

**UNIVERSITÀ DEGLI STUDI DI PADOVA**  
**DIPARTIMENTO DI INGEGNERIA CIVILE, EDILE E**  
**AMBIENTALE**  
Department of Civil, Environmental and Architectural Engineering

Corso di Laurea Magistrale in Ingegneria Civile



**TESI DI LAUREA**

**STIMA DELLE PIENE ESTREME**  
**UTILIZZANDO MODELLI STATISTICI E**  
**CONCETTUALI DELLA RISPOSTA**  
**IDROLOGICA**

Relatore:

Prof. Marco Marani

Correlatore:

Dott. Stefano Basso

Laureando:

Pietro Devò

Matr. 1242564

ANNO ACCADEMICO 2021-2022



A mia nonna Rosa



# INDICE

INDICE DELLE FIGURE.....	3
INDICE DELLE TABELLE.....	7
INTRODUZIONE.....	9
1 TEORIA DEGLI EVENTI ESTREMI.....	13
1.1 Inquadramento storico.....	13
1.2 La teoria univariata classica.....	14
1.2.1 Teorema di massima stabilità.....	15
1.2.2 Teorema delle tipologie estreme.....	16
1.3 Criterio di attrazione per le tipologie limite.....	16
1.4 La distribuzione GEV.....	19
1.4.1 Il metodo della soglia.....	21
1.5 Il metodo di Cramer.....	24
1.5.1 La distribuzione asintotica.....	25
1.6 Le distribuzioni di probabilità giornaliere.....	27
1.7 MEVD: un approccio probabilistico.....	28
1.8 PHEVD: un approccio fisicamente basato.....	31
1.9 Alcuni recenti studi applicativi.....	34
2 METODOLOGIA.....	37
2.1 Procedura di analisi.....	37
2.2 Estrazione e selezione dei picchi.....	38
2.3 Modello di stima dei parametri.....	40
2.3.1 Stima dei parametri legati alla precipitazione.....	40
2.3.2 Stima dei parametri legati alla portata.....	40
2.4 Modello di calibrazione numerica.....	41
2.4.1 Nomenclatura adottata.....	42
2.4.2 La funzione negative-log-likelihood.....	42
2.4.3 Procedura di ottimizzazione.....	44
2.5 Calcolo dei quantili.....	48
2.6 Definizione delle metriche.....	50
2.6.1 Errore relativo.....	50

2.6.2 Scarto quadratico medio.....	50
2.6.3 Coefficiente di determinazione.....	51
2.6.4 Indice di classificazione.....	51
2.7 Strumenti di calcolo.....	52
2.8 Set di dati.....	53
3 RISULTATI.....	57
3.1 Analisi globale.....	57
3.1.1 Gli estremi secondo PHEVD.....	57
3.1.2 Gli estremi secondo MEVD.....	59
3.1.3 Modello delle recessioni.....	65
3.1.4 Stagionalità.....	66
3.2 Analisi dei campioni.....	70
3.2.1 I parametri delle recessioni.....	73
3.4 Alcuni casi esemplari.....	77
CONCLUSIONI.....	83
APPENDICE A.....	85
APPENDICE B.....	89
BIBLIOGRAFIA.....	97
RINGRAZIAMENTI.....	101

## INDICE DELLE FIGURE

Figura 1: Dinamiche del volume di eccesso (V) e della portata (q) ottenute dal modello meccanico-stocastico utilizzato (Botter e altri, 2009).....	32
Figura 2: Algoritmo di selezione dei picchi. Il quadrato indica il più recente picco delle recessioni identificato. La stella rappresenta il successivo massimo locale, vuota se non verrà selezionato, piena se verrà selezionato, in base al soddisfacimento di tutte le condizioni del metodo (Dralle e altri, 2017).....	38
Figura 3: Convenzione di codifica dei set di parametri.....	42
Figura 4: Schema logico dell'algoritmo di ottimizzazione.....	47
Figura 5: Schema di calcolo dei quantili.....	49
Figura 6: Localizzazione dei 178 bacini oggetto dell'analisi.....	54
Figura 7: Diagrammi della mediana e dei quartili (boxplot) degli errori relativi sulle osservazioni dei massimi stagionali per tutte le combinazioni di valutazione dei parametri effettuate, utilizzando la formulazione dei massimi di portata di PHEV.....	58
Figura 8: Diagrammi della mediana e dei quartili (boxplot) degli errori relativi sulle osservazioni dei massimi stagionali per tutte le combinazioni di valutazione dei parametri effettuate, utilizzando la formulazione dei massimi di portata di MEV.....	60
Figura 9: Diagrammi della mediana e dei quartili degli errori relativi sulle osservazioni dei massimi stagionali nelle combinazioni di adattamento di uno dei quattro parametri della distribuzione alla volta, con alle estremità le combinazioni limite 0000 e 1111..	61
Figura 10: Diagrammi della mediana e dei quartili degli errori relativi sulle osservazioni dei massimi stagionali nelle combinazioni di adattamento di due dei quattro parametri della distribuzione alla volta, con alle estremità le combinazioni limite 0000 e 1111.....	62
Figura 11: Diagrammi della mediana e dei quartili degli errori relativi sulle osservazioni dei massimi stagionali nelle combinazioni di adattamento di tre dei quattro parametri della distribuzione alla volta, con alle estremità le combinazioni limite 0000 e 1111.....	63
Figura 12: Diagrammi della mediana e dei quartili degli errori relativi sulle osservazioni dei massimi stagionali in tutte le combinazioni di adattamento possibili, imposta la calibrazione del parametro a.....	64

Figura 13: Diagrammi della mediana e dei quartili degli errori relativi sulle osservazioni dei massimi stagionali in tutte le combinazioni di adattamento possibili, imposta la calibrazione del parametro K.....	64
Figura 14: Diagrammi della mediana e dei quartili degli errori relativi sulle osservazioni dei massimi stagionali in tutte le combinazioni di valutazione dei parametri, avendo utilizzato un modello delle recessioni non-lineare (1) e lineare (2) per la stima di a e K. ....	65
Figura 15: Diagrammi della mediana e dei quartili degli errori relativi sulle osservazioni dei massimi stagionali in tutte le combinazioni di valutazione dei parametri, raggruppati per stagione.....	67
Figura 16: Diagrammi della mediana e dei quartili degli errori relativi sulle osservazioni dei massimi stagionali in tutte le combinazioni di valutazione dei parametri, raggruppati per stagione.....	67
Figura 17: Diagrammi della mediana e dei quartili del numero di recessioni estratte dal modello nell'intero periodo di osservazione, per stagione.....	68
Figura 18: Diagrammi della mediana e dei quartili dei valori di a stimati nelle diverse stagioni.....	69
Figura 19: Diagrammi della mediana e dei quartili dei valori di K stimati nelle diverse stagioni.....	69
Figura 20: Grafico di dispersione osservazioni-previsioni dei massimi periodici di portata per il campione dei migliori casi studio.....	71
Figura 21: Grafico di dispersione osservazioni-previsioni dei massimi periodici di portata per il campione dei migliori casi studio (blu), peggiori sovrastime (arancione) e peggiori sottostime (verde).....	72
Figura 22: Diagrammi della mediana e dei quartili degli errori relativi sulle osservazioni dei massimi stagionali nel modello PHEV (blu) e PHEV-MEV (arancione) per i tre campioni di casi studio.....	73
Figura 23: Diagrammi della mediana e dei quartili dei valori assunti dal parametro a stimato da adattamento di funzione delle recessioni (sinistra) e per ottimizzazione della distribuzione di probabilità degli eventi sui picchi di portata (destra) nei tre campioni di casi studio.....	74



Figura 24: Diagrammi della mediana e dei quartili dei valori assunti dal parametro $K$ stimato da adattamento di funzione delle recessioni (sinistra) e per ottimizzazione della distribuzione di probabilità degli eventi sui picchi di portata (destra) nei tre campioni di casi studio.....	74
Figura 25: Grafico di dispersione del valore stimato con modello lineare contro modello non-lineare per $a$ .....	75
Figura 26: Grafico di dispersione del valore stimato con modello lineare contro modello non-lineare per $K$ .....	76
Figura 27: Curva di probabilità di inondazione (sopra) e funzioni di probabilità cumulata di eccedimento degli eventi ordinari e dei massimi (sotto), con valori osservati, per il caso 18662000 summer.....	78
Figura 28: Curva di probabilità di inondazione (sopra) e funzioni di probabilità cumulata di eccedimento degli eventi ordinari e dei massimi (sotto), con valori osservati, per il caso 27615100 autumn.....	79
Figura 29: Curva di probabilità di inondazione (sopra) e funzioni di probabilità cumulata di eccedimento degli eventi ordinari e dei massimi (sotto), con valori osservati, per il caso 1301 winter.....	80
Figura 30: Curva di probabilità di inondazione (sopra) e funzioni di probabilità cumulata di eccedimento degli eventi ordinari e dei massimi (sotto), con valori osservati, per il caso 125 spring.....	81



## **INDICE DELLE TABELLE**

Tabella 1: Limitazioni dei parametri della distribuzione.....	45
Tabella 2: Grado di copertura delle osservazioni per lunghezze temporali crescenti, espresso mediante numero di bacini e relativa percentuale.....	55



## INTRODUZIONE

Il tema degli eventi di piena estremi è di fondamentale importanza nell'ingegneria e costituisce una pietra miliare nel campo dell'idrologia. La progettazione di qualsiasi opera civile per la gestione delle risorse idriche muove da un'analisi dei rischi, cioè una attenta valutazione delle possibilità che si verifichino quegli eventi la cui entità è ritenuta non accettabile, ovvero quando questi possano causare danni alla salute umana, all'ambiente, nonché alle opere stesse. Risulta impossibile non menzionare l'importanza che la questione ha assunto negli ultimi decenni, in particolar modo nel nuovo millennio, per effetto della notevole serie di eventi di natura sempre più catastrofica, che ha coinvolto il pianeta e l'uomo. La sensibilità è inevitabilmente mutata e soprattutto negli ultimi anni il dibattito internazionale è estremamente animato, anche e particolarmente per le ormai incontrovertibili evidenze che continuano ad essere avvalorate in merito ai cambiamenti climatici. Il compito non è evidentemente facile e, come si può intuire, determinante per una valutazione accurata del problema è il patrimonio della "memoria", ovvero le informazioni contenute nelle osservazioni storiche. La quantità di osservazioni storiche disponibili è tuttavia quasi sempre limitata e, inoltre, le finalità della progettazione richiedono molto spesso un'estrapolazione al di fuori dell'intervallo dei valori contenuti nei dati a disposizione. Rispetto al primo punto è evidente che, ad eccezione di casi studio di studio particolare, l'assenza di strumentazione è la principale problematica, non superabile alla scala temporale di breve e medio termine. Per quanto riguarda l'estrapolazione, essa comporta evidentemente un aumento dell'incertezza finale. La metodologia tradizionale per affrontare il problema consiste nell'adozione di modelli asintotici per descrivere la distribuzione dei valori estremi del processo naturale. In questo modo, un modello idoneo viene adattato ai dati e l'estrapolazione a valori mai osservati è possibile sulla base di esso. Nonostante il vasto numero di studi sull'argomento susseguitisi nello scorso secolo, il problema è in qualche modo ancora aperto: I metodi tradizionali presentano molte limitazioni e utilizzano una piccola frazione dell'informazione disponibile nelle osservazioni. In queste metodologie sono insiti due difetti sostanziali: Un primo elemento di criticità risiede nel fatto che si concentri l'attenzione sulle code dei fenomeni, dal momento che effettivamente lì sono individuati gli eventi estremi, le

quali tuttavia costituiscono una parte numericamente marginale rispetto all'intera distribuzione di probabilità. Un secondo elemento di criticità consiste, indipendentemente dal grado di sofisticatezza del modello adottato, nello spettro di informazioni induttive utilizzato per l'adattamento, che per approcci puramente statistici non contempla altro dominio se non quello dei dati stessi da riprodurre. Per i processi di cui si tratta, e di interesse nel campo ingegneristico, le leggi fisiche che regolano tali processi possono invece fornire ulteriori elementi quantitativi opportunamente concettualizzabili. Da ciò le incertezze che possono condurre a gravi errori nelle stime della probabilità di non accadimento degli eventi da utilizzarsi nei dimensionamenti. Il presente lavoro si basa su un approccio che vuole contribuire a superare queste limitazioni: Si adottano e combinano due metodi per lo studio della risposta idrologica, con l'obiettivo di ampliare il dominio induttivo utilizzato dal modello previsionale, mediante informazioni relative ai processi fisici in gioco. L'ampliamento dell'insieme di osservazioni utilizzate nel processo di inferenza statistica avviene mediante l'impiego del metodo '*Metastatistical Extreme Values distribution*' (MEVD, Marani e Ignaccolo, 2014), che permette di rimuovere l'ipotesi asintotica e di considerare esplicitamente il tasso secondo il quale si presentano gli eventi nel tempo come una variabile causale. L'approccio proposto include le informazioni e i vincoli relativi ai processi fisici rilevanti utilizzando un recente modello idrologico-statistico, "*PHysically based Extreme Value distribution*" (PHEV, Basso e altri, 2016), che definisce la distribuzione di probabilità degli eventi di piena in un corso d'acqua sulla base di caratteristiche fisiche del bacino esaminato. La tesi si struttura come segue: Nel Capitolo 1 vengono brevemente riportati alcuni cenni sulla teoria degli eventi estremi e si espongono i modelli adottati. È in questo modo possibile comprendere le basi di questa affascinante scienza e seguirne l'evoluzione di cui è stata oggetto, che permette di giungere agli approcci contemporanei che il presente lavoro prende in esame. Nel Capitolo 2 si espone tutto quello che concerne la metodologia con cui si sono strutturate le analisi, ovvero la procedura seguita, gli strumenti matematico-fisici, l'implementazione numerica, e i set di dati a disposizione. Questa sezione è descrittiva di una parte notevole del lavoro svolto, che qui non può che essere presentata concettualmente, ma che è formalizzazione della attività di implementazione numerica dei modelli che sono stati analizzati. Nel Capitolo 3 vengono riportati e analizzati i risultati ottenuti,

supportati da opportune restituzioni grafiche, che hanno permesso poi di pervenire alle conclusioni del presente lavoro, dove si raccolgono le principali caratteristiche e si propongono le possibili applicazioni del metodo, in una chiave che è quindi da quelle finalità pratiche che contraddistinguono l'ingegneria, chiamata a fornire delle risposte concrete alle sfide che la natura mette di fronte all'uomo.





# 1 TEORIA DEGLI EVENTI ESTREMI

## 1.1 Inquadramento storico

L'analisi statistica dei valori estremi, o “*Exreme Values*” (EV) in lingua inglese, nasce con l'articolo di Fischer e Tippett (1928), in cui fu enunciato per la prima volta il teorema del valore estremo e vennero definite le tre distribuzioni limite fondamentali. Questa scoperta viene successivamente rafforzata da Von Mises, che nel 1936 propose per primo l'espressione della distribuzione generalizzata dei valori estremi, o “*Generalized Exreme Values*” (GEV), ed identificò delle condizioni di convergenza per le tre leggi limitanti. Il teorema del valore estremo è stato poi dimostrato nel caso generale da Gnedenko nel 1943. A quest'ultimo va anche attribuito il merito di aver enunciato più formalmente la condizione per la convergenza debole delle leggi limite, definendo i corrispondenti domini di attrazione. Seguirono molti altri contributi, che affinarono la teoria asintotica e stabilirono il tasso di convergenza per un vasta gamma di funzioni di distribuzione. Tra gli altri, il lavoro di Gumbel (1958) è degno di essere menzionato, dal momento che ha applicato i metodi statistici dei valori estremi alla modellazione delle piene, permettendo il diffondersi dell'interesse per questo argomento tra gli idrologi e gli ingegneri. Negli ultimi decenni l'analisi degli eventi estremi ha sperimentato una diffusione esponenziale ed estesa a nuovi campi delle scienze e delle attività antropiche, come ad esempio nel settore assicurativo, nella finanza, nella biologia, nella medicina e via discorrendo. Allo stato odierno il tema è più che mai vivo e in continua evoluzione: Numerosi sono i contributi e le formulazioni elaborate, alcune delle quali si candidano a valenza universale, altre invece all'impiego in mirati casi di studio. Si assiste in altre parole ad un vero e proprio scontro sul campo da parte delle distribuzioni di probabilità, in cui l'esito finale è tutt'altro che scontato. Di interesse è ripercorrere brevemente gli aspetti principali della teoria classica ed alcuni teoremi fondamentali, per meglio comprendere come si perviene alle moderne analisi degli eventi estremi.

## 1.2 La teoria univariata classica

Viene qui esposta la teoria univariata classica, che costituisce il fondamento dell'analisi degli eventi estremi. Si supponga che  $X_1, \dots, X_n$  siano variabili casuali con una funzione di distribuzione cumulativa di probabilità data da:

$$F(x) = P(X_j \leq x) \quad \forall j=1, \dots, n \quad \forall x \in \Omega \quad (1.1)$$

dove  $\Omega$  rappresenta la popolazione comune di tutte le variabili  $X_i$ . Se le variabili sono a popolazione indipendente, allora la di distribuzione dei massimi  $M_n = \max \{ X_1, \dots, X_n \}$  di cardinalità  $n$  è data dalla potenza  $n$ -esima di  $F(x)$ :

$$P(M_n \leq x) = F_n(x) = F(x)^n \quad (1.2)$$

La tradizionale teoria dei valori estremi si concentra sul comportamento asintotico di questa distribuzione, dove il termine 'asintotico' si riferisce alla distribuzione limite per  $n \rightarrow \infty$ . Si noti che per definizione di distribuzione cumulativa di probabilità deve essere::

$$0 < F(x) < 1 \quad (1.3)$$

Quindi, risulta che, al crescere della dimensione del campione:

$$F(x)^n \rightarrow 0 \quad \text{per } n \rightarrow \infty \quad (1.4)$$

Volendo studiare il comportamento di tale distribuzione per campioni di grandi dimensioni è necessaria una operazione di normalizzazione, definendo una famiglia di costanti di scala,  $a_n$  e  $b_n$ , che permettano di ottenere una distribuzione non de-genera:

$$P\left(\frac{M_n - b_n}{a_n} \leq x\right) = P(M_n \leq a_n x + b_n) = F^n(M_n \leq a_n x + b_n) \quad (1.5)$$

Ottenendo che:

$$F^n(M_n \leq a_n x + b_n) \rightarrow H(x) \quad \text{per } n \rightarrow \infty \quad (1.6)$$

Emerge che esistono solamente tre tipologie principali di distribuzioni limitanti  $H_n$ , ovvero quelle di Gumbel, Frechet e Weibull. Questa evidenza prende il nome di "Teorema dei Tipi Estremi", enunciata per la prima volta da Fischer e Tippett (1928) e

successivamente dimostrato da Gnedenko (1943). Seguono due importanti definizioni, fondamentali per la comprensione del teorema:

- Date due funzioni di distribuzione,  $H_1$  e  $H_2$ , queste sono dette dello stesso tipo se è possibile trasformare l'una nell'altra attraverso una trasformazione lineare, dove si siano definite  $a > 0$  e  $b \in \mathfrak{R}$  come due costanti opportunamente scelte:

$$H_1(x) = H_2(ax + b) \quad (1.7)$$

- Una distribuzione di probabilità non degenerata  $F$  è definita massimamente stabile se per una sequenza di variabili indipendenti e identicamente distribuite  $(X_i)_{i \in N}$ , con distribuzione comune  $F$ , e per ogni  $n \in N$ , esistono  $a > 0$  e  $b \in \mathfrak{R}$  tali che la quantità:

$$\chi = \frac{M_n - b_n}{a_n} \quad (1.8)$$

abbia anch'essa distribuzione  $F$ . In altre parole, si definisce la condizione di massimamente stabile se vale:

$$F^n(a_n x + b_n) = F(x) \quad \forall x \in \mathfrak{R} \quad (1.9)$$

Quest'ultima proprietà implica che qualsiasi potenza di  $F$  non determina una modifica del tipo di distribuzione. Come conseguenza, se cambia la cardinalità del massimo considerato, un modello di valore estremo resterà coerente, ovvero ci sarà solo un cambiamento nei parametri mentre la forma della funzione di distribuzione rimarrà inalterata. La proprietà di massima stabilità garantisce che i due modelli rimarranno reciprocamente coerenti. La stessa proprietà vale se si considerano un modello per i superamenti di una determinata soglia: il modello rimarrà in ogni caso consistente. Per queste ragioni il concetto di stabilità risulta di fondamentale importanza.

### 1.2.1 Teorema di massima stabilità

Una distribuzione di probabilità non degenerata  $F$  è definita massimamente stabile se e solo se esiste una sequenza di variabili casuali,  $(X_i)_{i \in N}$ , indipendenti e identicamente distribuite, e date due successioni in modo che:

$$a > 0 \quad b \in \mathfrak{R} \quad (1.10)$$

in modo che valga la convergenza:

$$\frac{M_n - b_n}{a_n} \rightarrow H \quad (1.11)$$

### 1.2.2 Teorema delle tipologie estreme

Ciascuna distribuzione massimamente stabile  $H$  è definita di valore estremo se è della stessa tipologia di una delle seguenti tre distribuzioni:

Gumbel

$$H(x) = e^{-e^{-x}} \quad \text{per } -\infty < x < \infty \quad (1.12)$$

Frechet

$$H(x) = \begin{cases} e^{-x^{-\alpha}} & \text{per } 0 < x < \infty \\ 0 & x < 0 \end{cases} \quad (1.13)$$

Weibull

$$H(x) = \begin{cases} e^{-(-x)^{-\alpha}} & \text{per } -\infty < x < 0 \\ 1 & x > 0 \end{cases} \quad (1.14)$$

Viceversa, ogni distribuzione della tipologia dei valori estremi è massimamente stabile.

I due teoremi combinati implicano che, data  $M_n = \max_{1 \leq i \leq n} \{X_i\}$  la sequenza di variabili indipendenti e identicamente distribuite, ogni limite non-degenere della (1.8) rientra in una delle tre tipologie di estremo. Al contrario, ogni distribuzione di valore estremo è il limite debole della (1.8) per due opportune successioni  $a > 0$  e  $b \in \mathfrak{R}$ , scelta  $H$  come distribuzione di  $(X_i)_{i \in \mathbb{N}}$ .

## 1.3 Criterio di attrazione per le tipologie limite

Per una tipologia  $H$ , il *Dominio di Attrazione* è definito come il set di funzioni di distribuzione  $F$  tali che, data una sequenza di variabili casuali  $(X_i)_{i \in \mathbb{N}}$ , indipendente e identicamente distribuite, quest'ultima ne sia la distribuzione cumulata, valga la convergenza debole. Per distribuzioni sufficientemente regolari è possibile determinare la tipologia limite  $H$  definendo la *Funzione Reciproca di Rischio* come:

$$r(x) = \frac{1-F(x)}{f(x)} \quad (1.15)$$

Inoltre definite le successioni:

$$a_n = r(b_n) \quad (1.16)$$

$$b_n = F^{-1}\left(1 - \frac{1}{n}\right) \quad (1.17)$$

La distribuzione limite risulta essere:

$$e^{-(1+\xi x)_+^{\frac{1}{3}}} \quad \text{per } \xi \neq 0 \quad (1.18)$$

$$e^{-e^{-x}} \quad \text{per } \xi = 0 \quad (1.19)$$

Il parametro di forma può essere determinato grazie al limite:

$$\xi = \lim_{x \rightarrow \infty} r'(x) \quad (1.20)$$

A seguire si vogliono derivare i criteri di attrazione per le tre distribuzioni definite dal Teorema delle tipologie estreme.

### Gumbel

Quando una distribuzione residuale presenta una coda con decadimento più veloce di qualsiasi funzione polinomiale quando la variabile indipendente tende ad infinito, essa appartiene al dominio di attrazione della funzione di Gumbel. In questo caso ricade ogni funzione esponenziale della forma:

$$1-F(x) = e^{-h(x)} \quad (1.21)$$

essendo  $h(x)$  una funzione derivabile positiva e monotona, di crescita positiva e più rapida di qualsivoglia legge logaritmica. Una distribuzione di famiglia esponenziale rientra nel dominio di attrazione di Gumbel se almeno una delle seguenti condizioni risulta soddisfatta:

- $h'(x) = x^{\alpha-1}L(x)$  per alcuni valori  $\alpha > 0$  e funzione  $L$  debolmente variabile, ovvero che verifichi:

$$\frac{\lim_{x \rightarrow \infty} L(xy)}{L(x)} \quad \forall y > 0 \quad (1.22)$$

- $h(x) = x^\alpha L(x)$  per alcuni valori  $\alpha > 0$  e funzione  $L$  debolmente variabile, con  $h'(x)$  monotona tra  $x_0$  e  $\infty$  per alcuni valori  $x_0 > 0$ .

Si può esprimere il tutto grazie alla condizione di Von Mises, la quale afferma che, data una densità cumulata di frequenza  $F(x)$  e la sua derivata funzione di densità di probabilità  $f(x)$ , Gumbel è la sua funzione limite se e solo se:

$$\frac{d}{dx} \left( \frac{1-F(x)}{f(x)} \right) \rightarrow 0 \quad \text{per } x \rightarrow \infty \quad (1.23)$$

In questo caso esistono due costanti, ovvero  $a_n$  e  $b_n$ , per cui Gumbel è la funzione limite per i valori estremi. Nella condizione più banale, dove si ponga  $h(x) = x$ , ovvero una funzione esponenziale con media unitaria, e siano:

$$a_n = 1 \text{ e } b_n = \log(n) \quad (1.24)$$

si perviene, per  $x \rightarrow \infty$ , a:

$$F^n(a_n x + b_n) = (1 - e^{-x - \log(n)})^n = \left( 1 - \frac{e^{-x}}{n} \right)^n \rightarrow e^{-e^{-x}} \quad (1.25)$$

ovverosia fornendo la distribuzione di Gumbel come funzione limite.

### Frechet

Il dominio di attrazione consiste in questo caso in tutte quelle funzioni di distribuzione  $F(x)$  tali che sia  $r'(x) > 0$ . Qualsiasi funzione densità cumulata di frequenza la cui coda sia espressa da una legge di potenza del tipo:

$$1 - F(x) \sim x^{-\alpha} \quad \text{per } \begin{matrix} x \rightarrow \infty \\ k > 0 \\ \alpha > 0 \end{matrix} \quad (1.26)$$

risulta nel dominio di attrazione di Frechet, con medesimo esponente  $\alpha$ . Nella condizione più banale di distribuzione di Pareto di primo tipo, dove si ponga  $1 - F(x) = k x^{-\alpha}$ , e siano:

$$a_n = (nk)^{\frac{1}{\alpha}} \text{ e } b_n = 0 \quad (1.27)$$

si perviene, per  $x \rightarrow \infty$ , a:

$$F^n(a_n x + b_n) = (1 - k(a_n x)^{-\alpha})^n = \left(1 - \frac{x^{-\alpha}}{n}\right) \rightarrow e^{-x^{-\alpha}} \quad (1.28)$$

ovverosia fornendo la distribuzione di Frechet, caratterizzata da parametro di forma  $\alpha$ , come funzione limite.

### Weibull

Preso una funzione di distribuzione  $F(x)$  che sia caratterizzata da un limite superiore  $\omega_F$  e da un comportamento che sia in generale riconducibile ad una legge di potenza per  $x \rightarrow \omega_F$ , in modo che risulti:

$$1 - F(\omega_F - y) \sim k y^\alpha \quad \text{per } x \rightarrow \infty \quad (1.29)$$

per alcuni valori delle costanti  $\alpha > 0$  e  $k > 0$ . Se risultano soddisfatte tali condizioni, è la funzione di distribuzione nel dominio di attrazione di Weibull.

## 1.4 La distribuzione GEV

Von Mises, nel 1936, ha proposto la formulazione di un'unica distribuzione che inglobasse tutte e tre le famiglie di limite fondamentali:

$$H(x; \xi; \psi; \mu) = e^{-\left(1 + \frac{\xi}{\sigma}(x - \mu)\right)_+^{-1/\xi}} \quad (1.30)$$

La funzione è definita per ogni valore di  $x$ , in modo che:

$$1 + \frac{\xi}{\sigma}(x - \mu) > 0 \quad (1.31)$$

Lo spazio dei parametri si caratterizza come segue:

- $\mu \in \mathcal{R}$  è il parametro di localizzazione ed indica il valore di  $x$  che costituisce il centro della *PDF*;
- $\sigma > 0$  è il parametro di scala, che controlla la dispersione della distribuzione attorno alla localizzazione  $\mu$ ;
- $\xi$  è il parametro di forma, che determina il tasso di decadimento della coda. Un valore maggiore di 0 riconduce al caso di coda pesante (Frechet), dove

quest'ultima è descritta da una legge di potenza con esponente  $\alpha=1/\xi$ ; Nel caso in cui  $\xi$  tende a 0 la distribuzione GEV corrisponde a Gumbel, ovvero caso di cosa leggera, e il decadimento risulta esponenziale. Valori minori di 0 portano infine ad una funzione di Weibull inversa.

