Are different $M = a + bI_0$ relationships due to a statistical bias?

Marco Mucciarelli
ISMES S.p.A., Seriate, Bergamo, Italy

Abstract
Relationships between magnitude and macroseismic intensity are widely used, even though intensity is a non-metric observable. Several proposed relationships for the Italian and surrounding regions show a significant correlation of the parameters of the widely used functional form $M = a+bI_0$. This could suggest either the common dependence on a third variable or that the different proposed relationships are due to a pivotal phenomenon around average values, mainly driven by data sampling and uncertainties. Synthetic simulations lend support to the last hypothesis.

Key words *intensity – magnitude – historical earthquakes*

1. Introduction

The large majority of data on destructive earthquakes are derived from macroseismic observations (see, for Italy, Boschi *et al.*, 1995). The magnitude values available in parametric catalogues are thus mainly derived from intensity data.

Recently, a more refined approach takes into account the distances of a given set of intensity values from the epicentre (*e.g.*, Albarello *et al.*, 1995; Gasperini and Ferrari, 1995; Ambraseys and Melville, 1982). Still, the most widely used approach is to seek for a relationship between magnitude and the maximum observed intensity for a given earthquake. Here again, a recently developed approach (Tabular intensity, see Sibol *et al.*, 1987; Di Maro and Tertulliani, 1990) is used in a minority of cases compared with the standard linear relationship between Magnitude (M) and Maximum Observed Intensity (I_0):

$$M = a + bI_0. \tag{1.1}$$

Intensity is a non-metric measure by definition: it is a ordinal scale which is more correctly described by the use of Roman ordinal figures (VI, VIII, XI) rather than by real-looking numbers (6, 8, 11). This means that the «distance» from, *e.g.*, intensity IX and VIII cannot be defined as equal to the «distance» from VI and VII. When we represent intensities equally spaced on the x-axis of a plot, we perform an arbitrary assumption not justified by intensity definition. Moreover, the practice of using macroseismic intensity as a quantitative measure of ground shaking is not endorsed by the author of most recent and advanced macroseismic scales (Grünthal *et al.*, 1993).

A clear description of the reasons why intensity does not give a good correlation with, *e.g.*, peak ground acceleration is given by Spence *et al.* (1992).

With magnitude, however, intensity seems to show a very reasonable linear fit. Thus, many researchers tend to adopt the more or

Mailing address: Dr. Marco Mucciarelli, ISMES, Via Pastrengo 9, 24100 Seriate (BG), Italy; e-mail: mmucciarelli@ismes.it

less hidden assumption that they must be somehow correlated. The only definition of correlation that can be correctly applied to intensity is the non-parametric one (Kendall Tau or Spearman Coefficient). This is because the ranking (equality/inequality) is the only operator that can be used on non-metric variables, which do not admit sum or product.

Moreover, we must note that Mercalli (and before him Rossi and Forel) gave the definition of macroseismic intensity years in advance with respect to the Gutenberg (1945) standard definition of magnitude and thus it is very surprising to observe a good fit between those two variables.

In conclusion, it is obvious that a larger magnitude gives larger intensity but linear regression should not be performed and non-parametric techniques (such as tabular relationships) should be used. Notwithstanding this, the practice of performing linear regressions with intensity is very widespread.

2. Data analysis

As for many empirical relationships, the formula (1.1) has been applied to a variety of data subsets: regional, at various depths, for different tectonic environments, different magnitudes definitions, etc.

The various relationships appear to be different, but they exhibit a common feature: their a and b parameters are strongly correlated.

Figure 1 shows the correlation between a and b parameters for five of the world-wide proposed relationships.

Figure 2 refer to different sets of relationships proposed for Italy and the surrounding regions, also reported in table I. The various a and b values are given either for geographic zonation (Karnik, 1969; Marcelli and Montecchi, 1962; Tinti *et al.*, 1986; Decanini *et al.*, 1993) or for different magnitude definitions (Albarello *et al.*, 1995). It can be clearly seen that they show a linear correlation.

