

Università degli studi di Padova
Dipartimento di Scienze Economiche e Aziendali "M. Fanno"

Corso di Laurea Triennale in
Economia



PROVA FINALE

STATISTICA E CLIMATOLOGIA

Relatrice: Prof.ssa Luisa Bisaglia
Dipartimento di Scienze Statistiche

Laureando: Alberto Calore
Matricola 1188599

Anno Accademico 2020/2021

*Ignoranti quem portum petat
nullus suus ventus est*

Lucius Annaeus Seneca

Indice

Introduzione	7
1 <i>I Dati Statistici</i> in ambito climatico	9
2 Analisi degli eventi estremi	17
3 Applicazioni	27
Bibliografia e Sitografia	37
Ringraziamenti	41

Introduzione

La Climatologia, per sua natura, fonda buona parte del suo progresso sulla diretta osservazione. Questa disciplina ha riscontrato molta considerazione tra la popolazione generale negli ultimi anni, dal momento che il tema del cambiamento climatico ha iniziato a coinvolgere il singolo individuo. La sua natura fortemente osservazionale la lega a doppio filo con la Statistica, i cui strumenti sono utilizzati per analizzare la mole di dati ad oggi in nostro possesso. La complessità del sistema clima si rivela però un'ardua sfida anche per le tecniche di raccolta di dati e di analisi moderne.

In questo lavoro verranno affrontate dapprima le fonti dell'informazione climatica, sia quelle relative al clima corrente che quelle paleoclimatiche, entrambe fondamentali per comprendere lo stato attuale del sistema. I dati raccolti prendono la forma di serie storiche generate da un processo stocastico. I concetti sopra menzionati verranno quindi approfonditi nel capitolo primo e si ripresenteranno anche nei successivi.

Successivamente nel capitolo secondo si tratteranno gli eventi estremi, manifestazioni violente sempre più frequenti a causa del cambiamento climatico, e i processi utilizzati per studiarne l'incidenza.

Infine, nel capitolo terzo, questi strumenti troveranno applicazione nel caso studio delle ondate di caldo in Europa, che hanno trovato culmine della disastrosa estate 2003.

Capitolo 1

I Dati Statistici in ambito climatico

"Il clima di un luogo è dato dalla funzione di distribuzione della probabilità che a un dato istante si osservi un determinato tempo meteorologico" (Treccani 2021). Possiamo interpretare questa definizione come una formalizzazione statistico-matematica dell'idea di clima nel senso comune: "Generalmente fa caldo", "Solitamente piove molto" sono frasi che riferite ad un determinato luogo riassumono la nostra conoscenza sul tempo meteorologico che sia lecito aspettarsi e su quello invece che sia insolito. Queste congetture affondano le loro radici nell'esperienza, diretta ed indiretta, degli eventi atmosferici che si sono manifestati in passato e che tendiamo a sintetizzare nel concetto di clima. La nostra conoscenza del clima passa quindi inevitabilmente attraverso un ricorso ai dati raccolti in passato non essendo quindi possibile determinare a priori la probabilità di un determinato evento meteorologico. Ad esempio l'Organizzazione Meteorologica Mondiale (OMM) raccomanda di studiare i dati giornalieri in un intervallo di 30 anni ai fini di prevedere la probabilità di un determinato evento atmosferico: non si potranno ottenere i valori assoluti delle probabilità, ma si potrà determinare cosa è lecito aspettarsi dal futuro e cosa invece rappresenta un evento inaspettato (Treccani 2007). Trent'anni di dati per delle "semplici" previsioni meteorologiche potrebbero sembrare una mole più che sufficiente e con le continue e accurate rilevazioni proprie di questo periodo quasi eccessiva: ma quanto spesso queste poi si rivelano inaccurate? Inoltre, lo scenario che ho presentato a fini esplicativi richiede

solo un'infinitesima parte dei dati necessari per comprendere problemi ancora più complessi, come lo stato attuale del sistema climatico, per cui si rende necessario estrapolare informazioni risalenti a decine di migliaia di anni fa.

Il clima è un sistema complesso e interagisce su scala globale e la stessa fondazione dell'OMM ne è testimonianza. L'Organizzazione Meteorologica Mondiale vede la sua costituzione nel 1947 con il fine di individuare e prevedere tutti quegli eventi atmosferici che hanno diretta influenza sulle attività umane e si pone al centro della collaborazione a questo fine. La costituzione dell'OMM presenta rilevanti ripercussioni storiche per il periodo in cui si colloca, il secondo dopoguerra: difatti si presenta come il primo rapporto di collaborazione scientifica tra i blocchi antagonisti emersi dopo la fine della Seconda Guerra Mondiale, il blocco sovietico e quello occidentale. Negli anni Cinquanta in piena guerra fredda viene effettivamente avviata la collaborazione attraverso il programma World Weather Watch (WWW), le cui finalità erano quelle di rendere disponibili osservazioni, temporalmente e spazialmente dettagliate, su scala globale ed in tempi congrui per utilizzarle a fini previsivi. Inoltre, la raccolta dei dati viene resa sostanzialmente ancor più di fondamentale importanza dalla non riproducibilità degli eventi atmosferici in laboratorio: un temporale di minore entità ad esempio prende atto "in volumi di chilometri cubi con scambi di energie valutabili in migliaia di miliardi di joule (10¹² J) e confrontabili con quelle delle esplosioni nucleari, ma che fortunatamente avvengono a temperature prossime a quella dell'ambiente" (Treccani 2007).

L'informazione climatica proviene da fonti molto diverse, in funzione anche della distanza temporale dal presente. La prima fonte è rappresentata dai dati meteorologici e oceanici misurati direttamente da strumenti specifici: è la fonte più accurata e diretta di conoscenza che si possa ottenere, si caratterizza per la sua connotazione globale e la sua raccolta è iniziata un centinaio di anni fa. I dati sono raccolti dal suolo fino ad almeno un'altitudine di trenta chilometri di quota, attraverso stazioni collocate in situ (dove possibile), attraverso reti di boe e attraverso costellazioni di satelliti, questi ultimi capaci anche di ovviare al problema di effettuare osservazioni nelle zone più complesse da monitorare (parti non ancora raggiunte da boe negli oceani o deserti ad esempio). Le stazioni terrestri sono

solite rilevare i parametri più generici, come la temperatura dell'aria, la pressione atmosferica e l'umidità, direzione ed intensità del vento. Alcune di queste stazioni inoltre effettuano misurazioni legate non solo al tempo meteorologico ma anche al clima e all'ambiente in generale, come ad esempio la concentrazione di CO₂ o altri gas serra. Un importante esempio di rete di boe è rappresentato dall'Argo Network, composto da 3000 dispositivi collocati negli oceani secondo una griglia indicativamente di circa 3° x 3° in latitudine / longitudine. Con cadenza di 10 giorni ogni boa misura pressione, temperatura e salinità lungo una colonna d'acqua di 2000 metri. Questa rete fornisce le caratteristiche di 100.000 sezioni verticali di oceano ogni anno, con una risoluzione che può essere, nelle boe di più recente installazione, di 2 metri (Argo 2000, Riser et al. 2016). Esempi moderni invece di satelliti meteorologici sono GOES-15 e GOES-16 lanciati dall'U.S. National Oceanic and Atmospheric Administration in collaborazione con la NASA, appartenenti alla categoria dei satelliti geostazionari cioè che osservano sempre la stessa porzione di Terra (categoria differente è rappresentata dai satelliti elisincroni, che sorvolano invece un territorio sempre alla stessa ora) (<https://www.ospo.noaa.gov/>).

Una seconda fonte proviene dall'interpretazione della documentazione storica che temporalmente ci permette di ricavare informazioni relative agli ultimi millenni: eventi come le esondazioni del fiume Elba (che attraversa il Nord Europa) sono ben documentate anche nel periodo pre-strumentale, cioè precedentemente al sedicesimo secolo, e la possibilità di perdere informazioni è minima, specialmente per gli eventi di maggiore entità (Mudelsee et. al. 2004).

Una terza fonte è rappresentata dai dati detti Proxi, "denominazione di dati relativi a epoche remote a partire da quella attuale, dedotti indirettamente" (Treccani 2008). Un esempio immediato per il lettore potrebbe essere l'analisi degli anelli arborei, detta dendrologia: le prime considerazioni si possono trarre dallo spessore degli anelli che permettono di comprendere l'andamento meteorologico generale dell'anno in cui si sono generati, ma analisi chimico-fisiche possono portare anche ad una comprensione più approfondita del clima in senso lato. Queste ultime comprendono analisi isotopiche di ossigeno, deuterio e carbonio, metodologia applicabile non solo alla dendrologia, ma anche a sedimenti oceanici, calotte polari,

stalattiti, strati geologici..., in quanto differenti composizioni corrispondono a differenti periodi climatici. Possono essere Proxi, ad esempio, anche le informazioni su fauna e flora del passato, i terrazzi marini (che indicano il massimo livello dei mari nei periodi interglaciali), le morene (che forniscono informazioni sulla massima estensione glaciale), il paleomagnetismo, *et cetera rerum*. Questa categoria di informazioni richiede un'intensa elaborazione e al fine di migliorarne l'attendibilità occorre che non vengano analizzate singolarmente.

