



## Theme IV – Understanding Seismicity Catalogs and their Problems

### Estimating the magnitude of completeness for earthquake catalogs

Arnaud Mignan • Jochen Woessner

Swiss Seismological Service, ETH Zurich

How to cite this article:

Mignan, A., J. Woessner (2012), Estimating the magnitude of completeness for earthquake catalogs, Community Online Resource for Statistical Seismicity Analysis, doi:10.5078/corssa-00180805. Available at <http://www.corssa.org>.

Document Information:

Issue date: 1 April 2012 Version: 1.0

## Contents

1	Motivation	3
2	Prerequisites	3
3	Expected Achievements	3
4	Theoretical Background	4
5	Methodologies: An Overview	5
5.1	Catalog-based Methods	5
5.1.1	Why is a correct estimate of $M_c$ important?	7
5.1.2	Should we dismiss information below $M_c$ or is this incomplete information of any value?	9
5.2	Network-based Methods - Using Phase-Pick or Waveform Data	9
5.2.1	Probabilistic Magnitude of Completeness (PMC)	10
5.2.2	Assumptions, Features and Limits of PMC	11
6	Catalog-based Techniques to Evaluate $M_c$	14
6.1	Maximum Curvature (MAXC) technique	14
6.2	Goodness-of-Fit Test (GFT)	16
6.3	$M_c$ by $b$ -value stability (MBS)	18
6.4	$M_c$ from the Entire Magnitude Range (OK1993, EMR)	18
6.5	Median-Based Analysis of the Segment Slope (MBASS)	21
6.6	Day-to-night noise modulation	23
6.7	Diagnostics from Synthetic Catalogs	25
6.8	Exercise: Strategy to estimate $M_c$ for a Single Catalog Data Sample	27
7	Catalog-based Techniques to Map $M_c(x, y, z, t)$	28
7.1	Temporal Variations in $M_c$	28
7.2	Standard Catalog-based $M_c$ Spatial Mapping	29
7.3	Bayesian Magnitude of Completeness (BMC) Spatial Mapping	30
7.4	Comparison of Network-based and Catalog-based $M_c$ Spatial Mapping	33
8	Software packages and Catalog Data	36
8.1	Software packages	36
8.2	Catalog Data	36
9	Appendix	37

## 1 Motivation

Assessing the [magnitude of completeness](#)  $M_c$  of instrumental earthquake catalogs is an essential and compulsory step for any seismicity analysis.  $M_c$  is defined as the lowest magnitude at which 100% of the earthquakes in a space-time volume are detected. A correct estimate of  $M_c$  is crucial since a value too high leads to under-sampling, by discarding usable data, while a value too low leads to erroneous seismicity parameter values and thus to a biased analysis, by using incomplete data. In this article, we describe peer-reviewed techniques to estimate and map  $M_c$ . We provide examples with real and [synthetic earthquake catalogs](#) to illustrate features of the various methods and give the pros and cons of each method. With this article at hand, the reader will get an overview of approaches to assess  $M_c$ , understand why  $M_c$  evaluation is essential and an a non-trivial task, and hopefully be able to select the most appropriate  $M_c$  method to include in his seismicity studies.

## 2 Prerequisites

In order to test the procedures presented in this article, the reader shall have access to an earthquake catalog. A number of earthquake catalogs are freely accessible from the webpages of the seismic networks or their associated data centers. Examples include global datasets from the [Advanced National Seismic System](#) (ANSS) and regional datasets from the [Southern California Earthquake Data Center](#) (SCEDC) or the [Italian Seismic Bulletin](#). The data sets used in this tutorial are from [Woessner and Wiemer \(2005\)](#).

The reader should already be familiar with statistical concepts and tools presented in [Theme III - Statistical Foundations](#) ([Naylor et al. 2010](#)), as well as with the basic features of seismicity catalogs as presented in other articles of [Theme IV - Understanding Seismicity Catalogs and their Problems](#) ([Husen and Hardebeck 2010](#); [Woessner et al. 2010](#)). We assume that the catalogs only include tectonic events and that man-made contaminations have already been removed. This can often already be selected when downloading a catalog.

## 3 Expected Achievements

After completing this tutorial, the reader should be able to perform various completeness magnitude analyses and identify the advantages and pitfalls of different techniques. By understanding the importance of a correct estimate of  $M_c$  prior to any seismicity analysis, the reader will have the tools to verify his scientific hypotheses based on a robust dataset.

In the appendix we provide simple functions to compute  $M_c$  in the R Language,

which is freely available on all platforms. We also provide a small set of functions to calculate  $M_c$  in MATLAB requiring the Matlab Statistic Toolbox. The function set is included in the software that accompanies this article.

## 4 Theoretical Background

Earthquake catalogs are one of the most important products that seismic networks provide to the scientific community and society. An earthquake catalog is an inferred product that is by no means an easy-to-calibrate entity as a long processing chain from the detection of the first seismic waves to the determination of the location and magnitude of an event has to be passed, including subjective *a priori* and post-processing decisions. The detection capability of a network depends on the density and distribution of seismic stations, their site conditions, their recording characteristics, and their data link to the processing center. Assessing and quantifying this capability is a complicated task, as almost none of the effects of the links in the recording chain can be quantified without making assumptions. The earthquake catalog carries all uncertainties that are intrinsic in the assumptions, making the evaluation of the catalog completeness challenging.

The completeness magnitude  $M_c$  is theoretically defined as the lowest magnitude at which 100% of the earthquakes in a space-time volume are detected ([Rydelek and Sacks 1989](#)). In practice, a relatively high magnitude threshold may provide a conservative estimate of  $M_c$  but would deprive the dataset of potentially valuable information. One may consider that less than 100% of detectability is enough to define  $M_c$ . The most important aspect to consider is that deviation from the theoretical definition of  $M_c$  should not influence the feasibility nor reliability of subsequent seismicity analyses.

Why are not all earthquakes detected in the first place? Reasons are multifold: (1) the event is too small and its signal undistinguishable from the background noise on the seismograph, (2) the event is too small to be recorded on a sufficiently large number of stations - a minimum number of stations must be triggered to initiate the location procedure and thus the report of the event, (3) network operators decided that events below a certain threshold are not of interest, or (4) in case of an [after-shock](#) sequence, some events are too small to be detected within the coda of larger events (i.e. increased noise). To learn more about seismicity catalog generation, the reader should refer to the CORSSA article by [Husen and Hardebeck \(2010\)](#).

There are two distinguished classes of methods to evaluate  $M_c$ : (1) catalog-based methods (e.g., [Rydelek and Sacks 1989](#); [Woessner and Wiemer 2005](#); [Amorèse 2007](#)) (Section 5.1) and (2) network-based methods (e.g., [Kvaerna and Ringdal 1999](#); [Schorlemmer and Woessner 2008](#); [D'Alessandro et al. 2011](#)) (Section 5.2). Although both classes are presented below, this tutorial focuses on the former one

(Sections 6 and 7), in which  $M_c$  computation is straightforward and based on readily accessible parametric catalog data. Network-based techniques are generally time-consuming to be practical for most seismicity analyses. Moreover, we only consider completeness issues in [instrumental catalogs](#), which in general start in the mid 1970s. For completeness analysis of historical or paleoseismic data, which is key to probabilistic seismic hazard assessment, the reader should refer to e.g. [Albarello et al. \(2001\)](#), [Stucchi et al. \(2004\)](#) and references therein.

## 5 Methodologies: An Overview

### 5.1 Catalog-based Methods

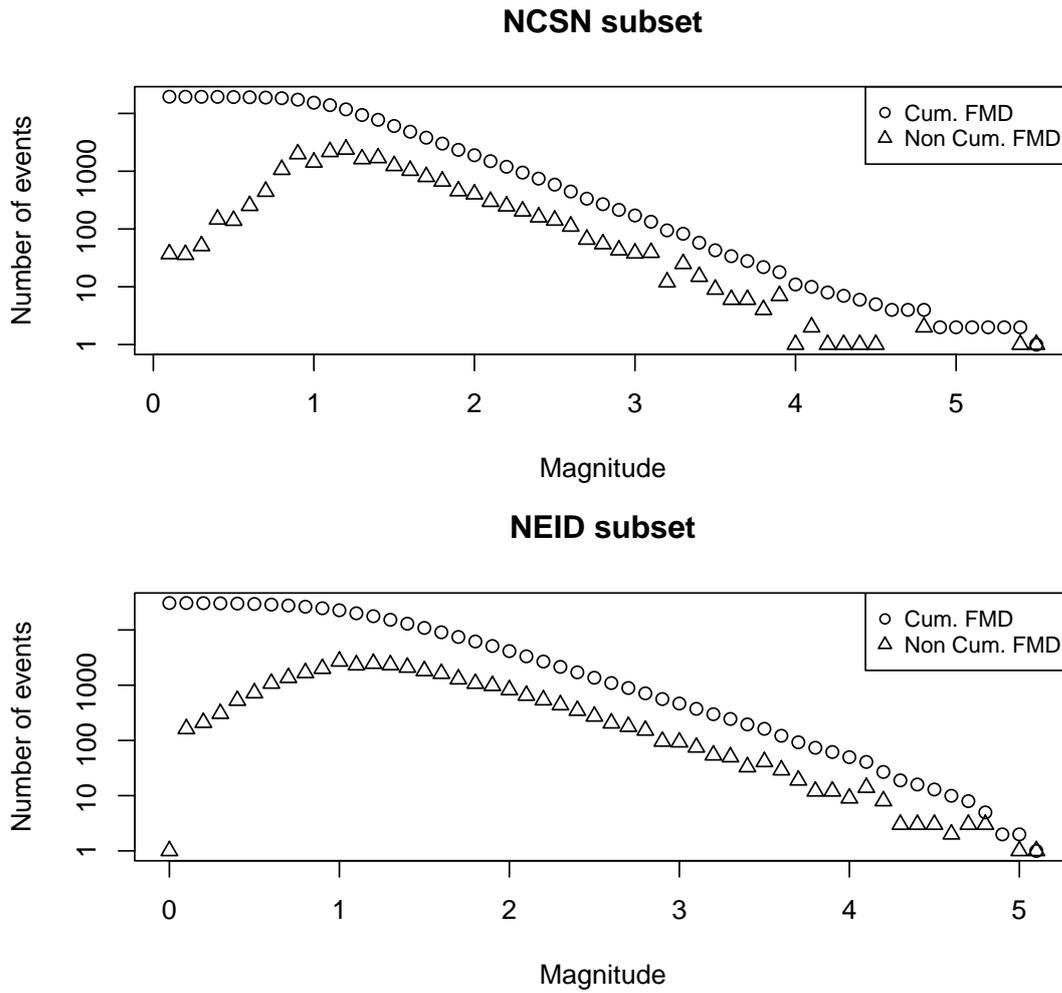
$M_c$  is often estimated by fitting a [Gutenberg-Richter](#) (G-R) model to the observed frequency-magnitude distribution (FMD). The magnitude at which the lower end of the FMD departs from the G-R law is taken as an estimate of  $M_c$  ([Zuniga and Wyss 1995](#)). It can be written as follows:

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

where  $N$  is the number of events with magnitude at least  $m$ ,  $a$  is the earthquake productivity and  $b$  describes the relative distribution of small and large earthquakes ([Gutenberg and Richter 1944](#); [Ishimoto and Iida 1939](#)). Events of magnitude  $m < M_c$  are discarded.

Figure 1 shows the FMD of a subset of the North California Seismic Network (NCSN) catalog and of a subset of the earthquake catalog maintained by the National Research Institute for Earth Science and Disaster Prevention (NIED); data sets are from [Woessner and Wiemer \(2005\)](#) who reviewed catalog-based methods to evaluate  $M_c$ . As discussed in Theme III - Statistical Foundations ([Naylor et al. 2010](#)), cumulative data should be treated with care. Each value in a cumulative quantity depends on all the preceding values. For this reason, we plotted the non-cumulative FMD in addition to the standard (cumulative) FMD. If we assume self-similarity (i.e. the G-R law),  $M_c$  is simply the magnitude increment at which the FMD departs from the linear trend in the log-lin plot. In the NCSN example, even a visual evaluation can lead to a correct estimate of the completeness magnitude ( $M_c \sim 1.2$ ). In contrast, the NIED example has a less pronounced departure from the linear trend which makes a visual estimate of  $M_c$  more difficult. It should be emphasized that we only consider deviations from linearity at small  $m$  (lower end of the FMD); deviations at large  $m$  (upper end of the FMD) can be due to statistical fluctuations due to under-sampling or to a real break in the G-R scaling (e.g., [Naylor et al. 2010](#); [Wesnousky 1994](#)).

A number of technical papers provide tools to compute  $M_c$ , based on the validity



**Fig. 1** FMD of the subset of the NCSN and NIED catalogs. If we assume self-similarity (i.e. the G-R law),  $M_c$  is simply the magnitude increment at which the lower end of the FMD departs from the linear trend in the log-lin plot. In the NCSN example, a visual evaluation can lead to a correct estimate of the completeness magnitude ( $M_c \sim 1.2$ ). This is however less trivial in the NIED example where the departure from the linear trend is less pronounced.

of the G-R law at small  $m$  (*Wyss et al. 1999; Wiemer and Wyss 2000; Cao and Gao 2002; Woessner and Wiemer 2005; Amorèse 2007*). Many variants have been developed in the scope of specific seismicity analyses (e.g., *Kagan 2003; Marsan 2003; Helmstetter et al. 2007*). Common techniques are described in Section 6. Noteworthy, the gradual curvature observed in some FMDs is commonly due to spatiotemporal heterogeneities in  $M_c$  (*Wiemer and Wyss 2000; Mignan et al. 2011*). These heterogeneities are due to the seismic network spatial configuration and its evolution through time (*Habermann 1987; Wiemer and Wyss 2000; Mignan et al.*

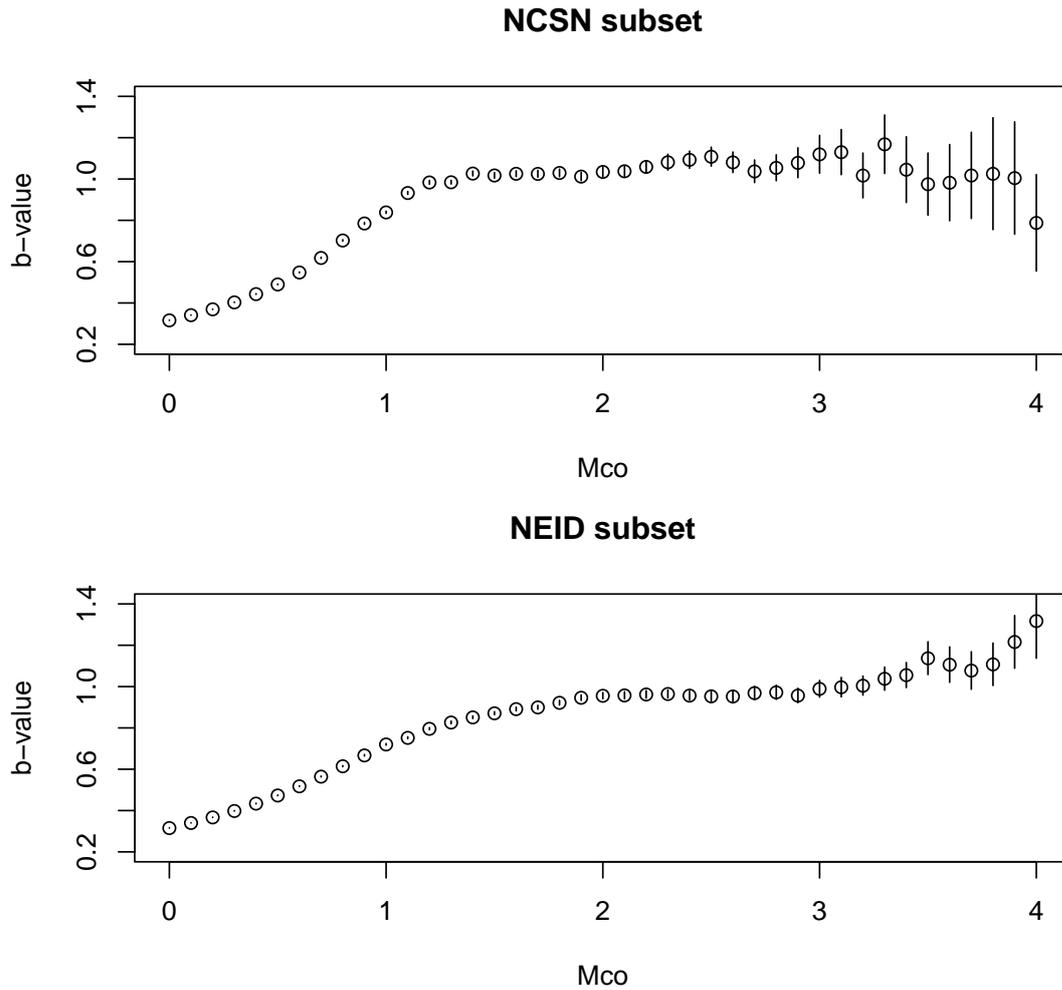
2011) and potentially to man-made effects such as the introduction of different magnitude scaling relationships (*Habermann 1987; Zuniga and Wyss 1995; Tormann et al. 2010*). Methods to evaluate  $M_c(x, y, z, t)$  are described in Section 7. Deviations from the G-R law could also be due to physical processes in the earth (e.g., *Rydelek and Sacks 1989*). For instance, the subset of the NIED catalog (Fig. 1) corresponds to a volcanic area, which might explain why the FMD does not strictly follow the G-R law. However any claim of a real deviation from self-similarity at low  $m$  based on bulk data is highly questionable as the most likely reason for a deviation from the G-R law is a spatiotemporal variation in  $M_c$  (*Wiemer and Wyss 2000*). An  $M_c$  test based on day-to-night noise modulation has been proposed by *Rydelek and Sacks (1989)*, an assumption that does not assume the self-similarity of seismicity. In this approach, the authors consider that the daytime noise level at a seismic station is increased relative to night-time due to human activity (see Section 6.6 for a detailed description of this technique). Contrary to other catalog-based methods in which  $M_c$  is inferred from the expected behavior of seismicity, the method by *Rydelek and Sacks (1989)* relates more closely to detectability issues of seismic stations. Methods directly based on seismic network information are described in Section 5.2.

