

PARTE III

Modellazioni innovative della pericolosità

10. ANALISI STATISTICA DI UNA ZONAZIONE UMBRIA-MARCHE E STUDIO DI VARIAZIONI DELLA DISTRIBUZIONE DI PROBABILITA' DEL TEMPO DI INTERCORRENZA

Renata Rotondi

IAMI, CNR, Milano

10.1 Introduzione

Alcuni studi geologici svolti a seguito del terremoto del 26 settembre 1997 (Boncio e Lavecchia, 1999) hanno identificato tre allineamenti principali di faglie in direzione NNW - SSE che, riferiti alla zonazione ZS4.0 (Scandone, 1997; Meletti et al., 2000), interessano principalmente le zone 45, 46, 47, 51, 52. Questi risultati hanno suggerito di analizzare, con metodi statistici, una nuova zonazione dell'area Umbria-Marche, ottenuta estendendo la zona 47 (z47) di ZS4.0 con l'aggiunta, rispettivamente, delle zone 45, 46 e 51, 52; in questo modo si sono ottenute due nuove aree che abbiamo indicato con z47.45-6 e z47.51-2. Dal catalogo NT4.1.1 (Camassi e Stucchi, 1997) abbiamo poi estratto gli insiemi degli eventi con magnitudo $M_S \geq 4.4$ appartenenti a queste due aree e a z47. Dovendo premettere a qualsiasi analisi statistica l'esame della qualità dei dati, abbiamo preso in considerazione il problema dell'incompletezza del catalogo con l'obiettivo di stimare la data da cui ciascuno degli insiemi di dati in esame può essere considerato completo. In NT4.1.1, come in ogni catalogo storico, convivono non-stazionarietà dovute alla mancata registrazione di eventi e quelle dovute a cambiamenti strutturali nel processo fisico; è però ragionevole pensare che le prime costituiscano l'aspetto predominante, soprattutto nella parte iniziale del catalogo. Abbiamo perciò assunto che in un'analisi iniziale si possa considerare la sequenza dei tempi di accadimento di un'area come una realizzazione di un processo di Poisson generalizzato la cui funzione intensità $\lambda(t)$ è una funzione a passi, cioè ha valore costante h_1 fino a un certo tempo s e poi assume un altro valore h_2 , in genere superiore ad h_1 (Rotondi e Garavaglia, 1998). Le variabili s e h_2 , indicanti rispettivamente l'inizio della parte completa del catalogo e il corrispondente tasso, sono state considerate come variabili casuali e stimate attraverso nuovi metodi di simulazione stocastica.

Abbiamo poi analizzato più in dettaglio la parte completa di ciascun insieme di dati al fine di identificarvi non-stazionarietà del processo fisico, riconoscibili da variazioni di indicatori del regime sismico quale la distribuzione di probabilità del tempo di intercor-

Contributo dell'UO IAMI

CNR - Istituto per le Applicazioni della Matematica e dell'Informatica, Via Ampère 56, 20131, Milano

Responsabile: R. Rotondi

e-mail 1° autore: reni@iami.mi.cnr.it

renza. Per far ciò, senza fissare la forma funzionale di tale distribuzione, abbiamo applicato un metodo di stima non parametrica basato su misture di processi stocastici di Dirichlet (Antoniak, 1974); tale metodo è stato inserito in un problema più ampio di identificazione di più punti di cambio corrispondenti, in questo caso, ai tempi in cui si è verificato un cambiamento di distribuzione.

