

# A test on methods for $M_C$ estimation based on earthquake catalog

YiJian Zhou<sup>1</sup>, ShiYong Zhou<sup>1\*</sup>, and JianCang Zhuang<sup>2</sup>

<sup>1</sup>Department of Geophysics, Peking University, Beijing 100871, China;

<sup>2</sup>Institute of Statistical Mathematics, 10-3 Midoori-Cho, Tachikawa, Tokyo, 190-8569, Japan

**Abstract:** This study tested five methods widely used in estimating the complete magnitudes ( $M_C$ ) of earthquake catalogs. Using catalogs of observed earthquake properties, we test the performance of these five algorithms under several challenging conditions, such as small volume of events and spatial-temporal heterogeneity, in order to see whether the algorithms are stable and in agreement with known data. We find that the maximum curvature method (MAXC) has perfect stability, but will significantly underestimate  $M_C$  unless heterogeneity is absent.  $M_C$  estimated by the  $b$ -value stability method (MBS) requires many events to reach a stable result. Results from the goodness of fit method (GFT) were unstable when heterogeneity lowered the fitness level. The entire magnitude range method (EMR) is relatively stable in most conditions, and can reflect the change in  $M_C$  when heterogeneity exists, but when the incomplete part of the earthquake catalog is dismissed, this method fails. The median-based analysis of the segment slope method (MBASS) can tolerate small sample size, but is incapable of reflecting the missing degree of small events in aftershock sequences. In conditions where  $M_C$  changes rapidly, such as in aftershock sequences, observing the time sequence directly can give a precise estimation of the complete sub-catalog, but only when the number of events available for study is large enough can the MAXC, GFT, and MBS methods give a similarly reliable estimation.

**Keywords:** complete magnitude of an earthquake catalog; G-R law;  $b$ -value; FMD

**Citation:** Zhou, Y. J., Zhou, S. Y., and Zhuang, J. C. (2018). A test on methods for  $M_C$  estimation based on earthquake catalog. *Earth Planet. Phys.*, 2, 150–162. <http://doi.org/10.26464/epp2018015>

## 1. Introduction

Earthquake catalogs are important for seismicity analysis and seismic hazard assessment. In compiling earthquake catalogs, not all events can be detected and thus cataloged, for example when the event is too small or the ambient noise is too large, and when the aftershock waveform is overwhelmed by the main shock (Iwata, 2008). The incompleteness of earthquake catalogs can significantly affect any subsequent analysis that relies on catalog data. Thus, assessing the completeness of an earthquake catalog is a necessary step in some researches, such as  $b$ -value estimation.

The complete magnitude ( $M_C$ ) of an earthquake catalog is defined as: the magnitude of the weakest event which is fully collected in an earthquake catalog. A catalog that includes all events whose magnitude is greater or equal to  $M_C$  is called a complete catalog. In empirical studies, complete catalogs have been found to exhibit a log-linear frequency magnitude distribution (FMD) as depicted by equation (1):

$$\log_{10} N = a - b(m - M_C), \quad m \geq M_C \quad (1)$$

This statistical law is called the G-R law (Gutenberg and Richter, 1944);  $N$  represents the number of events with magnitude larger than magnitude  $m$ , and  $M_C$  is the smallest magnitude in the com-

plete catalog.

Traditionally,  $M_C$  is estimated based on this relationship, where we estimate  $M_C$  as the point deviating from the G-R law. Such methods are less computationally intensive, and are still frequently used. Another kind of method, based on waveform information (e.g., Gomberg, 1991), is rarely applied due to the huge computation time it requires. Another important parameter limiting use of the G-R law is the  $b$ -value, which controls the relative number of large and small earthquakes. Relative studies (e.g., Schorlemmer et al., 2005) have shown that  $b$ -values may reflect the stress level of the particular region; we can thus consider the  $b$ -value as having real physical meaning. Aki (1965) introduced what has become the common method for  $b$ -value estimation, the maximum likelihood estimation, as is shown in equation (2):

$$b = \frac{\log_{10} e}{\langle M \rangle - \left( M_C - \frac{\Delta m}{2} \right)}. \quad (2)$$

In the above equation,  $\langle M \rangle$  is the mean of the magnitudes that are larger than  $M_C$ ;  $\Delta m$  is the width of the magnitude bin, commonly chosen to be 0.1. In this relationship,  $M_C$  directly affects the estimation of  $b$ -value, so we can reasonably assume that our estimation of  $M_C$  can affect the  $b$ -value estimate, and this is verified in some tests (e.g., Woessner and Wiemer, 2005).

The complete catalog is generally extracted from a raw catalog by dismissing its incomplete part. And since the frequency dimen-

Correspondence to: S. Y. Zhou, zsy@pku.edu.cn

Received 11 JAN 2018; Accepted 19 MAR 2018.

Accepted article online 27 MAR 2018.

Copyright © 2018 by Earth and Planetary Physics.

sion is log scaled in FMD, a small difference in  $M_C$  can significantly change  $N$ , the number of events: When  $M_C$  is over-estimated, the quantity of available data is considerably decreased; this reduction affects the stability of further calculations, especially if the raw catalog is small. On the other hand, an under-estimated  $M_C$  will result in biased results due to the uncollected (incomplete) events. The importance of correctly estimating  $M_C$  is thus clear on these grounds alone. But since equation (2) and related studies correlate  $M_C$  and  $b$ -value, an accurate  $M_C$  estimation is also crucial to correctly estimating the  $b$ -value.

In real applications, considering the spatiotemporal heterogeneity of  $M_C$ , investigators must consider the pattern of an event's spatiotemporal distribution, so as to determine its extension in space and time, when the catalog is relatively complete (Wiemer and Wyss, 2000). When spatial heterogeneity exists, in mapping  $M_C$  the data used to calculate  $M_C$  on each spatial grid can be heterogeneous. A robust technique for mapping  $M_C$  should give a median estimation if the data on a grid is composed of subsets with different  $M_C$ . On the other hand, due to a high spatial resolution or low seismicity, there can be only about 100 events in one grid (Mignan et al., 2011), which challenges the algorithm's stability. In conditions where  $M_C$  changes rapidly, such as in aftershock sequences,  $M_C$  will be quite large immediately following the main shock, since the waveform of aftershocks is affected or even overwhelmed by that of the main shock. However, in determining parameters of some models of seismicity (such as Omori-Utsu formula, Utsu, 1957), a complete sub-catalog is required (Zhuang JC et al., 2012). Thus, a standard test for a method's performance in temporal heterogeneity is whether it can detect sensitively the point at which  $M_C$  decreases to a stable level.

In brief, several alternative methods for calculating  $M_C$  are often used in such special conditions as: small volume of events, spatial  $M_C$  heterogeneity, and rapid change of  $M_C$  temporally. We will test the performance of the five existing methods, in these special situations, for  $M_C$  estimation based on earthquake catalogs, and proceed to give suggestions for when each might be most appropriately used. Similar work has been done using synthetic catalogs (such as Huang YL et al., 2016); our study provides further validation by testing on data from real catalogs of different characters and from different seismic networks.

## 2. Data

We used three groups of datasets for testing three aspects of  $M_C$  estimation performance. A detailed description follows.

### 2.1 Dependence on Sample Size

In this part of the test, we use the same data sets as in the original paper introducing the MBASS method (Amorese, 2007). To test

his new method, Amorese tried to reproduce the results in Woessner and Wiemer (2005) (abbreviated as W&W2005 hereafter), and his data sets are basically the same as in W&W2005. These are the data sets:

- (1) Regional catalog ECOS, which is a subset of the Earthquake Catalog of Swiss events from the Swiss Seismological Service. The subset we used records earthquakes in the southern province Wallis from 1992 to 1998; the magnitude scale is local magnitude.
- (2) Regional catalog NCSN, compiled by the Northern California Seismic Network. We use a subset from the San Francisco Bay area during 1998–2002; this catalog gives coda-duration magnitudes.
- (3) Volcanic region NIED, maintained by the National Research Institute for Earth Science and Disaster Prevention. The subset covers 1992–2002 for a volcanic region in Kanto province; magnitudes are locally reported.
- (4) Global catalog CMT, a subset of the Harvard Centroid Moment Tensor catalog, 1983–2002; only shallow depth earthquakes ( $d < 70$  km) are included.

A brief description of each catalog is summarized in Table 1, and their FMD are plotted in Figure 1. The catalogs are from different tectonic regions and of different magnitude scales, which increases the likelihood that our results will be reliable in a wide range of situations.

