Sensitivity analysis of a seismic hazard assessment for a Finnish nuclear power plant

Vili Visakorpi

School of science

Thesis submitted for examination for the degree of Master of Science in Technology.
May 27, 2022

Supervisor

Prof. Ahti Salo

Advisor

M.Sc. Jukka Koskenranta
M.Sc. 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 master's thesis

Author Vili Visakorpi

Title Sensitivity analysis of a seismic hazard assessment for a Finnish nuclear power plant

Degree programme Master's programme in Engineering Physics

Major Engineering Physics **Code of major** SCI3056

Supervisor Prof. Ahti Salo

Advisor M.Sc. Jukka Koskenranta
M.Sc. Timo Leppänen

Date May 27, 2022 **Number of pages** vii+42+XIV **Language** English

Abstract

Probabilistic seismic hazard analysis (PSHA) seeks to determine the frequency of occurrence of seismic events where ground movement exceeds a given threshold. It is a common first step in calculating seismic risk and engineering earthquake-resistant structures. Due to many uncertain parameters and modelling options, an estimate for seismic hazard can only be attained with significant uncertainty. Assessment of the uncertainties is an important part of the analysis.

One of the necessary subjects to model is the magnitude-recurrence relationship of eartquake events. The relationship is usually expressed by Gutenberg-Richter law, which approximates the earthquake magnitudes to follow an exponential distribution. Its two parameters are determined empirically from a historic earthquake catalog.

Operating a nuclear power plant (NPP) requires that the seismic risk to the plant is assessed. A new PSHA was developed for Loviisa NPP in 2021, taking various different modelling decisions compared to the last PSHA from 2018. In addition to the Loviisa NPP, the model created for the 2021 PSHA was used to produce a hazard estimate for Olkiluoto NPP. The aim of this thesis is to assess how much the new decisions contribute to the estimated seismic hazards of the Loviisa and Olkiluoto NPPs.

The general procedure of PSHA is reviewed and a picture of the methods and decisions involved with Loviisa 2021 PSHA is formed. The recurrence estimation methods are discussed with focus on maximum likelihood estimation of the Gutenberg-Richter parameters. An open source hazard calculation software HAZ is adopted for the analysis. Sensitivity is assessed through the primary output of PSHA, the annual frequency of exceedance.

Keywords PSHA, Seismic Hazard, Nuclear safety, Gutenberg-Richter, Sensitivity analysis, Loviisa, Olkiluoto

Aalto University
School of Science

Aalto-yliopisto, PL 11000, 00076 AALTO
www.aalto.fi
Diplomityön tiivistelmä

Tekijä Vili Visakorpi

Työn nimi Sensitivity analysis of a seismic hazard assessment for a Finnish nuclear power plant

Koulutusohjelma Teknillisen fysiikan maisteriohjelma

Pääaine Teknillinen fysiikka **Pääaineen koodi** SCI3056

Työn valvoja Prof. Ahti Salo

Työn ohjaaja M.Sc. Jukka Koskenranta
M.Sc. Timo Leppänen

Päivämäärä May 27, 2022 **Sivumäärä** vii+42+XIV **Kieli** Englanti

Tiivistelmä

Todennäköisyysperusteinen seisminen hasardianalyysi (PSHA) pyrkii määrittämään esiintymistaajuuden sellaisille seismisille tapahtumille, joilla jokin maanliikettä kuvaavan suureen taso ylittyy. Sen yleisiä käyttökohteita ovat seismisen riskin arvioiminen sekä maanjäristyskestävien laitosten suunnitteleminen. Seismiseen hasardianalyysiin liittyy paljon epävarmuuden lähteitä mallintamisessa sekä parametrien määrittelemisessä. Hasardiarvion aikaansaaminen ilman merkittäviä epävarmuuksia onkin mahdotonta. Epävarmuuksien huomioiminen on oleellinen osa seismistä hasardianalyysiä.