10.2 Analisi statistica di una zonazione Umbria-Marche

Abbiamo assunto che la sequenza dei tempi di accadimento y_1, y_2, \dots, y_n nell'intervallo $[T_1, T_2]$ costituisca una realizzazione di un processo di Poisson con funzione intensità $\lambda(t) = h_1 I_{[T_1, s]}(t) + h_2 I_{[s, T_2]}(t)$; quindi la verosimiglianza ha la forma:

$$p(\mathbf{y} | h_1, h_2, s) = h_1^{\sum_{i=1}^n \delta_{[T_1, s]}(y_i)} e^{-h_1(s-T_1)} h_2^{\sum_{i=1}^n \delta_{[s, T_2]}(y_i)} e^{-h_2(T_2-s)}.$$

Nel modello bayesiano gerarchico che abbiamo adottato sia il punto di cambio s che i tassi h_1, h_2 e un parametro delle loro distribuzioni iniziali sono variabili casuali, a cui abbiamo assegnato distribuzioni di probabilità sulla base della nostra conoscenza iniziale sul fenomeno. L'inferenza può agevolmente essere effettuata applicando metodi di simulazione stocastica, i cosiddetti metodi *Markov chain Monte Carlo* (MCMC, vedi Gelfand and Smith, 1990; Carlin et al., 1992); attraverso loro si genera un campione di m valori dipendenti per le variabili da stimare, $(s^{(j)}, h_1^{(j)}, h_2^{(j)})_{j=1}^m$, da cui poi si ricavano stime delle loro distribuzioni e degli indici sintetici: media, moda, varianza e quantili. Così, ad esempio, la stima della funzione densità $f(s)$ è data da:

$$\tilde{f}(s) = \sum_{j=1}^m f(s | h_1^{(j)}, h_2^{(j)}, \mathbf{y}) / m.$$

dove $f(s | \cdot)$ indica la densità condizionata di s , e le stime della media e della varianza sono date rispettivamente da:

$$\tilde{s} = \sum_{j=1}^m s^{(j)} / m \quad \tilde{\sigma}_s^2 = \sum_{j=1}^m (s^{(j)} - \tilde{s})^2 / m.$$

Lo stesso vale per h_1 and h_2 . Analogamente si ottiene anche la stima del tasso medio a posteriori del processo di Poisson:

$$\hat{\lambda}(t) = \frac{1}{m} \sum_{j=1}^m \left\{ h_1^{(j)} I_{[T_1, s^{(j)}]}(t) + h_2^{(j)} I_{[s^{(j)}, T_2]}(t) \right\}.$$

I valori ottenuti per la zona 47 sono raccolti nella Tab. 1; osserviamo che oltre ai suddetti indicatori sintetici abbiamo anche riportato l'intervallo di minima ampiezza in cui la variabile h_2 cade con probabilità 0.95 (Highest Posterior Density credible set - HPD). Non riportiamo la rappresentazione grafica dei risultati perchè formalmente simi-

Tab. 1 - Stime relative alla completezza del catalogo della zona 47 ($M_s \geq 4.4$).

	media	moda	deviazione standard				
s	1865.0	1873.1	30.28				
h_2	0.366	0.364	0.0542				
<i>quantili</i>	$q_{0.025}$	$q_{0.05}$	$q_{0.25}$	$q_{0.5}$	$q_{0.75}$	$q_{0.95}$	$q_{0.975}$
s	1744.8	1852.0	1868.7	1871.8	1874.7	1878.0	1878.6
h_2	0.227	0.261	0.327	0.366	0.407	0.468	0.490
<i>intervallo .95HPD</i>	q_l	q_h					
h_2	0.239	0.499					

le a quella per la zona z47.51-2.

