

ALMA MATER STUDIORUM · UNIVERSITÀ DI BOLOGNA

Scuola di Scienze
Corso di Laurea Magistrale in Fisica del Sistema Terra

**Applicazione di tecniche di clustering
spaziale su dati macrosismici di felt reports
per stimare i parametri dei terremoti**

Relatore:
Prof. Paolo Gasperini

Presentata da:
Bruno Martinico

Correlatore:
Dott. Gianfranco Vannucci

Sessione III
Anno Accademico 2019/2020

*A Carla,
ai miei Horcrux dispersi,
alle cose immortali.*

L'uomo ambizioso ritiene bene
personale un attività che non è
la propria

Marco Aurelio
Meditationes

Abstract

Questa tesi tratta lo sviluppo e l'applicazione di tecniche di clustering a dati macrosismici di risentimento individuali, ovvero dei singoli utenti, raccolti su piattaforme internet e l'uso di dati di intensità per ricavare informazioni su localizzazione e magnitudo dei terremoti. La procedura si articola in 2 step principali: raggruppamento delle singole intensità individuali (IDPs) attraverso metodi di clustering per ricavare intensità macrosismiche (MDPs) simili alle località secondo le definizioni delle scale macrosismiche (MCS, EMS); elaborazione di tali intensità (MDPs) mediante l'applicazione del codice Boxer (Gasperini et al.,1999; 2010). Sono stati analizzati 9 terremoti in totale avvenuti in Italia dal 2016. Sono state utilizzate 5 tecniche per ricavare intensità MDPs dalle intensità IDPs di EMSC e HSIT, piattaforme di raccolta di IDPs web-based. Le tecniche applicate sono state sviluppate in Matlab e sono: grid-based spaziali con celle equiareali quadrate ed esagonali; aree circolari con raggio proporzionale alla densità; tecniche miste (cerchi+celle equiareali); tecnica DBSCAN. Le MDPs costruite da dati EMSC sono state poi elaborate con il codice Boxer e i risultati ottenuti confrontati con i parametri ricavati per le MDPs di HSIT e con i dati strumentali. La tesi è strutturata in 4 capitoli: 1) trattazione generale sulla fisica del fenomeno sismico e descrizione di cause ed effetti. Vengono descritti i parametri sintetici strumentali e gli equivalenti macrosismici. Viene quindi descritto l'algoritmo di Boxer utilizzato. 2) trattazione sui felt report raccolti da Internet come dati macrosismici con proprie caratteristiche e problematiche specifiche. 3) definizione della trasformazione tra IDPs e MDPs e principali tipologie di tecniche di clustering per effettuarla. 4) descrizione dei dataset dei terremoti utilizzati, dell'elaborazione e dell'analisi condotte.

Indice

Introduzione	1
1 Sismologia macrosismica per la stima dei parametri dei terremoti	3
1.1 Il terremoto e la sorgente sismica	3
1.1.1 La sorgente sismica	3
1.1.2 Le onde sismiche	4
1.2 I parametri dei terremoti	6
1.2.1 Localizzazione dell'ipocentro e dell'epicentro	6
1.2.2 La magnitudo	7
1.2.3 I parametri di faglia	9
1.2.4 Il meccanismo focale	10
1.3 Sismologia macrosismica	12
1.3.1 Scale d'intensità	15
1.3.2 L'intensità	16
1.3.3 L'indagine macrosismica	19
1.3.4 Approccio macrosismico alla stima dei parametri sintetici	22
1.3.5 Cataloghi parametrici dei terremoti	26
1.4 BOXER: stima dei parametri sintetici mediante programma di elaborazione di dati macrosismici	28
1.4.1 Localizzazione macrosismica	30
1.4.2 Magnitudo equivalente	32
1.4.3 Dimensioni e orientazione della faglia	33
1.4.4 Incertezze	36
1.4.5 Accuratezza, affidabilità e confronti tra i metodi	37
2 Dati macrosismici di felt reports raccolti da Internet su base volontaria	39
2.1 Stato dell'arte dei felt reports raccolti da Internet per la sismologia macrosismica citizen-based	39
2.2 Caratteristiche specifiche e problematiche dei dati	42
3 Tecniche di clustering spaziale per dati macrosismici di felt reports	47
3.1 Generalità della "cluster analysis spaziale"	47
3.2 Scelta delle tecniche di clustering spaziale da utilizzare per i dati EMSC	51
3.3 Tecnica 1: griglia di quadrati equiareali	54
3.4 Tecnica 2: griglia di esagoni equiareali	58
3.5 Tecnica 3: cerchi attorno a località	60
3.6 Tecnica 4: mista, cerchi attorno a località e griglia esagonale	63

3.7	Tecnica 5: DBSCAN	65
4	Analisi dei risultati	69
4.1	Scelta dei datasets	69
4.2	Elaborazione dei dati	70
4.3	Discussione risultati	73
4.3.1	Magnitudo	74
4.3.2	Localizzazione epicentrale	88
	Conclusioni	107
	Riconoscimenti	109
	Riferimenti bibliografici	111
A	Costruzione delle griglie di quadrati ed di esagoni, equiareali, di passo variabile.	121

Introduzione

Questa tesi tratta lo sviluppo e l'applicazione di tecniche di clustering a dati macrosismici di risentimento individuali, ovvero dei singoli utenti, raccolti su piattaforme internet e l'uso di dati di intensità per ricavare informazioni parametriche dei terremoti, in particolare localizzazione e magnitudo. Tale procedura si articola in due step principali: il primo è il raggruppamento dei singoli dati/intensità individuali (IDPs) attraverso tecniche di clustering per ricavare intensità macrosismiche (MDPs) idealmente simili alle località secondo le definizioni delle scale macrosismiche (MCS, EMS); il secondo è l'elaborazione di tali intensità (MDPs) mediante l'applicazione del codice Boxer (Gasperini et al.,1999; 2010).

Sono stati analizzati 9 terremoti in totale avvenuti in Italia dal 2016.

Per quanto concerne il primo step, sono state utilizzate 5 tecniche per ricavare intensità MDPs partendo dalle IDPs di EMSC e HSIT. EMSC è un ente che fornisce a scala globale le IDPs raccolte tramite una apposita piattaforma EQN, basata su risposte degli utenti a disegni (thumbnails) di scenari macrosismici di scuotimento e danno; HSIT è invece un'applicazione dell'INGV che raccoglie per l'area italiana le risposte di utenti a questionari testuali. EMSC raccoglie intensità sull'intero range di definizione della scala EMS (1-12), mentre HSIT fornisce stime EMS e MCS nel solo range 1-8, inferiore quindi al danno strutturale sulle abitazioni.

HSIT elabora inoltre le risposte dei singoli questionari per ottenere a scala comunale l'intensità macrosismica, quindi utilizza un criterio di clustering basato sull'appartenenza o meno di un punto ad una polyline.

Per il clustering dei dati EMSC l'idea di base seguita è che un approccio basato sulla densità possa essere in grado di individuare clusters di punti di intensità relativi a luoghi anche piccoli purché densamente campionati, quindi attribuendo loro un'intensità propria, distinta da quella di eventuali centri abitati maggiori limitrofi, in genere più campionati. Le tecniche sono: grid-based spaziali con celle equiareali quadrate (1) ed esagonali (2); aree circolari con raggio proporzionale alla densità media di popolazione centrati su località amministrative prese da database (3); tecnica miste (aree circolari+celle equiareali) (4); tecnica DBSCAN (Ester, 1996) (5). Gli algoritmi per il clustering delle tecniche sono stati scritti in ambiente Matlab.

Le MDPs ottenute da EMSC sono state poi elaborate con il codice Boxer e i risultati ottenuti confrontati in termini di distanze epicentrali e di differenze di magnitudo con i parametri ricavati per le MDPs di HSIT e con i dati strumentali.

La tesi è strutturata come segue.

Il primo capitolo consta di una trattazione generale sulla fisica del fenomeno sismico e ne descrive cause ed effetti. L'attenzione viene posta sui parametri sintetici strumentali utilizzati per la caratterizzazione dei terremoti e delle faglie responsabili.

In particolare è messo in risalto l'approccio macrosismico alla stima dei parametri sintetici, a cui segue la descrizione dettagliata dell'algoritmo di Boxer utilizzato.

Il secondo capitolo è dedicato ai felt report raccolti da Internet i quali hanno caratteristiche e problematiche specifiche e differenti rispetto ai più tradizionali dati macrosismici.

Il capitolo tre tratta il problema dell'irrobustimento dei dati macrosismici sparsi, e cioè la trasformazione tra gli IDPs (Individual intensity Data Points), rilevazioni grezze delle piattaforme Internet-based, e gli MDPs (Macroseismic Data Points), punti di intensità consolidati costituenti il campo macrosismico utili per successive considerazioni ed elaborazioni. In questo capitolo sono inoltre descritte le principali tipologie di tecniche di clustering disponibili in letteratura e la loro applicazione specifica. Sono quindi esposti i criteri di scelta delle tecniche utilizzate.

Il quarto capitolo descrive i dataset dei terremoti utilizzati e le analisi condotte su di essi. I terremoti sono stati scelti in base ai seguenti criteri: (1) numero di IDPs dei dataset di EMSC più elevato possibile; questo è crescente nel tempo in relazione all'efficienza della piattaforma di raccolta, perciò si è stabilito l'anno 2017 come soglia inferiore; (2) epicentro strumentale localizzato nel territorio italiano o in prossimità di esso per avere disponibilità di dati della piattaforma di raccolta nazionale HSIT; (3) magnitudo strumentale più elevata possibile ($M \geq 5$) per ragioni di calibrazione di Boxer su grandi terremoti. Segue dunque l'analisi e l'interpretazione degli output ottenuti con Boxer per i diversi tipi di MDPs.

Capitolo 1

Sismologia macrosismica per la stima dei parametri dei terremoti

1.1 Il terremoto e la sorgente sismica

Il termine terremoto (dal latino *terrae motus*, “movimento della terra”) indica una serie di movimenti della superficie terrestre provocati da onde elastiche (dette onde sismiche) generate dall’improvviso rilascio di energia potenziale elastica accumulata nella litosfera per via dei movimenti tettonici.

Il meccanismo di generazione dei terremoti può essere descritto mediante la teoria del rimbalzo elastico (*elastic-rebound theory*) proposta da Harry Fielding Reid nel 1910 in seguito allo studio del terremoto di San Francisco del 1906 (Reid, 1910).

Un terremoto è il risultato di un lungo periodo preparatorio in cui ha luogo il lento accumulo dell’energia che verrà in seguito rilasciata. All’origine di questo processo c’è il movimento delle placche tettoniche le quali, muovendosi l’una rispetto all’altra, si deformano creando uno stato di sforzo, con conseguente accumulo di energia elastica. Le rocce della superficie terrestre presentano numerose fratture, dette faglie, in grado di subire uno scorrimento relativo dei due lembi. Se due placche adiacenti si muovono l’una rispetto all’altra tendono a produrre uno scorrimento delle faglie presenti nella zona di margine, le cui pareti rimangono tuttavia serrate l’una contro l’altra per azione della pressione litostatica fino a quando lo sforzo non supera la resistenza dell’attrito presente tra esse. A questo punto le pareti di faglia si svincolano improvvisamente causando il rimbalzo elastico e la produzione di onde sismiche che si propagano nel mezzo terrestre. In seguito allo scorrimento, lo sforzo e la deformazione si annullano e il processo ricomincia da capo.

Le onde sismiche vengono generate all’interno di un mezzo elastico per mezzo dell’applicazione di un sistema di forze che variano nel tempo. Tale sistema di forze definisce la sorgente sismica e la relazione che ne descrive la dipendenza dalla posizione e dal tempo è detta funzione sorgente. Si chiama durata della sorgente l’intervallo di tempo durante il quale la forza varia.

1.1.1 La sorgente sismica

Il caso più semplice che si possa considerare è quello di una forza singola applicata in un punto del mezzo elastico tuttavia un’ipotesi più plausibile per la rappresentazione

di un fenomeno naturale interno alla Terra è quello di una doppia coppia (Maruyama, 1963). Più in generale una sorgente sismica può essere considerata equivalente ad un sistema di doppie coppie di forze.

Lo scorrimento di una faglia può essere descritto da una dislocazione di taglio (*shear dislocation*). La superficie di dislocazione è l'insieme dei punti per i quali lo scorrimento è una funzione a due valori. Considerando l'intera superficie di dislocazione possiamo immaginare che in ciascun punto di essa sia collocata una doppia coppia di forze che produce lo stesso effetto della dislocazione, la quale è caratterizzata in maniera completa in ogni punto dalla densità di momento (momento per unità di area).

A grande distanza dalla sorgente la dislocazione può essere rappresentata da un momento complessivo scalare, detto momento sismico, indicato con M_0 e legato allo spostamento medio finale delle pareti della faglia D , all'area della superficie della faglia A e alla rigidità delle rocce μ dalla seguente relazione (Aki, 1966):

$$M_0 = \mu DA \quad (1.1)$$

Il momento sismico fornisce anche una misura della deformazione anelastica irreversibile nella regione di frattura.

1.1.2 Le onde sismiche

L'energia liberata dalla sorgente si propaga all'interno della Terra sotto forma di onde elastiche dette onde sismiche. Tali onde possono essere suddivise in:

- Onde di volume (*body waves*): vengono generate contemporaneamente dalla sorgente sismica e si propagano in tutte le direzioni; possono essere ulteriormente distinte in:
 - Onde P (primarie): sono onde longitudinali (o di compressione), ovvero provocano una deformazione del mezzo elastico nella stessa direzione in cui si propagano; si propagano sia attraverso materiali allo stato solido sia fluido con una velocità che va da 1,45 km/s a 13,7 km/s (Kennett e Engdahl, 1995);
 - Onde S (secondarie) sono onde trasversali (o di taglio), ovvero producono una deformazione perpendicolare alla direzione di propagazione dell'onda; si propagano attraverso mezzi solidi, in grado di sopportare gli sforzi di taglio, ma si smorzano rapidamente quando incontrano materiali fluidi; la velocità va da 1,0 km/s a 7,3 km/s (Kennett e Engdahl, 1995);
- Onde di superficie (*surface waves*); sono onde la cui ampiezza decresce rapidamente con la profondità; si propagano lungo la superficie della Terra ma anche lungo le superfici interne di discontinuità delle proprietà elastiche; possono essere distinte in:
 - Onde di Rayleigh: prendono il nome da Lord Rayleigh, che ne predisse l'esistenza nel 1885; fanno compiere alle particelle del mezzo una traiettoria ellittica verticale e retrograda;

- Onde di Love: prendono il nome da A. E. H. Love, che ne predisse l'esistenza nel 1911; impongono alle particelle del mezzo un moto orizzontale e trasversale (simile a quello delle onde S orizzontali).

Entrambi i tipi di onde superficiali sono più lenti delle onde P e S. Nei sismogrammi vengono registrate per prime le onde P, le più veloci, per questo chiamate primarie, poi le onde S e infine le onde superficiali. La rilevazione delle onde sismiche allorché raggiungono la superficie è la principale osservazione diretta praticabile dell'intero fenomeno sismico.

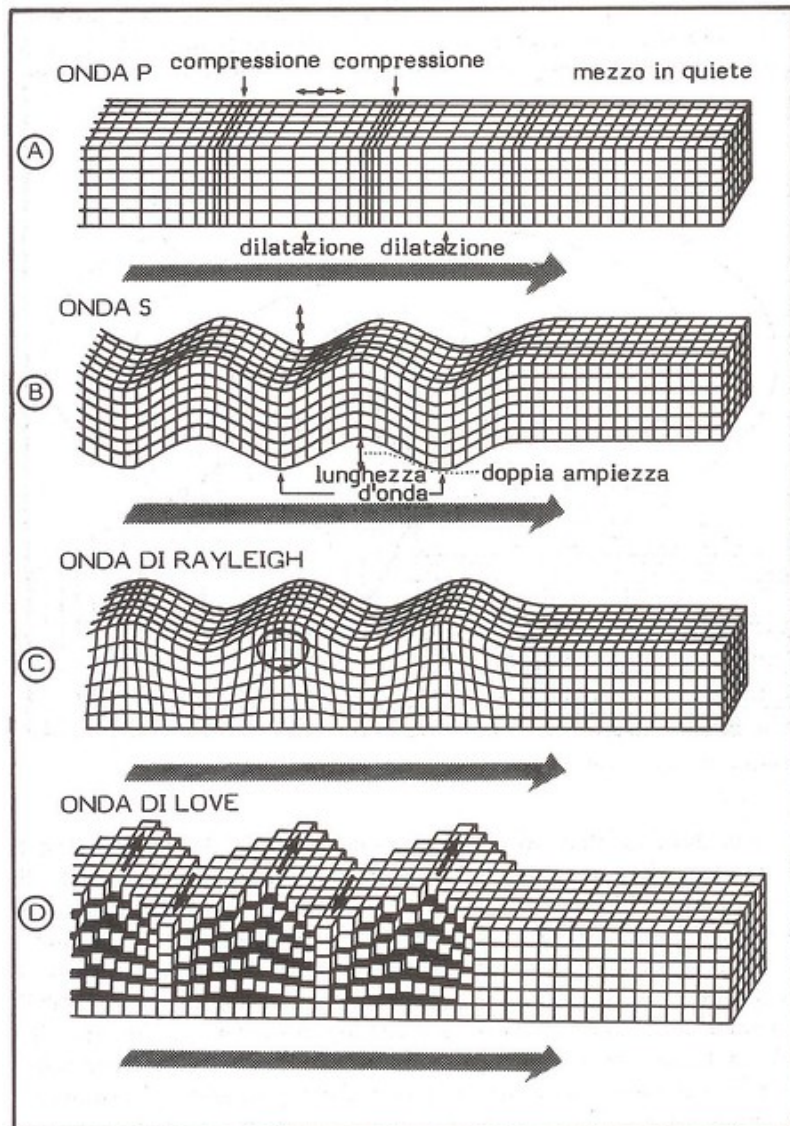


Figura 1.1: Oscillazioni del mezzo per i quattro principali tipi di onde sismiche (calcolostrutture.com)

Siccome la Terra è un mezzo eterogeneo in buona approssimazione stratificato, le onde di volume sono soggette a fenomeni di riflessione e rifrazione, dando origine, oltre alle già citate onde superficiali, ad una grande varietà di fasi sismiche (vengono indicate con pP e sS le onde P ed S riflesse dalla superficie libera prima di raggiungere la stazione, PKP le onde P che hanno attraversato il nucleo esterno ecc.).

È inoltre opportuno menzionare il fatto che, nonostante i modelli più semplici di propagazione delle onde sismiche assumano la Terra come isotropa, o quantomeno formata da strati isotropi, essa presenta un livello di anisotropia, in particolare nel mantello (*Beghein et al.*, 2006; *Panning and Romanowicz*, 2006), responsabile di variazioni delle velocità di propagazione delle onde sismiche maggiori di quelle dovute alle variazioni di temperatura e di composizione chimica (*Anderson*, 1989).

Una delle principali conseguenze dell'anisotropia sono i differenti tempi di arrivo delle due diverse polarizzazioni delle onde S (*shear wave splitting* o *shear wave birefringence*), ovvero le onde S polarizzate orizzontalmente, SH, e verticalmente SV (*Levin et al.*, 2007).

Un altro importante fenomeno che interessa le onde sismiche è quello dell'attenuazione, ovvero la diminuzione dell'ampiezza con la distanza dalla sorgente: essa è dovuta sia alla dissipazione anelastica dell'energia trasportata dall'onda (smorzamento interno o intrinseco), sia a ragioni geometriche di distribuzione di tale energia sull'area crescente della superficie del fronte d'onda (smorzamento geometrico), sia ai fenomeni di riflessione e rifrazione, citati poc'anzi, i quali redistribuiscono l'energia a diverse fasi sismiche (smorzamento per scattering).

L'attenuazione intrinseca delle onde sismiche dipende da molti fattori, quali il tipo di roccia, la temperatura, la granulometria e la presenza di materiale fuso o di acqua (*David et al.*, 2019).

A livello di immediato sottosuolo (profondità $d < 30Km$ circa) le condizioni geologiche, ma anche le condizioni morfologiche locali, possono modificare in maniera significativa la forma dello spettro a parità di energia e distanza dalla sorgente e si possono avere talvolta fenomeni di amplificazione (*Albarelo*, 2013). Si parla in questo caso di effetti di sito e di risposta sismica locale.

1.2 I parametri dei terremoti

Ogni terremoto può essere descritto da una serie di parametri, detti parametri sintetici del terremoto: l'ipocentro, l'epicentro, il tempo origine, il momento sismico e la magnitudo.

Viene chiamato ipocentro (o fuoco) il punto, situato all'interno della Terra, dove hanno origine le onde sismiche. La proiezione di questo punto sulla superficie terrestre viene detta epicentro. Queste definizioni presuppongono l'utilizzo di un modello puntiforme della sorgente sismica.

Il tempo origine è definito come l'istante di tempo in cui lo scorrimento ha inizio e le onde cominciano ad irradiarsi dall'ipocentro.

1.2.1 Localizzazione dell'ipocentro e dell'epicentro

Per determinare la posizione dell'epicentro di un terremoto si sfrutta la differente velocità delle onde P e delle onde S. In particolare, siccome le onde S si propagano più lentamente rispetto alle onde P, più il terremoto è distante dal ricevitore maggiore è l'intervallo di tempo che intercorre tra gli arrivi dei due tipi di onde. Mettendo in correlazione questo ritardo con le curve standard del tempo di propagazione delle onde P e delle onde S in funzione della distanza (dromòcrone) è possibile leggere la distanza dell'epicentro dalla stazione.

Per determinare la posizione è necessario conoscere la distanza da almeno tre stazioni, a partire dalle quali si tracciano circonferenze o archi di circonferenza di raggio pari alla distanza trovata e la loro intersezione indica l'ubicazione dell'epicentro. La profondità dell'ipocentro può essere determinata misurando la differenza tra l'arrivo del raggio diretto (raggio P) e quello del raggio che si riflette sulla superficie (raggio pP). Maggiore è la differenza tra i tempi di arrivo tra i due segnali maggiore è la profondità dell'ipocentro (Mussett e Aftab Khan, 2003).

La determinazione della profondità dell'ipocentro è il parametro più difficile da determinare per via del fatto che il tempo di viaggio varia molto lentamente con la profondità (Haskov *et al.*, 2011).

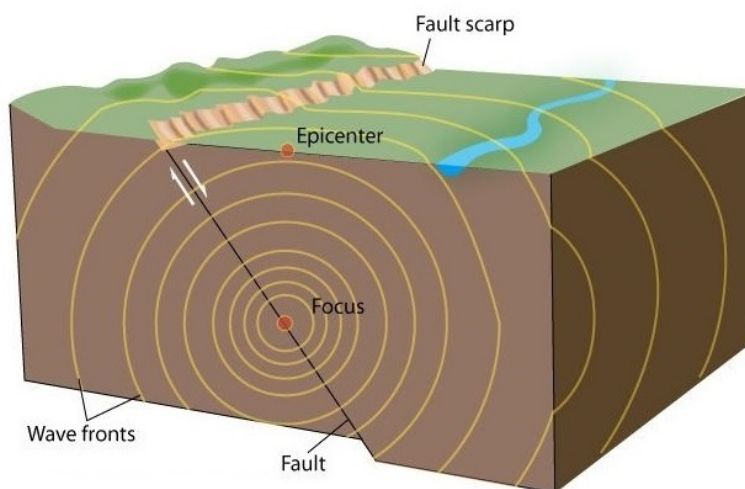


Figura 1.2: Rappresentazione schematica di alcuni parametri del terremoto; nella figura sono anche mostrati i fronti d'onda e la scarpata di faglia. Fonte: University of Waikato (NZ)

1.2.2 La magnitudo

La magnitudo è una grandezza adimensionale legata all'energia del terremoto alla sorgente. Quello della quantificazione della “grandezza” dei terremoti è uno dei problemi principali della sismologia fin dai primordi.

Il primo ad introdurre una valutazione basata su misure strumentali fu Charles F. Richter nel 1935 (C. Richter, 1935). La magnitudo introdotta da Richter, detta anche magnitudo locale in quanto valida solo per distanze epicentrali inferiori ai 1000 km, è una stima indiretta dell'energia elastica rilasciata dall'evento sismico alla sorgente basata sull'ampiezza massima dell'oscillazione prodotta dal moto del terreno su un sismografo di riferimento.

Una possibile definizione generale della magnitudo è data dalla seguente formula (Havskov e Lars, 2008):

$$M = \log_{10} \left(\frac{A}{T} \right) + f(\Delta, h) + C \quad (1.2)$$

dove A è l'ampiezza massima della traccia registrata per quel terremoto (in micrometri), T è il periodo dell'onda (in secondi), f è una funzione della distanza

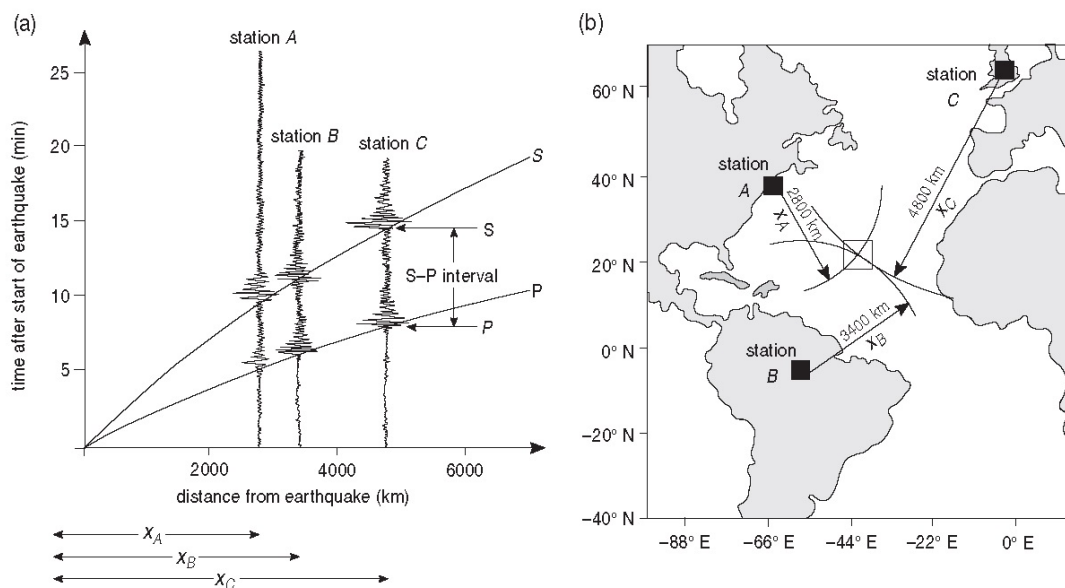


Figura 1.3: Localizzazione di un terremoto utilizzando le differenze nei tempi di arrivo delle onde P e delle onde S. (Mussett e Aftab Khan, 2003)

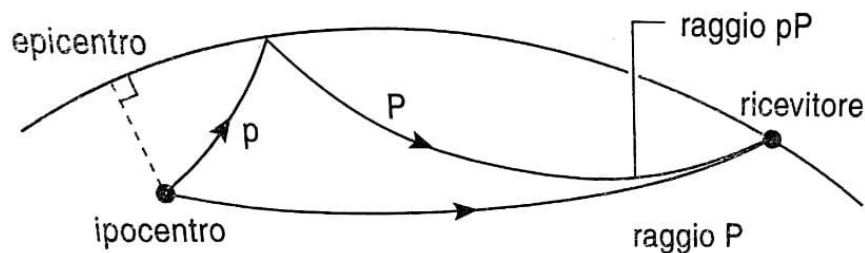


Figura 1.4: Percorso di un raggio P e di un raggio riflesso dalla superficie pP. (Mussett e Aftab Khan, 2003)

epicentrale Δ e della profondità dell'ipocentro h e C è un fattore di correzione che tiene conto della geologia locale della zona di interesse.

Successivamente sono state definite altre magnitudo, in modo tale da applicare il concetto a terremoti in un ampio raggio di distanze e profondità; tra queste vale la pena di citare la magnitudo delle onde di volume (*body-wave magnitude*), adatta ai terremoti più profondi, e la magnitudo delle onde superficiali (*surface-wave magnitude*), adatta a valutare terremoti con distanza epicentrale superiore a 600 km (Gutenberg, 1945a,b) e la magnitudo di durata, basata sulla durata totale del segnale sismico (Bisztricsany, 1958). Per terremoti molto grandi (con magnitudo maggiore di 8), le diverse scale tendono a saturarsi, ovvero ad attribuire valori non troppo differenti a terremoti che rilasciano quantità di energia elastica significativamente diverse. Ciò accade per via del fatto che, quando la superficie di dislocazione è molto estesa, l'emissione delle onde sismiche non avviene contemporaneamente da tutta la superficie ma solo da una sua parte, la quale varia con l'avanzare del fronte della dislocazione (Dragoni, 2005).

Per superare questi limiti è stata introdotta la magnitudo momento (*Kanamo-*

ri, 1977; (Hanks e Kanamori, 1979), basata sul momento sismico ed espressa dalla relazione

$$M_w = \frac{2}{3} \log_{10} M_0 - 6 \quad (1.3)$$

dove M_0 è il momento sismico misurato in newton per metro e il pedice w sta per “work” (lavoro meccanico). La magnitudo può essere calcolata dal momento sismico grazie al fatto che è legata all’energia elastica rilasciata e che quest’ultima è proporzionale al momento sismico.

La magnitudo momento è oggi ritenuta il parametro più significativo per valutare la “grandezza” di un terremoto per via del suo significato fisico legato alle dimensioni della sorgente, la deformazione cosismica e la rigidità del mezzo elastico in cui la sorgente agisce.

1.2.3 I parametri di faglia

Quanto discusso finora presuppone l’utilizzo del modello puntiforme di sorgente sismica. Le osservazioni geologiche mostrano tuttavia che le faglie sono variamente estese ed orientate nello spazio e queste caratteristiche influiscono sulle proprietà del treno d’onde emesso e rilevato dagli strumenti. Pertanto, una caratterizzazione più raffinata di un evento sismico richiede di specificare ulteriori parametri: i parametri geometrici e cinematici della faglia che lo ha prodotto.

Il piano di faglia, che rappresenta la superficie in corrispondenza della quale i due lembi di una roccia fratturata si spostano, può essere verticale o inclinato. Se il piano di faglia è più o meno inclinato, nella roccia si distinguono il tetto (o *hanging wall*, ovvero la parte che sta sopra il piano di faglia) ed il letto (o muro o *footwall*, ovvero la parte che sta sotto il piano di faglia); lo spostamento del tetto rispetto al letto permette di distinguere diversi tipi di faglie (Elmi e Diretto, 1996):

Se prevale la componente verticale dello spostamento si ha una:

- faglia diretta (o normale): in questo caso il tetto scende lungo il letto;
- faglia inversa (*reverse fault* o *thrust fault*): il tetto risale lungo il letto;

Se prevale la componente orizzontale dello spostamento si ha una:

- faglia trascorrente (*transform fault*): può essere destra o sinistra;

Quando sono significative entrambe le componenti si ha una:

- faglia obliqua;

L’orientazione nello spazio del piano di faglia viene detto giacitura. La giacitura del piano di faglia o di altre superfici geologiche planari è determinata da tre angoli:

- La direzione (*strike*) ϕ : è l’angolo che una retta orizzontale che giace sul piano di faglia forma con la direzione del Nord geografico; è misurato in verso antiorario rispetto al Nord geografico ($0 \leq \phi \leq 2\pi$);
- L’immersione (*dip*) δ : esprime la direzione verso cui la superficie si immerge, misurata rispetto al piano orizzontale ($0 \leq \delta \leq \pi/2$); l’immersione forma un angolo retto con la direzione;

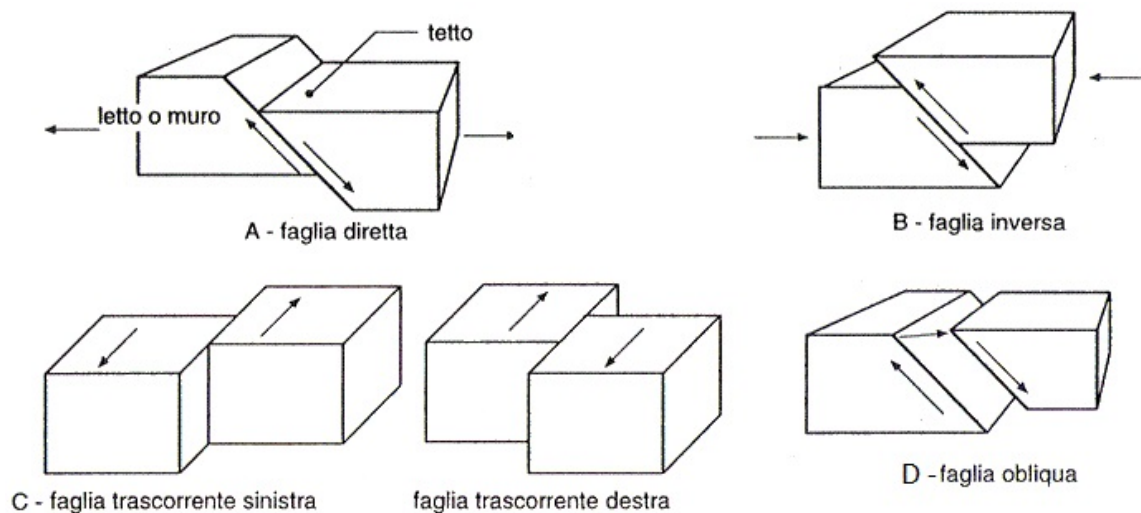


Figura 1.5: Rappresentazione schematica dei principali tipi di faglie. (Elmi e Diretto, 1996)

- L'inclinazione (*rake*) λ : esprime la direzione del vettore spostamento del tetto rispetto al letto misurata sul piano di faglia in senso antiorario rispetto alla direzione di strike ($-\pi \leq \lambda \leq \pi$).

Per ricavare l'orientamento del piano di faglia e la direzione dello spostamento su di esso occorre conoscere la direzione dei primi arrivi a un certo numero di ricevitori posti attorno all'epicentro in modo tale da ricoprire più uniformemente possibile l'angolo giro (Mussett e Aftab Khan, 2003).

Le faglie dirette e inverse vengono anche chiamate *dip-slip* in quanto lo scorrimento avviene nella direzione di *dip* mentre le faglie trascorrenti vengono anche dette *strike-slip* in quanto lo scorrimento avviene nella direzione di *strike*.

1.2.4 Il meccanismo focale

Dalle registrazioni di un sismogramma si può interpretare il primo arrivo delle onde P come una compressione o come una dilatazione. Osservando la distribuzione degli impulsi di dilatazione e compressione nei vari osservatori, che dipende dal "meccanismo" con cui agisce la sorgente sismica, è possibile stabilire la direzione di provenienza delle onde sismiche e da lì ricavare informazioni sulla geometria della faglia (orientazione del piano di faglia nello spazio e direzione del movimento dei due blocchi).

Per le onde P ed S è possibile rappresentare l'ampiezza dello spostamento del suolo al variare della direzione di osservazione ad una stessa distanza dalla sorgente. Tale rappresentazione prende il nome di diagramma di radiazione (*radiation pattern*) e assume una forma a quattro lobi (vedi figura 1.7).

Convenzionalmente si considera positiva la polarità del primo impulso quando il moto è diretto dalla sorgente verso la stazione sismometrica e negativa nel caso contrario. I lobi del quadrifoglio con polarità positiva sono disposti parallelamente alla direzione di sforzo distensivo, mentre quelli con polarità negativa sono paralleli alla direzione di sforzo compressivo.

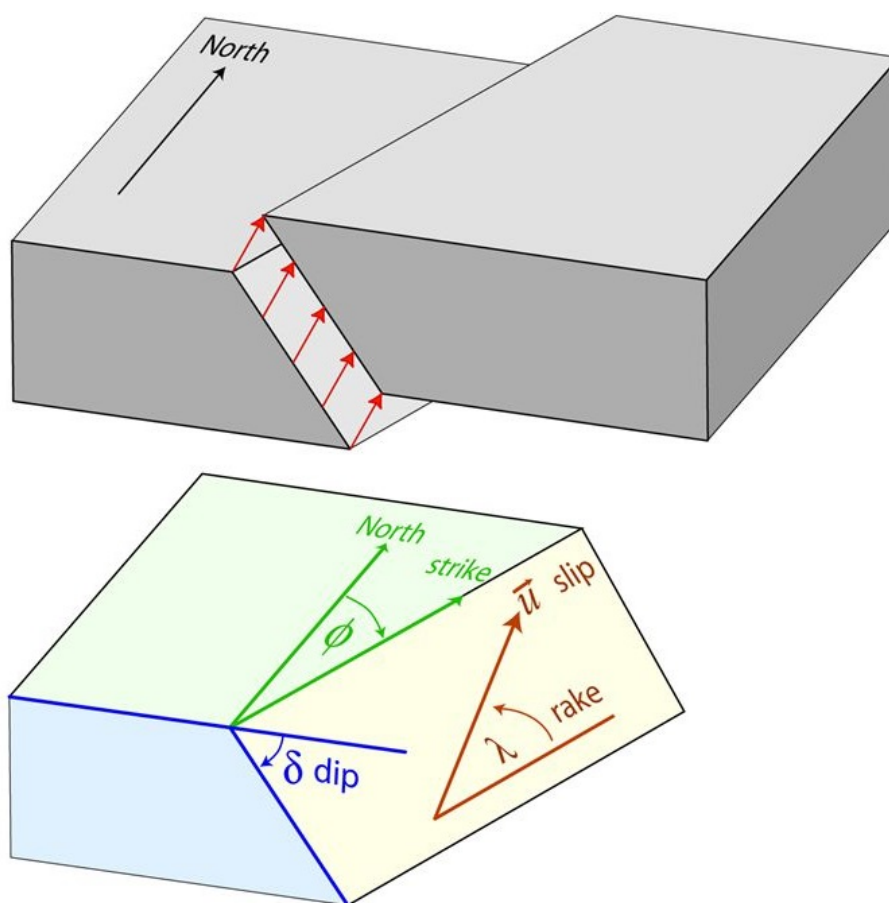


Figura 1.6: Rappresentazione degli angoli che individuano la giacitura del piano di faglia (Wilcock, 2016).

Per la simmetria del diagramma di radiazione rispetto alle direzioni N-S e E-O, non è possibile stabilire, solo sulla base di questo, quale delle due direzioni individuano il piano di faglia e quale un piano perpendicolare ad esso detto piano ausiliario (i due piani vengono anche detti piani nodali).

Tale ambiguità può essere risolta osservando le direzioni relative ad altri eventi della stessa area sismogenetica, in particolare le repliche (*aftershocks*) di un episodio sismico principale, in quanto è probabile che gli eventi che si susseguono siano localizzati sullo stesso piano di faglia (Yukutake & Iio, 2017).

Il piano di faglia può inoltre essere rivelato da spaccature superficiali, carte geologiche, fotografie aeree o dati GPS. (McClymont *et al.*, 2008).

Nell'ipotesi di sorgente puntiforme il sistema di forze è rappresentato da un'unica doppia coppia applicata all'ipocentro, ritenuto il punto rappresentativo dell'intera superficie di frattura. Si può definire allora la sfera focale come una sfera di raggio unitario centrata all'ipocentro (o fuoco).

Il piano di faglia e il piano ausiliario tagliano la sfera focale in quattro parti, ognuna delle quali sarà caratterizzata da primi arrivi compressivi o distensivi. Proiettando la semisfera inferiore su un piano si ottengono delle aree che possono essere colorate (solitamente di nero) oppure no a seconda che corrispondano a primi arrivi positivi (fase compressiva) oppure negativa (fase distensiva). Si ottengono in que-

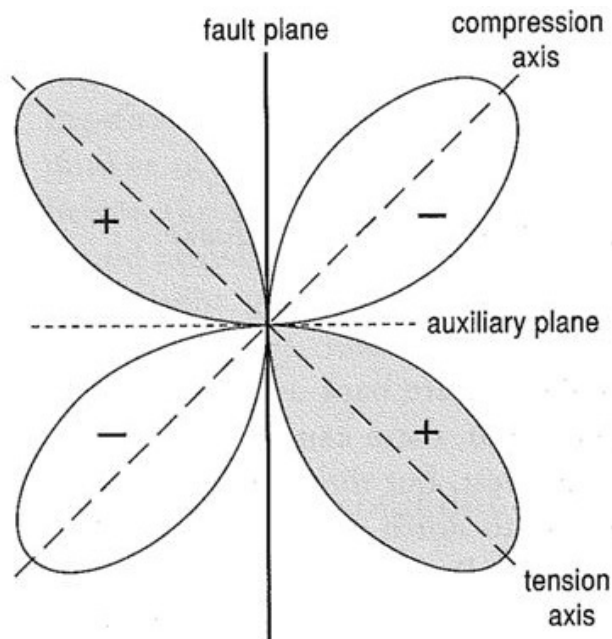


Figura 1.7: Diagramma di radiazione a quattro lobi, con primi movimenti alternativamente di compressione (+) e di dilatazione (-) per una faglia trascorrente destra con orientamento N-S. (Mussett e Aftab Khan, 2003)

sto modo i cosiddetti "palloni da spiaggia" (*beachballs*), ovvero la rappresentazione grafica del meccanismo focale (vedi figura 1.8)

Il meccanismo focale descrive l'orientazione del piano di faglia e la direzione dello spostamento attraverso la polarità (positiva o negativa) prevalente dei primi impulsi in ogni zona.

1.3 Sismologia macrosismica

La *sismologia macrosismica* è quella branca della sismologia che studia i terremoti sulla base degli effetti macroscopici che questi provocano sull'ambiente naturale ed antropico. Essa raccoglie, analizza e sintetizza dati non-strumentali descrittivi di tutto ciò che può mantenere memoria dell'intensità dell'oscillazione sismica, con particolare attenzione alle deformazioni permanenti (danni) di oggetti, edifici, ambiente naturale e agli effetti del fenomeno percepiti dalle persone.

I metodi macrosismici di analisi dei terremoti sono tra i più antichi metodi noti della sismologia e hanno subito trasformazioni nel tempo: le prime descrizioni degli effetti dei terremoti risalgono infatti ad alcuni millenni fa, queste appaiono poi in varie forme (cronache, trattati accademici, corrispondenze private, ...), attraverso i secoli, nei documenti storici di diverse civiltà del passato (Cecic e Musson, 2004; Boschi et al., 1997; Mantyniemi, 2017); relativamente ai terremoti italiani una prima metodicità nella classificazione dei danni in base alla gravità viene riportata nella mappa coeva del terremoto di Nizzardo del 1564 (1.9)(Moroni e Stucchi, 1989) e mappe analoghe ci pervengono da due diversi autori per il terremoto del Gargano del 1627 (Gaudiosi et al., 2014); mentre una catalogazione in 4 categorie, tra terremoti

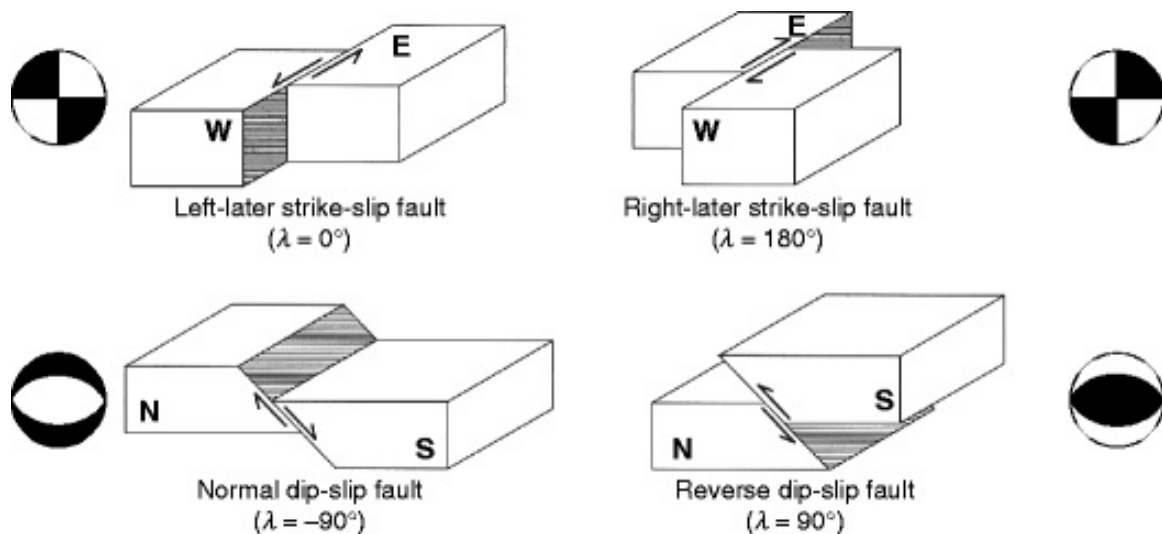


Figura 1.8: Le principali geometrie di faglia con i corrispondenti diagrammi “beach-ball”. Le regioni nere indicano le zone sottoposte a compressione, quelle bianche a distensione. Fonte: *Stein & Wysession, 2003*

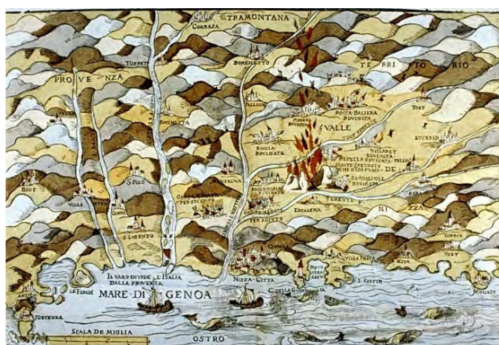


Figura 1.9: Mappa con classificazione dei danni del terremoto del Nizzardo del 1564, attribuibile al mercante genovese Magioli, 1564 (Gaudiosi et al., 2014; Almagia, 1914)

diversi –anch’essa coeva e operata in base alla gravità dei danni– ci perviene per gli eventi della Calabria occorsi tra il 1783 e il 1786 (Gaudiosi et al., 2014). Tuttavia, al di là di questi primi tentativi isolati di organizzazione di dati macrosismici, si è soliti sancire l’inizio della macrosismologia moderna con la pubblicazione della prima scala d’intensità macrosismica internazionalmente riconosciuta e utilizzata, la scala Rossi-Forel (Rossi, 1883). Ci si avvia dunque verso una sistematicità nella trattazione degli effetti. La scala Rossi-Forel infatti stabilisce una classificazione degli effetti di un terremoto in 10 gradi di intensità e costituisce perciò una formale linea di demarcazione tra un approccio allo studio degli effetti puramente descrittivo e occasionalmente classificativo e un approccio invece quantitativo che la macrosismologia perseguirà e potenzierà negli anni seguenti fino ai giorni nostri. Infatti nel XX secolo, sulle orme della scala Rossi-Forel, sono state elaborate le scale macrosismiche che vengono tutt’oggi utilizzate a livello globale, come l’MCS, l’MM, l’EMS e i metodi di ulteriore elaborazione dei dati con esse trattati (Musson e Ceciè, 2011). Le scale macrosismiche vengono approfondite al paragrafo successivo (1.3.1), invece qui si vuole sottolineare

cosa si intende per approccio macrosismico moderno quantitativo, che è naturalmente adottato in questo lavoro di tesi.

L'approccio quantitativo ha le sue fondamenta nella definizione di un parametro di *intensità macrosismica*, valutabile per mezzo delle scale macrosismiche (vedi 1.3.1), che sia concepito come misura del livello di scuotimento del suolo e perciò relazionabile ai parametri fisici strumentali del terremoto come la PGA ¹. Quindi i gradi della scala macrosismica non sono intesi solamente come delle liste in quantità e qualità (severità) degli effetti che potrebbero presentarsi con un terremoto, ma che gli effetti inclusi nei vari gradi siano “diagnostici” di un'informazione fisica e oggettiva sintetizzata nel numero corrispondente al grado stesso, l'*intensità*. Questa interpretazione dell'intensità era già implicita alla pubblicazione di Cancani del 1904 (Cancani, 1904) ed è oggi enormemente suffragata da studi su relazioni empiriche esistenti tra *intensità* e parametri strumentali del moto del suolo; in proposito si citano alcuni studi rappresentativi: Gutenberg e Richter (1942), Trifunac e Brady (1975), Murphy e O'Brien (1977) e Caprio et al. (2015). Gli studi sul parametro d'*intensità* sono stati dunque estesi alla ricerca di relazioni con altri parametri descrittivi della sorgente, i cosiddetti *parametri sintetici dei terremoti* (magnitudo, coordinate epicentrali, etc), i quali, dedotti, potessero essere confrontati con i parametri di sorgente ricavati invece con i metodi della sismologia strumentale.

Quest'approccio quantitativo della macrosismologia che produce risultati confrontabili con i risultati della sismologia strumentale è finalizzato proprio alla *stima dei parametri sintetici*, posta come obiettivo principale delle due discipline e delle branche ad esse afferenti (geodesia, geologia, ...), al fine di caratterizzare la *sismicità* di un territorio² e contribuire al completamento dei *cataloghi sismici*³ (Storchak et al., 2013; Storchak et al., 2015; Di Giacomo et al., 2018; Manchuel et al., 2018; USGS, n.d.; Rovida et al., 2020). Con queste informazioni è infatti possibile costruire modelli che stimano la *pericolosità sismica*⁴ di un territorio (Stucchi et al., 2011; M.-R. INGV, 2004). Per costruire questi modelli, poichè non è al momento possibile prevedere un terremoto in modo deterministico, ci si affida a metodi probabilistici che fanno uso della statistica e della sismicità del passato per le stime dei *tassi di ricorrenza* sui quali si basa la definizione di *pericolosità*. Tuttavia solo in tempi recenti si è avuta la disponibilità di reti sismiche dense e accurate che forniscono una buona copertura per la misurazione dei parametri fisici dei terremoti (Obara et al., 2005; Gee e Leith, 2011; INGV, 2020). Perciò per dare completezza ai *cataloghi sismici* in relazione a quei terremoti, recenti o passati, non adeguatamente registrati dalla rete le uniche informazioni disponibili sono spesso quelle descrittive degli effetti macrosismici sui centri urbani (Bakun e Wentworth, 1997). Infatti ciascun centro abitato rappresenta, da un punto di vista macrosismico, uno strumento di misura degli eventi che lo hanno interessato.

Anche se esula dagli scopi di questa tesi si vuole qui accennare al fatto che il respiro della macrosismologia, e in particolare il suo approccio quantitativo, non

¹Peak Ground Acceleration (PGA): massima accelerazione registrata in un sito da un sismometro all'occorrere di un evento sismico

²Per sismicità si intende il complesso delle caratteristiche di distribuzione spaziale, temporale e in energia dei terremoti verificatisi in una data regione.

³Cataloghi che riportano gli eventi sismici del passato in base ai *parametri sintetici*.

⁴Probabilità che un parametro rappresentativo dello scuotimento del suolo superi una certa soglia in un luogo, in un certo intervallo di tempo.

si limita ad applicazioni inerenti la disciplina stessa; infatti il dato di base di cui principalmente essa si occupa è relativo alle perdite e ai danni ad edifici e persone, quindi si presta ad esempio a studi ingegneristici sulla vulnerabilità⁵ delle strutture antropiche (Fasan et al., 2019; Lagomarsino e Giovinazzi, 2006), a pianificazioni e interventi che riguardano le allerte, la protezione civile e la gestione del territorio (Bossu et al., 2019; Landes et al., 2019; Dolce e Di Bucci, 2014; Zonno e al., 2011; Ceci e Musson, 2004).

1.3.1 Scale d'intensità

Lo strumento principale utilizzato per la quantificazione degli effetti macrosismici di un terremoto ovvero dei danni da esso provocati è costituito dalle *scale di intensità macrosismica*: ciascuna di esse cerca di stabilire criteri oggettivi di valutazione della qualità e della quantità dei danni e degli effetti riscontrati; è organizzata in livelli (o gradi) di *intensità*, ciascuno dei quali riporta una descrizione degli effetti, detti *diagnostici*, che verosimilmente ci si aspetta di osservare per il particolare livello.

I *diagnostici* che vengono considerati nelle scale possono essere suddivisi in base a “chi o cosa risente di essi” e cioè in base agli “oggetti-sensori”. Se ne individuano quattro tipologie (Grunthal, 1998): esseri viventi (persone, animali), oggetti comuni (quadri, stoviglie, mobili), edifici e ambiente naturale (argini, livello d'acqua nei pozzi). All'aumentare dell'*intensità* i primi avvertono lo scuotimento in numero crescente e più violentemente, gli oggetti comuni vibrano e cadono in numero maggiore, i danni agli edifici sono via via più severi, aumentano effetti sul terreno come crepe, fratture e frane. Generalmente si intende ciascun grado della scala come inclusivo degli effetti descritti ai gradi inferiori.

Numerose sono le scale macrosismiche redatte e utilizzate nel mondo; Davison nel 1921 ne contava 27 diverse (Davison, 1921), Musson et al. (2010) ne hanno poi stimato un numero dell'ordine del centinaio. Molte di meno sono quelle effettivamente utilizzate su larga scala. Si riportano qui la scala Mercalli Cancani Sieberg (MCS)(Sieberg, 1923; Sieberg, 1930), la scala Mercalli Modificata (MM)(Wood e Neumann, 1931), la scala Medvedev Sponheuer Kárník (MSK)(Medvedev et al., 1964), l'European Macroseismic Scale (EMS-98)(Grunthal, 1998), la scala ESI 2007 (Guerrieri et al., 2012; Michetti et al., 2007) e la scala JMA (JMA, 1996).

Le prime quattro citate sono legate da concezioni comuni; esse sono in effetti l'esito di reciproche elaborazioni operate da diversi autori. Musson et al. (2010) le ascrivono alla classe delle “Cancani scales”, caratterizzate da 12 gradi di intensità. L'MCS è basilare per le altre e viene pubblicata da Sieberg nel 1923 come sintesi tra la scala Mercalli a 10 gradi (Mercalli, 1902), la proposta di Cancani di aggiungere alla precedente 2 gradi ed espliciti riferimenti alla PGA (Cancani, 1904), e le descrizioni per ciascun grado fornite da Sieberg stesso; essa è ancora oggi molto utilizzata in Italia. Nel 1931 l'MCS viene tradotta in inglese e modificata da Wood e Neuman, i quali la pubblicano con il nome di “Modified Mercalli scale” (MM). L'MM viene successivamente modificata da Richter (C. F. Richter, 1958), ma anche molte altre volte da diversi autori per poter essere estesamente utilizzata nel mondo anglosassone. L'MSK è invece una sintesi dei lavori precedenti e degli studi paralleli di Medvedev

⁵propensione di un bene a subire danneggiamento

ed è stata molto usata in europa orientale. Infine l'EMS consiste in una radicale revisione della MSK ed è attualmente la più diffusa in Europa.