Eräs tärkeistä mallintamisen kohteista on järistysten magnitudin ja toistuvuuden välinen suhde. Tavallinen tapa kuvata tätä suhdetta on Guternberg-Richter (GR) laki, joka arvioi järistysten magnitudin eksponentiaalisesti jakautuneeksi. Siihen sisältyvät kaksi parametria arvioidaan empiirisesti maanjäristysluettelon perusteella.

Ydinvoimalaitoksilta vaaditaan seismisen riskin huomioiminen laitoksen turvallisuuden arvioinnissa. Loviisan ydinvoimalalle kehitettiin vuonne 2021 uusi PSHA, johon sisältyi useita entiseen vuoden 2018 arvioon eroavia mallintamispäätöksiä. Loviisan lisäksi uuteen arvioon kehitettyä mallia käytettiin laskemaan hasardiarvio myös Olkiluodon ydinvoimalalle. Tässä opinnäytetyössä tutkitaan näiden päätösten vaikutusta Loviisan sekä Olkiluodon hasardiarvioihin.

PSHA:n yleinen menetelmä esitellään ja 2021 hasardiarvion menetelmiin ja mallintamispäätöksiin tutustutaan kokonaiskuvaa muodostaen. GR-parametrien arvioimista käsitellään keskittyen erityisesti suurimman uskottavuuden menetelmään. Avoimen lähdekoodin hasardilaskentasovellus HAZ otetaan käyttöön Loviisan seismiseen hasardianalyysiin. Tulosten herkkyyttä arvioidaan käyttäen vuosittaisen ylitystaajuuden muutosta.

Avainsanat Todennäköisyysperusteinen seisminen hasardianalyysi, Seisminen hasardi, Ydinturvallisuus, Herkkyysanalyysi, Loviisa, Olkiluoto

Preface

Despite a difficult start to the project and a number of setbacks with the original plans, this thesis is to be finished about on time. I would like to express gratitude to my supervisor Professor Ahti Salo for making this possible. This could not happen without the quick responses I got to my messages and the helpful meetings. Thank you Jukka Koskenranta, Päivi Mäntyniemi and Marianne Malm for the valuable comments on this thesis. I would also like to thank the Fortum and TVO seismic hazard project group and the entire PRA group for the support and for providing me this opportunity.

Espoo, May 27, 2022

Vili M. Visakorpi

Contents

Symbols and Abbreviations

Symbols

$n(m)$	Earthquake recurrence as a function of magnitude m in the Gutenberg-Richter equation $\lg(n(m)) = a - bm$
b, β	b-value in the Gutenberg-Richter equation, $\beta = \ln(10)\, b$
a, α	Recurrence parameter in the Gutenberg-Richter equation, $\alpha = \ln(10)\, a$
$n_{\min}$	Recurrence calculated at minium magnitude
$f_X(x)$, $F(X \leq x)$	Probability density function, cumulative distribution function
$P(X > x)$	Probability of exceedance

Abbreviations

AFE	Annual frequency of exceedance
CDF	Cumulative distribution function
DBE	Design basis earthquake
GMPE	Ground movement prediction equation
GR	Gutenberg-Richter
GRS	Ground response spectrum
IAEA	International atomic energy agency
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
YVL	Finnish regulatory guide on nuclear safety

1 Introduction

In Finland, the operation of a nuclear power plant (NPP) requires that the external hazards that could result in spreading of radioactive materials are assessed. Radioactive safety rules are determined and overseen by Radiation and Nuclear Safety Authority (STUK) under the Finnish law, including general and plant-specific instructions. The regulatory guide on nuclear safety B.7 (YVL B.7) mandates taking into account seismic hazard in the design and risk analysis of NPPs [1]. A design basis earthquake (DBE) and a corresponding ground response spectrum (GRS) shall be developed to assess ground movement during earthquakes.

A DBE is described with a ground motion characteristic that is anticipated to be exceeded less frequently than a predetermined rate. For Finnish NPPs, YVL B.7 requires the developement of a DBE with frequency less than once in hundred thousand years (10^{-5} 1/a). A GRS is the response of a damped harmonic oscillator to ground movement as a function of the natural frequency of the oscillator (see e.g. [2]). The response is commonly given as spectral acceleration, which is the maximum acceleration the oscillator experiences with selected frequency and damping. The methodology for producing the DBE and GRS is currently not specified by YVL B.7, albeit it is required to be well-grounded. One approach is probabilistic seismic hazard analysis (PSHA).