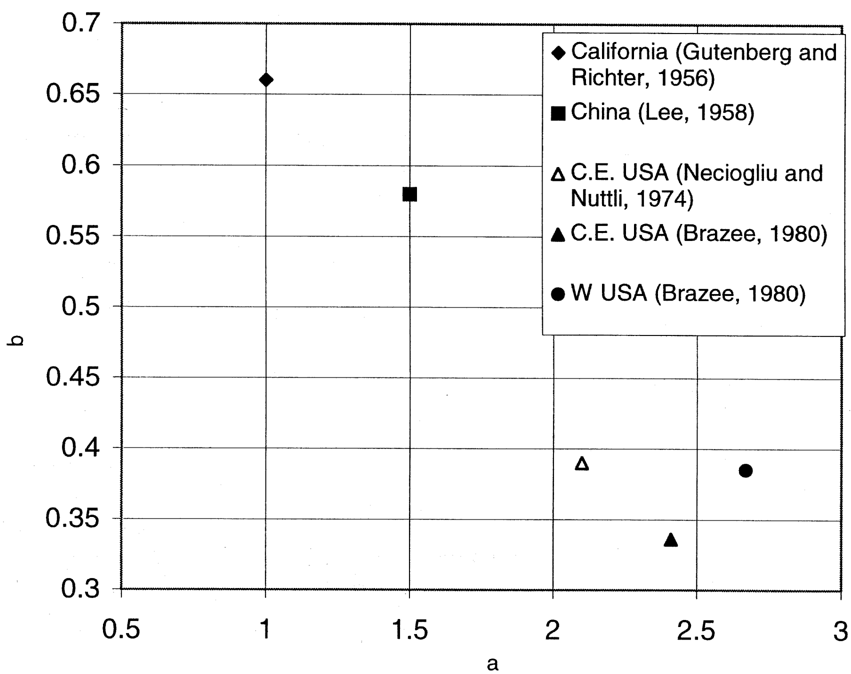

Fig. 1. Correlation between a and b parameters for 5 of the earliest proposed relationships between intensity and magnitude.

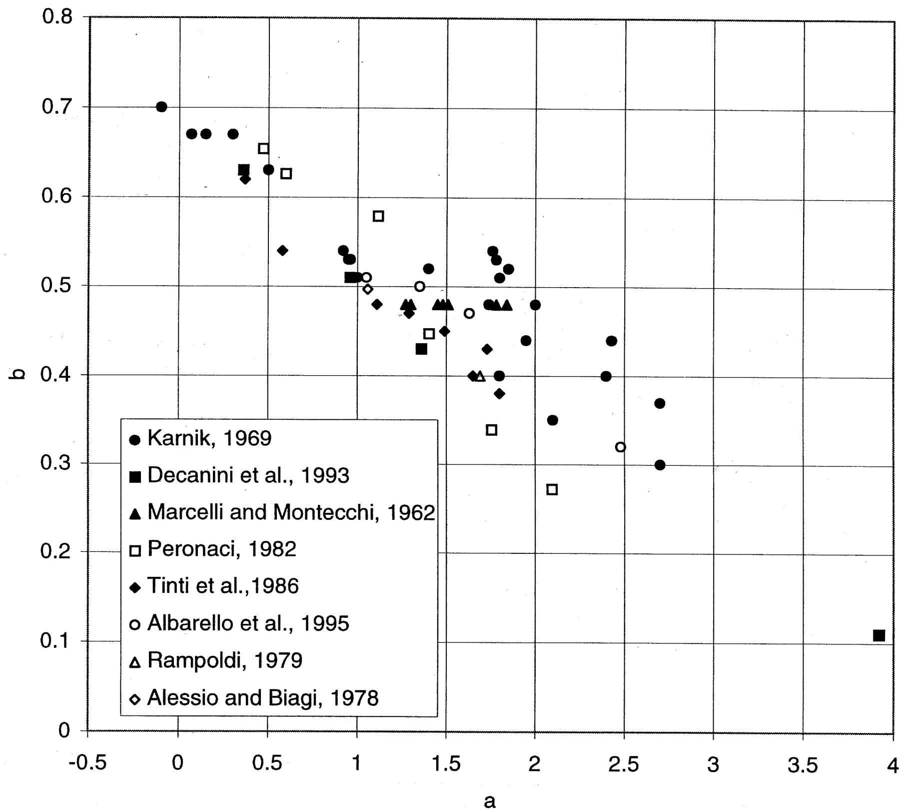

Fig. 2. Correlation between a and b parameters for $M = a + b\,I_0$ relationships for Italy and the surrounding regions.