Le diverse modificazioni apportate a queste scale riguardano le specificità di scelta e di dettaglio per la descrizione degli effetti *diagnostici* relativi ai vari gradi. In particolare, per queste scale, i *diagnostici* principali per la discriminazione tra i diversi gradi sono i danni agli edifici. Ma i danni dipendono dalle caratteristiche degli edifici stessi, perciò gli autori hanno via via adattato le descrizioni alle realtà costruttive della regione alla quale la scala doveva essere finalizzata. Già l'MCS tiene conto della vulnerabilità degli edifici; l'MSK introduce chiare classificazioni interne alla scala discriminando gli edifici in 3 tipologie a differente vulnerabilità e il danno in 5 differenti livelli; l'EMS ha a riguardo un dettaglio ancora superiore: 6 tipologie di fabbricati e 5 classi di danneggiamento. Precisazioni di questo tipo permettono una più accurata discriminazione del peso che i diversi livelli di danno osservati su strutture a diversa vulnerabilità hanno per la valutazione di ciascun grado. Infatti a parità di scuotimento ci si attende che una struttura più vulnerabile riporti un livello di danno maggiore rispetto ad una meno vulnerabile.

Infine un'altra importante classificazione interna alle scale è quella sulle quantità e cioè sulle percentuali approssimative di "oggetti-sensori" che reagiscono in un certo modo sul totale osservato (presente nell'MSK e nell'EMS). Questo elemento permette una più accurata discriminazione tra gradi d'*intensità* che includono nella descrizione lo stesso effetto su un "oggetto", che però si presenta con diverse frequenze approssimativamente definite per ciascun grado.

Vale la pena qui accennare alle altre due scale macrosismiche citate prima. La prima è la scala ESI 2007, anch'essa a 12 gradi, ma progettata unicamente per la valutazione dell'*intensità* dai soli effetti sull'ambiente; è perciò priva di riferimenti alle altre tipologie di "oggetti-sensori". Tale separazione degli effetti, attuata in pratica anche nella EMS, è dovuta al fatto che gli effetti sull'ambiente sono più complessi, meno evidenti e meno "grado-specifici" degli altri; essi sono perciò ritenuti inefficaci per una valutazione combinata delle intensità (Grunthal, 1998). La seconda è la scala JMA (Japanese Meteorological Agency) diffusa in Giappone e anch'essa di tradizione centenaria (Omori, 1900). Essa definisce 7 gradi con alcune sotto-suddivisioni degli stessi. Si riporta che attualmente il suo utilizzo è prevalentemente limitato alla conversione dei parametri del moto del suolo in valori di pseudo-intensità, quindi spesso non si ha idea se dato un set di valori d'intensità le relative descrizioni dei gradi della scala abbiano effettive corrispondenze con i danni occorsi sul territorio (Yamazaki et al., 1998; Musson et al., 2010).

1.3.2 L'intensità

L'*intensità* è il parametro principale della sismologia macrosismica. Esso è concepito come indicatore della severità dello *scuotimento del suolo* valutata per un'area limitata sulla base degli effetti osservati sull'ambiente naturale ed antropico (Grunthal, 1998). L'*intensità* è definita in seno alle scale macrosismiche come un parametro ordinale il cui valore è rappresentativo di ciascun grado descrittivo della scala, e permette così sia la classificazione che la quantificazione degli effetti di un terremoto in un sito. Essa assume valori in un range il cui minimo (e.g. 1, I, 0, a seconda della notazione della scala) corrisponde ad un risentimento impercettibile o nullo dello scuotimento

e il cui massimo rappresenta danni catastrofici e distruttivi sul paesaggio (e.g. crollo di edifici, dissesti ambientali).

Nella pratica *l'intensità*, in quanto rappresentativa dello *scuotimento del suolo*, viene trattata spesso come parametro continuo per ricavare relazioni empiriche tra essa e i parametri fisici del moto del suolo come la PGA e la velocità (Gomez-Capera et al., 2020). La ricerca di tali relazioni inizia con la pubblicazione di Cancani del 1904 che assegnava un'intervallo di valori di PGA per ogni grado d'*intensità* (Cancani, 1904). Essa ha poi proseguito con una serie di studi empirici come Gutenberg e Richter (1956) e Trifunac e Brady (1975). L'ipotesi di supporto a questa linea di ricerca è che valga la *legge di Fechner* secondo cui "l'intensità \mathcal{I} di una sensazione umana è approssimativamente proporzionale al logaritmo dell'intensità \mathcal{S} dello stimolo che la ha prodotta": $\mathcal{I} = \kappa \log(\mathcal{S})$ (Fechner, 1860; Weber, 1834). Quindi si può assumere che al primo ordine le relazioni abbiano la forma tipica seguente

$$\log_{10} A = c_1 I - c_2 \quad (1.4)$$

dove A è un parametro del moto del suolo, c_1 e c_2 sono parametri da determinare empiricamente e I è *l'intensità*. Da queste risulta approssimativamente un raddoppio dell'ampiezza del parametro del moto per ogni grado di *intensità* (Kovesligethy, 1906; Gutenberg e Richter, 1956; Trifunac e Brady, 1975; Panza et al., 1997).

Come per i parametri strumentali del moto del suolo anche per *l'intensità macrosismica* si osserva il fenomeno dell'attenuazione: la differenza tra *intensità* locale ed epicentrale diminuisce monotonicamente con la distanza dall'epicentro, con derivata negativa. Sono state ricavate diverse relazioni empiriche che descrivono l'attenuazione dell'*intensità* utilizzando diversi tipi di funzioni matematiche: lineari, bilineari, logaritmiche, dipendenti dalla seconda, terza o quarta potenza (Gupta e Nuttli, 1976; Berardi et al., 1993; Gasperini, 2001). A titolo di esempio si riporta qui il risultato di Gasperini (2001) per una relazione di attenuazione bilineare ricavata per il territorio italiano per distanze dalla sorgente $d < 180Km$ che descrive un'attenuazione di 1 grado ogni $20Km$ fino a $45Km$ di distanza (campo vicino), e di 1 grado ogni $50km$ per distanze superiori (campo lontano).

In generale le relazioni empiriche così ricavate per l'attenuazione dell'*intensità* presentano una dispersione molto ampia e risultano legate alla regionalità dei dati con i quali sono state ricavate. Analogamente si può dire delle leggi di attenuazione empiriche dei parametri strumentali del moto del suolo – si veda Castro et al. (1999) per un compendio di queste per il territorio italiano. Infatti il fenomeno dell'attenuazione delle onde sismiche è dovuto a diversi processi fisici come la dispersione geometrica, la dissipazione anelastica, la riflessione e la rifrazione delle onde sismiche e anche dalle caratteristiche della sorgente (dimensioni, orientazione) (Carletti e Gasperini, 2003). Il bilancio tra questi processi per la descrizione del fenomeno è un problema aperto ed in pratica gli approcci risolutivi sono stati di trascurare tutti o parte dei dettagli dei singoli processi fisici per ricavare relazioni semplici.

In Carletti e Gasperini, 2003 viene mostrato un generale accordo tra l'attenuazione dello scuotimento del suolo ricavata da dati macrosismici e quella ricavata da dati strumentali. Inoltre viene mostrato come con un'elevata densità di dati macrosismici disponibili per un territorio (quello italiano) sia possibile mappare le variazioni locali delle proprietà di propagazione delle onde sismiche. Da questo studio emerge anche una forte correlazione tra l'attenuazione di I e il flusso di calore terrestre su-

perficiale: alle regioni di maggiore attenuazione corrispondono regioni di flusso più elevato. Questa ulteriore evidenza rappresenta una conferma delle basi fisiche dell'interpretazione dei dati di *intensità macrosismica* e giustifica il moderno approccio quantitativo.

Infine per la descrizione dell'attenuazione dell'*intensità* è da tenere in considerazione il rumore dovuto alla dipendenza di I dalle caratteristiche eterogenee di vulnerabilità delle strutture esaminate per la sua valutazione, e la dipendenza della severità dei danni da parametri come il periodo e la durata della sollecitazione sismica, solitamente non esplicitati nelle relazioni empiriche.

1.3.3 L'indagine macrosismica

L'indagine macrosismica si differenzia nelle due operazioni di raccolta dei dati sugli effetti dei terremoti occorsi e, sulla base di questi, di valutazione delle *intensità*. Dalla letteratura possono essere individuate le modalità generali di conduzione delle indagini macrosismiche; tali modalità sono variamente adottate in base alle caratteristiche dei terremoti da studiare e delle risorse umane e tecnologiche che si hanno a disposizione (Ceciè e Musson, 2004).

Per guidare ed omogeneizzare la raccolta dei dati sugli effetti vengono utilizzate come riferimento le scale macrosismiche. In pratica, per un evento, scelta una scala macrosismica, si cerca di raccogliere informazioni relative ai *diagnostici* descritti nella scala stessa e quindi agevolare la successiva operazione di valutazione dell'*intensità*.

Nel caso di eventi sismici contemporanei può essere svolta in maniera diretta un'*indagine di campagna* effettuata da esperti sismologi e ingegneri nei giorni immediatamente successivi ad un evento (Sira, 2015; Rossi et al., 2019); questa modalità di indagine macrosismica consiste nella raccolta di dati tramite interviste dirette ai testimoni dell'evento e nell'osservazione e registrazione dei danni a strutture e ambiente: questo tipo d'indagine è oggi principalmente adottato per terremoti che producono effetti al di sopra della soglia di danno.

In alternativa, sempre per gli eventi contemporanei, possono essere acquisite le risposte ad appositi *questionari macrosismici* distribuiti attraverso diversi canali: telefono, posta ordinaria, posta elettronica, apposite piattaforme online, stampe cartacee diffuse in situ; questi sono appositamente progettati dagli esperti per raccogliere direttamente dalla popolazione dati sugli effetti del terremoto e possono essere compilati dalle autorità locali o da volontari. In questa tesi verranno analizzati dati raccolti con questa modalità, che verrà discussa in dettaglio al capitolo 2.

Invece nel caso di eventi del passato per i quali non è possibile avere esperienza diretta e per i quali le informazioni spesso non erano raccolte per scopi scientifici, quindi non standardizzate mediante l'ausilio delle scale di *intensità* (Vannucci et al., 2015), è necessario effettuare un'operazione di ricerca e analisi delle fonti che descrivono gli effetti. Questa si rivela solitamente tanto più onerosa quanto più l'evento è lontano nel tempo, e può richiedere in certi casi una collaborazione tra sismologi, storici e linguisti. Quest'ultima modalità d'indagine riguarda principalmente gli eventi occorsi in epoca pre-strumentale e rientra nel dominio di studio della *sismologia storica* (Boschi et al. (1997), cit altri articoli). Per questi eventi l'interpretazione delle descrizioni degli effetti riportate nelle fonti è un'operazione molto delicata che ha un ruolo decisivo per l'affidabilità delle successive operazioni di valutazione e quantificazione macrosismica.

In generale, per qualunque modalità di raccolta dati si deve tener conto che le fonti esaminate possono essere o contemporanee all'evento o successive: più sono posticipate rispetto al tempo di occorrenza dell'evento, maggiore è la possibilità che la descrizione di esso sia alterata o confusa.

A seguito della raccolta dei dati sugli effetti macrosismici si prosegue alla loro elaborazione per fornire una mappatura del parametro d'*intensità* per l'intera area in cui è stato avvertito il terremoto. In linea di principio per ciascun ambiente urbano, o comunque antropizzato, che ha risentito degli effetti è possibile assegnare un livello d'*intensità* macrosismica per quell'evento. In sintesi stabilire il grado d'*intensità* per

un luogo significa scegliere una scala macrosismica di riferimento e confrontare i dati di osservazione con le descrizioni dei gradi della scala scelta determinando quale tra queste offre la migliore corrispondenza (Grunthal, 1998). Tale operazione è tutt'altro che banale e nonostante l'oggettività dei *diagnostici* descritti nelle scale macrosismiche, nella pratica è caratterizzata da problematiche e disomogeneità.

Le modalità di *valutazione dell'intensità* sono varie a seconda del tipo di dato che è stato raccolto. In generale può essere fatta una distinzione tra

- *valutazione tradizionale e*
- *valutazione automatica.*

La prima viene effettuata da operatori i quali confrontano manualmente il dataset con le descrizioni della scala ed è essenzialmente soggettiva; la seconda consente di elaborare i dati attraverso opportuni algoritmi, si basa su procedure oggettive e riproducibili ma spesso è incerta l'effettiva corrispondenza con la scala scelta come riferimento.

Per un compendio esaustivo sulle modalità di *valutazione dell'intensità* si rimanda alla letteratura (Musson e Ceciè, 2011; D. J. Wald et al., 2011; Sbarra et al., 2010; Tosi et al., 2015), mentre al capitolo successivo (par:2.2) si descriverà la procedura di valutazione in relazione al solo tipo di dati utilizzati per questa tesi. Qui si vogliono invece discutere le caratteristiche generali, le problematiche e le accortezze pratiche comuni a tutte le modalità di valutazione.

In primo luogo si vuole mettere in evidenza il carattere intrinsecamente soggettivo dell'informazione macrosismica. Esso emerge ad esempio nelle forti inconsistenze nei risultati delle valutazioni effettuate da operatori differenti (Musson e Ceciè, 2011). Infatti spesso accade che nei rapporti macrosismici si riscontri una sovrapposizione di *diagnostici* rappresentativi di due o tre gradi differenti, oppure che si presentino nell'interpretazione dello stesso rapporto macrosismico descrittivo contraddizioni o ambiguità per un singolo diagnostico (può verificarsi che di due costruzioni "identiche" affiancate una subisca gravi danni mentre l'altra rimanga pressoché intatta). Dai diagnostici che un operatore ritiene più affidabili e rappresentativi dipende la valutazione finale (Ferrari e Guidoboni, 2000). L'elemento di soggettività può essere ridotto sia dall'inclusione, nelle definizioni della scala, di *diagnostici* che siano il più possibile grado-specifici sia dalla redazione di linee guida per l'interpretazione dei *diagnostici* stessi. Tuttavia la relazione effetti-gradi non sembra essere stata ancora studiata in maniera esaustiva (Vannucci et al., 2015), mentre solo di recente sono state introdotte linee guida per la valutazione (Grunthal, 1998; Lee et al., 2002; Borman, 2002).

In secondo luogo bisogna considerare la natura statistica del parametro di *intensità macrosismica*: esso deve essere valutato sulla base delle occorrenze degli effetti sull'intero campione di "oggetti" osservato. Ciò significa che in un qualsiasi luogo ci si aspetta che un particolare effetto venga osservato solamente in un dato numero di casi, la cui entità è rappresentativa della severità dello scuotimento (e.g. scala MCS, intensità V: "sporadici oggetti si rovesciano", intensità VI: "molti oggetti si rovesciano"); ci si aspetta dunque che si manifestino effetti più severi e meno severi del livello modale di *intensità* atteso. Se ad esempio dai dati per un luogo si evince $I = 4$ – al di sotto della soglia di danno – e si verifica la caduta di un camino, questo

non invalida la valutazione. Inoltre non è detto che il campione di effetti osservato sia statisticamente rappresentativo, in particolar modo questo può essere vero nel caso in cui i dati sono raccolti da un campione di volontari autoselezionante. L'idea è che l'assenza di un effetto nei dati macrosismici non assicura che questo non si sia verificato: l'informazione non è completa.

Perciò, ad esempio, è importante che nella fase di sintesi dei dati grezzi delle rilevazioni macrosismiche: 1) si sia cauti nel peso da attribuire agli *outliers*⁶ – sporadiche osservazioni estreme o effetti anomali isolati; 2) si consideri, per l'attribuzione dell'*intensità* ad un luogo, un'area di osservazione dei diagnostici sufficientemente grande per cercare di avere un buon campione statistico, ma tale da non perdere importanti variazioni locali dell'entità dello scuotimento.

A causa di queste problematiche sono frequenti i casi di ambiguità nell'attribuzione del grado (Boschi et al., 1997; Camassi e Stucchi, 1997; Stucchi et al., 2007; Musson e Ceciè, 2011); per questi solitamente viene riportata o l'ambiguità grezza (e.g. V-VI) oppure si attribuisce un grado di valore intermedio (e.g. 5.5). In questi casi un'interpretazione è che il grado inferiore è certamente occorso, ma che, nonostante alcuni elementi tipici del grado superiore, questo non possa essere asserito con certezza (Musson e Ceciè, 2011). È inoltre necessario tener conto che l'attribuzione di un grado intermedio implicherebbe di fatto la capacità di poter discernere tra un numero doppio di gradi rispetto a quelli della scala ordinaria, quindi di migliorarne inverosimilmente la risoluzione (Musson e Ceciè, 2011; Vannucci et al., 2015). Potrebbe non essere una coincidenza che le scale moderne convergano di fatto ad una discretizzazione a 10 gradi.

Infine sono le stesse caratteristiche degli eventi sismici che possono inficiare una corretta valutazione dell'intensità. Ad esempio le repliche che in breve tempo seguono una scossa principale possono essere tali da modificare e sporcare il dato macrosismico di un luogo: gli effetti cumulativi del danno tendono a rappresentare un'intensità sovrastimata; solo una rilevazione tempestiva, non sempre disponibile, precedente alla prima replica, può fornire informazioni esclusive sulla prima scossa.

Come prodotto finale della valutazione delle intensità per i luoghi che hanno risentito di un dato sisma si ottiene una mappa nota come *piano quotato delle intensità* che fornisce un immediato riscontro grafico della distribuzione spaziale dei valori. Talvolta per avere una migliore visualizzazione del risentimento complessivo del territorio interessato dall'evento o per effettuare successive analisi vengono tracciate le *isosieme*, curve di livello di intensità (Figura 1.10). Queste curve possono essere tracciate o manualmente da operatori esperti o mediante l'utilizzo di software che effettuano un'interpolazione dei dati in base ad alcuni parametri, per i quali tuttavia non esiste uno standard (Postpischl et al., 1985; Topozada et al., 1981; Bakun e Wentworth, 1997).

⁶per *outliers* si intendono quei record anomali che si discostano talmente tanto dalle altre osservazioni da far pensare che siano stati generati da un processo sostanzialmente diverso da quello generale. Per le caratteristiche specifiche degli *outliers* nell'ambito dei felt reports si rimanda al paragrafo 2.2

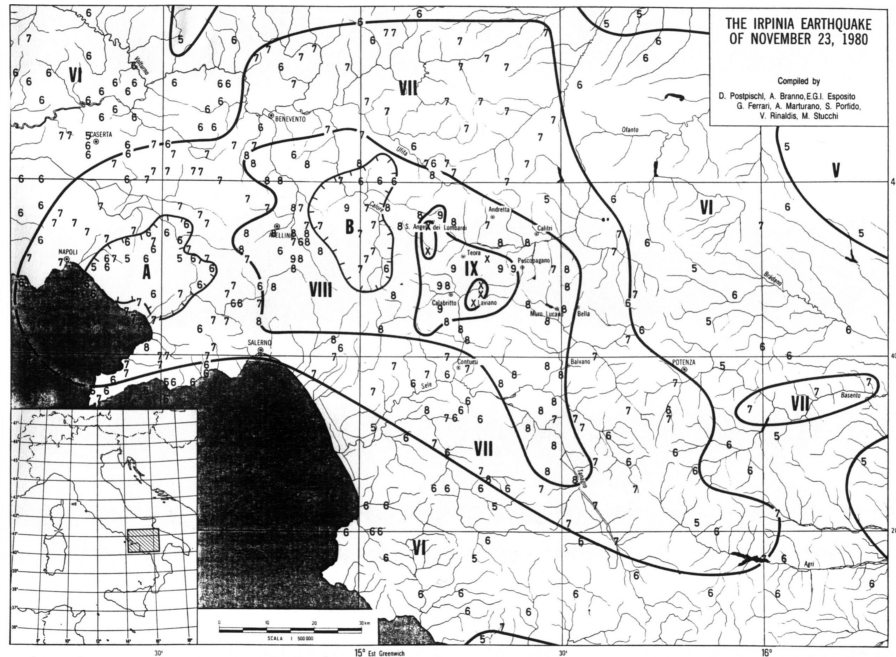


Figura 1.10: Esempio di *piano quotato delle intensità* e isosisme del terremoto dell'Irpinia 23/11/1981.

1.3.4 Approccio macrosismico alla stima dei parametri sintetici

I dati strumentali sono inevitabilmente più oggettivi rispetto ai dati macrosismici nonché indispensabili per determinare in modo attendibile i parametri della sorgente (epicentro, ipocentro, magnitudo, geometria della sorgente) tuttavia risulta sempre più frequente l'utilizzo di osservazioni macrosismiche al fine di stimare parametri e caratteristiche geofisiche delle sorgenti sismogenetiche.

Uno degli scopi è quello di risolvere il cosiddetto "problema inverso", ovvero di ricavare le caratteristiche della sorgente sulla base dei soli risentimenti superficiali. Tale operazione risulta molto complessa in quanto le variabili di interesse sono numerose e spesso di difficile interpretazione; in particolare parametri come l'angolo di immersione e la direzione di scorrimento della faglia non hanno una corrispondenza univoca con l'intensità osservata in superficie.

Numerosi studi (*Sirovich & Chiaruttini, 1993; Panza et al., 1997; Sirovich et al., 2001*) si basano sulla comparazione tra dati macrosismici e dati strumentali (disponibili per terremoti contemporanei) al fine di costruire modelli in grado di essere applicati al problema della determinazione dei parametri di sorgente del passato.

In generale, è largamente accettato che l'intensità macrosismica dipenda in maniera significativa dallo scuotimento del terreno. In particolare l'ampiezza, lo spettro della radiazione e la durata sembrano giocare un ruolo importante (*Sokolov & Chernov, 1998; Sokolov, 2002*), tuttavia non si dispone attualmente di un modello fisico capace di dare conto quantitativamente di queste relazioni. Pertanto, in assenza di un tale modello, il confronto fra queste due classi di grandezze (intensità macrosismica da un lato e parametri strumentali dall'altro) può essere effettuato solo mediante metodologie empiriche e di trattamento statistico dei dati macrosismici (*Gasperini*

& Ferrari, 1995; 1997; Gasperini et al., 1999; Amini et al., 2017). Tali metodologie verranno approfondite in seguito.

Data e tempo origine

L'esatta determinazione della data e del tempo origine di un evento sismico in ambito macrosismico è di competenza della sismologia storica ed interessa principalmente i terremoti di epoca prestrumentale. Se ne accenna qui per completezza.

La questione presenta molte problematiche legate alla non uniformità dei riferimenti per la datazione nelle varie epoche storiche e nelle diverse aree geografiche.

Ad esempio prima del XVII secolo nel solo territorio italiano sono stati usati 7 diversi stili di datazione, necessariamente da convertire allo standard d'uso corrente per operare confronti e successive analisi (Gasperini & Ferrari, 2000).

Per quanto riguarda il tempo di accadimento dell'evento sismico all'interno della giornata è possibile sfruttare indicazioni legate alle attività umane dell'epoca, al moto apparente del Sole rispetto alla Terra o alle ore suonate dal campanile.

Per la natura del problema di ricostruzione storica è evidente che il grado di incertezza può essere elevato. Va tuttavia osservato che i parametri di tempo non sono necessari per le successive indagini. Infatti se da un lato i terremoti che verranno analizzati in questo studio sono tutti recenti e con disponibilità di accurati dati strumentali di data e tempo origine, dall'altro le stime macrosismiche dei parametri sintetici non utilizzano parametri temporali, che risultano perciò completamente irrilevanti.

Epicentro

Mentre la definizione di epicentro ottenuto mediante dati strumentali è largamente condivisa nella letteratura sismologica attuale non si può dire lo stesso per quanto riguarda l'epicentro macrosismico. Ciò è dovuto alla carenza di standardizzazione presente negli studi macrosismici.

L'epicentro macrosismico può essere interpretato come la miglior stima della posizione dell'epicentro senza fare uso di dati strumentali (Musson & Cčić, 2002), tuttavia oggi è largamente accettato che l'epicentro macrosismico possa essere considerevolmente distante dall'epicentro strumentale (Gasperini & Ferrari, 2000).

Ciò è dovuto al fatto che, mentre l'epicentro strumentale corrisponde alla proiezione dell'ipocentro sulla superficie terrestre, l'epicentro macrosismico rappresenta il centro della distribuzione degli effetti del terremoto e può essere più verosimilmente messo in relazione con la proiezione sulla superficie terrestre del baricentro della regione di frattura.

Nella realizzazione del Catalogo Parametrico dei Terremoti Italiani (CPTI15), dove l'epicentro macrosismico è stato determinato per 3005 terremoti, la localizzazione dello stesso è stata effettuata in diversi modi:

- Per la maggior parte dei terremoti l'epicentro è stato calcolato come il baricentro dei punti con la massima intensità (Gasperini et al., 1999; Gasperini et al., 2010); tale metodo è risultato stabile anche in presenza di distribuzioni di dati macrosismici non ottimali;

- Per alcuni terremoti con epicentro verosimilmente in mare o prossimo alla costa è stata in caso eccezionale preferita la localizzazione dell'epicentro in termini di latitudine, longitudine, profondità e intensità attesa secondo la relazione di Pasolini *et al.* (2008), minimizzando la somma dei quadrati dei residui (per i dettagli si veda Gasperini *et al.*, 2010); metodi analoghi erano già stati proposti in precedenza da Peruzza (1992) e Bakun e Wentworth (1997);
- Per alcuni terremoti dell'area etnea, anche in presenza di dati relativi all'intensità, si è preferita una localizzazione basata sugli effetti di fagliazione cosismica superficiale (Azzaro *et al.*, 2000).

Laddove la localizzazione effettuata mediante uno dei metodi precedentemente descritti è apparsa poco o per nulla coerente con la distribuzione delle intensità, si è modificato “manualmente” l'epicentro, facendolo in genere coincidere con il punto di massima intensità oppure non considerando uno o più punti di intensità elevata presumibilmente anomali (*Gruppo di Lavoro CPTI*, 2016).

Ulteriori elementi che possono essere presi in considerazione nella stima dell'epicentro macrosismico sono la forma delle isosisme, la segnalazione di eventuali *foreshock* (le cosiddette “scosse premonitrici”) e *aftershock* (o repliche), la conoscenza della geologia locale ecc. (*Musson & Cčić*, 2002).

è opportuno osservare infine che parte del processo di valutazione può risentire di una certa soggettività e che non esiste una linea guida che possa essere applicata ugualmente in tutti i casi.

Profondità

I metodi per la stima della profondità ipocentrale mediante dati macrosismici fanno quasi tutti riferimento a quello proposto da Radó von Kövesligethy, il quale, nella sua opera in latino “*Seismonomia*” pubblicata nel 1906 deduce una relazione che sarà alla base di tutte le ricerche future della profondità ipocentrale in funzione dell'intensità (*Marcelli & Spadea*, 1981).

Il modello può essere descritto dalla formula:

$$I - I_0 = \log \left(\frac{r}{h} \right) + 3\alpha \log \left[\frac{e(r-h)}{R_T} \right] \quad (1.5)$$

dove I_0 è l'intensità del sisma all'epicentro, I l'intensità in un punto A posto ad una distanza r dall'epicentro (ovvero il raggio dell'isosisma di intensità I), h la profondità, e il numero di Eulero, α un coefficiente di assorbimento anelastico da determinare sperimentalmente e R_T il raggio terrestre.

Successivamente molti altri si sono serviti della stessa formula introducendo alcune modifiche non essenziali (*Grablovitz*, 1910; *Blake*, 1941; *Marcelli & Peronaci*, 1979; *Burton et al.*, 1985)

Il modello di Kövesligethy è usato tuttora nella stima dei parametri macrosismici dei terremoti: è infatti implementato nel software Meep sviluppato da Musson e Jimenez (*Musson & Jimenez*, 2008).

Magnitudo

La stima della magnitudo a partire da dati macrosismici (magnitudo macrosismica o magnitudo equivalente) viene effettuata in genere mediante due possibili approcci: utilizzando l'intensità epicentrale oppure la *felt area* (area di percezione), ovvero l'area della regione in cui il terremoto è stato avvertito.

Metodi basati sull'intensità epicentrale A causa della scarsità di dati disponibili, in passato era d'uso valutare la magnitudo di eventi sismici prestrumentali mediante relazioni empiriche tra la magnitudo e l'intensità epicentrale (*Karnik, 1969; Topozada, 1975*).

Al netto dell'incertezza intrinseca nella stima del parametro di intensità, è evidente che la profondità ipocentrale gioca un ruolo cruciale. Infatti, a parità di energia rilasciata, maggiore è la profondità della sorgente minori sono gli effetti sulla superficie. Ne consegue che, per terremoti relativamente profondi, la magnitudo calcolata esclusivamente mediante l'intensità epicentrale verrà sottostimata, mentre per terremoti relativamente superficiali verrà sovrastimata.

Basili *et al.* (1980) hanno ipotizzato una dipendenza lineare tra magnitudo M e intensità epicentrale I_0 e una dipendenza logaritmica tra magnitudo e profondità h esprimibili tramite una relazione del tipo

$$M = aI_0 + b \log h + c \quad (1.6)$$

con a , b e c coefficienti numerici da determinare sperimentalmente e che sono stati determinati per terremoti con epicentro in area italiana.

Tinti *et al.* (1986) hanno utilizzato metodi analoghi suddividendo l'area di interesse in aree più piccole da analizzare separatamente in base a criteri sismotettonici o statistici.

Un approccio alternativo all'uso di equazioni del tipo 1.6, è quello di trattare ogni valore separatamente e analizzare la distribuzione della magnitudo per terremoti con la stessa intensità massima (*Cavallini & Rebez, 1996*).

Szeliga *et al.* (2010) hanno utilizzato una relazione che tiene conto dell'attenuazione e lega l'intensità alla magnitudo momento e alla distanza ipocentrale secondo la forma

$$I = a + bM_w + cR + d \log R \quad (1.7)$$

dove I è l'intensità, M_w la magnitudo momento, R la distanza ipocentrale e a , b , c e d sono costanti determinate per terremoti dell'area indiana.

Metodi basati sulla *felt area* La *felt area* può essere definita come l'area della regione in cui il terremoto è stato effettivamente percepito dalla popolazione. Si tratta di un parametro che può dipendere anche dall'ora del giorno in cui è avvenuto il terremoto e da altri fattori che favoriscono o avversano la percezione. Tuttavia l'incertezza nella stima dovuta a fenomeni legati all'anisotropia, ad effetti di amplificazione locali o alla geografia antropica possono essere ridotti significativamente. (*Gasperini & Ferrari, 2000*)

Nuttli e Zollweg (1974) hanno ricavato una relazione empirica che lega la magnitudo delle onde di volume alla *felt area* A_f avente la seguente forma

$$m_b = a + b \log(A_f) + c \log^2(A_f) \quad (1.8)$$

con A_f misurata in chilometri quadrati e le costanti a , b e c riferite agli Stati Uniti centrali.

Per quanto riguarda il territorio italiano si citano gli studi di Westaway (1992) e di Termini *et al.* (2003b).

Metodi misti Galanopoulos (1961) ha proposto un metodo che combina i due approcci precedentemente descritti. La relazione usata da Galanopoulos ha la forma

$$M = a \log(I_0 A_f) + b \quad (1.9)$$

in cui i parametri a e b sono stati determinati per terremoti dell'area greca e californiana.

Studi successivi (*Sibol et al.*, 1987; *Gasperini & Ferrari*, 1997, 2000) sono stati eseguiti utilizzando formule del tipo

$$M = a + bI_0^2 + c \log^2 A_f \quad (1.10)$$

dove parametri a , b e c sono stati determinati da Gasperini e Ferrari mediante regressioni lineari con uno schema di peso, funzione dell'affidabilità dei dati di base.

Parametri di faglia

Gasperini *et al.* (1999) assumono che ogni terremoto storico analizzato nel loro studio rappresenti l'evento di intensità massima possibile per la rispettiva sorgente. Questo permette di considerare ogni terremoto come il risultato dell'azione di una specifica sorgente le cui caratteristiche possono essere utilizzate per costruire un modello di faglia o integrarne uno esistente.

Tuttavia sulla base dei dati macrosismici è possibile effettuare stime del solo parametro di *orientazione* dello strike della faglia e delle *dimensioni* della stessa.

Queste sono basate sull'assunzione che in campo vicino, a determinare la distribuzione delle *intensità* più elevate sia più la dimensione della faglia che il pattern di radiazione della sorgente. Si osserva infatti che le isosisme delle *intensità* più elevate tendono ad approssimare la geometria della proiezione della faglia sorgente sulla superficie terrestre come mostrato in Figura 1.11.

In particolare per le *dimensioni della faglia* vale l'ipotesi dell'esistenza di una relazione tra esse e il momento sismico, quindi con la relativa *magnitudo momento* 1.1 1.3. Un modo semplice e diffuso di descrivere una sorgente sismica estesa è quello di rappresentarla come un rettangolo orientato (*Haskell*, 1964) le cui dimensioni (lunghezza e larghezza) possono essere ottenute mediante formule empiriche come descritto da Wells e Coppersmith (1994).

Una volta localizzata la sorgente e valutate le dimensioni finite è possibile stimarne l'orientazione mediante un algoritmo elaborato da Gasperini *et al.* (1999).

Sia le relazioni per le *dimensioni* che quelle per l'*orientazione* della faglia sono presentate in seguito, nei paragrafi relativi al programma Boxer 1.4.3.

1.3.5 Cataloghi parametrici dei terremoti

Sin dai tempi più remoti l'incidenza talvolta distruttiva di un sisma su una comunità ha fatto sì che venissero registrate testimonianze e prodotti documenti tramandati tuttavia in modo spesso frammentario ed eterogeneo.

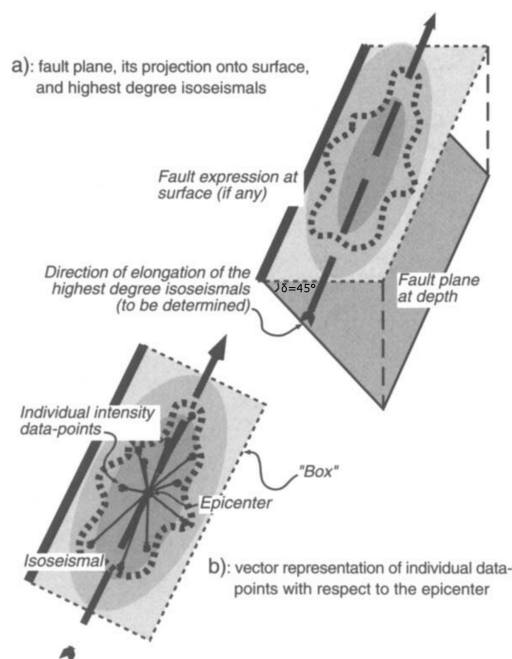


Figura 1.11: Geometry of the problem. (a) Fault plane dipping 45° , its surface projection, and isoseismals. (b) Vector representation of Mercalli intensity data points with respect to the macroseismic epicenter (i.e., the mid-point of the data distribution).

Il primo a realizzare un vero e proprio catalogo fu l'ingegnere irlandese Robert Mallet (1810-1881) il quale, verso la metà dell'Ottocento compilò il più grande elenco di terremoti dell'epoca corredato da tutte le informazioni disponibili.

La prima opera a prendere in considerazione l'intero territorio italiano è il catalogo presentato da Mario Baratta nel 1901 (*Baratta, 1901*).

Nei decenni successivi sono comparsi nuovi cataloghi in Europa, in particolare in Italia (*Cavasino, 1935; Caloi, 1942*). A partire dagli anni Settanta le nuove tecnologie hanno accelerato notevolmente la raccolta e la catalogazione di informazioni riguardanti i terremoti. Tra i cataloghi di questo periodo si segnala, per l'Italia, quello realizzato da Carozzo *et al.* (1973). Per maggiori dettagli a riguardo si rimanda a Molin *et al.* (2008).

Nel 1995 è stato realizzato il "Catalogo dei Forti Terremoti in Italia" (*Boschi et al., 1995*), un'opera dinamica e passibile di qualsiasi aggiunta futura, come effettivamente avvenuto nel 1997 (*Boschi et al., 1997*).

Nel 1999, lo sforzo congiunto di vari gruppi ed enti di ricerca nel settore della sismologia (ING, GNDT, SGA)⁷, ha sintetizzato il frutto delle conoscenze acquisite in vari anni di studio e ricerca in un "Catalogo Parametrico dei Terremoti Italiani" (CPTI) (*Gruppo di Lavoro CPTI, 1999*). Tale catalogo è stato profondamente aggiornato nel 2016 e denominato CPTI15, precedentemente citato. Una descrizione di tale catalogo, aggiornato al 2021, è consultabile a questo indirizzo web: <https://emidius.mi.ingv.it/CPTI15-DBMI15>. Esso riporta dati e parametri raccolti e stimati sia con i metodi della sismologia strumentale sia con i metodi della sismolo-

⁷Istituto Nazionale di Geofisica, Gruppo Nazionale per la Difesa dai Terremoti, Storia Geofisica Ambiente.

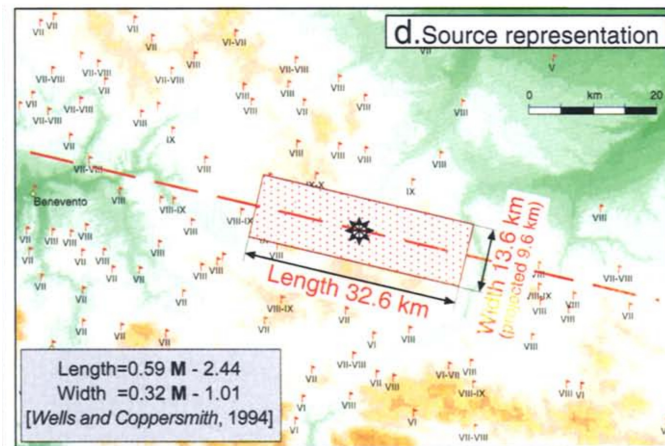


Figura 1.12: Rappresentazione della sorgente, dell'epicentro macrosismico e del piano quotato dell'intensità del terremoto dell'Irpinia del 23 Luglio 1930 con i metodi di Boxer (Gasperini et al., 1999): dimensioni calcolate con le relazioni empiriche di Wells e Coppersmith (1994), orientazione ed epicentro calcolati con i metodi di Gasperini et al. (1999) (vedi paragrafi successivi).

gia macrosismica e ne indica inoltre un'ordine di affidabilità terremoto-specifico non sempre a favore della sismologia strumentale.

Per concludere è opportuno sottolineare come l'Italia sia il paese con il più grande registro di dati macrosismici al mondo.

1.4 BOXER: stima dei parametri sintetici mediante programma di elaborazione di dati macrosismici

Boxer è un programma di elaborazione di dati macrosismici sviluppato in linguaggio Fortran⁸ (Gasperini et al., 1999; Gasperini et al., 2010). Esso consente di stimare, a partire dalla distribuzione geografica delle *intensità macrosismiche*, i *parametri sintetici* di un terremoto quali la *localizzazione epicentrale*, la *magnitudo* e le *dimensioni fisiche della sorgente* sismogenetica attraverso relazioni empiriche e, in aggiunta, permette di stimare l'*orientazione della faglia sorgente*. La ricerca e lo sviluppo degli algoritmi inclusi in Boxer si fondano sull'ipotesi che esista una connessione fisica tra la sorgente sismica in profondità e la distribuzione spaziale dello scuotimento alla superficie terrestre quantificata mediante il parametro di *intensità*. In particolare si assume che da tale distribuzione si possano disegnare i margini della proiezione della faglia sulla superficie terrestre, la quale viene rappresentata come una "box" rettangolare (Figura 1.12) – da cui il programma prende il nome.

La prima versione del codice Boxer nasce nel 1999 come aggregazione, riorganizzazione e standardizzazione di diversi codici sviluppati dal 1993 ad allora e utilizzati per calcolare i *parametri sintetici* dei terremoti ai fini della compilazione del *Catalogo*

⁸<https://emidius.mi.ingv.it/boxer/>

dei Forti Terremoti in Italia dal 461 a.C. al 1990 (Boschi et al., 1997), del *Catalogo Parametrico dei Terremoti Italiani* (CPTI99, 1999) e della pubblicazione di Gasperini et al. (1999) in cui sono stati descritti in dettaglio i metodi implementati nel codice (BOXER, 2010). Da allora il programma ha subito modifiche e migliorie, consistenti nell'aggiunta di metodi alternativi per la stima dei parametri e delle relative incertezze, in nuove procedure per la calibrazione delle relazioni empiriche già implementate (Gasperini et al., 2010), nello sviluppo di funzionalità per il plotting e la visualizzazione dei risultati con MapInfo (GIS) e GMT e nello sviluppo di un'interfaccia da cui l'utente potesse impostare i parametri di input e riprocessare i risultati ottenuti (Triponone et al., 2010). Gli aggiornamenti continui hanno fatto di Boxer uno strumento centrale per la compilazione delle successive versioni dei cataloghi CFTI⁹ e CPTI¹⁰, i quali costituiscono un riferimento per tutte le elaborazioni di pericolosità sismica a scala nazionale; inoltre le stime effettuate con questo metodo hanno contribuito alla definizione del database delle singole strutture sismogenetiche (DISS)¹¹. Per aver contezza dell'importanza del contributo si consideri che Boxer è stato usato per circa il 90% delle stime dei parametri sintetici da dati macrosismici contenute nel CPTI15, e inoltre che, in questo catalogo, i terremoti per i quali sono disponibili parametri sintetici da stime macrosismiche sono circa il 60% del totale, mentre circa il 40% riporta parametri sintetici da stime strumentali (Rovida et al., 2019). L'applicazione di Boxer è stata fondamentale per standardizzare le procedure di stima dei parametri sintetici non solo per i terremoti storici che sono privi di dati strumentali, ma anche per i terremoti recenti per i quali la copertura strumentale non è stata sufficiente a fornire stime affidabili, come nel caso di circa la metà (34) degli eventi del CPTI15 per i quali l'epicentro è verosimilmente in mare o sulla costa.

In Boxer sono implementati diversi *metodi per il calcolo dei parametri*, che sono di seguito riportati secondo la denominazione di Gasperini et al. (2010).

Per la *localizzazione* ve ne sono 7 differenti: il “metodo 0”, o “del baricentro”, è stato sviluppato da Gasperini et al. (1999) e stima l'*epicentro macrosismico*; i “metodi 1-6” stimano *epicentro* e *profondità* utilizzando una stessa procedura, sviluppata in 6 diverse varianti da Gasperini et al. (2010) sulla base degli studi di Pasolini et al. (2008); alle stime epicentrali del “metodo 0” è associata una stima di *profondità* secondo la procedura di R. M. W. Musson (1996).

Per la stima di *magnitudo* sono incluse in Boxer 2 procedure: l' “old method”, già descritto in Gasperini et al. (1999) e in Gasperini e Ferrari (1995) e Gasperini e Ferrari (2000), segue l'approccio di Sibol et al. (1987) con qualche modificazione; l' “alternative method” (Gasperini et al., 2010) segue invece l'approccio di Bakun e Wentworth (1997).

Le *dimensioni della faglia* vengono stimate mediante le relazioni empiriche di Wells e Coppersmith (1994).

La procedura di stima dell'*orientazione della faglia* è invece descritta già in Gasperini et al. (1999), poi affinata e meglio discussa in Gasperini et al. (2010).

Le *incertezze* sui parametri sono valutate 1) con metodi formali tramite il calcolo di una matrice di varianza/covarianza calcolata invertendo l'Hessiana alle differenze finite di una funzione di log-likelihood al suo massimo (Guo e Ogata, 1997) o calco-

⁹<https://emidius.mi.ingv.it/>

¹⁰<http://storing.ingv.it/cfti/cfti5/>

¹¹<http://diss.rm.ingv.it/diss/>

lando le deviazioni standard delle medie dei parametri, 2) con tecniche di simulazione bootstrap (Efron e Tibshirani, 1986; Hall, 1992).

Per l'utilizzo di Boxer su dati di terremoti al di fuori dell'area italiana è necessaria la *calibrazione* dei metodi e cioè la calibrazione dei coefficienti delle relazioni empiriche utilizzate che sono di default impostati per l'Italia. Mentre per il “metodo 0” di localizzazione e il metodo di stima dell'orientazione la *calibrazione* non è necessaria, nel programma è inclusa una procedura per calibrare i coefficienti della relazione di regressione usata nell' “old magnitude method”, scelto un set di terremoti con PQ noto e magnitudo nota. Gasperini et al. (2010) suggeriscono come regola generale di utilizzare almeno 20-30 terremoti con diverse intensità epicentrali I_E per avere una calibrazione affidabile. In alternativa, in caso di zone a debole sismicità sprovviste di un simile dataset l'unica possibilità è una *calibrazione a posteriori* tramite una regressione empirica con i dati strumentali, utilizzando per le stime macrosismiche i coefficienti di default per l'Italia.

Invece per i nuovi metodi di localizzazione (Gasperini et al., 2010) si proponevano di includere in Boxer la procedura di calibrazione dell'equazione di attenuazione $M - I_E$ sviluppata da Pasolini et al. (2008) e ad essi necessaria, la quale può comunque essere effettuata con codici indipendenti da Boxer.

1.4.1 Localizzazione macrosismica

Per la localizzazione macrosismica sono implementati in Boxer due differenti metodi. Il primo fornisce solamente stime delle coordinate epicentrali, mentre con il secondo, che è implementato in 6 varianti, si ottengono anche stime della profondità.

Il primo è detto “metodo del baricentro” (metodo 0) ed è sviluppato in accordo con la definizione di epicentro macrosismico più diffusamente accettata, almeno per terremoti al di sopra della soglia di danno e cioè con magnitudo circa $M > 5.5$ (Gasperini et al., 1999; Gasperini et al., 2010; Gasperini e Ferrari, 2000; Kondorskaya e Shebalin, 1982; Postpischl et al., 1985; Camassi e Stucchi, 1997): «baricentro dell'area caratterizzata dai massimi effetti».

In particolare per un terremoto il baricentro viene calcolato applicando la “trimmed mean” (TM)¹² alle coordinate degli MDPs che riportano il valore di *intensità* più elevato I_{\max} , includendo tuttavia nel calcolo le coordinate dei punti della classe immediatamente inferiore ($I_{\max} - K$ con $K = 1$) nel caso in cui si abbiano meno di $N = 3$ punti con I_{\max} . Lo stimatore statistico TM, e i parametri $K = 1$ e $N = 3$ sono stati scelti a seguito di un'analisi comparata tra diverse combinazioni di stimatori e valori dei parametri (Gasperini e Ferrari, 2000).

Per ragioni di confronto, agli epicentri calcolati con questo metodo sono associate stime di profondità effettuate calcolando soluzioni particolari con il metodo di R. M. W. Musson (1996) (Gasperini et al., 2010, pp. 2039-2040)

Il secondo metodo è implementato in 6 differenti varianti (“metodi 1-6”). È basato sulla relazione di attenuazione dell'*intensità* ricavata da Pasolini et al. (2008) per il territorio italiano:

$$I_i = I_E - a(D_i - h) - b(\ln D_i - \ln h) \quad (1.11)$$

¹²trimmed mean: media aritmetica dei soli dati inclusi tra il 20 -esimo e l'80 -esimo percentile

l'intensità dell'*i* –esimo sito del PQ può essere stimata da una relazione log-lineare dipendente dai parametri h , profondità media indicativa dei terremoti della regione oggetto d'analisi, a e b , coefficienti empirici di attenuazione, $D_i = \sqrt{R_i^2 + h^2}$, distanza ipocentrale dell'*i* –esimo sito con $R_i = R_i(lat_E, lon_E, lat_i, lon_i)$ distanza epicentrale calcolata sulla geodetica di una Terra sferica di raggio R_0 , I_E , intensità attesa all'e-picentro macrosismico.

Il problema della localizzazione è dunque posto nei termini della ricerca dei parametri a , b , I_E , lat_E , lon_E , h tali che sia minimizzata la somma dei quadrati dei residui di ogni sito:

$$SSR(lat_E, lon_E, h, I_E, a, b) = \sum_{j=1}^N \{I_j - I_E + a(D_j - h) + b(\ln D_j - \ln h)\}^2. \quad (1.12)$$

Il problema è affrontato con un approccio probabilistico, anche per trattare gli MDPs con *intensità* incerta (e.g. V-VI), come suggerito da Pasolini et al. (2008): si assegna, per un sito, una distribuzione di probabilità che assegna dei pesi, a somma unitaria, ai gradi della scala macrosismica (e.g. $p(I) = [0, 0, 0, 0, 0.5, 0.5, 0, 0, 0, 0, 0]$). Coerentemente si assume che i residui di *intensità* del campo macrosismico seguano una distribuzione quasi-normale (e.g. Gasperini (2001)), la cui funzione di log-likelihood per un dataset di N MDPs rispetto ad un predittore μ_j con varianza σ può essere scritta come

$$L = \sum_{j=1}^N \ln \left\{ \frac{1}{\sigma\sqrt{2\pi}} \sum_{I=1}^{12} p_j(I) \int_{I-0.5}^{I+0.5} \exp \left[-\frac{(i - \mu_j)^2}{2\sigma^2} \right] di \right\} \quad (1.13)$$

il cui predittore μ_j può essere scritto, esplicitando la dipendenza dai parametri, come

$$\mu_j(lat_E, lon_E, h, I_E, a, b) = I_E - a\{D_j(lat_E, lon_E, h) - h\} - b\{\ln [D_i(lat_E, lon_E, h)] - \ln h\} \quad (1.14)$$

Il problema può quindi essere risolto massimizzando la funzione di log-likelihood (Maximum Likelihood Estimation: MLE) rispetto ai 6 parametri incogniti. Per effettuare l'MLE viene usato l'algoritmo implementato nella subroutine di Fortran BCONF/DBCONF (IMSL, 1991).

I “metodi 1-6” si differenziano per diverse combinazioni di stima dei parametri incogniti, i quali possono essere lasciati liberi durante il run dell'algoritmo o assunti noti e pre-stimati con metodi esterni all'MLE:

metodo 1 si vincolano h , I_E , b e a con stime esterne all'MLE come in Pasolini et al. (2008); in particolare per l'*intensità epicentrale* viene utilizzato il predittore $I_E = \bar{I} + a[h - \bar{D}] + b[\ln h - \overline{\ln D}]$ dove \bar{I} , \bar{D} , $\overline{\ln D}$ sono le medie aritmetiche dei relativi parametri calcolate includendo tutti i punti del piano quotato del terremoto in questione; questa relazione viene quindi inserita nella relazione (1.14) per μ_j e il predittore risultante viene utilizzato nell'MLE per stimare solamente lat_E e lon_E ;

metodo 2 I_E , b e a sono vincolati come al “metodo 1”, mentre lat_E , lon_E e h vengono stimati contemporaneamente con l'MLE;

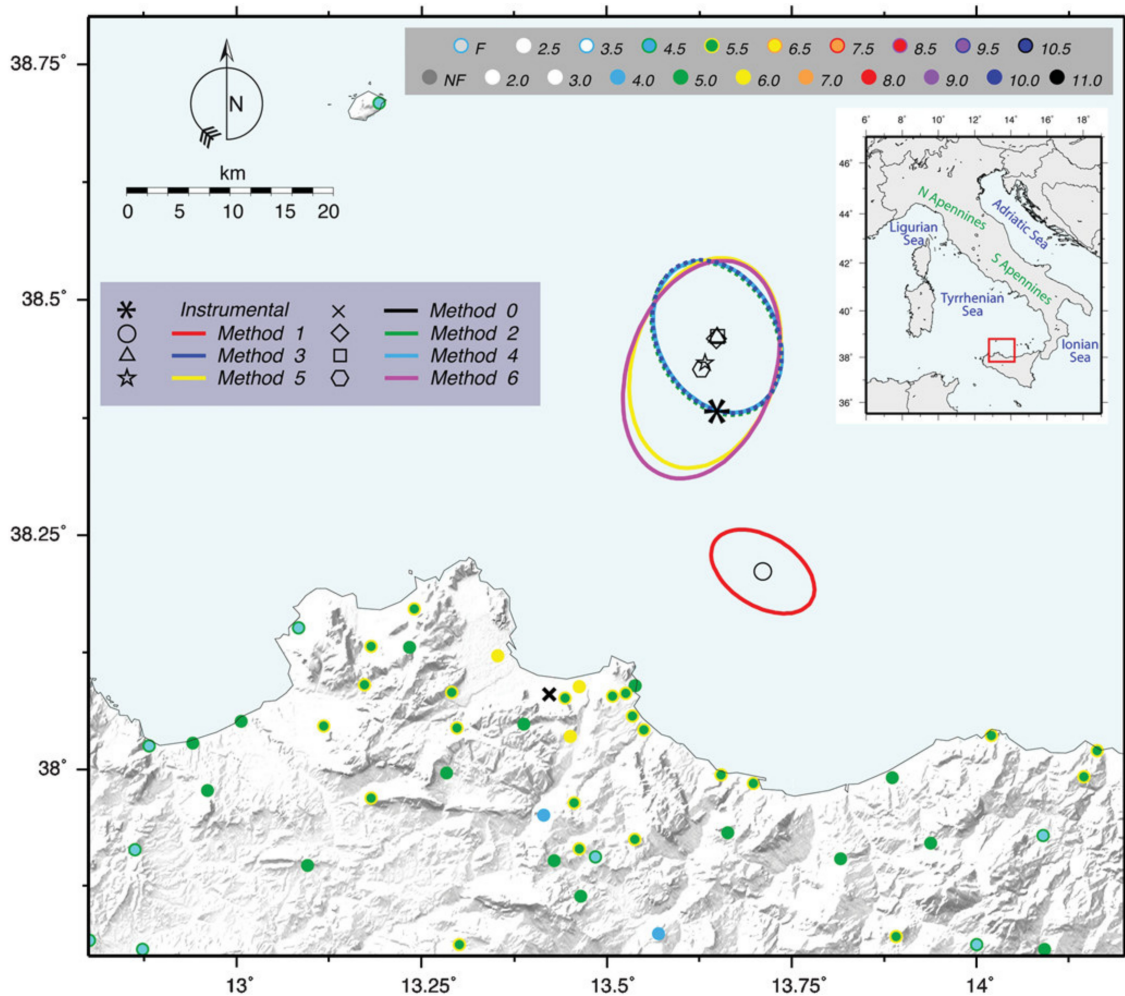


Figura 1.13: Intensity distribution, instrumental (asterisk) and macroseismic (symbols) epicenters, and ellipses of 90% uncertainty (computed using the Hessian approach) of the earthquake of 6 September 2002 that occurred in the southern Tyrrhenian Sea (southern Italy), close to Palermo.

metodo 3 si vincolano h , b e a , stimati come in Pasolini et al. (2008), e si usa l'MLE, con il predittore generale della (1.14), per stimare I_E , lat_E , lon_E ;

metodo 4 b e a sono fissati come in Pasolini et al. (2008) e gli altri 4 parametri sono stimati congiuntamente con l'MLE;

metodo 5 b viene fissato come in Pasolini et al. (2008) e gli altri vengono calcolati con l'MLE;

metodo 6 tutti e 6 i parametri sono stimati contemporaneamente con l'MLE;

1.4.2 Magnitudo equivalente

Per calcolare la *magnitudo equivalente* sono implementati due metodi differenti.