Sia  $f(x)$  la distribuzione GEV nel caso di una tipologia limite di Gumbel, quindi  $\xi=0$ . Si può osservare che coda destra della distribuzione decade seguendo una legge esponenziale di forma:

$$f(x) \sim e^{-x} \quad \text{per } x \rightarrow +\infty \quad (1.32)$$

mentre la coda sinistra decade seguendo una legge doppio-esponenziale:

$$f(x) \sim e^{-e^x} \quad \text{per } x \rightarrow -\infty \quad (1.33)$$

Nel caso di una tipologia limite di Frechet, la distribuzione GEV presenta  $x=0$  come limite sinistro e la coda destra decade con forma:

$$f(x) \sim x^{-\alpha} \quad \text{per } x \rightarrow +\infty \quad (1.34)$$

Nel caso di una tipologia limite di Weibull, la distribuzione GEV presenta  $x=0$  come limite destro e la coda sinistra decade con forma:

$$f(x) \sim x^{-\alpha} \quad \text{per } x \rightarrow -\infty \quad (1.35)$$

Se  $\xi < 1/k$ , esiste il momento di ordine  $k$  della distribuzione GEV. L'esistenza del valore medio e della varianza, le cui espressioni sono definite come:

$$E(X) = \mu + \frac{\sigma}{\xi} [\Gamma(1-\xi) - 1] \quad (1.36)$$

$$E([X - E(X)]^2) = \frac{\sigma^2}{\xi^2} [\Gamma(1-2\xi) \Gamma^2(1-\xi)] \quad (1.37)$$

è assicurata, rispettivamente, per  $\xi < 1$  e  $\xi < 1/2$ . Nel caso limite  $\xi \rightarrow 0$  queste si riducono al valore medio e alla varianza della distribuzione di Gumbel:

$$E(X) = \mu + \sigma \gamma \quad (1.38)$$

$$E([X - E(X)]^2) = \frac{\sigma^2 \pi^2}{6} \quad (1.39)$$

Per adattare la distribuzione GEV ad un campione di massimi di portata è necessario



definire cosa questi siano e come ottenerli dalle osservazioni di cui si dispone. Il modo più semplice e utilizzato è quello di utilizzare dei blocchi di pari lunghezza temporale per estrarre una sequenza di massimi. Nel caso di osservazioni giornaliere di portate questo blocco è definito dall'anno o dalla stagione, quindi i massimi così estratti prendono il nome di massimi annuali o stagionali.

### 1.4.1 Il metodo della soglia

L'utilizzo del blocco dei massimi per adattare la distribuzione GEV ad una serie di osservazioni ha il difetto di utilizzare un singolo valore per ciascun blocco (anno o stagione). Questo approccio può determinare la perdita di importanti informazioni, vista la dimensione contenuta del campione. Il campione dei massimi periodi potrebbe non essere affatto rappresentativo della coda effettiva della distribuzione degli eventi sottesa. Ad esempio potrebbero non essere considerati numerosi eventi di grande magnitudine verificatisi in uno stesso periodo, per il quale solo un massimo viene estratto, sebbene questi possano in assoluto risultare superiori a manifestazioni di periodi successivi. Per superare queste limitazioni, l'approccio alternativo principale dell'idrologia è quello di utilizzare un metodo denominato "*Peak Over Threshold*" (POT), ovvero di definire i picchi sulla base del superamento di una certa soglia (Balkema e De Haan, 1974; Pickand, 1975). Si è dimostrato che se esiste una distribuzione degli eccedimenti, opportunamente scalati linearmente, di forma:

$$Y_i = X_i - q \quad (1.39)$$

provenienti da una funzione di ri di osservazioni indipendenti e identicamente distribuite, che superano una definita soglia  $q$ , la distribuzione limite risultante è una distribuzione generalizzata di Pareto (GPD). In termini pratici, si fissa una soglia sufficientemente alta del valore di portata e si selezionano così tutti i suoi eccedimenti. In questo approccio, il numero degli eccedimenti e l'entità degli stessi rappresentano due variabili casuali. Si consideri la variabile aleatoria  $X$ , con funzione di ripartizione  $F$ , e si definisca la condizione di superamento (1.39) condizionata a  $X > u$ . Si ottiene in questo modo che:

$$P(Y \leq y) = P(x \leq x+q | X > q) = F_q(y) = \frac{F(q+y) - F(q)}{1 - F(q)} \quad (1.40)$$

Quando il valore soglia tende ad infinito, è possibile dimostrare l'esistenza di un set di parametri  $\xi$  e  $\sigma_q$  tali che la distribuzione generalizzata di Pareto rappresenti un'ottima approssimazione per la distribuzione degli eccedimenti:

$$F_q(y) \sim F(y; \xi; \sigma_q) = 1 - \left(1 + \frac{\xi}{\sigma_q} y\right)^{-\frac{1}{\xi}} \quad \text{per } x \rightarrow +\infty \quad (1.41)$$

Si distinguono tre diversi casi, a seconda del segno assunto dal parametro  $\xi$ :

- $\xi > 0$ : la funzione è definita nell'intervallo  $0 < y < \infty$  e la coda della distribuzione decade con legge di potenza. La corrispondenza che si riscontra è con la tipologia di estremo di Frechet;
- $\xi < 0$ : la funzione presenta un estremo superiore, similmente a quanto avveniva nella tipologia di Weibull, definito come:

$$\omega_F = \frac{\sigma_q}{|\xi|} \quad (1.42)$$

- $\xi \rightarrow 0$ : la funzione diviene distribuzione esponenziale con valore medio  $\sigma_q$ , analogamente a quanto avveniva in Gumbel:

$$F(y; \sigma_q; 0) = 1 - e^{-\frac{y}{\sigma_q}} \quad (1.43)$$

Così come per la GEV, anche per la GPD l'esistenza del valore medio e della varianza, le cui espressioni sono definite come:

$$E(Y) = \frac{\sigma}{1 - \xi} \quad (1.44)$$

$$E([Y - E(Y)]^2) = \frac{\sigma^2}{(1 - \xi)^2 (1 - 2\xi)} \quad (1.45)$$

è assicurata, rispettivamente, per  $\xi < 1$  e  $\xi < 1/2$ . Il numero  $n$  di eccedimenti in un anno, definita una certa soglia  $q$ , è una variabile casuale. L'approccio generalmente adottato prevede la modellazione della variabile casuale  $N$  utilizzando una distribuzione di Poisson, dove sia il valore atteso definito come:

$$E(N) = E([N - E(N)]^2) = \lambda \quad (1.46)$$

In questo modo si può scrivere la probabilità di accadimento di un numero  $n$  di

eccedimenti in un anno:

$$P(N=n)=f(n;\lambda)=\frac{\lambda^k e^{-\lambda}}{k!} \quad (1.47)$$

Assumendo che il numero annuo di eccedimenti abbia distribuzione di Poisson, caratterizzati da valore medio  $\lambda$ , e siano tutti realizzazioni indipendenti con distribuzione GPD, si perviene all'espressione della probabilità che un massimo annuale sia maggiorato da un valore di  $x$ :

$$P(\max_{1 \leq i \leq N} \leq x) = P(N=0) + \sum_{k=1}^{\infty} P(N=n; Y_1 \leq x, \dots, Y_n \leq x) = e^{-\lambda \left(1 + \frac{\xi x - q}{\sigma}\right)^{-\frac{1}{\xi}}} \quad (1.48)$$

Emerge quindi che la distribuzione di probabilità dei massimi annuali di una GPD con modello Poissoniano è la medesima di una GEV con parametri  $\xi, \psi$  e  $\mu$ , da cui consegue che i due modelli risultano tra loro sovrapponibili se:

$$\begin{cases} \xi = \xi \\ \sigma = \psi + \xi(q - \mu) \\ \lambda = \left(1 + \frac{\xi(q - \mu)}{\psi}\right)^{-\frac{1}{\xi}} \end{cases} \quad (1.49)$$

La risoluzione del sistema porge infine:

$$\begin{cases} \xi = \xi \\ \psi = \sigma \lambda^{\xi} \\ \mu = q - \frac{\sigma(1 - \lambda^{\xi})}{\xi} \end{cases} \quad (1.50)$$

La distribuzione generalizzata di Pareto soddisfa la proprietà di stabilità della soglia. Se viene fissato un valore più alto risulta lo stesso parametro di forma per la distribuzione, ma parametro di scala traslato cioè se si fissa una soglia più alta, ovvero:

$$\sigma_2 = \sigma_1 + \xi(u_2 - u_1) \quad (1.51)$$

Solo nel caso limite in cui si assume  $\xi=0$  la dipendenza del parametro di scala dalla soglia scompare, riducendosi per altro la GPD ad una distribuzione esponenziale. L'approccio appena esposto permette una stima dei parametri della GEV basata sul comportamento reale della distribuzione giornaliera degli eventi di piena, offrendo una migliore descrizione rispetto a quella ottenibile con i più semplici metodi tradizionali,

che considerano i massimi annuali. È però allo stesso importante tener presente che la scelta di valori di soglia opportuni è quanto mai fondamentale e, generalmente, questi non dovranno essere mai troppo bassi, per mantenere coerenza con l'adozione della distribuzione Poissoniana del numero di eccedimenti e la distribuzione di Pareto dei valori di superamento.

## 1.5 Il metodo di Cramer

L'interesse delle analisi è studiare la distribuzione di probabilità dei massimi di una qualche quantità naturale variabile tra blocchi di lunghezza temporale fissata  $T$ . Sia dunque  $n$  il numero di realizzazioni della varianza relativa nell'intervallo considerato. Si vuole cercare la distribuzione dei massimi degli  $n$  campioni indipendenti, ciascuno ottenuto dalla medesima funzione principale  $F(x) = P(X \leq x)$ . La teoria classica della probabilità, come si è visto in precedenza, indica che la funzione cumulata di probabilità risulta essere:

$$F_n(x) = [F(x)]^n \quad (1.52)$$

Si introduce la definizione più grande valore caratteristico di  $X$ :

$$Q(X_T) = P(X \geq X_T) = \frac{1}{n} \quad (1.53)$$

Il più grande valore caratteristico è quello che viene ecceduto con probabilità definita come reciproco di  $n$ . In questo modo, se si esprime l'equazione (1.52) secondo la funzione di probabilità residuale, ovvero:

$$P(X \geq x) = Q(x) = 1 - F(x) \quad (1.54)$$

si perviene alla espressione che segue:

$$F_n(x) = [1 - Q(x)]^n = \left[ 1 - \frac{Q(x)}{nQ(X_T)} \right] \quad (1.55)$$

Considerando un valore  $x > X_T$ , risulterà:

$$\frac{Q(x)}{nQ(X_T)} \leq 1 \quad (1.56)$$

che corrisponde cioè a casi in cui si prendono in esame tempi di ritorno superiori all'anno. Essendo poi, in ogni caso,  $n > 1$ , l'espressione (1.55) può essere semplificata invocando l'approssimazione di Cauchy, facendo riferimento all'espansione di primo ordine della funzione:

$$F_n(x) = \left[ 1 - \frac{Q(x)}{nQ(X_T)} \right] \simeq 1 - n \frac{Q(x)}{nQ(X_T)} \simeq e^{-\frac{Q(x)}{Q(X_T)}} \quad (1.57)$$

Questa espressione linearizzata viene frequentemente chiamata col nome di approssimazione Penultima. La dipendenza tra l'errore nel trascurare i termini di ordine superiore è direttamente proporzionale ad  $n$  ed inversamente proporzionale alla differenza  $x - X_T$ . Se nello specifico si considera la variabile:

$$z = -\frac{Q(x)}{Q(X_T)} \quad (1.58)$$

e si fissa un suo valore, ovvero se si va a fissare la distanza dal punto in cui è definito il centro della serie, l'errore viene a dipendere unicamente da  $n$ . Questa può essere dimostrata valutando l'errore associato al più grande valore caratteristico  $X_T$ . Risulta, utilizzando l'approssimazione di Cauchy, che:

$$F_n(X_T) = e^{-1} \quad (1.59)$$

mentre il risultato esatto viene a dipendere da  $n$ . L'errore si ottiene per differenza:

$$\delta(n) = e^{-1} - \left( 1 - \frac{1}{n} \right)^n \quad (1.60)$$

Risulta che:

$$Q(x) < Q(X_T) \quad \text{per } x > X_T \quad (1.61)$$

da cui ne consegue che l'equazione (1.59) tende a sovrastimare l'errore che deriva dall'approssimazione di Cauchy, mentre l'errore tende rapidamente a zero al tendere di  $n$  ad infinito, ovvero all'aumentare del campione.

### 1.5.1 La distribuzione asintotica

Utilizzando la linearizzazione fornita dalla (1.57), Cramer ha derivato la distribuzione

di Gumbel. La distribuzione principale degli eventi di piena è detta di tipo esponenziale se può essere scritta nella forma seguente:

$$F(x) = 1 - e^{-g(x)} \quad (1.62)$$

avendo definito la funzione  $g(x)$  come positiva dal comportamento monotono crescente più rapido di una funzione logaritmica. Dalla seconda forma dell'approssimazione di Cauchy per funzione esponenziale è possibile ottenere:

$$F(x) \simeq e^{-\frac{e^{-g(x)}}{e^{g(x_T)}}} \quad (1.63)$$

$$y_T = -\ln[-\ln(F_n(x))] = g(x) = -g(X_T) \quad (1.64)$$

Questa rappresenta la variabile ridotta di Gumbel. Espandendo in serie di Taylor quest'ultima equazione in un intorno del più grande valore caratteristico  $X_T$ , si ottiene:

$$y_T = g'(X_T)(x - X_T) + \frac{g''(X_T)}{2}(x - X_T)^2 + \dots \quad (1.65)$$

che rappresenta ancora l'approssimazione Penultima per una variabile caratterizzata da funzione di distribuzione esponenziale. Non è necessaria la conoscenza della funzione  $g(x)$  per la sua valutazione e l'errore deriva unicamente dall'applicazione dell'approssimazione di Cauchy. Trascurando i termini di ordine superiore al primo nell'equazione (1.65), si ottiene la forma asintotica finale per Gumbel:

$$y_T = \alpha(x - \mu) \quad (1.66)$$

avendo rinominato le quantità come segue:

$$\mu = X_T \quad (1.67)$$

$$\alpha = g'(X_T) \quad (1.68)$$

In questo modo si può notare immediatamente che, quanto è stato ottenuto è semplicemente la funzione cumulata della distribuzione di Gumbel:

$$F(x) = e^{-e^{-\alpha(x-\mu)}} \quad (1.69)$$

Essendo una distribuzione asintotica, la dipendenza con  $n$  cessa di esistere, dal momento che viene applicata al limite  $n \rightarrow \infty$ . Allo stesso modo, non vi alcuna

dipendenza dalla forma particolare di  $g(x)$ . Quanto appena visto può essere analogamente applicato alla distribuzione di Weibull, dove:

$$F(x) = P(X \leq x) = 1 - e^{-\left(\frac{x}{C}\right)^\omega} \quad (1.70)$$

L'errore commesso utilizzando la distribuzione asintotica dipende dal grado di convergenza all'asintoto, ovvero al valore di  $n$ . L'errore deriva dai termini di grado superiore al primo trascurati nella (1.65) per ottenere la distribuzione di Gumbel, che è possibile stimare come segue:

$$\frac{\epsilon(y_T)}{y_T} = \frac{h''(x_T)(x - X_T)^2}{2h'(x_T)(x - X_T)} \quad (1.71)$$

Nel caso di Weibull risulta la formulazione:

$$\frac{\epsilon(y_T)}{y_T} = \frac{\omega - 1}{2} \left( \frac{x}{x_T} - 1 \right) \quad (1.72)$$

Si osservi che:

- $\omega = 2$ , il valore di tutti i termini trascurati è zero, quindi la forma dell'errore esatta. Questo è il caso di una distribuzione di Rayleigh;
- $\omega = 1$  l'errore è nullo e la distribuzione è esponenziale.

Da quest'ultima considerazione si può concludere che, nel caso di una funzione di distribuzione principale esponenziale, la forma asintotica ultima di Gumbel è esatta e l'errore è semplicemente associato all'approssimazione di Cauchy. Nel caso generale in cui  $\omega \neq 1$ , la distribuzione asintotica è affetta sia dall'errore dell'approssimazione di Cauchy, sia da quello legato ai termini trascurati nella (1.65).

## 1.6 Le distribuzioni di probabilità giornaliere

Il criterio di attrazione ed il metodo di Cramer suggeriscono che la distribuzione principale scelta per eventi ordinari giochi un ruolo fondamentale nella determinazione della forma della distribuzione degli eventi estremi. Effettivamente, la coda destra della distribuzione governa la probabilità di permanenza di quegli eventi caratterizzati da un'intensità "sufficientemente elevata". Per quanto concerne gli eventi che abbiano una

scala inferiore, non è nota una legge generale per la distribuzione di probabilità. Si è visto che l'approccio più semplice ed utilizzato per affrontare il problema, assume che gli interarrivi degli eventi seguano una distribuzione Poissoniana, mentre le intensità degli stessi possano descriversi con una opportuna distribuzione appartenente alla famiglia esponenziale. Questo approccio ha mostrato dei limiti in molteplici situazioni: Spesso la coda della distribuzione risulta essere sotto-esponenziale e l'adozione di una funzione leggera conduce alla sottostima delle probabilità per gli eventi più estremi. Un secondo elemento di criticità è nell'ipotesi di distribuzione Poissoniana degli interarrivi per gli eventi, dal momento che gli eventi estremi, siano essi precipitazioni o piene fluviali, tendono a verificarsi a blocchi. La distribuzione di Poisson andrebbe usata esclusivamente per la modellazione degli interarrivi in caso di eccedimenti di una certa magnitudine (Leadbetter, 1983). Più alta è la soglia adottata e maggiore è il grado di accordo tra la distribuzione empirica dei dati ed il modello di Poisson

## 1.7 MEVD: un approccio probabilistico

L'espressione della probabilità cumulata per il massimo  $M_n$  di un campione di eventi annuale è, come evidenziato in precedenza:

$$P(M_n \leq x) = H_n(x; \vec{\theta}; n) = F_n(x; \vec{\theta}) \quad (1.73)$$

avendo definito con  $n$  il numero di eventi per il periodo, o blocco, di riferimento, generalmente pari ad un anno. In questo modo la distribuzione cumulativa dei massimi annuali viene a dipendere da  $n$  e dal vettore  $\vec{\theta}$  dei parametri della distribuzione dei "valori ordinari" principale. Dal momento che la procedura di stima dei parametri dipende dal set di dati, i parametri stessi sono da considerarsi delle variabili stocastiche. Analogamente il numero dei eventi annuali  $n$  altro non è che la realizzazione di una variabile discreta stocastica  $N$ . Da questo consegue che una definizione generale della cumulata dei massimi annuali dovrebbe tenere in considerazione questa natura stocastica di ordine superiore, e ciò può farsi considerando il valore atteso di  $H_n(x; n; \vec{\theta})$  computato su tutte le possibili realizzazioni di  $n$  e  $\vec{\theta}$ :

$$\zeta(x) = \sum_{n=1}^{\infty} \int_{\vec{\theta}} g(n; \vec{\theta}) H_n(x; n; \vec{\theta}) d\vec{\theta} \quad (1.74)$$



essendo  $g(n; \vec{\theta})$  la distribuzione di probabilità congiunta delle variabili aleatorie  $\{N, \Theta_1, \dots, \Theta_n\}$  e  $d\vec{\theta}$  il differenziale  $d\theta_1 \cdot d\theta_2 \cdot \dots \cdot d\theta_n$ . Questa espressione è stata proposta come “*Metastatistic Extreme Value distribution*” (MEVD, Marani e Ignaccolo, 2014). La formulazione generale (1.74) si presenta decisamente attraente, tuttavia può nella pratica risultare tutt’altro che banale individuare una espressione analitica per le distribuzioni  $g(n; \vec{\theta})$  e  $H_n(x; n; \vec{\theta})$ . In assenza di ciò è possibile usare unicamente delle distribuzioni empiriche dei parametri. La definizione della distribuzione MEV si fonda sull’approssimazione del valore atteso nell’equazione (1.62) attraverso la media campionaria. Data una funzione di variabile aleatoria  $f(\vec{\theta})$  basata su una funzione di densità di probabilità  $p(\theta)$ :

$$\int p(\vec{\theta})f(\vec{\theta})d\vec{\theta} \simeq \frac{1}{T} \sum_{j=1}^M f(\vec{\theta}_j) \quad (1.75)$$

La precisione di questa approssimazione è tanto maggiore quanto lo è il numero di elementi  $M$  campionati. Nel caso di campioni indipendenti, l’equazione (1.63) è conseguenza diretta della legge dei grandi numeri, ma il risultato ha valenza in termini generali. Per dimostrare il risultato è necessario riconoscere che il numero di volte che un dato valore  $f(\vec{\theta}_*)$  ricorre nella sommatoria è pressoché  $M p(\vec{\theta}_*)d\theta_*$  e che maggiore è  $M$  e più prossimo è il valore reale a questa quantità. Ciò sussiste per ogni valore di  $\theta_* \in \Omega_{\vec{\theta}}$  e dunque, assumendo che il set di valore reali possa essere trattato come contabile, senza perdere di alcuna generalità, è possibile esprimere il termine destro della (2.3) come:

$$\frac{1}{T} \sum_{j=1}^M f(\vec{\theta}_j) \simeq \sum_{\vec{\theta}_* = \inf(\Omega_{\vec{\theta}})}^{\sup(\Omega_{\vec{\theta}})} f(\vec{\theta}_*) p(\vec{\theta}_*) d(\vec{\theta}_*) \quad (1.76)$$

che costituisce una discretizzazione dell’integrale originariamente fornito nella (1.75). Conseguenza che, nel caso in cui non sia nota a priori la distribuzione congiunta  $g(n; \vec{\theta})$ , è ancora possibile una stima della (1.74). Fornita ad esempio una serie storica di  $T$  anni di eventi ordinari, è possibile scrivere  $\zeta(x)$  come una media di  $H_n(x)$  su tutti gli accadimenti annuali di  $n$  e  $\vec{\theta}$ :

$$\zeta(x) = \frac{1}{T} \sum_{j=1}^T H_{n_j}(x; n_j; \vec{\theta}_j) \quad (1.77)$$

Tale espressione può formalmente ricavarsi anche dalla (1.62), considerando la densità di probabilità congiunta di  $n$  e  $\vec{\theta}$  come una bivariata delta di Dirac, centrata in  $n$  e  $\vec{\theta}$ , ovvero nelle attuali realizzazioni delle variabili aleatorie associate all'anno  $j$ -esimo:

$$g(n; \vec{\zeta}) = \frac{1}{T} \sum_{j=1}^T \delta(n - n_j; \vec{\theta} - \vec{\theta}_j) \quad (1.78)$$

È possibile in questo modo, successivamente alla selezione di una opportuna distribuzione per la descrizione degli eventi ordinari, ottenere l'espressione della *CDF* dei massimi annuali. L'uso della media campionaria permette l'utilizzo della distribuzione empirica in luogo di un modello parametrico per  $\vec{\theta}$  e  $N$ . Un singolo anno può risultare insufficiente come campione per conseguire una buona stima del set di parametri. In questo senso risulta fondamentale una opportuna sensibilità nella selezione della dimensione del campione di adattamento in funzione della forma analitica particolare della distribuzione di probabilità adottata. La formulazione di MEV quindi permette di considerare in maniera sistematica la variabilità dei parametri della distribuzione nel dominio del problema. Nel presente studio si vuole concentrare l'attenzione sui processi di definizione degli eventi estremi a partire dal set di eventi ordinari e non si indagherà sulla variabilità temporale dei parametri. Una importante motivazione sottesa a questa scelta è anche insita nell'approccio perseguito: Adottando modelli concettuali, si introducono parametri caratterizzanti secondo alcune particolarità fisiche, la cui stima può essere negativamente condizionata dall'adozione di finestre temporali limitate. In altre parole il contributo positivo che la densità temporale parametrica apparentemente può offrire è potenzialmente vanificato dalle incertezze che affliggono la stima. La formulazione della (1.77) diviene:

$$\zeta(x) = \frac{1}{T} \sum_{j=1}^T H_{n_j}(x; n_j; \vec{\theta}) \quad (1.79)$$

Il vettore dei parametri perde cioè il pedice  $j$ , decadendo così la dipendenza temporale. Questo significa che il vettore degli  $m$  parametri è riferito all'intera finestra temporale considerata e, dunque, algebricamente è:

$$\vec{\theta} = \begin{pmatrix} \theta_1 \\ \cdot \\ \dot{\theta}_m \end{pmatrix} \in \mathfrak{R}^{m,1} \quad (1.80)$$

## 1.8 PHEVD: un approccio fisicamente basato

La curva di frequenza delle inondazioni oggetto di studio si deriva mediante la costruzione di un modello meccanico-stocastico degli eventi giornalieri (Botter e altri, 2009). Le dinamiche che sottendono tali eventi risultano da un bilancio, condotto alla scala di bacino, dei valori di umidità del suolo nell'area radicale (Laio e altri, 2001) secondo incrementi stocastici, per via delle infiltrazioni dovute alle precipitazioni giornaliere, assunte essere eventi Poissoniani, e delle perdite per evapotraspirazione. Nel momento in cui il flusso impulsivo meteorico bilancia il deficit creato dall'evapotraspirazione e determina il superamento una determinata soglia di umidità del suolo (ovvero la capacità di campo dello stesso), avviene l'innescò di un evento efficace, in grado di generare una portata. Quanto appena descritto viene modellato come evento Poissoniano con frequenza  $\lambda (T^{-1})$  ed esponenzialmente distribuito secondo un valore medio  $\alpha(L)$ . La precipitazione effettiva ricarica il volume di eccedimento del bacino (ovvero la frazione di riserva complessiva che può essere drenata) e contribuisce eventualmente alla portata generata. Secondo questa struttura, le dinamiche del volume di eccedimento  $V$  sono governate dall'equazione:

$$\frac{dV}{dt} = -q + \xi(t) \quad (1.81)$$

dove  $\xi(t)$  rappresenta una sequenza Poissoniana di incrementi non correlati di immagazzinamento dovuti agli eventi efficaci di precipitazione e  $q$  è una portata specifica. Quest'ultima viene concettualizzata secondo una relazione volume-deflusso del tipo  $q = \rho(V)$  (Botter e altri, 2009), che quantifica in che modo l'eccesso d'acqua sia drenato dal bacino e si traduca in portata. In questo caso viene impiegata una legge di potenza per la relazione di bilancio, che permette di legare la velocità di drenaggio del bacino al livello di immagazzinamento, implicitamente incorporando in questo modo gli effetti di diverse componenti che concorrono alla portata finale, ossia il deflusso sub-superficiale e superficiale (Basso e altri, 2015). In accordo con tali

assunzioni, la distribuzione di densità di probabilità del volume di eccedimento,  $p(V)$ , è stata derivata analiticamente come funzione di  $\alpha, \lambda$  e dei parametri di  $\rho(V)$ . L'esistenza di una relazione volume-deflusso monotona rende possibile un collegamento formale tra le caratteristiche statistiche di  $V$  e  $q$ . Le dinamiche del volume di eccesso definite dalla (1.81) e le rispettive dinamiche di portata, sono rappresentate in Figura 1:

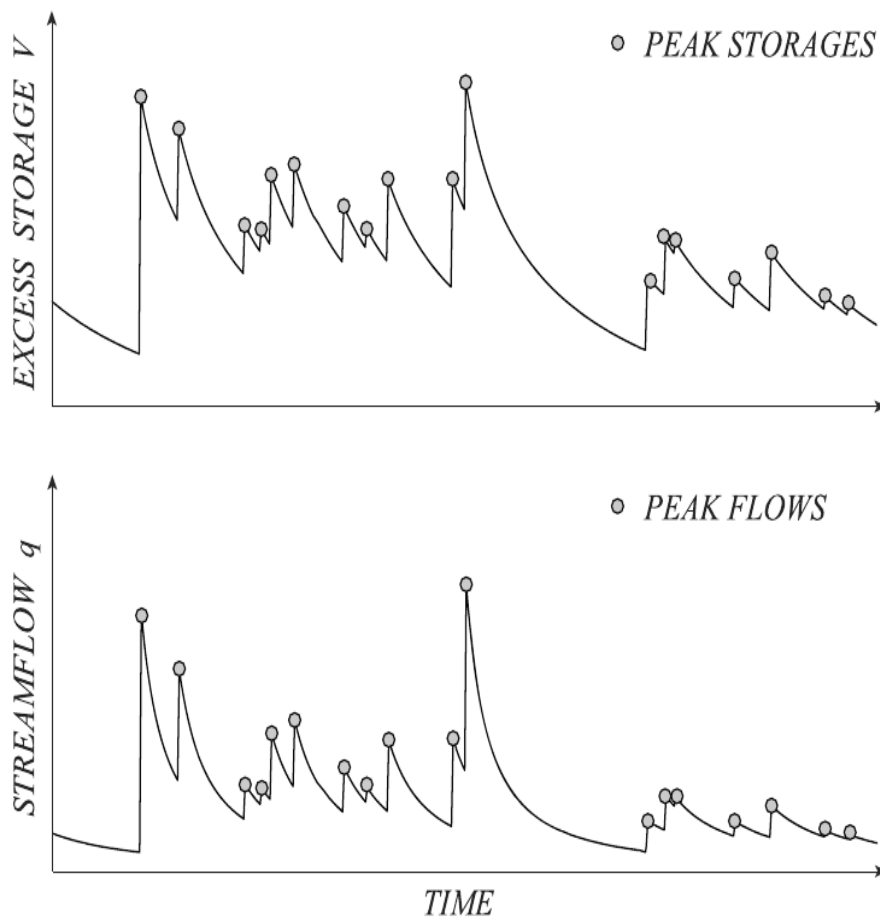


Figura 1: Dinamiche del volume di eccesso ( $V$ ) e della portata ( $q$ ) ottenute dal modello meccanico-stocastico utilizzato (Botter e altri, 2009).