This fact was noted by Tinti *et al.* (1986) for the parameters of their zoned relationship, and attributed to a possible dependence on earthquakes' different depth.

This explanation pertains to one of the possible general reasons for two parameters showing this correlated behaviour: both variables are dependent on a third one which is not explicit in the relationship. The other possible explanation relies on dispersion induced by error propagation.

For each proposed relationship, a set of expected magnitudes has been generated for intensities varying from III to X (the relationships proposed by Decanini *et al.* (1993) were not included due to their limitation to a smaller intensity range).

Figure 3 reports the standard deviation of magnitudes for each intensity value. Note the presence of heteroscedasticity: extreme intensity values tend to have larger variations than central ones. This could be a clue to a pivotal behaviour around a «true» value: the different a and b values reflect the fact that the linear relationships are similar in the medium intensity range and tend to diverge for extremes. This last property appears to be very important if one considers the following factors that tend to affect extremes of the I/M distribution:

– Magnitude tends to saturate unless M_w is used, which is not the case for the studied relationships.

Table I. List of *a* and *b* parameters for $M = a + b I_0$ relationships proposed for Italy and the surrounding regions.

Proposed by	Estimated for	*a*	*b*
Karnik (1969)	Zone 2	1.74	0.48
Karnik (1969)	Zone 3	2	0.48
Karnik (1969)	Zones 7-8-9	0.07	0.67
Karnik (1969)	Zone 11	1.95	0.44
Karnik (1969)	Zones 13-14-15	2.7	0.37
Karnik (1969)	Zone 16	0.15	0.67
Karnik (1969)	Zone 17	0.5	0.63
Karnik (1969)	Zone 18	0.95	0.53
Karnik (1969)	Austria	−0.1	0.7
Karnik (1969)	Switzerland	0.3	0.67
Karnik (1969)	Zone 19	1	0.51
Karnik (1969)	Zone 20	2.1	0.35
Karnik (1969)	Zone 21	0.96	0.53
Karnik (1969)	Zone 22	0.92	0.54
Karnik (1969)	Zone 23	1.4	0.52
Karnik (1969)	Zone 25	2.7	0.3
Karnik (1969)	Zone 26	2.4	0.4
Karnik (1969)	Zone 27	1.8	0.51
Karnik (1969)	Zone 29	1.76	0.54
Karnik (1969)	Zone 30	2	0.48
Karnik (1969)	Zone 32	1.85	0.52
Karnik (1969)	Zone 33	1.85	0.52
Karnik (1969)	Zone 34	1.78	0.53
Karnik (1969)	Zone 35	2.43	0.44
Karnik (1969)	Zone 7-8-9	1.8	0.4
Decanini *et al.* (1993)	Etna, I < V	0.96	0.51
Decanini *et al.* (1993)	Etna, I < VIII	1.36	0.43
Decanini *et al.* (1993)	Etna, I > VIII	3.92	0.11
Decanini *et al.* (1993)	Iblei Mts., I < V	0.96	0.51
Decanini *et al.* (1993)	Iblei Mts., I < VIII	0.36	0.63
Marcelli and Montecchi (1962)	E. Alps	1.84	0.48
Marcelli and Montecchi (1962)	Emilia-Romagna	1.78	0.48
Marcelli and Montecchi (1962)	Tuscany	1.3	0.48
Marcelli and Montecchi (1962)	Umbria-Marche	1.27	0.48
Marcelli and Montecchi (1962)	Abruzzo-Molise	1.48	0.48

Table I (*continued*).