La mole di dati in nostro possesso e le sue plurime fonti potrebbero sembrare sufficienti a comprendere lo stato attuale del clima e ad affrontare il problema contemporaneo del cambiamento climatico ma purtroppo la complessità del sistema clima in quanto risultato dell'interazione tra vari elementi (atmosfera, radiazione solare, idrosfera, litosfera e biosfera) secondo innumerevoli regole fisiche, chimiche e biologiche ci pone nella posizione di aver solo scalfito la superficie di ciò che sarebbe necessario sapere. Edward N. Lorenz, meteorologo statunitense attivo nella seconda metà del XX secolo, aveva ribadito attraverso i propri studi l'impossibilità di utilizzare metodi deterministici per prevedere l'evoluzione dell'atmosfera: questo non solo in relazione all'insormontabile complessità del sistema clima, ma anche all'incredibile sensibilità alle condizioni iniziali. A riguardo, il luminare introduce la teoria del cosiddetto "effetto farfalla": può il battito d'ali di una farfalla in Brasile provocare un ciclone in Texas? Per quanto possa risultare un'affermazione iperbolica si presenta in linea con la Teoria del Caos (Lorenz 1963).

Due concetti che inevitabilmente si incontreranno in studi climatici sono quello di "serie storica", nell'affrontare le fonti, e quello di "processo stocastico", in contrapposizione all'idea di processo deterministico non adatta al sistema clima.

Una serie storica può essere definita come un insieme di osservazioni effettuate in successione nel tempo. Questa può essere continua se la misurazione è continua nel tempo, o discreta se le osservazioni sono effettuate solo in determinati momenti, generalmente facendo intercorrere intervalli regolari (giorni, mesi, anni...). Se è possibile prevedere una serie storica attraverso le misurazioni passate questa viene detta deterministica, se i valori futuri sono solo parzialmente determinati dal passato è detta stocastica; in quest'ultimo caso le previsioni esatte verranno sostituite

dall'idea che i valori assumeranno una distribuzione di probabilità condizionata ai valori passati. L'analisi delle serie storiche può presentare diverse finalità quali descrivere, spiegare e prevedere. L'analisi descrittiva passa attraverso la rappresentazione dei dati tramite grafici e il calcolo di misure dette, appunto, descrittive (il calcolo della media ne è un esempio) con il fine di cogliere alcune proprietà della serie come stagionalità o stazionarietà, individuarne punti di svolta che potrebbero suggerire la suddivisione in sottoinsiemi omogenei; ed è anche possibile in questa fase effettuare trasformazioni dei dati. Tra queste operazioni è caratteristico delle serie storiche il calcolo del coefficiente di autocorrelazione ($r_k = \frac{\sum_{t=1}^{N-k}(x_t - \bar{x})(x_{t+k} - \bar{x})}{\sum_{t=1}^{N-k}(x_t - \bar{x})^2}$) e la sua rappresentazione al variare di k nel cosiddetto correlogramma. L'analisi esplicativa si pone obiettivo di investigare il meccanismo che ha originato la serie e di sintetizzarlo attraverso un processo stocastico e inoltre di individuare eventuali nessi correlativi o causali. Infine, l'analisi previsiva si prefigge di dare uno sguardo al futuro, attraverso proiezioni che generalmente assumono che i fattori determinanti dell'evoluzione passata agiranno anche in futuro. Un'ulteriore suddivisione può essere effettuata tra analisi univariata, se si descrive, spiega o prevede un singola serie, o multivariata se si cerca una relazione tra serie diverse (Chatfield 2003).

Altri concetti chiave relativi alle serie storiche sono il concetto di "trend" e il concetto di "stagionalità". Il trend può essere definito come: "Nel linguaggio statistico [...] qualsiasi tendenza di lungo periodo" (Chatfield 2003), ne è un esempio il surriscaldamento globale e l'incremento graduale delle temperature che sta comportando. La stagionalità, invece, può essere definita come: "Proprietà di alcuni fenomeni osservati nel tempo, generalmente modellati da una serie storica [...], che si evidenzia sotto forma di fluttuazioni a cadenza periodica più o meno regolare" (Chatfield 2003), concetto che non richiede ulteriori esempi.

Un processo stocastico, invece, può essere descritto come un fenomeno statistico che evolve nel tempo secondo leggi probabilistiche. Esempi ben noti sono la lunghezza di una lista d'attesa, le dimensioni di una colonia batterica e la temperatura dell'aria in giorni successivi in un determinato luogo (Chatfield 2003). Nel linguaggio matematico un processo stocastico può essere definito come un insieme

di variabili casuali ordinate nel tempo e definite in un insieme di punti temporali che possono essere continui o discreti. Nell'analisi delle serie storiche, per quanto possa essere possibile variare la lunghezza del campione, è solitamente impossibile effettuare più di un'osservazione in un dato istante di tempo. Quindi avremo un singolo valore e una singola osservazione della variabile casuale al tempo t . Ciononostante potremmo considerare la serie storica osservata come una sola manifestazione finita dell'insieme delle infinite possibili serie storiche che potrebbero essere osservate. Ogni serie storica può quindi essere considerata come una possibile realizzazione finita del processo stocastico. Al fine di descrivere un processo stocastico è utile definirne i momenti, in particolare i momenti primo e secondo, rappresentati dalla media, dalla varianza e dalle autocovarianze; queste possono essere formalmente definite in un orizzonte temporale continuo (con simili definizioni in caso di tempo discreto) come segue.

La funzione media è definita come:

$$\mu(t) = E[X(t)]$$

La funzione varianza è definita come:

$$\sigma^2(t) = Var[X(t)]$$

La sola varianza non è sufficiente per specificare il momento secondo di una sequenza di variabili casuali e si rende quindi necessario specificare la funzione autocovarianza $\gamma(t_1, t_2)$ che non è altro che la covarianza di $X(t_1)$ con $X(t_2)$:

$$\gamma(t_1, t_2) = E\{[X(t_1) - \mu(t_1)][X(t_2) - \mu(t_2)]\}$$

Un'importante proprietà dei processi stocastici è la cosiddetta stazionarietà. Da un punto di vista intuitivo, un processo stocastico si dice stazionario la sua media e la sua variabilità non presentano variazioni sistematiche e la sua dinamica non presenta variazioni strettamente periodiche. Utilizzando un linguaggio più rigoroso

si deve innanzitutto distinguere tra stazionarietà in senso stretto (o strettamente stazionario) e stazionarietà di secondo ordine (o debolmente stazionario). Un processo stocastico è detto stazionario in senso stretto se la distribuzione congiunta di $X(t_1), \dots, X(t_n)$ è la stessa distribuzione congiunta di $X(t_1 + \tau), \dots, X(t_n + \tau)$ per ogni t_1, \dots, t_n, τ . In altre parole traslare l'origine di una quantità τ non ha effetto sulla distribuzione congiunta. Da questo consegue che i primi due momenti sono finiti, costanti e non dipendenti dal valore di t : $\mu(t) = \mu$ $\sigma^2(t) = \sigma^2$ (Chatfield 2003). Poiché la condizione di stazionarietà in senso è molto forte, in quanto impone dei vincoli sull'intera distribuzione del processo (e di conseguenza su tutti i momenti), una forma di stazionarietà meno restrittiva è quella di stazionarietà debole o stazionarietà di secondo ordine. Un processo stocastico è detto debolmente stazionario se: media μ e varianza σ^2 siono (come da notazione) costanti e finite al variare di t e la funzione di autocovarianza tra t e $t + \tau$ sia $\gamma(\tau)$, dipendente quindi solo dal ritardo τ e non da t . Il correlogramma assumerà sempre la stessa forma indipendentemente dal punto in cui si inizia a calcolarlo (Chatfield 2003).