Il primo, l' "old magnitude method", descritto in Gasperini et al. (1999, Appendix 2) e in Gasperini e Ferrari (1995) e Boschi et al. (1997), segue l'approccio misto proposto da Galanopoulos (1961) che, per ridurre errori sistematici nella stima della magnitudo, prevede sia l'utilizzo dell'*intensità epicentrale* sia l'utilizzo dell'*estensione dell'area di risentimento* (FA). Dal dataset di MDPs di un dato terremoto viene effettuata una regressione utilizzando una relazione empirica del tipo proposto da Sibol et al. (1987); la relazione di Sibol utilizzata in Boxer è tuttavia modificata, come segue, per poter includere nel calcolo non soltanto l'FA complessiva, ma anche l'area delle singole isosime:

$$M = a + bI_0^2 + c \log^2(A_I) \quad (1.15)$$

a , b e c sono parametri di calibrazione che devono essere ricavati per un set di terremoti di cui sono note la magnitudo strumentale, il piano quotato delle intensità e l'epicentro macrosismico calcolato con il "metodo 0" descritto al paragrafo precedente; I_0 è l'intensità epicentrale assunta pari alla massima intensità I_{\max} del PQ se esistono almeno 2 MDPs con quel valore, altrimenti si assume $I_0 = I_{\max} - 1$; mentre A_I è l'area di un'isosisma circolare centrata nell'epicentro macrosismico, di raggio R_I calcolato come mediana delle distanze dall'epicentro macrosismico degli MDPs relativi a quel particolare valore di intensità per il quale si vuole calcolare A_I .

Nella fase di calibrazione si effettua una regressione per ogni isosisma d'intensità per la quale sono disponibili almeno 4 MDPs.

Viene infine stimata la *magnitudo equivalente* per un terremoto calcolando la media pesata dei valori di magnitudo stimati per ciascuna isosisma disponibile. I pesi sono gli inversi degli errori quadratici della regressione.

Il secondo metodo ("alternative method"), descritto in Gasperini et al. (2010), segue l'idea di Bakun e Wentworth (1997) di trattare la *magnitudo equivalente* similmente a quella strumentale (C. Richter, 1935). Per lo scopo è utilizzata una relazione che permette di stimare una magnitudo M_j per ogni MDP del dataset di un terremoto. Questa è ricavata dalla combinazione della relazione di attenuazione di Pasolini et al. (2008) (1.11) e della relazione empirica tra *intensità epicentrale* I_E e *magnitudo momento* M_w di Fuller (1987) $I_E = c + dM_w$, e risulta

$$M_j = \frac{1}{d} \{ I_j + \hat{a}(D_j - \hat{h}) + \hat{b} [\ln(D_j) - \ln(\hat{h})] - c \} \quad (1.16)$$

dove \hat{h} non è l'effettiva profondità del terremoto in questione, ma un valore di profondità tipico della regione del terremoto, stimato con metodo indipendente (e.g. Gasperini et al. (2010, Tabella 1, Extended)); \hat{h} viene usato per ottenere il best fit della relazione di Fuller. La profondità effettiva h dell'evento, ricavata con uno dei metodi descritti al paragrafo precedente, è invece usata per il calcolo delle distanze ipocentrali macrosismiche D_j . Anche i coefficienti \hat{a} e \hat{b} sono calcolati come secondo Pasolini et al. (2008).

La magnitudo finale del terremoto è calcolata operando una media su tutti i valori di M_j calcolati dalle osservazioni individuali di intensità I_j .

1.4.3 Dimensioni e orientazione della faglia

Sulla base delle considerazioni effettuate al paragrafo relativo all'approccio macrosismico nella stima delle *dimensioni* e dei *parametri di faglia* Gasperini et al. 2010

applicano il seguente metodo.

Si assume una faglia con geometria rettangolare le cui 2 dimensioni sono stimate utilizzando le relazioni log-lineari empiriche di Wells e Coppersmith (1994), una per la *lunghezza* (RLD) l'altra per la *larghezza* (RW), derivate come media tra diverse possibili relazioni di questo tipo per i diversi stili di faglia (normale, trascorrente, ...):

$$\log_{10}(RLD) = 0.59(\pm 0.02) M - 2.44(\pm 0.11), \quad (1.17)$$

$$\log_{10}(RW) = 0.32(\pm 0.02) M - 1.01(\pm 0.10), \quad (1.18)$$

dove M è la *magnitudo momento*, e i coefficienti sono quelli stimati da Gasperini et al. (1999). Dalla stima della *magnitudo equivalente* del terremoto, calcolata con uno dei metodi descritti al paragrafo precedente, si stimano con queste relazioni *lunghezza* RLD e *larghezza* RW della faglia.

Per la stima dell'*orientazione della faglia*, invece, considerati i segmenti di congiunzione tra epicentro macrosismico e ciascun MDP, si assume che la direzione di *strike* sia approssimata dagli azimut α_i degli MDPs con probabilità tanto maggiore quanto di essi è maggiore la distanza epicentrale. La direzione dello *strike* viene dunque stimata effettuando una media pesata assiale $\bar{\alpha}$ degli azimut di quei siti con intensità al di sopra di una certa soglia I_{\min} . I pesi w_i sono definiti come il rapporto tra la distanza epicentrale del sito e la distanza epicentrale predetta per lo stesso da una relazione di attenuazione cubica empirica (Berardi et al., 1993). La soglia I_{\min} è invece determinata in modo tale che i siti selezionati abbiano una distanza epicentrale media il più possibile prossima alla *lunghezza* della faglia RLD stimata con l'(1.17). Le relazioni esplicite utilizzate sono le seguenti:

$$\bar{\alpha} = \frac{1}{2} \cos^{-1} C/Q = \frac{1}{2} \sin^{-1} S/Q \quad (1.19)$$

con C , S , Q funzioni dei pesi w_i e degli azimut α_i :

$$C = \frac{\sum_{i=1}^N w_i \cos 2\alpha_i}{\sum_{i=1}^N w_i}, \quad S = \frac{\sum_{i=1}^N w_i \sin 2\alpha_i}{\sum_{i=1}^N w_i}, \quad Q = \sqrt{C^2 + S^2}, \quad (1.20)$$

$$\alpha_i = \begin{cases} \sin^{-1} \{ \cos(\text{lat}_i) \sin(\text{lon}_i - \text{lon}_E) / \sin(R_i/R_0) \}, & \text{if } \sin(\text{lat}_i) \geq \cos(R_i/R_0) \sin(\text{lat}_E) \\ -\sin^{-1} \{ \cos(\text{lat}_i) \sin(\text{lon}_i - \text{lon}_E) / \sin(R_i/R_0) \}, & \text{if } \sin(\text{lat}_i) < \cos(R_i/R_0) \sin(\text{lat}_E) \end{cases}, \quad (1.21)$$

$$w_i = \frac{R_i}{\sqrt[3]{\{(I_0 - I_i) + 0.46\}/0.93}}. \quad (1.22)$$

Dove lat_i , lon_i , lat_E , lon_E sono le coordinate dell' i -esimo sito ed epicentrali, R_i sono le distanze epicentrali geodetiche calcolate su una sfera di raggio $R_0 = 6371 \text{ Km}$.

Infine in Boxer la sorgente viene rappresentata (Figura 1.11) come un rettangolo centrato nell'epicentro macrosismico con dimensione maggiore pari alla RLD, parallela alla direzione di orientazione $\bar{\alpha}$, e dimensione minore corrispondente alla proiezione alla superficie dell' RW di una generica faglia con angolo di *dip* $\delta = 45^\circ$ (1.11; 1.14). Se invece non è possibile calcolare l'orientazione della faglia al posto della "box" viene disegnato un cerchio centrato nell'epicentro.

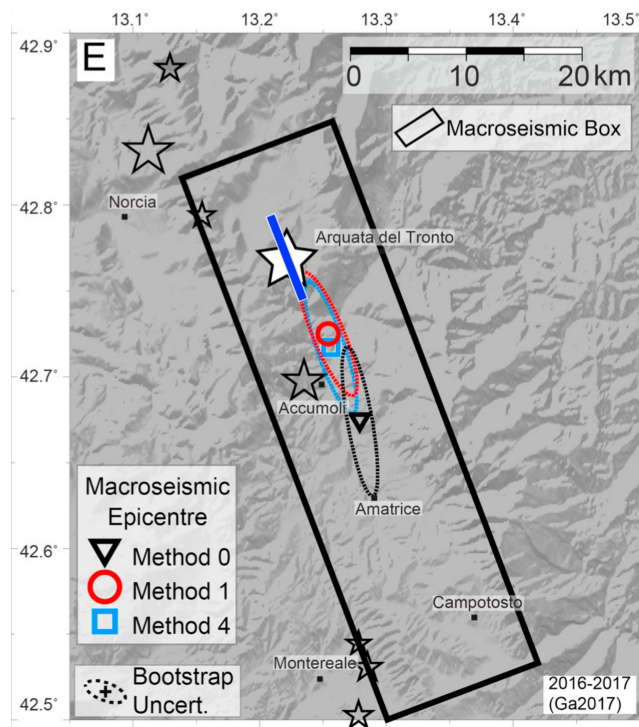


Figura 1.14: Terremoto del Centro Italia del 24 Agosto 2016. Stime da dati macrosismici di Galli et al., 2017. Dimensioni e orientazione della faglia sorgente sono rappresentate da un rettangolo centrato all'epicentro macrosismico calcolato con il "metodo 0" (e profondità calcolata come in R. M. W. Musson (1996)), quando il dataset consente il calcolo dell'orientazione. Sono mostrati gli epicentri calcolati con i "metodi 1 e 4"; i riferimenti strumentali di epicentro e orientazione della faglia (la stella bianca e il segmento in blu); le ellissi di incertezza bootstrap (90%); epicentri strumentali di altri eventi con $M_w > 5$ (stelle grigie).

1.4.4 Incertezze

Per la valutazione delle incertezze è bene effettuare alcune precisazioni e considerazioni (Gasperini et al., 2010, p. 2042).

L'*incertezza* è un indice della dispersione dei valori assunti da una variabile/parametro misurata; essa rappresenta un range di valori entro cui ci si aspetta di trovare una certa frazione dei valori che la variabile assume nella misurazione (Boxer permette di preimpostare il livello di confidenza delle incertezze).

L'*accuratezza* è invece definita come la discrepanza tra il valore di una misura e il valore vero (spesso ignoto) della variabile corrispondente.

Si considera l'incertezza un buon estimatore dell'accuratezza della stima di un parametro nell'assunzione che le misure non siano affette da bias o errori sistematici.

Le incertezze calcolate con Boxer sono rappresentative della dispersione delle stime dei parametri macrosismici, le quali tuttavia si basano su assunzioni che possono essere errate poiché l'informazione macrosismica è intrinsecamente soggettiva e l'intensità macrosismica non è basata su chiare definizioni fisiche. Quindi né è noto il valore vero, né è ovvio che le misurazioni non siano affette da bias o errori sistematici. Da ciò si evince che le incertezze dei parametri macrosismici potrebbero largamente sottostimare l'accuratezza nel riprodurre i rispettivi parametri strumentali, tanto più se, a rigore, non indicano la stessa grandezza fisica. Questa considerazione è particolarmente vera per l'incertezza sulla magnitudo, che è di per se un parametro empirico non basato su solide descrizioni fisiche anche se calcolato usando dati strumentali.

Rimane tuttavia utile stimare le incertezze per effettuare confronti tra i metodi e valutare i limiti superiori di accuratezza.

In Boxer vengono stimati due tipi di incertezze: *formali* ed *empiriche*, descritte in dettaglio in Gasperini et al. (2010), Appendice 3. Le *incertezze formali* sono stimate in base alla definizioni dei relativi parametri e via via le si descrive di seguito, mentre per le *incertezze empiriche* è implementato il metodo bootstrap che consiste nella costruzione di dataset sintetici ricampionando randomemente (con rimpiazzo) il dataset originale. In quest'ultimo caso le incertezze vengono quindi stimate dai parametri ricalcolati per ciascun dataset sintetico¹³. Ci si aspetta inoltre che le *incertezze bootstrap* saranno più grandi di quelle *formali*, poiché le prime dipendono da stime dell'epicentro macrosismico, effettuate con la medesima procedura, che è diverso per ciascun dataset ricampionato, mentre le seconde sono basate su una sola stima dell'epicentro. Questo aspetto della procedura *bootstrap* ha però il vantaggio di descrivere la variabilità dei parametri simulando il caso frequente in macrosismica in cui siano perduti o trascurati dati.

Per la *localizzazione epicentrale* effettuata con il “metodo del baricentro” l'*incertezza formale* σ_{formal} sulle coordinate è calcolata come deviazione standard delle coordinate dei siti selezionati dal campione per calcolare la trimmed mean; si tenga conto tuttavia che questa stima, a rigore, è un indicatore della variabilità del campione, e non della media, corretto con un fattore a denominatore, e inoltre che la formula

¹³Il principio sottostante al *metodo bootstrap* è che la distribuzione empirica delle frequenze del dataset originale sia un'ottima stima della distribuzione ignota della popolazione, in termini di massima likelihood. I dataset ottenuti con il ricampionamento manterranno inalterate le caratteristiche statistiche del dataset originale, e potranno quindi essere usati per valutare le proprietà di un campione della popolazione.

dipende dall'estensione dell'area mesosismica¹⁴ ; per i “metodi 1-6” σ_{formal} è invece valutata calcolando la matrice di varianza/covarianza con il metodo dell'Hessiana della funzione di log-likelihood calcolata al suo massimo (1.13) (Guo e Ogata, 1997). Inoltre per tutti i “metodi” (0-6) vengono calcolate le *incertezze empiriche* con il *metodo bootstrap*.

Per la stima dell'incertezza sulla *profondità* effettuata con il metodo di R. M. W. Musson (1996) sono usati sia il metodo dell'Hessiana della funzione di log-likelihood per ottenere l'*incertezza formale*, sia il *metodo bootstrap* per ottenere l'*incertezza empirica*.

Le *incertezze formali* dell' “old magnitude method” per il calcolo della *magnitudo equivalente* sono stimate in due differenti modi (Gasperini et al., 2010, Appendice 2): 1) radice quadrata dell'inversa della somma delle inverse delle deviazioni standard quadrate delle regressioni di ciascun'isosisma, calcolate come funzioni delle deviazioni standard delle regressioni $M - I$ e del numero di dati usati; 2) deviazioni standard pesate della media delle stime di magnitudo effettuate usando diverse isosisme.

L'*incertezza formale* sull' “alternative magnitude method” è definita come la deviazione standard della media delle stime della magnitudo calcolate con (1.16) per ogni valore di intensità.

Per entrambi i metodi di calcolo della magnitudo sono inoltre calcolate le incertezze con il *metodo bootstrap*.

È da notare che solamente il modo 1) prende in considerazione le incertezze inerenti alla magnitudo strumentale e alla regressione $M - I$ e può a ragione essere usato come indice di accuratezza della stima macrosismica nel riprodurre la stima strumentale, mentre tutti gli altri descrivono piuttosto la riproducibilità della procedura di calcolo.

Infine l'*incertezza formale* dell'*orientazione della faglia*, e cioè dell'orientazione assiale media è calcolata con l'approccio di Fisher (1993, p. 34) per il calcolo della deviazione standard circolare del campione.

Anche per l'orientazione assiale media è calcolata l'*incertezza bootstrap*.

1.4.5 Accuratezza, affidabilità e confronti tra i metodi

In Gasperini et al. (2010) sono stati confrontati i risultati dei metodi possibili (0-6) con dati strumentali per individuare una statistica di attendibilità.

È emerso che il “metodo 0” fornisce una stima robusta della *localizzazione*. Si è rivelato stabile e affidabile utilizzando dati macrosismici da differenti versioni dei cataloghi (Gasperini e Ferrari, 2000), anche per terremoti con numero di MDPs $< 5-10$; risulta invece inaffidabile nel caso di terremoti con epicentro in mare o privi di MDPs nell'area epicentrale.

Tutti e 7 i metodi di *localizzazione* sono consistenti per la maggior parte dei terremoti: il 50% delle localizzazioni effettuate con i nuovi metodi differiscono al più di $5Km$ dalla localizzazione col “metodo 0”, mentre il 90% di essi ne differiscono per non più di $15-20Km$.

Al confronto tra tutti i metodi di Boxer e le stime strumentali nel 20% dei casi le differenze tra gli epicentri non hanno superato i $5Km$ e solo nel 5% dei casi hanno superato i $30Km$.

Inoltre il “metodo 1” ha prodotto buoni risultati nella *localizzazione* di terremoti con

¹⁴meizoseismal area: area in cui gli effetti del terremoto sono stati avvertiti con maggiore severità

epicentro in mare.

Le stime di *profondità* hanno invece in generale debole corrispondenza con le stime strumentali.

L' "old magnitude method" per la stima della *magnitudo equivalente* si è mostrato efficace nel riprodurre le magnitudo momento strumentali: è stato ottenuto un valore del coefficiente $R^2 = 0.74$ nel confronto effettuato da Gasperini et al. (2010) e un valore $R^2 = 0.91$ nel confronto effettuato in Gasperini e Ferrari (2000) con magnitudo calcolate invertendo le forme d'onda.

Metodi per la *magnitudo* vecchi e nuovi si sono mostrati altrettanto accurati nei casi in cui venivano utilizzati i metodi più semplici per il calcolo dell'epicentro.

Si è riscontrato un ottimo accordo tra *magnitudo* macrosismiche e strumentali con differenze massime di 0.1 unità di magnitudo e deviazioni standard di 0.3 unità di magnitudo.

Gasperini et al. (1999) hanno stabilito una soglia di affidabilità della stima, individuata in base a quei terremoti per i quali la dimensione massima della faglia sia confrontabile con la risoluzione spaziale dei dati di intensità, e cioè per quei terremoti con magnitudo momento $M_w > 5.5$ o con lunghezza $RLD > 5-10Km$, e per quei dataset di almeno 5 MDPs. Spesso infatti il potere risolutivo della distribuzione degli MDPs di intensità più elevata utilizzati per il calcolo della media pesata assiale $\bar{\alpha}$ degli azimut non è sufficiente per risolvere lo *strike*.

In particolare la soglia di numerosità del dataset è necessaria per poter effettuare significativamente i test di Rayleigh e Kuiper per verificare la non uniformità della distribuzione degli azimut degli MDPs. In particolare, per verificare che la distribuzione degli azimut *non* sia uniforme e quindi che la media assiale calcolata *non* sia priva di senso, vengono effettuati i test di Rayleigh e di Kuiper sull'ipotesi H_0 di distribuzione uniforme. H_0 *non* viene rigettata solamente se entrambi i test sono positivi; in tutti gli altri casi si ritiene affidabile la stima di $\bar{\alpha}$.

Da un confronto tra le orientazioni di Boxer e le orientazioni dei meccanismi focali dei rispettivi terremoti è stata complessivamente riscontrata una massima deviazione di 20°-30°.

Infine, per ciò che riguarda le *incertezze*, le *formali* sono più rapide da calcolare delle *bootstrap*, e si mostrano in generale ben rappresentative della variabilità dei parametri. Per la magnitudo tuttavia si osserva che il metodo bootstrap ed anche le deviazioni standard di singola magnitudo sottostimano le incertezze.

Capitolo 2

Dati macrosismici di felt reports raccolti da Internet su base volontaria

Per *felt reports* si intendono le descrizioni qualitative degli effetti di un terremoto sulle strutture antropiche, sulle persone e sull'ambiente. L'indagine macrosismica si occupa di raccogliere questi dati, i quali: possono essere estrapolati da resoconti contemporanei all'evento o successivi; possono essere relativi ad una testimonianza diretta o indiretta; possono essere raccolti sia da esperti sia da cittadini comuni, e possono seguire protocolli di raccolta prestabiliti o fornire descrizioni completamente libere. Per queste ragioni l'informazione contenuta nei dati di *felt reports* è in generale molto disomogenea; esistono infatti diversi tipi di dato di *felt reports* definibili in base alle caratteristiche appena esposte. Per gli scopi di questa tesi verranno solamente analizzati *felt reports* raccolti da Internet su base volontaria.

Si vuole qui ricordare che i *felt reports*, indipendentemente dalle caratteristiche particolari di raccolta, messi in relazione con le descrizioni interne alle scale macrosismiche, permettono di quantificare il fenomeno macrosismico e sintetizzarlo nel parametro di *intensità*, che rappresenta il risentimento alla località. Un'adeguata distribuzione di intensità stimate nell'area di risentimento di un terremoto definisce il *piano quotato delle intensità* (PQ, o distribuzione di *Macroseismic Data Points*, MDPs) che può essere utilizzato per determinare i *parametri di sorgente*, come è descritto al capitolo precedente.

2.1 Stato dell'arte dei felt reports raccolti da Internet per la sismologia macrosismica citizen-based

Negli ultimi decenni lo sviluppo e la diffusione delle tecnologie di comunicazione basate su Internet hanno permesso un parallelo sviluppo della cosiddetta *citizen-based science*, espressione con cui ci si riferisce a tutte quelle attività scientifiche in cui scienziati non professionisti e volontari non addestrati partecipano a vari livelli alla raccolta, all'analisi e alla diffusione dei dati di un progetto scientifico (Cohn, 2008; Silvertown, 2009).

Tra le discipline scientifiche per le quali è stato attivato questo sistematico coinvolgimento del pubblico è anche la sismologia macrosismica, che in virtù della natura percettiva del suo parametro principale di interesse (l'intensità macrosismica) ne trae alcuni vantaggi computazionali. Infatti la *macrosismologia citizen-based* dispone di una grande mole di dati, detti *crowdsourced macroseismic data*, che vengono forniti dalla stessa popolazione che avverte gli eventi sismici e raccolti quasi in tempo reale. È eloquente il grafico di Figura 2.1 che mostra un tipico picco del numero di accessi degli utenti al sito di EMSC ¹ all'occorrere di un terremoto.

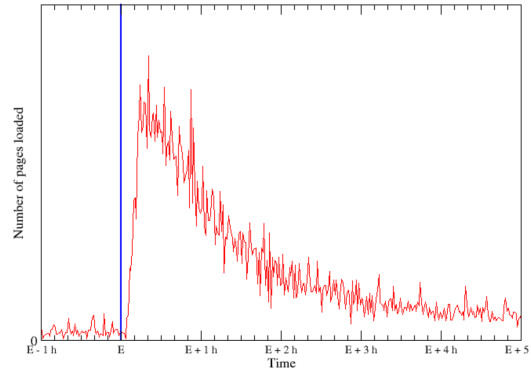


Figura 2.1: Typical surge of web traffic on EMSC web site after a felt earthquake (blue line) (Mazet-Roux et al., 2008).

Grazie ai sistemi automatici di elaborazione si riesce ad ottenere una rapida rilevazione dei terremoti avvertiti ed una loro mappatura con risoluzione talvolta superiore a quella ottenuta con i metodi tradizionali (indagini in situ condotte da team di esperti) (D. J. Wald et al., 2011). D'altra parte gli stessi utenti traggono immediato beneficio dal loro contributo: possono infatti accedere alle risorse informative sull'evento occorso, che vengono rapidamente sviluppate e pubblicate online. I sistemi di *citizen-based science* che utilizzano la rete Internet hanno dunque la caratteristica di essere bidirezionali ed interattivi.

Questa particolare modalità di indagine macrosismica è disposta su apposite piattaforme online sviluppate da istituzioni come la United States Geological Survey (USGS)², l'European Mediterranean Seismological Centre (EMSC)³ e l'Istituto Nazionale di Geofisica e Vulcanologia (INGV)⁴, con lo scopo di supportare i relativi sistemi di monitoraggio e prevenzione sismica. Con le reti di strumenti sismometrici globali e locali e con le modalità tradizionali di raccolta dei dati macrosismici, essa consolida la caratterizzazione sismica dei territori e fornisce informazioni quasi-real time utilizzabili, oltre che dai sismologi per scopi di ricerca, dai cittadini e dalle autorità per fini di gestione del rischio sismico (Steed et al., 2019). Con le varie tipologie di dati trattate assieme vengono prodotte mappe di risentimento, di danno, di pericolosità e di rischio, vengono stimati i parametri di sorgente dei singoli terremoti, ed estrapolate informazioni sugli agglomerati urbani vicini all'epicentro e sulla popolazione coinvolta nell'evento.

¹<http://www.emsc-csem.org>

²<https://www.usgs.gov/>

³<https://www.emsc-csem.org/#2>

⁴<http://www.ingv.it/it/>

Tra i *dati macrosismici crowdsourced web-based* si distinguono: i *felt reports* propriamente detti, le cosiddette “impronte digitali”, i liberi commenti e il materiale multimediale geolocalizzato (Bossu et al., 2018b). I primi consistono in valutazioni geolocalizzate di diagnostici di intensità macrosismica a cui si aggiungono sia i commenti liberi degli utenti rispetto all’esperienza vissuta, sia eventuali fotografie e video inviate dagli utenti. Per “impronte digitali” si intendono invece le tracce informatiche geolocalizzate sia dei semplici accessi degli utenti alle piattaforme online di monitoraggio sismico – anche senza rilascio di testimonianza – sia delle ricerche sui social e sui motori di ricerca di parole chiave relative al terremoto.

Tra i *crowdsourced macroseismic data* soltanto i *felt reports* sono oggetto di analisi in questo lavoro di tesi con lo scopo di stimare i parametri di sorgente del terremoto.

I sistemi di raccolta online dei *felt reports* hanno iniziato ad essere operativi dagli anni 90, seppur ancora a livello regionale. Essi costituivano un aggiornamento rispetto ai sistemi già attivi basati su interviste telefoniche e su questionari cartacei spediti per posta o via email (Cajka e Halchuk, 1998; D. Wald et al., 1999). Fu “Did You Feel It” (DYFI)⁵, inizialmente noto come CIIM (Community Internet Intensity Map), ad essere introdotto come primo sistema interamente automatizzato per la raccolta di report online da questionari macrosismici e per l’assegnazione dell’*intensità*. Esso venne attivato per il territorio della California nel 1997 (D. Wald et al., 1999). L’algoritmo per l’assegnazione delle *intensità* procedeva raggruppando i questionari secondo i codici postali e calcolando, in base alle risposte alle domande, un parametro interno al metodo che veniva poi relazionato alla scala MMI tramite una regressione lineare. DYFI venne diffuso nel 2000 in tutti gli USA e dal 2004 è disponibile per la raccolta di dati da tutto il Pianeta.

Inoltre dal 2007 è online il sistema di raccolta italiano automatizzato “Hai Sentito il Terremoto” (HSIT)⁶ sviluppato dall’INGV e basato anch’esso sui questionari (Tosi et al., 2007). In questo sistema la procedura per l’attribuzione dell’*intensità* prevede intanto di associarne una ad ogni questionario in base al principio della massima verosimiglianza delle risposte con i diagnostici grado-specifici della scala di riferimento, poi si effettua la media per città o paesi (Tosi et al., 2015). La procedura è sviluppata sia per le *intensità* della scala MCS sia per quelle definite dall’EMS.

Nel 2010 il French Seismological Central Office è stato il primo ad utilizzare per la raccolta, effettuata tramite l’app “SismoCom” e la piattaforma EQN, i *thumbnails* (Bossu et al., 2017; Finazzi, 2016; 2020; Finazzi e Fassò, 2017), consentendo l’immediata codifica in gradi di *intensità* della singola testimonianza di un utente. Infatti i *thumbnails* sono immagini di scenari di scuotimento, ovvero schizzi grafici stilizzati raffiguranti esempi di effetti percepibili durante una scossa in tipici contesti urbani o domestici, e a ciascuno di essi corrisponde un grado della scala macrosismica di riferimento scelta.

Si cita qui anche il sistema di raccolta dell’EMSC, operativo al livello globale dal 2004 con un questionario online con domande estese disponibile in diverse lingue. Il questionario esteso è stato poi sostituito da uno basato sui *thumbnails* disponibile sia

⁵<https://earthquake.usgs.gov/data/dyfi/>

⁶<https://www.hsit.it/>

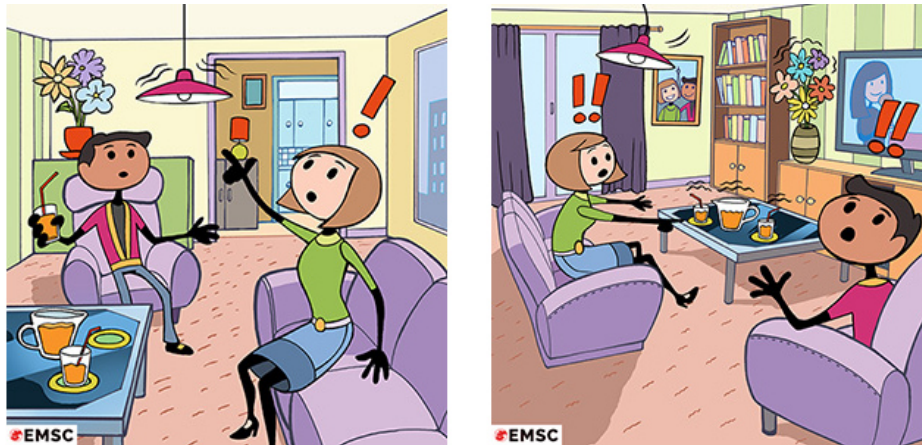


Figura 2.2: Esempio di *thumbnails*: immagini guida per l'utente per il rilascio del feedback di *felt report*; ciascuna delle due vignette corrisponde ad un unico grado di intensità (EMSC; <https://www.emsc-csem.org>)

su sito web⁷, sia, dal 2014, su un'applicazione per supporto mobile anch'essa basata sui *thumbnails* ("LastQuake" app scaricabile dai Google ed Apple store). La scala di riferimento utilizzata da questo sistema è l'EMS (Grunthal, 1998), tuttavia esso, a differenza di DYFI e HSIT, non include una procedura automatica per l'assegnazione di *intensità* statisticamente robuste che mappino adeguatamente il piano quotato (D. J. Wald et al., 2011).

Ormai l'*indagine macrosismica crowdsourced web-based* è ampiamente utilizzata nel mondo, esistono infatti analoghi sistemi di raccolta in paesi come il Belgio, l'Inghilterra e il Canada. Tuttavia a causa di alcune criticità intrinseche continua ad essere affiancato a questo metodo quello tradizionale di raccolta in situ, soprattutto per gli eventi al di sopra della soglia di danno (D. J. Wald et al., 2011).

2.2 Caratteristiche specifiche e problematiche dei dati

I dati di *felt reports crowdsourced web-based* sono oggetto specifico di studio di questa tesi; essi sono dati sparsi di natura macrosismica raccolti sulla base delle testimonianze di utenti volontari che dispongono di un dispositivo per l'accesso ad Internet.

Per questo tipo di dati la raccolta delle testimonianze dei risentimenti sismici degli utenti avviene attraverso i canali di raccolta delle istituzioni sismologiche competenti (USGS, EMSC, INGV, ...). Tali canali consistono in piattaforme online disponibili su siti web classici, su siti web per dispositivi mobili, oppure ancora su apposite applicazioni (apps). Gli utenti possono liberamente e gratuitamente accedere alle piattaforme e condividere la loro esperienza in maniera guidata.

Per guidare il rilascio delle testimonianze esistono due tipologie principali di *questionari*: estesi e dettagliati, oppure basati su *thumbnails*. Questi ultimi sono stati già introdotti nel paragrafo precedente (figura 2.2), mentre i questionari del

⁷https://www.emsc-csem.org/Earthquake/Contribute/choose_earthquake.php?lang=it;
<https://m.emsc.eu/index.php>

primo tipo si compongono di domande a risposta multipla relative sia agli effetti transienti dello scuotimento che ai danni permanenti agli edifici. Queste domande, in misura maggiore rispetto ai *thumbnails*, hanno lo scopo di indagare nel dettaglio precisi diagnostici il più possibile grado-specifici, in quanto permettono all'utente di dettagliare certe risposte, mentre i disegni sono solo una rappresentazione "rigida" dello scenario, con un certo grado intrinseco di ambiguità o di inesattezza rispetto allo scenario che osserva l'utente. Questo aspetto influisce sulla intensità IDP assegnata da ogni utente. Tuttavia nessuno dei due metodi è tale da poter acquisire ed elaborare informazioni sulla vulnerabilità delle strutture, necessarie per la discriminazione dei gradi più elevati delle scale d'intensità; infatti per valutare la vulnerabilità sarebbe necessario che tutti gli utenti avessero competenze ingegneristiche.

A questo punto si pongono due fondamentali problemi al fine di poter utilizzare le testimonianze per caratterizzare adeguatamente il campo macrosismico di un terremoto:

1. il primo consiste nell'attribuzione alle singole testimonianze di *caratteristiche* sintetiche individuali significative e cioè le *coordinate geografiche* e un valore d'intensità; questa operazione porta a definire un IDP (*Individual intensity Data Point*) per ciascuna testimonianza.
2. Il secondo problema si pone invece come diretta conseguenza della natura statistica del parametro di intensità e consiste in un irrobustimento dell'informazione sull'intensità contenuta già negli IDPs: di fatto si effettua una trasformazione dal campo degli IDPs al campo degli MDPs (*Macroseismic Data Points*) raggruppando i primi in gruppi detti *clusters* e calcolando per ciascun *cluster*, con operazioni statistiche sulle *caratteristiche* degli IDPs in esso contenuti, l'MDP più rappresentativo.

Il problema 2) può essere risolto con metodi e tecniche differenti che rientrano nel dominio della *clustering analysis*, oggetto di studio in questa tesi e trattata al capitolo successivo.

Il problema 1) invece, di seguito descritto, si divide a sua volta nei due seguenti sotto-problemi, variamente affrontati in letteratura per i diversi sistemi e canali di raccolta (D. J. Wald et al., 2011; Tosi et al., 2015; Amorese et al., 2015).

Attribuzione dell'intensità: In generale, per l'attribuzione dell'intensità si sceglie prima di tutto una scala macrosismica a cui riferirsi. In secondo luogo, nel caso di report da *thumbnail*, l'attribuzione è immediata e avviene in base alla scelta dell'utente di uno e uno solo tra i *thumbnail* disposti, tale scelta avrà corrispondenza univoca prestabilita con i gradi della scala; nel caso invece di questionario esteso l'attribuzione dell'intensità necessita di un'analisi a posteriori delle risposte volta ad individuare il grado della scala che ha maggiori corrispondenze con i diagnostici individuati dalle risposte stesse.

Localizzazione: La localizzazione di un singolo questionario dipende invece dalla tecnologia del dispositivo attraverso il quale il questionario viene compilato e dalle informazioni che l'utente sceglie di condividere. In particolare essa può avvenire automaticamente mediante localizzazione con tecnologia GPS (Global Positioning System) o mediante l'IP (Internet Protocol), oppure manualmente

in base al codice postale indicato dall'utente stesso. A seconda del metodo di localizzazione si ottiene dunque o un'univoca coppia di coordinate per ogni *felt report* o già dei raggruppamenti delle testimonianze per aree geografiche locali a ciascuna delle quali corrisponde convenzionalmente una coppia di coordinate rappresentative.

È bene ricordare che la rilevazione degli IDPs non costituisce di per sé prova dell'occorrenza di eventi sismici di origine tettonica e nemmeno dell'occorrenza di danni (D. J. Wald et al., 2011; Bossu et al., 2018b). Infatti effetti di scuotimento possono essere causati da onde elastiche generate da altri fenomeni come boom sonici, esplosioni antropogeniche, meteore, e le testimonianze rilasciate sulle piattaforme online possono essere arbitrarie e immotivate. I dati necessitano dunque di essere confermati mediante i diversi metodi di osservazione e di rilevazione, sia sismologico-strumentali sia macrosismici, indipendenti dai *felt reports crowdsourced web-based* e implementati negli stessi sistemi di monitoraggio sismico istituzionali.

Inoltre la possibilità di raccogliere i *felt reports crowdsourced web-based* dipende dalla funzionalità del metodo stesso di raccolta (automatizzato e basato sulla tecnologia Internet) che può presentare malfunzionamenti, anche importanti, soprattutto all'occorrere di eventi dannosi per le linee di comunicazione. Tuttavia, se confrontato con i metodi di raccolta tradizionali, questo si rivela essere un metodo di raccolta veloce, efficiente e a basso costo, il quale permette la mappatura dell'area di risentimento sismico già a poche decine di minuti dal tempo della sorgente (Bossu et al., 2017). È fondamentale la diffusione e la popolarità dei *canali di crowdsourcing* tra la popolazione della regione in cui avviene un terremoto. Dove vi è maggiore densità di popolazione e diffusione dei *canali web-based* aumenta la capacità degli IDPs di mappare il campo macrosismico con migliori risoluzione e rilevanza statistica. In generale si riscontra una tendenza all'aumento nel tempo del numero delle testimonianze raccolte con questi metodi. Si mostra ad esempio in Figura 2.3 il grafico rappresentativo di tale dipendenza temporale per gli IDPs del sistema EMSC raccolti su tutto il globo nell'intervallo 1 Luglio 2014 - 1 Maggio 2016.

Inoltre si osservano caratteristiche ricorrenti nel pattern degli IDPs di diversi terremoti. In particolare è stato mostrato che la percentuale di popolazione che accede ai canali di crowdsourcing macrosismico a seguito di un evento ha una qualche dipendenza dal livello di scuotimento prodotto nell'evento stesso (Bossu et al., 2014; Bossu et al., 2018a). Infatti si riscontra diffusamente che i gradi di *intensità* corrispondenti a livelli di scuotimento che provocano danni sono per lo più assenti, sottorappresentati o comunque non significativi, e perciò risultano impossibili da trattare statisticamente. La maggior parte degli IDP riporta invece *intensità* nell'intervallo compreso tra la soglia di “debolmente percepito” e la soglia di danneggiamento degli edifici, costituendo una sorta di “effetto ciambella” del campionamento attorno all'area di maggiore risentimento del terremoto (Bossu et al., 2018a). Analogamente ai gradi più elevati anche i gradi minimi si presentano in numero molto contenuto. Nel caso dei gradi più elevati la sottorappresentatività può essere spiegata sia in relazione alle priorità che gli utenti-testimoni hanno nelle circostanze di eventi con effetti dannosi, sia dalla legge di attenuazione del parametro d'*intensità*, la quale prevede per questi gradi una superficie di risentimento via via meno estesa, quindi una minore probabilità che in essa siano presenti centri urbani popolati; nel secondo caso invece la difficoltà nel

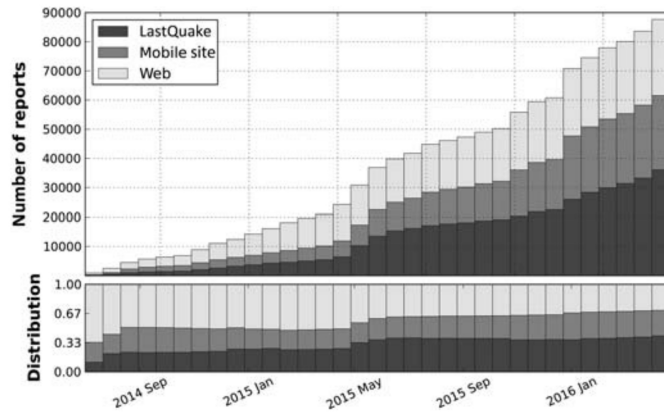


Figura 2.3: Time evolution of testimony collection through the classical website, mobile website, and smartphone app. The rate of collection has been increasing for these three channels over the whole period; the increase is faster for the mobile website and the app, which both use thumbnail-based questionnaires (Bossu et al., 2017). Testimonies for this histogram are collected all over the world.

percepire uno scuotimento estremamente debole – che non viene riconosciuto come evento sismico – tenderebbe appunto a non far riportare alcuna testimonianza (Bossu et al., 2017; Tosi et al., 2015).

Una considerazione particolare va fatta inoltre per le acquisizioni di *felt reports* che riportano basse *intensità*. Per queste, che si basano sui soli effetti transienti di scuotimento e non sui danni, lo studio di Sbarra et al., 2014 mette in luce che una valutazione oggettiva è più difficoltosa rispetto ai gradi più elevati. In particolare si osservano due aspetti della condizione dell’utente che influiscono sulla percezione del terremoto. Questi sono l’ubicazione (fuori dagli edifici, ai piani bassi, ai piani alti) e “cosa si sta facendo” (stato di quiete o di moto, di veglia o di sonno), e il secondo risulta più influente del primo. Se essi non sono opportunamente indagati dai questionari il bias nella valutazione dell’*intensità* può superare il mezzo grado. Notiamo che di essi non vi è menzione nei *thumbnail* dell’EMSC, mentre sono opportunamente indagati dai questionari di HSIT.

Infine per gli scopi dell’analisi degli IDPs è bene tenere in considerazione la possibilità della presenza di *outliers* nel dataset. Gli *outliers* sono definiti genericamente come record anomali e aberranti, cioè i cui parametri si discostano dai parametri degli altri record a tal punto da far pensare che non siano rappresentativi del fenomeno che si vuole misurare. Nel caso dei *felt reports* si possono distinguere due principali tipi di *outliers*:

- *outliers d’intensità*, e cioè records con *intensità* che differiscono notevolmente in valore (2-3 gradi) rispetto agli IDPs prossimi ad essi;
- *outliers spaziali*, che possono essere o IDPs localizzati a distanze di parecchie decine o centinaia di chilometri, improbabili a causa dell’attenuazione cui è soggetta l’*intensità*, oppure IDPs isolati che non possono essere utilizzati per la stima dell’*intensità* del loro luogo particolare per mancanza di rappresentatività statistica.

Gli *outliers* possono essere sia volontari, dovuti a scherzi degli utenti, che involontari, come esagerazioni o diaspora dei testimoni dal luogo di risentimento.

Per affrontare questa problematica possono essere applicate diverse convenzioni. Ad esempio si cita il lavoro di Tosi et al. del 2015 in cui si descrivono chiaramente le convenzioni applicate per ripulire le acquisizioni di HSIT. EMSC invece non applica alcun filtro sugli outliers.

Dalle caratteristiche dei *felt report crowdsourced web-based* esposte in questo paragrafo possono essere motivate le soglie che i diversi istituti di ricerca macrosismica pongono all'affidabilità degli IDPs sulla base del valore d'intensità ad essi associato e sulla capacità di questi di costituire MDPs di intensità effettivamente corrispondenti alle attribuzioni effettuate con i metodi macrosismici tradizionali.

Il sistema DYFI è calibrato per poter effettuare attribuzioni d'intensità della scala MMI fino al grado IX (D. J. Wald et al., 2011).

L'EMSC considera affidabili le sole intensità $I \leq 10$ (EMS), ponendo tale soglia come massimo per le intensità corrette I_{corr} ; tuttavia individua già $I = 7$ come prima soglia di riduzione dell'affidabilità (Bossu et al., 2017).

Il sistema di valutazione di HSIT (INGV), seguendo le valutazioni di Musson, accorpa le $I \geq 7$ in un'unica classe (R. Musson, 2007; Tosi et al., 2015).

Per ciò che riguarda invece le basse intensità (1, 2) tutte le precedenti istituzioni riportano una sottorappresentatività rispetto ai report attesi. In aggiunta INGV riscontra per i metodi di raccolta in questione l'indistinguibilità tra i primi due gradi, sia della scala EMS sia della MCS.

Capitolo 3

Tecniche di clustering spaziale per dati macrosismici di felt reports

In questo capitolo la tesi prosegue da un'introduzione generale sulla *cluster analysis* fino allo specifico problema del *clustering spaziale* dei dati di *felt reports*. Infatti, come mostrato al precedente capitolo, se da un lato i dati della piattaforma HSIT sono disponibili già con *intensità* stimate per località comunali, i dati di EMSC sono forniti in forma grezza e non aggregata (IDPs). Per questi ultimi è quindi necessario, per poterli utilizzare per la stima dei *parametri sintetici*, affrontare il problema della trasformazione degli IDPs in MDPs mediante opportune operazioni di irrobustimento statistico che verranno descritte di seguito.

3.1 Generalità della “cluster analysis spaziale”

Il crescente sviluppo delle tecnologie ha consentito di avere disponibilità di *raw data* di varia natura in quantità tali da rendere necessarie operazioni automatizzate di gestione degli stessi. Parallelamente hanno visto grande sviluppo i sistemi di *Data Mining* che hanno lo scopo di “estrarre conoscenza” da grandi moli di dati grezzi. Questi sistemi consistono in una serie di operazioni procedurali da effettuare sui dati volte ad espletare funzioni di campionamento, di ordinamento, di individuazione di regole. Le funzioni espletate da un sistema possono essere realizzate mediante l'uso di diversi metodi o tecniche. Ciascuno dei metodi definisce degli algoritmi, e cioè il dettaglio delle operazioni da effettuare sui dati (Milone, 2007).

La **cluster analysis** è una funzionalità del data mining che ha l'obiettivo di raggruppare i record di un dataset in gruppi detti *clusters* in base ad opportuni criteri di *similarità* tra i record. L'operazione di clustering propriamente detta è un' “operazione di apprendimento non supervisionato” e cioè, dato un record del dataset, non è possibile associarlo (classificarlo) ad uno dei gruppi di un insieme noto a priori (classi) senza riassetare l'intero raggruppamento; in altre parole può essere pensata come una classificazione di cui non si conoscono a priori le classi (Milone, 2007).

L'idea di base nell'effettuazione di un clustering è che gli elementi (o oggetti, o pattern, o record) di un cluster siano “più simili” tra loro “che simili” agli elementi appartenenti ad altri cluster. È cruciale dunque il concetto di similarità tra gli elementi, il quale viene definito in relazione al tipo di dato che si vuole trattare e agli obiettivi che si vogliono raggiungere. Nella pratica definire la *similarità* tra gli oggetti

significa definire una nozione di distanza tra i vettori descrittivi degli oggetti del dataset. Diverse definizioni di distanza e diversi criteri per stabilire se alcuni oggetti possono essere considerati “sufficientemente vicini (simili)” per formare un cluster e se altri possono essere considerati “sufficientemente lontani (dissimili)” per non esserne inclusi definiscono le diverse *tecniche di clustering* (Milone, 2007; Agostinelli et al., 2018).

Quindi il tipo di dato da trattare influisce sulla scelta delle tecniche da utilizzare e ciò ha portato alla specializzazione di tecniche generiche per particolari domini di dati. Tra i vari domini sono particolarmente rilevanti quelli di *spazio* e *tempo* che spesso caratterizzano inerentemente la raccolta di un qualsiasi altro tipo di dati. Altri esempi di dominio sono il *booleano*, le *stinghe di caratteri*, le *variabili ordinali* – come l’intensità macrosismica –, le *variabili razionali* – che naturalmente includono il dominio spaziale, ma possono rappresentare altre grandezze fisiche, parametri economici, ecc (Everitt, 2011).

Per gli scopi di questa tesi si circoscrive la trattazione al dominio della **spatial clustering analysis**, la quale riguarda le tecniche di elaborazione di dati di tipo spaziale. I dati di tipo spaziale includono tra le componenti (o attributi o caratteristiche) dei vettori descrittivi degli oggetti informazioni sullo spazio, come la latitudine, la longitudine e l’altitudine (Ansari et al., 2020). La *spatial clustering analysis* si basa su principi del tipo della Prima legge di Tobler della Geografia che afferma che “tutto è in relazione reciproca, ma che le cose vicine sono maggiormente in relazione tra loro rispetto alle cose lontane” (Tobler, 1979). Con questa idea di base si cerca dunque di agevolare l’operazione di estrazione dal dataset della conoscenza sugli attributi non spaziali e alla loro relazione con quelli spaziali. I clusters che risultano dall’operazione di raggruppamento possono essere utilizzati per diversi scopi, tra i quali si menzionano la classificazione, la predizione, ulteriori analisi sui clusters, processi di “decision making” e la semplice visualizzazione del dataset mediante la costruzione di mappe (Joshi, 2011).

Dalla letteratura relativa alle diverse *tecniche di clustering spaziale* si evince che esse possono risultare più o meno adeguate ed efficienti nella trattazione di dataset in base alle assunzioni su di essi, alla loro estensione, alla potenza computazionale di cui si dispone e alla presenza di *outliers*¹. In particolare gli *outliers*, se inclusi indiscriminatamente nei clusters, possono causare una clusterizzazione debolmente rappresentativa dell’informazione che si vuole estrapolare dai dati. Tuttavia l’individuazione di diversi *approcci e tecniche al clustering spaziale* non può prescindere da una suddivisione più generale delle tecniche di clustering, indipendentemente dal tipo di dato da trattare. Si propone qui una suddivisione generale basata su *sei grandi tipologie*, che permette di individuare speditivamente quale tra queste possa essere di maggior interesse (Han et al., 2011):

- partizionanti,
- gerarchiche,
- fuzzy,

¹per *outliers* si intendono quei record anomali che si discostano talmente tanto dalle altre osservazioni da far pensare che siano stati generati da un processo sostanzialmente diverso da quello generale. Per le caratteristiche specifiche degli *outliers* nell’ambito dei felt reports si rimanda al paragrafo 2.2

- basate sulla densità,
- basate sui modelli,
- basate sulle griglie.

Le *tecniche partizionanti* suddividono i dati in raggruppamenti in modo tale che un dato appartenga ad uno ed un solo cluster e che ciascun cluster contenga almeno un dato. I cluster sono ad intersezione nulla. Gli algoritmi che fanno parte di questa tipologia, seppur con variazioni di dettaglio, si basano tutti su un'inizializzazione che prevede la selezione casuale di k centroidi, oggetti rappresentativi di k cluster, e l'assegnazione degli altri punti in base a criteri di minimizzazione dei parametri di distanza dai centri o di varianza. Iterativamente si procede a ricalcolare i centroidi e riassegnare i punti fino a convergenza. Questo tipo di tecniche necessitano della conoscenza a priori del numero k di cluster, che le influenza fortemente; sono inoltre dipendenti dalla scelta iniziale dei centri e non gestiscono bene gli outliers (Sabeti e Javaherian, 2009; Rehman et al., 2014; Amorese et al., 2015).

Il *clustering gerarchico* consiste invece nel raggruppare i dati in una struttura a dendrogramma in cui ciascun dato fa parte di cluster via via più grandi (o più piccoli, a seconda se si proceda con un approccio bottom-up o top-down), dal cluster di singolo record al cluster che raggruppa l'intero dataset (o viceversa). Per i raggruppamenti (o separazioni) si procede iterativamente mediante opportune nozioni di distanza. La scelta finale del livello del dendrogramma corrispondente a un particolare raggruppamento è a discrezione dell'operatore, e ciò risulta un importante difetto della tecnica. Un ulteriore difetto del metodo è costituito dall'intrinseca sensibilità agli outlier. In generale gli algoritmi di *clustering gerarchico* si distinguono nettamente da tutti gli altri tipi che invece costruiscono clusters tutti dello stesso livello e sono perciò detti *piatti* (Czeczka e Bondar, 2019).

La logica delle *tecniche di fuzzy clustering*, dette anche di “soft clustering” risiede nella possibilità di attribuire ad un singolo record un'appartenenza multipla a diversi clusters, sulla base di opportune “funzioni di appartenenza”. Quest'approccio sottende l'assunzione di un'intrinseca imprecisione dei dati di cui si dispone e produce dei raggruppamenti, perciò detti “sfocati”, che la rispecchiano. Gli algoritmi di questo tipo necessitano di essere inizializzati su un set di centroidi rispetto ai quali viene assegnato un certo grado di appartenenza di ciascun record. Vengono definite, per un record, delle *funzioni di appartenenza* che rappresentano la probabilità che il record appartenga ai vari cluster sulla base di un'opportuna definizione di distanza dai centroidi. L'algoritmo può prevedere il ricalcolo dei centroidi e delle funzioni di appartenenza iterativamente, fino a convergenza dei centroidi (Anzari et al., 2015).

Le *tecniche di density-based clustering* formano i clusters individuandoli tra regioni del dataset ad elevata densità di record separate tra loro da regioni a più bassa densità. La nozione di densità è vincolata alla scelta di parametri iniziali di raggio (distanza) attorno ad un record e di numero minimo di punti in esso inclusi per poter iniziare a formare un cluster. L'inizializzazione avviene con un singolo record: dopo aver individuato i record sufficientemente densi attorno ad esso e costruito il primo cluster si procede allo stesso modo serialmente sui record non ancora assegnati. Questo tipo di algoritmo permette di individuare clusters di forma arbitraria e ha un'intrinseca buona gestione degli outliers (Ester et al., 1996; Georgoulas et al., 2013).

Il *model-based clustering* si fonda sull'assunzione che i dati appartengano ad un'insieme di distribuzioni di probabilità e che ciascun cluster sia una componente di tale insieme, dunque modellizzato da una particolare distribuzione di probabilità. Tali algoritmi necessitano pertanto di conoscere esplicitamente il modello da cui è verosimilmente estratto il dataset. Essi possono essere intesi come metodi di *soft clustering*, che costruiscono cioè clusters con assegnazioni non nette dei record, ma guidate appunto da probabilità (Fraley e Raftery, 1998).

L'ultimo approccio, noto come *grid-based*, invece di essere incentrato sulle caratteristiche dei record e le loro relazioni di similarità è basato sullo spazio stesso. Consiste cioè nel selezionare l'intero dominio del dataset e suddividerlo in celle mediante la costruzione di una griglia equispaziata. È possibile quindi calcolare parametri statistici empirici per ciascuna cella della griglia dai record che in essa si trovano automaticamente raggruppati. Questa operazione costituisce in pratica una partizione dello spazio del dataset, essa può tuttavia essere eseguita costruendo una sorta di gerarchia di griglie in modo tale che le statistiche delle celle a risoluzione superiore permettano di calcolare quelle a risoluzione inferiore: fissato il passo di griglia a risoluzione massima (passo minimo) si assegnano i record alle celle e si calcolano i parametri statistici rappresentativi di ciascuna; al risultato si applica quindi una seconda griglia con passo superiore per le cui celle le statistiche sono calcolate dalle statistiche ottenute all'iterazione precedente. In generale i metodi *grid-based* sono per lo più intuitivi per la logica di funzionamento, ma fortemente soggettivi per la scelta dei parametri. Infatti i clusters (celle) risultanti dipendono significativamente dai parametri di griglia (passo, forma, orientazione) e risentono dell'influenza degli outliers. Di fatto, scelti i parametri di griglia, l'operazione di raggruppamento è immediata, quindi *supervisionata* e più simile ad una *classificazione* (Wesseloo et al., 2014).

I diversi approcci al clustering qui esposti possono essere sviluppati e intrecciati variamente. Ciò da origine a moltissime tecniche ed algoritmi compositi e via via maggiormente specializzati ed appropriati a seconda delle diverse caratteristiche dei datasets che di volta in volta si vogliono trattare, ma anche degli obiettivi e della potenza di calcolo a disposizione. In letteratura sono accuratamente descritti molti algoritmi e la loro implementazione in diversi contesti (Milone, 2007; Manning et al., 2008) la cui trattazione esula però dal presente lavoro. Per i nostri scopi si rimandano al paragrafo successivo le considerazioni per la scelta delle tecniche da utilizzare.

