Estimating the Magnitude of Completeness and Its Uncertainty

Thesis submitted in partial fulfilment of the requirements for the
BS-MS Dual Degree Programme

Indian Institute of Science Education and Research, Pune

By

Vrushali Rajesh Sarwan

20141125

Under the Guidance of

Dr. Utsav Mannu, INSPIRE Faculty, IISER Pune

Dr. Shyam Nandan, Postdoctoral Researcher, ETH Zurich

Certificate

This is to certify that this dissertation entitled "**Estimating the magnitude of completeness and its uncertainties**" towards the partial fulfilment of the BSMS dual degree programme at the Indian Institute of Science Education and Research, Pune represents study/work carried out by "**Vrushali Rajesh Sarwan**" at "IISER Pune" under the supervision of "**Dr. Utsav Mannu**, Inspire Faculty, Department of Earth and Climate Sciences" during the academic year 2018-2019.

Utsav Mannu

Signature of Supervisor
Dr. Utsav Mannu

Date : 01/04/2019

Signature of Student
Vrushali Rajesh Sarwan

Date : 01|04|2019

Declaration

I hereby declare that the research work presented in the report entitled "**Estimating the magnitude of completeness and its uncertainties**" have been carried out by me at the Department of Earth and Climate Sciences, IISER Pune, under the supervision of **Dr. Utsav Mannu** and the same has not been submitted elsewhere for any other degree.

Utsav Mannu

Signature of Supervisor

Dr. Utsav Mannu

Date : 1/04/2019

Signature of Student

Vrushali Rajesh Sarwan

Date : 01/04/2019

Acknowledgements

First of all, I am most grateful to my supervisor, Dr. Utsav Mannu for his continuous concern and support since the first day I invaded the Geophysics Department of IISER Pune. The scientific assistance that he offered me together with his encouragement to move on whenever I was stuck somewhere and was not able to find a path of how to proceed further, significantly helped me to build my scientific profile. I genuinely appreciate his efforts and time he invested towards improving my understanding of the concepts of geophysics and statistical seismology along with clearing my silly doubts that use to rise because of being naive to the subject. He played a major role in developing my skills of MATLAB programming, bash scripting, and extending my horizons of understanding and developing algorithms. I admire the time that he devoted to examining my codes along with identifying and rectifying the errors even if they were trivial. I also acknowledge his attitude of constantly motivating me to apply for International Student Internship Programs as well as in seeking Ph.D. positions.

I wish to express my sincere gratitude to my TAC member Dr. Shyam Nandan, Postdoctoral Researcher, ETH Zurich, for the stream of knowledge he offered me and for insinuating at new ideas to address certain queries and exploring issues from a different point of views. I'm also very grateful to Professor Shyam Rai, Head of the Earth and Climate Science Department of IISER Pune for suggesting me to work under the guidance of Dr. Utsav Mannu and Dr. Shyam Nandan. I would like to thank Dr. Rahul Dehiya and Dr. Sudipta Das for letting me be a part of an on-field demo session of installation of geophones thereby providing me with the opportunity to earn important knowledge concerning field experiments, catalog elaboration and data acquisition process. In the end, I would like to appreciate the staff, all the Geophysics lab members, my friends and my family for supporting and motivating me to keep trying and have patience!

Table of Contents

Page no.

List of figures and Tables

Abstract

Earthquake catalogs are not complete over the entire range of magnitudes. A preliminary step that should be performed before any seismicity and hazard-related studies is to assess the quality, consistency and completeness of the earthquake catalogs. This can be achieved by assessing a threshold magnitude called magnitude of completeness, M_c, defined as the lowest magnitude above which all the magnitudes follow Gutenberg-Richter law(GR law). Assessing M_c have received considerable attention in the last few decades. In general, most of the catalog based methods are deployed by fitting GR law fit the observed Frequency magnitude distribution (FMD) of the earthquake magnitudes. Although, the limitation of these methods in estimating M_c is that they fail in the case of less number of events in the catalog. We propose new catalog-based methods would work even with less number of events in the catalog. The stochastic method used for generating the synthetics for testing the method was by modelling the FMD using the probability density function(pdf) of the normal distribution to model FMD below M_c and GR law for magnitudes greater than or equal to M_c. The best estimate of M_c was drawn from a set of assumed M_c by using two methods. We check which of these assumed $M_c's$ satisfies the criteria of method 1) KS distance approach and method 2) maximum probability approach, by comparing original FMD with the modelled FMD. A comparative analysis was carried out to check the performance of the proposed methods with those of three existing catalog based methods, using generated synthetics. Furthermore, we are planning to develop synthetic catalog by incorporating the uncertainties associated with the earthquake magnitudes. In addition, we are focussed to come up with realistic synthetic catalogs which carry the spatial and temporal similarity with the catalog.

CHAPTER 1

Introduction

1.1 Aim of the Study

Earthquake catalogs are considered to be one of the foremost necessary products of geophysical sciences. Seismological research highly depends on the use of earthquake catalogs as the source of data in regards to the spatial and temporal distribution of earthquakes. They are a primary result of the seismological network and a general source of information for varied studies such as earthquake physics, seismicity, seismotectonics and hazard analysis. Seismological network evolves over time as a result of improved instrumentation and the progress in better understanding of the earth's structure (Hutton, 2010). The spatial and temporal properties of the seismic network considerably affect the level of earthquake detections and lead to inhomogeneous earthquake catalogs. The question that arises here is why not all the earthquakes detected. The reasons responsible for scarcity of detection of smaller magnitude events as outlined by Mignan, A., J. Woessner (2012) such as, (1) not able to distinguish smaller events from the background noise on the seismograph, (2) for an event to be reported, a minimum number of stations should have received the signal in order to commence the location procedure, and (3) network operators have an authority to choose a lower bound and discard all events below it. As a result, the current catalogs available are considered to be only complete up to a certain magnitude(the magnitude of completeness, M_c) and for events greater than it. Using events with magnitude less than the magnitude of completeness, i.e. incomplete data, leads to inaccurate assessments of Gutenberg-Richter law(GR law) parameters and erroneous seismicity interpretations. Previous studies have been carried out to resolve the critical issue of completeness of catalog by estimating a completeness magnitude, M_c theoretically defined as the threshold magnitude above which 100% of the events in a space-time volume are detected (Rydelek and Sacks, 1989). The aim of the thesis is to propose a new catalog based method to estimate M_c and its uncertainty as well as perform a comparative analysis with the existing deployed methods. Although the estimation of M_c is performed routinely, these state of the art methods is based on different FMD assumptions and results in different values of the estimation of M_c. The uncertainties of the earthquake catalogs along with the intrinsic assumptions that go into the pre-processing of the earthquake catalog.

1.2 Theoretical Background

M_c is theoretically defined as the minimum magnitude above which all earthquakes are reliably recorded in a given space-time window. Methods used for Estimation of M_c can be classified into two categories Network-based methods(Schorlemmer and Woessner 2008; D'Alessandro et al. 2011) and catalog-based methods(Rydelek and Sacks, 1989; Woessner and Wiemer, 2005). Network-based methods are based on the detection and sensitivity properties of the seismic network with prior information of the density and distribution of stations.

This approach uses a probability-based magnitude of completeness $M_P(x,t)$, at a given location x, time t and a predefined probability level P based on the number of network stations available, $M_p(x,t)$ is defined as the lowest magnitude at which the probability of detection $P_E(m,x,t)$ is $1-Q$, where Q is the probability that an earthquake is not detected. This implies probabilistic magnitude of completeness is the function of x,t,Q is given by -

$M_P(x,t,Q) = min(m| P_E(m,x,t) = 1-Q)\ where\ m \,\varepsilon\, M$ and M is the interval of possible magnitudes of completeness.

Whereas the catalog-based methods obeys a different definition of the magnitude of completeness, M_c. It is defined as the lowest magnitude at which the FMD deviates from the GR law.

Comparing the aforementioned definitions of the magnitude of completeness we can say that Network-based methods are always better than Catalog based methods.

However, ensuring this fact is not a trivial task as it involves the understanding of seismicity as well as mixing of waveforms of two events which is very difficult to accomplish.

1.2.1 Frequency Magnitude Distribution(FMD)

Frequency Magnitude Distribution as the name suggests is the visual representation of the variation of the frequency of magnitudes in a specified bin size, with respect to the magnitudes in a given earthquake catalog.

1.2.2 Gutenberg Richter Law

In this section, the basic principles of earthquake frequency-magnitude distribution are presented along with the description and determination process of the involved parameters. Gutenberg Richter law (GR law) illustrates the relationship between the frequency of magnitudes and the occurrence of earthquakes. The GR law is given by-

$$log_{10}(N) = a - bm \qquad 1.1$$

where, N is the cumulative number of earthquakes having magnitudes larger than M, and a and b are constants. The parameter b commonly referred to as b-value is commonly closed to 1.0 in seismically active regions(Lay and Wallace, 1995). Fig 1.1 shows the FMD of the California catalog used for the study.