I risultati relativi alla zona z47.51-2 sono raccolti nella Tab. 2 e in Fig. 1; come per z47 l'inizio della parte di catalogo stimata completa è intorno al 1870 con deviazione standard molto ridotta. La rappresentazione grafica dei risultati di Fig. 1 comprende, riquadro superiore, la curva del numero cumulato degli eventi (linea punteggiata) e il tasso medio $\hat{\lambda}(t)$ (linea continua); nella parte inferiore vi sono invece le densità di probabilità stimate di s e h_2 . La media e la moda a posteriori di s sono indicate rispettivamente con un quadratino pieno e uno vuoto. Riteniamo opportuno usare come stimatore della variabile h_2 la media a posteriori, che minimizza l'errore quadratico medio, mentre la multimodalità della densità a posteriori di s ci ha indotto a preferire la moda, come stimatore, alla media. La Tab. 3 e Fig. 2 raccolgono i risultati relativi al catalogo della zona z47.45-6 nel periodo dal 1005 al 1992. In questo caso il catalogo risulta completo dal 1702; se però si esclude dall'analisi il periodo dal 1005 al 1240 perchè, con un unico terremoto nel 1005, appare palesemente incompleto, l'analisi conduce a un risultato diverso dal precedente (Tab. 4 e Fig. 3) e concorde con quelli ottenuti per z47.51-2 e z47. C'è comunque da sottolineare che tutte le curve che compongono Fig. 3 sono bimodali, il che indica un sensibile incremento del tasso di accadimento dopo il 1700. Di fronte a questa incertezza l'esistenza di informazioni storiche secondo cui la quasi totalità degli eventi accaduti nel XVIII e XIX secolo sono stati registrati, potrebbe anche condurre alla conclusione che la completezza inizia nel 1700

Tab. 2 - Stime relative alla completezza del catalogo di z47.51-2 ($M_s \geq 4.4$).

	media	moda	deviazione standard				
s	1875.0	1878.1	4.0201				
h_2	0.607	0.598	0.0726				
<i>quantili</i>	$q_{0.025}$	$q_{0.05}$	$q_{0.25}$	$q_{0.5}$	$q_{0.75}$	$q_{0.95}$	$q_{0.975}$
s	1867.7	1869.2	1872.4	1876.0	1877.6	1880.7	1881.9
h_2	0.473	0.493	0.556	0.604	0.654	0.731	0.757
<i>intervallo .95HPD</i>	q_l	q_h					
h_2	0.467	0.751					

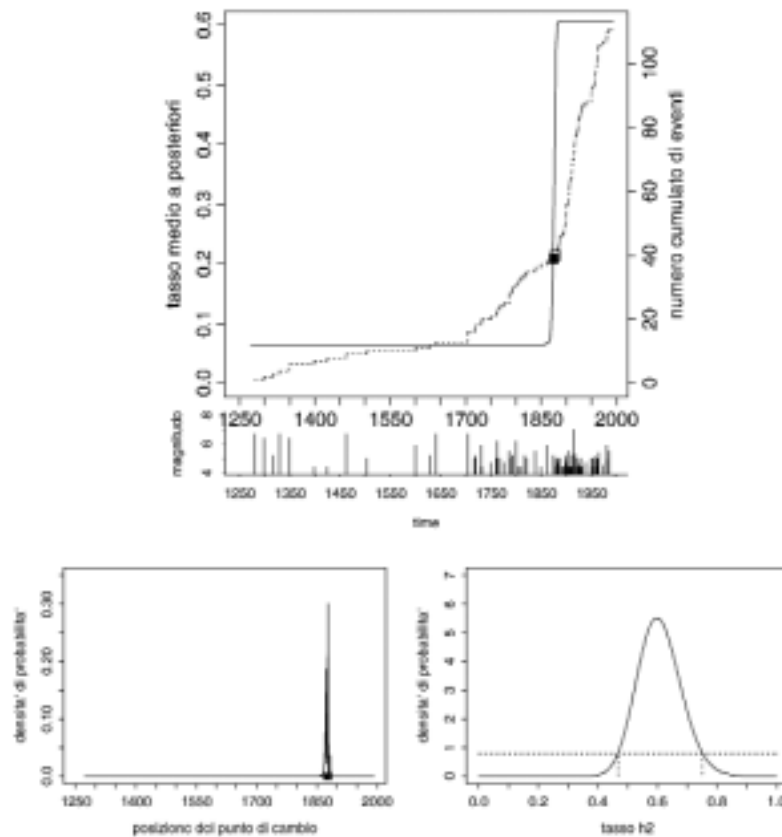


Fig. 1 - Rappresentazione grafica dei risultati sulla completezza del catalogo di $z_{47.51-2}$ ($M_s \geq 4.4$): tasso medio a posteriori (fig. superiore, linea continua), stima della densità di probabilità della posizione del punto di cambio (fig. inferiore, a sinistra), stima della densità di probabilità del tasso h_2 (fig. inferiore, a destra).

anzichè nel 1870, anche se la densità di probabilità $f(s)$ ha solo un massimo locale in questo punto.