Proposed by	Estimated for	a	b
Marcelli and Montecchi (1962)	Basilicata-Calabria	1.45	0.48
Marcelli and Montecchi (1962)	E. Sicily	1.51	0.48
Peronaci (1982)	Alps	0.598	0.626
Peronaci (1982)	Po Valley	1.115	0.579
Peronaci (1982)	Apennines	1.755	0.339
Peronaci (1982)	Ancona	2.094	0.272
Peronaci (1982)	Calabria	0.471	0.654
Peronaci (1982)	W. Sicily	1.402	0.447
Tinti *et al.* (1986)	Italy	1.29	0.47
Tinti *et al.* (1986)	Zone 1	0.58	0.54
Tinti *et al.* (1986)	Zone 2	1.73	0.43
Tinti *et al.* (1986)	Zone 3	0.37	0.62
Tinti *et al.* (1986)	Zone 4	1.65	0.4
Tinti *et al.* (1986)	Zone 5	1.8	0.38
Tinti *et al.* (1986)	Zone 6	1.11	0.48
Tinti *et al.* (1986)	Zone 7	1.49	0.45
Albarello *et al.* (1990)	Italy	1.05	0.51
Albarello *et al.* (1995)	M_l	1.63	0.47
Albarello *et al.* (1995)	M_s	1.35	0.5
Albarello *et al.* (1995)	M_b	2.48	0.32
Rampoldi (1979)	Italy	1.69	0.4
Alessio and Biagi (1978)	Latium	1.058	0.497

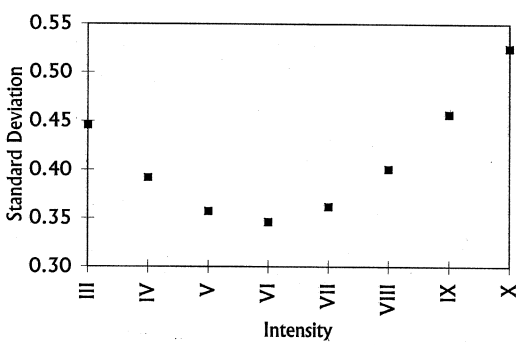

Fig. 3. Standard deviation of magnitudes calculated using all the relationship from table I (except Decanini *et al.*, 1993), as a function of intensity.

– Intensity is saturated by definition at XII degree. In Italy this degree has never been assigned, and even the most catastrophic earthquakes are upper-bounded at XI MCS.

– The relation between M and I is in turn saturated: above a certain magnitude, any event is «catastrophic» and it is not likely that, *e.g.*, a magnitude 7.0 event could give effects relevant to intensity distinguishable from a 7.5 event. On the lower edge, magnitude is a logarithmic scale that may assume negative values, while intensity lowest value is set to 1 for all the events that can be only instrumentally detected.

– Large events are rare, and depending on the method adopted for regression, they can influence the proposed relationships.

– Seismic networks have a detectability range that affects completeness for the lower magnitude range, and thus cause undersampling in this part of the relationship.

A synthetic test was carried out to check the influence of random errors as well as the fact that, due to Gutenberg-Richter law, higher magnitudes are less frequent than lower ones, as are intensities.

Hundred sets of events were generated. They follow a Gutenberg-Richter law with *b*-values centered at 1, with random variation ± 0.25. The relation between intensity and magnitude was taken equal for all the sets as $M = 0.7 + 0.4 I_0$. Both magnitude and intensity were scrambled with the addition of a «noise» uniformly distributed, with mean = 0 and standard deviation = 0.5 (for intensity the error was discretized at semi-integer values). The lower magnitudes were cut-off in order to simulate catalogue incompleteness due to a detection threshold. The value for this high-pass magnitude cut was a random value in the range 3.0 ± 0.25.

To simulate different sample dimensions in real data sets, three different subsets of *a* and *b* were calculated using 10, 50 or 100 events.

Figure 4 shows the 100 calculated *a* and *b* parameters (listed in table II) together with the real observed ones. It can be noted that the trend is the same. Synthetic data appear less scattered around the main trend. This could be due to the fact that real data sets are affected by further sources of uncertainty which were difficult to include in the simulation; among them we can name the following:

– Non-uniformity in magnitude (*e.g.*, M_s merged with M_l).