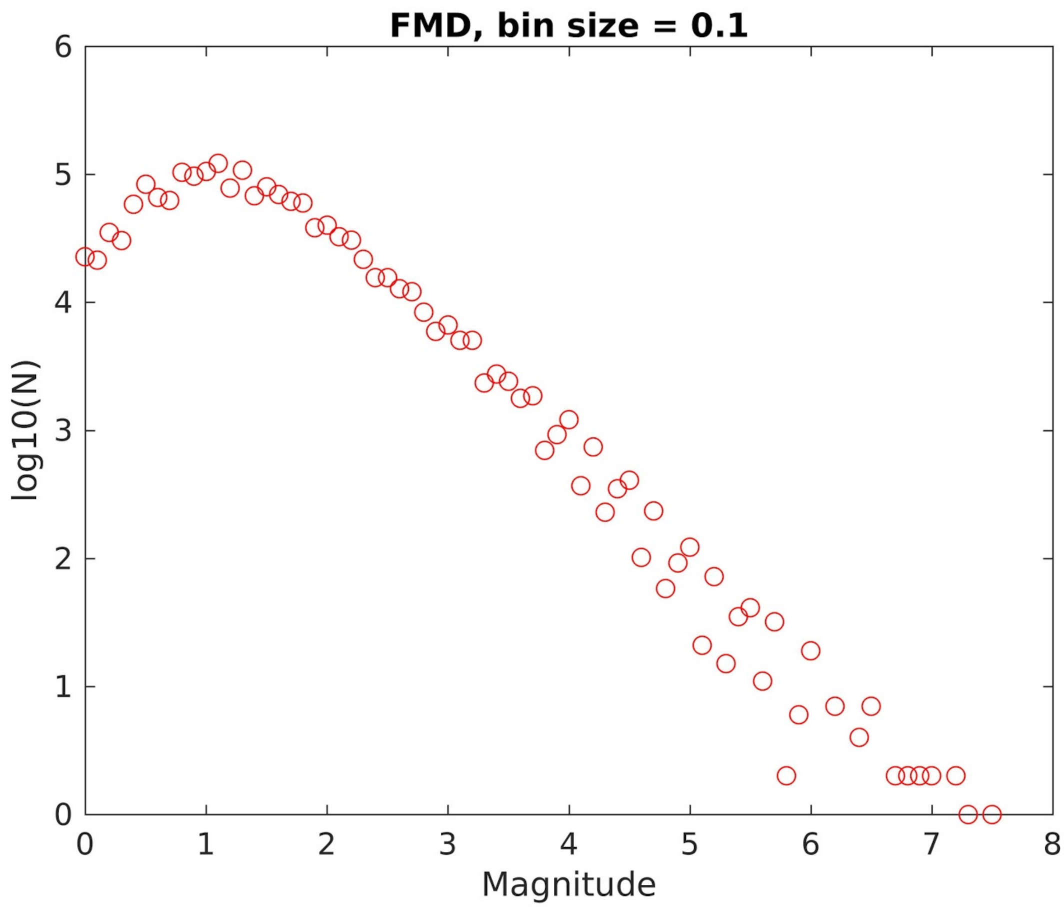

Fig 1.1 Frequency magnitude distribution(FMD) of the California Catalog

1.2.3 Significance of estimating GR law parameter b-value and M_c

Estimation of the magnitude of completeness, M_c has a direct influence on the evaluation of GR law parameter i.e. $b-value$ as reported by C. Godano, E. Lippiello, L. de Arcangelis, (2014). In general, the GR law parameters are the basis of seismic hazard studies (Cornell, 1968) and of earthquake forecast models (Wiemer and Schorlemmer 2007). The spatiotemporal variation of $b-value$ in a given region is highly linked to the characteristics of seismic hazard analysis, for instance, high b-values in the magma chambers indicates high seismicity in the region (Sanchez et al., 2004; Wiemer and McNutt, 1997). It has also been outlined by Schorlemmer et al., (2005) the regions with low $b - value$ implies that large differential stress of the earth's crust and thereby pointing towards the end of the seismic cycle. Hence, correct estimation of M_c and in turn GR law parameters is an essential task.

1.3 State of the art Methods: Review

1. Maximum Curvature (MAXC),(Wiemer and Wyss, 2000)

Maximum curvature method is a non-parametric method and is considered to be one of the fastest methods to estimate M_c. It determines the maximum of the first derivative of FMD. Also, the most common practice of using this method is to find the bin of magnitudes with the highest frequency in non-cumulative FMD, this matches with the former approach i.e. the maximum of the first derivative of FMD.

2. The M_c by b−value stability (MBS) method,(Cao and Gao, 2002)

In this approach, the M_c is estimated by studying the stability of b-value with respect to a cut off magnitude, M_{co}. Woessner and Wiemer, (2005) named the method as MBS. The author found that b-value increases for $M_{co} < M_c$ and for $M_{co} >= M_c$ it doesn't change. The objective was to stabilize the b-value numerically, the method was modified by Woessner and Wiemer, (2005) he defined an uncertainty measure of b-value given by

$$\delta b = 2.3b^2 \sqrt{\sum_{i=1}^{n} (M - \overline{M})^2 / N(N-1)} \qquad 1.2$$

Where $\overline{M}$ is the mean magnitude and N is the total no. of events in the catalog. Further, for the estimation of M_c is defined as the first magnitude at which $\Delta b = |bave - b| \le \delta b$, where $bave$ is the mean of the evaluated b-values for the successive cut off magnitude bins of size 0.5.

3. Goodness-of-Fit Test (GFT)

GFT was introduced by Wiemer and Wyss, (2000), the test estimated M_c by comparing the synthetic data with the FMD. The goodness of fit is computed using the following parameter.

$$R(a,b,M_{co}) = 100-(\sum_{Mco}^{Mmax}(|B_i - S_i| / \sum_i B_i) * 100) \quad 1.3$$

where B_i and S_i are the observed and predicted value cumulative number of events in each magnitude bin. The first cut off magnitude M_{co} at which the value of R comes out to be 90% or 95% is defined as estimated M_c .

All the above-mentioned methods have been implemented using various synthetics and compared with the proposed method(see Chapter 3).

CHAPTER 2

Method

Assessing the magnitude of completeness M_c is one of the prerequisites for acquiring a complete catalog, hence making it reliable for subsequent seismic analysis. Assuming that earthquakes follow GR law, M_c can be defined as the lowest magnitude at which the FMD deviates from the exponential decay (Zuniga and Wyss, 1995). In an ideal case, we can ensure that the above-defined definition of M_c reconciles with the actual definition of M_c (described in section 1.2) whereas in the case of real catalogs ensuring that this definition of M_c adapts the actual definition of M_c is a challenging task. Obeying this definition of M_c the chapter throws light on two different approaches that we propose to estimate the GR law parameter b-value followed by estimation of M_c.

2.1 Maximum probability Estimator(MPE)

To assess the completeness of catalog we first tested a new method to evaluate the magnitude of completeness, M_c. The following derivation of estimating the b-value describes all the steps involved in evaluating β.

Assuming that the magnitudes are discrete, the data greater than and equal to M_c is represented as

$\overline{m} = \{m_1,\ m_2,\ m_3, ..., m_n\}$ where n is the total number of events in the dataset.

The probability mass function, PMF is the exponential relation exhibiting GR law, given by:

$$f(m_i|\overline{m}) = N_c(\beta) * exp(-\beta m_i) \ , \quad \forall \ m_i \geq M_c \qquad 2.1$$

The summation of the PMF multiplied by a normalizing constant for all the events above in the catalog should be 1 i.e.

$$N_c(\beta) * \left[e^{-\beta m_i} + e^{-\beta m_{i+1}} + e^{-\beta m_{i+2}}, \ .... \right] = 1$$

$$N_c(\beta) * \sum_{i}^{\infty} exp(-\beta m_i) = 1 \qquad 2.2$$

$$\Rightarrow \quad N_c(\beta) = 1/ \sum_{i=1}^{\infty} exp(-\beta m_i) \qquad 2.3$$

The summation term in the denominator of the above equation is an infinite geometric series with first term $a = exp(-\beta m_1)$ where $m_1 = m_c$ and common ratio $r = exp(-\beta \Delta m)$, where $\Delta m = m_{i+1} - m_i$ i.e. the difference between two consecutive magnitudes in an ordered list of events.

Let us denote this sum by S_∞, given by $S_\infty = a/1 - r$

$$\text{Therefore, } S_\infty = exp(-\beta m_c)/ 1 - exp(-\beta \Delta m) \qquad 2.4$$

Substituting S_∞ in eq 2.3

$$\Rightarrow N_c(\beta) = 1 - exp(-\beta \Delta m)/ exp(-\beta m_c) \qquad 2.5$$

The likelihood of the PMF can be defined as

$L(\beta|\overline{m}) = \prod_{i=1}^{n} exp(-\beta m_i) * N_c(\beta)$, where $\overline{m}$ is the given set of events in the catalog

2.6

Eq 2.6 can further be written as

$$L(\beta|\overline{m}) = (N_c(\beta))^n \prod_{i=1}^{n} exp(-\beta m_i) \qquad 2.7$$

Taking log both the sides would result in log-likelihood given by

$$LL(\beta|\overline{m}) = n * log((N_c(\beta)) - \sum_{i=1}^{n} \beta m_i \tag{2.8}$$

Substituting the value of $N_c(\beta)$ in the above equation becomes

$$LL(\beta|\overline{m}) = n * log(1 - exp(-\beta \Delta m)) - \beta * \sum_{i=1}^{n} (m_i - m_c) \tag{2.9}$$

Differentiating Eq. 2.9 with respect to β and equating, further equating it to zero to solve for β will give an expression for the estimated value of β denoted by $\widehat{\beta}$ and is given by the following equation

$$\widehat{\beta} = (\Delta m)^{-1} * log(1 + n\Delta m / \sum_{i=1}^{n} (m_i - m_c)) \tag{2.10}$$

Furthermore, to evaluate M_c we compute the likelihood points which is a function of the data $\overline{m}, \beta \; and \; m_c$ given by equation 2.11

$$LP_i(\overline{m}, m_c, \beta) = log(1 - exp(-\beta \Delta m)) - \beta * (m_i - m_c) \;\; \forall \, i = 1 \; to \; n \tag{2.11}$$

Using the likelihood points a true log likelihood value is computed denoted by LL_{true}

$$LL_{true} = max(LP(\overline{m}, m_c, \beta)) \tag{2.12}$$

The pair of $m_c \; and \; \beta$ for which LL_{true} is defined is used to produce a GR law synthetics given by Eq. …. for N number of times. Again the same procedure is obeyed to estimate β (say β^*) using Eq 2.10 followed by calculating the true likelihoods (Eq 2.12) for all the generated N synthetics denoted by LL_{eff}

$$LL_{eff} = \{LL_{true1,} LL_{true2}, ... LL_{truej}, ..., LL_{trueN}\} \tag{2.13}$$

A 90% interquartile range of LL_{eff} is calculated, if LL_{true} (eq 2.12) lies in this range then m_c for which LL_{true} is defined is considered to be our estimated M_c .