10.3 Studio di variazioni della distribuzione di probabilità del tempo di intercorrenza

Il tempo che separa due eventi consecutivi è uno dei possibili indicatori del modello di accadimento; abbiamo perciò affrontato il problema dell'identificazione di diverse fasi sismiche assumendo che ogni fase sia caratterizzata da una diversa distribuzione di probabilità $F(t)$ del tempo di intercorrenza. Se (t_0, t_1, \dots, t_n) è la sequenza dei tempi di accadimento osservati in un'area, indichiamo con (y_1, \dots, y_n) , dove $y_i = t_i - t_{i-1}$, $i=1, \dots, n$, la sequenza dei tempi di intercorrenza. Supponiamo che siano state riconosciute $(k+1)$ fasi e che i k cambiamenti avvengano in corrispondenza di altrettanti eventi; cioè, se indichiamo con $(y_{r_1}, \dots, y_{r_k})$ il primo dei tempi di intercorrenza appartenenti a ciascuna delle

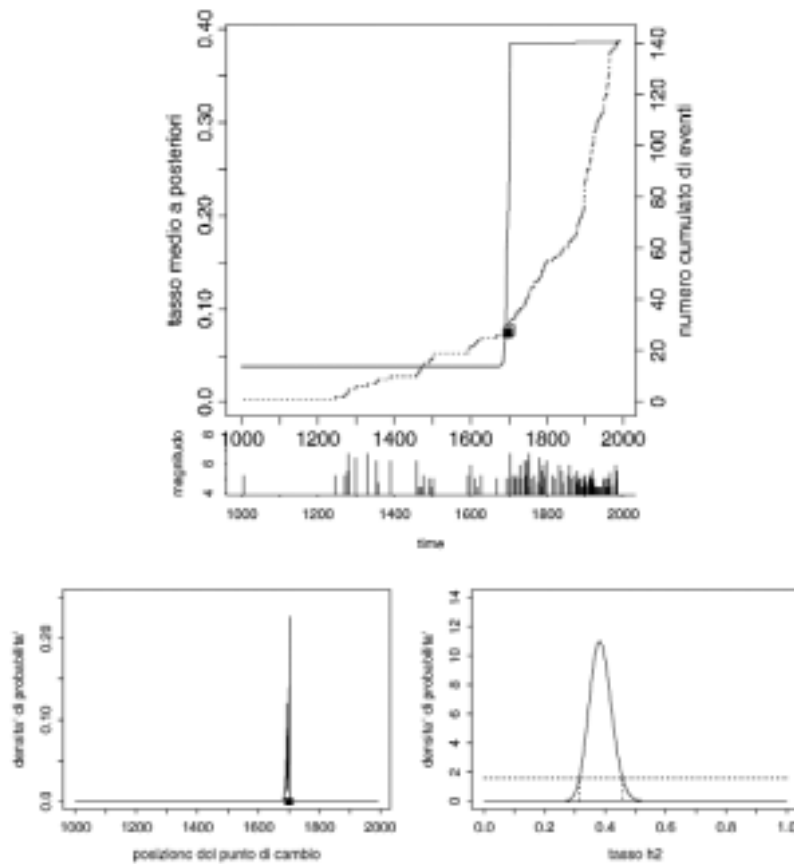


Fig. 2 - Rappresentazione grafica dei risultati sulla completezza del catalogo di z47.45-6 ($M_s \geq 4.4$), periodo 1005-1992: tasso medio a posteriori (fig. superiore, linea continua), stima della densità di probabilità della posizione del punto di cambio (fig. inferiore, a sinistra), stima della densità di probabilità del tasso h_2 (fig. inferiore, a destra).

