Verification and Sensitivity Analysis of Maximum Likelihood Estimation for Loviisa NPP Seismic Hazard

Laura Heikkilä

School of Science

Special Assignment Report
Espoo 8.9.2023

Supervisor

Jarmo Ala-Heikkilä

Instructors

Jukka Koskenranta

Timo Leppänen

Aalto University
School of Science

Aalto University
School of Science

Aalto University, P.O. BOX 11000, 00076 AALTO
www.aalto.fi
Abstract of the special assignment

Author Laura Heikkilä

Title Verification and Sensitivity Analysis of Maximum Likelihood Estimation for Loviisa NPP Seismic Hazard

Degree programme Engineering Physics and Mathematics

Major Engineering physics **Code of major** SCI3056

Supervisor Jarmo Ala-Heikkilä

Instructors Jukka Koskenranta, Timo Leppänen

Date 8.9.2023 **Number of pages** iv+26+VIII **Language** English

Abstract

Probabilistic seismic hazard analysis (PSHA) is used for estimating the risk of earthquakes to nuclear power plants. To achieve a required level of safety, the recurrence of earthquakes of different magnitudes has to be known. This is done by estimating the parameters a and b in the Gutenberg-Richter equation. The purpose of this work is to verify the use of maximum likelihood estimation (MLE) method in the estimation of earthquake recurrence in Finland and perform a sensitivity analysis for this method. The verification and the sensitivity analysis are performed by comparing the estimated values of parameters a and b. The least squares (LS) method is included for comparison.

The verification gives results similar to those from previous studies for both of the used methods. For the sensitivity analysis, changing completeness times, minimum and maximum magnitude and the width of the magnitude bins is tested. Both of the methods seem to be the most sensitive to changing the minimum magnitude, but the lack of high-magnitude events and the incompleteness of the data from the smallest earthquakes increase the uncertainty of the estimation. At the end of this paper, parameters giving the most reliable results are suggested.

Keywords seismicity, PSHA, Gutenberg-Richter, MLE, LS, SSA, seismic activity rates, nuclear safety

Contents

Symbols and abbreviations

Symbols

$v(c)$	Rate of exceedance
$P(C > c)$	Probability of exceedance
$n(m)$	Number of earthquakes of magnitude m or greater per unit of time
n_i	Number of earthquakes in magnitude bin i
a, α	Intercept of the Gutenberg-Richter equation, $\alpha = \ln(10)a$
b, β	Slope of the Gutenberg-Richter equation, $\beta = \ln(10)b$
$n_{\min}$	Recurrence at minimum magnitude

Abbreviations

GMPE	Ground motion prediction equation
GR	Gutenberg-Richter
LS	Least squares
MLE	Maximum likelihood estimation
NPP	Nuclear power plant
PSHA	Probabilistic seismic hazard analysis
SSA	Seismic source area
STUK	Radiation and Nuclear Safety Authority

1 Introduction

A high level of safety is a priority during the whole lifetime of a nuclear power plant (NPP) and in all steps of nuclear energy production. To ensure the safety and security of nuclear energy, both national and international laws, regulations and standards guide the design and usage of the NPPs. This includes being prepared for different accidents and disasters that could possibly harm the NPPs. One example of a natural disaster causing a nuclear accident is from Fukushima in 2011, when an earthquake followed by a tsunami caused the failure of the electricity supply and the emergency generators in the Fukushima Daiichi NPPs. This is an extreme example of how the seismic activity of the Earth can cause major risks to the operation and safety of NPPs.

Most of the largest earthquakes around the world happen close to the edges of the tectonic plates. Geologically, Finland is located on the Fennoscandian Shield, which is an interplate area far from the edges of the Eurasian plate it belongs to. This is why most of the earthquakes in this area are small, and larger earthquakes are very rare. Thus, the seismic activity usually does not cause major risks to most parts of the infrastructure. However, in NPPs, even small fractures or other damages can in some cases be very harmful, if critical parts or safety systems of the NPP are damaged. To ensure that appropriate safety systems can be designed and built, the seismic risk has to be known. This is why using the right methods to estimate the seismic risk is important.

When the NPPs in Loviisa were built, there were no requirements for seismic hazard analysis or taking seismic hazard into consideration in the plant design [1]. However, before it was required, the first probabilistic seismic hazard analysis (PSHA) in Finland was started at the end of the 1980's for the Loviisa NPPs. The PSHA was completed and submitted to Radiation and Nuclear Safety Authority (STUK) a few years later, in 1992. As the PSHA was not yet a requirement at that time, STUK did not give a detailed review of the PSHA until much later. In their review, STUK set some requirements about how the PSHA should be updated. After that, the Loviisa PSHA has been updated based on requirements and suggestions from STUK, revised methods and new seismic data.

The latest Loviisa PSHA was conducted by an American consulting company Slate [2] in 2021, which was then submitted to STUK in the same year. STUK has reviewed the document and requested more justification for the parameters chosen for the final determination of the risk. A sensitivity analysis of these parameters and the method used was performed in a Master's thesis "Sensitivity analysis of a seismic hazard assessment for a Finnish nuclear power plant" [3]. According to the results from the thesis, using maximum likelihood estimation (MLE) instead of least squares (LS) method used in the original PSHA gives different results. The goal of this special assignment is to further analyse the MLE method by verifying it, comparing it to alternative methods and performing a sensitivity analysis.

2 Theory

2.1 Probabilistic seismic hazard analysis

The purpose of a PSHA is to estimate the seismic activity of a specific site so that a numerical value to the risk can be given. Typical hazard products are hazard curves and uniform hazard response spectra, which describe ground accelerations at the site investigated. The first PSHA method was presented by Cornell in 1968, and it has been widely used ever since, still working as a basis for the PSHAs carried out today. PSHA methods have also been a subject of criticism, and highly destructive earthquakes have sometimes occurred in areas predicted to have low seismicity using PSHA [4].

The seismic risk calculated with the PSHA is often given as a rate of exceedance, $v(c)$. The rate of exceedance tells the frequency with which an earthquake characteristics C exceeds a given value c. By assuming that the number of earthquakes is Poisson distributed, the probability of exceedance in a specific time period can be calculated. The probability of an earthquake event with C exceeding c within time t is given by formula

$$P(C > c) = 1 - e^{-v(c)t}. \tag{1}$$

The STUK regulatory guide on nuclear safety B.7 (YVL B.7) states that the NPPs in Finland must be able to tolerate earthquakes which have a frequency of 10^{-5} 1/a or greater [5]. It is also stated in the YVL B.7 that the minimum value for peak ground acceleration is 0.1 g, which generally sets the limit for earthquake tolerance in Finland due to the low seismicity.

2.2 Gutenberg-Richter equation

Most applications of PSHA use an exponential probability distribution to represent the amount of earthquakes of different magnitudes [6]. Probably the most common equation to describe this distribution is Gutenberg-Richter (GR) equation, which shows the relation of magnitude and the number of earthquake events. The general form of the GR equation is

$$\log_{10}(n(m)) = a - bm, \tag{2}$$

where n is the number of earthquakes in time interval t with a magnitude of m or greater and a and b are parameters. For PSHA, the GR equation is often given in an exponential form

$$n(m) = e^{\alpha - \beta m}, \tag{3}$$

where $\alpha = \ln(10)\,a$ and $\beta = \ln(10)\,b$. Existing earthquake data can then be used to solve parameters α and β, which can be used to estimate the number of earthquake occurrences in the future and thereby the seismic hazard. Most commonly used methods to estimate the parameters are maximum likelihood estimation and least squares method.

2.3 Maximum likelihood estimation

2.3.1 Truncation

The exponential form of the GR equation (3) is often truncated to exclude both the smallest and the greatest magnitudes. As $n(m)$ gives the number of earthquakes with a magnitude of m or greater, leaving out events with the lowest magnitudes does not affect its value, as long as m is greater than than the minimum magnitude $m_{\min}$. However, $m_{\min}$ is often included in the definition of $n(m)$:

$$n(m) = n_{\min} e^{-\beta(m - m_{\min})}, \tag{4}$$

where $n_{\min} = n(m_{\min})$. Generally, the largest magnitudes are cut off as well, as it is often assumed that seismic source areas cannot produce earthquakes of a magnitude greater than a certain maximum magnitude $m_{\max}$ [6]. The maximum magnitude can be added to the formula of $n(m)$ by subtracting the number of earthquakes with $m > m_{\max}$. In this case, the difference has to be divided by $1 - e^{-\beta(m_{\max} - m_{\min})}$ to normalize the value so that $n(m_{\min}) = n_{\min}$. This results in the following formula for $n(m)$:

$$n(m) = e^{\alpha} \frac{e^{-\beta m} - e^{-\beta m_{\max}}}{1 - e^{-\beta(m_{\max} - m_{\min})}}. \tag{5}$$

2.3.2 Derivation

To use the MLE method, the probability function of having a specific number of events in a time interval t has to be formed. We can assume that the earthquakes follow a Poisson distribution, which requires an assumption of independently occurring events. The probability function of events following the Poisson distribution is

$$P(n \text{ events in interval } t) = \frac{(rt)^n e^{-rt}}{n!}, \tag{6}$$

where r is the average rate at which the events occur. As the rate depends on the magnitude, the rate has to be calculated for each magnitude range separately. This can be done by calculating the number of events between magnitudes $m_i - \frac{1}{2}\delta m$ and $m_i + \frac{1}{2}\delta m$, where $i \in [1, I]$ is the index of the magnitude interval and δm_i is the width of that interval. Thus, the rate parameter is of the form

$$r_i = \delta n_i = n\left(m_i - \frac{1}{2}\delta m_i\right) - n\left(m_i + \frac{1}{2}\delta m_i\right) \tag{7}$$
$$= \frac{e^{\alpha - \beta m_i}}{1 - e^{-\beta(m_{\max} - m_{\min})}}\left(e^{\frac{1}{2}\beta\delta m_i} - e^{-\frac{1}{2}\beta\delta m_i}\right) \tag{8}$$
$$= 2\sinh\left(\frac{1}{2}\beta\delta m_i\right)\frac{e^{\alpha - \beta m_i}}{1 - e^{-\beta(m_{\max} - m_{\min})}}. \tag{9}$$