3.2 Scelta delle tecniche di clustering spaziale da utilizzare per i dati EMSC

Per ciò che riguarda il carico computazionale delle tecniche di clustering i datasets utilizzati per questa tesi non hanno dimensioni tali da costituire un problema e condizionare perciò la scelta degli algoritmi da utilizzare. Si tratta infatti di *set* che vanno da poche centinaia a poche migliaia di *oggetti* (IDPs), ciascuno dei quali è descritto da 3 sole *caratteristiche* (latitudine, longitudine, intensità). Questi numeri rispettano ampiamente i consigli di G. Milone (Milone, 2007), la quale suggerisce di trattare vettori (*oggetti*) composti mediamente da 1480 pattern (*caratteristiche*).

Dalla clusterizzazione degli IDPs ci si aspetta di “estrarre conoscenza” sui livelli di severità dello scuotimento avverito conformemente alle interpretazioni del parametro di *intensità* delle scale macrosismiche. Si vuole quindi che ciascun cluster rappresenti una località pertinente alla definizione di intensità e che questo sia a sua volta rappresentato da un solo *Macroseismic Data Point* (MDP) le cui *caratteristiche* sono calcolate dalle *caratteristiche* dei soli IDPs del cluster utilizzando opportuni indici statistici (media, mediana, ecc.). Si vuole dunque in primo luogo un’assegnazione univoca di ciascun IDP ad un cluster; in secondo luogo cercare di raggruppare in cluster tutti gli IDPs “sufficientemente vicini” ed in “numero sufficiente” per costruire statistiche robuste che siano espressione dell’intensità del luogo, senza eccessiva perdita di risoluzione e senza assunzioni a-priori sulla distribuzione dei dati. Si assume inoltre la presenza nel dataset di outliers che potrebbero negativamente inficiare le statistiche dei singoli clusters, soprattutto per quei clusters che risultano poco popolati (con pochi IDPs).

Questa idea porta immediatamente a scartare sia le tecniche di *clustering gerarchico*, per la loro struttura a dendrogramma e l’ingestibilità automatica degli outliers, sia le *tecniche fuzzy*, per le quali un IDP sarebbe condiviso da più cluster ² e la cui implementazione sarebbe di una complessità tale da esulare dalle risorse disponibili per questa tesi.

Un’ulteriore rapida considerazione può essere fatta sui *metodi model-based* il cui utilizzo sui nostri datasets sarebbe inappropriato: infatti, sebbene si possa disporre di una legge empirica di attenuazione dell’intensità con la distanza epicentrale, non esistono modelli consolidati che riproducano localmente il campionamento delle intensità per un terremoto; infatti l’intensità dipende anche dalla vulnerabilità (qualità) delle costruzioni e dagli effetti di sito, che sono variabili non facilmente modellabili. Non è perciò possibile introdurre forti assunzioni sulla distribuzione dei dati, quindi, per i nostri scopi, si scartano anche i metodi model-based.

Per ciò che riguarda invece i *metodi partizionanti* si può notare intanto che assolvono alla richiesta di associare ciascun’osservazione ad un solo cluster, adeguata per i dataset di IDPs, tuttavia essi sono fortemente dipendenti dal numero di clusters di inizializzazione, che nel nostro caso non è noto a-priori; l’unico modo sarebbe sviluppare clustering con differenti inizializzazioni per poi confrontarli e valutare tra essi il migliore. Ad esempio un’inizializzazione ragionevole potrebbe essere effettuata su centroidi corrispondenti alle coordinate di riferimento delle località amministrati-

²Le tecniche fuzzy sono tuttavia efficacemente utilizzate quando l’informazione originaria sull’intensità di un luogo è suddivisa in diagnostici (singoli effetti sugli “oggetti sensori”) che sono contemporaneamente espressione di diversi gradi di una scala (Vannucci et al., 2015)

ve della regione del terremoto, per poi unire o eliminare i clusters che riportano un numero di IDPs troppo basso per un'adeguata statistica. Inoltre, quanto alla risoluzione dei metodi partizionanti applicati a dati macrosismici, sembra che si riescano ad ottenere clusters con la massima risoluzione possibile (Amorese et al., 2015; Hough, 2013). Si riporta qui un riferimento allo studio di Amorese et al. del 2015 per il quale è stata implementata efficacemente una tecnica di clustering partizionante, nota come PAM (Partition Around Medoids), la quale ha permesso di ottenere clusters da dati sintetici di felt reports sia con una risoluzione maggiore rispetto ai metodi più elementari *grid-based*, sia con maggiore accuratezza della stima delle intensità degli MDPs. Il confronto tra le due tecniche ha però portato agli stessi risultati in maniera indipendente dalla densità di popolazione. La tecnica PAM affronta implicitamente anche il problema della sensibilità agli outliers.

In ultima analisi la scelta dei tipi di metodo di clustering da utilizzare per il presente lavoro di tesi ricade sui metodi *grid-based* e *density-based*. Di seguito le ragioni e ulteriori considerazioni sui metodi.

- I metodi *grid-based* forniscono raggruppamenti elementari e speditivi che possono fungere da confronto con clusterizzazioni effettuate con metodi alternativi come i *partizionanti* (Amorese et al., 2015) o i *density-based*. In aggiunta la velocità di calcolo non dipende dal numero di pattern ma dal numero di celle (Milone, 2007). Quest'ultima caratteristica tuttavia può rendere più o meno rapido il metodo a seconda dell'estensione dei datasets che verranno trattati in questo studio. Per l'applicazione di queste tecniche ci si attende inoltre che i clusters siano dipendenti dalle geometrie di griglia, oltre che dalla posizione e dall'orientazione della griglia stessa. Infatti la geometria delle celle, verosimilmente, non rispecchia né la forma dei paesi e delle città, né risulta essa legata in alcun modo a caratteristiche geometriche della distribuzione spaziale empirica dell'*intensità*. Per questo motivo la scelta della griglia rimane arbitraria e l'efficacia potrà essere valutata a-posteriori in base alla stabilità dei risultati. Per questo studio si sceglie di utilizzare due tipi di **griglia poligonale**: una a **celle quadrate** e l'altra a **celle esagonali**. È inoltre da tenere in considerazione che la risoluzione spaziale di questo tipo di metodo è direttamente vincolata al passo della griglia (Helmstetter et al., 2007; Wesseloo et al., 2014).
- I metodi *density-based* (DB) sono scelti in primo luogo per cercare di costruire clusters "più naturali possibile" e cioè rispondenti alla distribuzione empirica dei dati. L'idea è che un approccio basato sulla densità possa essere in grado di individuare clusters relativi a luoghi anche piccoli purché densamente campionati, quindi attribuendo loro un'*intensità* propria, distinta da quella di eventuali centri abitati maggiori limitrofi, in genere più campionati: l'individuazione automatica della regione a bassa densità tra i due luoghi da parte di un algoritmo DB ne consentirebbe la discriminazione in due differenti clusters. In questo modo si renderebbe il clustering meno affetto dall'estensione delle località campionate (Amorese et al., 2015). Un importante pregio dei DB più comuni è di svincolare gli algoritmi di clusterizzazione dalle inizializzazioni artificiali dei clusters, infatti essi sono inizializzati ad un singolo record e procedono seguendo la sola nozione di densità (Milone, 2007). In secondo luogo essi sono tali da rimuovere il vincolo sulla geometria dei clusters imposto da nozioni di distanza

spaziale o da parametri di griglia. Ci si aspetta dunque che la forma dei clusters rispecchi la forma dei luoghi – se ben campionati –, qualsiasi essa sia, purché sia rispettato il vincolo sulla densità. Il vincolo della densità dei DB è espresso mediante due parametri fortemente caratterizzanti il risultato del clustering, determinati a discrezione dell’operatore: numero minimo di punti *MinPts* presenti in un’area circolare di raggio *eps* che sono necessari per iniziare a formare un cluster. Infine con questo tipo di tecnica gli outliers spaziali, e cioè gli IDPs isolati dagli agglomerati densi di campionamenti, per i quali non è possibile effettuare un’adeguata statistica che consolidi la stima d’*intensità*, sono automaticamente individuati e non clusterizzati. Tuttavia se essi si presentano in gran numero si può cercare di riassegnarli con i metodi *grid-based* (Nanda e Panda, 2015; Ester et al., 1996).

In sintesi le tecniche che sono state applicate sono cinque: *due grid-based* (una a **celle quadrate** e una a **celle esagonali**); *due density-based* (un raggruppamento in **cerchi costruiti attorno alle località** e una tecnica basata su un elementare **algoritmo DBSCAN**) e *una mista* (che consiste nell’uso sequenziale del **metodo dei cerchi** e della **griglia di esagoni**).

Si vuole qui evidenziare la differenza principale nell’approccio al clustering tra la tecnica basata sull’algoritmo DBSCAN e le altre 4 tecniche: la DBSCAN è *non supervisionata*, cioè costruisce clusters empirici, basati solo sul dataset in elaborazione; le altre sono inecce *supervisionate*, cioè costruiscono clusters rigidi in base ai parametri a-priori che le definiscono. Inoltre in particolare le due tecniche density-based si differenziano tra loro per il fatto che la tecnica dei cerchi si basa sulla densità di abitanti – non di IDPs – di specifiche località, il che costituisce un’assunzione a priori sulla distribuzione dei dati, mentre la seconda è basata sulla densità stessa dei dati acquisiti, a meno di una soglia limite. Per ulteriori dettagli si vedano i paragrafi relativi alle singole tecniche.

In Figura 3.1 si mostra uno schema riassuntivo delle tecniche applicate in questo lavoro di tesi per la trasformazione dei datasets di IDPs in distribuzioni di MDPs da utilizzare con il programma Boxer.

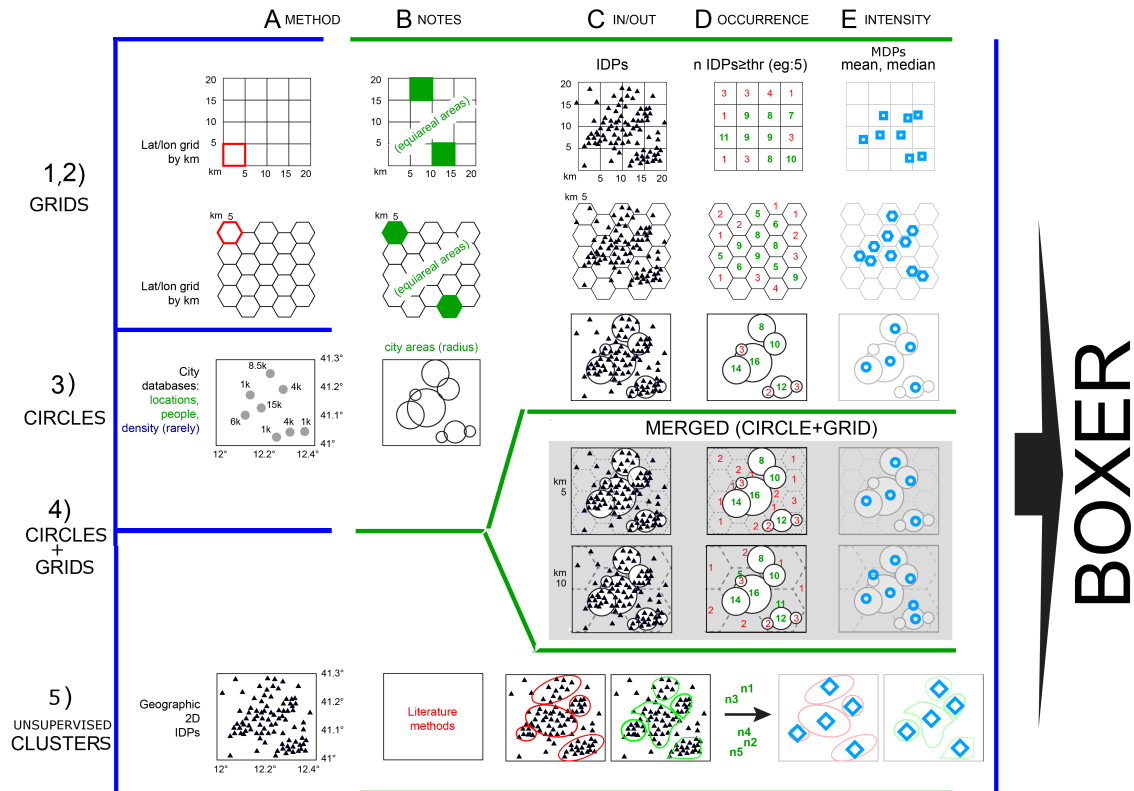


Figura 3.1: Sintesi delle tecniche sviluppate in questa tesi per la trasformazione di IDPs in MDPs ai fini dell’elaborazione con Boxer. Nelle colonne A e B i metodi di clustering e le rispettive note caratteristiche. Le tecniche di griglia sono tali da avere celle equiareali in una stessa griglia. I cerchi sono costruiti con raggio dipendente dalla densità di abitanti, a partire da un database di località. Il cluster non supervisionato è eseguito con il metodo DBSCAN basato sulla densità empirica e tratto dalla letteratura (par.3.7). Le colonne C, D, E mostrano il processo di costruzione degli MDPs (E) a partire dalla distribuzione spaziale degli IDPs (C) selezionando tra i clusters quelli formati da un adeguato numero di IDPs (D). Per ulteriori dettagli si vedano i paragrafi relativi alle singole tecniche.

3.3 Tecnica 1: griglia di quadrati equiareali

Questa tecnica di clustering *grid-based* fornisce una divisione dello spazio detto *spazio delle caratteristiche* in celle multidimensionali (iperprismi) entro cui raggruppare gli *oggetti* (IDPs) componenti il dataset. Questa divisione prende il nome di *tassellazione*. Per i nostri scopi viene effettuato un clustering spaziale, perciò la tecnica viene applicata solamente alle *caratteristiche* di latitudine e longitudine degli IDPs, quindi su uno spazio bidimensionale; perciò gli iperprismi sono di fatto dei poligoni e in particolare si scelgono i quadrati. Inoltre si sceglie di non costruire una gerarchia di griglie annidate a diversa risoluzione ma di costruire solamente griglie di singolo

livello le quali generano clusters tutti dello stesso livello; l'algoritmo è perciò definibile come *piatto* (Milone, 2007; Manning et al., 2008).

Questo tipo di tecnica non ha la necessità di un'esplicita definizione di *similarità* tra i *pattern* da raggruppare poiché la definizione della griglia, determinata dal passo p , dal *punto origine* per l'inizializzazione della procedura di costruzione e dalla sua orientazione, stabilisce automaticamente gli IDPs che saranno raggruppati in ciascuna cella.

In Appendice A si riporta l'esatta procedura utilizzata per la costruzione della griglia di quadrati equiareali, con passo variabile, utilizzata nel presente lavoro.

Come parametro di griglia è stato utilizzato un passo $p = 10Km$ corrispondente ad un'area nominale della superficie di singola cella di $100Km^2$ (con incertezza dello 0.03%, vedi Appendice A), quindi più estesa rispetto alla definizione di località di EMS-98³, alla luce del fatto che i datasets di EMSC a nostra disposizione non hanno molte risposte (IDPs), le quali inoltre, raccolte su base volontaria, sono irregolari e casuali (e.g. Figura 3.2). L'idea è che un parametro più largo possa fugere da soglia base per un confronto di questa tecnica speditiva con altre tecniche più sofisticate; ed in particolare che un parametro così scelto permetta la costruzione di un accettabile numero minimo di MDPs seppur con il rischio di perdita di risoluzione in corrispondenza dei centri abitati maggiormente campionati con IDPs. Il passo di griglia è stato uniformato per tutte le griglie di quadrati costruite per l'elaborazione dei terremoti analizzati in questa tesi: scelta effettuata per finalità speditive e per non incorrere nel rischio di overfitting.

Per l'assegnazione degli IDPs del dataset alle celle della griglia è stata utilizzata la routine di Matlab "inpolygon" (https://it.mathworks.com/help/matlab/ref/inpolygon.html?s_tid=doc_ta) che confronta direttamente coordinate dei vertici di poligono con coordinate di singolo IDP. Vengono dunque memorizzate le celle che contengono almeno un IDP: ciascuna di queste definirà uno specifico cluster.

Per ragioni di significatività l'analisi statistica viene effettuata per quei soli clusters che contengono un numero di IDPs non inferiore a 5. Questa scelta è in linea con la convenzione adottata anche da altri per poter considerare affidabile e stabile l'attribuzione dell'*intensità* ad un *Macroseismic Data Point* (MDP) a partire da informazioni macrosismiche raccolte con i questionari estesi (D. Wald et al., 1999; D. J. Wald et al., 2011; Mazet-Roux et al., 2010; Tosi et al., 2015). I questionari estesi HSIT indagano i *diagnostici*⁴ dei gradi della scala macrosismica con maggior dettaglio rispetto ai *thumbnails* di EMSC, i quali invece forniscono per costruzione un'informazione meno precisa quindi più facilmente soggetta ad attribuzioni erranee di *intensità* (par.2.2); di conseguenza in linea di principio, per i *thumbnails* il numero minimo di IDPs necessario per costruire un MDP aumenterebbe. Tuttavia al momento non sono state effettuate analisi statistiche per individuare una soglia preferibile per i *thumbnails*.

Per i clusters individuati dalla griglia sono quindi calcolati, per le 3 *caratteristiche* (lat, lon, int), quattro differenti valori statistici che rappresentano il rispettivo centroide (MDP): *media*, *mediana* e *medie* decimate usando il sub-dataset compreso negli intervalli percentili 15°-85° e 25°-75°, in seguito indicate come media 15/85 e 25/75.

³“superficie geografica che abbia l'estensione al massimo di una cittadina europea di medie dimensioni e al minimo di un piccolo villaggio” (Grunthal, 1998).

⁴effetti macrosismici, definiti al par. 1.3.1 (Grunthal, 1998)

Sono inoltre calcolati per ogni cluster *deviazione standard della media del campione* σ e relativo *errore standard* come:

$$errstd = \frac{\sigma}{\sqrt{N}}, \quad \sigma = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (A_i - \mu)^2}, \quad \mu = \frac{1}{N} \sum_{i=1}^N A_i \quad (3.1)$$

dove A_i è il vettore delle caratteristiche del cluster (lat o lon o int), N è il numero degli elementi di A_i (#IDPs), μ è la media aritmetica.

In Figura 3.3 è mostrato un esempio di MDPs costruiti con la tecnica delle celle quadrate a partire dagli IDPs dello stesso terremoto di Figura 3.2. Si osserva come nonostante la scarsità di IDPs a disposizione la tecnica riesca a trovare clusters accettabili secondo i criteri stabiliti (#MDPs = 49); di questi MDPs circa il 70% ha un numero di IDPs costituenti inferiore alla decina.

Si osserva inoltre la suddivisione della città di Roma (ellisse verde, Figura 3.3), in 6 principali MDPs, ciascuno composto da un numero di IDPs che va da qualche decina al centinaio; in particolare 1 dei 6 è costituito da #IDPs = 155, mentre gli altri 5 sono formati al più da #IDPs = 30. La rigidità della griglia causa la separazione di agglomerati più densi di IDPs, parte dei quali viene scartata o inclusa in un diverso raggruppamento adiacente (ellissi rosse); questo si manifesta quindi nell'assegnazione degli indicatori rappresentativi degli MDPs (media, mediana, ecc.) in corrispondenza dei margini delle celle.

Infine per alcuni MDPs non vengono calcolate le medie decimate 15/85 e 25/75 poiché la base dati non raggiunge il numero minimo di IDPs stabilito.

Relativamente agli *errori standard* degli MDPs si considera la proporzionalità inversa con la radice del numero degli IDPs costituenti e si riporta che, per quegli MDPs con #IDPs ≥ 10 gli *errori* sono risultati dell'ordine del millesimo di grado per le coordinate e al più di 3 decimi di unità per le *intensità*; mentre per gli MDPs con #IDPs < 10 l'*errore standard* sulle coordinate rimane per lo più dello stesso ordine del caso precedente, solo per #IDPs = 5-7 si registra talvolta un *errore* sul centesimo di grado, invece quello sulle *intensità* supera il mezzo grado pur rimanendo inferiore all'unità.

Si evidenzia infine che, per ragioni legate alla formulazione della *scala macrosismica*, le *intensità* degli MDPs sono state tutte arrotondate al mezzo grado più vicino.

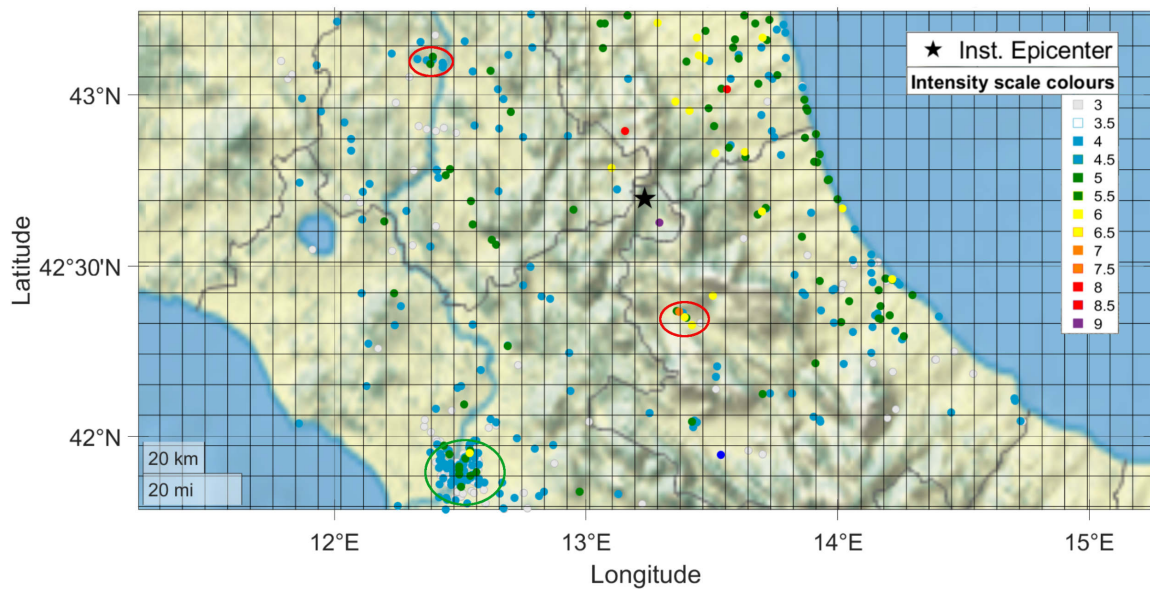


Figura 3.2: Terremoto 24/08/2016. IDPs (I_{raw} , EMSC) su griglia di quadrati equiareali di passo $p = 10Km$. La griglia è deformata poichè visualizzata con una proiezione cartografica non-equiareale. Vedere testo per ulteriori dettagli.

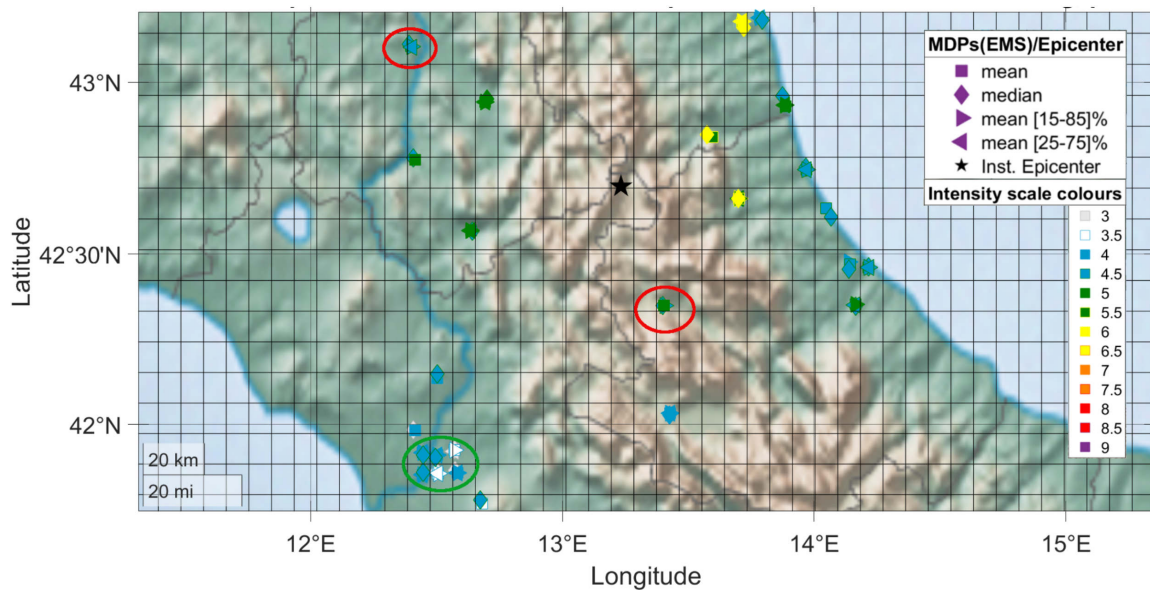


Figura 3.3: Terremoto 24/08/2016. MDPs (da I_{corr} EMSC) costruiti con la tecnica di clustering griglia di quadrati equiareali di passo $p = 10Km$. La griglia è deformata poichè visualizzata con una proiezione cartografica non-equiareale. Vedere testo per ulteriori dettagli.

3.4 Tecnica 2: griglia di esagoni equiareali

Questa tecnica di clustering è analoga alla precedente, con la sola differenza nella geometria e nella dimensione delle celle utilizzate per la tassellazione dello spazio. La forma esagonale delle celle riduce in linea di principio l'artificio della partizione dello spazio con la forma quadrata e approssima meglio la forma equidistanziata del cerchio.

La griglia di esagoni equiareali è dunque una tecnica *grid based*, partizionante, effettuata sulle due *caratteristiche* spaziali del dataset, con celle tutte dello stesso livello, quindi *piatta*. In questo caso il parametro di riferimento della griglia è il lato di singola cella, che equivale alla distanza r tra centro e vertici dell'esagono. Anche qui, in assenza di strategie di ottimizzazione del parametro di griglia e data l'intenzione di applicare una tecnica speditiva che costruisca MDPs anche con datasets di IDPs poco densi e poco numerosi, è stato impostato un raggio $r = 5Km$ uniforme per tutte le griglie esagonali utili all'elaborazione dei terremoti studiati per questa tesi; a questo parametro corrispondono esagoni equiareali di area nominale di circa $65Km^2$.

Per la costruzione di questa griglia si è seguita una procedura analoga a quella utilizzata per la griglia di quadrati. I dettagli sono riportati in Appendice A.

Anche in questo caso l'assegnazione degli IDPs alle celle è effettuata tramite confronto diretto tra coordinate di IDPs e coordinate dei vertici di cella con la funzione "inpolygon" di Matlab.

Per la costruzione degli MDPs si considerano, per ragioni statistiche, i soli esagoni a cui sono stati assegnati almeno 5 IDPs, come indicato al par.3.3. Vengono dunque caratterizzati gli MDPs relativi a tali esagoni con gli stessi indicatori (media, mediana, ecc.) utilizzati per la tecnica dei quadrati (par. 3.3).

Come per la tecnica dei quadrati, osservando le mappe riportate nelle Figure 3.4 e 3.5 ottenute dai dati dello stesso terremoto di cui alle Figure 3.2 e 3.3, appare evidente come nonostante la scarsità di IDPs a disposizione la tecnica degli esagoni riesca a trovare clusters accettabili secondo i criteri stabiliti ($\#MDPs = 47$); anche qui circa il 70% degli MDPs ha un numero di IDPs inferiore alla decina; inoltre gli *errori standard* delle *caratteristiche* degli MDPs presentano le stesse peculiarità descritte al paragrafo precedente per la tecnica di quadrati.

Se si confrontano le mappe degli MDPs costruiti con le due diverse griglie (Figure 3.3 e 3.5) si osserva una discreta sovrapposibilità delle coordinate degli MDPs con la presenza di alcune fluttuazioni, inoltre con la griglia di esagoni si ottiene un numero di MDPs inferiore di sole due unità rispetto a quello ottenuto con la griglia di quadrati.

Un altro aspetto da considerare è che per questa tecnica l'area nominale delle celle è inferiore all'area nominale dei quadrati ($100Km$) della tecnica precedente, quindi in linea di principio la griglia di esagoni qui costruita avrebbe maggior risoluzione della griglia di quadrati. Tuttavia questo non si verifica in generale a causa dei rigidi vincoli della griglia (par.3.3). Infatti confrontando la mappa degli MDPs costruiti con le due diverse griglie (Figure 3.3 e 3.5) si osserva per la griglia di esagoni un effettivo aumento di risoluzione nella zona ben campionata della capitale (7 MDPs per la griglia di esagoni e 6 MDPs per l'altra), mentre nelle zone in cui densità e numero di IDPs sono inferiori si osserva l'assenza (o la presenza) di MDPs che la griglia di quadrati aveva (non aveva) invece costruito. Questa fluttuazione nella costruzione degli MDPs sembra tuttavia dipendere dall'arbitrarietà della posizione

e dell'orientazione della griglia, che può rigidamente separare IDPs vicini in celle adiacenti, piuttosto che dallo scarto di risoluzione delle due griglie (Figure 3.2, 3.4).

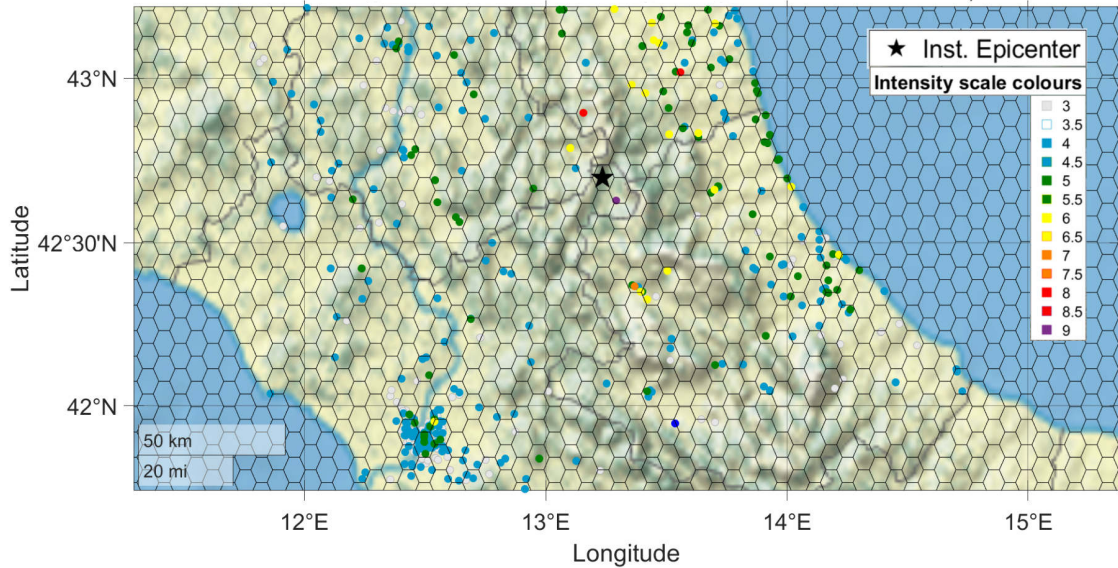


Figura 3.4: Terremoto 24/08/2016. IDPs (I_{raw} , EMSC) su griglia di esagoni equiareali di lato $r = 5Km$. La griglia è deformata poichè visualizzata con una proiezione cartografica non-equiareale. Vedere testo per ulteriori dettagli.

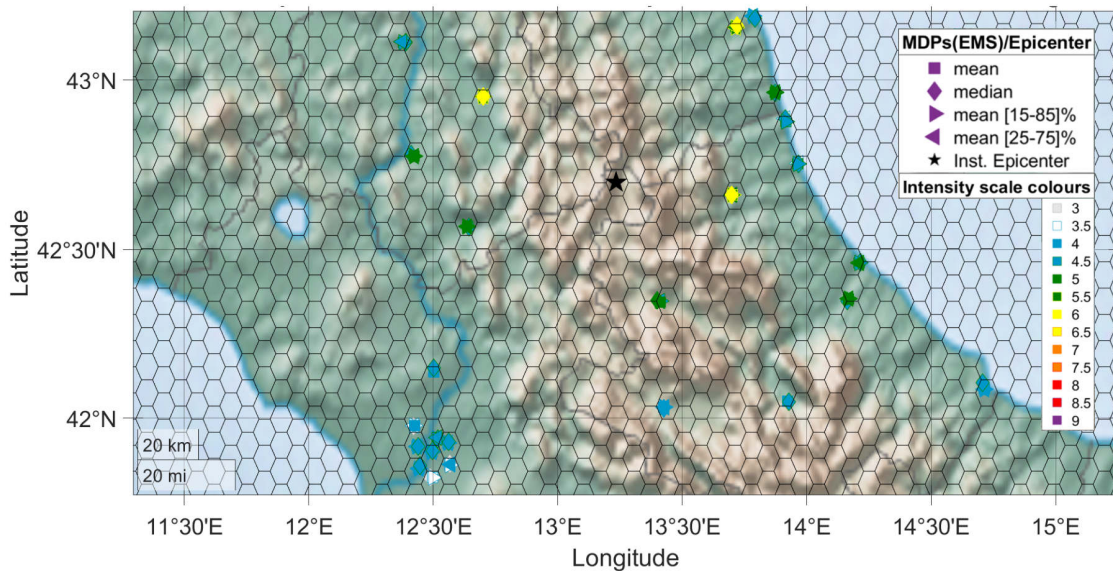


Figura 3.5: Terremoto 24/08/2016. MDPs (da I_{corr} EMSC) costruiti con la tecnica di clustering griglia di esagoni equiareali di lato $r = 5Km$. La griglia è deformata poichè visualizzata con una proiezione cartografica non-equiareale. Vedere testo per ulteriori dettagli.

3.5 Tecnica 3: cerchi attorno a località

Questa tecnica di raggruppamento spaziale di dati costituisce un ulteriore approccio speditivo per la costruzione di MDPs, in alternativa alle tecniche di quadrati ed esagoni con cui confrontare i risultati ottenuti. Questa, come le tecniche di griglia, è basata sullo spazio, ma in aggiunta introduce una “*dipendenza dello spazio da una densità a-priori*”. Con quest’ultima espressione ci si riferisce all’uso, per la determinazione dell’estensione spaziale dei potenziali clusters, di un parametro di densità di abitanti residenti, per l’appunto *a-priori* rispetto alla densità dei punti (IDPs) del dataset. Inoltre anche per questa tecnica come per le *grid-based* l’operazione di associazione di un IDP ad un cluster, noti a-priori centri ed estensione dei cerchi, è supervisionata, quindi più simile ad una classificazione che ad un clustering propriamente detto (definizione al par.3.1).

In pratica utilizzando il database “*cities500.zip*” (<https://www.geonames.org/>), relativo a località di tutto il mondo con numero di abitanti maggiore di 500, vengono definite aree circolari centrate sulle località amministrative del territorio che ha risentito del terremoto. Si costituiscono perciò clusters rigidi, di *area determinata a-priori*, tutti dello stesso livello, localizzati sui centroidi delle aree amministrative. L’area di ciascun cerchio è determinata in base alla semplice proporzionalità con il parametro di densità di popolazione residente $dens = pop/\pi R_{city}^2$, assunto uniforme su tutto il territorio. Il valore di densità imposto è $dens = 3500ab/Km^2$; è scelto arbitrariamente e uniformemente per tutti i datasets EMSC processati in questa tesi.

La definizione dei clusters con questa tecnica comporta che: non tutta la superficie di risentimento è racchiusa in un qualche cerchio, quindi alcuni IDPs saranno a-priori estromessi da qualunque cluster e si avrà perdita di dati (Figura 3.6); i cerchi corrispondenti a località limitrofe e prossime possono avere intersezione non nulla e, in particolare, cerchi di località più piccole possono essere interamente inclusi nei cerchi di più grandi città vicine (Figura 3.6), causando sovrapposizione dei clusters. Nel primo caso una certa quantità a-priori di *outliers*, IDPs lontani da qualsiasi centro di località, ma non necessariamente inadeguati (pochi o poco vicini tra loro), non saranno utilizzati per caratterizzare lo scuotimento del luogo; questo inconveniente si cerca di limitarlo con la tecnica mista descritta al paragrafo successivo (par 3.6). Nel secondo caso invece la sovrapposizione può causare un’importante perdita di risoluzione in relazione all’ordine in cui vengono considerati i cerchi ai fini dell’assegnazione degli IDPs; perciò il raggruppamento degli IDPs avviene effettuando un’operazione di “*sorting*” dei cerchi, e cioè ordinandoli in base all’area e procedendo dal cerchio di area minima al cerchio di area massima.

L’assegnazione di un IDP ad un cerchio è effettuata confrontando il raggio del cerchio con la distanza geodetica *centro-IDP* calcolata su una superficie sferica di raggio $R_0 = 6371Km$.

Per la selezione dei clusters significativi di IDPs ed operare quindi il calcolo delle statistiche viene utilizzata la stessa soglia delle precedenti tecniche di raggruppamento (numero minimo di IDPs inclusi=5; par.3.3); sono inoltre utilizzati gli stessi indicatori (media, mediana, ecc.) e le stesse approssimazioni per la caratterizzazione degli MDPs di cui al par.3.3.

Dalle elaborazioni per uno stesso terremoto risulta che in generale questa tecnica costruisce un numero di MDPs nettamente inferiore a quelli risultanti dall’applica-

zione delle due tecniche precedenti. Questo emerge già dal confronto tra le figure 3.3, 3.5, 3.7 relative al terremoto del 24/08/2016; con la tecnica dei cerchi se ne ottengono 30, e cioè circa il 35% in meno; le località di piccola dimensione vengono rappresentate meno frequentemente. Tuttavia risultano generalmente sovrapponibili gli MDPs relativi alle località di media dimensione, ma con fluttuazioni nell'assegnazione dell'intensità. Da ciò si evince l'adeguatezza nella scelta del valore del parametro *dens* almeno allo scopo di rilevare MDPs nei centri abitati principali. Si conferma dunque tale capacità di rilevazione anche per le due tecniche precedenti. È altresì evidente la perdita di risoluzione in corrispondenza dei grandi centri abitati: la capitale è infatti rappresentata con un solo MDP, a confronto con i 6-7 MDPs rilevati con quadrati ed esagoni rispettivamente.

Anche per questa tecnica la percentuale di MDPs costituiti da $\#IDPs < 10$ risulta prossima al 70% anche se leggermente inferiore a quella delle due tecniche di griglia. Gli *errori standard* di latitudine, longitudine e intensità degli MDPs presentano le stesse caratteristiche descritte ai par. 3.3, 3.4.

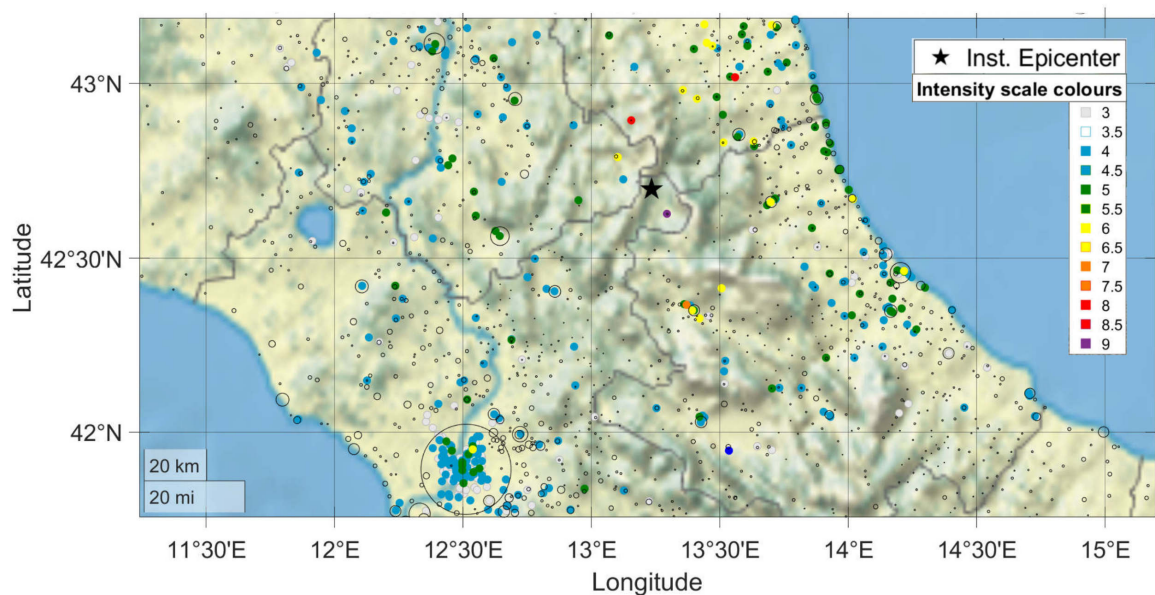


Figura 3.6: Terremoto 24/08/2016. IDPs (I_{raw} , EMSC) su cerchi attorno a località con parametro $dens = 3500ab/Km^2$. Vedere testo per ulteriori dettagli.

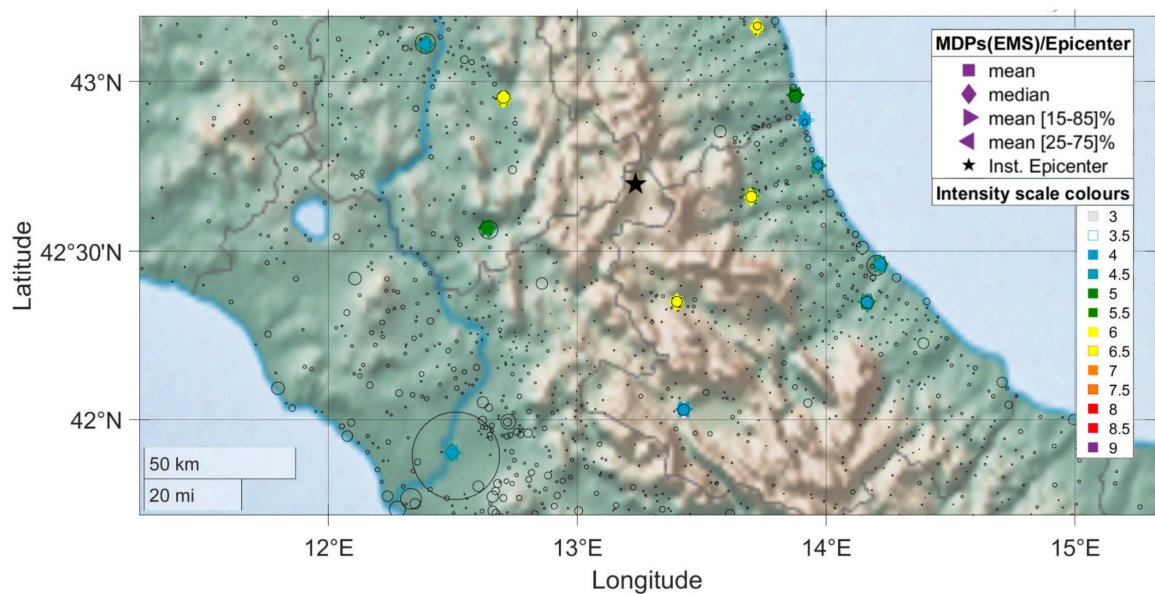


Figura 3.7: Terremoto 24/08/2016. MDPs (da I_{corr} EMSC) costruiti con la tecnica di clustering cerchi attorno a località con parametro $dens = 3500ab/Km^2$. Vedere testo per ulteriori dettagli.

3.6 Tecnica 4: mista, cerchi attorno a località e griglia esagonale

Questa tecnica consiste nell'uso sequenziale della tecnica 3 dei cerchi e della tecnica 2 di griglia esagonale equispaziata descritte ai paragrafi 3.4 e 3.5. L'idea è di ridurre l'inconveniente di una presenza eccessiva di *outliers* dovuti all'assunzione di proporzionalità tra area del cluster/cerchio e densità di popolazione residente (assunta uniforme per tutte le località, come già detto). Gli IDPs che non ricadono in nessun cerchio ma che ricadono in numero sufficiente (almeno 5, come da convenzione, par.3.3) in un poligono della griglia di esagoni che invece partiziona l'intera superficie del dataset, sono poi assegnati e utilizzati per la costruzione di MDPs.

Per la formazione dei clusters si applica quindi sequenzialmente 1) la procedura dei cerchi e 2) la tecnica degli esagoni sugli IDPs non ancora assegnati, entrambe nelle stesse modalità descritte precedentemente (par.3.4, 3.5). In questo modo, individuando ulteriori MDPs, si cerca di limitare la perdita di informazione e di risoluzione implicita alla sola tecnica dei cerchi. I valori statistici degli MDPs sono calcolati come per le tecniche precedenti.

Confrontando le mappe delle figure 3.5, 3.7 e 3.9 relative al terremoto 24/08/2016 si riscontra l'efficacia della "riassegnazione" con la tecnica 4 degli IDPs rilevati come *outliers* dalla sola tecnica 3 (cerchi attorno a località). Alcuni degli MDPs non costruiti con la tecnica 3 ma costruiti con la tecnica 2 (griglia di esagoni) sono assegnati dalla tecnica 4 (cerchi+esagoni) e risultano generalmente sovrapponibili (ellissi viola di figura 3.9). L'informazione perduta con la tecnica 3 (cerchi) viene dunque parzialmente recuperata con la tecnica 4 (cerchi+esagoni). Per quest'ultima tecnica aumenta di circa il 5%, rispetto a soli quadrati ed esagoni (tecniche 1 e 2), la percentuale di MDPs costituiti da $\#IDPs < 10$. Nel caso specifico del terremoto 24/08/2016 gli MDPs totali sono 40, dei quali i 10 "riassegnati" sono tutti costituiti da 5-6 IDPs: qui la "riassegnazione" avviene solamente per IDPs associabili ad agglomerati di piccole e piccolissime località (ellissi viola di figura 3.8 e 3.9); la risoluzione in corrispondenza della città di Roma (ellissi verde) non è recuperata nemmeno parzialmente.

Le incertezze sulle *caratteristiche* degli MDPs hanno le stesse peculiarità descritte ai paragrafi 3.3 e 3.4.

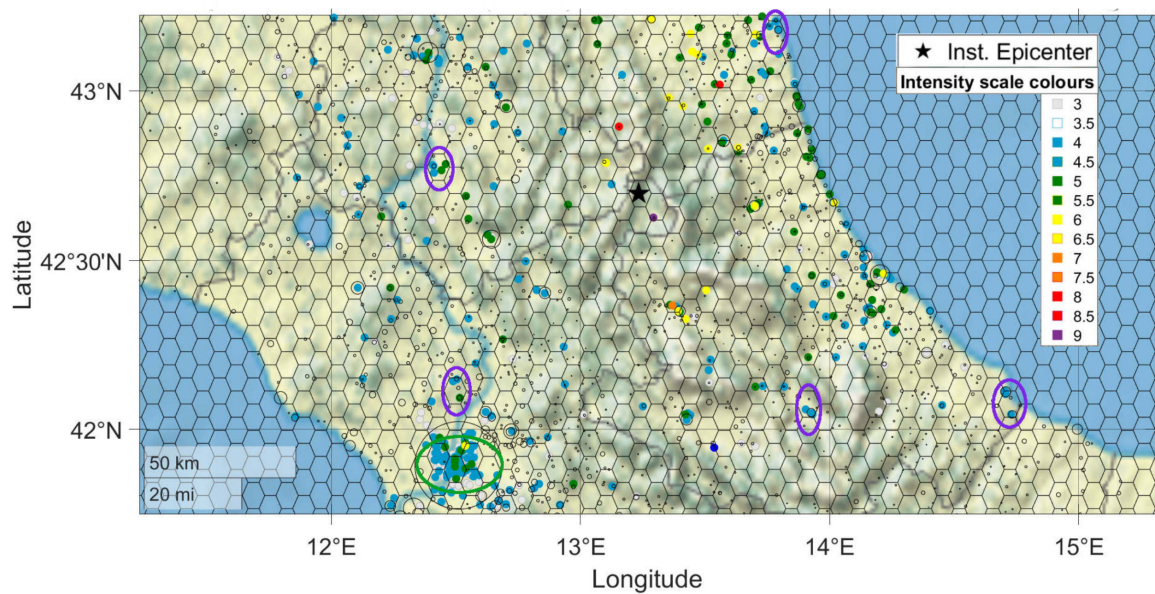


Figura 3.8: Terremoto 24/08/2016. IDPs (I_{raw} , EMSC) su cerchi attorno a località con parametro $dens = 3500ab/Km^2$ e griglia di esagoni equiareali di lato $r = 5Km$. Vedere testo per ulteriori dettagli.

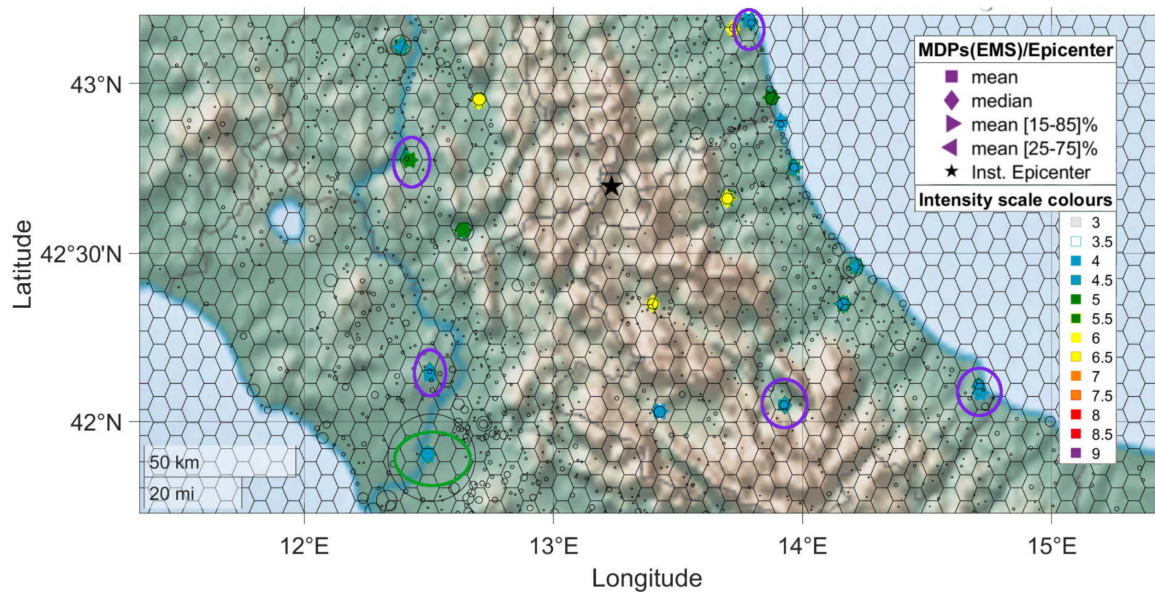


Figura 3.9: Terremoto 24/08/2016. MDPs (da I_{corr} EMSC) costruiti con la tecnica di clustering mista (cerchi+esagoni) con parametri $dens = 3500ab/Km^2$ e $r = 5Km$. Vedere testo per ulteriori dettagli.

3.7 Tecnica 5: DBSCAN

La DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*) è una tecnica di clustering spaziale basata sulla densità proposta da Ester, Kriegel e altri nel 1996 (Ester et al., 1996). Ad oggi è la più diffusa tecnica di clustering poiché si è rivelata efficiente nell'individuare forme e strutture arbitrarie di punti di datasets inerenti diversi ambiti di applicazione; è ampiamente utilizzata per dati di tipo spaziale e geografico (Li, J. W. et al., 2020; Rajat, 2020; Mai, 2014) e, in particolare, è utilizzata anche in ambito sismologico (Fan e Xu, 2019; Karri et al., 2018; Kazemi-Beydokhti et al., 2017).

A differenza della tecnica 3 per formare i clusters la DBSCAN utilizza la *densità a-posteriori*, e cioè l'effettiva densità dei punti del dataset.

Le sue caratteristiche principali e teoricamente vantaggiose per i dataset di IDPs trattati per questa tesi, sono 1) la capacità di individuare clusters anche piccoli ma sufficientemente densi 2) di gestire automaticamente gli outliers e 3) la limitazione degli effetti vincolanti della metrica, utilizzata per le distanze tra i punti, nel determinare clusters di forme arbitrarie.

Per l'applicazione di questa tecnica è stata utilizzata la routine "dbscan" (*The Statistics and Machine Learning Toolbox*) di Matlab ⁵ (The MathWorks, 2019a; The MathWorks, 2019b). L'algoritmo della routine implementa la tecnica DBSCAN secondo la procedura seguente:

dato un dataset, specificati i due parametri *eps*, distanza che definisce il raggio di *prossimità* (*neighborhood*) e *minpts*, numero minimo necessario di punti *prossimi* per iniziare a formare un cluster,

1. viene selezionato il primo *oggetto* che non è ancora stato clusterizzato e viene verificato se vi siano *minpts oggetti* entro un raggio *eps* centrato su esso;
2. se la condizione non è verificata l'*oggetto* è etichettato come *outlier* e si ripete il punto precedente selezionando l'*oggetto* successivo del dataset;
3. se la condizione è verificata l'*oggetto* viene etichettato come *punto di nucleo* (*core point*), inizia la formazione di un cluster, e il punto 1 viene eseguito per ciascuno degli *oggetti* presenti nell'*eps* del *punto di nucleo*;
4. se per i punti nell'*eps* del *punto di nucleo* vale la condizione di *punto di nucleo* anch'essi vengono etichettati come *core points* dello stesso cluster, altrimenti sono etichettati come *border points* e comunque inclusi nel cluster;
5. assegnati tutti i *border points* e i *core points* del cluster in costruzione la procedura reinizia al punto 1 fino all'avvenuta etichettatura di tutti gli *oggetti* del dataset.

I cluster sono quindi formati sia da *core points* sia da *border points*.

In questa procedura gli *oggetti* etichettati come *outliers* ad un qualche step possono essere etichettati come *border points* se rientrano nell'*eps* di un *oggetto* processato in seguito.

⁵https://it.mathworks.com/help/stats/dbscan.html?searchHighlight=dbscan&s_tid=srchtitle

La metrica qui utilizzata per l'algoritmo corrisponde alla distanza geodetica misurata su una superficie sferica di raggio $R_0 = 6371Km$; dati due *oggetti* i e j è definita come

$$D_{ij} = R_0 \arccos (\sin (lat_j) \sin (lat_i) + \cos (lat_j) \cos (lat_i) \cos (lon_j - lon_i)). \quad (3.2)$$

La scelta dei parametri *eps* e *minpts* dell'algoritmo non è stata ottimizzata seguendo una procedura automatica di selezione. Qui è invece stato adottato un approccio euristico che ha selezionato parametri confrontabili con quelli delle altre tecniche di raggruppamento utilizzate per l'analisi degli IDPs e descritte ai paragrafi 3.3, 3.4, 3.5 e 3.6. La soglia *minpts* = 5 riflette un numero minimo di IDPs sulle cui *intensità* si possa calcolare una media o una mediana minimamente stabili, come da convenzione (par.3.3). Il raggio *eps* = 5Km, come le altre scelte per i parametri di griglia, è sicuramente rappresentativo di una città di medie dimensioni.