k fasi, si ha che il primo cambiamento è accaduto in t_{r_1-1} , il secondo in t_{r_2-1} e così via. Indichiamo inoltre con F_0, F_1, \dots, F_k le distribuzioni di probabilità del tempo di intercorrenza che caratterizzano le varie fasi e, per garantire flessibilità al metodo, assumiamo che siano distribuzioni casuali; in particolare abbiamo supposto che ciascuna F_i segua una

Tab. 3 - Stime relative alla completezza del catalogo di z47.45-6 ($M_s \geq 4.4$), periodo 1005-1992.

	media	moda	deviazione standard				
s	1697.2	1702.7	6.81				
h_2	0.385	0.382	0.0360				
<i>quantili</i>	$q_{0.025}$	$q_{0.05}$	$q_{0.25}$	$q_{0.5}$	$q_{0.75}$	$q_{0.95}$	$q_{0.975}$
s	1686.3	1688.4	1693.2	1699.0	1701.5	1702.6	1702.7
h_2	0.317	0.327	0.360	0.384	0.409	0.447	0.460
<i>intervallo .95HPD</i>	q_l	q_h					
h_2	0.314	0.457					

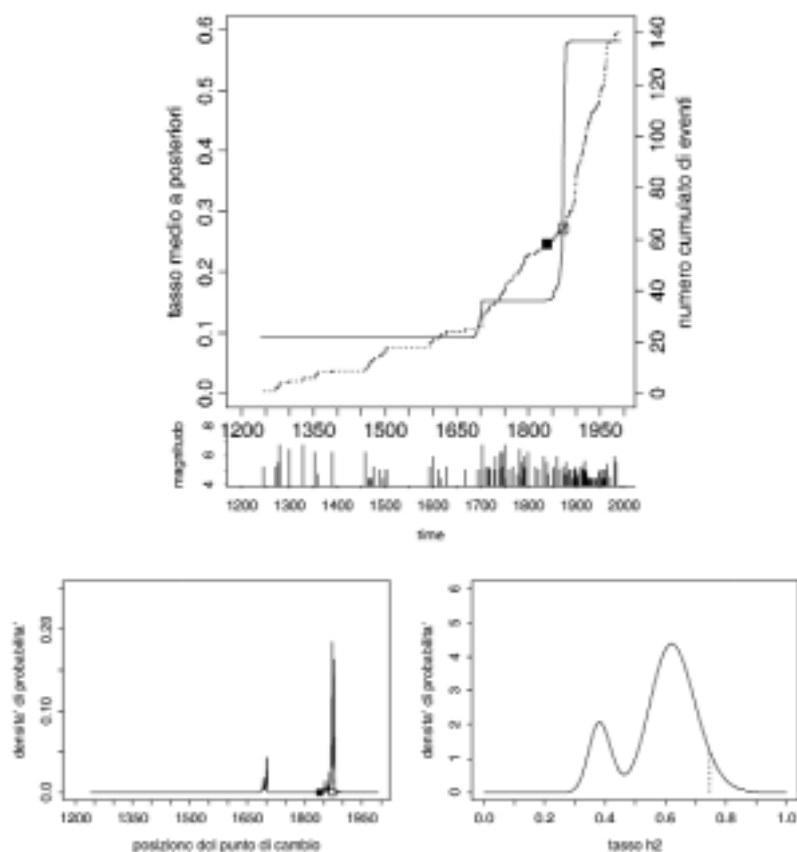


Fig. 3 - Rappresentazione grafica dei risultati sulla completezza del catalogo di z47.45-6 ($M_s \geq 4.4$), periodo 1240-1992: tasso medio a posteriori (fig. superiore, linea continua), stima della densità di probabilità della posizione del punto di cambio (fig. inferiore, a sinistra), stima della densità di probabilità del tasso h_2 (fig. inferiore, a destra).

mistura di processi di Dirichlet. Ricordiamo che F_i è un processo di Dirichlet se, data una partizione misurabile $\{B_1, \dots, B_m\}$ del suo supporto, si ha

$$(F_i(B_1) \leq u_{i1}, \dots, F_i(B_m) \leq u_{im}) \sim D(u_{i1}, \dots, u_{im}; MG(B_1; \theta_i), \dots, MG(B_m; \theta_i))$$