In the literature, PSHA is often divided into four or five steps [3, 4]. The first step is seismotectonic study, where seismic sources are defined. The sources can be specifically determined faults or general areas where the seismicity is assumed to be uniform. These areas are known as seismic source areas (SSAs). The second step is the assessment of earthquake recurrence with appropriate parameters: the magnitude-recurrence relationships and maximum magnitudes for each source. The dependency of magnitude on the frequency of occurrence is commonly estimated via Gutenberg-Righter (GR) law. The third step is the definition of ground motion attenuation from the source to the site. The functions describing this attenuation are known as ground motion prediction equations (GMPEs). The fourth step is the integration of hazard from different seismic sources to the site. International atomic energy agency (IAEA) adds a fifth step where on-site ground movement is studied [4], although this is not necessary for constructions founded on bedrock.

Each step of PSHA involves a mixture of empiric data and expert modelling decisions, which carry uncertainty. This uncertainty related to lack of knowlege is known as epistemic uncertainty. On the other hand, the unpredictability of outcome of a probabilistic model is known as aleatory uncertainty. Studying these uncertainties is a major part of contemporary PSHA [4–6]. The recommended technique to propagate epistemic uncertanty from the different steps is by the developement of a so called logic tree [3, 4, 6].

Empirical determination of recurrence parameters relies on a seismic catalog with complete listing of earthquakes and their magnitudes within a timeframe for the area of interest. Open earthquake catalog FENCAT is one of the most comprehensive lists of seismic events in the northern Europe region. While the first listed event is from 1375 [7], instrumental data becomes available only after mid 20th century. This makes the useful data shorter for low-magnitude events which are rarely felt without instrumentation. The sensitivity and span of the seismic measurement network has continued to improve to this day, hence assessment of the completeness of the data becomes important for recurrence estimation. FENCAT provides the events with various magnitude scales which should be homogenized for recurrence estimation, normally to moment magnitude [4]. Probabilistic hazard studies also usually require the removal of aftershocks and induced earthquakes to satisfy the assumption of independent occurrences.

A Finnish nuclear company Fortum commissioned the developement of a new PSHA by a geotechnical consulting company for Loviisa NPP. The project was joined by TVO and the same model was used to estimate seismic hazard for Olkiluoto NPP. The assessment covers the steps two to four of PSHA, leaning on a previous study by the University of Helsinki [8] for the seismic source characterization. Various decisions and updates to previous estimates create the need for a sensitivity analysis on the most significant features of the assessment.

Some sensitivity analysis regarding the modelling choices was already conducted within the new PSHA. In addition, preliminary results of an ongoing seismic hazard sensitivity study by STUK [9] have been provided to Fortum. It has been recognized that Finnish PSHA is sensitive to the GR parameters. Coincidentally the parameters have had significant changes in the new report due to both dataset and estimation method differences. This thesis seeks to assess the influence of selected modelling

decisions on the estimated hazard with focus on recurrence estimation. The addition of new earthquake data to the estimation is also investigated.

The integration of seismic hazard to NPP site is achieved with a modified version of an open-source program HAZ by Norman Abrahamson. The original source code written in Fortran is available on Github [10]. The modifications include a new magnitude-recurrence relation and additional GMPEs and earthquake depth distributions. This version was provided to Fortum as a compiled program by the geotechnical consultants. HAZ outputs annual frequency of exceedance (AFE) as a function of spectral acceleration for all variations of the input parameters. Further processing such as calculating confidence bounds is done with Excel. Sensitivity analysis is performed for Loviisa NPP and repeated for Olkiluoto NPP. All given magnitudes are in moment magnitude scale.

