



UNIVERSITA' DEGLI STUDI "G. D'ANNUNZIO"

CHIETI – PESCARA

DIPARTIMENTO DI INGEGNERIA E GEOLOGIA

CORSO DI LAUREA IN SCIENZE GEOLOGICHE

**ANALISI STATISTICHE DEI DATI DEL CATALOGO SISMICO
ITALIANO DAL 1985 AD OGGI**

RELATORE

Chiar.mo Prof. 

LAUREANDO

Diego Eustachio Farchione

ANNO ACCADEMICO 2019/20

INDICE

1 - Introduzione.....	2
2 - Il Catalogo.....	6
2.1 - Tipologia di magnitudo	9
3 - Analisi sulle profondità.....	13
4 - Analisi della Gutenberg-Richter	19
4.1 - La legge di Gutenberg-Richter	19
4.2 - Stima del b-value	20
4.3 - Stima della Gutenberg-Richter	25
4.4 - Analisi probabilistiche e test statistici sulla Gutenberg-Richter.....	27
5 - Analisi temporale	38
5.1 - Istogrammi magnitudo-frequenza nel tempo.....	38
5.2 - Grafici tempo-frequenza nel tempo	40
5.3 - Analisi degli <i>interevent times</i>	41
6 - Il <i>declustering</i>	45
6.1 - Metodi di <i>declustering</i>	45
6.2 - <i>Declustering</i> con il metodo di Zaliapin (2008)	48
6.3 - Test di “Goodness of fit”	53
6.4 - Confronto tra gli <i>interevent times</i> successivi e gli eventi nel tempo prima e dopo il <i>declustering</i>	59
7 - Conclusioni.....	62
8 - Bibliografia	66
8.1 - Pubblicazioni	66
8.2 - Libri	69
8.3 - Sitografia	69

1 - INTRODUZIONE

Il presente lavoro di Tesi ha l'obiettivo di presentare analisi statistiche, con relativi grafici, dei dati del catalogo sismico italiano dal 1985 ad aprile 2020, e di interpretare i risultati per elaborare informazioni che potranno eventualmente essere utili allo studio della sismicità italiana.

L'elaborato comincia con un'introduzione sui cataloghi sismici, passando ad un'analisi grafica della tipologia di magnitudo e della distribuzione di esse nel tempo. Il terzo capitolo riporta le analisi effettuate degli eventi in profondità, studiando come essa varia in funzione di altri parametri, quali frequenza, latitudine, longitudine e magnitudo, e il quarto all'analisi della distribuzione magnitudo-frequenza, tramite l'equazione di Gutenberg-Richter. In particolare, in quest'ultimo capitolo ho analizzato come varia il coefficiente angolare dell'equazione per diverse magnitudo limite, ho creato dei cataloghi sintetici e effettuato statistiche su di essi.

Successivamente ho analizzato la distribuzione degli eventi nel tempo, la completezza del catalogo e ho dedicato un paragrafo alla differenza temporale tra eventi consecutivi, analizzando la loro distribuzione nel dettaglio.

È stato inoltre spiegato lo scopo del *declustering*, ossia il processo di eliminazione delle repliche dai cataloghi, e sono stati in breve esposti alcuni metodi per effettuarlo per poi applicare due di questi metodi nell'analisi. In particolare, mi sono soffermato sul metodo di Zaliapin et al. (2008), che è quello che ho implementato personalmente dall'inizio alla fine utilizzando uno specifico codice di calcolo.

Ho infine applicato dei test statistici a questi “nuovi” cataloghi ottenuti a seguito del *declustering*, con lo scopo di comprendere se essi seguissero un processo di Poisson nel tempo e ho mostrato la differenza tra l'occorrenza di due eventi consecutivi tra i “nuovi” cataloghi e quello originario. Ho poi confrontato nelle conclusioni i risultati dei test statistici, con altri ottenuti in letteratura sulla sismicità della California del sud (Gardner, J. K., and Knopoff, L. (1974), Luen, B., and Stark, P. B. (2012)).

Tutti i dati, raccolti dal sito dell'INGV, sono stati analizzati in linguaggio Python utilizzando l'IDE (cioè l'ambiente di sviluppo) denominato Jupyter Notebook. Questa applicazione è già incorporata nella piattaforma Anaconda, la quale è una distribuzione (ossia una raccolta di software) per l'analisi di dati e la scienza computazionale¹.

Ho utilizzato diverse librerie² del linguaggio Python, le principali sono:

- Numpy 1.18.1: utilizzata soprattutto con funzioni matematiche.
- Pandas 1.0.1: quella che ho più utilizzato. È di fondamentale importanza per gestire grandi quantità di dati, per creare tabelle (in modo simile ad Excel) ed effettuare modifiche nella struttura dei dati.
- Matplotlib 3.2.0: sfruttata per produrre grafici.
- Seaborn 0.10.0: libreria simile a Matplotlib ma con alcune caratteristiche aggiuntive.
- SciPy 1.4.1: in particolare il modulo `scipy.stats`, utile per effettuare test statistici e analizzare distribuzioni di dati.

Tutte queste librerie sono state installate a parte, mentre alcune come Datetime, Math e Random sono incluse nel pacchetto di installazione.

Per consultare le funzioni di queste librerie, ho utilizzato le pagine web ufficiali delle librerie e il libro *Python for Data Analysis* di W. McKenney, fondatore di Pandas.

Ho utilizzato anche MATLAB (versione 2019b) per scaricare ZMAP, un pacchetto che permette di eseguire statistiche sui terremoti che ho adoperato per effettuare il *declustering* con il metodo di Reasenberg, P. (1985) (*Capitolo 6*).

Prima di cominciare ad avvicinarmi all'analisi e allo sviluppo di codici per lo studio del catalogo, ho studiato la letteratura scientifica disponibile sull'argomento e sull'evoluzione della ricerca in ambito della sismologia statistica (tra le principali pubblicazioni da cui sono partito per compiere il lavoro di ricerca, si possono citare: Gardner, J. K., and Knopoff, L. (1974); Luen, B., and Stark, P. B. (2012); Naylor, M.,

¹ scienza che si occupa dell'utilizzo di computer e sistemi di elaborazione per risolvere problemi in ambito scientifico.

² Sono un insieme di funzioni che possono essere già presenti nel linguaggio di programmazione oppure installate successivamente, che permettono di svolgere funzioni molto specifiche o di facilitare le operazioni di sviluppo del codice.

Orfanogiannaki, K., and Harte, D. (2010); Van Stiphout, T., J. Zhuang, and D. Marsan (2012)).

La maggior parte del tempo è stato speso per familiarizzare con le numerose funzioni specifiche incluse nelle librerie di Python ed in particolare per implementare il *declustering* con il metodo di Zaliapin. Anche il tempo di calcolo ha avuto una grossa influenza soprattutto nell'esecuzione del *declustering*, infatti sono state necessarie molte ore di elaborazione da parte del computer per portarlo a termine. Ho più volte trovato difficoltà nello scrivere il codice necessario, ma anche per applicare i test statistici nel modo giusto. All'inizio ho trovato particolari problemi anche nell'utilizzo del Catalogo della Sismicità Italiana CSI 1.1 (Castello, B., Selvaggi, G., Chiarabba, C., Amato, A. (2006)), poiché alcuni eventi del centro Italia apparivano "traslati" in latitudine e longitudine, forse per un errore nella lettura dei dati o per il trasferimento da un formato ad un altro. Inoltre, prima di poter utilizzare il catalogo, ho dovuto procedere nella eliminazione dei dati "corrotti" e nella correzione di alcune misure (ad esempio trasformare latitudine e longitudine in formato decimale o trasformare le date in un formato adeguato al fine di poter analizzare nel tempo gli eventi).

Perciò ho deciso di considerare il catalogo ISIDe (ISIDe Working Group. (2007)). In questo caso non ho riscontrato particolari difficoltà e ho dovuto solo effettuare poche conversioni, tra cui quella di convertire in un formato più idoneo la data e l'ora di ogni evento. Anche per trasformare il catalogo successivo al *declustering* di Reasenberg dal formato `.mat` di MATLAB ad un formato `.csv` ho riscontrato problemi, che ho poi risolto portando il catalogo su Excel e poi su Python, applicando delle funzioni che trasformassero i dati.

Di fondamentale importanza è stato acquisire una discreta conoscenza della statistica che sono poi andato ad utilizzare. Per poter comprendere le pubblicazioni è stato necessario approfondire alcuni argomenti di statistica, quali la statistica inferenziale, le distribuzioni di probabilità, i test statistici e la statistica di base.

Ovviamente per sapere come poter applicare i codici di calcolo e la statistica è fondamentale avere una buona conoscenza della sismologia, poiché è necessario sapere quali dati bisogna considerare, come sono strutturati i cataloghi e quali cataloghi

considerare, come interpretare i risultati ottenuti, come procedere nelle analisi e come si possono spiegare eventuali *pattern* presenti nei grafici rappresentati.

Il codice che ho implementato per produrre i grafici ed effettuare i test statistici è visibile online al seguente indirizzo: <https://github.com/Theironwilledguy/bachelor-thesis/tree/master>, cliccando sulle icone con il prefisso “THESIS”.

2 - IL CATALOGO

Il catalogo sismico che si è analizzato è il catalogo ISIDe (Italian Seismological Instrumental and Parametric Data-Base), disponibile sul sito dell'INGV (<http://terremoti.ingv.it/en>).

Ho impostato una ricerca personalizzata, considerando tutti i terremoti con magnitudo maggiore di 1.5. C'è da specificare che il criterio di ricerca nella pagina web non fa riferimento ad un tipo di magnitudo particolare, bensì al tipo di magnitudo con cui gli eventi sono stati registrati, compresi tra una latitudine di 36.1N e 47.1N e una longitudine di 6.3E e 18.4E, includendo perciò gli eventi avvenuti su tutto il territorio italiano e quelli registrati lungo le coste della Dalmazia e parte dell'Arco Alpino (Figura 2). L'intervallo temporale preso in considerazione copre il periodo compreso tra il primo gennaio 1985 e il ventotto aprile 2020.

I dati sono stati raccolti esportando 10000 eventi alla volta, che corrisponde al limite massimo di esportazione previsto. Questi dati sono disponibili in formato Text, Kml, Atom e Quakeml; quest'ultimo è un formato nato con l'obiettivo di contenere dati sismologici e che può fornire ulteriori informazioni anche riguardo l'incertezza nell'identificazione della magnitudo e della localizzazione. Nel caso del catalogo utilizzato nel presente lavoro ho considerato il formato Text. Successivamente ho unito tutti questi sotto-cataloghi di 10000 eventi eliminando qualora fossero presenti, i terremoti in comune nei cataloghi consecutivi. Ho finalizzato questa procedura tramite la libreria Pandas di Python, in particolare utilizzando la funzione `concatenate` per unire tutti i cataloghi e la funzione `read_csv` per leggere i cataloghi in formato .txt, considerando come separatore tra i vari campi la barra '/'.

Il catalogo presenta i seguenti campi:

1. **EventId**: ovvero il numero che identifica il terremoto;
2. **Time**: il tempo sottoforma di anno con 4 cifre (ad esempio: 1994, 2003, 2015 ecc..), mese (2 cifre), giorni (2 cifre), ore (considerando il sistema orario a 24 ore e non a 12), minuti, secondi e centesimi di secondo;
3. **Latitude**: la latitudine (composta di 5 cifre);

4. **Longitude:** la longitudine (anch'essa di 5 cifre);
5. **Depth/Km:** profondità in chilometri (con precisione al decimo di chilometro);
6. **Author:** l'ente che ha acquisito i dati;
7. **Catalog,Contributor,ContributorID:** campi che non hanno valori contenenti informazioni;
8. **MagType:** il tipo di magnitudo calcolata;
9. **MagAuthor:** campo anch'esso presente, ma non contenente informazioni;
10. **EventLocationName:** il luogo scritto per esteso, dove si ipotizza sia avvenuto il terremoto (ad esempio: 9 km SE Longobucco (CS)).

Come descritto da Woessner et al. (2010), un catalogo per essere tale deve avere un codice di identificazione dell'evento, formato da numeri o da lettere e numeri, da un sistema di coordinate per posizionarlo nello spazio (comprendendo la profondità), il tempo di origine del terremoto e la magnitudo.

Altri parametri che possono essere aggiunti e che aggiungono ulteriori informazioni sono l'incertezza sulla magnitudo e sullo spazio, il numero di osservazioni, il GAP (Greatest Azimuthal Gap), cioè l'angolo maggiore per cui non ci sono stazioni che hanno registrato il terremoto, il GAP secondario, ovvero il secondo angolo maggiore. Poi ancora si può avere lo scarto quadratico medio dei residui dei tempi di arrivo osservati, la distanza tra la stazione più vicina, parametri per la determinazione dei meccanismi focali (immersione, direzione ecc..) e la loro incertezza, l'identificazione delle varie fasi, il tipo di evento e la sorgente che lo ha causato e stime di massima intensità e intensità ipocentrale.

Il risultato finale che ho ottenuto caricando e visualizzando è il seguente (Figura 1):

#EventID	Time	Latitude	Longitude	Depth/Km	Author	Catalog	Contributor	ContributorID	MagType	Magnitude	MagAuthor	EventLocationName
0	79 1985-01-02T22:57:43.090000	41.715	13.941	5.8	BULLETIN-VAX	NaN	NaN	NaN	Id	2.6	--	6 km S Civitella Alfedena (AQ)
1	89 1985-01-03T00:46:16.990000	41.677	13.931	6.1	BULLETIN-VAX	NaN	NaN	NaN	Id	2.5	--	6 km NE Picinisco (FR)
2	199 1985-01-04T16:57:40.240000	46.158	7.174	10.0	BULLETIN-VAX	NaN	NaN	NaN	M	3.8	--	Confine Italia-Svizzera (SVIZZERA)
3	229 1985-01-04T23:59:54.820000	45.338	6.641	5.0	BULLETIN-VAX	NaN	NaN	NaN	M	3.7	--	Confine Italia-Francia (FRANCIA)
4	239 1985-01-05T03:47:03.640000	45.409	6.627	10.0	BULLETIN-VAX	NaN	NaN	NaN	M	3.2	--	Confine Italia-Francia (FRANCIA)
...

Figura 1 - prime 5 righe del catalogo sismico visualizzate con Pandas

Dove i valori corrispondenti alla sigla “NaN” sono valori senza un dato. Gli eventi totali importati sono 181778.

Oltre a caricare tutto il database, ho successivamente proceduto nel processo denominato anche *data cleaning*, che consiste nel rilevare eventuali dati corrotti, verificare l’affidabilità dei dati e modificare il formato. Ho eseguito questi passaggi più volte e in più occasioni, a seconda della tipologia di dati che ho analizzato e rappresentato.

A partire dal catalogo ho rappresentato graficamente i vari dati, attuando quella che viene definita “*analisi esplorativa di dati*”. Per prima cosa ho importato il catalogo su ZMAP³ (Wiemer, S. (2001)), un software GUI (Graphical User Interface) ovvero un tipo di programma in grado di interagire con l’utente tramite un’interfaccia grafica. Esso è spesso utilizzato dai sismologi per poter interpretare meglio i cataloghi sismici, valutare la completezza di un catalogo, realizzare semplici grafici ed effettuare il *declustering* dei cataloghi. Per poter usufruire del software, è necessario possedere una licenza MATLAB, perciò ho per prima cosa ottenuto la licenza e scaricato tutti i pacchetti esterni opzionali. I pacchetti (ovvero un insieme di *tool* esterni che permettono di svolgere funzioni particolari o che permettono l’avvio di un determinato software), necessari per l’installazione di ZMAP sono i pacchetti “*Mapping Toolbox*” e “*Statistics and Machine Learning Toolbox*”. Ho quindi avviato il software attraverso del codice che mi permettesse di entrare nella directory (ossia nella cartella) dove è presente ZMAP. Ho quindi importato il catalogo considerando gli eventi con magnitudo maggiore o uguale a 3 (11194 eventi), per ragioni di visibilità dei dati e di tempo di processamento. Automaticamente questo software esegue la rappresentazione grafica di questi eventi (Figura 2).

³ <http://www.seismo.ethz.ch/en/research-and-teaching/products-software/software/ZMAP/>

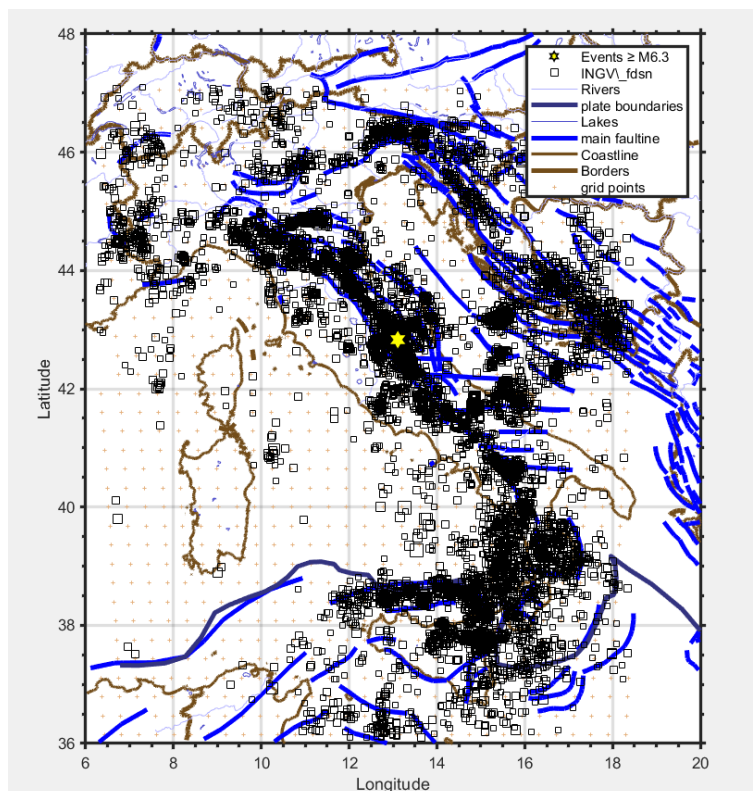


Figura 2 - eventi con magnitudo maggiore o uguale a 3 in ZMAP

L'evento raffigurato in Figura 2 con una stella bianca, rappresenta la scossa di magnitudo momento 6.5 del 30 ottobre 2016 nei pressi di Norcia.

2.1 - TIPOLOGIA DI MAGNITUDO

Una volta visualizzati i terremoti su una mappa, ho analizzato più nel dettaglio tutti i terremoti del catalogo (tutti quelli maggiori o uguali di 1.5). Pertanto, ho proceduto partendo dall'analisi della magnitudo utilizzata e della frequenza di ognuna, attraverso rappresentazioni ad istogrammi (Figura 3) o numericamente (Figura 4).

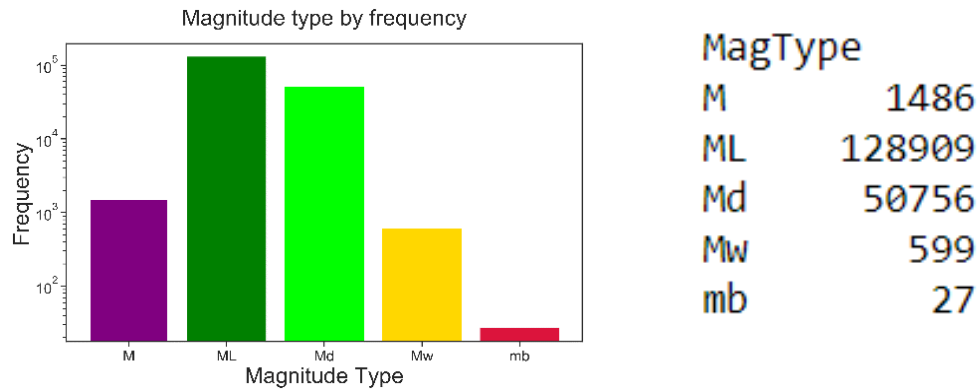


Figura 3 (a sinistra) - tipo di magnitudo e rispettiva frequenza

Figura 4 (a destra) - Numero di eventi per tipologia di magnitudo

Si nota che la tipologia di magnitudo più utilizzata è quella locale (ML) con 128909 eventi. La magnitudo locale si calcola con la seguente formula:

$$M_L = \log A + 2.76 \log \Delta - 2.48 \quad (2.11)$$

In cui Δ è la distanza epicentrale e A è l'ampiezza massima visibile sulla traccia del sismogramma, registrata da un sismometro Wood-Anderson.

Nel catalogo ritroviamo poi la magnitudo di durata (Md), basata molto semplicemente sulla durata dell'evento sismico, quella di volume (mb) che si utilizza per terremoti a distanza superiore a 600 km e che prende in considerazione le onde di volume. Infine, la magnitudo generica (M), calcolata anch'essa tenendo conto dell'ampiezza del segnale, del periodo dell'onda e di ulteriori correzioni regionali, di distanza, di stazione e di profondità e la magnitudo momento (Mw) più complessa da ricavare, che si basa sul momento sismico e di conseguenza necessita la conoscenza di altri parametri come l'area di rottura della faglia, la dislocazione cosismica media e il modulo di taglio.

Si ha la possibilità inoltre di compiere ulteriori analisi, confrontando il tipo di magnitudo con il valore stesso della magnitudo e con il tempo. Per poter interpretare questi dati, è molto utile sfruttare il *grafico a scatola* oppure il *grafico a violino*. I due sono molto simili tra di loro, infatti entrambi riescono a visualizzare in un solo grafico la mediana, il 25° e il 75° percentile della distribuzione (spesso indicati rispettivamente con Q1 e Q3). Il primo permette, oltre di visualizzare la posizione di Q1 e Q3, anche

questo range compreso tra il 25° e il 75° percentile denominato range interquartile (IQR). Solitamente esso è rappresentato con una barra, mentre la mediana con un segmento all'interno di questa barra. In seguito, prolungando questa barra, vengono raffigurati ulteriori due segmenti alle estremità che si interrompono in corrispondenza dei valori $Q1-1.5IQR$ e $Q3+1.5IQR$. In alcuni casi possono essere anche rappresentati degli *outliers*, ovvero quei dati che sono esterni all'intervallo $Q1-1.5IQR$ e $Q3+1.5IQR$.