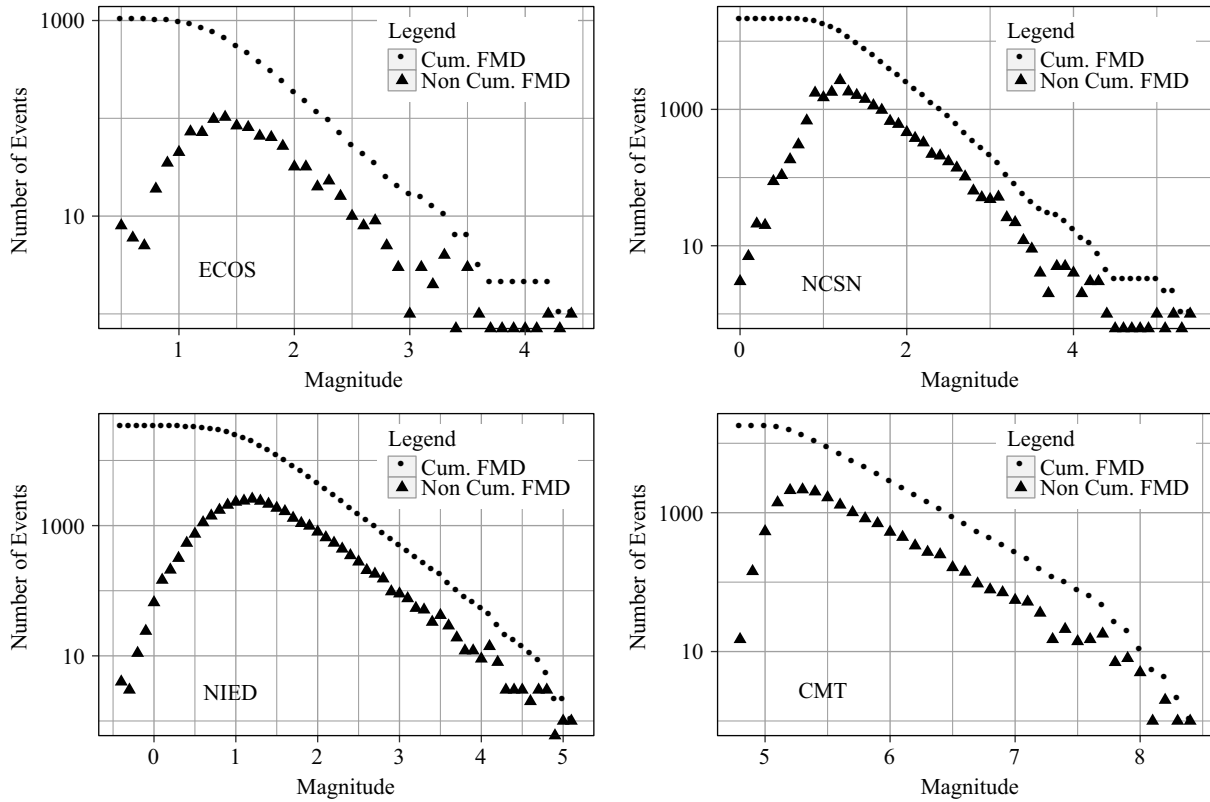
Here we chose the data set from Schorlemmer and Woessner (2008), a subset from the Southern California Seismic Network (SCSN), with a time span of January 2001 to July 2007. We chose this earthquake catalog because previous studies have mapped its  $M_C$  distribution (Schorlemmer and Woessner, 2008). Two sub-regions with significant different  $M_C$  values are selected, whose data are mixed to create heterogeneity, which is what can happen in real  $M_C$  mapping procedures. We fixed the size of the mixed catalog, and changed the mixing ratio to see the effect of heterogeneity on each algorithm. The expectation is that the resulting  $M_C$  will change linearly with increasing mixing ratio.

### 2.2 Detectability of Short Term Missing of Small Events

A complete sub catalog is needed in fitting an event to models of seismicity. In terms of aftershock sequence, the starting time for a complete catalog needs to be estimated. Zhuang JC et al., (2012) suggests that this starting time can be estimated by observing directly the time sequence, which is magnitude in respect of events' index. We wish to see if the algorithms for  $M_C$  estimation can reflect the missing of events, and whether the results can be compared to that of the 'visual inspection method'. We select four sets of aftershock sequences.

**Table 1.** Detailed information for the four catalogs used here

	ECOS	NCSN	NIED	CMT
Property	Regional	Regional	Volcanic	Global
Time Span	1992–2002	1998–2002	1992–2002	1983–2002
Spatial Range	6.8°E–8.4°E 45.9°N–46.65°N	123°W–120.5°W 36.0°N–39.0°N	138.95°E–139.35°E 34.08°N–35.05°N	
Sample Size	985	19690	31240	16472



**Figure 1.** FMD of four real catalogs used here, which is from the origin paper for MBASS method (Amorese, 2007).

(1) Wenchuan earthquake: the aftershock sequence occurred on 12 May 2008 in Sichuan Province, China, with a magnitude  $M8.0$ . The data set is from the appendix of an article on temporal seismicity model on CORSSA (Zhuang JC et al., 2012). There are 198 events in the catalog.

(2) Fukuoka earthquake: 16 August 2005 offshore of the western area of Fukuoka-Ken, Kyushu, Japan. The aftershock data are the test data set in the software SASeis2006 (Ogata, 2006). There are 403 events in the catalog.

(3) Miyagi earthquake:  $M6.2$ , 26 July 2003 in northern Miyaji-Ken, northern Japan. The test data set is in the software SASeis2006 (Ogata, 2006). We extract a subset with magnitude larger than 1.5, 1459 events in this catalog.

(4) Landers earthquake: a 1992  $M_W7.3$  aftershock sequence extracted from the SCSN catalog. We use the subset of events of magnitude larger than 1. The selected subset includes 19707 events.

These aftershock sequences are from different tectonic regions and the data were collected with different detecting abilities (thus characterized by differing numbers of events). This variability in our data sets can give us important knowledge of how well the methods work in different situations.

### 3. Method

#### 3.1 Algorithms for $M_C$ Estimation Based on Earthquake Catalog

The five methods that are used for estimating  $M_C$ :

(1) MAXC, maximum curvature method by Wiemer and Wyss (2000). The magnitude corresponding to the point of the maximum curvature deviating G-R law is considered an estimate of  $M_C$ , i.e., the magnitude connecting to the linear cumulative part of FMD.

(2) GFT, goodness-of-fit method by Wiemer and Wyss (2000). This method extracts subsets with ascending cut-off magnitude  $M_{CO}$ , and measures the deviation between these subsets and the G-R law by making linear fits to the G-R law, which will finally output a series of goodness-of-fit values. The goodness-of-fit  $R$  is defined as:

$$R(a, b, M_{CO}) = 100 - \left( \frac{\sum_{M_{CO}}^{M_{max}} |B_i - S_i|}{\sum_i B_i} 100 \right), \quad (3)$$

where  $B_i$  and  $S_i$  are the observed and predicted number of events in  $i$ -th magnitude bin. With larger  $M_{CO}$ , fewer events are missed in the catalog, and the corresponding  $R$  value tends to grow. Once this  $R$  reaches a confidence level of 95% or 90%, the corresponding  $M_{CO}$  is determined as  $M_C$ . Note that if a 90% goodness-of-fit level cannot be reached, the GFT method fails.

(3) MBS,  $M_C$  by  $b$ -value stability by Cao AM and Gao SS (2002) and modified by Woessner and Wiemer (2005). This method, too, extracts subsets with ascending cut-off magnitude  $M_{CO}$ , and calculates the corresponding  $b$ -values for each subset. When small events are not recorded, larger earthquakes constitute a relatively heavier part of the subset of small  $M_{CO}$ , and the corresponding  $b$ -value is too low; with  $M_{CO}$  ascending, the estimated  $b$ -value becomes higher and finally reaches a ‘plateau’. The first  $M_{CO}$  result-

ing in a stable  $b$ -value is the estimated  $M_C$ .

With development of this method, the criterion for stability is modified. Cao AM and Gao SS (2002) measures the change of  $b$ -value for two successive  $M_{CO}$ , and  $b$ -value is considered stable if the difference is smaller than 0.03. But Woessner and Wiemer (2005) found that this criterion will lead to poor numerical stability of the MBS method. Their criterion measures the difference between the  $b$ -value for a certain  $M_{CO}$  and the average  $b$ -value for this  $M_{CO}$ , which is defined by equation (4):

$$b_{ave} = \sum_{M_{CO}}^{M_{CO}+win\_M} b(M_{CO}) \Delta m / (win\_M + 0.1), \quad (4)$$

where  $\Delta m$  is the size of magnitude bin, and  $win\_M$  is the window length for calculating.