I grafici mettono in rilievo il fatto che i massimi stagionali delle portate giornaliere appartengono al sottoinsieme di portate che si verificano immediatamente al seguito di eventi di precipitazione efficaci, rappresentati dai pallini grigi e d'ora in poi definiti come picchi di portata. La distribuzione di probabilità dei picchi di volume,  $p_j(V)$ , è qui ottenuta come convoluzione tra la funzione densità di probabilità del volume di eccedimento e la funzione densità di probabilità degli incrementi volumetrici, assunta

esponenziale con valore medio pari ad  $\alpha$ . Si assume che il fenomeno dei salti abbia natura Markoviana. Cioè quella di un processo aleatorio in cui la probabilità di transizione che determina il passaggio a uno stato di sistema dipende solo dallo stato del sistema immediatamente precedente e non da come si è giunti a questo stato. La funzione densità di probabilità del volume di eccesso immediatamente antecedente ogni salto è pari alla funzione densità di probabilità del volume di eccesso stessa,  $p(V)$ . mediante la risoluzione della convoluzione, si ottiene la seguente formulazione per la funzione di probabilità dei picchi di volume:

$$p_j(V) = C_1 \exp\left(-\frac{V}{\alpha}\right) \exp\left[\frac{\lambda(2-a)}{(1-a)[K(2-a)]^{\frac{1}{2-a}}}(V-V_0)^{\frac{1-a}{2-a}}\right] \quad (1.82)$$

essendo  $C_1$  una costante di normalizzazione,  $a$  e  $K$  sono esponente e coefficiente della legge di potenza che definisce la recessione idrografica tra due eventi generanti portata consecutivi,  $V_0$  è un valore minimo o massimo di immagazzinamento, a seconda della relazione volume-deflusso adottata. La distribuzione di probabilità dei picchi di portata è pertanto ottenuta dall'applicazione dell'approccio della funzione derivata all'equazione (1.82):

$$p_j(q) = C_2 q^{1-a} \exp\left[\frac{-q^{2-a}}{\alpha K(2-a)} + \frac{\lambda q^{1-a}}{K(1-a)}\right] \quad (1.83)$$

essendo  $C_2$  una costante di normalizzazione. Questa espressione è stata proposta recentemente come “*PHysically based Extreme Value distribution*” (PHEVD, Basso e altri, 2021) L'integrazione dell'equazione (1.83) fornisce la probabilità cumulata di non eccedimento dei picchi di portata:

$$P_j(q) = \int_0^q p_j(q) dq \quad (1.84)$$

Con l'obiettivo di ottenere la distribuzione di probabilità dei massimi stagionali, si assuma un numero costante  $m$  di picchi ordinari indipendenti in un dato periodo di riferimento  $\tau$ , che può essere ad esempio un anno o una stagione. Dal momento che i picchi di portata si sono assunti indipendenti, la probabilità che il massimo valore di  $q$  tra gli  $m$  picchi di portata non venga superato è il prodotto della probabilità di ciascun

picco di portata di avere una magnitudine inferiore di  $q$ , ovvero:

$$P_j[q_1 \leq q] \dots P_j[q_m \leq q] = P_j(q)^m \quad (1.85)$$

Posto che il numero di picchi di portata nell'intervallo temporale considerato sia distribuito secondo una legge di Poisson, con valore medio  $\lambda \tau$ , come assunto per le precipitazioni efficaci, e ricordando che:

$$e^z = \sum_{m=0}^{\infty} \frac{z^m}{m!} \quad (1.86)$$

la seguente espressione per la probabilità cumulata di non eccedimento dei massimi stagionali può essere ottenuta:

$$P_M(q) = \sum_{m=0}^{\infty} \frac{(\lambda \tau)^m e^{-\lambda \tau}}{m!} P_j(q)^m = e^{-\lambda \tau [1 - P_j(q)]} = e^{-\lambda \tau D_j(q)} \quad (1.87)$$

avendo definito le curve di durata come:

$$D_j(q) = 1 - P_j(q) \quad (1.88)$$

La distribuzione densità di probabilità dei massimi si può ottenere come:

$$p_M(q) = \lambda \tau e^{-\lambda \tau D_j(q)} p_j(q) \quad (1.89)$$

Il tempo di ritorno,  $T_R$ , può essere calcolato come l'inverso della probabilità cumulata di accadimento dei massimi di portata. Risulta cioè ottenibile la seguente espressione analitica per la curva di frequenza delle inondazioni:

$$T_R(q) = \frac{1}{1 - P_M(q)} = \frac{1}{1 - \exp[-\lambda \tau D_j(q)]} \quad (1.90)$$

dove  $\lambda \tau$  rappresenta il numero medio degli eventi di precipitazione efficace durante la stagione, l'anno o il periodo di riferimento considerato.

## 1.9 Alcuni recenti studi applicativi

Una applicazione dei metodi statistici e concettuali proposti è stata condotta negli ultimi anni ed è degna di menzione. Marani e Ignaccolo (2014) hanno effettuato

sperimentazioni con metodo Montecarlo sulla serie di osservazioni di precipitazione di Padova, nota per la sua lunga storia, al fine di testare le potenzialità di MEV, utilizzando la distribuzione di Weibull per gli eventi ordinari e confrontando questo approccio con GEV ed il metodo tradizionale di Gumbel. Un importante studio condotto da Zorzetto (2016) estende le analisi su più vasta scala, confrontando MEV e GEV e mettendo ancora una volta in mostra le qualità del metodo in oggetto. È poi la volta dell'applicazione agli eventi di piena e qui si cita il lavoro di Miniussi (2020), che ha applicato MEV ad un vastissimo set di dati negli Stati Uniti. In questo lavoro si sono valutate le prestazioni di MEV utilizzando tre differenti funzioni di distribuzione principale per la descrizione degli eventi ordinari, ossia la distribuzione di Weibull, Pareto e Gamma, e concludendo che con l'adozione di quest'ultima si possono ottenere i migliori risultati. Di altrettanto rilievo è quanto riscontrato da Mushtaq (2020), che ha condotto uno studio di analoga entità in Germania, scegliendo per la descrizione degli eventi ordinari la distribuzione Lognormale e la distribuzione Gamma, adottate nella MEV, riportando risultati in linea con gli studi precedenti. Il modello proposto PHEV ha visto l'applicazione da parte di Basso (2016) su alcuni bacini in territorio nordamericano, evidenziando le potenzialità del metodo meccanico-stocastico e la sua stabilità predittiva in condizione di dati limitati. Analoghi contributi da segnalare sono quelli di Miniussi (2021), che ha esteso le indagini a più vasta scala, tra Stati Uniti e Germania, con lo scopo di valutare la capacità del metodo di identificare i bruschi cambiamenti di andamento nelle curve di probabilità di inondazione.





# 2 METODOLOGIA

## 2.1 Procedura di analisi

Nella presente trattazione si conduce una ampia analisi di sensibilità alla totalità del set di dati disponibile, considerando separatamente le stagioni dell'anno ed utilizzando l'intera serie temporale. In questo modo è possibile procedere ad una valutazione completa dell'efficacia e delle criticità del modello proposto PHEV, della distribuzione MEV, nonché dei metodi offerti dalla letteratura per la valutazione dei parametri. I punti salienti possono sintetizzarsi in:

- analisi delle recessioni;
- stima dei parametri;
- calcolo delle metriche;
- classificazione.

L'analisi delle recessioni consiste nel prendere in esame l'intera sequenza temporale di osservazioni per individuare gli eventi ordinari all'interno del periodo di riferimento. È a partire da questi ultimi che l'approccio utilizzato permette di derivare gli eventi estremi. La stima dei parametri  $\alpha$ ,  $\lambda$ ,  $a$  e  $K$  avviene per analisi dei dati di precipitazione e per adattamento di una legge di potenza ai dati dell'idrogramma, come verrà esposto nel seguito. A questa procedura è affiancato un metodo di stima che avviene per minimizzazione numerica, adattando la funzione densità di probabilità della distribuzione direttamente alle osservazioni di portata. Si procede quindi alla costruzione delle curve di frequenza di inondazione per i massimi stagionali e al calcolo dei valori di portata predetti dal modello associati alle medesime frequenze dei dati osservati (in altre parole agli stessi tempi di ritorno). Così facendo è possibile, mediante l'utilizzo di opportune metriche, valutare le prestazioni del modello e condurre l'analisi di sensitività.

## 2.2 Estrazione e selezione dei picchi

Il processo di estrazione e selezione dei picchi di portata, a partire dagli idrogrammi delle osservazioni, costituisce un passaggio fondamentale per l'adattamento del modello. Lo schema adottato (Dralle e altri, 2017) seleziona un picco di portata se risultano verificate contemporaneamente le seguenti condizioni:

- costituisce un massimo locale;
- il minimo locale compreso tra esso ed il picco precedente viene superato di una quantità superiore a un valore soglia;
- dello stesso valore di soglia deve almeno decadere la portata prima che venga identificato un nuovo massimo locale.

L'algoritmo viene illustrato graficamente in Figura 2:

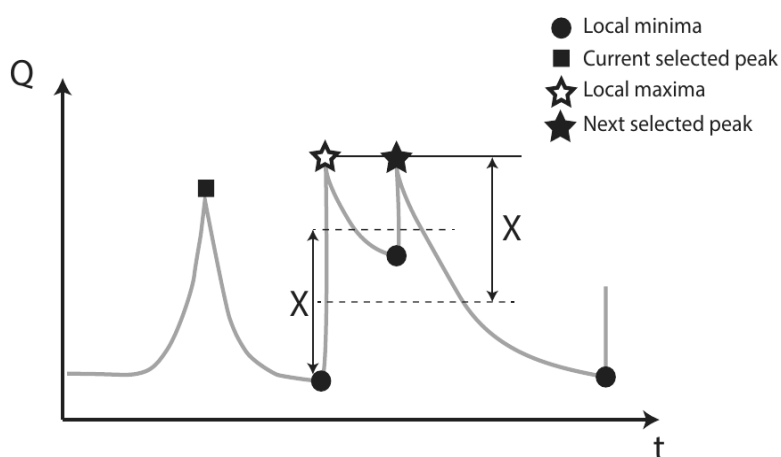


Figura 2: Algoritmo di selezione dei picchi. Il quadrato indica il più recente picco delle recessioni identificato. La stella rappresenta il successivo massimo locale, vuota se non verrà selezionato, piena se verrà selezionato, in base al soddisfacimento di tutte le condizioni del metodo (Dralle e altri, 2017).

Il valore soglia viene definito secondo la semplice formulazione:

$$X = \frac{Q_{max} - Q_{min}}{d} \quad (2.1)$$

essendo  $d$  un parametro che esprime la tolleranza dell'estrazione, generalmente posto pari a 50 per estrazioni altamente selettive, pari a 500 per estrazioni meno selettive. Una volta identificati i picchi di portata, vengono definite le recessioni per selezione di un istante temporale iniziale ed uno finale. Per quanto concerne il primo, che è il picco di portata verificatosi, è possibile valutare l'aggiunta di un certo ritardo e la ragione più comunemente avvalorata a giustificazione di ciò è quella di garantire che sia preponderante la dinamica delle acque sotterranee nel segnale di recessione su quella del deflusso superficiale (Biswall e Marani, 2014). Nel presente studio si decide di non applicare tale ritardo, vista l'incertezza di interpretare le dinamiche alla scala di evento. Il tempo che sancisce la fine della recessione può individuarsi secondo diversi criteri, ed in generale una recessione può essere definita tale se la funzione di portata risulta negativa tra due picchi contigui, ovvero:

$$\frac{dQ}{dt} < 0 \quad (2.2)$$

Un criterio più stringente, adottato in letteratura, aggiunge alla condizione espressa dalla (2.2) anche la richiesta di concavità positiva, dunque:

$$\frac{d^2Q}{dt^2} > 0 \quad (2.3)$$

Questo secondo criterio può talvolta risultare eccessivamente stringente, dal momento che anche minime oscillazioni di portata tra passi temporali verrebbero a determinarne un non soddisfacimento. Alcuni aggiustamenti sono stati suggeriti (Dralle e altri, 2017), come ad esempio l'applicare tale condizione a valori mediati su finestra temporale mobile di dimensione opportuna. Nello studio in esame il modello è stato implementato in accordo con i criteri espressi dalla (2.2) e dalla (2.3), adoperando una finestra temporale mobile di 3 passi, quindi 3 giorni, per la mediazione della concavità. Oltre a ciò va definita una lunghezza temporale minima complessiva della recessione. Questo è giustificato dal fatto che si vogliono eliminare fenomeni eccessivamente brevi e possibilmente dominati da rumore ed incertezze di osservazioni. Nel modello implementato si sceglie una durata minima di 5 giorni, ridotta a 3 quando il numero di recessioni risulta insufficiente.

## 2.3 Modello di stima dei parametri

I quattro parametri della distribuzione PHEV sono facilmente stimabili alla scala di bacino, dipendendo essenzialmente dalle osservazioni di precipitazione e portata.

### 2.3.1 Stima dei parametri legati alla precipitazione

I parametri legati alla precipitazione sono valutati in accordo con le indicazioni di Botter (2007). Il parametro  $\alpha$  viene calcolato come l'altezza di precipitazione giornaliera media durante gli eventi meteorici:

$$\alpha = \frac{\sum_{i=1}^w h_i}{W} \quad (2.4)$$

dove  $w$  è il numero di giorni in cui si verifica un evento di precipitazione, dunque un sottoinsieme del numero totale di giorni, privato dei giorni asciutti. Dimensionalmente è una lunghezza e viene qui espresso in *cm*. Il parametro  $\lambda$  è invece definito come la frequenza degli eventi meteorici generanti portate e viene calcolato come il rapporto tra la portata media di lungo termine e  $\alpha$ :

$$\lambda = \frac{\bar{q}}{\alpha} \quad (2.5)$$

Dimensionalmente è l'inverso di un tempo, cioè una frequenza, che è in questo caso espresso in giorni, dunque  $d^{-1}$ .

### 2.3.2 Stima dei parametri legati alla portata

Per quanto concerne i parametri  $a$  e  $K$ , che compaiono nelle formulazioni di PHEV, questi discendono dalla legge di potenza della relazione volume-deflusso, la cui formulazione tipica si scrive come (Boussinesq, 1877,; Hall, 1968; Tallaksen, 1995):

$$\frac{dq}{dt} = -K q^a \quad (2.6)$$

Rappresentano, dunque, rispettivamente, esponente e coefficiente, e la loro stima avviene per analisi degli idrogrammi, secondo l'approccio proposto da Biswal e Marani (2010). In particolare è possibile seguire uno schema semplificato, che prevede il

passaggio ai logaritmi e l'adattamento della funzione:

$$\log\left(-\frac{dq}{dt}\right) = a \log(Kq) \quad (2.7)$$

o un approccio che prevede l'adattamento diretto della soluzione analitica non-lineare della funzione (2.6):

$$q = [(a-1)(c_1 - kt)]^{\frac{1}{1-a}} \quad (2.8)$$

Utilizzare un adattamento di tipo non-lineare, sebbene in via teorica potrebbe sembrare la scelta più diretta, non sempre comporta risultati migliori ed affidabili, dal momento che, come si può intuire, la sua implementazione può risentire di maggiori problemi di convergenza in fase di risoluzione, per via numerica. Lo studio svolto da Dralle e altri (2017) entra nello specifico della questione e confronta estensivamente le peculiarità dei differenti approcci, mostrando come effettivamente si possano ottenere migliori risultati con l'adattamento della (2.8). È anche bene sottolineare che la tipologia di adattamento scelto è da coniugare con le dimensioni ed il regime idrologico del bacino in esame: Bacini di grandi dimensioni, caratterizzati da eventi lunghi ed inerziali presentano delle recessioni che ben si prestano ad un adattamento lineare, mentre bacini di piccole dimensioni, dai deflussi brevi ed impulsivi, necessitano di un altro tipo di valutazione. Nel modello implementato vengono valutati e confrontati entrambi gli approcci qui esposti e si farà poi uso del secondo, che ha presentato i migliori risultati.

## 2.4 Modello di calibrazione numerica

Nel modello implementato sono prese in esame tutte le possibili combinazioni di stima e calibrazione dei parametri che compaiono nella formulazione di PHEV. Un parametro è definito come stimato quando il valore è assegnato con metodo diretto, ovvero in accordo con le procedure proposte nel Capitolo 2.3. È invece da considerarsi come calibrato quando il valore è ottenuto per adattamento, ovvero con metodo numerico di ottimizzazione della distribuzione di probabilità ai dati di portata, secondo quanto qui verrà esposto. L'adattamento di uno o più parametri della distribuzione PHEV è ottenuto per ottimizzazione numerica, sfruttando il metodo della massima verosimiglianza.

### 2.4.1 Nomenclatura adottata

Per poter definire univocamente tutti i casi, si decide di optare come segue: Si invoca la teoria del calcolo combinatorio, in particolare delle possibili disposizioni con ripetizione di  $n$  oggetti in classe  $k$ :

$$D_{n,k}^r = n^k \quad (2.9)$$

ovvero le due possibili valutazioni dei parametri assegnate alle relative posizioni, per un totale di 16 set distinti. Per identificare univocamente tali set si decide di utilizzare una logica binaria posizionale, che per altro è coerente con gli operatori booleani sottesi al codice di calcolo (0 per FALSO e 1 per VERO), definendo la nomenclatura così come segue in Figura 3:

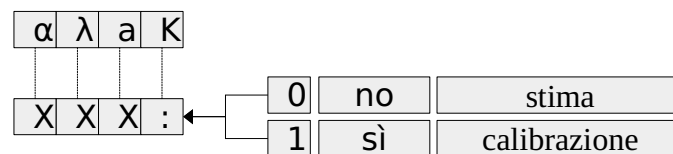


Figura 3: Convenzione di codifica dei set di parametri.

In questo modo, ad esempio, il set di parametri ricavato per stima totale con procedimenti diretti è definito 0000, mentre quello ottenuto per adattamento con metodo di minimizzazione è definito 1111. Tutte le altre combinazioni, che costituiscono casi intermedi, sono facilmente derivabili di conseguenza.

### 2.4.2 La funzione negative-log-likelihood

Il metodo della massima verosimiglianza, in lingua inglese “*Maximum Likelihood Estimation*” (MLE), è un procedimento matematico per determinare uno stimatore. Caso particolare della più ampia classe di metodi di stima basata sugli stimatori d'estremo, il metodo consiste nel massimizzare la funzione di verosimiglianza, definita in base alla probabilità di osservare una data realizzazione campionaria, condizionatamente ai valori assunti dai parametri statistici oggetto di stima. Data una distribuzione di probabilità  $D$ , con funzione di massa (o densità, se continua) di probabilità  $L_D$  caratterizzata da un parametro  $\theta$ , dato un campione di dati osservati  $\{x_i\}_{i=1}^n$  di dimensione  $n$  si può calcolare la probabilità associata ai dati osservati:

$$P(\{x_i\}_{i=1}^n | \theta) = L_D(\theta | \{x_i\}_{i=1}^n) = \prod_{i=1}^n L_{D,i}(\theta | x_i) \quad (2.10)$$

D'altra parte può darsi che il parametro  $\theta$  sia ignoto, sebbene sia noto che il campione è estratto dalla distribuzione  $D$ . Una possibilità per stimare  $\theta$  è allora utilizzare i dati a disposizione. Il metodo della massima verosimiglianza ricerca il valore più verosimile di  $\theta$ , ossia ricerca all'interno dello spazio  $\Theta$  di tutti i possibili valori di  $\theta$ , il valore del parametro che massimizza la probabilità di aver ottenuto il campione dato. Da un punto di vista matematico,  $L_D(\theta | \{x_i\}_{i=1}^n)$  è detta funzione di verosimiglianza e lo stimatore di massima verosimiglianza è ottenuto come:

$$\vec{\theta} = \underset{\theta \in \Theta}{\operatorname{argmax}} L_D(\theta | \{x_i\}_{i=1}^n) \quad (2.11)$$

La funzione di verosimiglianza, sfruttando le proprietà dei logaritmi, può facilmente essere riformulata come segue:

$$l(\theta | \{x_i\}_{i=1}^n) = \log L_D(\theta | \{x_i\}_{i=1}^n) = \sum_{i=1}^n \log L_{D,i}(\theta | x_i) \quad (2.12)$$

Da cui ne deriva lo stimatore di massima verosimiglianza logaritmica:

$$\vec{\theta} = \underset{\theta \in \Theta}{\operatorname{argmax}} \sum_{i=1}^n \log L_{D,i}(\theta | x_i) \quad (2.13)$$

È possibile pervenire ad una formulazione particolarmente utile quando la popolazione dei dati è ragionevolmente riconducibile ad una Gaussiana, ovvero considerando, in luogo del massimo, il minimo della funzione opposta, ottenendo la funzione di verosimiglianza logaritmica negativa, che presenta lo stimatore:

$$\vec{\theta} = \underset{\theta \in \Theta}{\operatorname{argmin}} \left( - \sum_{i=1}^n \log L_{D,i}(\theta | x_i) \right) \quad (2.14)$$

L'approccio logaritmico, trasformando una produttoria in una sommatoria, implica notevoli benefici in termini numerici. La funzione di verosimiglianza diviene:

$$l_D(\theta | \{x_i\}_{i=1}^n) = - \sum_{i=1}^n \log p_i(q_i | \theta) \quad (2.15)$$

essendo dunque  $p_i$  la funzione densità di probabilità adottata degli eventi adottata.

### 2.4.3 Procedura di ottimizzazione

La minimizzazione implementata nel codice del modello ha in ingresso la funzione densità di probabilità della distribuzione adottata, ovvero una formulazione della PHEV (1.83), riferita agli eventi ordinari, o (1.89), riferita ai massimi annuali. Nella presente tesi si è utilizzata la prima formulazione. Gli elementi costituenti la procedura di ottimizzazione sono:

- $\vec{q}$  è il vettore dei picchi di portata selezionati secondo i criteri precedentemente definiti al Capitolo 2.2;
- $\vec{\Theta}$  è il vettore dei parametri della distribuzione, dove ciascuna componente può essere costituita da un valore nullo, di primo tentativo o fissato, a seconda del passo di iterazione e della richiesta specifica di calibrazione;
- $\vec{f}$  è il vettore che contrassegna per ciascuna posizione la calibrazione, mediante un valore booleano. Se la calibrazione è richiesta, il parametro posizionale è inserito tra le variabili del problema di ottimizzazione, altrimenti esso è considerato come una costante e non è oggetto di risoluzione;
- $\vec{e}$  è il vettore dei parametri della distribuzione stimati in accordo con i metodi esposti nel Capitolo 2.3, eventualmente utilizzati come valori di primo tentativo o fissati, a seconda della richiesta specifica di calibrazione;
- $D$  è il dizionario che contiene valori di riferimento, estremi di dominio ed altre caratteristiche funzionali ai criteri di scelta del modello, in fase di verifica dei risultati, di modifica dei parametri, di valutazione del dominio esplorato.

I parametri  $\Theta$  della distribuzione sono singolarmente verificati all'inizio della routine principale, affinché non siano incompatibili analiticamente, in accordo con la funzione matematica della distribuzione di probabilità, e fisicamente, ovvero congruenti ad un principio di realtà materiale. Questo secondo aspetto è perseguito tramite l'impiego di un dizionario  $D$  opportunamente costruito sulla base di risultanze di studi precedenti (Botter, 2016; Miniussi, 2020; Basso, 2021) e in ogni caso sempre fondato su criteri di ragionevolezza e significato fisico dei parametri. Il vettore  $\vec{f}$  ha le stesse dimensioni del vettore  $\vec{\Theta}$  e definisce posizionalmente uno stato logico booleano per la procedura di



ottimizzazione, cioè se il parametro associato a tale posizione è da considerarsi variabile del problema (VERO), quindi oggetto di modificazione, o costante (FALSO), non alterabile, nella fase di minimizzazione. Sulla base di quando così indicato, l'algoritmo costruisce la funzione di densità di probabilità specifica per il caso in esame, elaborata iterativamente dal risolutore numerico. Le variabili adattate richiedono dei valori iniziali, che possono essere forniti dall'iterazione precedente, in fasi successive, o in generale da valori stimati o arbitrariamente forniti. Il modello utilizza inizialmente in ingresso, se specificati, che possono essere quelli provenienti dalle procedure di stima diretta, e ricorre in secondo luogo a valori di riferimento, quando disponibili. Quando questi tipi di informazione falliscono viene utilizzato il dizionario  $D$ . L'algoritmo esegue un'esplorazione del dominio dei parametri, necessaria sia alla risoluzione di eventuali di condizioni localizzate di stallo, sia ad una effettiva e corretta valutazione dei minimi assoluti. Per questa ragione, sono forzatamente imposte diverse esplorazioni per alterazione ripetuta delle condizioni iniziali del risolutore. Per quanto concerne il risolutore, la libreria utilizzata mette a disposizione numerose procedure di letteratura, collaudate e storicamente utilizzate nel campo dei problemi numerici complessi. Nella fattispecie ci si avvale di più schemi risolutivi, sia per superare le eventuali difficoltà che, in determinate condizioni, possono affliggere un approccio particolare, sia per assicurare, ancora, una soluzione stabile. In altre parole, l'implementazione posta in essere vuole garantire il più possibile una validità generale. I risolutori numerici utilizzati sono:

- *Nelder Mead* è un metodo che fa uso delle derivate e si basa sul concetto di semplice, un particolare tipo di politopo con  $n+1$  vertici in uno spazio di dominio a  $n$  dimensioni;
- *SLSQP* è un metodo iterativo di ottimizzazione ampiamente utilizzato in problemi matematici per i quali la funzione obiettivo ed i vincoli sono due volte continuamente differenziabili;
- *COBYLA* è un metodo che sfrutta un criterio di miglioramento del problema approssimativo utilizzato per la successiva iterazione dell'algoritmo, affinando la ricerca sulla dimensione del passo.

Ai parametri ottimizzati vengono poste delle condizioni di vincolo. La ragione fondamentale è evitare di ottenere dei risultati non sensati analiticamente o fisicamente, evitando adattamenti del modello privi di significato fisico. In secondo luogo vi è una ragione estremamente funzionale: ridurre il dominio esplorato, ottimizzando e facilitando la convergenza. Si utilizzano le limitazioni per i parametri riportate in Tabella 1.

parametro	minimo	massimo
$\alpha$ (cm)	0	10
$\lambda$ (d <sup>-1</sup> )	0	1
a (/)	1	50
K (cm <sup>1-a</sup> /d <sup>2-a</sup> )	0	100

Tabella 1: Limitazioni dei parametri della distribuzione.

In fase di ottimizzazione l'algoritmo può incontrare difficoltà puntuali tali da determinare il fallimento del metodo, ad esempio per mancata convergenza. In queste circostanze la procedura applicata prevede alcune strategie particolari:

- modifica dei valori iniziali;
- aumento del numero delle iterazioni massime;
- diminuzione della precisione richiesta.

L'applicazione congiunta delle tre operazioni appena elencate garantisce il funzionamento del metodo di ottimizzazione per la maggior parte dei casi, a parità di risolutore adoperato. Qualora, dopo un prefissato numero di tentativi, non sia possibile uscire dalla condizione di stallo, si stabilisce il fallimento dell'esplorazione di questo e si procede con il risolutore successivo. Il criterio finale di valutazione confronta i valori ottenuti nella funzione di verosimiglianza associata a ciascun procedimento di minimizzazione effettuato e seleziona il set di parametri relativo a quello migliore, ovvero:

$$l_D^* = \min[l_D(\theta|\{x_i\}_{i=1}^n)]_k$$

avendo indicato con  $k$  le distinte ottimizzazioni completate. In Figura 4 si schematizza il modello appena illustrato.

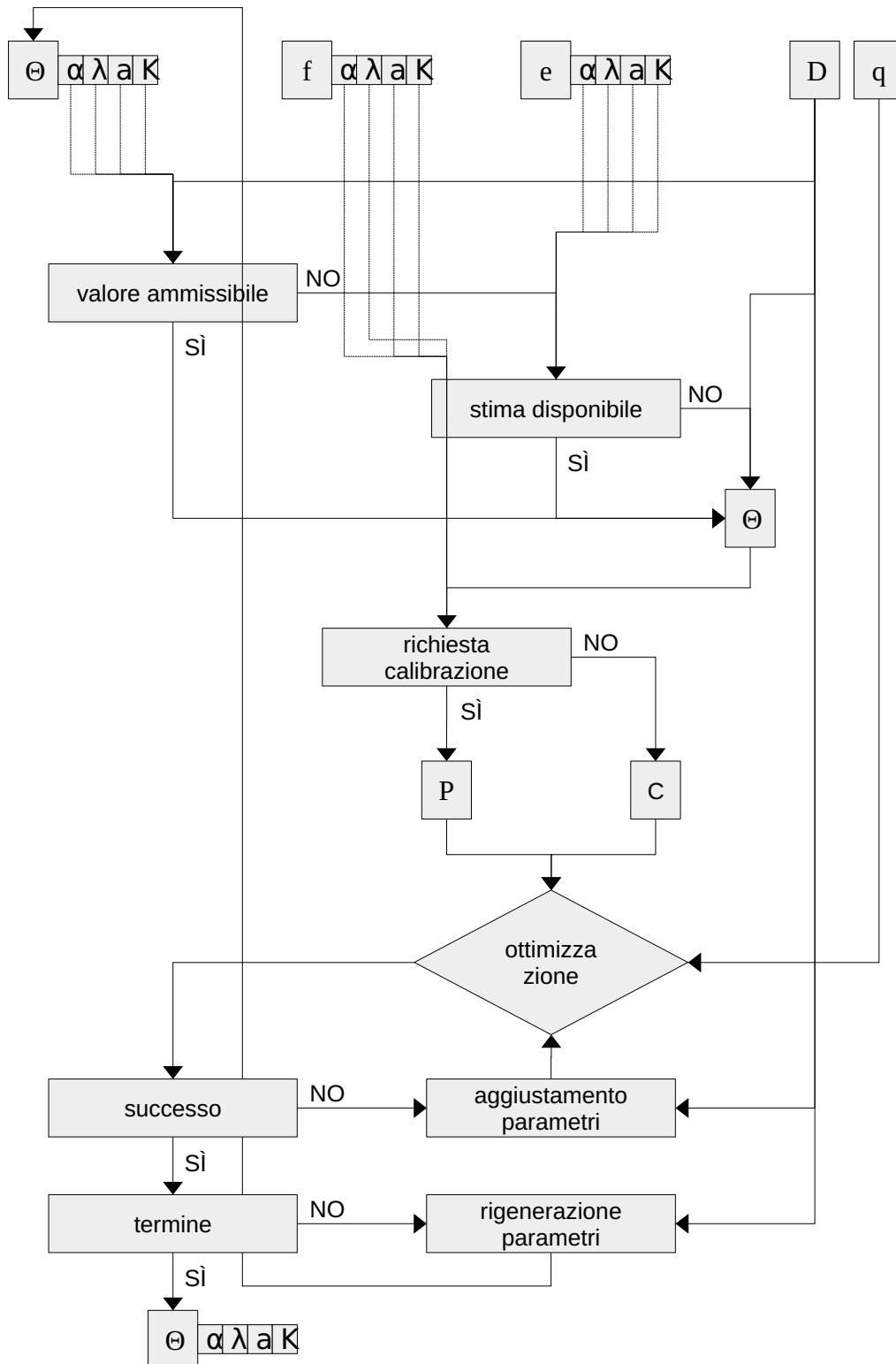


Figura 4: Schema logico dell'algoritmo di ottimizzazione.

