punti al bias mediano
    geom_point(aes(y = .data[[obs]]),
      shape = 21, fill = "white", color = "black", size = 1) +
    scale_x_continuous(
      name = "Site index",
      limits = c(1, nrow(df)),
      breaks = seq(0, nrow(df), by = 100)
    ) +
    scale_y_continuous(name = y_lab) +
    labs(title = title) +
    theme_minimal(base_size = 12) +
    theme(
      panel.grid.major.x = element_blank(),
      panel.grid.minor  = element_blank(),
      plot.title = element_text(hjust = 0.5))
}

```

```

p_bias_mean <- make_bias_plot(
  df = df_bias,
  obs = "bias_mean", lo = "bias_mean_lo", hi = "bias_mean_hi",
  y_lab = "Relative bias of mean",
  title = "Bias relativo media"
)

```

```

p_bias_sd <- make_bias_plot(
  df = df_bias,

```

```

obs = "bias_sd", lo = "bias_sd_lo", hi = "bias_sd_hi",
y_lab = "Relative bias of SD",
title = "Bias relativo SD"
)

p_bias_q90 <- make_bias_plot(
  df = df_bias,
  obs = "bias_q90", lo = "bias_q90_lo", hi = "bias_q90_hi",
  y_lab = "Relative bias of 90th percentile",
  title = "Bias relativo 90° percentile"
)

plots_bias <- list(
  "Bias relativo media" = p_bias_mean,
  "Bias relativo SD" = p_bias_sd,
  "Bias 90° Percentile" = p_bias_q90)

for (nm in names(plots_bias)) {
  cat("\n--- Adesso il grafico:", nm, "---\n")
  print(plots_bias[[nm]])
  readline(prompt = "Premi <Invio> per passare al prossimo grafico")
}

rho_events <- cor(events, use='pairwise.complete.obs', method='pearson')
rho_events <- rho_events[lower.tri(rho_events, diag=FALSE)]
rho_simevents <- cor(simevents, method='pearson')
rho_simevents <- rho_simevents[lower.tri(rho_simevents, diag=FALSE)]
spe_events <- cor(events, use='pairwise.complete.obs', method='spearman')
spe_events <- spe_events[lower.tri(spe_events, diag=FALSE)]
spe_simevents <- cor(simevents, method='spearman')
spe_simevents <- spe_simevents[lower.tri(spe_simevents, diag=FALSE)]

df_pear <- data.frame(
  obs = rho_events,
  sim = rho_simevents)

df_spe <- data.frame(
  obs = spe_events,
  sim = spe_simevents)

# Funzione per generare il plot
make_cor_plot <- function(df, title) {
  ggplot(df, aes(x = obs, y = sim)) +
    geom_point(alpha = 0.3, size = 1) +
    geom_abline(slope = 1, intercept = 0, linetype = "dashed", color = "red") +
    coord_fixed(xlim = c(-1, 1), ylim = c(-1, 1)) +
    labs(
      x = "r observed",
      y = "r simulated",
      title = title
    )
}

```



```

) +
theme_minimal(base_size = 12) +
theme(
  plot.title = element_text(hjust = 0.5)
)
}

p_pearson <- make_cor_plot(df_pear, "Pearson: osservate vs simulate")
p_spearman <- make_cor_plot(df_spe, "Spearman: osservate vs simulate")

print(p_pearson)
print(p_spearman)

# funzione che calcola le metriche riassuntive
summary_cor_metrics <- function(df) {
  obs <- df$obs
  sim <- df$sim
  ok <- !(is.na(obs) | is.na(sim))
  obs <- obs[ok]; sim <- sim[ok]
  n <- length(obs)
  cor_xy <- if(n>1) cor(obs, sim, use = "complete.obs") else NA_real_
  rmse <- sqrt(mean((sim - obs)^2, na.rm = TRUE))
  bias <- mean(sim - obs, na.rm = TRUE)
  rel_bias <- mean((sim - obs) / ifelse(obs==0, NA, obs), na.rm = TRUE) # relativo
  r2 <- if(n>2) summary(lm(sim ~ obs))$r.squared else NA_real_
  list(n = n, cor = cor_xy, RMSE = rmse, bias = bias, rel_bias = rel_bias, R2 = r2)
}

# calcolo le metriche per Pearson e Spearman
metrics_pear <- summary_cor_metrics(df_pear)
metrics_spe <- summary_cor_metrics(df_spe)