Nel grafico a violino, in particolare in quelli considerati in Figura 5 e Figura 6, riusciamo invece a rappresentare la mediana (rappresentata come un punto bianco), l'IQR (rappresentato con una barra nera) e la frequenza di un particolare insieme di dati in funzione dei valori in ordinata: più è spessa la parte colorata, più sono frequenti quei valori.

Nel grafico a destra (Figura 5), si osserva come per bassi valori di magnitudo venga preferito l'utilizzo della magnitudo locale, al contrario per alti valori abbiamo un utilizzo più frequente della magnitudo momento e della magnitudo di volume. Nel primo caso perché sono necessarie stime più accurate per la determinazione di queste misure, nel secondo invece abbiamo solo 27 eventi con questa misura, perciò è utilizzata molto

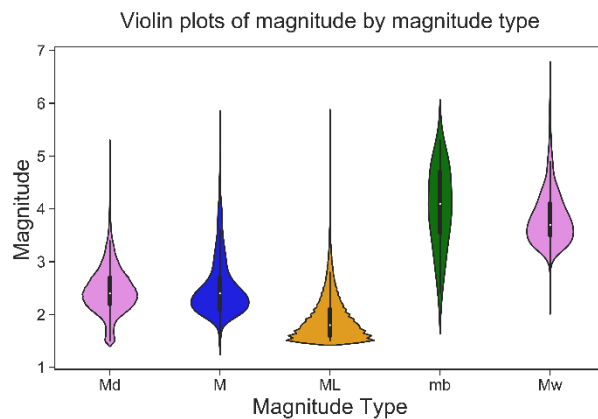


Figura 5 - tipologia di magnitudo e loro distribuzione per magnitudo

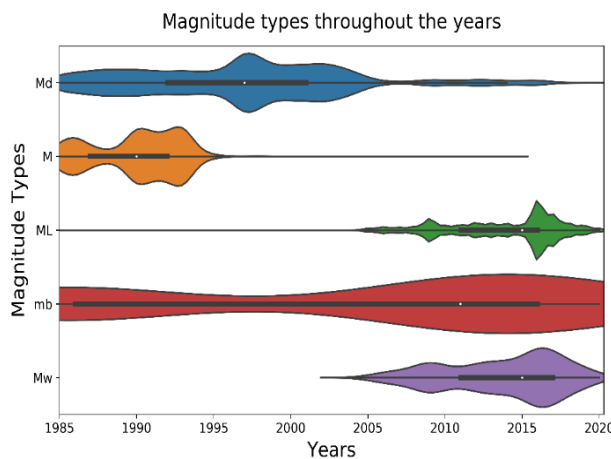


Figura 6 - tipologia di magnitudo nel tempo

raramente e non possiamo interpretare con precisione. L'utilizzo della magnitudo di durata e della magnitudo generica è invece frequente per la maggior parte per gli eventi tra 2 e 3.

In quello qui a sinistra (Figura 6), si è compiuta una visualizzazione nel tempo per ogni tipologia di magnitudo. Si nota che la magnitudo locale e la

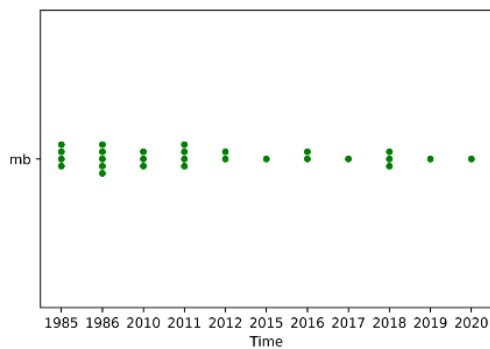


Figura 7 - magnitudo di volume per numero di eventi negli anni

magnitudo momento sono state utilizzate negli ultimi anni, in particolare la Mw che è stata introdotta per la prima volta molto più di recente. I picchi nella distribuzione di questi due diagrammi sono evidenti in corrispondenza del 2009 e del 2016, in cui si sono verificati rispettivamente il terremoto dell'Aquila (06/04/2009) e la sequenza

sismica del Centro Italia (i cui principali eventi si sono verificati il 24 agosto 2016, il 26 e il 30 ottobre 2016 e il 18 gennaio 2017). Si può anche constatare con un ulteriore grafico a sciame (Figura 7) avendo pochi dati, che la Mb è in uso nei primi due anni e negli ultimi dieci del catalogo. Dato che essa viene utilizzata per misurare terremoti a distanza maggiore di 600 km, si può facilmente osservare rappresentando questi eventi su una mappa o semplicemente controllando il catalogo, che questi eventi avvenuti dopo il 2010 sono tutti avvenuti al largo delle coste croate, della Tunisia o della Francia, luoghi molto distanti dal territorio italiano (e di magnitudo superiore rispetto alla media) e perciò elegibili per poter calcolare questo tipo di magnitudo. I 9 eventi avvenuti nel 1985 e 1986 sono invece stati registrati nel Tirreno Meridionale tra i 180 e i 611 km di profondità, di conseguenza tutti molto lontani dalle stazioni sismiche.

Le magnitudo di durata e quella generica sono state invece utilizzate maggiormente nel passato: la prima mostra un picco in corrispondenza della sequenza sismica di Colfiorito del 1997 e ancora oggi, sebbene venga raramente utilizzata; mentre la seconda non è oggi giorno utilizzata.

3 - ANALISI SULLE PROFONDITA'

Il catalogo ci fornisce i dati sulla profondità in chilometri di ciascun evento sismico, pertanto si può analizzarne la distribuzione in frequenza, verificare se ci sono correlazioni tra i vari campi del catalogo o il motivo per cui i vari terremoti si distribuiscono con una determinata disposizione in profondità.

La prima rappresentazione che permette di visualizzare gli eventi in profondità è la distribuzione spaziale lungo i tre assi, che nel caso della terra sono indicati dalla longitudine, dalla latitudine e dalla profondità (Figura 8).

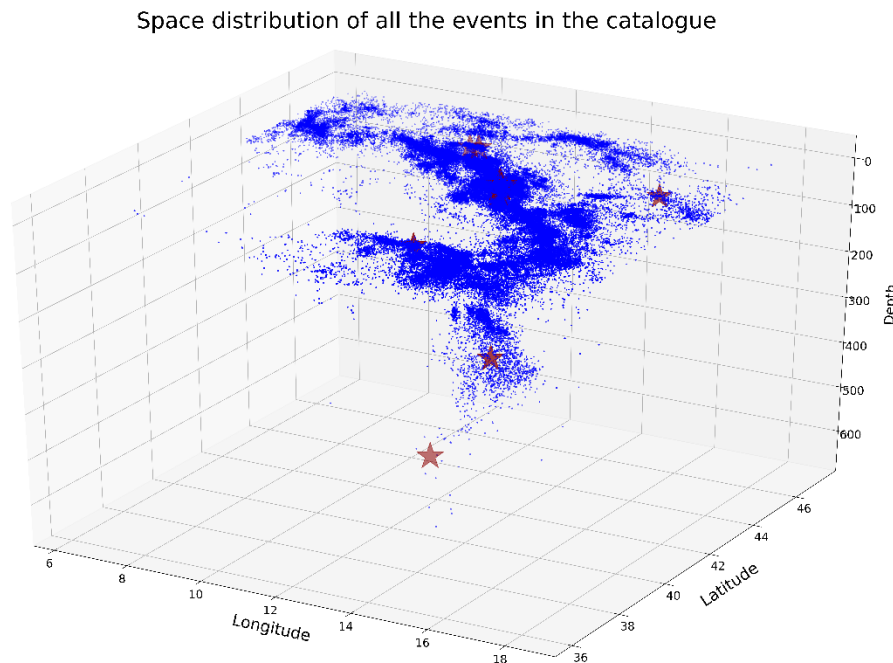


Figura 8 - rappresentazione tridimensionale di tutti gli eventi nel catalogo

Gli eventi indicati con una stella in rosso sono quelli più importanti di tutto il catalogo, ossia maggiori o pari a una magnitudo di 5.6. Si può facilmente notare che gli eventi superficiali si distribuiscono lungo la catena appenninica e nel Tirreno Meridionale a nord della Sicilia, e ulteriori eventi lungo la catena alpina in particolare tra Italia e Francia, in Friuli e tra Slovenia e Italia. Si sono verificati nell'arco di questo lungo periodo ulteriori eventi al di fuori dei confini nazionali lungo le Dinaridi, tra Croazia, Slovenia e Bosnia, come si può constatare in Markušić et al. (1998) e Moulin et al. (2016), e alcuni nel Mar Adriatico. Terremoti più profondi sono quasi esclusivamente

concentrati fino a 650 km di profondità nel Mar Tirreno e correlati alla subduzione della litosfera oceanica, corrispondente al vecchio oceano della Tetide, al di sotto della crosta continentale (In Sartori, R. (2003) viene descritta più dettagliatamente questa regione tettonica).

È possibile visualizzare il precedente grafico tridimensionale in ulteriori due modi, considerando un grafico profondità-longitudine (Figura 9) e uno profondità-latitudine (Figura 10)

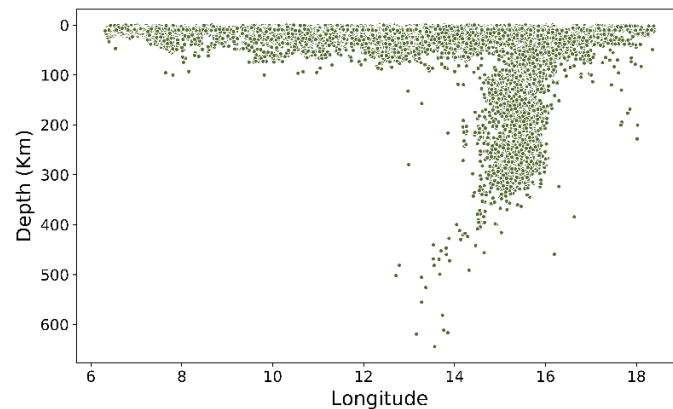


Figura 9 - grafico longitudine-profondità

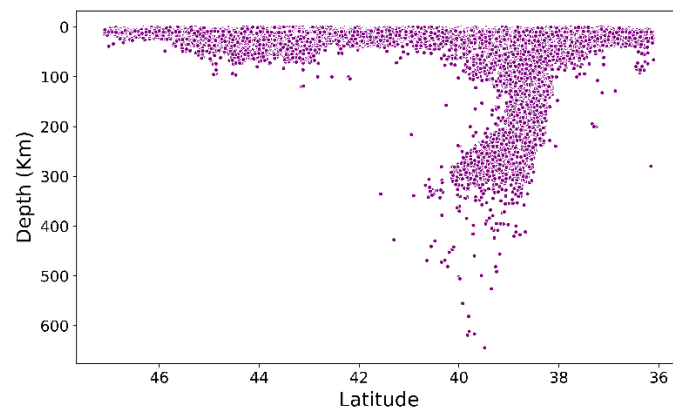


Figura 10 - grafico latitudine-profondità

Attraverso questo modo di visualizzare i dati, si può denotare più chiaramente il piano di Benhioff lungo cui si dispongono i terremoti della crosta oceanica in subduzione.

Oltre alla rappresentazione a dispersione può essere effettuata una rappresentazione tramite griglie di frequenza (Figura 11) per osservare la distribuzione del numero di eventi in profondità, utilizzando una scala logaritmica in modo tale da poter meglio analizzare la distribuzione a distanze più superficiali.

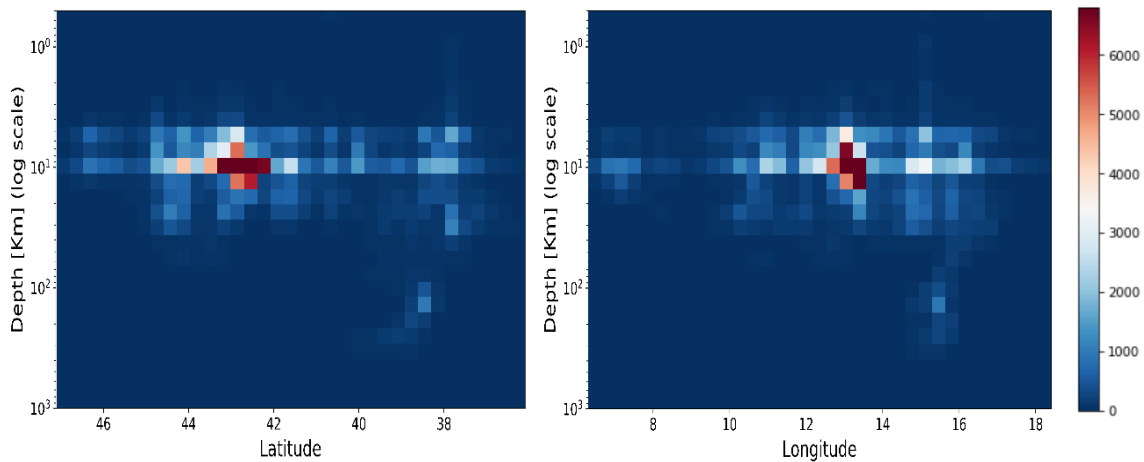


Figura 11 - frequenza degli eventi in profondità con la latitudine (a sinistra) e con la longitudine (a destra)

La maggior parte degli accadimenti si collocano tra i 43.5N e i 42N gradi di latitudine e tra i 12.5E e 14E gradi di longitudine. Ciò vuol dire che la stragrande maggioranza di terremoti in tutto il catalogo si sono verificati nel centro Italia, precisamente tra Umbria, Marche, Lazio e Abruzzo. Una dimostrazione di ciò è che i più tre grandi *mainshocks*, ossia i terremoti più importanti all'interno di una sequenza (Colfiorito, L'Aquila e sequenza del Centro Italia), si collocano in corrispondenza di questo range di latitudine e longitudine. Quindi anche i successivi *aftershocks*, ovvero i terremoti che seguono la scossa principale, essendo molto numerosi, si concentrano in quest'area. È importante evidenziare la correlazione esistente tra i *mainshocks* e gli *aftershocks*. Questi ultimi sono certamente di minore importanza rispetto alla scossa principale, ma dipendono in gran parte dalla magnitudo del *mainshock*: maggiore è la magnitudo, maggiore è il numero e la magnitudo dei terremoti che si verificano successivamente e maggiore è l'intervallo temporale in cui si verificheranno gli *aftershocks*.

Un'altra interpretazione che possiamo evincere da queste due rappresentazioni, è la profondità in cui si verificano la maggior parte dei terremoti. Infatti, la maggior parte ha

un ipocentro compreso tra 3 e 13 km di profondità, con un picco in corrispondenza di circa 10 km di profondità.

Avendo osservato relazioni tra frequenza di eventi in funzione della profondità e delle coordinate, dato che la frequenza del numero di terremoti è proporzionale alla magnitudo per la legge di Gutenberg-Richter, si osserva una magnitudo più alta sempre in corrispondenza di 10 km di profondità, dove si ha il picco di accadimenti (Figura 12).

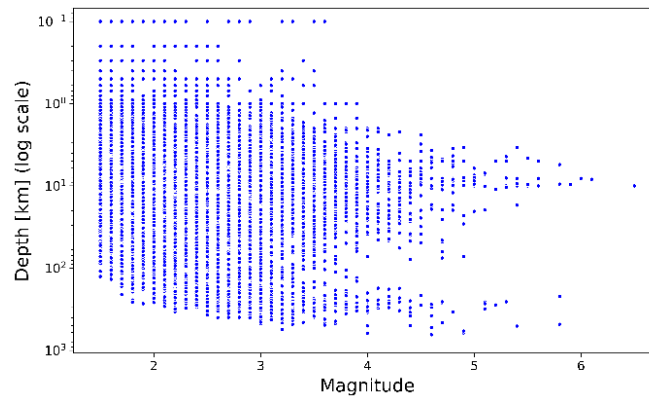


Figura 12 - relazione tra magnitudo e profondità

Ancora una volta sono visibili le due popolazioni di terremoti: quella più profonda, che inizia a distinguersi a partire dai 100 km di profondità, in cui ricadono la maggior parte dei terremoti avvenuti nel Tirreno Meridionale, e la seconda comprendente eventi più superficiali, che è collegata principalmente alla sismicità delle restanti aree coperte dal catalogo.

Nei successivi grafici si sono realizzati degli istogrammi che rappresentano più nel dettaglio il numero di eventi in funzione della profondità.

Number of earthquakes ≥ 1.5 by depth

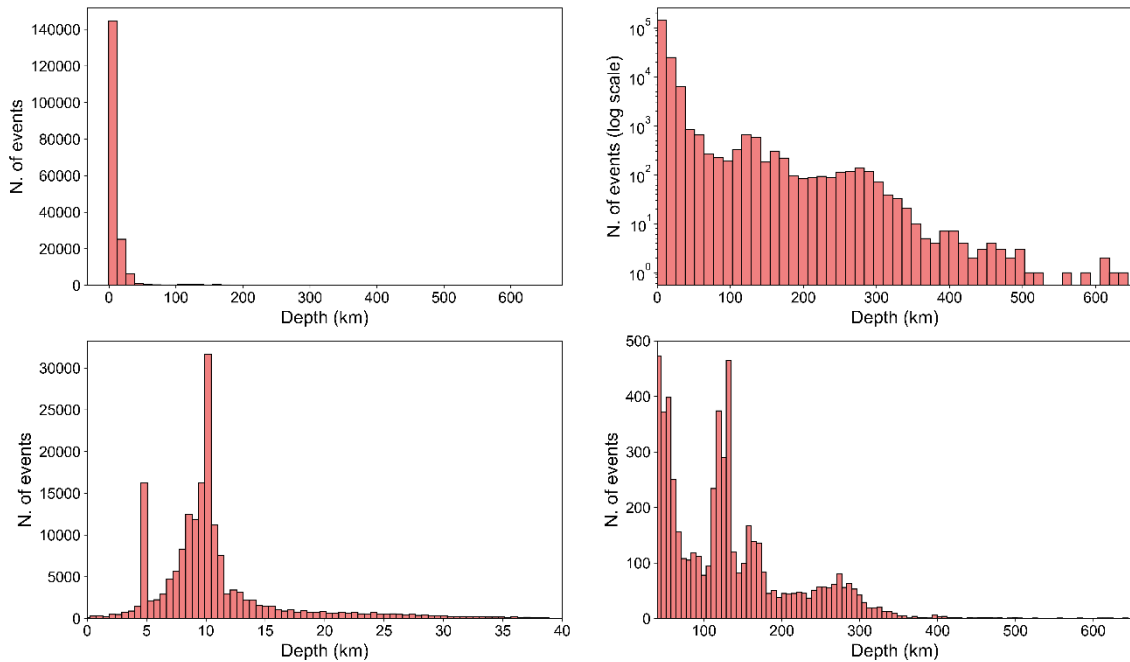


Figura 13a (in alto a sinistra), 13b (in alto a destra), 13c (in basso a sinistra), 13d (in basso a destra) - istogrammi sulla profondità

Nella Figura 13a si è ottenuto un grafico con entrambi gli assi lineari e con l'ascissa che copre tutto l'intervallo di profondità in cui si sono verificati i terremoti. È evidente che gli eventi verificatisi nei primi chilometri di profondità sono preponderanti, mentre quelli profondi sono molto meno frequenti e non osservabili in scala lineare. Nel grafico della Figura 13b avente ordinata logaritmica, si vede ora chiaramente che più si raggiungono profondità elevate, più la frequenza decade esponenzialmente. Per analizzare più nel dettaglio si può suddividere l'intervallo di profondità considerato, dapprima considerando eventi fino ad una profondità di 40 km (Figura 13c) e successivamente a partire dai 40 ai 650 km (Figura 13d). Nel primo di questi, si ha un picco in corrispondenza di 10 km. Ciò vuol dire che la maggior parte degli eventi (come già dimostrato in Figura 11a e 11b) si concentra in questo intervallo.

Oltre a un picco molto pronunciato in corrispondenza di 10km, si ha un picco inusuale anche in corrispondenza di 5 km. Questo non è associato a ragioni sismotettoniche, ma principalmente è dovuta ad una approssimazione artificiale. Come si può osservare in *Figura 14*, diminuendo la grandezza dell'intervallo di campionamento dell'istogramma

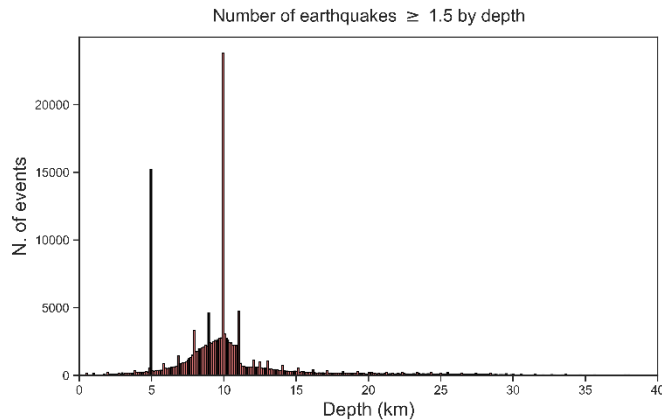


Figura 14 - numero di eventi in profondità con campionamento ristretto

prodotto in precedenza, si riesce a vedere che gli eventi di magnitudo 5 e 10 sono estremamente maggiori agli altri e allo stesso modo tutti quelli che seguono la successione dei numeri naturali (1,2,3,4,5,6,7,8,9,10,11 ecc... km di profondità). La ragione, come già

citato, è probabilmente dovuta ad una approssimazione all'unità che viene effettuata

nel momento in cui viene registrata la scossa.

Si può altresì vedere (*Figura 15*) che se aumenta l'intervallo di campionamento, l'approssimazione diventa impercettibile e si riesce in modo migliore a visualizzare la distribuzione in profondità.