Tab. 4 - Stime relative alla completezza del catalogo di z47.45-6 ($M_s \geq 4.4$), periodo 1240-1992.

	media	moda	deviazione standard				
s	1839.2	1873.1	68.36				
h_2	0.583	0.620	0.0666				
<i>quantili</i>	$q_{0.025}$	$q_{0.05}$	$q_{0.25}$	$q_{0.5}$	$q_{0.75}$	$q_{0.95}$	$q_{0.975}$
s	1691.6	1693.8	1863.8	1872.3	1875.3	1877.9	1878.7
h_2	0.345	0.362	0.526	0.604	0.664	0.747	0.775
<i>intervallo .95HPD</i>	q_l	q_h					
h_2	0.	0.747					

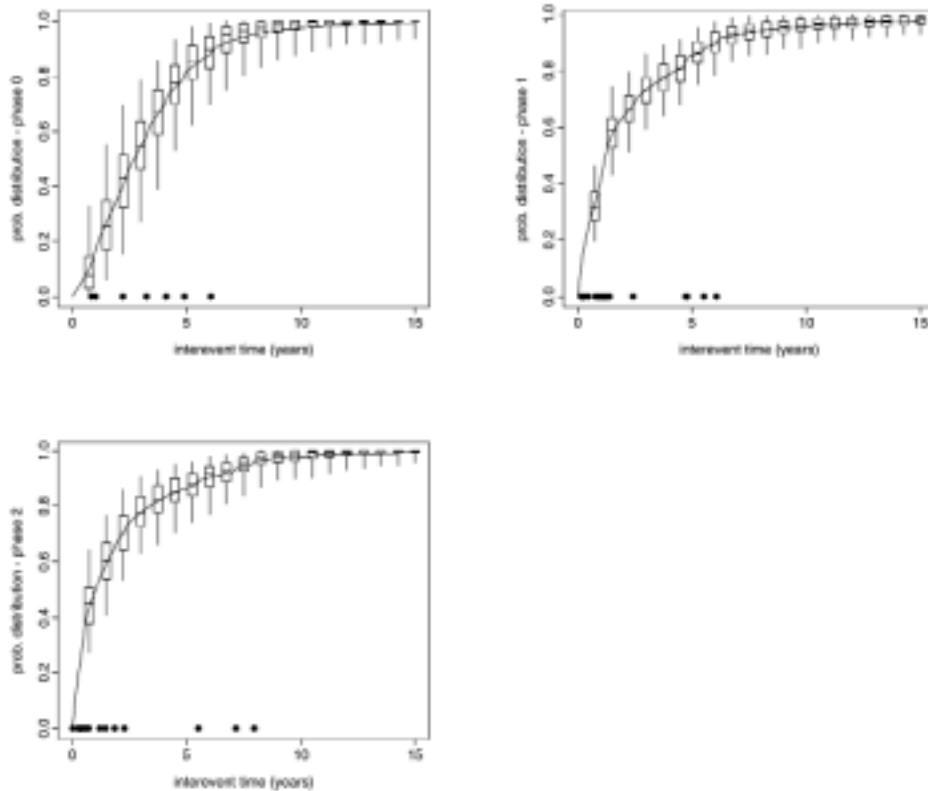


Fig. 4 - Distribuzione di probabilità del tempo di intercorrenza stimata in ciascuna delle tre fasi sismiche identificate nella zona 47 ($M_s \geq 4.4$); i punti corrispondono agli intertempi di ciascuna fase.

dove $D(u_{i1}, \dots, u_{im}; G_1, \dots, G_m)$ indica la distribuzione di Dirichlet di parametri (G_1, \dots, G_m) . Poichè la misura $G(\cdot)$ rappresenta il valore atteso iniziale di $F(\cdot)$, cioè esprime la nostra opinione sul presunto andamento della variabile temporale, abbiamo posto G uguale a una gamma generalizzata con densità

$$g(y; \eta, \xi, \rho) = \frac{\eta \xi^\rho y^{\rho-1} e^{-\xi y^\eta}}{\Gamma(\rho)} .$$