Capitolo 2

Analisi degli eventi estremi

Cosa sono gli eventi estremi e perché ricoprono un ruolo così centrale nella climatologia (ma non solo)? Come si legge nel report dell'IPPC del 2012 si definisce evento estremo l'accadimento di un valore relativo ad un tempo atmosferico o una variabile climatica al di sopra (o al di sotto) di un valore limite collocato nella parte alta (o bassa) dell'insieme dei valori osservati della variabile. Ne sono un esempio delle piogge torrenziali, dei giorni molto freddi o molto caldi, dei forti venti, etc. Questi esempi risultano un po' incolore in quanto senza fonti, come i telegiornali che mettono l'accento sul valore in sé, all'attenzione comune non arrivano gli eventi estremi, per quanto non escludo comunque non vengano percepiti: quello che notiamo sono gli eventuali disastri comportati da questi valori anomali. Una forte pioggia potrebbe portare ad un'inondazione, dei giorni con temperature troppo basse o troppo alte potrebbero rovinare i raccolti, i forti venti potrebbero diventare tifoni ed uragani e così via. I disastri sono definiti come gravi alterazioni nel normale funzionamento di una comunità o della società dovuti a pericolosi eventi fisici che, interagendo con condizioni di vulnerabilità, conducono a vasti danni umani, materiali, economici o ambientali che richiedono un immediato intervento d'emergenza al fine di soddisfare bisogni primari e che possono richiedere supporto esterno per la ripresa (IPPC 2012). Dovrebbe essere quindi immediato comprendere l'importanza dell'analisi della componente estrema del clima anche e soprattutto al fine di prevedere. Per l'analisi degli eventi estremi un approc-

cio utilizzato è quello di scomporre la variabile casuale in differenti componenti. Il metodo di cui sopra trae origine dai modelli tradizionali di scomposizione delle serie storiche, la cui ipotesi di base è che il comportamento della serie sia il risultato della sovrapposizione di queste componenti, prefissate ma non osservabili singolarmente se non dopo averle isolate (Chatfield 2003). Un approccio di questo genere, adeguato al sistema clima, è quello della cosiddetta formulazione additiva, in cui ogni componente viene considerata come esprimersi in autonomia; il modello che verrà adottato in questo scritto è quello esposto da Mudelsee (2020):

$$X(T) = X_{trend}(T) + X_{ext}(T) + S(T) \cdot X_{noise}(T)$$

Il processo stocastico $X(T)$ (dove T rappresenta un orizzonte temporale continuo) viene scomposto nella componente $X_{trend}(T)$ che rappresenta il trend, nella componente $X_{ext}(T)$ che rappresenta gli eventi estremi ed infine $X_{noise}(T)$, la componente noise, scalata da una funzione di variabilità $S(T)$.

Il trend segue la definizione data nel capitolo precedente e quindi racchiude ogni tendenza di lungo periodo, sistematica o deterministica che sia, come ad esempio variazioni stagionali o incrementi lineari. Gli eventi estremi, al centro di questo studio, sono solitamente rari e con un grande valore assoluto. La componente di disturbo si assume debolmente stazionaria con media zero e assenza di autocorrelazione, inoltre, si assume presenti deviazione standard unitaria al fine di introdurre $S(T)$ per raggiungere una maggiore attinenza alla definizione di clima come non solo la media ma anche la variabilità dello stato dell'atmosfera e delle altre componenti (Bruckner 1890).

Sempre seguendo la notazione adottata da Mudelsee (2020), usando una notazione compatta dove $X(i) \equiv X(T(i))$, possiamo scrivere l'equazione climatica come:

$$X(i) = X_{trend}(i) + X_{ext}(i) + S(i) \cdot X_{noise}(i)$$

Deriva quindi che la notazione della serie storica discreta osservata come risultato del processo di $X(i)$ sia $\{t(i), x(i)\}_{i=1}^n$, un insieme di coppie di valori $t(i)$ e

$x(i)$.

L'analisi di $X_{ext}(T)$ richiederà quindi di isolare il processo dalle altre componenti di $X(T)$, problema statistico di una certa rilevanza: il trend dovrà essere stimato o se ne dovrà assumere l'assenza e successivamente si dovrà distinguere la componente estrema dalla componente di disturbo $S(T) \cdot X_{noise}(T)$. Il primo passo per inferire le proprietà della componente estrema richiede di identificare ed isolare dalla serie storica $\{t(i), x(i)\}_{i=1}^n$ i valori estremi, che formeranno un sottoinsieme formalmente descritto come $\{t_{ext}(j), x_{ext}(j)\}_{j=1}^m$.

Al fine di fare inferenza è necessario conoscere i valori temporali di quando si sono manifestati gli eventi estremi. Questi sono matematicamente definiti come:

$$\{T_{ext}(j)\}_{j=1}^m \{T(i) | X_{ext}(i) \neq 0\}_{i=1}^n$$

quindi come i valori temporali quando la componente estrema assume valori diversi da zero. Ulteriore condizione è che i valori siano indipendenti, cioè che gli estremi non dipendano da un evento precedentemente accaduto (Mudelsee 2020). All'atto pratico sono percorribili due strade per identificare ed isolare gli eventi estremi, denominate "Peak over threshold" e "Block extremes". Iniziando ad analizzare la prima, una traduzione che potrebbe rendere l'idea è "picchi che superano un limite": come intuibile vengono selezionati tutti gli eventi che superano un determinato limite prefissato. Questa tecnica si rivela particolarmente utile se l'intensità dell'evento è rilevante ai fini dell'identificazione, ma richiede che dei dati si conoscano sia i valori temporali che l'intensità. Matematicamente questo approccio si descrive come:

$$\{T_{ext}(j), X'_{ext}(j)\}_{j=1}^m = \{T(i), X(i) | X(i) > u\}_{i=1}^n$$

condizionandoli quindi a osservazioni $X(i)$ maggiori di una costante u . Nella scienza climatica però sarebbe più adeguato adottare un limite variabile in funzione del tempo, per tenere in considerazione l'andamento comportato dalla componente trend X_{trend} e dalla variabilità S . Anche in questo caso è fatta l'assunzione che i

valori siano mutualmente indipendenti.

La seconda modalità detta "Block extremes" come suggerisce il nome consiste nel dividere la serie storica in blocchi e prendere il valore massimo di ogni blocco. Nel caso della ricerca di eventi estremi caratterizzati da valori più piccoli o negativi è possibile prendere il minimo di ogni blocco o semplicemente moltiplicare tutte le osservazioni per -1 e prendere valore massimo. L'espressione matematica che sintetizza questo metodo (scritta nell'esempio di blocchi annuali) è:

$$X'_{ext}(j) = (\{X(i)\}|T(i) \text{ nel } j - \text{esimo anno della serie storica}),$$

$$T_{ext}(j) = j - \text{esimo anno della serie storica}$$

Al fine di far comprendere la logica alla base di questo processo verrà utilizzato lo stesso esempio utilizzato da Mudelsee (2020). Si consideri una variabile casuale e la sua funzione di densità di probabilità $f(x)$. Se si chiedesse la probabilità che X assuma un valore maggiore o uguale di un determinato limite u la risposta sarebbe il risultato dell'integrale definito $\int_u^\infty f(x)dx$. Ora si considerino due variabili X_1 e X_2 , tra loro indipendenti ed identicamente distribuite, e ci si ponga la domanda di quale sia la probabilità di osservare il massimo del blocco delle due variabili $\max(X_1, X_2)$ assumere un valore maggiore o uguale ad u . Estendendo lo stesso processo a k variabili indipendenti ed identicamente distribuite il massimo del blocco segue approssimativamente una distribuzione, detta Generalized Extreme Value, indipendentemente dalle distribuzioni delle singole variabili. Affinché l'approssimazione sia accurata il numero di variabili k deve essere sufficientemente grande: ad esempio i dati giornalieri lungo il periodo di un anno, quindi $k = 365$ o $k = 366$, forniscono una buona approssimazione, mentre utilizzare il periodo di un mese, quindi $k = 30$, sarebbe meno adeguato (Mudelsee 2020). Nel caso del clima, inoltre, viene a mancare l'indipendenza tra tutte le osservazioni, attraverso un fenomeno detto persistenza. Al fine di ovviare a questo problema sarà necessario ridurre il dataset ai soli dati indipendenti. Una volta estratti dalla serie storica $\{t(i), x(i)\}_{i=1}^n$ i valori estremi $\{t_{ext}(j), x_{ext}(j)\}_{j=1}^m$ si può procedere con l'analisi. Quest'ultima differisce sensibilmente se si considera il processo generatore

$X(t)$ come stazionario o come non stazionario, in quanto l'uno non si presenta come una mera estensione dell'altro. In caso di stazionarietà i due modelli utilizzati sono la distribuzione GEV per gli estremi raccolti attraverso un processo di "blocking" e la distribuzione GP per quelli raccolti tramite posizionamento di un limite. La teoria matematica sottostante è ben elaborata e pone le sue radici negli anni venti del secolo scorso. La condizione di stazionarietà è però poco adeguata allo studio del sistema climatico e del processo che lo genera. Per quanto possa sembrare logico adattare i due modelli appena citati alla presenza del cambiamento questo approccio comporta due problemi: il problema di ottimizzazione richiesto per adattare la curva ai dati risulterebbe molto complesso ed inoltre si presenterebbe un problema legato alla "ristretta forma funzionale della dipendenza temporale" (Mudelsee 2020). I processi di Poisson utilizzati al posto di GEV e GP si rivelano quindi più adeguati. I modelli citati verranno ora approfonditi con ordine. Si inizi con la distribuzione GEV, acronimo di "Generalized Extreme Value". Come detto precedentemente questa distribuzione è usata per i dati raccolti nella forma di block maxima e si presenta come segue:

$$F_{GEV}(x_{ext}) = \begin{cases} \exp\{-[1 + \xi(x_{ext} - \mu)/\sigma]^{-1/\xi}\} & (\xi \neq 0), \\ \exp\{-\exp[-(x_{ext} - \mu)/\sigma]\} & (\xi = 0), \end{cases}$$

dove $1 + \xi(x_{ext} - \mu)/\sigma > 0$, $-\infty < \mu < \infty$, $\sigma > 0$ e $-\infty < \xi < \infty$.