# creo stringhe formattate per l'annotazione (3 cifre decimali)
make_label_text <- function(m) {
  sprintf("n = %d\ncor = %.3f\nRMSE = %.3f\nbias = %.3f\nrel.bias = %.2f%%\nR2 = %.3f",
    m$n, m$cor, m$RMSE, m$bias, 100*m$rel_bias, m$R2)
}

label_pear <- make_label_text(metrics_pear)
label_spe <- make_label_text(metrics_spe)

# posizione dell'annotazione
xpos <- -0.95
ypos <- 0.90

#plot
p_pearson_annot <- p_pearson +
  annotate("label", x = xpos, y = ypos,
    label = label_pear, hjust = 0, vjust = 1,
    size = 3.2, fill = "white", color = "black")

```

```

p_spearman_annot <- p_spearman +
  annotate("label", x = xpos, y = ypos,
    label = label_spe, hjust = 0, vjust = 1,
    size = 3.2, fill = "white", color = "black")

p_pearson_annot <- p_pearson_annot +
  geom_smooth(data = df_pear, aes(x = obs, y = sim), method = "lm",
    se = FALSE, color = "blue", size = 0.6, linetype = "solid")

p_spearman_annot <- p_spearman_annot +
  geom_smooth(data = df_spe, aes(x = obs, y = sim), method = "lm",
    se = FALSE, color = "blue", size = 0.6, linetype = "solid")

print(p_pearson_annot)
print(p_spearman_annot)

# --- Calcolo differenze e statistiche ---#
df_pear$diff <- df_pear$sim - df_pear$obs
df_spe$diff <- df_spe$sim - df_spe$obs

stats_from_vec <- function(x, y) {
  ok <- !(is.na(x) | is.na(y))
  diff <- y[ok] - x[ok]
  rmse <- sqrt(mean(diff^2))
  list(mean = mean(diff), sd = sd(diff), RMSE = rmse, n = length(diff))
}

stats_pear <- stats_from_vec(df_pear$obs, df_pear$sim)
stats_spe <- stats_from_vec(df_spe$obs, df_spe$sim)

# --- Funzione per creare histogram + density (annotato) ---
make_diff_hist <- function(df, stats, title, bins = 40) {
  lab_text <- sprintf("n = %d\nmean = %.4f\nsd = %.4f\nRMSE = %.4f",
    stats$n, stats$mean, stats$sd, stats$RMSE)
  p <- ggplot(df, aes(x = diff)) +
    geom_histogram(aes(y = ..density..), bins = bins, fill = "grey80", color = "black") +
    geom_density(alpha = 0.3, size = 0.7) +
    geom_vline(xintercept = stats$mean, color = "red", linetype = "dashed", size = 0.6) +
    geom_vline(xintercept = 0, color = "blue", linetype = "dotted", size = 0.6) +
    annotate("label", x = Inf, y = Inf, label = lab_text,
      hjust = 1.05, vjust = 1.1, size = 3.2, fill = "white") +
    labs(x = "Difference (sim - obs)", y = "Density", title = title) +
    theme_minimal(base_size = 12)
  return(p)
}

```

```

hist_pear <- make_diff_hist(df_pear, stats_pear, title = "Histogram of differences (Pearson: sim
- obs)")
hist_spe <- make_diff_hist(df_spe, stats_spe, title = "Histogram of differences (Spearman: sim
- obs)")

# --- Affianca scatter annotato e istogramma ---
p_pearson_combined <- grid.arrange(p_pearson_annot, hist_pear, ncol = 2, widths = c(0.6,
0.4))
p_spearman_combined <- grid.arrange(p_spearman_annot, hist_spe, ncol = 2, widths = c(0.6,
0.4))

#-----Eventi alluvionali simulati sulla GAR-----#
P_u_sim <- simevents*NA
for(j in 1:ncol(P_u_sim)) {
  print(j)
  dummy <- as.numeric(simevents[,j])
  low=paramParetoGML[j,'x0']; upp=+Inf;
  if(paramParetoGML[j,'kappa'] > 0) upp=paramParetoGML[j,'x0'] +
paramParetoGML[j,'alpha']/paramParetoGML[j,'kappa']
  quali <- (!(is.na(dummy)) & (dummy > low) & (dummy < upp))
  P_u_sim[quali,j] <- pGP(dummy[quali], x0=paramParetoGML[j,'x0'],
alpha=paramParetoGML[j,'alpha'], kappa=paramParetoGML[j,'kappa'])
  P_u_sim[(dummy <= low),j] <- 0
  P_u_sim[(dummy >= upp),j] <- 1
}

