Geophysical Journal International

Geophys. J. Int. (2010) **180,** 347–360 doi: 10.1111/j.1365-246X.2009.04414.x

On varying *b*-values with depth: results from computer-intensive tests for Southern California

D. Amorèse,[1] J.-R. Grasso[2] and P. A. Rydelek[3]

[1]*Laboratoire de M2C, Université de Caen,* UMR CNRS/INSU 6143, *Caen,* F-14000, *France. E-mail: daniel.amorese@unicaen.fr*
[2]*Laboratoire de Géophysique Interne et Tectonophysique, CNRS, Observatoire de Grenoble, Université J. Fourier, Maison des Géosciences,* BP53, 38041 *Grenoble Cedex* 9, *France*
[3]*Korea Ocean Research and Development Institute, Ansan* 425-600, *Korea*

Accepted 2009 October 9. Received 2009 October 9; in original form 2008 November 28

SUMMARY

The Gutenberg–Richter *b*-value is thought to reflect the stress conditions in the crust; therefore, spatial and/or temporal variations of the *b*-value can provide important information regarding crustal tectonics. We investigate the variation of *b*-value with depth in seven selected areas of Southern California. A previous study provided a detailed mapping of the variations of *b* with depth in California; our study is less systematic than this study. Our approach is more similar to the regional one used by Mori & Abercrombie. In comparison to these previous studies, our investigation indicates that the variability of *b* is often not statistically significant and that the decrease of *b* with depth should be interpreted with caution. The seismic catalogues used are subsets of a set of about 100 000 seismic events recorded by the Southern California Seismic Network (SCSN) and relocated by Richards-Dinger & Shearer. We study the performance of Utsu's test compared to bootstrap tests for comparison of *b*-values. The results of our investigation also raise the question of the relevancy of Utsu's test when comparing *b*-values. Both simulations and real cases show that the Utsu's test is biased towards rejection of the null hypothesis in favour of the hypothesis that *b*-values are significantly different.

Key words: Numerical approximations and analysis; Spatial analysis; Statistical seismology; North America.

1 INTRODUCTION

A fundamental statistical description of seismicity is the Gutenberg–Richter (G–R) law (Gutenberg & Richter 1944)

$$\log_{10} N(M) = a - bM, \qquad (1)$$

where a and b are constants, M is the magnitude, and $N(M)$ is the number of earthquakes that occur in a specific time window with magnitude $\geq M$. The *b*-value characterizes each frequency–magnitude distribution (FMD) by reflecting the relative proportion of the number of large and small earthquakes in a time interval for a given region. Because *b* is often interpreted as an indicator of applied shear stress and material heterogeneity (Mogi 1962; Scholz 1968; Main *et al.* 1992; Mori & Abercrombie 1997; Schorlemmer *et al.* 2005), an important question is whether significant spatial or temporal variations of the *b*-value exist. In short, when the confining pressure is increased with depth, the *b*-value is expected to decrease whereas the average earthquake magnitude is expected to increase. Therefore this inferred link of *b*-value with the applied stress explains why *b*-value variations are considered valuable data in statistical seismology and seismic hazard studies (estimation of recurrence times of large earthquakes) and in volcanic seismology (mapping of the magma chamber) (Wiemer & Wyss 2002; Zuniga & Wyss 2001; McNutt 2005; Murru *et al.* 2007). Many results are presented to support the spatial and/or temporal heterogeneity of *b* (Wiemer & Benoit 1996; Amelung & King 1997; Molchan *et al.* 1997; Mori & Abercrombie 1997; Wiemer & Wyss 1997; Wyss *et al.* 1997; Power *et al.* 1998; Wiemer *et al.* 1998; Wyss *et al.* 2000; Gerstenberger *et al.* 2001; Zuniga & Wyss 2001; Rydelek *et al.* 2002; Schorlemmer *et al.* 2003; Murru *et al.* 2004; Bridges & Gao 2006; Murru *et al.* 2007). Nevertheless, some seismologists have shown that the apparent variability of the *b*-value was in some cases not significant (Del Pezzo *et al.* 2003) and others interpreted the *b*-value variability as an artefact due to a lack of statistical rigour (Kagan 1999). Such viewpoints are not surprising, when one acknowledges the numerous sources of instabilities and uncertainties in *b*-value determinations. Following 'Ockham's Razor' principle, the simplest explanations are the most likely. Thus, when temporal changes or spatial variations in *b*-value are observed, one may suspect that these are due to changes in network operating parameters or to spatial variations in seismic station density. Nevertheless, errors in *b*-value calculation can also result in potentially misleading *b*-value variations. Wiemer & Wyss (2000) showed that a significant bias may be introduced in *b*-value computation by a wrong choice of the threshold magnitude. Marzocchi & Sandri (2003) recently pointed out the biases introduced by the use of binned magnitudes and the effect of measurement errors on magnitude. But other problems and sources of error are possible when computing the b-value

(i) insufficient sample size;
(ii) the mixing of magnitude types;
(iii) the contamination by blasts and
(iv) earthquake clusters.

It should be noted that some studies (Wyss *et al.* 2000; Rydelek *et al.* 2002) supporting a b-value variation drew their conclusions about this possible phenomenon without performing any statistical test of hypothesis. Most studies (Mori & Abercrombie 1997; Bridges & Gao 2006; Gerstenberger *et al.* 2001; Zuniga & Wyss 2001; Murru *et al.* 2004) only used a single test method to support their results regarding b-value changes, namely the Utsu's test (Utsu 1992, 1999). Unfortunately, this test is a goodness-of-fit and homogeneity test; therefore, it does not evaluate the significance of a one-sided null hypothesis. In this study, we propose additional tests for the statistical significance of b-value differences. Our first approach is based on bootstrapping to estimate the relevancy of the b-value comparison statistic, when b-values are calculated through the Aki–Utsu formula (Aki 1965; Utsu 1965). Our results are cross-checked, when possible, using a non-parametric linear regression method for the computation of b-values.

After a brief summary of previous results on b-values for Southern California, these methods are presented. Several numerical simulations are then performed to compare the different computation techniques of b-value (Aki–Utsu formula and the non-parametric method). Next we perform numerical simulations to compare the results of various significance tests on the differences in b-value. Finally, the dependence of b-value on depth is investigated for seven subzones of Southern California, with areas ranging from 900 to 2500 km^2.

2 SOME PREVIOUS RESULTS ON *b*-VALUES FOR SOUTHERN CALIFORNIA

This section is mostly focused on the Gerstenberger *et al.*'s work (Gerstenberger *et al.* 2001). In their paper, Gerstenberger *et al.* (2001) tested the hypothesis that the b-value for the shallowest part of the crust is significantly higher than the bottom part of the crust. They systematically mapped the ratio (r_b) of shallow to deep b-values in California and tested to see if the difference was significant at the 99 per cent level. Their tests were carried out both on Northern and Southern California data. For Southern California, Gerstenberger *et al.* stated that they used the high quality relocated Southern California catalogue of Richards-Dinger & Shearer (2000) from 1981.3 to 1998.2. As several quarries are located in Southern California, these authors removed all daytime events between the hours of 16:00 and 3:00 GMT. Their study areas were sampled at 5 km node spacing by using cylinders containing 1000 earthquakes.

To compute the ratio of shallow to deep b-values (r_b) each cylinder with radius smaller or equal to 40 km was divided into a a top zone from 0 to 5 km and a bottom zone from 8 to 15 km. A maximum magnitude cut-off of M5.5 was adopted by these authors without explanation and calculations of b-values were done using the Aki–Utsu (Aki 1965; Utsu 1965) equation. A minimum of 50 events of magnitude greater than m_0 (threshold magnitude) was required for each b-value computation. To determine if the difference between the top and bottom zone b-value was significant, Gerstenberger *et al.* applied the Utsu's test (Utsu 1992).

Whereas the study of Gerstenberger *et al.* was concerned with both Northern and Southern California, our study is focused on Southern California where Gerstenberger *et al.* found 34.4 per cent of the area showing r_b values greater than 1.0 (top zone b-value greater than bottom zone b-value). They compared this value with the small area showing $r_b < 1.0$ (1.1 per cent of the Southern California sample area). From their analysis of both the Southern and the Northern California catalogues, Gerstenberger *et al.* were in agreement with previous studies (Mori & Abercrombie 1997; Wiemer & Wyss 1997; Wiemer *et al.* 1998) and concluded that the general trend was a decrease in b-value with depth. Moreover, combining random simulations and the Utsu's test, Gerstenberger *et al.* suggested that r_b values smaller than 1.0 might simply represent random fluctuation of b.

3 METHODS

3.1 Computation of the *b*-values

Many equations have been proposed to represent the distribution of earthquake magnitudes, although the original G–R relation seems to be the most suitable for many cases (Utsu 1999). Therefore, we have calculated b-values for California under the assumption that magnitudes were distributed in accordance with the G–R formula. Although some investigators (Marzocchi & Sandri 2003) recently noted an improvement provided by Tinti and Mulargia (1985, 1987) to the estimation of the b-value, we have followed Gerstenberger *et al.*'s (2001) and employed the Aki–Utsu (Aki 1965; Utsu 1965) equation. This choice was made to use the most widely used equation for the computation of the b-value. The Aki–Utsu (Aki 1965; Utsu 1965) equation with continuity correction (in this study, magnitude values have been rounded to one decimal place) reads

$$b^{\mathrm{AU}} = \log_{10}(e)/[\bar{m} - (m_0 - 0.05)]. \tag{2}$$

Here $\bar{m}$ is the mean magnitude and m_0 is the threshold magnitude. Thus, a high mean magnitude is equivalent to a low b-value. In this study, m_0 was determined following the change-point method proposed by Amorèse (2007). Aki (1965) and Shi & Bolt (1982) gave error estimates for maximum likelihood b-values

$$SE_{\mathrm{SB}} = 2.30 b^{\mathrm{AU}^2} \sqrt{\frac{\sum_{i=1}^{N} (M_i - \bar{m})^2}{N(N-1)}}. \tag{3}$$

This error estimate is too low because possible sources of errors are neglected by the underlying statistical model. The most cited error source is from the uncertainty in the threshold magnitude, but uncertainties in magnitude estimation should also be considered (Rhoades 1996). This was not directly possible in this study as the data set that was used did not include the information on magnitude uncertainties. Thus, like Schorlemmer *et al.* (2003) and Amorèse (2007), we simply adopted a bootstrap approach to estimate more realistic errors in b. The bootstrap approach is intended to simulate the variability of FMDs.