### 5.1.1 Why is a correct estimate of $M_c$ important?

The choice of the minimum magnitude cutoff  $M_{co}$  has a direct impact on the evaluation of the  $b$ -value of the G-R law (Eq. 1), which in turn influences the evaluation of the  $a$ -value (i.e. overall *seismicity rate*). The G-R parameters are in general the basis of seismic hazard studies (e.g., *Cornell 1968; Wiemer et al. 2009*) and of earthquake forecast models (e.g., *Wiemer and Schorlemmer 2007*). Moreover a correct estimate of  $a$  and  $b$  is crucial to better understand the physics of the earth's crust (e.g., *Mignan 2011*). Figure 2 illustrates how the  $b$ -value estimate depends on the choice of the magnitude cutoff  $M_{co}$ . The  $b$ -value is computed by using the *maximum likelihood* technique:

$$b = \frac{\log_{10}(e)}{\langle M \rangle - (M_c - \frac{\Delta M}{2})} \quad (2)$$

where  $\langle M \rangle$  is the mean magnitude of the sample (with  $m \geq M_c$ ) and  $\Delta M$  is the binning width of the catalog (*Aki 1965*). For more details, we refer the reader to the CORSSA article by *Naylor et al. (2010)*. For  $M_{co} < M_c$ , the  $b$ -value is erroneous because the FMD is not a strict power-law (compare Figs. 1 and 2). For  $M_{co} \geq M_c$ , the  $b$ -value estimate stabilizes (around 1.0 in the NCSN and NIED data sets) before fluctuating again due to under-sampling at the higher end of the FMD. Looking at fluctuations in  $b$  is one possible approach to determine  $M_c$ , as proposed by *Wiemer*



**Fig. 2**  $b$ -value estimate as a function of magnitude cutoff  $M_{co}$  for the subset of the NCSN and NIED catalogs. For  $M_{co} < M_c$ , the  $b$ -value is erroneous because the FMD is not a strict power-law. For  $M_{co} \geq M_c$ , the  $b$ -value estimate starts to stabilize (around 1.0 in the 2 examples) before fluctuating again due to under-sampling at the higher end of the FMD. Error bars represent  $\pm 1\sigma$ , standard deviation obtained from bootstrapping (Efron 1979).

and Wyss (2000) or Cao and Gao (2002) (see Section 6). Note that seismic networks in general improve with time (e.g., Hutton *et al.* 2010) and more smaller events are detected, leading to an apparent increase in  $a$ -value if the change in  $M_c$  is not considered.

### 5.1.2 Should we dismiss information below $M_c$ or is this incomplete information of any value?

While some models have been proposed to define the detection rate of events (e.g., [Ogata and Katsura 1993](#)) (see Section 6), standard statistical methods may result in doubtful interpretations because events below the  $M_c$  threshold experience many unreported significant changes and variations ([Kagan 2002](#)), in particular due to subjective decisions in the event processing chain. Thus disregarding data below completeness is good practice when drawing conclusions about the dynamics seismicity. However, for some techniques the information is necessary ([Ogata and Katsura 1993](#); [Woessner and Wiemer 2005](#)) in the modeling process and to model the entire frequency-magnitude distribution.

## 5.2 Network-based Methods - Using Phase-Pick or Waveform Data

Completeness can also be assessed from a network detection perspective; in this view, the primary focus is not to define the completeness level of the data set - the target is rather to define the probability level with which an earthquake can be detected given the station (sensor) sensitivity and station distribution. The result is then not anymore a function of the earthquake sample but a function of the seismic network properties. The approaches can again be separated in the type of data and methodology used:

1. Waveform based techniques have been developed taking advantage of the signal-to-noise ratio measured at each network station. These types of approaches have been developed mainly in the framework of monitoring the global nuclear test ban treaty. The literature relating to these methods is too vast to provide a complete review (e.g., [Ringdal 1975, 1986](#); [von Seggern and Blandford 1976](#); [Ringdal and Kvaerna 1989](#); [Sereno and Bratt 1989](#); [Gomberg 1991](#); [Kvaerna and Ringdal 1999](#); [Kvaerna et al. 2002a,2002b](#); [von Seggern 2004](#)). Moreover, in most of these studies, the completeness magnitude is not the main focus or not explicitly considered. Thus we will not describe this further.
2. One method uses a waveform forward simulation technique ([D'Alessandro et al. 2011](#)) that can also be used to estimate completeness though it is primary designed to assess [hypocenter](#) location accuracy. We do not describe this method.
3. Using phase-pick data, [Schorlemmer and Woessner \(2008\)](#) proposed the probabilistic magnitude of completeness (PMC) method to assess detection capabilities of a seismic network using empirical data. This method has been applied to various data sets (e.g., [Nanjo et al. 2010b](#); [Schorlemmer et al. 2010b](#); [Plenkers et al. 2011](#)) and is illustrated in short in the next section.

All these methods are broader in the sense that they are not targeted to the completeness magnitude itself. These methods can be used to design and to evaluate particular performance measure of the seismic network.

### 5.2.1 Probabilistic Magnitude of Completeness (PMC)

[Schorlemmer and Woessner \(2008\)](#) proposed the probabilistic magnitude of completeness (PMC) method to assess detection capabilities of a seismic network. Within this method, the detection probability of each network station for any magnitude at a given distance from the station needs to be defined, then the probability of detecting an earthquake occurring at a specific location can be computed by basic combinatorics. The method uses empirical data (phase data, station metadata, and attenuation relation used for magnitude determination) and results in three essential products:

1. Station detection probabilities  $P_D(m, L)$  of a seismic station as a function of magnitude  $m$  and hypocenter-station distance  $L$ : Characteristic probability distributions are derived from phase picks for each station describing probabilities of detecting an earthquake of a given magnitude at a given distance from the station. Detection means to have a phase determined at a particular station. For each earthquake during the on-time of a station, a data triplet is generated, consisting of three parameters: magnitude  $m$  of the earthquake, hypocentral distance  $L$  and a binary value for phase picked or not picked, plotted as a green dot or red dot, respectively (Figure 3).
2. Earthquake detection probabilities  $P_E(m, \mathbf{x}, t)$  of an earthquake of magnitude  $m$  at a given location  $\mathbf{x}$  and time  $t$  based on the available network stations: Combining the probability distributions  $P_D(i, m, L)$  of all stations  $i$  using basic combinatoric principles, probabilities of detecting events of a given magnitude at the minimum number of required stations can be mapped. An example for detection probabilities for different magnitude levels is shown in Figure 4.
3. The probability-based magnitude of completeness  $M_P(\mathbf{x}, t)$  for a predefined probability level  $P$  at a given location  $\mathbf{x}$  and time  $t$  based on the available network stations: It equals the lowest magnitude for which the probability of detection  $P_E(m, \mathbf{x}, t)$  is  $1 - Q$ , with  $Q$  being the complementary probability that an event will not be detected:

$$M_p(\mathbf{x}, t, Q) = \min_{m \in M} m \mid P_E(m, \mathbf{x}, t) = 1 - Q \quad (3)$$

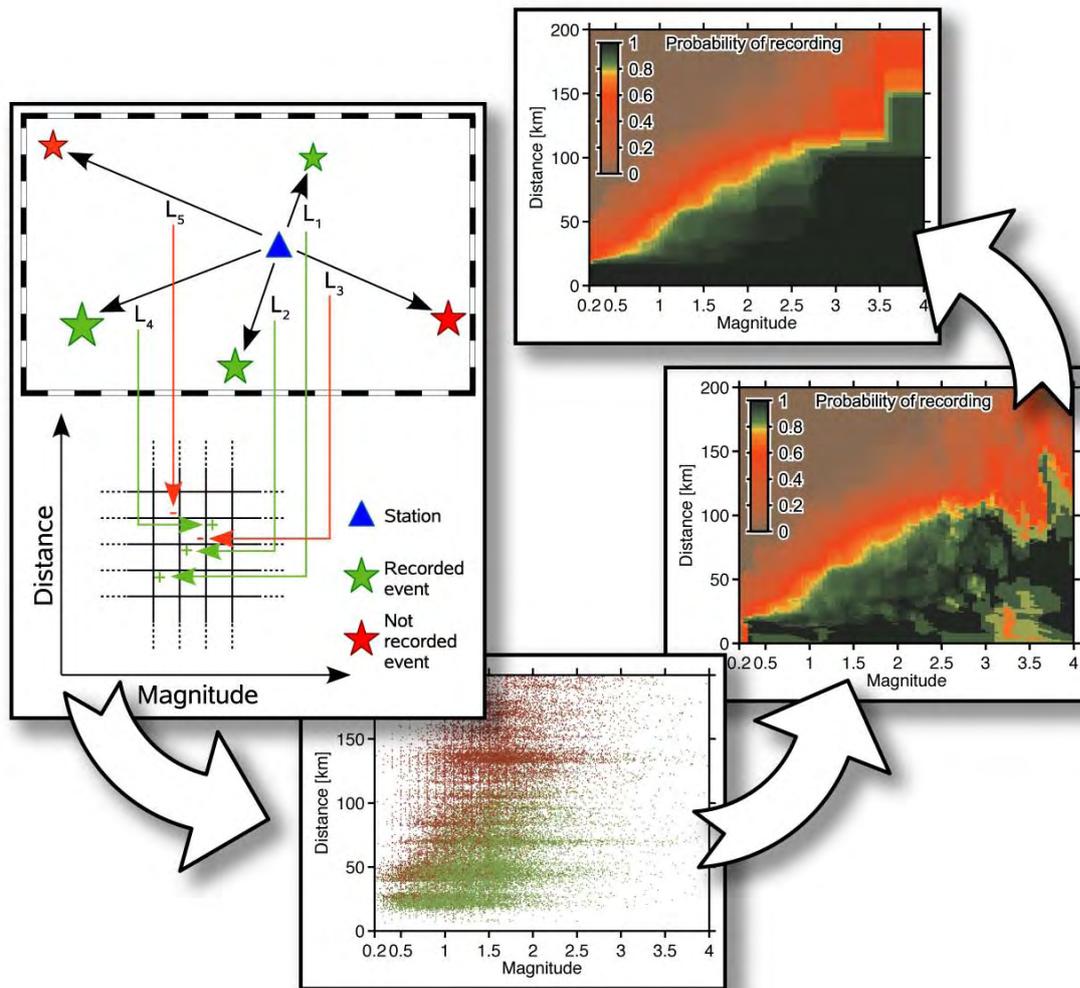
where  $M$  is the interval of possible completeness magnitudes.

Seismic networks are subject to changes, e.g. new stations are deployed while others are removed or exchanged. Each new configuration leads to a change in detection probabilities, such that detection probabilities are only valid for a particular network configuration. PMC, as well as other network based methods, are able to account for this and thus monthly  $M_P(\mathbf{x}, t, Q)$  maps are calculated for Southern California, Italy and Switzerland (accessible at <http://completeness.usc.edu/>). In general every time the network is changed, maps should be recomputed, yet this is impossible to track for an independent resource outside the authoritative agency.

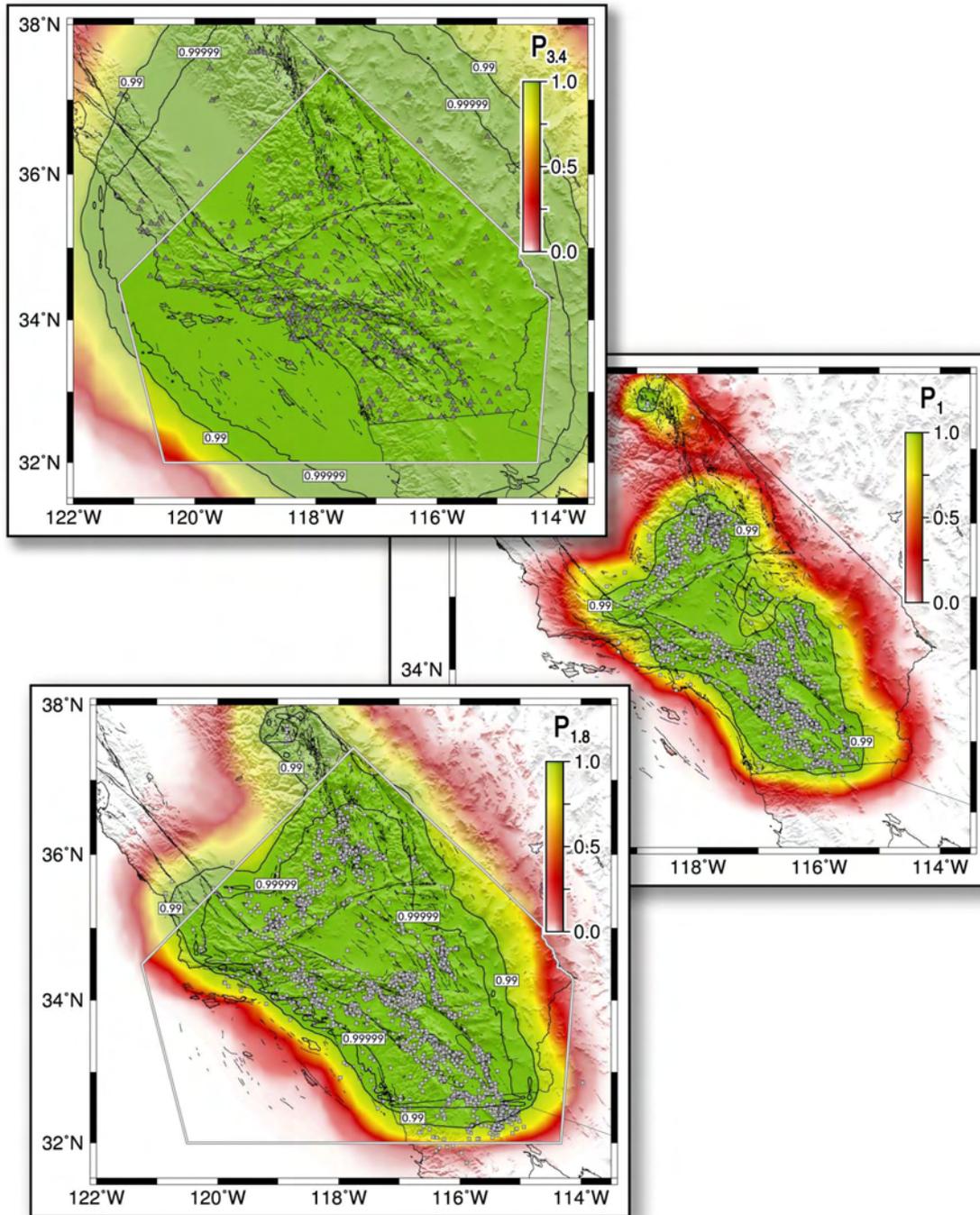
### 5.2.2 Assumptions, Features and Limits of PMC

The PMC approach is considered the faster out of the network-based approaches since it only uses empirical and parametric data, not waveforms. Yet, there are some assumptions and open issues that need to be considered when applying the method:

1. PMC assumes isotropic attenuation while it has been shown to be anisotropic. PMC can account for this when accounting for azimuthal dependency as shown in *Plenkers et al. (2011)*. The amount of data may however play a limiting factor here.
2. The initial amount of data to generate the station detection probabilities  $P_D(m, L)$  has a lower limit yet the lower limit is not quantified as well as its influence on the overall earthquake detection probability  $P_E$  nor the  $M_P$ .
3. The smoothing approach to create  $P_D$  tends to overestimate the station detection probabilities for very low magnitudes and short distances which can lead to small values of  $M_P(\mathbf{x}, t, Q)$ . *Bachmann (2011)* (Chapter 6) evaluated this issue and describes how to perform sensitivity analyses to assess an optimal value of  $Q$  considering the lower detection probability thresholds of stations that should still be considered in the detection process.
4. Computation time may become an issue in case of large networks since the number of combinations to compute  $P_E$  increases dramatically (*Bachmann 2011*).
5. Large aftershock sequences or induced seismicity may alter the station detection probabilities  $P_D$ , mainly reducing detection probabilities in narrow distance bands over the entire magnitude range of interest. Removing this feature during data analysis is a requisite but cumbersome and an automated solution has yet to be found (*Bachmann 2011*).
6. Comparison between completeness estimates from catalog-based approaches and  $M_P(\mathbf{x}, t)$  is not trivial since the latter is explicitly *time-dependent*. *Nanjo et al. (2010b)*; *Schorlemmer et al. (2010b)* show how to compare  $M_P$  and  $M_c$  maps, shortly discussed in Section 7.