T_n_sim <- 1/(ev_x_yr * (1 - P_u_sim))

D_n_sim <- T_n_sim*NA
for(j in 1:ncol(D_n_sim)) {
  print(j)
  D_n_sim[,j] <- ramp(log(as.numeric(T_n_sim[,j])), minlT=parramp[1], maxlT=parramp[2])
}
# media pesata sulle aree
aree=rep(NA,nrow(paramParetoGML))
for (j in 1:nrow(paramParetoGML)) {
  aree[j]=round(area_cats_GAR_filtrata[[j]]/1000000,2)
}

D_m_sim <- rep(NA, nrow(D_n_sim))
for (ii in 1:nrow(D_n_sim)) {
  D_m_sim[ii] <- weighted.mean(as.numeric(D_n_sim[ii,])^p, w=log(aree))^(1/p)
}

round(quantile(D_m_sim, prob=c(0,.01,.05,.1,.25,.5,.75,.9,.95,.99,1)), 4)
ranks <- rank(D_m_sim)
RP_m_sim <- round(nyr/nrow(events), 1) * 1/(1 - ranks/(max(ranks) + 1))

#-----plot di eventi alluvionali generati, random-----#

```

```

quali <- 1:15
some_RP_m_sim <- RP_m_sim[quali]
some_D_m_sim <- D_m_sim[quali]
some_simevents <- simevents[quali,]
some_RPln3_simevents <- T_n_sim[quali,]
t(round(apply(some_RPln3_simevents, 1, quantile, prob=c(0,.01,.05,.1,.25,.5,.75,.9,.95,.99,1))))

# Palette, breaks e labels fissi
breaks <- c(0, 10, 20, 50, 100, 200, 500, 1000, 10000, 20000, 50000, Inf)
labels_classi <- c("<= 10", "<= 20", "<= 50", "<= 100", "<= 200", "<= 500", "<= 1000", "<= 10000", "<= 20000", "<= 50000", "> 50000")
palette_classi_sim <- c("white", "lightgreen", "green", "darkgreen", "yellow", "orange",
"red", "darkred", "#762a83", "violetred", "violet")
radius_classi <- c(2, 4, 5, 6, 7, 8, 9, 10, 12, 14, 16) # una dimensione per classe
names(radius_classi) <- labels_classi

for (i in seq_len(nrow(some_simevents))) {
  df_ev_sim_some <- data.frame(
    ID = df_ev$ID,
    lon = df_ev$lon,
    lat = df_ev$lat,
    flow = as.numeric(some_simevents[i, ]),
    RP_sim = as.numeric(some_RPln3_simevents[i, ]),
    Dm = some_D_m_sim[i])

  df_ev_sim_some <- df_ev_sim_some %>%
    mutate(
      classe_RP = cut(RP_sim,
        breaks = breaks,
        labels = labels_classi,
        include.lowest = TRUE,
        right = TRUE),
      classe_RP = factor(classe_RP, levels = labels_classi),
      radius = case_when(
        classe_RP == labels_classi[1] ~ 2,
        classe_RP == labels_classi[2] ~ 4,
        classe_RP == labels_classi[3] ~ 5,
        classe_RP == labels_classi[4] ~ 6,
        classe_RP == labels_classi[5] ~ 7,
        classe_RP == labels_classi[6] ~ 8,
        classe_RP == labels_classi[7] ~ 9,
        classe_RP == labels_classi[8] ~ 10,
        classe_RP == labels_classi[9] ~ 12,
        classe_RP == labels_classi[10] ~ 14,
        classe_RP == labels_classi[11] ~ 16,
        TRUE ~ 4
      ))

  pal <- colorFactor(palette = palette_classi_sim, domain = labels_classi, ordered = TRUE)

```