Now, the probability of observing n_i events from magnitude range $[m_i - \frac{1}{2}\delta m, m_i + \frac{1}{2}\delta m]$ in time period t_i is

$$P(n_i) = \frac{(\delta n_i t_i)^{n_i}}{n_i!} e^{(-\delta n_i t_i)}. \tag{10}$$

It should be noted that in the equation above, n_i is the absolute value of earthquakes in magnitude bin i, not the cumulative value. The likelihood function can be formed by calculating the probability of observing n_i events for each magnitude interval $[1, I]$, which is the product of the separate probabilities:

$$\mathcal{L} = \prod_{i=1}^{I} P(n_i). \tag{11}$$

To make the differentiation easier, it is simpler to use the natural logarithm of the likelihood function as it has its maximum at the same values of α and β as the original likelihood but gets rid of the product:

$$\ln(\mathcal{L}) = \sum_{i=1}^{I} \ln\left(\frac{(\delta n_i t_i)^{n_i}}{n_i!} e^{-\delta n_i t_i}\right) \tag{12}$$
$$= \sum_{i=1}^{I} \left[n_i \ln(\delta n_i t_i) - \ln(n_i!) - \delta n_i t_i\right]. \tag{13}$$

The values of α and β maximizing the likelihood can now be found by calculating the derivatives of this function with respect to α and β and finding the roots of the derivatives. For α, the derivative is

$$\frac{\partial \ln(\mathcal{L})}{\partial \alpha} = \sum_{i=1}^{I} \frac{\partial}{\partial \alpha}\left[n_i(\ln(\delta n_i) + \ln(t_i)) - \ln(n_i!) - \delta n_i t_i\right] \tag{14}$$
$$= \sum_{i=1}^{I} \left[n_i - \delta n_i t_i\right] \tag{15}$$
$$= N - \sum_{i=1}^{I} \delta n_i t_i = 0, \tag{16}$$

where N is the total number of earthquakes in the interval $[m_{\min}, m_{\max}]$. By inserting δn_i from equation (9), we can reformulate equation (16):

$$N = \sum_{i=1}^{I} \left[2\sinh\left(\frac{1}{2}\beta\delta m_i\right) \frac{e^{\alpha-\beta m_i}}{1-e^{-\beta(m_{\max}-m_{\min})}} t_i \right] \tag{17}$$

$$\frac{e^{\alpha}}{1-e^{-\beta(m_{\max}-m_{\min})}} = \frac{N}{\sum_{i=1}^{I} 2e^{-\beta m_i}\sinh(\frac{1}{2}\beta\delta m_i) t_i} \tag{18}$$

$$\alpha = \ln\left(\frac{N(1-e^{-\beta(m_{\max}-m_{\min})})}{\sum_{i=1}^{I} 2e^{-\beta m_i}\sinh(\frac{1}{2}\beta\delta m_i) t_i} \right). \tag{19}$$

For β, the derivative is

$$\frac{\partial \ln(\mathcal{L})}{\partial \beta} = \sum_{i=1}^{I} \left[\frac{n_i}{\delta n_i t_i} \cdot t_i \frac{\partial}{\partial \beta}\delta n_i - t_i \frac{\partial}{\partial \beta}\delta n_i \right] \tag{20}$$

$$= \sum_{i=1}^{I} \left[\frac{\partial}{\partial \beta}\delta n_i \left(\frac{n_i}{\delta n_i} - t_i \right) \right] = 0, \tag{21}$$

where

$$\frac{\partial}{\partial \beta}\delta n_i = \frac{2e^{\alpha-\beta m_i}}{1-e^{-\beta(m_{\max}-m_{\min})}} \left(\frac{\delta m_i}{2}\cosh\left(\frac{\beta\delta m_i}{2}\right) - m_i \sinh\left(\frac{\beta\delta m_i}{2}\right) - \frac{(m_{\max}-m_{\min})\sinh(\frac{\beta\delta m_i}{2})e^{-\beta(m_{\max}-m_{\min})}}{1-e^{-\beta(m_{\max}-m_{\min})}} \right) \tag{22}$$

$$= \delta n_i \left(\frac{\delta m_i}{2}\coth\left(\frac{\beta\delta m_i}{2}\right) - m_i - \frac{(m_{\max}-m_{\min})e^{-\beta(m_{\max}-m_{\min})}}{1-e^{-\beta(m_{\max}-m_{\min})}} \right) \tag{23}$$

$$= \delta n_i \phi_i. \tag{24}$$

Rearranging equation (9) as

$$\frac{e^{\alpha}}{1-e^{-\beta(m_{\max}-m_{\min})}} = \frac{\delta n_i}{2\sinh(\frac{\beta\delta m_i}{2})e^{-\beta m_i}} \tag{25}$$

and inserting that into (18) gives

$$\delta n_i = \frac{N\sinh(\frac{\beta\delta m_i}{2})e^{-\beta m_i}}{\sum_{i=1}^{I}\sinh(\frac{\beta\delta m_i}{2})e^{-\beta m_i} t_i}. \tag{26}$$

The equation above can then be inserted into (21), yielding equation

$$\sum_{i=1}^{I}\left[\left(n_i - \frac{N\sinh(\frac{\beta\delta m_i}{2})e^{-\beta m_i}}{\sum_{j=1}^{I}\sinh(\frac{\beta\delta m_j}{2})e^{-\beta m_j}t_j}t_i\right)\phi_i\right] = 0, \tag{27}$$

which can now be used to solve β. However, this equation cannot be solved analytically, so numerical solution must be used [7].

According to [6], the last term of ϕ_i is close to zero when the difference of $m_{\max}$ and $m_{\min}$ is more than about 2 magnitude units, which in this work is always the case. This approximation in addition with an assumption of equally wide magnitude bins ($\delta m_i = \delta m$) gives us an updated formula for equation (27):

$$\sum_{i=1}^{I}\left[\left(n_i - \frac{Ne^{-\beta m_i}}{\sum_{j=1}^{I}e^{-\beta m_j}t_j}t_i\right)\left(\frac{\delta m}{2}\coth\left(\frac{\beta\delta m}{2}\right) - m_i\right)\right] \tag{28}$$

$$= -\sum_{i=1}^{I}n_i m_i + \frac{N}{\sum_{i=1}^{I}e^{-\beta m_i}t_i}\left(\sum_{i=1}^{I}e^{-\beta m_i}t_i m_i - \frac{\delta m}{2}\coth\left(\frac{\beta\delta m}{2}\right)\sum_{i=1}^{I}e^{-\beta m_i}t_i\right)$$

$$+\frac{\delta m}{2}\coth\left(\frac{\beta\delta m}{2}\right)\sum_{i=1}^{I}n_i \tag{29}$$

$$= -\sum_{i=1}^{I}n_i m_i + \frac{N\sum_{i=1}^{I}e^{-\beta m_i}t_i m_i}{\sum_{i=1}^{I}e^{-\beta m_i}t_i} = 0 \tag{30}$$

The equation on the last row can also be found from literature, e.g. [6, 7, 8]. This equation can now be used to solve β instead of equation (27). Similarly, in the formula of α (16), $e^{-\beta(m_{\max}-m_{\min})}$ is close to zero for a large enough difference of $m_{\min}$ and $m_{\max}$ and can thus be replaced with zero.

2.3.3 Uncertainty

To calculate the margins of error of α and β, a Fisher information matrix is first formed. This is done my forming the Hessian of the negative log-likelihood function. By taking an inverse of the Fisher information matrix, we get the covariance matrix [9]. Thus, the covariance matrix is

$$C = \mathcal{I}^{-1} = -\begin{pmatrix}\frac{\partial^2 \ln\mathcal{L}}{\partial\alpha^2} & \frac{\partial^2 \ln\mathcal{L}}{\partial\alpha\partial\beta}\\ \frac{\partial^2 \ln\mathcal{L}}{\partial\alpha\partial\beta} & \frac{\partial^2 \ln\mathcal{L}}{\partial\beta^2}\end{pmatrix}^{-1} = -\frac{1}{D}\begin{pmatrix}\frac{\partial^2 \ln\mathcal{L}}{\partial\beta^2} & -\frac{\partial^2 \ln\mathcal{L}}{\partial\alpha\partial\beta}\\ -\frac{\partial^2 \ln\mathcal{L}}{\partial\alpha\partial\beta} & \frac{\partial^2 \ln\mathcal{L}}{\partial\alpha^2}\end{pmatrix}, \tag{31}$$

where D is the determinant of the Hessian matrix, calculated as

$$D = \frac{\partial^2 \ln\mathcal{L}}{\partial\alpha^2}\cdot\frac{\partial^2 \ln\mathcal{L}}{\partial\beta^2} - \left(\frac{\partial^2 \ln\mathcal{L}}{\partial\alpha\partial\beta}\right)^2. \tag{32}$$

The second derivatives follow these formulas:

$$\frac{\partial^2 \ln \mathcal{L}}{\partial \alpha^2} = \frac{\partial}{\partial \alpha}(N - \sum_{i=1}^{I} \delta n_i t_i) = -\sum_{i=1}^{I} \delta n_i t_i = -N, \tag{33}$$

$$\begin{aligned}\frac{\partial^2 \ln \mathcal{L}}{\partial \beta^2} &= \frac{\partial}{\partial \beta} \sum_{i=1}^{I} \left[\left(\frac{\delta m_i}{2} \coth\left(\frac{\beta \delta m_i}{2}\right) - m_i \right) (n_i - \delta n_i t_i) \right] \\ &= \sum_{i=1}^{I} \left[(n_i - \delta n_i t_i) \left(\frac{\delta m_i}{2}\right)^2 \left(1 - \coth^2\left(\frac{\beta \delta m_i}{2}\right)\right) \right. \\ &\quad \left. - \delta n_i t_i \left(\frac{\delta m_1}{2} \coth\left(\frac{\beta \delta m_i}{2}\right) - m_i \right) \right], \end{aligned} \tag{34}$$

$$\frac{\partial^2 \ln \mathcal{L}}{\partial \alpha \partial \beta} = \frac{\partial}{\partial \beta}\left(N - \sum_{i=1}^{I} \delta n_i t_i\right) = \sum_{i=1}^{I} t_i \frac{\partial}{\partial \beta} \delta n_i = \sum_{i=1}^{I} \delta n_i \phi_i t_i. \tag{35}$$

In the previous equations, the same approximation of the last term of ϕ_i being close to zero is used again. Now, the variances and the covariance of a and b can be calculated from the elements of the covariance matrix as

$$\mathrm{var}(a) = \frac{1}{\ln(10)^2} \cdot \mathrm{var}(\alpha) = \frac{1}{\ln(10)^2} \cdot \left(-\frac{1}{D} \frac{\partial^2 \ln \mathcal{L}}{\partial \beta^2} \right), \tag{36}$$

$$\mathrm{var}(b) = \frac{1}{\ln(10)^2} \cdot \mathrm{var}(\beta) = \frac{1}{\ln(10)^2} \cdot \left(-\frac{1}{D} \frac{\partial^2 \ln \mathcal{L}}{\partial \alpha^2} \right), \tag{37}$$

$$\mathrm{cov}(a, b) = \frac{1}{\ln(10)^2} \cdot \mathrm{cov}(\alpha, \beta) = \frac{1}{\ln(10)^2} \cdot \frac{1}{D} \frac{\partial^2 \ln \mathcal{L}}{\partial \alpha \partial \beta}. \tag{38}$$

The margins of error can then be calculated as usual using the variance:

$$\hat{a} \pm 1.65 \sqrt{\mathrm{var}(a)}, \tag{39}$$

$$\hat{b} \pm 1.65 \sqrt{\mathrm{var}(b)}, \tag{40}$$

where the hat (ˆ) denotes the estimated value.