A ciascun cluster individuato dall'algoritmo è quindi associata un'*intensità* (MDP) stimata esattamente come per i clusters costruiti con le precedenti tecniche (paragrafi 3.3, 3.4, 3.5 e 3.6).

Sono riportate di seguito le mappe dei clusters e degli MDPs costruiti con l'algoritmo DBSCAN relative al terremoto 24/08/2016 (Figure 3.10, 3.11) per un confronto esemplificativo con le altre tecniche (Figure 3.1-3.8). Il DBSCAN riesce a costruire i principali MDPs riconosciuti dalle altre tecniche relativi a località sia medio-grandi sia piccole. Tuttavia per alcuni di questi si osservano fluttuazioni sia nella posizione, che può variare di qualche chilometro, sia nell'assegnazione del valore di *intensità*, anche dell'ordine del grado. Il DBSCAN riconosce come IDPs di uno stesso cluster gli IDPs segregati invece rigidamente in poligoni/cerchi adiacenti delle tecniche precedenti (ellissi rosse); d'altra parte si osserva la capacità di segregazione in MDPs differenti di IDPs relativi a clusters esagonali adiacenti e cerchi di città di medie dimensioni limitrofe se tra questi è insufficiente la densità di IDPs (ellissi ocra). Sono inoltre costruiti MDPs che accorpano IDPs associabili a piccole e piccolissime località limitrofe e che le altre tecniche non avevano riconosciuto (ellissi ciano). Infine la risoluzione per la città metropolitana di Roma è analoga alla tecnica 3 (MDP unico, ellisse verde).

In totale la tecnica DBSCAN permette di costruire 38 MDPs, 10 in meno delle griglie, 8 in più dei cerchi, 2 in meno della tecnica mista, con una percentuale inferiore al 65% di MDPs costituiti da $\#IDPs < 10$, la più bassa tra tutte le tecniche.

Le incertezze sulle *intensità* risultano tutte non superiori al mezzo grado, mentre quelle su *lat* e *lon* sono tutte nell'ordine del millesimo di grado decimale ad eccezione di qualche MDPs con $\#IDPs < 10$ per cui si presentano di 0.01 gradi decimali. Per questa tecnica le incertezze sono dunque generalmente inferiori rispetto a tutte le altre tecniche.

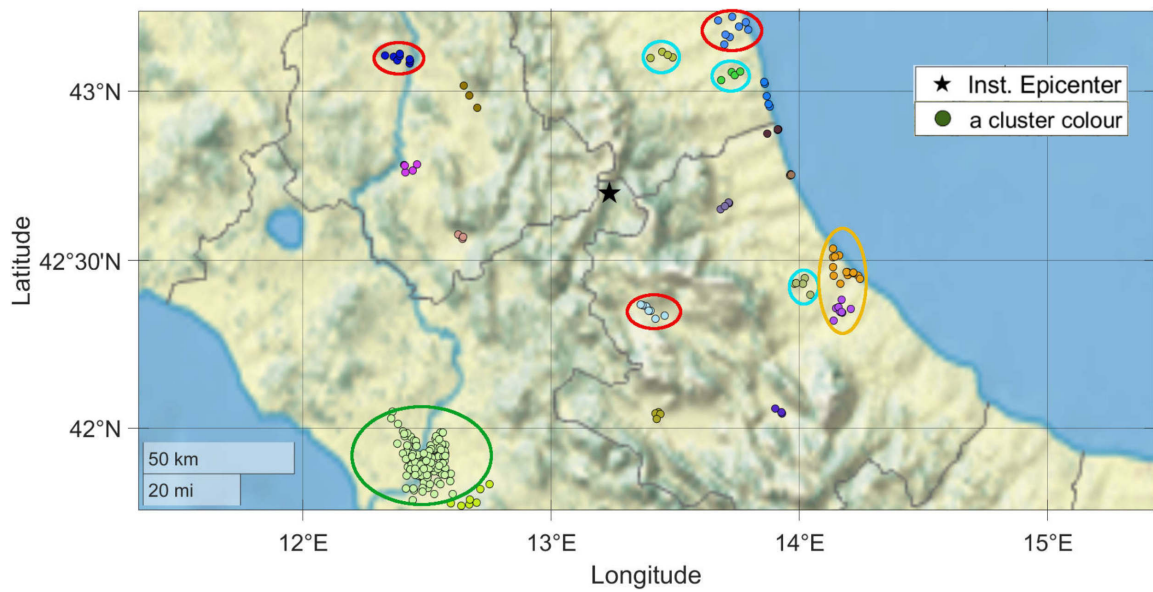


Figura 3.10: Terremoto 24/08/2016. Clusters di IDPs (I_{raw} , EMSC) costruiti con la tecnica di clustering DBSCAN con parametri $eps = 5Km$ and $minpts = 5$. Vedere testo per ulteriori dettagli.

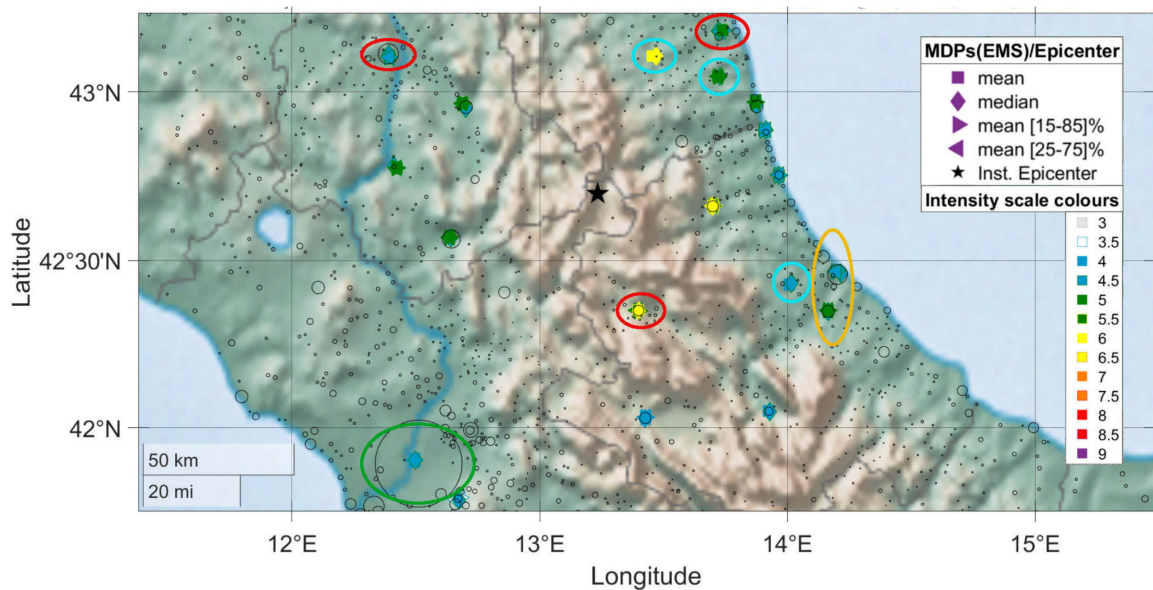


Figura 3.11: Terremoto 24/08/2016. MDPs (da I_{corr} EMSC) costruiti con la tecnica di clustering DBSCAN con parametri $eps = 5Km$ e $minpts = 5$. Per riferimento sono mostrati i cerchi costruiti attorno alle località con parametro $dens = 3500ab/Km^2$ utilizzati per la tecnica 3. Vedere testo per ulteriori dettagli.

Capitolo 4

Analisi dei risultati

4.1 Scelta dei datasets

In questa tesi sono stati esaminati 9 terremoti con le tecniche di clustering analysis descritte al capitolo 3. Tutti questi eventi sismici sono stati rilevati e confermati con le tecniche della sismologia strumentale, per i cui confronti si utilizzano le stime strumentali fornite da INGV e riportate nell'header stesso dei files dei datasets macrosismici.

I set di dati macrosismici relativi ai 9 eventi sono stati forniti sia da INGV sia da EMSC, e per ciascuno degli eventi erano disponibili i dati di entrambe le istituzioni. Questo ha permesso di effettuare confronti anche in base ai due differenti sistemi di raccolta degli IDPs. Si ricorda infatti che, come mostrato in precedenza al capitolo 2, i datasets di *felt reports* di EMSC sono stati raccolti mediante i *thumbnails web-based*, mentre i datasets dell'INGV mediante i questionari online di HSIT.

I datasets dell'INGV sono disponibili sul sito <https://www.hsit.it/> nelle sezioni relative a ciascun terremoto o nel database generale “*HSIT macroseismic intensity database 2007-2018, version 1*” (De Rubeis et al., 2019); quelli EMSC sono disponibili sul sito <https://www.seismicportal.eu/>, come descritto nelle specifiche del servizio “*Specifications of the EMSC testimony's Service*” (Landes, 2017).

In Tabella 4.1 si riporta l'elenco dei terremoti indicando l'identificativo (*ID*) del terremoto utilizzato, *data e ora UTC*, *magnitudo*, *latitudine* e *longitudine epicentrale*, *numero di IDPs* del dataset di EMSC corrispondente.

I 9 terremoti sono stati scelti adottando i seguenti criteri

- Eventi occorsi sul territorio italiano, o prossimi ad esso, poichè HSIT raccoglie web-based felt reports solamente in Italia. Si possono così effettuare confronti con i dati di EMSC relativi agli stessi eventi.
- Data non precedente al 2017 per ragioni di efficienza del sistema EMSC nella raccolta dei *thumbnail-based felt reports* nel territorio Italiano (vedi par.2.1).
- Numero di IDPs dei dataset di EMSC più elevato possibile (si sono preferiti quelli con $Num_{IDPs} \simeq 10^3$).
- Magnitudo strumentale più elevata possibile ($M \geq 5.0$) per poter avere un elevato numero di MDPs e permettere le elaborazioni con il software Boxer.

Analysed Events						
Earthquake ID	Date	Time (UTC)	Magnitude	Lat	Lon	#IDPs (EMSC)
24-08-2016 _	24/08/2016	01:36:32	6.0	42.698	13.234	1082
18-01-2017 A	18/01/2017	09:25:40	5.1	42.545	13.277	916
18-01-2017 B	18/01/2017	10:14:10	5.5	42.531	13.284	621
18-01-2017 C	18/01/2017	10:25:25	5.4	42.503	13.277	339
18-01-2017 D	18/01/2017	13:33:36	5.0	42.473	13.275	772
16-08-2018 _	16/08/2018	18:19:04	5.1	41.874	14.865	541
25-10-2018 _	25/10/2018	22:54:51	6.8	37.492	20.595	1883
07-11-2019 _	07/11/2019	17:35:21	4.4	41.776	13.604	343
26-11-2019 _	26/11/2019	02:54:11	6.2	41.371	19.467	4036

Tabella 4.1: Lista dei terremoti analizzati. Sono riportati i parametri strumentali forniti da INGV e il numero di IDPs del dataset di EMSC corrispondente.

La sole eccezioni sono l'evento dell'Italia Centrale 24/08/2016 ore 01:36:32, antecedente all'anno soglia (2017) ma scelto poichè di magnitudo elevata ($M = 6.0$) e con $Num_{IDPs} = 1082$, e il terremoto del 07/11/2019 ore 17:35:21 di bassa magnitudo $M = 4.4$ e con pochi IDPs $Num_{IDPs} = 343$.

Tra i restanti terremoti 5 sono avvenuti in area italiana, hanno magnitudo $5.0 < M < 5.5$ e $Num_{IDPs} < 10^3$, 1 ha epicentro sulla costa albanese (26/11/2019), $M = 6.2$ e $Num_{IDPs} = 4036$, 1 ha epicentro in mare in territorio greco (25/10/2018), $M = 6.8$ e $Num_{IDPs} = 1883$.

4.2 Elaborazione dei dati

Per l'analisi utilizzeremo i seguenti termini per distinguere gli approcci adottati per raggruppare ("clusterizzare") le intensità IDPs e derivare in ogni unità di raggruppamento (MDP) una intensità di riferimento. Dato un approccio (per celle, aree ecc), l'insieme dei raggruppamenti con esso costituiti sostituisce l'intensità classica assegnata alle singole località e viene definito come "distribuzione macrosismica di intensità MDPs". In particolare con il termine "tecnica (di clustering)" verrà indicato l'approccio per raggruppare gli IDPs all'interno di un'area di riferimento (cella, cerchio ecc), mentre con "metodo (statistico)" saranno indicati i parametri usati per derivare le intensità dalla distribuzione MDPs (media, mediana ecc).

L'elaborazione dei dati relativi ai 9 terremoti selezionati per l'analisi (Tabella 4.1) è suddivisa in due fasi:

1. la prima consolida l'informazione macrosismica in essi contenuta mediante l'operazione di trasformazione dal campo degli IDPs (*Individual intensity Data Points*) al campo degli MDPs (*Macroseismic Data Points*) applicando metodi e tecniche descritti al capitolo 3;
2. la seconda applica il software Boxer (par. 1.4) alle distribuzioni di MDPs ottenute nella prima fase, allo scopo di stimare i *parametri sismici sintetici*.

Relativamente alla prima fase le differenze tra i dati di base forniti da EMSC e quelli forniti da INGV (HSIT) (capitolo 2) riguardano i dati di EMSC che sono

costituiti da IDPs sparsi rappresentativi delle singole testimonianze geolocalizzate (lat, lon) degli utenti, ciascuna caratterizzata da I_{raw} , intensità scelta dall'utente e I_{corr} , intensità corretta (Bossu et al., 2017)¹. Questi sono stati elaborati con le 5 tecniche di clustering utilizzate in questo lavoro di tesi per la costruzione di distribuzioni di MDPs da IDPs sparsi. Le tecniche sono descritte in dettaglio al capitolo 3. I dati di base forniti da HSIT sono già distribuzioni di MDPs costruiti a partire dall'associazione delle testimonianze (IDPs) ad una stessa località amministrativa (Comune), secondo la procedura descritta da Tosi et al., 2015. Per ciascuno di questi MDPs è riportata la coppia di coordinate di riferimento della località (lat, lon), l'intensità ad essa assegnata (sia in unità di EMS-98 sia in unità di MCS) e il numero di IDPs costituenti (Tosi et al., 2007). Questi dati non hanno quindi subito sostanziali modificazioni prima di essere elaborati con Boxer. L'unica modificazione è stata la selezione di quei soli MDPs costruiti a partire da almeno 5 IDPs: questa convenzione è applicata anche agli MDPs costruiti dai dati EMSC per ragioni statistiche e di uniformità, esposte ai paragrafi 2.2 e 3.3.

Un'ulteriore convenzione è stata applicata ai valori di intensità, sia di EMSC sia di HSIT. Sono state considerate le sole intensità nel range $2 \leq I \leq 10$. La soglia massima segue la convenzione di Bossu et al., 2017 che ritiene inaffidabili le valutazioni di intensità superiori ad essa da parte di utenti volontari per ragioni di competenza ed di priorità (R. Musson, 2007). D'altra parte HSIT in base a medesime considerazioni fornisce i dati con intensità elevate raggruppati in un'unica classe (Tosi et al., 2015). L'intensità minima è scelta in relazione a considerazioni di sottorappresentatività e d'indistinguibilità dei gradi più bassi della scala macrosismica (D. J. Wald et al., 2011; Tosi et al., 2015). Inoltre le relazioni empiriche Magnitudo-Intensità utilizzate da Boxer sono calibrate con terremoti di magnitudo elevata con prevalenza di intensità elevate, perciò le basse intensità dei nostri datasets sono generalmente meno rilevanti ai fini delle stime di magnitudo.

Per ogni terremoto l'output della prima fase di elaborazione dei dati consta di distribuzioni di MDPs differentemente costruite. 40 sono costruite a partire dai dati EMSC applicando le 5 tecniche di clustering precedenti sia alle I_{raw} sia alle I_{corr} . Le intensità degli MDPs costruiti con le 5 tecniche sono inoltre rappresentate con 4 differenti metodi statistici applicati agli IDPs che li formano (media, mediana, media 15/85, media 25/75) (paragrafo 3.3); le differenti rappresentazioni del campo macrosismico prodotte da dati EMSC sono pertanto 40 (4 metodi, 5 tecniche, 2 tipi di intensità). Altre 2 distribuzioni di MDPs sono costruite a partire dai dati HSIT – già clusterizzati con la tecnica di Tosi et al., 2015– espressi in 2 tipi di intensità, EMS-98 ed MCS. Quindi per ogni evento si hanno 42 differenti distribuzioni di intensità MDPs.

Infine le intensità degli MDPs costruiti con le precedenti tecniche sono state arrotondate al mezzo grado più vicino per ragioni legate alla formulazione della *scala macrosismica*.

Nella seconda fase dell'elaborazione le distribuzioni di MDPs sono state utilizzate come input del programma Boxer per ottenere per ciascuno dei 9 terremoti 42 differenti stime dei *parametri sismici sintetici*.

¹la correzione è di tipo lineare ed è stata suggerita in via preliminare, assumendo coincidenza tra la scala EMS-98 e la scala MMI utilizzata dal sistema DYFI, a seguito di un confronto tra le due intensità che ha evidenziato una sottostima da parte dei *thumbnails* basati sull'EMS-98.

I *parametri sismici sintetici* presi in considerazione per quest’analisi sono solamente le *coordinate epicentrali* e la *magnitudo*.

La profondità non è stata considerata per ragioni di generale inaffidabilità delle stime di questo parametro sia con metodi macrosismici sia con metodi strumentali (Lomnitz, 2006; Gasperini et al., 2010; Musson e Ceciè, 2011).

Le stime dell’orientazione e delle dimensioni della faglia non sono invece state effettuate per mancanza di dati di campionamento del near-field macrosismico di tutti e 9 i terremoti analizzati; inoltre 6 dei terremoti in esame hanno magnitudo al di sotto della soglia minima di affidabilità ($M = 5.7$) individuata da Gasperini et al., 2010 per le stime dell’orientazione; già Gasperini et al., 1999 suggerivano come soglia $M = 5.5$ a cui corrisponderebbe una faglia di poco meno di $7Km$ di lunghezza (Wells e Coppersmith, 1994).

I *parametri sintetici* di tutti e 9 i terremoti sono stati stimati con il “metodo 0” di Boxer che è stato scelto tra gli altri per la stabilità dei risultati ottenuti in precedenti applicazioni, anche con datasets macrosismici costituiti da distribuzioni di MDPs non ottimali (Gasperini et al., 2010; Rovida et al., 2019). Il “metodo 0” consiste nell’applicazione del “metodo del baricentro” per la stima delle coordinate epicentrali e dell’ “old magnitude method” per la stima della magnitudo, come precedentemente descritti ai paragrafi 1.4.1 e 1.4.2.

Inoltre per i 2 terremoti con epicentro in mare o prossimo alla costa (25/10/2018 e 26/11/2019, Tabella 4.1), caratterizzati quindi a-priori da una distribuzione azimutale degli MDPs non uniforme attorno all’epicentro, è stata avviata anche la stima dei *parametri* con il “metodo 1” di Boxer. Tale metodo ottimizza una funzione di verosimiglianza della distribuzione delle MDPs per le coordinate geografiche e consente la localizzazione anche in mare o aree scarsamente abitate (Gasperini et al., 2010) (par. 1.4.1); ad esso è associato l’ “alternative magnitude method” per la stima delle magnitudo (par. 1.4.2).

Le incertezze sulle stime dei parametri considerate per quest’analisi sono le *incertezze formali* come descritte al paragrafo 1.4.4. Per la “old magnitude” è utilizzata l’incertezza del modo 1) (par.1.4.4), detta “type-I magnitude uncertainty” in Gasperini et al., 2010, Appendix 2. Inoltre il “metodo 1” di stima dei *parametri*, ha fornito risultati anche per gli *ellissoidi di incertezza* relativi alla matrice di varianza e covarianza degli errori.

4.3 Discussione risultati

In questo paragrafo vengono presentati i risultati delle stime di *epicentri* e *magnitudo* effettuate con Boxer a partire dalle distribuzioni di MDPs costruite con le diverse tecniche e metodi. Tali stime vengono presentate separatamente, in due differenti sezioni: “Magnitudo” e “Localizzazione Epicentrale”. In ciascuna sezione, per ogni terremoto vengono riportati:

- plot riassuntivo con le differenze delle stime macrosismiche rispetto alla stima di riferimento strumentale;
- tabella riassuntiva delle stime macrosismiche e delle differenze rispetto al riferimento strumentale;
- tabelle di “score/pungeggi” assegnati alle tecniche di clustering e ai metodi statistici di assegnazione dell’intensità (media, mediana, ecc) in base alle migliori o peggiori stime ottenute;
- mappa delle migliori localizzazioni epicentrali ottenute con le diverse tecniche e metodi (in sezione “Localizzazione epicentrale”).

Nelle tabelle riassuntive (4.4, 4.7), delle stime effettuate da dati EMSC sono riportate solamente quelle relative alla tipologia di intensità I_{corr} , che corrisponde all’intensità EMS riportata dagli utenti (I_{raw}) corretta con la relazione di Bossu et al., 2017. Questa scelta si effettua poichè, come si osserva dai plot delle differenze (in cui sono riportate le stime da ambo i tipi di intensità, Figure 4.2-4.10), con le I_{corr} si ottengono generalmente risultati migliori rispetto che con le I_{raw} . Questo risultato è particolarmente evidente dai plot delle magnitudo, è invece meno evidente nei plot degli epicentri. Per questa ragione, per tutto il resto dell’analisi delle stime da dati EMSC vengono considerate solamente quelle con intensità corrette I_{corr} .

A seguire, nelle relative sezioni di parametro, vengono mostrati plot, tabelle e mappe con i risultati dei 2 terremoti con epicentro in mare o prossimo alla costa (25/10/2018, 26/11/2019), stimati con il “metodo 1” di Boxer (par. 4.2).

4.3.1 Magnitudo

In questa sezione vengono presentati per i 9 terremoti analizzati i risultati delle stime di magnitudo equivalente effettuate con il “metodo 0” (MT0) di Boxer a partire dalle distribuzioni di MDPs costruite da dati macrosismici EMSC e HSIT. A seguire, nella sottosezione “Boxer MT1”, sono presentate le stime di magnitudo equivalente effettuate con il “metodo 1” per i soli 2 terremoti con epicentro prossimo alla costa.

Per ciascuno dei terremoti si pone un plot con le differenze delle stime macrosismiche, sia da dati EMSC sia da dati HSIT, rispetto alla magnitudo strumentale di riferimento (Figure 4.2-4.10). Nel plot sono rappresentate tutte le 42 differenze di magnitudo relative alle 42 distribuzioni di MDPs (40 da dati EMSC, 2 da dati HSIT) costruite come sintetizzato al paragrafo 4.2. Sull’asse delle ordinate si pongono le differenze, sull’asse delle ascisse si pongono le etichette dei metodi statistici utilizzati per la rappresentazione degli MDPs da dati EMSC e le etichette delle stime da dati HSIT. Per completezza, delle stime da dati EMSC sono mostrate sia quelle effettuate a partire da intensità corrette I_{corr} secondo Bossu et al., 2017 (simboli blu) che quelle effettuate a partire da intensità I_{raw} selezionate dagli utenti (simboli rossi), nonostante le prime stimino sistematicamente meglio delle seconde; la forma dei simboli si riferisce alle diverse tecniche di clustering utilizzate per la costruzione degli MDPs (descritte ai paragrafi 3.3-3.7) (\square =celle quadrate, \diamond =celle esagonali, \circ =cerchi attorno a località, \times =mista cerchi ed esagoni, $+$ =DBSCAN, \triangle =HSIT); delle stime HSIT sono colorate in verde chiaro quelle ottenute da intensità espresse con la scala macrosismica EMS, in verde scuro quelle ottenute da intensità espresse con la scala macrosismica MCS.

Dopo i plot si pone una tabella riassuntiva (Tabella 4.2) dei valori (colonna sinistra) e delle differenze di magnitudo (colonna destra), suddivisa in sezioni orizzontali terremoto-specifiche. Le incertezze sui valori sono quelle formali calcolate con Boxer. Le caselle in viola indicano ID , *parametri strumentali* e $\#IDPs(EMSC)$ del terremoto. La sezione di ciascun terremoto è suddivisa in stime da dati EMSC (sopra) e stime da dati HSIT (sotto).

In questa tabella delle stime da dati EMSC si mostrano solamente quelle effettuate da intensità corrette I_{corr} poichè generalmente migliori delle stime effettuate da intensità I_{raw} , come emerge osservando i plot. Da questa evidenza si è scelto di circoscrivere il seguito dell’analisi delle stime da dati EMSC alle sole stime da intensità I_{corr} .

Per ciascuna stima è indicata lungo la stessa riga sia la tecnica di clustering usata per la costruzione della relativa distribuzione di MDPs

1. SQ = \square , griglia di quadrati
2. HE = \diamond , griglia di esagoni
3. CI = \circ , cerchi attorno a località
4. CH = \times , mista cerchi ed esagoni
5. DB = $+$, DBSCAN
6. HSIT_EMS e HSIT_MCS = \triangle , tecnica HSIT con differente scala macrosismica di riferimento

sia il relativo numero di MDPs ($\#MDPs$). Le stime da dati EMSC sono suddivise in colonne in base ai metodi statistici di assegnazione delle intensità alle MDPs della relativa distribuzione

1. mean $([0, 100]\%)$,
2. median $([0, 100]\%)$,
3. mean $[15, 85]\%$,
4. mean $[25, 75]\%$.

A parità di tecnica le 2 medie decimate sono tali da ridurre il $\#MDPs$ utile alle stime. Inoltre per ciascun terremoto sono evidenziate in verde e in rosso rispettivamente la migliore e la peggiore stima ottenute con le 5 tecniche sviluppate in questa tesi. Con gli stessi colori ma in gradazione più scura sono evidenziate le stime migliori e peggiori di HSIT solamente se esse risultano tali rispetto a tutte quelle effettuate con le 5 tecniche. Per ogni terremoto, per ciascuna delle 5 tecniche sono riportate in *corsivo* e in sottolineato rispettivamente le stime migliori e peggiori in relazione ai metodi statistici (media, mediana, ecc).

Per la sintesi delle stime migliori e peggiori in relazione alle tecniche di clustering e ai metodi statistici utilizzati per la rappresentazione degli MDPs sono mostrate 2 tabelle di score.

La prima (Tabella 4.3a) è divisa in due sezioni e mostra nella sezione superiore il conteggio, per ciascuna delle 5 tecniche, per aver fornito i valori migliori (sfondo verde chiaro di Tabella 4.2) e peggiori (sfondo rosso chiaro di Tabella 4.2) tra le altre, considerando ciascuno dei 9 terremoti. In verde e in rosso sono evidenziati i punteggi più alti per la migliore e la peggiore stima, rispettivamente. Nella sezione inferiore è mostrato il conteggio delle stime migliori (sfondo verde scuro di Tabella 4.2) e peggiori (sfondo rosso scuro di Tabella 4.2) effettuate da dati HSIT confrontate, separatamente EMS ed MCS, con tutte e 5 le tecniche, considerando ciascuno dei 9 terremoti.

La seconda tabella di score (Tabella 4.3b) riporta i conteggi dei migliori e dei peggiori risultati delle stime ottenuti utilizzando i diversi metodi statistici per rappresentare gli MDPs (media, mediana, ecc). Il confronto è effettuato per ciascun terremoto, per ciascuna delle 5 tecniche: si assegna un punto ad un indice per ogni valore in corsivo (migliore) e ogni valore sottolineato (peggiore) di Tabella 4.2. In verde e in rosso sono evidenziati i punteggi più alti, migliore e peggiore rispettivamente.

Dai plot delle differenze di magnitudo equivalente rispetto al riferimento strumentale (Figure 4.2-4.10) emerge per tutti i terremoti una sistematica sottostima dei valori macrosismici, generalmente non superiore a 2 unità di magnitudo. L'unica eccezione è il terremoto 07/11/2019 dove le 2 stime da dati HSIT sono in accordo con il riferimento strumentale.

Generalmente nei singoli plot si possono individuare 3 gruppi di stabilità, coincidenti con: stime da intensità I_{raw} (EMSC) scelte dagli utenti (simboli rossi), stime da intensità corrette I_{corr} (EMSC) (simboli blu), stime da dati HSIT (simboli verdi). Le prime sottostimano sistematicamente in misura maggiore delle seconde, le terze subiscono fluttuazioni sia positive che negative rispetto alle altre due. Questa evidenza risultata più debole nei plot dei terremoti 18/01/2017 C, 16/08/2018, 07/11/2019

(Figure 4.4, 4.6, 4.8) dove comunque, a parità di tecnica e di metodo statistico, le stime da intensità I_{corr} sono migliori delle stime da intensità I_{raw} . Considerando solamente le I_{corr} , oltre ad ottenere risultati migliori in valore assoluto, la dispersione delle stime (differenza tra le stime massima e minima) si riduce a circa 0.4 u.m..

Per ogni terremoto si osserva che il gruppo delle stime da dati HSIT, per le quali varia il tipo di intensità (EMS, MCS) ma non la tecnica, presenta meno dispersione rispetto agli altri 2 gruppi. La differenza tra le stime HSIT è stabile entro le 0.2 u.m..

Per ogni terremoto, relativamente ai metodi statistici si può distinguere il gruppo delle stime che utilizzano il metodo della mediana e il gruppo delle stime che utilizzano i 3 diversi metodi della media.

Considerando complessivamente le 5 tecniche, il gruppo delle medie è tale da presentare risultati generalmente stabili tra loro, con fluttuazioni prevalenti delle stime effettuate con le medie decimate, alle quali corrisponde anche una lieve maggiore dispersione (e.g. Figure 4.1, 4.3, 4.4, 4.5, 4.6, 4.7, 4.9).

Il gruppo delle mediane si discosta generalmente dal gruppo delle medie e produce, a parità di tecnica, risultati migliori (e.g. Figure 4.1, 4.7, 4.9), ma non per tutti i terremoti (e.g. Figure 4.2, 4.4).

Nessuna delle 5 tecniche fornisce risultati nettamente migliori delle altre. Considerando complessivamente i 9 terremoti la tecnica dei cerchi presenta le fluttuazioni maggiori e le stime peggiori (e.g. Figure 4.3, 4.4, 4.8); la tecnica di griglia esagonale si mostra la più stabile tra le 5, presentando fluttuazioni in range più piccoli. I risultati delle altre 3 tecniche sono generalmente coerenti nel range del gruppo di stabilità delle stime da intensità I_{corr} di ciascun terremoto, tuttavia ciascuna di esse presenta casi di ampie fluttuazioni: celle di quadrati (Figure 4.4, 4.5), mista cerchi ed esagoni (Figura 4.8), DBSCAN (Figure 4.5, 4.6). In 3 soli terremoti (Figure 4.3, 4.5, 4.8), indipendentemente dai metodi statistici, si può individuare una gerarchia tra le tecniche che è tuttavia differente per ciascuno degli eventi.

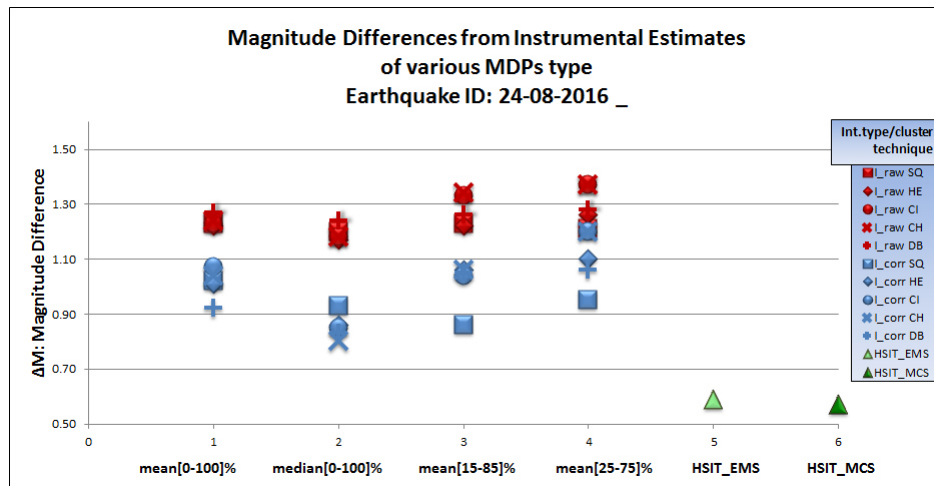


Figura 4.1: Risultati per la magnitudo dell'evento 24/08/2016. Per dettagli simbolologia vedere testo. Le stime da I_{raw} sono comprese circa tra 1.2 e 1.4 unità di magnitudo (u.m.); la sottostima delle I_{corr} è generalmente compresa tra le 0.8 e le 1.1 u.m. ; le differenze HSIT sono le migliori e non superano le 0.6 u.m.. Si osserva minore dispersione tra le stime con le 5 tecniche se si utilizzano i metodi statistici con la base dati completa [0-100]%

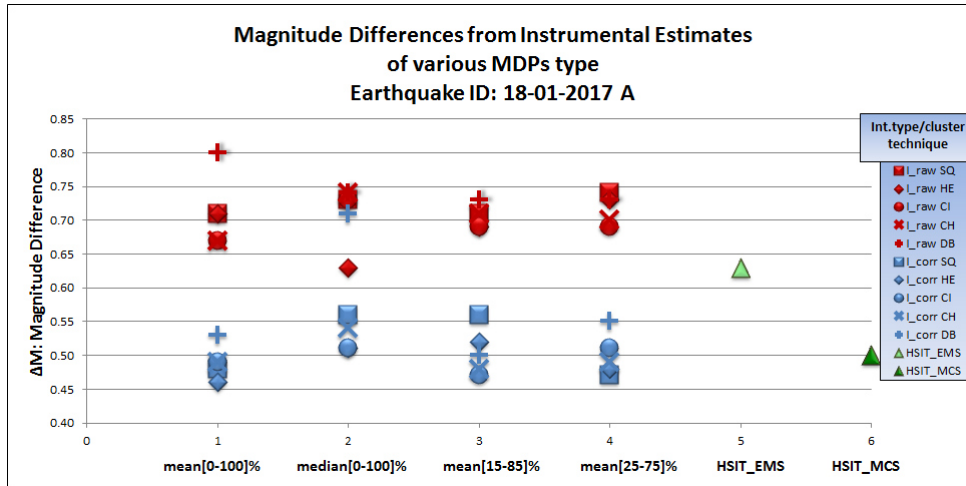


Figura 4.2: Risultati per la magnitudo dell'evento 18/01/2017 A. Per dettagli simbologia vedere testo. La sottostima delle I_{raw} è generalmente compresa in $[0.65, 0.75]$ unità di magnitudo, la sottostima delle I_{corr} in $[0.45, 0.55]$ u.m.. Le differenze HSIT sono comprese in $[0.5, 0.65]$ u.m.; $HSIT_{MCS}$ è nel range delle stime da I_{corr} , mentre $HSIT_{EMS}$ è peggiore di queste. Per le I_{corr} , tra le 5 tecniche, la tecnica DBSCAN presenta generalmente sottostime maggiori, con un massimo in corrispondenza della mediana.

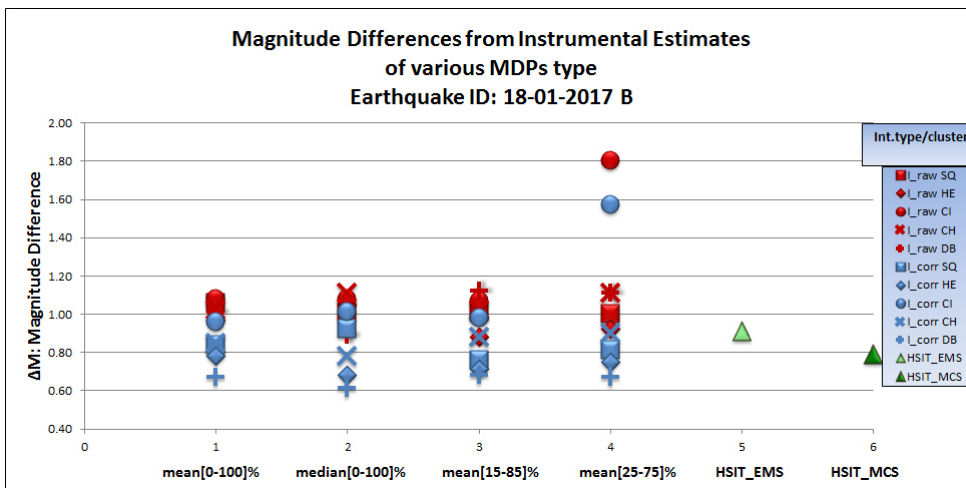


Figura 4.3: Risultati per la magnitudo per l'evento 18/01/2017 B. Per dettagli simbologia vedere testo. La sottostima delle I_{raw} è indicativamente compresa in $[0.9, 1.1]$ unità di magnitudo, la sottostima delle I_{corr} in $[0.6, 1.0]$ u.m.. Entrambe le stime HSIT sono nel range di quelle da I_{corr} . Indipendentemente dall'indice statistico usato, tra le 5 tecniche la tecnica DBSCAN stima meglio e la tecnica dei cerchi stima peggio. Rispetto ai gruppi di stabilità si osserva una fluttuazione della stima effettuata con la tecnica dei cerchi con MDPs da media 25/75.

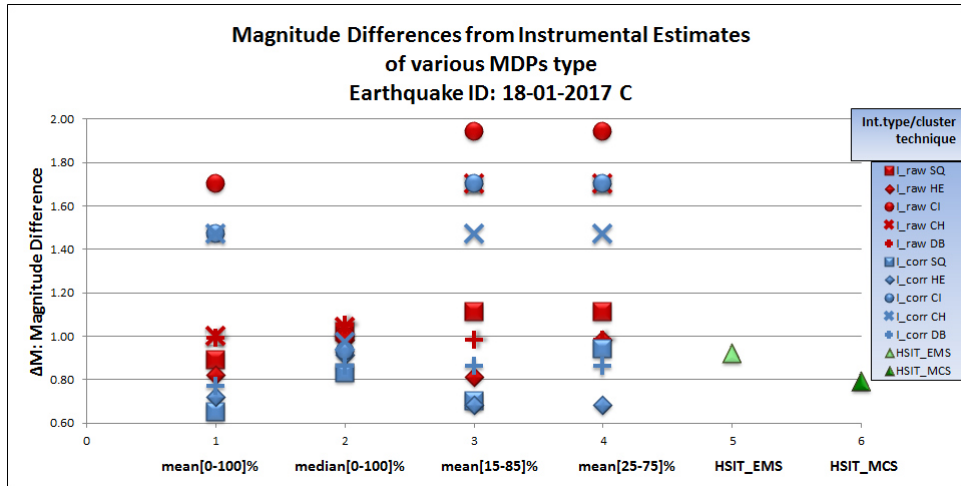


Figura 4.4: Risultati per la magnitudo dell'evento 18/01/2017 C. Per dettagli simbologia vedere testo. La tecnica dei cerchi e la tecnica mista, entrambe con MDPs espressi dalle medie sui diversi percentili, sottostimano sistematicamente la magnitudo molto più che le altre tecniche, da 1.4 fino a 2.0 unità di magnitudo; questo non si verifica per le stime da MDPs espressi con la mediana, indice per il quale si ha dispersione minore tra le stime effettuate con tutte e 5 le tecniche. Le stime con i diversi tipi di media risultano più disperse che quelle con la mediana, ma forniscono valori migliori. Le tecniche di griglia di quadrati e di esagoni e la tecnica DBSCAN forniscono stime comprese in $[0.6, 1.0]$ u.m.; anche le stime HSIT sono comprese in questo range.

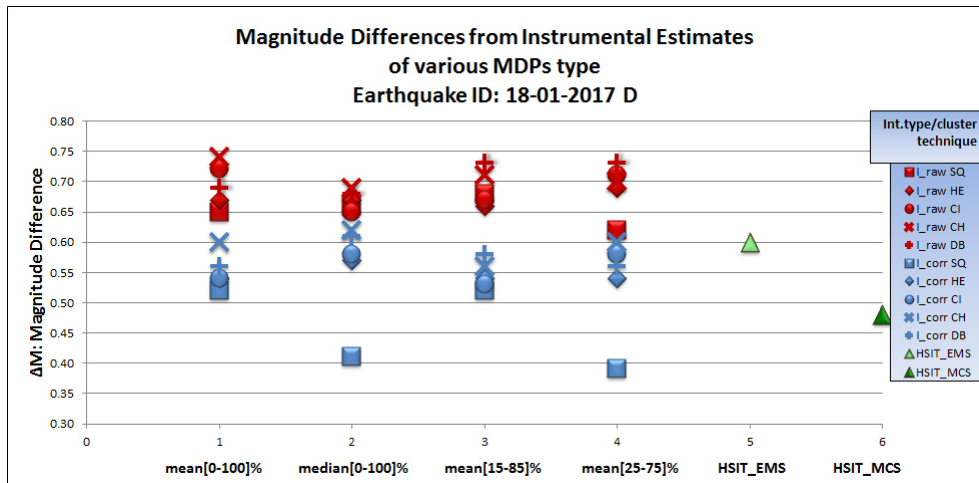


Figura 4.5: Risultati per la magnitudo per l'evento 18/01/2017 D. Per dettagli simbologia vedere testo. Le differenze delle I_{raw} sono indicativamente comprese in $[0.60, 0.75]$ unità di magnitudo, le differenze delle I_{corr} in $[0.50, 0.60]$ u.m., con l'eccezione della tecnica di quadrati che in corrispondenza di mediana e media 25/75 migliora la stima di circa 0.1 u.m.. Entrambe le stime HSIT sono nel range di quelle da I_{corr} .

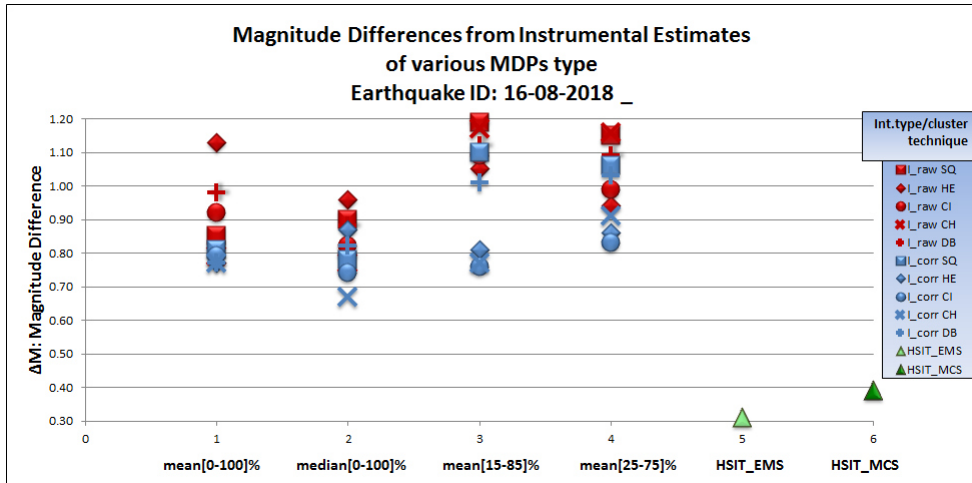


Figura 4.6: Risultati per la magnitudo per l'evento 16/08/2018. Per dettagli simbologia vedere testo. Le differenze di magnitudo da I_{corr} sono generalmente comprese in $[0.7,0.9]$ unità di magnitudo; le differenze di magnitudo da I_{raw} sono invece maggiormente disperse e comprese in $[0.7,1.2]$ u.m.. Si osserva maggiore dispersione per le stime relative alle medie decimate che per le stime relative a media e mediana semplici. In corrispondenza del metodo della media semplice si ha un ottimo accordo tra le 5 tecniche attorno a 0.8 u.m.. Entrambe le stime HSIT sono nettamente migliori delle stime EMCS e non superano le 0.4 u.m..

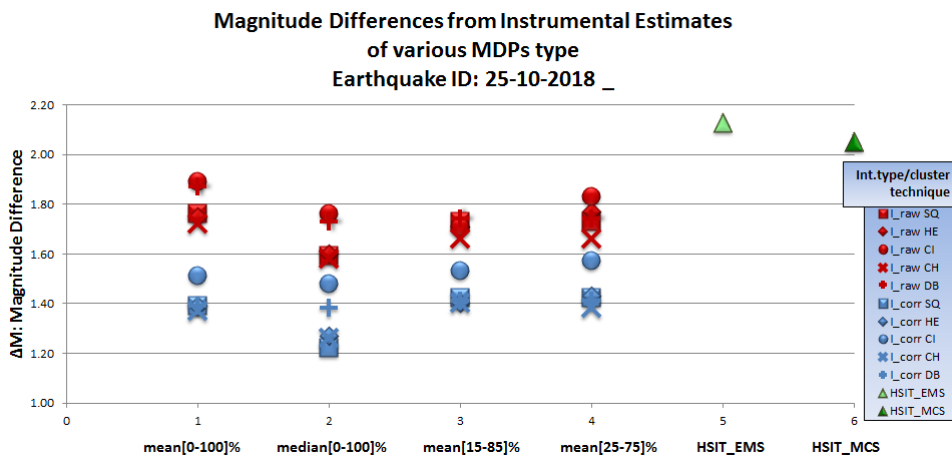


Figura 4.7: Risultati per la magnitudo dell'evento 25/10/2018. Per dettagli simbologia vedere testo. Le differenze da I_{raw} sono circa comprese tra 1.6 e 1.9 unità di magnitudo; le differenze da I_{corr} sono generalmente comprese tra le 1.2 e le 1.6 unità; le differenze HSIT sono le peggiori e superano le 2 unità di magnitudo. Per le stime da I_{corr} effettuate con le 2 tecniche di griglia, con la tecnica mista e con la DBSCAN sono generalmente in accordo; la tecnica dei cerchi stima sistematicamente peggio che le altre. I 3 metodi statistici delle media, a parità di tecnica di clustering forniscono risultati quasi sovrapponibili.

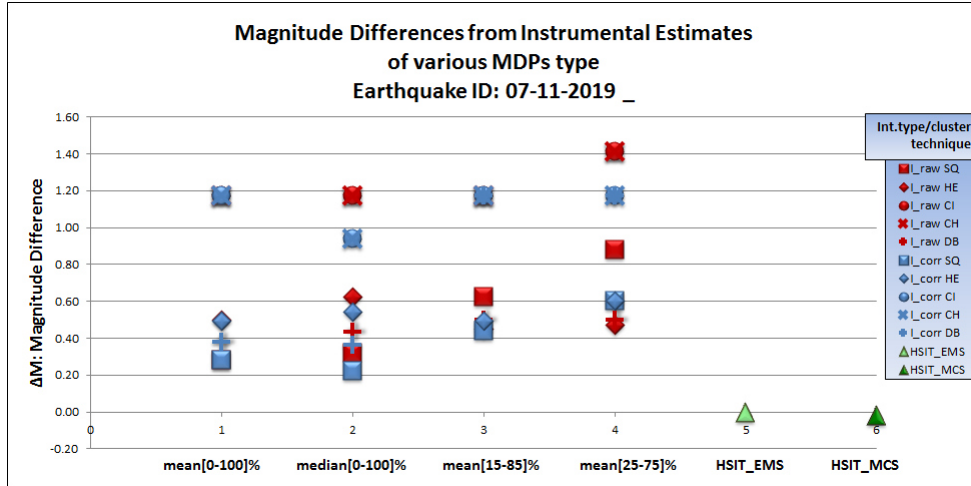


Figura 4.8: Risultati per la magnitudo dell'evento 07/11/2019. Per dettagli simbologia vedere testo. Le stime HSIT sono sovrapponibili alla stima strumentale; in particolare $HSIT_{EMS}$ coincide con questa e $HSIT_{MCS}$ la sovrastima di 0.02 unità. Per ciascun metodo statistico i valori risultano in un range di ampiezza di circa 1 unità di magnitudo. Considerando le I_{corr} , con la tecnica dei cerchi (CI) e la tecnica mista (CH) si ottengono le stime peggiori; invece con le 2 tecniche di griglia (SQ, HE) il range delle differenze si assesta in $[0.2, 0.6]$ u.m.; anche la tecnica DBSCAN (DB) fornisce stime in quest'ultimo range se si usano medie e mediane semplici.

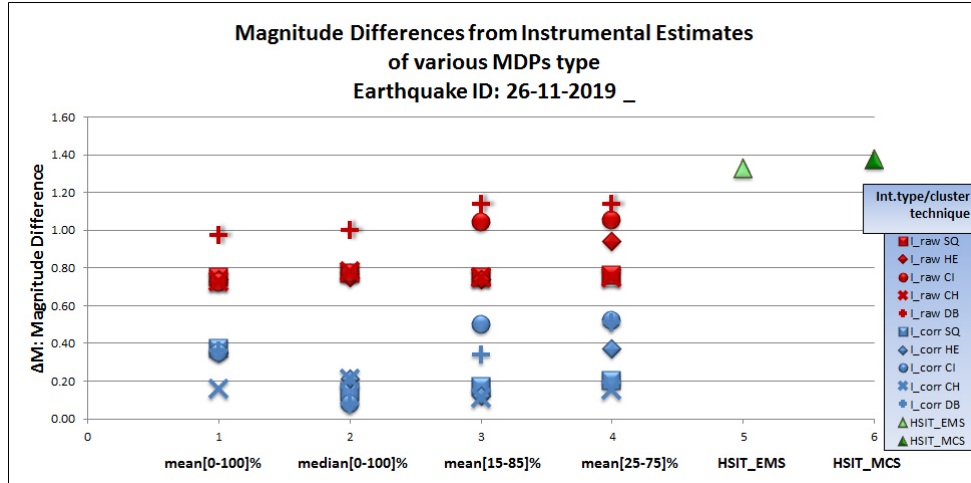


Figura 4.9: Risultati per la magnitudo dell'evento 26/11/2019. Per dettagli simbologia vedere testo. Le differenze da I_{raw} sono circa comprese tra 0.7 e 1.0 unità di magnitudo; le differenze da I_{corr} sono generalmente comprese tra 0.1 e 0.4 u.m.; le differenze HSIT sono le peggiori e superano le 1.3 u.m.. Tra le stime ottenute da intensità I_{corr} quelle effettuate con il metodo delle mediane sono generalmente migliori e mostrano maggiore stabilità tra le tecniche; buona stabilità mostrano anche le stime effettuate con il metodo di media; la dispersione è maggiore per gli MDPs rappresentati con le medie decimate 15/85, 25/75.

La Tabella 4.2 mostra i valori delle stime di magnitudo equivalente e delle differenze rispetto al riferimento strumentale. Considerando le sezioni della tabella relative

alle stime da dati EMSC di tutti i terremoti, si nota una prevalenza di migliori stime di terremoto (sfondo verde chiaro) in corrispondenza della colonna delle *mediane*, come metodo utilizzato per la rappresentazione degli MDPs; l'uso del metodo di *media* corrisponde alla migliore stima in soli 2 terremoti (18/01/2017 A e 18/01/2017 C); mentre nel terremoto 18/01/2017 D la migliore stima si ottiene utilizzando la *media 25/75*. Le stime peggiori (sfondo rosso chiaro) si ottengono prevalentemente con MDPs rappresentati dalle *medie decimate 15/85* e *25/75*, ad eccezione dei terremoti 18/01/2017 A e 18/01/2017 D per i quali la peggiore stima si ottiene utilizzando la *mediana*; per l'evento 07/11/2019, assieme alle medie decimate, anche l'indice *media* semplice è associato alle stime peggiori.

Tabella 4.2:
 Risultati per le magnitudo
 ottenuti con il “metodo
 0” di Boxer applicato agli
 MDPs costruiti con le
 diverse tecniche di
 clustering e metodi
 statistici. Per dettagli
 simbologia e contenuto si
 fa riferimento al testo di
 inizio sezione
 “Magnitudo”.