## 2.5 Calcolo dei quantili

La forma del legame tra valore di portata e probabilità di non superamento, o alternativamente superamento, è del tipo:

$$F=1-D=f(q) \quad (2.16)$$

dove la relazione che lega le due grandezze non presenta una forma lineare esplicita. Per la costruzione delle curve di frequenza delle inondazioni, è sufficiente utilizzare una serie opportunamente densa di valori di portata per calcolare le corrispettive probabilità, dunque eventualmente i tempi di ritorno. È tuttavia di interesse la soluzione della (2.16) per la determinazione dei valori di portata associati ad una specifica probabilità, in modo da poter confrontare i dati osservati, di cui sono calcolate le probabilità di accadimento, ed applicare le metriche desiderate. Questo è perseguibile numericamente, in modo banale per punti legati dall'applicazione diretta della (2.16), ovvero probabilità associate a portate arbitrariamente e densamente generate, e successiva selezione delle migliori corrispondenze ai dati osservati. Nel modello del presente lavoro si è invece optato per un processo risolutivo a quantili. A livello elaborativo, si cerca la soluzione per  $q$  della (1.71), a partire da un certo valore di  $F$  o  $D$ , qui riscritta come:

$$F(q)=\int_0^q p(q) dq=1-\int_q^\infty p(q) dq=1-D(q) \quad (2.17)$$

Essendo d'altra parte poi ottenuto il tempo di ritorno:

$$T_R=\frac{1}{1-F(q)}=\frac{1}{D(q)} \quad (2.18)$$

La soluzione inversa della (2.17) avviene numericamente e richiede alcuni accorgimenti finalizzati a garantirne l'applicazione. Definito nel dominio un numero  $n$  congruo di quantili di probabilità  $P$  secondo il criterio matematico:

$$\{P \in \mathbb{R} : 0 \leq P \leq 1\} \quad (2.19)$$

si procede alla soluzione numerica dell'equazione non-lineare per passi all'indietro, ovvero a partire dal quantile cui è associata la probabilità più elevata (di non superamento), che trova la corrispondenza con valori di portata più grandi. La ragione alla base di ciò è essenzialmente numerica. In generale si utilizza un passo pari a 0.005

lungo l'asse delle probabilità, ma questa dimensione è variabile in funzione del grado di dettaglio che si vuole perseguire per la costruzione della curva di probabilità di inondazione. Minore il valore, maggiore la risoluzione e, in generale, anche la stabilità numerica, ma anche il tempo richiesto dall'elaborazione. È dunque opportuno valutare opportunamente la scelta e bilanciarla in base alle necessità del caso e, ad esempio, è possibile ottimizzare al meglio il processo concentrando il dettaglio dove effettivamente necessario, sulla base sia della deformazione della restituzione grafica adottata, che dell'ordine di grandezza dei valori, inferiore nella parte sinistra del dominio. In fase risolutiva, come valore di primo tentativo si utilizza la portata media del bacino, ossia  $\hat{q}$ , che viene incrementato progressivamente se viene a manifestarsi una qualche difficoltà puntuale. Nei passi successivi è poi utilizzata come soluzione di primo tentativo quella individuata in precedenza, eventualmente ridotta, nei casi di mancata convergenza. Al vettore  $\vec{P}$  dei quantili fissati, in questo modo utilizzati per la costruzione delle curve di probabilità di inondazione sono aggiunti, in accordo con il valore, i quantili di probabilità dei massimi periodici osservati  $\vec{P}_{oss}$ , ottenendo così il dominio dell'algorithm di calcolo. Lo schema appena esposto è concettualmente rappresentato in Figura 5.

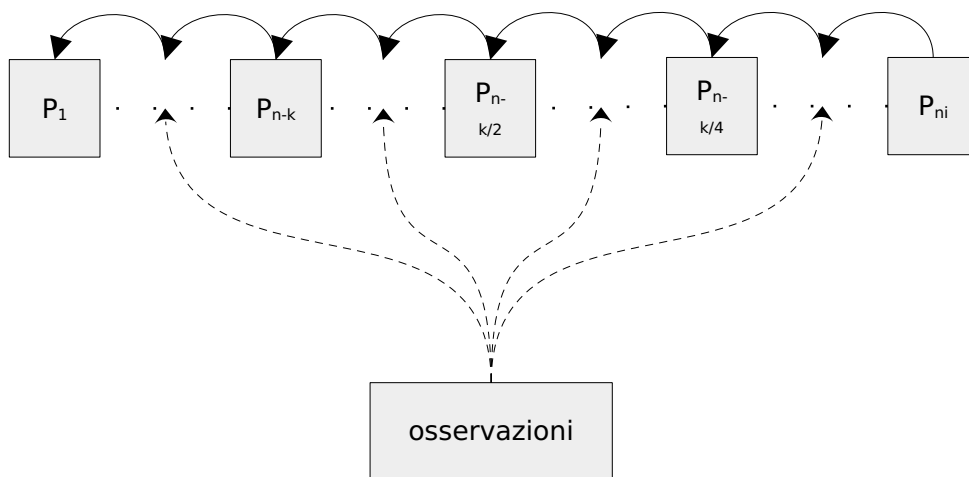


Figura 5: Schema di calcolo dei quantili.

Seguendo questa procedura è possibile ottenere le soluzioni puntuali necessarie alle analisi delle metriche in maniera efficiente e stabile.

## 2.6 Definizione delle metriche

Nel seguito si definiscono ed espongono i criteri metrici quantitativi utilizzati per fasi e finalità differenti nel corso delle analisi.

### 2.6.1 Errore relativo

Il criterio metrico principale utilizzato per la quantificazione della qualità previsionale del modello è l'errore relativo, definito come:

$$\epsilon = \frac{y - \hat{y}}{y} \quad (2.20)$$

A comprovare la validità di questo indicatore, di semplice ed esplicita efficacia, è il vasto uso che trova spazio in letteratura. Misura, in un qualsiasi modello che contempli una serie di dati reali ed una funzione di predizione, il rapporto tra l'errore di quest'ultima ed il valore reale. In questo modo l'errore è reso relativo, ovvero pesato rispetto alla magnitudine del dato reale, nonché adimensionalizzato, quindi opportunamente confrontabile. Risulta che:

- valore nullo corrisponde ad un modello predittivo perfetto, perché nulla è la differenza tra valore reale e calcolato;
- valori positivi indicano una sovrastima da parte della funzione di previsione, e non sono limitati superiormente;
- valori negativi indicano una sottostima da parte della funzione di previsione, e sono limitati inferiormente, dal momento che, nel caso specifico, non sono fisicamente possibile portate negative, ma al più possono essere nulle.

### 2.6.2 Scarto quadratico medio

Un'importante metrica utilizzata è l'errore quadratico medio, che si esprime secondo la classica formulazione di letteratura:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (2.21)$$

e utilizzato nella presente trattazione per lo più con lo scopo di esprimere una qualità

globale di stima, necessaria nelle procedure numeriche ad istruire l'algoritmo nelle scelte intraprese, ad esempio in fase di calibrazione dei parametri.

### 2.6.3 Coefficiente di determinazione

In statistica, il coefficiente di determinazione, più comunemente  $R^2$ , è una statistica utilizzata nel contesto di modelli il cui scopo principale è la previsione di risultati futuri o la verifica di ipotesi, sulla base di altre informazioni correlate. Fornisce una misura di come i risultati ben osservati vengono replicati dal modello, in base alla frazione della varianza non spiegata. La definizione più generica è la seguente:

$$R^2 = 1 - \frac{RSS}{TSS} \quad (2.22)$$

avendo definito la devianza residua:

$$RSS = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (2.23)$$

e la devianza totale:

$$TSS = \sum_{i=1}^n (y_i - \bar{y}_i)^2 \quad (2.24)$$

Se  $R^2$  è prossimo al valore unitario, significa che i regressori predicono bene il valore della variabile dipendente in campione. Il coefficiente di determinazione viene usato in sede di analisi delle recessioni per valutare l'adattamento della legge di potenza.

### 2.6.4 Indice di classificazione

Per la classificazione ed il campionamento qualitativo dei risultati ottenuti, è stato necessario avvalersi di un criterio che permettesse una misura efficace dell'efficienza del modello implementato, come ad esempio, la definizione di classi di insiemi contrassegnate da un comportamento simile, ovvero una buona o cattiva stima. Si definisce una metrica di punteggio come segue:

$$S = \frac{\sum_{i=1}^n \delta_i}{n} \quad (2.25)$$

dove  $\delta$  rappresenta la generica metrica utilizzata per la valutazione della qualità predittiva sul valore  $i$  considerato. Evidentemente l'assegnazione di un punteggio complessivo può seguire diversi approcci, a seconda dell'oggetto di valutazione. Nel caso in cui, ad esempio nelle procedure di adattamento di curve a dati reali, si voglia misurare specificatamente la congruenza assoluta tra dati e funzione descrittiva, è da considerarsi un valore assoluto per  $\delta$ , nonché eventualmente regolarizzato in accordo con un criterio di specificità della magnitudine degli errori, ovvero una analisi che possa essere quantitativa e qualitativa. Nella fattispecie, si definisce un errore normalizzato:

$$\epsilon_n = \log \left( 1 + \frac{\hat{y} - y}{y} \right) \quad (2.26)$$

La relazione così definita è una riformulazione dell'errore relativo che permette la costruzione di una metrica bilanciata tra errori positivi, di sovrastima, e negativi, di sottostima. Adottando la (2.26), presa in modulo, nella (2.25), si perviene al metodo di classificazione utilizzato per il campionamento.

## 2.7 Strumenti di calcolo

Ci si avvale del linguaggio di programmazione Python per la realizzazione del codice principale del modello e tutte le sue funzioni. Alcune librerie specifiche sono utilizzate:

- *scipy* è una libreria di algoritmi e strumenti matematici che contiene moduli per l'ottimizzazione, per l'algebra lineare, l'integrazione, funzioni speciali, FFT, elaborazione di segnali ed immagini, risolutori ODE e altri strumenti comuni nelle scienze e nell'ingegneria;
- *statsmodels* è una libreria che fornisce classi e funzioni per la stima di molti modelli statistici diversi, nonché per condurre test ed esplorazione di dati;
- *pandas* è una libreria per la manipolazione e l'analisi dei dati, offrendo in particolare le strutture e le funzioni per tabelle numeriche e serie temporali;
- *seaborn* è una libreria di visualizzazione dati, che fornisce un'interfaccia di alto livello per l'elaborazione di grafici statistici complessi ed informativi.

In ragione della potenza elaborativa richiesta dalle seguenti analisi, si è utilizzato il



cluster computing, ovvero un sistema in grado di distribuire un algoritmo complesso su un numero elevato di nodi elaborativi, aumentando la potenza di calcolo del sistema grazie al calcolo parallelo implementato opportunamente nel codice del modello. Il cluster di UFZ è gestito dallo Slurm Workload Manager, precedentemente noto come Simple Linux Utility for Resource Management (SLURM), o semplicemente Slurm, ovvero un programma di pianificazione dei lavori gratuito e open source per kernel simili a Linux e Unix, utilizzato da molti dei supercomputer e cluster di computer del mondo. Esso fornisce tre funzioni chiave:

- assegnazione di accesso esclusivo e/o non esclusivo alle risorse (nodi informatici) agli utenti per un certo periodo di tempo in modo che possano svolgere il lavoro;
- fornitura di un sistema comune per avviare, eseguire e monitorare il lavoro, in genere un lavoro parallelo come l'interfaccia di passaggio dei messaggi (MPI) su un insieme di nodi allocati;
- gestire la contesa per le risorse utilizzando una coda di lavori in sospeso.

Slurm utilizza un algoritmo best fit per ottimizzare la località delle assegnazioni delle attività su computer paralleli. È ampiamente diffuso tra i più grandi sistemi di elaborazione parallela a livello internazionale.

## **2.8 Set di dati**

I dati utilizzati nelle analisi sono forniti dall'*Helmholtz Centre for Environmental Research – UFZ*, con sede ad Halle (DE), dove è stata condotta l'intera attività di ricerca, e sono costituiti dalle serie storiche di precipitazione e portata giornaliere registrate per 178 bacini idrografici ubicati in Germania, il cui elenco totale è disponibile in Appendice A. Il gruppo di bacini a disposizione è evidentemente molto consistente. In Figura 6 si mostrano tutti i bacini considerati.

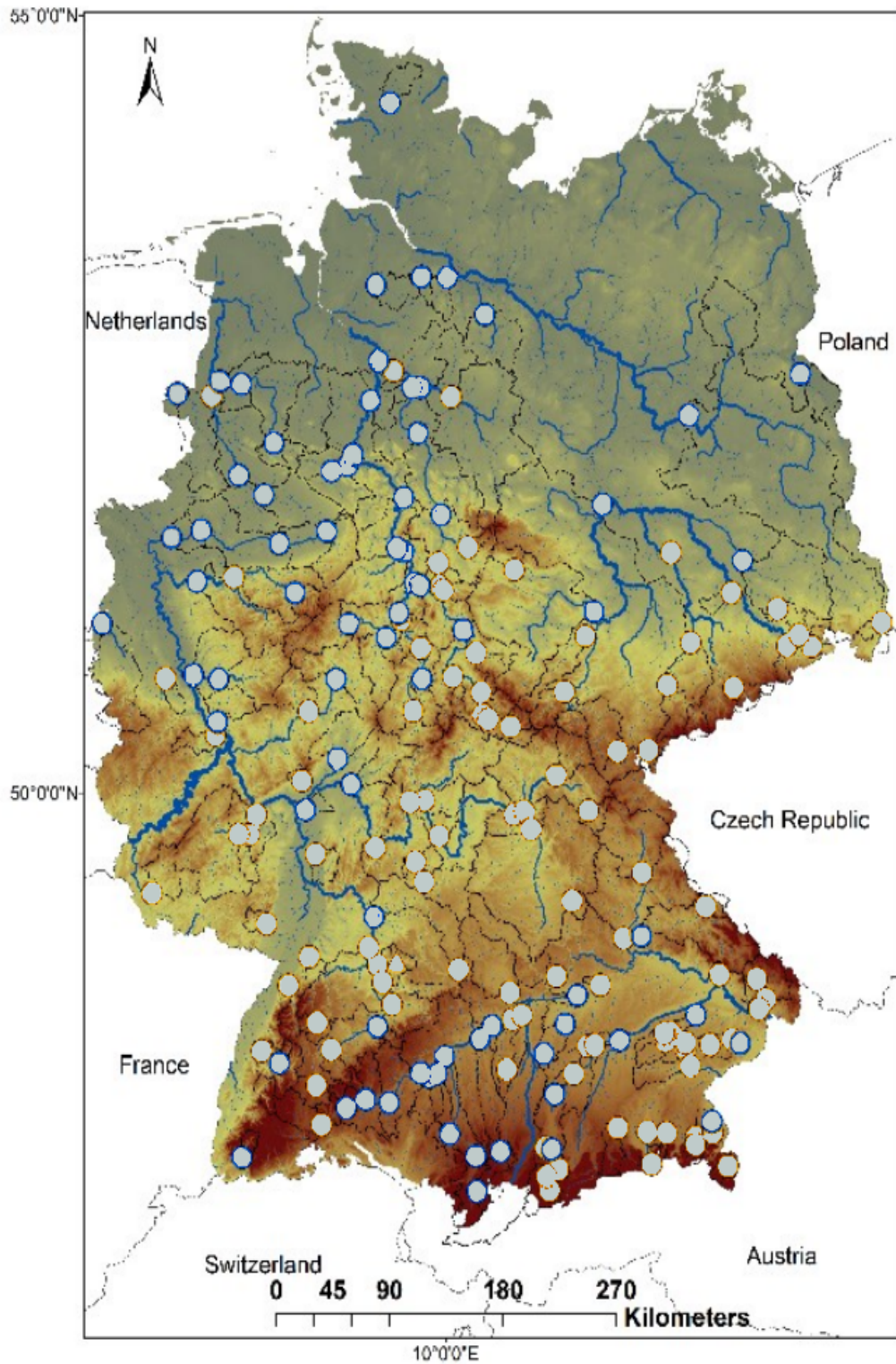


Figura 6: Localizzazione dei 178 bacini oggetto dell'analisi.

Il set di dati registrati riguarda un intervallo temporale vasto, da circa metà del 1900 fino ai giorni d'oggi. In Tabella 2 si riporta la copertura per finestre temporali.

anni	bacini	%
>30	178	100%
>40	175	98%
>50	162	91%
>60	147	82%

*Tabella 2: Grado di copertura delle osservazioni per lunghezze temporali crescenti, espresso mediante numero di bacini e relativa percentuale.*

Tutto il set di dati garantisce una copertura temporale per le osservazioni di precipitazione e portata all'interno dei bacini che è almeno superiore a 30 anni.



# 3 RISULTATI

## 3.1 Analisi globale

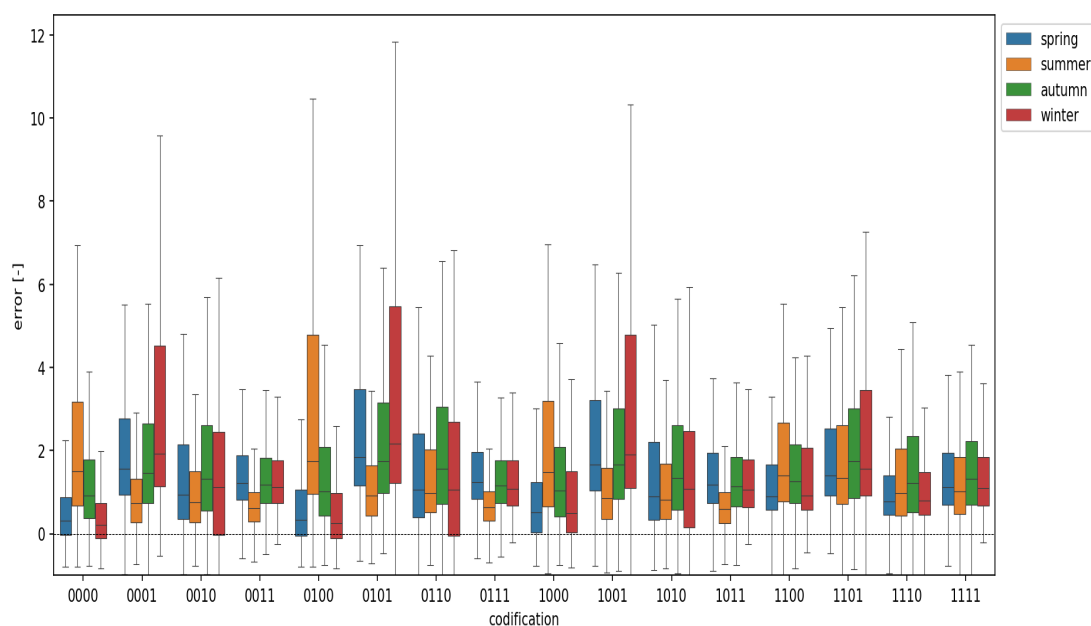
Per tutte le 16 combinazioni studiate si compongono le coppie osservazione-previsione dei colmi di piena, formate da:

- valore di portata massima osservato;
- valore di portata massima predetto dal modello.

I due valori sono caratterizzati dalla stessa probabilità cumulata di non eccedimento. Preso un valore osservato, si calcola la probabilità di non eccedimento ad esso associata e si individua il valore predetto che presenti il medesimo valore di quest'ultima. Ciò è stato possibile grazie allo schema di calcolo esposto nel Capitolo 2.5, che permette la determinazione dei quantili di probabilità di non eccedimento delle portate massime predette dal modello. Si calcola quindi per ogni coppia l'errore relativo, metrica esposta nel Capitolo 2.6, ottenendo quindi una metrica associata ad ogni singolo accadimento. Il numero di valori a cui si perviene, per ogni singolo bacino e per ogni singola stagione è pari al numero di anni di osservazioni.

### 3.1.1 Gli estremi secondo PHEVD

Si è posta preliminarmente l'attenzione sulla distribuzione di probabilità degli eventi massimi di piena generata da modello concettuale PHEV. Gli eventi estremi derivano dai picchi ordinari e, in questo caso, si è per quest'ultimi ricorsi all'ipotesi fondamentale che, nell'intervallo temporale considerato, gli inter-arrivi tra due eccedimenti successivi del valore arbitrario scelto seguano una legge di Poisson. Gli errori relativi sono rappresentati graficamente grazie ai diagrammi della mediana e dei quartili, o "boxplot" in lingua inglese, di Figura 7.



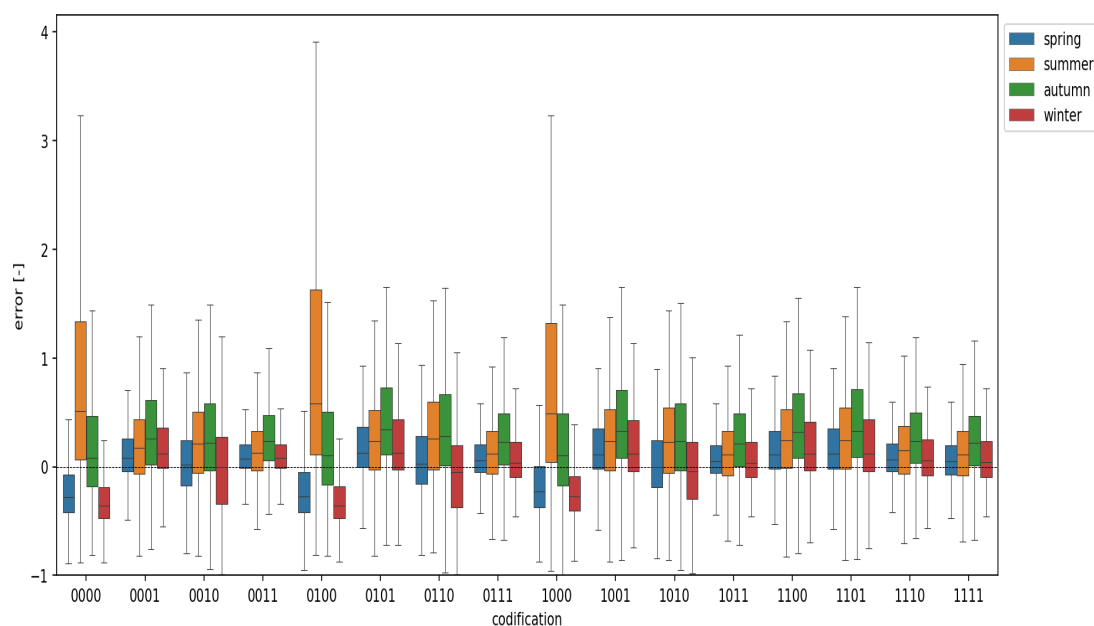
*Figura 7: Diagrammi della mediana e dei quartili (boxplot) degli errori relativi sulle osservazioni dei massimi stagionali per tutte le combinazioni di valutazione dei parametri effettuate, utilizzando la formulazione dei massimi di portata di PHEV.*

Si può anzitutto osservare in maniera chiara che il modello concettuale PHEV, indipendentemente dalla combinazione di parametri e dalla stagionalità, conduce a valori predetti delle portate stagionali che costituiscono una sovrastima rispetto ai valori effettivamente osservati. L'entità di questa sovrastima è significativa. L'errore relativo mediano è compreso tra circa il 50% e il 200%. Il set denominato "0000", che corrisponde a valori dei parametri interamente stimati con i semplici metodi di letteratura, comporta una qualità in termini di risultati previsionali che non è sostanzialmente surclassata dalla calibrazione per uno o più parametri del set, ma si possono ottenere alcuni modesti benefici quando la quantità di parametri adattati è superiore a quella dei parametri stimati, limitatamente ad alcuni casi. La modalità di adattamento del modello secondo il set 0000 mostra le maggiori discrepanze rispetto alle osservazioni nei periodi estivi ed autunnali, nei quali si verificano le sovrastime più rilevanti. Per la stagione invernale, e quella primaverile si ottengono invece risultati prossimi alle osservazioni, con un errore relativo che non supera il 100%. Come si discute più estesamente nel seguito tale circostanza è attribuibile alle caratteristiche

degli idrogrammi di piena, ovvero al numero diverso di recessioni estratte. Il set 0011, ossia quello in cui avviene la calibrazione di  $a$  e  $K$  (esponente e coefficiente della funzione che adatta le recessioni), se confrontato con il set 0111 e 1011, dove agli appena citati si aggiunge la calibrazione di  $\lambda$  o  $\alpha$  (altezza di precipitazione giornaliera e frequenza degli eventi generanti portata), rispettivamente, pur presentando una dispersione dei valori inferiore, mostra dei rapporti tra i grafici delle stagioni che sono equivalenti. In altre parole, la calibrazione di uno di questi ultimi due parametri non è in grado di incidere in modo sostanziale sul risultato finale. Quanto appena affermato può essere avvalorato confrontando il set 0000, ovvero nessun parametro è calibrato sulla distribuzione dei picchi di portata, ai sei 1000 e 0100, che corrispondono ai casi in cui i parametri  $\alpha$  e  $\lambda$ , uno alla volta, sono ottimizzati. I grafici degli errori mostrano dei valori che, sebbene siano generalmente minori e più compattamente distribuiti per il primo, hanno valore medio e proporzione tra le dispersioni stagionali analoghi. Ancora, lo stesso schema di analogia è riscontrabile tra 1010 e 0110 e tra 1001 e 0101, dove cioè, fissato il blocco destro dei parametri  $a$  e  $K$ , si alterna l'adattamento e la stima di  $\alpha$  e  $\lambda$ . La prossima analisi rende ancora più tangibili questi aspetti. Il set 1111, dove tutti i parametri vengono adattati senza utilizzare alcuna procedura di stima diretta, conduce a previsioni dei massimi stagionali di portata con una sovrastima mediana del 100%. Gli errori si presentano inoltre solo moderatamente dipendenti dalla stagione.

### **3.1.2 Gli estremi secondo MEVD**

È possibile ora entrare nel vivo delle analisi, adottando la distribuzione PHEV come distribuzione degli eventi ordinari all'interno del metodo MEV, così da calcolare i massimi stagionali di portata. L'applicazione del modello alle 16 combinazioni di parametri permette di valutare l'errore relativo calcolato su ogni coppia osservazione-previsione, rappresentato graficamente grazie ai diagrammi della mediana e dei quartili di Figura 8.



*Figura 8: Diagrammi della mediana e dei quartili (boxplot) degli errori relativi sulle osservazioni dei massimi stagionali per tutte le combinazioni di valutazione dei parametri effettuate, utilizzando la formulazione dei massimi di portata di MEV.*

Il modello combinato PHEV-MEV, in cui PHEVD è la distribuzione ordinaria adottata in MEVD, permette di ottenere un chiaro e netto miglioramento della qualità di stima dei massimi periodici del modello rispetto al caso in cui si ricorre alla descrizione semplificata degli stessi fornita da PHEVD. L'errore relativo, confrontando la Figura 3.2 con la Figura 3.1, salvo alcune eccezioni di cui si tratterà, è drasticamente ridotto, mostrando valori medi contenuti in un intorno ragionevole del valore nullo e valori di estremo superiore ampiamente al di sotto del valore unitario. La tendenza alla sovrastima delle portate previste rimane visibile, caratteristica quindi della distribuzione principale adottata. Allo stesso modo permangono alcune delle valutazioni fatte in precedenza. Si consideri ora quanto in Figura 9, dove, assieme ai casi limite 0000 e 1111, si concentra l'attenzione alle sole combinazioni dove si è adattato numericamente uno solo dei quattro parametri della distribuzione PHEV alla volta.



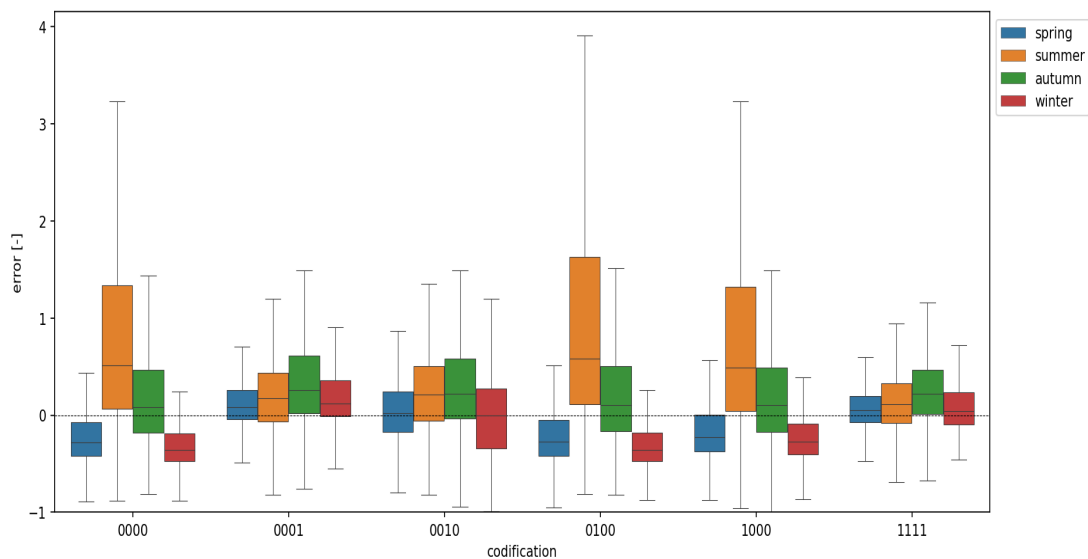


Figura 9: Diagrammi della mediana e dei quartili degli errori relativi sulle osservazioni dei massimi stagionali nelle combinazioni di adattamento di uno dei quattro parametri della distribuzione alla volta, con alle estremità le combinazioni limite 0000 e 1111.