2.4 Least squares method

Another way to estimate the parameters a and b is with least squares method, which minimizes the sum of vertical distances from observation points to the predicted values. Now, the goal is to fit a linear function to the logarithm of $n(m)$ according to equation (2). The equations of the LS method are not proved here, but can be found from e.g. [10] and [11]. The parameter b can be calculated using equation

$$b = -\frac{\sum_{i=1}^{I} (m_i - \overline{m})(\log_{10}(n(m_i)) - \overline{\log_{10}(n)})}{\sum_{i=1}^{I} (m_i - \overline{m})^2}, \tag{41}$$

where the bar ($\bar{\ }$) denotes mean value. It should be noted that now i only goes through the bins having a nonzero value of n, as a logarithm of zero cannot be calculated. This means that the empty magnitude bins in the high-magnitude end cannot be taken into account. After solving b, a can be calculated with equation

$$a = \overline{\log_{10}(n)} + b\overline{m}. \tag{42}$$

The variances and covariance of parameters a and b can be calculated using the following formulas:

$$\mathrm{var}(a) = \frac{\sigma^2 \sum_{i=1}^{I} m_i^2}{I \sum_{i=1}^{I}(m_i - \overline{m})^2}, \tag{43}$$

$$\mathrm{var}(b) = \frac{\sigma^2}{\sum_{i=1}^{I}(m_i - \overline{m})^2}, \tag{44}$$

$$\mathrm{cov}(a, b) = -\frac{\sigma^2 \overline{m}}{\sum_{i=1}^{I}(m_i - \overline{m})^2}, \tag{45}$$

where σ is the standard deviation calculated as

$$\sigma = \sqrt{\frac{\sum_{i=1}^{I}(y - \hat{y})^2}{I - 2}} = \sqrt{\frac{\sum_{i=1}^{I}(\log_{10}(n(m_i)) - a - bm_i)^2}{I - 2}} \tag{46}$$

and I is the number of magnitude bins taken into account in the calculation, only including the bins with a nonzero numbers of n.

After this, the margins of error can be calculated using the formulas (39) and (40). In the Master's thesis, the error margin of b is calculated by replacing the square root of $\mathrm{var}(b)$ with the covariance divided by the square root of $\mathrm{var}(a)$ to avoid assuming that the error of parameter b follows normal distribution. However, in this work the margins of error are computed in the same way for both MLE and LS methods to ease their comparison.

In the Slate PSHA and in the Master's thesis, the variances and the covariance of a and b seem to be calculated using bivariate normal distribution theory. However, the regular method more commonly used in mathematics is used in this work. This together with the difference in the error formulas might cause some differences in the margins of error of the results from the Slate PSHA and the Master's thesis compared to the results presented in this work.

Based on the definition of the GR equation, it should be noted that when using the LS method, the values of $n(m_i)$ should be normalised by dividing them with the corresponding time intervals. This way the the earthquakes of higher magnitudes with a longer measurement history do not have more weight than the low-magnitude earthquakes. In addition, it is important to use the cumulative values of $n(m)$ required in the GR equation.

While being a rather simple method to determine the parameters a and b, LS method has its downsides. The most commonly recognized shortcoming is the inability to take the empty magnitude bins in the high-magnitude end into consideration [8], ignoring the valuable information about the lack of earthquakes of certain magnitudes. Especially in Finland, where the seismic activity is quite low, this can have a significant impact on the results. Because of this, many sources suggest that LS method should not be used when estimating the seismic hazard [7, 8, 12].

3 Methods

3.1 Earthquake catalog

An earthquake catalog is a list of earthquakes observed containing information about where and when earthquakes have occurred and how strong they were. The earthquake catalog used in this work is the same as the one used in the Master's thesis and in the Slate PSHA. This catalog is based on two catalogs collected by the University of Helsinki in 2016 and 2020. The catalog from 2016 is not publicly available, but information about it can be found from Appendix 2 of [13]. The catalog from 2020 can be found from [14]. To make the catalog suitable for the ground motion prediction equations (GMPEs) and recurrence methods selected, the earthquake magnitudes must be converted to moment magnitudes [2]. This has already been done for the catalog used in the Slate PSHA. In addition, the original catalog was declustered by removing foreshocks and aftershocks and only leaving the mainshocks. This is also a requirement for the MLE method, as using the Poisson distribution expects independently occurring events.

After these changes, the catalog has been assessed for completeness. This was done in the Slate PSHA by dividing the earthquakes into half-magnitude bins and on visual examination assigning each bin a completeness time [2]. Completeness time represents the time after which all the earthquakes from a chosen magnitude range have been captured. This means that larger earthquakes generally have longer completeness times, as they have been measured for a longer time. Completeness times for each magnitude bin can be found from Table 1. After determining the completeness times, the seismic events outside these time frames were filtered out. This left the final catalog with 360 events. However, the catalog used in the Slate report and in the Master's thesis still included three seismic events with magnitudes 0.7, 0.7 and 0.2 in area 5 from 2012, which are outside the completeness times and were thus removed from the catalog used in this work. Due to the large number of earthquakes in area 5, removing these three earthquakes did not affect the parameters of this area very much in the tests performed.

Even though the catalog used has been assessed for completeness, the earthquakes of a magnitude lower than 1.0 are still excluded from the Slate PSHA and the Master's thesis. This decision was made because of the insufficient sampling of events in this magnitude range [2]. As can be seen from Table 1, the earthquakes of magnitudes under 1.0 were assigned with the same completeness times as the earthquakes of magnitude 1.0-1.5, but the number of these earthquakes is still lower. The first magnitude bin only has seven occurrences, which is clearly against the expected count based on the GR equation. Thus, these earthquakes of magnitude 0.0-0.5 are left out from this work as well. However, though excluded from the Slate PSHA, the 0.5-1.0 magnitude bin is used for the sensitivity analysis.

3.2 Seismic source areas

To categorize areas by which kind of earthquakes they produce and how often, the interplate area around Finland is divided into seismic source areas (SSAs). The division used in this work is executed in a study by the University of Helsinki [13] in 2016 and is used in the Slate PSHA as well. A map of this SSA division is presented in Figure A1 in Appendix A. There are also other possible divisions but those are left out from this work. Parameters a and b are calculated for each SSA separately, and together they are used for estimating the total seismic hazard.

In the SSA division used, there are in total of 11 source areas, and the Loviisa NPP is located within Zone 10. Zone 6 is divided into three Subzones, 6a, 6b and 6c. The Olkiluoto NPP is located within Subzone 6a. Only a few seismic events have been measured from Zones 7 and 11, and these events were all removed during the catalog completeness evaluation. Thus, parameters of Zone 6 are assumed for Zones 7 and 11 as well. Zone 9 is not included at all due to its distance and lack of seismic events. Zones 1-5 are located outside a 300 km radius of the Loviisa NPP and do not have a significant impact on the seismic hazard, which is why they can be left out from the Loviisa PSHA. However, these areas are in the vicinity of the Olkiluoto NPP and are thus included in this work as well. The earthquake counts from each area and magnitude bin can be found from Table 1.

Table 1: Earthquake counts for each SSA and the completeness times of each magnitude bin based on the catalog from the Slate PSHA [2]. Earthquakes of a magnitude equal to the magnitude bin boundary belong to the higher magnitude bin.

Magnitude	0.0-0.5*	0.5-1-0*	1.0-1.5	1.5-2.0	2.0-2.5	2.5-3.0	3.0-3.5	3.5-4.0	4.0-4.5
Start year	2013	2013	2013	2012	1994	1994	1944	1944	1909
End year	2014	2014	2014	2014	2014	2014	2014	2014	2014
1	0	2	3	1	8	2	2	2	0
2	0	6	9	8	16	1	1	0	0
3	0	0	3	6	16	1	0	0	0
4	0	0	0	6	7	1	2	0	0
5	5**	34**	37	22	22	5	6	0	1
6	0	0	5	3	4	0	0	3	1
6a	0	0	1	1	2	0	0	0	0
6b	0	0	3	1	1	0	0	0	0
6c	0	0	1	1	1	0	0	3	1
8	0	18	9	3	6	1	2	0	1
10	2	19	34	9	1	0	1	0	0
Total	7	79	100	58	80	11	14	5	3

* The earthquakes of a magnitude less than 1.0 are not included in the Slate PSHA.
** The three earthquakes outside the completeness times in Zone 5 are excluded from the table.

3.3 Implementation

The calculations for the parameter estimation in the Master's thesis are performed using Excel. To make sure that the differences between the MLE and LS methods are not caused by any computational errors, the code is rewritten for this work using Python. When solving the value of b numerically, Newton's method is used in the Master's thesis. In this study, bisection method is used instead to avoid calculating the derivatives for different versions of the formulas. Using another method to solve the root should, however, not have an impact to the estimated values within the given precision. The required accuracy of 10^{-9} for the value of β used for the MLE method in the Master's thesis is also used in this work. For the values of b estimated in this work, this accuracy results in a relative error of approximately 10^{-7} %, which is not significant in the determination of the seismic hazard.

To ensure that the verification gives reliable results, it is performed with the same parameter values as used in the Slate PSHA and in the Master's thesis. In the Master's thesis, a minimum magnitude of 1.0 and a maximum magnitude of 5.0 are used. In the Slate PSHA, the maximum magnitude is not specified, so the same value of 5.0 is assumed. However, as the empty magnitude bins from the high-magnitude end cannot be included in the LS method, any maximum magnitude higher than the strongest observed earthquake leads to the same result. Instead, in the MLE method, the maximum magnitude can affect the estimated values. The width of the magnitude bins is 0.5, as this is also the bin width used in the estimation of the completeness times. For the LS method, the observed cumulative values are set to the middle points of the magnitude intervals, which is also done in the Master's thesis and in the Slate PSHA.

For the sensitivity analysis, these parameters are changed. In addition to using 0.5 bin width, 0.1 bins are tried as well, which is also the precision of the given earthquake data. As the completeness times are only given for the half-magnitude bins, the same completeness times must be assumed individually for each 0.1 bin inside the original interval. For some magnitude bins, this may lead to using completeness times that are shorter than the actual ones, as with the half-magnitude bins the completeness times must be estimated based on the smallest magnitude of each bin. However, even though possibly leaving out some useful information, this should not distort the data. Another method used is checking the completeness times for each 0.1 magnitude bin individually by finding the first year an earthquake with a magnitude of that bin or smaller first occurs in the catalog. Like with the first method, this may not give realistic completeness times, but the method is included to get a better view of how sensitive the parameters are to the completeness times. The full original catalog and the details of the completeness time evaluation process are not available, so new evaluation of the completeness times cannot be performed in this work.