A  $b$ -value is considered stable if the difference is smaller than the  $b$ -value uncertainty by Shi YL and Bolt (1982), given by the following equation:

$$\delta b = 2.3b^2 \sqrt{\frac{\sum_{i=1}^N (M_i - \langle M \rangle)}{N(N-1)}}, \quad (5)$$

where  $b$  is the calculated  $b$ -value;  $N$  is the number of events to calculate  $b$ ;  $\langle M \rangle$  is the average magnitude.

(4) EMR, entire magnitude range method by Woessner and Wiemer (2005). This method tries to fit the entire non-cumulative part of FMD, instead of merely the linear complete part of FMD. Woessner and Wiemer (2005) used the normal cumulative distribution function (CDF) to model the incomplete part of the seismic intensity, and the G-R law for the complete part, which is depicted by equation (6). They compared normal CDF with other CDFs, and relative tests show that normal CDF fits best.

$$\lambda(m|\mu, \sigma) = \begin{cases} \lambda(m|\mu, \sigma, m < M_C) = \int_{-\infty}^m \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx \\ \lambda(m|\mu, \sigma, m \geq M_C) = \exp(-\beta(m - M_C)). \end{cases} \quad (6)$$

The detailed procedure is: first divide the catalog into complete and incomplete parts by an ascending cut-off magnitude, and then fit the data separately. For each  $M_{CO}$ , there are four parameters changeable:  $\sigma$  and  $\mu$  of the normal CDF, and  $a$ -value and  $b$ -value in the G-R law. The  $M_{CO}$  that leads to the best fit is the final  $M_C$  estimation.

(5) MBASS, the median-based analysis of the segment slope by Amorese (2007). The MBASS method calculates  $M_C$  by detecting the multiple change-point in the non-cumulative FMD's slope series. The segment slope at magnitude  $m$  in FMD is defined as equation:

$$s(m) = \frac{\log(N(m + \Delta m)) - \log(N(m))}{\Delta m}, \quad (7)$$

where  $N$  is the number of events;  $\Delta m$  is the size of magnitude bin.

After each segment slope at consecutive magnitudes was calculated, the median was subtracted from this slope series. The Wilcoxon-Mann-Whitney (WMW) test (Mann and Whitney, 1947; Wilcoxon, 1945) was then applied to find a notable and stable change point. These two steps are done iteratively, and the mag-

nitude corresponding to the main change in the median of the slope series is found, which is the output  $M_C$  value. The detailed algorithm and corresponding program is provided in Amorese (2007).

All the programs implementing these five methods can be found in the package ZMAP (Wiemer, 2001). Our results were obtained from this package, which is used also by other scholars, such as Woessner and Wiemer (2005).

### 3.2 The Uncertainty for $M_C$ and $b$ -value Estimation

The real catalogs are generated following a certain probability distribution. We assume that the FMD of a catalog can represent its real probability distribution, while some random fluctuations are inevitable. To assess the influence of these fluctuations, or the reliability of the results, we calculate  $M_C$  and  $b$ -values and their uncertainties by these steps:

We use a Monte Carlo approximation of the bootstrap method (Efron, 1979), which is to draw, with replacement from the original catalog, a certain size bootstrap sample. The sampling procedure is repeated several hundreds of times, and we estimate  $M_C$  for each bootstrap sample. The mean value of these  $M_C$  is the final result for  $M_C$  estimation, and its standard deviation gives the uncertainty. The corresponding  $b$ -value is estimated by MLE via the equation (2), with the uncertainty estimated by Shi YL and Bolt's (1982) criteria (equation (5)).

### 3.3 Generation of Synthetic Catalogs

In real catalogs, we do not know the true  $M_C$ . In order to find the accuracy of each method, we can take advantage of synthetic tests, where we can set the "true" value of  $M_C$ .

The model introduced by Ogata and Katsura's (1993) (abbreviated as OK1993 hereafter) was chosen as the basis for synthetic catalog generation. OK1993 models the detectability of station by normal CDF, following equation (10). The final observed magnitude frequency intensity (equation (8)) is given by multiplication of the detection rate probability and the real intensity from the G-R law (equation (9)).

$$\lambda(m) = \lambda_0(m) q(m), \quad (8)$$

$$\lambda_0(m|\beta) = \exp(-\beta m), \quad (9)$$

$$q(x|\mu, \sigma) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx. \quad (10)$$

The difference between the OK1993 model and the EMR method lies in the fact that no separation of complete and incomplete parts is made. This avoids introduction of a discontinuous point in FMD, which is unreasonable physically. The magnitude of each event is generated stochastically by the rejection method (Zhuang JC and Touati, 2015). The rejection method is a basic technique for simulating a random variable from a known distribution (equation (8) in our case). The brief procedure is, first, to generate a random number, and then determine whether to adopt this number, according to the known probability density function.

The estimation of theoretical  $M_C$  value follows Huang YL et al.

(2016). Theoretical  $M_C$  is estimated as the magnitude above which one event in every 500 events is expected to be unrecorded. To be specific, it is the resulting  $M$  value when  $N$  equals 500 in equation (11), where  $q(m)$  has the same definition as in equation (10).

$$N \times \left\{ \int_M^{\infty} [1 - q(m)] dm \right\} = 1. \quad (11)$$

The detailed discussion for choosing the criteria of theoretical  $M_C$  is presented in Huang YL et al. (2016).

### 3.4 Dependence on Sample Size

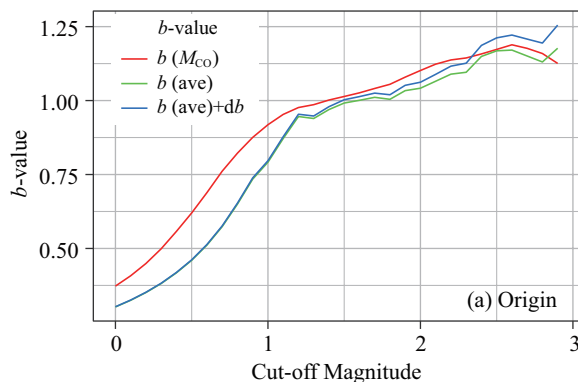
For real catalogs, we used the bootstrap method mentioned above to get  $M_C$  and  $b$ -values (with their uncertainties) under different sampling size. This is done by changing the size of bootstrap samples to get catalogs of different volumes of events that nevertheless come from the same FMD.

This method will leave or even magnify a certain pattern in the FMD to bootstrap samples, which can cause systematic deviation in the final result. Thus, in tests on synthetic catalogs, we tried two different approaches. One is to generate catalogs of the same size as the corresponding real catalogs, then repeat the same procedure as to the real catalogs; the other is to generate directly different sizes of catalogs. We tested the difference between these two approaches.

In generating synthetic catalogs, we cite the results from Mignan and Woessner (2012), who used the OK1993 model to fit the data-set in W&W2005; fitted parameters are presented.

### 3.5 Tolerance to Heterogeneity

The bootstrap method mentioned before is used to select randomly a certain number of events from both subsets. Considering the stochastic fluctuation, selected events may come mainly from large or small magnitude events of one subset, and thus the final results of  $M_C$  to mixing ratio will not change monotonously. To attenuate this effect, for each mixing ratio the procedure was repeated five times and the resulting  $M_C$  values were averaged. Apart from consuming an unreasonable amount of time, increasing uncertainty is an important reason why we do not repeat more than five times.



### 3.6 Detectability of Small Missed Events

We used a sliding window approach, calculating  $M_C$  for events in the moving 'window'. A relation between  $M_C$  and starting index is plotted to see whether the  $M_C$  algorithms can detect the missing of small events in aftershock sequences, and to compare the results to that of the 'visual inspection method'.

## 4. Results

### 4.1 Problems Found in Reproducing the Results of W&W2005 and Relative Modification in Programs

Before conducting our tests, we tried to reproduce the results of W&W2005 for inspection of the programs and data sets we used. We found that the MBS method gave much larger results than W&W2005 for the four catalogs we used, as shown in Table 2. We then plotted  $b$ -value in respect to  $M_{CO}$  for the NCSN catalog. Though a 'plateau' is found when  $M_{CO}$  equals about 1.4, the  $b$ -value uncertainty is too big by Shi and Bolt's (1982) criteria, so the  $b$ -value is thus not considered stable (Fig. 2a).