		Values				Differences			
		Mean	Median	Mean [15-85]%	Mean [25-75]%	Mean	Median	Mean [15-85]%	Mean [25-75]%
MAGNITUDE									
Boxer: MTO									
24/08/2016 #IDPs= 1082									
Lat_Inst= 42.698	Lon_Inst= 13.234								
M_Inst=6.0	#MDPs	M	M	M	M	ΔM	ΔM	ΔM	ΔM
SQ	49	4.98±0.19	5.07±0.16	5.14±0.21	5.05±0.32	1.02	0.93	0.86	0.95
HE	47	4.99±0.19	5.14±0.16	4.94±0.27	4.90±0.28	1.01	0.86	1.06	1.10
CI	30	4.93±0.22	5.15±0.21	4.96±0.30	4.80±0.30	1.07	0.85	1.04	1.20
CH	40	4.96±0.23	5.20±0.18	4.94±0.24	4.80±0.26	1.04	0.80	1.06	1.20
DB	38	5.08±0.23	5.17±0.18	4.95±0.27	4.94±0.28	0.92	0.83	1.05	1.06
HSIT ESTIMATES									
		M				ΔM			
HSIT_EMS	339	5.41±0.06				0.59			
HSIT_MCS	336	5.43±0.06				0.57			
18/01/2017 A #IDPs= 916									
Lat_Inst= 42.545	Lon_Inst= 13.277								
M_Inst=5.1	#MDPs	M	M	M	M	ΔM	ΔM	ΔM	ΔM
SQ	46	4.62±0.18	4.54±0.16	4.54±0.23	4.63±0.19	0.48	0.56	0.56	0.47
HE	44	4.64±0.20	4.59±0.17	4.58±0.20	4.62±0.20	0.46	0.51	0.52	0.48
CI	25	4.61±0.32	4.59±0.22	4.63±0.28	4.59±0.27	0.49	0.51	0.47	0.51
CH	40	4.61±0.22	4.56±0.17	4.62±0.26	4.61±0.29	0.49	0.54	0.48	0.49
DB	34	4.57±0.22	4.39±0.18	4.60±0.22	4.55±0.23	0.53	0.71	0.50	0.55
HSIT ESTIMATES									
		M				ΔM			
HSIT_EMS	97	4.47±0.10				0.63			
HSIT_MCS	101	4.60±0.10				0.50			
18/01/2017 B #IDPs= 621									
Lat_Inst= 42.531	Lon_Inst= 13.284								
M_Inst=5.5	#MDPs	M	M	M	M	ΔM	ΔM	ΔM	ΔM
SQ	22	4.66±0.32	4.58±0.23	4.74±0.39	4.69±0.31	0.84	0.92	0.76	0.81
HE	22	4.72±0.33	4.82±0.23	4.79±0.33	4.75±0.35	0.78	0.68	0.71	0.75
CI	9	4.54±0.36	4.49±0.35	4.52±0.44	3.93	0.96	1.01	0.98	1.57
CH	15	4.65±0.39	4.72±0.28	4.62±0.49	4.60±0.43	0.85	0.78	0.88	0.90
DB	14	4.83±0.50	4.89±0.29	4.82±0.49	4.83±0.41	0.67	0.61	0.68	0.67
HSIT ESTIMATES									
		M				ΔM			
HSIT_EMS	69	4.59±0.13				0.91			
HSIT_MCS	72	4.71±0.12				0.79			
18/01/2017 C #IDPs= 339									
Lat_Inst= 42.503	Lon_Inst= 13.277								
M_Inst=5.4	#MDPs	M	M	M	M	ΔM	ΔM	ΔM	ΔM
SQ	11	4.75±0.46	4.57±0.35	4.70±0.37	4.46±0.38	0.65	0.83	0.70	0.94
HE	11	4.68±0.58	4.49±0.33	4.72±0.39	4.72±0.39	0.72	0.91	0.68	0.68
CI	4	3.93	4.47±0.53	3.70±0.53	3.70	1.47	0.93	1.70	1.70
CH	7	3.93	4.43±0.40	3.93±0.40	3.93	1.47	0.97	1.47	1.47
DB	11	4.63±0.46	4.54±0.35	4.54±0.47	4.54±0.47	0.77	0.86	0.86	0.86
HSIT ESTIMATES									
		M				ΔM			
HSIT_EMS	73	4.48±0.12				0.92			
HSIT_MCS	81	4.61±0.11				0.79			
18/01/2017 D #IDPs= 772									
Lat_Inst= 42.473	Lon_Inst= 13.275								
M_Inst=5.0	#MDPs	M	M	M	M	ΔM	ΔM	ΔM	ΔM
SQ	37	4.48±0.21	4.59±0.18	4.48±0.23	4.61±0.23	0.52	0.41	0.52	0.39
HE	43	4.46±0.18	4.43±0.17	4.46±0.21	4.46±0.25	0.54	0.57	0.54	0.54
CI	23	4.46±0.26	4.42±0.23	4.47±0.30	4.42±0.32	0.54	0.58	0.53	0.58
CH	36	4.40±0.20	4.38±0.18	4.44±0.25	4.40±0.30	0.60	0.62	0.56	0.60
DB	28	4.44±0.22	4.39±0.20	4.42±0.24	4.44±0.25	0.56	0.61	0.58	0.56
HSIT ESTIMATES									
		M				ΔM			
HSIT_EMS	37	4.40±0.17				0.60			
HSIT_MCS	38	4.52±0.18				0.48			
16/08/2018 #IDPs= 541									
Lat_Inst= 41.874	Lon_Inst= 14.865								
M_Inst=5.1	#MDPs	M	M	M	M	ΔM	ΔM	ΔM	ΔM
SQ	22	4.29±0.36	4.32±0.21	4.00±0.22	4.04±0.22	0.81	0.78	1.10	1.06
HE	17	4.33±0.26	4.23±0.26	4.29±0.30	4.24±0.25	0.77	0.87	0.81	0.86
CI	13	4.31±0.30	4.36±0.30	4.34±0.27	4.27±0.49	0.79	0.74	0.76	0.83
CH	19	4.33±0.27	4.43±0.23	4.33±0.23	4.19±0.43	0.77	0.67	0.77	0.91
DB	15	4.33±0.29	4.28±0.26	4.09±0.23	4.07±0.24	0.77	0.82	1.01	1.03
HSIT ESTIMATES									
		M				ΔM			
HSIT_EMS	199	4.79±0.13				0.31			
HSIT_MCS	201	4.71±0.07				0.39			
25/10/2018 #IDPs= 1883									
Lat_Inst= 37.492	Lon_Inst= 20.595								
M_Inst=6.8	#MDPs	M	M	M	M	ΔM	ΔM	ΔM	ΔM
SQ	70	5.41±0.13	5.58±0.15	5.38±0.15	5.38±0.15	1.39	1.22	1.42	1.42
HE	74	5.41±0.13	5.53±0.13	5.40±0.15	5.37±0.15	1.39	1.27	1.40	1.43
CI	63	5.29±0.14	5.32±0.15	5.27±0.15	5.23±0.16	1.51	1.48	1.53	1.57
CH	89	5.43±0.11	5.54±0.12	5.40±0.13	5.42±0.12	1.37	1.26	1.40	1.38
DB	52	5.41±0.16	5.42±0.16	5.39±0.17	5.39±0.17	1.39	1.38	1.41	1.41
HSIT ESTIMATES									
		M				ΔM			
HSIT_EMS	124	4.67±0.09				2.13			
HSIT_MCS	125	4.75±0.11				2.05			
07/11/2019 #IDPs= 343									
Lat_Inst= 41.776	Lon_Inst= 13.604								
M_Inst=4.4	#MDPs	M	M	M	M	ΔM	ΔM	ΔM	ΔM
SQ	14	4.12±0.35	4.18±0.26	3.96±0.30	3.80±0.27	0.28	0.22	0.44	0.60
HE	11	3.91±0.26	3.86±0.32	3.91±0.24	3.80±0.23	0.49	0.54	0.49	0.60
CI	4	3.23	3.46	3.23	3.23	1.17	0.94	1.17	1.17
CH	4	3.23	3.46	3.23	3.23	1.17	0.94	1.17	1.17
DB	6	4.02±0.34	4.04±0.52	3.23	3.23	0.38	0.36	1.17	1.17
HSIT ESTIMATES									
		M				ΔM			
HSIT_EMS	104	4.40±0.12				0			
HSIT_MCS	105	4.42±0.11				-0.02			
26/11/2019 #IDPs= 4036									
Lat_Inst= 41.371	Lon_Inst= 19.467								
M_Inst=6.2	#MDPs	M	M	M	M	ΔM	ΔM	ΔM	ΔM
SQ	87	5.83±0.13	6.06±0.17	6.03±0.12	6.00±0.11	0.37	0.14	0.17	0.20
HE	92	5.85±0.13	5.99±0.13	6.08±0.13	5.83±0.13	0.35	0.21	0.12	0.37
CI	54	5.85±0.18	6.12±0.15	5.70±0.17	5.68±0.18	0.35	0.08	0.50	0.52
CH	84	6.04±0.11	5.99±0.13	6.09±0.12	6.05±0.13	0.16	0.21	0.11	0.15
DB	58	5.84±0.18	6.12±0.15	5.86±0.18	5.69±0.17	0.36	0.08	0.34	0.51
HSIT ESTIMATES									
		M				ΔM			
HSIT_EMS	207	4.87±0.10				1.33			
HSIT_MCS	202	4.82±0.07				1.38			

Analizzando, per ciascun metodo statistico (media, mediana, ecc), le stime migliori e peggiori ottenute considerando complessivamente tutte e 5 le tecniche e tutti e 9 i terremoti, la *mediana* fornisce i risultati migliori (Tabella 4.3b). Da questo conteggio la *media decimata 25/75* totalizza il peggior rapporto “migliore”/”peggiore” ($r_s = 0.11$), la *media decimata 15/85* è al penultimo posto ($r_s = 0.79$), la semplice *media* è al secondo posto ($r_s = 1.14$).

Per alcuni terremoti si osserva coincidenza tra stime effettuate con tecniche o metodi differenti.

a		
Score stime di Magnitudo Boxer: MTO		
Tecnica	Migliore	Peggior
QUADRATI	4	1
ESAGONI	1	0
CERCHI	1	6
MISTA	2	3
DBSCAN	2	2
Tecnica	Migliore	Peggior
HSIT_EMS	3	2
HSIT_MCS	3	2

b		
Score stime di Magnitudo Boxer: MTO		
Metodo statistico	Migliore	Peggior
Media	8	7
Mediana	23	13
Media [15-85]%	11	14
Media [25-75]%	3	27

Tabelle 4.3a, b: Score delle stime di magnitudo migliori e peggiori in relazione alle tecniche di clustering e ai metodi statistici utilizzati per la rappresentazione dei relativi MDPs. Per dettagli sulle modalità di conteggio vedere testo di inizio sezione “Magnitudo”.

Le differenze di magnitudo (Tabella 4.2) sono generalmente comprese nell’intervallo $[0, 2]$ unità, da cui emerge la sistematica sottostima sia da dati HSIT che da dati EMSC; tuttavia considerando i soli metodi di *media* e *mediana* semplici e trascurando le stime di HSIT per i due terremoti con epicentro strumentale localizzato al di fuori dal territorio italiano (Grecia 25/10/2018, Albania 26/11/2019) l’intervallo di valori si stabilizza in $[0.5, 1.0]$. Fa eccezione il terremoto dell’Albania 26/11/2019 (Figura 4.9) per il quale le stime da dati EMSC approssimano meglio la magnitudo strumentale che per tutti gli altri eventi: in particolare si ha complessivamente $0.08 \leq \Delta M \leq 0.52$, mentre per le sole stime da mediana $\Delta M \leq 0.21$. Questo evento ha infatti magnitudo strumentale più elevata degli altri, $M = 6.2$, alla quale solitamente, a brevi distanze dai centri abitati (circa $3Km$ da Durazzo e circa $20Km$ dalla capitale Tirana), corrispondono intensità al di sopra della soglia di danno. Queste caratteristiche, assieme al numero elevato di dati (IDPs) raccolti (Bossu et al., 2020), hanno permesso la costruzione di un numero elevato di MDPs per tecnica (fino a 92, per la griglia esagonale). Per gli altri 2 eventi di elevata magnitudo del dataset ($M=6.0$ 24/08/2016, $M=6.8$ 25/10/2018) non si sono ottenuti risultati altrettanto accurati poichè la regione del near-field non era sufficientemente rappresentata dagli

IDPs: in particolare il primo si è verificato nel periodo iniziale di attivazione dell'app LastQuake di EMSC per la raccolta degli IDPs, allora meno diffusa del sito HSIT ($\#IDPs_{EMSC} = 1082$, Figura 3.6; $\#IDPs_{HSIT} = 12073$, <https://www.hsit.it/>); il secondo invece ha avuto epicentro in mare a circa 30 Km dalla costa, per cui anche nel far-field non si hanno IDPs per un ampio angolo attorno all'epicentro (Figura 4.10a).

In generale nessuna tra le tecniche di clustering produce risultati nettamente migliori delle altre. Tutte forniscono stime tra loro consistenti entro le incertezze stimate a meno di qualche fluttuazione, prevalente per la tecnica dei cerchi. Tra le due intensità MCS e EMS disponibili per HSIT, la prima dà risultati più vicini alla magnitudo di riferimento rispetto a quanto avviene usando le intensità EMS per 6 terremoti su 9. Se si considera la Tabella 4.3a, sezione superiore, emerge che tra le 5 tecniche di clustering applicate ai dati EMSC la tecnica a celle quadrate totalizza il numero maggiore di migliori stime (4 su 9 possibili); la tecnica dei cerchi attorno alle località totalizza il numero maggiore di peggiori stime (6 su 9 possibili). La tecnica DBSCAN in generale fornisce stime confrontabili con tutti gli altri metodi, e totalizza 2 peggiori stime e 2 migliori stime; di queste ultime 2 stime, 1 coincide con l'evento statisticamente favorevole del terremoto d'Albania 26/11/2019, l'altra coincide con l'evento 18/01/2017 B.

Confrontando per ogni terremoto le stime effettuate da dati HSIT con quelle effettuate da dati EMSC si osserva un generale accordo tra queste. Le prime sono peggiori delle seconde solamente per i 2 terremoti con epicentro prossimo alla costa e al di fuori del territorio italiano, 25/10/2018 (Figura 4.29) e 26/11/2019 (Figura 4.31); le stime da dati HSIT sono invece migliori di quelle da dati EMSC per i 3 eventi 24/08/2016, 16/08/2018, 07/11/2019. I terremoti con stime $HSIT_{EMS}$ migliori rispetto alle stime da dati EMSC coincidono con quelli con stime migliori $HSIT_{MCS}$; lo stesso vale per le stime peggiori. Questo risultato è sintetizzato nella tabella di score 4.3a, sezione inferiore.

In generale, dato un terremoto, il $\#MDPs$ da dati HSIT è di molto superiore al $\#MDPs$ da dati EMSC: per i diversi terremoti il $\#MDPs_{HSIT}$ varia rispetto a $\#MDPs_{EMSC}$ da un valore di circa il doppio fino all'ordine di grandezza superiore, con l'eccezione del terremoto 18/01/2017 D per il quale i 2 $\#MDPs$ sono molto prossimi. Per i 3 terremoti per i quali la stima di HSIT è migliore delle altre il $\#MDPs$ da dati EMSC risulta particolarmente contenuto, ed è compreso circa tra la 7-ma e la 26-esima parte del rispettivo $\#MDPs$ da dati HSIT; per gli altri terremoti invece non risulta inferiore all'8-va parte; inoltre per questi 3 terremoti si osserva una discrepanza di un ordine di grandezza tra i $\#IDPs$ raccolti dalle due istituzioni, rispettivamente $\#IDPs_{EMSC} = 1082, 541, 343$ e $\#IDPs_{HSIT} = 12073, 5181, 3674$.

Considerando gli errori sulle stime da dati EMSC, questi risultano compresi nell'intervallo [0.1, 0.5] unità di magnitudo; rispetto a questi gli errori sulle stime HSIT sono generalmente inferiori e talvolta risultano sul centesimo di u.m.. Generalmente l'incertezza sulla stima diminuisce per numeri di IDPs e di MDPs crescenti: questa si stabilizza in [0.1,0.2] unità di magnitudo per gli eventi maggiormente campionati ($\#IDPs > 1000$), 24/08/2016, 25/10/2018, 26/11/2019. Nei casi di $\#MDPs$ di poche unità Boxer non stima gli errori.

Per uno stesso terremoto le incertezze maggiori si registrano generalmente in corrispondenza delle medie decimate e delle tecniche dei cerchi (CI) e mista (CH), le quali

costruiscono numeri di MDPs inferiori rispetto agli altri metodi e alle altre tecniche (e.g. 18/01/2017 B, 18/01/2017 C, 18/01/2017 D, 16/08/2018).

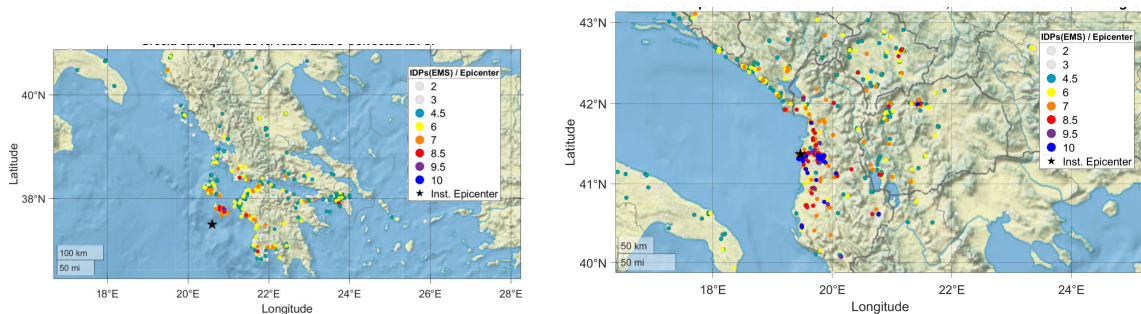
Si osserva inoltre che al crescere del $\#IDPs$ migliora anche la precisione delle diverse stime: lo scarto tra le differenze massima e minima (caselle rosse e verdi, colonna sinistra) supera l'unità per terremoti con $\#IDPs \sim 300$, si assesta intorno a 0.4 u.m. per gli altri; se invece si considerano i soli terremoti con $\#IDPs > 700$ e le sole stime effettuate con i metodi di media e mediana semplici, lo scarto tra le differenze massima e minima è inferiore a 0.3 u.m.. Per tutti i terremoti con $\#IDPs > 500$, considerando separatamente i metodi di media e mediana semplici, le stime delle 5 tecniche risultano coerenti entro le incertezze calcolate.

Magnitudo: Boxer MT1

In questa sezione vengono mostrati i risultati ottenuti con il “metodo 1” di Boxer (MT1, descritto in sintesi al par. 4.2, in dettaglio al par. 1.4) per le magnitudo equivalenti dei 2 terremoti che hanno epicentro strumentale localizzato in mare o prossimo alla costa (25/10/2018 Figura 4.29, 26/11/2019 Figura 4.31). Entrambi gli epicentri strumentali sono localizzati al di fuori del territorio italiano ed in particolare in territorio greco (25/10/2018) e in territorio albanese (26/11/2019).

L'epicentro localizzato in prossimità della linea di costa comporta un'intrinseca assenza di dati macrosismici per un ampio angolo attorno ad esso, sia nel near-field che nel far-field. Questo è riscontrabile nelle distribuzioni di IDPs EMSC e HSIT (Figure 4.10, 4.11). Inoltre la compilazione dei questionari (IDPs) HSIT è attiva solamente per il territorio italiano, perciò per questi si ottiene maggiore disuniformità nella copertura del campo macrosismico.

Per questa ragione per i 2 eventi in questione sono state avviate le stime dei *parametri sintetici* con il “metodo 1” di Boxer, che può essere applicato a terremoti con tali caratteristiche (Gasperini et al., 2010).



(a) Terremoto 25/10/2018.

(b) Terremoto 26/11/2019.

Figura 4.10: Distribuzione di IDPs (intensità I_{corr} , EMSC) per gli eventi indicati.

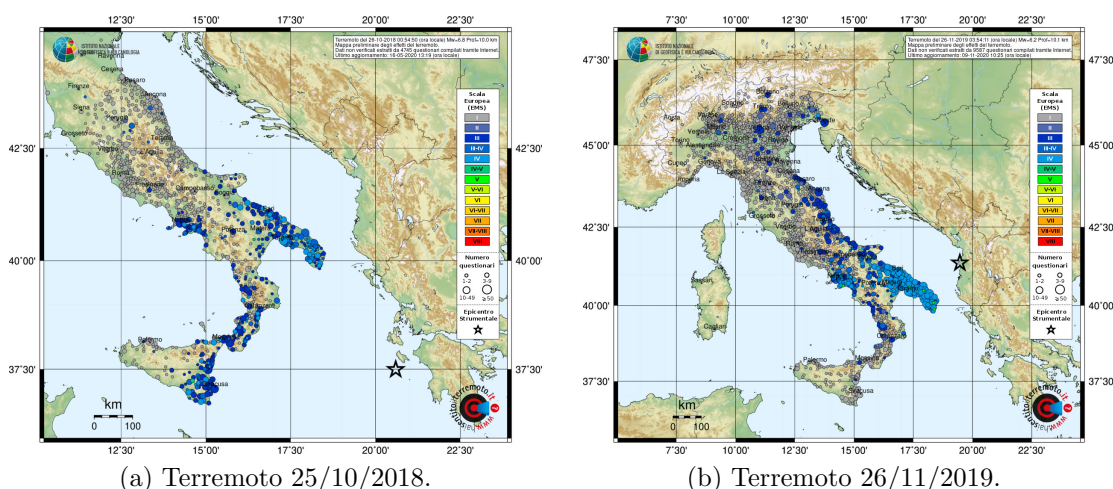


Figura 4.11: Distribuzione di questionari HSIT (intensità EMS) per gli eventi indicati. I dati sono raccolti solo sul territorio italiano e presentano grande disuniformità nella copertura del campo macrosismico.

I risultati dell'applicazione del “metodo 1” sono mostrati in plot e tabelle con le medesime modalità e caratteristiche utilizzate per i risultati del “metodo 0” (per descrizione legende e simbologia vedere inizio sezione “Magnitudo” 4.3.1).

Per entrambi i terremoti con epicentro prossimo alla costa, dal confronto dei risultati ottenuti con il “metodo 1” di Boxer (Figure 4.12, 4.13 e Tabella 4.4) con i rispettivi risultati ottenuti con il “metodo 0” (Figure 4.7, 4.9 e Tabella 4.2) emerge che le stime di magnitudo equivalente sono sovrapponibili. Si confermano la generale sottostima e i migliori risultati delle intensità EMSC corrette I_{corr} rispetto alle I_{raw} . Il “metodo 1” non stima la magnitudo equivalente da dati HSIT a causa del campionamento fortemente disomogeneo del campo macrosismico. Risultato nullo (“-”) si ottiene anche per il terremoto 25/10/2018 in corrispondenza della tecnica dei cerchi usata congiuntamente al metodo statistico *media*.

Gli MDPs rappresentati con il metodo statistico *mediana* sono quelli a cui corrispondono generalmente le stime migliori, sia considerando stime relative ad una singola tecnica di clustering, sia considerando complessivamente tutte le stime per un terremoto.

Le stime effettuate con le 2 tecniche di griglia (di quadrati SQ e di esagoni HE) e con la tecnica mista (CH) sono in ottimo accordo a parità di indice statistico e risultano inoltre le più stabili al variare dell'indice. Rispetto alle stime di queste 3 tecniche (SQ, HE, CH) le stime della tecnica DBSCAN sono maggiormente soggette a fluttuazioni, sebbene spesso coincidano con esse.

La tecnica dei cerchi (CI) e le *medie decimate 15/85*, *25/75* forniscono generalmente i risultati peggiori; inoltre le stime relative alle *medie decimate* presentano maggiore dispersione rispetto alle stime effettuate con gli altri metodi statistici.

La tecnica di griglia di quadrati (SQ) fornisce il miglior risultato per il terremoto 25/10/2018; la tecnica DBSCAN fornisce il miglior risultato per il terremoto 26/11/2019.

Nessuna tra le tecniche di clustering produce risultati nettamente migliori delle altre.

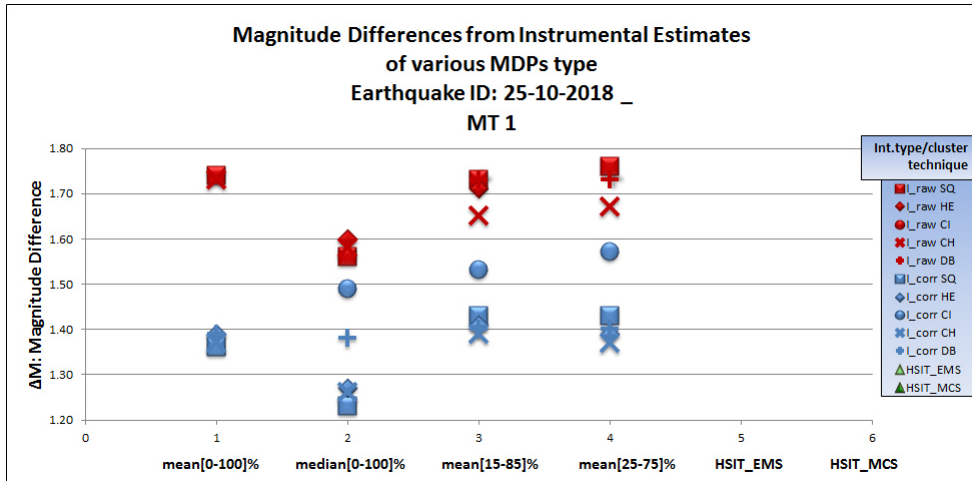


Figura 4.12: Risultati per la magnitudo dell'evento 25/10/2018 ottenuti con il “metodo 1” di Boxer. Per simbologia legenda fare riferimento al testo di inizio sezione “Magnitudo”. Trascurando la tecnica dei cerchi e la tecnica DBSCAN (DB) usata congiuntamente alla mediana, le altre stime a parità di metodo statistico sono in generale in ottimo accordo; sono inoltre in accordo le stime con i 3 diversi tipi di media.

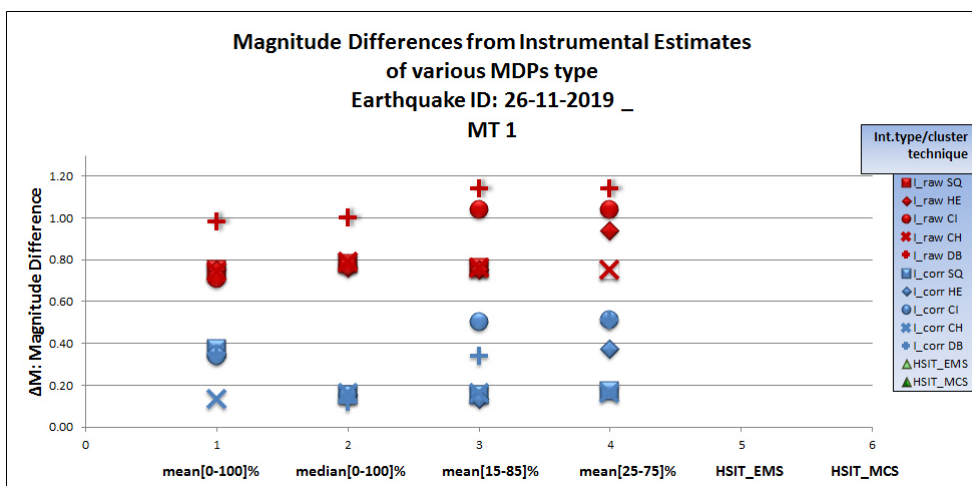


Figura 4.13: Risultati per la magnitudo dell'evento 26/11/2019 con il “metodo 1” di Boxer. Per simbologia legenda fare riferimento al testo di inizio sezione “Magnitudo”. Considerando le sole I_{corr} , la tecnica mista (CH) fornisce le stime più stabili al variare del metodo statistico; segue per stabilità la tecnica di griglia di quadrati (SQ).

MAGNITUDE Boxer: MT1									
Values					Differences				
25/10/2018		#IDPs= 1883	Greece						
Lat_inst= 37.492	Lon_inst= 20.595	Mean	Median	Mean [15-85]%	Mean [25-75]%	Mean	Median	Mean [15-85]%	Mean [25-75]%
M_inst=6.8	#MDPs	M	M	M	M	ΔM	ΔM	ΔM	ΔM
SQ	70	5.44±0.13	5.57±0.13	5.37±0.15	5.37±0.15	1.36	1.23	1.43	1.43
HE	74	5.41±0.13	5.53±0.13	5.39±0.15	5.41±0.14	1.39	1.27	1.41	1.39
CI	63	-	5.31±0.15	5.27±0.15	5.23±0.16	-	1.49	1.53	1.57
CH	89	5.43±0.11	5.54±0.12	5.41±0.13	5.43±0.12	1.37	1.26	1.39	1.37
DB	52	5.41±0.16	5.42±0.16	5.40±0.17	5.40±0.17	1.39	1.38	1.40	1.40
HSIT ESTIMATES									
	#MDPs	M				ΔM			
HSIT_EMS	124	-				-			
HSIT_MCS	125	-				-			
26/11/2019		#IDPs= 4036	Albania						
Lat_inst= 41.371	Lon_inst= 19.467	Mean	Median	Mean [15-85]%	Mean [25-75]%	Mean	Median	Mean [15-85]%	Mean [25-75]%
M_inst=6.2	#MDPs	M	M	M	M	ΔM	ΔM	ΔM	ΔM
SQ	87	5.83±0.13	6.05±0.13	6.05±0.12	6.03±0.12	0.37	0.15	0.15	0.17
HE	92	5.85±0.13	6.05±0.13	6.07±0.12	5.83±0.13	0.35	0.15	0.13	0.37
CI	54	5.86±0.18	6.05±0.15	5.70±0.17	5.69±0.18	0.34	0.15	0.50	0.51
CH	84	6.07±0.11	6.04±0.13	6.04±0.12	6.04±0.14	0.13	0.16	0.16	0.16
DB	58	5.85±0.18	6.08±0.15	5.86±0.18	5.69±0.17	0.35	0.12	0.34	0.51
HSIT ESTIMATES									
	#MDPs	M				ΔM			
HSIT_EMS	207	-				-			
HSIT_MCS	202	-				-			

Tabella 4.4: Risultati per le magnitudo ottenuti con il “metodo 1” di Boxer applicato agli MDPs costruiti con le diverse tecniche di clustering e metodi statistici. Per dettagli simbologia e contenuto si fa riferimento al testo di inizio sezione “Magnitudo”.

4.3.2 Localizzazione epicentrale

In questa sezione vengono presentati per i 9 terremoti analizzati i risultati delle stime delle coordinate epicentrali effettuate con il “metodo 0” (MT0) di Boxer a partire dalle distribuzioni di MDPs costruite dai dati macrosismici EMSC e HSIT. A seguire, nella sottosezione “Localizzazione epicentrale: Boxer MT1”, sono presentate le stime delle coordinate epicentrali effettuate con il “metodo 1” per i soli 2 terremoti con epicentro strumentale prossimo alla costa.

Per ciascun terremoto sono elaborate con Boxer 42 distribuzioni di MDPs (Macroseismic Data Points) diversamente costruite, a cui corrispondono altrettante stime di localizzazione epicentrale (coppie di coordinate (lat, lon)).

2 di queste distribuzioni sono costruite a partire da dati IDPs raccolti tramite i questionari HSIT, espressi sia in intensità EMS sia in MCS, e pubblicati da HSIT già clusterizzati in MDPs con la medesima procedura (“tecnica” HSIT= Δ) descritta in Tosi et al., 2015.

Le altre 40 distribuzioni sono costruite a partire dai dati IDPs di EMSC che sono espressi in 2 tipologie di intensità: intensità EMS I_{raw} scelte dagli utenti mediante l’app LastQuake² e intensità I_{corr} , corrispondenti alle I_{raw} corrette secondo Bossu et al., 2017. Per ogni terremoto le IDPs acquisite da EMSC sono quindi clusterizzate (raggruppate) in MDPs, mediante l’utilizzo delle 5 “tecniche (di clustering)” descritte al cap. 3 (\square =celle quadrate, \diamond =celle esagonali, \circ =cerchi attorno a località, \times =mista cerchi ed esagoni, $+$ =DBSCAN), costituendo una “distribuzione di MDPs”. A ciascu-

²<https://www.emsc-csem.org/>

na distribuzione è assegnata un'intensità che sostituisce l'intensità classica assegnata alle località dell'area interessata dagli effetti macrosismici; per l'assegnazione dell'intensità sono applicati 4 differenti “metodi statistici” (media, mediana, media[15,85]%, media[25/75]%) alle intensità IDPs costituenti gli MDPs. Si hanno quindi per ogni terremoto 40 (2 intensità, 5 tecniche, 4 metodi) distribuzioni di MDPs da dati EMSC e 2 distribuzioni da dati HSIT (EMS, MCS) elaborate con il “metodo 0” di Boxer.

I risultati sono presentati nella stessa forma e modalità dei risultati per le magnitudo equivalenti (par. 4.3.1) e cioè mostrando per ciascun terremoto: 1 plot delle distanze tra gli epicentri macrosismici e l'epicentro strumentale di riferimento (Figure 4.14-4.22); 1 tabella riassuntiva delle stime delle coordinate e delle distanze epicentrali rispetto al riferimento strumentale (Tabella 4.5); 2 tabelle di “score/punteggi” assegnati alle tecniche di clustering e ai metodi statistici – utilizzati per la costruzione delle distribuzioni di MDPs – in relazione alle stime migliori e peggiori a cui sono associati (Tabelle 4.6a, b); 1 mappa delle migliori localizzazioni epicentrali ottenute con le diverse tecniche e metodi statistici (Figure 4.24-4.34).

Per descrizione legende e simbologia vedere inizio sezione “Magnitudo” 4.3.1.

Analizzando i plot delle distanze epicentrali (Figure 4.14-4.22) e confrontando le sole stime ottenute a partire dalle intensità EMSC non si osservano, come avviene invece per i risultati delle magnitudo, i 2 netti gruppi di stabilità relativi alle intensità corrette I_{corr} (simboli blu) e alle intensità grezze I_{raw} (simboli rossi) (par. 4.3.1; Bossu et al., 2017), né si osservano sistematici risultati migliori di una o dell'altra intensità di base a parità di tecnica di clustering e di metodo statistico.

Rispetto alle 2 tipologie di intensità EMSC i risultati delle distanze epicentrali presentano diverse fluttuazioni, grande variabilità e ampia dispersione (in un range di circa 50-70Km per ogni terremoto); inoltre vi sono diversi casi di sovrapposibilità a parità di tecnica e di metodo statistico (e.g. Figure 4.14, 4.20, 4.21, 4.22), con una leggera prevalenza di risultati migliori per le intensità I_{corr} (e.g. Figura 4.19).

Per questa ragione, in coerenza con la scelta operata per l'analisi dei risultati delle magnitudo, anche per il resto dell'analisi sugli epicentri ci si riferisce alle sole stime effettuate a partire dalle intensità corrette I_{corr} .

Osservando per tutti i terremoti le stime ottenute utilizzando i dati EMSC si notano diversi casi di sovrapposibilità tra stime effettuate da MDPs costruiti con tecniche differenti a parità di metodo statistico, e viceversa (e.g. Figure 4.19, 4.21). I casi di sovrapposibilità diventano molto frequenti per i terremoti con valori di magnitudo strumentale e numero di IDPs (EMSC) più elevati (Tabella 4.1, Figure 4.20, 4.22) e per alcuni terremoti in corrispondenza del metodo statistico di mediana (Figure 4.14, 4.19, 4.21).

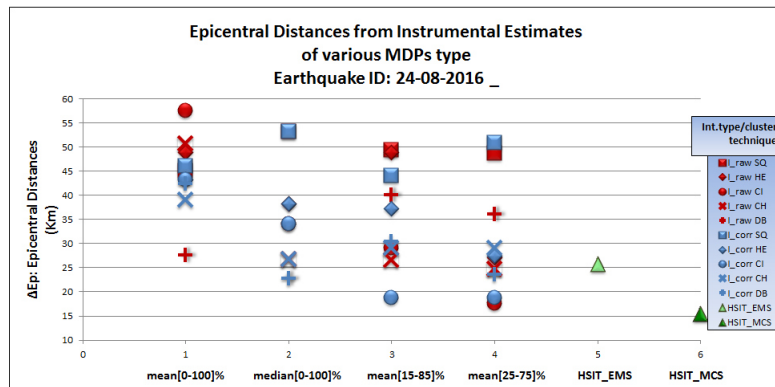


Figura 4.14: Terremoto 24/08/2016. Risultati per le localizzazioni epicentrali ottenuti con il “metodo 0” di Boxer. Per descrizione legenda e simbologia vedere inizio sezione “Magnitudo” 4.3.1.

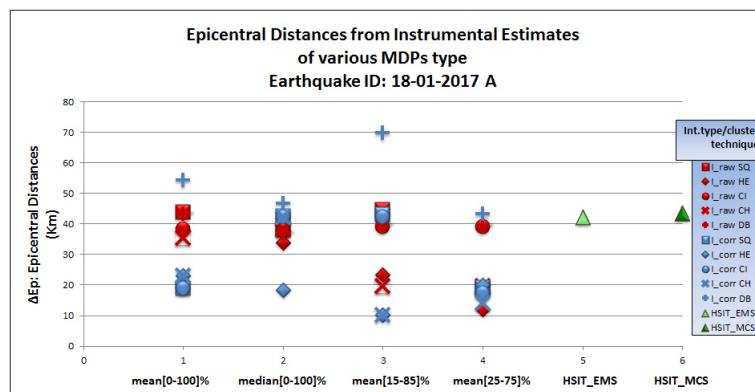


Figura 4.15: Terremoto 18/01/2017 A. Risultati per le localizzazioni epicentrali ottenuti con il “metodo 0” di Boxer. Per descrizione legenda e simbologia vedere inizio sezione “Magnitudo” 4.3.1.

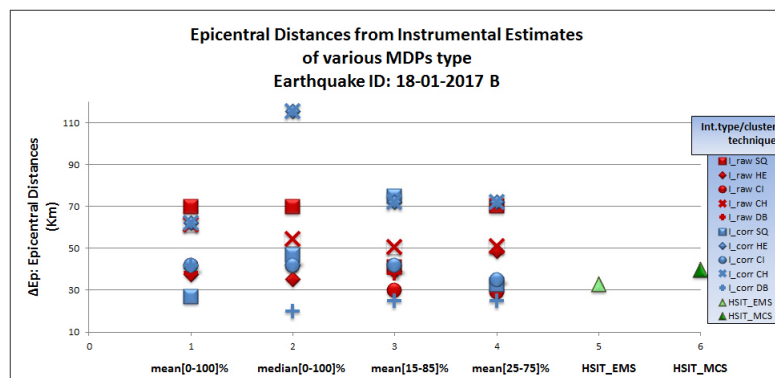


Figura 4.16: Terremoto 18/01/2017 B. Risultati per le localizzazioni epicentrali ottenuti con il “metodo 0” di Boxer. Per descrizione legenda e simbologia vedere inizio sezione “Magnitudo” 4.3.1.

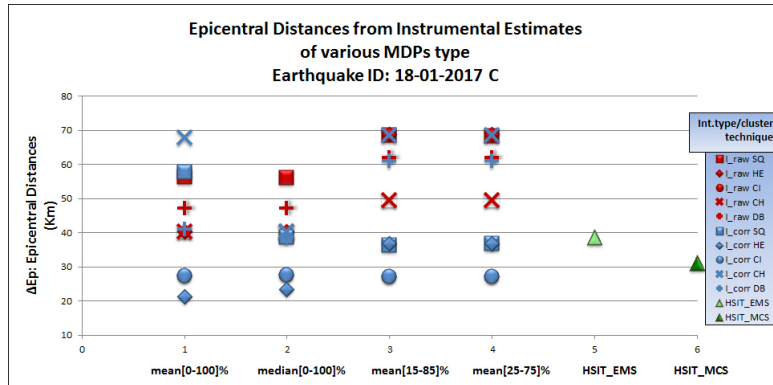


Figura 4.17: Terremoto 18/01/2017 C. Risultati per le localizzazioni epicentrali ottenuti con il “metodo 0” di Boxer. Per descrizione legenda e simbologia vedere inizio sezione “Magnitudo” 4.3.1.

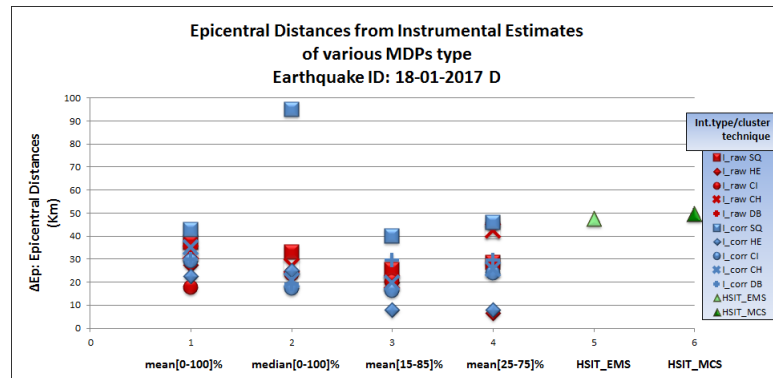


Figura 4.18: Terremoto 18/01/2017 D. Risultati per le localizzazioni epicentrali ottenuti con il “metodo 0” di Boxer. Per descrizione legenda e simbologia vedere inizio sezione “Magnitudo” 4.3.1.

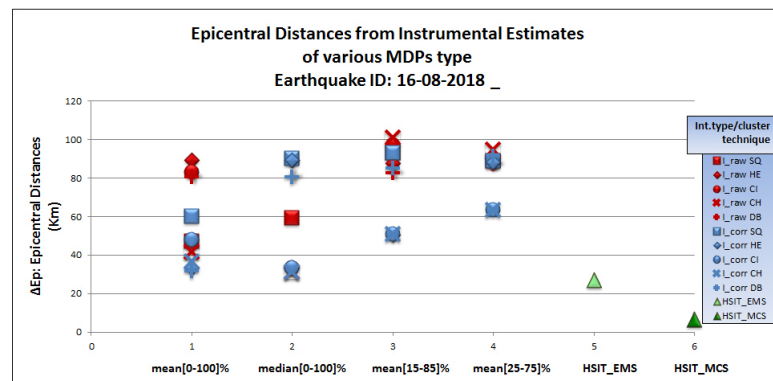


Figura 4.19: Terremoto 16/08/2018. Risultati per le localizzazioni epicentrali ottenuti con il “metodo 0” di Boxer. Per descrizione legenda e simbologia vedere inizio sezione “Magnitudo” 4.3.1.

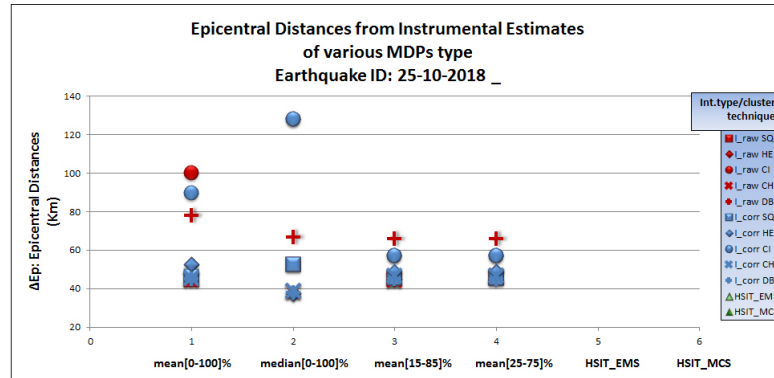


Figura 4.20: Terremoto 25/10/2018. Risultati per le localizzazioni epicentrali ottenuti con il “metodo 0” di Boxer. Per descrizione legenda e simbologia vedere inizio sezione “Magnitudo” 4.3.1.

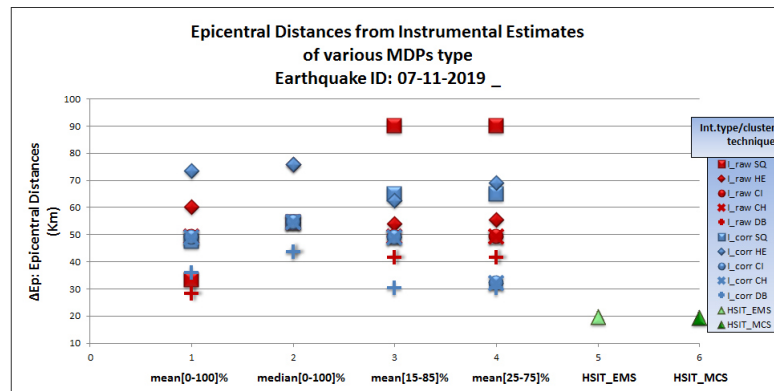


Figura 4.21: Terremoto 07/11/2019. Risultati per le localizzazioni epicentrali ottenuti con il “metodo 0” di Boxer. Per descrizione legenda e simbologia vedere inizio sezione “Magnitudo” 4.3.1.

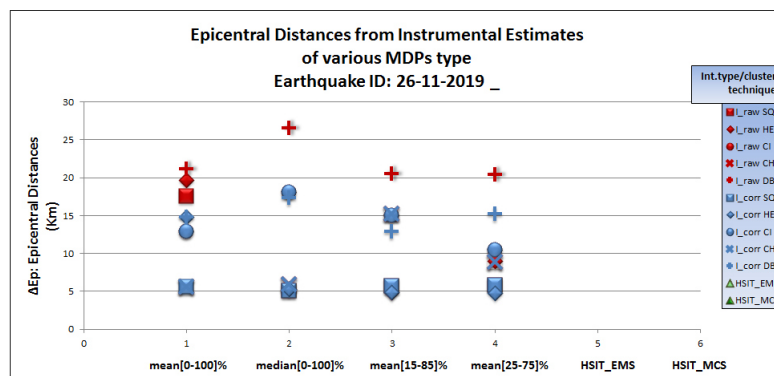


Figura 4.22: Terremoto 26/11/2019. Risultati per le localizzazioni epicentrali ottenuti con il “metodo 0” di Boxer. Per descrizione legenda e simbologia vedere inizio sezione “Magnitudo” 4.3.1.

Soffermandosi sul range di variabilità delle distanze epicentrali ΔEp tra stima macrosismica e stima strumentale, e considerando ancora le stime da dati EMSC

di tutti e 9 i terremoti complessivamente, emerge che questo risulta molto ampio ($[5, 100]Km$); inoltre non emergono regole chiare per l'identificazione di alcune tra le tecniche di clustering, tra i metodi statistici o tra le tipologie di intensità che siano migliori o peggiori delle altre le cui stime relative si concentrino in un sottoinsieme di valori di tale range.

È tuttavia possibile individuare per alcuni degli eventi analizzati un range ridotto $\Delta R \sim 20Km$ entro cui ricade un gruppo più cospicuo di stime variamente ottenute: per l'evento 18/01/2017 C (Figura 4.17) il range ridotto interessa principalmente le stime effettuate con le 2 tecniche di griglia e con la tecnica dei cerchi (SQ, HE, CI), per questo e per gli altri eventi della sequenza sismica del 18/01/2017 (Figure 4.15, 4.16, 4.18) la maggior parte delle stime si concentra ad una distanza epicentrale ΔEp di $[20, 40]Km$; per l'evento 25/10/2018 (Figura 4.20) il range ridotto è circa compreso tra $[40, 60]Km$, con ottima stabilità e precisione delle stime effettuate con i 3 metodi di media e, separatamente, con le stime effettuate con il metodo di mediana; per l'evento 07/11/2019 (Figura 4.21) si osserva ΔR circa compreso tra $[50, 70]Km$, in cui si ha per le tecniche di quadri (SQ), cerchi (CI) e mista (CH) buona stabilità e ottima precisione per i due metodi di media e mediana semplici; per l'evento 26/11/2019 (Figura 4.22) tutte le stime sono comprese nel range di distanza epicentrale $\Delta Ep \subseteq [0, 20]Km$, le tecniche di griglia (SQ e HE) producono le stime più stabili e precise ($\Delta E \sim 5Km$), mentre le stime effettuate con la tecnica dei cerchi (CI) e la tecnica DBSCAN (DB) si stabilizzano intorno a $\Delta E = 15Km$.

Le distanze epicentrali ottenute da dati HSIT sono generalmente incluse nel range di variabilità delle rispettive stime da dati EMSC con l'eccezione di 3 terremoti per i quali ottengono risultati migliori (Figure 4.14, 4.19, 4.21) e 2 terremoti per i quali ottengono risultati peggiori (Figure 4.20, 4.22). Anch'esse subiscono fluttuazioni, tuttavia più contenute rispetto alle stime da dati EMSC: per uno stesso evento il range di variabilità non supera i $20Km$. Inoltre la distanza dall'epicentro strumentale è generalmente inferiore ai $40Km$.

In Tabella 4.5 sono riportate le incertezze sulle stime delle coordinate epicentrali calcolate con Boxer con le formule relative alle *incertezze formali* (par. 1.4.4); sono espresse in Km . Sia quelle sulle latitudini che quelle sulle longitudini sono generalmente comprese in $[10, 30]Km$, ma per l'evento del 26/11/2019 si riducono sistematicamente a qualche Km .

Nelle Figure 4.23-4.31 sono mostrate per ogni terremoto le mappe con le migliori stime di localizzazione epicentrale ottenute con Boxer relativamente a ciascuna delle tecniche di clustering utilizzate per la costruzione delle distribuzioni di MDPs. Sono mostrate le relative barre d'errore ove disponibili.

Tabella 4.5:
Risultati per le localizzazioni epicentrali ottenuti con il “metodo 0” di Boxer applicato agli MDPs costruiti con le diverse tecniche di clustering e metodi statistici. Per dettagli simbologia e contenuto si fa riferimento al testo di inizio sezione “Magnitudo”.

		Values (°±Km)								Epicentral Distances (Km)			
		Mean		Median		Mean [15-85]%		Mean [25-75]%		Mean	Median	Mean [15-85]%	Mean [25-75]%
24/08/2016	#IDPs= 1082	Lat	Lon	Lat	Lon	Lat	Lon	Lat	Lon	ΔE	ΔE	ΔE	ΔE
Lot_Inst=42.698	Lon_Inst=13.234												
M_Inst=6.0	#MDPs	Lat	Lon	Lat	Lon	Lat	Lon	Lat	Lon	ΔE	ΔE	ΔE	ΔE
SQ	49	43.0969±12	13.3859±16	43.1014±15	13.5840±6	43.0571±20	13.4625±23	43.1361±20	13.4124±24	46.049	53.149	44.058	50.835
HE	47	43.0142±15	13.5419±20	42.9236±16	13.5884±30	43.0142±12	13.3869±20	42.9089±12	13.4084±22	43.198	38.276	37.303	27.430
CI	30	43.0058±19	13.5582±17	42.9236±31	13.5172±27	42.8577±11	13.3066±31	42.8578±11	13.3068±31	43.242	34.102	18.720	18.736
CH	40	43.0206±28	13.4256±17	42.7802±25	13.5412±20	42.9377±19	13.0927±24	42.9377±19	13.0928±24	39.124	26.701	29.038	29.035
DB	38	43.0032±21	13.5456±12	42.7055±24	13.5123±7	42.9474±16	13.3846±17	42.8961±16	13.3351±19	42.390	22.757	30.330	23.522
HSIT ESTIMATES													
		Lat				Lon				ΔE			
HSIT_EMS	339	42.9272±8				13.2798±7				25.758			
HSIT_MCS	336	42.811±16				13.344±5				15.445			
18/01/2017 A	#IDPs= 916												
Lot_Inst=42.545	Lon_Inst=13.277												
M_Inst=5.1	#MDPs	Lat	Lon	Lat	Lon	Lat	Lon	Lat	Lon	ΔE	ΔE	ΔE	ΔE
SQ	46	42.5594±14	13.5072±7	42.8568	13.5749	42.8993±10	13.4896±15	42.5597±14	13.5097±7	18.924	42.363	43.054	19.131
HE	44	42.7095±14	13.4490±19	42.6023±28	13.4872±7	42.5568±22	13.1552±19	42.4583±11	13.4892±9	23.078	18.354	10.063	19.889
CI	25	42.7057±15	13.3456±18	42.8548	13.5750	42.7906±25	12.8838±21	42.6928±23	13.2180±28	18.730	42.187	42.182	17.129
CH	40	42.7092±18	13.4476±18	42.8548	13.5750	42.5575±31	13.1532±27	42.4597±28	13.4026±28	22.982	42.187	10.236	13.999
DB	34	43.0276±9	13.3689±12	42.9625±11	13.3349±13	43.050±6	12.7779±12	42.9325±14	13.3089±11	54.184	46.664	69.814	43.167
HSIT ESTIMATES													
		Lat				Lon				ΔE			
HSIT_EMS	97	42.9237±14				13.3206±7				42.260			
HSIT_MCS	101	42.9374±12				13.264±8				43.646			
18/01/2017 B	#IDPs= 621												
Lot_Inst=42.531	Lon_Inst=13.284												
M_Inst=5.5	#MDPs	Lat	Lon	Lat	Lon	Lat	Lon	Lat	Lon	ΔE	ΔE	ΔE	ΔE
SQ	22	42.7638±22	13.1956±21	42.9491±14	13.3389±17	43.1717±30	13.5640±6	42.8076±32	13.1905±29	26.877	46.706	74.810	32.636
HE	22	43.0898±28	13.2560±29	43.5569	13.5147	43.1208±28	13.6525±6	43.1208±28	13.6525±6	62.178	115.605	72.140	72.140
CI	9	42.8406±30	13.5702±38	42.8412±28	13.5708±22	42.8406±30	13.5702±38	42.7874±25	13.5276±27	41.622	41.705	41.622	34.780
CH	15	43.0898±28	13.2560±29	43.5569	13.5147	43.1208±28	13.6525±6	43.1208±28	13.6525±6	62.178	115.605	72.140	72.140
DB	14	42.8168±18	13.6162±21	42.6003±28	13.5069±6	42.6936±31	13.4922±8	42.6937±31	13.4922±8	41.803	19.814	24.843	24.851
HSIT ESTIMATES													
		Lat				Lon				ΔE			
HSIT_EMS	69	42.8269±16				13.2811±10				32.903			
HSIT_MCS	72	42.8869±15				13.3139±10				39.650			
18/01/2017 C	#IDPs= 339												
Lot_Inst=42.503	Lon_Inst=13.277												
M_Inst=5.4	#MDPs	Lat	Lon	Lat	Lon	Lat	Lon	Lat	Lon	ΔE	ΔE	ΔE	ΔE
SQ	11	42.8526±17	13.7996±7	42.6590	13.6995	42.2442±32	13.5465±51	42.2387±23	13.5452±53	57.760	38.698	36.308	36.731
HE	11	42.3791±19	13.4720±33	42.5150±13	13.5611±33	42.2350±32	13.5388±53	42.2350±32	13.5388±53	21.115	23.326	36.751	36.751
CI	4	42.5768±40	13.5937±33	42.5811±40	13.5974±33	42.5761±40	13.5927±33	42.5761±40	13.5927±33	27.214	27.648	27.112	27.112
CH	7	42.5449±11	14.1022±5	42.6535±18	13.7248±32	42.5148±13	14.1115±6	42.5148±13	14.1115±6	67.785	40.304	68.416	68.416
DB	11	42.7569±22	13.6418±10	42.6590	13.6995	42.8696±24	13.8313±12	42.8696±24	13.8313±12	41.082	38.698	60.946	60.946
HSIT ESTIMATES													
		Lat				Lon				ΔE			
HSIT_EMS	73	42.8455±15				13.1919±10				38.714			
HSIT_MCS	81	42.7568±14				13.1169±10				31.113			
18/01/2017 D	#IDPs= 772												
Lot_Inst=42.473	Lon_Inst=13.275												
M_Inst=5.0	#MDPs	Lat	Lon	Lat	Lon	Lat	Lon	Lat	Lon	ΔE	ΔE	ΔE	ΔE
SQ	37	42.8058±9	13.5329±12	43.2808	13.6481	42.8174±10	13.4159±13	42.8648±14	13.4454±14	42.596	94.829	39.992	45.740
HE	43	42.6674±8	13.3454±19	42.6447±8	13.4717±23	42.5430±8	13.2777±23	42.5436±8	13.2774±22	22.372	24.981	7.787	7.853
CI	23	42.7031±18	13.4444±12	42.6275±17	13.2909±16	42.6022±21	13.3677±15	42.6118±27	13.4955±18	29.103	17.229	16.250	23.760
CH	36	42.7405±16	13.4998±8	42.6283±19	13.4159±11	42.6092±20	13.4267±11	42.6167±23	13.5217±12	34.975	20.771	19.592	25.764
DB	28	42.7352±21	13.3385±11	42.6436±20	13.3043±14	42.7357±20	13.2479±13	42.7357±20	13.2486±13	29.615	19.121	29.295	29.291
HSIT ESTIMATES													
		Lat				Lon				ΔE			
HSIT_EMS	37	42.8871±17				13.4175±11				47.496			
HSIT_MCS	38	42.9118±17				13.1724±12				49.508			
16/08/2018	#IDPs= 541												
Lot_Inst=41.874	Lon_Inst=14.865												
M_Inst=5.1	#MDPs	Lat	Lon	Lat	Lon	Lat	Lon	Lat	Lon	ΔE	ΔE	ΔE	ΔE
SQ	22	42.1880±12	14.2723±19	42.3727±10	14.0061±11	42.480±10	14.0878±6	42.4410±9	14.1061±7	60.129	89.959	92.987	88.815
HE	17	42.0494±3	14.5427±26	42.4313±23	14.0870±11	42.0750±3	14.3201±32	42.4137±29	14.0842±13	33.024	89.181	50.285	88.008
CI	13	42.1679±13	14.4431±20	42.0519±3	14.5413±26	42.0807±3	14.3151±32	42.2245±16	14.2611±19	47.777	33.281	50.937	63.288
CH	19	42.0837±20	14.5226±28	42.0623±3	14.5866±19	42.0807±3	14.3151±32	42.2245±16	14.2611±19	36.671	31.115	50.937	63.288
DB	15	42.1041±7	14.6457±18	42.3315±29	14.1116±13	42.3767±16	14.0857±15	42.4103±20	14.0270±19	31.355	80.319	85.176	91.268
HSIT ESTIMATES													
		Lat				Lon				ΔE			
HSIT_EMS	139	41.9587±8				14.5609±8				26.867			
HSIT_MCS	201	41.888±5				14.79±6				6.401			
25/10/2018	#IDPs= 1883												
Lot_Inst=37.492	Lon_Inst=20.595												
M_Inst=6.8	#MDPs	Lat	Lon	Lat	Lon	Lat	Lon	Lat	Lon	ΔE	ΔE	ΔE	ΔE
SQ	70	37.7555±4	21.0010±14	37.8649±22	20.9599±24	37.7576±4	20.9825±14	37.7582±4	20.9823±14	46.228	52.446	45.132	45.162
HE	74	37.7006±4	21.1272±16	37.7696±2	20.8332±3	37.7351±6	21.0541±26	37.7358±4	21.0543±19	52.312	37.321	48.641	4.000
CI	63	38.2141±74	21.0462±47	38.6416±102	20.6096±50	37.7284±6	21.1687±24	37.7284±6	21.1686±24	89.534	127.836	56.963	56.955
CH	89	37.7225±22	21.0163±30	37.7749±2	20.8498±3	37.7491±5	21.0140±21	37.7491±5	21.0139±21	45.103	38.639	46.681	46.674
DB	52	37.7512±4	20.9840±13	37.7753±122	20.8340±53	37.7515±4	20.9831±14	37.7522±3	20.9828±14	44.771	37.885	44.732	44.762
HSIT ESTIMATES													
		Lat				Lon				ΔE			
HSIT_EMS	124	39.6315±34				16.4956±22				428.459			
HSIT_MCS	125	39.719±32				16.4896±22				435.629			
07/11/2019	#IDPs= 343												
Lot_Inst=41.776	Lon_Inst=13.604												
M_Inst=4.4	#MDPs	Lat	Lon	Lat	Lon	Lat	Lon	Lat	Lon	ΔE	ΔE	ΔE	ΔE
SQ	14	41.9763±13	13.0984±25	41.9636±10	12.9968±20	41.9303±6	12.8483±24	41.9296±6	12.8478±24	47.417	54.433	64.899	64.919
HE	11	41.9694±4	12.7549±12	41.9480±3	12.7163±8	42.0645±11	12.9574±22	41.9675±4	12.8128±26	73.520	75.960	62.379	68.885
CI	4	42.1934±18	13.4141±1	42.0977±15	13.1071±25	42.1936±18	13.4154±1	42.0347	13.4315	48.995	54.488	48.982	32.114
CH	4	42.1934±18	13.4141±1	42.0977±15	13.1071±25	42.1936±18	13.4154±1	42.0347	13.4315	48.995	54.488	48.982	32.114
DB	6	42.0400±15	13.3530±21	41.9982±17	13.1712±34	42.0447±2	13.6643±19	42.0447±2	13.6643±19	35.961	43.521	30.292	30.292
HSIT ESTIMATES													
		Lat				Lon				ΔE			
HSIT_EMS	104	41.7111±4				13.3855±5				19.512			
HSIT_MCS	105	41.7365±4				13.3759±4				19.424			
26/11/2019	#IDPs= 4036												
Lot_Inst=41.371	Lon_Inst=19.467												
M_Inst=6.2	#MDPs	Lat	Lon	Lat	Lon	Lat	Lon	Lat	Lon	ΔE	ΔE	ΔE	ΔE
SQ	87	41.3249±1	19.4923±3	41.3283±1	19.4879±3	41.3235±0.4	19.4930±3	41.3232±0.3	19.4944±3	5.544	5.058	5.710	5.786
HE	92	41.3315±3	19.6364±7	41.3304±1	19.4983±3	41.3341±1	19.4967±3	41.3348±1	19.4963±3	14.806	5.216	4.794	4.751
CI	54	41.3319±1	19.6119±9	41.3513±2	19.6816±10	41.3234±0.3	19.6354±15	41.4001±9					

Dalla Tabella 4.5 in cui si riportano i valori delle stime delle coordinate (in gradi decimali) e delle distanze epicentrali (in *Km*) rispetto all'epicentro di riferimento strumentale emerge la disomogeneità dei risultati già evidente all'osservazione dei plot (Figure 4.14-4.22). Considerando le stime in relazione ai metodi statistici (media, mediana ecc) con cui sono costruite le relative distribuzioni di MDPs e conteggiando per ciascuno di questi metodi i migliori e i peggiori risultati ottenuti per ogni tecnica per ogni terremoto (valori in *corsivo* e in sottolineato di Tabella 4.5; conteggi in Tabella 4.6b) emerge che: la *media 15/85* totalizza il numero maggiore di migliori stime (13) che è di sole 4 unità più alto del numero minore di migliori stime, 9 totalizzato dalle *medie*; la *mediana* totalizza il numero più elevato di peggiori stime (17), mentre per le 2 *medie decimate* si conteggia, separatamente, il minor numero (9); i rapporti “migliore”/“peggiore” sono inferiori all'unità e molto vicini per *media* e *mediana semplici*, maggiori di 1 e molto vicini per le 2 *medie decimate*.