Different values for the minimum and maximum magnitudes are also tested for the sensitivity analysis. For the minimum magnitude, the tested values are 0.5 and 1.5. In addition, we try defining the minimum magnitude for each SSA separately by checking the smallest observed earthquake from each area, which requires using the 0.1 bin width. It should be noted that the smallest observed magnitudes might be smaller for many areas if they were checked from the original catalog. Due to the small number of occurrences in the first magnitude bins, all minimum magnitudes are set to be at least 0.5. For the maximum value, a magnitude of 8.0 is tested as well.

The differences between the methods and different parameters are tracked by comparing the absolute differences of a and b. Relative differences are not used, as they may give a biased view of the differences due to the use of logarithmic scale. In some sources, like in the Master's thesis, the activity rate at the smallest magnitude of interest, $n_{\min}$, is given instead of parameter a. These values are not presented in this work, as they can easily be calculated when the parameters a and b are known. In the Slate PSHA, the smallest magnitude of interest is set to 4.5. It should be noted that this is a different value from the minimum magnitude $m_{\min}$ used for the parameter estimation, as the smallest magnitude of interest is used for the estimation of the hazard, which only includes earthquakes strong enough to possibly cause harm.

In the Slate PSHA and in the Master's thesis, the parameter b is set to 2.0 for magnitudes greater than 5.75. This decision has no effect to the parameter sensitivity analysis and thus is not included in the plots of this work. However, when calculating the total seismic hazard, this change needs to be taken into account.

Both in the Slate PSHA and in the LS method used in the Master's thesis, all magnitude bins with zero occurrences were left out from the line fitting even when the cumulative value n is nonzero. However, only the empty bins from the high-magnitude end need to be left out due to the logarithmic scale. To see whether including the empty middle bins gives more reliable results, the impact of these magnitude bins is taken into account for the sensitivity analysis.

In this work, the time periods t_i are measured in years, which leaves out the possible effect of leap days. In the Master's thesis, the leap days are taken into account, but the length of the year has been computed only until 30.12., leaving out one day and resulting in a year length of 364 days. These differences can cause small inequalities in the parameter comparison.

3.4 HAZ code

To calculate the annual frequency of exceedance, e.g. an open-source code HAZ by Norman Abrahamson [15] can be used, though many other programs exist as well. The version 45.2 of the HAZ code has been modified by Slate, and Fortum has been provided with the modified code. Modifications include additional ground motion prediction equations (GMPEs) and depth distributions in addition to a modified b-value approach [2]. To avoid running the code repeatedly for different parameters, the verification and sensitivity analysis are performed for the parameters a and b, as they are the only parameters defining the seismic acticity of each SSA.

4 Results

4.1 Verification

The verification is performed for both the LS and the MLE method. The estimated values of parameters a and b are compared to the corresponding values from the Slate PSHA and from the Master's thesis. For the LS method, the parameters a and b are estimated using equations (42) and (41), and the variances of these parameters are calculated with (43) and (44). For the MLE method, the parameters α and β are estimated using equations (19) and (30), and these parameters can then be used for solving a and b. The variances of a and b are calculated with (36) and (37). For both methods, the margins off error are calculated based on the variances according to equations (39) and (40). These equations are used throughout the work.

4.1.1 LS method

First, we shall verify the results calculated with the LS method. The results from the Slate PSHA are presented in Table 2, including the 90 % error margins. In Table 3, the results calculated with the LS method from the Master's thesis and this work are presented. The estimated parameter values between the two works are nearly the same: the differences are mostly under 0.001 units for both the values of a and b, and the relative difference compared to the values from the Master's thesis is less than 0.08 % for all areas. These inequalities are caused by the difference in the calculation of the time period t. Again, the 90 % error margins are presented as well. These have more variety between the two works, as the method used for calculating them is different.

Table 2: Estimated a and b values and 90 % error margins using the LS method from the Slate PSHA [2].

SSA	a	$a_{5\%}$	$a_{95\%}$	b	$b_{5\%}$	$b_{95\%}$
1	1.361	0.790	1.933	0.784	0.579	0.988
2	2.925	1.778	4.071	1.451	0.988	1.915
3	3.665	1.380	5.951	1.773	0.790	2.756
4	2.635	1.222	4.049	1.322	0.784	1.861
5	2.864	1.968	3.761	1.176	0.873	1.479
6	1.441	0.673	2.210	0.790	0.547	1.034
6a	1.260	0.263	2.257	0.991	0.451	1.532
6b	2.308	0.723	3.894	1.599	0.740	2.458
6c	0.593	0.015	1.172	0.579	0.396	0.763
8	1.698	0.986	2.409	0.923	0.683	1.164
10	3.056	1.495	4.617	1.596	0.941	2.252

Table 3: Estimated a and b values and 90 % error margins using the LS method from the Master's thesis [3] compared to the results calculated in this work. The values of a from the Master's thesis are calculated based on the values of $n_{\min}$ given.

	Master's thesis				This work			
SSA	a	a error	b	b error	a	a error	b	b error
1	1.3645	± 0.5726	0.7846	± 0.2051	1.3641	± 0.1638	0.7844	± 0.0620
2	2.9276	± 1.1471	1.4523	± 0.4640	2.9271	± 0.5649	1.4520	± 0.2395
3	2.4819	± 1.1705	1.2823	± 0.5428	2.4814	± 1.1435	1.2821	± 0.5507
4	2.5952	± 1.3904	1.3059	± 0.5297	2.5950	± 0.4814	1.3057	± 0.1879
5	2.8792	± 0.9004	1.1802	± 0.3042	2.8788	± 0.2195	1.1800	± 0.0794
6	1.4432	± 0.7689	0.7907	± 0.2437	1.4428	± 0.4266	0.7905	± 0.1475
6a	1.2579	± 0.9946	0.9895	± 0.5390	1.2570	± 0.5434	0.9890	± 0.3024
6b	2.3063	± 1.5832	1.5973	± 0.8580	2.3051	± 0.3588	1.5966	± 0.1997
6c	0.5949	± 0.5786	0.5800	± 0.1833	0.5947	± 0.4867	0.5798	± 0.1683
8	1.7617	± 0.7280	0.9429	± 0.2460	1.7613	± 0.3781	0.9427	± 0.1368
10	3.1367	± 1.5901	1.6258	± 0.6675	3.1359	± 1.4181	1.6255	± 0.6303

When comparing the results from this work to the Slate PSHA results, the absolute difference is under 0.08 units and under 4 % for a and under 0.03 units and under 3 % for b for most of the SSAs. However, in Zone 3, the difference is 1.18 for a and 0.49 for b. This difference is relatively larger than the others, which indicates either an error in the calculation process or a difference in the calculation method. It seems that, in the Slate PSHA, the first observation point from Zone 3 is excluded from the fit. If the same decision is done, the value of a increases to 3.6626 and the value of b to 1.7709. This reduces the differences to -0.024 units for a and -0.0021 units for b, which correspond to relative differences of -0.07 % and -0.12 %. These are now of the same order of magnitude as the other differences. Thus, it can be assumed that the first value was left out from the Slate PSHA, though it was not mentioned in the report. This seems like a reasonable decision, as the first observation point is quite far from the fitted line. However, this point was still included in this work to avoid making similar examination for other situations as well. In addition, the number of data points is too small to definitely state that one point should be excluded. Including the first observation point also makes it easier to compare the results with the values from the Master's thesis. To demonstrate the effect of the first observation point, the graphs for Zone 3 are plotted in Figure (1).

The Slate PSHA claims to use MLE method from [8] and [16] for the parameter estimation. However, it is not specified which formulas from these sources are used, and none of these formulas tested for this work seem to give results close to the Slate PSHA values. Instead, as stated above, the Slate PSHA results are very close to the values calculated with the LS method, so it seems that the parameter estimation is actually based on this method.

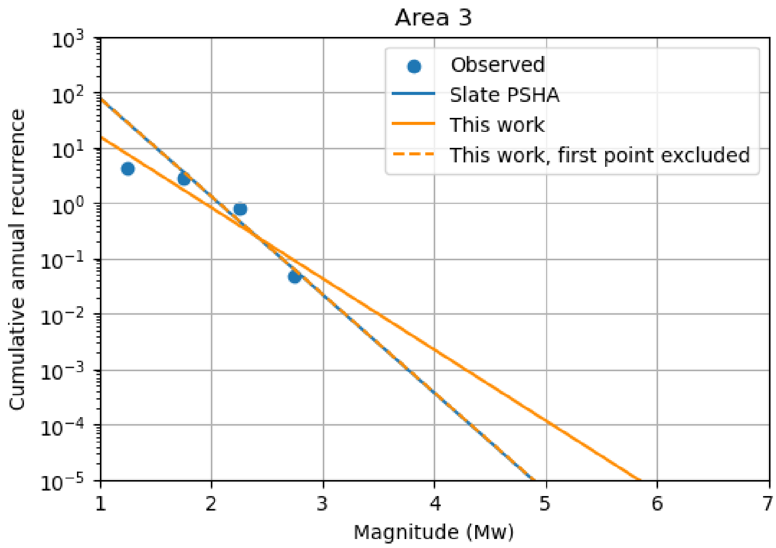


Figure 1: Comparison of GR equations from Slate PSHA and this work using LS method for area 3.

4.1.2 MLE method

Next, the verification is performed for the MLE method. In Table 4, the results calculated with the MLE method in the Master's thesis and this work are presented. The results are again very similar, the difference in the values of a being less than 0.002 units (0.07 %) and in the values of b less than 0.001 units (0.05 %) for all SSAs. Again, the differences are caused by the difference in the calculation of the time period t. The 90 % error margins, also presented in the table, are almost identical as well. The relative differences of the LS and MLE results from this work compared to the Slate PSHA values are presented in Table 5.

Table 4: Estimated a and b values and 90 % error margins using the MLE method from the Master's thesis [3] compared to the results calculated in this work. The values of a from the Master's thesis are calculated based on the values of $n_{\min}$ given.