```

m <- leaflet(df_ev_sim_some) %>%
  addProviderTiles(providers$OpenStreetMap) %>%
  addCircleMarkers(
    ~lon, ~lat,
    radius = ~radius,
    color = "black",
    weight = 0.9,
    fillColor = ~pal(classe_RP),
    fillOpacity = 0.8,
    popup = ~paste0(
      "<b>Stazione:</b> ", ID, "<br>",
      "<b>RP simulato:</b> ", round(RP_sim, 2), " anni"),
    popupOptions = popupOptions(autoClose = FALSE, closeOnClick = FALSE)
  ) %>%
  addLegend(
    "bottomright", pal = pal, values = ~classe_RP,
    title = "RP simulato (anni)"
  ) %>%
  addControl(
    html = paste0(
      "<div style='padding:4px;background:white;border:1px solid #ccc;'>",
      "<b>Evento n. </b> ", i, "<br>",
      "<b>Dm dello scenario:</b> ", round(some_D_m_sim[i], 4), "</div>"
    ),
    position = "topright")
print(m)
readline(prompt = "Premi <Invio> per vedere la prossima mappa...")
}

#-----plot eventi simulati più gravosi-----#
quali <- which(rank(-D_m_sim) <= 15)
large_RP_m_sim <- RP_m_sim[quali]
large_D_m_sim <- D_m_sim[quali]
large_simevents <- simevents[quali,]
large_RPln3_simevents <- T_n_sim[quali,]
t(round(apply(large_RPln3_simevents, 1, quantile, prob=c(0,.01,.05,.1,.25,.5,.75,.9,.95,.99,1))))

# Loop sui 15 eventi "più grandi"
for (i in seq_len(nrow(large_simevents))) {
  df_ev_sim_large <- data.frame(
    ID   = df_ev$ID,
    lon  = df_ev$lon,
    lat  = df_ev$lat,
    flow = as.numeric(large_simevents[i, ]),
    RP_sim = as.numeric(large_RPln3_simevents[i, ])
  )

  df_ev_sim_large <- df_ev_sim_large %>%
    mutate(
      classe_RP = cut(RP_sim,
        breaks = breaks,

```



```

      labels = labels_classi,
      include.lowest = TRUE,
      right = TRUE),
  classe_RP = factor(classe_RP, levels = labels_classi),
  radius = case_when(
    classe_RP == labels_classi[1] ~ 2,
    classe_RP == labels_classi[2] ~ 4,
    classe_RP == labels_classi[3] ~ 5,
    classe_RP == labels_classi[4] ~ 6,
    classe_RP == labels_classi[5] ~ 7,
    classe_RP == labels_classi[6] ~ 8,
    classe_RP == labels_classi[7] ~ 9,
    classe_RP == labels_classi[8] ~ 10,
    classe_RP == labels_classi[9] ~ 12,
    classe_RP == labels_classi[10] ~ 14,
    classe_RP == labels_classi[11] ~ 16,
    TRUE ~ 4
  ))

pal <- colorFactor(palette = palette_classi_sim, domain = labels_classi, ordered = TRUE)

m <- leaflet(df_ev_sim_large) %>%
  addProviderTiles(providers$OpenStreetMap) %>%
  addCircleMarkers(
    ~lon, ~lat,
    radius = ~radius,
    color = "black",
    weight = 0.9,
    fillColor = ~pal(classe_RP),
    fillOpacity = 0.8,
    popup = ~paste0(
      "<b>Stazione:</b> ", ID, "<br>",
      "<b>RP simulato:</b> ", round(RP_sim, 2), " anni"),
    popupOptions = popupOptions(autoClose = FALSE, closeOnClick = FALSE)
  ) %>%
  addLegend(
    "bottomright", pal = pal, values = ~classe_RP,
    title = "RP simulato (anni)"
  ) %>%
  addControl(
    html = paste0(
      "<div style='padding:4px;background:white;border:1px solid #ccc;'>",
      "<b>Evento n. </b> ", i, "<br>",
      "<b>Dm dello scenario:</b> ", round(large_D_m_sim[i], 4), "</div>"
    ),
    position = "topright"
  )
print(m)
readline(prompt = "Premi <Invio> per vedere la prossima mappa...")
}

```

```

#-----plot eventi con periodi di ritorno 100 anni,random-----#
quali <- sort(sample(which((RP_m_sim > 90)&(RP_m_sim < 110)), 15))
some100_RP_m_sim <- RP_m_sim[quali]
some100_D_m_sim <- D_m_sim[quali]
some100_simevents <- simevents[quali,]
some100_RPln3_simevents <- T_n_sim[quali,]
t(round(apply(some100_RPln3_simevents, 1, quantile,
prob=c(0,.01,.05,.1,.25,.5,.75,.9,.95,.99,1))))