Our algorithm to estimate M_c is as follows,

1. Given an earthquake catalogue of the region of study, extract the magnitude attribute from the catalogue.
2. Bin the magnitudes, by fixing the bin width (eg $\delta m = 0.1$).
3. Assume an M_c, for instance, $M_c = 1.1$
4. Using the assumed M_c and the binned magnitudes estimate the b-value(eq. 2.10) and remove all the magnitudes which are less than this assumed M_c.
5. Compute the log likelihood points of the binned data, given by eq 2.11, followed by calculating the true likelihood value(eq 2.12)
6. Repeat the steps 4 and 5 for all assumed of M_c.
7. Sort the set of assumed M_c and b-values with respect to the log likelihood (in descending order).
8. Generate GR synthetic data(see section 2.3.1) for the first pair sorted M_c and b-value(Note: this pair of M_c and b-value have the max log likelihood value) which follows GR law (eq 2.1).
9. Repeat steps 4 and 5 for the synthetics generated in step 8.
10. Reiterate the 8th and 9th steps for 10,000 times, and store the log likelihood for each iteration. This will result in a distribution of log-likelihoods(say LL_eff).
11. The M_c for which the true likelihood lies in the 90% confidence interval of LL_eff is our estimated M_c.absolute difference between assumed Mc and the Mc used for generating GR synthetics is the estimated Mc i.e. the desired output, else repeat this step until the result is obtained.

2.2 Kolmogorov–Smirnov test (KS - test)

The goal of this method is to estimate the value of M_c by making the probability distributions of the observed data and the best fit GR model as similar as possible for all the magnitudes above M_c. To compute the distance between two probability distributions, we used Kolmogorov - Smirnov or KS statistic, defined as the maximum distance between the cumulative distribution functions (CDFs) of the observed data and the fitted model, given by eq 2.14

$$D = max\ |S(x) - P(x)| \qquad 2.14$$

where $S(x)$ is the CDF of the observed data and $P(x)$ is the CDF of the GR model that best fits the data for all the magnitudes above M_c.

The CDF of the GR model is given by-

$$F(m) = 1 - exp(-\beta(m - m_c)) \qquad 2.15$$

The method can be applied to both discrete and continuous data. The value of β in the above equation is given by the formula-

$$\beta = 1/ < M > - M_c \qquad 2.16$$

where $< M >$ is the mean of all the magnitudes greater than or equal to M_c

The steps involved in this approach to estimate M_c are explained by the help of a flowchart (Fig 2.1)

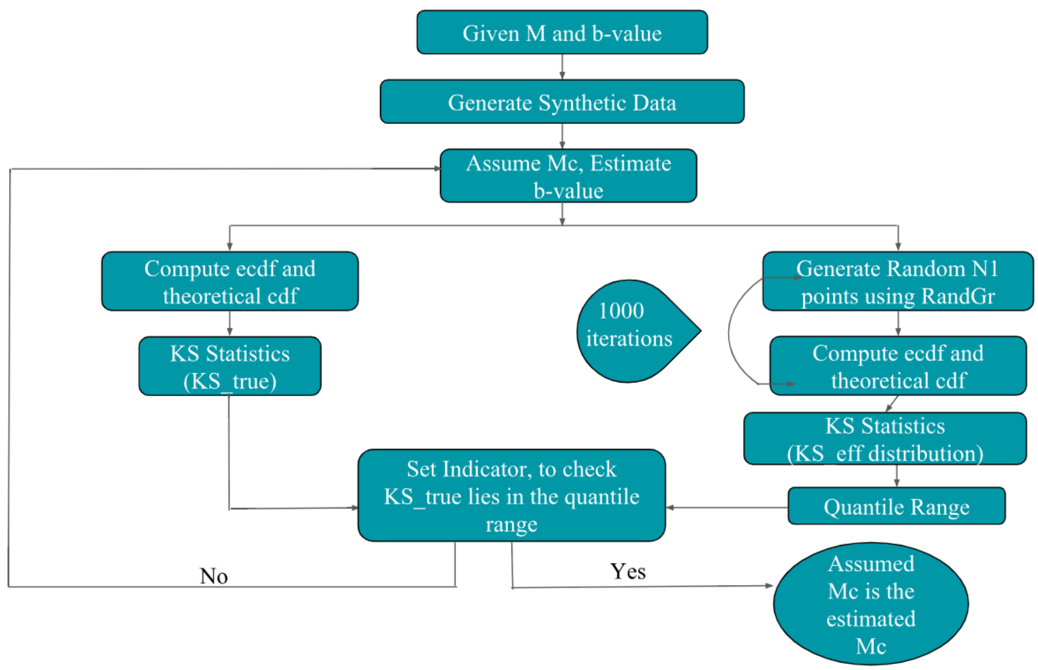

Fig 2.1 Algorithm- Flowchart of the testing the KS test

The method is explained in brief in the following algorithm.

To begin with, we assume an M_c . And estimate the b-value using the MLE(eq 2.11).

The basic idea of our method is to use the Kolmogorov-Smirnov test (KS test).

Algorithm

The general algorithm of our method is the following:

1. Using the KS test compute Kolmogorov-Smirnov statistic (eq 2.13) for the original catalogue and call it KStrue.
2. Next, we compute the distribution of magnitudes with the same β and M_c values, which represents a perfect fit to a power law, using the random uniform function, U(0,1) and define this function to be a random generator (RandGR hereafter).

$$M = m_c - (log(1 - U(0,1))/\beta) \qquad 2.14$$

3. Computing the Kolmogorov-Smirnov statistic for the distribution generated in step

2, and name it KSeff.

4. Iterating steps 2 and 3 for 10,000 times and forms a KSeff distribution.
5. Using the quantile MATLAB built-in function, we compute the 90% confidence interval of the KSeff distribution.
6. The M_c for which the difference between KStrue and the confidence interval computed in step 5 is less than zero, we claim that to be our estimated M_c (EM_c).
7. To find the error in estimation is computed by calculating the absolute difference between the EM_c and M_c.

2.3 Log-Likelihood test(LL-test)

After estimating the value of β (eq. 2.10), the maximum likelihood estimator is been used to compute the log-likelihood by using the following formula:

$$LL_{true} = N * log(1 - exp(-\beta\Delta m)) - \beta * \sum_{i=1}^{N}(m_i - m_c) \quad 2.15$$

where N is the number of events greater than or equal to m_c in a given catalog.
The LL_{true} is computed for all the assumed $m_c's$.
The pair of m_c *and* β for which LL_{true} is defined is used to produce a GR law synthetics given by Eq. …. N times. Again the same procedure is obeyed to estimate β (say β^*) using Eq 2.10 followed by calculating the true likelihoods (Eq 2.15) for all the generated N synthetics denoted by LL_{eff}

$$LL_{eff} = \{LL_{true1,} LL_{true2}, ... LL_{truej}, ..., LL_{trueN}\} \quad 2.16$$

A 90% interquartile range of LL_{eff} is calculated, if LL_{true} (eq 2.15) lies in this range then m_c for which LL_{true} is defined is considered to be our estimated M_c.

2.4 Extension of MLE (EMLE)

Furthermore, in order to take into account the uncertainties of the magnitudes recorded at the seismic stations while estimating the magnitude of completeness, an algorithm has been designed.

1) The first step of the method is to deal with the uncertainties of the magnitudes by binning the magnitudes of the catalog by the following formula:

$$m^*_i = [(m_i - m_o)/\delta m] * \delta m + m_o + \delta m/2 \quad 2.17$$

where, m^*_i is the ith binned magnitude, δm is the bin width and m_o is the minimum magnitude of the catalog.

2) The unique values of the binned magnitudes are used as a set of assumed M_c compute the further calculations, let these unique values of magnitudes be denoted by $\{m^*_1, m^*_2, ..., m^*_J\}$ and let the no. of events in these non-empty bins be $\{n^*_1, n^*_2, ..., n^*_J\}$

3) Assuming m^*_j to be the magnitude of completeness using the equations 1.3 and 2.10 the estimate of $b - value$ can be computed using the following equation

$$b = (\delta m * log10)^{-1} * log((1 + \delta m) * N) / \sum_{i=1}^{N}(m^*_i - m^*_j)) \quad 2.18$$

where m^*_i is the set of magnitudes which are greater than m^*_j and $N = \sum_{k=j}^{J} n^*_k$ is the total number of magnitudes greater than m^*_j in the binned catalog.

4) The expected number of events in the kth non-empty bin $[m^*_k - \delta m/2, m^*_k + \delta m/2]$ is given by:

$E[n^*_k] = N * (10^{-b(m^*_k - m^*_j)} - 10^{-b(m^*_k + \delta m - m^*_j)})$, the process of finding the expected number of events in the kth bin is reiterated for all the non-empty bins defined above for which $m^*_k \geq m^*_j$.

5) For each of the above defined non-empty bins, we plan to check whether the observed number of events i.e. n^*_k falls within the 95%ile of a Poissonian distribution whose mean is given by $E[n^*_k]$. If this is true then the magnitude bin will be considered as consistent. A fraction of such consistent bins is computed say f, if f is greater than 90%, then the

combination of m^{*}_{j} , δm and b will be an appropriate combination for the given earthquake catalog.