	Master's thesis				This work			
SSA	a	a error	b	b error	a	a error	b	b error
1	1.3023	± 0.3977	0.8353	± 0.1968	1.3018	± 0.3981	0.8350	± 0.1969
2	2.1402	± 0.2864	1.1672	± 0.1736	2.1393	± 0.2864	1.1667	± 0.1736
3	1.8614	± 0.3290	1.0699	± 0.1880	1.8607	± 0.3290	1.0695	± 0.1880
4	1.4387	± 0.4186	0.9412	± 0.2208	1.4380	± 0.4186	0.9408	± 0.2209
5	2.6675	± 0.1778	1.2374	± 0.1123	2.6666	± 0.1778	1.2369	± 0.1123
6	1.4035	± 0.4189	0.9208	± 0.2183	1.4029	± 0.4190	0.9205	± 0.2183
6a	1.2613	± 0.8530	1.2100	± 0.5303	1.2605	± 0.8528	1.2095	± 0.5302
6b	1.9228	± 0.8773	1.6374	± 0.6747	1.9216	± 0.8771	1.6367	± 0.6745
6c	0.2850	± 0.6731	0.5301	± 0.2907	0.2849	± 0.6883	0.5299	± 0.2963
8	1.8656	± 0.3592	1.1191	± 0.2115	1.8648	± 0.3591	1.1186	± 0.2115
10	3.2716	± 0.3543	1.9791	± 0.3045	3.2703	± 0.3543	1.9783	± 0.3044

Table 5: Differences of the a and b values estimated in this work with the LS and MLE methods compared to the Slate PSHA values. Negative value indicates that the estimated value is smaller than the Slate PSHA result and positive that it is greater. Total difference is calculated as a sum of the absolute values of the differences. The asterisk (*) indicates values which do not take the first data point of Zone 3 into account.

	LS		MLE	
SSA	a	b	a	b
1	0.0031	0.0004	-0.0592	0.0510
2	0.0021	0.0010	-0.7857	-0.2843
3	-1.1836	-0.4909	-1.8043	-0.7035
3*	-0.0024	-0.0021	-	-
4	-0.0400	-0.0163	-1.1970	-0.3812
5	0.0148	0.0040	-0.1974	0.0609
6	0.0018	0.0005	-0.0381	0.1305
6a	-0.0030	-0.0020	0.0005	0.2185
6b	-0.0029	-0.0024	-0.3864	0.0377
6c	0.0017	0.0008	-0.3081	-0.0491
8	0.0633	0.0197	0.1668	0.1956
10	0.0799	0.0295	0.2143	0.3823
Total	1.3962	0.5675	5.1578	2.4946
Total*	0.2150	0.0787	-	-

4.2 Sensitivity analysis

Next, a sensitivity analysis is performed for different input parameters. The goal of the sensitivity analysis is to determine how sensitive the values of a and b are to different parameters and to get information about which parameters would be the most suitable for the hazard estimation.

In most of the calculations performed, the sensitivity is estimated by calculating the difference of the new values compared to the values calculated with the original parameters. The original parameters are the same as those used in the verification above and also in the sensitivity analysis performed in the Master's thesis: a bin width of 0.5, completeness times from table 1, a minimum magnitude of 1.0 and a maximum magnitude of 5.0. However, some of the tests performed require a shorter bin width of 0.1, e.g. when a precise minimum magnitude is wanted. In these cases, the results are compared to the values calculated with the 0.1 bins to omit the impact of the bin width when the sensitivity of another parameter is estimated.

4.2.1 Bin width

For the sensitivity analysis, a shorter width for the magnitude bins is tested first. Instead of the 0.5 magnitude bins, 0.1 magnitude bins are used. This causes the number of data points used to increase significantly, so now single abnormal values affect the estimated parameters less. However, the number of empty magnitude bins increases simultaneously.

The resulting parameters from this test are presented in Table B1 of Appendix B. By looking at the differences compared to the original values, the changes do not seem consistent between the MLE and the LS method. As the number of data points with the 0.5 magnitude bins is quite small, the arrangement of those points can have a large impact, especially when LS method is used. Thus, increasing the number of data points can cause significant changes in the parameter values, especially in the areas where the number of earthquakes is particularly low.

4.2.2 Completeness time

Next, the completeness times are redefined by checking them for each 0.1 magnitude bin separately. However, this does not change the completeness times for every bin, and some of the changes are relatively small. 0.1 magnitude bins are still used, as otherwise the new completeness times would not affect the results. The original and the redefined completeness times are presented in Table 6.

Table 6: Original an redefined completeness times for 0.1 magnitude bins, given in years. Red color indicates bins where the completeness time has changed.

Magnitude	-1.4	1.5-1.9	2.0-2.3	2.4-2.9	3.0-3.9	4.0-4.2	4.3-4.4	4.5-
Original	2	3	21	21	71	106	106	106
Redefined	2	3	21	22	71	71	84	106

The estimated parameter values are presented in Table B2, which also includes the comparison to the results from Table B1. For both methods and all SSAs, the difference is less than 0.05 units for a and less than 0.03 units for b. Thus, the impact of the redefined completeness times is small, though more significant changes in the completeness times would likely cause larger differences.

4.2.3 Maximum magnitude

Next, the maximum magnitude is switched from 5.0 to 8.0. This is only performed for the MLE method, as the LS method cannot take into account the empty bins in the high-magnitude end and thus they do not affect the results. The estimated parameter values using the MLE method and the updated maximum magnitude are given in Table B3.

For most of the SSAs, increasing the maximum magnitude by 3.0 units has little to no effect on the parameter values. For both a and b the new values are higher, meaning that the high-magnitude earthquakes are less likely. This makes sense, as extending the interval from the high-magnitude end shows that there has not been earthquakes of a magnitude higher than 5.0. Only for the SSA 6c the values of a and b are significantly higher than before, which is probably caused by the small number of earthquakes on the area with more high-magnitude earthquakes than any other area. Increasing the maximum magnitude on this area also brings the parameter values a little closer to the Slate PSHA values.

In most of the SSAs, the largest observed earthquake is significantly smaller than the original maximum magnitude of 5.0. If the maximum magnitude is set to a value much closer to the largest observed magnitude, the difference compared to the original value increases, as the important information about the lack of high-magnitude earthquakes is left out. However, the results from these tests are not included in this work, as taking into account the empty magnitude bins from the high-magnitude end is one of the main advantages of the MLE method.

4.2.4 Minimum magnitude

Now, different values for the minimum magnitude are tested. First, the minimum magnitude is decreased from the original value of 1.0 to 0.5. The estimated parameter values using both the MLE and the LS method and the updated minimum magnitude are given in Table B4.

When the LS method is used, there is no difference in the a and b values for most of the SSAs. This is due to the fact that for these areas the magnitude of the smallest observed earthquake is greater or equal to 1.0, and the empty magnitude bins are left out from the estimation. For the other SSAs, both parameters a and b are mostly smaller than the original values, as the first observation point is generally a little lower than the estimated value based on the GR equation.

When using the MLE method, having either few or no earthquakes in the first magnitude bin cause the values of a and b to be lower, explaining the negative changes in Table B4. This means that the GR plot has lower values in the low-magnitude end and that it is less steep. This results to a greater estimated recurrence of the high-magnitude earthquakes. An example of this kind of situation is given in Figure 2, where the GR plots of Zone 6 are compared for the minimum magnitudes of 0.5 and 1.2, the latter being the magnitude of the smallest observed earthquake in that area. In the figure, a bin width of 0.1 is used to better show the impact of the empty magnitude bins, though the same effect happens with wider magnitude bins. With a smaller minimum magnitude, the high-magnitude recurrence is significantly higher. This can have a great impact on the total seismic hazard, as the recurrence of the stronger earthquakes is used for determining it.

Next, a minimum magnitude of 1.5 is tested, which is higher than the original value an thus leaves the first data point off from all SSAs except Zone 4, which does not have earthquakes of a magnitude smaller than 1.5. The estimated parameter values using both the MLE and the LS method and a minimum magnitude of 1.5 are given in Table B5.

As can be seen from the table, the differences compared to the original values are really large, which can be explained by the fewness of the observation points. For many areas, the number of data points was already small, and thus leaving out just one observation point can drastically change the estimated parameters.

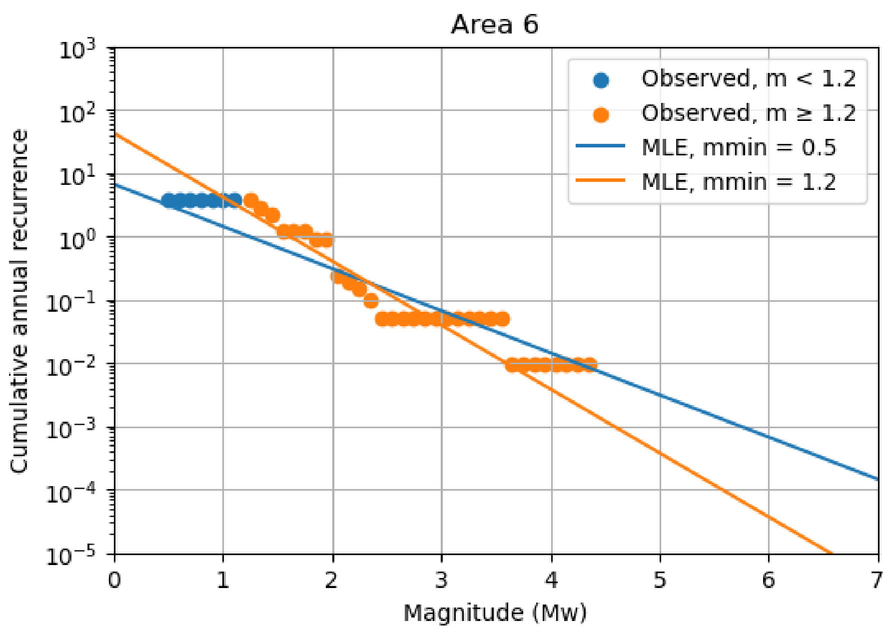


Figure 2: Comparison of two different minimum magnitudes for SSA 6 with MLE method.

The last minimum magnitudes tested are based on the smallest observed earthquakes from each SSA. Now, the minimum magnitude is specific for each SSA and is set to the same value as the smallest observed earthquake, though at least to the value of 0.5. For this, the bin width is set to 0.1 to maximize the accuracy with which the minimum magnitudes can be chosen. The results from this test are presented in Table B6.

When using the LS method, the estimated parameters are different only when the minimum magnitude is smaller than 1.0. In this case, the changes are similar to the ones observed with the minimum magnitude of 0.5, as the empty magnitude bins in the beginning do not affect the parameter estimation.

When using the MLE method, the estimated parameter values increase for areas 3, 4 and 6 including its subzones. For these areas, the smallest observed earthquake is greater than 1.0, and now the empty magnitude bins in the beginning are left out. This means that the empty magnitude bins in the beginning are not pulling the line down, so the values of both a and b increase and the estimated recurrence of the high-magnitude earthquakes decreases.

The greatest decrease in the parameter values for the MLE method can be seen on area 10. This can be explained by the low number of earthquakes in the range of 0.5-1.0, which is now taken into account, making the graph less steep and lower in the beginning. This increases the estimated likelihood of high-magnitude earthquakes occurring. The same problem, though not as clear, can be seen on areas 2 and 5, where the parameter values decrease as well.