Questa famiglia di distribuzioni contiene infatti propriamente tutte quelle proposte in letteratura per il tempo di intercorrenza, ad es. a $\eta = 1$ corrisponde una gamma e a $\rho = 1$ una Weibull. Per una dettagliata descrizione del metodo rimandiamo a Pievatolo and Rotondi (1999); qui vogliamo solo sottolineare che la procedura seguita stima, oltre alle distribuzioni F_i , tutti i parametri del modello, cioè il numero k dei cambiamenti, la loro posizione data dal vettore degli indici \mathbf{r} , e per ogni fase il vettore di parametri $\theta_i = (\eta_i, \xi_i, \rho_i)$, $i=0, \dots, k$ della gamma generalizzata. Naturalmente la potenza del metodo comporta uno sforzo computazionale tale da richiedere l'applicazione di metodi di calcolo intensivo quali gli MCMC.

Come risultato collaterale dell'analisi si ottiene per ogni evento la probabilità che esso segni l'inizio di una nuova fase; i valori di tale probabilità che superano una soglia prefissata danno il numero k dei cambiamenti accaduti. In ciascuna delle tre zone esaminate, z47, z47.45-6, z47.51-2, sono state identificate tre fasi separate da due punti di cambio localizzati il primo tra il 1895 e il 1897 e il secondo tra il 1938 e il 1949. In Fig. 4 sono riportate le distribuzioni di probabilità stimate per ciascuna delle fasi identificate nella zona 47; in ciascun grafico i box-plots danno una misura della dispersione della stima.

10.4 Osservazioni

La discrepanza osservata nel grado di completezza dei vari cataloghi è spiegabile con l'elevata attività sismica registrata nel XVIII secolo nelle zone 47, 52 e non nella 45 e 46. Ciò induce a proporre un'altra zonazione dell'area in esame in senso longitudinale che riunisca in un'area le zone 46, 52 e parte della zona 47, e in un'altra la parte rimanente di z47 con le zone 45, 51 seguendo meglio, in tal modo, i sistemi di faglia recentemente individuati.

Bibliografia

- Antoniak C. E. (1974): Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *Annals of Statistics*, **2**, 1152-1174.
- Boncio P. e Lavecchia G. (1999): I terremoti di Colfiorito (Appennino umbro-marchigiano) del Settembre-Ottobre 1997: contesto tettonico e prime considerazioni sismogenetiche. *Boll. Soc. Geol. It.*, **118**.
- Camassi R. e Stucchi M. (1997): *NT4.1, un catalogo parametrico di terremoti di area italiana al di sopra della soglia di danno (versione NT4.1.1)*. GNDT, Milano, 66 + XXVII pp. e Internet: <http://emidius.itim.mi.cnr.it/NT/home.html>.
- Carlin B. P., Gelfand A. E. and Smith A. F. M. (1992): Hierarchical Bayesian analysis of change-point problems. *Applied Statistics*, **41**, 389-405.
- Gelfand A. E. and Smith A. F. M. (1990): Sampling based approaches to calculating marginal densities. *J. American Statistical Association*, **85**, 398-409.
- Meletti C., Patacca E., Scandone P. (2000): Construction of a seismotectonic model: the case of Italy. *Pageoph*, **157**, 11-35.
- Pievatolo A. and Rotondi R. (1999): Analysing the interevent time distribution to identify seismicity phases: a Bayesian nonparametric approach to the multiple change-point problem. *Applied Statistics*, in stampa.
- Rotondi R. e Garavaglia E. (1998): *Analisi statistica della completezza di un catalogo sismico*. Quaderno IAMI-98.16
- Scandone P. (1997): Linea di ricerca 2 "Sismotettonica". In: Corsanego A., Faccioli E., Gavarini C., Scandone P., Slejko D. e Stucchi M. (a cura di), *L'attività del GNDT nel triennio 1993 - 1995*, CNR - GNDT, Roma, 67-96.