for (i in seq_len(nrow(some100_simevents))) {
  df_ev_sim_some100 <- data.frame(
    ID   = df_ev$ID,
    lon  = df_ev$lon,
    lat  = df_ev$lat,
    flow = as.numeric(some100_simevents[i, ]),
    RP_sim = as.numeric(some100_RPln3_simevents[i, ]))

  df_ev_sim_some100 <- df_ev_sim_some100 %>%
    mutate(
      classe_RP = cut(RP_sim,
        breaks = breaks,
        labels = labels_classi,
        include.lowest = TRUE,
        right = TRUE),
      classe_RP = factor(classe_RP, levels = labels_classi),
      radius = case_when(
        classe_RP == labels_classi[1] ~ 2,
        classe_RP == labels_classi[2] ~ 4,
        classe_RP == labels_classi[3] ~ 5,
        classe_RP == labels_classi[4] ~ 6,
        classe_RP == labels_classi[5] ~ 7,
        classe_RP == labels_classi[6] ~ 8,
        classe_RP == labels_classi[7] ~ 9,
        classe_RP == labels_classi[8] ~ 10,
        classe_RP == labels_classi[9] ~ 12,
        classe_RP == labels_classi[10] ~ 14,
        classe_RP == labels_classi[11] ~ 16,
        TRUE ~ 4
      ))

  pal <- colorFactor(palette = palette_classi_sim, domain = labels_classi, ordered = TRUE)

  m <- leaflet(df_ev_sim_some100) %>%
    addProviderTiles(providers$OpenStreetMap) %>%
    addCircleMarkers(
      ~lon, ~lat,
      radius = ~radius,
      color = "black",
      weight = 0.9,
      fillColor = ~pal(classe_RP),

```

```

fillOpacity = 0.8,
popup = ~paste0(
  "<b>Stazione:</b> ", ID, "<br>",
  "<b>RP simulato:</b> ", round(RP_sim, 2), " anni"),
popupOptions = popupOptions(autoClose = FALSE, closeOnClick = FALSE)
)%>%
addLegend(
  "bottomright", pal = pal, values = ~classe_RP,
  title = "RP simulato (anni)"
)%>%
addControl(
  html = paste0(
    "<div style='padding:4px;background:white;border:1px solid #ccc;'>",
    "<b>Evento n. </b>", i, "<br>",
    "<b>Dm dello scenario:</b> ", round(some100_D_m_sim[i], 4), "</div>"
  ),
  position = "topright")
print(m)
readline(prompt = "Premi <Invio> per vedere la prossima mappa...")
}

```

```

#-----plot eventi con periodi di ritorno 10 anni,random-----#
quali <- sort(sample(which((RP_m_sim > 9)&(RP_m_sim < 11)), 15))
some10_RP_m_sim <- RP_m_sim[quali]
some10_D_m_sim <- D_m_sim[quali]
some10_simevents <- simevents[quali,]
some10_RPln3_simevents <- T_n_sim[quali,]
t(round(apply(some10_RPln3_simevents, 1, quantile,
prob=c(0,.01,.05,.1,.25,.5,.75,.9,.95,.99,1))))

```

```

for (i in seq_len(nrow(some10_simevents))) {
  df_ev_sim_some10 <- data.frame(
    ID = df_ev$ID,
    lon = df_ev$lon,
    lat = df_ev$lat,
    flow = as.numeric(some10_simevents[i, ]),
    RP_sim = as.numeric(some10_RPln3_simevents[i, ]))
}

```

```

df_ev_sim_some10 <- df_ev_sim_some10 %>%
mutate(
  classe_RP = cut(RP_sim,
    breaks = breaks,
    labels = labels_classi,
    include.lowest = TRUE,
    right = TRUE),
  classe_RP = factor(classe_RP, levels = labels_classi),
  radius = case_when(
    classe_RP == labels_classi[1] ~ 2,
    classe_RP == labels_classi[2] ~ 4,
    classe_RP == labels_classi[3] ~ 5,

```

```

classe_RP == labels_classi[4] ~ 6,
classe_RP == labels_classi[5] ~ 7,
classe_RP == labels_classi[6] ~ 8,
classe_RP == labels_classi[7] ~ 9,
classe_RP == labels_classi[8] ~ 10,
classe_RP == labels_classi[9] ~ 12,
classe_RP == labels_classi[10] ~ 14,
classe_RP == labels_classi[11] ~ 16,
TRUE ~ 4
))

```

```

pal <- colorFactor(palette = palette_classi_sim, domain = labels_classi, ordered = TRUE)

```

