Geophys. J. Int. (2006) doi: 10.1111/j.1365-246X.2006.03207.x

The waveform similarity approach to identify dependent events in instrumental seismic catalogues

S. Barani,[1] G. Ferretti,[1] M. Massa[2] and D. Spallarossa[1]

[1]*Dipartimento per lo Studio del Territorio e delle sue Risorse, Università di Genova, Viale Benedetto XV,* 5, 16132 *Genova, Italy.*
E-mail: barani@dipteris.unige.it
[2]*Istituto Nazionale di Geofisica e Vulcanologia, via Bassini* 15, 20133 *Milano, Italy*

Accepted 2006 September 15. Received 2006 September 7; in original form 2006 February 27

SUMMARY

In this paper, waveform similarity analysis is adapted and implemented in a declustering procedure to identify foreshocks and aftershocks, to obtain instrumental catalogues that are cleaned of dependent events and to perform an independent check of the results of traditional declustering techniques.

Unlike other traditional declustering methods (i.e. windowing techniques), the application of cross-correlation analysis allows definition of groups of dependent events (multiplets) characterized by similar location, fault mechanism and propagation pattern. In this way the chain of intervening related events is led by the seismogenetic features of earthquakes. Furthermore, a time-selection criterion is used to define time-independent seismic episodes eventually joined (on the basis of waveform similarity) into a single multiplet. The results, obtained applying our procedure to a test data set, show that the declustered catalogue is drawn by the Poisson distribution with a degree of confidence higher than using the Gardner and Knopoff method (1974). The declustered catalogues, applying these two approaches, are similar with respect to the frequency–magnitude distribution and the number of earthquakes.

Nevertheless, the application of our approach leads to declustered catalogues properly related to the seismotectonic background and the reology of the investigated area and the success of the procedure is ensured by the independence of the results on estimated location errors of the events collected in the raw catalogue.

Key words: declustering, dependent events, statistical methods, waveform analysis.

INTRODUCTION

The problem of identifying dependent events, such as aftershocks and foreshocks, has long been studied in seismology in order to obtain declustered catalogues that contain only main shocks and isolated earthquakes. The use of declustering methods is necessary for many seismological applications such as the quantification of foreshock occurrence probabilities (e.g. Savage & Depolo 1993) and analysis of seismicity patterns (e.g. Reasenberg 1985; Console *et al.* 2003). Moreover, declustering techniques are used in probabilistic seismic hazard analyses (PSHA), based on both conventional procedures (Cornell 1968) and smoothing approaches (Frankel 1995; Lapajne *et al.* 2003), in order to provide earthquake catalogues that are drawn from the Poisson distribution. In PSHA, the assumption of earthquakes occurring randomly, with ‘no memory’ of the time, size or location of any preceding events, allows the application of simple probability models (i.e. Cornell’s approach, Cornell 1968) that use a Poisson process to evaluate the probability of occurrence of earthquakes.

Several methods have been suggested in order to identify and remove dependent events from a seismic data set such as spatio-temporal windowing (e.g. Knopoff 1964; Gardner & Knopoff 1974; Keilis-Borok *et al.* 1982; Öncel & Alptekin 1999), interaction zone modelling (e.g. Reasenberg 1985; Reasenberg & Jones 1989) and multifractal analysis (e.g. Godano & Caruso 1995; Godano *et al.* 1999). In particular, Gardner & Knopoff (1974) have proposed a technique that consists of removing dependent events from a catalogue by specifying the spatial and temporal extent of aftershock sequences as a function of the magnitude of the main shock. To simplify and speed up this declustering process, the catalogues are often deprived of the presence of aftershocks and foreshocks using a non-dependent magnitude space–time window. This approach tends to overestimate the aftershock population since the spatial and temporal extent of aftershock sequences varies widely with respect to the magnitude of the main shock.

A more complicated approach to identify aftershocks was proposed by Reasenberg (1985). It assumes that an earthquake is part of a cluster if it falls within the interaction zone of a prior earthquake. The dimension of the interaction zone is a function of spatial and temporal parameters. The spatial extent is based on the stress distribution associated with any earthquake and it is estimated as the radius of a circular-crack (Kanamori & Anderson 1975) corresponding to the event's seismic moment. The temporal extent of the interaction zone is calculated using a probabilistic approach based on the time one should wait for the next earthquake in the cluster (look-ahead time). The application of this technique requires a careful definition of the dynamic source parameters of a region even if the subjective setting of some parameters (degree of confidence desired to be sure of observing the next event in a sequence, P; scaling parameter, c; stress drop value, $\Delta\sigma$; etc.) does not seem to have a great influence on the final results (Savage & Depolo 1993).

These methods are the most widely used because of their simple applicability and because they require catalogues that list just a few parameters such as origin time, location and magnitude. Thus they can be applied to both macroseismic and instrumental catalogues. However, the quality of instrumental data strongly differs from the quality of the historical ones and, as a consequence, it is reasonable to develop and apply different approaches to identify and remove dependent events from instrumental or macroseismic catalogues.

In this paper, we propose an alternative method to decluster instrumental catalogues applying the waveform similarity analysis. This approach, generally used for other kinds of data processing such as the cluster analysis (e.g. Maurer & Deichmann 1995; Cattaneo *et al.* 1999; Ferretti *et al.* 2005) and the automatic phase picking (e.g. Rowe *et al.* 2002), is opportunely modified and implemented in order to make it useful to employ in a declustering procedure. The waveform similarity analysis, allowing definition of groups of events characterized by similar location, fault mechanism and propagation pattern, is adapted to identify sequences and to remove dependent events (foreshocks and aftershocks).

In detail, the proposed procedure allows the identification of clusters of dependent events defined as groups of earthquakes that satisfy two requirements:

(1) Earthquakes must belong to the same seismogenetic source, as assessed by the waveform similarity analysis (spatial requirement).

(2) Earthquakes that meet the spatial requirement must occur within time windows probabilistically determined assuming Omori's law (temporal requirement).

Catalogues declustered by applying our procedure contain isolated earthquakes (not belonging to any cluster) and the highest magnitude event of each cluster. The application to an actual data set points out that our declustering method leads to a catalogue that fits the Poisson distribution.