In this study, to strengthen our results, we were concerned with an alternative way to estimate b-value. Many authors show that the estimation of the b-value by a least squares technique is statistically unjustified and/or biased (Weichert 1980; Bender 1983; Sandri & Marzocchi 2007). For instance, b-value calculated by the method of linear least squares regression has the major drawback of being strongly influenced by the distribution of a few large earthquakes in the FMD. Thus, our attention was directed towards a non-parametric regression technique using Siegel's repeated medians (Siegel 1982). This method is highly resistant to outliers and large measurement errors (Smirnov 2003) and requires limited *a priori* information regarding errors. For the G–R relation, if n magnitude intervals are considered, for each point i, $n-1$ slopes are computed, between i and the other points

$$b_{ij} = \frac{\log_{10}(N_j) - \log_{10}(N_i)}{M_j - M_i}, \tag{4}$$

where $M_i \neq M_j$.

Then, for each point i, the median of slopes is taken. This results in n median values. In the repeated medians technique, the slope estimator is the median from these n medians. Therefore, the b-value is

$$b^{RM} = -\mathrm{Med}[\mathrm{Med}(b_{ij})]. \tag{5}$$

The standard error (SE) of this slope can be estimated through a bootstrap approach as follows (Siegel, pers. comm., 2007 November 27):

(i) the sample of n $(M, \log_{10}(N))$ pairs is treated as a population and sampled with replacement.

(ii) the b^{RM} value is calculated for this bootstrap sample.

(iii) stages #1 and #2 are repeated B times [B is large enough when $B > 1000$ (Efron & Tibshirani 1993; Mooney & Duval 1993)].

(iv) the bootstrap standard error of b^{RM} is the standard deviation of the B bootstrap b^{RM} values.

Thus, if B is the number of replicates, the bootstrap standard error of b^{RM} is estimated by

$$SE(b^{\mathrm{RM}}) = \sqrt{\frac{1}{B-1}\sum_{i=1}^{B}\left(b_i^{\mathrm{RM}} - \bar{b}^{\mathrm{RM}}\right)^2}. \tag{6}$$

In this study, the repeated medians regression technique was used to fit incremental FMDs. The bin size was taken to be equal to 0.1 magnitude unit.

3.2 Significance test of the difference in *b*-value between two groups

This question has been addressed by Utsu (1966, 1992, 1999). The first solution proposed by Utsu (1966) was based on the F-distribution. Given two earthquake groups A and B, this test consisted in testing the hypothesis $b_{0\mathrm{A}} = b_{0\mathrm{B}}$ (b-values of populations) by comparing $b_{\mathrm{B}}/b_{\mathrm{A}}$($b$-values of the samples, $b_{\mathrm{A}} < b_{\mathrm{B}}$) with the F-value with $2s_{\mathrm{A}}$ and $2s_{\mathrm{B}}$ degrees of freedom at a given confidence level, where s_{A} and s_{B} are the numbers of earthquakes in A and B.

Utsu also proposed (Utsu 1992, 1999) a similar test based on the Akaike information criterion (AIC). This later test is certainly the most widely used when b-values have to be compared (Mori & Abercrombie 1997; Zuniga & Wyss 2001). For two earthquake groups A and B, it consists in computing the difference in AIC

$$\begin{aligned}\Delta AIC = & -2(N_{\mathrm{A}} + N_{\mathrm{B}})\ln(N_{\mathrm{A}} + N_{\mathrm{B}}) + 2N_{\mathrm{A}}\ln(N_{\mathrm{A}} + N_{\mathrm{B}}b_{\mathrm{A}}/b_{\mathrm{B}}) \\ & + 2N_{\mathrm{B}}\ln(N_{\mathrm{A}}b_{\mathrm{B}}/b_{\mathrm{A}} + N_{\mathrm{B}}) - 2,\end{aligned} \tag{7}$$

where N_{A} and N_{B} are the number of earthquakes used to calculate b-values b_{A} and b_{B}, respectively. Usually, the difference in AIC is considered significant if it exceeds about 2. A relationship exists to derive a probability value from ΔAIC (Utsu 1992)

$$p_{\Delta AIC} = \exp\left(\frac{-\Delta AIC}{2} - 2\right). \tag{8}$$

Therefore, if $\Delta AIC = 2$, $p_{\Delta AIC} \simeq 0.05$. It should be noted that two data sets with different magnitude thresholds could only be tested by this ΔAIC test if the assumed Gutenberg–Richter relation is similar (Utsu 1999). Until now, this assumption is still a debatable point (Rydelek & Sacks 2003; Wiemer *et al.* 2003). But the most important limitation for both the F-test and the AIC test consists in their null hypothesis ('the data follow a specified distribution'), because these two tests are goodness-of-fit tests. Therefore, these procedures do not provide strong indications on the significance of comparison null hypotheses ('$b_1 - b_2 = 0$'), particularly when one-sided alternative hypotheses are tested ('$b_1 - b_2 > 0$').

In this study, we used the bootstrap resampling procedure to test the difference in b^{AU}. The bootstrap method (Efron 1979; Efron & Tibshirani 1993) allows to carry out crude statistical inference while empirically incorporating all the sources of uncertainty in the computation of b-value. The fundamental assumption of the bootstrap method is that the observed data are representative of the underlying population. Schorlemmer *et al.* (2003) used a bootstrap method to compute the stability and significance of a b-value contrast.

Several statistical quantities can be estimated with the bootstrap. In this study, the bootstrap was focused on obtaining estimates of

(i) standard errors of b^{RM} (see Section 3.1);

(ii) p-values for test statistics under the null hypothesis that two population b^{AU}-values are equal.

The method principle consists in generating replicates (resamples) from the original sample with a Monte-Carlo simulation. The bootstrap procedure is computer-time consuming and requires a reliable pseudo-random number generator. With modern computers, these two following requirements are easily satisfied:

(i) computer power is progressively growing and is inexpensive;

(ii) the free R statistical environment (Ihaka & Gentleman 1996) provides several efficient pseudo-random number generators.

The procedure for obtaining estimates of standard errors of b^{RM} has been previously described in Section 3.1. The procedure for estimating p-values for test statistics is as follows. For the correct estimation of p-values, resampling must be performed under an appropriate null hypothesis. We consider the case of two independent random samples of earthquake magnitudes and we are interested in the difference in population b-value. To get the null distribution of the difference, we need to force the b-values to be equal when creating replicates. This can be achieved by drawing both samples with replacement from the pooled magnitude set. This is the resampling version of the R.A. Fisher's permutation test (Fisher 1934), but using b-values instead of means as a comparison statistic. The algorithm of our 'seismological' version of the Fisher's permutation test statistic is as follows:

(i) All the $n + m$ magnitude values from both FMDs are combined together to form x.

(ii) B samples of size $n + m$ are drawn with replacement from the pooled magnitude set x. The first n observations are called z^{*i} and the remaining m observations are called y^{*i}, for $i = 1, 2, \ldots, B$.

(iii) On each sample, $T(.)$ is evaluated, $T(x^{*i}) = b(z^{*i}) - b(y_{*i})$, $i = 1, 2, \ldots, B$.

(iv) The bootstrap one-sided p-value of the test is defined as

$$p_{\mathrm{B}} = \frac{\text{No. of } T(x^{*i}) \geq T_0}{B}, \tag{9}$$

where T_0 is the value of T computed from the observed samples (i.e. the observed difference in b-value).

To test the difference in b^{RM}, we followed A.F. Siegel's recommendation (Siegel, personal communication, 2007 November 27) of a bootstrap t-test and formed the statistics

$$t = \frac{b_{\mathrm{A}}^{\mathrm{RM}} - b_{\mathrm{B}}^{\mathrm{RM}}}{\sqrt{SE_{\mathrm{A}}^2 + SE_{\mathrm{B}}^2}}, \tag{10}$$

where SE_{A} and SE_{B} are the bootstrap-derived standard errors for each b^{RM} value. If $|t| > 1.96$, then the difference is significant at the 5 per cent level.

4 TESTS ON SYNTHETIC SEISMIC CATALOGUES

We compared the performance of b-value calculation techniques and of the difference test procedures on synthetic seismic catalogues. The complete part of the catalogue was modeled by the G–R relation and the incomplete part of the FMD was modeled by a normal cumulative distribution function

$$F(M|\mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{m_0} \exp(-\frac{(M-\mu)^2}{2\sigma^2})\mathrm{d}M. \tag{11}$$

In this equation, μ denotes the magnitude at which 50 per cent of the earthquakes are detected and σ is the standard deviation describing the magnitude range of the incomplete part of the FMD. This choice was not based on physical reasoning but instead based on 'shape likeness' used by Ogata & Katsura (1993) and Woessner & Wiemer (2005). After generating the synthetic catalogues, the magnitude values were sorted with a fixed bin width of 0.1 to mimic the discrete magnitude values in a real catalogue.

Synthetic seismic catalogues were used to investigate only random fluctuations and to eliminate effects that may be due to spatial and/or temporal variations in the natural seismicity. Two synthetic seismic catalogues were generated (Fig. 1), each including one and a half million events with 90 per cent of them above $M = 0.98$. Theoretical $b = 1.0$ and 0.7 FMDs were simulated, respectively.

4.1 Comparison of the performance of *b*-value calculation techniques

Our regression technique has not been previously used in b-value estimation; therefore, initial tests were concerned with the ability of the repeated medians regression to compute correct values of b and the associated uncertainties. The opportunity was also taken to check the accuracy of the standard error estimates for maximum likelihood b-values. From each of our two one and a half million events catalogues, we randomly extracted 10 000 different seismic catalogues. The sizes of these 10 000 samples (i.e. 'subcatalogues') were trimmed to get the number of data above or equal to the threshold magnitude uniformly ranging from 50 to 2500 events. These size limits were not arbitrary because various studies adopted 50 as the minimum number of events for stable b-value results (Murru *et al.* 1999; Gerstenberger *et al.* 2001; Schorlemmer *et al.* 2003; Bridges & Gao 2006). In the same studies, it was rare when regional b-values were computed with more than 2500 events. For each of the 10 000 catalogues, m_0 (threshold magnitude) was determined from 1000 bootstrap replicates of each FMD, following Amorese's procedure (2007). Then, we calculated b^{AU}, b^{RM} and their error estimates

(i) SE_{SB} for b^{AU} (eq. 3).
(ii) SE_{AU}, the bootstrap standard error from 1000 replicates for b^{AU}
(iii) SE_{RM}, the bootstrap standard error from 1000 replicates for b^{RM}.

To investigate the bias in the estimated b-value, in Fig. 2, we reported the different b-values as a function of the number of data above or equal to the threshold magnitude, for the cases $b = 1$ and 0.7. Fig. 2 shows that b^{AU} values are quite close to the theoretical b-values whereas b^{RM} and especially b^{LS} values seem to underestimate true b-values (Fig. 2). We find that b^{RM} values appear to be reliable estimates only when N is larger than 600. This is the reason why in the following, the repeated median regression is only applied on the largest data sets. The repeated median regression is certainly not the most efficient technique for computing the b-value. Anyhow, this is always a better method than the least-squares regression. Under favourable conditions (large N), the repeated median regression has the advantage of being an efficient alternative approach to the Aki–Utsu equation.