```

m <- leaflet(df_ev_sim_some10) %>%
  addProviderTiles(providers$OpenStreetMap) %>%
  addCircleMarkers(
    ~lon, ~lat,
    radius = ~radius,
    color = "black",
    weight = 0.9,
    fillColor = ~pal(classe_RP),
    fillOpacity = 0.8,
    popup = ~paste0(
      "<b>Stazione:</b> ", ID, "<br>",
      "<b>RP simulato:</b> ", round(RP_sim, 2), " anni"),
    popupOptions = popupOptions(autoClose = FALSE, closeOnClick = FALSE)
  ) %>%
  addLegend(
    "bottomright", pal = pal, values = ~classe_RP,
    title = "RP simulato (anni)"
  ) %>%
  addControl(
    html = paste0(
      "<div style='padding:4px;background:white;border:1px solid #ccc;'>",
      "<b>Evento n. </b>", i, "<br>",
      "<b>Dm dello scenario:</b> ", round(some10_D_m_sim[i], 4), "</div>"
    ),
    position = "topright")
print(m)
readline(prompt = "Premi <Invio> per vedere la prossima mappa...")
}

```

```

#-----Distribuzione spaziale degli eventi simulati-----#
dummy_sim <- log(T_n_sim)
dummy_sim[is.infinite(dummy_sim)] <- NA
meanRPln3_simevents <- exp(apply(dummy_sim, 2, mean, na.rm=TRUE))
meanD_n_simevents <- apply(D_n_sim, 2, mean)
nsimevRPgr2 <- apply(T_n_sim > 2, 2, sum)
nsimevDgr0 <- apply(D_n_sim > 0.00000001, 2, sum)
Nsim <- nrow(T_n_sim)

```

```

breaks_RP_geom <- c(0,1000, 2000, 3000, 4000, 5000, 6000, 8000, 10000, 20000, 50000, Inf)
labels_classi_RP_geom <- c("<= 1000", "<= 2000", "<= 3000", "<= 4000", "<= 5000", "<= 6000", "<= 8000", "<= 10000", "<= 20000", "<= 50000", "> 50000")
palette_classi_sim <- c("white", "lightgreen", "green", "darkgreen", "yellow", "orange", "red", "darkred", "#762a83", "violetred", "violet")
df_map_sim <- data.frame(
  ID = names(meanRPln3_simevents),
  RP_geom = meanRPln3_simevents,
  D_med = meanD_n_simevents,
  RP_gt2 = nsimevRPgr2,
  D_gt0 = nsimevDgr0
) %>%
left_join(df_mappa, by = "ID")

df_map_sim <- df_map_sim %>%
mutate(
  classe_RP = cut(RP_gt2,
    breaks = breaks_RP_geom,
    labels = labels_classi_RP_geom,
    include.lowest = TRUE,
    right = TRUE),
  classe_RP = factor(classe_RP, levels = labels_classi_RP_geom),
  radius = case_when(
    classe_RP == labels_classi_RP_geom[1] ~ 2,
    classe_RP == labels_classi_RP_geom[2] ~ 4,
    classe_RP == labels_classi_RP_geom[3] ~ 5,
    classe_RP == labels_classi_RP_geom[4] ~ 6,
    classe_RP == labels_classi_RP_geom[5] ~ 7,
    classe_RP == labels_classi_RP_geom[6] ~ 8,
    classe_RP == labels_classi_RP_geom[7] ~ 9,
    classe_RP == labels_classi_RP_geom[8] ~ 10,
    classe_RP == labels_classi_RP_geom[9] ~ 12,
    classe_RP == labels_classi_RP_geom[10] ~ 14,
    classe_RP == labels_classi_RP_geom[11] ~ 16,
    TRUE ~ 4))
#-----plot con n eventi con RP >2-----#

ggplot(df_map_sim, aes(x = lon, y = lat)) +
  borders("world", xlim = c(5, 20), ylim = c(44, 50), fill = "gray95", colour = "gray60") +
  geom_point(aes(fill = classe_RP, size = classe_RP), shape = 21, color = "black", stroke = 0.4) +
  scale_fill_manual(values = palette_classi_sim, name = "n. of events\nhaving RP > 2:") +
  scale_size_manual(
    values = c(5, 6, 7, 8, 9, 10, 11, 12),
    name = "n. of events\nhaving RP > 2:") +
  coord_fixed(
    xlim = c(5, 17),
    ylim = c(43.5, 49.5)) +
  theme_minimal() +
  labs(

```