METHOD

In the last 20 yr the waveform similarity analysis has been adopted by many authors to investigate the spatial and temporal evolution of foreshock—main shock—aftershock seismic sequences, swarms (e.g. Poupinet *et al.* 1984; Frechet 1985; Console & Di Giovanbattista 1987; Deichmann & Garcia-Fernandez 1992; Augliera *et al.* 1995; Dodge *et al.* 1995; Maurer & Deichmann 1995; Cattaneo *et al.* 1997; Shearer 1997; Cattaneo *et al.* 1999, Waldhauser *et al.* 1999; Shaff *et al.* 2002; Scarfı *et al.* 2003; Shaff *et al.* 2004; Ferretti *et al.* 2005; Massa *et al.* 2006) and volcanic activity (e.g. Got *et al.* 1994; Fremont & Malone 1987). In seismology the waveform similarity method represents a very powerful tool to identify and characterize the seismotectonic structures of an area and to study the propagation of seismic waves. In particular, this approach is largely used both to identify similar events (recorded by common stations) and to increase the precision in the computation of phase arrival time-shift between seismograms. Geller & Mueller (1980) found that earthquakes characterized by very similar seismograms (i.e. multiplets), called 'earthquake families' by Tsujiura (1983), are caused by the same source mechanism. Highly accurate locations can be obtained considering the results of the waveform similarity analysis (i.e. cross-correlation and time-shift values) and using 'master' (e.g. Cattaneo *et al.* 1999) and/or 'double difference' relative location methods (Waldhauser & Ellsworth 2000) in order to define the active seismic structures of a region with high precision.

In this work the performances and the potentiality of the waveform similarity analysis are used with an alternative aim. The detection of earthquake families or multiplets as sets of similar events is finalized to the identification of groups of dependent earthquakes clustered in space that are related to the same fault. In this context the detection of multiplets is made to identify distinct sources.

In order to define couples of similar events the normalized cross correlation function is applied (e.g. Augliera *et al.* 1995; Cattaneo *et al.* 1997, 1999):

$$C'_{12} = \frac{C_{12}(\tau)}{\sqrt{C_{11}(0)C_{22}(0)}} \tag{1}$$

the cross-correlation function being defined as:

$$C_{12}(\tau) = \int_{-\infty}^{+\infty} a_1(t)a_2(t+\tau)\,dt. \tag{2}$$

Since instrumental catalogues collect events localized in relatively wide areas (Northwestern Italy in our test application), the waveform similarity analysis is performed considering several stations which ensure an optimal coverage of the area under study. The use of more than one station allows us to subdivide the whole catalogue into different data sets collecting events with hypocentral distance from each station up to 100 km. In such a way we can cross-correlate recordings with an average signal to noise ratio greater than 10 dB, improving the reliability of the outcomes.

All waveforms are filtered by a bandpass filter from 1 to 12 Hz to reduce the bias of the noise and the high frequency wiggles, and some of the waveform differences between similar events due to differences in magnitude, focal mechanisms and small-scale heterogeneities. Then, a continuous signal of 6 seconds after the P onset is considered for each recording in order to perform cross-correlation analyses on signals including the first pulse (dependent on the focal mechanism and, mainly determined by the radiation pattern) and, for hypocentral distances smaller than about 40 km, the S waves (dependent on the focal mechanism) and the first part of the coda (mainly determined by propagation).

The detection of multiplets requires the definition of the minimum value of the cross-correlation coefficient, $C_{\min}$, beyond which events are assumed to belong to the same multiplet. For a given station, $C_{\min}$ value is selected by a trial and error procedure in order to ensure both the maximum number of families and the greatest number of events for each of them (e.g. Cattaneo *et al.* 1999; Ferretti *et al.* 2005; Massa *et al.* 2006). The bridging technique is used in order to overcome bias if events differ from each other by more than one order of magnitude (e.g. Deichmann & Garcia-Fernandez 1992; Ferretti *et al.* 2005). The bridging algorithm is based on the Equivalence Class approach (Press *et al.* 1988) and has already been applied to local earthquake data sets by Aster & Scott (1993), Cattaneo *et al.* (1997, 1999) and Ferretti *et al.* (2005). If two couples of events, (A, B) and (B, C) share a common quake (B) then all the events are attributed to the same family, even if the match between A and C is below the minimum cross-correlation value for similarity. Thus,

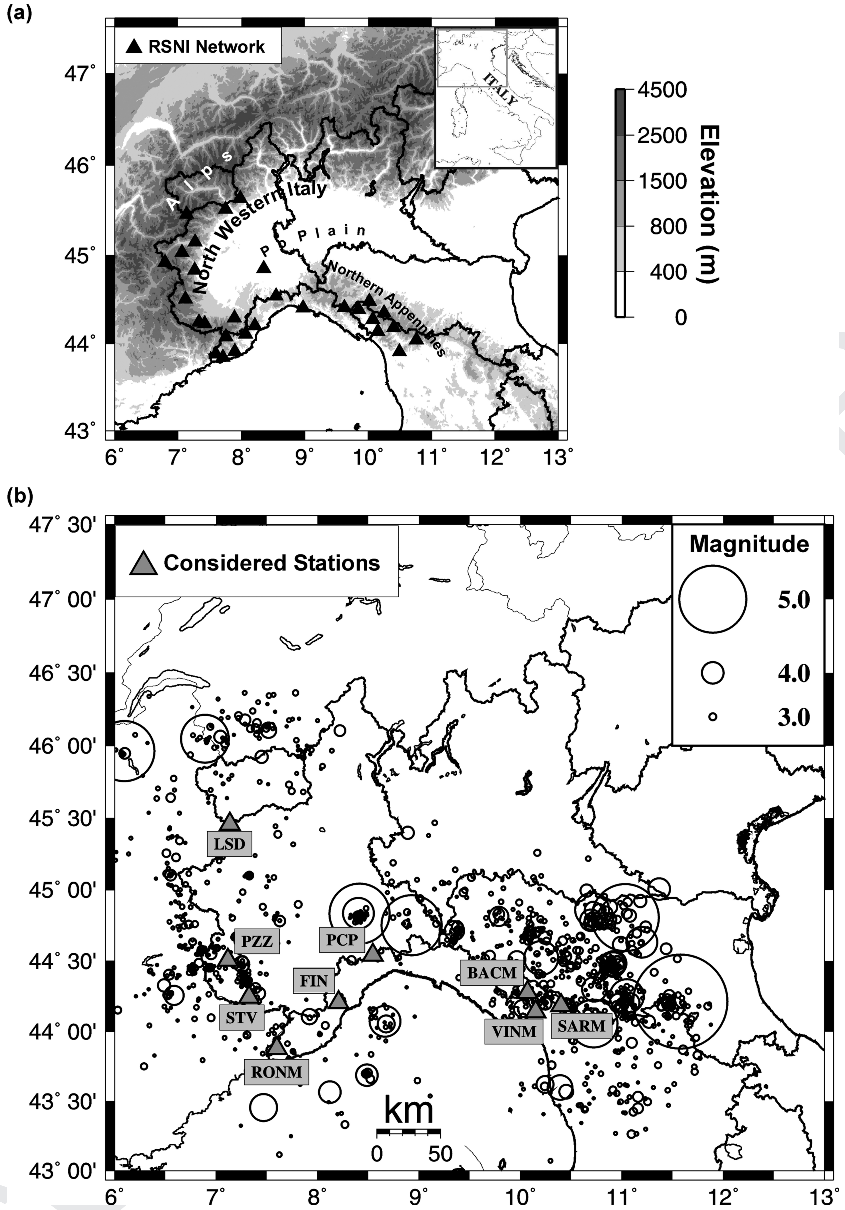

Figure 1. (a) Regional seismic network of Northwestern Italy (RSNI network, black triangles). (b) Epicentre map of the raw catalogue (black circles). The location of the station belonging to the RSNI network selected for the waveform similarity analysis is also reported (grey triangles).