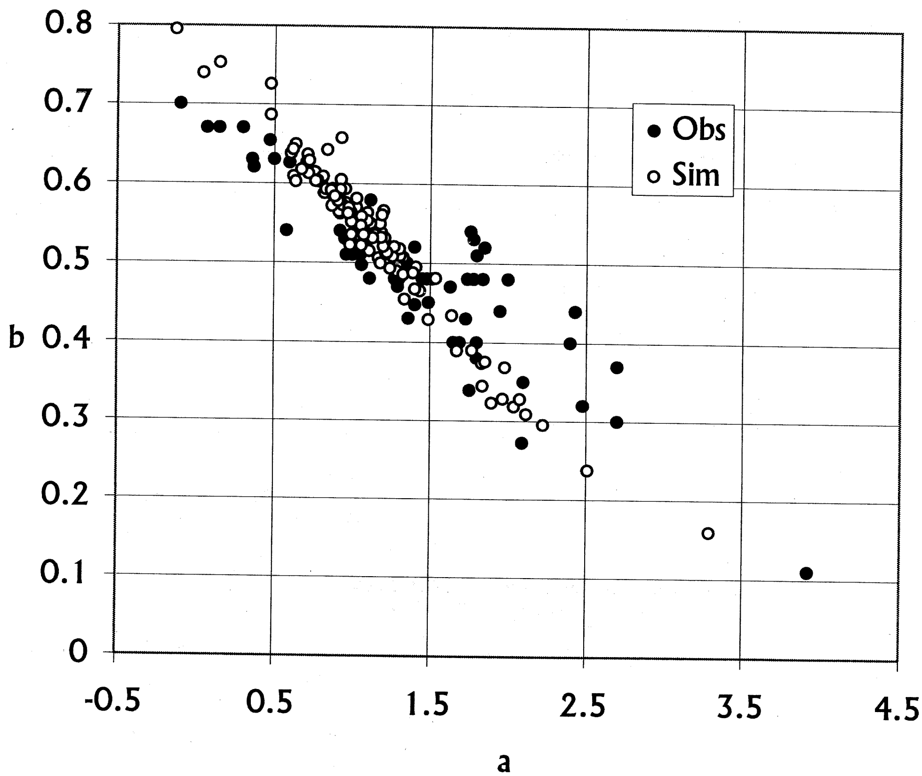

Fig. 4. Correlation between *a* and *b* values for 100 synthetic data sets and the real observed ones.

Table II. List of a and b parameters generated with stochastic simulation (see text for details).

b-value	Number of realisations	Magnitude cut	a	b	b-value	Number of realisations	Magnitude cut	a	b
1.06	10	2.81	−0.13	0.79	1.15	50	2.95	1.06	0.52
0.76	10	2.98	0.04	0.74	0.85	100	2.95	1.06	0.56
0.93	10	2.79	0.15	0.75	1.11	50	2.81	1.07	0.54
1.18	10	3.20	0.47	0.73	1.15	100	2.95	1.07	0.53
1.15	10	2.95	0.48	0.69	0.83	10	2.90	1.10	0.56
0.83	100	2.90	0.61	0.64	0.79	50	3.19	1.10	0.55
1.02	100	2.97	0.62	0.64	0.85	10	2.95	1.10	0.55
1.11	10	2.81	0.62	0.61	0.84	50	2.99	1.10	0.53
0.83	50	2.90	0.64	0.65	1.11	100	2.81	1.11	0.51
1.10	100	2.87	0.64	0.60	0.92	100	3.13	1.12	0.55
1.06	100	2.89	0.67	0.62	0.98	10	2.97	1.13	0.53
1.06	100	3.02	0.71	0.63	1.06	50	3.02	1.14	0.53
0.77	50	3.11	0.71	0.64	1.15	50	2.89	1.16	0.51
0.97	50	2.90	0.72	0.61	0.83	10	3.11	1.18	0.55
0.92	10	3.13	0.72	0.63	1.22	100	2.82	1.18	0.50
0.77	100	3.11	0.76	0.61	0.88	100	3.22	1.18	0.53
0.98	50	2.97	0.76	0.60	0.99	50	3.13	1.19	0.54
0.85	50	2.95	0.77	0.60	0.82	10	3.21	1.19	0.56
0.93	100	2.79	0.79	0.60	0.76	100	2.98	1.20	0.52
0.85	50	3.21	0.80	0.60	1.13	10	3.14	1.20	0.57
0.92	50	3.13	0.81	0.61	0.89	50	3.24	1.21	0.53
1.13	100	3.14	0.83	0.59	1.13	50	3.14	1.22	0.51
0.82	100	3.21	0.84	0.64	1.06	50	2.81	1.24	0.49
1.11	100	2.77	0.86	0.59	0.89	100	3.24	1.27	0.52
0.77	10	3.06	0.86	0.59	1.02	10	2.76	1.28	0.49
1.22	50	2.82	0.87	0.57	1.22	10	2.82	1.30	0.52
0.77	100	3.06	0.87	0.59	0.79	100	3.19	1.31	0.51
0.98	100	2.97	0.88	0.58	1.02	50	2.97	1.33	0.49
0.83	100	3.11	0.90	0.58	0.99	10	3.13	1.33	0.51
1.02	50	3.02	0.91	0.56	1.02	10	3.02	1.34	0.45
0.83	50	3.11	0.91	0.58	1.21	100	3.07	1.39	0.49
1.02	50	2.76	0.92	0.56	1.13	50	3.01	1.40	0.47
0.85	100	3.21	0.92	0.59	0.77	10	3.11	1.41	0.49
0.82	50	3.21	0.93	0.66	1.15	100	2.89	1.49	0.43
1.06	100	2.81	0.93	0.57	0.88	10	3.22	1.53	0.48
1.18	50	2.97	0.95	0.56	0.79	10	3.19	1.64	0.43
0.78	100	3.18	0.95	0.59	1.11	10	2.77	1.67	0.39
1.10	50	2.87	0.96	0.58	1.18	50	3.20	1.77	0.39
0.76	50	2.98	0.97	0.56	0.99	100	3.13	1.83	0.37
0.93	50	2.79	0.97	0.57	1.21	10	3.07	1.84	0.34
1.13	10	3.01	0.98	0.52	1.06	10	2.89	1.85	0.38
1.02	100	2.76	0.99	0.56	1.18	10	2.97	1.90	0.32
1.02	100	3.02	0.99	0.56	0.97	10	2.90	1.97	0.33
0.97	100	2.90	0.99	0.54	0.85	10	3.21	1.98	0.37
1.18	100	2.97	0.99	0.55	1.15	10	2.89	2.04	0.32
0.78	50	3.18	1.02	0.57	1.18	100	3.20	2.08	0.33
0.88	50	3.22	1.03	0.58	1.10	10	2.87	2.12	0.31
0.84	10	2.99	1.04	0.53	0.89	10	3.24	2.23	0.30
0.77	50	3.06	1.04	0.55	1.06	10	3.02	2.51	0.24
1.11	50	2.77	1.05	0.55	0.78	10	3.18	3.29	0.16