```

  title = "Frequenza eventi con RP > 2 simulati per stazione (GAR)", subtitle = paste0("Numero
totale di simulazioni per stazione: ", Nsim),
  x = "Longitudine", y = "Latitudine") +
  theme(legend.position = "right", plot.title = element_text(hjust = 0.5), plot.subtitle =
element_text(hjust = 0.5))

#-----plot con n eventi con D > 0-----#
df_map_sim <- df_map_sim %>%
  mutate(
    classe_D = cut(
      D_gt0,
      breaks = c(0, 100, 200, 500, 700, 900, 1000, 1100, 1200, 1300, 1500, Inf),
      labels = c("<= 100", "<= 200", "<= 500", "<= 700", "<= 900", "<= 1000", "<= 1100", "<= 1200", "<=
1300", "<= 1500", "> 1500"),
      include.lowest = TRUE,
      right = TRUE)
  )
ggplot(df_map_sim, aes(x = lon, y = lat)) +
  borders("world", xlim = c(5, 20), ylim = c(44, 50), fill = "gray95", colour = "gray60") +
  geom_point(aes(fill = classe_D, size = classe_D), shape = 21, color = "black", stroke = 0.4) +
  scale_fill_manual(values = palette_classi_sim, name = "n. of events\nhaving D > 0:") +
  scale_size_manual(
    values = c(3, 3, 3, 4, 5, 6, 7, 8, 9, 10, 11),
    name = "n. of events\nhaving D > 0:") +
  coord_fixed(
    xlim = c(5, 17),
    ylim = c(43.5, 49.5)) +
  theme_minimal() +
  labs(title = "Frequenza eventi simulati con danno > 0 per stazione (GAR)", subtitle =
paste0("Numero totale di simulazioni per stazione: ", Nsim), x = "Longitudine", y =
"Latitudine") +
  theme(legend.position = "right", plot.title = element_text(hjust = 0.5), plot.subtitle =
element_text(hjust = 0.5))

#-----RP_geom (classi quantile)-----#
# usare quantili per ottenere classi bilanciate
n_classes <- 7
qbreaks_RP <- unique(quantile(df_map_sim$RP_geom, probs = seq(0,1,length.out =
n_classes+1), na.rm=TRUE))
# se ci fossero duplicati nel quantile
if(length(qbreaks_RP) - 1 < n_classes) {
  qbreaks_RP <- pretty(df_map_sim$RP_geom, n = n_classes)
}
# genero labels
labels_RP_geom <- paste0("Q", 1:(length(qbreaks_RP)-1))

df_map_sim <- df_map_sim %>%
  mutate(
    classe_RP_geom = cut(RP_geom, breaks = qbreaks_RP, labels = labels_RP_geom,
      include.lowest = TRUE, right = TRUE),

```

```

  classe_RP_geom = factor(classe_RP_geom, levels = labels_RP_geom)
)

palette_geom <- brewer.pal(n = max(length(labels_RP_geom),3), name = "PuBu")
sizes_geom <- seq(3, 10, length.out = length(labels_RP_geom))

p3 <- ggplot(df_map_sim, aes(x = lon, y = lat)) +
  borders("world", xlim = c(5, 20), ylim = c(44, 50), fill = "gray95", colour = "gray60") +
  geom_point(aes(fill = classe_RP_geom, size = classe_RP_geom), shape = 21, color = "black",
stroke = 0.35) +
  scale_fill_manual(values = palette_geom, name = "RP_geom") +
  scale_size_manual(values = sizes_geom, name = "RP_geom") +
  coord_fixed(xlim = c(5, 17), ylim = c(43.5, 49.5)) +
  theme_minimal() +
  labs(
    title = "Media geometrica locale dei tempi di ritorno (RP_geom)",
    subtitle = paste0("Nsim = ", Nsim),
    x = "Longitudine", y = "Latitudine",
    caption = paste0("Breaks: ", paste(round(qbreaks_RP,3), collapse = " | "))
  ) +
  theme(legend.position = "right", plot.title = element_text(hjust = 0.5), plot.subtitle =
element_text(hjust = 0.5))
p3