**Fig. 3** Methodology sketch to assess station detection probability  $P_D(m, L)$ . (Left) Data triplets (picked [plus triplet, green] / not picked [minus triplet, red], hypocentral distance to station, magnitude) are generated for each event that occurred while a station (blue triangle) was in operation. If an event was picked (green star) or not picked (red star) at the station, we generate a plus triplet or minus triplet, respectively. (Bottom) Plot of all data triplets of station POB from the Southern California Seismic Network (SCSN) for the period 1 January 2001 to 1 July 2007. Green dots indicate plus triplets (picked events); red dots indicate minus triplets (not picked events). (Right) Station detection probability matrix as function of magnitude and distance for station POB derived from raw data triplets. (Top) Smoothed  $P_D(m, L)$  that enters further calculations. Modified with courtesy from [Schorlemmer and Woessner \(2008\)](#).



**Fig. 4** Map of detection probabilities,  $P_E(M, x, t)$ , for different magnitudes on 1 July 2007 at a depth of 7.5 km. Gray boxes mark events of magnitude under consideration from the period 1 January 2001–1 July 2007. Gray triangles mark stations in operation on 1 July 2007. The white-gray polygon indicates the authoritative region of SCSN. (Top) Map of  $P_E(m = 3.4)$ . The two black lines mark the  $P = 0.99$  and  $P = 0.99999$  contours. (Center) Map of  $P_E(m = 1)$ . The black line marks the  $P = 0.99$  contour. (Bottom) Map of  $P_E(m = 1.8)$ . The two black lines mark the  $P = 0.99$  and  $P = 0.99999$  contours. The SCSN intends to be complete at the  $M_L = 1.8$  level for its authoritative region. Courtesy of [Schorlemmer and Woessner \(2008\)](#).

## 6 Catalog-based Techniques to Evaluate $M_c$

We present a non-exhaustive but representative list of catalog-based techniques to assess  $M_c$ :

1. the Maximum Curvature (MAXC) technique (e.g., [Wiemer and Wyss 2000](#)) (Section 6.1),
2. the Goodness-of-Fit Test (GFT) ([Wiemer and Wyss 2000](#)) (Section 6.2),
3. the  $M_c$  by  $b$ -value stability (MBS) approach ([Cao and Gao 2002](#)) (Section 6.3),
4.  $M_c$  from the Entire Magnitude Range (OK1993, EMR) ([Ogata and Katsura 1993](#); [Woessner and Wiemer 2005](#)) (Section 6.4),
5. the Median-based analysis of the segment slope (MBASS) ([Amorèse 2007](#)) (Section 6.5)
6. and the day-to-night noise modulation (day/night method) ([Rydelek and Sacks 1989](#)) (Section 6.6).

All techniques, except for the [Rydelek and Sacks \(1989\)](#) method, are based on the validity of the G-R law (Fig. 5). The main distinction is whether they are parametric (GFT, MBS, EMR) or non-parametric (MAXC, MBASS). Parametric techniques are based on fitting the FMD while non-parametric techniques are based on the evaluation of changes in the FMD (e.g., possible breaks in the slope). At the present time, there is no consensus on which technique to use to compute  $M_c$  and different techniques may provide significantly different results (Fig. 5, Table 1). Table 2 summarizes the pros and cons of each technique. General issues linked to  $M_c$  reliability and uncertainty are discussed using synthetic data in Section 6.7.

The different techniques are tested on the data sets from [Woessner and Wiemer \(2005\)](#), which correspond to subsets of the Earthquake Catalog of Switzerland (ECOS), the Northern California Seismic Network (NCSN) regional catalog, the earthquake catalog of a volcanic area maintained by the National Research Institute for Earth Science and Disaster Prevention (NIED), the Harvard Centroid Moment Tensor (CMT) global catalog and the International Seismological Centre (ISC) global catalog (see [Woessner and Wiemer \(2005\)](#) for more details).  $M_c$  estimates from FMD-based techniques are listed in Table 1. Uncertainties are also considered, using the [Monte Carlo](#) approximation of the bootstrap method ([Efron 1979](#); [Woessner and Wiemer 2005](#)).

### 6.1 Maximum Curvature (MAXC) technique

The Maximum Curvature (MAXC) technique ([Wyss et al. 1999](#); [Wiemer and Wyss 2000](#)) is a fast and straightforward way to estimate  $M_c$  and consists in defining the point of the maximum curvature by computing the maximum value of the first

	ECOS	NCSN	NIED	CMT	ISC
MAXC	$1.16 \pm 0.10$	$1.20 \pm 0.00$	$1.00 \pm 0.00$	$5.31 \pm 0.03$	$4.30 \pm 0.01$
GFT	$1.39 \pm 0.17$	$1.12 \pm 0.04$	$1.38 \pm 0.06$	$5.30 \pm 0.00$	$4.66 \pm 0.05$
MBS	$1.53 \pm 0.16$	$1.48 \pm 0.10$	$1.97 \pm 0.10$	$5.80 \pm 0.09$	$4.80 \pm 0.00$
EMR	$1.48 \pm 0.08$	$1.29 \pm 0.03$	$1.40 \pm 0.01$	$5.49 \pm 0.02$	-
MBASS	$1.28 \pm 0.29$	$1.25 \pm 0.34$	$1.42 \pm 0.13$	$5.44 \pm 0.07$	$5.02 \pm 0.45$

**Table 1**  $M_c$  estimates and standard deviation  $\sigma_0$  obtained from 200 bootstrap samples (Efron 1979) for different FMD-based techniques and data sets. We test the following techniques: Maximum Curvature (MAXC) technique, Goodness-of-Fit Test (GFT),  $M_c$  by  $b$ -value stability (MBS), Entire-magnitude range (EMR) technique and Median-based analysis of the segment slope (MBASS) - We use the same earthquake catalog subsets as in Woessner and Wiemer (2005). All FMDs are shown in Figure 5.

Technique	Main ref.	Pros	Cons
MAXC	Wiemer and Wyss (2000)	non-parametric straightforward statistically robust	underestimates $M_c$ in bulk data
GFT	Wiemer and Wyss (2000)	G-R deviation definition	90% conf. not always reached may underestimate $M_c$
MBS	Cao and Gao (2002)	based on $b$ -value stability	may overestimate $M_c$ relatively high uncertainty
EMR	Woessner and Wiemer (2005)	complete FMD model	assumption below $M_c$ 4 parameters to fit
MBASS	Amorèse (2007)	non-parametric	main discontinuity may not be $M_c$ relatively high uncertainty
Day/Night	Rydelek and Sacks (1989)	does not assume G-R law	requires declustering unstable (on bulk data)

**Table 2** Pros and cons of some common catalog-based techniques to evaluate  $M_c$ . See text for more details about each one of these techniques (and other ones).

derivative of the frequency-magnitude curve. In practice, this matches the magnitude bin with the highest frequency of events in the non-cumulative FMD.

Despite the easy applicability of this approach,  $M_c$  is underestimated in the case of gradually curved FMDs and the use of other techniques providing more conservative estimates, such as GFT or EMR, has been suggested (Wiemer and Wyss 2000; Woessner and Wiemer 2005). MAXC provides most of the lowest  $M_c$  estimates in Table 1. However Mignan et al. (2011) showed that MAXC does not underestimate  $M_c$  (compared to the MBASS technique) when considering a local data set in which heterogeneities in  $M_c$  are minimized. It would suggest that the gradual curvature of a bulk FMD is indeed due to  $M_c$  heterogeneities and that the MAXC technique is valid in high-resolution  $M_c$  mapping (see Section 7). The MAXC technique has also the advantage to require fewer events than other techniques to reach a stable result (Mignan et al. 2011). This is discussed in more details in Section 6.7.

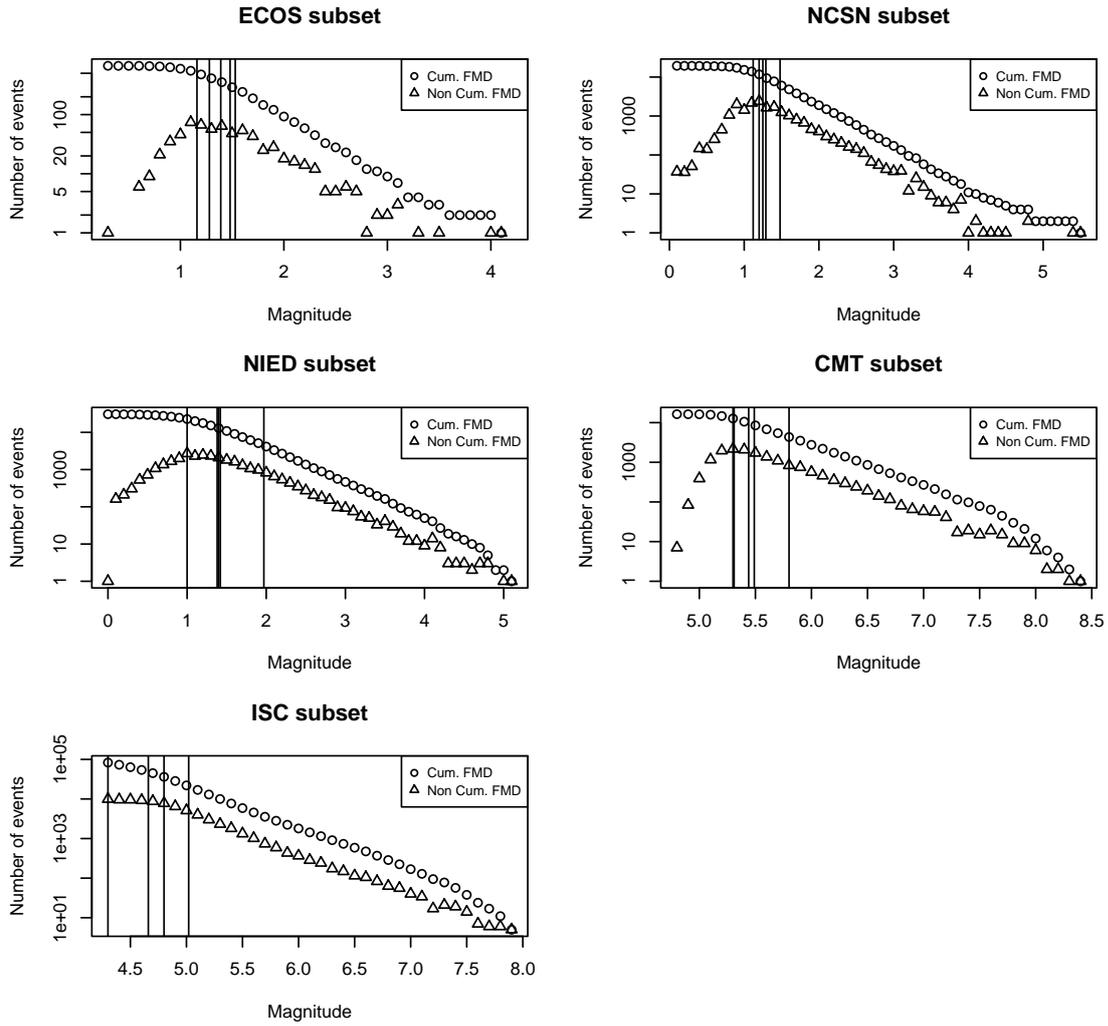
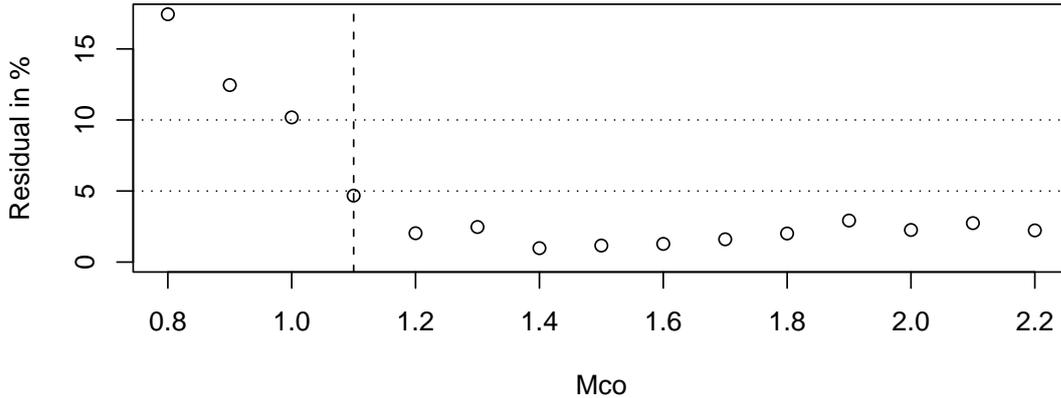


Fig. 5 FMDs of the data sets listed in Table 1 and corresponding  $M_c$  estimates (vertical lines).

## 6.2 Goodness-of-Fit Test (GFT)

The Goodness-of-fit test (GFT), proposed by *Wiemer and Wyss (2000)*, calculates  $M_c$  by comparing the observed FMD with synthetic ones. The goodness-of-fit is evaluated by the parameter  $R$ , absolute difference of the number of events in each magnitude bin between the observed and synthetic G-R distributions. Synthetic distributions are calculated using estimated  $a$ - and  $b$ -values of the observed dataset



**Fig. 6** GFT technique (*Wiemer and Wyss 2000*) - Residual ( $100 - R$ ) as a function of minimum magnitude cutoff  $M_{co}$  for the NCSN catalog subset.  $M_c$  (vertical line) is found at the first magnitude cutoff at which the confidence  $R = 95\%$  is reached.

for  $M \geq M_{co}$  as a function of ascending cutoff magnitude  $M_{co}$ .

$$R(a, b, M_{co}) = 100 - \left( \frac{\sum_{M_{co}}^{M_{max}} |B_i - S_i|}{\sum_i B_i} 100 \right) \quad (4)$$

where  $B_i$  and  $S_i$  are the observed and predicted cumulative number of events in each magnitude bin.

$M_c$  is found at the first magnitude cutoff at which the observed data for  $M \geq M_{co}$  is modeled by a straight line (in log-lin plot) for a fixed confidence level, e.g.  $R = 90\%$  or  $95\%$ . Figure 6 shows the residual (or  $100 - R$ ) at different magnitude cutoffs for the NCSN subset.  $M_c$  is defined in this example as the first magnitude bin at which the residual falls below the horizontal line of the 95% fit. It is interesting to note that for higher magnitude cutoffs, the residual remains low, which corresponds to the G-R part of the FMD. When the 95% level of fit is not obtained, a 90% level is a compromise. If a 90% confidence cannot be reached, the MAXC estimate is used instead (ZMAP software (*Wiemer 2001*) definition). *Woessner and Wiemer (2005)* showed that the GFT-90% approach leads to  $M_c$  estimates on the lower end of the  $M_c$  distribution compared to other techniques, with results close to the MAXC ones (see also Table 1).

*Kagan (2003)* proposed an approach similar to the GFT but based on a more conventional statistical method.

### 6.3 $M_c$ by $b$ -value stability (MBS)

*Cao and Gao* (2002) estimated  $M_c$  using the stability of the  $b$ -value as a function of cutoff magnitude  $M_{co}$ , referred to as MBS by *Woessner and Wiemer* (2005). This model is based on the assumption that  $b$ -value estimates ascend for  $M_{co} < M_c$  and remain constant for  $M_{co} \geq M_c$ . If  $M_{co} < M_c$ , the resulting  $b$ -value is incorrect. As  $M_{co}$  approaches  $M_c$ , the  $b$ -value approaches its true value and remains constant for  $M_{co} > M_c$ , forming a plateau (see Figure 2).