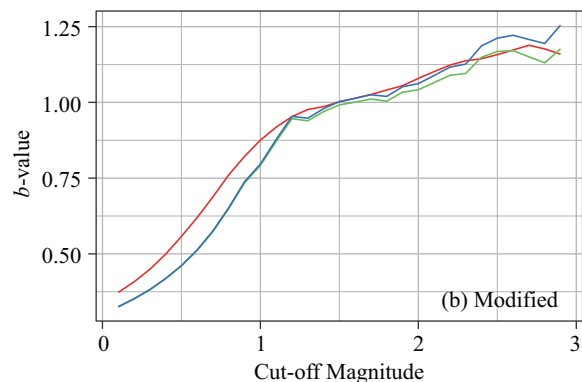
In W&W2005, the author pointed out that 'window length' has a strong influence on the result, and suggested that it be set to 5. Our tests show that too short a 'window', such as 2 or 3, will lead to large uncertainty, which means poor computational stability. We also found in the code that average  $b$ -value is estimated from  $b$ -values for magnitudes from a certain  $M_{CO}$  forward (equation 4), which will give too much weight to  $b$ -values for larger  $M_{CO}$ .

So we tried to change the starting point, and finally found that a  $-0.1$  shift will give better results (following equation (12)). Modified results are shown in Figure 2 and Table 2.

$$b_{ave} = \sum_{M_{CO}-0.1}^{M_{CO}+win\_M-0.1} b(M_{CO}) \Delta m / (win\_M + 0.1). \quad (12)$$

**Table 2.**  $M_C$  estimation by MBS method for four real catalogs

	ECOS	NCSN	NIED	CMT
ORIGIN	1.64±0.12	2.24±0.33	1.98±0.11	5.93±0.34
MODIFIED	1.55±0.12	1.47±0.37	1.91±0.07	5.46±0.16
W&W2005	1.64±0.11	1.44±0.12	1.44±0.12	5.94±0.34
CORSSA	1.53±0.16	1.48±0.10	1.97±0.10	5.80±0.09



**Figure 2.** The relationship between  $b$ -value and cut-off magnitude for NCSN catalog, where the average  $b$ -value is estimated by (a) Original method in equation (4), and (b) modified method in equation (12).  $b(M_{CO})$  is the  $b$ -value corresponding to cut-off magnitude  $M_{CO}$ ;  $b(ave)$  is the averaged  $b$ -value calculated by equation (4) or (12);  $db$  is the uncertainty of estimated  $b$ -value.

## 4.2 Dependence on Sample Size

### 4.2.1 Resample from real catalog

Applying the bootstrap method mentioned before to real catalogs, setting the sample size  $N$  of the bootstrap samples to 100, 200, 500, 1000, 5000, 10000, 100000, respectively, the resulting relation between  $M_C$  and  $b$ -value to  $N$  is shown in Figure 3. We find in the four real catalogs that, for small  $N$ , the MBS method gives a substantially lower  $M_C$  estimate than other methods and the  $b$ -value is also seriously under-estimated; conversely, when  $N$  is large, the resulting  $M_C$  and  $b$ -values are much larger than from other methods. This shows that the MBS method is sensitive to  $N$ . In contrast, results from the MAXC method are quite stable, but are always the lowest of all five methods. Estimation of EMR will

decrease slightly with ascending  $N$ , and will finally be similar to that of the MAXC method, which tends to underestimate  $M_C$ ; thus a modification value of about 0.2 is needed (Huang et al., 2016). The GFT and MBASS methods have stable performance in this part of the test, but give estimates that are smaller than those of MBS.

### 4.2.2 Resample from synthetic catalogs

As a comparison to last part, synthetic catalogs of the same sizes as the original ones were processed in the same way. The FMD of synthetic catalogs and their theoretical  $M_C$  values are shown in Figure 4, and the results are shown in Figure 5. It is worth mention that additional treatment needs to be done for the ‘tail’ of FMD, since large earthquakes rarely happen and tend to deviate

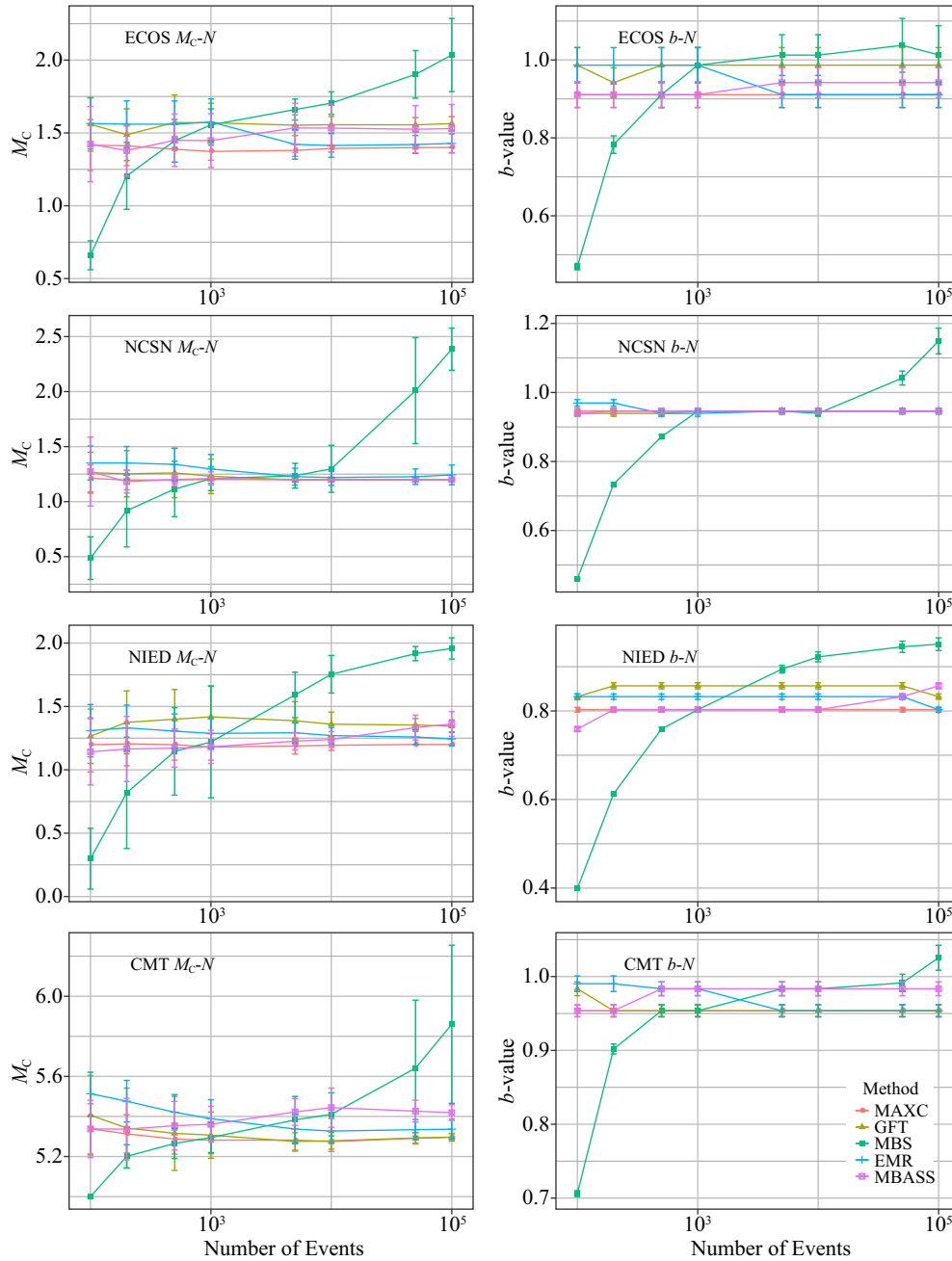
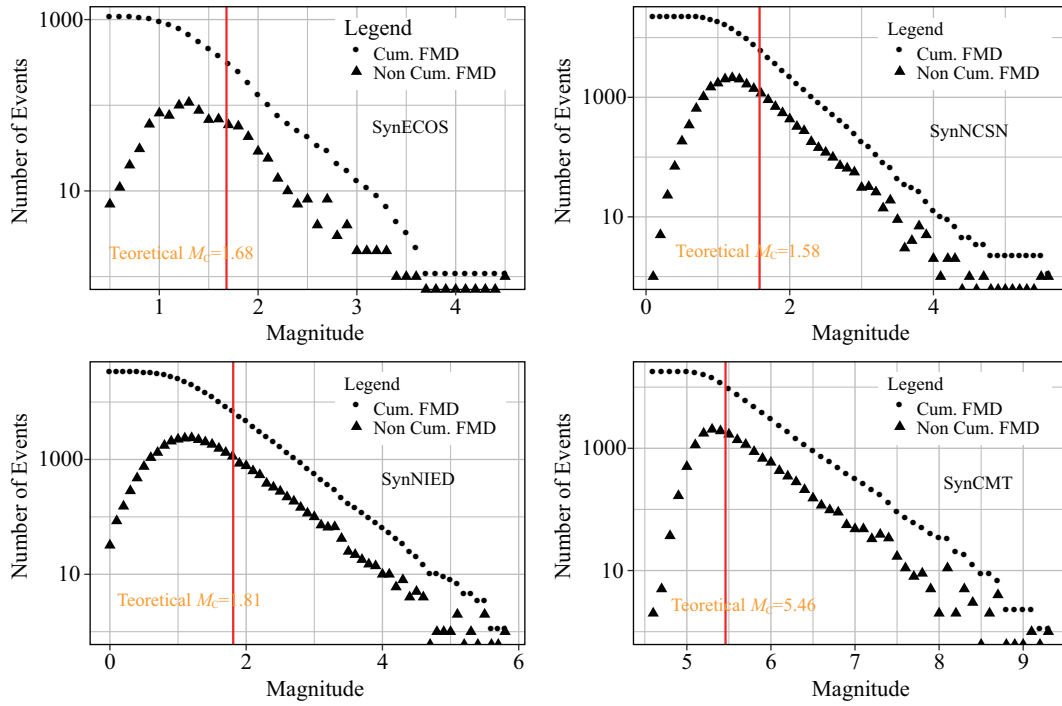