Il parametro μ identifica la posizione della distribuzione, mentre σ è il parametro di scala. Il parametro ξ determina invece il comportamento delle code (Figura 2.1). Il nostro obiettivo è quello di fare inferenza sui i parametri μ , σ e ξ partendo da un sottoinsieme m di dati raccolti sotto forma di block maxima, che annotiamo come $\{x_{ext}\}_{j=1}^m$. Il processo di inferenza statistica utilizzato è quello detto di massima verosimiglianza. La funzione che si vuole massimizzare è il logaritmo della funzione di verosimiglianza:

$$l(\mu, \sigma, \xi) = -m \ln(\sigma) - (1 + 1/\xi) \sum_{j=1}^m \ln[y(j)] - \sum_{j=1}^m y(j),$$

dove:

$$y(j) = 1 + \xi \left[\frac{x_{ext}(j) - \mu}{\sigma} \right]$$

Vengono fatte alcune assunzioni: che siano presenti sufficienti valori indipendenti per ogni blocco, i vincoli sui parametri precedentemente annotati e le cosiddette "condizioni di regolarità" (Coles 2001). Soddisfatte queste condizioni possono essere calcolati anche gli standard error dei valori stimati (Mudelsee 2020). Se gli eventi estremi sono invece stati selezionati attraverso il piazzamento di un limite dovranno essere analizzati attraverso l'utilizzo della distribuzione GP, definita come segue:

$$F_{GP}(x_{ext}) = \begin{cases} 1 - \{1 + \xi(x_{ext} - \mu)/[\sigma + \xi(u - \mu)]\}^{-1/\xi} & (\xi \neq 0), \\ 1 - \exp[-(x_{ext} - u)/\sigma] & (\xi = 0), \end{cases}$$

dove $(x_{ext} > u, \sigma > 0, \{1 + \xi(x_{ext} - \mu)/[\sigma + \xi(u - \mu)]\} > 0$ e $-\infty < \xi < \infty$.

Il modello è descritto anche in questo caso da tre parametri e il parametro ξ svolge lo stesso ruolo di descrivere il comportamento delle code (Figura 2.1). La stima di μ , σ e ξ avviene nuovamente attraverso il metodo della massima verosimiglianza sulla funzione logaritmica:

$$l(\mu, \tilde{\sigma}) = -m \ln(\tilde{\sigma}) - (1 + 1/\xi) \sum_{j=1}^m \ln[y(j)],$$

dove:

$$y(j) = 1 + \xi \left[\frac{x_{ext}(j) - \mu}{\tilde{\sigma}} \right]$$

e:

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

Ulteriori condizioni oltre a quelle imposte sopra sono che u non sia troppo piccolo in relazione al centro della della distribuzione iniziale e che $y(j) > 0 \forall j$. Come per la distribuzione GEV possono essere calcolati gli standard error dei parametri stimati (Mudelsee 2020). Supporre però che il sistema clima sia stazionario, oltre che indurre in errore, non ci permetterebbe di cogliere ciò che effettivamente stiamo cercando di studiare: il cambiamento climatico in sé e gli effetti che comporta sulla manifestazione degli eventi estremi.

Per questo genere di studi bisognerà come detto in precedenza abbandonare i modelli parametrici GEV e GP. Se provassimo ad espanderli ad esempio riscrivendo i tre parametri come funzioni lineari del tempo:

$$\mu(T) = \beta_0 + \beta_1 T$$

$$\sigma(T) = \exp(\gamma_0 + \gamma_1 T)$$

$$\xi(T) = \delta_0 + \delta_1 T$$

si presenterebbero due problemi: il primo è che i parametri da stimare sono ora sei invece che tre, il secondo è che in questo esempio si è supposta una relazione lineare con il tempo T che non è detto sia realistica (Mudelsee 2020).

Dovranno quindi essere utilizzati processi di Poisson, processi che in caso di stazionarietà presentano un solo parametro γ e che possono essere estesi in caso di assenza di stazionarietà stimando invece $\gamma(T)$. γ rappresenta il tasso di incidenza dell'evento estremo, cioè il numero di eventi indipendenti per unità di tempo. Al fine di stimare $\gamma(T)$ partendo da dati raccolti come $\{T_{ext}(j)\}_{j=1}^m$ è possibile utilizzare una stima di densità kernel. L'idea alla base è quella di contare il numero di eventi in una finestra di tempo continuamente spostata, creando così una curva continua di tassi di incidenza, invece dei valori discreti che si otterrebbero utilizzando finestre contigue (Mudelsee 2020).

La formula matematica dello stimatore kernel è:

$$\hat{\gamma}(T) = h^{-1} \sum_{j=1}^m K([T - T_{ext}(j)]/h)$$

dove h è chiamata l'ampiezza della finestra e K rappresenta la funzione kernel scelta. Vi sono diverse famiglie di funzioni kernel che possono essere utilizzate, la più comune è quella di densità gaussiana standardizzata $K(y) = (2\pi)^{-1/2} \exp(-y^2/2)$ (Silverman 2017). Di maggiore importanza è la scelta dell'ampiezza della finestra h : all'aumentare del valore aumentano le misurazioni che contribuiscono alla stima, riducendo così gli standard error, però al contempo comporta una perdita di risoluzione, perdendo i dettagli della distribuzione e portando ad un aumento della distorsione (Figura 2.2) (Mudelsee 2020).