Analizzando l'ultimo grafico in *Figura 13d*, dove si sono presi in considerazione tutti i terremoti con ipocentro maggiore di 40 km di profondità, si ha una graduale

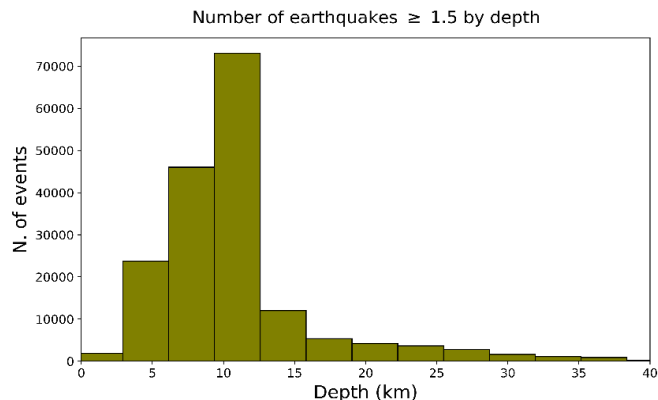


Figura 15 - numero di eventi in profondità con intervallo di campionamento più largo

diminuzione del numero di terremoti. Poiché si hanno due popolazioni di terremoti (quelli più superficiali e quelli del Tirreno Meridionale molto più profondi), si può notare come a partire da circa 100 km di profondità il numero aumenti nuovamente, per poi decadere quasi esponenzialmente: comportamento simile a quello osservato nei primi km di profondità.

4 - ANALISI DELLA GUTENBERG RICHTER

4.1 - LA RELAZIONE DI GUTENBERG-RICHTER

L'equazione di Gutenberg-Richter è una delle relazioni magnitudo-frequenza più importanti nella sismologia. Essa esprime la relazione tra il numero totale di terremoti e la magnitudo in un catalogo sismico che comprende eventi in un particolare intervallo di tempo e in una data area. La relazione può valere sia in forma incrementale, sia in forma cumulativa. Nel primo caso vengono considerati tutti i terremoti di una certa magnitudo, mentre nel secondo tutti i terremoti maggiori di quella magnitudo. Quando si parla di magnitudo, in questo caso si intendono tutti gli eventi compresi all'interno dell'intervallo di campionamento (*bin*). Ad esempio, nel caso di un intervallo di campionamento di 0.2 e per terremoti maggiori di magnitudo 1.5, i valori della frequenza in ordinata saranno calcolati considerando il numero di eventi tra 1.50 e 1.69, tra 1.70 e 1.89, tra 1.90 e 2.09 ecc...

In ordinata la frequenza può essere espressa in vari modi: con il numero medio di terremoti in un anno, con il numero di eventi considerando un dato periodo di tempo oppure come probabilità che un evento con una determinata magnitudo si verifichi (in questo modo si ottengono le PDF, ossia le funzioni di densità di probabilità)

La relazione di Gutenberg-Richter è la seguente:

$$\log(N(M)) = a - bM \quad (4.11)$$

Dove:

- N è il numero di terremoti ad una certa magnitudo (o maggiori di una certa magnitudo se consideriamo la forma cumulativa);
- M è la magnitudo considerata;

- a è un coefficiente denominato anche “produttività” e corrisponde all’intercetta della retta ossia al valore dell’ordinata quando la magnitudo equivale a 0. Varia a livello regionale, dato che dipende dal numero di terremoti presenti nel catalogo totale;
- b corrisponde al coefficiente angolare della retta ed è un parametro che è circa uguale ad 1.

Quest’ultimo è un parametro molto importante poiché ci fornisce informazioni importanti riguardo la frequenza degli eventi. Dato che questo valore rappresenta il coefficiente angolare della retta, esso ci indica la pendenza della retta e ne consegue che un basso valore è associato un largo intervallo di magnitudo, mentre un alto valore abbiamo una maggiore frequenza di eventi per magnitudo basse.

4.2 - STIMA DEL b-VALUE

Nella letteratura scientifica sono stati concepiti molteplici modi per stimare il b-value. Un metodo comune, ma che non dovrebbe essere utilizzato per la stima del parametro è il metodo dei minimi quadrati (OLS). Questo metodo si basa sul ricavare una retta che più si avvicina ai dati riportati sul grafico magnitudo-frequenza. Tale retta è denominata retta di regressione lineare e solitamente viene definita come:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad (4.21)$$

Dove:

- Y_i è la variabile dipendente, corrispondente al logaritmo della frequenza;
- β_0 che equivale all’intercetta;
- β_1 corrispondente al coefficiente angolare della retta;
- X_i è la variabile indipendente, ossia la magnitudo;
- ε_i è l’errore associato alla misura;
- Il pedice i che varia con il numero di osservazioni.

I parametri β_0 e β_1 possono essere determinati in modo da minimizzare la somma dei quadrati dei residui, ovvero la somma delle differenze tra i valori di ordinata reali dei

dati e quelli ottenuti tramite interpolazione. Ciò si effettua ricavando la derivata parziale della somma dei quadrati dei residui rispetto al coefficiente angolare e rispetto all'intercetta. Successivamente si pongono le due derivate parziali uguali a zero (per poter trovare il punto di minimo che minimizza l'errore) e si ottengono i valori:

$$\beta_1 = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2} \quad (4.22)$$

$$\beta_0 = \bar{Y} - \beta_1 \bar{X} \quad (4.23)$$

In cui \bar{X} e \bar{Y} sono rispettivamente la media di tutti i valori X_i e Y_i .

Il metodo dei minimi quadrati è però impreciso nel determinare il b-value e di conseguenza nella stima della retta di regressione lineare. Infatti, le condizioni per poter applicare questa relazione sono:

- La linearità tra le due variabili;
- L'indipendenza: ovvero le osservazioni devono essere indipendenti le une dalle altre;
- Per ogni valore delle ascisse i valori delle ordinate devono essere distribuiti normalmente;
- L'omoschedasticità: per ogni valore della variabile indipendente, la varianza dei residui deve essere la stessa.

Nel caso di studio, la varianza dei residui cambia, come verrà spiegato più avanti, e quindi decade l'ipotesi di omoschedasticità. Inoltre, anche la terza ipotesi decade, dato che i valori delle ordinate non sono distribuiti normalmente ma seguendo una distribuzione di Poisson. Se poi consideriamo i dati della distribuzione cumulativa, anche l'ipotesi di indipendenza non viene rispettata, poiché i dati ottenuti sono influenzati dai dati relativi alle magnitudo più elevate e di conseguenza le variabili sono dipendenti tra loro.

Un metodo molto efficace per la stima del b-value è quello calcolato da Aki nel 1965, che si basa sul metodo della massima verosomiglianza. Si calcola dapprima la funzione di verosomiglianza, in questo caso la probabilità di osservare una data magnitudo al variare del b-value, ossia:

$$L(b) = \prod_{i=1}^N f(M_i | \beta) \quad (4.24)$$

In cui $\beta = \frac{b}{\log_{10} e}$.

Successivamente, si prende il b-value che massimizza questa funzione di verosomiglianza. Il valore ottenuto è il seguente:

$$\hat{b} = \frac{1}{\ln(10)(\mu - M_{thresh})} \quad (4.25)$$

Dove:

- \hat{b} è il b-value ottenuto dalla funzione di massima verosomiglianza;
- μ è la media campionaria delle magnitudo;
- M_{thresh} corrisponde alla minima magnitudo per cui può definirsi completo un catalogo.

La completezza di un catalogo è definita come la magnitudo minima per cui possiamo considerare il catalogo completo: in altre parole al di sopra di questa magnitudo minima si sono registrati tutti i terremoti avvenuti nell'intervallo spazio-temporale che il catalogo copre. Essa può aumentare o diminuire in funzione dell'intervallo temporale e delle coordinate spaziali che si prendono in considerazione.

Per poter applicare la formula la magnitudo massima deve essere molto maggiore di quella minima consentita ($M_{max} - M_{min} \geq 3$).

Si può anche utilizzare una variante di questa formula:

$$\hat{b} = \frac{1}{\ln(10)[\mu - (M_{thresh} - \frac{\Delta M}{2})]} = \frac{\log(e)}{\mu - (M_{thresh} - \frac{\Delta M}{2})} \quad (4.26)$$

Anche se non molto utilizzata, questa relazione esprime il b-value anche in funzione dell'intervallo di campionamento ed è più precisa della prima. Infatti, solitamente la magnitudo minima per cui può definirsi un catalogo completo M_{thresh} è diversa da M_{min} . La M_{thresh} contiene tutti i valori compresi nell'intervallo $M_{thresh} - \frac{\Delta M}{2} \leq M \leq M_{thresh} + \frac{\Delta M}{2}$, mentre $M_{min} = M_{thresh} - \frac{\Delta M}{2}$. Questa variazione aumenta se aumenta l'intervallo di campionamento.

L'errore associato alla stima, secondo Aki (1965), è il seguente:

$$\sigma_{st-err} = \frac{\hat{b}}{\sqrt{N}} \quad (4.27)$$

Un'altra formula è quella ricavata da Shi e Bolt (1982):

$$\sigma_{st-err} = 2.30\hat{b}^2 \sqrt{\frac{\sum_{i=1}^N (M_i - \mu)^2}{N(N-1)}} \quad (4.28)$$

In cui:

- σ_{st-err} è l'errore standard corrispondente al 68% dell'intervallo di confidenza (ad 1 deviazione standard dal b-value calcolato);
- \hat{b} è il b-value calcolato con il metodo della massima verosomiglianza;
- N è il numero di terremoti;
- M_i è la magnitudo del terremoto considerato;
- μ è la media della magnitudo di tutti gli eventi.

L'errore considerato è associato al 68% dell'intervallo di confidenza, corrispondente ad una distanza dal b-value di una deviazione standard, considerando che la distribuzione dell'errore non segue necessariamente una distribuzione Gaussiana (Naylor et al. (2010)).

Dunque, ho importato il catalogo e calcolato il b-value (metodo della massima verosomiglianza) e l'errore associato con il metodo di Aki (1965) e con quello di Shi & Bolt (1982).

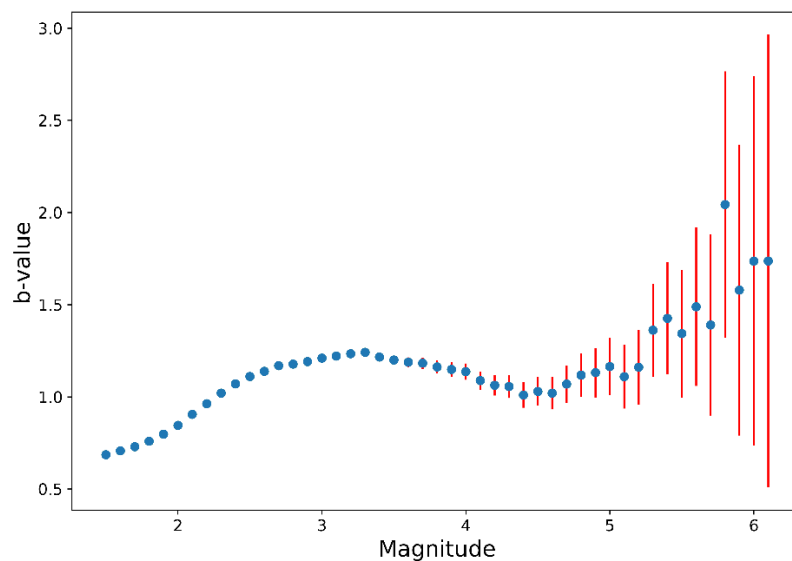


Figura 16 - valori di b-value ottenuti per il catalogo sismico variando la magnitudo di threshold (in blu) ed errori standard calcolati con il metodo di Aki (in rosso, rappresentati con delle barre).

Si osserva come il b-value può variare in modo considerevole a seconda della magnitudo di *threshold* che si prende in considerazione. Per basse magnitudo, a causa dell'incompletezza del catalogo, si hanno valori bassi del b-value. Per magnitudo comprese tra 2.5 e 4 esso si stabilizza tra 1.1 e 1.25 e successivamente varia, subendo dapprima un modesto calo e poi un aumento. L'errore associato (per ragioni di visibilità è stato rappresentato solo quello col metodo di Aki) è quasi nullo per basse magnitudo, poiché diminuisce il numero di eventi registrati e aumenta l'incertezza associata (aumenta il denominatore in ciascuno dei metodi utilizzati per calcolare l'errore standard), mentre aumenta man mano che la magnitudo cresce.

Nella tabella sottostante sono visualizzati tutti i valori calcolati.

Magnitude	B_value	$\sigma_{st-error}$ Aki	$\sigma_{st-error}$ Shi & Bolt				
1.5	0.686260	0.001610	0.001281	4.0	1.136860	0.044387	0.045986
1.6	0.707021	0.001770	0.001398	4.1	1.088261	0.049415	0.049440
1.7	0.729553	0.001954	0.001528	4.2	1.062344	0.055303	0.053993
1.8	0.758944	0.002171	0.001696	4.3	1.055845	0.062325	0.060156
1.9	0.797301	0.002434	0.001919	4.4	1.009770	0.068706	0.062386
2.0	0.845379	0.002755	0.002220	4.5	1.029526	0.078048	0.070720
2.1	0.904676	0.003152	0.002633	4.6	1.020104	0.087473	0.076200
2.2	0.963178	0.003621	0.003130	4.7	1.069033	0.101014	0.089755
2.3	1.019706	0.004177	0.003731	4.8	1.117986	0.117197	0.106381
2.4	1.070730	0.004829	0.004441	4.9	1.131556	0.134291	0.120195
2.5	1.110043	0.005576	0.005236	5.0	1.163660	0.155501	0.139280
2.6	1.138723	0.006432	0.006122	5.1	1.109413	0.173261	0.140145
2.7	1.169282	0.007448	0.007223	5.2	1.160463	0.202011	0.161208
2.8	1.177423	0.008561	0.008316	5.3	1.361572	0.252838	0.230945
2.9	1.190919	0.009875	0.009662	5.4	1.426042	0.304033	0.284935
3.0	1.210169	0.011438	0.011358	5.5	1.343179	0.346807	0.293321
3.1	1.221397	0.013228	0.013290	5.6	1.489010	0.429840	0.389266
3.2	1.233552	0.015325	0.015647	5.7	1.389742	0.491348	0.384131
3.3	1.241494	0.017745	0.018431	5.8	2.043739	0.722571	0.830734
3.4	1.215782	0.020252	0.020883	5.9	1.579253	0.789626	0.754309
3.5	1.200024	0.023146	0.023836	6.0	1.737178	1.002960	1.060242
3.6	1.187678	0.026445	0.027259	6.1	1.737178	1.228370	1.388182
3.7	1.182608	0.030274	0.031446				
3.8	1.162662	0.034390	0.035579				
3.9	1.148899	0.039086	0.040440				

Tabella 1 - valori di b-value, errore standard con il metodo di Aki, di Shi & Bolt per ogni magnitudo

Da qui è possibile osservare che i valori ottenuti per i due errori sono molto simili soprattutto per basse magnitudo mentre, se aumenta la magnitudo, essi tendono a discostarsi a causa della più elevata incertezza.

4.3 - STIMA DELLA DISTRIBUZIONE GUTENBERG-RICHTER

Dopo aver stimato il valore del b-value, ho proceduto nella rappresentazione della Gutenberg-Richter, sia con ZMAP, sia con le librerie Matplotlib e Pandas di Python.

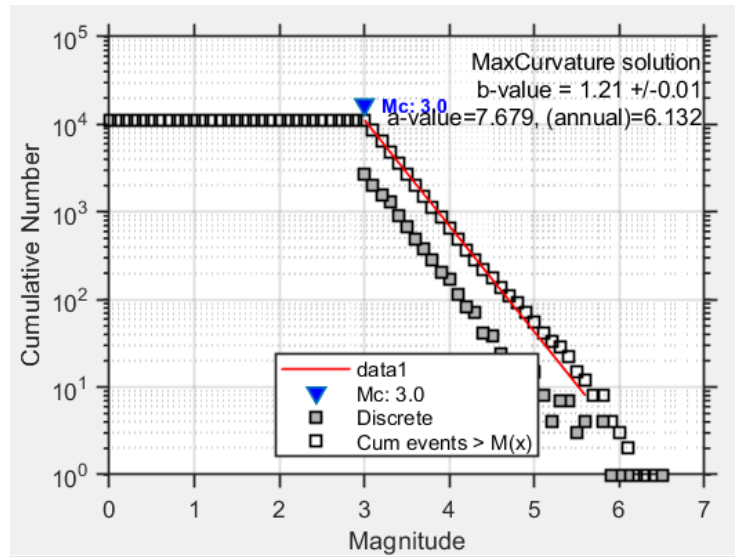


Figura 17 - Gutenberg-Richter calcolata con il software ZMAP per magnitudo maggiori di 3, cumulativi e con un binning di 0.1

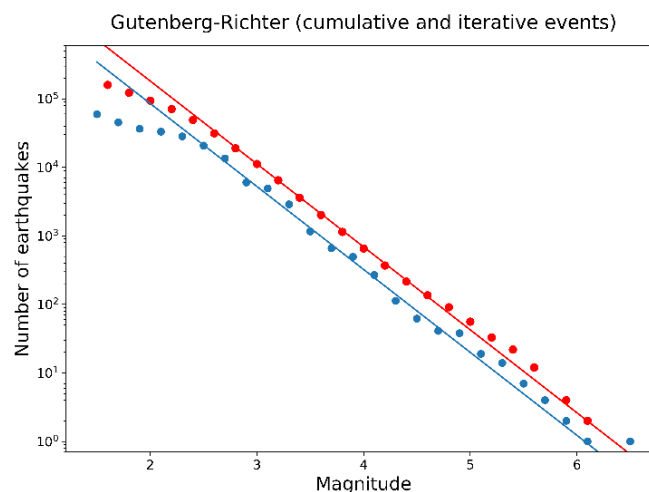


Figura 18 - Gutenberg-Richter con medesimo b-value utilizzato nella Figura 17, con un binning di 0.2 e per valori sia cumulativi, sia incrementali con la libreria Matplotlib di Python

Nella Figura 17, i dati sono stati rappresentati nel grafico dal programma e sono stati calcolati anche il b-value e il valore dell'intercetta (a-value). I terremoti che ho preso in considerazione sono tutti quelli superiori a magnitudo 3, sia perché i valori del b-value sono stabili in corrispondenza di quella magnitudo come si osserva dalla Figura 16 (perciò non avrebbe senso considerare tutti i terremoti nel catalogo), sia perché il programma al di sopra dei 100.000 eventi subisce dei bug e dei rallentamenti evidenti che non permettono di visualizzare il grafico ottenuto (probabilmente anche a causa della poca memoria RAM a disposizione). La retta rossa corrisponde all'equazione:

$$\log(N(M)) = 7.679 - 1.21M \quad (4.31)$$

L'equazione della retta è stata ricavata basandosi sulla curva cumulativa ed è stata riportata allo stesso modo anche nella Figura 18, dove il grafico raffigura i valori ogni 0.2 decimi di magnitudo. Come si può osservare, in questo caso la retta di interpolazione dei valori incrementali è più distanziata rispetto a quella dei cumulativi, dal momento che i valori incrementali nel caso della Figura 18 prendono in considerazione un intervallo di campionamento più ampio (il doppio) e dunque ne risulta una frequenza più elevata dei valori incrementali, più ravvicinati ai valori cumulativi.

4.4 - ANALISI PROBABILISTICHE E TEST STATISTICI SULLA GUTENBERG-RICHTER

Per comprendere meglio la variazione dei dati lungo la Gutenberg-Richter si può ricorrere a dei cataloghi sintetici. Studiando la distribuzione dei dati in questi cataloghi sintetici, si riesce a studiare la distribuzione del numero di terremoti per una data equazione di Gutenberg-Richter e una data magnitudo, ma si ha anche la possibilità di compiere test statistici su queste distribuzioni. Per generare i cataloghi sintetici, ho utilizzato lo stesso b-value calcolato per il catalogo sismico italiano con lo scopo di calcolare la probabilità che si verifichi un evento con una determinata magnitudo. Ho ripetuto ciò per tutte le magnitudo comprese tra 1.5 e 6.5. Le probabilità ottenute, le quali sono indipendenti dall'a-value del catalogo ma dipendenti dal b-value (dato che se il b-value è un valore più alto gli eventi con più basse magnitudo sono più probabili, al contrario lo sono meno), sono state calcolate nel modo seguente:

$$p_M = \frac{f_M}{u} = \frac{\text{numero di eventi con magnitudo } M}{\text{numero di eventi totali}} \quad (4.41)$$

I valori di p_M che ci forniscono la probabilità che si verifichi un evento con una determinata magnitudo M , sono stati utilizzati per generare 1000 cataloghi sintetici, ciascuno con 100000 eventi sia in forma cumulativa, sia in forma incrementale. Nel

passo successivo ho trasformato i dati ottenuti dal formato `array` della libreria Numpy, al formato `dataframe` della libreria Pandas ed ho effettuato ulteriori procedimenti di *data manipulation*⁴, racchiudendo tutti i dati ottenuti in tabelle facilmente interpretabili.