The precision of results was investigated by plotting the relationships between the different b-values standard errors as a function of the number of events above or equal to the threshold magnitude (Fig. 3). As expected, in each case, the standard error gets smaller as the sample size gets larger (Fig. 3). An interesting observation is that standard errors that are derived from the Shi and Bolt equation (eq. 3) (Shi & Bolt 1982) seem to underestimate true standard errors: in Fig. 3, root mean square error values (taken to represent true standard errors) are systematically larger than the Shi and Bolt's estimates for standard errors. In contrast, root mean square error values and bootstrap standard errors show a good agreement (Fig. 3).

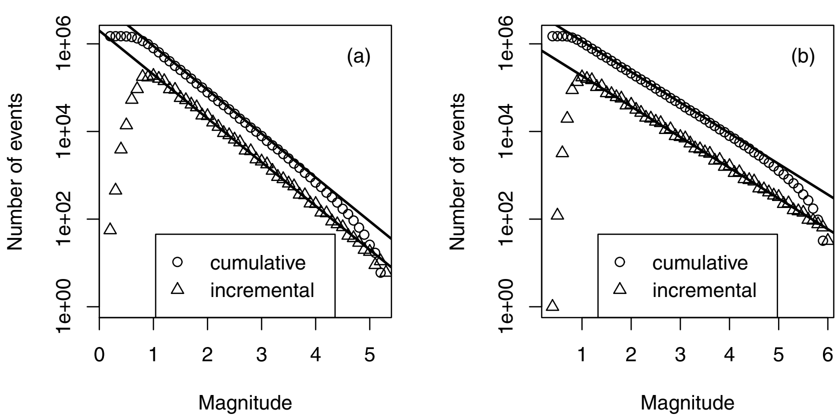

Figure 1. Synthetic frequency magnitude distributions, for the cases $b = 1$ (a) and $b = 0.7$ (b).

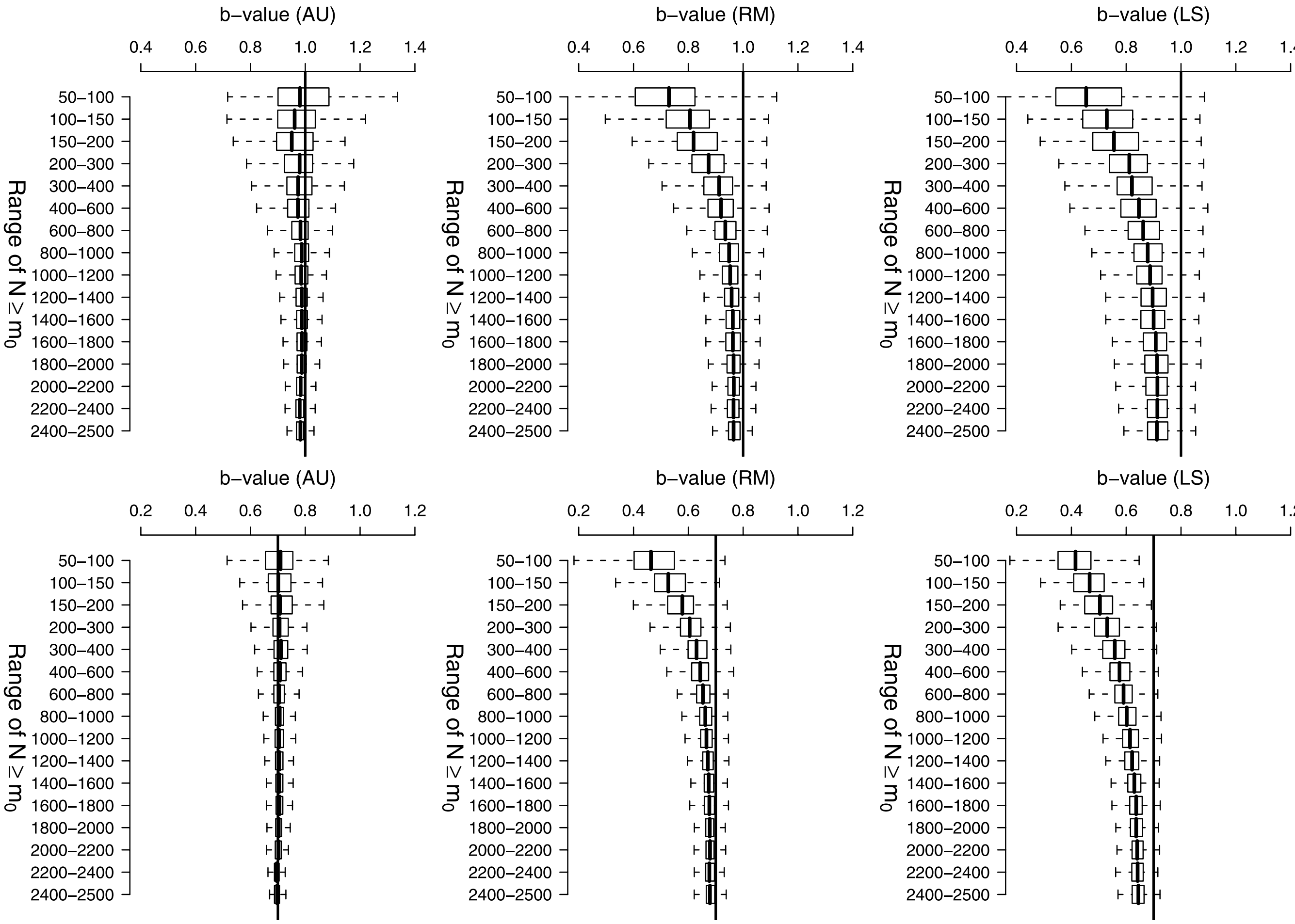

Figure 2. Plots of b-value versus number of data above or equal to the threshold magnitude for the Monte-Carlo simulation of various b-value estimation methods, for the cases true $b = 1$ (top) and true $b = 0.7$ (bottom). Three fitting techniques are compared: the Aki–Utsu equation (AU), the repeated median regression (RM) and the orthogonal least squares regression (LS). Each box-and-whisker plot includes at least 140 simulations. Into each box, the horizontal line shows the median. The bottom and top of the box show the 25th and 75th percentiles, respectively. The whiskers (dashed lines) show the range of the data, excluding outliers. An outlier is any value that lies more than 1.5 times the interquartile range (length of the box) from either end of the box. Numbers of data above or equal to the threshold magnitude are grouped into 16 class intervals.

4.2 Performance of tests for determining the statistical significance of a difference between *b*-values

Numerical simulations were also used to compare p-values associated with the various significance tests. Our comparative experiment was performed as follows:

(i) From our synthetic catalogue of 1 500 000 events with a b-value of 1.0, we randomly extracted two samples of size N_1 and N_2.

(ii) For each sample, the b-value was computed both through the Aki–Utsu formula and the repeated median regression.

(iii) Type I error (namely, rejecting null hypothesis when it is true) p-values of the difference in b-value (two-sided hypothesis) between the two earthquake samples were calculated from the two-sample bootstrap test (comparison of b^{AU} values) and the (ΔAIC) Utsu's test. Calculations were performed on 1000 replicates.

(iv) The procedure was repeated 10 000 times.

We performed two kinds of runs: $N_1 = N_2 = 1000$ and $N_1 = N_2 = 300$. Results are shown in Fig. 4. In all cases, Utsu's test p-values are never larger than 0.4; this surprisingly small value is inherent in eq. (8): when b-values are equal, it can be shown that $p = \exp(-1) = 0.37$. Utsu's test p-values are overall smaller than bootstrap test p-values (Figs 4a and b). Moreover, whereas data samples comes from the same artificial population, the Utsu's test always shows a high rate of rejection of the null hypothesis. Actually, few small p-values are expected as the null hypothesis is actually 'true': the two compared samples are coming from one 'population' only. For the Utsu's test about 12 per cent of the p-values are smaller than 0.05 (Figs 4a and b) and about 3 or 4 per cent of the p-values are smaller than 0.01. The Utsu's test is biased, whereas the bootstrap test is apparently correct, for example, for the bootstrap test, more than 90 per cent of the p-values are larger than 0.1, more than 95 per cent of the p-values are larger than 0.05 (Figs 4a and b).

5 DATA

The study that was made by Gerstenberger *et al.* (2001) is one of the fundamental works about the variation of b-value with depth. Their study partly inspired us. Whereas these authors investigated both Southern and Northern California, our study was only restricted to Southern California because this area, according to Gerstenberger