2 Theoretical background

2.1 Probabilistic seismic hazard analysis

The roots of PSHA lie in a 1968s seminal paper by Cornell [11]. Guides, books and descriptions of modern PSHA implementations have been published by multiple sources, including nuclear authorities [3, 4, 6, 12, 13]. Naturally PSHA has also been a target of criticism regarding its ability to predict strong ground motions [14]. While progress has been made with its methods, the basic formula presented by Cornell remains.

Generally, PSHA for a specific site seeks to find the rate $\nu(c)$ with which a seismic characteristic C exceeds some value c at the location. This can be turned into probability measure with an assumption of Poisson distribution of events. The probability of one or more events occurring within time t with C exceeding c is

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

The rate $\nu(c)$ is often given as events per year, ergo the name annual frequency of exceedance. The inverse of this is known as return period. When small, the probability of exceedance is close to $\nu(c)t$, hence the annual frequency and probability are sometimes used interchangeably [6]. Different used characteristics include spectral acceleration, peak ground velocity, fault displacement and others. Plots of $\nu(c)$ as a function of the characteristic in question are known as hazard curves.

The research starts with the determination of seismic sources. The sources can consist of areas of uniform recurrence or specifically defined faults. Since each source can be assessed separately, PSHA gives the researcher a lot of freedom in the modelling. The choice of source type is then based on the amount and type of information available on the seismicity. After this step earthquake recurrence and ground attenuation can be studied.

2.1.1 Earthquake recurrence

The widely known Gutenberg-Richter (GR) law relates the number of earthquake occurrences to magnitude:

$$\lg\left(n(m)\right) = a - bm\,. \tag{2}$$

Here $n(m)$ is predicted number of earthquake events with magnitudes greater-than or equal to m and a and b are parameters. The notation lg() stands for logarithm with base 10. The intercept a is known as the recurrence parameter, as it determines the number of earthquake events within the observation period. The slope parameter b, called b-value, describes the distribution of magnitudes.

Many studies have sought to find a physical meaning for the b-values, and some have proposed its universality with value close to 1 [15–17]. Conclusions about a physical meaning have not been achieved and significant deviations of b-values have been observed [15–17]. For PSHA it is often preferred to fit the b-value to local earthquake data.To this aim, one needs a catalog containing earthquakes of a certain area or fault, classified by magnitudes. Various different methods are used to estimate the parameters, frequently prioritizing simplicity instead of accuracy.

The GR equation is often given with exponential base

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

where $\beta = \ln(10)\ b$ and $\alpha = \ln(10)\ a$. Maximum likelihood estimation (MLE) of β has a long history. The often cited first estimate published in 1965 by Aki is [18]

$$\frac{1}{\beta} = \bar{m} - m_0\,.$$

Here $\bar{m}$ is the mean observed magnitude and m_0 the lowest observed magnitude. This approach has two severe limitations: the magnitudes are assumed to come from a continuous distribution and there is only a single observation period for all seismic activity. In practice, most datasets divide earthquakes into magnitude intervals omitting precise magnitude values. In addition, these magnitude bins often have unequal observation periods due to improvement in seismic measurements. Despite

its inaccuracies, the method has remained popular in practical applications [16].

The MLE estimate of b with magnitude bins with unequal observation periods was first presented by Weichert in 1980 [19]. This estimate is limited to small, equally large magnitude intervals and the variance of a assumes known b-value. The former issue is easy to ameliorate, and the approximation remains good up to ~ 0.5 magnitude interval width [13, 20]. The latter error is likely a result of seismic studies focusing on the magnitude distribution $e^{-\beta m}$, omitting recurrence rate and the parameter a [20]. On the other hand, assessing hazard requires both a and b to be estimated simultaneously. The joint estimation of recurrence parameters by developing covariance matrices are presented in derivations by Kijko and Sellevoll [21] and by Johnston [22].