The authors arbitrarily defined  $M_c$  as the magnitude for which the change in  $b$ -value  $\Delta b$  between two successive magnitude bins is smaller than 0.03. *Woessner and Wiemer* (2005) found this criterion to be unstable since the frequency of events in single magnitude bins can vary strongly. To base the approach on an objective measure and to stabilize it numerically, *Woessner and Wiemer* (2005) used the  $b$ -value uncertainty  $\delta b$  according to *Shi and Bolt* (1982) as criterion:

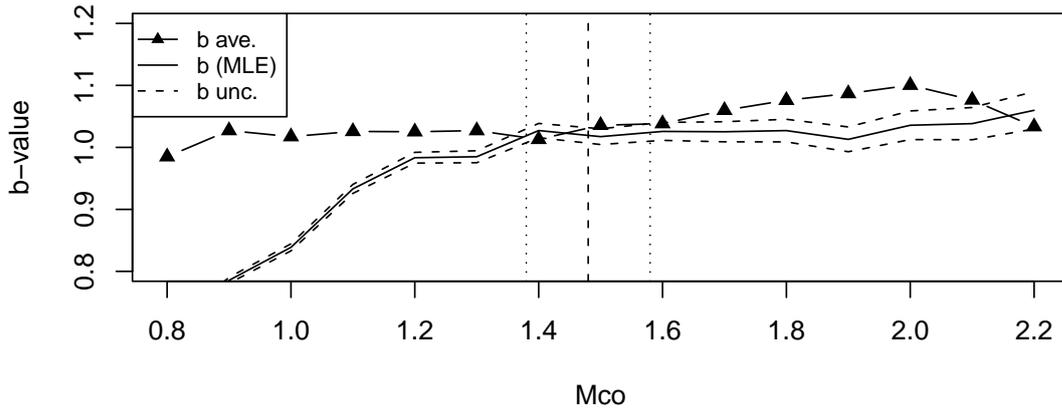
$$\delta b = 2.3b^2 \sqrt{\frac{\sum_{i=1}^N (M_i - \langle M \rangle)^2}{N(N-1)}} \quad (5)$$

with  $\langle M \rangle$  being the mean magnitude and  $N$  the number of events (note that the second square exponent is missing in Eq. 4 of *Woessner and Wiemer* (2005)).  $M_c$  is then defined as the first magnitude increment at which  $\Delta b = |b_{ave} - b| \leq \delta b$  (Figure 7). The arithmetic mean  $b_{ave}$  is calculated from  $b$ -values of successive cutoff magnitudes  $M_{co}$  in half a magnitude range  $dM = 0.5$  such as  $b_{ave} = \sum_{M_{co}}^{M_{co}+dM} b(M_{co}) \Delta m / dM$  for a bin size  $\Delta m = 0.1$ . Large magnitude ranges are preferable, and would be justified for FMDs that perfectly obey a power-law. *Woessner and Wiemer* (2005) found that the MBS method leads to the highest  $M_c$  values compared to other techniques, in agreement with results of Table 1.

*Marsan* (2003) introduced a method computing the  $b$ -value and the log-likelihood of completeness for earthquakes above a certain cutoff magnitude. The log-likelihood of completeness is defined as the logarithmic probability that the G-R law fitted to the data above the cutoff magnitude can predict the number of earthquakes in the magnitude bin just below the cutoff magnitude.  $M_c$  is chosen so that the  $b$ -value drops for  $m < M_c$  and the log-likelihood drops for  $m = M_c$ . The method is similar to the MBS one.

### 6.4 $M_c$ from the Entire Magnitude Range (OK1993, EMR)

*Woessner and Wiemer* (2005) proposed a method to estimate  $M_c$  that uses the entire magnitude range (EMR), thus including events below  $M_c$ . They provided a model consisting of two parts: the G-R law for the complete part, and the cumulative



**Fig. 7** MBS technique (*Cao and Gao 2002*) modified by *Woessner and Wiemer (2005)* -  $b$ -value estimate as a function of minimum magnitude cutoff  $M_{co}$  for the NCSN catalog subset.  $M_c$  is the first magnitude increment at which  $\Delta b = |b_{ave} - b| \leq \delta b$  (dashed uncertainty envelope). The dashed vertical line represents the mean  $M_c$  and the dotted lines  $\pm 1\sigma$  obtained from bootstrapping (see Table 1).

normal distribution for the incomplete part of the non-cumulative FMD. The model attempts to reproduce the entire frequency-magnitude distribution, thus fits the incompletely observed part, a technique which has been questioned (*Kagan 2002*). The EMR approach is similar to that of *Ogata and Katsura (1993)* (hereafter referred to as OK1993).

The non-cumulative FMD can be described by the intensity  $\lambda$  (normalized number of events) at magnitude  $m$  as

$$\lambda(m) = \lambda_0(m)q(m) \quad (6)$$

with

$$\lambda_0(m | \beta) = \exp(-\beta m) \quad (7)$$

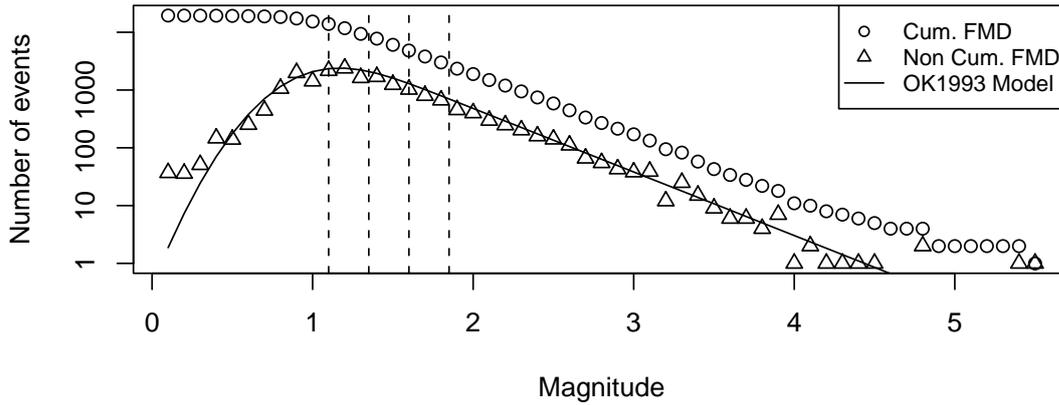
the G-R law,  $\beta = b \log 10$  and  $q(m)$  a detection function with  $0 \leq q \leq 1$ .  $q$  is commonly defined as the cumulative normal distribution of mean  $\mu$  and standard deviation  $\sigma$  (e.g., *Ogata and Katsura 1993, 2006; Iwata 2008*)

$$q(m | \mu, \sigma) = \int_{-\infty}^m \frac{1}{\sqrt{2\pi}\sigma} \exp \frac{-(x - \mu)^2}{2\sigma^2} dx \quad (8)$$

By substituting Eqs. 7 and 8 in Eq. 6, OK1993 provided a model to fit the FMD over the entire magnitude range. The completeness magnitude is only implicit with

$$M_c(n) = \mu + n\sigma \quad (9)$$

where  $n$  indicates the confidence level.  $n = 0$  means that 50% of the events are detected above  $M_c$ .  $n = (1, 2, 3)$  means that 68%, 95% and 99% of the events are



**Fig. 8** Observed and predicted FMD for the subset of the NCSN catalog. Parameters of the OK1993 model are determined from the maximum-likelihood method (by using Eq. 10):  $b = 1.1$ ,  $\mu = 1.1$  and  $\sigma = 0.25$ . It yields  $M_c(3) = 1.60$  (95% confidence) and  $M_c(3) = 1.85$  (99% confidence). Vertical dashed lines represent  $\mu$ ,  $\mu + 1\sigma$ ,  $\mu + 2\sigma$  and  $\mu + 3\sigma$ .

detected respectively. The parameters  $\theta = (\beta, \mu, \sigma)$  are simultaneously obtained by maximizing the log-likelihood function  $\log L(\theta) = \sum_i \log f(m_i | \theta)$  with the normalized density function  $f(m | \theta) = c\lambda(m | \theta)$ ,  $c$  being a normalization factor. In the OK1993 model:

$$f(m | \beta, \mu, \sigma) = \beta \exp(-\beta(m - \mu) - \beta^2 \frac{\sigma^2}{2}) q(m | \mu, \sigma) \quad (10)$$

(*Ogata and Katsura 2006*). Figure 8 shows the observed and predicted FMD in the case of the NCSN subset, with the OK1993 model parameters determined by using Eq. 10. The OK1993 model parameters for the ECOS, NCSN, NIED and CMT subsets are listed in Table 3. For the ISC subset, the OK1993 model is not valid as the incomplete part of the FMD does not show a decrease in the number of events for low  $m$  (Fig. 5).

	ECOS	NCSN	NIED	CMT	ISC
$\mu$	1.2	1.1	1.1	5.2	-
$\sigma$	0.25	0.25	0.35	0.15	-
$b$	1.1	1.1	0.9	1.0	-
$M_c(95\%)$	1.70	1.60	1.80	5.50	-
$M_c(99\%)$	1.95	1.85	2.15	5.65	-
$M_c^{EMR}$	$1.48 \pm 0.08$	$1.29 \pm 0.03$	$1.40 \pm 0.01$	$5.49 \pm 0.02$	-

**Table 3** OK1993 model parameter estimates for different catalog subsets. Parameters are obtained from a maximum likelihood approach (Eq. 10) with tested values incremented every 0.1 for  $b$  and  $\mu$  and every 0.05 for  $\sigma$ .  $M_c$  is determined from Eq. 9.  $M_c^{EMR}$  is given for comparison.

In the EMR method, [Woessner and Wiemer \(2005\)](#) use the following model, where  $M_c$  is explicit:

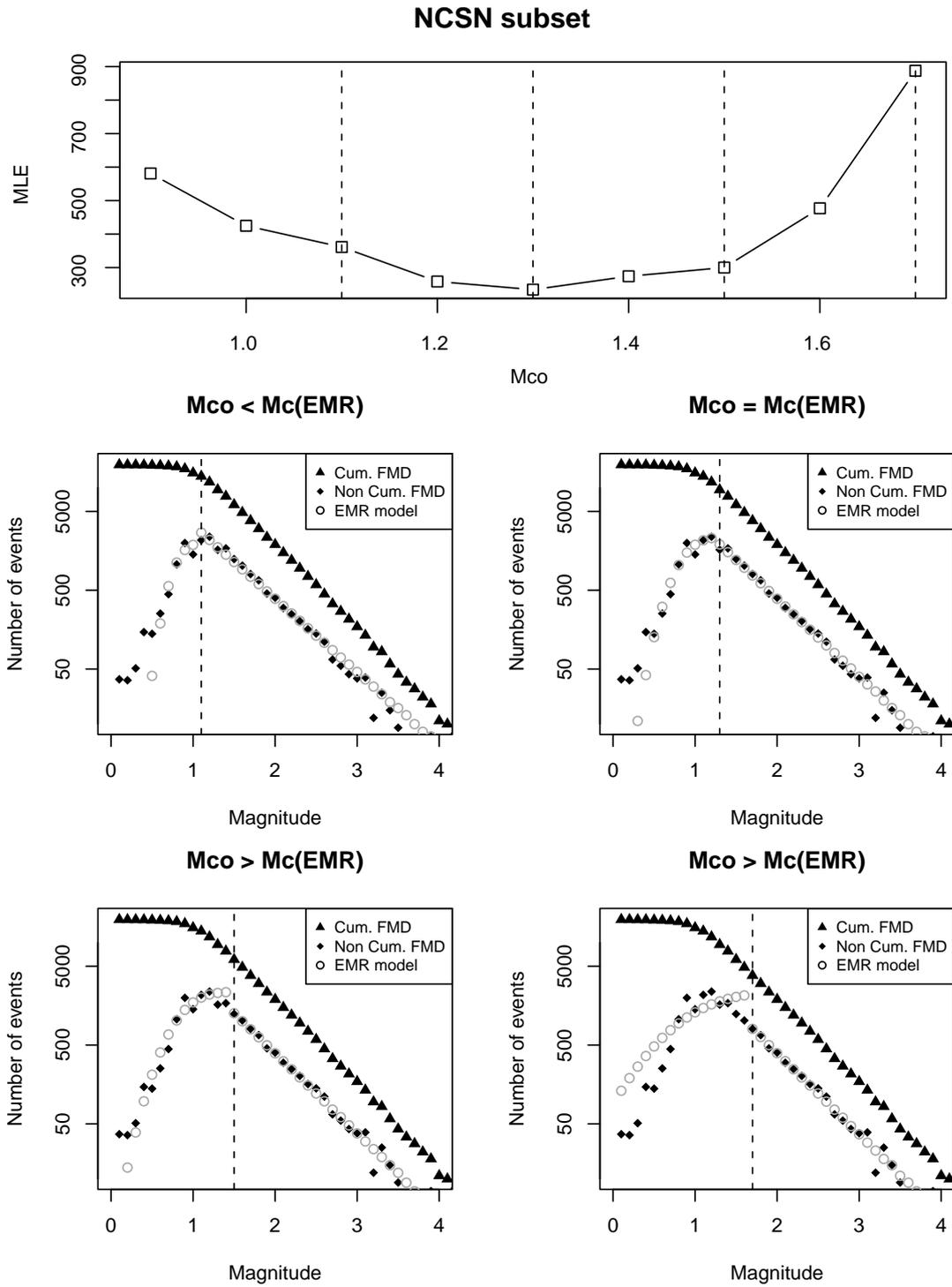
$$\lambda(m | \mu, \sigma, m) = \begin{cases} \lambda(m | \mu, \sigma, m < M_c) = \int_{-\infty}^{M_c} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx \\ \lambda(m | \mu, \sigma, m \geq M_c) = \exp(-\beta(m - M_c)) \end{cases} \quad (11)$$

Figure 9 shows the maximum likelihood estimates and sample frequency-magnitude plots for ascending cut-off magnitudes of the EMR-model for estimate for the NCSN catalog subset. For thresholds  $M_{co}$  above the best completeness value, the EMR-method does not describe the gradual curvature of the FMD correctly by not multiplying the detection function to the theoretical G-R law (compare Eqs. 6 and 11). Nonetheless, EMR results are in the range of  $M_c$  values expected by other techniques (Table 1). Noteworthy,  $M_c$  estimates derived from the OK1993 model are more conservative (Table 3), i.e. the  $M_c$  values are generally larger than the ones from EMR.

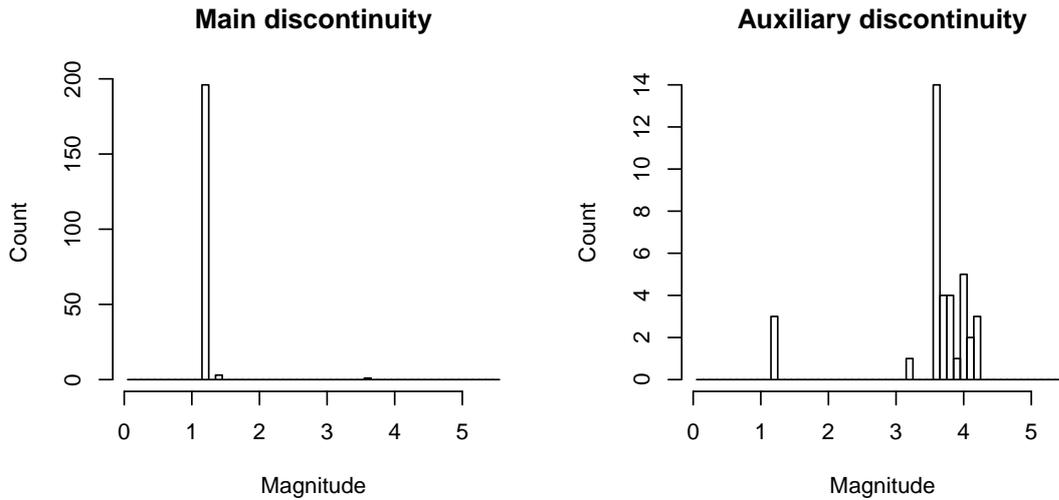
### 6.5 Median-Based Analysis of the Segment Slope (MBASS)

The Median-Based Analysis of the Segment Slope (MBASS), proposed by [Amorèse \(2007\)](#), is a non-parametric technique to estimate  $M_c$ . It is based on an iterative method designed to search for multiple change points in the non-cumulative FMD. The null hypothesis acceptance or rejection is based on the Wilcoxon rank sum test. Different discontinuities in the FMD can be found by the MBASS technique, with the main discontinuity corresponding to  $M_c$ . Other discontinuities may correspond to upper-magnitude breakpoints (e.g., [Wesnowsky 1994](#)). Figure 10 shows the MBASS statistics, i.e. the distribution of the main and auxiliary discontinuities, for the NCSN catalog subset. [Amorèse \(2007\)](#) tested the MBASS method on the same catalogs used by [Woessner and Wiemer \(2005\)](#) and showed that the MBASS and EMR techniques give similar results when uncertainty is taken into account. This is verified in Table 1, which also indicates that MBASS uncertainties are larger compared to other techniques. The MBASS algorithm (in R language) is published in [Amorèse \(2007\)](#).

The MBASS technique is used in [Mignan et al. \(2011\)](#) where results are compared with the ones from the non-parametric MAXC technique. The authors used the difference  $\Delta = M_c^{MBASS} - M_c^{MAXC}$  as a proxy to the degree of curvature of the FMD and showed that  $\Delta$  tends to zero when spatiotemporal heterogeneities in  $M_c$  are minimized. For bulk data sets, the MBASS technique provides a conservative estimate compared to the MAXC technique.



**Fig. 9** EMR technique (*Woessner and Wiemer 2005*) - Maximum likelihood estimate of  $M_c$  for the subset of the NCSN catalog based on the EMR model (Eq. 11), not to confound with the original OK1993 model (Figure 8). By using Eq. 11 instead of Eq. 6, the EMR method fails to describe the gradual curvature of the FMD above  $m = M_c^{MAXC} = 1.2$ . Vertical lines represent  $M_{co} = 1.1, 1.3, 1.5$  and  $1.7$ .