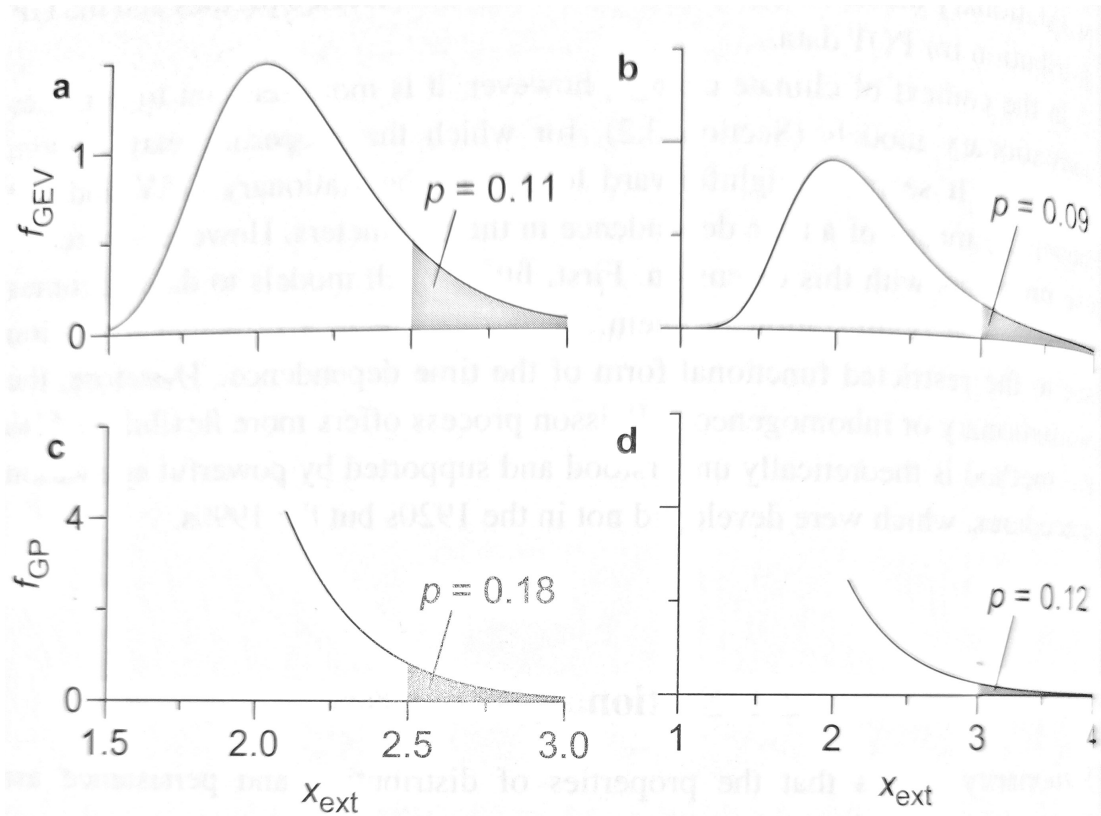


Figure 3.1 GEV and GP density functions. The parameter setting is (a) $\mu = 2.0$, $\sigma = 0.25$, $\xi = -0.05$; (b) $\mu = 2.0$, $\sigma = 0.4$, $\xi = 0.05$; (c) $\mu = 2.0$, $\sigma = 0.25$, $\xi = -0.05$, $u = 2.1$; and (d) $\mu = 2.0$, $\sigma = 0.25$, $\xi = 0.05$, $u = 2.1$. Also shown (shaded areas) are the tail probabilities, p , for two return levels: (a, c) $x_p = 2.5$ and (b, d) $x_p = 3.0$.

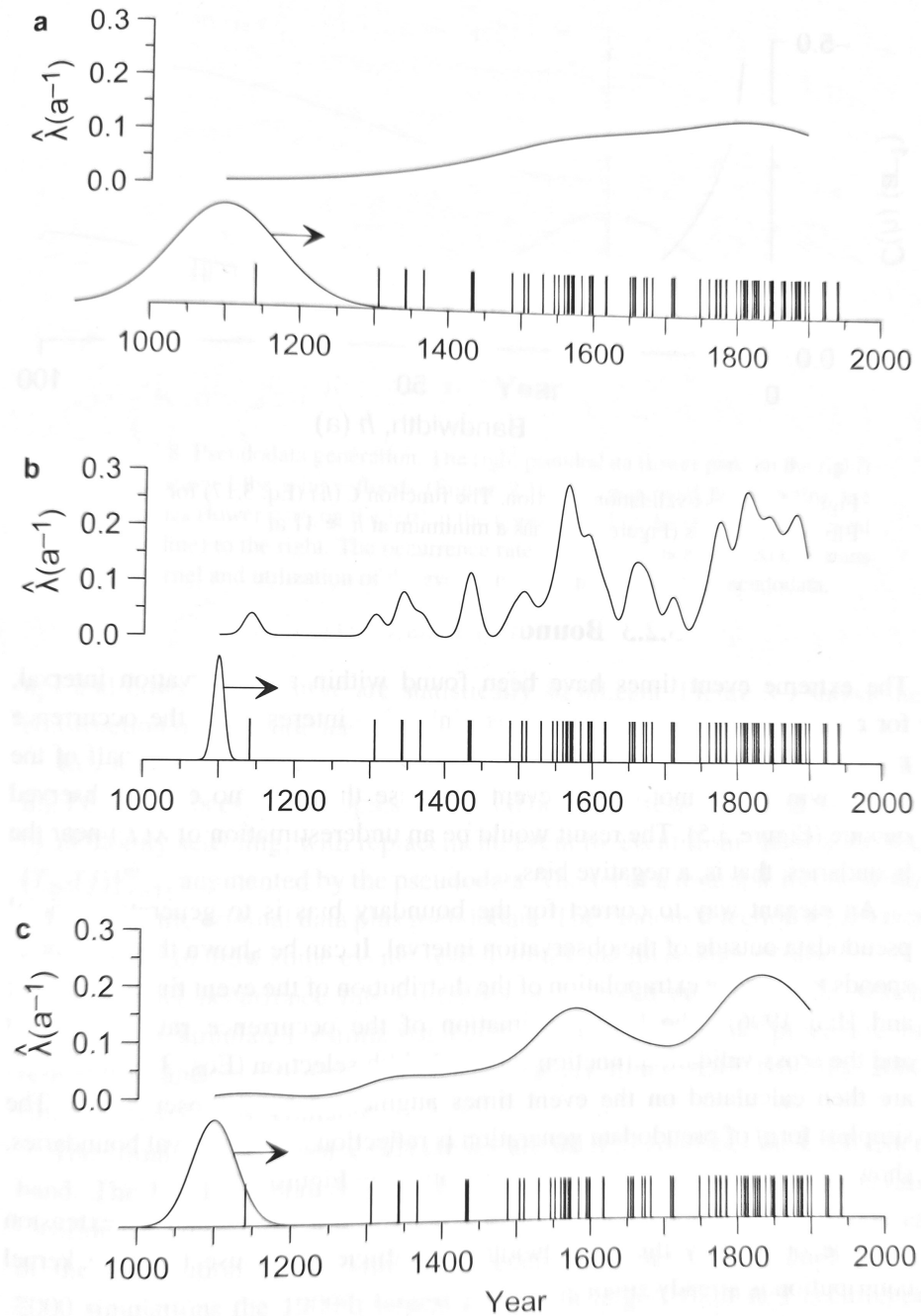


Figure 3.6 Bandwidth selection. Occurrence rate estimate for the heavy Elbe winter floods (Figure 2.1b) based on a Gaussian kernel with bandwidth 100 years (a), 10 years (b), and 41 years (c).

Figura 2.2: Fonte Mudelsee 2020

Capitolo 3

Applicazioni

Le analisi degli eventi estremi possono essere applicate a situazioni differenti ma forse la più diretta riguarda le ondate di caldo, manifestazioni violente dell'aumento delle temperature, che caratterizzano il fenomeno del cambiamento climatico tanto da farlo, alternativamente, chiamare riscaldamento globale. Come è stato evidenziato nel capitolo secondo, il primo passo per analizzare degli eventi estremi è quello di identificarli ed isolarli dall'intera serie storica. Le definizioni di ondate di caldo sono diverse ma spesso fanno riferimento ad indici specifici, calcolati per individuare questi fenomeni e distinguerli da dei semplici giorni con temperature anomale. Mudelsee (2020) ha sintetizzato, attingendo a vari autori che si sono interessati all'argomento (si veda, per esempio Zhang et al. 2011, Hartmann et al. 2013, Perkins e Alexander 2013), quattro requisiti a cui deve sottostare un indice affinché possa essere ritenuto utile ed informativo.

Requisito 1: L'indice deve essere applicabile in una varietà di settori. Deve essere in grado di catturare entrambi gli aspetti dell'evento estremo, intensità e durata. Le basi del suo calcolo devono essere comprensibili a ricercatori, tecnici o altri utilizzatori.

Requisito 2: L'indice deve essere adeguato ad essere applicato a diverse regioni geografiche.

Requisito 3: I dati devono essere disponibili per un calcolo significativo dell'indice. Il numero di valori assenti deve essere piccolo, ma non necessariamente zero.

Requisito 4: La definizione data dell'indice deve essere chiara e concisa facilitando così la riproducibilità dei risultati. Può rendersi necessario introdurre dei parametri per la definizione, come ad esempio un limite superiore. In ogni caso è meglio limitare il numero di parametri discrezionali al minimo necessario.

Per spiegare la ratio di questi requisiti si procederà con alcuni esempi che fanno riferimento agli indici in figura 3.1.

Si consideri, al fine di rilevare questo genere di eventi estremi, l'utilizzo della temperatura massima annuale o stagionale: per quanto questo indice rispetti i requisiti due, tre e quattro, non veicola alcun tipo di informazione riguardo alla durata, venendo quindi a violare il requisito numero uno. La durata infatti ricopre un ruolo importante in questa tipologia di eventi in quanto da questa dipendono molti degli effetti negativi su salute, ambiente ed attività economiche.