4.2.5 Including empty magnitude bins

Finally, including the empty middle bins in the LS method is tested. Now, the only empty magnitude bins excluded are those smaller than the smallest observed earthquake and larger than the greatest observed earthquake in each SSA. As before, the earthquakes of a magnitude smaller than 0.5 are left out. The bin width of 0.1 is used so that more data points can be included, and the results are thus comparable to those presented in Table B6. The updated parameter values are presented in Table B7. The MLE method is not included as it already takes into account the empty magnitude bins.

Based on the results, the value of parameter b increases for almost all of the areas as the empty magnitude bins make the line steeper. Especially for parameter a, the difference on each SSA depends on the number of empty magnitude bins included and where those bins are located. This causes various differences depending on the SSA. The margins of errors stay on about the same level as before, except in SSA 6 and its subzones. In these areas, the margins of error are mainly smaller than before, which is likely caused by the small number of data points before adding the empty bins. However, in Subzone 6b the margins of error increase, as the GR plot previously aligned so well with the data points in that area.

5 Discussion and conclusions

The verification shows that both MLE and LS methods give results similar to those calculated in the Master's thesis, which indicates that the calculations and the implementation are correctly done. The small differences between the results from these two works can be explained by the different values of parameter t. The results from the LS method also matched well with the results from the Slate PSHA.

The sensitivity analysis performed gives an insight of the impact of different parameters. Even though a shorter bin width of 0.1 does not give consistent results, it is a useful tool when more data points or a more precise minimum magnitude are wanted. The redefined completeness times do not have a major impact, but further analysis on the topic would be useful when an updated catalog is available. Increasing the maximum magnitude only has an impact when the area has a lot of high-magnitude earthquakes, which in our case only implies in Zone 6c. Instead, changing the minimum magnitude has a major impact on the estimated parameter values, especially when using the MLE method, and it seems that the estimated values are the most sensitive to the minimum magnitude. When using the LS method, the estimated parameters can vary significantly depending on whether the empty magnitude bins in between are included or not.

Now that the effect of different parameters is known, the most suitable parameters can be suggested for our situation, where there is not much data, completeness times are short and most earthquakes are small. As the full earthquake catalog is not available for this work, it is best to use the original completeness times and assume them for each 0.1 magnitude bin as well. Shorter magnitude bins of 0.1 would probably be better, because this offers more data points and a better accuracy for e.g. the minimum magnitude. With more data points, fluctuations of single data points do not have as much impact and the scattering of the data points can be obtained with a higher precision. However, defining completeness times separately for each 0.1 magnitude bin may be difficult, especially when there is not much data.

For the maximum magnitude, setting the value high enough is likely to give more accurate results. This way, the MLE method gets more information about the previous occurrence of high-magnitude earthquakes. For the LS method, changing the maximum magnitude does not have an impact. The minimum magnitude is recommended to be set to the value of the smallest observed earthquake in each SSA, because the completeness times given in the Slate PSHA are not very accurate and it is not certain whether the earthquake data is complete before those values. Due to the same reason we cannot be sure whether the data before the magnitude of 1.0 is actually complete even within the given completeness times, so all minimum magnitudes used should be at least 1.0. All earthquakes of a magnitude less than 1.0 are left out from the Slate PSHA as well.

In the LS method, including the empty magnitude bins mimics the way of MLE method taking those bins into account and is thus used for the final comparison of the two methods with the suggested parameters. Without clear conclusions or information from other sources, it is good to keep the inclusion of the empty middle bins as an option when the LS method is used. In the future or in more active areas, when more earthquake data is available, the number of empty magnitude bins decreases, making the decision of whether to include those bins or not less significant.

The values of a and b estimated with these parameters, including the 90 % margins of error, are presented in Table C1 of Appendix C. In Table C2, the differences with the values calculated with the MLE and LS methods are presented. The GR plots based on these values can be found from Figure C1. With these parameter values, the total difference of parameters a and b between the two methods is smaller than in any of the cases tested for the sensitivity analysis. In some of the SSAs, significant differences between the two methods still exist. Even though having more data could decrease the differences even more, getting very similar results with the two methods is not necessarily something that should be achieved, as the methods have different operating principles and limitations.

It should be noted that these parameters presented might not be the only justifiable option, and the most suitable parameters may depend on the area and the earthquake catalog used. The logic tree used for the hazard calculation usually includes a lot of different branches, and adding new branches for different parameters is a viable option.

The study completed does not give a clear answer on which of the tested methods gives more reliable results in the specific area examined in this work. However, as multiple sources suggest the use of MLE method instead of LS method, it is advisable to take the parameters estimated with the MLE method into account as well instead of using only the LS parameters as done now. This could be done by e.g. including the MLE method into the logic tree used in the hazard calculation process.

The margins of error presented in this study are calculated with methods chosen partly arbitrarily based on assumptions and easing the calculation and comparison processes. These error estimation methods are not the only suitable options, and other methods might lead to significantly different results. Confidence bounds of the parameters are usually included in the estimation of the seismic hazard, so the margins of error have a concrete impact. To ensure that the margins of error are justified, further analysis would be recommended.

The parameters estimated in this study are based on a relatively narrow catalog, which was already declustered and filtered based on completeness evaluation. The original catalog is not available, so the declustering and filtering decisions together with the possible randomness and measurement errors can have a large impact regardless of the estimation method. 360 seismic events divided into 8 zones does not leave too many events for each SSA, so great uncertainties remain. For the smallest earthquakes, the catalog only covers two years, even though in a low-seismic area such as Finland even those smallest earthquakes are important in the parameter estimation. By collecting more data and including that into the catalog, the estimates could be made more reliable. In addition, by re-assessing the catalog for completeness especially for the 0.1 magnitude bins, the recurrence could be better estimated.

In addition to the earthquake catalog, a different version of the SSA division could be used. The division is updated based on newest information, and a different division might have SSAs better gathering seismically similar areas to the same SSA. The new SSAs could then be used to calculate even more accurate parameters.

References

[1] P. Varpasuo. "The Finnish Approach to Seismic Hazard Analysis – Case Loviisa". In: *OECD-NEA Workshop on Recent Findings and Developments in PSHA Methodologies and Applications*. Vol. 1. Lyon, France, 2008.

[2] Slate Geotechnical Consultants, Inc. *Finland Probabilistic Seismic Hazard Analysis*. 2021. LO1-T84252-00016Liite1.

[3] V. Visakorpi. "Sensitivity analysis of a seismic hazard assessment for a Finnish nuclear power plant". Master's thesis. Espoo: Aalto University, 2022.

[4] S. Stein, R.J. Geller, and M. Liu. "Why earthquake hazard maps often fail and what to do about it". In: *Tectonophysics* 562–563 (2012), pp. 1–25. DOI: 10.1016/j.tecto.2012.06.047.

[5] Radiation and Nuclear Safety Authority. *Guide YVL B.7: Provisions for Internal and External Hazards at a Nuclear Facility*. 2019. URL: https://www.finlex.fi/data/normit/41791-YVL_B.7e.pdf. (Accessed on 28.6.2023).

[6] R. McGuire. *Seismic Hazard and Risk Analysis*. Earthquake Engineering Research Institute, 2004.

[7] D. Stromeyer and G. Grünthal. "Capturing the Uncertainty of Seismic Activity Rates in Probabilistic Seismic Hazard Assessments". In: *Bulletin of the Seismological Society of America* 105 (2015), pp. 580–589. DOI: 10.1785/0120140185.

[8] D. Weichert. "Estimation of the Earthquake Recurrence Parameters for Unequal Observation Periods for Different Magnitudes". In: *Bulletin of the Seismological Society of America* 70 (1980), pp. 1337–1346. DOI: 10.1785/0120140185.

[9] D. Coe. *Fisher Matrices and Confidence Ellipses: A Quick-Start Guide and Software*. Jet Propulsion Laboratory, California Institute of Technology. 2009.

[10] N. Draper. *Applied Regression Analysis*. John Wiley & Sons, Incorporated, 1998.

[11] M. Kendall. *The advanced theory of statistics. Vol. 2*. Charles Griffin & Co. Ltd., 1948.

[12] L. Sandri and W. Marzocchi. "A technical note on the bias in the estimation of the b-value and its uncertainty through the Least Squares technique". In: *Annals of geophysics* 50 (2007), pp. 329–339. DOI: 10.4401/ag-4432.

[13] P. Mäntyniemi et al. "Seismic Source Areas in Central Fennoscandia". In: Institute of Seismology, Report S 64 (2016). Ed. by A. Korja, S. Kihlman, and K. Oinonen.

[14] University of Helsinki Institute of Seismology. *Catalog of earthquakes in Northern Europe (FENCAT) 1375-*. Web catalog compiled by Uski, M. and Oinonen, K. Last updated 28.2.2020. URL: https://www.seismo.helsinki.fi/bulletin/list/catalog/Scandia_updated.html. (Accessed on 5.7.2023.)

[15] N. Abrahamson. *HAZ*. Version 45.2. 2017. URL: https://github.com/abrahamson/HAZ. (Accessed on 6.7.2023.)

[16] K. Aki. “Maximum likelihood estimate of b in the formula log N = a – bM and its confidence limits”. In: *Bulletin of Earthquake Research* 43 (1965), pp. 237–239.