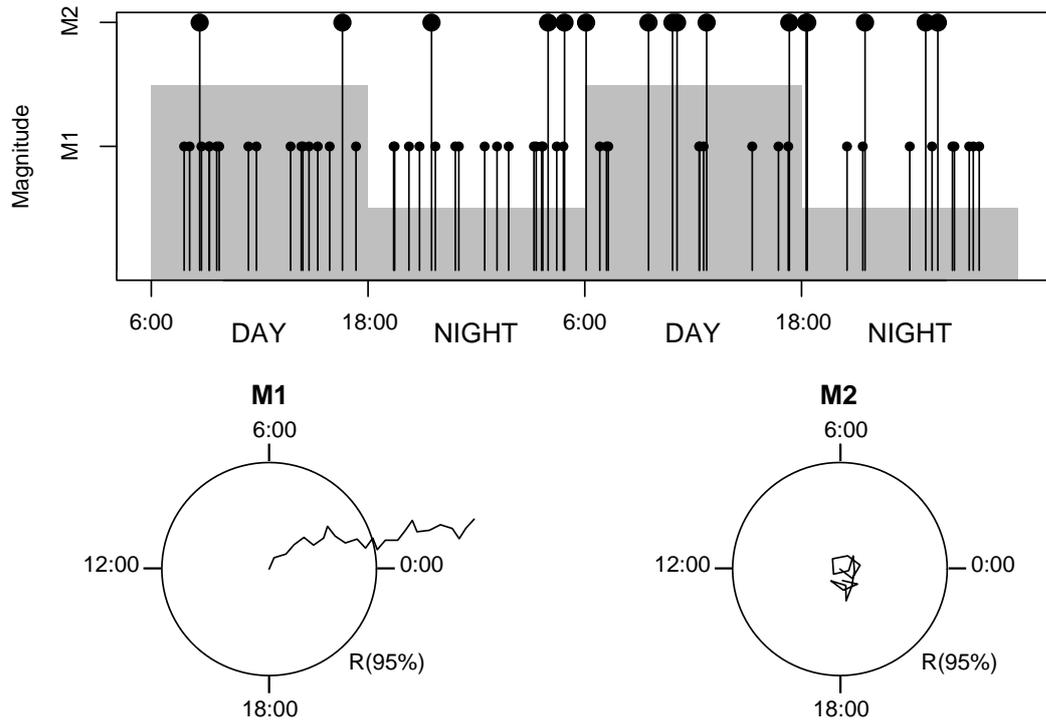
**Fig. 10** Distribution of the main and auxiliary discontinuities computed using the MBASS technique (Amorèse 2007) on 200 bootstrap samples of the NCSN catalog subset. The main discontinuity corresponds to  $M_c^{MBASS}$ .

## 6.6 Day-to-night noise modulation

The day-to-night noise modulation method (hereafter referred to as day/night method), introduced by *Rydelek and Sacks (1989)*, estimates  $M_c$  based on detectability changes expected between day and night. This is the only catalog-based method which does not assume self-similarity of the earthquake process (i.e. the G-R law). The test is based on two assumptions: (1) that earthquakes follow a [Poisson distribution](#) and (2) that the daytime noise level is increased relative to night-time due to anthropogenic activity and wind.

Figure 11 - similar to Figure 1 of *Rydelek and Sacks (1989)* - illustrates how the day/night method works. In this example, earthquakes follow a Poisson distribution and events below  $M_c$  are not detected during day hours (from 6:00 to 18:00) with  $M_1 < M_c < M_2$ . For each magnitude increment (here  $M_1$  and  $M_2$ ), the occurrence time of each event is used to form a phase angle on the local 24-hour clock. Each phase angle is assigned unit amplitude and then a phasor sum is performed over the set of individual unit phasors. The resulting phasor sum or 'walkout',  $R$ , is compared with that expected from a [random](#) process. The authors considered that a significant bias in day-to-night recording is obtained when  $R$  exceeds  $R_{95} = 1.73\sqrt{N}$ , the 95% confidence level with  $N$  the number of events. This is verified in Figure 11 for simulated data that follow the two assumptions of the day/night method.

In practice, the seismicity catalog must first be declustered to be consistent with assumption 1. Declustering should however always be used with caution. We tested the method on a declustered version of the NCSN catalog subset using the win-



**Fig. 11** Day-to-night noise modulation (day/night) method of [Rydelek and Sacks \(1989\)](#) applied to a synthetic earthquake sequence - Earthquakes follow a Poisson distribution and events below  $M_c$  are not detected during day hours (from 6:00 to 18:00) when the noise is higher (within grey threshold). With  $M_1 < M_c < M_2$ , the phasor plot for  $M_1$  shows a significant walkout, extending outside of the 95% confidence level for a Poisson process. For  $M_2$ , there is no bias in the walkout. This figure is similar to Figure 1 of the original 1989 paper.

dow method of [Gardner and Knopoff \(1974\)](#). This declustering method and other ones are described in detail in the [CORSSA article](#) by [van Stiphout et al. \(2010\)](#). We found that there is no clear magnitude threshold above which  $R < R_{95}$ . For instance the condition is not verified for  $M_{co} = 1.0, 1.7$  or  $2.1$  but is for other magnitude increments. Noteworthy, anthropogenic noise varies in space and a bulk analysis may suffer from this. This is not investigated in the present article. It should be emphasized that any claim of a real deviation from self-similarity at low  $m$  based on bulk data is highly questionable as the most likely reason for a deviation from the G-R law is a spatiotemporal variation in  $M_c$  ([Wiemer and Wyss 2000](#); [Mignan et al. 2011](#)).

## 6.7 Diagnostics from Synthetic Catalogs

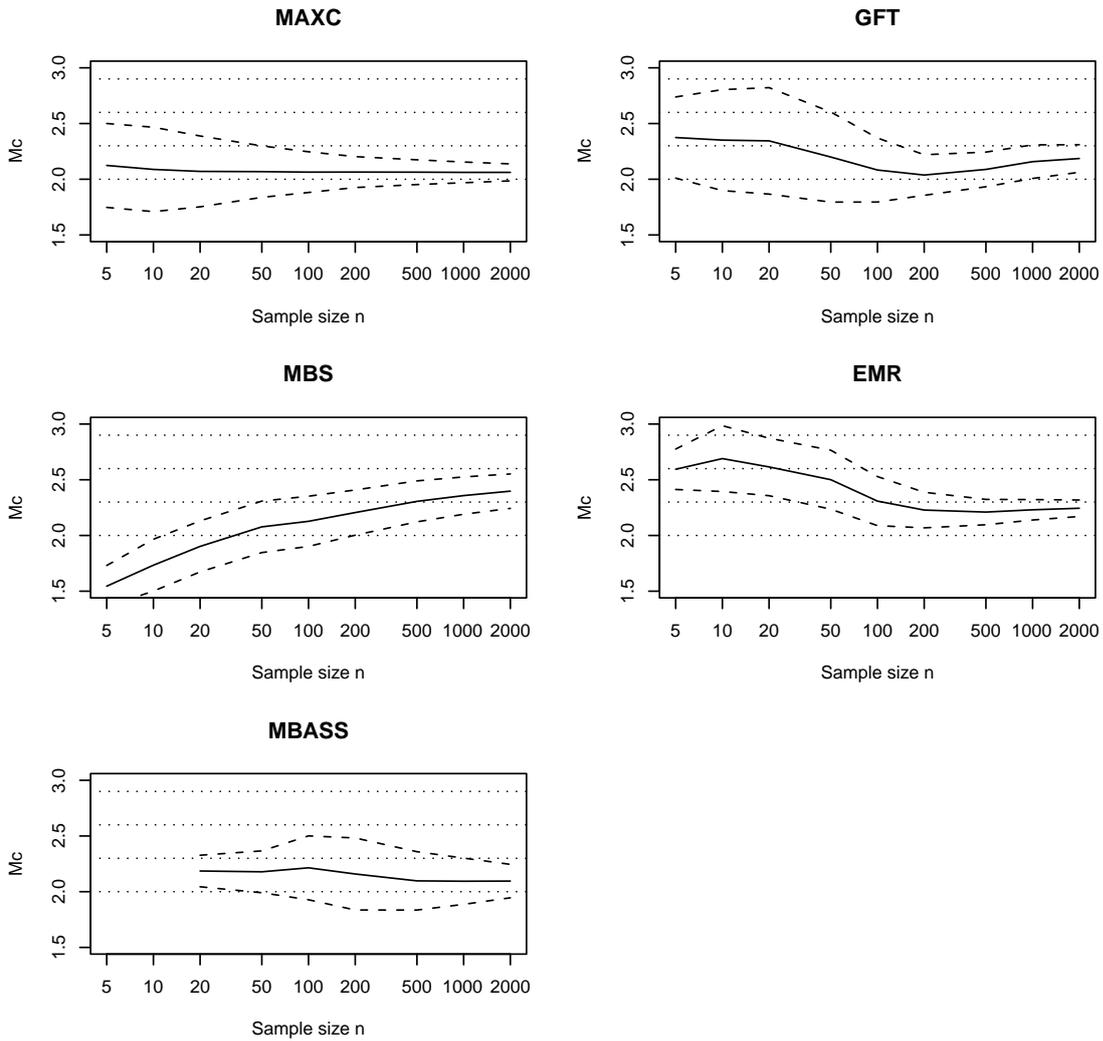
Whether or not a technique gives a correct estimate of  $M_c$  is something to assess. We showed above from real seismicity catalogs that the different FMD-based techniques provide different results with the MAXC and GFT estimates usually on the lower bound and the MBS estimates on the higher bound (Fig. 5, Table 1). The reliability of  $M_c$  techniques can also be investigated from synthetic catalogs in which the true  $M_c$  is known (Woessner and Wiemer 2005; Mignan et al. 2011).

In Woessner and Wiemer (2005), the authors defined a synthetic catalog by using the EMR model (Eq. 11) with parameters  $\theta = (M_c = 1, b = 1, \mu = 0.5, \sigma = 0.25)$ . From the synthetic dataset, random samples of ascending size  $20 \leq n \leq 1500$  were drawn and each time,  $M_c$  computed for 1000 bootstrap samples. Noteworthy, the fact that the authors used only one unique random sample per size  $n$  may explain the variability observed in their results (their Figure 5). The authors showed that: the EMR method is well capable of recovering the true  $M_c$ , the MBS approach underestimates  $M_c$  substantially for small sample sizes ( $n \leq 250$ ) and shows the strongest dependence on  $n$ , both the MAXC and GFT-95% approaches underestimate the true  $M_c$  by about 0.1 with MAXC consistently calculating the smallest value. The authors found that bootstrap uncertainty decreases with increasing sample size, except for the MBS technique, and that both MAXC and EMR approaches result in reasonable values in small data sets. For consistency between techniques, Woessner and Wiemer (2005) fixed  $n_{min} = 200$ . In regards to the reliability of the bootstrap approach, they found that estimates of uncertainties stabilize above 200 samples.

In Mignan et al. (2011), the authors defined a synthetic catalog by using the original model of Ogata and Katsura (1993) with parameters  $\theta = (b = 1, \mu = 2, \sigma = 0.3)$ . They then randomly selected 1000 samples ( $i_n$ ) of size  $n$ , repeating the operation for  $n = (4, 5, 6, 7, 8, 9, 10, 20, 50, 100, 500, 1000)$ . For each sample  $i_n$ , the authors estimated the mean completeness magnitude  $M_c^{i,n}$  and its standard deviation  $\sigma_0^{i,n}$  by using the MAXC technique on bootstrap samples. They also defined the metric  $\epsilon^{i,n}$

$$\epsilon^{i,n} = \frac{|M_c^{i,n} - M_c|}{\sigma_0^{i,n}} \quad (12)$$

which is the absolute error of the  $M_c$  estimate based on the  $i$ th sample of size  $n$ , normalized by the corresponding estimated standard deviation. Bootstrap uncertainty on  $M_c$  estimates is tacitly assumed to follow a normal distribution with mean  $M_c$  and standard deviation  $\sigma_0$  (Woessner and Wiemer 2005). If this is the case, the 0.68, 0.95 and 0.99 percentiles of the distribution  $P_n(\epsilon)$  should be equal to 1, 2 and 3 respectively. Mignan et al. (2011) observed that these percentiles are lower than the values expected by a normal distribution indicating that bootstrapping provides



**Fig. 12**  $M_c$  as function of the sample size  $n$  for different FMD-based techniques. Standard deviation  $\pm\sigma_0$  is represented by dashed curves. Values are averaged over 100 random samples per sample size. Horizontal lines represent  $M_c(50\%)$ ,  $M_c(68\%)$ ,  $M_c(95\%)$  and  $M_c(99\%)$ , following the OK1993 model.

conservative uncertainty estimates for  $M_c^{MAXC}$  for  $n_{min} = 4$ . For other techniques (GFT, EMR and MBASS), they found  $n_{min} \sim 100 - 200$ . The authors obtained similar results when using 200 or 1000 bootstrap samples. Here  $n_{min}$  means that the uncertainty estimate is reliable for samples of this size or larger. The uncertainty may however be very large for small samples and a higher  $n_{min}$  considered to get a more robust  $M_c$  estimate.

Similarly to [Woessner and Wiemer \(2005\)](#) but with a synthetic catalog based on

the OK1993 model with parameters  $\theta = (b = 1, \mu = 2, \sigma = 0.3)$ , we investigated the variations of  $M_c$  and its standard deviation  $\sigma_0$  with the sample size  $n$ , for different FMD-based techniques (MAXC, GFT, MBS, EMR and MBASS). We assumed that bootstrapping provides reliable uncertainty estimates for any technique. We drew 100 random samples of size  $n$ , with  $n = (5, 10, 20, 50, 100, 200, 500, 1000, 2000)$ . Figure 12 shows the  $M_c$  and its standard deviation  $\sigma_0$ , averaged over the 100 random samples, for the different sample sizes  $n$ . For all techniques,  $\sigma_0$  decreases with increasing  $n$ . In agreement with [Woessner and Wiemer \(2005\)](#), the MBS approach underestimates  $M_c$  substantially for small sample sizes ( $n \leq 50$ ) and shows the strongest dependence on  $n$ . The MBASS technique cannot compute  $M_c$  for  $n < 20$ . The  $n_{min} = 200$  threshold proposed by [Woessner and Wiemer \(2005\)](#) appears as a reasonable choice for most techniques, although the averaged  $M_c^{MAXC}$  remains stable at very low  $n$ . As in real catalogs, MAXC provides the lower  $M_c$  estimates ( $M_c^{MAXC} \sim \mu$ ) and MBS the higher ones ( $M_c^{MBS} > \mu + \sigma$ ). Other techniques provide intermediary results with all results significantly lower than the 95% confidence level ( $\mu + 2\sigma$ ) expected by the OK1993 model.

We emphasize that these results assume the validity of the OK1993 model to describe the FMD. Other models may favor other techniques. For instance, [Woessner and Wiemer \(2005\)](#) found that the EMR method is well capable of recovering the true  $M_c$ , which is to be expected for an FMD described by the EMR model. Similarly, MAXC by definition will always underestimate  $M_c$  in the OK1993 model, which describes well bulk data sets. Following [Mignan et al. \(2011\)](#), the OK1993 model may not be appropriate when  $M_c$  heterogeneities are minimized (see Section 7). Moreover, even if the uncertainty estimate obtained for a given technique is low, it does not mean that the  $M_c$  estimate is reliable, only that the output estimate is robust in regards of the data set considered. This is in agreement with the fact that the ranges of possible  $M_c$  obtained from different techniques may not overlap, as they do not necessarily correspond to the same definition of  $M_c$ .

### 6.8 Exercise: Strategy to estimate $M_c$ for a Single Catalog Data Sample

Given the variety of catalog-based approaches, one strategy to estimate  $M_c$  for a given data sample is to start with the simplest method, namely the maximum curvature approach (MAXC) for the first estimate. To understand possible differences, do the same with some other techniques. Depending on the purpose of a given study, one then can estimate the uncertainty using a Monte-Carlo bootstrap technique and then further explore with additional methods.

As an exercise with the MATLAB software package, download the archive *FMDcalc.tgz* or *FMDcalc.zip* for this article and reproduce the values and figures. The

usage is provided in the Scripts *Samplescript.m* and *SampleScriptBootstrap.m*. Then try other data sets.