Utilizzare il conteggio dei giorni in cui la temperatura supera un determinato limite potrebbe sembrare un soluzione adeguata, ma sono presenti alcune complicazioni. L'utilizzo di un limite assoluto, come nell'esempio in tabella >20 °C, renderebbe complesso il confronto fra diverse regioni: superare il limite dei 20° nei Paesi Scandinavi non ha lo stesso valore di superarlo in Kenya, lungo la linea dell'Equatore. Si rivela più adeguato quindi l'utilizzo di un limite relativo, solitamente espresso come il superamento di un determinato percentile della distribuzione empirica. La scelta di quale percentile utilizzare non è, tuttavia, semplice: alcuni ricercatori come Perkins e Alexander (2013) ritengono che l'utilizzo del 90° percentile sia il giusto compromesso, mentre altri sostengono sia sempre importante provare differenti parametri, soprattutto in analisi esplorative (Mudelsee 2020). Inoltre, se il 90° percentile viene calcolato non prendendo in considerazione l'intero anno, ma il singolo giorno in anni differenti, è possibile concentrare l'analisi sul conteggio dei valori anomali al netto del ciclo stagionale. La metodologia appena descritta (90° percentile del giorno) rappresenta la convenzione per identificare i giorni caldi, ma non le ondate di caldo, in quanto anche qui viene trascurato l'aspetto della durata. Al fine di conciliare i vari requisiti e creare quindi un classe di indici adeguati a questo genere di eventi vengono introdotti gli "action measures". Tali indici sono calcolati come l'integrale definito della curva dei valori che superano un limite

Table 5.1. *Indices for heatwaves. The list is a selection from the Fifth Assessment Report of the IPCC (Hartmann et al. 2013: box 2.4 therein), augmented by the index “WSDI” (Zhang et al. 2011); also shown are the new “action measures.” For TX90p and TN90p, also relative units are employed (percentage of season/year instead of days). These two indices are calculated with calendar day percentiles. We also consider absolute percentile values (TX|90p| and TN|90p|). Stronger extremes, such as TX95p based on the 95th percentile, are also used. T_{max} , maximum daily temperature; T_{min} , minimum daily temperature.*

Index	Description	Definition	Unit
TXx	Warmest T_{max}	Seasonal/annual maximum of T_{max}	°C
TNx	Warmest T_{min}	Seasonal/annual maximum of T_{min}	°C
TX90p	Warm days	Seasonal/annual count of days when $T_{max} >$ calendar day 90th percentile	d
TN90p	Warm nights	Seasonal/annual count of days when $T_{min} >$ calendar day 90th percentile	d
TR	Tropical nights	Seasonal/annual count of days when $T_{min} > 20$ °C	d
WSDI	Warm spell duration index	Seasonal/annual count of days when $T_{max} >$ 90th percentile on ≥ 6 consecutive days	d
ATX 90p	Action measure for warm days	Integral of exceedance ($T_{max} - 90$ th percentile) over duration on ≥ 3 consecutive days	°C · d
ATN 90p	Action measure for warm nights	Integral of exceedance ($T_{min} - 90$ th percentile) over duration on ≥ 3 consecutive days	°C · d

Figura 3.1: Fonte Mudelsee 2020

percentile lungo un intervallo di tre o più giorni. In questo modo, tali indici rappresentano il perfetto compromesso, catturando durata ed intensità dell'evento (come suggerisce anche la stessa unità di misura $C \cdot d$, gradi centigradi per giorni), non perdendo l'informazione di quanto venga superato il limite e utilizzando un limite relativo che permette la confrontabilità tra regioni differenti.

L'attenzione riguardo alle ondate di caldo è particolarmente legata alla memoria che i cittadini europei hanno dell'estate 2003. In quell'anno l'Europa è stata colpita da un'intensa ondata di caldo le cui ripercussioni sulla salute sono difficili da stimare, ma un tentativo effettuato da Foullier et al. (2006) ritiene abbia causato un incremento di mortalità di quindicimila decessi nella sola Francia. Questo aumento delle temperature è legato al fenomeno del cambiamento climatico, più precisamente all'aumento della concentrazione di gas serra nell'atmosfera, e si ritiene, inoltre, che eventi di questa portata siano più frequenti che in passato e che possano diventare ancora più frequenti in futuro. L'ipotesi che l'incidenza di questi fenomeni estremi stia seguendo un trend crescente verrà commentata attraverso i dati forniti delle stazioni di Orléans e Potsdam, situate rispettivamente in Francia e Germania, e dalle analisi eseguite da Mudelsee (2020). Per stessa affermazione dell'autore due stazioni non sono sufficienti a comprendere l'estensione spaziale delle ondate di caldo, ma è possibile in futuro espandere il campione per un'analisi più approfondita del rischio che questi eventi disastrosi si verifichino in Europa. Gli indici più adeguati per condurre una prima analisi descrittiva sono quelli appartenenti alla categoria precedentemente introdotta degli *action measures*, in quanto in grado, come spiegato, di catturare intensità e durata delle ondate, informazioni fondamentali per comprendere le ripercussioni che possono portare. Questa scelta implica che siano due i parametri da decidere, il limite superiore e la durata minima. Iniziando ad analizzare le scelte possibili, il 99° percentile o una durata troppo lunga, come ad esempio di sei giorni, porterebbero ad isolare troppi pochi eventi; utilizzare invece una durata di un singolo giorno finirebbe per rimuovere completamente questo aspetto dall'analisi. La migliore combinazione ricade sul 90° o 95° percentile e su una durata di almeno 3 giorni (Mudelsee 2020). Un altro aspetto positivo delle *action measures* è la comparabilità tra diverse zone geografiche: l'e-

vento di massima entità rilevato a Orléans è stimato con un $ATX|90p| = 98.1C \cdot d$ (calcolato come in figura 3.1) ed è possibile confrontarlo con il più severo registrato a Potsdam, che ha fatto invece registrare $ATX|90p| = 50.2C \cdot d$ (figura 3.2). È curioso che il record di Potsdam non abbia origine nell'estate del 2003 ma in quella del 1994, più precisamente il 27 Luglio.

Si noti come l'utilizzo di indici *action measures* vada sostanzialmente a selezionare gli eventi estremi dalla serie storica attraverso il posizionamento di un limite, rendendo chiara l'appartenenza alla categoria introdotta nel capitolo secondo dei *peak over threshold* $\{T_{ext}(j), X'_{ext}(j)\}_{j=1}^m = \{T(i), X(i)|X(i) > u\}_{i=1}^n$. Questa caratteristica rende il sottoinsieme così ottenuto analizzabile attraverso i processi precedentemente spiegati: la distribuzione GP e i processi di Poisson. Volendo però verificare l'ipotesi della presenza di un trend crescente nel manifestarsi di questo genere di eventi non è adatta un'analisi in cui si assume stazionarietà, ma è più adeguata quella di assenza di stazionarietà. Calcolato l'indice di incidenza $\hat{\lambda}$ come documentato da Mudelsee (2020), attraverso una funzione Kernel Gaussiana e un bandwidth $h = 5$, il grafico per la stazione di Orléans si presenta come rappresentato in figura 3.3.

Si può notare come l'incidenza raggiunga un massimo locale in prossimità degli anni Quaranta, raggiungendo un valore di $\hat{\lambda} \approx 1.5a^{-1}$. Questo valore indica che fosse lecito aspettarsi un evento identificabile come ondata di caldo con cadenza di uno ogni una o due stagioni estive meteorologiche (che durano dal primo giorno di Giugno all'ultimo giorno di Agosto). Però questa stima potrebbe non essere del tutto corretta in quanto la serie storica della stazione di Orléans ha origine nel 1946 e quindi potrebbe risentire della correzione effettuata ai margini della finestra Gaussiana. Confrontando il grafico dell'incidenza con quello della stazione di Potsdam (figura 3.4) i due si rivelano molto simili, compreso l'incremento degli anni Quaranta che qui non presenta problemi nella stima poiché la serie inizia a fine del XIX secolo. Nel caso della stazione francese i valori decrescono fino agli anni ottanta per poi subire un incremento fino ad raggiungere $\hat{\lambda} \approx 2a^{-1}$ e poi stabilizzarsi, con un comportamento definito "plateau", negli ultimi quindici anni di misurazioni effettuate. La stazione tedesca, invece, registra un incremento conti-