A Seismic source areas

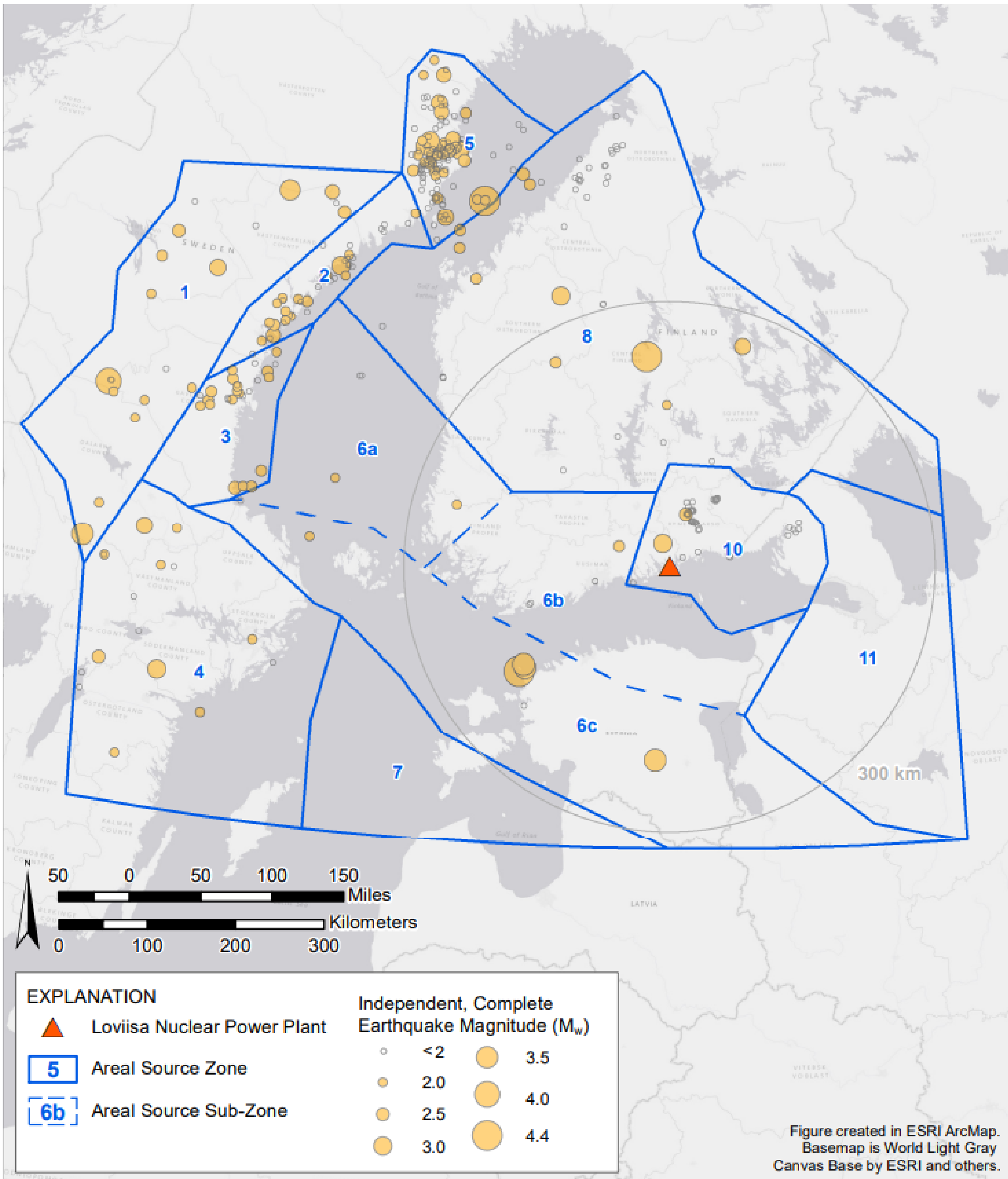


Figure A1: SSA division used in this work. Image source: [2]

B Results for sensitivity analysis

Table B1: Estimated values of a and b using both the MLE and LS methods and a bin width of 0.1. Delta (Δ) gives the difference compared to the original MLE and LS values.

	MLE				LS			
SSA	a	Δa	b	Δb	a	Δa	b	Δb
1	1.2970	-0.0048	0.8324	-0.0026	1.2568	-0.1073	0.7886	0.0042
2	2.1384	-0.0009	1.1661	-0.0006	2.5889	-0.3382	1.3623	-0.0897
3	1.8448	-0.0159	1.0595	-0.0100	2.1692	-0.3122	1.1667	-0.1154
4	1.4879	0.0499	0.9702	0.0294	2.3675	-0.2275	1.3129	0.0072
5	2.6142	-0.0524	1.2008	-0.0361	2.6769	-0.2019	1.1817	0.0017
6	1.3446	-0.0583	0.8872	-0.0333	1.3427	-0.1001	0.8111	0.0206
6a	1.1271	-0.1334	1.1204	-0.0891	2.3680	1.1110	1.5864	0.5974
6b	1.6386	-0.2830	1.4123	-0.2244	2.0733	-0.2318	1.4415	-0.1551
6c	0.3189	0.0340	0.5460	0.0161	0.6326	0.0379	0.5941	0.0143
8	1.8148	-0.0500	1.0864	-0.0322	1.6311	-0.1302	0.9375	-0.0052
10	3.0809	-0.1894	1.8105	-0.1678	3.0258	-0.1101	1.6429	0.0174

Table B2: Estimated values of a and b using both the MLE and LS methods and a bin width of 0.1 with redefined completeness times. Delta (Δ) gives the difference compared to the MLE and LS values estimated with 0.1 magnitude bins and original completeness times.

	MLE				LS			
SSA	a	Δa	b	Δb	a	Δa	b	Δb
1	1.2824	-0.0146	0.8240	-0.0084	1.2508	-0.0060	0.7879	-0.0007
2	2.1365	-0.0019	1.1658	-0.0003	2.5941	0.0052	1.3661	0.0038
3	1.8398	-0.0050	1.0571	-0.0024	2.1875	0.0183	1.1803	0.0136
4	1.4796	-0.0083	0.9657	-0.0045	2.3659	-0.0016	1.3134	0.0005
5	2.6130	-0.0012	1.2009	0.0001	2.6754	-0.0015	1.1830	0.0013
6	1.3326	-0.0120	0.8804	-0.0068	1.2968	-0.0459	0.7842	-0.0269
6a	1.1240	-0.0031	1.1192	-0.0012	2.3680	0.0000	1.5864	0.0000
6b	1.6404	0.0018	1.4146	0.0023	2.0733	0.0000	1.4415	0.0000
6c	0.2909	-0.0280	0.5290	-0.0170	0.5846	-0.0480	0.5658	-0.0283
8	1.8107	-0.0041	1.0846	-0.0018	1.5862	-0.0449	0.9117	-0.0258
10	3.0837	0.0028	1.8134	0.0029	3.0292	0.0034	1.6459	0.0030

Table B3: Estimated values of a and b using MLE method and a maximum magnitude of 8.0. Delta (Δ) gives the difference compared to the original MLE values.

	MLE			
SSA	a	Δa	b	Δb
1	1.3230	0.0212	0.8483	0.0133
2	2.1419	0.0026	1.1686	0.0019
3	1.8656	0.0049	1.0729	0.0034
4	1.4492	0.0112	0.9482	0.0074
5	2.6682	0.0016	1.2381	0.0012
6	1.4155	0.0126	0.9287	0.0082
6a	1.2624	0.0019	1.2109	0.0014
6b	1.9217	0.0001	1.6368	0.0001
6c	0.3917	0.1068	0.5895	0.0596
8	1.8683	0.0035	1.1212	0.0026
10	3.2703	0.0000	1.9783	0.0000

Table B4: Estimated values of a and b using both the MLE and LS methods and a minimum magnitude of 0.5. Delta (Δ) gives the difference compared to the original MLE and LS values.

	MLE				LS			
SSA	a	Δa	b	Δb	a	Δa	b	Δb
1	0.9577	-0.3441	0.6888	-0.1462	1.2087	-0.1554	0.7311	-0.0533
2	1.5132	-0.6261	0.8653	-0.3014	2.3388	-0.5883	1.2243	-0.2277
3	1.1692	-0.6915	0.7541	-0.3154	2.4814	0.0000	1.2821	0.0000
4	0.8641	-0.5739	0.6909	-0.2499	2.5950	0.0000	1.3057	0.0000
5	2.1048	-0.5618	0.9544	-0.2825	2.6833	-0.1955	1.1172	-0.0628
6	0.8474	-0.5555	0.6803	-0.2402	1.4428	0.0000	0.7905	0.0000
6a	0.4395	-0.8210	0.8150	-0.3945	1.2570	0.0000	0.9890	0.0000
6b	0.7016	-1.2200	0.9560	-0.6807	2.3051	0.0000	1.5966	0.0000
6c	0.0525	-0.2324	0.4373	-0.0926	0.5947	0.0000	0.5798	0.0000
8	1.6715	-0.1933	1.0227	-0.0959	1.8276	0.0663	0.9640	0.0213
10	1.9978	-1.2725	1.1636	-0.8147	2.8014	-0.3345	1.4928	-0.1327

Table B5: Estimated values of a and b using both the MLE and LS methods and a minimum magnitude of 1.5. Delta (Δ) gives the difference compared to the original MLE and LS values.

	MLE				LS			
SSA	a	Δa	b	Δb	a	Δa	b	Δb
1	1.4281	0.1263	0.8864	0.0514	1.3820	0.0179	0.7902	0.0058
2	2.9285	0.7892	1.5418	0.3751	3.5033	0.5762	1.6616	0.2096
3	2.9301	1.0694	1.5723	0.5028	3.6626	1.1812	1.7709	0.4888
4	2.4623	1.0243	1.4017	0.4609	2.5950	0.0000	1.3057	0.0000
5	2.9279	0.2613	1.3594	0.1225	2.8346	-0.0442	1.1666	-0.0134
6	1.1340	-0.2689	0.8108	-0.1097	1.2668	-0.1760	0.7417	-0.0488
6a	2.3272	1.0667	1.7312	0.5217	1.9183	0.6613	1.3064	0.3174
6b	2.6625	0.7409	2.0278	0.3911	2.7417	0.4366	1.8062	0.2096
6c	0.0431	-0.2418	0.4374	-0.0925	0.4464	-0.1483	0.5387	-0.0411
8	1.6426	-0.2222	1.0216	-0.0970	1.4951	-0.2662	0.8623	-0.0804
10	3.9594	0.6891	2.3610	0.3827	2.5908	-0.5451	1.4272	-0.1983

Table B6: Estimated values of a and b using both the MLE and LS methods and a minimum magnitude equal to the smallest observed earthquake in each SSA. Delta (Δ) gives the difference compared to the MLE and LS values estimated with 0.1 magnitude bins and the original minimum magnitude.

	MLE				LS			
SSA	a	Δa	b	Δb	a	Δa	b	Δb
1	1.2450	-0.0520	0.8103	-0.0221	1.2270	-0.0298	0.7783	-0.0103
2	1.6396	-0.4988	0.9253	-0.2408	2.1222	-0.4667	1.1525	-0.2098
3	2.0451	0.2003	1.1520	0.0925	2.1692	0.0000	1.1667	0.0000
4	2.5507	1.0628	1.4499	0.4797	2.3675	0.0000	1.3129	0.0000
5	2.0882	-0.5260	0.9388	-0.2620	2.5237	-0.1532	1.1235	-0.0582
6	1.6316	0.2870	1.0103	0.1231	1.3427	0.0000	0.8111	0.0000
6a	2.3828	1.2557	1.7317	0.6113	2.3680	0.0000	1.5864	0.0000
6b	2.5144	0.8758	1.8910	0.4787	2.0733	0.0000	1.4415	0.0000
6c	0.4210	0.1021	0.5854	0.0394	0.6326	0.0000	0.5941	0.0000
8	1.6510	-0.1638	1.0016	-0.0848	1.7120	0.0809	0.9665	0.0290
10	1.9640	-1.1169	1.1220	-0.6885	2.6470	-0.3788	1.4553	-0.1876

Table B7: Estimated values of a and b using the LS method and including empty middle bins. Bin width is 0.1 and the minimum magnitude is the same as used in Table B6. Delta (Δ) gives the difference compared to the LS values presented in that table.