Some general results of maximum likelihood estimation of poisson-variates can be found in statistics textbooks, eg. [23]. Here is presented a MLE method for producing GR parameter estimates which is close to that by Weichert [19]. A more generalized result is given in Appendix A with close resemblance to derivation in [20]. Equation 3 implies that the number of events between magnitudes $m_i - \frac{1}{2}\delta m$ and $m_i + \frac{1}{2}\delta m$ is

$$\delta n_i = \left[e^{\alpha-\beta m}\right]\Big|_{m=m_i+\frac{1}{2}\delta m}^{m_i-\frac{1}{2}\delta m} \tag{4}$$

$$= \left(e^{\beta\frac{1}{2}\delta m} - e^{-\beta\frac{1}{2}\delta m}\right)e^{\alpha-\beta m_i} \tag{5}$$

$$= 2\sinh\left(\beta\frac{1}{2}\delta m\right)e^{\alpha-\beta m_i}\,. \tag{6}$$

If $\beta\delta m_i/2$ is sufficiently small equation (6) can be approximated by $\delta m_i \beta e^{\alpha-\beta m_i}$. Let us assume earthquakes recur following a Poisson process with rate parameter $\delta n_i t$. The probability of observing n events within magnitude interval $m_i - \frac{1}{2}\delta m$ to $m_i + \frac{1}{2}\delta m$ in time period t_i is

$$P_i(n) = \frac{(\delta m_i \beta e^{\alpha-\beta m_i} t_i)^n}{n!} e^{-\delta m_i \beta e^{\alpha-\beta m_i} t_i}\,. \tag{7}$$

Assuming independent observations, the likelihood of observing n_i observations for each magnitude bin $i \in [1,I]$ with completeness periods t_i is

$$\mathcal{L} = \prod_{i=1}^{I} P_i(n_i)\,. \tag{8}$$

Following a common practice of MLE, the likelihood function is maximized by maximizing its logarithm

$$\ln \mathcal{L} = \sum_{i=1}^{I} \left[n_i(\alpha - \beta m_i + \ln(\delta m_i \beta t_i)) - \log(n_i!) - \delta m_i \beta t_i e^{\alpha - \beta m_i} \right] . \tag{9}$$

To maximize this, it suffices to find the two roots

$$\frac{\partial \ln \mathcal{L}}{\partial \alpha} = \sum_{i=1}^{I} \left[n_i - \delta m_i \beta t_i e^{\alpha - \beta m_i} \right] = 0 \, , \tag{10}$$

$$\frac{\partial \ln \mathcal{L}}{\partial \beta} = \sum_{i=1}^{I} \left[n_i(-m_i + \frac{1}{\beta}) - \delta m_i t_i e^{\alpha - \beta m_i}(1 - m_i \beta) \right] = 0 \, . \tag{11}$$

Defining the total observed earthquakes $\sum_{i=1}^{I} n_i = N$ equation (10) gives

$$e^{\alpha} = \frac{N}{\beta \sum_{i=1}^{I} (\delta m_i t_i e^{-\beta m_i})} \, .$$

Inserting into (11) gives

$$\frac{N}{\beta} - \sum_{i=1}^{I}(n_i m_i) - \sum_{i=1}^{I} \left(\delta m_i t_i e^{-\beta m_i}(1 - m_i \beta) \right) \frac{N}{\beta \sum_{i=1}^{I} (\delta m_i t_i e^{-\beta m_i})} = 0$$

$$\sum_{i=1}^{I}(n_i m_i) - N \frac{\sum_{i=1}^{I} (m_i \delta m_i t_i e^{-\beta m_i})}{\sum_{i=1}^{I} (\delta m_i t_i e^{-\beta m_i})} = 0 \quad .$$

The last equation can be solved for β numerically. If the magnitude intervals are equal in size, Weichert's original result is redeemed as

$$\frac{\sum_{i=1}^{I} (m_i t_i e^{-\beta m_i})}{\sum_{i=1}^{I} (t_i e^{-\beta m_i})} = \frac{\sum_{i=1}^{I} (n_i m_i)}{N} \, . \tag{12}$$

For a sufficiently large number of events N, GR parameter uncertainty can be assessed by inverting Fisher-information matrix to acquire the covariance matrix for a and b-value, a method familiar in maximum likelihood estimation. A distribution of the predicted number of earthquake events can then be sampled from one-dimensional normal distribution [20]. The covariance matrix of the GR parameter estimates is given in appendix A.