6) Reiterate steps 3-5 for all the unique magnitudes assumed in step 2.

7) Repeat the steps 2-6 for any typical value of δm.

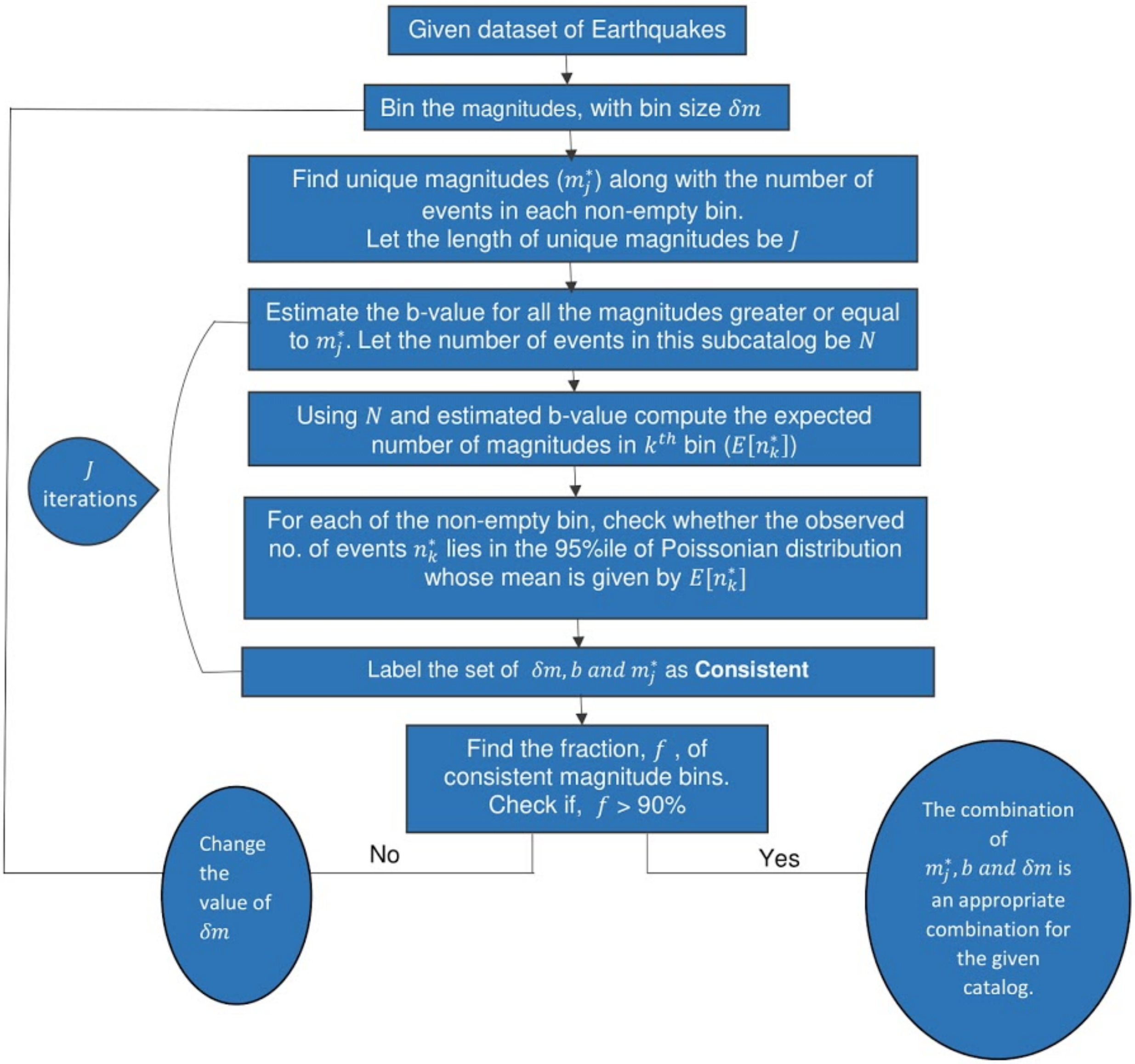

Fig 2.2 Algorithm- Flowchart of the testing the EMLE test

The steps involved in this approach to jointly estimate M_{c}, b and δm are explained in a simpler way with the help of a flowchart (Fig 2.2)

2.5 Generating Synthetic Data

For testing the robustness of the proposed algorithm, we have generated 2 different sets of synthetic data, described in the following subsections.

2.5.1 GR Law Synthetics

A unimodal distribution function, that follows GR law frequency relation for M_c and all the magnitudes above it. The goal of testing the method with GR law synthetics was performing a controlled experiment since this is the most trivial way of generating the desired dataset. The recipe that goes into simulating the GR law synthetics is as follows-

The cumulative distribution function of GR law(eq 1.1) is given by-

$$F(m) = 1 - exp(-\beta(m - m_c)) \quad 2.19$$

$F(m)$ in eq 2.19 are generated using random uniform function $U(0, 1)$. Then the eq. 2.19 is rewritten as

$$U(0, 1) = 1 - exp(-\beta(m - m_c)) \quad 2.20$$

The equation 2.21 is the inverse function of Eq. 2.20 to generate synthetic data given by -

$$m = m_c - (log(1 - U(0, 1)))/\beta \quad 2.21$$

The histogram plot the synthetics hence generated using eq 2.20 is shown in fig 2.1 for $M_c = 3$ *and* $b - val = 1$

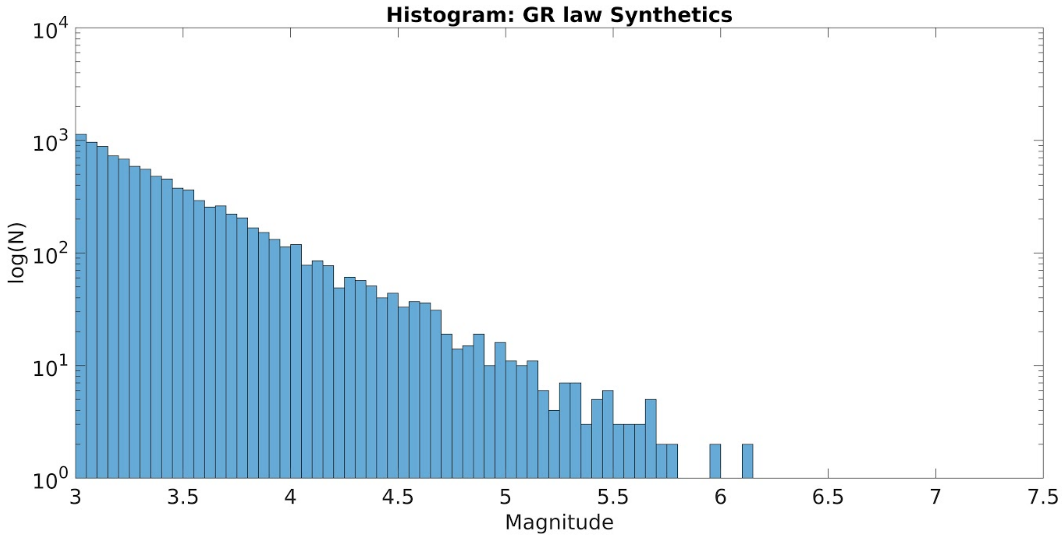

Fig 2.2 Histogram plot of GR Law Synthetic for $M_c = 3$ *and* $b - val = 1$

2.5.2 Normal and GR Synthetics (NGR Synthetics)

Unlike GR law model, this is a bimodal distribution function of magnitudes where M_c and all the magnitudes above it follow GR law and the magnitudes below M_c obeys the Normal distribution function. The pdf of the synthetic data hence simulated is given by:

$$f(m|\sigma, \beta) = \left\{ 1/\sqrt{2\pi\sigma^2}\, exp(-(m - \mu)2)/2\sigma^2,\ \ for\ m < M_c \right\} \qquad 2.21$$

$$= \{exp(-\beta m),\ \ for\ m >= M_c\}$$

This is a parametric way of generating data below M_c. The comparison of the real and synthetic data computed using Eq. 2.21 is being shown in fig 2.3

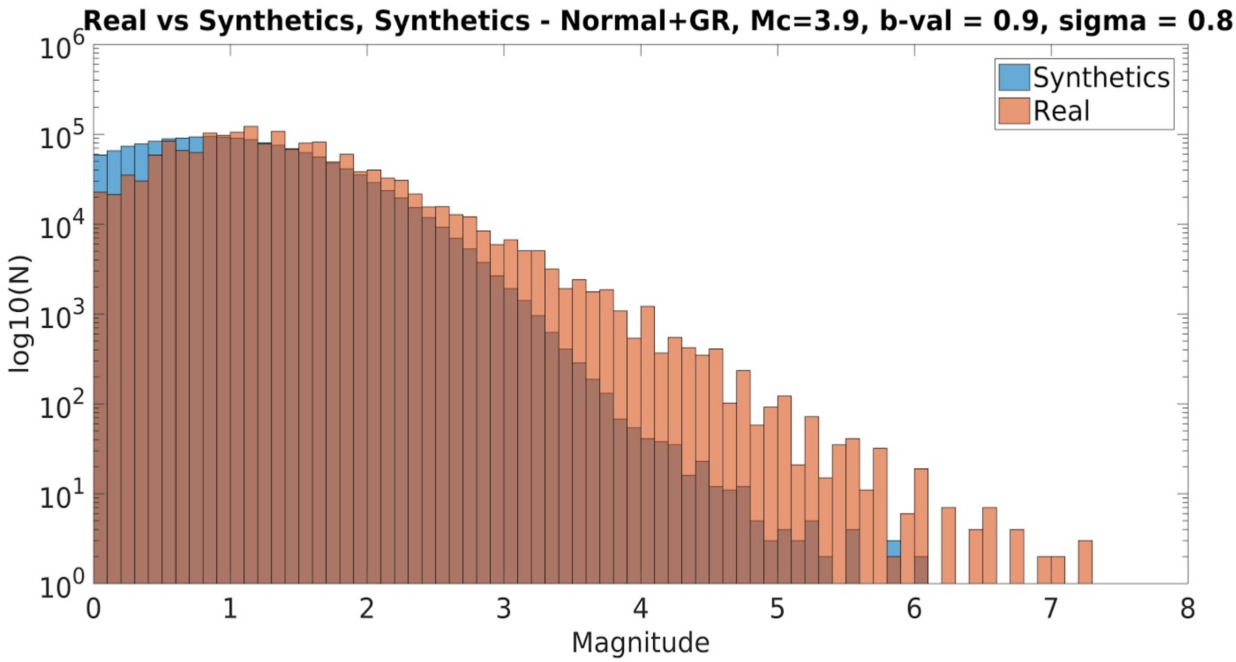

Fig 2.3 Histogram plot of Real vs Synthetics data: NGR, $M_c = 3.9,\ b-value = 0.9,\ sigma = 0.8$