	1.5	1.6	1.7	1.8	1.9	2.0	2.1	2.2	2.3	2.4	...	5.6	5.7	5.8	5.9	6.0	6.1	6.2	6.3	6.4	6.5
1	24374.0	18377.0	13864.0	10465.0	7989.0	6114.0	4542.0	3472.0	2630.0	1962.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	24346.0	18449.0	14065.0	10509.0	8121.0	6012.0	4507.0	3364.0	2545.0	1978.0	...	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	1.0	0.0
3	24305.0	18390.0	14025.0	10650.0	8033.0	6030.0	4486.0	3519.0	2648.0	1927.0	...	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0
4	24272.0	18509.0	13846.0	10532.0	7969.0	6018.0	4480.0	3485.0	2683.0	1954.0	...	0.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
5	24433.0	18478.0	13892.0	10382.0	7864.0	6120.0	4574.0	3470.0	2591.0	1987.0	...	0.0	1.0	0.0	1.0	0.0	0.0	1.0	0.0	0.0	0.0
...
996	24455.0	18188.0	13934.0	10594.0	8034.0	6090.0	4580.0	3448.0	2552.0	2012.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
997	24299.0	18456.0	14226.0	10435.0	8156.0	5888.0	4527.0	3477.0	2630.0	1946.0	...	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0
998	24249.0	18499.0	13942.0	10547.0	7784.0	6021.0	4729.0	3442.0	2598.0	1975.0	...	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0
999	24471.0	18279.0	13969.0	10565.0	7908.0	6032.0	4647.0	3506.0	2612.0	1931.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1000	24472.0	18479.0	13807.0	10462.0	7966.0	6036.0	4587.0	3400.0	2508.0	2036.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

1000 rows × 51 columns

Tabella 2 - numero di eventi incrementali calcolati per ogni magnitudo e per 1000 cataloghi generati sinteticamente, ciascuno avente 100000 eventi al suo interno

	1.5	1.6	1.7	1.8	1.9	2.0	2.1	2.2	2.3	2.4	...	5.6	5.7	5.8	5.9	6.0	6.1	6.2	6.3	6.4	6.5
1	100000.0	75577.0	57208.0	43333.0	32671.0	24683.0	18630.0	14127.0	10761.0	8118.0	...	1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	100000.0	75939.0	57537.0	43636.0	32973.0	24872.0	18809.0	14233.0	10710.0	8056.0	...	3.0	2.0	2.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0
3	100000.0	75545.0	57165.0	43333.0	32831.0	24884.0	18771.0	14326.0	10870.0	8151.0	...	4.0	3.0	2.0	2.0	2.0	2.0	2.0	2.0	0.0	0.0
4	100000.0	76007.0	57686.0	43581.0	32953.0	25002.0	18793.0	14206.0	10779.0	8170.0	...	4.0	3.0	3.0	2.0	1.0	1.0	0.0	0.0	0.0	0.0
5	100000.0	75512.0	57164.0	43402.0	32952.0	24852.0	18779.0	14199.0	10806.0	8209.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
...
996	100000.0	75644.0	57164.0	43370.0	32905.0	24901.0	18941.0	14241.0	10803.0	8205.0	...	2.0	2.0	1.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0
997	100000.0	75971.0	57623.0	43742.0	33042.0	24965.0	18967.0	14404.0	10918.0	8340.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
998	100000.0	75673.0	57224.0	43269.0	32678.0	24640.0	18500.0	13965.0	10579.0	7943.0	...	2.0	2.0	2.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0
999	100000.0	75759.0	57524.0	43510.0	32872.0	24777.0	18677.0	14100.0	10757.0	8087.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1000	100000.0	75506.0	57185.0	43050.0	32563.0	24567.0	18538.0	14060.0	10656.0	8102.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

1000 rows × 51 columns

Tabella 3 - numero di eventi cumulativi calcolati per ogni classe di magnitudo (da 1.5 a 6.5) di 1000 cataloghi generati sinteticamente, ciascuno con 100000 eventi

⁴ Consiste nella “manipolazione” dei dati, ossia nella modifica di essi all’interno di un database e nell’eliminazione o aggiunta di nuovi parametri.

Una volta generati, ho prelevato due cataloghi sintetici tra i 1000 presenti lungo le righe delle due tabelle e ho calcolato i b-value con i rispettivi errori per diverse magnitudo di cut-off.

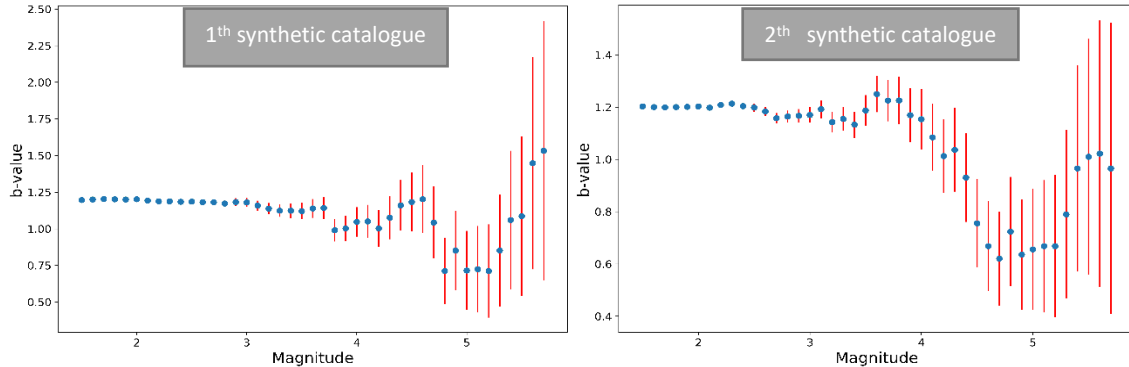


Figura 19 - valori di b-value ottenuti per due cataloghi sintetici variando la magnitudo di threshold (in blu) ed errori standard calcolati con il metodo di Aki (rappresentati in rosso con delle barre).

Si può osservare come questi due cataloghi sintetici siano differenti da quello ottenuto con i dati reali nella Figura 16; infatti nel caso dei cataloghi sintetici non si ha un valore al di sopra del quale il catalogo può essere definito completo, essendo stati generati direttamente dalla Gutenberg-Richter. Inoltre, anche l'influenza dell'errore strumentale può far variare i valori effettivi.

In Figura 20 sono invece rappresentate le forme incrementali e cumulative dei due cataloghi in un grafico magnitudo-frequenza.

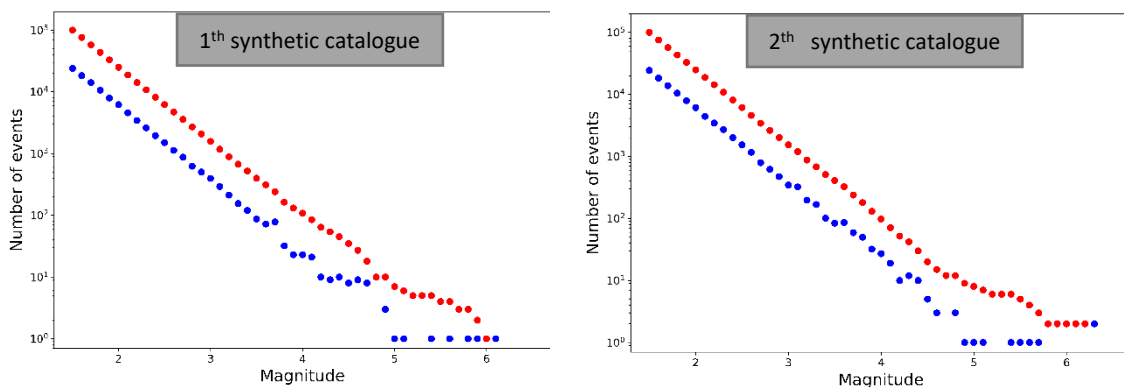


Figura 20 - grafici magnitudo-frequenza per due cataloghi sintetici. In rosso sono raffigurate le forme cumulative, in blu quelle incrementali

Una volta che si ha a disposizione un numero elevato di cataloghi ciascuno contenente un numero molto grande di eventi, è possibile effettuare ulteriori analisi statistiche sulla

distribuzione del numero di eventi per ogni magnitudo, calcolata per tutti i cataloghi sintetici. Per effettuare ciò si deve supporre che i dati sulla frequenza del numero di eventi siano distribuiti secondo una distribuzione di Poisson, in cui sull'ascissa abbiamo il numero di eventi e in ordinata la frequenza per ogni numero di eventi.

Essa ha un'equazione del tipo:

$$p(X = x) = \frac{\lambda^x e^{-\lambda}}{x!} \quad (4.42)$$

Dove:

- λ è la media del numero di eventi;
- x è il numero di eventi;
- $e \approx 2.71828$ è il numero di Nepero;
- $p(X = x)$ è la probabilità che il numero di eventi x si verifichi.

Per poter visualizzare la dispersione dei dati si può determinare un intervallo di confidenza, ovvero un intervallo di valori in cui è più frequente trovare il numero di eventi per ciascuna magnitudo. Nel caso in questione, ho calcolato il 90% dell'intervallo di confidenza con limite inferiore al 5% di probabilità di incontrare un determinato numero di eventi e superiore al 95%. Ho calcolato questi valori per ogni magnitudo e ho utilizzato la funzione `poisson.ppf(q, μ)` all'interno della libreria `scipy.stats` di Python. I parametri passati alla funzione sono q e μ , dove q è la probabilità che si abbia un numero di eventi inferiore al limite inferiore dell'intervallo di confidenza e μ (ossia λ) corrispondente alla somma degli eventi diviso il numero di cataloghi. La funzione deve essere applicata anche per determinare la probabilità relativa al limite superiore dell'intervallo di confidenza. Si dimostra che λ è uguale alla somma degli eventi diviso il numero N di cataloghi osservati, attraverso la funzione di massima verosimiglianza (in Naylor et al. (2010) viene semplicemente citato questo metodo ma non viene dimostrato)

$$L(\lambda) = \prod_{i=1}^N f(x_i | \lambda) \quad (4.43)$$

Si ricava quindi la funzione di massima somiglianza a partire dall'equazione della distribuzione di Poisson e successivamente si trasforma la produttoria delle probabilità calcolate per ciascun catalogo N in sommatorie, applicando il logaritmo (in modo che successivamente sia più semplice risolvere la derivata prima):

$$L(\lambda) = \prod_{i=1}^N \frac{\lambda^{x_i} e^{-\lambda}}{x_i!} = \sum_{i=1}^N \log \frac{\lambda^{x_i} e^{-\lambda}}{x_i!} \quad (4.44)$$

Nel passaggio successivo, per le proprietà dei logaritmi, i prodotti e le divisioni diventano rispettivamente somme e differenze tra valori.

$$L(\lambda) = \sum_{i=1}^N [X_i \log(\lambda) - \lambda - \log(X_i!)] = \quad (4.45)$$

$$\log(\lambda) \sum_{i=1}^N X_i - N\lambda - \sum_{i=1}^N \log(X_i!) \quad (4.46)$$

Applicando la derivata prima e considerando che $\sum_{i=1}^N \log(X_i!)$ è una costante (poiché X_i è un valore fisso che indica il numero di eventi) e per questo motivo si annulla, si ottiene:

$$L'(\lambda) = \frac{1}{\lambda} \sum_{i=1}^N X_i - N = 0 \quad (4.47)$$

Ricavando λ si ottiene:

$$\hat{\lambda} = \frac{\sum_{i=1}^N X_i}{N} \quad (4.48)$$

Ho utilizzato il valore di λ ottenuto per calcolare l'intervallo di confidenza tra il 5° e il 95° percentile, ottenendo:

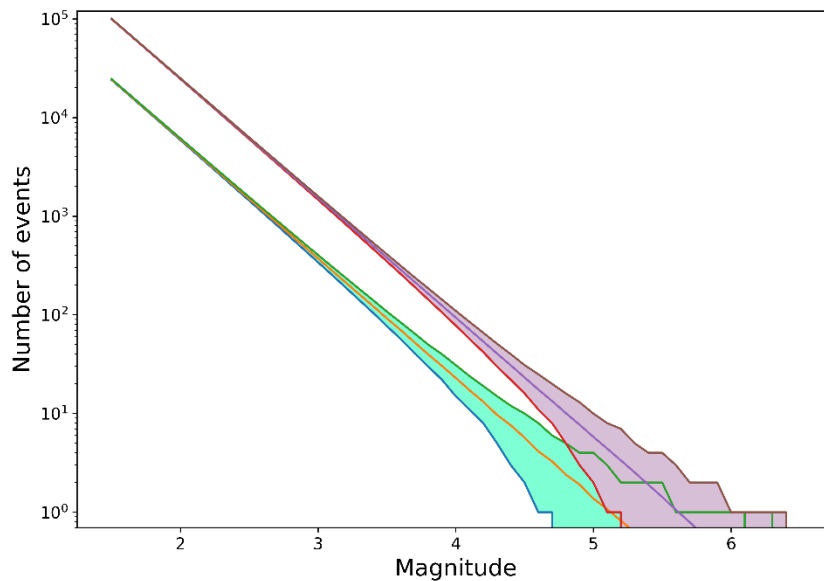


Figura 21 - localizzazione dell'intervallo compreso tra il 5° e il 95° della distribuzione Poisson del numero di eventi. In turchese si ha l'area compresa in questo intervallo per la forma incrementale, in violetto l'area compresa nell'intervallo di confidenza per la forma cumulativa.

Dalla Figura 21 è evidente come l'intervallo di confidenza vari notevolmente. Per basse magnitudi si osserva come è più ristretto questo *range* dal momento che si ha un numero esponenzialmente più elevato di eventi. Al contrario, per elevate magnitudi, diminuendo il numero di terremoti analizzati e contemporaneamente aumentando l'incertezza, i valori di frequenza sono più ampi.

Nella Figura 22 sono raffigurati i due cataloghi sintetici con il relativo intervallo di confidenza.

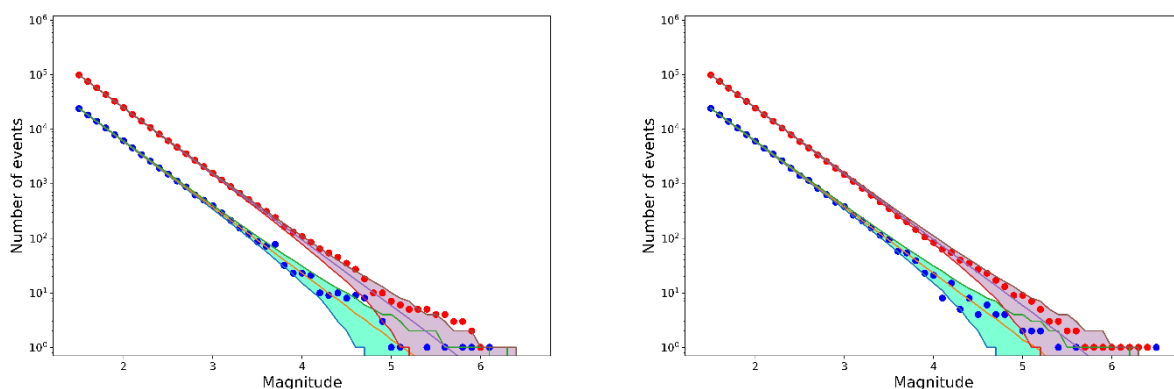


Figura 22 - localizzazione dell'intervallo di confidenza. In turchese si ha l'area compresa in questo intervallo per la forma incrementale, in violetto l'area compresa nell'intervallo di confidenza per la forma cumulativa.

Per riuscire a percepire meglio l'eteroschedasticità, ossia la variazione della deviazione standard (e della varianza) al variare della magnitudo, è possibile rappresentare degli istogrammi, in cui sull'ascissa si ha il numero di eventi e in ordinata la frequenza del numero di eventi. Ho effettuato ciò prendendo in considerazione la distribuzione per magnitudo pari a 1.5, 4.1, 4.8 e 5.6.

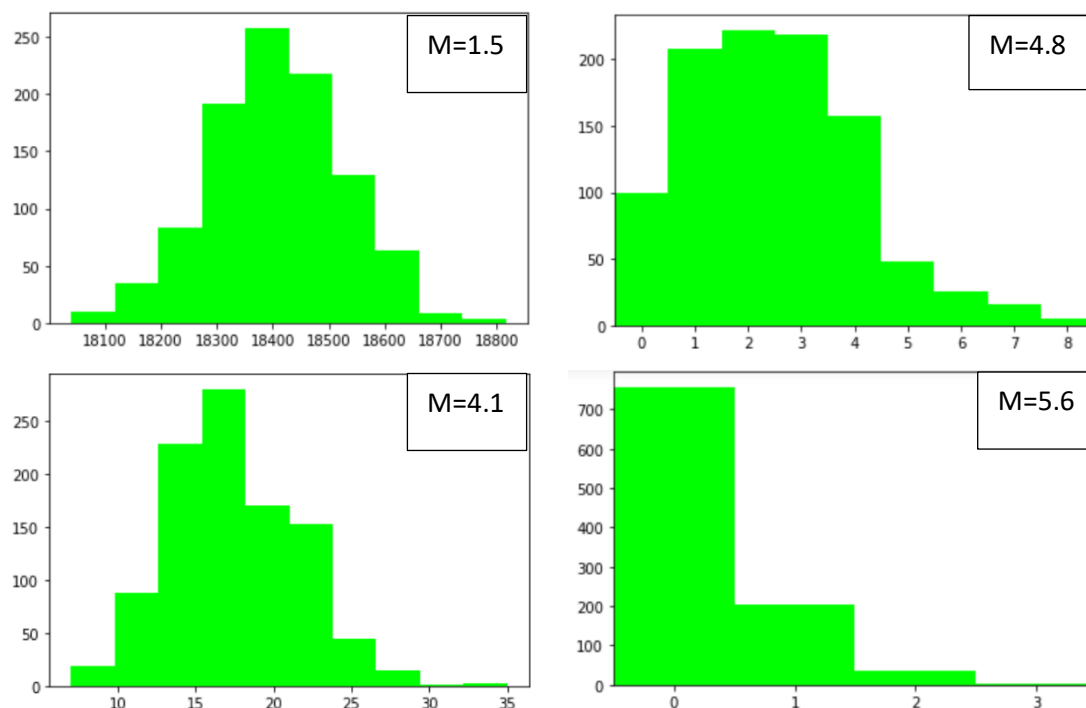


Figura 23 - distribuzione del numero di eventi per magnitudo di 1.5, 4.1, 4.8 e 5.6

In Figura 23 è evidente come vari la distribuzione. Nel caso dei terremoti di magnitudo 1.5 l'istogramma segue una curva Gaussiana, simmetrica da entrambi i lati. Nel caso dei terremoti di magnitudo pari a 4.1 si comincia a scorgere un certo *skewness*

nella distribuzione dei dati. Lo *skewness* è un indice che definisce quanto sia asimmetrica una determinata distribuzione.

Esso si calcola utilizzando la seguente formula:

$$S_K = \frac{\sum_i^N (X_i - \bar{\mu})}{(N - 1)\sigma^3} \quad (4.49)$$

$$\text{Con: } \sigma = \sqrt{\frac{\sum_i^N (X_i - \bar{\mu})^2}{(N-1)}}$$

In cui σ è la deviazione standard campionaria, $\bar{\mu}$ la media, N è il numero di cataloghi e X_i il numero di eventi in un determinato catalogo. Valori intorno a 0 indicano una distribuzione simmetrica, valori negativi molto minori di 0 (*negative skew*), ci suggeriscono una tendenza dei dati a concentrarsi verso valori più alti, mentre risultati positivi (*positive skew*), indicano un picco di osservazioni per bassi valori delle ascisse.

Tornando alla Figura 23, si nota che per magnitudo pari a 4.8 e 5.6 la distribuzione è molto differente da una Gaussiana e gli istogrammi seguono sempre più una distribuzione Poisson classica. Ciò si può spiegare nel modo seguente: più aumentano i dati a disposizione (il numero di eventi osservato, da non confondere con il numero di cataloghi) più aumenta λ e più la distribuzione Poisson può essere approssimata ad una Gaussiana. In simboli:

$$\text{per } x \rightarrow \infty \Rightarrow \lambda \rightarrow \infty$$

↓

$$f(x) = \frac{\lambda^x e^{-\lambda}}{x!} \approx \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x - \bar{\mu})^2}{2\sigma^2}\right] \quad (4.410)$$

dove l'ultimo membro è l'equazione della distribuzione Gaussiana.

Un test spesso utilizzato per verificare se la distribuzione dei dati segue una distribuzione normale, è il test di Kolmogorov-Smirnov. Per poter effettuare questo test

si considerano i dati del numero di eventi standardizzati, ossia ordinati in modo che la media sia pari a 0 e la deviazione standard pari a 1. Per compiere questa operazione ho applicato la seguente formula:

$$Z_i = \frac{X_i - \mu}{\sigma} \quad (4.411)$$

Dove X_i è il valore che si vuole convertire, μ è la media, σ la deviazione standard e Z_i il valore X_i standardizzato.

A partire da questi dati standardizzati si costruiscono le CDF (*cumulative density functions*, le funzioni di ripartizioni o funzioni cumulate). Perciò dagli istogrammi ottenuti per ogni magnitudo, si calcolano i valori delle CDF partendo dal numero di eventi più basso fino ad arrivare a quello più alto, sommando man mano i valori precedenti minori o uguali.

Si costruisce così un grafico con il numero di eventi sulle ascisse e la probabilità cumulata sulle ordinate. Dopo aver rappresentato la CDF della distribuzione empirica, viene rappresentata la funzione cumulata della curva teorica per cui si vuole effettuare il test (nel caso del catalogo, si rappresenta la funzione di ripartizione della Gaussiana).

Una volta rappresentate le due funzioni si calcola:

$$D_n = \sup_x |F_n(x) - F(x)| \quad (4.412)$$

In cui:

- $F_n(x)$ è il valore in ordinata della funzione sperimentale;
- $F(x)$ è il valore in ordinata della funzione teorica.

Si calcola D_n considerando l'estremo superiore del valore assoluto della differenza tra la funzione CDF sperimentale e quella teorica. Con Python non è necessario effettuare questi passaggi e richiamando la libreria `scipy.stats`, basta applicare la funzione `kstest(x, dist.)`, in cui come argomento ho inserito i dati standardizzati e la

stringa 'norm'. Ciò permette di effettuare il test considerando come funzione teorica la distribuzione normale.

Per lo *skewness* ho adoperato la funzione `skew(x)` dove nell'argomento ho inserito i dati non standardizzati.

I valori ottenuti per lo *skewness* e il test di Kolmogorov-Smirnov (KS test) sono stati rielaborati in formato `dataframe` e visibili nella Tabella 4.