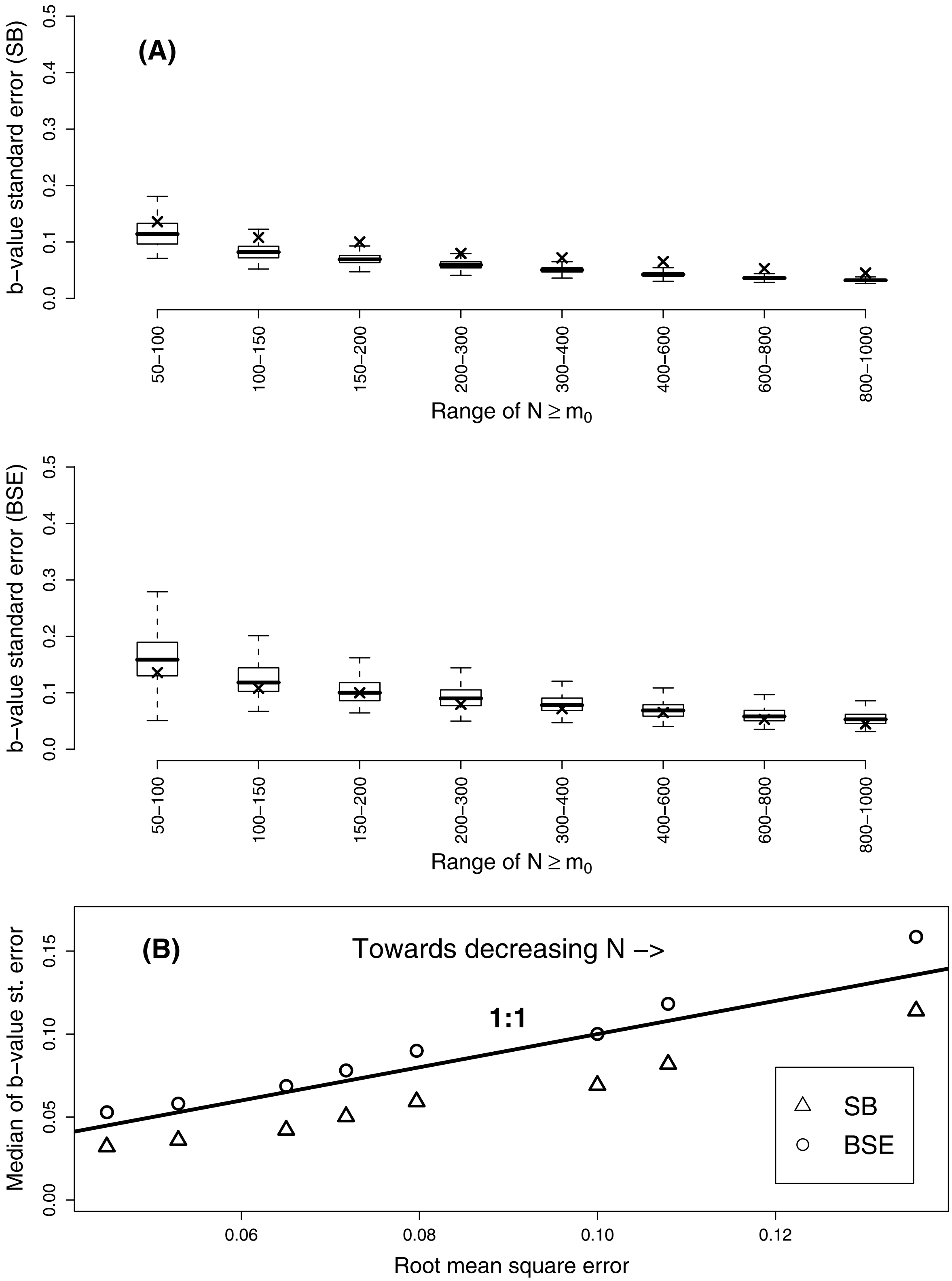

Figure 3. Comparisons of b-value standard errors. (a) Plots of b-value standard error versus number of data above or equal to the threshold magnitude for the Monte-Carlo simulation of b-value estimation methods, for the case true $b = 1$. Two computation techniques are compared : the standard error for the Aki–Utsu b-value estimate following Shi and Bolt (Shi & Bolt 1982) (SB) and the bootstrap standard error for the Aki–Utsu b-value estimate (BSE). Each box-and-whisker plot includes at least 140 simulations. Into each box, the horizontal line shows the median. The bottom and top of the box show the 25th and 75th percentiles, respectively. The whiskers (dashed lines) show the range of the data, excluding outliers. An outlier is any value that lies more than 1.5 times the interquartile range (length of the box) from either end of the box. In each N interval, the cross represents the root mean square error (RMSE), as defined as $RMSE = \sqrt{\frac{1}{n}\sum_i^n(\hat{b} - 1.0)^2}$, where $\hat{b}$ is the estimated b-value and n is the number of simulations in the interval. To keep the figure clear, only eight class intervals for N are plotted. (b) Plot of medians of SB and BSE versus RMSE.

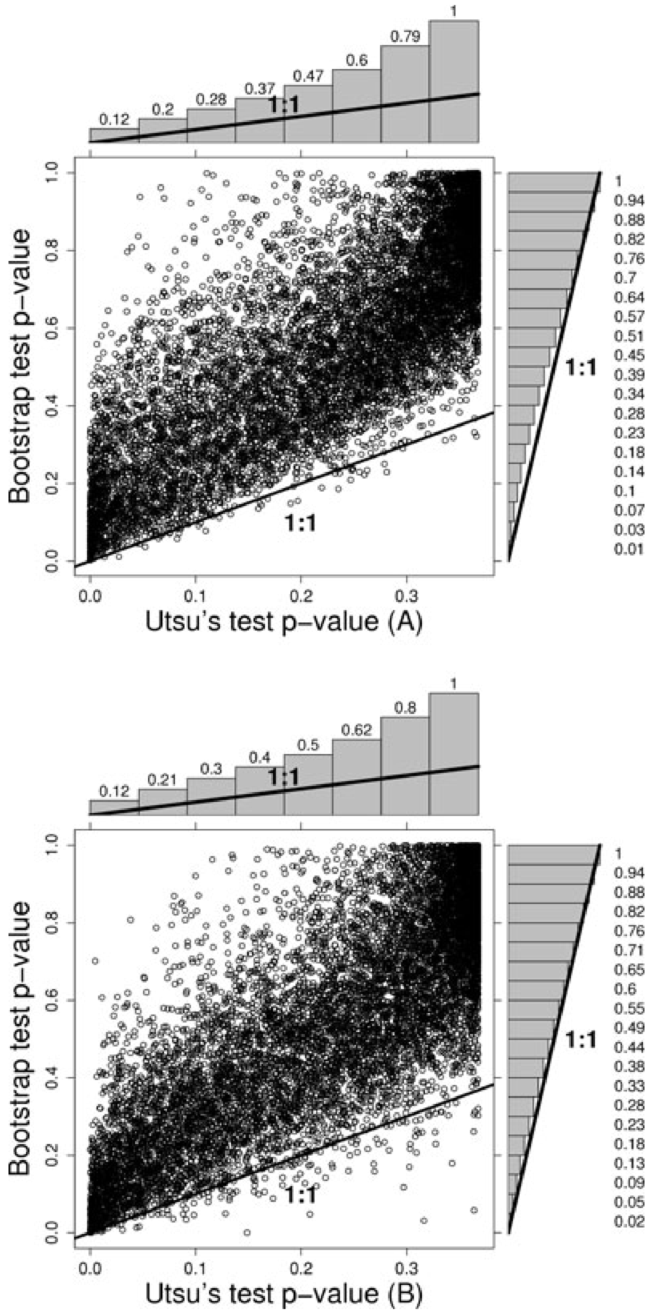

Figure 4. Scatterplots with marginal histograms showing the relationship between the two-sample bootstrap (comparison of b^{AU} values) and the (ΔAIC) Utsu's tests' p-values. Reported p-values are type I error p-values for 10 000 pairs of samples being compared in each graph. The line 1:1 is shown for reference. (a) The size of both samples is 1000. (b) The size of both samples is 300.

et al.'s (2001) Plate 1, showed 'wide' zones of constant or almost constant r_b value.

Gerstenberger *et al.* (2001) showed that while individual areas with high r_b values may exist, they may be difficult to detect when a b-value was calculated for a large area. Therefore, they used sampling cylinders with radii of 40 km or less to map r_b values. Though r_b mapping was not the aim of our study, we used sampling volumes with roughly the same size as Gerstenberger *et al.*'s cylinder radii: we investigated seven quadrangular areas showing a minimum and a maximum side of 28 and 67 km, respectively. These seven subareas both showed sufficient density of seismic events and $r_b > 1$, according to Gerstenberger *et al.*'s (2001, Plate 1b). In our study, the choice of exactly demarcated areas (their limits are given in Table 1 and shown in Fig. 5) for the computation of b-values guaranteed the reproducibility of the calculations. Our investigations were performed on high-quality data provided by the relocated Southern California catalogue (Richards-Dinger & Shearer 2000), which is practically the same data set from Richards-Dinger & Shearer (2000) (http://www.data.scec.org/ftp/catalogs/dinger-shearer) that was used by Gerstenberger *et al.* (2001): the time interval analyzed spanned from 1981.3 (April 1981) to 1998.2 (March 1998). Because we were interested in natural seismicity, we used only nighttime events by discarding all events between 16:00 and 3:00 GMT from our analysis. This daytime period was also adopted by Gerstenberger *et al.* (2001). Nevertheless, we note that another choice was possible for Southern California: Richards-Dinger & Shearer (2000) discarded daytime events between 7:00 and 19:00 PST (14:00–2:00 GMT).

Using a small number of events can lower the robustness of the calculations (Howell 1985; Pickering *et al.* 1995; Mori & Abercrombie 1997). This fact is illustrated in Fig. 2. Therefore, we decided to compute b-values from data sets that included at least about 200 events (Table 1). According to results from Section 4.1, the repeated median regression was used only for data sets that included at least about 600 events (Idyllwild, Salton City; Table 1). Following Gerstenberger *et al.*, for each zone, we selected events in depth ranges of 0–5 and 8–15 km. According to Richards-Dinger & Shearer (2000), median vertical location uncertainties (standard errors) for relocated events are about 740 m, which is much smaller than the 3 km depth gap between the 0–5 km and the 8–15 km depth ranges. By using the highest quality catalogue available, we were somewhat assured that uncertainties in hypocentral locations were not responsible for apparent variations of the b-value with depth. Information on the data samples used in the analysis are presented in Table 1. And because the studied seismic regions are relatively small, a significant spatial variation in completeness in each quadrangle is not expected. In each quadrangle, completeness was calculated following the Amorese's procedure (2007). We did not directly take into account possible temporal variations of the completeness magnitude. Nevertheless, these variations (http://completeness.usc.edu) are always smaller than uncertainties in m_0 values (Table 2).

The definition of aftershocks notwithstanding (De Rubeis *et al.* 2007), many authors (Oncel & Alptekin 1999; Knopoff 2000; Chan & Chandler 2001) suggested the use of declustered catalogues in the computation of local b-values. Gerstenberger *et al.* (2001) did not declustered their data before they calculated b-values. In our study, four (Santa Paula, Glendale, Salton City, Imperial, Fig. 5) of our seven areas under investigation may be affected by possible contamination from the aftershocks of larger events. Following Mori & Abercrombie (1997), we did suspect that contamination by clusters was a key issue. Therefore, to create an additional 'declustered' data set for these four areas, we adopted the same kind of declustering procedure as used by Mori & Abercrombie (1997). It consists in separating out the larger aftershock sequences for earthquakes of magnitude about 6.0 and above. Thus, we removed events that were located within 30 km radius around the Whittier Narrows (1987 October 1, 34.1°N, 118.1°W, M_L5.9), the Elmore Ranch/Superstition Hills (November 24, 1987, 33°N, 115.8°W, M_w 6.1 and 6.6) and the Northridge (1994 January 17, 34.2°N, 118.5°W, M_w 6.7) mainshocks. Actually, the 30 km threshold roughly corresponds to one

Table 1. Information on the data samples used in the analysis.