event B represents the 'bridge event' between these couples. The potential and the success of the approach are estimated in Ferretti *et al.* (2005). However, if the magnitude difference between the main shock and the highest magnitude aftershock/foreshock is too wide (e.g. $\Delta M > 1.5$) and, consequently, the 'bridge event' may be missing, the bridging technique might fail to join to a same family aftershocks/foreshocks of strong events (e.g. $M \geq 4.5$) that might have different signal shapes with respect to the main shock. Therefore, the following conditional statement is implemented in the algorithm: if the i_{th} event in the data set has magnitude greater than a minimum value M_0 and $\Delta M = M_i - M_j$ (for $j = 1, \ldots,$ number of events in the data set) is greater than a reference value ΔM_0, then $C_{\min}$ (minimum value of the cross-correlation coefficient) is lowered to a value $C_0 < C_{\min}$. M_0, ΔM_0, and C_0 are tuned in order to allow the waveform similarity approach to identify each main sequence as a family. The main sequences used to fix the values of the above mentioned parameters and to test the procedure are taken out of the whole catalogue by a simple visual inspection (sets of events close in space and time). It is worth noting that C_0, even if it is used as minimum correlation threshold for stronger earthquakes only, has not to be set to too low a value (e.g. $C_0 > 0.60$) in order to avoid meaningless associations of events into families. Obviously, since C_0 is defined independently of $C_{\min}$, if $C_0 > C_{\min}$ then the conditional statement is not applied.

The proposed method associates events belonging to the same source (spatial requirement) that may be spaced in time by months or years (temporal multiplets as defined by Geller & Mueller 1980) to the same family. These multiplets represent repeated energy releases related to a single source reactivated in separate episodes (Cattaneo *et al.* 1999), and therefore, could collect time-independent seismic sequences. For this reason, a probabilistic temporal criterion is included in our declustering procedure to isolate time-independent

Table 1. Waveform similarity results. For each station, the number of available recordings, the selected minimum cross-correlation threshold ($C_{\min}$), the number of identified multiplets and the total number of earthquakes associated into multiplets are listed.

Station code	Recordings	$C_{\min}$ (per cent)	Multiplets	Components
FIN	378	70	21	168
LSD	331	70	13	112
PCP	345	70	15	110
PZZ	393	70	15	146
STV	387	70	22	157
RONM	216	65	12	77
BACM	453	75	19	146
SARM	586	70	27	254
VINM	468	70	22	160

seismic sequences or swarms joined into a single multiplet on the basis of the seismogram similarity. The identified families are checked by a temporal criterion based on the look-ahead time, T, already used in Reasenberg's declustering procedure (Reasenberg 1985)

$$\mathrm{T} = \frac{-\ln(1-P)t}{10^{c(\Delta-1)}}, \qquad (3)$$

where P is the degree of confidence of observing the next event, t is the time between the largest and the last event in the sequence, Δ is the magnitude difference between the largest event in the sequence and the minimum completeness threshold of the catalogue (Savage & Depolo 1993). The value of coefficient$c$ is empirically calibrated using the main sequences taken out of the whole catalogue by a visual inspection (Reasenberg 1985). This equation, based on Omori's law, determines the time interval necessary to wait in order to be P confident of observing the next event in the sequence.

Finally, each group of events, satisfying both the spatial (as derived from the waveform analysis) and the temporal requirements (time difference between two events belonging to the same family less than T), identifies a cluster of dependent events. From each sequence, a master event is selected as the highest magnitude earthquake. The master events and the earthquakes not associated with any of the identified families (isolated events) constitute the declustered catalogue.

APPLICATION IN NORTHWESTERN ITALY

The proposed declustering method is tested considering an instrumental catalogue of 1235 selected earthquakes recorded in Northwestern Italy by the RSNI network (Regional Seismic Network of Northwestern Italy) in the period 1996–2005 (Fig. 1). A preliminary data selection is carried out considering local magnitude values greater than 2.0. In the last ten years the seismicity that has occurred in Northwestern Italy is mainly characterized by a large percentage of low energy events, seldom characterized by M_l values greater than 3.0.

The main seismic sequences (afterwards used to tune the values of M_0, ΔM_0, and C_0, and the value of the c coefficient, Eq. 3) are localized in the Northern Apennines (e.g. 1999 July, with main shock of M_l 4.4; 2002 June, with main shock of M_l 4.1; 2003 September, with main shock of 5.3) and in the southwestern corner of the Po Plain (e.g. 2000 August, with main shock of M_l 4.9; 2001 July, with main shock of M_l 4.3; 2003 April, with main shock of M_l 4.9), as shown in Fig. 1b. The similarity analysis is performed taking into account the vertical component of the seismic signals recorded by nine RSNI seismic stations. These stations (Fig. 1b) are selected to ensure both an even coverage of the area (as a function of the seismicity distribution) and the completeness (the greatest number of recordings) and the quality of the data set (sets of waveforms characterized by the highest signal to noise ratio). In order to identify multiplets, all available recordings of the selected RSNI stations are compared in the search for waveform similarities by the application of the cross correlation technique joined to the performance of the bridging algorithm (Aster & Scott 1993). For each receiver a data set collecting events with hypocentral distance up to 100 km is selected and the minimum value of the correlation coefficient ($C_{\min}$) is set as explained in the previous paragraph. For all the stations, $M_0 = 4.3$, $1.9 \leq \Delta M_0 \leq 2.2$, and $C_0 = 0.66$. The chosen thresholds $C_{\min}$ and the number of families identified by the waveform similarity analysis for each station are reported in Table 1.

It is worth noting that many families are identified by more than one receiver, indicating highly reliable multiplets. This 'multiple' definition of a family could represent an important constraint to define the number of events belonging to each multiplet (Ferretti *et al.* 2005). Nevertheless, the detection of a family at different stations is a sufficient but not a necessary condition because several factors might degrade the similarity (station failures, instrumental limitation, occurrence of strong, incoherent noise and superposition of different events very closely spaced in time) (Cattaneo *et al.* 1997). As an example, Fig. 2 shows some seismograms relative to a multiplet that occurred in the Northern Apennines identified by the waveform similarity at VINM station.

By application of the waveform similarity analysis a data set collecting all the identified families is obtained for each station. As stated in the previous paragraph, a temporal criterion based on the look-ahead time (Reasenberg 1985) is applied to each data set to discriminate in time seismic sequences previously joined (on the basis of waveform similarity) into a single multiplet. For this purpose, eq. (3) is used with the coefficient c equal to 0.47 and P equal to 0.95. In Fig. 3, the capability of our method to detect spatial and temporal dependent events is checked by analysing the 2000–2001 Monferrato sequences (Massa *et al.* 2006). This example shows the capability of the proposed procedure in detecting two separate seismic episodes generated by the same seismogenetic source activated in 2000 August and re-activated after about 8 months.