Score stime epicentrali Boxer: MTO		
Tecnica	Migliore	Peggior
QUADRATI	0	3
ESAGONI	5	2
CERCHI	1	2
MISTA	1	2
DBSCAN	2	1
Tecnica	Migliore	Peggior
HSIT_EMS	2	2
HSIT_MCS	3	2

Score stime epicentrali Boxer: MTO		
Metodo statistico	Migliore	Peggior
Media	9	13
Mediana	11	17
Media [15-85]%	13	9
Media [25-75]%	11	9

Tabelle 4.6a, b: Score delle stime di localizzazione epicentrale migliori e peggiori in relazione alle tecniche di clustering e ai metodi statistici utilizzati per la rappresentazione dei relativi MDPs. Per dettagli sulle modalità di conteggio vedere testo di inizio sezione “Magnitudo”.

La Tabella 4.6a riporta i conteggi per ciascuna delle tecniche di clustering, utilizzate per la costruzione delle distribuzioni di MDPs dai dati IDPs, in relazione alle stime migliori e peggiori ottenute. Per la sezione superiore della tabella si assegnano i punteggi considerando solamente le stime relative alle 5 tecniche applicate agli IDPs EMSC, per ogni terremoto (valori verde chiaro e rosso chiaro di Tabella 4.5). Risulta una prevalenza di stime migliori (5 su 9) per la tecnica di griglia di esagoni; la tecnica DBSCAN totalizza 2 stime migliori; 1 stima migliore si ottiene con la tecnica dei cerchi e 1 con la tecnica mista; con la tecnica dei quadrati non si ottiene alcuna migliore stima. Il maggior numero di stime peggiori si ottiene con la tecnica dei quadrati (3 su 9), la DBSCAN ne totalizza 2, le altre 3 tecniche ne totalizzano 2 ciascuna (con una coincidenza tra stime effettuate con la griglia di esagoni (HE) e con la tecnica mista (CH) per il terremoto 18/01/2017 B).

La sezione inferiore della Tabella 4.6a assegna un'unità di punteggio alla tecnica HSIT, separatamente per i due tipi di intensità utilizzati (EMS e MCS), solamente

per quei terremoti in cui la stima con questa tecnica risulta migliore/peggiore di tutte le stime effettuate con le altre 5 tecniche (applicate ai dati EMSC) considerate complessivamente. In 3 terremoti (24/08/2016, 16/08/2018 e 07/11/2019, Tabella 4.5, Figure 4.14, 4.19, 4.21) la stima $HSIT_{MCS}$ è migliore di tutte quelle effettuate da dati EMSC e in 2 di questi (16/08/2018 e 07/11/2019) anche $HSIT_{EMS}$ stima meglio delle 5 tecniche. Per 2 terremoti la stima relativa ad HSIT, sia EMS sia MCS, risulta peggiore che le altre 5 tecniche (25/10/2018, 26/11/2019).

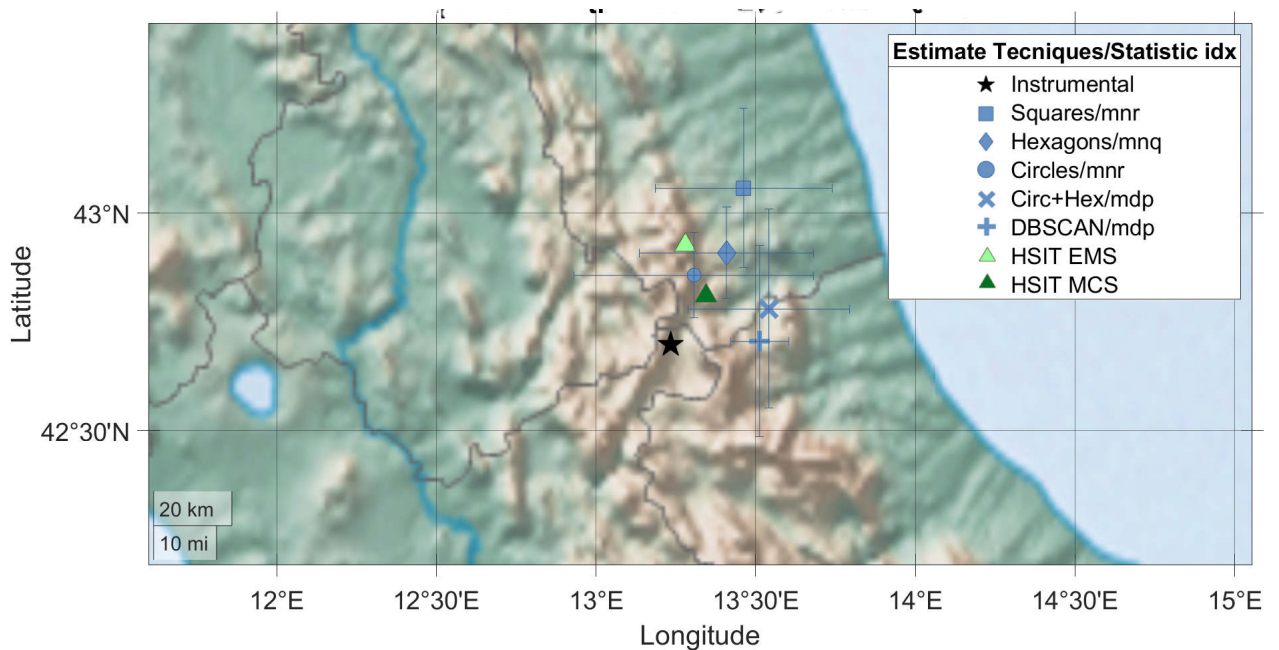


Figura 4.23: Terremoto 24/08/2016. Risultati per le migliori localizzazioni epicentrali ottenute con il “metodo 0” di Boxer per ciascuna delle tecniche di clustering utilizzate per la costruzione delle distribuzioni di MDPs. Le stime sono ottenute da IDPs EMSC di intensità corretta I_{corr} (simboli blu) (par. 4.3.1; Bossu et al., 2017) e da dati HSIT espressi sia in intensità EMS (verde chiaro) sia in intensità MSC (verde scuro). Per ciascun risultato da dati EMSC è riportato anche il metodo statistico (“Statistic idx”) utilizzato per il calcolo dell’intensità delle MDPs: mnp=media, mdp=mediana, mnq=media[15,85]%, mnr=media[25/75]%. Sono mostrate le barre d’errore corrispondenti alle incertezze formali o le ellissi della matrice di varianza/covarianza degli errori ove calcolate (par. 1.4.4).

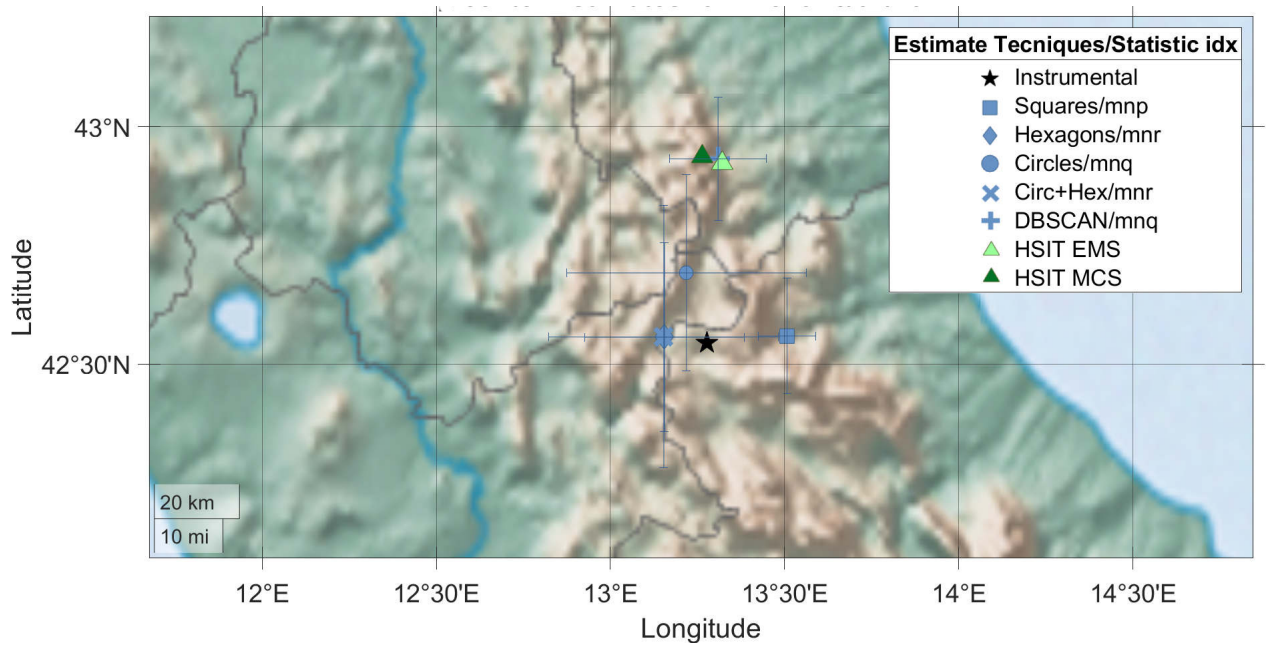


Figura 4.24: Terremoto 18/01/2017 A. Risultati per le migliori localizzazioni epicentrali ottenute con il “metodo 0” di Boxer. Per simbologia legenda fare riferimento a 4.23.

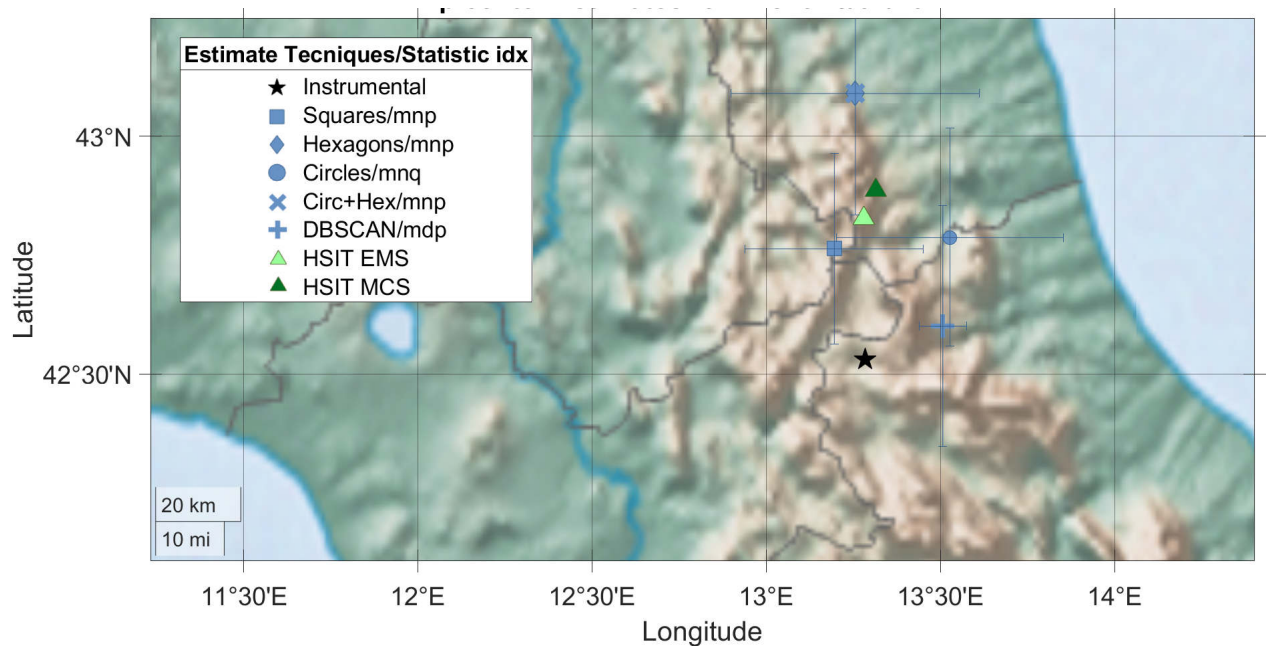


Figura 4.25: Terremoto 18/01/2017 B. Risultati per le migliori localizzazioni epicentrali ottenute con il “metodo 0” di Boxer. Per simbologia legenda fare riferimento a 4.23.

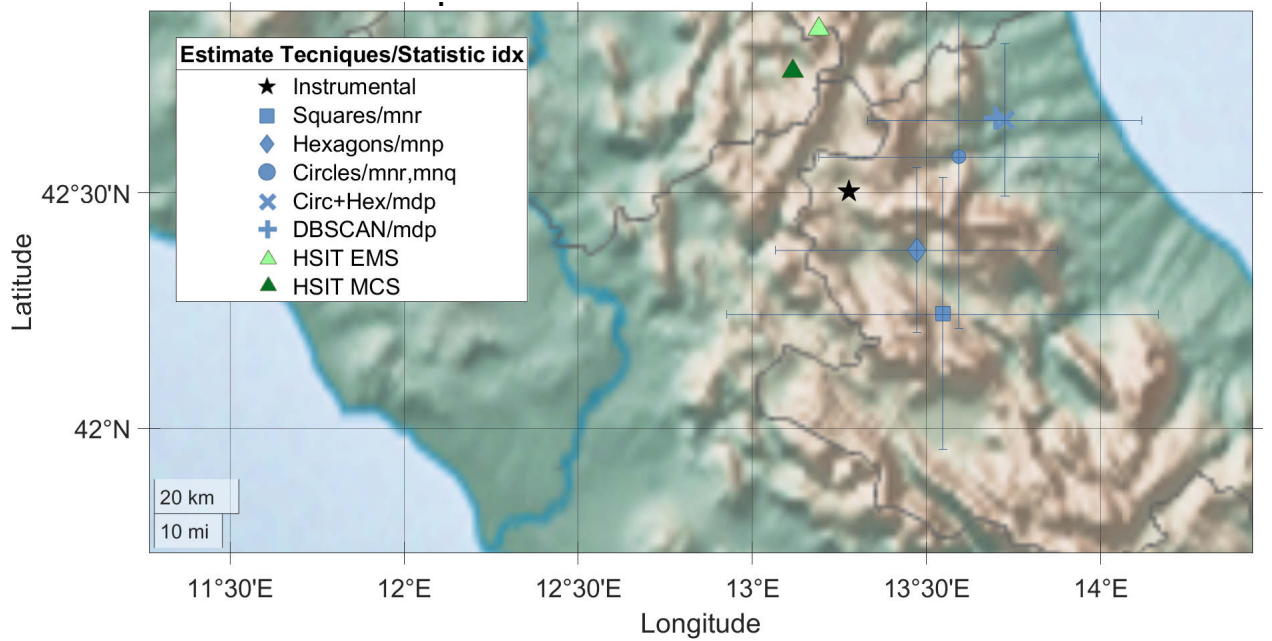


Figura 4.26: Terremoto 18/01/2017 C. Risultati per le migliori localizzazioni epicentrali ottenute con il “metodo 0” di Boxer. Per simbologia legenda fare riferimento a 4.23.

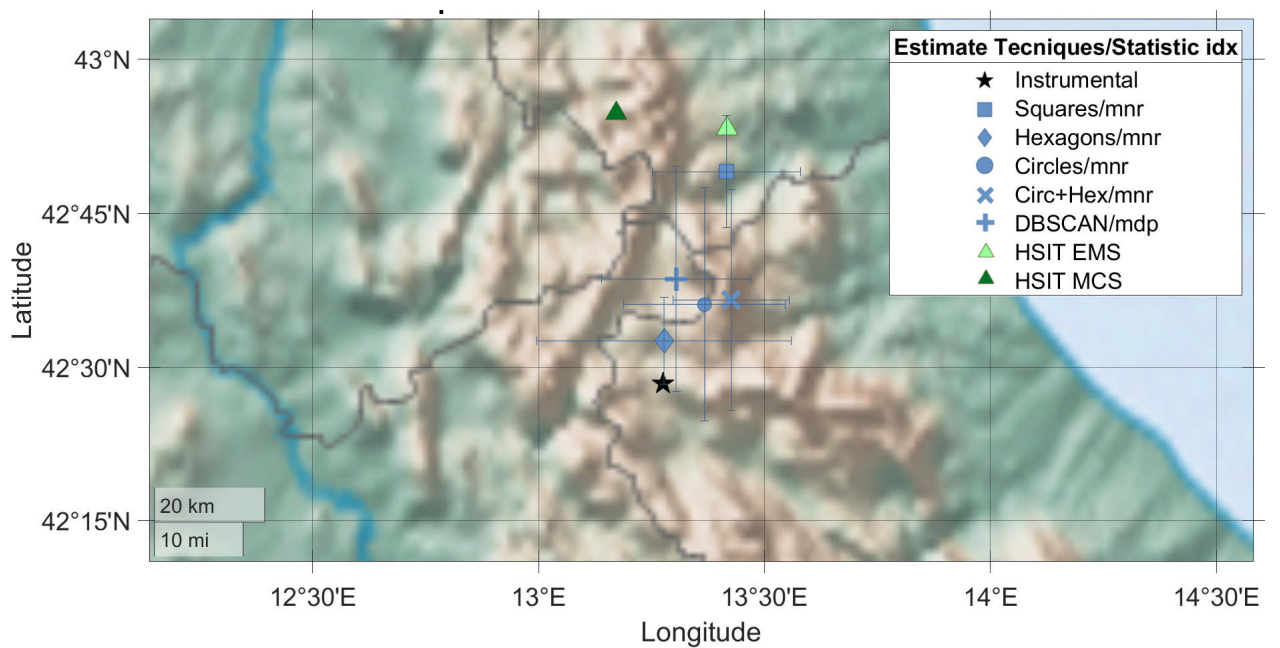


Figura 4.27: Terremoto 18/01/2017 D. Risultati per le migliori localizzazioni epicentrali ottenute con il “metodo 0” di Boxer. Per simbologia legenda fare riferimento a 4.23.

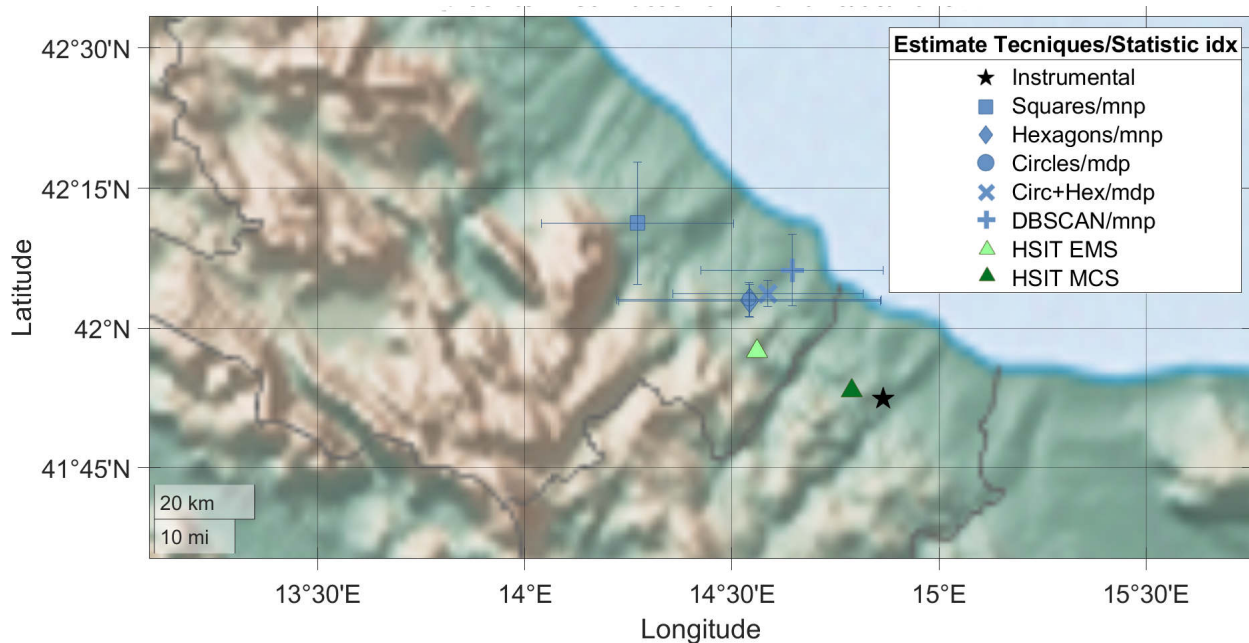


Figura 4.28: Terremoto 16/08/2018. Risultati per le migliori localizzazioni epicentrali ottenute con il “metodo 0” di Boxer. Per simbologia legenda fare riferimento a 4.23.

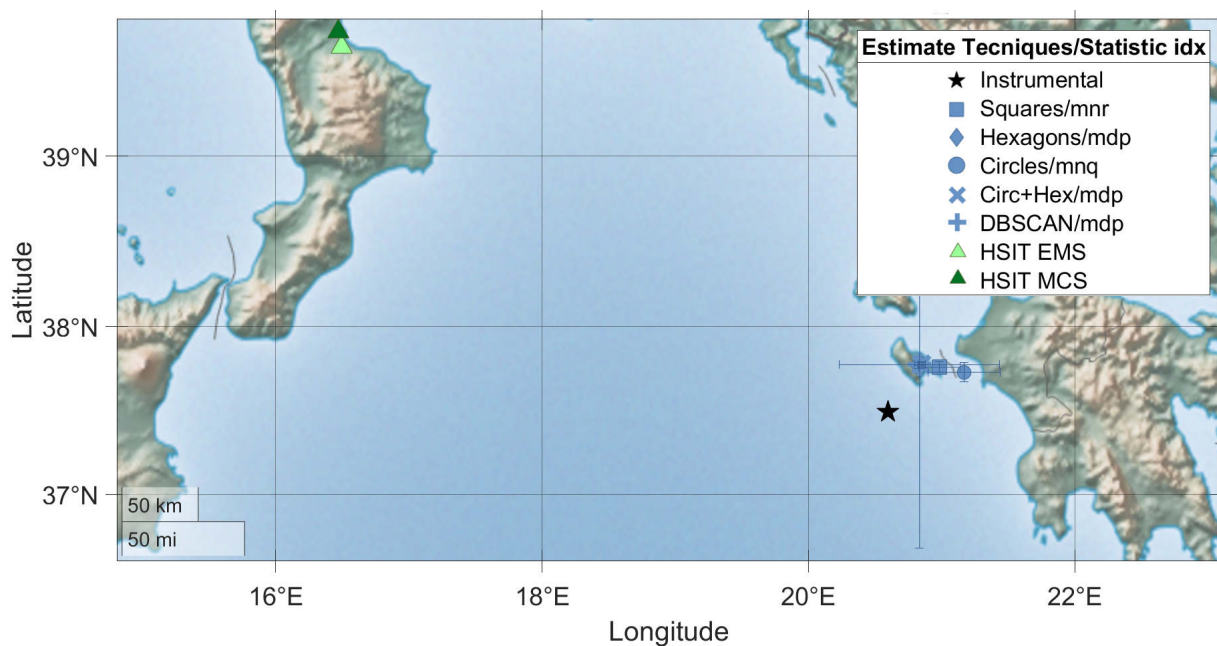


Figura 4.29: Terremoto 25/10/2018. Risultati per le migliori localizzazioni epicentrali ottenute con il “metodo 0” di Boxer. Per simbologia legenda fare riferimento a 4.23.

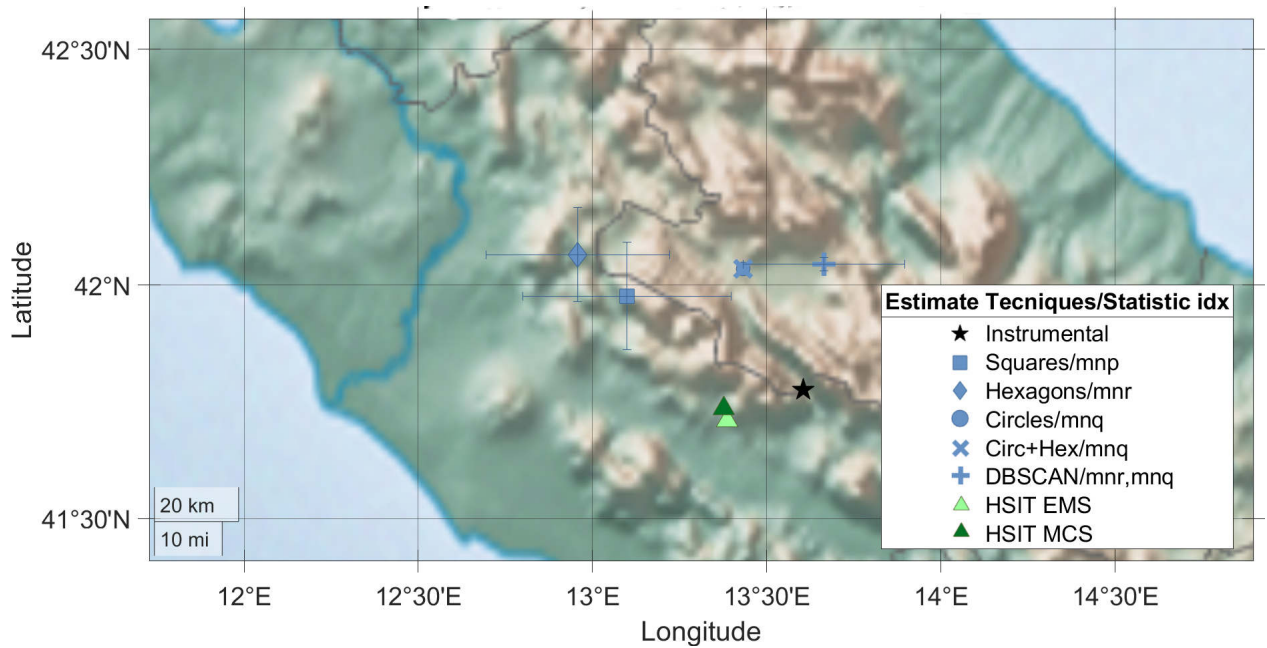


Figura 4.30: Terremoto 07/11/2019. Risultati per le migliori localizzazioni epicentrali ottenute con il “metodo 0” di Boxer. Per simbologia legenda fare riferimento a 4.23.

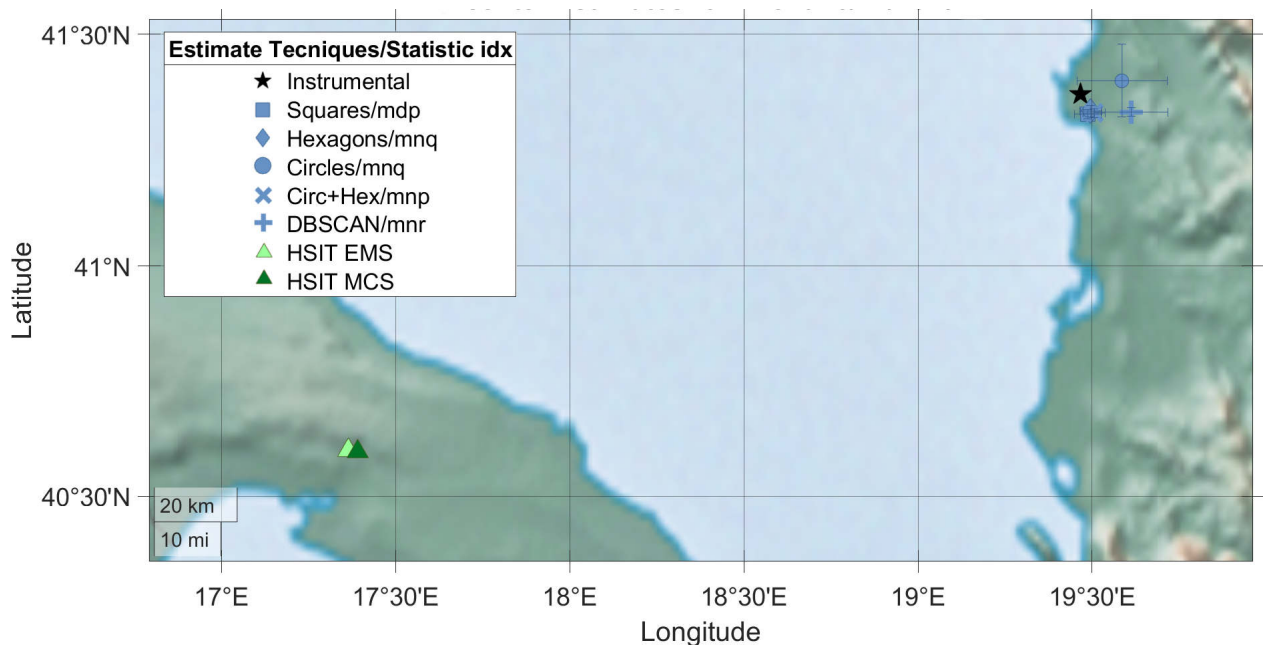


Figura 4.31: Terremoto 26/11/2019. Risultati per le migliori localizzazioni epicentrali ottenute con il “metodo 0” di Boxer. Per simbologia legenda fare riferimento a 4.23.

In ultima analisi per le stime epicentrali, come per le stime di magnitudo, effettuate a partire da dati macrosismici risulta determinante per migliorare l’accuratezza e la precisione avere un cospicuo numero di IDPs nella regione del *near-field macrosismico* da cui costruire con le tecniche di clustering un maggiore numero di MDPs per l’elaborazione con Boxer. Confermano tale necessità le stime per la localizzazione degli eventi 24/08/2016,

16/08/2018 e 07/11/2019 (Tabella 4.5, Figure 4.14, 4.19, 4.21). Per questi eventi le stime effettuate da dati HSIT risultano nettamente migliori di quelle effettuate da dati EMSC. In corrispondenza delle prime si osserva un numero di IDPs raccolti dalla relativa istituzione (HSIT, INGV) superiore di un ordine di grandezza al numero di IDPs raccolti da EMSC, rispettivamente $\#IDPs_{HSIT} = 12073, 5181, 3674$ (<https://www.hsit.it/>) e $\#IDPs_{EMSC} = 1082, 541, 343$ (Tabella 4.1); inoltre i numeri di MDPs relativi alle stime da dati EMSC risultano particolarmente contenuti, e sono compresi circa tra la 7-ma e la 26-esima parte del rispettivo $\#MDPs$ da dati HSIT. Per tutti gli altri terremoti invece il $\#MDPs_{EMSC}$ non risulta inferiore all'8-va parte del $\#MDPs_{HSIT}$ e le stime relative a queste 2 tipologie di dato (HSIT e EMSC) risultano più vicine.

Inoltre i 2 terremoti per cui le stime da dati HSIT risultano peggiori che le stime da dati EMSC (Tabella 4.5) corrispondono a quelli con epicentro strumentale localizzato al di fuori del territorio italiano, dove HSIT non raccoglie testimonianze (IDPs), con conseguente disuniformità intrinseca nella copertura del campo macrosismico (Figure 4.29, 4.31).

Infine è da considerare che la maggior parte dei terremoti analizzati in questa tesi (Tabella 4.1) hanno magnitudo strumentale in prossimità della soglia di affidabilità ($M > 5.5$) del “metodo del baricentro” di Boxer (relativo al “metodo 0”; par. 1.4.1) con cui sono state effettuate le stime epicentrali. Tuttavia l'elevata magnitudo degli eventi 2018/10/25 e 2016/08/24, rispettivamente 6.8 e 6.0, superiore alla soglia di accettabilità, non è sufficiente per ottenere una stima affidabile se gli IDPs di base rimangono poco numerosi (1883 per il primo dei due eventi in questione, 1082 per il secondo) e il *near-field* sottorappresentato (Figure 4.10a e 3.6 rispettivamente).

In generale gli IDPs dei dataset di EMSC utilizzati per questa tesi sono caratterizzati da bassa numerosità (Tabella 4.1) e sono geograficamente distribuiti con grande disuniformità azimutale indipendentemente dal valore di magnitudo strumentale. L'unica eccezione è il terremoto 26/11/2019 per il quale sono stati ottenuti ottimi risultati.

A fronte della fatto che la numerosità dei datasets raccolti è crescente nel tempo (Bossu et al., 2020) questi metodi di localizzazione, in associazione con le tecniche di clustering di IDPs, hanno buone prospettive per essere ulteriormente testati e consolidati.

Localizzazione epicentrale: Boxer MT1

In questa sezione vengono mostrati i risultati ottenuti con il “metodo 1” di Boxer (MT1, descritto in sintesi al par. 4.2, in dettaglio al par. 1.4) per gli epicentri macrosismici dei 2 terremoti che hanno epicentro strumentale localizzato in mare o prossimo alla costa (25/10/2018 Figura 4.29, 26/11/2019 Figura 4.31). Entrambi gli epicentri strumentali sono localizzati al di fuori del territorio italiano ed in particolare in territorio greco (25/10/2018) e in territorio albanese (26/11/2019).

L’epicentro localizzato in prossimità della linea di costa comporta un’intrinseca assenza di dati macrosismici per un ampio angolo attorno ad esso, sia nel near-field che nel far-field. Questo è riscontrabile nelle distribuzioni di IDPs EMSC e HSIT (Figure 4.10, 4.11). Inoltre la compilazione dei questionari (IDPs) HSIT è attiva solamente per il territorio italiano, perciò per questi si ottiene maggiore disuniformità nella copertura del campo macrosismico.

Per questa ragione per i 2 eventi in questione sono state avviate le stime dei *parametri sintetici* con il “metodo 1” di Boxer, che può essere applicato a terremoti con tali caratteristiche (Gasperini et al., 2010).

I risultati dell’applicazione del “metodo 1” sono mostrati in plot, tabelle e mappe con le medesime modalità e caratteristiche utilizzate per i risultati del “metodo 0” (per descrizione legende e simbologia vedere inizio sezione “Magnitudo” 4.3.1).

Confrontando per l’evento 25/10/2018 i risultati ottenuti con il “metodo 1” di Boxer (Figura 4.32, 4.34, Tabella 4.7) con i rispettivi risultati ottenuti con il “metodo 0” (Figure 4.20, 4.29 e Tabella 4.5) emerge che le distanze epicentrali ΔEp delle stime macrosismiche rispetto all’epicentro di riferimento strumentale rimangono comprese circa nello stesso range di valori $[40, 60] Km$. Per le stime effettuate con il “metodo 1” si osservano generali lievi riduzioni di ΔEp rispetto alle stime relative al “metodo 0”, ma un aumento della dispersione di queste; la stabilità e la precisione delle stime effettuate con il “metodo 0”, relative ai 3 metodi di media e, separatamente, al metodo di mediana non è confermata. Si osservano le fluttuazioni maggiori per il “metodo 1” in corrispondenza delle 2 tecniche di griglia (quadrati SQ, esagoni HE). Come per la magnitudo equivalente, anche per le coordinate epicentrali il “metodo 1” produce risultato nullo (“-”) sia in corrispondenza delle MDPs costruite a partire dai dati HSIT, sia per le MDPs costruite a partire da dati EMSC con la tecnica dei cerchi attorno alle località congiuntamente al metodo statistico media.

Considerando il terremoto 26/11/2019 i risultati ottenuti con il “metodo 1” (Figure 4.33 e 4.35, Tabella 4.7) sono nettamente migliori dei risultati ottenuti con il “metodo 0” (Figure 4.22, 4.31, Tabella 4.5). Osservando i plot si riscontra una sistematica riduzione della distanza epicentrale ΔEp delle stime macrosismiche rispetto al riferimento strumentale. Quasi tutte le stime presentano $\Delta E < 4 Km$ indipendentemente dalla tecnica di clustering o dal metodo statistico utilizzato per la costruzione delle relative distribuzioni di MDPs. Queste sono consistenti con la stima epicentrale di riferimento strumentale entro le incertezze calcolate da Boxer (Tabella 4.7, ellissi d’errore di Figura 4.31).

In particolare rispetto alle stime ottenute con il “metodo 0”, tra quelle ottenute con il “metodo 1” migliorano quelle relative alla tecnica DBSCAN (DB) e alla tecnica dei cerchi (CI). Tuttavia, come per l’evento 25/10/2018, anche per 26/11/2019 si riscontrano un’aumento della dispersione delle stime da “metodo 1” rispetto alle stime da

“metodo 0” e risultati nulli in corrispondenza dei dati HSIT.

Infine è possibile individuare nel plot di Figura 4.33 2 gruppi di stabilità delle stime differenziati in relazione alla tipologia d’intensità con cui vengono espressi gli IDPs di EMSC: intensità I_{raw} scelte dagli utenti secondo la scala EMS (simboli rossi) e intensità corrette I_{corr} come da Bossu et al., 2017 (simboli blu). Al gruppo delle I_{corr} corrispondono stime sistematicamente migliori delle stime del gruppo delle I_{raw} . Questi 2 gruppi di stime sono già stati individuati nei plot d’analisi per le magnitudo (par. 4.3.1); pertanto il riscontro in quest’ulteriore plot consolida la scelta operata, ed esposta ad inizio capitolo, di considerare le sole intensità corrette I_{corr} per la conduzione della presente analisi.

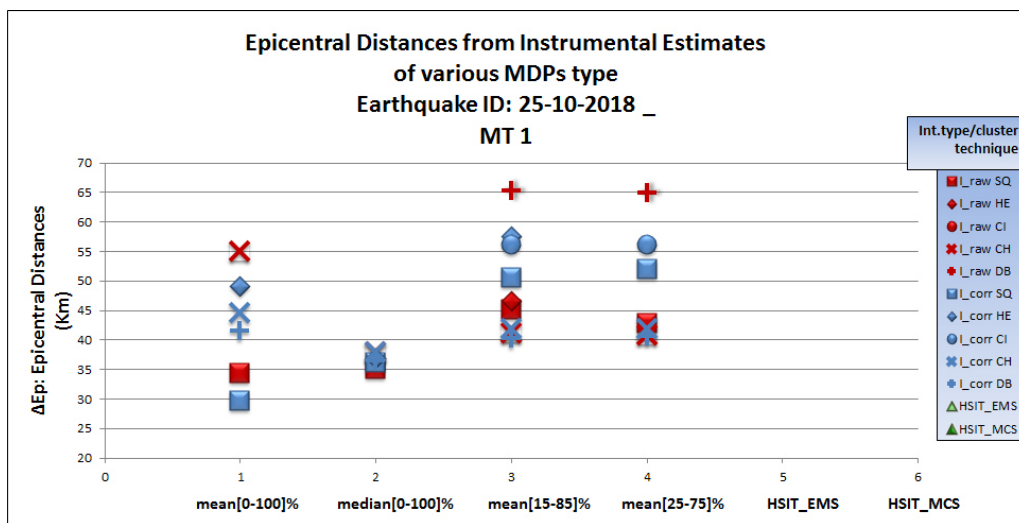


Figura 4.32: Risultati per le localizzazioni epicentri dell’evento 25/10/2018 ottenuti con il “metodo 1” di Boxer. Per simbologia legenda fare riferimento al testo di inizio sezione “Magnitudo”.

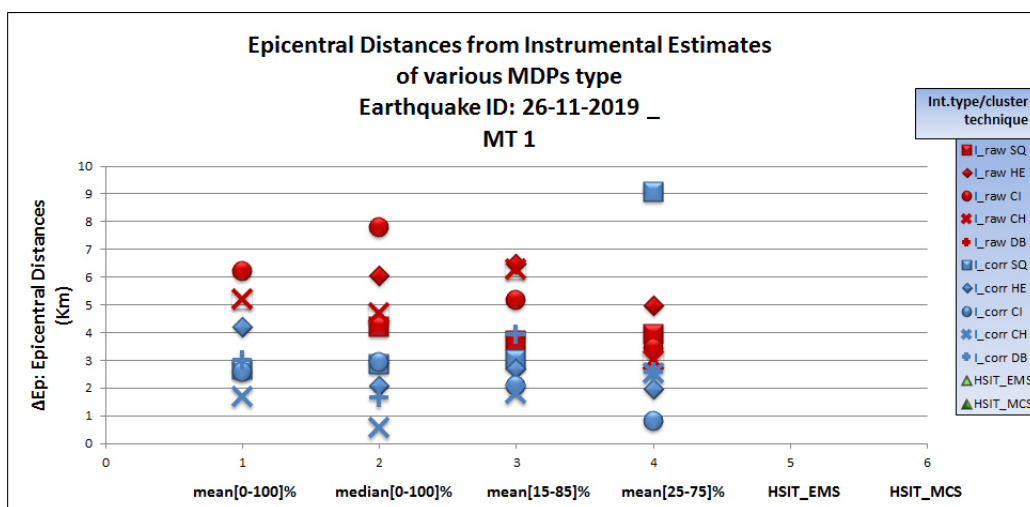


Figura 4.33: Risultati per le localizzazioni epicentri dell’evento 26/11/2019 ottenuti con il “metodo 1” di Boxer. Per simbologia legenda fare riferimento al testo di inizio sezione “Magnitudo”.

EPICENTRAL LOCATION													
Boxer: MT1													
Values (\pm Km)										Epicentral Distances (Km)			
25/10/2018		#IDPs= 1883											
Lat_Inst= 37.492		Lon_Inst= 20.595											
M_Inst=6.8		#MDPs											
		Mean		Median		Mean [15-85]%		Mean [25-75]%		Mean	Median	Mean [15-85]%	Mean [25-75]%
		Lat	Lon	Lat	Lon	Lat	Lon	Lat	Lon	ΔE	ΔE	ΔE	ΔE
SQ	70	37.6623±4	20.8523±16	37.8038±5	20.7124±2	37.8382±9	20.9674±6	37.8884±7	20.9058±10	29.542	36.178	50.561	51.872
HE	74	37.6697±19	21.1053±8	37.7645±2	20.8309±3	37.8900±17	21.0108±14	37.6328±24	20.7009±33	49.118	36.738	57.421	18.228
CI	63	-	-	38.6395	20.6166	37.8170±4	21.0820±1	37.8170±4	21.0819±1	-	127.6102	56.072	56.065
CH	89	37.7165±10	21.0141±6	37.7659±2	20.8503±2	37.7036±12	20.9875±10	37.7037±12	20.9874±10	44.567	37.856	41.825	41.824
DB	52	37.7153±14	20.9714±9	37.7837±3	20.7839±2	37.7072±13	20.9629±10	37.7097±13	20.9644±9	41.425	36.452	40.288	40.559
HSIT ESTIMATES													
	#MDPs	Lat		Lon						ΔE			
HSIT_EMS	124	-		-		-		-		-			
HSIT_MCS	125	-		-		-		-		-			
26/11/2019		#IDPs= 4036											
Lat_Inst= 41.371		Lon_Inst= 19.467											
M_Inst=6.2		#MDPs											
		Mean		Median		Mean [15-85]%		Mean [25-75]%		Mean	Median	Mean [15-85]%	Mean [25-75]%
		Lat	Lon	Lat	Lon	Lat	Lon	Lat	Lon	ΔE	ΔE	ΔE	ΔE
SQ	87	41.3878±4	19.4444±5	41.3798±4	19.4351±4	41.3484±4	19.4474±5	41.2949±5	19.4290±5	2.654	2.836	2.999	9.037
HE	92	41.3789±4	19.5159±4	41.3888±4	19.4749±4	41.3906±4	19.4856±4	41.3839±3	19.4832±3	4.174	2.086	2.675	1.971
CI	54	41.3772±5	19.4968±5	41.3610±5	19.4993±5	41.3700±6	19.4917±7	41.3710±4	19.4763±6	2.580	2.916	2.064	0.776
CH	84	41.3584±4	19.4557±4	41.3742±4	19.4617±4	41.3636±4	19.4862±4	41.3516±5	19.4508±5	1.689	0.568	1.801	2.546
DB	58	41.3909±4	19.4910±5	41.3780±4	19.4845±5	41.4029±6	19.4870±6	41.3780±6	19.4963±8	2.984	1.655	3.920	4.000
HSIT ESTIMATES													
	#MDPs	Lat		Lon						ΔE			
HSIT_EMS	207	-		-		-		-		-			
HSIT_MCS	202	-		-		-		-		-			

Tabella 4.7: Risultati per le localizzazioni epicentrali ottenuti con il “metodo 1” di Boxer applicato agli MDPs costruiti con le diverse tecniche di clustering e metodi statistici. Per dettagli simbologia e contenuto si fa riferimento al testo di inizio sezione “Magnitudo”.

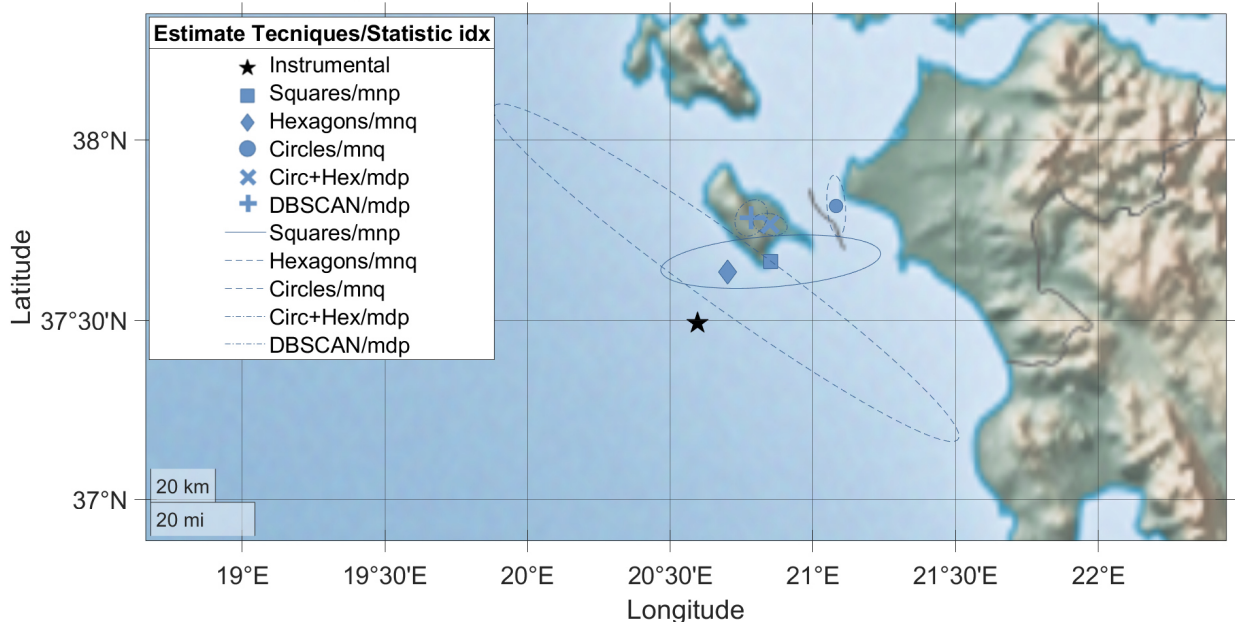


Figura 4.34: Terremoto 25/10/2018. Risultati per le migliori localizzazioni epicentrali ottenute con il “metodo 1” di Boxer per ciascuna delle tecniche di clustering utilizzate per la costruzione delle distribuzioni di MDPs. Per simbologia legenda fare riferimento a Figura 4.23.

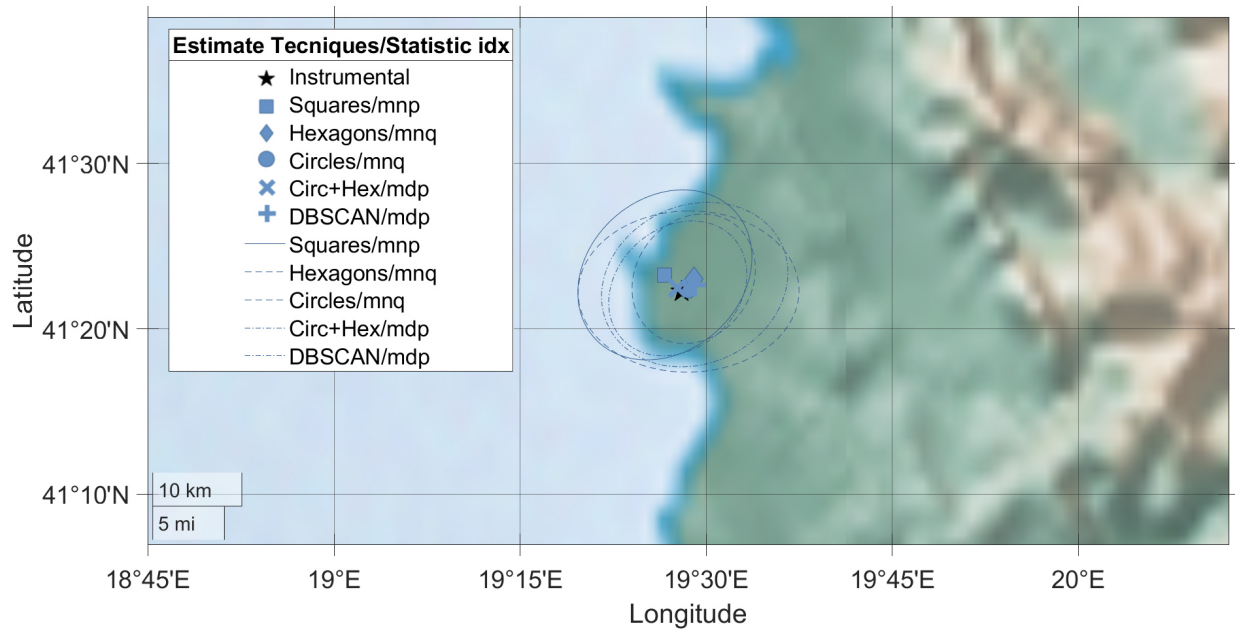


Figura 4.35: Terremoto 26/11/2019. Risultati per le migliori localizzazioni epicentrali ottenute con il “metodo 1” di Boxer per ciascuna delle tecniche di clustering utilizzate per la costruzione delle distribuzioni di MDPs. Per simbologia legenda fare riferimento a Figura 4.23.