La combinazione del set 0000 presenta, in misura anche maggiore di prima, dei risultati sovrapponibili a quelli dei set 1000 e 0100, a riprova del fatto che i parametri legati alle precipitazioni,  $\alpha$  e  $\lambda$ , se calibrati numericamente, non migliorano le prestazioni del modello. Inoltre in questi tre casi si registrano i maggiori errori di sottostima delle previsioni. Allo stesso modo, nelle combinazioni 0001 e 0010, dove si è adattato solamente un parametro, tra quelli relativi alle recessioni  $a$  e  $K$ , si riscontra un effettivo ed immediato beneficio, tanto che il comportamento del modello si mostra molto simile a quello del caso limite 1111, dove si è adattato l'intero set di parametri della distribuzione indipendentemente dalle concettualizzazioni fisiche poste in essere. È possibile da questo inferire che i parametri  $a$  e  $K$  sono critici per il modello, quindi ne risulta fondamentale una stima il più possibile accurata e basata su osservazioni. La calibrazione di questi contribuisce ad aumentare le prestazioni del modello perché permette di mitigare l'incertezza ad essi associata. Si procede ancora ad analizzare le dinamiche incrementando il numero di parametri determinati per calibrazione e, in Figura 10, assieme ai casi limite 0000 e 1111, si concentra l'attenzione alle sole

combinazioni dove si sono adattati numericamente due dei quattro parametri della distribuzione PHEV alla volta.

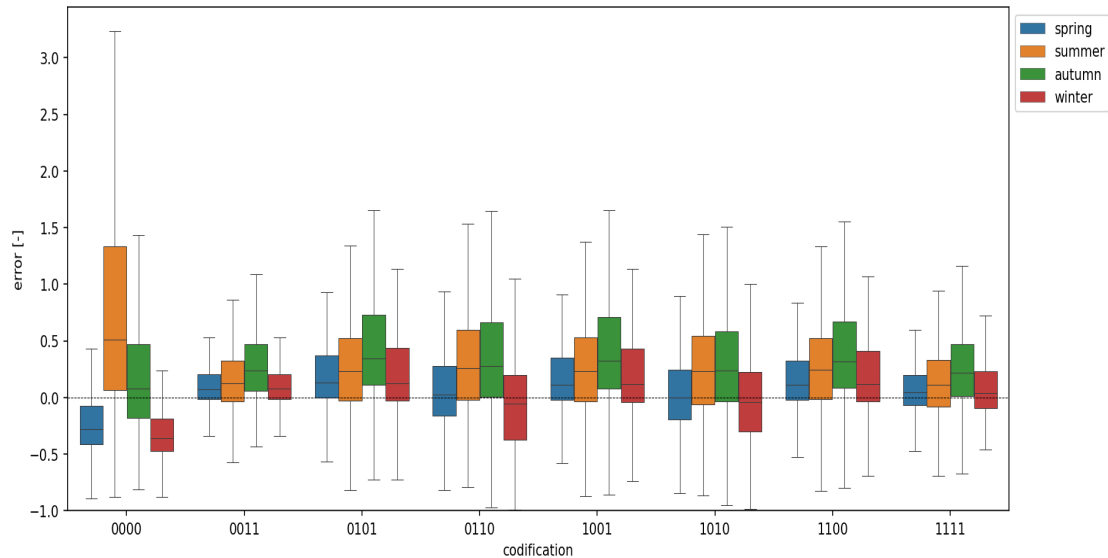


Figura 10: Diagrammi della mediana e dei quartili degli errori relativi sulle osservazioni dei massimi stagionali nelle combinazioni di adattamento di due dei quattro parametri della distribuzione alla volta, con alle estremità le combinazioni limite 0000 e 1111.

L'aumento del numero di gradi di libertà soggetti ad adattamento, esteso a due parametri, mostra qualche beneficio globale, ma l'aspetto di maggior rilievo è quello che si può visualizzare confrontando i risultati per i casi 0011 e 1111. I migliori risultati in assoluto ottenuti nella modellazione corrispondono ad un adattamento dei due parametri legati alle recessioni, utilizzando per i parametri delle precipitazioni i valori che si stimano con i semplici metodi di letteratura. Il set 1100, che invece opera al contrario, cioè adatta numericamente questi due ultimi parametri, viene superato qualitativamente dal caso 1111. Si registra comunque un miglioramento dei risultati in diverse condizioni, dal momento che viene offerto al calibratore un numero di gradi di libertà maggiore per meglio rappresentare la realtà dei dati osservati. Infine, in Figura 11, assieme ai casi limite 0000 e 1111, si analizzano le combinazioni dove si sono adattati numericamente tre dei quattro parametri della distribuzione PHEV.

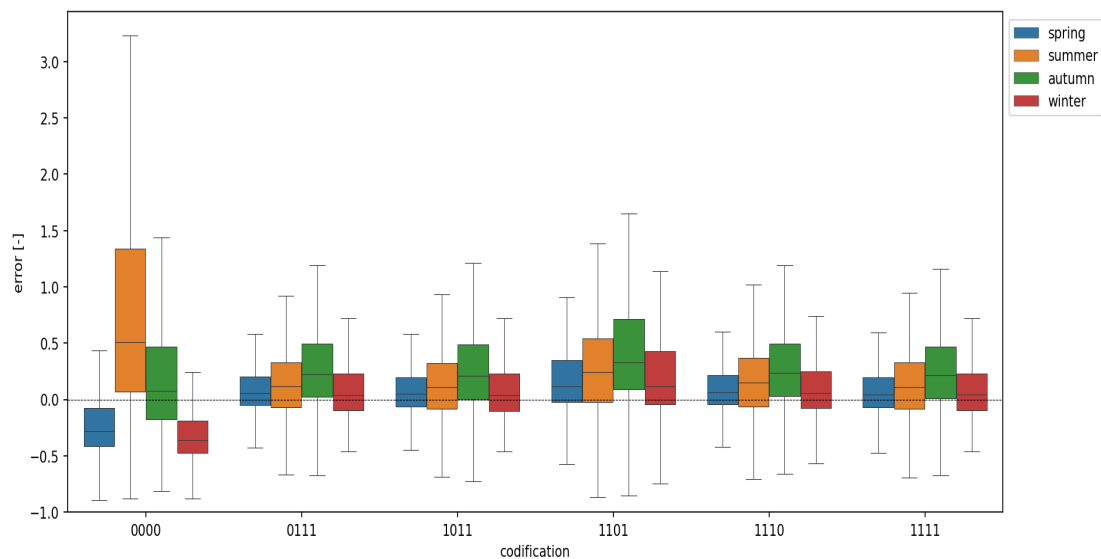


Figura 11: Diagrammi della mediana e dei quartili degli errori relativi sulle osservazioni dei massimi stagionali nelle combinazioni di adattamento di tre dei quattro parametri della distribuzione alla volta, con alle estremità le combinazioni limite 0000 e 1111.

In queste condizioni il modello mostra un comportamento sostanzialmente immutato, se non nel caso 1101, dove si registra per lo più un lieve effetto di amplificazione dei valori di errore relativo. La ragione di questo fenomeno sarà nel seguito meglio analizzata, e qui si anticipa che è legata alla maggiore sensibilità del modello alle modeste variazioni del parametro  $a$ , cioè il ruolo di quest'ultimo nella funzione di distribuzione, dal momento che è responsabile della forma che essa assume. I parametri  $a$  e  $K$ , come anticipato, condizionano fortemente la precisione del modello e si vuole mostrare con maggiore efficacia questo aspetto. Viene fissato l'adattamento numerico per uno di questi due parametri alla volta,  $a$  in Figura 12 e  $K$  in Figura 13, dove sono valutate tutte le possibili combinazioni di stima dei restanti parametri.

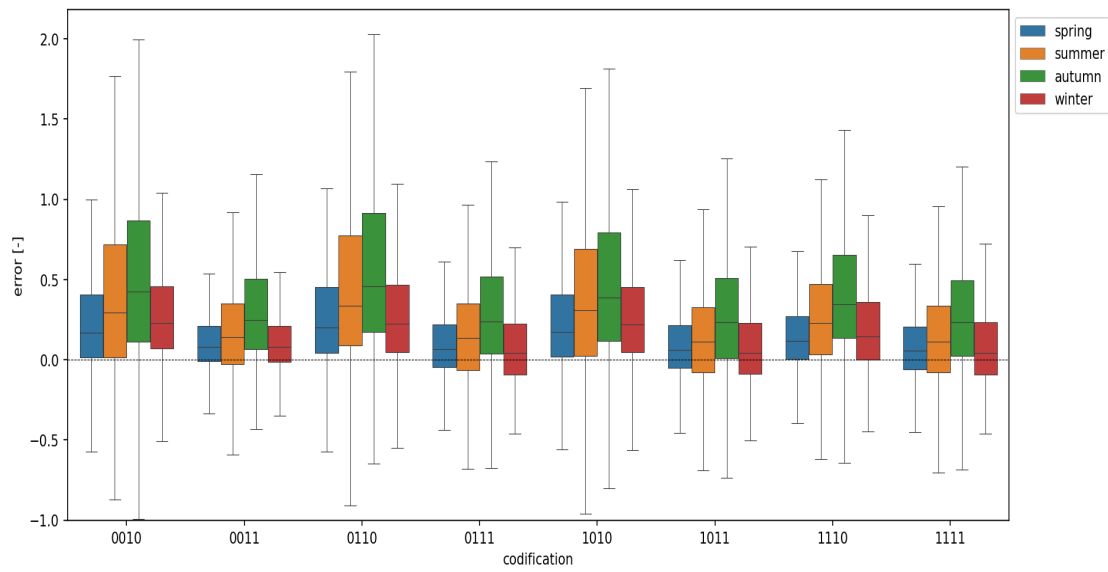


Figura 12: Diagrammi della mediana e dei quartili degli errori relativi sulle osservazioni dei massimi stagionali in tutte le combinazioni di adattamento possibili, imposta la calibrazione del parametro  $a$ .

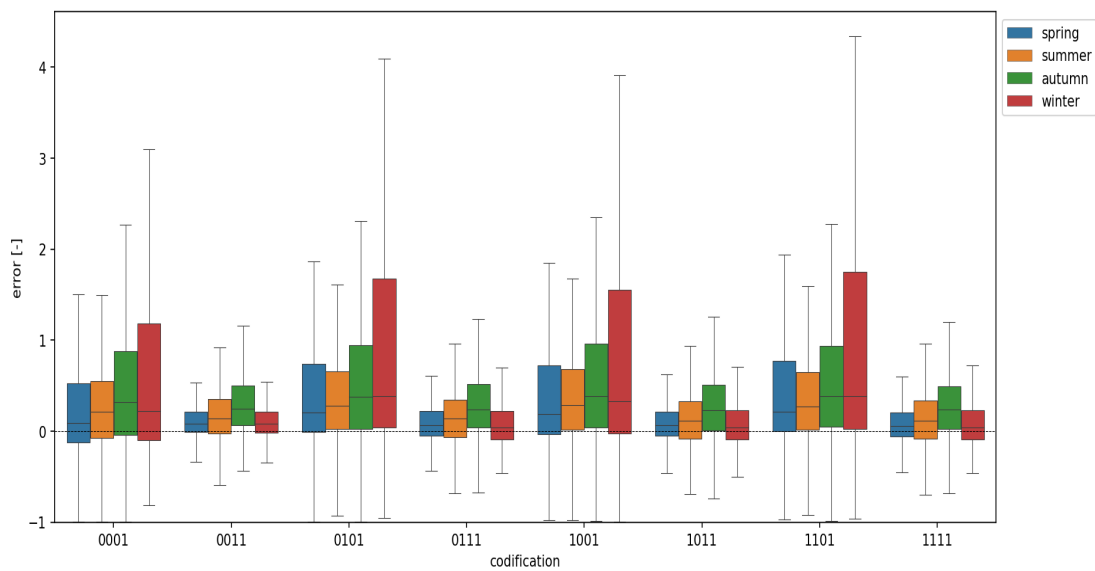


Figura 13: Diagrammi della mediana e dei quartili degli errori relativi sulle osservazioni dei massimi stagionali in tutte le combinazioni di adattamento possibili, imposta la calibrazione del parametro  $K$ .

È nelle combinazioni in cui anche il secondo parametro legato alle recessioni compare tra quelli adattati numericamente per minimizzazione della funzione della probabilità ai dati osservati, che si possono conseguire i migliori risultati assoluti, in modo sostanzialmente indipendentemente dal numero di parametri secondari ottimizzati.

### 3.1.3 Modello delle recessioni

Si vuole ora valutare l'effetto che il metodo di stima utilizzato per i parametri  $a$  e  $K$  può determinare nei risultati finali. Come si è visto nel Capitolo 2.3, è possibile stimare questi parametri adattando direttamente l'equazione (2.6), per via numerica, o utilizzare una procedura di linearizzazione che ne permetta una soluzione più diretta. Il primo metodo è quello adottato, alla luce di quanto si discuterà nelle prossime pagine. Entrambi gli approcci sono stati sistematicamente applicati e si è proceduto alle medesime valutazioni condotte in precedenza, valutando l'errore relativo calcolato su ogni coppia osservazione-previsione del modello PHEV-MEV, rappresentato graficamente nei diagrammi degli estremi e dei quartili di Figura 14.

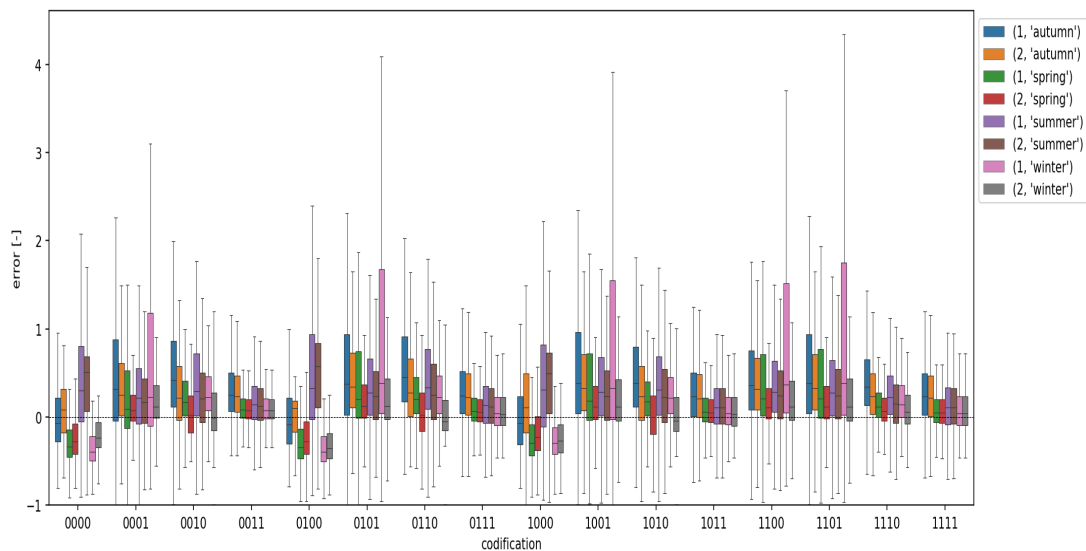


Figura 14: Diagrammi della mediana e dei quartili degli errori relativi sulle osservazioni dei massimi stagionali in tutte le combinazioni di valutazione dei parametri, avendo utilizzato un modello delle recessioni non-lineare (1) e lineare (2) per la stima di  $a$  e  $K$ .

Nel grafico e nella trattazione si utilizza la notazione:

- 1 quando si adotta il modello non-lineare;
- 2 quando si adotta il modello lineare.

L'adozione del modello 1 per la stima dei parametri delle recessioni porta ad un significativo miglioramento complessivo delle prestazioni del modello PHEV-MEV, e a dei diagrammi graficamente più compatti. In altre parole si riduce la dispersione dei valori degli errori attorno alla mediana. Per quanto riguarda il modello di ottimizzazione numerica, sono formulabili alcune importanti osservazioni. Nel processo di adattamento dei parametri della distribuzione di probabilità, come si è visto, sono utilizzati, quando disponibili, i valori stimati con i metodi del Capitolo 2.3, che nel caso di  $a$  e  $K$  dipendono dal modello delle recessioni adottato. Questo significa che la variazione delle stime per via del differente modello adottato ha effetto sui valori di primo tentativo utilizzati in fase di calibrazione. Si evince che il processo di ottimizzazione sostanzialmente non risente del tipo di procedura scelta. Per verificare ciò è necessario confrontare tutti quei casi di combinazione di parametri in cui almeno  $a$  e  $K$  vengono ottimizzati, cioè tutti quei casi che presentano una nomenclatura -11. In tutte queste ultime combinazioni, dove di fatto si ignora la concettualizzazione fisica dei parametri delle recessioni, ovvero si rinuncia a quella porzione di informazioni che provengono dall'analisi degli idrogrammi giornalieri, l'adattamento può effettuarsi a partire da valori iniziali semplicemente imposti o stimati con il metodo che assicuri la massima stabilità ed efficienza.

### **3.1.4 Stagionalità**

Per meglio comprendere quali sono i fattori che condizionano in modo significativo le capacità predittive del modello studiato, si analizza con maggior dettaglio l'effetto della stagionalità. Seguono gli errori relativi calcolati sugli estremi periodici di portata, rappresentati in Figura 15 per il modello PHEV ed in Figura 16 per il modello combinato PHEV-MEV.

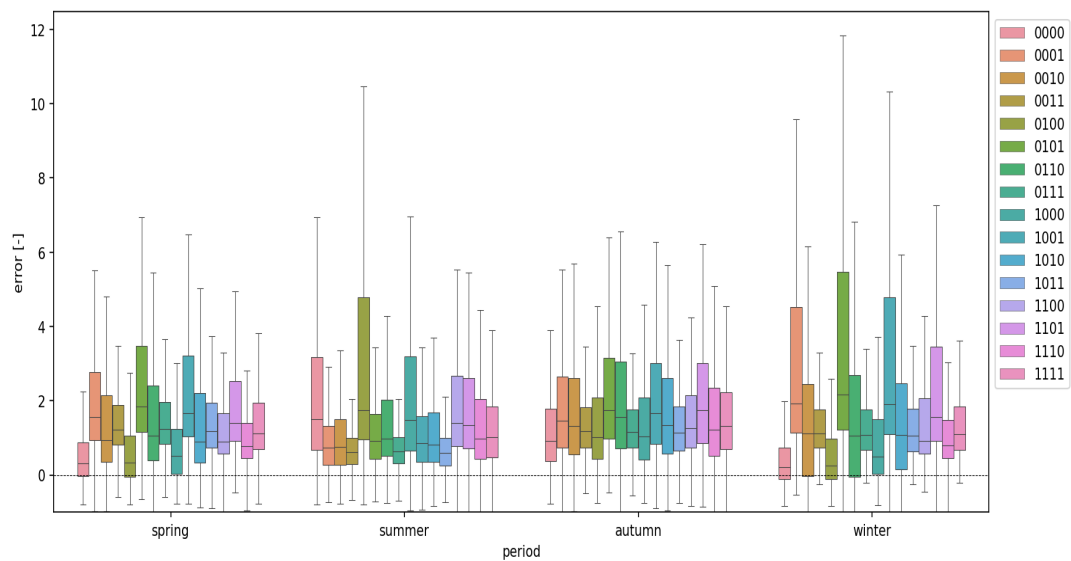


Figura 15: Diagrammi della mediana e dei quartili degli errori relativi sulle osservazioni dei massimi stagionali in tutte le combinazioni di valutazione dei parametri, raggruppati per stagione.

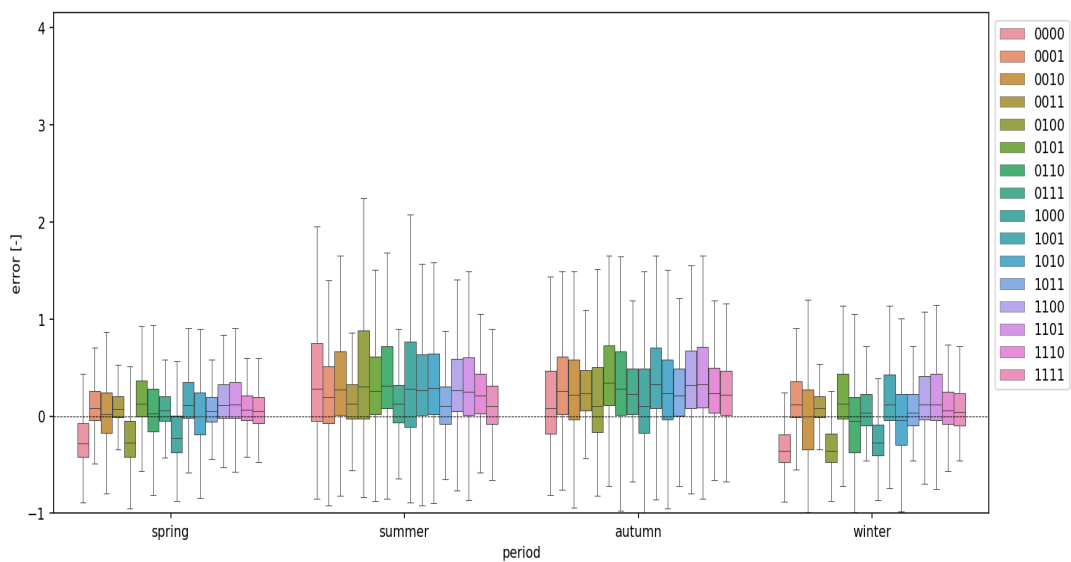
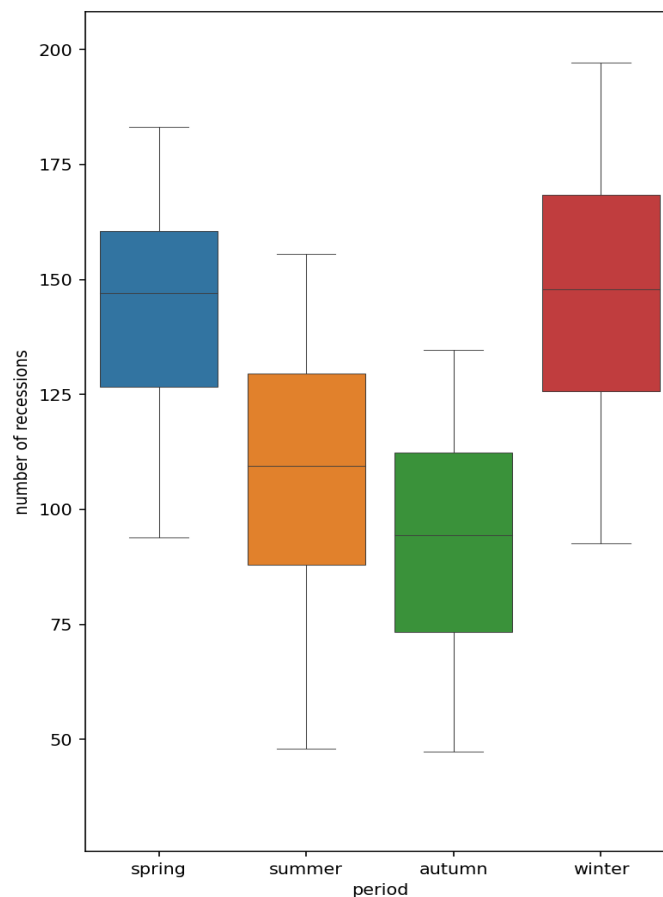


Figura 16: Diagrammi della mediana e dei quartili degli errori relativi sulle osservazioni dei massimi stagionali in tutte le combinazioni di valutazione dei parametri, raggruppati per stagione.

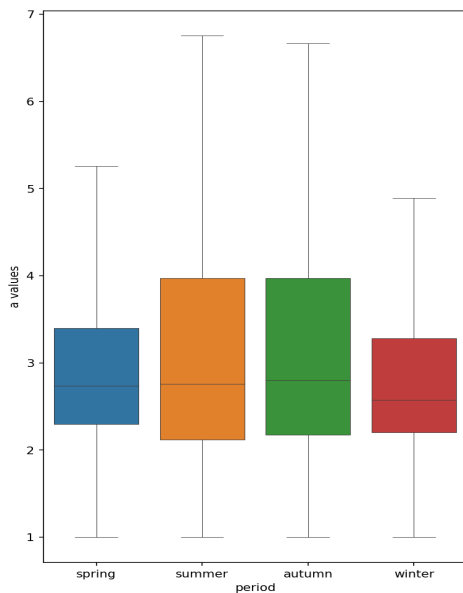
I grafici mostrano anzitutto che, raggruppando tutte le possibili combinazioni di valutazione dei parametri per stagione, è possibile individuare in modo visivamente chiaro due principali tendenze. Le stagioni primaverile e invernale sono contraddistinte da una minore variabilità dell'errore e valori mediani contenuti. Le stagioni estiva e autunnale registrano errori più elevati e fondamentalmente di sovrastima. Le principali anomalie si riscontrano nei set 0000, 1000 e 0100, ovvero in quei casi in cui sia  $a$  che  $K$  sono stimati dalle recessioni. In primavera ed inverno questi corrispondono alle sottostime delle portate massime periodiche predette. Si prendano quindi in esame i diagrammi della mediana e dei quartili del numero di recessioni estratte dai bacini, nell'intero periodo delle osservazioni e nelle quattro stagioni considerate, illustrati in Figura 17.



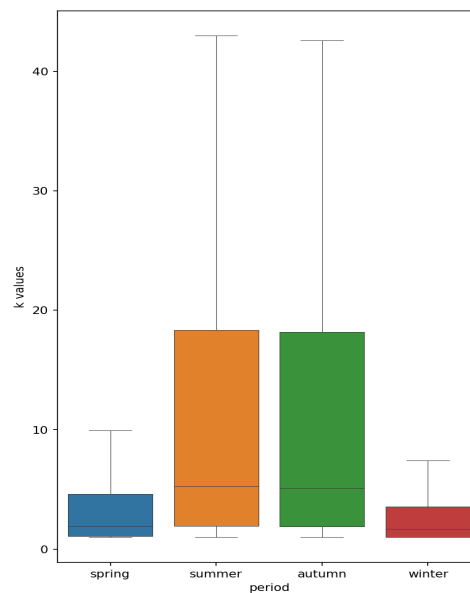
*Figura 17: Diagrammi della mediana e dei quartili del numero di recessioni estratte dal modello nell'intero periodo di osservazione, per stagione.*



Il grafico mostra che il numero di recessioni complessivamente estratte è superiore in inverno e primavera rispetto a quello che si ha in estate ed in autunno. Nelle prime due stagioni si riscontrano le principali anomalie quando la stima dei parametri  $a$  e  $K$  segue esclusivamente dall'adattamento della (2.6) alle recessioni estratte. Un maggior numero di recessioni può portare con sé anche una incertezza di stima, nell'individuazione di un valore opportunamente rappresentativo dell'intero insieme di curve adattate. Stando a quanto suggerito dai risultati, l'incertezza si traduce in fenomeni di sottostima dei massimi di portata. Il numero delle recessioni è anche indicazione diretta del numero di picchi ordinari di portata e qui si può registrare l'effetto opposto, perché è proprio per le stagioni che presentano il numero più consistente di recessioni che i set adattati numericamente riducono sensibilmente gli errori previsionali del modello. Questa evidenza conferma l'efficacia del modello di adattamento, che all'aumentare dei picchi di portata estratti riesci a meglio ottimizzare i parametri. Se si valutano i valori di  $a$  e  $K$  stimati nelle diverse stagioni, rappresentati nei diagrammi di Figura 18, è possibile caratterizzare ulteriormente quanto fino ad ora riscontrato.



*Figura 18: Diagrammi della mediana e dei quartili dei valori di  $a$  stimati nelle diverse stagioni.*



*Figura 19: Diagrammi della mediana e dei quartili dei valori di  $K$  stimati nelle diverse stagioni.*

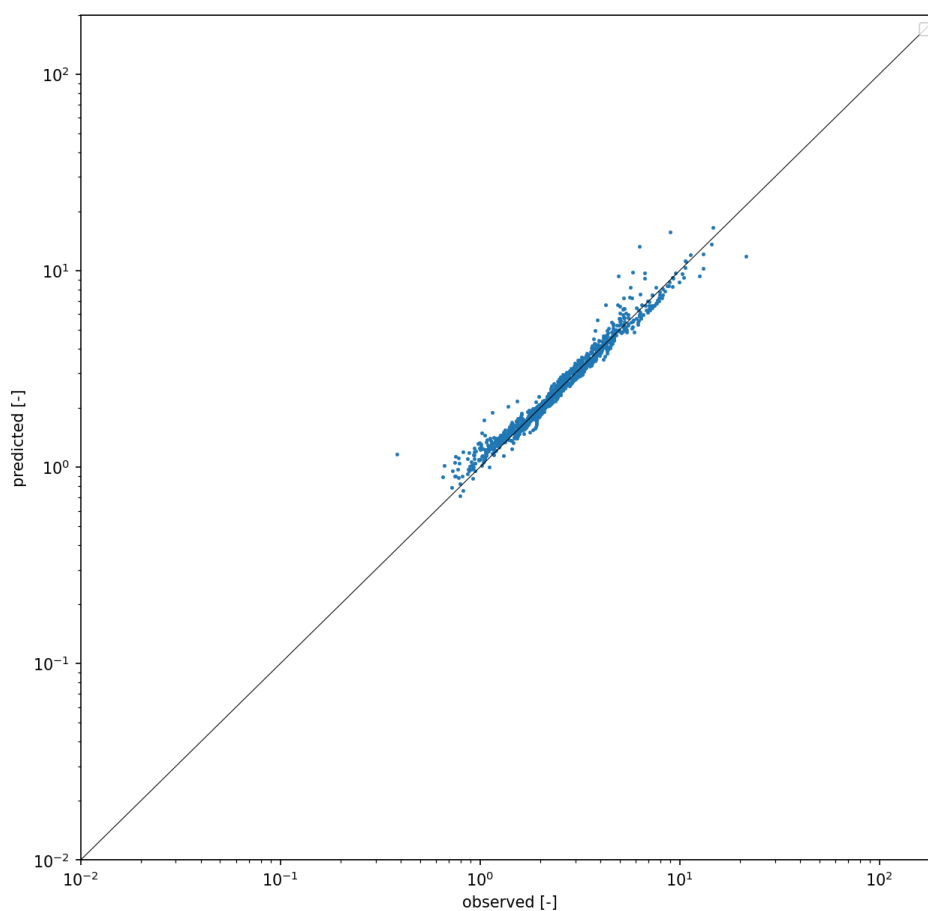
Sempre in accordo con le due tendenze principali individuate in precedenza, i valori dei parametri presentano una chiara disposizione. Essi sono maggiormente dispersi attorno al valore medio in estate ed autunno rispetto a quelli di inverno e primavera. Nel caso di  $K$  presentano anche valori significativamente maggiori.