Nine declustered data sets, one for each seismic station, are obtained by removing aftershocks and foreshocks and collecting both the master events and the isolated ones.

As a result, a declustered catalogue is obtained selecting and merging master and isolated events that are common to each of the nine data sets. The final declustered catalogue includes 799 earthquakes.

STATISTICS AND INFLUENCE ON SEISMIC RECURRENCE PARAMETERS

The accuracy and the reliability of the proposed procedure (called *CrossCorr* hereinafter) are assessed evaluating the Poissonian characteristics of the declustered catalogue and comparing our method with other traditional ones.

To test whether the catalogue collects independent events we have to answer the question: can we disprove, to a certain level of significance, the null hypothesis that the declustered data set is drawn from the Poisson distribution? Failing to disprove the null hypothesis shows that the data set can be consistent with the expected model. The accepted tests for differences between an observed, binned

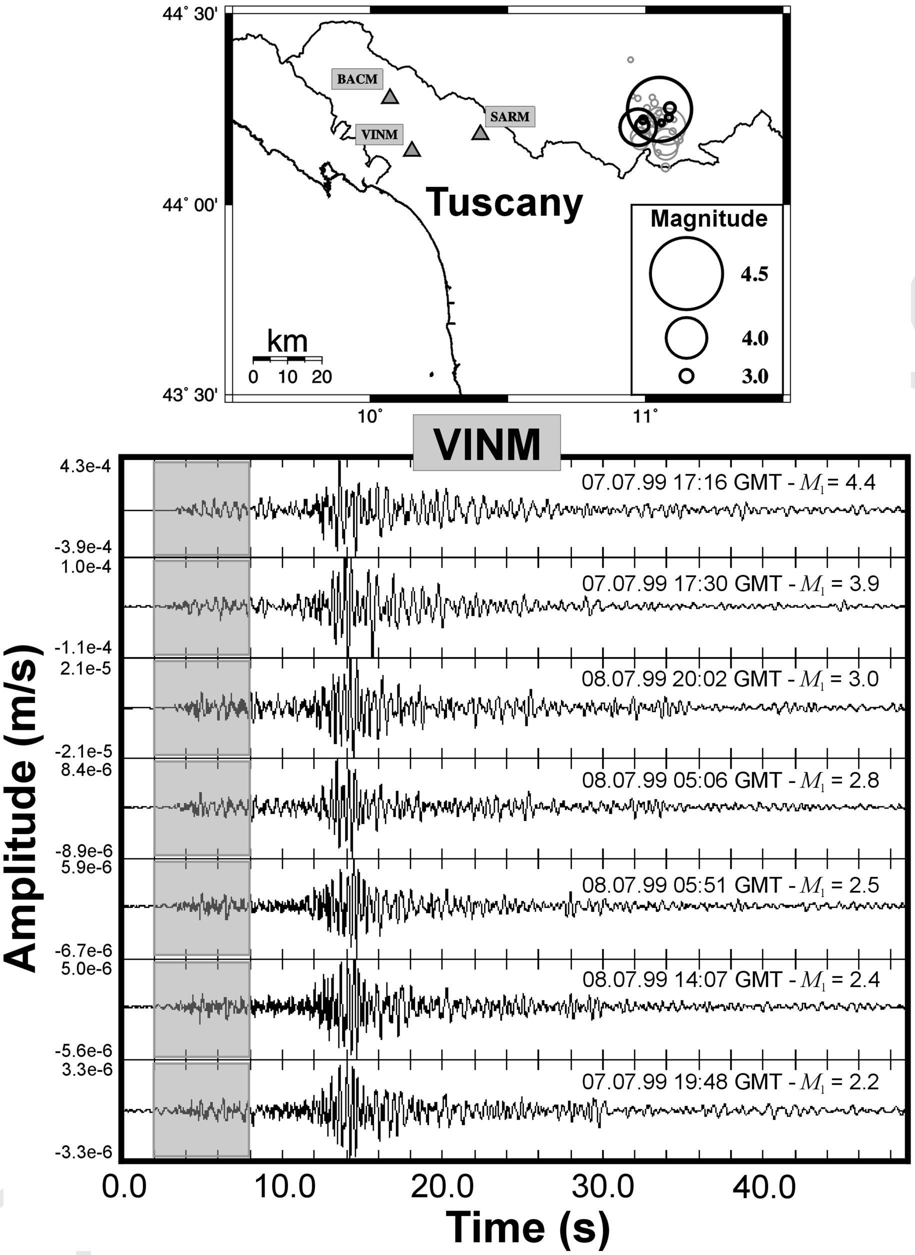

Figure 2. Top panel: example of a group of events (grey and black circles) with magnitude M_l spanning from 2.0 to 4.4 attributed to the same family via the waveform similarity technique considering VINM station with a minimum cross-correlation threshold of 70 per cent. Bottom panel: shape of the waveforms of the events indicated by black circles (top panel); the shaded area indicates the cross-correlation window (6 s).

distribution (the seismic catalogue) and an expected one (the Poisson distribution) are the *chi-square test* and the *reduced chi-square* test. The chi-square statistic is:

$$\chi^2 = \sum_k \frac{(O_k - E_k)}{E_k}, \tag{4}$$

where O_k is the number of discrete intervals with k events and E_k is the number envisaged according to the expected distribution. All the terms with $O_k = E_k = 0$ are omitted from the sum. A low value of χ^2 indicates that our hypothesis is quite likely. In order to define the χ^2 significance, the probability that the chi-square (χ^2) should exceed the observed value (χ_o^2) is computed (Press *et al.* 1988):

$$Q(\chi^2 \geq \chi_o^2) \equiv Q\left(\frac{\upsilon}{2}; \frac{\chi^2}{2}\right) \equiv \frac{\Gamma(\chi^2, \upsilon)}{\Gamma(\chi^2)} \equiv \frac{1}{\Gamma(\chi^2)} \int_\upsilon^\infty e^{-t} t^{\chi^2 - 1}\, dt \qquad (\chi^2 > 0), \tag{5}$$

where Γ is the gamma function and υ is the number of degrees of freedom. Any models with $Q > 0.001$ (Press *et al.* 1988) are considered acceptable. Q values $\cong 1$ indicate that the data set is consistent with the model.