#-----D_med (classi quantile) -----#
n_classes_d <- 7
qbreaks_D <- unique(quantile(df_map_sim$D_med, probs = seq(0,1,length.out =
n_classes_d+1), na.rm=TRUE))
if(length(qbreaks_D) - 1 < n_classes_d) {
  qbreaks_D <- pretty(df_map_sim$D_med, n = n_classes_d)
}
labels_D_med <- paste0("Q", 1:(length(qbreaks_D)-1))

df_map_sim <- df_map_sim %>%
  mutate(
    classe_D_med = cut(D_med, breaks = qbreaks_D, labels = labels_D_med,
      include.lowest = TRUE, right = TRUE),
    classe_D_med = factor(classe_D_med, levels = labels_D_med)
  )

palette_D <- brewer.pal(n = max(length(labels_D_med),3), name = "YlGnBu")
sizes_D <- seq(3, 10, length.out = length(labels_D_med))

p4 <- ggplot(df_map_sim, aes(x = lon, y = lat)) +
  borders("world", xlim = c(5, 20), ylim = c(44, 50), fill = "gray95", colour = "gray60") +
  geom_point(aes(fill = classe_D_med, size = classe_D_med), shape = 21, color = "black", stroke
= 0.35) +
  scale_fill_manual(values = palette_D, name = "D_med") +
  scale_size_manual(values = sizes_D, name = "D_med") +
  coord_fixed(xlim = c(5, 17), ylim = c(43.5, 49.5)) +

```

```

theme_minimal() +
labs(
  title = "Danno medio locale (D_med)",
  subtitle = paste0("Nsim = ", Nsim),
  x = "Longitudine", y = "Latitudine",
  caption = paste0("Breaks: ", paste(round(qbreaks_D,5), collapse = " | "))
) +
theme(legend.position = "right", plot.title = element_text(hjust = 0.5), plot.subtitle =
element_text(hjust = 0.5))
p4

```

#-----Grafici confronto tra eventi osservati e simulati-----#

```

# Funzione per aggiungere R2 al grafico
add_r2 <- function(x, y, pos="topright") {
  r2 <- cor(x, y, use="complete.obs")^2
  legend(pos, legend = bquote(R^2 == .(round(r2, 3))), bty="n")
}

```

```

# Layout 2 x 3
par(mfrow = c(2, 3), mar = c(4,4,2,1))

```

```

# 1. RP geom: simulato vs osservato
plot(meanRPIn3_events, meanRPIn3_simevents,
  xlab = "observed local geom-mean RP (yrs)",
  ylab = "simulated local geom-mean RP (yrs)",
  pch = 21, bg = "white", cex = 1.2)
abline(0,1)
add_r2(meanRPIn3_events, meanRPIn3_simevents)

```

```

# 2. Danno medio: simulato vs osservato
plot(meand_n_events, meanD_n_simevents,
  xlab = "observed local mean damage",
  ylab = "simulated local mean damage",
  pch = 21, bg = "white", cex = 1.2)
abline(0,1)
add_r2(meand_n_events, meanD_n_simevents)

```

```

# 3. N. eventi con RP > 2: simulato vs osservato
plot(nevRPgr2, nsimevRPgr2,
  xlab = "observed n. of events having RP > 2",
  ylab = "simulated n. of events having RP > 2",
  pch = 21, bg = "white", cex = 1.2)
abline(0,1)
add_r2(nevRPgr2, nsimevRPgr2)

```

```

# 4. N. eventi con D > 0: simulato vs osservato
plot(nevDgr0, nsimevDgr0,
  xlab = "observed n. of events having D > 0",
  ylab = "simulated n. of events having D > 0",

```

```

    pch = 21, bg = "white", cex = 1.2)
abline(0,1)
add_r2(nevDgr0, nsimevDgr0)

# 5. Relazione osservata: RP geom vs Danno medio
plot(meanRPln3_events, meand_n_events,
     xlab = "observed local geom-mean RP (yrs)",
     ylab = "observed local mean damage",
     pch = 21, bg = "white", cex = 1.2)
abline(lm(meand_n_events ~ meanRPln3_events), col="black")
add_r2(meanRPln3_events, meand_n_events)

# 6. Relazione simulata: RP geom vs Danno medio
plot(meanRPln3_simevents, meanD_n_simevents,
     xlab = "simulated local geom-mean RP (yrs)",
     ylab = "simulated local mean damage",
     pch = 21, bg = "white", cex = 1.2)
abline(lm(meanD_n_simevents ~ meanRPln3_simevents), col="black")
add_r2(meanRPln3_simevents, meanD_n_simevents)

#####

```