Figure 3. Relationship between  $M_C$ ,  $b$ -value and events number  $N$  for four real catalogs, using bootstrap method.



**Figure 4.** FMD of four synthetic catalogs. The vertical line indicates the theoretical  $M_C$  values estimated by the method in Huang et al. (2016).

from the G-R law. This happens in the synthetic CMT catalog. But we keep the original one since this problem would not affect our main conclusion.

We find the same phenomenon of stability for the MBS method; but for large  $N$ ,  $b$ -value estimation is not substantially as large as for  $M_C$ . The difference exists for the GFT method when  $M_C$  is over estimated for large  $N$ , but has no influence on  $b$ -value. Generally, all methods tend to under-estimate  $M_C$  compared to the theoretical  $M_C$  value estimated by Yilei Huang's criteria (Huang et al., 2016).

#### 4.2.3 Generating direct catalogs of different sample sizes

The results are shown in Figure 6. Note that all five methods are applied to the same catalogs, so the statistical fluctuations in creating synthetic catalogs are the same for all methods. The general patterns are the same for the last part of the test, but MBS performs more stable than other methods.

#### 4.3 Tolerance to Spatial Heterogeneity

We select from Schorlemmer and Woessner (2008) two polygon sub regions of significantly different  $M_C$ : (35.8°, 36.35°N) × (117.88°, 118.4°W) and (33.35°, 33.7°N) × (116.55°, 119.95°W), whose sample size are similar (6337 and 6471 events); their FMD and EMR fittings are plotted in Figure 7. Note that the first subset has a flat non-cumulative FMD, which is caused by heterogeneity (Wiemer and Wyss, 2000; Mignan et al., 2011).

The mixing ratio of the two sub catalogs changes from 0 to 1, and the sizes of mixed catalogs are set to 500, 1000, 3000, 5000, 7000, 10000, 15000, 20000, respectively. Affected by heterogeneity, the goodness-of-fit criterion for the GFT method is set to 90%; other side, the method would be numerically unstable. The results are

shown in Figure 8.

More fluctuation exists when  $N$  is no more than 1000; below that threshold, computation is not stable. For samples large enough, results for the five methods are divided into three groups: the MAXC and MBASS methods provide the lower estimations; they rise and become closer with increasing ratios; the GFT and EMR methods lead to a median and relatively stable estimation; the MBS method leads to the largest fluctuations, and even the downward trend is ambiguous.

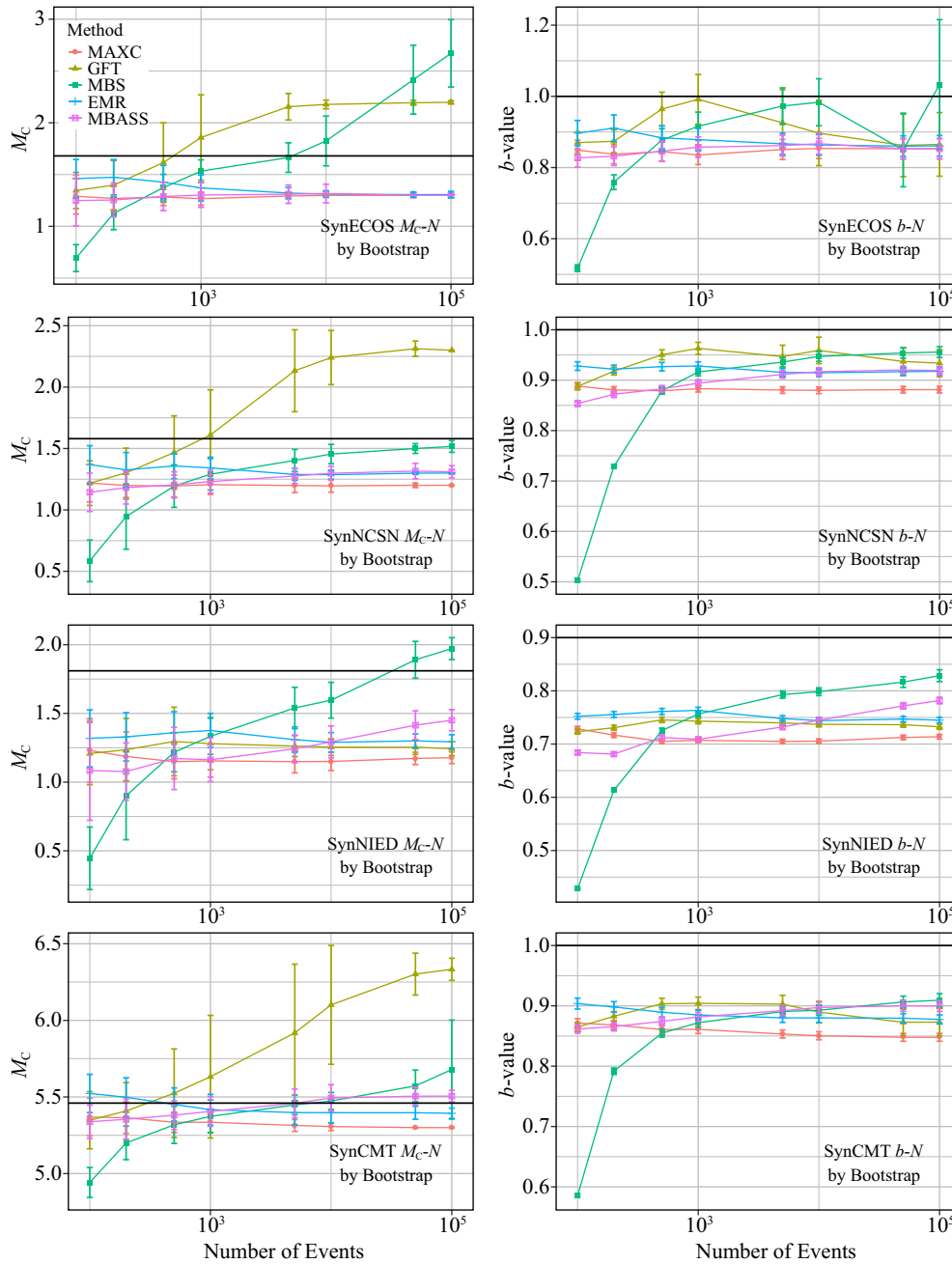
#### 4.4 Detectability of Small Missed Events

##### 4.4.1 Aftershock sequence for Wenchuan and Fukuoka earthquake

This set of aftershock sequences contain a small number of events. The results are shown in Figure 9. For the Wenchuan earthquake, since the sample size is small, MBS is unstable and the criterion for the GFT method is set to 90%. However, the visual inspection estimates the 'starting time' to be 40, which is not big enough for window length and would lead to an unstable estimation. So we choose a window length considering the tradeoff between window length and numerical stability. If the  $M_C$  algorithms work, the resulting  $M_C$  value should decrease with a time lapse. But as we can see in the figure, the usable MAXC, GFT and MBASS methods cannot reflect the missing of small events. The same phenomenon appears in the Fukuoka earthquake results.