Area name	Area limits	Depth range (km)	Sample size	Magnitude range	Nearby large events
Arvin	35.2°–35.6°N	0–5	1250	0.1–3.7	
	118.9°–118.5°W	8–15	288	0.2–3.4	
Santa Paula	34.3°–34.6°N	0–5	689 [540]	0.3–5.2 [0.3–3.4]	Northridge
	119.2°–118.6°W	8–15	1666 [1491]	0.4–5.1 [0.4–5.1]	
Glendale	33.9°–34.2°N	0–5	272 [268]	0.9–3.6 [0.9–3.6]	Whittier Narrows
	118.6°–118.0°W	8–15	453 [293]	0.7–5.9 [0.7–5.9]	Northridge
Santa Ana	33.6°–34.0°N	0–5	507	0.7–3.1	
	118.0°–117.5°W	8–15	193	0.8–3.5	
Idyllwild	33.4°–33.8°N	0–5	1307	0.9–3.3	
	117.2°–116.7°W	8–15	1214	0.9–3.9	
Salton City	33.1°–33.4°N	0–5	1753 [1750]	0.2–4.1 [0.2–4.1]	Elmore Ranch/
	116.2°–115.9°W	8–15	900 [900]	0.1–4.4 [0.1–4.4]	Superstition Hills
Imperial	32.8°–33.4°N	0–5	5301 [4239]	0.2–6.6 [0.2–5.7]	Elmore Ranch/
	115.9°–115.5°W	8–15	583 [525]	0.2–4.7 [0.2–4.1]	Superstition Hills

Note: Bracketed values are for 'declustered' samples.

fault length distance to mainshock for a M6.7 event (Wells & Coppersmith 1994). Whereas Mori and Abercrombie removed events for 1 year following the date of each large mainshock, we removed events for only 12 days after each occurence. This smaller time interval results in the same kind of temporal pattern in Schuster's diagrams (Figs 6b and c) (Rydelek & Hass 1994) that an entire year removal. Thus, some large 'pig tail' features (Fig. 6a) were removed without reducing data sizes too far. Obviously, after removals, some curly patterns remain in Schuster's diagrams (Figs 6b–d) showing that our declustering procedure is certainly not perfect. Indeed, in our declustering procedure, we do not remove aftershocks that are linked to smaller mainshocks ($M < 6$). A perfect declustering is beyond the scope of our study. We did not use declustering algorithms (Reasenberg & Jones 1989; Frohlich & Davis 1993) because they remove the smaller aftershocks, resulting in systematically lower b-values (Mori & Abercrombie 1997). After this coarse declustering, we were left with 65 113 and 34 492 night-time events for Southern California (32.6–36°N; 119.5–115°W) for the depth ranges 0-5 and 8-15 km, respectively. The decluster procedure removed only about 2.6 per cent of the total number of seismic events (2619 'aftershocks') but 43 per cent of these aftershocks were located in one of our study areas (Imperial area). As stated in the following section, we confirmed our initial suspicion and found that the clustering of earthquakes in our data did substantially alter our results. Actually, aftershocks do not show a different magnitude-frequency distribution than any other set of earthquakes but they happen so rapidly that they overwhelm the detection system. This results in a higher completeness threshold and then leads to a smaller b-value than the rest of the data set (Helmstetter *et al.* 2005).

6 RESULTS

6.1 Arvin area

In this area, the top data sample includes 1250 earthquakes, whereas the bottom data sample shows only 288 events. The Aki–Utsu equation provides $b = 0.90$ and 0.70 for the top and bottom zone, respectively (Table 2 and Fig. 7). The Utsu's test suggests that the probability that the two depth zones have the same b-values can be rejected at a confidence limit in excess of 99 per cent (Table 2). The two-sample bootstrap test for the comparison of b^{AU} values fails to reject the null hypothesis ($p = 0.16$, Table 2). According to the results of Section 4.2, we reject any significant changes of b-value with depth in the Arvin area.

6.2 Santa Paula

The b-bottom value (0.77) is smaller than the b-top value (0.97). The null hypothesis can be rejected at the 1 per cent significance level by any of the test statistics (Table 2). FMDs are shown in Fig. 7.

Table 1 shows that both the top and the bottom zones can be significantly influenced by possible aftershocks (22 and 10 per cent of the seismicity, respectively). Actually, the raw data samples resulted in lower b-values for both the bottom ($b = 0.68$) and top data sets ($b = 0.72$). In this case, this did radically change the results from the comparison tests, that is, the b-values are no more significantly different from each other ($p = 0.32$ for the two-sample bootstrap test). Because we have chosen to draw our conclusions from declustered data's results, we suggest that the b-value is decreasing with depth in the Santa Paula area. However, one should bear in mind that this inference is not free from the declustering procedure.

6.3 Glendale area

Very few events are available to compute b-values in the depth ranges 0–5 and 8–15 km in the Glendale area (Table 1); the bottom zone b-value is calculated with only 140 magnitude values (Table 2). The two FMDs are seen to be identical in their low magnitude parts (Fig. 7). The FMD for the 8–15 km depth sample shows a tail in the higher magnitude range. This feature is due to a M5 event that occurred on 1988 December 3 in the epicentral area of the Whittier Narrows mainshock, thus suggesting that our declustering procedure is possibly not adequate. We are facing the data quality-versus-quantity dilemma: too strict of a declustering procedure would impede stable calculations. Therefore, we choose to keep this M5 event and probably several other aftershocks in our data sets. The results from the three tests are consistent with the hypothesis that the bottom b-value is not significantly smaller

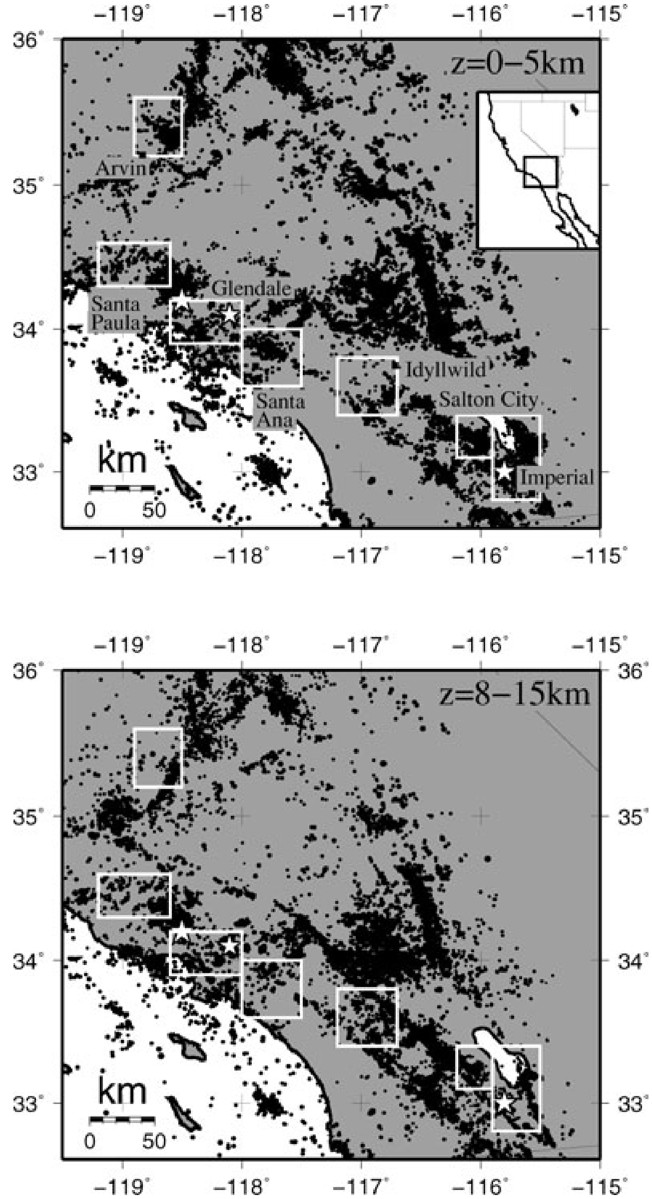

Figure 5. Southern California map showing the seismicity from the relocated Southern California catalogue (Richards-Dinger & Shearer 2000) and the seven investigated areas of this study. Only nighttime events of the 'declustered' catalogue are plotted (all events between the hours of 16:00 and 3:00 GMT have been removed) from 1981.3 to 1998.2 in the depth ranges (0–5 km) and (8–15 km). White stars locate the epicenters of the Whitthier Narrows, Elmore Ranch/Superstition Hills and Northridge earthquakes.

than or different from the top b-value (Tables 2 and 3). The seismicity in the Glendale area shows that clustering of earthquakes can substantially alter b-value comparison results; if computations are performed with the raw data sets, b-top and b-bottom values are 1.13 and 0.71, respectively and both the Utsu's ΔAIC and the two-sample bootstrap tests agree with the hypothesis of a decreasing b-value with depth, with p-values smaller than 10^{-2}. Actually, in this case, the comparison is mostly between seismic events in the 0–5 km depth range and aftershocks of the Whittier Narrows earthquake in 8–15 km depths where 35 per cent of the seismic events can be considered as possible aftershocks (Table 1). From the declustered data sets, we conclude that the b-value is constant with depth in the Glendale area.

6.4 Santa Ana area

The Santa Ana area also includes relatively few seismic events and the bottom b-values are calculated with less than 150 magnitude values (Table 2). In this region, the two Utsu's tests indicate rejection of the null hypothesis of b-value homogeneity, whereas the bootstrap test suggests that there is not enough evidence for rejection (Table 2). The bootstrap test for the comparison of b^{AU} values also fails to reject the null hypothesis at the 1 per cent significance level against the alternative hypothesis that the b-bottom value is smaller than the b-top value (Table 3). By looking at the FMDs (Fig. 7), we can surmise that the slopes are not accurately known. For the Santa Ana area, we offer that there is no reliable basis to conclude that b-values are varying with depth.

6.5 Idyllwild area

More than 700 magnitude values are used to compute b-values in each of the two depth intervals in the Idyllwild area (Table 2); therefore, the most reliable results are expected from these two data sets. As data sets are large, tests on repeated median b-values are possible and can be convincing. Once again, as for the Glendale area, the null hypothesis cannot be rejected at the 1 per cent significance level by any of the test statistics (Table 2). Even the bootstrap one-sided tests cannot reject the null hypothesis in favour of the alternative that 'the b-bottom value is smaller than the b-top value' in the Idyllwild area (Table 3). In reality, there is not enough evidence to reject the null hypothesis at even the 5 per cent level of significance. The two FMDs overlap almost perfectly up to $M = 2$ (Fig. 7). For the Idyllwild area, we therefore infer that the b-value is constant with depth.

6.6 Salton City area

The Salton City area also has a relatively large number of data used to calculate b-values: 1281 and 575 events in the top and the bottom zones, respectively. Under these conditions, it is reasonable to presume that b-values and their uncertainties are reliably estimated. In the Salton City area, the bottom zone b-value is apparently larger than the top zone b-value but none of the statistical tests suggest a significant difference in b-values (Table 2). Of course, when the test is one-sided (alternate hypothesis is 'b-bottom is smaller than b-top'), a very high p-value is obtained (Table 3). By looking at the FMDs (Fig. 7), it appears that the data are fitted as well by lines with b^{AU} slopes, as by lines with b^{RM} slopes. Moreover, the figure clearly illustrates that the lines are almost parallel. Once again, our results tend to disagree with those of Gerstenberger *et al.* in Plate 1 (Gerstenberger *et al.* 2001).