– Non uniformity in intensity assessment (*e.g.*, MCS mixed with MSK).

– Choice of the regression algorithm (*e.g.*, minimization of I or M residuals).

– Assignment of semi-integer intensity classes to adjacent ones.

– Fake independence between the two variables (in some old catalogues, macroseismic magnitude were erroneously labeled as instrumental ones, see Paciello, 1988).

To verify if there are third variables that may explain the a and b correlation, the distribution of the simulated parameters *versus* other variables was examined. The parameters derived from synthetic data sets do not show any correlation either with b-values (fig. 5) or magnitude cut-off (fig. 6) used in the simulations. Grouping parameters following the samples dimension (fig. 7), it can be seen that using 50 or 100 events one obtains the same distribution: if a small sample is used (10 events) the variance increases sharply.

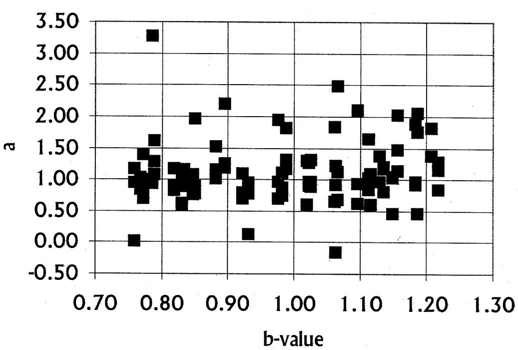

Fig. 5. Relation between synthetic a parameters and b-values used in simulation.

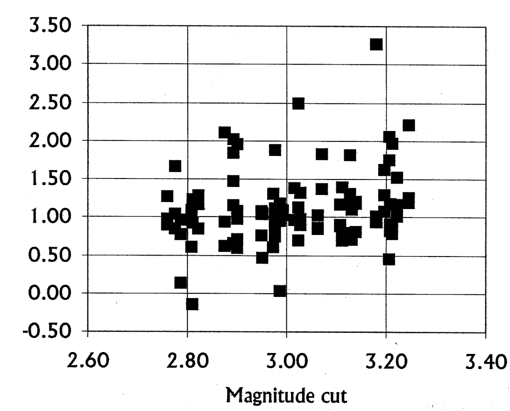

Fig. 6. Relation between synthetic a parameters and magnitude cut-off values used in simulation.