## 3.2 Analisi dei campioni

Per approfondire le analisi, si sono classificati i risultati ottenuti, con i metodi esposti al Capitolo 2.6, identificando tre diversi campioni di dimensione  $n$ :

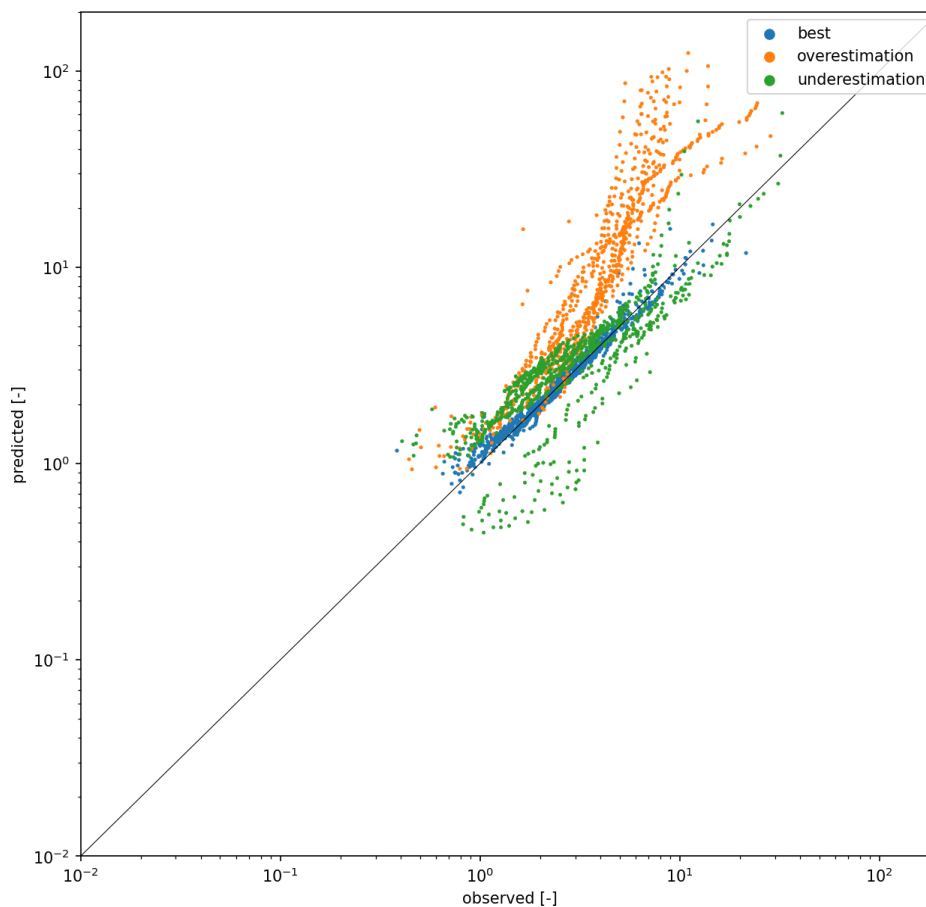
- i migliori casi studio, ovvero quelli in cui si verifica la più significativa corrispondenza tra estremi di portata previsti e osservati;
- i casi studio di più importante sovrastima delle portate;
- i casi studio di più importante sottostima delle portate.

Per caso studio si intende l'oggetto minimo analizzato, che è costituito da bacino e stagione. Essendo 178 i bacini, valutati nelle 4 stagioni dell'anno, i casi complessivi risultano essere 712. Si sceglie di campionare con una dimensione di  $n$  pari a 20. Nei grafici di dispersione che seguono i punti hanno per coordinate i valori adimensionali massimi stagionali di portata osservati, sull'asse delle ascisse, ed i valori predetti dal modello, sull'asse delle ordinate. Ogni valore massimo osservato è caratterizzato da una probabilità di non superamento, calcolata empiricamente, e, in altri termini, da un tempo di ritorno. Per la stessa probabilità di non superamento è calcolato il valore predetto dal modello. Dal momento che le portate assumono solo valori positivi, i grafici sono rappresentati nel primo quadrante e la bisettrice è sostanzialmente il luogo dei punti in cui si verifica la perfetta corrispondenza tra valore osservato e valore predetto dal modello. Più sono i punti disposti in un intorno della bisettrice, migliore è il modello che descrive i dati. Se i dati si trovano nella regione confinata superiormente, significa che il modello previsionale sovrastima i valori di portata rispetto alle osservazioni, al contrario, se nella regione inferiore, il modello li sottostima. In figura 20 si riporta il grafico di dispersione osservazioni-previsioni dei massimi per il campione dei migliori casi studio.



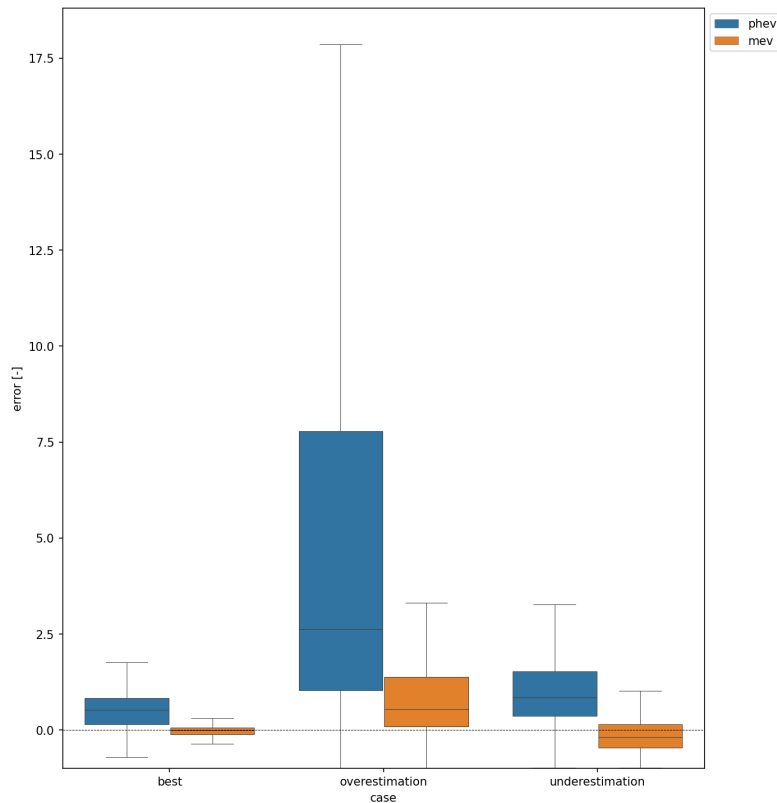
*Figura 20: Grafico di dispersione osservazioni-previsioni dei massimi periodici di portata per il campione dei migliori casi studio.*

Si può apprezzare la disposizione delle coppie osservazioni-previsioni, che presenta una chiara tendenza lineare. Sempre grazie a questo grafico è possibile verificare che la dispersione è minima nella parte più baricentrica della popolazione dei dati, dove per altro si concentra anche il numero maggiore di eventi. L'incertezza cresce allontanandosi dalla porzione centrale, verso le estremità. In questi casi, per i valori più bassi e più alti delle portate massime periodiche è possibile leggere una propensione alla sovrastima da parte del modello. Si considerino ora anche i casi studio meno prestanti, contraddistinti dalle più significative sovrastime e sottostime, nel diagramma di dispersione di Figura 21.



*Figura 21: Grafico di dispersione osservazioni-previsioni dei massimi periodici di portata per il campione dei migliori casi studio (blu), peggiori sovrastime (arancione) e peggiori sottostime (verde).*

I casi affetti dalle sovrastime più significative delle portate estreme sono dispersi nella parte superiore destra del grafico, cioè la discrepanza tra previsioni del modello e dati osservati è crescente al crescere della probabilità di non superamento degli eventi considerati, ovvero dell'entità stessa della piena. In altre parole, la coda delle curve di probabilità di inondazione cresce con una pendenza superiore a quella che in realtà sarebbe necessaria. Per quanto concerne i casi affetti dalle più gravi sottostime è possibile osservare che una dispersione significativa avviene nella parte inferiore sinistra della nuvola di punti, cioè per tempi di ritorno più limitati si verificano gli errori maggiori. Si confrontano infine gli errori relativi sui massimi stagionali per PHEV e PHEV-MEV, all'interno dei tre campioni di casi utilizzati, in Figura 22.



*Figura 22: Diagrammi della mediana e dei quartili degli errori relativi sulle osservazioni dei massimi stagionali nel modello PHEV (blu) e PHEV-MEV (arancione) per i tre campioni di casi studio.*

Gli errori relativi sui massimi di portata calcolati con la distribuzione semplificata degli estremi stagionali proposta da PHEV si presentano sostanzialmente scalati rispetto a quelli di MEV, mantenendo un rapporto

### **3.2.1 I parametri delle recessioni**

Sono analizzati i parametri  $a$  e  $K$  nei i tre campioni al fine di confrontare il valore da essi assunto, relativamente alla procedura di stima adottata, nei diagrammi in Figura 23, per  $a$ , e Figura 24, per  $K$ . Nei casi peggiori l’ottimizzazione numerica dei suddetti parametri ha sempre fornito risultati previsionali di qualità superiore. Questa operazione è necessaria per approfondire l’esistenza di una relazione tra intervalli di valore e prestazioni del modello.

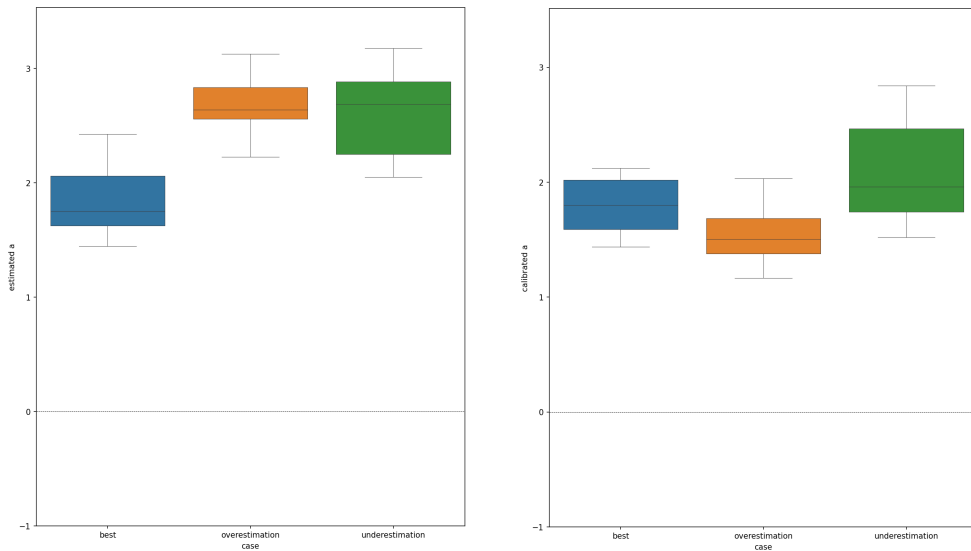


Figura 23: Diagrammi della mediana e dei quartili dei valori assunti dal parametro  $a$  stimato da adattamento di funzione delle recessioni (sinistra) e per ottimizzazione della distribuzione di probabilità degli eventi sui picchi di portata (destra) nei tre campioni di casi studio.

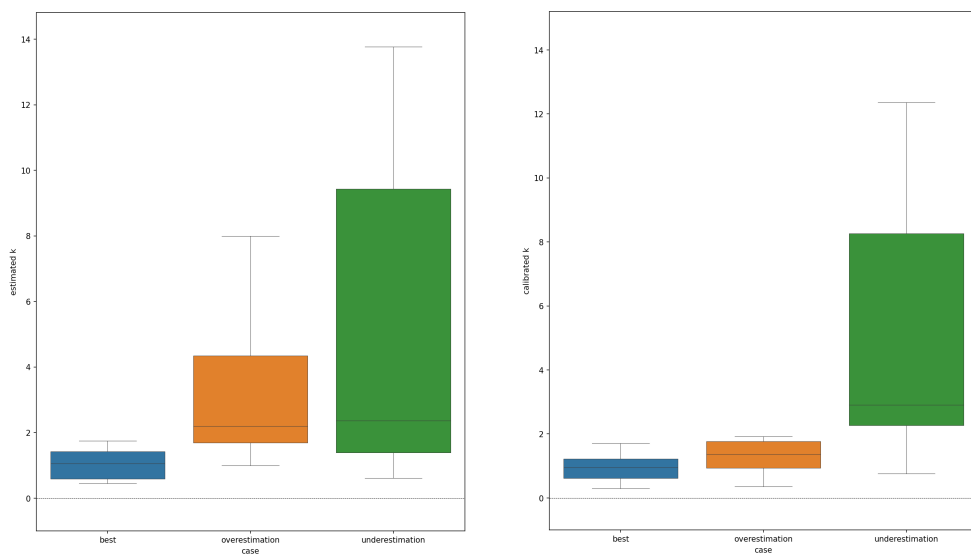
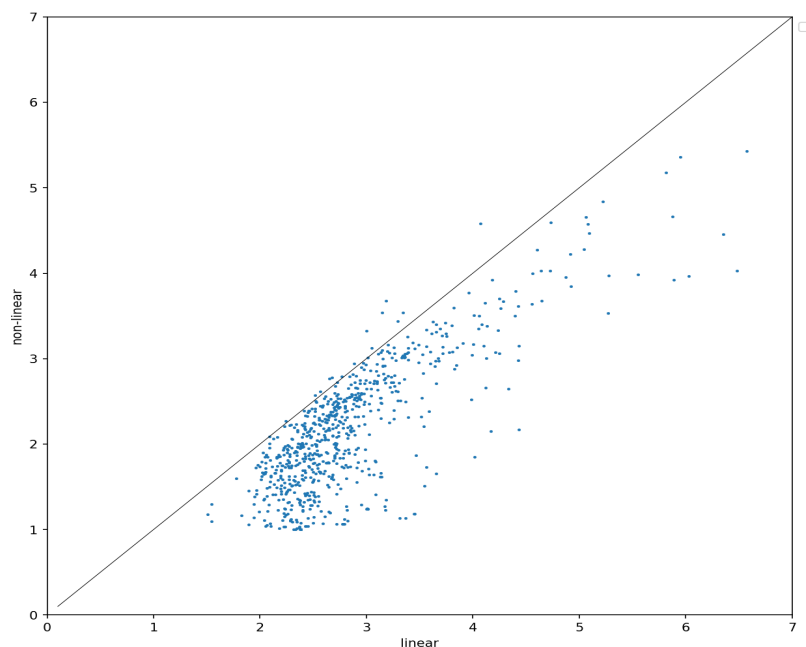


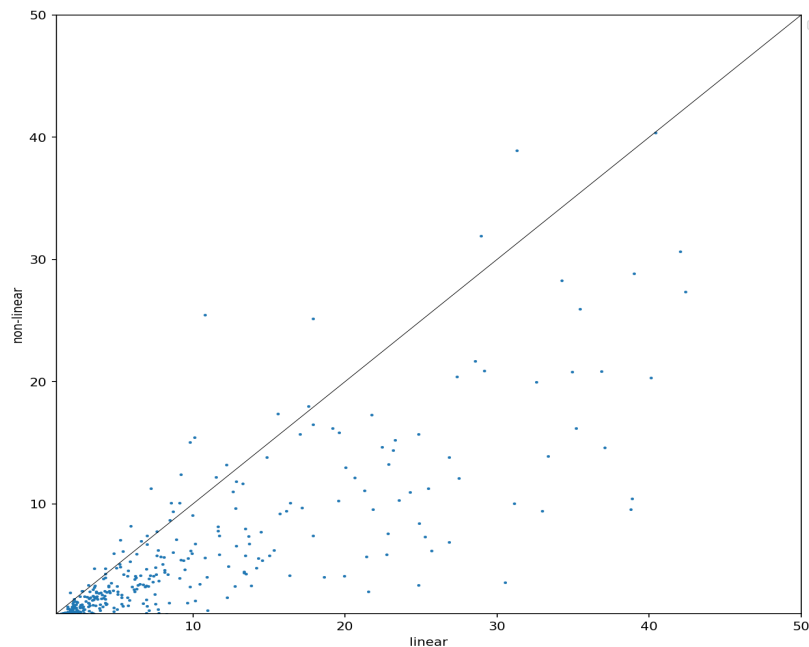
Figura 24: Diagrammi della mediana e dei quartili dei valori assunti dal parametro  $K$  stimato da adattamento di funzione delle recessioni (sinistra) e per ottimizzazione della distribuzione di probabilità degli eventi sui picchi di portata (destra) nei tre campioni di casi studio.

Si evidenzia innanzitutto che, come è ragionevole attendersi, i casi in cui il modello presenta le prestazioni più soddisfacenti sono quelli in cui è maggiore l'accordo tra i valori dei parametri stimati con i due metodi. È poi da notare che, nei casi di sovrastima delle portate, il valore di  $a$  ottimale, calibrato dal modello di ottimizzazione, è sensibilmente più basso di quello che il modello di stima più semplice aveva fornito, e lo stesso, seppur in misura minore, vale per i casi in cui avviene sottostima. Analoghe considerazioni valgono per  $K$ , indicando che valori troppo grandi di tale parametro sembrano concorrere al verificarsi delle principali anomalie. Nel Capitolo 3.1 si è mostrato come il modello delle recessioni adottato per stimare  $a$  e  $K$  sia in grado di incidere sulle previsioni finali. Si consideri ora il campione dei risultati migliori e si confrontino con dei diagrammi di dispersione i valori dei parametri ottenuti perseguendo i due approcci proposti. Ciascun punto nel piano cartesiano è cioè costituito dal valore stimato adattando il modello delle recessioni lineare, in ascissa, e non-lineare, in ordinata. Anche in questo caso, se la nuvola di punti è concentrata lungo la bisettrice del quadrante significa che i due modelli sono sostanzialmente in accordo tra loro, al contrario, pongono risultati diversi.



*Figura 25: Grafico di dispersione del valore stimato con modello lineare contro modello non-lineare per  $a$ .*

In figura 25 si può osservare che adattando il modello delle recessioni lineare si tende sostanzialmente in tutti i casi a stimare dei valori di  $a$  maggiori di quelli che si ottengono nel caso di modello non-lineare. Lo stesso tipo di fenomeni si riscontra per i valori di  $K$ , nel grafico di Figura 26.



*Figura 26: Grafico di dispersione del valore stimato con modello lineare contro modello non-lineare per  $K$ .*

Nel Capitolo 3.1 si è osservato che i risultati riscontrati l'adattamento dell'equazione (2.8) permette di ottenere risultati sensibilmente migliori in tutti quei casi in cui i set di parametri utilizzati  $a$  e  $K$  preservano il significato fisico che l'approccio del modello PHEV prevede, cioè non si opta per una calibrazione numerica. In accordo con quanto visto nelle ultime pagine, la sovrastima dei valori di questi parametri condiziona negativamente le prestazioni del modello previsionale. Si conferma quindi la fondamentale importanza che l'approccio di stima ricopre al fine di ottenere una corretta descrizione degli eventi di piena e che le principali incertezze di quest'ultimi sono fortemente legate ad esso.



### 3.4 Alcuni casi esemplari

Si considerano ora alcuni casi esemplari, al fine di apprezzare nel particolare i risultati ottenuti dal modello studiato. Sono scelti quattro bacini distinti, ciascuno in una stagione differente. Per ognuno si propone qui:

- il grafico delle portate massime in funzione del tempo di ritorno, in scala semi-logaritmica;
- il grafico delle probabilità cumulate di eccedimento in funzione delle portate, per gli eventi ordinari e per i massimi stagionali, in scala bi-logaritmica.

Della prima tipologia di grafico è disponibile un campione più corposo di esempi e si rimanda all'Appendice B per la sua consultazione. In merito alle curve rappresentate, non si riportano tutte le 16 combinazioni di set di parametri valutate nelle analisi, sia per l'impossibilità grafica di restituire i risultati in modo chiaro ed efficiente, sia per la poca rilevanza che ciò comporterebbe alla luce delle principali deduzioni sviluppate. Si prendono in esame i quattro set seguenti, di cui si ricordano le caratteristiche:

- 0000 set di parametri stimati direttamente con procedure di letteratura;
- 1100 set in cui  $\alpha$  e  $\lambda$  sono variabili adattate numericamente;
- 0011 set in cui  $a$  e  $K$  sono variabili adattate numericamente;
- 1111 set di parametri adattati numericamente.

Queste combinazioni individuano cioè i principali scenari di utilizzo del modello idrologico, perché i due blocchi di parametri dipendono, nelle procedure di stima diretta, uno dai dati di precipitazione ( $\alpha$  e  $\lambda$ ), uno da quelli di portata ( $a$  e  $K$ ). Questo significa che, operativamente, sono condizionati dalla quantità di informazioni di cui si dispone. Le possibili altre combinazioni ottenibili a piacere possono essere arbitrariamente adottate, ma risultano poco significative per le analisi.

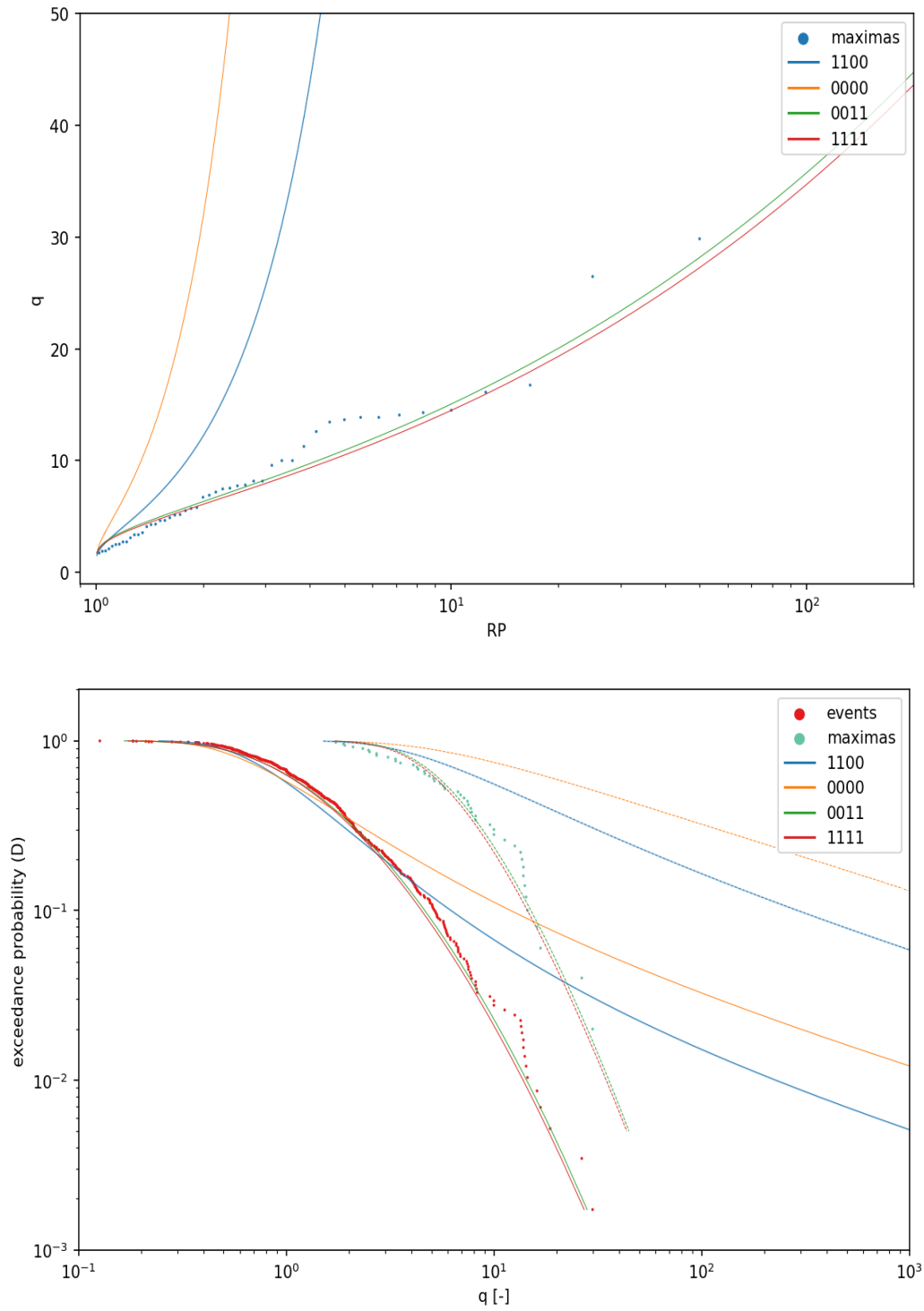


Figura 27: Curva di probabilità di inondazione (sopra) e funzioni di probabilità cumulata di eccedimento degli eventi ordinari e dei massimi (sotto), con valori osservati, per il caso 18662000 summer.

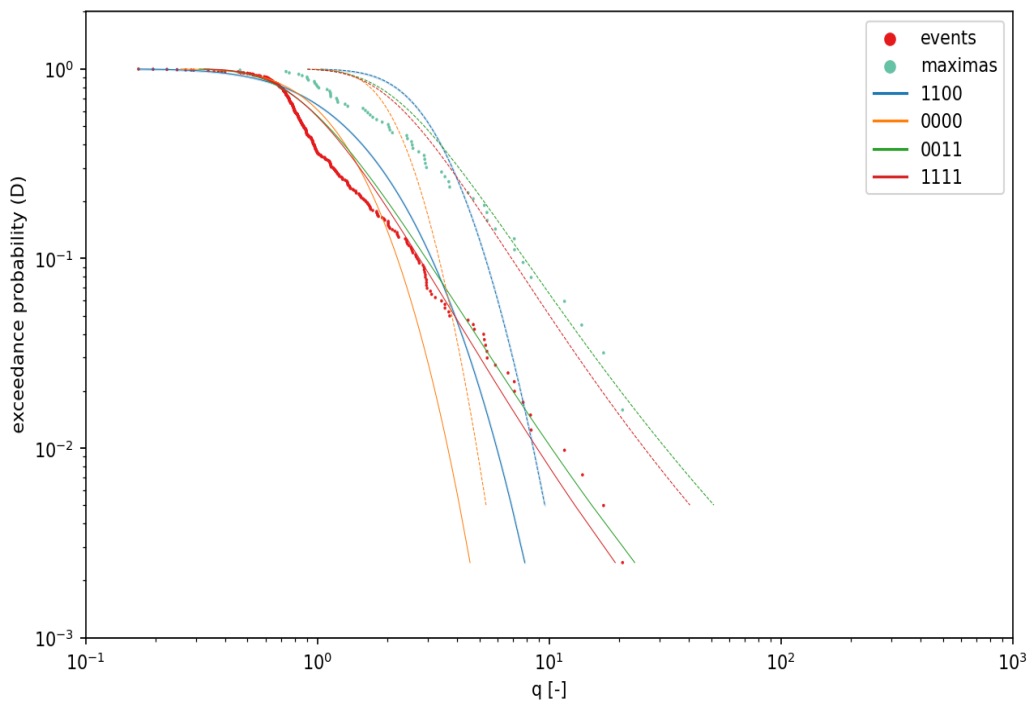
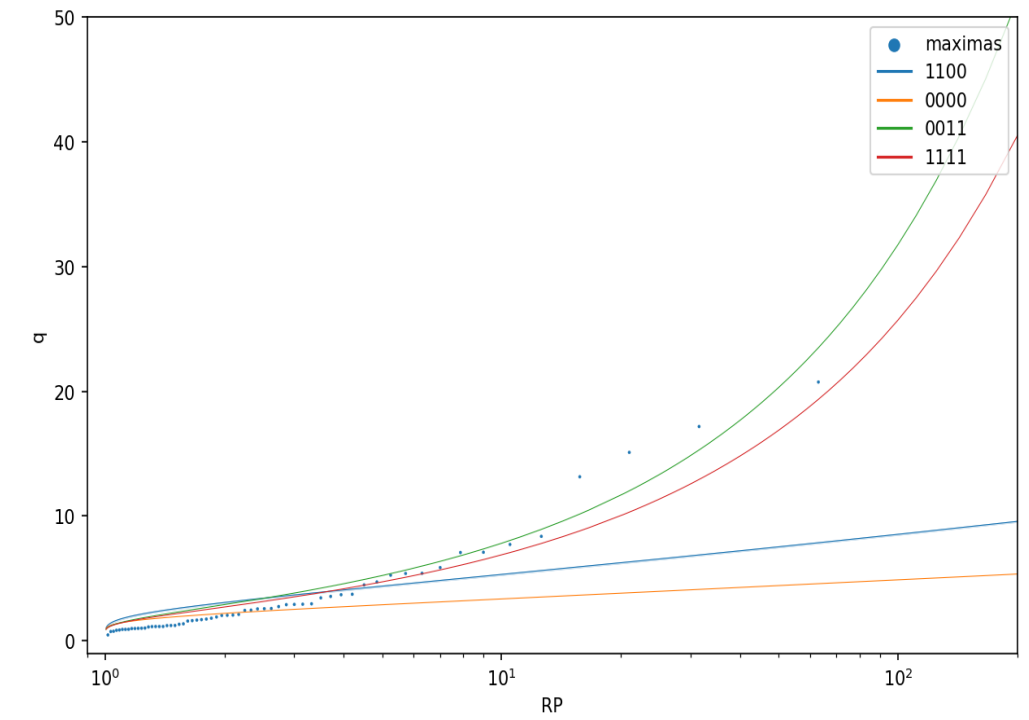


Figura 28: Curva di probabilità di inondazione (sopra) e funzioni di probabilità cumulata di eccedimento degli eventi ordinari e dei massimi (sotto), con valori osservati, per il caso 27615100 autumn.