6.7 Imperial area

The variation of the number of earthquakes with depth is remarkable in the Imperial area: whereas b-top values are calculated from a relatively large data set (3162 magnitude values), b-bottom values are estimated from less than 400 events (Table 2). The results from the Utsu's tests are consistent with the hypothesis that the bottom b-value is different (Table 2) from the top b-value. Nevertheless, a different conclusion is obtained from the bootstrap test (Table 2) or when raw data sets are used for the b-value computations. From raw data sets, the b-value of the bottom zone is 0.86, whereas the b-value of the top zone is 0.94. In that case, all the comparison

Table 2. Results summary.

Area	z range (km)	m_0[a]	N[b]	b^{AU}[c]	b^{RM}[d]	Two-sided test p-value: Ut.1[e]	Ut.2[f]	boot.1[g]	boot.2[h]
Arvin	0–5	1.1[0.9–1.4]	798	0.90 ± 0.10	–	1.5×10^{-3}	4.3×10^{-4}	0.16	–
	8–15	1.0[0.8–1.3]	210	0.70 ± 0.13	–				
Santa Paula	0–5	1.2[1.1–1.5]	366	0.97 ± 0.08		3×10^{-4}	8.4×10^{-5}	4×10^{-3}	–
	8–15	1.1[1.1–1.1]	1161	0.77 ± 0.02	–				
Glendale	0–5	1.6[1.4–1.8]	172	1.13 ± 0.16	–	0.06	0.03	0.23	–
	8–15	1.8[1.4–1.8]	140	0.91 ± 0.11	–				
Santa Ana	0–5	1.5[1.4–1.8]	312	1.12 ± 0.14	–	2.6×10^{-3}	8.3×10^{-4}	0.10	–
	8–15	1.4[1.1–1.8]	142	0.82 ± 0.15	–				
Idyllwild	0–5	1.1[1.1–1.1]	774	1.02 ± 0.04	1.06 ± 0.08	0.34	0.35	0.82	0.55
	8–15	1.1[0.8–1.4]	747	1.00 ± 0.11	0.97 ± 0.13				
Salton City	0–5	1.1[1.0–1.4]	1281	1.00 ± 0.08	0.99 ± 0.11	0.27	0.22	0.58	0.67
	8–15	1.1[1.1–1.3]	575	1.04 ± 0.06	1.07 ± 0.16				
Imperial	0–5	1.4[1.4–1.6]	3162	1.01 ± 0.05	–	7.7×10^{-3}	2.6×10^{-3}	0.09	–
	8–15	1.4[1.3–1.6]	373	0.87 ± 0.05	–				

[a]Threshold magnitude for the computation of b-values. Bracketed values are respectively the 5th and 95th percentiles of the bootstrap distribution of m_0.
[b]Number of data above or equal to the threshold magnitude.
[c]Aki–Utsu b-value.
[d]Repeated median b-value.
[e]Utsu's ΔAIC test.
[f]Utsu's F-test.
[g]Two-sample bootstrap test.
[h]Two-sample bootstrap t-test.

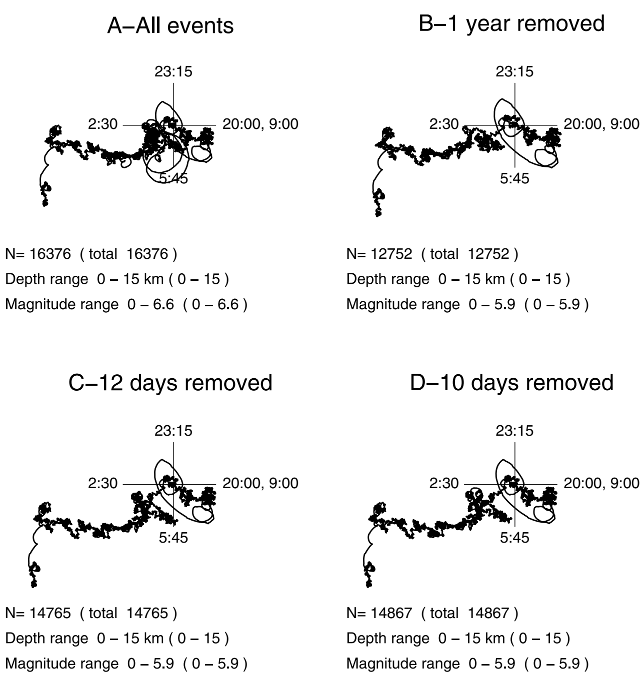

Figure 6. Schuster's diagrams for the seismicity of the seven investigated areas. Only nighttime events are summed. Aftershock sequences are suggested to induce 'pig tail' features in phasor walks. In plots b, c and d, events that were located within 30 km radius around largest mainshocks ($M \geq 6$) have been removed. The removing time interval after each mainshock is respectively 1 year, 12 days and 10 days long for plots b, c and d.

tests fail to reject the null hypothesis. The clustering of earthquakes alters the b-value in the same direction as for the Santa Paula and the Glendale areas (decrease) but for the Imperial area this bias does not allow the rejection of the null hypothesis because aftershocks are mostly contaminating the top part (20 per cent of the seismic events; Table 1). On the basis of these observations, we tend to conclude that there is no significant decrease in the b-value as a function of depth in the Imperial area.

7 DISCUSSION

There appears to be a distinct low b-value in the depth range 8–15 km in the Santa Paula area. Other results given here, however, indicate that the decrease of b-value with depth in Southern California is a less common phenomenon than previously believed. Considering that the seven investigated subzones cover about 12 700 km^2, the percentage of the area where the b-value is significantly decreasing with depth is 14 per cent. This value takes into consideration that the b-value is varying in the Santa Paula area (1815 km^2). This results is far from the 34.4 per cent of the sample area in the reference study by Gerstenberger *et al.* (2001). Of course, there is a problem in the geographical representation of our results: the seven studied subzones cover only about one fifth (Fig. 5) of the area sampled by Gerstenberger *et al.* in Southern California (Plate 1b). Nevertheless, one should keep in mind that the seven studied subzones are supposed to be homogeneous regions where there is a significant decrease in b-value with increasing depth, according to Plate 1b in Gerstenberger *et al.*. Therefore, our results are clearly at odds with the results of Gerstenberger *et al.*

Our results of the variability of b-value with depth also disagree with the study of Mori & Abercrombie (1997), where a somewhat

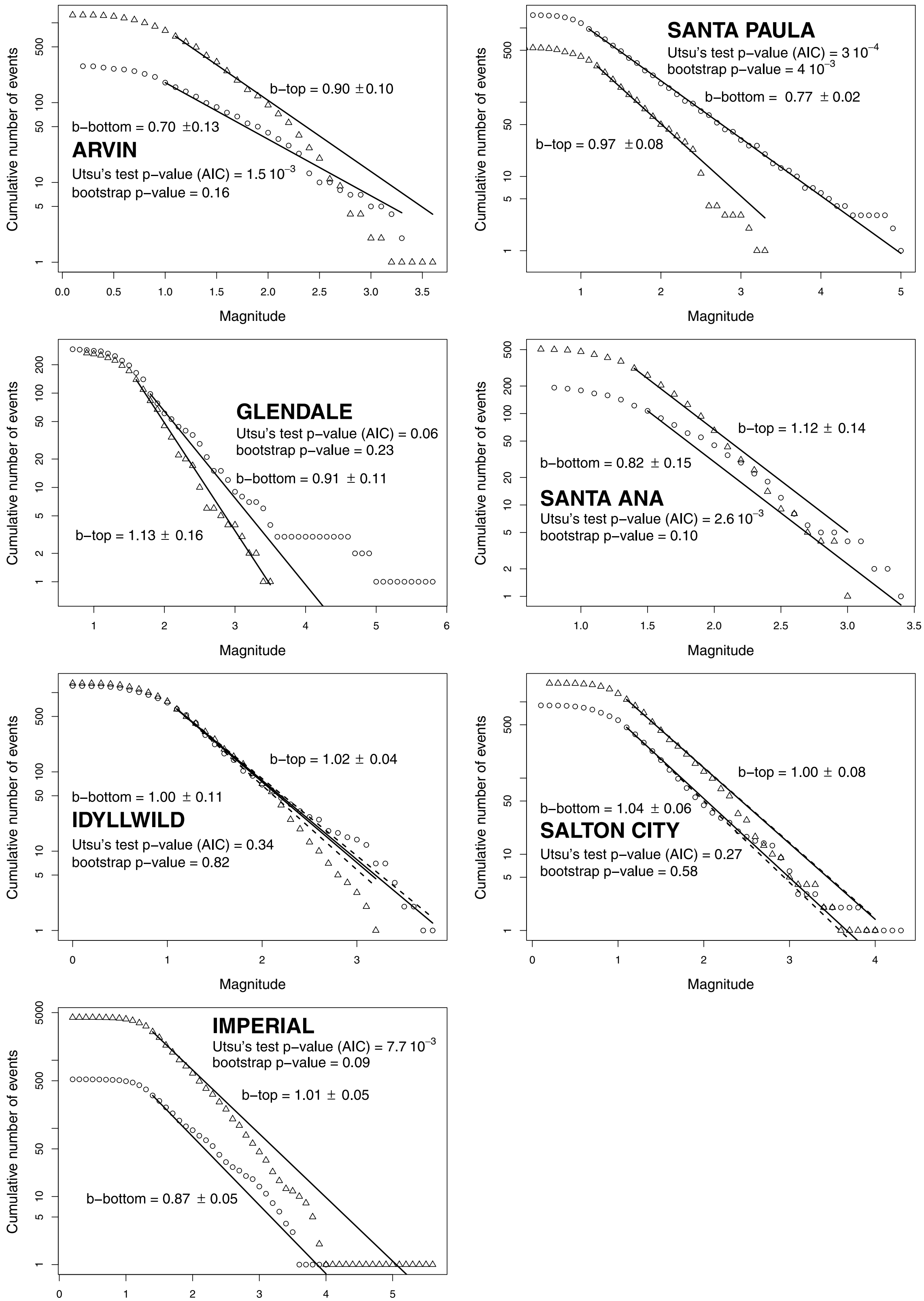

Figure 7. Frequency magnitude distributions (FMDs) of top 0–5 km depth range (triangles) and bottom 8–15 km depth range (circles) earthquakes for the investigated areas. Solid lines show the maximum likelihood estimates of the *b*-value. Dashed lines show the repeated median estimates of the *b*-value. To facilitate comparison, the *a*-value is fitted to force lines to pass through a common point, which is the $[m_0, N(m_0)]$ point of each FMD.