Conclusioni

In questa tesi è stata trattata la trasformazione delle intensità macrosismiche individuali (IDPs) raccolte da testimonianze di singoli utenti attraverso piattaforme web in intensità macrosismiche (MDPs) idealmente simili alle località secondo le definizioni delle scale macrosismiche (MCS, EMS). In particolare sono state usate le IDPs raccolte attraverso la piattaforma EQN di EMSC espresse in 2 differenti tipologie di intensità EMS (I_{raw} e I_{corr}). La conversione di IDPs in MDPs è stata eseguita attraverso l'applicazione di 5 “tecniche” di clustering spaziale per raggruppare le IDPs e di 4 “metodi” statistici per assegnare le intensità MDPs ai raggruppamenti. Le 5 tecniche sviluppate e testate per questa tesi sono: tecniche grid-based con celle equiareali quadrate (1) ed esagonali (2); aree circolari centrate su località amministrative con raggio proporzionale alla densità media di popolazione (3); tecnica mista (aree circolari+celle equiareali) (4); tecnica density-based DBSCAN (Ester, 1996) (5). I 4 metodi statistici sono: media, mediana, media [15, 85]%, media [25, 75]%. Le distribuzioni di MDPs ottenute con queste tecniche e metodi sono quindi state processate con il codice Boxer (Gasperini et al., 1999 e 2010) per ottenere stime di localizzazione epicentrale e magnitudo. I parametri ottenuti per le 5 tecniche di clustering sono stati confrontati con stime strumentali e con ulteriori parametri (localizzazione e magnitudo) di derivazione macrosismica ricavati sempre tramite Boxer utilizzando le MDPs direttamente fornite da HSIT (INGV), piattaforma di raccolta di risposte di utenti a questionari testuali per l'area italiana. La tecnica di HSIT per la costruzione delle MDPs differisce dalle precedenti ed è basata sull'appartenenza o meno di un punto ad una polyline corrispondente con i limiti amministrativi delle località comunali.

Sono stati analizzati dati per 9 terremoti avvenuti tra il 2016 e il 2019, 7 in territorio italiano, 1 in territorio albanese con epicentro strumentale sulla costa, 1 in territorio greco con epicentro in mare.

L'analisi dei risultati è stata diretta ad individuare regole in relazione alla tipologia di intensità, alla tecnica di clustering e al metodo statistico utilizzati per l'elaborazione delle distribuzioni di MDPs da cui si sono ottenute le stime dei parametri di localizzazione e magnitudo.

I risultati mostrano differenze non trascurabili tra i parametri strumentali e i parametri macrosismici.

Per le magnitudo si ottiene una sistematica sottostima compresa nell'intervallo $[0.0, 2.0]u.m.$, ridotto a $[0.5, 1.0]u.m.$ per gli MDPs costruiti con i metodi statistici di *media e mediana semplici*.

Per le stime epicentrali la distanza dall'epicentro strumentale è compresa in un ampio range, circa $[5, 100]Km$, ma in diversi terremoti si individua un range ridotto di circa $20Km$ di ampiezza in cui ricade un gruppo più cospicuo di esse, tra loro autoconsistenti.

Inoltre nessuna delle tecniche di clustering applicate è risultata in assoluto migliore rispetto alle altre. Solamente per la stima della magnitudo la statistica raccolta, sebbene limitata a 9 eventi, mostra uno score maggiore per la tecnica di clustering basata sulle celle quadrate; mentre per la stima epicentrale lo score maggiore si ottiene per la tecnica delle celle esagonali.

Relativamente alla statistica sulle stime peggiori di magnitudo si ottiene invece uno score maggiore per la tecnica delle aree circolari proporzionali alla densità di popolazione, le cui stime mostrano inoltre generalmente le fluttuazioni più ampie.

Nemmeno tra i metodi statistici utilizzati per l'assegnazione delle intensità agli MDPs è emersa una netta corrispondenza di alcuno di questi con le stime migliori. Si riscontra solamente per le stime delle magnitudo uno score maggiore per il metodo di mediana e una minore dispersione di queste sia con la mediana che con la semplice media.

Un'evidenza riscontrabile per le magnitudo è la sistematicità con cui le stime derivate da IDPs espressi in intensità corretta I_{corr} (Bossu et al., 2017) siano migliori di quelle derivate da IDPs espressi in intensità I_{raw} .

Dal confronto tra stime ottenute da dati HSIT e stime ottenute da dati EMSC è emerso sia per la magnitudo che per l'epicentro un generale accordo, con la differenza di una maggiore stabilità delle stime HSIT, i cui range di variabilità per uno stesso evento sono rispettivamente di $0.2u.m.$ e $20Km$.

È stato inoltre riscontrato un netto miglioramento delle stime da dati HSIT rispetto a quelle da dati EMSC, sia per le magnitudo sia per gli epicentri, in corrispondenza di rapporti crescenti tra i numeri di MDPs $\#MDPs_{HSIT}/\#MDPs_{EMSC}$. Per 3 terremoti $\#MDPs_{HSIT}$ è di un ordine di grandezza superiore rispetto a $\#MDPs_{EMSC}$. Il confronto non è stato possibile per i 2 terremoti con epicentro strumentale al di fuori del territorio italiano, per i quali il sistema nazionale HSIT per la raccolta degli IDPs è insufficiente.

Una possibile spiegazione delle differenze dei risultati ottenuti sia per la localizzazione che per la magnitudo è da individuare nelle caratteristiche dei terremoti utilizzati per l'analisi. La limitazione dell'indagine all'area italiana e la recente attivazione (2016) della piattaforma EQN dell'EMSC per la raccolta degli IDPs hanno comportato la prevalente selezione di terremoti di magnitudo $M < 6$ con numero di IDPs non molto elevati.

Sia per EMSC che per HSIT i dataset mostrano una sottostima del numero delle IDPs nell'area del near-field (in prossimità della sorgente) e un'intrinseca sottorappresentatività delle intensità più elevate (limitata a 8 per HSIT) da cui segue una diretta diminuzione della stima della magnitudo.

Infatti per migliorare l'accuratezza e la precisione delle stime è determinante avere un cospicuo numero di dati di base (IDPs) che campionino il più possibile la regione del near-field, da cui costruire con le tecniche di clustering un maggior numero di MDPs utili all'elaborazione con Boxer.

Tra i 9 terremoti, si sono verificate condizioni favorevoli per la stima solamente per il terremoto d'Albania del 2019 per il quale tutti i risultati ottenuti, sia per l'epicentro sia per la magnitudo, sono in ottimo accordo con i riferimenti strumentali.

I risultati ottenuti per questo evento e l'incremento del numero di IDPs nel tempo, anche in funzione della magnitudo, mostrano come la piattaforma EQN offra possibilità future migliori.

Riconoscimenti

All'ambiente del Corso di Laurea Magistrale in Didattica della Fisica e ai suoi abitanti, ma in modo particolare alla Professoressa Olivia Levrini generatrice di “momenti brevi-lunghi” di Erfahrung (Walter Benjamin) e di spunti per l'introspezione e la riflessione personale prima che professionale.

Ai miei tre vecchi relatori per averci provato insieme a me. Ai nuovi perchè ce l'abbiamo fatta.

Claudia, per i dolci e le leccornie.

Antonio per le risate e gli infusi di entusiasmo tecnico.

A Lorenzo per i riflessi chiari e nitidi, per il valore e il riconoscimento.

A Sergio per le lunghe lezioni di grammatica, di sintassi e di punteggiatura; per la forma, perchè la forma, se è “sforma”, sforma e distorce il contenuto e la sostanza, per non parlare del pensiero...

Ad Eleonora per gli esperimenti sulla trasmissione di ben altro *calore* a distanze prossime ai $280Km$, e per quelli sulla contrazione delle lunghezze; il dettaglio di questi non è pubblicato per motivi di riservatezza.

Ad Alejandra.

A Simone e a messer Friis per il tetto, la scrivania, la luce elettrica, l'accoglienza.

A quegli (in)coscienti dei miei genitori che hanno sostenuto questa folle impresa.

Sebbene non abbia apportato contributo diretto alcuno per la riuscita di questo lavoro, alla dedicataria di questa tesi spetta il riconoscimento per aver permesso la mia permanenza in queste attuali dimensioni dell'esistente negli anni più lunghi, bui e duri della vita.

Grazie per tutto ciò di essenziale che queste e altre persone che ho incontrato non mi hanno dato, perchè ho dovuto crearlo e costruirlo da solo.

Ah! Quasi dimenticavo... Alla musica e a chi la fa.

Bibliografia

- Agostinelli, C., R. Rotondi e E. Varini (2018). «Clustering Macroseismic Fields by Statistical Data Depth Functions.» In: *Springer, Cham*. In: Mola F., Conversano C., Vichi M. (eds) *Classification, (Big) Data Analysis and Statistical Learning. Studies in Classification, Data Analysis, and Knowledge Organization.*, pp. 1649–1658. DOI: https://doi.org/10.1007/978-3-319-55708-3_16.
- Aki, Keiiti (1966). «Generation and Propagation of G Waves from the Niigata Earthquake of June 16, 1964. : Part 2. Estimation of earthquake moment, released energy, and stress-strain drop from the G wave spectrum». In: *Bulletin of the Earthquake Research Institute, University of Tokyo* 44.1, pp. 73–88. URL: <https://ci.nii.ac.jp/naid/120000871042/en/>.
- Almagaia, R. (1914). «Intorno ai primi saggi di carte sismiche». In: *Rivista geografica italiana* 21.7, pp. 463–468.
- Amorese, D., R. Bossu e G. Mazet-Roux (2015). «Automatic Clustering of Macroseismic Intensity Data Points from Internet Questionnaires: Efficiency of the Partitioning around Medoids (PAM)». In: *Seismological Research Letters* 86.4, pp. 1171–1177. URL: <https://doi.org/10.1785/0220140140>.
- Ansari, M.Y., A. Ahmad, S.S. Khan, G Bhushan e Mainuddin (2020). «Spatiotemporal clustering: a review.» In: *Artificial Intelligence Review* 53.10, pp. 2381–2423. URL: <https://doi.org/10.1007/s10462-019-09736-1>.
- Anzari, A., L. Firuzi e L. Etemadsaeed (2015). «Delineation of Seismic Sources in Probabilistic Seismic-Hazard Analysis Using Fuzzy Cluster Analysis and Monte Carlo Simulation.» In: *Bulletin of the Seismological Society of America* 105.4, pp. 2174–2191. DOI: <http://dx.doi.org/10.1785/0120140256>.
- Bakun, W. H. e C. M. Wentworth (1997). «Estimating earthquake location and magnitude from seismic intensity data». In: *Bulletin of the Seismological Society of America* 87.6, pp. 1502–1521.
- Berardi, R., C. Petrungrara, C. Zonetti, L. Magri e M. Mucciarelli (1993). *Mappe di sismicità per l'area italiana*. ISMES. Bergamo, Italy.
- Bisztricsany, E. (1958). «A new method for the determination of the magnitude of earthquakes». In: *Geofiz. Kozlemen* 7.2, pp. 69–76.
- Borman, P. (2002). *IASPEI New Manual of Seismological Observatory Practice (NM-SOP)*. A cura di GeoForschungsZentrum Potsdam. ISBN: 3-9808780-0-7.
- Boschi, E., E. Guidoboni, G. Ferrari, G. Valensise e P. Gasperini (1997). *Catalogo dei Forti Terremoti in Italia dal 461a.C al 1990*. A cura di SGA Sotia Geofisica Ambiente. via Bellombra 24/2 40136 Bologna. ISBN 88-85213-08-1.
- Bossu, R., L. Fallou, M Landes, F. Roussel, S Julien-Laferrriere, J. Roch e R. Steed (2020). «Rapid Public Information and Situational Awareness After the November

- 26, 2019, Albania Earthquake: Lessons Learned From the LastQuake System». In: *Front. Earth Sci.* 8.235. DOI: 10.3389/feart.2020.00235.
- Bossu, R., M. Landes, F. Roussel e R. Steed (2018a). «Felt Reports for Tapid Mapping of Global Earthquake Damage: The Doughnut Effect?». In: *Seismological Research Letters* 89.1, pp. 138–144. URL: <https://doi.org/10.1785/0220170129>.
- Bossu, R., M. Landes, F. Roussel, R. Steed, G. Mazet-Roux, Stacey S. Martin e S. Hough (2017). «Thumbnail-Based Questionnaires for the Rapid and Efficient Collection of Macroseismic Data from Global Earthquakes». In: *Seismological Research Letters* 88.1, pp. 72–81. URL: <https://doi.org/10.1785/0220160120>.
- Bossu, R., S. Lefebvre, Y. Cansi e G. Mazet-Roux (2014). «Characterization of the 2011 Mineral, Virginia, Earthquake Effects and Epicenter from Website Traffic Analysis». In: *Seismological Research Letters* 85.1, pp. 91–97. URL: <https://doi.org/10.1785/0220130106>.
- Bossu, R., F. Roussel, L. Fallou, M. Landes, R. Steed, G. Mazet-Roux, A. Dupont, Frobert L. e Petersen L. (2018b). «LastQuake: From rapid information to global seismic risk reduction». In: *Elsevier, International Journal of Disaster Risk Reduction* 28, pp. 32–42. URL: <https://doi.org/10.1016/j.ijdr.2018.02.024>.
- Bossu, R., R. Steed, F. Roussel, M. Landes, A. Fuenzalida, E. Matrullo, A. Dupont, J. Roch e L. Fallou (2019). «App Earthquake Detection and Automatic Mapping of Felt Area». In: *Seismological Research Letters* 90.1, pp. 305–312. URL: <https://doi.org/10.1785/0220180185>.
- BOXER (2010). *BOXER User Guide. Version 4.0 (beta)*. URL: <https://emidius.mi.ingv.it/boxer/manuals/manual40.pdf>.
- Cajka, Mary G. e S. Halchuk (1998). «Collecting Intensity Data via the Internet: The Cap-Rouge, Quebec Earthquake». In: *Seismological Research Letters, Eastern Section* 69.6, pp. 585–587. URL: <https://doi.org/10.1785/gssr1.69.6.585>.
- Camassi, R. e M. Stucchi (1997). «NT4.1.1, un catalogo parametrico di terremoti di area italiana al di sopra della soglia del danno. GNDT, Milano». In: p. 95.
- Cancani, A. (1904). «Sur l'employ d'une double echelle sismique des intensités, empirique et absolute». In: *G. Beitr. Ergänzungsband* 2, pp. 281–283.
- Caprio, M., B. Tarigan, B. Worden, S. Wiemer e D. J. Wald (2015). «Ground Motion to Intensity Conversion Equations (GMICEs): A Global Relationship and Evaluation of Regional Dependency». In: *Bull. Seismol. Soc. Am.* 105.3. DOI: 10.1785/0120140286.
- Carletti, F. e P. Gasperini (2003). «Lateral variations of seismic intensity attenuation in Italy». In: *Geophys. J. Int.* 155, pp. 839–856.
- Castro, R. R., M. Mucciarelli, G. Monachesi, F. Pacor e R. Berardi (1999). «A review of nonparametric attenuation functions computed for different regions of Italy». In: *Ann. Geofis.* 42, pp. 734–748. URL: <https://doi.org/10.4401/ag-3753>.
- Ceciè, I. e R. Musson (2004). «Macroseismic Surveys in Theory and Practice». In: *Natural Hazards* 31, pp. 39–61. URL: <https://doi.org/10.1023/B:NHAZ.0000020255.00986.37>.
- Cohn, J.P. (2008). «Citizen Science: Can Volunteers Do Real Research?». In: *University of California Press on behalf of the American Institute of Biological Sciences, BioScience* 58.3, pp. 192–197. DOI: 10.1641/B580303.

- CPTI99 (1999). «Catalogo Parametrico dei Terremoti Italiani». In: *Catalogo Parametrico dei Terremoti Italiani, versione 1 (CPTI99, luglio 1999)*. ING-GNDT-SGA-SSN, Bologna, 92 pp. URL: <https://doi.org/10.6092/INGV.IT-CPTI99>.
- Czeczce, B. e I. Bondar (2019). «Hierarchical cluster analysis and multiple event relocation of seismic event clusters in Hungary between 2000 and 2016». In: *Journal of Seismology* 23, pp. 1313–1326. DOI: <https://doi.org/10.1007/s10950-019-09868-5>.
- Davison, C. (1921). «On scales of seismic intensity and on the construction of isoseismal lines». In: *Bull. Seismol. Soc. Am.* 11, pp. 95–129.
- De Rubeis, V., P. Sbarra, P. Tosi e D. Sorrentino (2019). *Hai Sentito Il Terremoto (HSIT) - Macroseismic intensity database 2007-2018, version 1*. URL: <https://doi.org/10.13127/HSIT/I.1>.
- Di Giacomo, D., E. R. Engdahl e D. A. Storchak (2018). «The ISC-GEM Earthquake Catalogue (1904-2014): status after the Extension Project». In: *Earth Syst. Sci. Data* 10, pp. 1877–1899. DOI: [10.5194/essd-10-1877-2018](https://doi.org/10.5194/essd-10-1877-2018).
- Dolce, M. e D. Di Bucci (2014). «National Civil Protection Organization and technical activities in the 2012 Emilia earthquakes (Italy)». In: *Bull Earthquake Eng* 12, pp. 2231–2253. URL: <https://doi.org/10.1007/s10518-014-9597-x>.
- Efron, B. e R. Tibshirani (1986). «Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy». In: *Statist. Sci.* 1.1, pp. 54–75. DOI: [10.1214/ss/1177013815](https://doi.org/10.1214/ss/1177013815). URL: <https://projecteuclid.org/euclid.ss/1177013815>.
- Elmi, C. e M. Diretto (1996). *Geologia, pagine 217*. A cura di Pitagora Editrice.
- Ester, M., H.P. Kriegel, J. Sander e X. Xiaowei (1996). «A density-based algorithm for discovering clusters in large spatial databases with noise». In: *Proceedings of the Second International Conference on Knowledge Discovery in Databases and Data Mining*. A cura di OR: AAAI Press Portland, pp. 226–231.
- Everitt, B. S. (2011). *Cluster analysis*. A cura di U.K: Wiley. analysis. Chichester West Sussex. Wiley series in probability and statistics.
- Fan, Z. e X. Xu (2019). «Application and visualization of typical clustering algorithms in seismic data analysis». In: *Procedia Computer Science* 151. The 10th International Conference on Ambient Systems, Networks and Technologies (ANT 2019) / The 2nd International Conference on Emerging Data and Industry 4.0 (EDI40 2019) / Affiliated Workshops, pp. 171–178. ISSN: 1877-0509. DOI: <https://doi.org/10.1016/j.procs.2019.04.026>. URL: <https://www.sciencedirect.com/science/article/pii/S1877050919304879>.
- Fasan, M., P. L. A. Renault e J. U. Klugel (2019). *A database of damage-consistent (intensity-based) natural and synthetic accelerograms for seismic risk assessment*. A cura di International Association for Structural Mechanics in Reactor Technology (IASMiRT). URL: <http://hdl.handle.net/11368/2955372>.
- Fechner, G. T. (1860). «Elemente der psychophysik [Elements of psychophysics]». In: 1. Leipzig, Austria: Breitkopf und Harterl.
- Ferrari, G. e E. Guidoboni (2000). «Seismic scenarios and assesment of intensity: some criteria for the use of the MCS scale.» In: *Ann Geofis* 43, pp. 707–720.
- Fisher, N. I. (1993). *Statistical analisys of circular data*. A cura di United Kingdom Cambridge University Press, p. 277.

- Fraley, C. e A. E. Raftery (1998). «MCLUST: Software for Model-Based Cluster and Discriminant Analysis». In: *The computer journal* 41.8.
- Fuller, W. A. (1987). *Mesurment Error Models*. A cura di Wiley. New York, p. 440.
- Galanopoulos, A.G. (1961). «On Magnitude Determination by Using Macro seismic Data». In: *INGV Annals of Geophysics* 14.3, pp. 225–253. DOI: 10.4401/ag-5297. URL: <https://www.annalsofgeophysics.eu/index.php/annals/article/download/5297/5352>.
- Gasparini, P. (2001). «The attenuation of seismic intensity in Italy: a bilinear shape indicates dominance of deep phases at epicentral distances longer than 45 km». In: *Bull. Seismol. Soc. Am.* 91, pp. 826–841.
- Gasparini, P., F. Bernardini, G. Valensise e E. Boschi (1999). «Defining seismogenic sources from historical felt reports». In: *Bulletin of the Seismological Society of America* 89.1, pp. 94–110. URL: <https://pubs.geoscienceworld.org/ssa/bssa/article-abstract/89/1/94/340514/Defining-seismogenic-sources-from-historical?redirectedFrom=fulltext>.
- Gasparini, P. e G. Ferrari (1995). «Stima dei parametri sintetici». In: *Catalogo dei Forti Terremoti in Italia dal 461 a.C. al 1980*. E. Boschi et al., (a cura di) ING-SGA, Bologna, pp. 96–111. URL: <https://emidius.mi.ingv.it/ASMI/study/CFTI>.
- (2000). «Deriving numerical estimates from descriptive information: the computation of earthquake parameters». In: *INGV Annals of Geophysics* 43.4, pp. 729–746. URL: <https://doi.org/10.4401/ag-3670>.
- Gasparini, P., G. Vannucci, D. Tripone e E. Boschi (2010). «The Location and Sizing of Historical Earthquakes Using the Attenuation of Macro seismic Intensity with Distance». In: *Bulletin of the Seismological Society of America* Vol. 100.No. 5A, pp. 2035–2066. doi:10.1785/0120090330.
- Gaudiosi, G., R. NAppi, G Alessio e S. Porfido (2014). «Breve storia delle misurazioni dell’Intensit  Macro sismica in Italia da Giuseppe Mercalli fino ai giorni nostri». In: Giuseppe Mercalli, una vita tra insegnamento e ricerca. Contributi presentati per l’inaugurazione dell’Anno Mercalliano – Napoli, 19 Marzo 2014.
- Gee, L. S. e W. S. Leith (2011). *The Global Seismographic Network: U.S. Geological Survey Fact Sheet*, USGS. URL: <https://pubs.usgs.gov/fs/2011/3021/>.
- Georgoulas, G., A. Konstantaras, E. Katsifarakis, C. D. Stylios, E. Maravelakis e G. J. Vachtsevanos (2013). «“Seismic-mass” density-based algorithm for spatio-temporal clustering, Expert Systems with Applications,» in: *Expert Systems with Applications* 40.10, pp. 4183–4189. DOI: <https://doi.org/10.1016/j.eswa.2013.01.028>. URL: <https://www.sciencedirect.com/science/article/pii/S0957417413000419>.
- Gomez-Capera, A. A., M. D’Amico, G. Lanzano, M. Locati e M. Santulin (2020). «Relationships between ground motion parameters and macro seismic intensity for Italy». In: *Bull Earthquake Eng* 18, pp. 5143–5164. URL: <https://doi.org/10.1007/s10518-020-00905-0>.
- Grunthal, G. (1998). *European Macro seismic Scale 1998 (EMS-98)*. Conseil de l’Europe. Cahiers du Centre Europ  @en de G  odyamique et de S  ismologie, Edizione italiana Tertulliani A., Azzaro R., Buffarini G., Luxembourg 2019.
- Guerrieri, L., E. Esposito, A.M. Michetti, S. Porfido e E. Vittori (2012). *La scala di intensita sismica ESI 2007*. URL: <http://eprints.bice.rm.cnr.it/4027/>.

- Guo, Z e Y Ogata (1997). «Statistical relations between the parameters of aftershocks in time, space, and magnitude». In: *Journal of Geophysical Research, Solid Earth* 102. issue B2, pp. 2857–2873. URL: <https://doi.org/10.1029/96JB02946>.
- Gupta, I. N. e O. W. Nuttli (1976). «Spatial Attenuation of Intensities for Central U.S. Earthquakes». In: *Bul. Seism. Sec. Am.* 66, pp. 743–751.
- Gutenberg, B. e C. F. Richter (1942). «Earthquake magnitude, intensity, energy and acceleration». In: *Bull. Seismol. Soc. Am.* 32, pp. 163–191.
- (1956). «Earthquake magnitude, intensity, energy and acceleration». In: *Sull. Seism. Soc. Am.* 46, pp. 105–145.
- Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*. A cura di Springer-Verlag New York, p. 372.
- Han, J., M. Kamber e J. Pei (2011). *Data Mining: Concepts and Techniques 3rd Edition*. A cura di Morgan Kaufmann Publishers The Morgan Kaufmann Series in Data Management Systems. International Geophysics, Academic Press, p. 744.
- Hanks, Thomas C. e Hiroo Kanamori (1979). «A moment magnitude scale». In: *Journal of Geophysical Research* 84.B5, pp. 2348–2350. URL: <https://doi.org/10.1029/JB084iB05p02348>.
- Havskov, Jens e Ottemoller Lars (2008). *Routine Data Processing in Earthquake Seismology with Sample Data, Exercises and Software, pages 347*. A cura di Dordrecht Springer. URL: <https://doi.org/10.1007/978-90-481-8697-6>.
- Helmstetter, A., Y. Kagan e D. Jackson (2007). «High-resolution Time-independent Grid-based Forecast for $M \geq 5$ Earthquakes in California.» In: *Seismological Research Letters, Seismological Society of America* 78.1, pp. 78–86. URL: https://hal.archives-ouvertes.fr/hal-00195399/file/HKL_SRL2006_final.pdf.
- IMSL (1991). *International Mathematic Statistic Library IMSL 1991 MATHLibrary User's Manual Version 2.0*. A cura di Visual Numerics Inc. Huston, Texas, p. 1313.
- INGV (2020). *RSN (Rete Sismica Nazionale)*. Istituto Nazionale di Geofisica e Vulcanologia. URL: <http://www.ingv.it/it/monitoraggio-e-infrastrutture-per-la-ricerca/reti-di-monitoraggio/1-ingv-e-le-sue-reti/rns>.
- INGV, Milano-Roma, cur. (2004). *Gruppo di Lavoro MPS2004. Redazione della mappa di pericolosità sismica prevista dall'Ordinanza PCM 3274 del 20 marzo 2003. Rapporto Conclusivo per il Dipartimento della Protezione Civile*. 65pp., 5 appendici.
- JMA (1996). *Tables explaining the JMA Seismic Intensity Scale*. URL: <http://www.jma.go.jp/jma/en/Activities/inttable.html>.
- Karri, N. A., M. Yousuf Ansari e A. Pathak (2018). «Identification of Seismic Zones of India using DBSCAN». In: *2018 International Conference on Computing, Power and Communication Technologies (GUCON)*, pp. 65–69. DOI: 10.1109/GUCON.2018.8674964.
- Kazemi-Beydokhti, M., R. Ali Abbaspour e M. Mojarab (2017). «Spatio-Temporal Modeling of Seismic Provinces of Iran Using DBSCAN Algorithm». In: *Pure and Applied Geophysics* 174.5. Provided by the SAO/NASA Astrophysics Data System, pp. 1937–1952. DOI: 10.1007/s00024-017-1507-0. URL: <https://ui.adsabs.harvard.edu/abs/2017PApGe.174.1937K>.
- Kennett, B.L.N. e E.R. Engdahl (1995). «Constraints on seismic velocities in the Earth from traveltimes». In: *Geophysical Journal International* 122.1, pp. 108–124. URL: <https://doi.org/10.1111/j.1365-246X.1995.tb03540.x>.

- Kondorskaya, N.V. e N.V. Shebalin (1982). *New Catalogue of strong earthquakes in the USSR from Ancient Times through 1975. Report SE-31, 2nd edition, Boulder, Colorado, 608 pp., 1st edition, 1977, Moscow, 536 pp., in Russian.* URL: <https://www.emidius.eu/AHEAD/study/KOSH982>.
- Kovesligethy, R. (1906). «A makroszeizmikus rengesek feldolgozasa». In: *Math. es Termeszettudományi Ertesito* 24, pp. 349–368.
- Lagomarsino, S. e S. Giovinazzi (2006). «Macroseismic and mechanical models for the vulnerability and damage assessment of current buildings». In: *Bull Earthquake Eng* 4, pp. 415–443. URL: <https://doi.org/10.1007/s10518-006-9024-z>.
- Landes, M., R. Bossu, L. Fallou, J. Roch, F. Roussel, R. Steed e E. Careno (2019). *Report on 2018 Operational Activities.* URL: https://www.emsc-csem.org/Doc/EMSC_DOCS/EMSC_RT_activities_2018.pdf.
- Landes, Matthieu (2017). *Specifications of the EMSC testimony's Service.* Version 1.0. Project EPOS, Grant agreement n676564, WP 8.5. URL: <https://www.emsc-csem.org/Documents/>.
- Lee, W., P. Jennings, C. Kisslinger e H. Kanamori (2002). *IASPEI International Handbook of Earthquake and Engineering Seismology.*
- Lomnitz, C. (2006). «Three theorems of earthquake location». In: *Bull. Seismol. Soc. Am.* 96, pp. 306–312.
- Mai, T. S. (2014). *Density-based Algorithms for Active and Anytime Clustering.* A cura di Munchen Core Universitat Maxim Sigill. URL: <https://core.ac.uk/download/pdf/79054726.pdf>.
- Manchuel, K., P. Traversa, D. Baumont, M. Cara, E. Nayman e C. Durouchoux (2018). «The French seismic CATalogue (FCAT-17)». In: *Bull. Earthquake. Eng.* 16, pp. 2227–2251. URL: <https://doi.org/10.1007/s10518-017-0236-1>.
- Manning, C.D., P. Raghavan e H. Schutze (2008). *Introduction to Information Retrieval.* A cura di Cambridge University Press. chapter 16. ISBN 0521865719. URL: <https://nlp.stanford.edu/IR-book/>.
- Mantyniemi, P. (2017). «Macroseismology in Finland from the 1730s to yhe 2000s. Part 2: From an Obligation of thee Learned Elite to Citizen Science». In: *Geophysica* 52.2, pp. 23–41.
- Maruyama, Takuo (1963). «On the force equivalents of dynamic elastic dislocations with reference to the earthquake mechanism». In: *Bull. Earthquake Res. Inst. Univ. Tokyo* 41.3, pp. 467–486. ISSN: 0040-8972. URL: <https://ci.nii.ac.jp/naid/120000866527/en/>.
- Mazet-Roux, G., R. Bossu, E. Carreno e J. Guilbert (2008). *EMSC Real Time Earthquake Information Services.* EMSC.
- (2010). *EMSC real time earthquake information services, European-Mediterranean Seismological Centre (EMSC) Report, 31 pp.* URL: www.emsc-csem.org/Files/news/EMSC/EMSC_RTEI_services.pdf (last accessed 20December 202014) ..
- Medvedev, S., W. Sponheuer e V. Karnik (1964). *Neue seismische Skala Intensity scale of earthquakes. 7.* Tagung der Europäischen Seismologischen Kommission vom 24.9. bis 30.9.1962. In: Jena, Veroff. Institut fur Bodendynamik und Erdbebenforschung in Jena, vol 77. Deutsche Akademie der Wissenschaften zu Berlin, pp 69-76.

- Mercalli, G. (1902). «Sulle modificazioni proposte alla scala sismica De Rossi-Forel». In: *Bollettino della Societa Sismologica Italiana* 8, pp. 184–191.
- Michetti, A. M., E. Esposito, L. Guerrieri, S. Porfido, L. Serva, R. Tatevossian, E. Vittori, F. Audemard, T. Azuma, J. Clague, V. Commerci, A. Gurpinar, J. Mc Calpin, B. Mohammadioun, N. A. Morner, Y. Ota e E. Roghazin (2007). *Environmental Seismic Intensity scale - ESI 2007 La scala di Intensita Sismica basata sugli effetti ambientali - ESI 2007*. Memorie descrittive della carta geologica D'Italia, LXXIV. APAT. URL: <http://www.isprambiente.gov.it/it/progetti/inquascale/documenti>.
- Milone, G. (2007). *Temporal Data Mining: tecniche e algoritmi di clustering*. PhD thesis. Universita degli studi di Napoli Federico II.
- Moroni, A. e M. Stucchi (1989). «Materials for the investigation of the 1564, Maritime Alps Earthquake». In: *EC project "Review of Historical Seismicity in Europe" (RHISE)*.
- Murphy, J. R. e L. J. O'Brien (1977). «The correlation of peak ground acceleration amplitude with seismic intensity and other physical variable». In: *Bull. Seismol. Soc. Am.* 67.3, pp. 877–915.
- Mussett, A.E. e M. Aftab Khan (2003). *Esplorazione del sottosuolo, una introduzione alla geofisica applicata, pages 440*. A cura di Bologna Zanichelli.
- Musson, R. M. W. (1996). «Determination of parameters for historical British earthquakes». In: *INGV Annals of Geophysics* 39.5, pp. 1041–1047. DOI: 10.4401/ag-4035. URL: <https://www.annalsofgeophysics.eu/index.php/annals/article/download/4035/4100>.
- Musson, R. e I. Cecià (2011). *Intensity and Intensity Scales*. June. Chapter 12. DOI: DOI:10.2312/GFZ.NMSOP-2_ch12.
- Musson, R., G. Grunthal e M. Stucchi (2010). «The comparison of macroseismic intensity scale». In: *J Seismol* 14, pp. 413–428. DOI: 10.1007/s10950-009-9172-0.
- Musson, R.M.W. (2007). *Macroseismic effects of the 2007 Cape St Vincent earthquake from EMSC online questionnaire*. CSEM/EMSC Newsletter, 22 May 2007.
- Nanda, S.J. e G. Panda (2015). «Design of computationally efficient density-based clustering algorithms». In: *Data and Knowledge Engineering* 95.10, pp. 23–38. DOI: <https://doi.org/10.1016/j.datak.2014.11.004>. URL: <https://www.sciencedirect.com/science/article/pii/S0169023X14001219>.
- Obara, K., K. Kasahara, S. Hori e Y. Okadaa (2005). «A densely distributed high-sensitivity seismograph network in Japan:Hi-net by National Research Institute for Earth Science and DisasterPrevention». In: *Review of Scientific Instruments* 76.021301. URL: <https://doi.org/10.1063/1.1854197>.
- Omori, F. (1900). «Intensity Scale». In: *Earthquake investigation committee publication in foreing languages* 4, pp. 137–141.
- Panza, G. F., R. Cazzaro e F. Vaccari (1997). «Correlation between macroseismic intensity and seismic groundmotion parameters». In: *Ann. Geofis.* 40, pp. 1371–1382.
- Pasolini, C., P. Gasperini, D. Albarello, B. Lolli e V. D'Amico (2008). «The attenuation of seismic intensity in Italy, part I: Theoretical and empirical backgrounds». In: *Bulletin of the Seismological Society of America* 98.2, pp. 682–691. URL: <https://doi.org/10.1785/0120070020>.

- Postpischl, D., A. Branno, E. Esposito e G. Ferrari (1985). «The Irpinia earthquake of 23 november 1980». In: *Atlas of isoseismal maps of Italian Earthquakes*. CNR, ITaly.
- Rajat, P. (2020). *Clustering Geo-location: DBSCAN*. A cura di community of Analytics Vidhya e Data Science professionals. URL: <https://medium.com/analytics-vidhya/clustering-geo-location-dbscan-cadb33b0442e>.
- Rehman, K., P.W. Burton e G.A. Weatherill (2014). «K-means cluster analysis and seismicity partitioning for Pakistan.» In: *Journal of Seismology* 18, pp. 401–419. DOI: <https://doi.org/10.1007/s10950-013-9415-y>.
- Reid, Harry Fielding (1910). *The California Earthquake of April 18, 1906. Volume II. The Mechanics of the Earthquake*. A cura di Carnegie Institution of Washington. Washington, Washington D.C. URL: http://activetectonics.asu.edu/ActiveFaultingSeminar/Papers/Reid_1910.pdf.
- Richter, C. F. (1958). «Elementary Seismology». In: *San Francisco: W. H. Freeman*, p. 768.
- Richter, C.F. (1935). «An instrumental earthquake magnitude scale». In: *Bulletin of the Seismological Society of America* 25.1, pp. 1–32. URL: <https://pubs.geoscienceworld.org/ssa/bssa/article-abstract/25/1/1/115102/An-instrumental-earthquake-magnitude-scale>.
- Rossi, A., A. Tertulliani, R. Azzaro, L. Graziani, A. Rovida, A. Maramai, V. Pessina, S. Hailemikael, G. Buffarini, F. Bernardini, R. Camassi, S. Del Mese, E. Ercolani, A. Fodarella, M. Locati, G. Martini, A. Paciello, S. Paolini, L. Arcoraci, C. Castellano, V. Verrubbi e M. Stucchi (2019). «The 2016-2017 earthquake sequence in Central Italy: macroseismic survey and damage scenario through the EMS-98 intensity assessment». In: *Bulletin of Earthquake Engineering* 17, pp. 2407–2431. URL: <https://doi.org/10.1007/s10518-019-00556-w>.
- Rossi, M.S.L. de (1883). «Programma dell’Osservatorio ed Archivio Centrale Geodinamico presso il R. Comitato Geologico d’Italia.» In: *Bollettino del Vulcanismo Italiano* 10, pp. 3–128.
- Rovida, A., M. Locati, R. Camassi, B. Lolli e P. Gasperini (2019). *Catalogo Parametrico dei Terremoti Italiani (CPTI15), versione 2.0*. Istituto Nazionale di Geofisica e Vulcanologia (INGV). URL: <https://doi.org/10.13127/CPTI/CPTI15.2>.
- (2020). «The Italian earthquake catalogue CPTI15». In: *Bulletin of Earthquake Engineering* 18.7, pp. 2953–2984. URL: <https://doi.org/10.1007/s10518-020-00818-y>.
- Sabeti, H. e A. Javaherian (2009). «Seismic Facies Analysis Based on K-means Clustering Algorithm Using 3D Seismic Attributes.» In: *European Association of Geoscientists & Engineers*. DOI: <https://doi.org/10.3997/2214-4609.20145876>. URL: <https://www.earthdoc.org/content/papers/10.3997/2214-4609.20145876>.
- Sbarra, P., P. Tosi e V. De Rubeis (2010). «Web-based macroseismic survey in Italy: method, validation and results». In: *Nat Hazards* 54, pp. 563–581. DOI: [10.1007/s11069-009-9488-7](https://doi.org/10.1007/s11069-009-9488-7).
- (2014). «How Observer Conditions Impact Earthquake Perception». In: *Seismological Research Letters* 85.2, pp. 306–313. URL: <https://www.doi.org/10.1785/0220130080>.

- Sibol, M.S., G.A. Bollinger e J.B. Birch (1987). «Estimation of magnitudes in central and eastern North America using intensity and felt area». In: *Bulletin of the Seismological Society of America* 77.5, pp. 1635–1654. URL: <https://pubs.geoscienceworld.org/ssa/bssa/article-abstract/77/5/1635/119003/Estimation-of-magnitudes-in-central-and-eastern>.
- Sieberg, A. (1923). «Geologische, Physikalische und Angewandte Erdbebenkunde». In: *G. Fischer, Jena*.
- (1930). «Geologie der Erdbeben». In: *Handbuch der Geophysik* 2.4, pp. 552–555.
- Silvertown, J. (2009). «A new dawn for citizen science». In: *Elsevier Ltd Cell Press* 24.9, pp. 467–471. URL: <https://doi.org/10.1016/j.tree.2009.03.017>.
- Sira, C. (2015). «Macroseismic Intervention Group: The Necessary Field Observation». In: *Ansal A. eds Perspectives on European Earthquake Engineering and Seismology. Geotechnical, Geological and Earthquake Engineering* 39. A cura di Cham Springer. https://doi.org/10.1007/978-3-319-16964-4_16.
- Steed, R.J., A. Fuenzalida, R. Bossu, I. Bondar, Heinloo A., A. Dupont, J. Saul e A. Strollo (2019). «Crowdsourcing triggers rapid, reliable earthquake locations». In: *Science Advances* 5.4, eaau9824. URL: <https://www.doi.org/10.1126/sciadv.aau9824>.
- Storchak, D. A., D. Di Giacomo, I. Bondar, E. R. Engdahl, J. Harris, W. H. K. Lee, A. Villasenor e P. Bormann (2013). «Public Release of the ISC-GEM Global Instrumental Earthquake Catalogue (1900-2009)». In: *Seism. Res. Lett.* 84.5, pp. 810–815. DOI: 10.1785/0220130034.
- Storchak, D. A., D. Di Giacomo, E. R. Engdahl, J. Harris, I. Bondar, W. H. K. Lee, P. Bormann e A. Villasenor (2015). «The ISC-GEM Global Instrumental Earthquake Catalogue (1900-2009): Introduction». In: *Phys. Earth Planet. Int.* 239, pp. 48–63. DOI: 10.1016/j.pepi.2014.06.009.
- Stucchi, M., R. Camassi, A. Rovida, M. Locati, E. Ercolani, C. Meletti, P. Migliavacca, F. Bernardini e R. Azzaro (2007). «DBMI04 il database delle osservazioni macrosismiche dei terremoti italiani utilizzate per la compilazione del catalogo parametrico CPTI04». In: *Quaderni di Geofisica* 49, p. 38. URL: <https://doi.org/10.6092/INGV.IT-DBMI04>.
- Stucchi, M., C. Meletti, V. Montaldo, H. Crowley, G. M. Calvi e E. Boschi (2011). «Seismic Hazard Assessment (2003-2009) for the Italian Building Code». In: *Bull. Seismol. Soc. Am.* 101.4, pp. 1885–1911. DOI: <https://doi.org/10.1785/0120100130>.
- The MathWorks, Inc, cur. (2019a). *DBSCAN*. URL: <https://it.mathworks.com/help/stats/dbscan-clustering.html>.
- cur. (2019b). *dbscan. Density-based spatial clustering of applications with noise (DBSCAN)*. Introduced in R2019a. URL: <https://it.mathworks.com/help/stats/dbscan.html>.
- Topozada, T.R., C. R. Real e D. L. Parke (1981). «Preparation of isoseismal maps and summaries of reported effects for pre-1900 California earthquakes». In: *Calif. Div. Mines Geol. Open File Rept. 81-11 SAC, 182pp*.
- Tosi, P., V. De Rubeis, P. Sbarra e D. Sorrentino (2007). *Hai Sentito Il Terremoto (HSIT)*. URL: <https://doi.org/10.13127/HSIT>.

- Tosi, P., P. Sbarra, V. De Rubeis e C. Ferrari (2015). «Macroseismic Intensity Assessment Method for Web Questionnaires». In: *Seismological Research Letters* 86.3. DOI: 10.1785/0220140229.
- Trifunac, M. D. e A. G. Brady (1975). «On the correlation of seismic intensity with peaks of recorded ground motion». In: *Bull. Seismol. Soc. Am.* 65.1, pp. 139–162.
- Tripone, D., G. Vannucci e P. Gasperini (2010). *Win-Boxer, a user-friendly interface to set option and parameter values of the Boxer code for computing earthquake parameterf from macroseismic data*. In preparation.
- USGS, cur. *Earthquake Hazards Program. ANSS Comprehensive Earthquake Catalog (ComCat) Documentation*. URL: <https://earthquake.usgs.gov/data/comcat/>.
- Vannucci, G., D. Tripone, P. Gasperini, G. Ferrari e B. Lolli (2015). «Automated assessment of macroseismic intensity from written sources using the fuzzy sets». In: *Bull Earthquake Eng.* DOI: 10.1007/s10518-015-9759-5.
- Wald, D. J., V. Quinteriano, B. Worden, M. Hopper e J. W. Dewey (2011). «USGS “Did You Feel It?” Internet-based macroseismic intensity maps». In: *Annals of Geophysics* 54.6. DOI: 10.4401/ag-5354.
- Wald, D.J., V. Quinteriano, L.A. Dengler e J.W. Dewey (1999). «Utilization of the Internet for Rapid Community Intensity Maps». In: *Seismological Research Letters* 70.6, pp. 680–697. URL: <https://doi.org/10.1785/gssr1.70.6.680>.
- Weber, E. H. (1834). *De pulsu, resorptione, auditu et tactu [On simulation, response, hearing and touch]*. Annotationes Anatomical et physiological. Leipzig, Austria: Koehler.
- Wells, D.L. e K.J. Coppersmith (1994). «New empirical relationships among magnitude, rupture length, rupture width, rupture area, and surface displacement.» In: *Bulletin of the Seismological Society of America* 84.4, pp. 974–1002. URL: <https://pubs.geoscienceworld.org/ssa/bssa/article-abstract/84/4/974/119792/New-empirical-relationships-among-magnitude>.
- Wesseloo, J., K. Woodward e J. Pereira (2014). «Grid-based analysis of seismic data.» In: *Journal of the Southern African Institute of Mining and Metallurgy, Johannesburg* 114.10, pp. 815–822.
- Wilcock, W. (2016). *Faulting and Earthquake focal mechanism*. URL: https://www.ocean.washington.edu/courses/oc410/labs/lab11_focal_mechanisms.pdf.
- Wood, H.O. e F. Neumann (1931). «Modified Mercalli Intensity Scale of 1931». In: *Bulletin of the Seismological Society of America* 21, pp. 277–283.
- Yamazaki, F., S. Noda e K. Meguro (1998). «Developments of early earthquake damage assessment systems in Japan». In: *7th international conference on structural safety and reliability*, pp. 1573–1580.
- Zonno, G. e et al. (2011). *Urban Prevention Strategies using Macroseismic and Fault Sources (UPStrat-MAFA)*. URL: https://ec.europa.eu/echo/funding-evaluations/financing-civil-protection-europe/selected-projects/urban-prevention-strategies_en. Brussels, 6 February 2012 - Joint kick-off meeting of the selected projects under the Call OJ 2011/C49/09.

Appendice A

Costruzione delle griglie di quadrati ed di esagoni, equiareali, di passo variabile.

Per ciascun terremoto lo spazio geografico relativo alla particolare area di analisi viene *tassellato* mediante la costruzione di due griglie regolari equiareali, una di quadrati e una di esagoni, entrambe con passo variabile (Figure A.1, A.2). I calcoli sono eseguiti approssimando la Terra ad una superficie sferica di raggio $6371Km$.

Per ciascuna griglia si è proceduto a costruire file parallele, di latitudine crescente, di punti equidistanziati in longitudine.

Per la griglia di quadrati di passo p la distanza in longitudine è pari a p_{Deg} , corrispondente al lato di una cella quadrata in unità di grado decimale; per la griglia di esagoni si impone invece una distanza in longitudine di r_{DegHEX} , corrispondente al lato r di una cella esagonale espresso in gradi decimali ed equivalente alla distanza *centro – vertice* di esagono.

Ciascuna fila parallela è distanziata in latitudine dalla precedente: per la griglia di quadrati di una misura pari ad un semi-lato $p_{Deg}/2$, per la griglia di esagoni di una misura pari a ap_{DegHEX} corrispondente all'apotema di un esagono; inoltre per entrambe le costruzioni di griglia le file dispari sono sfasate in longitudine dalle file pari di una misura uguale al semi-lato del rispettivo poligono di griglia.

Si procede dunque a classificare i punti costruiti come *centri* o come *vertici* di cella, rispettivamente punti rossi e punti blu in Figura A.1, A.2.

Scelto il *punto origine* per l'inizializzazione (esplicitamente indicato nelle figure A.1, A.2) questo viene classificato come *centro di cella*: per la griglia di quadrati il *punto origine* è $(minlat - p_{Deg}, minlon - p_{Deg})$ ¹, per la griglia di esagoni è $(minlat - 2ap_{DegHEX}, minlon - 2r_{DegHEX})$.

Per la griglia di quadrati si classificano i punti delle file parallele pari come *centri di cella*, ed i punti delle file dispari come *vertici di cella*; invece per la griglia di esagoni la classificazione procede lungo una stessa fila per longitudine crescente seguendo il pattern *centro-vertice-vertice* per le file pari e, tenendo conto dello sfasamento, seguendo il pattern *vertice-vertice-centro* per le file dispari.

¹dove *minlat* e *minlon* sono i minimi delle coordinate dell'intero dataset di IDPs(EMSC) da clusterizzare per un particolare terremoto.

Per ciascun *centro* si trovano i relativi *vertici* di poligono cercandoli tra i punti a distanza d da esso a diversi angoli azimutali. Per una cella quadrata si avranno 4 vertici agli angoli azimutali (45, 135, 225, 315) gradi decimali, con $p_{Deg}\sqrt{2}/2 < d < p_{Deg}\sqrt{10}/2$; per una cella esagonale si avranno 6 vertici agli angoli azimutali (30, 90, 150, 210, 270, 330) gradi decimali con $r_{DegHEX} < d < 2r_{DegHEX}$.

La conoscenza delle coordinate di tutti i vertici di una cella determina la cella stessa.

Con questa procedura è possibile costruire celle equiareali.

La funzione “areaint” di Matlab (<https://it.mathworks.com/help/map/ref/areaint.html>) consente di calcolare l’area della superficie di ciascuna cella della griglia mediante una procedura d’integrazione lineare che ha un’accuratezza inversamente proporzionale alla distanza tra i punti perimetrali forniti come argomento della funzione stessa. Come punti perimetrali sono stati forniti i soli vertici di poligono.

Le celle quadrate costruite con passo/lato $p = 10Km$ hanno grossomodo un’area di $75Km^2$, mentre quelle esagonali costruite con lato $r = 5Km$ hanno area di circa $50Km^2$.

Sia per la griglia di quadrati che per la griglia di esagoni si è ottenuto un errore percentuale sull’area di cella di circa lo 0.03%, a cui corrisponde un errore lineare dell’ordine del metro. L’errore è stimato per le celle di una griglia con le formule 3.1, dove A_i è il vettore delle aree delle celle, μ la relativa media, ecc.

Infine per la visualizzazione della griglia è stato utilizzato il protocollo necessario a poter applicare la funzione “geoplot” (<https://it.mathworks.com/help/matlab/ref/geoplot.html>) di Matlab: definire una coppia di matrici le quali hanno per colonne le coordinate dei vertici dei singoli poligoni, una per le latitudini e l’altra per le longitudini.

Per tutte le mappe del presente lavoro la rappresentazione grafica delle griglie risulta leggermente deformata a causa della proiezione cartografica non-equiareale.

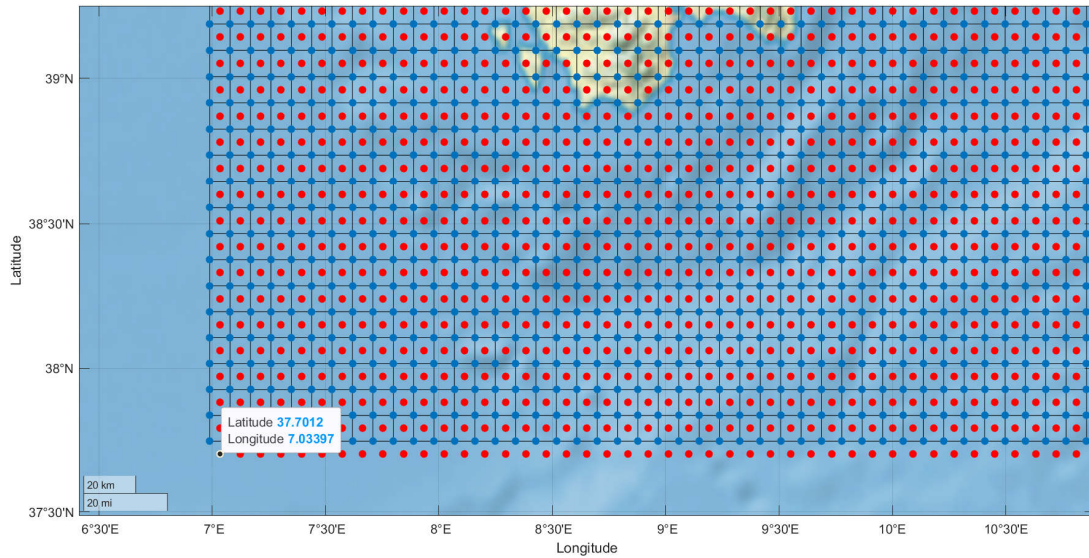


Figura A.1: Tassellazione dello spazio con griglia di quadrati equiareali. In rosso i centri dei poligoni di griglia, in blu i vertici. È evidenziato il punto origine con cui si inizializza la costruzione. Sono rappresentati i lati delle sole celle per le quali sono calcolati tutti i vertici. La griglia risulta leggermente deformata a causa della proiezione cartografica non-equiareale.

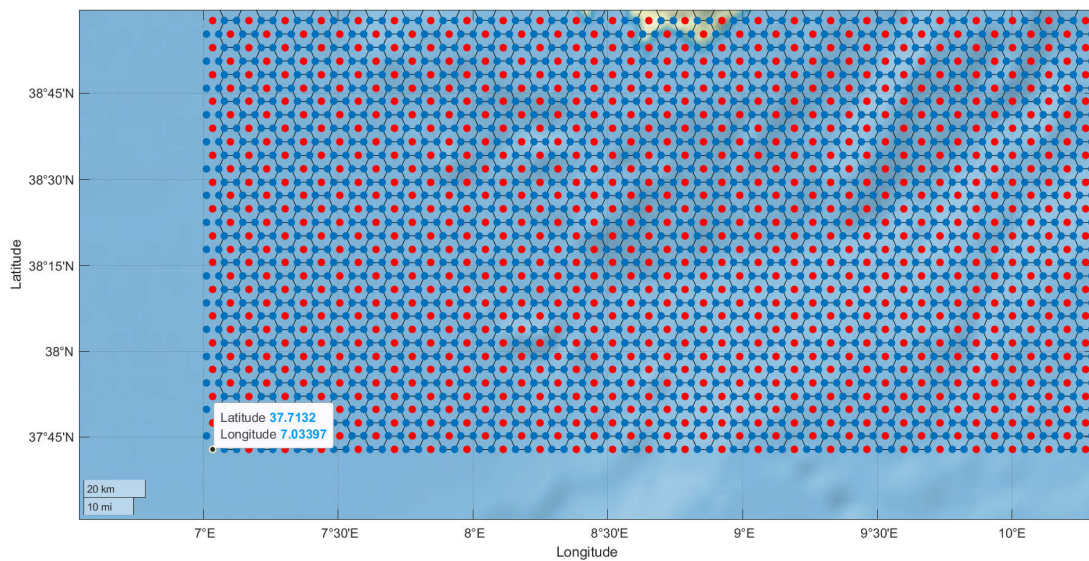


Figura A.2: Tassellazione dello spazio con griglia di quadrati equiareali. Per i dettagli vedere figura A.1.