##### 4.4.2 Aftershock sequence for Miyaji and Landers earthquake

When a catalog contains a large volume of events, a longer window length can be applied. The results are shown in Figure 10. For both of these earthquakes, the MBASS method gives a large  $M_C$  with strong fluctuations, so for the Landers aftershock sequence



**Figure 5.** Relationship between  $M_C$ ,  $b$ -value and sample size  $N$  for four synthetic catalogs, using bootstrap method. The horizontal line indicates the theoretical  $M_C$  estimations by the method in [Huang et al. \(2016\)](#).

we plotted results only for the other three methods. It can be seen in this part of the test that  $M_C$  estimated by the MAXC, GFT and MBS methods decreases to a rather stable value at the index close to that estimated by visual inspection. This means that they can detect the missing of small events.

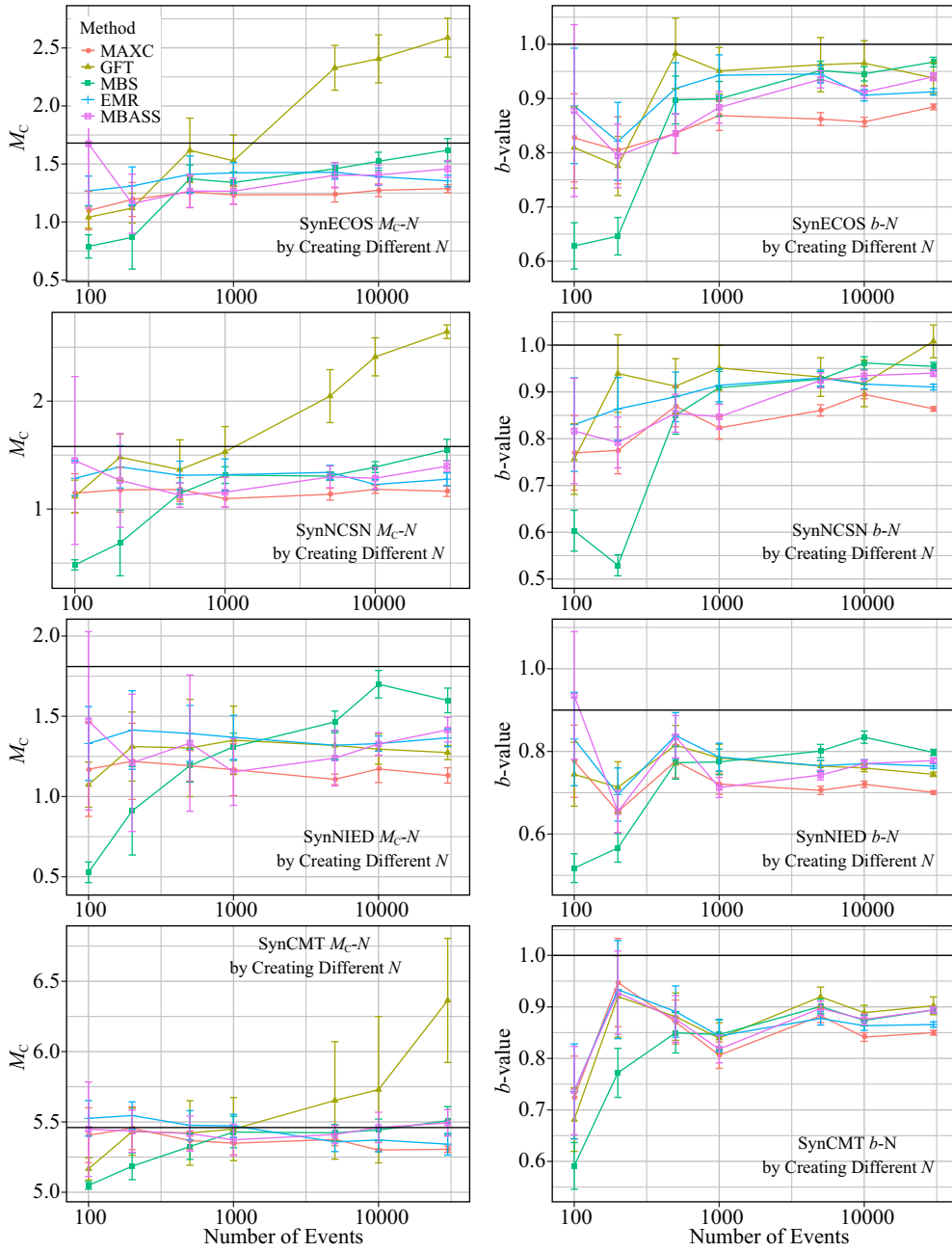
**5. Discussion and Conclusions**

In  $M_C$  estimation, all methods except the MBS method can give reasonable estimates when applied to data from small sample catalogs. When  $N$  is large, the MBS method will lead to a large  $M_C$  in tests based on real catalogs, but not when used on synthetic catalogs of different sample sizes. This fact reveals some kind of

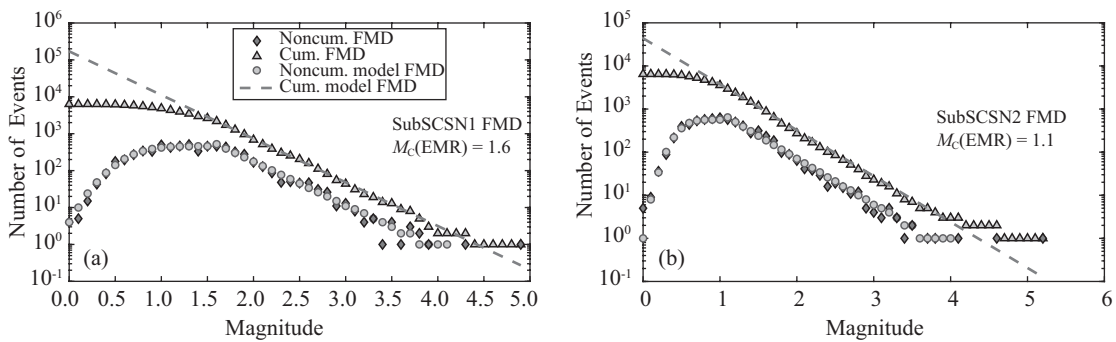
instability of the MBS method, but also shows that controlling bootstrap size is not a perfect way to study a method’s dependence on sample size. Perhaps a systematic error is magnified in the process of resampling, by some ‘bad’ pattern. In synthetic tests, the GFT method overestimates  $M_C$  significantly, which may result from the method used in creating synthetic catalogs. However, it is possible that the GFT method does lead to a substantially larger  $M_C$  value when the sample size is large.

In terms of  $b$ -value estimation, though  $M_C$  values can be significantly overestimated in synthetic tests,  $b$ -values are not. This may be explained if  $b$ -value as function of cutoff magnitude for synthetic catalogs has a long ‘plateau’, and even a large  $M_C$  estima-