	LS			
SSA	a	Δa	b	Δb
1	1.2370	0.0100	0.8081	0.0298
2	2.2960	0.1738	1.2900	0.1375
3	2.2873	0.1181	1.2501	0.0834
4	2.4138	0.0463	1.3749	0.0620
5	2.5379	0.0142	1.1506	0.0271
6	1.3162	-0.0265	0.8411	0.0300
6a	2.2612	-0.1068	1.5549	-0.0315
6b	2.2676	0.1943	1.6686	0.2271
6c	0.5184	-0.1142	0.6182	0.0241
8	1.6485	-0.0635	0.9730	0.0065
10	2.7835	0.1365	1.6372	0.1819

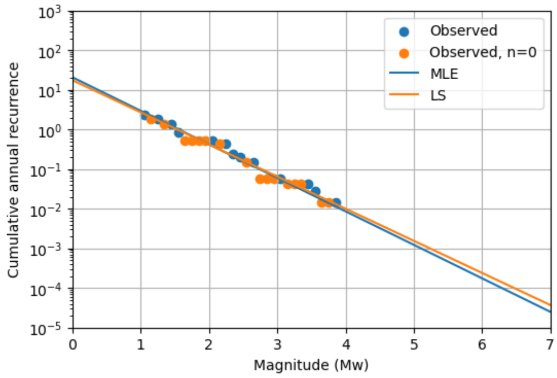
C Results with suggested parameters

Table C1: Estimated values of a and b using both the MLE and LS methods and parameters described in Section 5. The table includes 90 % error margins and minimum magnitudes specified for each SSA.

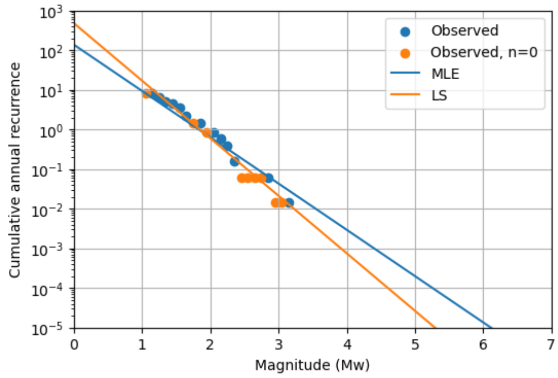
		MLE				LS			
SSA	m_{min}	a	a error	b	b error	a	a error	b	b error
1	1.0	1.3180	0.3832	0.8456	0.1869	1.2448	0.1075	0.8108	0.0415
2	1.0	2.1409	0.2808	1.1680	0.1694	2.6841	0.2033	1.4517	0.0927
3	1.1	2.0486	0.3574	1.1544	0.2022	2.2873	0.3338	1.2501	0.1707
4	1.5	2.5522	0.7129	1.4507	0.3773	2.4138	0.2969	1.3749	0.1209
5	1.0	2.6161	0.1732	1.2023	0.1066	2.6386	0.1900	1.1818	0.0649
6	1.2	1.6422	0.4811	1.0166	0.2436	1.3162	0.2467	0.8411	0.0837
6a	1.4	2.3830	1.4726	1.7318	0.8548	2.2612	0.9017	1.5549	0.4827
6b	1.2	2.5144	1.1345	1.8910	0.7792	2.2676	0.6367	1.6686	0.3474
6c	1.2	0.5178	0.6629	0.6370	0.2720	0.5184	0.2057	0.6182	0.0698
8	1.0	1.8190	0.3507	1.0894	0.2014	1.5385	0.1873	0.9387	0.0652
10	1.0	3.0809	0.3071	1.8106	0.2502	3.0475	0.3904	1.7492	0.1779

Table C2: Differences between the values of a and b estimated with the MLE and LS methods from Table C1. Total difference is calculated as a sum of the absolute values of the differences.

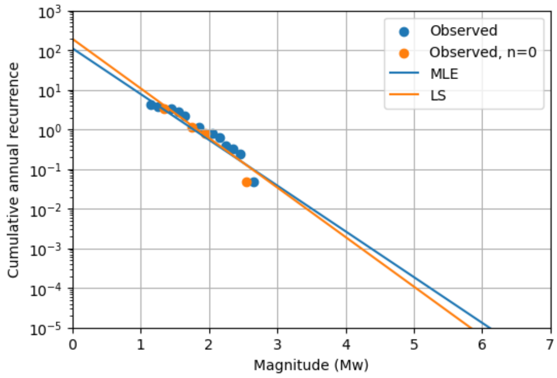
	Difference	
SSA	a	b
1	0.0732	0.0348
2	-0.5432	-0.2837
3	-0.2387	-0.0957
4	0.1384	0.0758
5	-0.0225	0.0205
6	0.3260	0.1755
6a	0.1218	0.1769
6b	0.2468	0.2224
6c	-0.0006	0.0188
8	0.2805	0.1507
10	0.0334	0.0614
Total	2.0251	1.3162



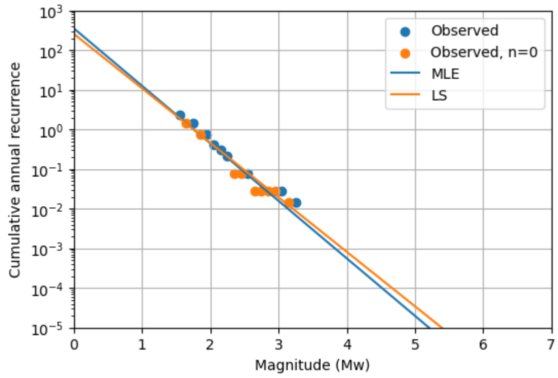
(a) SSA 1



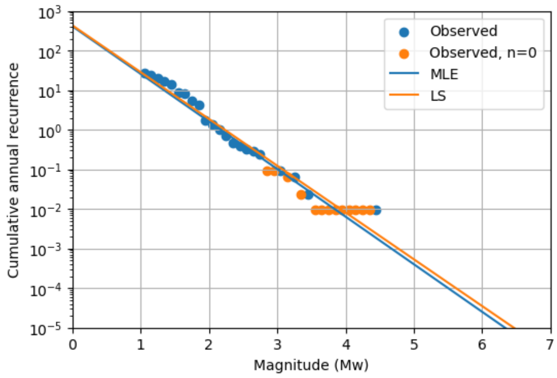
(b) SSA 2



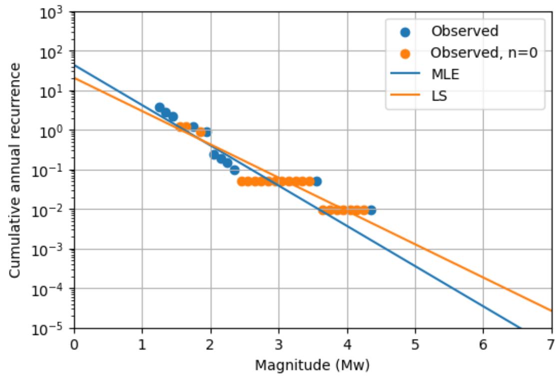
(c) SSA 3



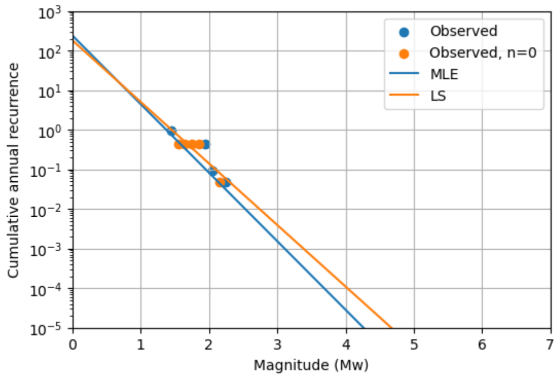
(d) SSA 4



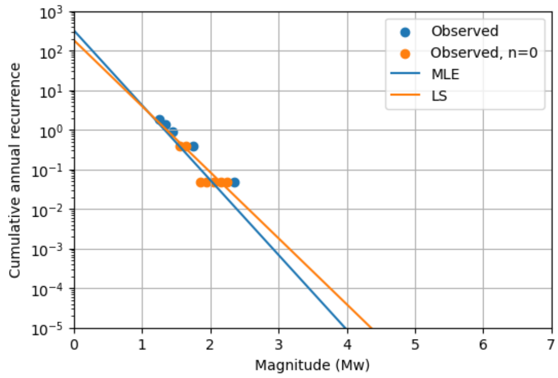
(e) SSA 5



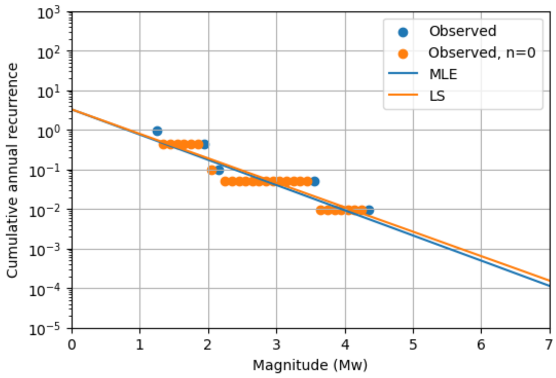
(f) SSA 6



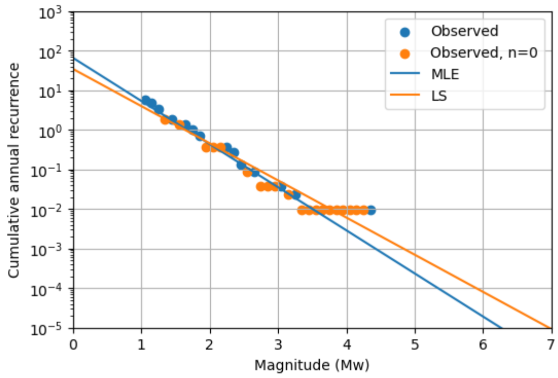
(g) SSA 6a



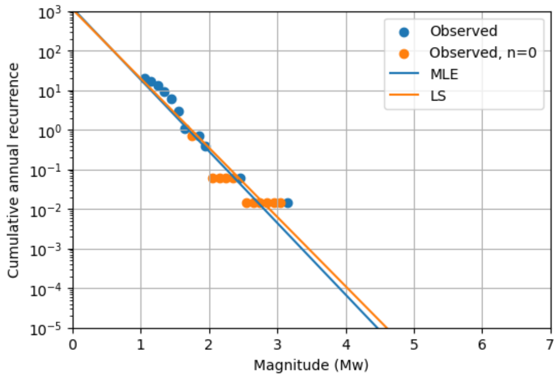
(h) SSA 6b



(i) SSA 6c



(j) SSA 8



(k) SSA 10

Figure C1: GR plots from each SSA based on the parameters presented in Table C1.