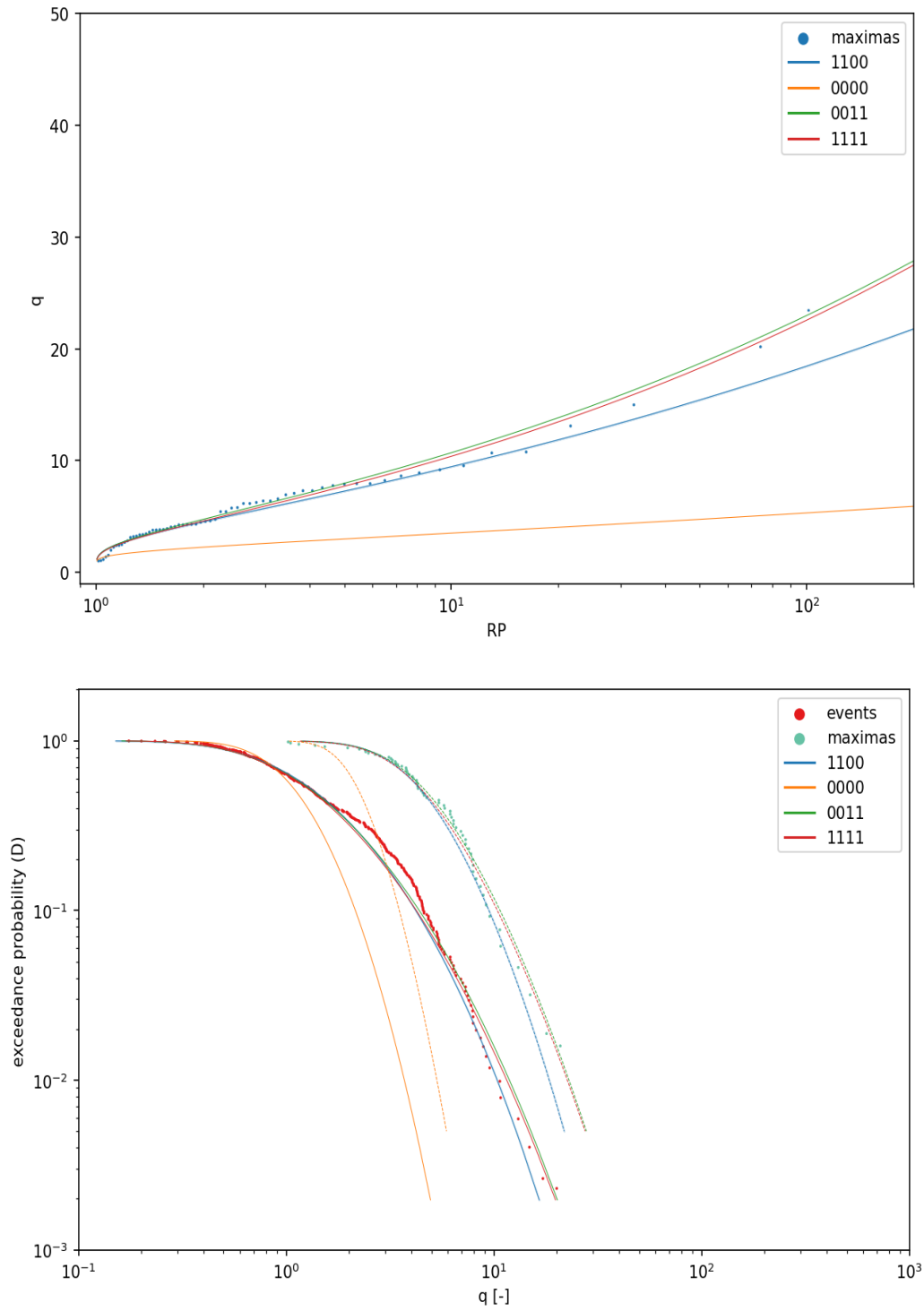


Figura 29: Curva di probabilità di inondazione (sopra) e funzioni di probabilità cumulata di eccedimento degli eventi ordinari e dei massimi (sotto), con valori osservati, per il caso 1301 winter.

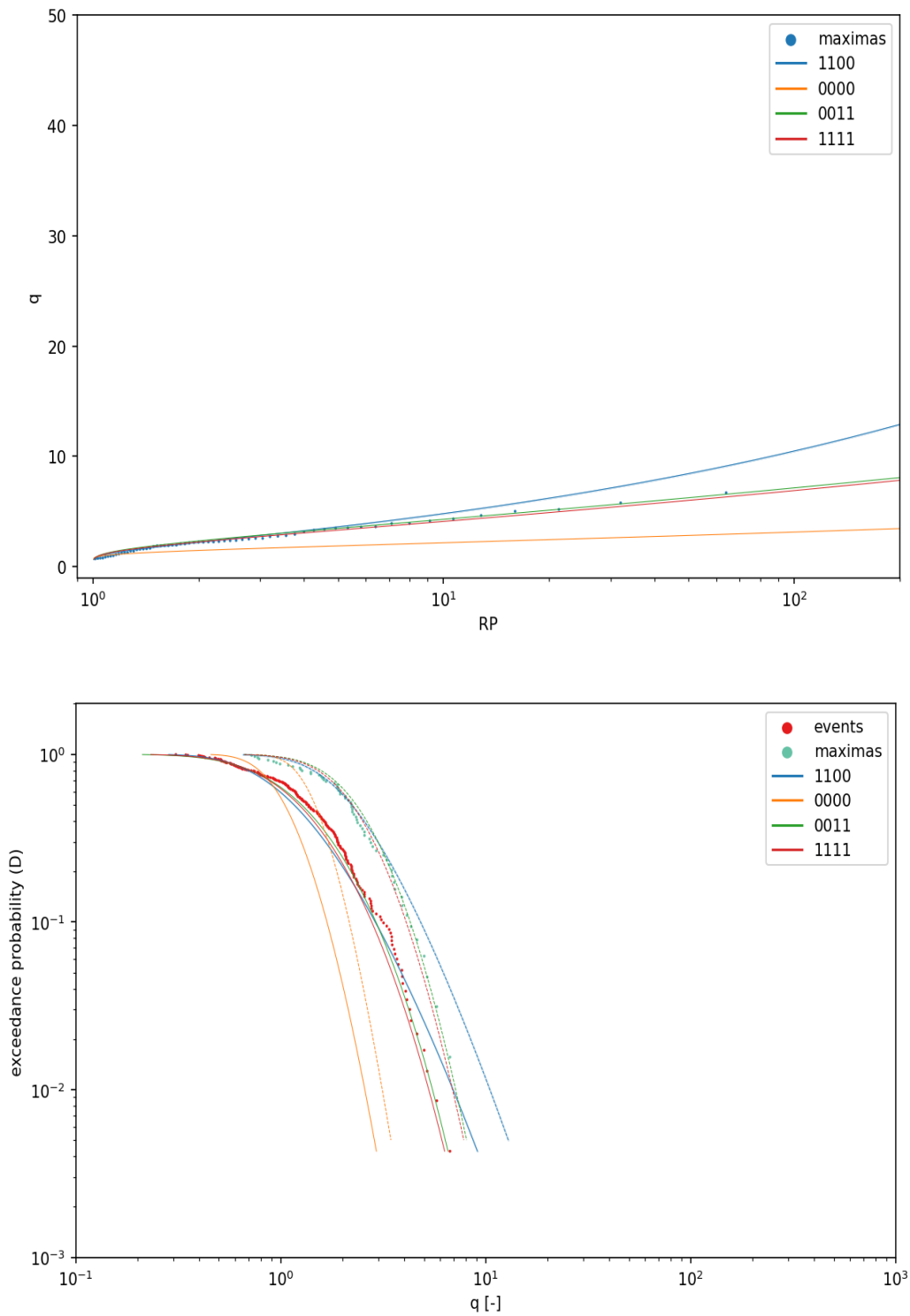


Figura 30: Curva di probabilità di inondazione (sopra) e funzioni di probabilità cumulata di eccedimento degli eventi ordinari e dei massimi (sotto), con valori osservati, per il caso 125 spring.

I grafici mostrano che il modello PHEV-MEV, se opportunamente valutati i parametri, è in grado di prevedere i massimi di portata stagionali con una accuratezza significativa. Nei casi in cui set 0000 si rivela inadeguato a descrivere correttamente i massimi stagionali di portata, passare al set 1100 può tradursi in miglioramenti più o meno sostanziali, in virtù del fatto che, in questo modo, si offrono all'algoritmo di calibrazione due gradi di libertà utili a ridurre le incongruenze del modello. È poi con i set 0011 e 1111 che si raggiungono le migliori prestazioni del modello. quanto si è discusso precedentemente circa il comportamento del set semi-adattato 0011, che tende ad essere succedaneo, quando non migliore del set completamente adattato 1111, è qui riscontrabile e conferma che la stima fisicamente basata dei parametri  $\alpha$  e  $\lambda$  si rivela efficace quanto quella ottenibile per ottimizzazione. I grafici inferiori permettono di visualizzare la trasformazione che il modello qui adottato pone in essere tra i picchi ordinari di portata e i massimi periodici, estraendo dagli eventi ordinari quelli massimi, dove è possibile saggiare visivamente che la quantità di dati disponibili per i primi è notevolmente superiore. Il set semi-adattato 0011 in sostanza preserva la componente del modello che recepisce le informazioni in merito alle precipitazioni, mentre altera quella derivante dai dati di portata, perché in luogo della stima dei parametri da recessioni adatta direttamente la funzione di probabilità degli eventi di portata.

## CONCLUSIONI

La modellazione degli eventi di piena estremi è elemento fondante della progettazione e si basa su una letteratura scientifica ed applicazioni che si apprestano a compiere l'età di un secolo. I modelli qui esaminati rappresentano alcuni dei nuovi sviluppi in materia e il metodo che si è qui proposto vuole contribuire a superare le limitazioni delle procedure tradizionali di stima delle portate dei corsi d'acqua, grazie alla combinazione di approcci statistici e fisici. I principali risultati a cui si è pervenuti sono:

- l'estrazione delle portate estreme annuali è possibile a partire dalla serie di picchi di portata ordinari;
- l'approccio fisicamente basato del modello concettuale adottato si presta a fornire una distribuzione di probabilità adeguata dei picchi di portata ordinari;
- i dati di precipitazioni sono descrittori fisici consistenti ed i parametri ad essi associati sono facilmente stimabili con espressioni chiuse;
- la modellazione delle recessioni mostra particolare incertezza e la calibrazione implementata dei parametri ad esse associati ne permette una valida stima;
- il metodo di semi-calibrazione dei parametri del modello concettuale permette di ottenere previsioni delle portate estreme annuali esaurienti.

Si è mostrato come sia possibile ottenere una modellazione idrologica grazie ad un approccio in grado di estrarre le portate estreme annuali dall'intera serie di picchi di portata ordinari, ampliando notevolmente il dominio di informazioni a disposizione. Questo risultato è di grande interesse ingegneristico perché promette di fornire stime di portate di progetto, complete di incertezza, nei bacini nei quali la quantità di osservazioni di portata è insufficiente per i metodi tradizionali, che si basano sull'adattamento dei soli massimi annuali. Il modello che si è adottato per descrivere la distribuzione di probabilità degli eventi ordinari propone una parametrizzazione fisicamente basata, che mira a superare un approccio meramente statistico, oltre a recepire un più ampio spettro di informazioni utili per la definizione di uno schema fenomenologico aderente alla realtà. Questo avviene con l'introduzione nel modello di

schemi descrittivi dei dati storici osservati sia delle precipitazioni, sia delle portate. Si è visto che la stima dei parametri è cruciale per pervenire ad una corretta previsione degli estremi di portata e la funzione analitica del modello gode di gradi di libertà e forma in grado di adattarsi opportunamente a dispersioni di dati di diversa natura. Le procedure proposte in letteratura sono attraenti per la semplicità di applicazione. I risultati mostrano come la valutazione dei parametri  $\alpha$  e  $\lambda$  risulti più efficace, mentre più complessa è quella di  $a$  e  $K$ . L'incertezza nella determinazione di questi ultimi parametri è imputabile allo schema di modellazione adottato per descrivere le recessioni dell'idrogramma. Quando aumentano i picchi ordinari di portata identificati, e conseguentemente anche il numero di recessioni, la stima di  $a$  e  $K$  per adattamento della legge di potenza della relazione volume-deflusso conduce a previsioni delle portate estreme annuali affette da errori significativi. L'approccio di calibrazione di questi due parametri sui dati di portata consente di ottenere dei set di parametri che mostrano previsioni delle portate estreme annuali esaurienti e questo effetto è tanto più riscontrabile quanto più aumenta il numero di dati. L'adattamento numerico implementato di questi ultimi due parametri permette di superare le incertezze della stima e preserva la significatività fisica che contraddistingue il modello idrologico, dal momento che avviene sui dati di portata, mentre le precipitazioni continuano a concorrere alla caratterizzazione concettuale. L'indicazione è quella di utilizzare il modello proposto prevedendo la stima di  $\alpha$  e  $\lambda$  con i semplici metodi suggeriti dalla letteratura e la calibrazione di  $a$  e  $K$  su dati di portata, condizione che corrisponde al set di parametri che è stato denominato 0011.



## APPENDICE A

In questa appendice si riportano in tabella tutti i bacini utilizzati nelle analisi, con l'indicazione di nome, ID, superficie e finestra temporale delle osservazioni.

nome	ID	area (km <sup>2</sup> )	anni
KIRCHEN-HAUSEN	105	773	1951-2008
HUNDERSINGEN	120	2620	1951-2013
BERG	125	4046	1951-2013
ACHSTETTEN	129	276.6	1951-2010
LAUBE	144	312.9	1951-2013
BAD MERGENTHEIM	212	1020	1951-2011
ALTENSTEIG	416	135.3	1951-2010
RIEDERICH	422	169	1951-2011
OPPENWEILER	436	178.9	1951-2009
BAD IMNAU	462	332.7	1968-2007
UNTERSCHMEIEN	1120	143	1951-2011
NIEDERKIRC	1144	410.4	1958-2011
FRIDINGE	1156	132.9	1951-2011
BAD ROTENFELS	1301	469.6	1951-2011
SCHWABSBERG	1411	168.7	1954-2003
TAUBERBISCHOFSCHEIM	44602	1591	1951-2013
BERGHAUSEN	60682	232.1	1951-2010
MEININGEN	420020	1166	1951-2013
RAPPELSDORF	421510	251.6	1951-2011
ELLINGSHAUSEN	422000	321.6	1951-2010
MITTELSCHMALKALDEN	424000	154.1	1951-2009
DORNDORF 2	426000	212.8	1951-2009
EISENACH-PETERSBER	429010	327.9	1951-2011
ARENSHAUSEN	447000	272.2	1956-2009
PORSCHDORF 1	550190	165.5	1951-2011
ELBERSDORF	550810	236.3	1951-2011
DOHN	550940	172.6	1961-2009
MERZDOR	552210	209.2	1968-2001
GROSSDITTMANNSDO	554520	284.2	1951-2013
ZWICKAU-POELBITZ	562070	931.9	1951-2011
WECHSELBURG 1	562115	1995	1951-2010
AUE 1	563790	5658	1951-2011
RUDOLSTADT	570270	2623	1951-2011
CALBE-GRIZEHNE	570940	23700	1951-2011
KAULSDORF-EICHICHT	572010	363.2	1961-2006
NIEDERTREBRA	572920	887.4	1951-2011
NORDHAUSEN	575500	311.8	1951-2011
TAUCHRIT	663090	164.9	1951-2011
HERZLAKE	3671101	2312	1951-2010
GOETTINGEN	4881142	658.3	1951-2011
HATTORF	4882168	122.4	1951-2009
GREENE	4885118	2918	1951-2011
COLNRADE OP	4965116	1594	1956-2009
WEILERSWISRT	6335040	261.6	1951-2011
FRIEDRICHSTHAL	6335076	682.5	1953-2010
SCHWAIBACH	6335125	959.6	1952-2010
TRUNSTADT	6335303	11990	1951-2013
WUERZBURG	6335500	14000	1954-2010
SACHSENHEIM	6335510	595.5	1951-2013
ROTTWEIL	6335603	410.8	1951-2008
GUTACH	6335621	302.6	1951-2013
MURR	6335675	504.4	1968-2002

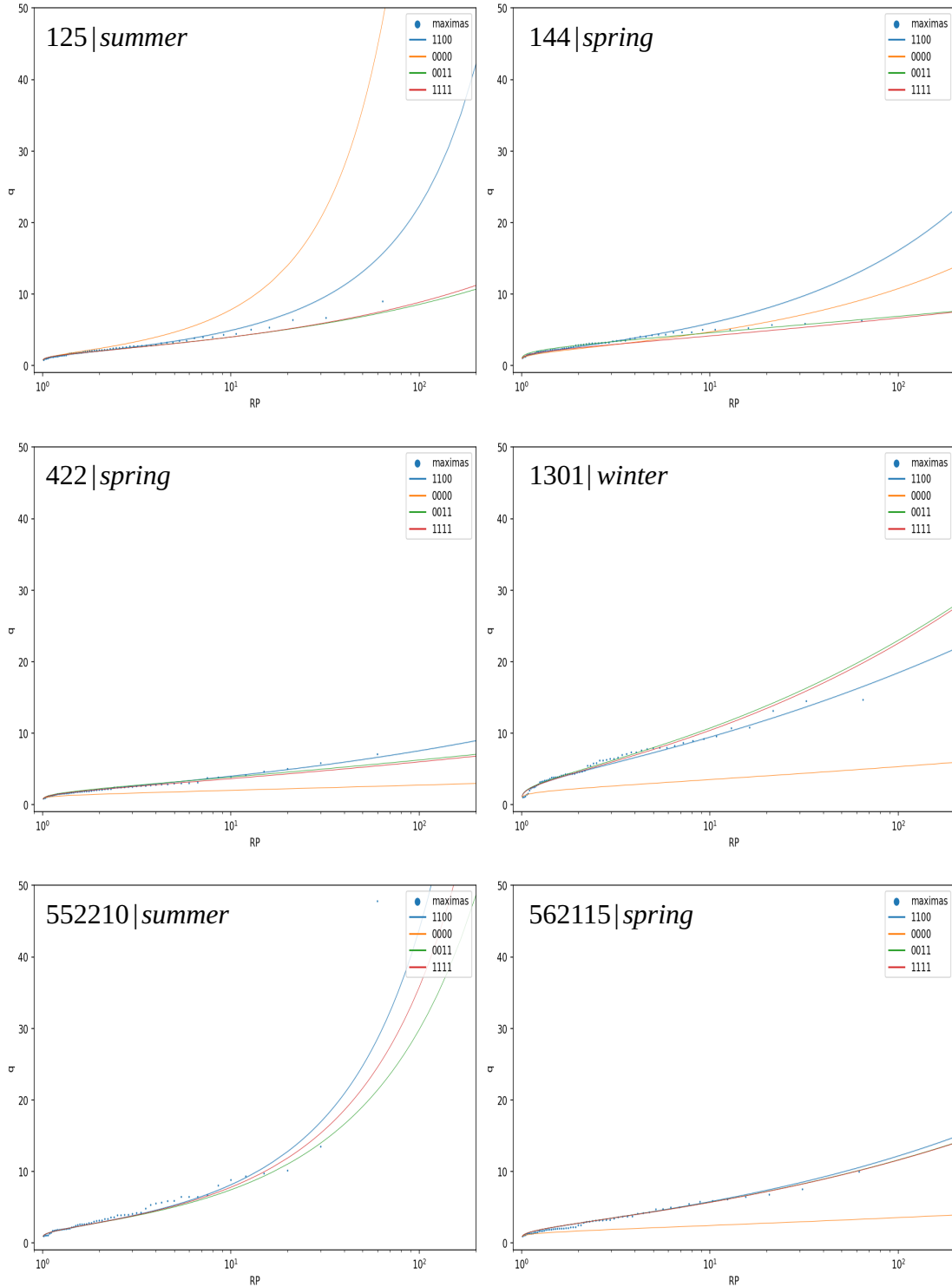
UNTERGRIESHEIM	6335681	1827	1951-2013
NEUSTADT	6335690	570.9	1951-2011
OBERWOLFACH	6335725	124.5	1956-2009
ZELL	6335730	207.9	1951-2011
BAYREUTH	6335810	323.3	1952-2010
RASCH	6335820	208.5	1951-2011
DOERVERDEN	6337515	21860	1956-2009
RECKERSHAUSEN	6337541	317.2	1951-2011
EINEN	6338150	1502	1951-2009
BOKELOH	6338160	326.8	1951-2011
EVERSBURG	6338163	3096	1954-2011
ROCKSTEDT	6338250	590.1	1956-2009
EMMEN	6338260	171	1951-2010
TREIA	6338800	503.1	1951-2010
BIENENBUETTEL	6340050	1415	1951-2013
LAUCHA	6340200	6222	1951-2011
JEHRDEN	6340218	410.1	1951-2011
KETZIN	6340510	14990	1957-2013
ADOLF	6340670	115.9	1968-2005
BAD LIEBENWERDA	6340700	3242	1951-2013
PFÄFFENHOFFEN	6342120	516.3	1951-2007
FISCHACH	6342125	131.1	1954-2010
WIBLINGEN	6342201	2113	1951-2013
RECKENBERG	6342210	125.5	1951-2010
HOHENKAMMER	6342655	388.1	1951-2013
VILSBIBURG	6342675	303.8	1951-2010
HOERMANNENBERG	6342820	211.8	1958-2013
EBERHARDSREUTH	6342960	114.2	1951-2011
SIEGSDORF	6343535	135.6	1951-2005
BLIESDORF	6941100	701.3	1951-2010
DILLINGEN	10035801	11220	1952-2010
INGOLSTADT	10046003	18740	1951-2013
NALBACH	10922200	710.5	1951-2011
KEMPTEN	11402001	842.7	1951-2010
OFFINGEN	11609000	956.9	1956-2009
HARBUR	11809009	1578	1968-2002
DRUISHEIM	11946000	486.1	1951-2013
LEUTERSCHACH	12425002	291.7	1954-2012
DASING	13202003	79.01	1956-2009
M??HLRIED	13206004	513.4	1951-2010
AUNKOFEN	13322005	51.46	1951-2010
EICHSTAET	13407200	1394	1951-2013
UNTERKOEBLITZ	14002305	2008	1951-2010
HEITZENHOFEN	14008006	5095	1953-2007
REGENSTAUF	15208500	2468	1968-2002
DEGGENDORF	15993001	36.12	1951-2013
SCHLEHDORF	16404900	352.2	1951-2011
GRAFRATH	16603000	1187	1956-2009
INKOFEN	16607001	2814	1968-2002
OBERAMMERGAU	16610709	109.1	1951-2013
PEI??FENBERG	16612001	286.1	1951-2011
OBERHAUSEN	16618008	112.8	1956-2009
BERGKIRCHEN	16658002	217.8	1953-2011
ROTTERS DORF	17204204	723.6	1952-2010
VILSBIBURG	17215007	518.9	1952-2011
DIETELSKIRCHEN	17225000	143.2	1956-2009
KALTENECK	17406005	834.3	1951-2013
BAD TRI??FL	18193001	38.72	1951-2011
STEPHANSKIRCHEN	18199008	100.9	1951-2010
FELDOLLING	18204006	494.8	1954-2010
ENGFURT	18384509	573.1	1951-2012
PRIEN	18463004	81.87	1952-2010
FRITZ AM SAND	18484503	106.9	1951-2011
ILSANK	18622006	126.2	1951-2011

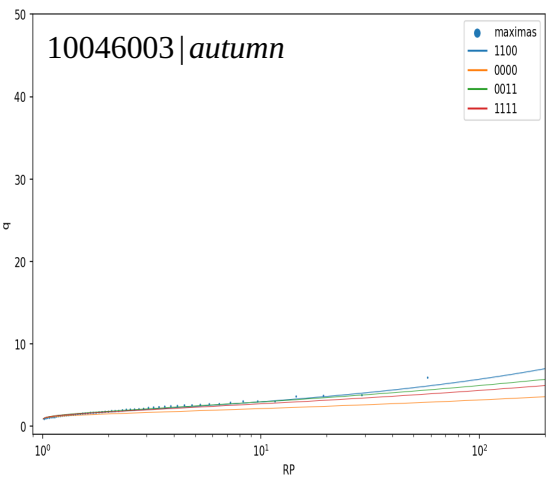
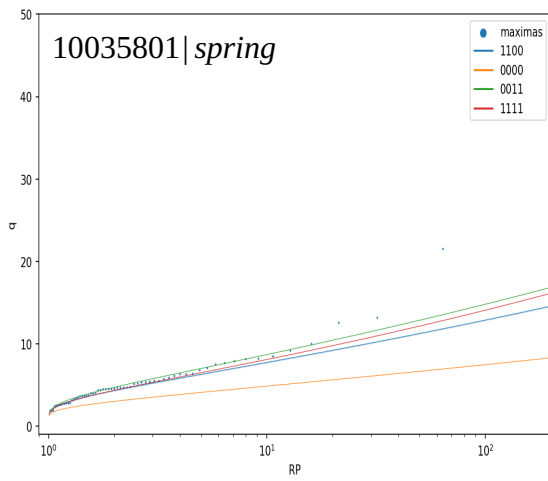
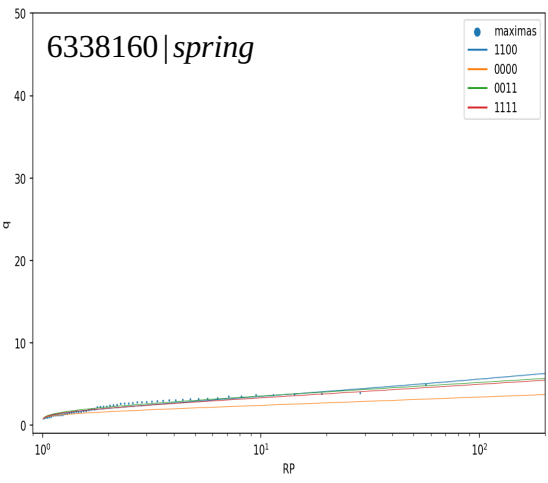
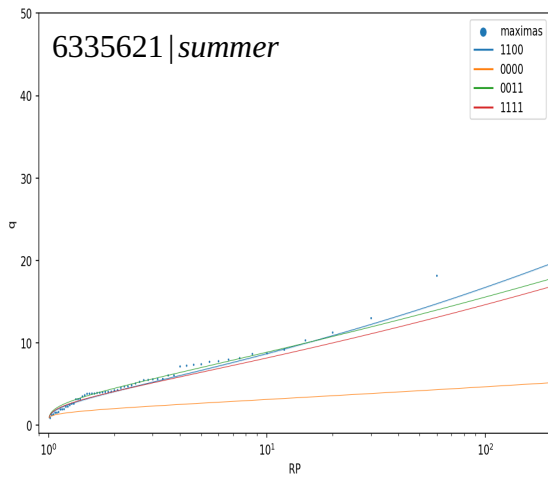
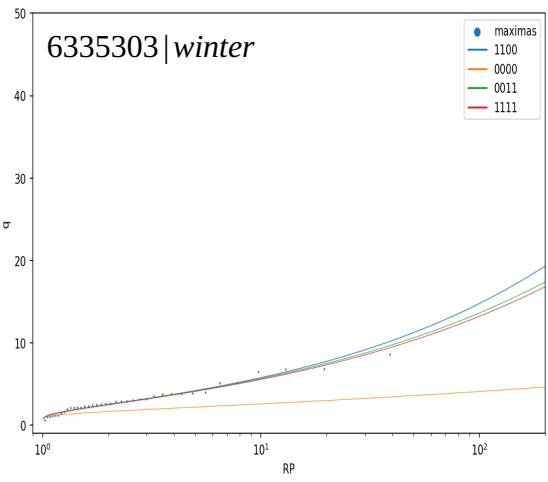
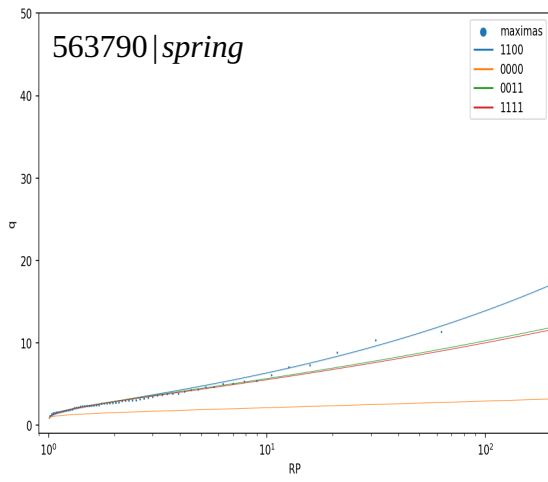
TEISENDORF	18662000	53.56	1951-2011
UNVERZUG	18683000	137.2	1951-2013
KINNING	18801005	152.2	1951-2011
LINDEN	18803805	528.9	1952-2008
BIRNBACH	18806406	856.8	1952-2010
RUHSTORF	18808004	936.8	1951-2009
PANZING	18825003	114.2	1951-2011
SIEBELDINGEN	23770507	204.7	1956-2009
PLOCHINGEN	23800100	3988	1951-2011
LAUFFEN	23800500	7900	1962-2011
LORSCH	23942300	413	1951-2010
NAUHEIM	23980353	170.9	1951-2013
KEMMERN	24010004	4236	1951-2011
STEINBACH	24050009	17890	1954-2009
KLEINHEUBACH	24064003	21500	1968-2001
UNTERLANGENSTADT	24143008	714.7	1951-2013
PETTSTADT	24209004	6981	1951-2010
HANAU	24784259	925	1951-2010
ILBENSTADT	24850058	1076	1951-2008
EPPSTEIN	24960307	110	1951-2013
BOOS	25400600	2835	1953-2010
GROLSHEIM	25400804	4012	1951-2013
ALTENBAMBERG	25480304	319.2	1951-2011
MARBURG	25830056	1663	1952-2010
ASSLAR	25842500	694.3	1951-2011
NETTEGUT	27140500	367.6	1951-2009
EITORF	27259100	1473	1951-2011
MENDEN 1	27291000	2829	1956-2009
MESCHEDE	27615100	96.11	1951-2011
VILLIGST	27655900	2027	1951-2011
HATTINGEN	27695100	4121	1951-2011
BENTFELD	27833900	1025	1951-2011
KESSELER 3	27857900	2050	1968-2001
HALTERN	27891000	4294	1951-2013
SCHERMBECK 1	27897700	4836	1951-2011
STAH	28291000	1778	1951-2008
GREVEN	33300101	2557	1951-2011
DALUM	35500407	4848	1951-2010
HELDRA	41700105	4271	1951-2013
LETZTER HELLER	41900206	5475	1951-2010
KAEMMERZELL	42350057	556.6	1958-2013
HERMANNSPIEGEL	42670557	419.2	1951-2013
ROTENBURG	42700100	2523	1953-2010
GREBENAU	42700202	2977	1951-2013
SCHMITTLOTHEIM	42800309	1195	1951-2011
UTTERSHAUSEN	42882806	988.5	1951-2012
GUNTERSHAUSEN	42900100	6367	1969-2001
HANN.-MUENDEN	43100109	12440	1951-2013
WAHMBECK	43900105	13000	1951-2012
KARLSHAFEN	45100100	14780	1952-2010
BODENWERDER	45300200	15930	1956-2009
VLOTHO	45900208	17620	1956-2011
HERFORD	46510000	874.3	1952-2008
PORTA	47100100	19140	1951-2011
LIEBENAU	47500200	21180	1956-2009
CELLE	48300105	4925	1956-2011
MARKLENDORF	48700103	6168	1951-2009
HERRENHAUSEN	48800108	5301	1951-2011
SCHWARMBSTEDT	48800301	6741	1954-2011
RETHEM	48900204	14180	1951-2011
HOF	56001502	460.6	1951-2011