## 7 Catalog-based Techniques to Map $M_c(x, y, z, t)$

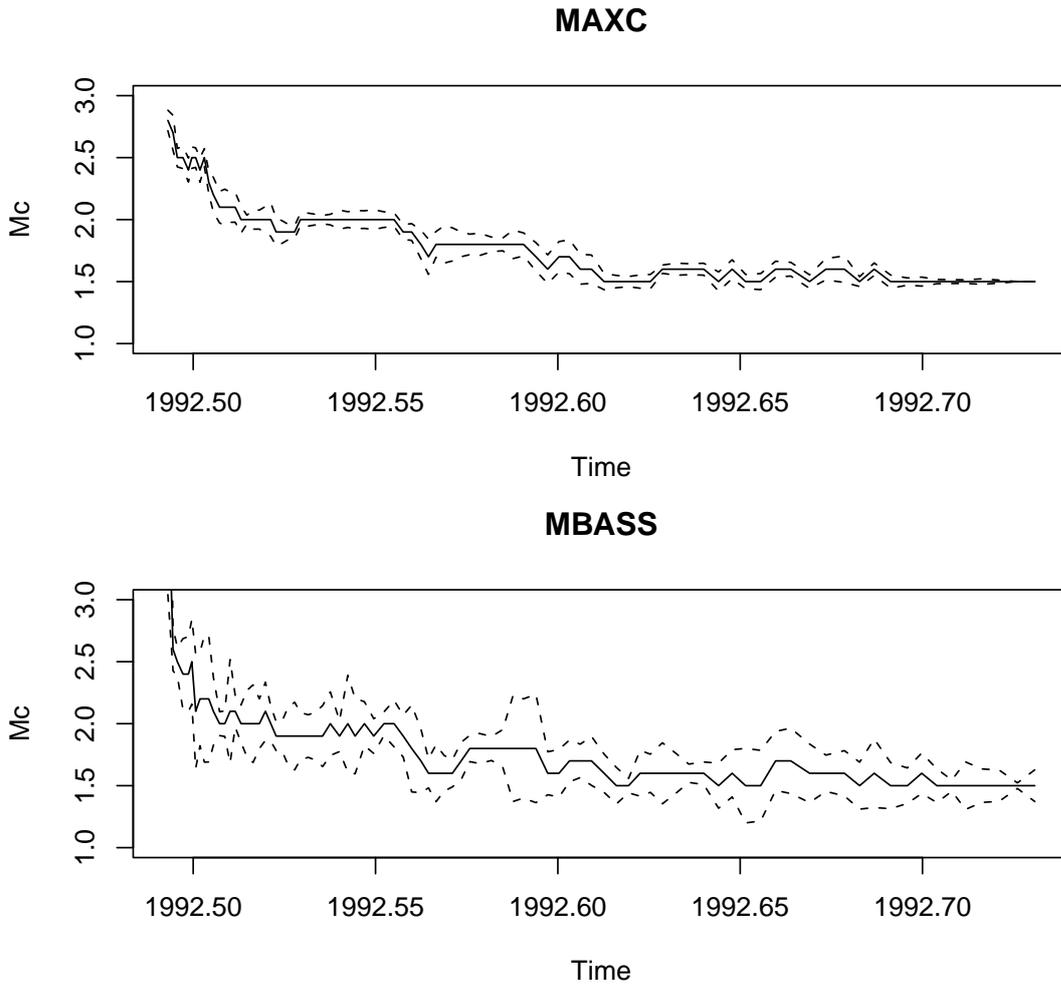
An important challenge is the understanding and assessment of possible spatiotemporal changes in  $M_c$ . Unless the space-time history of  $M_c(x, y, z, t)$  is taken into consideration, a study would have to conservatively assume the highest  $M_c$  observed (*Wiemer and Wyss 2000*). This maximum is difficult accessible from a simple bulk analysis, which is by definition based on an average data sample. Temporal changes are discussed in Section 7.1 and spatial changes in Sections 7.2, 7.3 and 7.4, respectively.

### 7.1 Temporal Variations in $M_c$

Temporal changes in  $M_c$  origin from the evolution of the seismic network (pseudo-permanent changes) or are due to large earthquakes sequences, aftershock or swarm activity (transient changes). Aftershock sequences often introduce the largest fluctuations in  $M_c$ .

Changes in seismic networks often correspond to the addition of new stations or to the update of existing hardware and software. Transitional phases in a network are usually well identified and show clear shifts in  $M_c$  estimates (e.g., *Hutton et al. 2010*). With the  $M_c(t)$  information, one can select a data set over a time period in which  $M_c(t)$  remains almost constant. If the analysis requires a longer time period and a fixed estimate for completeness, then  $M_c = \max M_c(t)$ , which can significantly alter the number of events available.

In case of an aftershock sequence, some events are too small to be detected within the coda of larger events (i.e. increased noise), and thus  $M_c$  increases. Figure 13 shows the evolution of  $M_c$  through time for the aftershock sequence of the 1992 Landers earthquake. We used the data set of *Woessner and Wiemer (2005)* and the same moving window approach. To generate the time series, we used a window size of 1000 events to compute  $M_c$  while moving the window by 250 events.  $M_c$  is computed using the MAXC and MBASS techniques. Similar results are obtained for both techniques, suggesting homogeneity in  $M_c$ , which is to be expected in data sets representing a small spatiotemporal volume (*Mignan et al. 2011*). In this example,  $M_c(t)$  increases up to two units in the first days after the main shock. This phenomenon can significantly impact aftershock statistics studies if not evaluated correctly. *Iwata (2008)* found a similar trend when investigating parameter  $\mu(t)$  of the OK1993 model (see Section 6.4).



**Fig. 13**  $M_c$  as a function of time for the 1992 Landers aftershock sequence. Moving window approach with a window of 1000 events, moved by 250 events. Data set from [Woessner and Wiemer \(2005\)](#). Dashed curves represent  $\pm\sigma_0$  obtained from 200 bootstrap samples.

## 7.2 Standard Catalog-based $M_c$ Spatial Mapping

The commonly used catalog-based mapping approach (hereafter referred to as standard method) consists in estimating  $M_c(x, y)$  from the FMD generated from events located in a cylindrical volume of fixed radius  $R$  or of fixed number of events  $N$ , centered on each node of a spatial grid ([Wyss et al. 1999](#); [Wiemer and Wyss 2000](#)). Any FMD-based  $M_c$  technique can be implemented (e.g., MAXC, GFT, EMR, MBASS). For the constant- $R$  approach, the minimum number of events  $N_{min}$  is fixed to avoid unstable  $M_c$  results.  $R$  must be large enough to have a sufficient number of nodes

with  $M_c$  estimates but small enough to avoid over-smoothing. For the constant- $N$  approach, the maximum radius  $R_{max}$  is fixed to avoid over-smoothing.  $N$  must be large enough to obtain reliable  $M_c$  estimates but small enough to avoid gaps. In both cases, finding the best compromise may be challenging with selected values based on an educated guess. Noteworthy, while variations with depth  $z$  are averaged over the cylindrical volumes in mapping on an horizontal plane, various  $M_c(x, y, z)$  cross-sections are possible (e.g., [Wiemer and Wyss 2000](#)). This approach has been used in numerous studies. Recent applications include [Woessner and Wiemer \(2005\)](#) for a global catalog, [Hutton et al. \(2010\)](#) for Southern California, [Schorlemmer et al. \(2010a\)](#) for Italy, [Nanjo et al. \(2010a\)](#) for Japan and [Mignan et al. \(2011\)](#) for Taiwan.

This approach has however two basic limitations: (1) the choice of a fixed cylinder radius is arbitrary compared to the true  $M_c$  resolution ([Mignan et al. 2011](#)) while the use of a constant number of events results in an unrealistic and variable relationship to actual seismicity ([Rydelek and Sacks 2003](#)) and (2) gaps remain in regions of low seismicity (i.e.  $N < N_{min}$ ) where  $M_c$  is not computed. To illustrate these limitations, Figure 14 shows the seismicity map for the NCSN subset and 3  $M_c$  maps created by using the constant radius approach with  $R = 5, 20$  and  $50$  km and  $N_{min} = 50$ . Note that over-smoothing may be identified by the presence of artificial circular constant  $M_c$  patches of radius  $R$ . A detailed description of the standard  $M_c$  mapping approach and of its limits is given by [Mignan et al. \(2011\)](#).

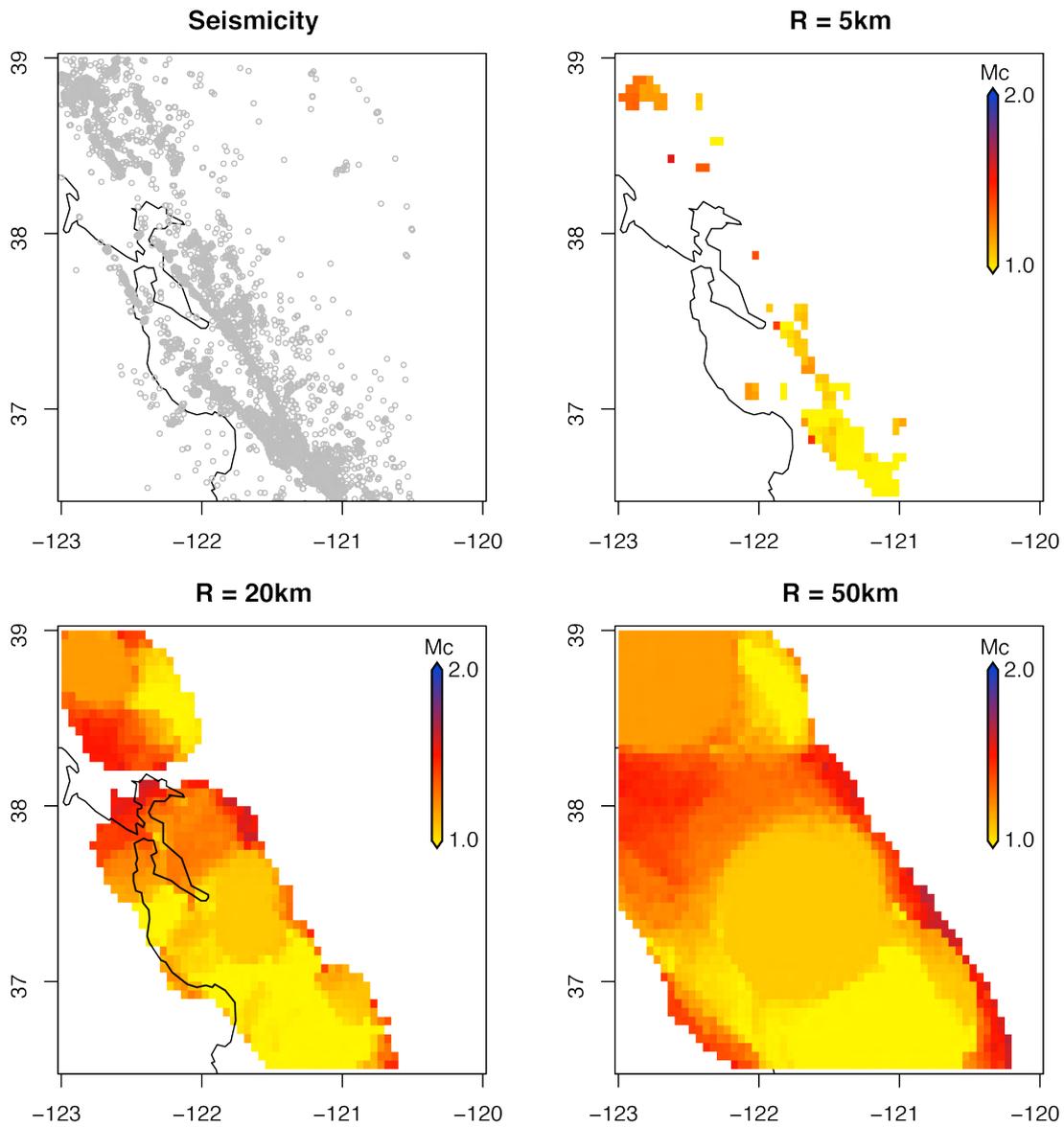
### 7.3 Bayesian Magnitude of Completeness (BMC) Spatial Mapping

[Mignan et al. \(2011\)](#) proposed the Bayesian Magnitude of Completeness (BMC) method to map  $M_c$  by avoiding the shortcomings of the standard mapping approach of [Wiemer and Wyss \(2000\)](#). BMC requires knowledge of the seismic stations locations in addition to access to the earthquake catalog. It is a two-step procedure consisting in (1) a spatial resolution optimization to minimize spatial heterogeneities and uncertainties in  $M_c$  estimates and (2) a Bayesian approach to merge prior information on the relationship between  $M_c$  and the density of seismic stations with locally observed  $M_c$ , weighted by their respective uncertainties.

The BMC spatial optimization procedure consists in estimating  $M_c$  from the FMD that corresponds to events located in a cylindrical volume of radius  $R$

$$R(d) = \frac{1}{2} \left[ \left( \frac{c_1 d^{c_2} + \sigma}{c_1} \right)^{\frac{1}{c_2}} - \left( \frac{c_1 d^{c_2} - \sigma}{c_1} \right)^{\frac{1}{c_2}} \right] \quad (13)$$

where  $d$  is the distance to the  $k$ th nearest seismic station (with  $k$  usually being between 3 and 5) and  $\sigma$  is the  $M_c$  interval below which variations cannot be resolved.



**Fig. 14** Standard  $M_c$  mapping (constant radius approach) for the NCSN subset.  $M_c$  is evaluated using the MAXC technique (see Section 6.1). Many locations have no  $M_c$  estimate due to low seismicity ( $N_{min} = 50$ ). Increasing the radius from  $R = 5$  km to 50km reduces the number of gaps but also increases spatial smoothing. Over-smoothing may be identified by the presence of artificial circular constant  $M_c$  patches of radius  $R$ .

Eq. 13 is derived from the prior model

$$M_c^{pred} = c_1 d^{c_2} + c_3 \quad (14)$$

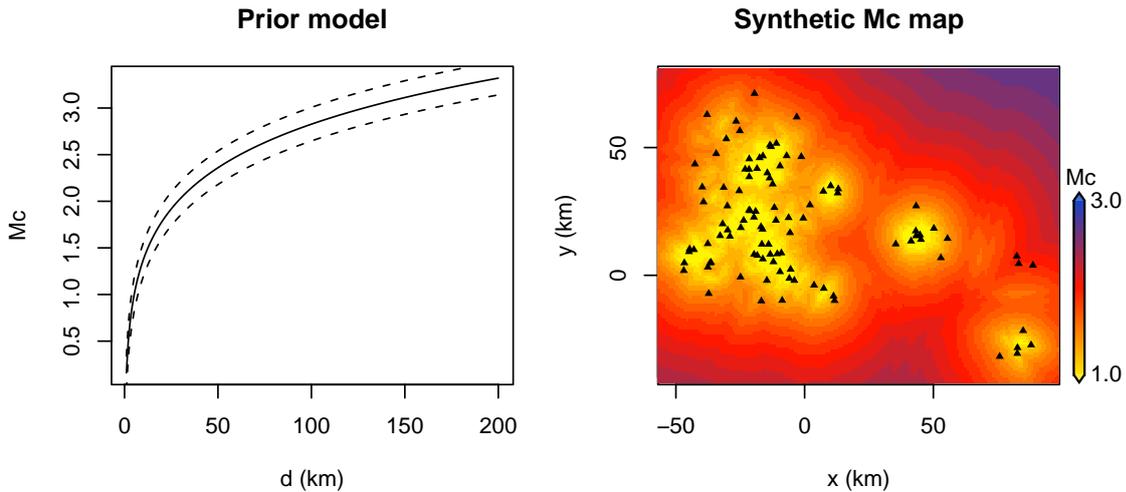
where parameters  $c_1$ ,  $c_2$  and  $c_3$  are determined empirically.  $\sigma$  is the standard deviation of the residual between prior and observed  $M_c$  values. For Taiwan, [Mignan et al. \(2011\)](#) found  $c_1 = 5.96$ ,  $c_2 = 0.0803$ ,  $c_3 = -5.80$  and  $\sigma = 0.18$  for  $k = 4$  with  $d$  and  $R$  in km. The parameter  $k$  is the minimum number of stations to be triggered for initiating the location procedure in the network. Figure 15 shows a plot of the prior model (Eq. 14) as well as an  $M_c^{pred}$  map for a synthetic seismic network with seismic stations locations described by a random walk. An increase of  $M_c$  with distance to the  $k$ th seismic station is also observed in California, Alaska and Japan ([Wiemer and Wyss 2000](#); [Nanjo et al. 2010a](#)). Eq. 14 additionally indicates that  $M_c$  evolves faster in the dense parts of the seismic network than in the outer regions. This is the basis of the variable- $R$  approach (Eq. 13): At a node located  $d = 20$  km from the 4th nearest seismic station,  $R = 6$  km, for  $d = 100$  km,  $R = 26$  km and for  $d = 200$  km,  $R = 50$  km. This approach avoids having to make an educated guess on the value of  $R$  and avoids over-smoothing by making  $R$  variable in space. While any FMD-based method can be implemented in BMC to compute  $M_c$ , [Mignan et al. \(2011\)](#) suggested using MAXC, which has been shown to not underestimate  $M_c$  in data sets where  $M_c$  heterogeneities are minimized.

The second step of the BMC approach consists in merging prior information (Eq. 14) with observations, based on Bayes Theorem, with

$$M_c^{post} = \frac{M_c^{pred} \sigma_0^2 + M_c^{obs} \sigma^2}{\sigma^2 + \sigma_0^2} \quad (15)$$

which is a weighted average of the predicted and observed completeness magnitude, where the weights are proportional to their respective uncertainties ([Mignan et al. 2011](#)). In contrast with the standard mapping approach (Section 7.2), the resulting  $M_c^{post}$  map (or BMC map) shows no gap as the prior model provides a continuous estimate of  $M_c$  in space (Fig. 15).

One limitation of the BMC method is that the prior model (Eq. 14) is not well constrained in regions of low seismicity or where seismicity is concentrated in dense areas of the seismic network. If the prior model cannot be defined from regional data, one may try to use the prior model defined for another region by assuming that the relationship  $M_c = f(d)$  is region-independent. This remains to be thoroughly investigated. Moreover, the current version of the BMC method does not include temporal changes in  $M_c$  and thus requires the use of a dataset defined over a time period of stable seismic network configuration (see [Mignan et al. \(2011\)](#) for more details).