The reduced chi-square, $\tilde{\chi}^2 = \chi^2/\upsilon$, and its probability, $P(\tilde{\chi}^2 \geq \tilde{\chi}_o^2)$, are also computed. The $\tilde{\chi}_o^2$ significance is provided taking into account a 5 per cent confidence threshold, as suggested by

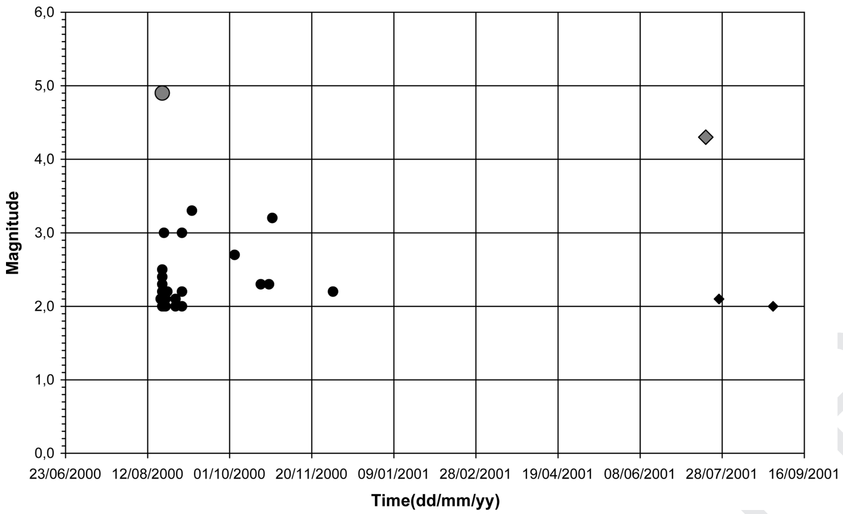

Figure 3. Temporal evolution of the Monferrato sequence: the waveform similarity analysis joined with the adopted temporal criterion allows the distinguishing of two sequences (black circles and black diamonds). Grey symbols indicate the two master events.

Table 2. Statistics: the degrees of freedom (*DF*), the observed chi-square (χ_o^2) and its probability (*Q*), the observed reduced chi-square ($\tilde{\chi}_o^2$) and its probability (*P*) are listed for each declustering procedure (*FW*, *GK* and *CrossCorr*) taking into account four different bin amplitudes (15, 20, 25 and 30 days).

Bin amplitude (days)	*FW* (Number of events: 471)					*GK* (Number of events: 830)					*CrossCorr* (Number of events: 799)				
	DF	χ_o^2	*Q*	$\tilde{\chi}_o^2$	*P*	*DF*	χ_o^2	*Q*	$\tilde{\chi}_o^2$	*P*	*DF*	χ_o^2	*Q*	$\tilde{\chi}_o^2$	*P*
15	6	13.24	0.04	2.21	≈ 11 per cent	6	12.88	0.05	2.15	≈ 5 per cent	10	9.46	0.49	0.95	≈44 per cent
20	6	6.26	0.40	1.04	≈ 42 per cent	7	13.17	0.07	1.88	≈ 7 per cent	7	10.92	0.14	1.56	≈13 per cent
25	8	1.37	0.99	0.17	≈ 99 per cent	13	20.12	0.09	1.55	≈ 8 per cent	9	9.28	0.41	1.03	≈44 per cent
30	7	8.28	0.31	1.18	≈ 30 per cent	7	15.84	0.03	2.26	≈ 3 per cent	9	13.04	0.16	1.45	≈18 per cent

Taylor (1986). If $P(\tilde{\chi}^2 \geq \tilde{\chi}_o^2) < 5$ per cent, the data set is not drawn from the expected distribution.

In order to compare C*rossCorr* with traditional declustering techniques, our raw catalogue (Fig. 1b) is cleaned of dependent events by using the windowing method proposed by Gardner & Knopoff (1974) (called *GK* hereinafter) and a fixed windowing approach (called *FW* hereinafter). The *GK* procedure identifies seismic sequences within magnitude-dependent space–time windows:

$$(t_a - t_m) < T_i,\ |g_a - g_m| < D_i,\ M_a < M_m, \tag{6}$$

where each tern t_a, g_a, M_a and t_m, g_m, M_m represents time, epicentre location and magnitude of aftershock and main shock, respectively. D_i and T_i are the spatial and temporal spans between a main shock and its aftershocks. The *GK* method is applied to the test data set using the D_i and T_i values proposed by Gardner & Knopoff (1974) for Southern California. The application of such values to the Northwestern Italy data set is proved to be suitable for the identification of the main sequences (main shock with $M_l \geq 3.4$) in the catalogue. However, a more accurate definition of D_i and T_i values would require a larger number of sequences (characterized by main shocks of different magnitude) than those in our data set and, besides, it is not the purpose of this work.

The *FW* approach is based on a magnitude-independent space–time windowing technique and consists of keeping the largest magnitude from each 30-km—90-day dimensioned window (Gruppo di lavoro CPTI 2004).

We divide the declustered catalogues into discrete temporal intervals (bins), and count the number of bins with *k* independent events; four different bin amplitudes (*Amp*: 15, 20, 25 and 30 days) are considered in order to show possible fluctuation of the values of the test-parameters that are listed in Table 2 ($P(\tilde{\chi}^2 \geq \tilde{\chi}_o^2)$ values are taken from Taylor 1986).

CrossCorr and *FW* techniques provide declustered data sets that agree with the assumption of a Poisson distribution. In fact both the $Q(\chi^2 \geq \chi_o^2)$ and $P(\tilde{\chi}^2 \geq \tilde{\chi}_o^2)$ probabilities are greater than the confidence thresholds of 0.001 and 5 per cent, respectively. Also the declustered catalogue provided by *GK* fit the Poisson distribution but with a lower value of *Q* and *P*. This result may be biased by a not optimal calibration of D_i and T_i.

In Fig. 4 the cumulative frequency-magnitude distributions and the Gutenberg–Richter curves for each of the three declustered catalogues are shown.

The frequency-magnitude distributions change with respect to the applied declustering method with the exception of the largest earthquakes ($M_l \geq 3.5$). The *GK* and the *CrossCorr* methods provide quite similar declustered data sets (considering the frequency-magnitude distribution and the number of earthquakes), reducing the number of events from 1235 to 830 and to 799, respectively. The *FW* approach greatly reduces the number of earthquakes from 1235 to 471. This is due to the spatial and temporal extent of the fixed window that seems too large for low energy earthquakes ($M_l < 3.5$). The Gutenberg–Richter curves, defined by the recurrence parameters (or frequency-magnitude parameters) *a* (10^a is the mean number of the earthquakes of magnitude greater than or equal to zero) and *b* (negative slope of the Gutenberg–Richter curve) whose values and uncertainties (standard deviation), are listed in Table 3, are coincident when using the *GK* and the *CrossCorr* approaches. *FW*, instead, provides lower values of the recurrence parameters. Obviously, all the declustered catalogues are characterized by lower values of *a* and *b* than the raw catalogue because of the strong reduction of events with magnitudes $M_l < 3.5$.