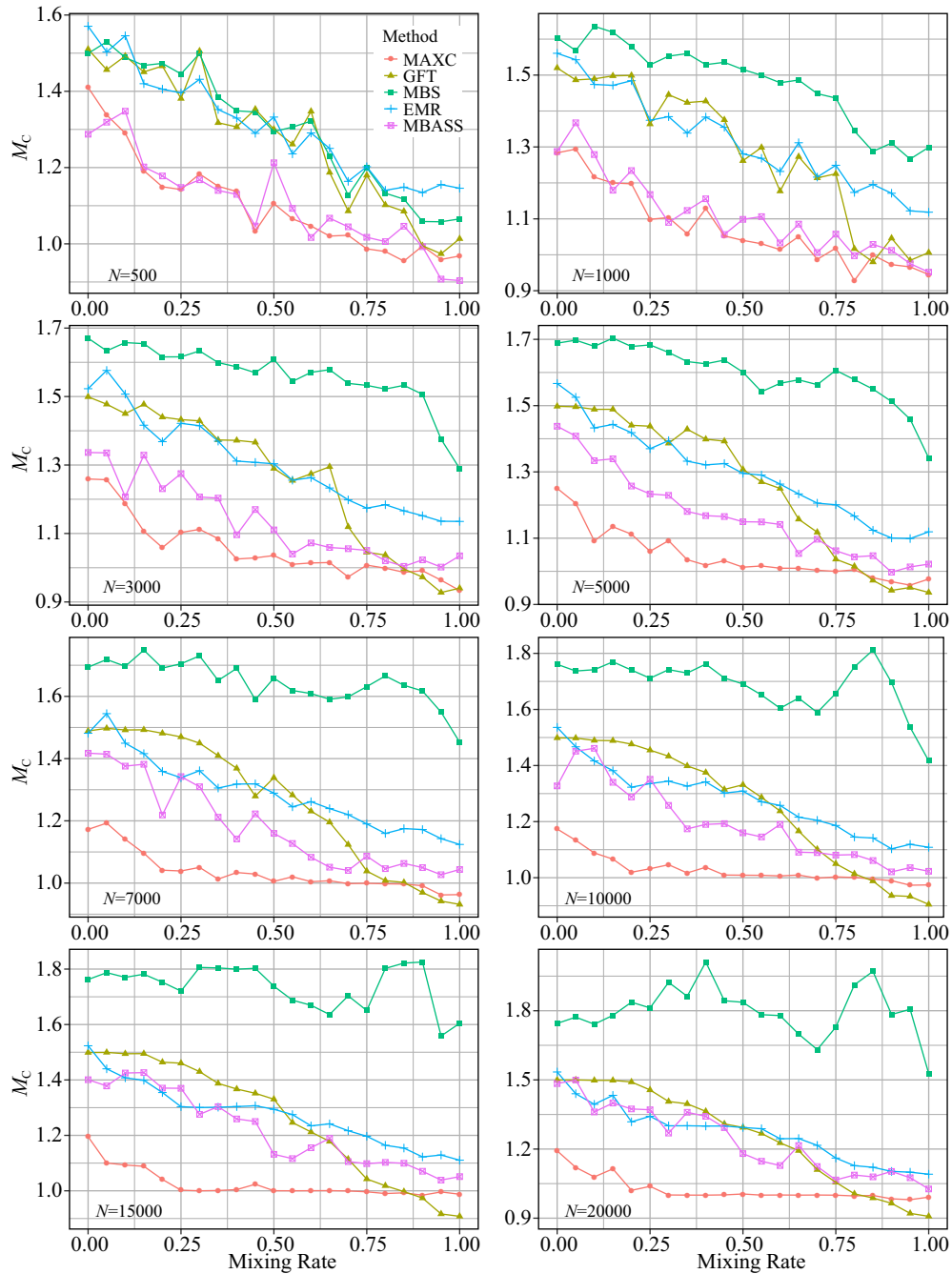




**Figure 6.** Relationship between  $M_C$ ,  $b$ -value and sample size  $N$  for four synthetic catalogs. Catalogs with different sizes was created directly. The horizontal line indicates the theoretical  $M_C$  estimations by the method in Huang et al. (2016).



**Figure 7.** FMD and EMR fitting results for two sub regions in Schorlemmer and Woessner (2008) from SCSN catalog.



**Figure 8.** Relationship between  $M_C$  and mixing rate for different sizes of mixed catalogs. ‘vRate’ is the ratio of subcatalog2 to the mixed catalog.

tion stays in this ‘plateau’. These observations reflect the over-idealization of synthetic catalogs; some ‘flaw’ in real catalogs is not being modeled. Thus, for  $b$ -value estimation, we suggest that the figure of  $b$ -value to cutoff magnitude should be inspected to get a visual impression.

Heterogeneity exists in our first selected subset. Results show that when the first subset takes a small ratio in the mixed catalog,  $M_C$  estimated by the MAXC method will be significantly lower than for the other four methods, which reveals a poor tolerance of heterogeneity by the MAXC method. Though  $M_C$  generally decreases by the MBASS method, fluctuations of the curve show an incapability of MBASS to reflect changes of  $M_C$  when heterogeneity exists.

On another hand, the conclusion of Mignan et al. (2011) is verified, since the separation between the results of MAXC and MBASS decreases when the ratio of the second subset increases. Results for the MBS method show even stronger fluctuation, and the trend of  $M_C$  decreasing can hardly be seen. This means that  $M_C$  estimation by the MBS method is untrustworthy for heterogeneous catalogs. GFT and EMR perform better in this test; the results show a linear change of  $M_C$  with increasing mixing ratio. This property enables these methods to be applied to studies of  $M_C$  mapping, as has been done by the EMR method in Schorlemmer and Woessner (2008).

In tests for aftershock sequences ranging from a few hundreds of

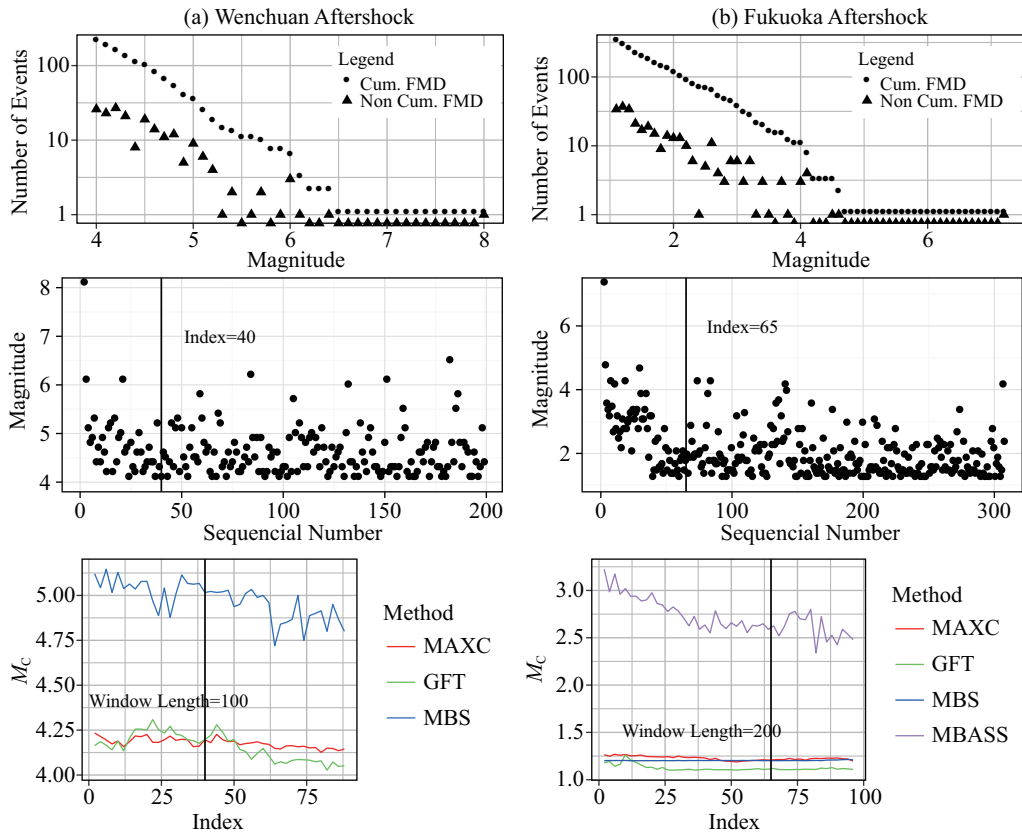


Figure 9. FMD, aftershock sequence and the relation between  $M_C$  and events' sequential index for Wenchuan and Fukuoka aftershock.

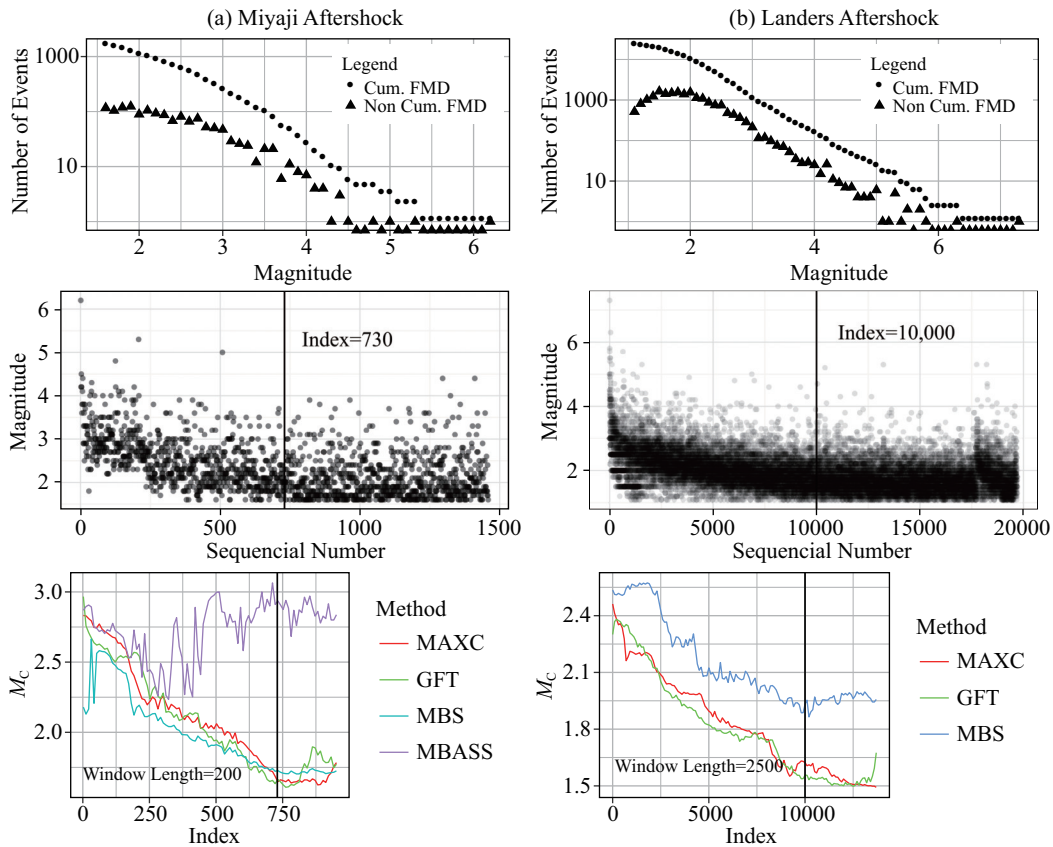


Figure 10. FMD, aftershock sequence and the relation between  $M_C$  and events' sequential index for Miyaji and Landers aftershock.