Table 3. Results summary.

Area name	One-sided test p-value	
	boot.1	boot.2
Arvin	0.08	–
Santa Paula	$2\ 10^{-3}$	–
Glendale	0.11	–
Santa Ana	0.05	–
Idyllwild	0.40	0.27
Salton City	0.69	0.67
Imperial	0.04	–

Note: The alternative hypothesis is that *b*-top is greater than *b*-bottom.

different approach than Gerstenberger *et al.* was used. Whereas Gerstenberger *et al.* only required a minimum of 50 events with magnitude greater than *Mc* for their computation of *b*-value in each (relatively small) zone, Mori and Abercrombie considered larger regions of California. Mori and Abercrombie were aware that ample data were required for statistically reliable results, while Gerstenberger *et al.* suggested that individual areas of *b*-value change may be difficult to detect when a single *b*-value was calculated for a large area. In this respect, we took an in-between approach: small areas and *b*-values were always computed with at least 140 magnitude values in each area. For a data set from Southern California, Mori and Abercrombie reported a collective *b*-top value ($z = 0$–3 km) equal to 1.29 ± 0.01 and a *b*-bottom value ($z = 12$–15 km) equal to 0.97 ± 0.02 (Table 1 Mori & Abercrombie 1997), with reported uncertainties corresponding to 95 per cent confidence limits. They had removed aftershocks from their catalogue and used a minimum magnitude of 2.0. We did the same calculations on our reference data set for the area within 32.6–36°N latitude and 115–119.5°W longitude. We found no significant difference between the top *b*-value (0.92 ± 0.03) and the bottom *b*-value (0.87 ± 0.02), with reported uncertainties obtained from the bootstrap *b*-value standard errors. Thus, we agree with Gerstenberger *et al.*'s statement that *b*-value variations may be difficult to detect when a collective *b*-value is calculated for a large area. For Southern California, Gerstenberger *et al.* found 1.08 ± 0.13 and 0.90 ± 0.12 for the *b*-top and the *b*-bottom, respectively. All these results are different (Table 4), which is not unexpected given the different data sets and diversity of the calculation parameters. For information purposes, our *b*-top value was calculated from a data set including 65 113 seismic events, using a threshold magnitude $m_0 = 1.6$ and our *b*-bottom value was calculated from a set of 34 492 events, using $m_0 = 1.4$.

Several factors may affect the comparison of *b*-values. The first is an erroneous estimation of FMD slopes. We have checked the ability of two independent computation methods to estimate *b*-values (Fig. 2). Because aftershocks are known to affect the calculation of background *b*-values, we used both clustered and declustered data sets. In addition, we have taken great care to plot each FMD that we have studied (Fig. 7). No significant departure from the linear G–R relation was observed in any of these plots, which is fortunate because our analysis does not take into account the possible nonlinearity of the FMDs (Aki 1987; Knopoff 2000). A key issue is how the uncertainties in magnitude measurements may propagate into the estimates of *b*-value (Marzocchi & Sandri 2003; Del Pezzo *et al.* 2003), which was not addressed in our study. However, regardless of this possible additional source of bias, we have no compelling reasons to suppose that our *b*-value calculations are overly corrupted and that reliable comparisons are impossible to perform. The same argument may probably apply to *b*-values computed by Gerstenberger *et al.*

The second factor that can affect comparisons lies in methods of comparisons. One of the most common ways to assess differences in *b*-values is to rely on the (ΔAIC) Utsu's test *p*-value. Converging evidence from simulations (Fig. 4) and examples (Tables 2 and 3) indicated that the Utsu's comparison tests tend to favour rejection of the null hypothesis of no change in *b*-values. Thus, the decrease in *b*-value with increasing depth reported by Gerstenberger *et al.* could be explained by random fluctuations in the errors in *b*-value calculations. In other words, the simplest explanation for an apparent decreases of *b* with depth in many places of Southern California may not be related to depth-dependent earthquake physics.

Our results do not lend support of the view that a decrease in *b*-value with increasing depth is a widespread phenomenon in Southern California; however, we observe an anomaly in the Santa Paula area that indicates there are relatively more smaller events at the shallowest depths. A simple interpretation of this observation in terms of mechanical variations in the crust may suggest a more homogeneous stress field and fewer fractures at depth (Mogi 1962; Scholz 1968; Mori & Abercrombie 1997) in this area.

To our best knowledge, this is the first time that the performance of Utsu's test has been thoroughly examined and that spatial variations in *b*-value have been investigated using various methods, that is, we have explored alternative means of determining *b*-values, uncertainties in *b*-values and the significance of the difference between *b*-values. The results of Gerstenberger *et al.*'s study has been partly compared here but our research suggests that other studies could have overestimated the degree of variation of the *b*-value, which may argue for a more rigorous analysis and careful interpretation of results.

8 CONCLUSIONS

Using bootstrap and non-parametric statistics, we investigated the variations in *b*-value with increasing depth in seven selected areas of Southern California. Statistically reliable results are mostly expected from the relatively large sample and high quality data obtained from the Southern California earthquake catalogue. In contrast with previous studies, we found that there was no firm evidence for the decrease of the *b*-value with depth in several areas of Southern California with the exception of a clear effect in the Santa Paula area. Thus, the change in the *b*-value as a function of depth appears to be less common than the literature suggests.

It is important to note that previous investigations were mostly based on the Utsu's test to infer the significance of differences in *b*-values. Here we have shown by numerical simulations that Utsu's test does not provide the correct Type I error rate and it is

Table 4. Collective *b*-values for the top (0–3 or 0–5 km) and bottom zones (12–15 or 8–15 km) in Southern California.

Depth range	Mori & Abercrombie (1997)	Gerstenberger *et al.* (2001)	This study
Top	1.29 ± 0.01	1.08 ± 0.13	0.92 ± 0.03
Bottom	0.97 ± 0.02	0.90 ± 0.12	0.87 ± 0.02

biased towards rejection of the null hypothesis; therefore, it may lead to erroneous conclusions and interpretations regarding the spatial variability of *b*-values and should be used with caution.

Regardless of possible errors involved in estimating earthquake magnitude, we have also confirmed results from Woessner & Wiemer (2005): by bootstrap analysis, it is possible to show that uncertainties in *b*-values may be underestimated by the usual formula in eq. (3) because it does not take into account the influence of the uncertainty associated with the determination of the threshold magnitude; therefore, the significance of spatial (or temporal) variations of the *b*-value may have been overestimated in many studies.

Finally, we submit that many previous results based on *b*-value variations, and especially those at borderline significance levels, should be viewed with caution simply because their authors did not acknowledge sources of errors in the calculation of *b*-values, among which the overwhelming of the detection system by clusters, is a key issue.

The results of our study suggest that the significance of spatial and/or temporal variations in *b*-value may be somewhat overestimated. These variations could be just localized anomalies, rather than a widespread phenomenon that would have important implications for crustal structure, stress modelling, and earthquake physics. This is consistent with the results of a detailed study of seismicity from the very dense and high-quality seismic networks located throughout Japan (Ishibe *et al.* 2008). Whereas it was found that *b*-value was dependent on focal mechanisms (normal faulting events had higher *b*-value than strike slip events), there was no indication of depth dependence in *b*-value for the shallow (declustered) seismicity in the Japan catalogue. It was further suggested that unknown biases in magnitude determination and/or in location may be factors in the apparent depth dependence of *b*-values.

We propose new quantitative methods that seismologists can use to compute *b*-values (repeated median regression technique) or test and evaluate the significance of changes in *b*-values (bootstrap tests). When the number of magnitude values is large enough, the repeated median regression provides a reliable estimate of the *b*-value. Bootstrap hypothesis testing or, to a lesser degree, non-parametric statistical procedures could be used as surrogate tools for improving the reliability and quality of crustal imaging via *b*-value variations.

ACKNOWLEDGMENTS

Maps were prepared with the GMT software package (Wessel & Smith 1991). Statistical calculations were performed using R (2.5.1 version) statistical programming environment (Ihaka & Gentleman 1996) available from http://www.r-project.org/ under the General Public License. DA is grateful to Prof. Andrew F. Siegel (Michael G. Foster School of Business, University of Washington, Seattle) for the information about the repeated medians regression. The authors thank Changjiang Wu for helpful comments for improving the manuscript. PAR is thankful for support from a Brain Pool grant of the Korea Research Foundation. The authors are grateful to reviewers and the associate editor for their helpful reviews of the paper.

REFERENCES

Aki, K., 1965. Maximum likelihood estimate of b in the formula log n=a-bm and its confidence limits, *Bull. Earthq. Res. Inst. Tokyo Univ.,* **43,** 237–239.

Aki, K., 1987. Magnitude-frequency relation for small earthquakes: a clue to theorigin of f_{max} of large earthquakes, *J. geophys. Res.,* **92,** 1349–1355.

Amelung, F. & King, G., 1997. Earthquake scaling laws for creeping and non-creeping faults, *Geophys. Res. Lett.,* **24,** 507–510.

Amorèse, D., 2007. Applying a change-point detection method on frequency-magnitude distributions, *Bull. seism. Soc. Am.,* doi:10.1785/0120060181.

Bender, B., 1983. Maximum likelihood estimation of b values for magnitude-grouped data, *Bull. seism. Soc. Am.,* **73,** 831–851.

Bridges, D.L. & Gao, S.S., 2006. Spatial variation of seismic *b*-values beneath Makushin Volcano, Unalaska Island, Alaska, *Earth planet. Sci. Lett.,* **245,** 408–415.

Chan, L.S. & Chandler, A.M., 2001. Spatial bias in *b*-value of the frequency-magnitude relation for the Hong Kong region, *J. Asian Earth Sci.,* **20,** 73–81.

De Rubeis, V., Loreto, V., Pietronero, L. & Tosi, P., 2007. Space-time combined correlation between earthquakes and a new, self-consistent definition of aftershocks, *Lect. Notes Phys.,* **705,** 259–279.

Del Pezzo, E., Bianco, F. & Saccorotti, G., 2003. Duration magnitude uncertainty due to seismic noise: inferences on the temporal pattern of G-R *b*-value at Mt. Vesuvius, Italy, *Bull. seism. Soc. Am.,* **93,** 1847–1853.

Efron, B., 1979. Bootstrap methods: another look at the jackknife, *Ann. Statis.,* **7,** 1–26.

Efron, B. & Tibshirani, R., 1993. *An Introduction to the Bootstrap,* Chapman and Hall, New York.

Fisher, R.A., 1934. *Statistical Methods for Research Workers,* 5th edn, Oliver and Boyd, Edinburgh.