Magnitude	KS_test	p-value	Skewness	4.2	0.074343	2.978903e-05	0.387578
1.5	0.013494	9.932914e-01	-0.027819	4.3	0.084084	1.340627e-06	0.187324
1.6	0.014802	9.808003e-01	0.007952	4.4	0.098347	7.163147e-09	0.340189
1.7	0.022897	6.708743e-01	0.150867	4.5	0.115079	5.409877e-12	0.500547
1.8	0.014290	9.868076e-01	-0.051815	4.6	0.141100	7.824873e-18	0.540921
1.9	0.014781	9.810754e-01	-0.010402	4.7	0.131749	1.350612e-15	0.496530
2.0	0.023507	6.383821e-01	0.024368	4.8	0.146370	3.664455e-19	0.658550
2.1	0.022016	7.175397e-01	0.079084	4.9	0.181332	3.032335e-29	0.646976
2.2	0.036220	1.415670e-01	-0.048392	5.0	0.215178	4.031371e-41	0.797469
2.3	0.019348	8.482159e-01	0.022039	5.1	0.238038	2.437411e-50	1.067624
2.4	0.018606	8.793121e-01	-0.035535	5.2	0.241464	8.337047e-52	1.011539
2.5	0.023153	6.572262e-01	0.023920	5.3	0.326586	1.787124e-95	1.214788
2.6	0.015823	9.637115e-01	-0.011666	5.4	0.373215	1.509310e-125	1.431748
2.7	0.034313	1.855322e-01	0.114523	5.5	0.439307	3.392530e-176	1.734529
2.8	0.033419	2.095336e-01	-0.105112	5.6	0.455374	5.153172e-190	2.040446
2.9	0.026643	4.750926e-01	0.014474	5.7	0.492335	3.496834e-224	2.181094
3.0	0.020614	7.891107e-01	-0.022477	5.8	0.501455	3.827521e-233	2.317372
3.1	0.027412	4.369539e-01	0.077585	5.9	0.525332	1.218014e-257	3.130761
3.2	0.028942	3.673662e-01	0.051875	6.0	0.536584	1.008874e-269	3.625914
3.3	0.031497	2.692819e-01	0.007265	6.1	0.535224	3.040419e-268	3.771272
3.4	0.029156	3.582936e-01	0.114589	6.2	0.539937	2.162675e-273	3.781924
3.5	0.030229	3.151886e-01	0.099948	6.3	0.537791	4.860457e-271	5.649292
3.6	0.050600	1.152659e-02	0.268643	6.4	0.538278	1.426440e-271	6.176931
3.7	0.057563	2.539895e-03	0.157334	6.5	0.536314	1.984660e-269	7.046343
3.8	0.047710	2.039436e-02	0.150258				
3.9	0.048473	1.760055e-02	0.036292				
4.0	0.054860	4.675797e-03	0.178349				
4.1	0.067599	2.037780e-04	0.275436				

Tabella 4 - elenco dei valori ottenuti per ogni magnitudo dal test di Kolmogorov-Smirnov (a sinistra), con rispettivo p-value associato al test KS (al centro) e valore dello skewness (a destra)

D_n è una grandezza che varia tra 0 e 1. Se il valore è molto vicino a 0, la distribuzione può essere approssimata ad una Gaussiana, al contrario vicino ad 1 si ha un istogramma che non può essere affatto associato ad una distribuzione normale. Il p-value ricavato a partire da D_n , al contrario, per valori vicini ad 1 ci fornisce una stima affidabile per poter considerare la distribuzione Gaussiana, mentre per valori più vicini a 0 la probabilità che i dati siano distribuiti normalmente diviene molto bassa.

Questi valori ottenuti possono essere presi in considerazione per poter effettuare dei test inferenziali. Ad esempio, si potrebbe definire una soglia al di sotto del quale si rifiuta l'ipotesi che i dati seguano una distribuzione normale e al di sopra del quale si ritenga che essi siano normalmente distribuiti.

Una soglia spesso utilizzata corrisponde ad un p-value di 0.95. Se ci si basa su questo limite, si ha che solamente gli eventi di magnitudo 1.5, 1.6, 1.8, 1.9 e 2.6 hanno una distribuzione del numero di eventi approssimabile ad una Gaussiana, mentre i restanti sono riconducibili ad una distribuzione Poisson. Questi dati non devono essere presi come dati effettivi, poiché i cataloghi sono stati generati sinteticamente, ma servono per una stima approssimativa dei risultati ottenuti.

Infatti, per basse magnitudo, si ha una tendenza all'accettazione dell'ipotesi di normalità, al contrario per alte magnitudo il p-value diviene bassissimo ed è praticamente impossibile che i dati siano normalmente distribuiti.

I risultati ottenuti per lo *skewness* indicano invece una grande asimmetria nei dati per magnitudo più elevate, mentre per basse magnitudo i dati sono simmetrici o con un leggero *skew* talvolta anche negativo. Si hanno dunque dei risultati che ci segnalano un alto grado di simmetria fino a magnitudo comprese tra 3.0 e 3.5. Per magnitudo superiori a 5.9, lo *skewness* è perfino maggiore di 3 e ciò indica che si ha una completa asimmetria dei dati.

5 - ANALISI TEMPORALE

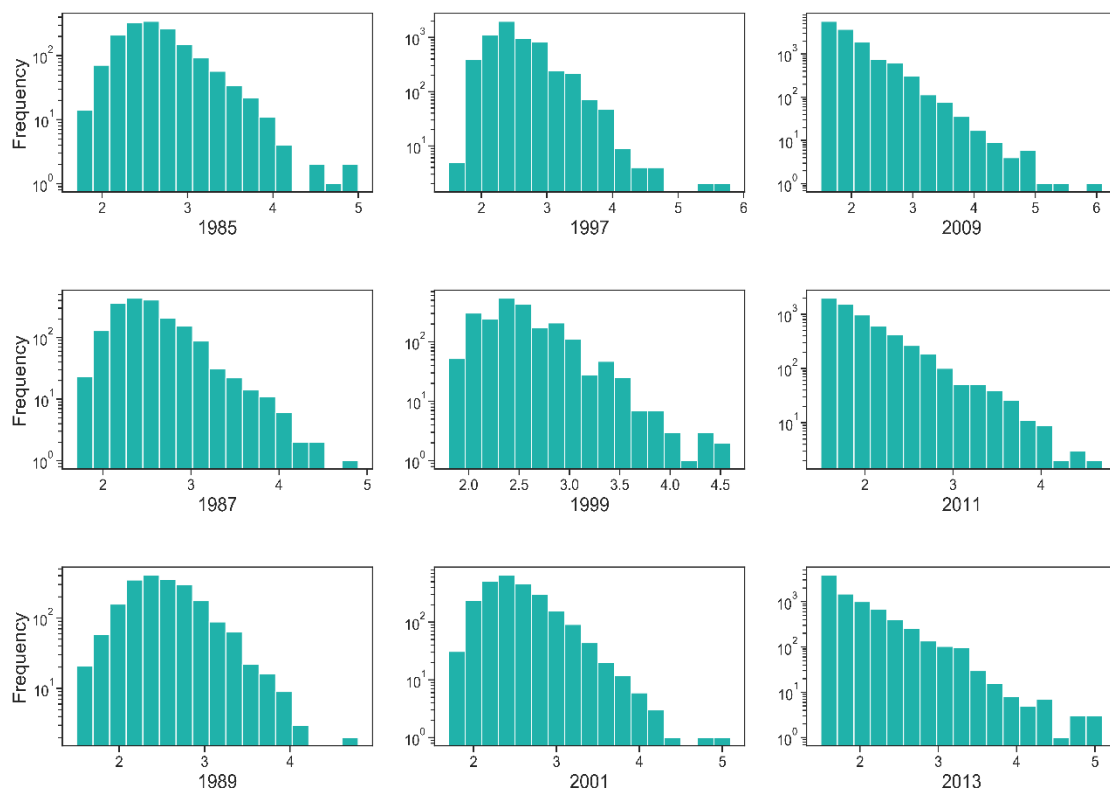
Fino a questo momento si sono illustrate le analisi che ho effettuato dal punto di vista della magnitudo, del numero di eventi, dello spazio (in profondità, latitudine e longitudine) e del tipo di magnitudo. Da questo momento si considereranno le analisi realizzate considerando il tempo o degli intervalli temporali.

5.1 - ISTOGRAMMI MAGNITUDO-FREQUENZA NEL TEMPO

Oltre a rappresentare la Gutenberg-Richter, è possibile ricorrere ad istogrammi che permettono di osservare come varia la distribuzione delle magnitudo negli anni con il fine principale di poter osservare la variazione della completezza negli anni.

Nel grafico sottostante (Figura 24) ho preso in considerazione gli anni dispari nel catalogo (per ragioni di spazio) e ho prodotto gli istogrammi magnitudo-frequenza per ciascuno di questi anni.

Completeness by year



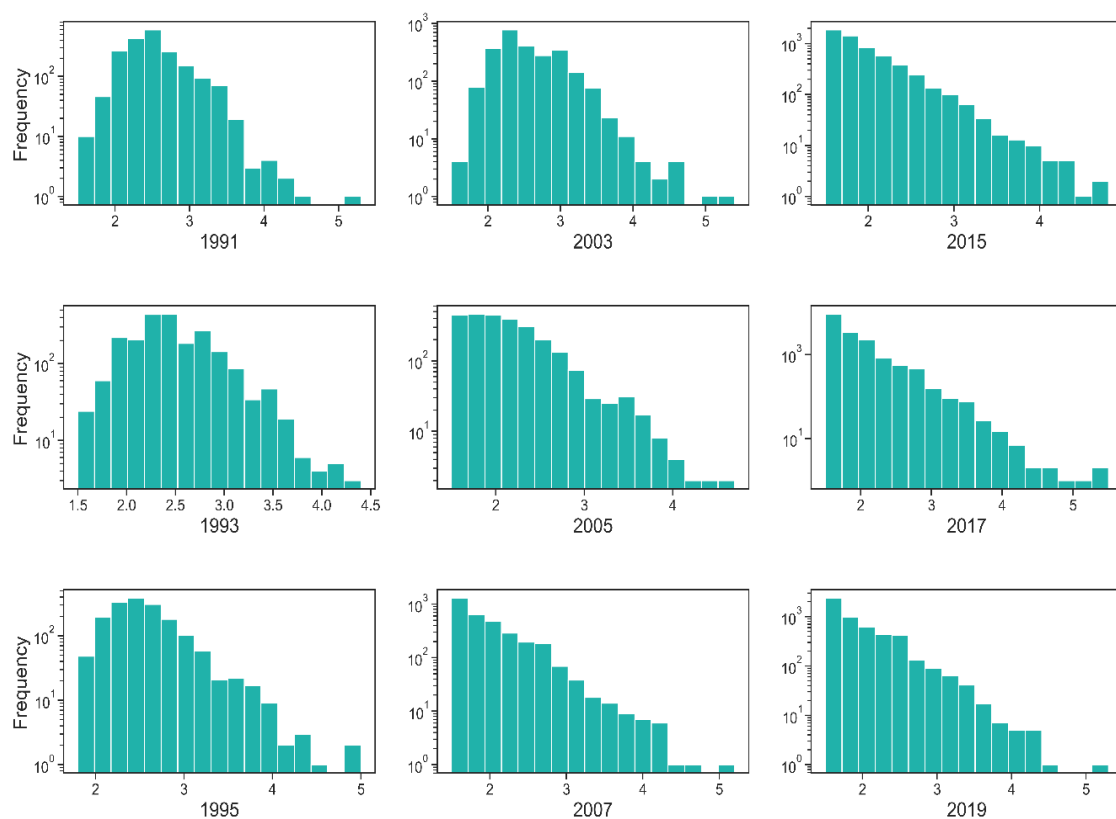


Figura 24 - istogrammi magnitudo-frequenza nel tempo.

È evidente che, osservando i grafici nel tempo, la completezza partendo dal 1985 e arrivando al 2019, ha un limite inferiore molto più basso man mano che si arriva a tempi più recenti. È altresì possibile vedere questa variazione considerando soltanto tre cataloghi lungo una riga orizzontale, ossia con un intervallo tra un grafico ed un altro di 12 anni.

L'aumento della completezza è relativo al fatto che gli strumenti col passare degli anni hanno aumentato la loro sensibilità. Infatti si può osservare come nel 1985 la completezza sia di poco inferiore a 5, mentre nel 2019 tutto il catalogo è completo.

5.2 - GRAFICI TEMPO-FREQUENZA

Una semplice analisi che è possibile effettuare è quella del numero di eventi nel tempo. Questo ci permette di individuare i terremoti più importanti avvenuti in questo periodo di tempo e come varia il numero di eventi negli anni.

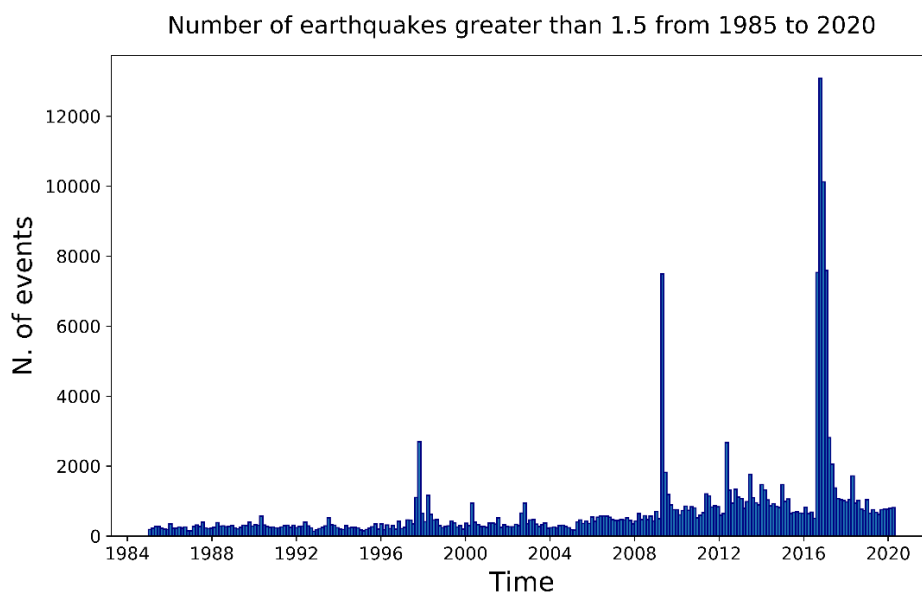


Figura 25 - numero di eventi nel tempo

Nella Figura 25, si vedono in modo ancora più evidente rispetto alla Figura 6 la sequenza sismica dell'Aquila (a partire dal 6 aprile 2009), la sequenza sismica del Centro Italia (con i *mainshocks* del 24 agosto 2016 - 26 e il 30 ottobre 2016 - 18 gennaio 2017), di Colfiorito (con la scossa principale del 26 settembre 1997) e in misura minore la sequenza sismica dell'Emilia (maggio 2012) e quella del Molise (a partire dal 31 ottobre 2002).

Il fatto che il picco che si ha in seguito alle scosse del terremoto di Colfiorito sia meno accentuato è dovuto alla completezza. Infatti (come si osserva dalla Figura 24 in corrispondenza dell'anno 1997) la bassa completezza del catalogo implica che non sono stati registrati tutti gli eventi di bassa magnitudo (che sono proprio quelli più frequenti). Anche osservando nel tempo tutte le altre scosse, ci si rende conto che gli eventi registrati sono più del doppio nel 2020 rispetto al 1985.

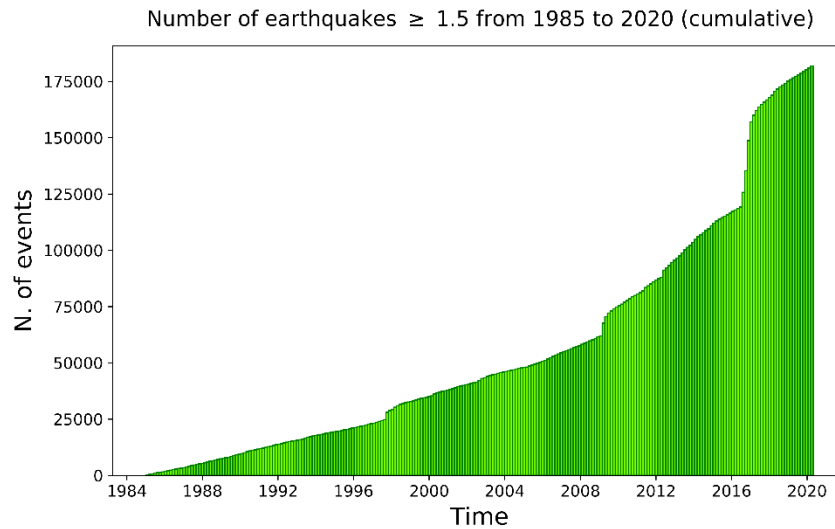


Figura 26 - numero di eventi nel tempo in forma cumulativa

Nella figura 26 ho rappresentato gli eventi in forma cumulativa. Per ogni *bin* si sono sommati gli eventi avvenuti in quel momento e tutti quelli precedenti. In questo grafico è la pendenza che indica quanti eventi sono stati registrati in quell'intervallo di tempo. Si possono notare pendenze elevate in corrispondenza delle principali sequenze sismiche. Inoltre, seguendo l'incremento della pendenza del diagramma negli anni, anche da questa rappresentazione si osserva un graduale aumento degli eventi registrati negli anni.

5.3 – ANALISI DEGLI *INTEREVENT TIMES*

Gli *interevent times* sono le differenze tra i tempi di occorrenza di due eventi successivi. Ho rappresentato questi attraverso degli istogrammi in cui sulle ascisse ho riportato la differenza tra due eventi consecutivi e sulle ordinate la frequenza.

Frequency by interevent time

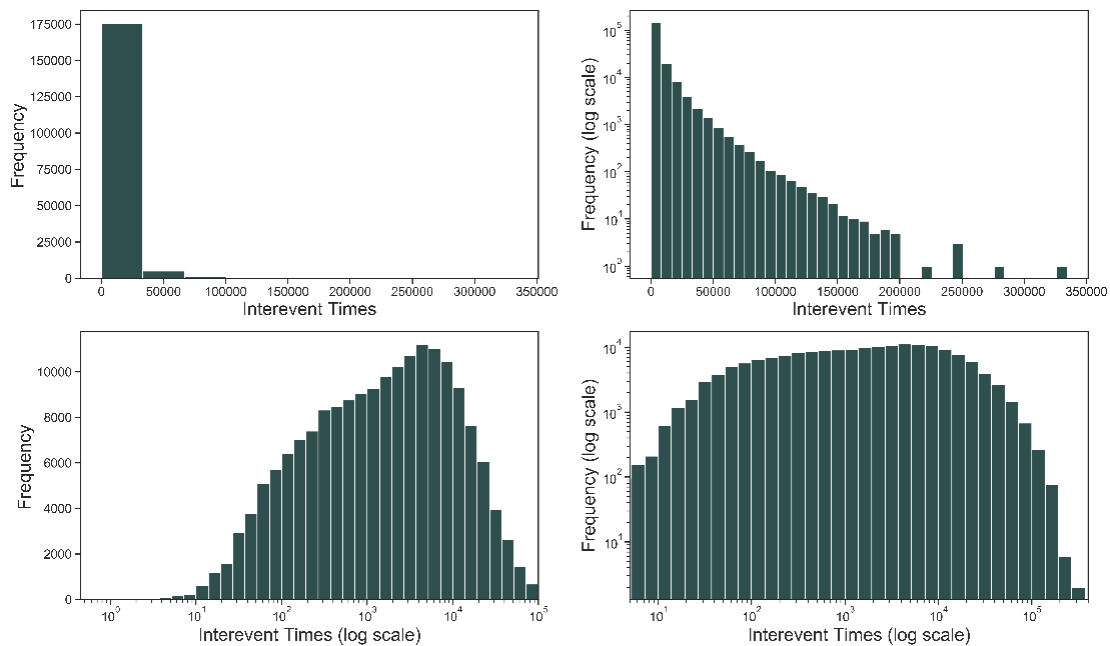


Figura 27 - vari modi di raffigurare la relazione tra tempi tra due terremoti successivi (in secondi) e rispettiva frequenza

Nella Figura 27 si sono rappresentati in quattro modi diversi i grafici *interevent times*/frequenze. In alto a sinistra sono stati rappresentati i tempi in scala lineare. Si osserva che i tempi inferiori ai 100000 secondi sono la quasi totalità, mentre i restanti sono solamente una piccolissima quantità. Nel grafico in alto a destra, la frequenza è in scala logaritmica e questo ci permette di osservare la distribuzione anche per *interevent times* maggiori di 100000. In quello in basso a sinistra invece, l'ascissa è logaritmica ma le ordinate sono lineari: ciò vuol dire che il *binning* è logaritmico e dunque permette di visualizzare più in dettaglio rispetto ai due grafici presenti, intervalli temporali più piccoli (al di sotto dei 10000 s), i quali sono allo stesso tempo i più frequenti. L'ultimo diagramma è bilogaritmico e permette di calcolare, convertendo le frequenze in probabilità, la PDF (funzione di densità di probabilità). A ogni *bin* è associata una probabilità di accadimento di quel particolare *range* di intervalli temporali. La somma di tutti i *bin* nell'istogramma deve essere, ovviamente, pari ad 1. Da qui si potrebbero calcolare anche numericamente la probabilità che si ottenga un certo valore temporale tra due terremoti consecutivi, tenendo in considerazione che il *binning* è in scala logaritmica.

Un altro grafico che ho rappresentato è quello tra due *interevent times* successivi (Figura 28).

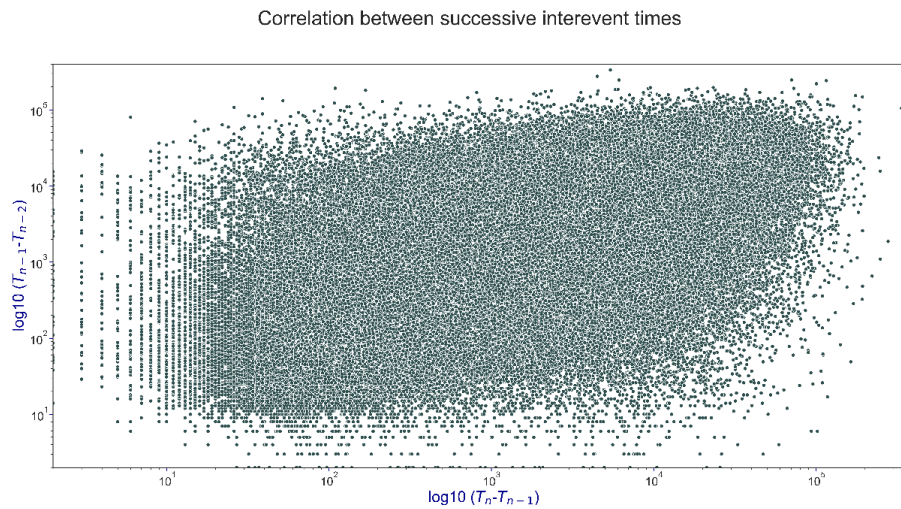


Figura 28 - rappresentazione tramite diagramma a sciame (scatter plot) di interevent times successivi. Sulle ascisse abbiamo la differenza temporale tra un determinato evento e quello precedente, sulle ordinate la differenza tra l'evento precedente e quello precedente ancora.