The relation between the parameters involved along with the constraints used to generate this synthetic data are given by equations 2.22 and 2.23

In order to achieve the continuity of the FMD at M_c, we constrained the FMD for magnitudes below M_c by introducing a normalizing constant given by,

$$N_c = f(m_n)/f(m_{n+1}) \text{ where } n+1 \text{ is the index of } M_c \qquad 2.22$$

The parameter of the normal distribution μ is evaluated using

$$\mu = m_c - \sigma * \sqrt{-2log(\sqrt{2\pi} * \sigma * exp(-\beta m_c))} \qquad 2.23$$

The equations 2.13 and 2.14 are derived using the continuity condition of the equation 2.12 which also defines the relationship between all the three parameters of the complete pdf.

CHAPTER 3

Results

The results obtained by implementing the proposed methods along with state of the art on various synthetic datasets are presented in this chapter. The plots that we will be illustrating hereafter are error maps. The error maps provide an inference of how do the errors(absolute difference between the presumed M_c and the estimated M_c) in the estimation of M_c vary with respect to presumed M_c and b-value hinting towards how robust are the proposed methods, how well they perform on varying synthetics as compared to the state of the art method.

3.1 Synthetic Tests

This section will demonstrate a variety of error maps obtained by experimenting our proposed methods as well as the state of the art (discussed in section 1.3) on different synthetics(described in section 2.4) i.e. GR and NGR synthetics.

Figure 3.1 is the output of the error maps of MPE estimator tested on the parametric synthetics i.e. GR and NGR. It can be observed in the case of Fig 3.1(a), the map indicates that the MPE method estimates M_c with zero error for 80% of the pairs of presumed b-value and M_c.

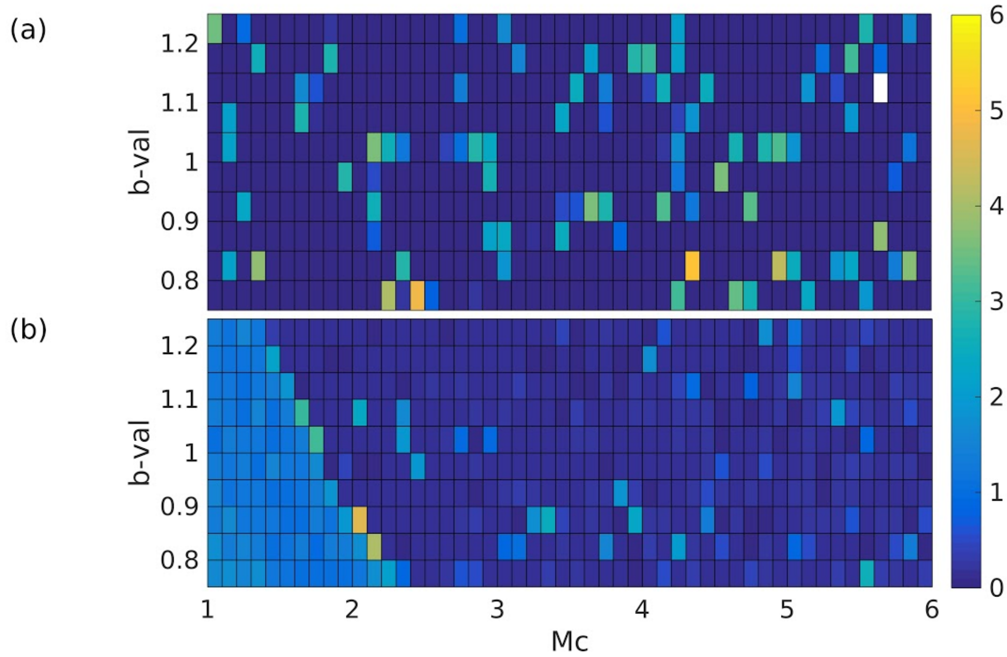

Fig 3.1. Error Maps of a) GR, b) NGR ($\sigma = 20.0$ *) and tested on MPE method to estimate* M_c

Fig 3.1(b) is an error map of NGR synthetics examined by using the MPE method illustrates for the presumed value of M_c in the range of $1\ to\ 2$ the estimate converges with an error of ± 2 respect to presumed $b-value$ ranging from $0.75\ to\ 1.25$. For presumed $M_c > 2$ the errors in estimation converge to zero.

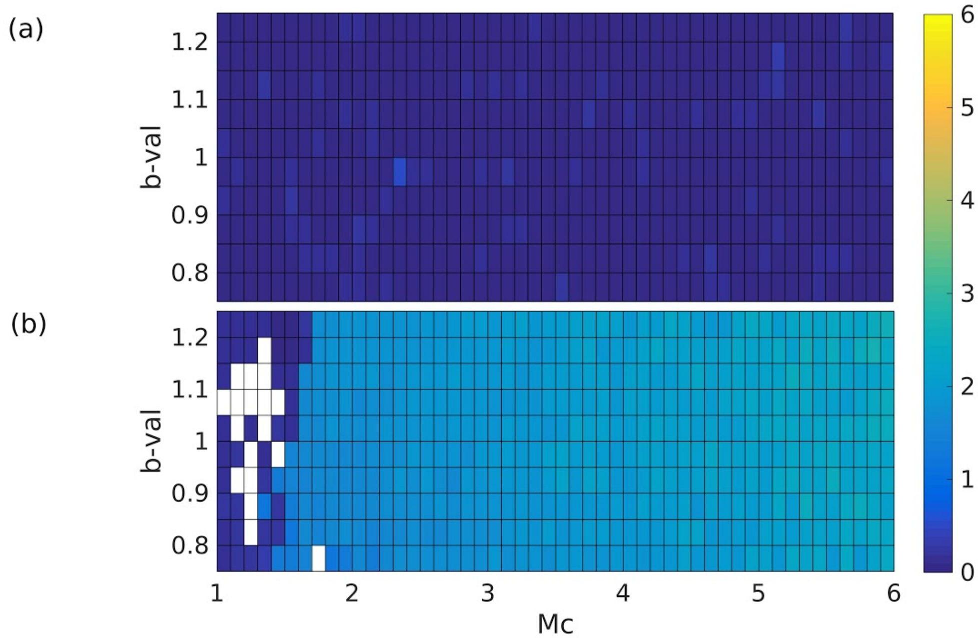

Fig 3.2. Error Maps of a) GR, b) NGR ($\sigma = 20.0$) and tested on KS method to estimate M_c

The plots in Fig 3.2 demonstrates the performance of the KS method when implemented on GR and NGR synthetics. It can be witnessed from the maps in the case of GR synthetics(Fig 3.2(a)) KS method have an accuracy of 100% in the estimation of M_c. Whereas in the case of NGR (Fig 3.2(b)), the estimations converge with an offset of ± 2 for 95% of the grid.

Fig 3.3(a) represents the error map of the MAXC method generated for the GR synthetics, the results are estimated with 100% accuracy for all the pairs of M_c *and* $b - value$. It can be observed from Fig 3.3(b) the errors in the estimation of M_c on NGR synthetics for the pairs of presumed M_c *and* $b - value$, the errors gradually rise in the order of $\pm (0 - 6)$.

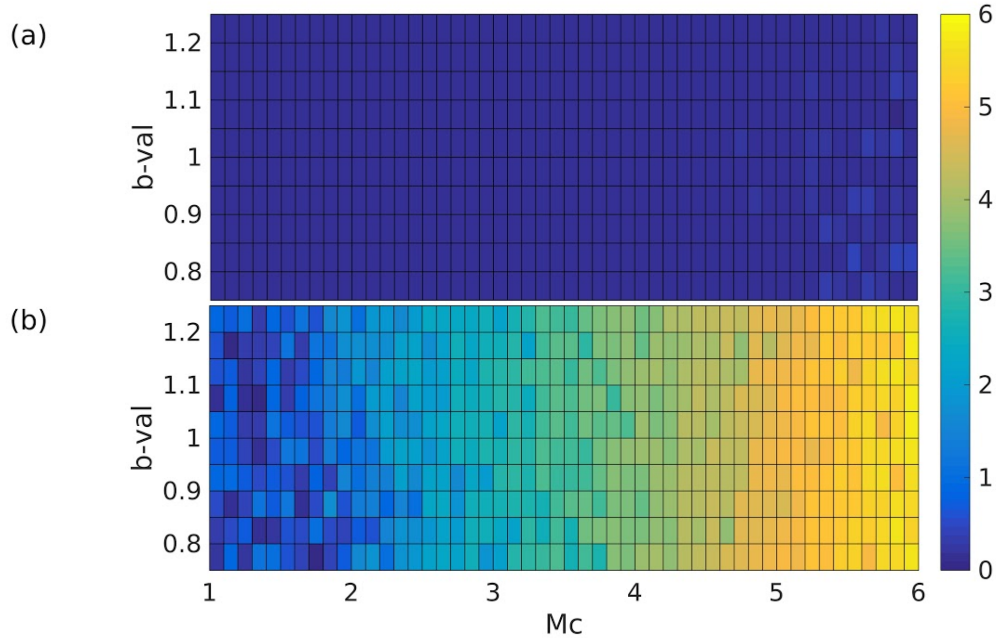

Fig 3.3. Error Maps of a) GR, b) NGR ($\sigma = 20.0$ *) and tested on MAXC method to estimate* M_c