**Table 3.** Characters of the five methods

	Sample Size $N$	Spatial Heterogeneity	Missed Events	Accuracy
MAXC	Insensitive to sample size	Under-estimated $M_C$ substantially for heterogeneous catalog	Capable of detecting missed small events with enough large volume of events	A modification value is needed for it lead to stably a too low $M_C$ estimation
GFT	A certain amount of events are needed (about 200) to meet the confidence level	Capable of reflecting the change of $M_C$ when heterogeneity exists	Capable of detecting missed small events with enough large volume of events	Relatively high estimation for heterogeneous catalogs, and overestimation may happen for too large sample size
MBS	More events are required (about 500 in this test) to reach a stable result	Not capable of reflecting correctly the change of $M_C$ when heterogeneity exists	Capable of detecting missed small events with enough large volume of events	Resulting a high $M_C$ estimation, but the $b$ -value is more accurate
EMR	Stable with small volume of events (100 is enough)	Capable of reflecting the change of $M_C$ when heterogeneity exists	Not valid in absence of incomplete part of catalog	A moderate $M_C$ estimation is given
MBASS	Stable with small volume of events (100 is enough)	Capable of reflecting the change of $M_C$ when heterogeneity exists, but with less accuracy	Incapable of detecting missed small events	Similar result to MAXC method for catalogs without heterogeneity, or a much higher estimation is resulted

events to thousands of events, the ‘visual inspection method’ can always be effectively applied. The MAXC, GFT and MBS methods can detect missed small events only when the sample size is large; the MBASS method does not have this property. Since the EMR method needs a whole range of magnitudes to fit, it is not valid in this condition, where the incomplete part of FMD is discarded in extracting aftershock sequences.

The properties of the five methods are summarized in Table 3, based on the discussion above.

### Acknowledgments

The research is financially supported jointly by China Earthquake Science Experiment Project (grant 2016 CESE 0104) & Nature Science Foundation of China (grant 41474033). Also the paper accepted many suggestions from Yilei Huang, many thanks are expressed here. The earthquake catalogs used in this paper are from: (1) Four catalogs in Amorese (2007) are downloaded from the author’s website: <http://danielamorese.com/2016/09/19/the-data-sets-of-my-paper-applying-a-change-point-detection-method-on-frequency-magnitude-distributions-bssa-2007/> (Last accessed November 2016). (2) SCSN catalog downloaded from [http://www.data.scec.org/ftp/catalogs/SCSN\\_pre2008/](http://www.data.scec.org/ftp/catalogs/SCSN_pre2008/) (Last accessed November 2016).

### References

- Aki, K. (1965). Maximum likelihood estimate of  $b$  in the formula  $\log N = a - bM$  and its confidence limits. *Bull. Earthq. Res. Inst.*, 43, 237–239.
- Amorese, D. (2007). Applying a change-point detection method on frequency-magnitude distributions. *Bull. Seismol. Soc. Am.*, 97(5), 1742–1749. <https://doi.org/10.1785/0120060181>
- Cao, A. M., and Gao, S. S. (2002). Temporal variation of seismic  $b$  - values beneath northeastern Japan island arc. *Geophys. Res. Lett.*, 29(9), 1334. <https://doi.org/10.1029/2001GL013775>
- Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Ann. Statist.*, 7(1), 1–26. <https://doi.org/10.1214/aos/1176344552>

- Gomberg, J. (1991). Seismicity and detection/location threshold in the Southern Great Basin Seismic Network. *J. Geophys. Res.*, 96(B10), 16401–16414. <https://doi.org/10.1029/91JB01593>
- Gutenberg, B., and Richter, C. F. (1944). Frequency of earthquakes in California. *Bull. Seismol. Soc. Am.*, 34(4), 185–188.
- Huang, Y. L., Zhou, S. Y., and Zhuang, J. C. (2016). Numerical tests on catalog-based methods to estimate magnitude of completeness. *Chin. J. Geophys.*, 59(3), 266–275. <https://doi.org/10.1002/cjg2.20232>
- Iwata, T. (2013). Estimation of completeness magnitude considering daily variation in earthquake detection capability. *Geophys. J. Int.*, 194(3), 1909–1919. <https://doi.org/10.1093/gji/ggt208>
- Mann, H. B., and Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *Ann. Math. Stat.*, 18(1), 50–60. <https://doi.org/10.1214/aoms/1177730491>
- Mignan, A., Werner, M. J., Wiemer, S., Chen, C. C., and Wu, Y. M. (2011). Bayesian estimation of the spatially varying completeness magnitude of earthquake catalogs. *Bull. Seismol. Soc. Am.*, 101(3), 1371–1385. <https://doi.org/10.1785/0120100223>
- Mignan, A., and Woessner, J. (2012). Estimating the magnitude of completeness for earthquake catalogs. Community Online Resource for Statistical Seismicity Analysis. <http://dx.doi.org/10.5078/corssa-00180805>
- Ogata, Y., and Katsura, K. (1993). Analysis of temporal and spatial heterogeneity of magnitude frequency distribution inferred from earthquake catalogues. *Geophys. J. Int.*, 113(3), 727–738. <https://doi.org/10.1111/j.1365-246X.1993.tb04663.x>
- Ogata, Y. (2006). Statistical analysis of seismicity - updated version (SASeis2006). In Computer Science Monograph (pp. 1–28). Tokyo, Japan: The Institute of Statistical Mathematics.
- Schorlemmer, D., Wiemer, S., and Wyss, M. (2005). Variations in earthquake-size distribution across different stress regimes. *Nature*, 437(7058), 539–542. <https://doi.org/10.1038/nature04094>
- Schorlemmer, D., and Woessner, J. (2008). Probability of detecting an earthquake. *Bull. Seismol. Soc. Am.*, 98(5), 2103–2117. <https://doi.org/10.1785/0120070105>
- Shi, Y. L., and Bolt, B. A. (1982). The standard error of the magnitude-frequency  $b$  value. *Bull. Seismol. Soc. Am.*, 72(5), 1677–1687.
- Utsu, T. (1957). Magnitudes of earthquakes and occurrence of their aftershocks. *Zisin*, 10, 35–45. [https://doi.org/10.4294/zisin1948.10.1\\_35](https://doi.org/10.4294/zisin1948.10.1_35)

- Wiemer, S., and Wyss, M. (2000). Minimum magnitude of completeness in earthquake catalogs: examples from Alaska, the Western United States, and Japan. *Bull. Seismol. Soc. Am.*, 90(4), 859–869. <https://doi.org/10.1785/0119990114>
- Wiemer, S. (2001). A software package to analyze seismicity: ZMAP. *Seismol. Res. Lett.*, 72(3), 373–382. <https://doi.org/10.1785/gssrl.72.3.373>
- Wilcoxon, F. (1992). Individual comparisons by ranking methods. In S. Kotz, N. L. Johnson (Eds.), *Breakthroughs in Statistics* (pp. 196–202). New York, NY: Springer.
- Woessner, J., and Wiemer, S. (2005). Assessing the quality of earthquake catalogues: estimating the magnitude of completeness and its uncertainty. *Bull. Seismol. Soc. Am.*, 95(2), 684–698. <https://doi.org/10.1785/0120040007>
- Zhuang, J. C., Harte, D., Werner, M. J., Hainzl, S., and Zhou, S. Y. (2012). Basic models of seismicity: temporal models, Community Online Resource for Statistical Seismicity Analysis. <http://dx.doi.org/10.5078/corssa-79905851>
- Zhuang, J. C., and Touati, S. (2015). Stochastic simulation of earthquake catalogs, Community Online Resource for Statistical Seismicity Analysis. <http://dx.doi.org/10.5078/corssa-43806322>