Da questo grafico non si riesce bene a percepire la correlazione tra le due variabili. Questa correlazione può essere meglio visualizzata con un *heat plot* e calcolando e raffigurando la retta di regressione lineare.

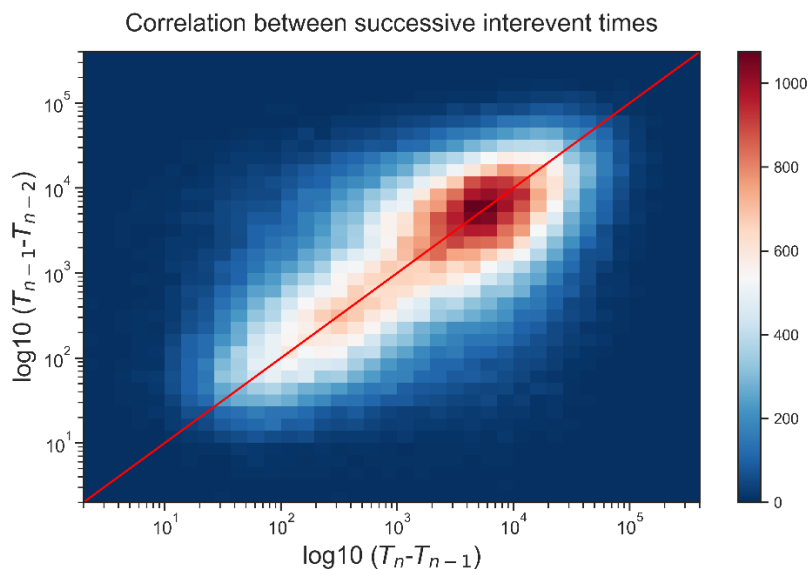


Figura 29 - rappresentazione tramite "heat plot" di interevent times successivi con retta di regressione lineare. A destra del grafico abbiamo la scala di colori a cui è associato il numero di eventi. La linea retta in rosso è la retta di regressione lineare

L'equazione della retta e l' R^2 associato sono rispettivamente:

$$y = -0.780111 + 1.0001239x$$

$$R^2: 0.10104090355274477$$

Questo basso valore del coefficiente di determinazione, prossimo allo zero, indica che c'è una correlazione non significativa. Ciò è probabilmente dovuto al fatto che il catalogo copre un'area molto ampia, con aree tettoniche molto diverse e non connesse tra di loro. Di conseguenza nella maggior parte dei casi non c'è una forte connessione tra due *interevent times* successivi, ma si osserva comunque un andamento preferenziale nella disposizione dei dati. Per calcolare i coefficienti retta di regressione β_0 e β_1 , ho applicato le formule 4.22 e 4.23 e ho inserito i valori ottenuti nella 4.21. L' R^2 invece equivale al coefficiente di correlazione al quadrato. Ho calcolato quest'ultimo nel modo seguente:

$$R = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2 \sum_{i=1}^N (y_i - \bar{y})^2}} \quad (5.31)$$

“Mescolando” gli *interevent times* e successivamente riproducendo graficamente gli stessi diagrammi, si può osservare come non si ha alcuna disposizione preferenziale dei dati (Figura 30 e Figura 31).

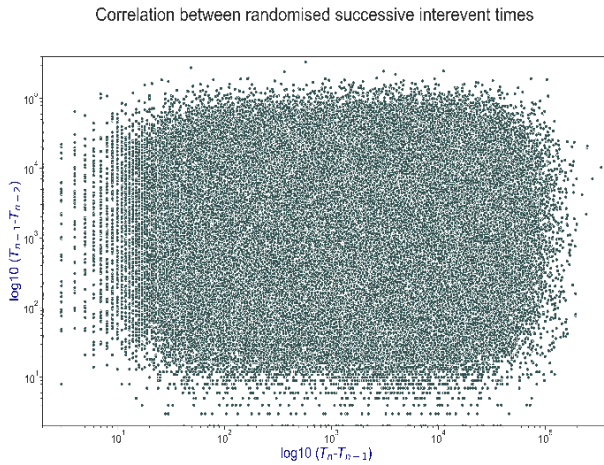


Figura 30 - rappresentazione di interevent times “mescolati” con un diagramma a sciame (scatter plot)

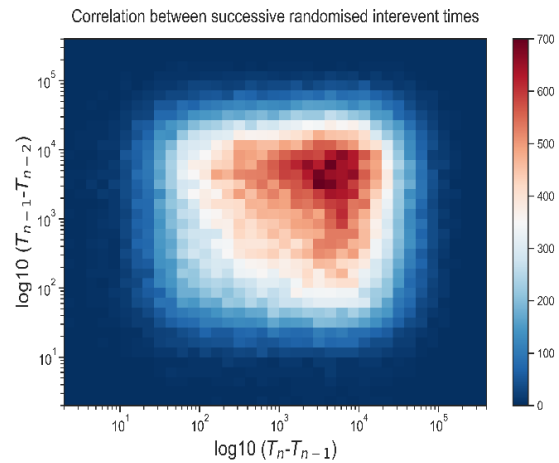


Figura 31 - rappresentazione di interevent times “mescolati” tramite “heat plot”

Anche in questo caso si nota una tendenza dei dati a disporsi nella parte in alto a sinistra per valori di ascissa compresi tra 300 e 20000 secondi (5 ore e mezzo circa). L' R^2 nella figura 31 è praticamente uguale a 0 con una totale assenza di correlazione:

$$R^2: 9.260672918339278e-07$$

6 - IL *DECLUSTERING*

In un catalogo sismico gli *aftershocks* sono dipendenti nello spazio e nel tempo dalle scosse principali (*mainshocks*). Perciò si possono distinguere all'interno di un catalogo i terremoti indipendenti tra di loro, ovvero i *mainshocks*, e i restanti ossia gli *aftershocks* ma anche i *foreshocks* (gli eventi che precedono le scosse principali). Gli eventi indipendenti (denominati anche *background earthquakes*) sono causati dalla tettonica oppure da stress transitori che non dipendono da altri terremoti. Al contrario quelli dipendenti sono dovuti a processi meccanici che sono almeno in parte indotti da eventi avvenuti precedentemente.

Lo scopo del *declustering* è quello di rimuovere dal catalogo i terremoti dipendenti (che solitamente formano dei “cluster” che si generano in corrispondenza degli eventi principali), in modo da ottenere un nuovo catalogo in cui sono presenti solo i terremoti indipendenti tra di loro.

6.1 - METODI DI *DECLUSTERING*

Esistono due tipologie di algoritmi di *declustering*: quelli deterministici e quelli stocastici. I primi classificano gli eventi o come *mainshocks* o come *aftershocks*, mentre i secondi sono più complessi e si basano sulla ricostruzione dei *mainshocks* che più probabilmente hanno generato gli *aftershocks*. Per effettuare ciò, ci si basa su una ricostruzione effettuata applicando degli algoritmi che misurano la distanza tra due eventi nello spazio e nel tempo e successivamente applicando dei criteri che devono essere soddisfatti per poter considerare un evento come *aftershock* di un altro evento precedente o come *mainshock*.

Tra le tecniche di *declustering* deterministiche si ha il metodo delle finestre. Esso consiste nell'identificare gli *aftershocks* a partire da un terremoto con una data magnitudo M , prendendo in considerazione tutti gli eventi che occorrono in uno specifico intervallo di tempo t da esso e, contemporaneamente, ad una distanza massima d dal terremoto. Il metodo non distingue gli *aftershocks* dai *foreshocks*: ciò vuol dire che considera, nel tempo e nello spazio, anche le scosse precedenti al terremoto

principale. L'algoritmo di *declustering* è stato proposto da Gardner e Knopoff nel 1972 ed è mostrato nelle due equazioni (6.11, 6.12).

$$t \text{ [days]} = \begin{cases} 10^{0.032*M+2.7389}, & \text{if } M \geq 6.5 \\ 10^{0.5409*M-0.547}, & \text{else} \end{cases} \quad (6.11)$$

$$d \text{ [km]} = 10^{0.1238*M+0.983} \quad (6.12)$$

Si può osservare in particolare che, mentre per la distanza viene utilizzata una sola relazione, per il tempo vengono considerati due casi diversi: il primo in cui si considerano magnitudo maggiori o uguali a 6.5 e il secondo per i casi restanti (minori di 6.5).

In Figura 32 ho rappresentato, in funzione della magnitudo, il valore massimo nel tempo (curva blu) e nello spazio (curva arancione) per cui un evento è considerato dipendente dall'evento principale. Per raffigurare le curve ho applicato l'algoritmo di Gardner e Knopoff del 1972 a intervalli di 0.1 di magnitudo, senza effettuare poi il *declustering* effettivo.

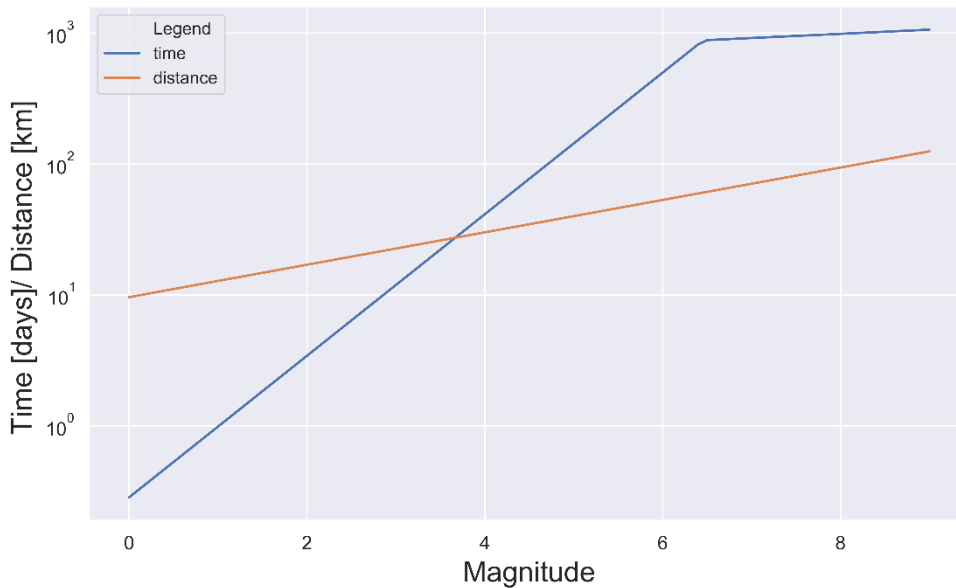


Figura 32 - rappresentazione dei valori massimi nello spazio e nel tempo per i quali si considera un mainshock dipendente dall'aftershock.

Altre equazioni alternative sono state proposte in seguito nel 1974 (da Gardner e Knopoff) e più avanti da Urhammer (1986).

Un altro metodo deterministico per effettuare il *declustering* di un catalogo è quello di Reasenberg (1985). Anche questo metodo si basa sull'identificazione degli *aftershocks* tramite zone di interazione nello spazio e nel tempo. In quest'ultimo caso, la zona di interazione ha come valore limite per poter considerare un evento come dipendente da un *mainshock*, quello ottenuto dalla retta $\log d(\text{km}) = 0.4M_0 - 1.943 + k$, in cui M_0 è la magnitudo del *mainshock*, k è un parametro uguale a 0 per l'ultimo evento nel cluster e pari ad 1 per l'evento più grande. Per quanto riguarda la “distanza” temporale, questa zona di interazione è basata sulla legge di Omori.

Perciò tutti gli eventi all'interno di questa zona di interazione che seguono o procedono il *mainshock* sono considerati rispettivamente *aftershocks* e *foreshocks* e fanno parte dello stesso cluster.

I parametri principali che vengono considerati per sono

- τ_{\min} , cioè il minimo valore temporale per cui si può elaborare un cluster;
- τ_{\max} , ossia il massimo valore nel tempo per poter costruire un cluster;
- p_1 : la probabilità di rilevare il prossimo evento nel cluster;
- X_{meff} , ossia la magnitudo più bassa di cutoff;
- X_k è l'incremento della magnitudo più bassa di cutoff all'interno del cluster X_{meff}
 $= X_{\text{meff}} + X_k M$, dove M è la magnitudo dell'evento principale nel cluster;
- R_{fact} è il numero di rotture (crack) circolari (Kanamori & Anderson, 1975) che circonda ciascun terremoto, all'interno del quale si può considerare un evento nel cluster (van Stiphout et al., 2012)

I parametri che ho inserito per effettuare il *declustering* sono indicati nella figura e sono quelli esistenti di default nel software (Figura 33)

Reasenbergl Declustering parameters :	
look-ahead times :	
(min) for UNclustered events [days] :	1
(max) for clustered events [days] :	10
Confidence Level :	0.95
XK factor :	0.5
Effective min mag cutoff :	1.5
Interation radius factor: :	10
Epicenter error :	1.5
Depth error :	2

Figura 33 - parametri utilizzati per il declustering con il metodo di Reasenbergl prendendo in considerazione (per ragioni di tempi e risorse di calcolo) gli eventi maggiori o uguali a magnitudo 3.

6.2 - DECLUSTERING CON IL METODO DI ZALIAPIN (2008)

I metodi stocastici, come già menzionato, permettono di ricostruire i *mainshocks* che più probabilmente hanno generato gli *aftershocks*.

Tra questi, il metodo che ho utilizzato per effettuare il *declustering* è quello di Baiesi e Paczuski (2004) successivamente rielaborato da Zaliapin et al. (2008). Come per il *declustering* con il metodo di Reasenbergl (1985), ho considerato solo i terremoti con magnitudo maggiore o uguale a 3 e ho applicato la seguente formula a tutti gli eventi nel catalogo:

$$\eta_{ij} = (t_j - t_i) r_{ij}^{df} 10^{-b m_i} \quad (6.21)$$

- η_{ij} è una distanza spazio-temporale che indica la distanza tra due eventi i e j (dove j è un evento successivo a i);
- r_{ij} è la distanza spaziale tra i due eventi (nel caso dell'algoritmo che ho implementato, la distanza tra i due ipocentri). Ho calcolato questo valore tramite delle formule trigonometriche che, avendo a disposizione i valori di latitudine,

longitudine e profondità dei due eventi e conoscendo il raggio della Terra, permettono di risalire alla distanza tra questi due punti all'interno della Terra;

- d_f è la dimensione frattale, ovvero una misura statistica della complessità del sistema;
- b è il b-value;
- m_i la magnitudo dell'evento precedente;
- $t_j - t_i$ la differenza temporale tra l'evento j e l'evento i .

I due terremoti che si considerano (facenti parti dello stesso catalogo e di conseguenza avendo gli stessi valori di d_f e b) avranno perciò un valore di η_{ij} maggiore se sono avvenuti a poco tempo di distanza l'uno dall'altro, se la magnitudo del primo terremoto è molto elevata e se la distanza spaziale tra i due ipocentri è ridotta.

Ho effettuato questa procedura per ogni terremoto j , calcolando per ciascuno di esso tutte le combinazioni possibili η_{ij} . Invece di considerare tutti i risultati ottenuti di questa matrice di valori, per ogni evento j ho prelevato la minima distanza η_{j*} tra tutte le distanze calcolate. L'evento a cui è associata questa minima distanza è indicato con i^* , equivalente all'evento che ha generato l'evento j .

Qui sotto, in dettaglio, è raffigurato l'algoritmo che ho sviluppato e implementato in Python.

```

d=1.6
b=1.21
a_time=pd.Series(a['Time'])
a['Time']=a_time.astype(np.datetime64)
space_time_total=[]
start_time=time.time()
R = 6373.0 # approximate earth radius in km
for j in range(int(a.size/len(a.columns))-3):
    timej=a_time.iloc[j]
    latj = radians(a.iloc[j]['Latitude'])
    lonj = radians(a.iloc[j]['Longitude'])
    i=1
    values=[]
    while i+j <= int(a.size/len(a.columns))-2 :
        #timei=a_time.iloc[i+j]
        dttime=((timej-a_time.iloc[i+j])).total_seconds()
        lati = radians(a.iloc[i+j]['Latitude'])
        loni = radians(a.iloc[i+j]['Longitude'])
        a2 = sin((lati-latj) / 2)**2 + cos(lati) * cos(lati) * sin((loni-lonj) / 2)**2
        distance = R*(sqrt((2 * atan2(sqrt(a2), sqrt(1 - a2))))**2+
            (a.iloc[i+j]['Depth/Km']-a.iloc[j]['Depth/Km'])**2))
        space_time=round(dttime*distance*d*pow(10, -b*float(a.iloc[i+j]['Magnitude'])),12)
        values.append(space_time)
        i+=1
    min_value=min(values)
    i_value=values.index(min_value)+1+j
    space_time_total.append([min_value,i_value,j])
end_time=time.time()
total_time=end_time-start_time
print(space_time_total,total_time)

```

Figura 34 - algoritmo di Baiesi e Paczuski (2004) per il calcolo di η_{ij} , implementato in Python. In rosso sono rappresentati i parametri in input (latitudine, longitudine, profondità e tempo in cui gli eventi si sono verificati; dimensione frattale, raggio della terra e b-value), in giallo l'output che ci fornisce il valore η_{ij} , assieme alla posizione nel catalogo degli eventi i^* e j (i_value , j) e al tempo impiegato per l'esecuzione dell'algoritmo ($total_time$).

L'algoritmo ha impiegato 99269 secondi (più di 27 ore e mezzo) a restituire i valori η_{ij} per ogni evento e ha effettuato più di 6200000 di iterazioni. Questi valori in output sono stati restituiti sottoforma di matrice e successivamente trasformati in una tabella in cui nella prima colonna si ha la distanza spazio-temporale, nella seconda e nella terza rispettivamente l'indice nel catalogo dell'evento i^* e j .

In Figura 35 è rappresentata la distribuzione di η_{j^*} .

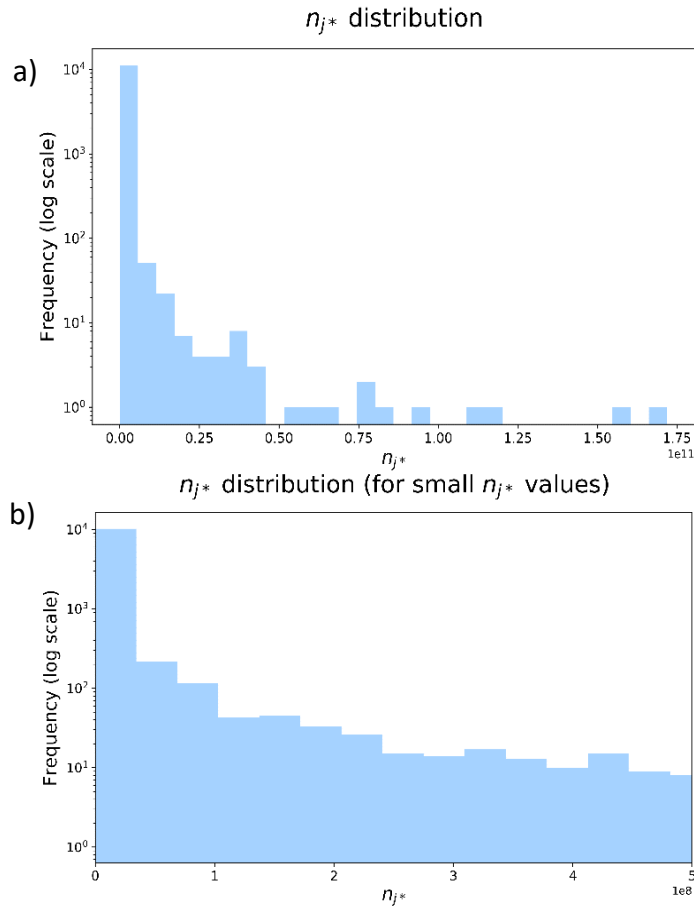


Figura 35 – Istogrammi di η_{j^*} per valori prossimi a 10^{11} (figura a) e per valori intorno a 10^8 (figura b)).

È evidente come la maggior parte dei valori abbia valori molto piccoli e la distribuzione dei dati sia di tipo esponenziale.

Il metodo di Baiesi e Paczuski (2004) prevede che si considerino solo i terremoti al di sotto di un certo η_{j^*} limite e si scartino i valori al di sopra di questa soglia: ciò si effettua per eliminare le deboli connessioni tra *mainshocks* e *aftershocks* e permette di dividere il catalogo in clusters. Per ogni cluster viene poi considerato l'evento principale. Il metodo di Zaliapin et al. (2008), successiva implementazione del metodo di Baiesi e Paczuski (2004), è quello che ho applicato al catalogo. A differenza del metodo precedentemente descritto, in questo caso non si considera un valore limite al di sopra del quale si scartano gli eventi, ma si utilizzano gli eventi i^* e j per calcolare altri due valori: T_j e R_j (equazioni 6.22, 6.23).

$$T_j = (t_j - t_{i^*}) 10^{-\frac{b m_{i^*}}{2}} \quad (6.22)$$

$$R_j = r_{i^*j}^{d_f} 10^{-\frac{b m_{i^*}}{2}} \quad (6.23)$$

Dopo aver calcolato tutti i valori di T_j e R_j , si raffigurano questi valori in un grafico in cui sull'ascissa viene riportato il valore di T_j e sull'ordinata di R_j (Figura 36).

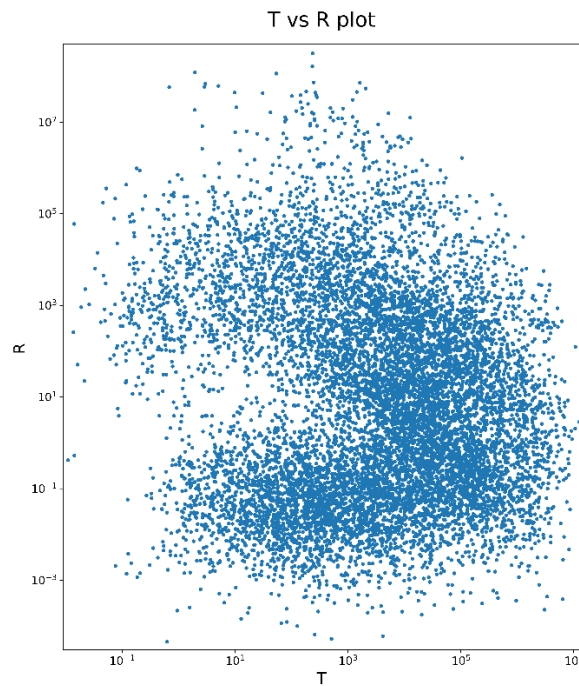


Figura 36 - grafico T_j - R_j

Nel grafico si osserva una distribuzione bimodale dei dati: un primo cluster più in basso con valori bassi di T_j e R_j e un altro con valori più elevati nella parte più in alto. Il primo corrisponde agli *aftershocks*, dunque a tutti gli eventi dipendenti; mentre il secondo corrisponde agli eventi che sono indipendenti, ossia i *mainshocks*. Di seguito (Figura 37), ho rappresentato con un *heatmap* il grafico della Figura 36.