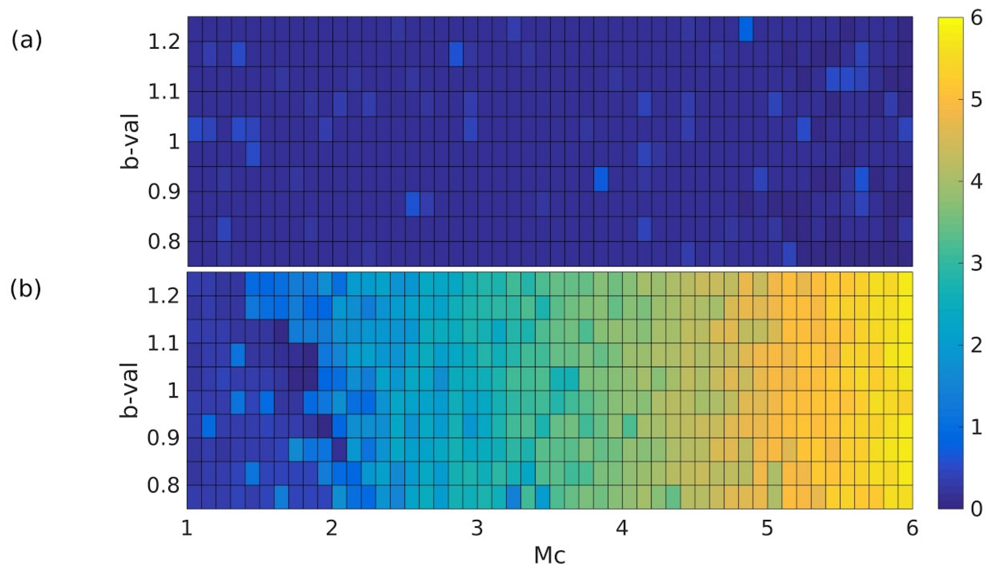

Fig 3.4. Error Maps of a) GR, b) NGR ($\sigma = 20.0$ *) and tested on MBS method to estimate* M_c

Fig 3.4 is an illustration of the MBS method examined on the synthetics. It can be observed from the figures 3.4(a) that the method estimates the values of M_c with error bounds of zero over the entire grid. In the case of NGR synthetics, the MBS method has error bounds lying in the range of $\pm(0-6)$ spread across the complete grid of Mc *and* $b-values$ as observed in Fig 3.4(b).

The error maps in Fig 3.5 are obtained as a result of the GFT method implemented using GR and NGR synthetics. Fig 3.5(a) represents the output of error in estimation using GR synthetic data set where the distribution of errors is zero for the complete grid of presumed M_c *and* $b-value$. In the case of NGR synthetics, the errors in estimation gradually increase with the increase in the presumed M_c.

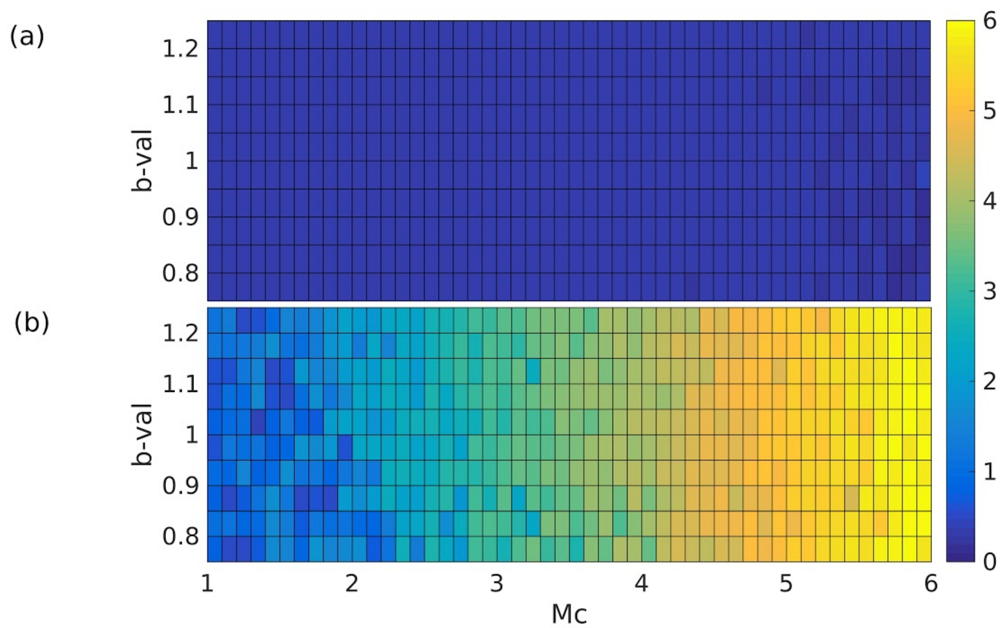

Fig 3.5. Error Maps of a) GR, b) NGR ($\sigma = 20.0$ *) and tested on GFT method to estimate* M_c

3.2 Comparative Analysis

The figures represented in this section are designed with an aim to compare the proposed methods with the state of the art over various synthetics(section 2.3) and will be described in more details in chapter 4.

Fig 3.6 represents error maps of the methods MPE, KS, MAXC, MBS, and GFT respectively when implemented using NGR synthetics. It can be observed from the plots that the error in estimation decreases with the increase in the presumed M_c in the case of MPE and KS methods(Fig 3.6(a), (b)). Whereas in the case of state of the art methods i.e MAXC, MBS and GFT the errors in estimation gradually increases with the increase in the presumed M_c shown in figures 3.6 a), b) and c) respectively.

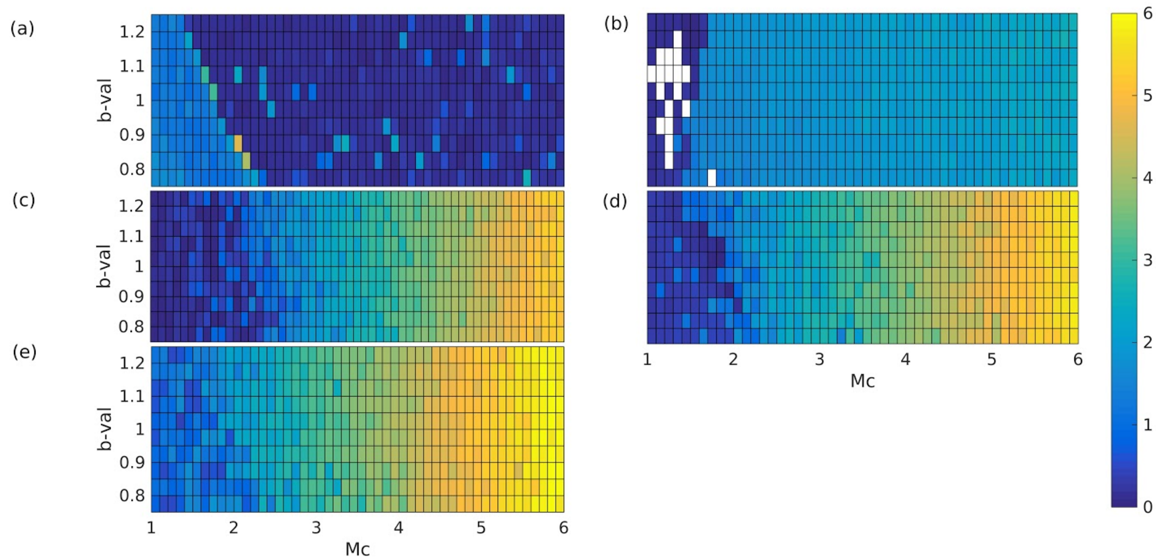

Fig 3.6. Error Maps of NGR Synthetics($\sigma = 20.0$) tested on a) MPE, b) KS, c) MAXC, d) MBS and e) GFT methods to estimate M_c

Implementing the proposed methods along with the state of the art methods using binned NGR synthetic illustrates the effect of binning the data on the estimation of M_c. Fig 3.7 shows how does the binned dataset affect the estimations of M_c when tested on MPE and KS methods. It can be observed that the errors in estimation decrease with the increase in the bin size in the case of KS method. The estimations using the MPE method are not affected much and produces almost the same results of estimation.