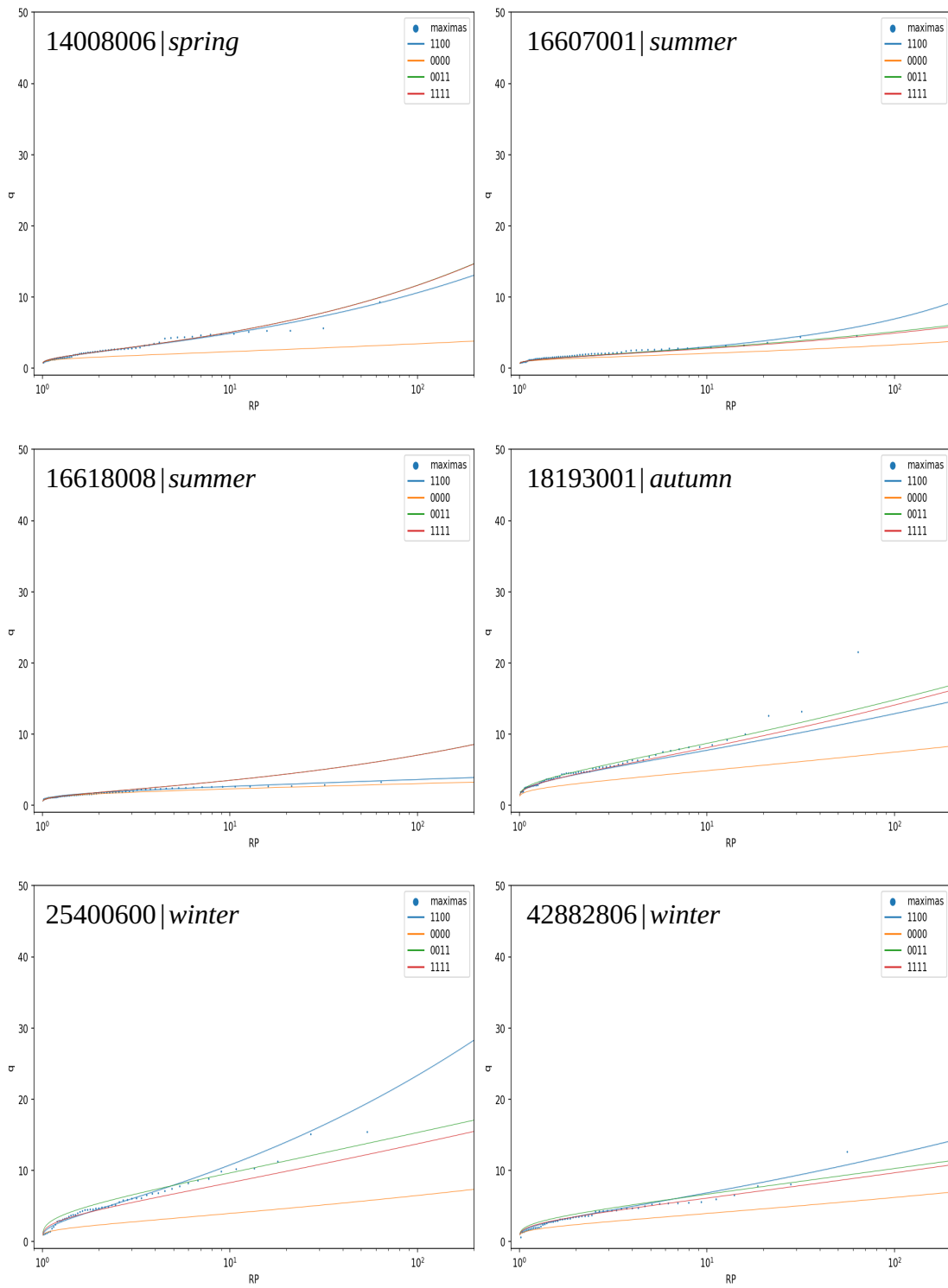


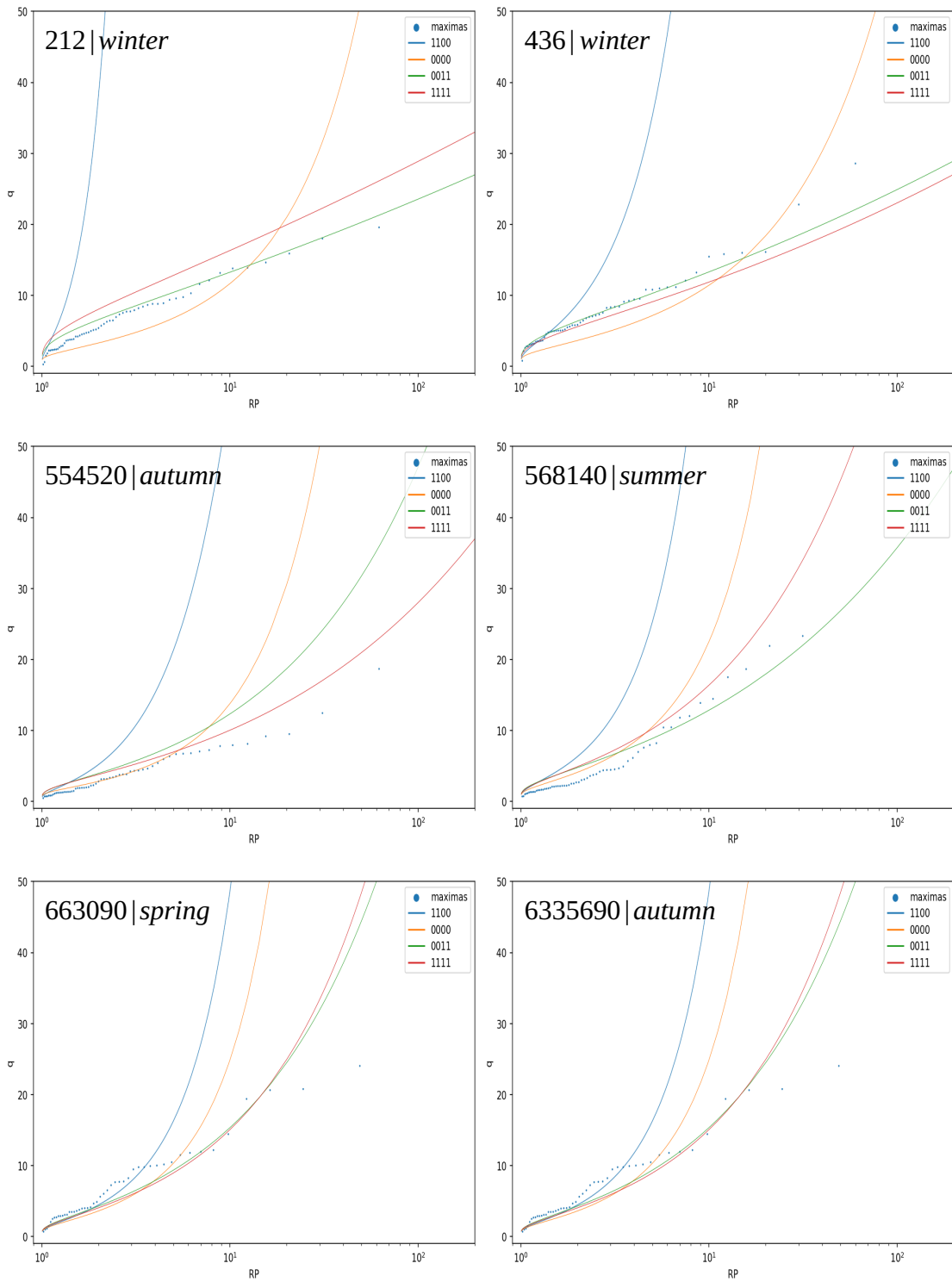
## APPENDICE B

In questa appendice si riportano i grafici delle curve di possibilità di inondazione per alcuni campioni significativi di casi studio.

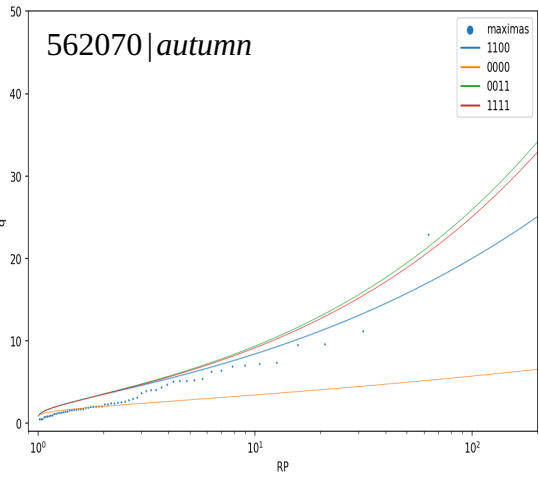
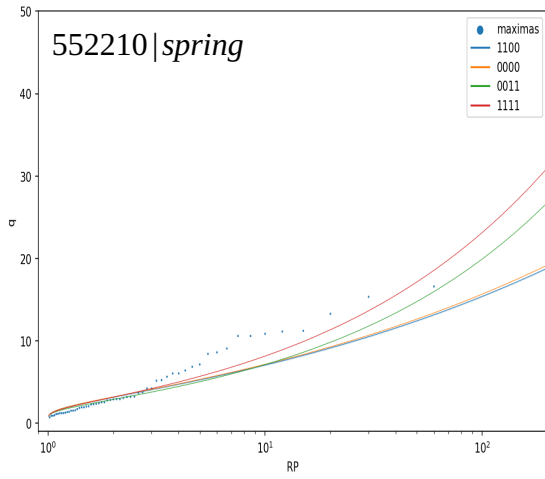
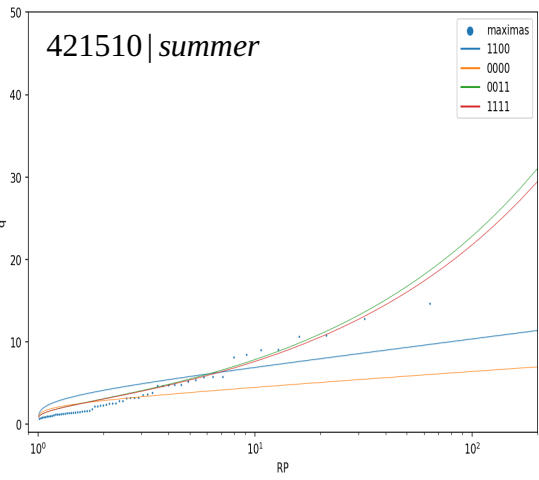
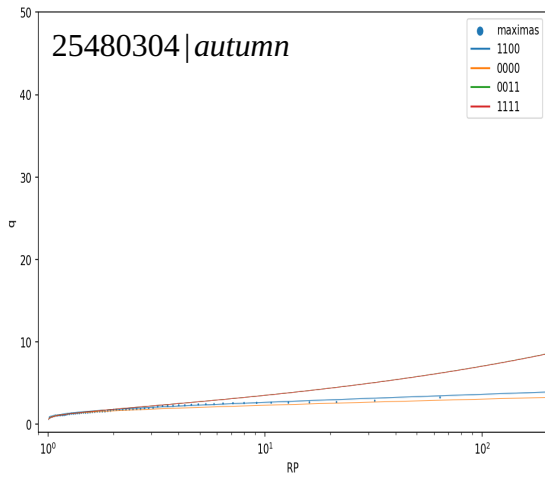
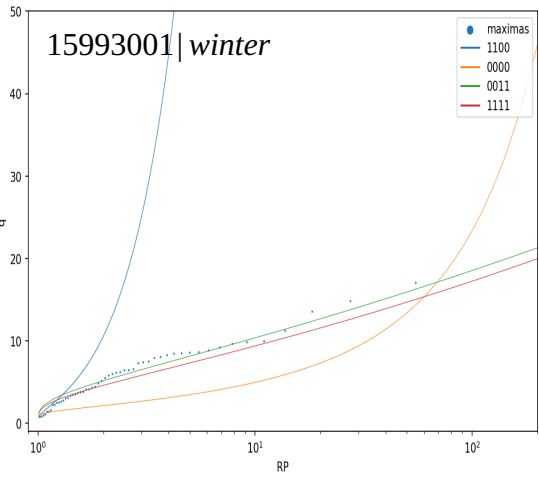
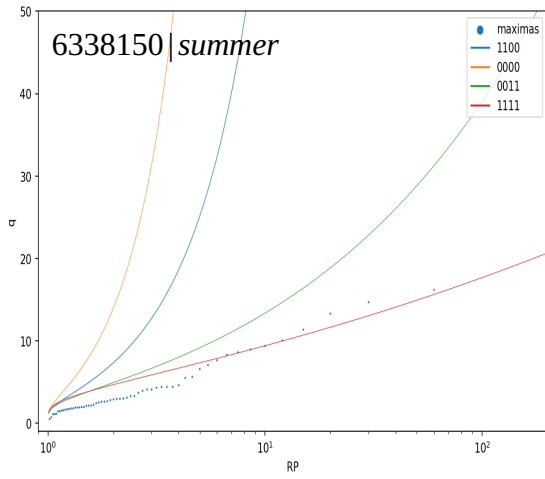


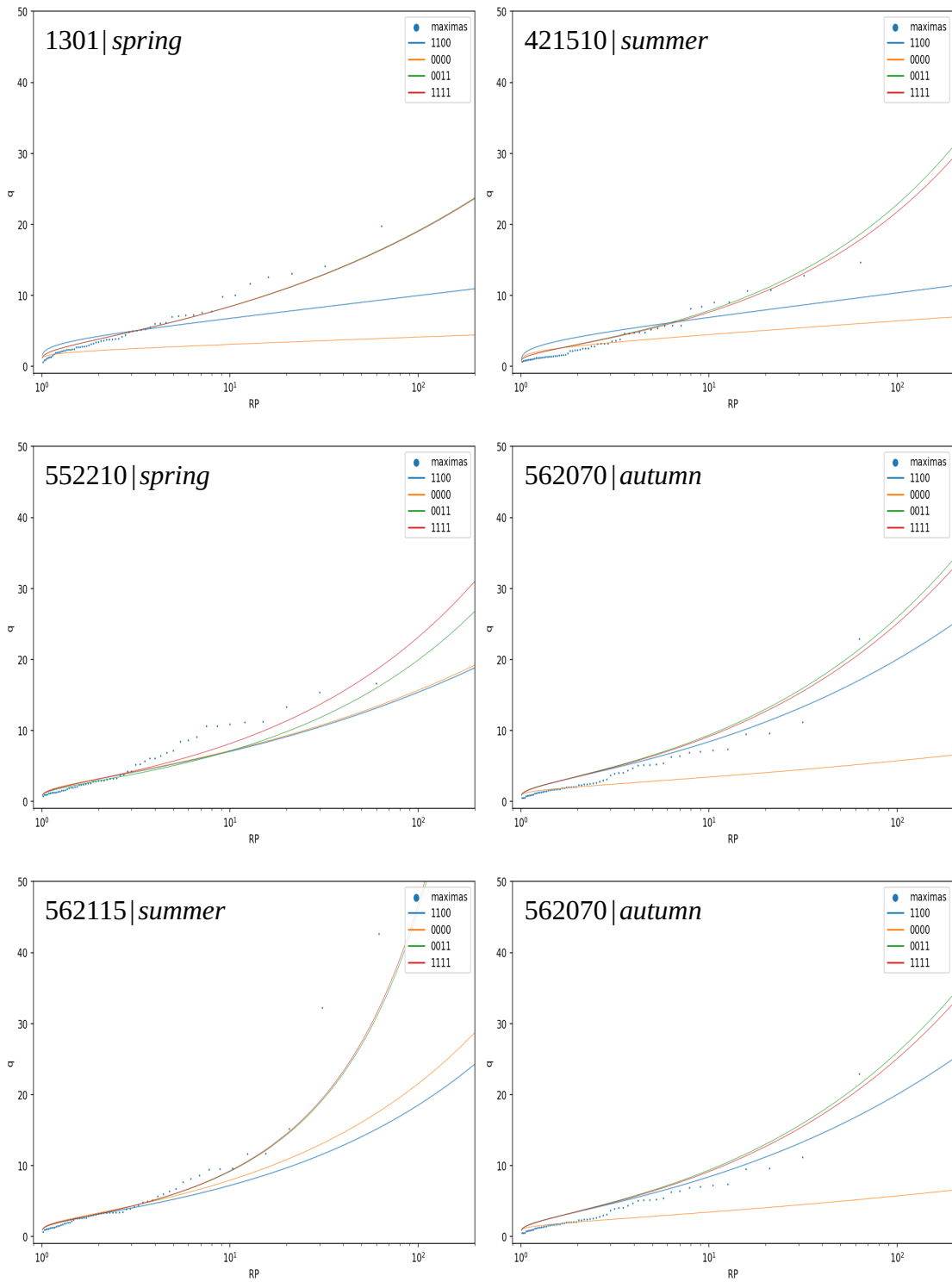


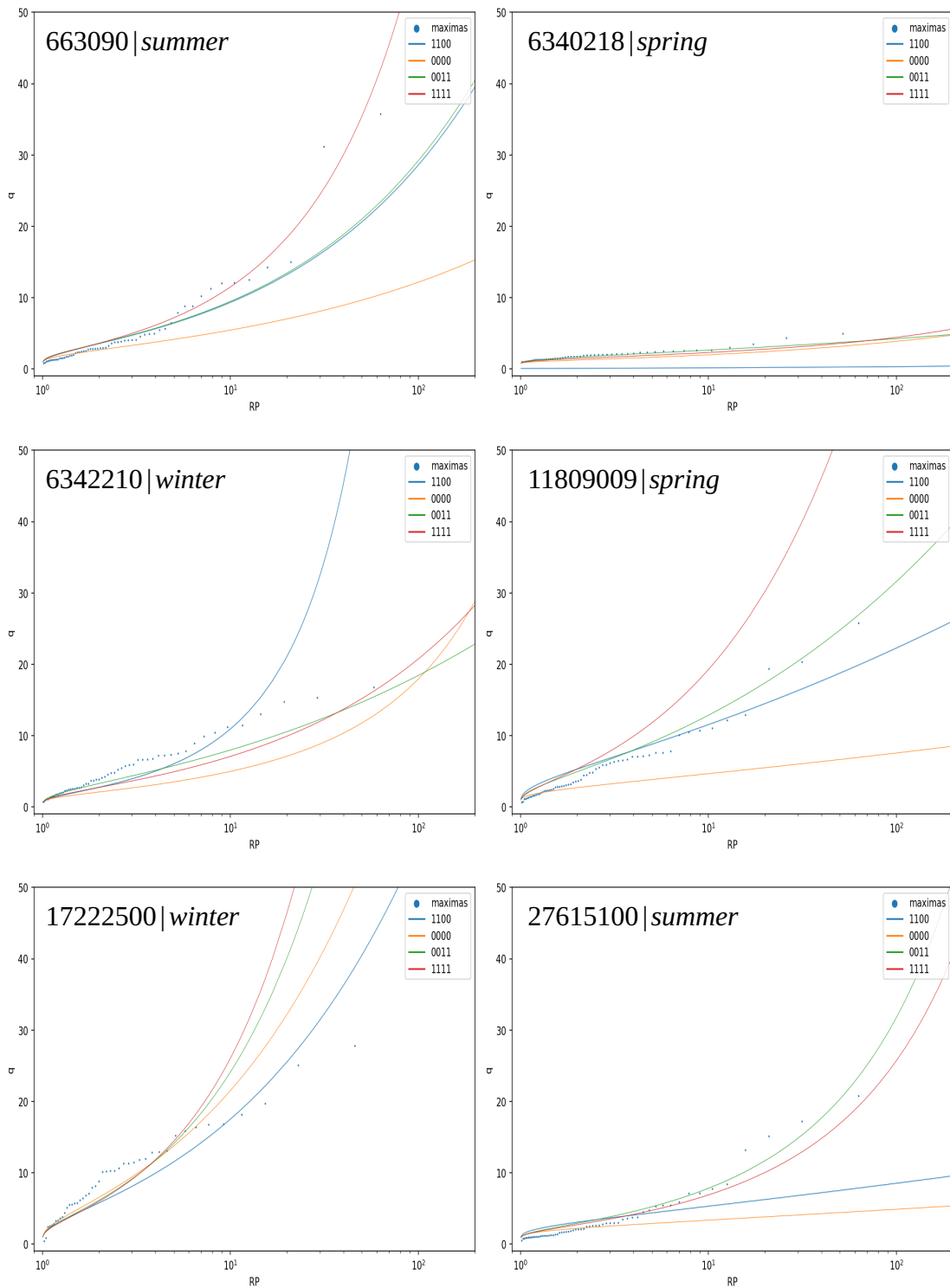














## BIBLIOGRAFIA

1. Coles S. (2001), *An Introduction to Statistical Modeling of Extreme Values*, Springer Series in Statistics, Springer, London.
2. Coles, S., Tawn J. (1990), Statistics of coastal flood prevention, *Philosophical Transactions of the Royal Society*, 332, 457–476.
3. Cramer H. (1946), *Mathematical methods of statistics*, Princeton University Press, Princeton (NJ).
4. Cramer H., Leadbetter M. F. (1967), *Stationary and related stochastic processes*, John Wiley, New York.
5. Gumbel, E. (1958), *Statistics of Extremes*, Columbia University Press, New York.
6. Resnick S. (2006), *Heavy-Tailed Phenomena: Probabilistic and Statistical Modeling*, Springer-Verlag, New York.
7. Katz, R., Parlange M., Naveau P. (2002), Statistics of extremes in hydrology, *Advances in Water Resources*, 25, 1287–1304.
8. Martins, E., Stedinger J. (2000), *Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data*, *Water Resources Research*, 36, 737–744.
9. Ignaccolo M., Marani M. (2015), *The Metastatistics of Daily Rainfall extremes: Non-asymptotic and Non stationary Extreme Value Analysis*, *Advances in Water Resources* 79, 121-126.
10. Von Mises R. (1936), *La distribution de la plus grande de n valeurs*, *Rev. math. Union interbalcanique*, 1.
11. Wilson P.S., Tuomi R. (2005), *A fundamental probability distribution for heavy rainfall*, *Geophysical Research Letters* 32(14): 1-4.
12. Kobierska, F., Engeland, K., Thorarinsdottir, T. (2018), *Evaluation of design flood estimates – a case study for Norway*, *Hydrology research*, 49 (2)

13. Marani M. (2017), *Processi e modelli dell'idrometeorologia: Un' introduzione*, DICEA, Universita' degli studi di Padova.
14. Scarrott, C., Mac Donald, A. (2012), *A review of extreme value thresholds estimation and uncertainty quantification*, *Statistical Journal* Vol. 10, 33-60.
15. Basso S., Lazzaro G., Bovo M., Soulsby C., Botter G. (2015), *Water-energy-ecosystem nexus in small run-of-river hydropower: Optimal design and policy*, *Applied Energy* 280, 115936.
16. Pattyn F., Huele W. V. (1998), *A Power law or power flow?*, *Earth Surface Processes and Landforms*, Vol. 3, Issue 8, 761-767.
17. Dralle D. N., Karst N. J., Charalampous K., Veenstra A., Thompson S. E. (2017), *Event-scale power law recession analysis: quantifying methodological uncertainty*, *Hydrology and Earth System Sciences*, 21, 65–81.
18. Mushtaq S., Miniussi A., Merz R., Basso S. (2020), *Reliable estimation of high floods: a method to select the most suitable ordinary distribution in the Metastatistical Extreme Value framework*, *Advances in Water Resources* 161.
19. Miniussi A., Marani M., Villarini G. (2020), *Metastatistical Extreme Value Distribution applied to floods across the continental United States*, *Advances in Water Resources* 136.
20. Basso S., Botter G., Merz R., Miniussi A. (2021), *PHEV! The PHysically-based Extreme Value distribution of river flows*, *Environmental Research Letters* 16.
21. Tarasova L., Basso S., Poncelet C., Merz R. (2018), *Exploring Controls on Rainfall-Runoff Events: 2. Regional Patterns and Spatial Controls of Event Characteristics in Germany*, *Water Resources Research*, Vol 54, 7688–7710.
22. Miniussi A., Kaule L., Merz R., Basso S. (2022), *Metastatistical Extreme Value Distribution applied to floods across the continental United States*, submitted to *Journal of Hydrology*.

23. Basso S., Schirmer M., Botter G. (2015), *On the emergence of heavy-tailed streamflow distributions*, *Advances in Water Resources* 82, 98-105.
24. Kirchner J. W. (2009), *Catchments as simple dynamical systems: Catchment characterization, rainfall-runoff modeling, and doing hydrology backward*, *Water Resources Research.*, Vol. 45, No. 2.
25. Biswal, B., Marani, M. (2010), *Geomorphological origin of recession curves*, *Geophysical Research Letters* 37.
26. Mutzner, R., Bertuzzo, E., Tarolli, P., Weijs, S. V., Nicotina, L., Ceola, S., Tomasic, N., Rodríguez-Iturbe, I., Parlange, M. B., and Rinaldo, A. (2013), *Geomorphic signatures on Brutsaert base flow recession analysis*, *Water Resources Research*, No. 49, 5462–5472.
27. Zorzetto E., Botter G., Marani M. (2016), *On the emergence of rainfall extremes from ordinary events*, *Geophysical Research Letters* 32(14): 1-4.
28. Lamb R., Beven K. (1997) *Using interactive recession curve analysis to specify a general catchment storage model*, *Hydrology and Earth System Sciences*, No. 1, 101–113.
29. Morrison J.E., Smith J.A. (2002), *Stochastic modeling of flood peaks using the generalized extreme value distribution*, *Water Resources Research*, Vol. 38, No. 12.
30. Marra, F., Nikolopoulos, E. I., Anagnostou, E. N., & Morin, E. (2018), *Metastatistical Extreme Value analysis of hourly rainfall from short records: Estimation of high quantiles and impact of measurement errors*, *Advances in Water Resources*, 117, 27–39.





## RINGRAZIAMENTI

La scrittura di queste righe è per me un momento davvero emozionante e lungamente atteso, da solo sufficiente a giustificare l'intero lavoro svolto in quest'ultimo anno.

Voglio ringraziare la mia famiglia tutta, che mi ha supportato, sostenuto e stimolato, costantemente ed incondizionatamente, lasciandomi libero di fare di testa mia e seguire le passioni, coi modi e coi tempi che queste comportavano. I mesi di permanenza all'estero sono stati un periodo denso di sfide ed avventure uniche non per me solamente: per mio padre, che preparandosi al viaggio d'oltralpe ha rispolverato il suo fluente inglese e poco è mancato che imparasse il tedesco, di cui comunque conosce più parole di quante non ne abbia apprese io; per mia madre, informatrice costante di ciò che avveniva in terra natia e destinataria delle testimonianze delle mie audaci sperimentazioni culinarie; per mia sorella, che c'è stata dal principio al termine e ha viaggiato in lungo ed in largo a bordo dei treni di mezza Europa; per mia nonna, che ha pazientemente atteso la fine di un lungo percorso e mi ha sempre amato con un figlio.

Un grazie va all'Helmoltz Centre for Enviromental Research (UFZ) ed in particolare al dipartimento di idrologia di Halle, che ha messo a disposizione gli strumenti per permettere le mie analisi, dove ho trovato un ambiente energetico, stimolante e ricco di opportunità, e soprattutto un gruppo di persone con cui condividere esperienze che vanno ben al di là dell'attività di ricerca. Ringrazio Felipe Saavedra, complice di scorribande alla macchinetta del caffè, fonte di ispirazione, compagno di avventure, amico sincero. Ringrazio Sumra Mushtaq, che dal suo ufficio ha sopportato la mia musica ed il mio cantare, regalandomi ogni giorno la voglia di sorridere e testimoniando i momenti assieme con sua inarrivabile passione per la fotografia.

Ringrazio Stefano Basso, che chiamare supervisore sarebbe un eufemismo riduttivo, per avermi guidato, supportato e spronato in tutto il lavoro, lasciandomi grande libertà e permettendomi di arrivare al termine con le mie forze.

Ringrazio il professor Marani per avermi tramesso coi suoi insegnamenti l'interesse per la materia, per aver personalmente reso possibile questa opportunità, e per avermi pedissequamente seguito nell'intensa e frenetica staffetta di stesura della tesi.