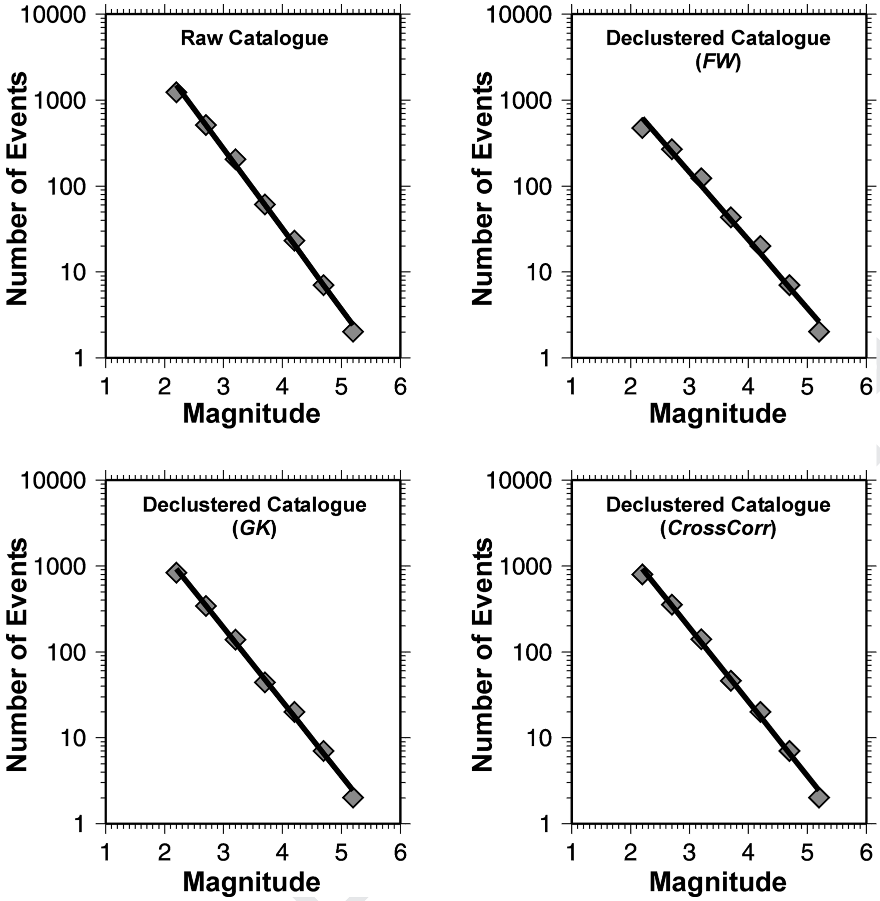

Figure 4. Cumulative frequency-magnitude distributions for the raw catalogue and the three declustered ones using the *FW*, *GK*, and *CrossCorr* procedures. The Gutenberg–Richter curves (black lines) are superimposed.

Table 3. Earthquake recurrence parameters (a; b) computed for the raw earthquake catalogue and for the declustered ones. a-values and b-values are estimated by using the least mean square method.

Method	Number of events	a-value ± standard deviation	b-value ± standard deviation
Raw catalogue	1235	5.22 ± 0.09	0.93 ± 0.02
CrossCorr	799	4.86 ± 0.09	0.86 ± 0.02
GK	830	4.86 ± 0.08	0.86 ± 0.02
FW	471	4.5 ± 0.1	0.79 ± 0.04

In Fig. 5, a comparison between the declustered catalogue provided by *CrossCorr* and those by *GK* (Fig. 5a) and *FW* (Fig. 5b) is shown by plotting the epicentre maps. For each comparison one map with the common independent events and a separate map with the independent events identified by the two methods separately are shown.

The largest difference between the catalogues declustered by *CrossCorr* and *GK* is shown in the Northern Apennines, mainly in the eastern sector where the coverage of the seismic network might be defective (Fig. 1). Moreover, *GK* identifies a greater number of independent events than *CrossCorr* near the Italian–French and the Italian–Swiss borders. It is clear that *FW* strongly reduces the number of independent events everywhere.

CONCLUSIONS

This paper presents an alternative procedure to identify dependent events based on the waveform similarity approach. The method is applied to identify and remove foreshocks and aftershocks from a test instrumental data set including 1235 earthquakes recorded by the RSNI seismic network from 1996 to 2005 ($M_l > 2.0$). The performances of the method and the reliability of the results are evaluated by applying several statistic tests and by comparison with other traditional windowing methodologies.

The declustered catalogue, collecting 799 independent earthquakes, fits the Poisson distribution with a degree of confidence greater than 5 per cent. The proposed declustering approach could allow the overcoming of the overestimation of aftershock and foreshock populations generally related to the application of windowing techniques (as stated by Reasenberg 1985). For example, in our test, the *FW* method (magnitude-independent space–time windowing procedure) mostly removes low magnitude events (smaller than 3.5), and, thus, it should be correctly used for declustering catalogues that collect earthquakes greater than a threshold magnitude $M_l = 3.5$. The *CrossCorr* method, instead, provides Poissonian catalogues without wrongly removing independent, low magnitude