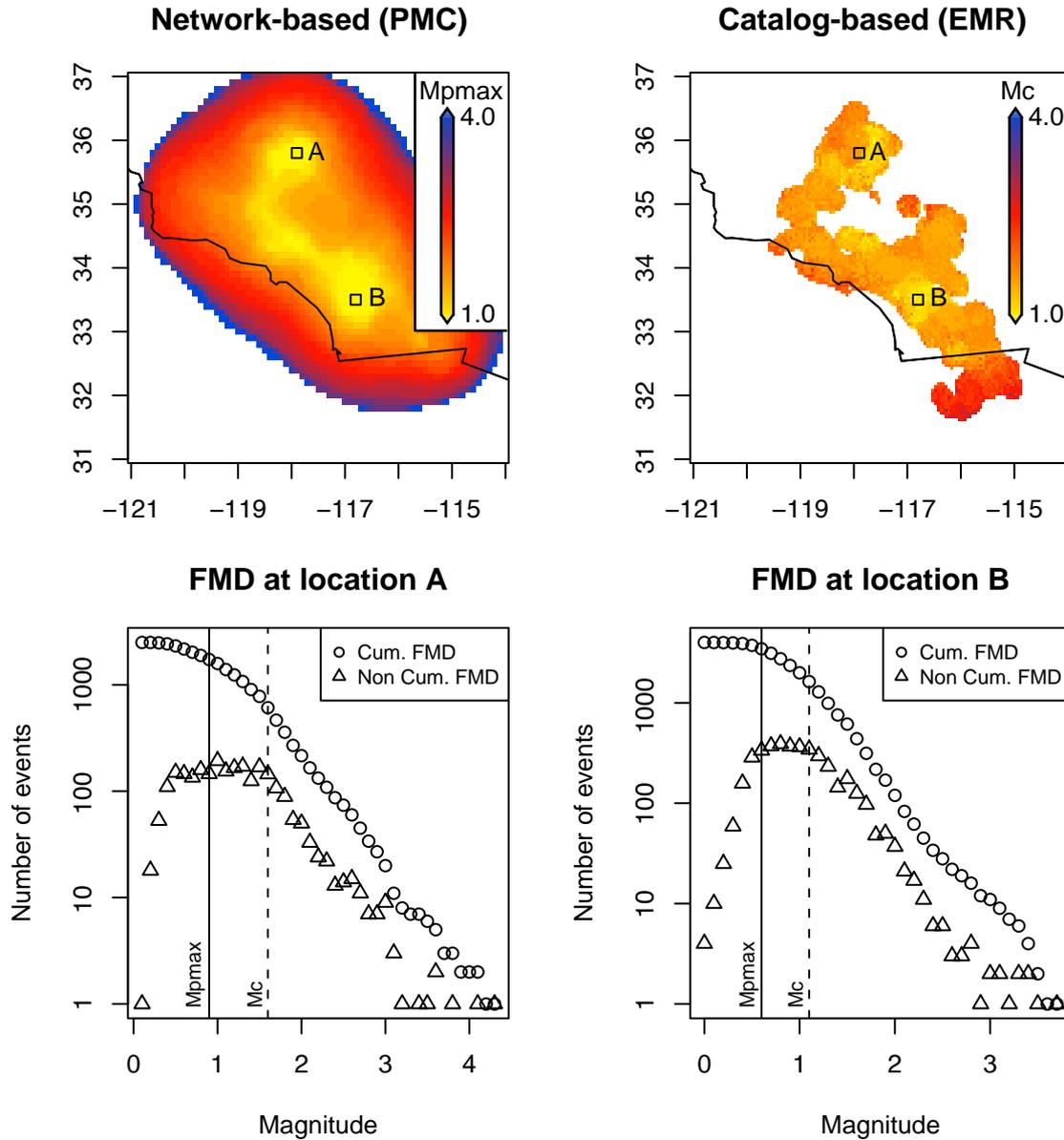
**Fig. 15** BMC mapping approach of [Mignan et al. \(2011\)](#). The prior model (Eq. 14) indicates that  $M_c$  increases with the distance to the  $k$ th nearest seismic station and that it evolves faster in the dense parts of the seismic network than in the outer regions. Dashed lines represent  $\pm\sigma = 0.18$ . An  $M_c^{pred}$  map can be generated from Eq. 14 using the seismic stations locations represented by the triangles. The locations of the stations of the synthetic seismic network are described by a random walk.

#### 7.4 Comparison of Network-based and Catalog-based $M_c$ Spatial Mapping

Do catalog-based techniques and network-based techniques (see Section 5.2) provide similar results? It should first be noted that a comparison is not straightforward as the definitions of the completeness are different. Comparison of traditional FMD-based  $M_c$  mapping and PMC ( $M_P$ ) mapping is given by [Schorlemmer and Woessner \(2008\)](#) for Southern California and by [Nanjo et al. \(2010b\)](#) for Switzerland. For Southern California, the authors compared  $M_P$  obtained for 1 July 2007 and  $M_c^{EMR}$  obtained for the period 1 January 2001 - 1 July 2007 and found that  $M_P$  estimates were lower than  $M_c^{EMR}$  in 77% of nodes, and in some areas were one unit lower. Underestimating the completeness is problematic as it leads to biased seismicity analyses (see Section 5.1). The comparison made by [Schorlemmer and Woessner \(2008\)](#) was however a rough one as both maps were not based on the same space-time volume:  $M_P$  is an estimate of completeness for the network at one specific point in space and time, whereas  $M_c$  is a completeness estimate of an earthquake sample for a specific space-time volume projected onto one grid node. For Switzerland, [Nanjo et al. \(2010b\)](#) computed  $M_{Pmax}(x, y)$ , the maximum  $M_P$  value observed in the space-time volume used for estimating  $M_c(x, y)$ . For an overall completeness, it provides the most conservative estimate of  $M_P$  as this strategy ensures to take the value from the time period and cylindrical volume in which the detection capabilities are worst. The authors found that  $M_P$  values are generally higher than

$M_c^{EMR}$  in the case of Switzerland.

We here repeat the procedure for Southern California with Figure 16 illustrating the difference between  $M_{Pmax}(x, y)$  and  $M_c^{EMR}(x, y)$ . The  $M_c^{EMR}$  map is the one obtained by *Schorlemmer and Woessner* (2008) for the period 1 January 2001 - 1 July 2007 and for a constant cylindrical radius of 20 km. The  $M_{Pmax}$  map is obtained from the combination of 79  $M_P$  maps, one per month from 1 January 2001 to 1 July 2007, downloadable at <http://completeness.usc.edu/>. At each node,  $M_{Pmax}(x, y)$  corresponds to the maximum  $M_P$  value observed between 1 January 2001 and 1 July 2007 in the cylinder of radius  $R = 20\text{km}$  centered on  $(x, y)$ . We find that  $M_{Pmax}$  is lower than  $M_c^{EMR}$  in central parts of Southern California and higher in outer regions. Noteworthy, only  $M_P$  maps at the depth of 7.5 km are available for download for Southern California. Taking into account possible variations with depth could only increase  $M_{Pmax}$ . Comparison is also limited in space as the standard  $M_c$  map shows many gaps (see Section 7.2). At the present time, there is no comparison available between the PMC and BMC methods (see Section 7.3). Figure 16 also shows some local FMDs at two locations, generated from events in a cylindrical volume of radius  $R = 20\text{km}$ . We used the same earthquake data set, the legacy Southern California seismicity catalog for 1 January 2001 - 1 July 2007 available at [http://www.data.scec.org/ftp/catalogs/SCSN\\_pre2008/](http://www.data.scec.org/ftp/catalogs/SCSN_pre2008/). While the EMR estimates are consistent with the G-R law, the PMC estimates are not. The  $M_P$  metrics quantifies some degree of detectability but does not seem to relate to catalog-based completeness. The assumption  $M_{Pmax} = M_c$  should thus be used with caution in seismicity analyses.



**Fig. 16** Comparison of network-based and catalog-based completeness magnitude in Southern California for the period 1 January 2001 - 1 July 2007.  $M_{Pmax}(x,y)$  is the maximum  $M_P$  value observed in the cylinder of radius  $R = 20\text{km}$  centered on  $(x,y)$  over 79 monthly  $M_P$  maps (<http://completeness.usc.edu/>). The  $M_c^{EMR}$  map is based on the standard catalog-based mapping approach with  $R = 20\text{km}$  and  $N_{min} = 100$  (Schorlemmer and Woessner 2008). The two local FMDs are generated from events in a cylindrical volume of radius  $R = 20\text{km}$  and centered on the black squares shown in both maps (location A:  $117.9^\circ\text{W}$ ;  $35.8^\circ\text{N}$ , location B:  $116.8^\circ\text{W}$ ;  $33.5^\circ\text{N}$ ).  $M_{Pmax}$  and  $M_c^{EMR}$  are represented by a solid line and dashed line, respectively. The Legacy Southern California seismicity catalog for 1 January 2001 - 1 July 2007 is available at [http://www.data.scec.org/ftp/catalogs/SCSN\\_pre2008/](http://www.data.scec.org/ftp/catalogs/SCSN_pre2008/).

## 8 Software packages and Catalog Data

### 8.1 Software packages

Most techniques to compute  $M_c$  (Section 6) as well as the standard  $M_c(x, y, z, t)$  mapping (Sections 7.1 and 7.2) are available in the MATLAB packages

- ZMAP package (*Wiemer 2001*), available at <http://www.earthquake.ethz.ch/software/zmap>
- The accompanying code archive *FMDcalc.zip* is a collection of functions to calculate  $M_c$  for the catalog based techniques (MAXC, GFT, MBS and EMR) together with uncertainty calculation using bootstrapping. This is available from the CORSSA website including example scripts. The package contains plotting of the FMD. The usage is described with in the scripts *Samplescript.m* and *SampleScriptBootstrap.m*.

The following techniques are also available in the R language in the Appendix of this tutorial (Section 9): MAXC, GFT, MBS and EMR. The MBASS R algorithm is published in *Amorèse (2007)*.

There is yet no package available online for the BMC method (Section 7.3). Two R programs (for optimized  $M_c$  mapping and for the Bayesian approach) are available directly from the lead author of this tutorial (inquiries by e-mail).

For the Probability-based Magnitude of Completeness (PMC) method (Sections 5.2 and 7.4), the website <http://completeness.usc.edu/> provides a time history of results for the Southern California Seismic Network, the Italian National Seismic Network and the Swiss Digital Seismic Network, downloadable in raster files. The package QuakePy, available at <https://quake.ethz.ch/quakepy/>, contains the PMC module (programs in Python). MATLAB programs are also available from the authors of the published references.

### 8.2 Catalog Data

The catalog data used for the examples in this tutorial are available from the CORSSA website.

**Acknowledgements** We are grateful to CORSSA Editor Jeanne Hardebeck and to an anonymous referee for their valuable comments.

## 9 Appendix

Disclaimer: the following functions are R-Language versions of the ZMAP functions, written in Matlab (see [Wiemer \(2001\)](#) and <http://www.earthquake.ethz.ch/software/zmap>). These have been tested on various data sets and shown to work properly. We however recommend to the user to contact us in case any glitch or any more serious bug would be encountered. Some basic R programs are also provided for plotting and bootstrap analysis - See also the CORSSA article by [Naylor et al. \(2010\)](#) for more tools in R. The R algorithm for the MBASS technique is given by [Amorèse \(2007\)](#).

```
# LIBRARIES
library(nlstools)      #nlsfit() for emr() function

# FUNCTIONS
fmd <- function(mag,mbin){
  mi <- seq(min(round(mag/mbin)*mbin), max(round(mag/mbin)*mbin), mbin)
  nbm <- length(mi)
  cumnbmag <- numeric(nbm)
  nbmag <- numeric(nbm)
  for(i in 1:nbm) cumnbmag[i] <- length(which(mag > mi[i]-mbin/2))
  cumnbmagtmp <- c(cumnbmag,0)
  nbmag <- abs(diff(cumnbmagtmp))
  res <- list(m=mi, cum=cumnbmag, noncum=nbmag)
  return(res)
}

#Maximum Curvature (MAXC) [e.g., Wiemer & Wyss, 2000]
maxc <- function(mag,mbin){
  FMD <- fmd(mag,mbin)
  Mc <- FMD$m[which(FMD$noncum == max(FMD$noncum))][1]
  return(list(Mc=Mc))
}

#Goodness-of-fit test (GFT) [Wiemer & Wyss, 2000]
gft <- function(mag,mbin){
  FMD <- fmd(mag,mbin)
  McBound <- maxc(mag,mbin)$Mc
  Mco <- McBound-0.4+(seq(15)-1)/10
  R <- numeric(15)
```

```

for(i in 1:15){
  indmag <- which(mag > Mco[i]-mbin/2)
  b <- log10(exp(1))/(mean(mag[indmag])-(Mco[i]-mbin/2))
  a <- log10(length(indmag))+b*Mco[i]
  FMDcum_model <- 10^(a-b*FMD$m)
  indmi <- which(FMD$m >= Mco[i])
  R[i] <- sum(abs(FMD$cum[indmi]-FMDcum_model[indmi]))/sum(FMD$cum[indmi])*100
  #in Wiemer&Wyss [2000]: 100-R
}
indGFT <- which(R <= 5)          #95% confidence
if(length(indGFT) != 0){
  Mc <- Mco[indGFT[1]]
  best <- "95%"
} else{
  indGFT <- which(R <= 10)      #90% confidence
  if(length(indGFT) != 0){
    Mc <- Mco[indGFT[1]]
    best <- "90%"
  } else{
    Mc <- McBound
    best <- "MAXC"
  }
}
return(list(Mc=Mc, best=best, Mco=Mco, R=R))
}

#Mc by b-val Stability (MBS) [Cao & Gao, 2002]
#Modification with Shi & Bolt [1982] uncertainty [Woesner & Wiemer, 2005]
mbs <- function(mag,mbin){
  McBound <- maxc(mag,mbin)$Mc
  Mco <- McBound-0.7+(seq(20)-1)/10
  bi <- numeric(20); unc <- numeric(20)
  for(i in 1:20){
    indmag <- which(mag > Mco[i]-mbin/2)
    nbev <- length(indmag)
    bi[i] <- log10(exp(1))/(mean(mag[indmag])-(Mco[i]-mbin/2))
    unc[i] <- 2.3*bi[i]^2*sqrt(sum((mag[indmag]-
      mean(mag[indmag]))^2)/(nbev*(nbev-1)))
  }
  bave <- numeric(15)
}

```

```

for(i in 1:15) bave[i] <- mean(bi[i:i+5])

dbi_old <- abs(diff(bi))
indMBS_old <- which(dbi_old <= 0.03)

dbi <- abs(bave[1:15]-bi[1:15])
indMBS <- which(dbi <= unc[1:15])
Mc <- Mco[indMBS[1]]
return(list(Mc=Mc, Mco=Mco, bi=bi, unc=unc, bave=bave))
}

#Entire Magnitude Range method (EMR) [Woesner & Wiemer, 2005]
emr <- function(mag,mbin){
  FMD <- fmd(mag,mbin)
  nbm <- length(FMD$m)

  McMAXC <- maxc(mag,mbin)$Mc
  mu <- abs(McMAXC/2); sig <- abs(McMAXC/4)
  if(mu > 1)mu <- abs(McMAXC/10); sig <- abs(McMAXC/20)
  McBound <- McMAXC
  Mco <- McBound-0.3+(seq(9)-1)/10
  params <- numeric(9*4); dim(params) <- c(9,4)    #a, b, mu, sigma
  prob <- numeric(9)
  savedmodel <- numeric(9*nbm); dim(savedmodel) <- c(9,nbm)
  for(i in 1:9){
    indmag <- which(mag > Mco[i]-mbin/2)
    nbev <- length(indmag)
    b <- log10(exp(1))/(mean(mag[indmag])-(Mco[i]-mbin/2))
    a <- log10(length(indmag))+b*Mco[i]
    cumN <- 10^(a-b*FMD$m)
    params[i,1] <- a; params[i,2] <- b

    cumNtmp <- 10^(a-b*(max(FMD$m)+mbin))
    cumNtmp <- c(cumN, cumNtmp)
    N <- abs(diff(cumNtmp))
    data <- data.frame(N=N, m=FMD$m, Nd=FMD$noncum)
    indLow <- which(FMD$m < Mco[i]); indHigh <- which(FMD$m >= Mco[i])
    dataTest <- data.frame(N=data$N[indLow], m=data$m[indLow], Nd=data$Nd[indLow])
    dataTmp <- data.frame(N=data$N[indHigh], m=data$m[indHigh], Nd=data$Nd[indHigh])
  }
}