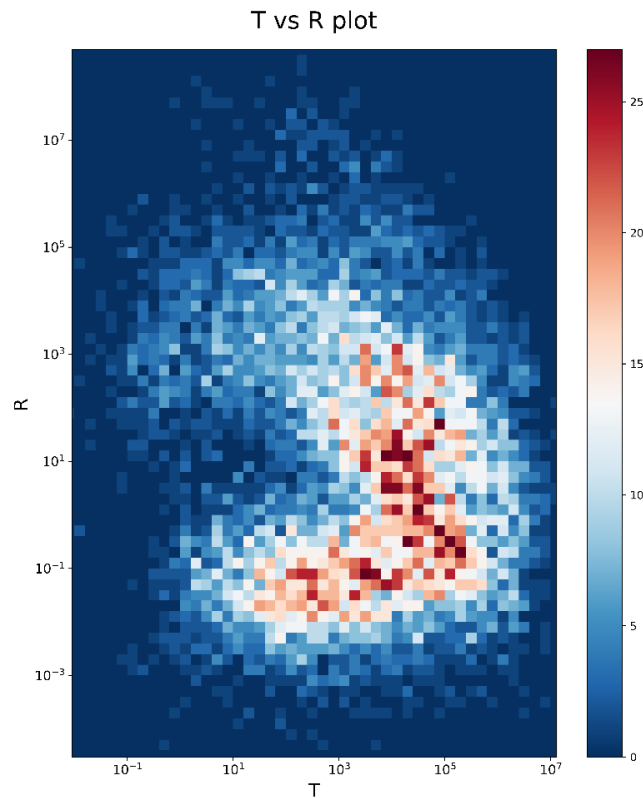


Figura 37 - heatmap delle due variabili T_j e R_j

Da questo grafico si osserva in modo ancora più evidente la distribuzione bimodale e inoltre si nota dove si concentrano maggiormente gli eventi.

Per effettuare il *declustering* ho dunque preso gli eventi indipendenti della distribuzione più in alto.

Nel prossimo paragrafo verrà spiegato più nel dettaglio come si sono selezionati i terremoti del catalogo per ottenere un catalogo privo di eventi dipendenti e verificare se esso segue un processo di Poisson omogeneo nel tempo.

6.3 - TEST DI “GOODNESS OF FIT”

In letteratura sono stati effettuati dei test per verificare se i cataloghi successivi alla rimozione degli eventi dipendenti avessero al proprio interno eventi la cui distribuzione

nel tempo e nello spazio potesse essere considerata approssimativamente casuale (*Spatially Inhomogeneous and Temporally Homogeneous Poisson Process*, SITHP).

Sia K. Aki nel 1956, considerando cataloghi mondiali, sia L. Knopoff nel 1964 per cataloghi locali dimostrarono che la distribuzione degli eventi nel tempo fosse non casuale e dunque non seguisse un modello Poisson.

J.K. Gardner e L. Knopoff nel 1974 hanno ottenuto per la sequenza della California del Sud, dopo aver applicato il test di *goodness of fit*, valori per magnitudo maggiori di 3.8, 4.3, 4.8 e 5.3 che indicano un comportamento Poisson nel tempo nella distribuzione dei terremoti del catalogo . Successivamente B. Luen e P. B. Stark nel 2012 analizzarono più nel dettaglio il problema della distribuzione degli eventi utilizzando vari test statistici, sia effettuando test di verifica spazio-temporali, sia test solo temporali. Vennero utilizzati in questo caso per il *declustering* il metodo di Reasenberg (1985), attraverso il pacchetto ZMAP di MATLAB (Wiemer, S. (2001)) con i parametri di default del programma e tre varianti del metodo di Gardner & Knopoff (1974).

Per i test temporali utilizzarono il test chi quadrato condizionale (CC) e multinomiale (MC), il test di Brown-Zhao (BZ) e quello di Kolmogorov-Smirnov (KS).

In questa tesi, ho applicato il test chi quadrato multinomiale e il test di Brown-Zhao.

J.K. Gardner e L. Knopoff nel 1974 scelsero nel loro lavoro di utilizzare il test chi quadrato multinomiale che, come descritto da Luen e Stark, non tiene conto della sovradisersione (*overdispersion*), ossia dell'osservazione di una maggiore variabilità dei dati rispetto al modello che ci si aspetterebbe.

Il test CC invece prende in considerazione la variabilità del numero di eventi osservato per ogni intervallo e non la distribuzione dettagliata del numero di eventi per intervallo (B. Luen and P.B. Stark 2012).

La formula per effettuare il test è la seguente:

$$\chi_c^2 = \sum_{k=1}^K \frac{(N_k - \hat{\lambda})^2}{\hat{\lambda}} \quad (6.31)$$

In cui:

- K è il numero di intervalli. Deve essere maggiore o uguale a 1.
- k è uno degli intervalli presi in considerazione. Esso ha lunghezza $\frac{T}{K}$, con T corrispondente all'intervallo temporale del catalogo e K il numero di intervalli. Ogni intervallo non deve avere elementi in comune con gli altri intervalli nel catalogo.
- N_k è il numero di eventi in un determinato intervallo k .
- $\hat{\lambda}$ è la media del numero di eventi per intervallo.

Il test di Brown-Zhao (BZ) è stato proposto nel 2002 e tiene conto della sovradisersione, in modo simile al test CC:

$$\chi_{BZ}^2 = 4 \sum_{k=1}^K (Y_k - \bar{Y})^2 \quad (6.32)$$

dove:

- $Y_k = \sqrt{N_k + \frac{3}{8}}$
- $\bar{Y} = \sum \frac{Y_k}{K}$

Dopo aver calcolato χ_{BZ}^2 e χ_c^2 , si procede alla verifica. Ho calcolato il χ_{95}^2 , ovvero il valore che ci permette definire con un intervallo di confidenza del 95 per cento che la distribuzione sia Poisson. Al di sopra di questo si considera il catalogo non Poisson nel tempo e al di sotto si può accettare l'ipotesi che esso sia Poisson nel tempo. In statistica, la prima ipotesi è chiamata ipotesi alternativa (H_a), la seconda ipotesi nulla (H_0). In sintesi:

H₀: gli eventi nel catalogo dopo aver effettuato il *declustering* rispecchiano una distribuzione Poisson (χ_{BZ}^2 e $\chi_c^2 < \chi_{95}^2$);

H_a: gli eventi nel catalogo dopo aver effettuato il *declustering* rispecchiano una distribuzione Poisson (χ_{BZ}^2 e $\chi_c^2 > \chi_{95}^2$);

Il χ_{95}^2 si calcola normalmente con le tavole di distribuzione chi-quadrato, considerando il valore χ_{95}^2 avente K-1 gradi di libertà (dove K è il numero di intervalli). Nella Tabella 5 e 6 si può osservare che esso equivale a 9.488 per K=5 e 5.991 per K=3.

Sono stati poi calcolati χ_{BZ}^2 e χ_c^2 , sia per il catalogo originario, sia dopo aver applicato il *declustering* con il metodo di Reasenberg e di Zaliapin. Le celle al cui interno sono presenti valori minori di χ_{95}^2 sono state evidenziate in rosso, al contrario in verde quelle con all'interno valori maggiori di χ_{95}^2 .

ZALIAPIN (2008)									
n=4504		straight line 1 (TR>x)	straight line 2 (T>x)	K=5 (d.f=4)			K=3 (d.f=2)		
M	d			p-value (CC)	CC	BZ	p-value (CC)	CC	BZ
>2.9	1.85	19200	3400	0.309	4.791	4.699	0.309	16.718	16.521
>3.1	1.8	9980	690	0.137	6.987	7.005	0.07	5.308	5.241
>3.3	1.83	6340	420	0.023	11.378	11.42	0.00012	18.07	17.512
>3.5	1.79	3200	360	0.0008	18.907	18.84	5.35E-5	19.673	18.842
>3.7	1.74	1320	460	0.002	16.718	16.521	0.002	12.269	11.882
>3.9	1.74	1280	460	0.003	15.859	15.321	0.0001	18.362	18.032
>4.1	1.74	1280	300	0.0013	17.962	17.525	0.00022	16.856	16.708
>4.3	1.84	3200	555	0.109	7.573	7.205	0.018	8.0	7.573
>4.5	1.82	7400	120	0.592	2.8	2.66	0.179	3.44	3.35
>4.7	1.8	600	1100	0.938	0.8	0.748	0.375	1.96	1.89
>4.9	1.8	600	900	0.955	0.666	0.632	0.407	1.8	1.776
>5.1	1.8	400	400	0.836	1.44	1.41	0.846	0.333	0.316
>5.3	1.8	550	200	0.985	0.364	0.28	0.368	2.0	1.945
>5.5	1.8	3550	100	0.955	0.666	0.435	1	0.0	0.0
>5.7	1.8	500	100	0.558	2.999	1.507	0.135	4.0	2.3
					$\chi_{95}^2 = 9.488$			$\chi_{95}^2 = 5.991$	

Tabella 5 - valori di chi quadrato ottenuti per i test CC e BZ per il catalogo ottenuto a seguito del declustering di Zaliapin (2008). I valori in verde sono quelli inferiori a χ_{95}^2 , in rosso i valori maggiori a χ_{95}^2 ; entrambi calcolati sia per K=5 e poi per K=3 (dove K è il numero di intervalli). I valori del p-value del test CC, sono stati calcolati da Python automaticamente e inseriti nella quinta e nella ottava colonna. M è la magnitudo minima considerata, d la dimensione frattale, n il numero di terremoti nel catalogo a seguito del declustering, d.f. sono i gradi di libertà. Straight line 1 e 2 sono le due rette che rappresentano i limiti inferiori del nuovo catalogo. La prima è un'iperbole

equilatera (una retta in scala logaritmica) con valori della costante x maggiori o uguali a quelli elencati in terza colonna; la seconda è una retta parallela all'asse R , con valori uguali a quelli della quarta colonna.

REASENBERG (1985)						
n=7722	K=5 (d.f=4)			K=3 (d.f=2)		
K=5	p-value (CC)	CC	BZ	p-value (CC)	CC	BZ
>2.9	7.66E-44	207.87	217.78	1.25E-22	100.875	104.718
>3.1	2.04E-14	70.219	71.114	1.26E-6	27.173	27.197
>3.3	0.021	11.469	11.469	0.33	2.214	2.214
>3.5	0.126	7.186	7.544	0.104	4.529	4.477
>3.7	0.14	6.919	6.955	0.103	4.552	4.434
>3.9	0.029	10.835	10.369	0.005	10.442	10.06
>4.1	0.001	18.381	17.539	0.0002	17.64	17.39
>4.3	0.076	8.452	8.086	0.007	10.0	9.358
>4.5	0.391	4.11	4.108	0.051	5.956	5.993
>4.7	0.543	3.088	3.281	0.128	4.105	3.846
>4.9	0.783	0.783	0.783	0.25	2.774	3.05
>5.1	0.774	0.774	1.784	0.504	1.368	1.42
>5.3	0.736	0.736	1.977	0.368	2	2.153
>5.5	0.736	0.736	1.456	0.607	1.0	0.808
>5.7	0.674	2.333	2.317	0.607	1.0	0.887
		$\chi^2_{95}= 9,488$			$\chi^2_{95}= 5.991$	
NO DECLUSTERED CATALOGUE						
n=11194	K=5 (d.f=4)			K=3 (d.f=2)		
K=5	p-value (CC)	CC	BZ	p-value (CC)	CC	BZ
>2.9	5.2E-37	176.07	179.116	2.87E-25	113.02	117.519
>3.1	8.66E-24	114.333	112.064	2,03E-22	99.9	102.919
>3.3	4,34E-18	87.562	84.902	1.47E-26	118.96	119.49
>3.5	5,63E-14	68.132	68.102	1.32E-23	105.36	104.708
>3.7	4.16E-10	49.708	49.185	2.11E-17	76.799	75.115
>3.9	4,36E-06	44.809	43.579	9.07E-16	69.27	67.761
>4.1	1,45E-10	51.891	50.569	2.38E-15	67.34	66.043
>4.3	5.83E-6	29.628	28.201	5.32E-10	42.707	39.681
>4.5	0.00053	19.852	18.263	5,00E-07	28.98	28.326
>4.7	0.00203	16.889	15.244	4.12E-6	24.8	22.574
>4.9	0.068	8.727	8.146	0.0013	13.345	13.705
>5.1	0.611	2.688	2.563	0.261	2.688	2.429
>5.3	0.292	4.952	4.046	0.0499	6.0	7.219
>5.5	0.297	4.909	3.502	0.307	2.364	2.038
>5.7	0.446	3.714	3.386	0.368	2.0	1.711
		$\chi^2_{95}= 9.448$			$\chi^2_{95}= 5.991$	

Tabella 6 - valori di chi quadrato ottenuti per i test CC e BZ per il catalogo iniziale e quello ottenuto a seguito del declustering di Reasenberg (1985). I valori in verde sono quelli inferiori a χ^2_{95} e in rosso i valori maggiori a χ^2_{95} , entrambi calcolati sia per K=5 e poi per K=3 (dove K è il numero di intervalli). I valori del p-value del test CC, sono stati calcolati da Python automaticamente e inseriti nella seconda e nella quinta colonna. M è la magnitudo minima considerata, n il numero di terremoti nel catalogo prima e dopo il declustering di Reasenberg, d.f. sono i gradi di libertà (K-1).

Per il *declustering* di Zaliapin (2008), sono stati calcolati i valori dei test CC e BZ variando manualmente i parametri (raffigurati in Tabella 5) per ogni magnitudo limite. Per il catalogo ottenuto dal *declustering* di Reasenberg (1985), sono stati calcolati i valori senza variare alcun parametro (poiché il catalogo tramite ZMAP è restituito con valori fissi e non è possibile modificare i parametri tramite codice). I valori in verde sono quelli inferiori a χ_{95}^2 , in rosso i valori maggiori a χ_{95}^2 , entrambi calcolati sia per $K=5$, sia per $K=3$ (dove K è il numero di intervalli).

Si osserva che per il catalogo originario si hanno valori che non rispecchiano per basse magnitudo un processo Poisson nel tempo, al contrario per alte magnitudo i bassi valori ottenuti dai test CC e BZ ci farebbero pensare ad un comportamento Poisson. Questi valori per magnitudo molto alte non sono affidabili, dato che per ogni intervallo sono pochi gli eventi presenti. Invece per i cataloghi con il metodo di Reasenberg (1985) e di Zaliapin (2008), si trovano valori inferiori a χ_{95}^2 anche per magnitudo più basse considerando comunque un catalogo con un numero elevato di eventi (dell'ordine di migliaia di eventi). Per il *declustering* di Zaliapin (2008), si è visto che i valori variano molto a seconda dei parametri che si considerano, in particolare le costanti che determinano le rette limite che rappresentano il limite inferiore del catalogo a seguito del *declustering*.

Un fattore importante che influenza il test è sicuramente la completezza. Essa può influenzare i risultati, infatti potrebbero essere presenti *mainshocks* non registrati oppure potrebbero esserci eventi classificati come *mainshocks* che invece andrebbero inseriti tra gli *aftershocks* (o viceversa) a causa della mancata registrazione degli altri eventi. Ciò può far variare i parametri per l'elaborazione del catalogo privo di eventi dipendenti (nel *declustering* di Zaliapin questi parametri possono essere η_{j^*} , T_j e R_j).

6.4 - CONFRONTO TRA GLI *INTEREVENT TIMES* SUCCESSIVI E GLI EVENTI NEL TEMPO PRIMA E DOPO IL *DECLUSTERING*

L'ultima analisi che ho effettuato è stata quella confrontare gli *interevent times* successivi per il catalogo a seguito del *declustering* con i due metodi. Ho deciso di rappresentare lo stesso grafico della Figura 29 e della Figura 31 con il fine di mostrare se gli *interevent times* all'interno del catalogo ottenuto non fossero correlati tra loro, ma avvenissero casualmente. Sono stati poi rimescolati gli eventi e osservato che, anche in questo caso, non si ha una correlazione (Figura 38, Figura 39).

ZALIAPIN (2008)

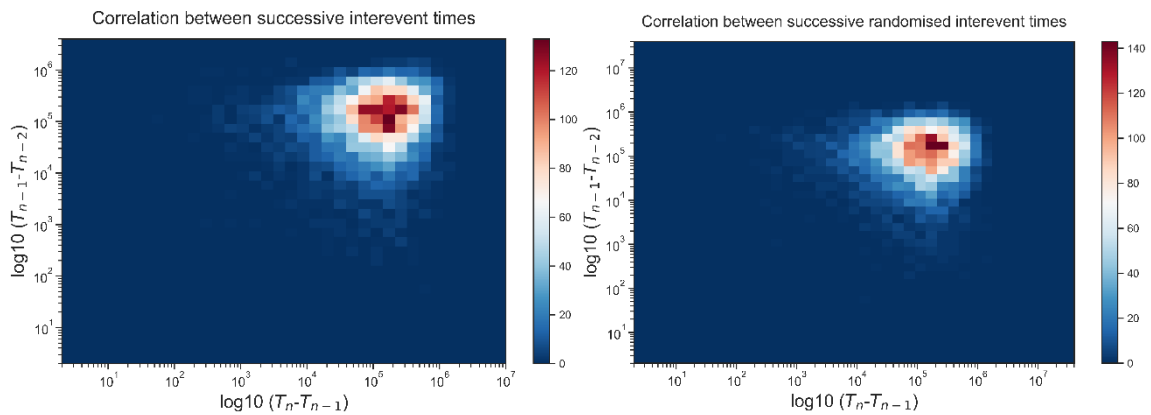


Figura 38 –Rappresentazione tramite “heat plot” di interevent times successivi con il metodo della correlation metrics di Zaliapin (2008) (a sinistra). Grafico mostrante gli stessi interevent times “rimescolati” (a destra).

REASENBERG (1985)

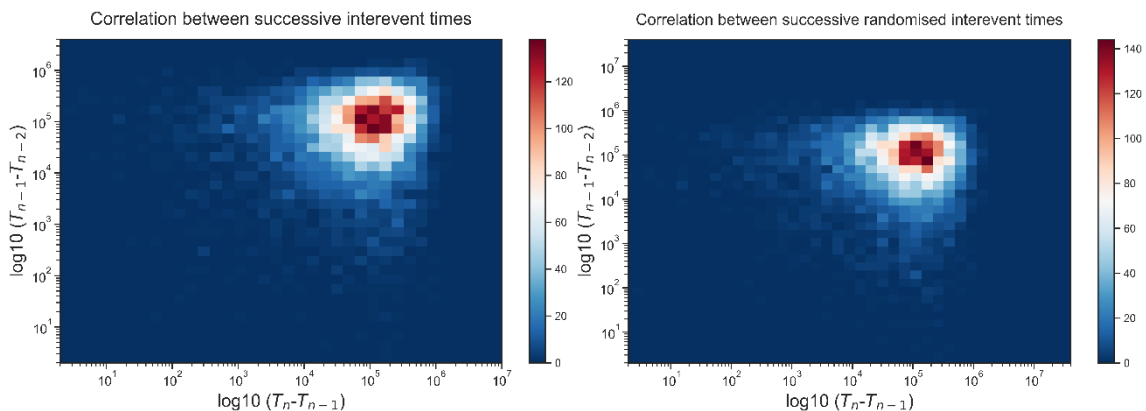


Figura 39 - Rappresentazione tramite “heat plot” di interevent times successivi con il metodo di Reasenber (1985) (a sinistra). Grafico mostrante gli stessi interevent times “rimescolati” (a destra).

Gli ultimi due grafici (Figura 40, Figura 41) rappresentano il numero di eventi nel tempo, sia in forma ordinaria che cumulata, similmente a quanto compiuto per la Figura 25 e per la Figura 26 in cui ho raffigurato tutti gli eventi del catalogo iniziale. In questo caso ho considerato gli eventi maggiori o uguali a magnitudo 3 e ho raffigurato il catalogo iniziale e quello ottenuto a seguito del *declustering* con il metodo di Reasenberg (1985) e di Zaliapin (2008).

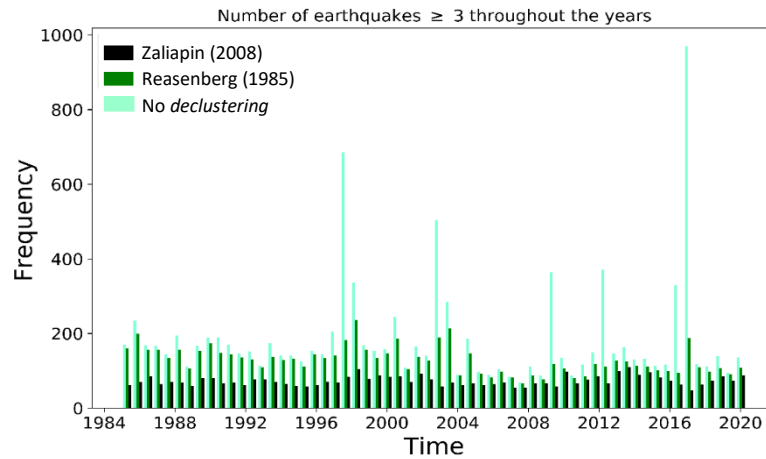


Figura 40 – numero di eventi maggiori o uguali a 3 nel tempo, considerando il catalogo originario e quelli ottenuti a seguito del declustering di Zaliapin e di Reasenberg

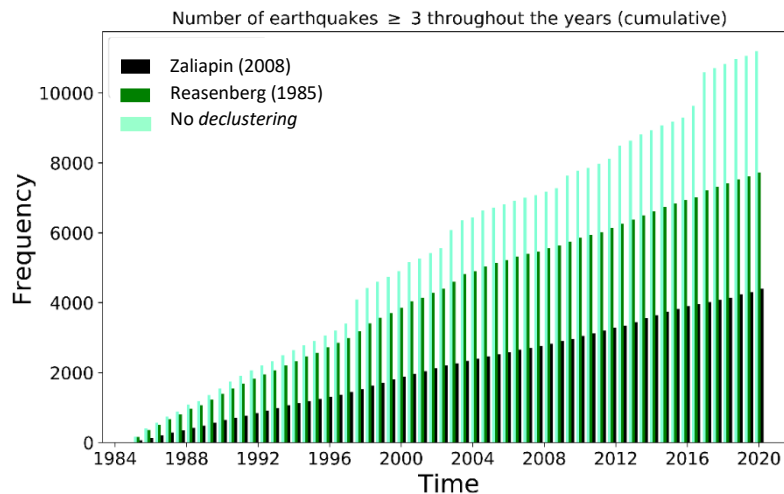


Figura 41 - numero di eventi maggiori o uguali a magitudo 3 nel tempo. È stata calcolata la frequenza cumulata per il catalogo originario e per i cataloghi ottenuti successivamente al declustering di Zaliapin e di Reasenberg.