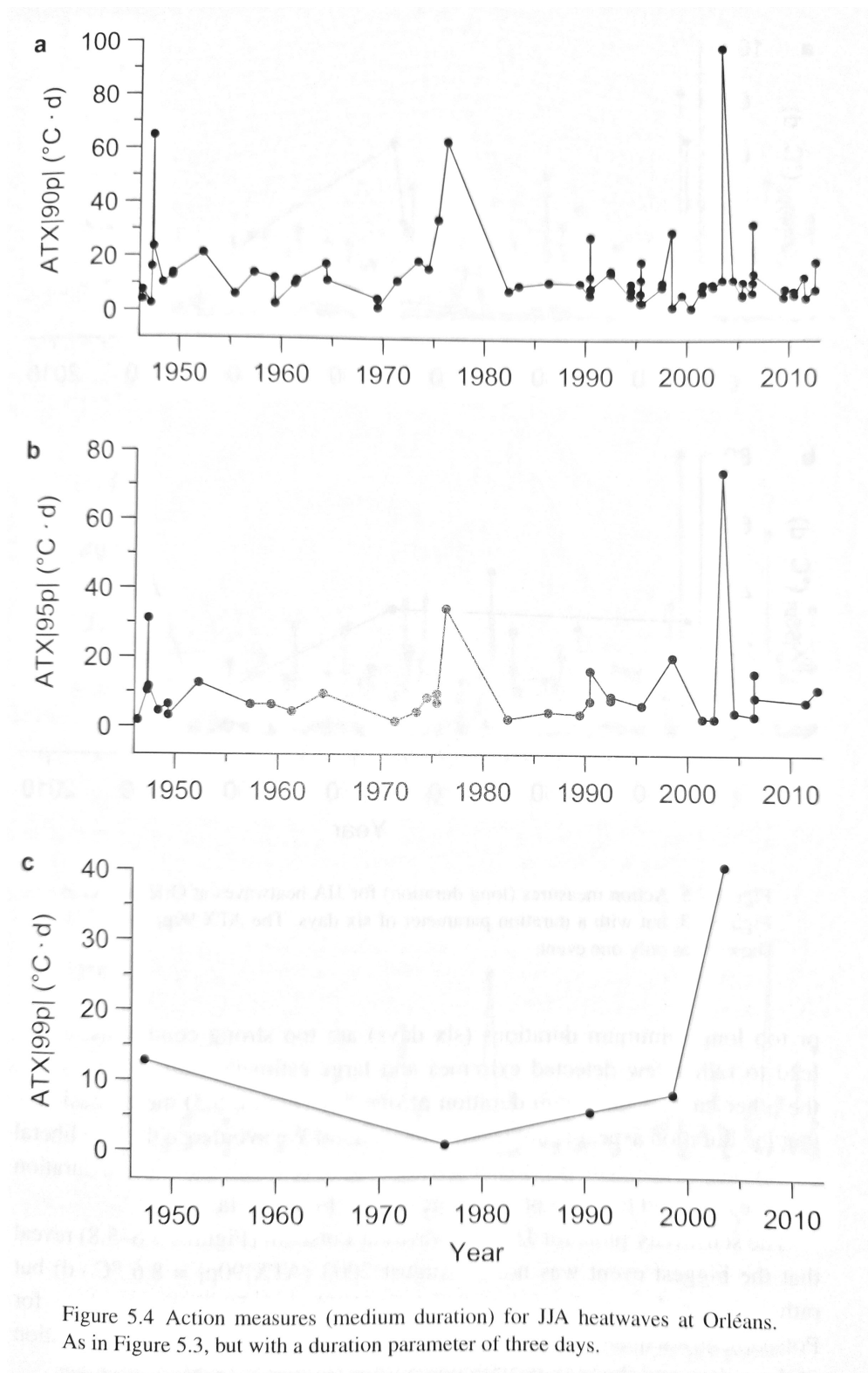


Figure 5.4 Action measures (medium duration) for JJA heatwaves at Orléans. As in Figure 5.3, but with a duration parameter of three days.

Figura 3.2: Fonte Mudelsee 2020

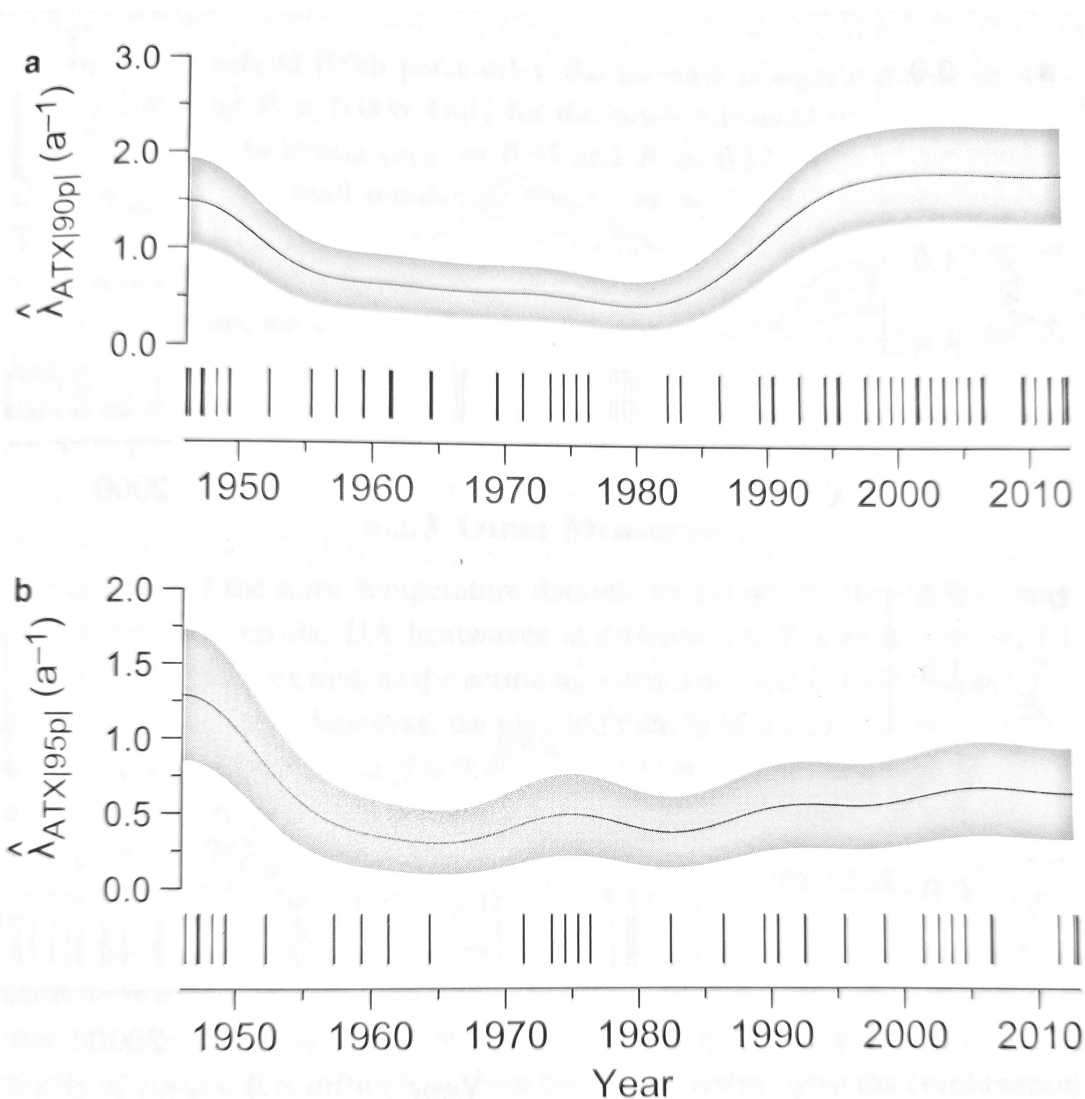


Figure 5.9 European heatwaves during the instrumental period: occurrence rates for Orléans. The detection index is the action measure, which is calculated with absolute percentile thresholds (a, 90th; b, 95th) and a duration parameter of three days (Figure 5.4a, b). The number of events is (a) $m = 72$ and (b) $m = 37$. The occurrence rates (solid lines) with 90% confidence bands (shaded) are estimated by means of a Gaussian kernel with a bandwidth of $h = 5$ a and 2000 bootstrap resamples.