Frohlich, C. & Davis, S.D., 1993. Teleseismic *b* values: or, much ado about 1.0, *J. geophys. Res.,* **98,** 631–644.

Gerstenberger, M., Wiemer, S. & Giardini, D., 2001. A systematic test of the hypothesis that the *b* value varies with depth in California, *Geophys. Res. Lett.,* **28,** 57–60.

Gutenberg, B. & Richter, F., 1944. Frequency of earthquakes in California, *Bull. seism. Soc. Am.,* **34,** 185–188.

Helmstetter, A., Kagan, Y.Y. & Jackson, D.D., 2005. Importance of small earthquakes for stress transfers and earthquake triggering, *J. geophys. Res.,* **110,** doi:10.1029/2004JB003286.

Howell, B.F.J., 1985. On the effect of too small a data base on earthquake frequency diagrams, *Bull. seism. Soc. Am.,* **75,** 1205–1207.

Ihaka, R. & Gentleman, R., 1996. R: A language for data analysis and graphics, *J. Comput. Graph. Stat.,* **5,** 299–314.

Ishibe, T., Tsuruoka, H. & Shimazaki, K., 2008. The dependency of the *b*-value on the focal mechanism and on the hypocentral depth, Japan Geoscience Union Meeting, Chiba, Japan, May 25–30.

Kagan, Y., 1999. The universality of the frequency-moment relationship, *Pure appl. Geophys.,* **155,** 537–574.

Knopoff, L., 2000. The magnitude distribution of declustered earthquakes in Southern California, *Proc. Natl. Acad. Sci. USA,* **97,** 11 880–11 884.

Main, I.G., Meredith, P.G. & Sammonds, P.R., 1992. Temporal variations in seismic event rate and *b*-values from stress corrosion constitutive laws, *Tectonophysics,* **211,** 233–246.

Marzocchi, W. & Sandri, L., 2003. A review and new insights on the estimation of the *b*-value and its uncertainty, *Ann. Geophys.,* **46,** 1271–1282.

McNutt, S.R., 2005. Volcanic seismology, *Annu. Rev. Earth planet. Sci.,* **32,** 461–491.

Mogi, K., 1962. Study of the elastic shocks caused by the fracture of heterogeneous materials and its relation to earthquake phenomena, *Bull. Earthq. Res. Inst. Tokyo Univ.,* **40,** 125–173.

Molchan, G., Kronrod, T. & Panza, G., 1997. Multi-scale seismicity model for seismic risk, *Bull. seism. Soc. Am.,* **87,** 1220–1229.

Mooney, C.Z. & Duval, R.D., 1993. *Bootstrapping: A Nonparametric Approach to Statistical Inference,* Sage Publications, Newbury Park.

Mori, J. & Abercrombie, R.E., 1997. Depth dependence of earthquake frequency-magnitude distributions in California: implications for the rupture initiation, *J. geophys. Res.,* **102,** 15 081–15 090.

Murru, M., Montuori, C., Wyss, M. & Privitera, E., 1999. The locations of magma chambers at Mt. Etna, Italy, mapped by *b*-values, *Geophys. Res. Lett.,* **26,** 2553–2556.

Murru, M., Console, R. & Lisi, A., 2004. Seismicity and mean magnitude variations correlated to the strongest earthquakes of the 1997 Umbria-Marche sequence (Central Italy), *J. geophys. Res.,* **109,** doi:10.1029/2002JB002276.

Murru, M., Console, R., Falcone, G., Montuori, C. & Sgroi, T., 2007. Spatial mapping of the *b* value at Mount Etna, Italy, using earthquake data recorded from 1999 to 2005, *J. geophys. Res.,* **112,** doi:10.1029/2006JB004791.

Ogata, Y. & Katsura, K., 1993. Analysis of temporal and spatial heterogeneity of magnitude frequency distribution inferred from earthquake catalogs, *Geophys. J. Int.,* **113,** 727–738.

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

Pickering, G., Bull, J.M. & Sanderson, D.J., 1995. Sampling power-law distributions, *Tectonophysics,* **248,** 1–20.

Power, J.A., Wyss, M. & Latchman, J.L., 1998. Spatial variations in frequency-magnitude distribution of earthquakes at Soufriere Hills volcano, Montserrat, West Indies, *Geophys. Res. Lett.,* **25,** 3653–3656.

Reasenberg, P.A. & Jones, L.M., 1989. Earthquake hazard after a mainshock in california, *Science,* **243,** 1173–1176.

Rhoades, D.A., 1996. Estimation of the Gutenberg-Richter relation allowing for individual earthquake magnitude uncertainties, *Tectonophysics,* **258,** 71–83.

Richards-Dinger, K. & Shearer, P., 2000. Earthquake locations in southern California obtained using source-specific station terms, *J. geophys. Res.,* **105,** doi:10 939–10 960.

Rydelek, P.A. & Hass, L., 1994. On estimating the amount of blasts in seismic catalogs with schuster's method, *Bull. seism. Soc. Am.,* **84,** 1256–1259.

Rydelek, P.A. & Sacks, I.S., 2003. Comment on 'minimum magnitude of completeness in earthquake catalogs: Examples from Alaska, the Western United States, and Japan,' by stefan wiemer and max wyss, *Bull. seism. Soc. Am.,* **93,** 1862–1867.

Rydelek, P.A., Horiuchi, S. & Iio, Y., 2002. Spatial and temporal characteristics of low-magnitude seismicity from a dense array in western Nagano Prefecture, Japan, *Earth Planets Space,* **54,** 81–89.

Sandri, L. & Marzocchi, W., 2007. A technical note on the bias in the estimation of the *b*-value and its uncertainty through the least squares technique, *Ann. Geophys.,* **50,** 329–339.

Scholz, C.H., 1968. The frequency-magnitude relation of microfracturing in rock and its relation to earthquakes, *Bull. seism. Soc. Am.,* **58,** 399–415.

Schorlemmer, D., Neri, G., Wiemer, S. & Mostaccio, A., 2003. Stability and significance tests for *b*-value anomalies: example from the Tyrrhenian Sea, *Geophys. Res. Lett.,* **30,** doi:10.1029/2003GL017335.

Schorlemmer, D., Wiemer, S. & Wyss, M., 2005. Variations in earthquake-size distribution across different stress regimes, *Nature,* **437,** doi:10.1038/nature04094.

Shi, Y. & Bolt, B.A., 1982. The standard error of the magnitude-frequency *b* value, *Bull. seism. Soc. Am.,* **72,** 1677–1687.

Siegel, A.F., 1982. Robust regression using repeated medians, *Biometrika,* **69,** 242–244.

Smirnov, M.Y., 2003. Magnetotelluric data processing with a robust statistical procedure having a high breakdown point, *Geophys. J. Int.,* doi:10.1046/j.1365-246X.2003.01733.x.

Tinti, S. & Mulargia, F., 1985. Effects of magnitude uncertainties on estimating the parameters in the Gutenberg-Richter frequency-magnitude law, *Bull. seism. Soc. Am.,* **75,** 1681–1697.

Tinti, S. & Mulargia, F., 1987. Confidence intervals of *b*-values for grouped magnitudes, *Bull. seism. Soc. Am.,* **77,** 2125–2134.

Utsu, T., 1965. A method for determining the value of b in a formula logn=a-bm showing the magnitude-frequency relation for earthquakes (in Japanese with english abstract), *Geophys. Bull. Hokkaido Univ.,* **13,** 99–103.

Utsu, T., 1966. A statistical significance test of the difference in *b* value between two earthquake groups, *J. Phys. Earth,* **14,** 37–40.

Utsu, T., 1992. Introduction to seismicity, *Surijishingaku (Mathematical Seismology)*, *Inst. Statis. Math,* **34,** (VII), 139–157, (in Japanese).

Utsu, T., 1999. Representation and analysis of the earthquake size distribution: a historical review and some new approaches, *Pure appl. Geophys.,* **155,** 509–535.

Weichert, D.H., 1980. Estimation of the earthquake recurrence parameters for unequal observation periods for different magnitudes, *Bull. seism. Soc. Am.,* **70,** 1337–1346.

Wells, D.L. & Coppersmith, K.J., 1994. New empirical relationship among magnitude, rupture length, rupture width, rupture area, and surface displacement, *Bull. seism. Soc. Am.,* **84,** 974–1002.

Wessel, P. & Smith, W.H.F., 1991. Free software helps map and display data, *EOS, Trans. Am. geophys. Un.,* **72,** 445–446.

Wiemer, S. & Benoit, J., 1996. Mapping the b-value anomaly at 100 km depth in the Alaska and New Zealand subduction zones, *Geophys. Res. Lett.,* **23,** 1557–1560.

Wiemer, S. & Wyss, M., 1997. Mapping the frequency-magnitude distribution in asperities : an improved technique to calculate recurrence time?, *J. geophys. Res.,* **102,** 15 115–15 128.

Wiemer, S. & Wyss, M., 2000. Minimum magnitude of complete reporting in earthquake catalogs: examples from Alaska, the Western United States, and Japan, *Bull. seism. Soc. Am.,* **90,** 859–869.

Wiemer, S. & Wyss, M., 2002. Mapping spatial variability of the frequency-magnitude distribution of earthquakes, *Adv. Geophys.,* **45,** 259–302.

Wiemer, S., McNutt, S.R. & Wyss, M., 1998. Temporal and three-dimensional spatial analysis of the frequency-magnitude distribution near Long Valley caldera, California, *Geophys. J. Int.,* **134,** 409–421.

Wiemer, S. & Wyss, M., 2003. Reply to "Comment on 'Minimum magnitude of completeness in earthquake catalogs: examples from Alaska, the Western United States and Japan by Stefan Wiemer and Max Wyss' by Paul A. Rydelek and I.S. Sacks", *Bull. seism. Soc. Am.,* **93,** 1868–1871.

Woessner, J. & Wiemer, S., 2005. Assessing the quality of earthquake catalogues: estimating the magnitude of completeness and its uncertainty, *Bull. seism. Soc. Am.,* **95,** 684–698.

Wyss, M., Shimazaki, K. & Wiemer, S., 1997. Mapping active magma chambers beneath the off-Ito volcano, Japan, *J. geophys. Res.,* **102,** 413–420.

Wyss, M., Schorlemmer, D. & Wiemer, S., 2000. Mapping asperities by minima of local recurrence time: the San Jacinto-Elsinore fault zones, *J. geophys. Res.,* **105,** 7829–7844.

Zuniga, R. & Wyss, M., 2001. Most- and least-likely locations of large to great earthquakes along the Pacific coast of Mexico estimated from local recurrence times based on *b*-values, *Bull. seism. Soc. Am.,* **91,** 1717–1728.