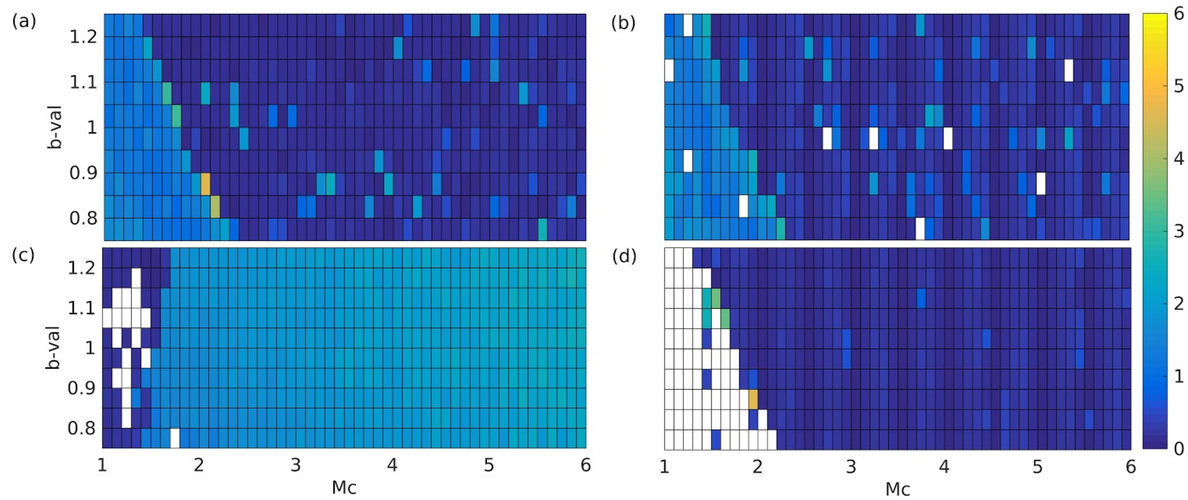

Fig 3.7. Error Maps of NGR Synthetics($\sigma = 20.0$ *) tested on a) MPE bin size = 0.1, b) MPE bin size = 0.5, c) KS, bin size = 0.1, d) KS, bin size = 0.5 to test the impact of binning on the estimation of* M_c

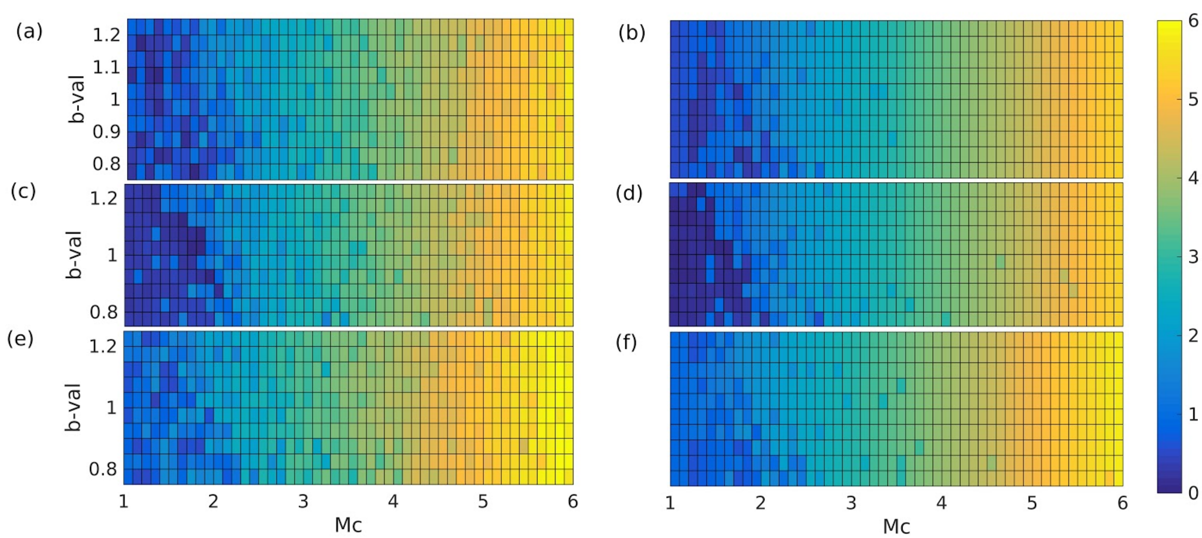

Fig 3.8. Error Maps of NGR Synthetics($\sigma = 20.0$ *) tested on a) MAXC bin size = 0.1, b) MAXC bin size = 0.5, c) MBS, bin size = 0.1, d) MBS, bin size = 0.5 , e) GFT bin size = 0.1, f) GFT bin size = 0.5 to test the impact of binning on the estimation of* M_c

It can be observed from Fig 3.8 increasing the bin size from $0.1\ to\ 0.5$ does not affect the estimations in M_c in the case of MAXC, MBS and GFT.

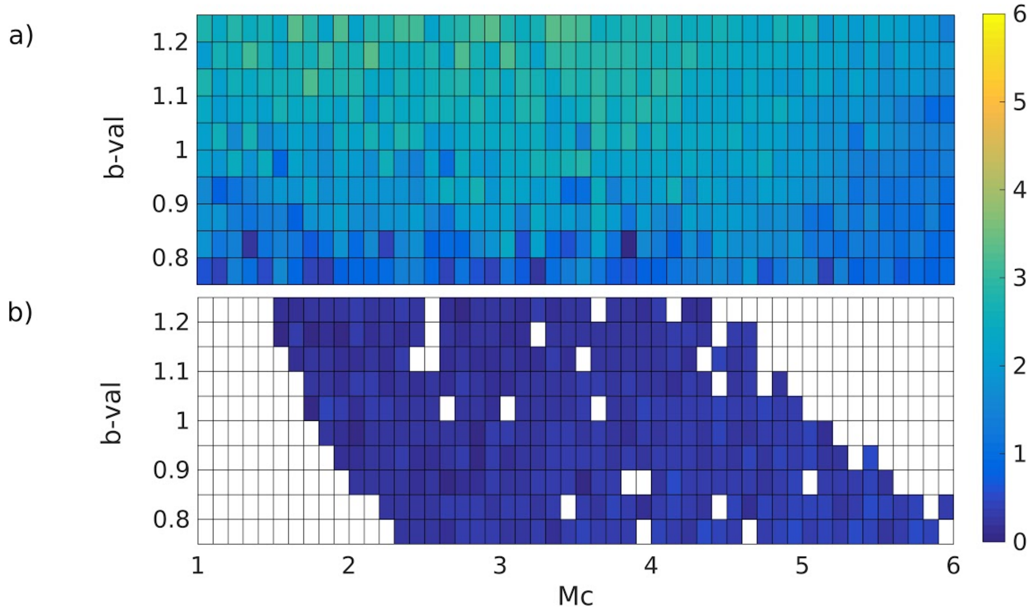

Fig 3.9 KS method tested on Synthetics NGR a) With binning and $\sigma = 0.8$*, b) Without binning and* $\sigma = 20.0$

Testing the KS method on parametric synthetic i.e. NGR synthetic. Fig 3.9(a) is a result of binning the synthetic data, it can be observed from the figure that the error distribution is of order $\pm(0-2)$ and is spread across the grid of $M_c\ and\ b-value$. When the synthetic data is continuous i.e. not binned Fig 3.6(b) is generated which produces errors in the order of $\pm(0-1)$.

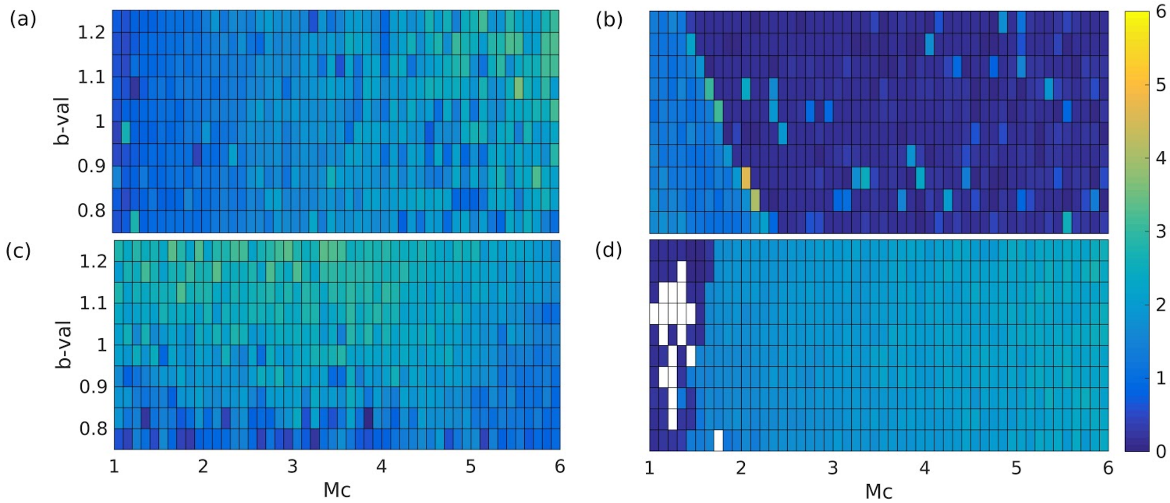

Fig 3.10. Error Maps of NGR Synthetics(bin size = 0.1) tested on a) MPE $\sigma = 0.8$, b) MPE $\sigma = 20.0$, c) KS, $\sigma = 0.8$, d) KS, $\sigma = 20.0$ to test the impact of σ on the estimation of M_c

The error maps in Fig 3.9 illustrates the effect of changing the value of σ *from* 0.8 *to* 20.0 while generating the NGR synthetics on the estimations of M_c. In the case of MPE method, it can be observed that from Fig 3.9(a) and (b) that due to the increase in the value of σ the errors in estimation reduces from ± 2 and converges to zero for presumed M_c varying from 2 *to* 6. In the case of KS method, the changing value of σ not has much effect on the errors in estimation except for some smaller values of M_c lying in the range of 1 *to* 1.5.

Fig 3.9 shows the impact of the increasing value of σ on errors in the estimation of M_c by state of the art methods. It can be observed that with the increasing value of σ the error are more pronounced in the case of MAXC, MBS and GFT method, shown in fig 3.9 b), d) and f) respectively.