3. Discussion and conclusions

The determination of the so called «Macroseismic Magnitude» is strongly affected by the non-metric characteristic of intensity and the intrinsic uncertainties of it.

This applies to all the proposed methodologies (including Tabular Intensity and the use of the whole macroseismic map) but in particular to the most used relationship, the linear one.

Great care should be taken before invoking physical reasons underlying the difference in a and b parameters obtained by empirical studies. The strong correlation observed among the a and b parameters was previously attributed in one case to different foci depth. But the fact that different relationships proposed using different data sets for the same area still show this phenomenon lends support to the idea of a statistical bias induced by a pivotal effect around some «true» or simply most probable value.

A simple synthetic test shows that this effect can be obtained simulating three very common causes:

1) Errors in magnitude estimates and intensity assessments.

2) Data incompleteness for lower magnitudes.

3) Different number of events in data sets.

Moreover, further uncertainties can affect the results, and namely

– saturation both in M and I scales;

– non-uniformity in magnitude and/or in intensity assessment;

– choice of the regression algorithm;

– assignment of semi-integer intensity classes to adjacent ones;

– fake independence between the two variables.

The simulation test cannot prove the uniqueness of the solution proposed. In princi-

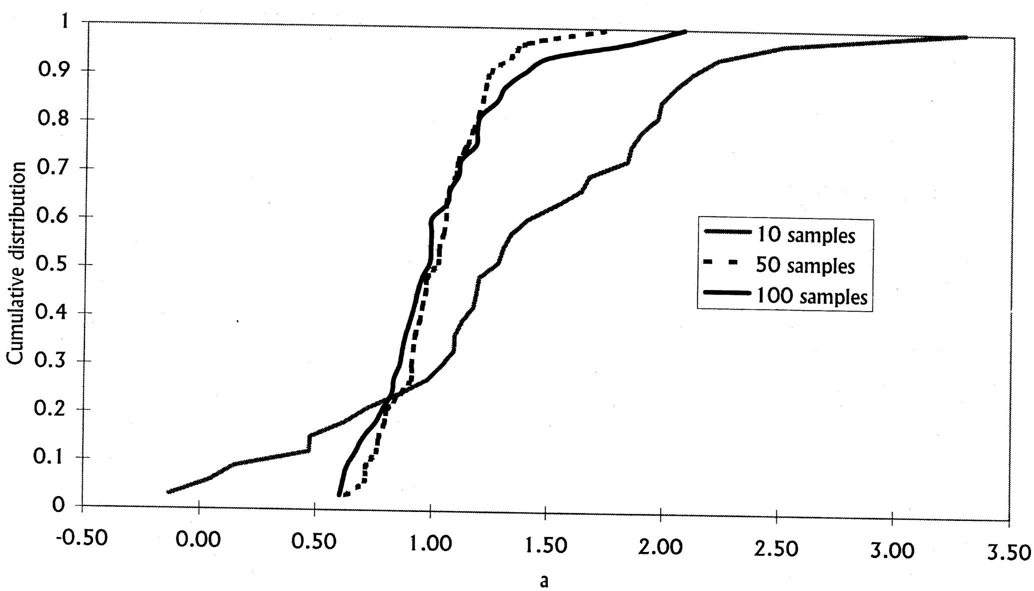

Fig. 7. Distributions of synthetic *a* parameters grouped according to number of realisation in the data sets used.

ple, a still unknown and unchecked variable may exist which might be the cause of the correlation observed between the *a* and *b* parameters. Nevertheless, before proposing a variable as the cause of this observed phenomenon all the above mentioned causes of uncertainties must be carefully inspected and ruled out.

REFERENCES

ALBARELLO, D., M. MUCCIARELLI and E. MANTOVANI (1990): Empirical magnitude-intensity relationships: a non-parametric approach, in *Proceedings XXI ESC General Assembly, Barcelona*, 871-874.

ALBARELLO, D., R. BERARDI, C. MARGOTTINI and M. MUCCIARELLI (1995): Macroseismic estimates of magnitude in Italy, *Pageoph*, **145** (2), 1122-1137.