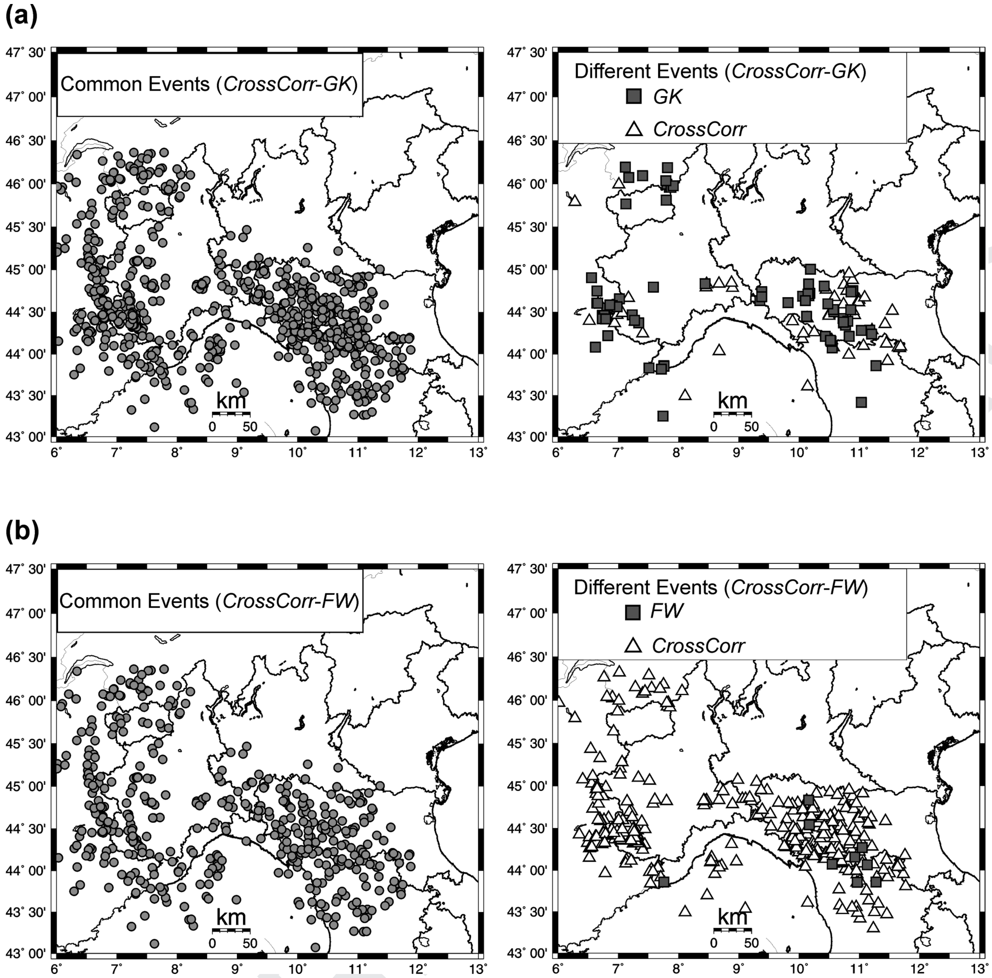

Figure 5. (a) Comparison between catalogues declustered by *GK* and *CrossCorr*; left panel: epicentre map with common independent events (grey circles); right panel: epicentre map with the independent events identified by *CrossCorr* (white triangles) and *GK* (grey squares) separately. (b) Comparison between catalogues declustered by *FW* and *CrossCorr*; left panel: epicentre map with common independent events (grey circles); right panel: epicentre map with the independent events identified by *CrossCorr* (white triangles) and *FW* (grey squares) separately.

events. Therefore, it can be useful for quantification of foreshock and aftershock occurrence probabilities, analysis of seismicity patterns, and hazard analyses of areas characterized by low magnitude seismicity. It is worth noting that declustered catalogues including also low energy earthquakes could allow a better and more accurate evaluation of seismic recurrence parameters, whose values could be affected by the incompleteness of medium—high energy events.

Moreover, the *CrossCorr* approach allows identification of aftershock and foreshock populations independently of the location errors associated with each recorded earthquake. Indeed the identification of seismic sequences is related to the waveform similarity and it does not depend on the accuracy of the epicentral coordinates. If the window spatial extent is smaller than location errors, the windowing techniques (such as *GK* and *FW* methods) could provide unreliable declustered catalogues; on the contrary the *CrossCorr* approach leads to chains of intervening related events independently of location parameters and dependently on the seismogenetic features of earthquakes. In fact, as shown by the comparison between the epicentre maps of the declustered catalogues (Fig. 5), the number of independent events, identified by the *CrossCorr* method and the windowing ones (above all *GK*) separately, strongly differs in areas where the accuracy of locations might be biased by the poor coverage of the seismic network.

We make no claim to be presenting a substitutive declustering procedure better than traditional ones; the results obtained for the 1996–2005 Northwestern Italy data set are very encouraging but further research is necessary to test the reliability of the proposed technique taking into account several data sets enclosing wider magnitude ranges.

ACKNOWLEDGMENTS

This work was substantially improved by the suggestions of two reviewers and the editor. We are particularly indebted to Dr N. Deichmann for his helpful and constructive comments. Special thanks are due to Prof. C. Eva and the RSNI staff. The authors are also very grateful for the stimulating discussions with Dr D. Bindi and Dr P. Augliera.

REFERENCES

Aster, R.C. & Scott, J., 1993. Comprehensive characterization of waveform similarity in microearthquake datasets, *Bull. Soc. Am.,* **83,** 1307–1314.

Augliera, P., Cattaneo, M. & Eva, C., 1995. Seismic multiplet analysis and its implication in seismotectonics, *Tectonophysics,* **248,** 219–234.

Cattaneo, M., Augliera, P., Spallarossa, D. & Eva, C., 1997. Reconstruction of a seismogenetic structures by multiplet analysis: an example of Western Liguria, Italy, *Bull. seism. Soc. Am.,* **87,** 971–986.

Cattaneo, M., Augliera, P., Spallarossa, D. & Lanza, V., 1999. A waveform similarity approach to investigate seismicity patterns, *Natural Hazards,* **19,** 123–138.

Console, R. & Di Giovanbattista, R., 1987. Local earthquake relative locations by digital record, *Phys. Earth planet. Inter.,* **47,** 43–49.

Console, R., Lombardi, A.M. & Murru, M., 2003. Bath's law and self-similarity of earthquakes, *J. geophys. Res.,* **108**(B2), 2128.

Cornell, C.A., 1968. Engineering seismic risk analysis, *Bull. seism. Soc. Am.,* **58,** 1583–1606.

Deichmann, N. & Garcia-Fernandez, M., 1992. Rupture geometry from high-precision relative hypocenter locations of micro-earthquake clusters, *Geophys. J. Int.,* **110,** 501–517.

Dodge, D., Beroza, G.C. & Ellsworth, W.L., 1995. Evolution of the 1992 Landers, California, foreshock sequence and its implications for earthquake nucleation, *J. geophys. Res.,* **100,** 9865–9880.

Ferretti, G., Massa, M. & Solarino, S., 2005. An improved method for the identification of seismic families: application to the Garfagnana-Lunigiana area (Italy), *Bull. seism. Soc. Am.,* **95**(5), 1903–1915.

Frankel, A., 1995. Mapping seismic hazard in the Central and Eastern United States, *Seism. Res. Lett.,* **66**(4), 8–21.

Frechet, J., 1985. Sismogenese et doublets sismiques, These d'Etat, Universite Scientifique et Medicale de Grenoble, 206.

Fremont, M.J. & Malone, D., 1987. High precision relative locations of earthquakes at Mount St. Helens, Washington, *J. geophys. Res.,* **92,** 10 223–10 236.

Gardener, J.K. & Knopoff, L., 1974. Is the sequence of earthquakes in southern California, with aftershocks removed, Poissonian?, *Bull. seism. Soc. Am.,* **64,** 1363–1367. Q1

Geller, R.J. & Mueller, C.S., 1980. Four similar earthquakes in central California, *Geophys. Res. Lett.,* **7,** 821–824.

Godano, C. & Caruso, V., 1995. Multifractal analysis of earthquake catalogues, *Geophys. J. Int.,* **121,** 385–392.

Godano, C., Tosi, P., Derubeis, V. & Augliera, P., 1999. Scaling properties of the spatio-temporal distribution of earthquakes: a multifractal approach applied to a Californian catalogue, *Geophys. J. Int.,* **136,** 99–108.

Got, J., Fréchet, J. & Klein, F.W., 1994. Deep fault plane geometry inferred from multiplet relative relocation beneath the south flank of Kilauea, *J. geophys. Res.,* **99,** 15 375–15 386.

Gruppo di lavoro CPTI., 2004. Catalogo Parametrico dei Terremoti Italiani, versione 2004 (CPTI04). INGV, Bologna. http://emidius.mi.ingv.it/CPTI/

Kanamori, H. & Anderson, D.L., 1975. Theoretical basis of some empirical relations in seismology, *Bull. seism. Soc. Am.,* **65,** 1073–1095.