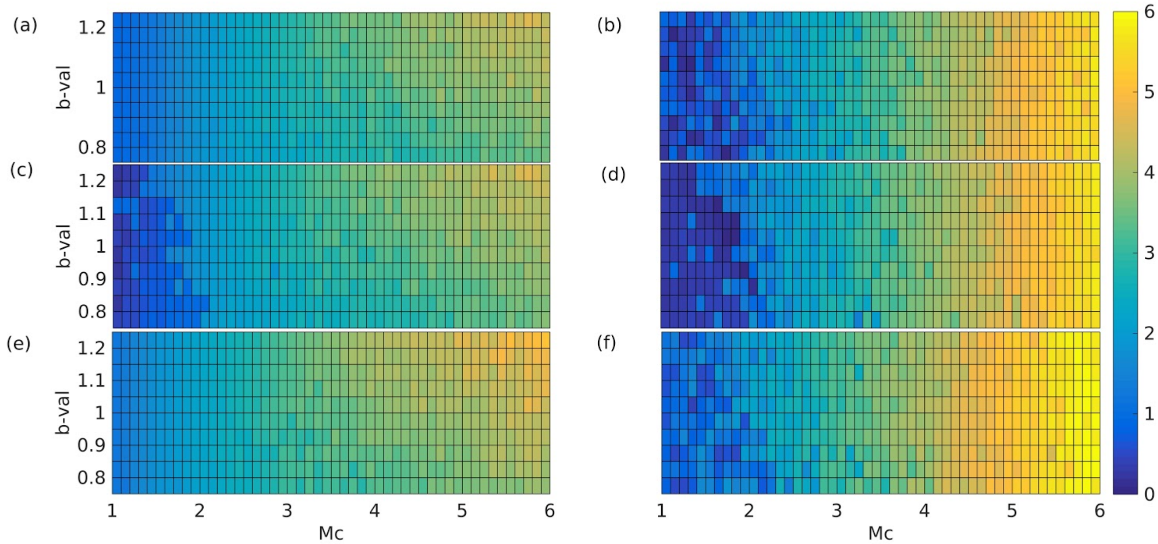

Fig 3.11. Error Maps of NGR Synthetic(bin size = 0.1) tested on a) MAXC $\sigma = 0.8$, *b) MAXC* $\sigma = 20.0$, *c) MBS,* $\sigma = 0.8$, *d) MBS,* $\sigma = 20.0$, *e) GFT* $\sigma = 0.8$, *f) GFT,* $\sigma = 20.0$ *to test the impact of* σ *on the estimation of* M_c

CHAPTER 4

Discussion

Highlighting the key features of the various methods implemented to evaluate M_c, this chapter will focus on deciphering the results in more details. The results have been characterised in such a way so as to discuss the pros and cons of different methods when compared over NGR synthetics, increasing in bin size, and increasing the value of σ. Synthetics analysis have been carried out to devise a comparative analysis. Starting with the MPE method tested on all the parametric synthetics i.e. GR and NGR, referring to the fig 3.1, it can be observed that both the synthetics perform well. The GR synthetics was just a controlled experiment to test how well the proposed methods operate. Looking at fig 3.2 the results for the KS method, it estimates the value of M_c with an error of ± 2 for more than 90% of the pairs of presumed M_c *and* $b-value$. Comparing these two proposed methods with the state of the art methods-

1) Over the NGR synthetics (fig 3.6) - it can be observed that our proposed methods are less prone to error over the entire grid of presumed M_c *and* $b-val$, unlike the other three methods where the errors increase gradually over the grid.

2) Over increasing size of binning (fig 3.7, 3.8) - Even though the MPE method does not have any effect of increasing size on the estimations of M_c, the KS method have shown improvement in the estimations of M_c by errors converging to zero for 90% of the pairs of presumed M_c *and* $b-value$ in the grid. Moreover, the KS results with binned and without binned synthetics (Fig 3.9) i.e discrete and continuous datasets produces errors converging to zero in the case of without binned case. Whereas, in the case of state of the art methods there is no effect of the increase in the bin size on the estimations of M_c.

3) Over increasing value of σ - MPE method have shown improvements in the estimations of M_c by producing errors that converge to zero as the value of σ increases

from $0.8\ to\ 20.0$. While the other method has no pronounced effect in the estimations of M_c due to the increase in the value of σ.

4.1 Limitations and Future prospects

Firstly, on implementing the LL-test described in section 2.4 we realised that the true log-likelihoods computed using Eq 2.15 for all the pairs of assumed $M_c\ and\ estimated\ b-value$ lies in the LL_{eff} 90% confidence interval. This implies the method does not converge to one value M_c and hence we cannot use this method for estimation of M_c. Secondly, catalog-based methods do not show the temporal and spatial aspect of the catalog for the estimation of $M_c\ and\ b-value$. Thirdly, the non-availability of the errors associated with the recorded magnitudes of the catalog leads to the arbitrary assumption of bin size. Varying the value of bin size is acceptable but not a very good way of introducing uncertainty in the magnitude of earthquake catalog. To overcome the drawbacks of the current methods we are planning to improve all the proposed methods by testing them on revising the real catalog magnitudes with the errors associated with them. The advantage of using these revised catalog magnitudes would be to estimate the value of M_c along with the uncertainties in estimation. Our next goal would be to come up with more realistic synthetic catalogs one of the ways by which this can be achieved is looking at the area under the curve of NGR distribution this would give us the proportion of the data above and below $assumed\ M_c$. We would also be implementing the EMLE method described in section 2.5. EMLE method incorporates the uncertainty in the magnitudes of the real/synthetic catalog by discretization of data i.e. binning the data and furthermore, jointly estimates the $bin\ size,\ M_c\ and\ b-value$ which would lead to the completeness of the catalog up to M_c.

CHAPTER 5

Conclusion

As nature doesn't provide us with the clear segregation boundary line at a particular magnitude to divide the set of magnitudes of the catalog into two subsets of complete and incomplete datasets the goal of the study was to achieve an adequate earthquake catalog suitable enough for several seismic hazard-related studies by estimating the magnitude of completeness, M_c. The proposed i.e using maximum likelihood estimator evaluating M_c was implemented by carrying out synthetic catalog tests. From the results obtained it can be concluded that the method performs convincingly when tested by using all the synthetics discussed in the previous chapter. However, the method requires some advancements to overcome with the drawbacks of the current method such as estimation of M_c for smaller b-values, reducing the time complexity of the codes, the modifications such that it produces appreciable results with any kind of synthetics data and hence be a trustworthy estimator of M_c for any real catalog. This can be achieved by investigating both the spatial and temporal variations of the catalog simultaneously. Furthermore, we'll be testing the method on a realistic and complex synthetic catalog which would incorporate the spatial and temporal features and help us in the cope up with the understanding of how does M_c varies with respect to time. And investigate how the sensitivity of M_c varies with the sampling size of the dataset.

References

1. Cao, A. M., and S. S. Gao (2002), Temporal variations of seismic b-values beneath northeastern Japan island arc, Geophysics. Res. Lett., 29, doi:10.1029/2001GL013775

2. Cornell, C. A. (1968), Engineering seismic risk analysis., Bull. Seismol. Soc. Am., 58 (5), 1583 1606.

3. 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. Spada, Matteo & Tormann, Thessa & Wiemer, Stefan & Enescu, Bogdan. (2012). Generic dependence of the frequency-size distribution of earthquakes on depth and its relation to the strength profile of the crust. Geophysical Research Letters. 40. 10.1029/2012GL054198.

5. Habermann, R. E. (1987). Man-made changes in seismicity rates. Bull. Seism. Soc. Am. 77, 141–159.

6. Habermann, R. E. (1991). Seismicity rate variations and systematic changes in magnitudes in teleseismic catalogs, Tectonophysics 193, 277–289

7. Habermann, R. E., and F. Creamer (1994). Catalog errors and the M8 earthquake prediction algorithm, Bull. Seism. Soc. Am. 84, 1551–1559.

8. Jochen Woessner and Stefan Wiemer(2005) Assessing the Quality of Earth See you around 14:30 then.quake Catalogues: Estimating the Magnitude of Completeness and Its Uncertainty, Bull. Seism. Soc. Am. 95, 684–698

9. Leptokaropoulos, K. M., V. G. Karakostas, E. E. Papadimitriou, A. K. Adamaki, O. Tan, and S. İnan (2013), A homogeneous earthquake catalogue compilation for western turkey and magnitude of completeness determination, Bull. Seismol. Soc. Am., 103, 5, 2739-2751.

10. 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

11. 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

12. Rydelek, P. A., and I. S. Sacks (1989). Testing the completeness of earthquake catalogs and the hypothesis of self-similarity, Nature 337, 251–253.

13. Schorlemmer, D., and J. Woessner (2008), Probability of detecting an earthquake, Bull. Seismol. Soc. Am., 98, doi:10.1785/0120070105.

14. Schorlemmer, D., Wiemer, S., Wyss, M., 2004. Earthquake statistics at Parkfield: 1. Stationarity of b-values. Journal of Geophysical Research 109, B12307. doi:10.1029/2004JB003234.

15. Taylor, D. W. A., J. A. Snoke, I. S. Sacks, and T. Takanami (1990). Nonlinear frequency-magnitude relationship for the Hokkaido corner, Japan, Bull. Seism. Soc. Am. 80, 340–353.

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

17. Wiemer, S., and M. Wyss (2002). Mapping spatial variability of the frequency-magnitude distribution of earthquakes, Adv. Geophys. 45, 259–302.

18. 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.

19. Wiemer, S., McNutt, S., 1997. Variations in frequency–magnitude distribution with depth in two volcanic areas: Mount St. Helens, Washington, and Mt. Spurr, Alaska. Geophysical Research Letters 24, 189–192

20. Wiemer, S., and D. Schorlemmer (2007), Alm: An asperity-based likelihood model for california, Seismol. Res. Lett.,78, 134–140.

21. Zuniga, F. R., and S. Wiemer (1999). Seismicity patterns: are they always related to natural causes? Pure Appl. Geophys. 155, 713–726

22. Lay, T., Wallace, T.C., 1995. Modern Global Seismology. Academic Press, p. 393.