```

```

checkNo0 <- which(dataTest$Nd != 0)
dataTest <- data.frame(N=dataTest$N[checkNo0], m=dataTest$m[checkNo0],
  Nd=dataTest$Nd[checkNo0])
#Nmax <- max(dataTmp$Nd)
Nmax <- max(dataTest$Nd)
#Nmax <- dataTest$Nd[length(dataTest$Nd)]
Mmintmp <- min(dataTest$m)
dataTest$Nd <- dataTest$Nd/Nmax
dataTest$m <- dataTest$m-Mmintmp
data4fit <- data.frame(N=dataTest$Nd, m=dataTest$m)

#non-linear least squares fit
nlsfit <- nls(N~pnorm(m, mean=mean, sd=sd), data=data4fit,
  start=list(mean=mu, sd=sig), control=list(maxiter=100, warnOnly = TRUE))
params[i,3] <- coef(nlsfit)["mean"]; params[i,4] <- coef(nlsfit)["sd"]
dataTest$N <- pnorm(dataTest$m, mean=coef(nlsfit)["mean"],
  sd=coef(nlsfit)["sd"])*Nmax
dataTest$m <- dataTest$m+Mmintmp
dataTest$Nd <- dataTest$Nd*Nmax

dataPred <- data.frame(N=c(dataTest$N, dataTmp$N), m=c(dataTest$m, dataTmp$m),
  Nd=c(dataTest$Nd, dataTmp$Nd))
dataPred$N <- round(dataPred$N)
savedmodel[i,c(checkNo0,indHigh)] <- dataPred$N

#CHECK EMR METHOD#
#pdf(paste(wd,"plot_NonCumModel_",Mco[i],".pdf", sep=""))
#plot(dataPred$m, dataPred$Nd, pch=18, xlab="Magnitude",
  ylab="Cumulative Number", log="y")
#points(dataPred$m, dataPred$N, pch=1)
#abline(v=Mco[i], lty="dashed")
#legend("topright", c("Data","EMR model"), cex=0.8, lty=c(0,0), pch=c(18,1))
#dev.off()
#write.table(dataPred, file=paste(wd, "file_NonCumModel_",Mco[i],
  ".txt", sep=""))

#Logarithm to the basis of 10 of Poisson probability density
probtmp <- numeric(nbm)
CheckNo0 <- which(dataPred$N != 0)
Pmodel <- dataPred$N[CheckNo0]; Pdata <- dataPred$Nd[CheckNo0]

```

```

    probtmp[CheckNo0] <- 1/log(10)*(-Pmodel+Pdata*log(Pmodel)-lgamma(Pdata+1))
    prob[i] <- -sum(probtmp)
  }
  indbestfit <- which(prob == min(prob, na.rm=TRUE))
  res <- list(Mc=Mco[indbestfit], a=params[indbestfit,1], b=params[indbestfit,2],
    mu=params[indbestfit,3], sigma=params[indbestfit,4],
    model=savedmodel[indbestfit,], Mco=Mco, prob=prob)
  return(res)
}

# INPUT PARAMETERS
wd <- "/Path/WorkingDirectory/"
cat_file <- "SeismicityCatalog.txt"
mbin <- 0.1          #Magnitude bin
nbsample <- 200     #Bootstrapping

## READ CATALOG ##
#For a catalog with data listed in columns separated by space or tab
#(Longitude, Latitude, Magnitude, etc...)
#Other formats may require a different R function
cat <- read.table(paste(wd, cat_file, sep=""), header=TRUE)
mag <- cat$Magnitude

## COMPUTE Mc ##
Mc_bootstrap <- numeric(nbsample)
#select function: maxc(), gft(), mbs(), emr()
#For mbass(), see algorithm Amorese [2007]
for(i in 1:nbsample) Mc_bootstrap[i] <- maxc(sample(mag, replace=TRUE),mbin)$Mc
#when using emr(), the loop may break due to failure of nlsfit(),
#in this case use:
for(i in 1:nbsample) Mc_bootstrap[i] <-
  as.numeric(try(emr(sample(mag, replace=TRUE),mbin)$Mc))

Mc_mean <- mean(Mc_bootstrap, na.rm=TRUE)
Mc_sd <- sd(Mc_bootstrap, na.rm=TRUE)
print(paste("Mc (mean): ", Mc_mean, sep=""))

```

```
print(paste("Sigma0 (std. dev.): ", Mc_sd, sep=""))

## PLOT FMD ##
FMD <- fmd(mag,mbin)

pdf(paste(wd,"CORSSA_Mc_plot_FMD.pdf", sep=""))
plot(FMD$m, FMD$cum, log="y", xlab="Magnitude", ylab="Number of events",
     main="Frequency-Magnitude Distribution")
points(FMD$m, FMD$noncum, pch=2)
abline(v=Mc_mean)
legend("topright", c("Cum. FMD", "Non Cum. FMD"), cex=0.8, pch=c(1, 2))
dev.off()
```

## References

- Aki, K. (1965), Maximum likelihood estimate of  $b$  in the formula  $\log N = a - bM$  and its confidence limits, *Bull. Earthq. Res. Inst., Univ. Tokyo*, *43*, 237–239. [7](#)
- Albarelo, D., R. Camassi, and A. Rebez (2001), Detection of Space and Time Heterogeneity in the Completeness of the Seismic Catalog by a Statistical Approach: An Application of the Italian Area, *Bull. Seismol. Soc. Am.*, *91*, 1694–1703. [5](#)
- Amor ese, D. (2007), Applying a change-point detection method on frequency-magnitude distributions, *Bull. Seismol. Soc. Am.*, *97*, doi:10.1785/0120060181. [4](#), [6](#), [14](#), [15](#), [21](#), [23](#), [36](#), [37](#)
- Bachmann, C. E. (2011), New approaches towards understanding and forecasting induced seismicity, Ph.D. thesis, ETH Zurich, doi:10.3929/ethz-a-006715437. [11](#)
- Cao, A. M., and S. S. Gao (2002), Temporal variations of seismic  $b$ -values beneath northeastern japan island arc, *Geophys. Res. Lett.*, *29*, doi:10.1029/2001GL013775. [6](#), [8](#), [14](#), [15](#), [18](#), [19](#)
- Cornell, C. A. (1968), Engineering seismic risk analysis., *Bull. Seismol. Soc. Am.*, *58*(5), 1583–1606. [7](#)
- D’Alessandro, A., D. Luzio, G. D’Anna, and G. Mangano (2011), Seismic Network Evaluation through Simulation: An Application to the Italian National Seismic Network, *Bull. Seismol. Soc. Am.*, *101*, doi:10.1785/0120100066. [4](#), [9](#)
- Efron, B. (1979), 1977 rietz lecture, bootstrap methods – another look at the jackknife, *Ann. Statist.*, *7*, 1–26. [8](#), [14](#), [15](#)
- Gardner, J. K., and L. Knopoff (1974), Is the sequence of earthquakes in Southern California, with aftershocks removed, Poissonian?, *Bull. Seismol. Soc. Am.*, *64*, 1363–1367. [24](#)
- Gomberg, J. (1991), Seismicity and Detection/Location Threshold in the Southern Great Basin Seismic Network, *J. Geophys. Res.*, *96*, 16,401–16,414. [9](#)
- Gutenberg, B., and C. F. Richter (1944), Frequency of earthquakes in california, *Bull. Seismol. Soc. Am.*, *34*, 184–188. [5](#)
- Habermann, R. E. (1987), Man-made changes of seismicity rates, *Bull. Seismol. Soc. Am.*, *77*, 141–159. [6](#), [7](#)
- Helmstetter, A., Y. Y. Kagan, and D. D. Jackson (2007), High-resolution time-independent grid-based forecast for  $M \geq 5$  earthquakes in california, *Seismol. Res. Lett.*, *78*, 78–86. [6](#)
- Husen, S., and J. L. Hardebeck (2010), Earthquake location accuracy, *Community Online Resource for Statistical Seismicity Analysis*, Available at <http://www.corssa.org>, doi:10.5078/corssa-55815573. [3](#), [4](#)
- Hutton, K., J. Woessner, and E. Hauksson (2010), Earthquake monitoring in southern california for seventy-seven years (1932–2008), *Bull. Seismol. Soc. Am.*, *100*, 423–446, doi:10.1785/0120090130. [8](#), [28](#), [30](#)
- Ishimoto, M., and K. Iida (1939), Observations of earthquakes registered with the microseismograph constructed recently, *Bull. Earthq. Res. Inst.*, *17*, 443–478. [5](#)
- Iwata, T. (2008), Low detection capability of global earthquakes after the occurrence of large earthquakes: investigation of the harvard cmt catalogue, *Geophys. J. Int.*, *174*, doi:10.1111/j.1365-246X.2008.03864.x. [19](#), [28](#)
- Kagan, Y. Y. (2002), Seismic moment distribution revisited: I. statistical results, *Geophys. J. Int.*, *148*, 520–541. [9](#), [19](#)
- Kagan, Y. Y. (2003), Accuracy of modern global earthquake catalogs, *Phys. Earth Planet. Int.*, *135*, 173–209. [6](#), [17](#)
- Kvaerna, T., and F. Ringdal (1999), Seismic Threshold Monitoring for Continuous Assessment of Global Detection Capability, *Bull. Seismol. Soc. Am.*, *89*, 946–959. [4](#), [9](#)
- Kvaerna, T., F. Ringdal, J. Schweitzer, and L. Taylor (2002a), Optimized seismic threshold monitoring - part 1: Regional processing, *Pure Appl. Geophys.*, *159*(5), 969–987. [9](#)
- Kvaerna, T., F. Ringdal, J. Schweitzer, and L. Taylor (2002b), Optimized seismic threshold monitoring - part 2: Teleseismic processing, *Pure Appl. Geophys.*, *159*(5), 989–1004. [9](#)
- Marsan, D. (2003), Triggering of seismicity at short timescales following californian earthquakes, *J. Geophys. Res.*, *108*, doi:10.1029/2002JB001946. [6](#), [18](#)
- Mignan, A. (2011), Retrospective on the Accelerating Seismic Release (ASR) hypothesis: Controversy and new horizons, *Tectonophysics*, *505*, 1–16, doi:10.1016/j.tecto.2011.03.010. [7](#)
- Mignan, A., M. J. Werner, S. Wiemer, C.-C. Chen, and Y.-M. Wu (2011), Bayesian estimation of the spatially varying completeness magnitude of earthquake catalogs, *Bull. Seismol. Soc. Am.*, *101*, doi:10.1785/0120100223. [6](#), [15](#), [21](#), [24](#), [25](#), [27](#), [28](#), [30](#), [32](#), [33](#)
- Nanjo, K. Z., T. Ishibe, H. Tsuruoka, D. Schorlemmer, Y. Ishigaki, and N. Hirata (2010a), Analysis of the Completeness Magnitude and Seismic Network Coverage of Japan, *Bull. Seismol. Soc. Am.*, *100*, doi:10.1785/0120100077. [30](#), [32](#)

- Nanjo, K. Z., D. Schorlemmer, J. Woessner, S. Wiemer, and D. Giardini (2010b), Earthquake detection capability of the Swiss Seismic Network, *Geophys. J. Int.*, *181*, 1713–1724, doi:10.1111/j.1365-246X.2010.04593.x. [9](#), [11](#), [33](#)
- Naylor, M., K. Orfanogiannaki, and D. Harte (2010), Exploratory data analysis: magnitude, space, and time, *Community Online Resource for Statistical Seismicity Analysis*, Available at <http://www.corssa.org>, doi:10.5078/corssa-92330203. [3](#), [5](#), [7](#), [37](#)
- Ogata, Y., and K. Katsura (1993), Analysis of temporal and spatial heterogeneity of magnitude frequency distribution inferred from earthquake catalogues, *Geophys. J. Int.*, *113*, 727–738. [9](#), [14](#), [19](#), [25](#)
- Ogata, Y., and K. Katsura (2006), Immediate and updated forecasting of aftershock hazard, *Geophys. Res. Lett.*, *33*, doi:10.1029/2006GL025888. [19](#), [20](#)
- Plenkens, K., D. Schorlemmer, G. Kwiatek, and J. R. Group (2011), On the probability of detecting picoseismicity, *Bulletin of the Seismological Society of America*, *101*(6), 2579–2591, doi:10.1785/0120110017. [9](#), [11](#)
- Ringdal, F. (1975), On the estimation of seismic detection thresholds, *Bull. Seismol. Soc. Am.*, *65*, 1631–1642. [9](#)
- Ringdal, F. (1986), Study of magnitudes, seismicity, and earthquake detectability using a global network, *Bull. Seismol. Soc. Am.*, *76*, 1641–1659. [9](#)
- Ringdal, F., and T. Kvaerna (1989), A multi-channel processing approach to real time network detection, phase association, and threshold monitoring, *Bull. Seismol. Soc. Am.*, *79*, 1927–1940. [9](#)
- Rydelek, P. A., and I. S. Sacks (1989), Testing the completeness of earthquake catalogs and the hypothesis of self-similarity, *Nature*, *337*, 251–253. [4](#), [7](#), [14](#), [15](#), [23](#), [24](#)
- Rydelek, P. A., and I. S. Sacks (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. Seismol. Soc. Am.*, *93*, 1862–1867. [30](#)
- Schorlemmer, D., and J. Woessner (2008), Probability of detecting an earthquake, *Bull. Seismol. Soc. Am.*, *98*, doi:10.1785/0120070105. [4](#), [9](#), [10](#), [12](#), [13](#), [33](#), [34](#), [35](#)
- Schorlemmer, D., A. Christophersen, A. Rovida, F. Mele, M. Stucchi, and W. Marzocchi (2010a), Setting up an earthquake forecast experiment in Italy, *Ann. Geophys.*, *53*, doi:10.4401/ag-4844. [30](#)
- Schorlemmer, D., F. Mele, and W. Marzocchi (2010b), A completeness analysis of the national seismic network of italy, *J. Geophys. Res.*, *115*, doi:10.1029/2008JB006097. [9](#), [11](#)
- Sereno, T. J. J., and S. R. Bratt (1989), Seismic Detection Capability at NORESS and Implications for the Detection Threshold of a Hypothetical Network in the Soviet Union, *J. Geophys. Res.*, *94*, 10,397–10,414. [9](#)
- Shi, Y., and B. A. Bolt (1982), The standard error of the magnitude- frequency b-value, *Bull. Seismol. Soc. Am.*, *72*, 1677–1687. [18](#)
- Stucchi, M., P. Albini, C. Mirto, and A. Rebez (2004), Assessing the completeness of Italian historical earthquake data, *Ann. Geophys.*, *47*, 659–673. [5](#)
- Tormann, T., S. Wiemer, and E. Hauksson (2010), Changes of reporting rates in the southern california earthquake catalog, introduced by a new definition of  $m_l$ , *Bull. Seismol. Soc. Am.*, *100*, doi:10.1785/0120090124. [7](#)
- van Stiphout, T., J. Zhuang, and D. Marsan (2010), Seismicity declustering, *Community Online Resource for Statistical Seismicity Analysis*, Available at <http://www.corssa.org>, doi:doi:10.3929/ethz-a-xxxxxxx. [24](#)
- von Seggern, D. H. (2004), Seismic Background Noise and Detection Threshold in the Southern Great Basin Digital Seismic Network, *Bull. Seismol. Soc. Am.*, *94*, 2280–2298. [9](#)
- von Seggern, D. H., and R. Blandford (1976), Seismic Threshold Determination, *Bull. Seismol. Soc. Am.*, *66*, 753–788. [9](#)
- Wesnousky, S. G. (1994), The Gutenberg-Richter or Characteristic Earthquake Distribution, Which Is It?, *Bull. Seismol. Soc. Am.*, *84*, 1940–1958. [5](#), [21](#)
- Wiemer, S. (2001), A software package to analyze seismicity: ZMAP, *Seismol. Res. Lett.*, *72*, 374–383. [17](#), [36](#), [37](#)
- Wiemer, S., and D. Schorlemmer (2007), Alm: An asperity-based likelihood model for california, *Seismol. Res. Lett.*, *78*, 134–140. [7](#)
- Wiemer, S., and M. Wyss (2000), Minimum magnitude of complete reporting in earthquake catalogs: examples from alaska, the western united states, and japan, *Bull. Seismol. Soc. Am.*, *90*, 859–869. [6](#), [7](#), [14](#), [15](#), [16](#), [17](#), [24](#), [28](#), [29](#), [30](#), [32](#)
- Wiemer, S., D. Giardin, D. Fäh, N. Deichmann, and S. Sellami (2009), Proababilistic seismic hazard assessment for Switzerland: best estimates and uncertainties, *J. of Seismology*, doi:10.1007/s10950-008-9138-7. [7](#)
- Woessner, J., and S. Wiemer (2005), Assessing the quality of earthquake catalogues: Estimating the magnitude of completeness and its uncertainty, *Bull. Seismol. Soc. Am.*, *95*, doi:10.1785/012040007. [3](#), [4](#), [5](#), [6](#), [9](#), [14](#), [15](#), [17](#), [18](#), [19](#), [21](#), [22](#), [25](#), [26](#), [27](#), [28](#), [29](#), [30](#)

- 
- Woessner, J., J. L. Hardebeck, and E. Hauksson (2010), What is an instrumental seismicity catalog - draft, *Community Online Resource for Statistical Seismicity Analysis*, Available at <http://www.corssa.org>, doi: 10.5078/corssa-38784307. 3
- Wyss, M., A. Hasegawa, S. Wiemer, and N. Umino (1999), Quantitative mapping of precursory seismic quiescence before the 1989, m7.1 off-sanriku earthquake, japan, *Annali Di Geofisica*, 42, 851–869. 6, 14, 29
- Zuniga, F. R., and M. Wyss (1995), Inadvertent changes in magnitude reported in earthquake catalogs: Their evaluation through b-value estimates, *Bull. Seismol. Soc. Am.*, 85, 1858–1866. 5, 7