Keilis-Borok, V.I., Knopoff, L. & Rotwain, I.M., 1982. Burst of aftershocks, long term precursors of strong earthquakes, *Nature,* **283,** 259–263 pp.

Knopoff, L., 1964. Statistics of earthquakes in Southern California, *Bull. seism. Soc. Am.,* **54,** 1871–1873.

Lapajne, J., Šket Motnikar, B. & Zupančič, P. , 2003. Probabilistic seismic hazard assessment methodology for distributed seismicity, *Bull. seism. Soc. Am.,* **93,** 2502–2515.

Massa, M., Eva, E., Spallarossa, D. & Eva, C., 2006. Detection of earthquake clusters on the basis of waveform similarity: an application in the Monferrato Region (Piedmont, Italy), *J. Seismol.,* **10,** 1–22.

Maurer, H. & Deichmann, N., 1995. Micro-earthquake cluster detection based on waveform similarities, with an application to the western Swiss Alps, *Gephys. J. Int.,* **123,** 588–600.

Öncel, A.O. & Alptekin, Ö., 1999. Effects of aftershocks on earthquake hazard estimation: an example from the North Anatolian fault zone, *Natural Hazards,* **19,** 1–11.

Press, W.H., Flannery, P.B., Teukolsky, S.A. & Wetterling, W.T., 1988. *Numerical Recipes in C, The Art of Scientific Computing*, Cambridge University Press, Cambridge, UK.

Poupinet, G., Ellsworth, W.L. & Fréchet, J., 1984. Monitoring velocity variations in the crust using earthquake doublets: an application to the Calaverals Fault, California, *J. geophys. Res.,* **89,** 5719–5731.

Reasenberg, P., 1985. Second-order moment of Central California seismicity, 1969–1982, *J. geophys. Res.,* **90,** 5478–5495.

Reasenberg, P.A. & Jones, L.M., 1989. Earthquake Hazard after a main shock in California, *Science,* **243,** 1173–1176.

Rowe, C.A., Aster, R.C., Borchers, B. & Young, C.J., 2002. An automatic, adaptive algorithm for refining phase picks in large seismic datasets, *Bull. seism. Soc. Am.,* **92,** 1660–1674.

Savage, M.K. & Depolo, D.M., 1993. Foreshock probabilities in the Western Great-Basin Eastern Sierra Nevada, *Bull. seism. Soc. Am.,* **83,** 1910–1938.

Scarfı, L., Langer, H. & Gresta, S., 2003. High-precision relative locations of two micro-earthquake clusters in South-eastern Sicily, Italy, *Bull. seism. Soc. Am.,* **93,** 1479–1497.

Shaff, D.P., Bokelmann, G.H.R., Beroza, G.C., Waldhauser, F. & Ellsworth, W.L., 2002. High resolution image of Calaverals Fault seismicity, *J. geophys. Res.,* **107**(B9), 2186, doi:10.1029/2001JB000633.

Shaff, D.P., Bokelmann, G.H.R., Ellsworth, W.L., Zanzerkia, E., Waldhauser, F. & Beroza, G.C., 2004. Optimizing correlation technique for improved earthquake location, *Bull. seism. Soc. Am.,* **94,**2, 705–721. Q2

Shearer, P.M., 1997. Improving local earthquake locations using the L1 norm and waveform cross correlation: application to the Whittier Narrows, California, aftershock sequence, *J. geophys. Res.,* **102,** 8269–8283.

Solarino, S., Ferretti, G. & Eva, C., 2002. Seismicity of Garfagnana-Lunigiana (Tuscany, Italy) as recorded by a network of semi-broad-band instruments, *J. Seismol.,* **6,** 145–152.

Taylor, J.R., 1986. Introduzione all'analisi degli errori, Ed. Zanichelli. Q3

Tsujiura, M., 1983. Characteristic frequencies for earthquake families and their tectonic implications: evidence from earthquake swarms in the Kanto District, Japan, *Pure appl. Geophys.,* **121,** 574–600.

Waldhauser, F. & Ellsworth, W., 2000. A double-difference Earthquake location Algorithm: method and application to the northern Hayward fault, California, *Bull. seism. Soc. Am.,* **90,** 1353–1368.

Waldhauser, F., Ellsworth, W.L. & Cole, A., 1999. Slip-parallel seismic lineations along the northern Hayward fault, California, *Geophys. Res. Lett.,* **26,** 3525–3528.

QUERIES

Journal: GJI
Paper: gji3207

Dear Author

During the copy-editing of your paper, the following queries arose. Please respond to these by marking up your proofs with the necessary changes/additions. Please write your answers on the query sheet if there is insufficient space on the page proofs. Please write clearly and follow the conventions shown on the corrections sheet. If returning the proof by fax do not write too close to the paper's edge. Please remember that illegible mark-ups may delay publication.

Query Reference	**Query**	**Remarks**
Q1	Au: Please check spelling of author Gardener for correctness in Reference Gardner &Knopoff (1974).	
Q2	Au: Please check the spelling of author Shaff for correctness in Reference Shaff *et al.* (2004).	
Q3	Au: Please provide complete information in Reference Taylor (1986).	

MARKED PROOF

Please correct and return this set

Please use the proof correction marks shown below for all alterations and corrections. If you wish to return your proof by fax you should ensure that all amendments are written clearly in dark ink and are made well within the page margins.

Instruction to printer	***Textual mark***	***Marginal mark***
Leave unchanged	• • • under matter to remain	Ⓥ
Insert in text the matter indicated in the margin	‸	New matter followed by ‸ or ‸ⓧ2
Delete	/ through single character, rule or underline or ⊢⊣ through all characters to be deleted	δ or δⓧ2
Substitute character or substitute part of one or more word(s)	/ through letter or ⊢⊣ through characters	new character / or new characters /
Change to italics	— under matter to be changed	⎵
Change to capitals	≡ under matter to be changed	≡
Change to small capitals	═ under matter to be changed	═
Change to bold type	〰 under matter to be changed	〰
Change to bold italic	≂ under matter to be changed	≂
Change to lower case	Encircle matter to be changed	≢
Change italic to upright type	(As above)	+
Change bold to non-bold type	(As above)	ǂ
Insert 'superior' character	/ through character or ‸ where required	ɣ or ⅄ under character e.g. ²ɣ or ²⅄
Insert 'inferior' character	(As above)	‸ over character e.g. ‸₂
Insert full stop	(As above)	⊙
Insert comma	(As above)	,
Insert single quotation marks	(As above)	‘ɣ or ‘⅄ and/or ’ɣ or ’⅄
Insert double quotation marks	(As above)	“ɣ or “⅄ and/or ”ɣ or ”⅄
Insert hyphen	(As above)	⊢-⊣
Start new paragraph	_⌐	_⌐
No new paragraph	↩	↩
Transpose	⊔⊓	⊔⊓
Close up	linking ⁀‿ characters	⁀‿
Insert or substitute space between characters or words	/ through character or ‸ where required	Y
Reduce space between characters or words	| between characters or words affected	⊤