Si nota come nei cataloghi successivi al *declustering* si hanno picchi meno accentuati e la curva cumulata è molto più lineare, soprattutto per quanto riguarda il catalogo ottenuto con il metodo di Zaliapin (2008) che, come si vede in tabella 5, per terremoti di magnitudo maggiore o uguale a 3 è un processo di Poisson nel tempo.

7 - CONCLUSIONI

La Tesi ha avuto come finalità quella di analizzare il catalogo sismico ISIDe, prendendo in considerazione tutti i parametri presenti nel catalogo. Le analisi sul tipo di magnitudo presente hanno mostrato una chiara dipendenza della tipologia di magnitudo con il tempo, dovuto al fatto che, nel corso degli anni, le tecniche si sono evolute e si sono impiegate le varie scale in contesti differenti. Si è osservata inoltre una correlazione tra l'energia del sisma e la magnitudo utilizzata e in alcuni casi specifici anche una dipendenza dalla profondità e dalla distanza dai sismografi.

Le analisi sulle profondità hanno posto in evidenza le aree in cui si concentrano la maggior parte dei terremoti. Si è osservata una sostanziale differenza in termini di profondità tra le regioni di subduzione (Mar Tirreno), in cui si hanno eventi anche molto profondi (fino a 600-650 km), e la catena Appenninica e Alpina in cui la quasi totalità degli eventi sono superficiali.

È stata ricavata l'equazione di Gutenberg-Richter, la quale ha permesso di visualizzare la distribuzione del numero di eventi al variare della magnitudo e ha consentito di determinare, considerando dei cataloghi sintetici, le distribuzioni del numero di eventi in frequenza per ogni magnitudo. Si è poi dimostrato che queste distribuzioni sono distribuzioni di Poisson.

Nel quinto capitolo dedicato alle analisi temporali, si è osservata una differenza sostanziale nella completezza del catalogo nel tempo: la magnitudo per cui il catalogo può definirsi completo è maggiore nella prima parte del catalogo, risultando di poco inferiore ad una magnitudo 3 e decresce col tempo fino ad arrivare negli ultimi anni a magnitudo inferiori a 1.5. Si sono poi individuate le maggiori sequenze sismiche italiane dal 1985 e poi studiati gli *interevent times*, ossia i tempi trascorsi tra due eventi successivi. È stato possibile osservare come nel grafico tra due *interevent times* successivi si abbia un'orientazione preferenziale dei dati, mentre rimescolando gli eventi non si ha alcuna correlazione.

È stato poi effettuato il *declustering* del catalogo sismico con il metodo di Reasenber (1985) e di Zaliapin (2008) e sono stati calcolati per il catalogo sismico e per i due

cataloghi ottenuti con il *declustering*, il test chi quadrato condizionale (CC) e il test di Brown-Zhao (BZ) dimostrando che solo per alcune magnitudo essi possono essere considerati processi di Poisson nel tempo. Rappresentando graficamente gli eventi nei cataloghi ottenuti tramite il *declustering* e considerando anche gli *interevent times*, è stato possibile inoltre visualizzare la casualità con cui avvengono gli eventi nel tempo e confrontare i grafici con quelli del catalogo originario.

I risultati ottenuti dal test di *goodness of fit*, mostrano risultati simili a quelli ottenuti da B. Luen e P. M. Stark (2012). Tra i metodi di *declustering* applicati da Luen e Stark, vengono utilizzati il metodo di Reasenberg (1985) e tre varianti del metodo di Gardner & Knopoff (1974). Sono stati presi in considerazione da Luen e Stark solo gli eventi maggiori di 3.8 e di 4.0 dal 1932 al 2010, dell'area della California del Sud, ma allo stesso tempo sono stati effettuati test anche nello spazio (quindi spazio-temporali) per verificare se anche nello spazio la distribuzione seguisse un processo di Poisson.

Le conclusioni a cui giungono B. Luen e P.M. Stark (2012) sono in linea con quelle a cui sono arrivato: i risultati del test dipendono da molti fattori, tra cui il metodo utilizzato per il *declustering*, i parametri utilizzati, l'area considerata, l'intervallo di tempo considerato, il numero di eventi e il test statistico utilizzato. I risultati precedenti di Gardner & Knopoff (1974), indicavano invece che per magnitudo maggiori a 3.8 e 2.8 per il catalogo della California del sud dal 1932 al 1971, la distribuzione fosse Poisson nel tempo. La spiegazione di ciò è stata esposta da B. Luen e P.M. Stark che ritengono che il test utilizzato non tenga conto di diversi fattori, tra cui la sovradisersione e la variazione a lungo termine nel tempo degli eventi.

Il problema del *declustering* che, si ricorda, viene effettuato per ottenere la migliore stima possibile dei *mainshocks* presenti, si riduce ad un problema di ottimizzazione statistica. Ovvero si cerca di variare i parametri del *declustering* con il fine di ottenere risultati migliori nel test, avendo comunque all'interno del catalogo migliaia di eventi.

In ogni caso, per poter ottenere risultati più affidabili sarebbe necessario svolgere ulteriori ricerche. Ad esempio, invece di prendere soltanto i terremoti maggiori di magnitudo 3, avendo più dati in futuro potranno essere effettuate misure più precise considerando tutti gli eventi maggiori di magnitudo 1.5 (o un altro valore limite più basso) e considerando un catalogo con una migliore completezza. Interessante sarebbe

studiare come quest'ultima influenzi il *declustering* finale. Anche l'utilizzo di sistemi di calcolo sempre più avanzati, potrà sicuramente aiutare nel procedimento. Così come l'implementazione di algoritmi in grado di svolgere più *declustering* contemporaneamente ed effettuare le operazioni in tempi più rapidi, facendo allo stesso tempo variare i parametri automaticamente e non manualmente, come nel caso del catalogo analizzato nella presente Tesi. Infatti, l'implementazione di codici in grado di effettuare automaticamente tutte queste operazioni nello stesso istante ancora non è presente ma potrebbe essere una soluzione che sicuramente porterà vantaggi notevoli, dato che si riuscirebbe allo stesso tempo ad acquisire i risultati del *declustering* di una determinata area, in un certo intervallo di tempo e di più metodi, variando i parametri per ciascun metodo considerato. Questa elaborazione, sicuramente necessiterà di una potenza di calcolo molto grande, ma anche di un grande spazio di archiviazione.

Di fondamentale importanza nel futuro, sarà osservare come la geologia locale influenzi il *declustering* e la distribuzione degli eventi. Infatti, ci potrà aiutare a comprendere se ci sono dei particolari *pattern* che si ripetono in zone geologicamente simili. Ad esempio, se in aree vulcaniche gli eventi indipendenti possiedono, nel tempo e nello spazio, una distribuzione differente da quella delle zone tettoniche, quali ad esempio le aree di subduzione o di *rifting*. Potranno perciò essere introdotti ulteriori parametri nelle equazioni di *declustering*, come parametri geologici, geofisici o geologico-strutturali, con lo scopo di poter al meglio isolare i terremoti indipendenti da quelli dipendenti, con la speranza di poter anche individuare *pattern* specifici nell'occorrenza dei *mainshocks* e giungere eventualmente un giorno ad una previsione un po' più attendibile dell'occorrenza degli eventi sismici.

Inoltre, un grande aiuto potrebbe derivare dall'utilizzo di reti neurali per compiere il *declustering*, o per tentare di prevedere con un'accuratezza maggiore la magnitudo e l'intervallo di tempo e spazio in cui potrebbero avvenire i successivi *mainshocks* (Asim, K.M., Idris, A., Iqbal, T., Martínez-Álvarez, F. (2018); Mignan, A., Broccardo, M. (2019); Panakkat, A., and Adeli, H. (2007)). Per effettuare ciò, ovviamente, è necessario avere a disposizione una quantità vastissima di dati, una strumentazione adeguata ed effettuare un numero elevato di esperimenti.

Innegabilmente, lo studio è molto complesso dato l'elevato numero di fattori che intervengono nel processo e c'è bisogno ancora di comprendere e definire tutti i meccanismi che interagiscono, per riuscire poi a quantificarli ed elaborare nuove soluzioni. Sicuramente una cooperazione tra figure anche di discipline differenti, come la geologia, la fisica, l'informatica, la statistica e molte altre, aiuterà nel futuro a trovare soluzioni alternative e a studiare da diversi punti di vista dei fenomeni così complessi.

8 - BIBLIOGRAFIA

8.1 - PUBBLICAZIONI

- Aki, K. (1956). Some problems in statistical seismology, *Zisin* 8, 205–228 (in Japanese). (English translation: A. S. Furimoto, Hawaii, 1963).
- Aki, K. (1965). Maximum likelihood estimate of b in the formula $\log N = a - bm$ and its confidence limits, *Bull. Earthquake Res. Inst.*, Tokyo Univ. 43, 237–238.
- Amini, H. (2014). COMPARING REASENBERG AND GRUENTHAL DECLUSTERING METHODS FOR NORTH OF IRAN, *Second European Conference on earthquake engineering and seismology*, Istanbul Aug. 25-29, 2014.
- Asim, K.M., Idris, A., Iqbal, T., Martínez-Álvarez, F. (2018). Earthquake prediction model using support vector regressor and hybrid neural networks, *PLoS ONE*, 13(7): e0199004, <https://doi.org/10.1371/journal.pone.0199004>.
- Baiesi, M. and Paczuski, M. (2004). Scale-free networks of earthquakes and aftershocks, *Phys. Physical review, E, Statistical, nonlinear, and soft matter physics*, 69. 066106. 10.1103/PhysRevE.69.066106.
- Belardinelli, M. E., Bizzarri, A., and Cocco, M. (2003). Earthquake triggering by static and dynamic stress changes, *JOURNAL OF GEOPHYSICAL RESEARCH*, VOL. 108, NO. B3, 2135, doi:10.1029/2002JB001779.
- Bhattacharya, P., Chakrabarti, B. K., Kamal, Samanta, D. (2009). Fractal Models of Earthquake Dynamics, *Reviews of Nonlinear Dynamics and Complexity*, ed. by H. G. Schuster, pp. 107 - 158, Wiley - VCH Verlag GmbH & Co. KGaA, Weinheim.
- Brown, L. D., and Zhao L. H. (2002) A TEST FOR THE POISSON DISTRIBUTION, *The Indian Journal of Statistics*, Volume 64, Series A, Pt. 3, pp 611-625.
- Castello, B., Selvaggi, G., Chiarabba, C., Amato, A. (2006). Catalogo della sismicità italiana 1981-2002, *Istituto nazionale di Geofisica e Vulcanologia (INGV)*, <https://doi.org/10.13127/CSI.1.1>.
- Castello, B., Olivieri, M., and Selvaggi, G. (2007). Local and Duration Magnitude Determination for the Italian Earthquake Catalog, 1981–2002, *Bulletin of the Seismological Society of America*, Vol. 97, No. 1B, pp. 128–139, February 2007, doi: 10.1785/0120050258.
- Daub, G. E., Trugman, D. T., and Johnson, P. A. (2015). Statistical tests on clustered global earthquake synthetic data sets, *J. Geophys. Res. Solid Earth*, 120, 5693–5716, doi:10.1002/2014JB011777.
- Gardner, J. K., and Knopoff, L. (1974). IS THE SEQUENCE OF EARTHQUAKES IN SOUTHERN CALIFORNIA, WITH AFTERSHOCKS REMOVED, POISSONIAN?, *Bulletin of the Seismological Society of America*, Vol. 64, October 1974, No. 5.

- Gugliemi, A. V. (2016). Interpretation of the Omori law, *Izvestiya, Physics of the Solid Earth*, 2016, No. 5.
- Hainzl, S., Scherbaum, F., and Beauval, C. (2016). Estimating Background Activity Based on Interevent-Time Distribution, *Bulletin of the Seismological Society of America*, Vol. 96, No. 1, pp. 313–320, February 2006, doi: 10.1785/0120050053.
- Han, Q., Wang, L., Xu, J., Carpinteri, A., Lacidogna, G. (2015). A robust method to estimate the b-value of the magnitude-frequency distribution of earthquakes, *Elsevier Ltd., Chaos, Solitons and Fractals* 81 (2015), 103-110.
- Improta, L., Latorre, D., Margheriti, L. *et al.* (2019). Multi-segment rupture of the 2016 Amatrice-Visso-Norcia seismic sequence (central Italy) constrained by the first high-quality catalog of Early Aftershocks, *Sci Rep* 9, 6921, <https://doi.org/10.1038/s41598-019-43393-2>.
- ISIDE Working Group. (2007). Italian Seismological Instrumental and Parametric Database (ISIDE), *Istituto Nazionale di Geofisica e Vulcanologia (INGV)*, <https://doi.org/10.13127/ISIDE>.
- Kagan, Y., and Knopoff, L. (1978). Statistical study of the occurrence of shallow earthquakes, *Geophys. J. K. astr. Soc.* (1978) 55, 67-86.
- Kanamori, H. and Anderson, D.L. (1975) Theoretical Basis of Some Empirical Relations in Seismology, *Bulletin of the Seismological Society of America*, 65, 1073-1095.
- Kagan, Y. Y. and Knopoff, L. (1979). Spatial distribution of earthquakes: the two-point correlation function, *Geophys. J. R. astr. SOC.* (1980) 62, 303-320.
- Knopoff, L. (1964). Statistics of earthquakes in Southern California, *Bull. Seism. Soc. Am.* 54, 1871-1873.
- Knopoff, L., and J. Gardner (1972), Higher Seismic Activity During Local Night on the Raw Worldwide Earthquake Catalogue, *Geophys. J. R. astr. Soc.*, 28, 311–313.
- Knopoff, L. (2000). The magnitude distribution of declustered earthquakes in Southern California, *PNAS*, October 24, 2000 97 (22) 11880-11884, <https://doi.org/10.1073/pnas.190241297>.
- Lombardi, A. M. (2016). SEDA a software package for the Statistical Earthquake Data Analysis: a tutorial application to the 2009 L'Aquila and the 2012 Emilia (Italia) sequences, *ANNALS OF GEOPHYSICS*, 60, 2, 2017; S0216; doi:10.4401/ag-7077.
- Luen, B., and Stark, P. B. (2012). Poisson Tests of Declustered Catalogs, *Geophys. Journal International*, Volume 189, Issue1, pp. 691-700.
- Markušić, S., Herak, D., Ivačić, Savić, I., Herak, M., and Prelogović (1998). Seismicity of Croatia in the period 1993-1996 and the Ston-Salno earthquake of 1996, *GEOFIZIKA*, VOL.15, 1998.
- Marzocchi, M., and Sandra L. (2003). A review and new insights on the estimation of the b-value and its uncertainty, *ANNALS OF GEOPHYSICS*, VOL. 46, N. 6, December 2003.
- Mignan, A., Broccardo, M. (2019). Neural Network Applications in Earthquake Prediction (1994–2019): Meta - Analytic and Statistical Insights on Their Limitations, *Seismological Research Letters* ; 91 (4): 2330–2342. doi: <https://doi.org/10.1785/0220200021>.
- Molchan, G. M., and Dmitrieva, O. E. (1992). Aftershock identification: methods and new approaches, *Geophys. J. Int.* (1992) 109, 501-516.

- Moro R., Basili A., Bono A., Marocci C., Mazza S., Mele F. M.. ISIDe, an Italian Seismic Data Base with minimum magnitude of completeness ML 1.8, *Italian Seismological Instrumental and Parametric Data-Base*, <http://hdl.handle.net/2122/5062>.
- Moulin A., Benedetti L., Rizza M., Rupnik P., Gosar A., et al. (2016). The Dinaric fault system: Large-scale structure, rates of slip, and Plio-Pleistocene evolution of the transpressive northeastern boundary of the Adria microplate, *Tectonics, American Geophysical Union (AGU)*, 2016, 35, pp.2258 - 2292. [ff10.1002/2016TC004188](https://doi.org/10.1002/2016TC004188)[ff. fahal-01420780](https://doi.org/10.1002/2016TC004188).
- Naylor, M., Orfanogiannaki, K., and Harte, D. (2010). Exploratory Data Analysis: Magnitude, Space, and Time, *Community Online Resource for Statistical Seismicity Analysis*, doi: 10.5078/corssa-92330203, <http://www.corssa.org>.
- Panakkat, A., and Adeli, H. (2007). Neural Network Models for Earthquake Magnitude Prediction Using Multiple seismicity Indicators, *International journal of neural systems* 17 1: 13-33.
- Reasenber, P. (1985). "Second-order moment of central California seismicity", *J. Geophys. Res.*, 90, 5479-5495.
- Sartori, R. (2003). Episodes, The Tyrrhenian back-arc basin and subduction of the Ionian lithosphere, Dipartimento di Scienze della Terra e Geologico-Ambientali, University of Bologna, *Episodes*, Vol. 26, no. 3.
- Shi, Y., and Bolt, B.A. (1982). The standard error of the magnitude-frequency *b* value, *Bulletin of the Seismological Society of America*, 72 (5): 1677–1687.
- Uhrhammer, R. (1986), Characteristics of Northern and Central California Seismicity, *Earthquake Notes*, 57(1).
- Van Stiphout, T., J. Zhuang, and D. Marsan (2012). Seismicity declustering, *Community Online Resource for Statistical Seismicity Analysis*, doi:10.5078/corssa52382934, <http://www.corssa.org>.
- Vorobieva, I. A., Peresan, A., Soloviev, A. A., and Panza, G.F. (2007). Study of strong seismicity in the Alps and Dinarides Junction zone: block structure modeling and historical observations, *Pure and Applied Geophysics*, December 2007, 164(11):2193-2234, doi: 10.1007/s00024-007-0273-9.
- Wiemer, S. (2001). A software package to analyze seismicity: ZMAP, *Seismological Research Letters*, 72(3), pp.373-382, <https://doi.org/10.1785/gssrl.72.3.373>.
- Woessner, J., J.L. Hardebeck, and E. Hauksson (2010), What is an instrumental seismicity catalog, *Community Online Resource for Statistical Seismicity Analysis*, doi:10.5078/corssa-38784307, <http://www.corssa.org>.
- Yaolin S., and Bolt, B.A. (1982). The standard error of the Magnitude-frequency, *Bulletin of the Seismological Society of America*, Vol. 72, No. 5, pp. 1677-1687, October 1982.
- Zaliapin, I., Gabrielov, A., Keilisi-Borok, V., and Wong, H. (2008). Clustering Analysis of Seismicity and Aftershock Identification, *Physical Review Letters*, 101(1):018501 DOI: 10.1103/PHYSREVLETT.101.018501.

8.2 - LIBRI

- McKinney W. (2017). Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython 2nd Edition, *O'Reilly Media, Inc, USA*.
- Riley, K. F. (2006). Mathematical Methods for Physics and Engineering, pp. 1119-1298, *Cambridge University Press*.

8.3 - SITOGRAFIA

- Anuradha Bhatia (2017). Kolmogorov - Smirnov Test [K-S Test], <https://www.youtube.com/watch?v=MEcAn5yzTwA>.
- GeoDa Software (2016). Point Pattern Analysis: Point Process Models, <https://www.youtube.com/watch?v=8v3C5023GIE&list=LLU84SO7V9r54Ssd1HcrW0ug&index=53>.
- Holmes, S. (2001). 9 Maximum Likelihood Estimation, <http://statweb.stanford.edu/~susan/courses/s200/lectures/lect11.pdf>.
- Jbstatistics (2014). Poisson or Not? (When does a random variable have a Poisson distribution?), https://www.youtube.com/watch?v=sv_KXSiorFk.
- Khan Academy (2010). Hypothesis testing and p-values | Inferential statistics | Probability and Statistics | Khan Academy, <https://www.youtube.com/watch?v=-FtlH4svqx4>.
- Matthew E. Clapham (2016). 10: Kolmogorov-Smirnov test, <https://www.youtube.com/watch?v=ZO2RmSkXK3c>.
- Professor Knudson (2017). 4. MLE Example: Poisson, <https://www.youtube.com/watch?v=n8YcOUZRZy8>.
- Roger D. Peng (2020). 7.1 Stationary Poisson Process, <https://bookdown.org/rdpeng/timeseriesbook/stationary-poisson-process.html>.
- Sarkar, D. (2018) The Art of Effective Visualization of Multi-dimensional Data Strategies for Effective Data Visualization, <https://towardsdatascience.com/the-art-of-effective-visualization-of-multi-dimensional-data-6c7202990c57>.
- The SciPy community, (2020). Statistical functions ([scipy.stats](https://docs.scipy.org/doc/scipy/reference/stats.html)), <https://docs.scipy.org/doc/scipy/reference/stats.html>.
- Zedstatistics (2018). Likelihood | Log likelihood | Sufficiency | Multiple parameters, <https://www.youtube.com/watch?v=ScdunwtrMzc&list=LLU84SO7V9r54Ssd1HcrW0ug&index=52>.
- Zedstatistics (2019), What are "moments" in statistics? An intuitive video!, https://www.youtube.com/watch?v=ISaVvSO_3Sg&list=LLU84SO7V9r54Ssd1HcrW0ug&index=54.