ALESSIO, M. and P.F. BIAGI (1978): Phenomenological aspect of the seismicity of Latium, *Ann. Geofis.*, **31**, 177-314.

AMBRASEYS, N.N. and C.P. MELVILLE (1982): *A History of Persian Earthquakes* (Cambridge University Press), pp. 219.

BOSCHI, E., P. GASPERINI, G. SMRIGLIO and G. VALENSISE (1995): Un catalogo dei terremoti di nuova generazione, in *Catalogo dei Forti Terremoti in Italia dal 461 a.C. al 1980*, edited by E. BOSCHI, G. FERRARI, P. GASPERINI, F. GUIDOBONI, G. SMRIGLIO and G. VALENSISE, 9-19.

BRAZEE, R.J. (1980): *An Analysis of Earthquake Intensity with Respect to Attenuation, Magnitude and Rate of Recurrence*, NUREG/CR 1085, Washington, pp. 99.

DECANINI, L., C. GAVARINI and G. OLIVETO (1993): Rivalutazione dei terremoti storici della Sicilia Sud Orientale, in *Atti VI Convegno Nazionale di Ingegneria Sismica, Perugia*, 1101-1109.

DI MARO, R. and A. TERTULLIANI (1990): The relation between intensity and magnitude for Italian earthquakes, *Pure Appl. Geophys.*, **132**, 711-718.

GASPERINI, P. and G. FERRARI (1995): Stima dei parametri sintetici, in *Catalogo dei Forti Terremoti in Italia dal 461 a.C. al 1980*, edited by E. BOSCHI, G. FERRARI, P. GASPERINI, E. GUIDOBONI, G. SMRIGLIO and G. VALENSISE, 96-111.

GRÜNTHAL, G., R.M.W. MUSSON, J. SCHWARZ and M. STUCCHI (1993): European Macroseismic Scale 1992 (up-dated MSK scale), *Cahiers du Centre Européen de Géodynamique et de Séismologie*, vol. 7.

GUTENBERG, B. (1945): Amplitude of surface waves magnitude of shallow earthquakes, *Bull. Seism. Soc. Am.*, **35**, 57-69.

GUTENBERG, B. and C.F. RICHTER (1956): Earthquake magnitude, intensity, energy and acceleration, *Bull. Seism. Soc. Am.*, **46**, 105-145.

KARNIK, V. (1969): *The Seismicity of the European area* (Reidel Publ. Co.), pp. 364.

LEE, S.P. (1958): A practical magnitude scale, *Acta Geophys. Sin.*, **9** (2), 98-102.

MARCELLI, I. and A. MONTECCHI (1962): Contributi per uno studio della sismicità dell'Italia, *Ann. Geofis.*, **15**, 2-3.

NECIOGLIU, A. and O.W. NUTTLI (1974): Some ground motion and intensity relations for the Central United States, *Earth. Eng. Struct. Dyn.*, **3**, 111-119.

PACIELLO, A. (1988): Some aspect of the hazard analysis for an Italian region, in *Historical Seismicity in Central-Eastern Mediterranean Region*, edited by L. SERVA and C. MARGOTTINI, 245-258.

PERONACI, M. (1982): Intensity-magnitude relationships for Italian regions, *Boll. Geofis. Teor. Appl.*, **24**, 121-128.

RAMPOLDI, R. (1979): *Rapporto sul Gruppo di Lavoro Scuotibilità*, CNR-Progetto Finalizzato Geodinamica.

SIBOL, M.S., P.A. BOLLINGER and J.B. BIRCH (1987): Estimation of magnitudes in Central and Eastern North America using intensity and felt area, *Bull. Seism. Soc. Am.*, **77** (5), 1635-1654.

SPENCE, R.J.S., A.W. COBURN, A. POMONIS and S. SAKAI (1992): Correlation of ground motion with building damage: the definition of a new damage based seismic intensity scale, in *Proceedings of X WCEE, Madrid*, vol. 1, 551-556

TINTI, S., T. VITTORI and F. MULARGIA (1986): Regional intensity-magnitude relationships for the Italian territory, *Tectonophysics*, **127**, 129-154.

(received September 10, 1997;
accepted March 27, 1998)