Figura 3.3: Fonte Mudelsee 2020

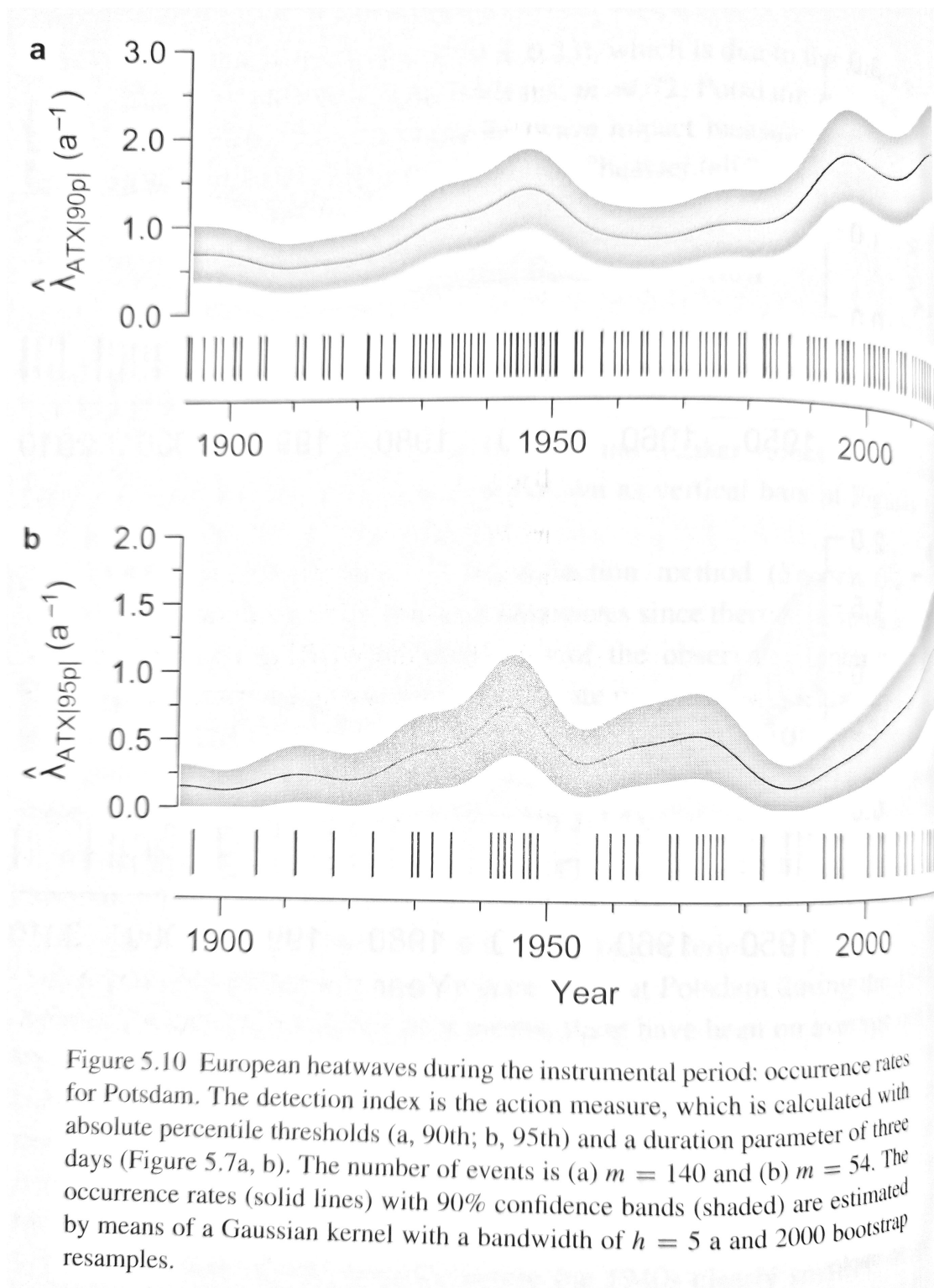


Figura 3.4: Fonte Mudelsee 2020

nuo fino alla fine della serie che trae comunque sempre origine negli anni Ottanta. La differenza di comportamento nella parte destra del grafico può essere spiegata osservando gli intervalli temporali: mentre Orléans termina nel 2012, Potsdam continua fino al 2018.

È quindi possibile notare già attraverso la sola osservazione del grafico la presenza di un trend crescente negli ultimi decenni. Questo fenomeno potrà essere confermato empiricamente attraverso l'esecuzione di un test di Cox - Lewis, utilizzabile per determinare se il tasso di incidenza aumenti, come effettuato da Mudelsee (2020) di cui riporto le conclusioni: "Per riassumere, abbiamo trovato che per entrambe le stazioni il tasso di incidenza manifesta un *plateau* o un recente trend incrementale, che rendono le stime di rischio per il presente verosimilmente maggiori di quanto è stato osservato in precedenza".

Bibliografia e Sitografia

1. Argo (2021). Argo float data and metadata from Global Data Assembly Centre (Argo GDAC). SEANOE. <https://doi.org/10.17882/42182>
2. Brückner, E., 1890. Klimaschwankungen seit 1700 nebst Bemerkungen über die Klimaschwankungen der Diluvialzeit, Geog. Abh. 4(2).
3. Chatfield, C. (2003). The analysis of time series: an introduction. Chapman and Hall/CRC.
4. Coles, S. (2001) An Introduction to Statistical Modeling of Extreme Values. Springer Verlag, Berlin. <http://dx.doi.org/10.1007/978-1-4471-3675-0>
5. Fouillet A, Rey G, Laurent F, Pavillon G, Bellec S, Guihenneuc-Jouyaux C, Clavel J, Jouglà E, Hémon D. Excess mortality related to the August 2003 heat wave in France. *Int Arch Occup Environ Health*. 2006 Oct;80(1):16-24. doi: 10.1007/s00420-006-0089-4. Epub 2006 Mar 8. PMID: 16523319; PMCID: PMC1950160.
6. Hartmann, D.L., Klein Tank, A.M.G., Rusticucci, M., Alexander, L.V., Brönnimann, S., Charabi, Y., Dentener, F.J., Dlugokencky, E.J., Easterling, D.R., Kaplan, A. and Soden, B.J., 2013. Climate change 2013: the physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Observations: Atmosphere and Surface, edited by T. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P. Midgley (Cambridge University Press, 2013).

7. IPCC, 2012 – Field, C.B., V. Barros, T.F. Stocker, D. Qin, D.J. Dokken, K.L. Ebi, M.D. Mastrandrea, K.J. Mach, G.-K. Plattner, S.K. Allen, M. Tignor, and P.M. Midgley (Eds.) Available from Cambridge University Press, The Edinburgh Building, Shaftesbury Road, Cambridge CB2 8RU ENGLAND, 582 pp. Available from June 2012
8. Mudelsee, M., Börngen, M., Tetzlaff, G., and Grünewald, U. (2004), Extreme floods in central Europe over the past 500 years: Role of cyclone pathway “Zugstrasse Vb”, *J. Geophys. Res.*, 109, D23101, doi:10.1029/2004JD005034.
9. Mudelsee, M. (2020). Frontmatter. In *Statistical Analysis of Climate Extremes* (pp. I-IV). Cambridge: Cambridge University Press.
10. Perkins, Sarah E., and Lisa V. Alexander. "On the measurement of heat waves." *Journal of Climate* 26.13 (2013): 4500-4517.
11. Resler LM. Edward N Lorenz’s 1963 paper, “Deterministic nonperiodic flow”, in *Journal of the Atmospheric Sciences*, Vol 20, pages 130–141: Its history and relevance to physical geography. *Progress in Physical Geography: Earth and Environment*. 2016;40(1):175-180. doi:10.1177/0309133315623099
12. Riser, Stephen et Freeland, Howard et Roemmich, Dean et Wijffels, Susan et Troisi, Ariel et Belbéoch, Mathieu et Gilbert, Denis et Xu, Jianping et Pouliquen, Sylvie et Thresher, Ann et Traon, Pierre-Yves et Maze, Guillaume et Klein, Birgit et Ravichandran, Muthalagu et Grant, Fiona et Poulain, Pierre-Marie et Suga, Toshio et Lim, Byunghwan et Sterl, Andreas et Jayne, Steven. (2016). Fifteen years of ocean observations with the global Argo array. *Nature Clim. Change*. 6. 145-153. 10.1038/nclimate2872.
13. Silverman, B.W. (1998). *Density Estimation for Statistics and Data Analysis* (1st ed.). Routledge. <https://doi.org/10.1201/9781315140919>
14. Zhang, X., Alexander, L., Hegerl, G.C., Jones, P., Tank, A.K., Peterson, T.C., Trewin, B. and Zwiers, F.W. (2011), Indices for monitoring changes in

extremes based on daily temperature and precipitation data. *WIREs Clim Change*, 2: 851-870. <https://doi.org/10.1002/wcc.147>

15. <https://www.ospo.noaa.gov/>

16. <https://www.treccani.it/vocabolario/clima1/>

17. https://www.treccani.it/enciclopedia/tempo-e-clima_%28Enciclopedia-della-Scienza-e-della-Tecnica%29/

Ringraziamenti

Alla mia famiglia, i miei genitori Cinzia e Luca e le mie sorelle Claudia e Chiara, i quali mi hanno sempre sostenuto non solo negli studi ma anche nella mia maturazione personale. In particolare ringrazio mia sorella Chiara senza la quale non sarei mai riuscito a superare il muro burocratico per laurearmi o addirittura iscrivermi all'Università. Ringrazio inoltre quel tipo di supporto psicologico che solo un cane può darti (Grazie Emy!) e quella saggezza che solo delle nonne possono passarti. Dedico inoltre la mia laurea ai miei nonni, che sarebbero stati sicuramente molto fieri di questo traguardo. Inoltre ringrazio la mia relatrice prof. Luisa Bisaglia che mi ha accompagnato nei primi passi verso il mondo della Statistica.