For ease of computation, least-squares (LS) regression is often used in place of MLE. At its simplest, the number of earthquakes is given in incremental or cumulative

form on magnitude axis, and a LS line is fitted on the logarithmic scale, yielding the GR parameters. To account for unequal observation periods, the incremental number of earthquakes is divided by time, resulting in earthquake rates for each magnitude interval. To change to cumulative plots, it is these rates that can be summed. Let $n_{m \geq m_{\text{low}}}$ denote the earthquake recurrence rate above magnitude m_{low} and δn_i the rate for magnitude bin i. By using the independence of occurrences between magnitude bins, the expectation of $n_{m \geq m_{\text{low}}}$ can be written as

$$\mathbb{E}[n_{m \geq m_{\text{low}}}] = \sum_i \mathbb{E}[\delta n_i] \ , \ \ \{i \in I \mid m_{\text{low}} \leq m_i \leq m_{\text{max}}\}$$

While the estimates by the LS method can be close to MLE with some datasets [19], the statistical assumptions behind LS are commonly violated and many suggest against using it [13, 19, 20, 24]. Its inability to consider magnitude bins with $n = 0$ was already noted by Weichert [19], while the logarithmic transformation and the nonindependent observations add to the inaccuracy of the method [13, 24]. Importantly, the variability from these issues is not covered by the common LS uncertainty estimates, resulting in underestimated confidence intervals [24].

Several modifications to the GR relation exist [13], such as characteristic magnitude distribution [13], separate b-values for main and aftershocks [15, 17] and quadratic magnitude-frequency law [25]. These modifications seek to better describe non-linear earthquake data, but with more complicated models, more data is needed to determine their applicability. The assumption of Poisson process prohibits the inclusion of fore- and aftershocks in the earthquake dataset, however after excluding these from the data the assumption is the most appropriate one to make [11].

2.1.2 Minimum and maximum magnitudes

In seismic hazard studies the GR law is often dealt like a magnitude distribution, which necessitates the determination of a minimum magnitude, below which the integration is stopped. The choice of minimum magnitude is not however purely a computational issue, it is also an important parameter when the results are used for seismic risk [26]. As a consequence, decreasing the minimum magnitude of a PSHA can still change the shape of the hazard curves [26]. It is recommended by IAEA that

the omitted magnitudes should not contribute risk and that the minimum magnitude should not exceed the value 5 [4]. Including a minimum magnitude requires no modifications to the GR law, but often the equations are written with a shifted magnitude $m \to m - m_{\min}$, so that

$$n_{\min} = e^{\alpha - \beta m_{\min}} \tag{13}$$

$$n(m) = n_{\min} e^{-\beta(m - m_{\min})} \,. \tag{14}$$

While GR law predicts the occurrence of arbitrarily large magnitudes, seismic studies and physical reasoning suggest otherwise [27]. Seismic hazard studies therefore incorporate the concept of a maximum magnitude. Assessing the maximum magnitude is a difficult task. Various estimation methods exist, including physical studies on the faults and earthquake history based methods [22, 28, 29]. A physical study has been attempted for the Finnish area in a thesis work resulting in maximum magnitude 7, although it is noted that the uncertainties remain large [28].

The standard way of incorporating maximum magnitude is by truncation of the GR law

$$n(m) = n_{\min} \frac{e^{-\beta(m - m_{\min})} - e^{-\beta(m_{\max} - m_{\min})}}{1 - e^{-\beta(m_{\max} - m_{\min})}} \,, \quad m_{\min} \le m \le m_{\max} \,. \tag{15}$$

The shift in the numerator can be understood as removing the contribution of the tail $m \ge m_{\max}$ from $n(m)$. This leaves the MLE estimates unchanged, as seen from equation (A2). The scaling by $1 - e^{-\beta(m_{\max} - m_{\min})}$ is necessary to normalize the value at $m_{\min}$ so that $n(m_{\min}) = n_{\min}$. It can be included into the definition of $n_{\min}$ through change of variables, resulting in no error to the estimated parameters, but deviating the estimated variance of a slightly. In the literature these estimates are often taken to be equivalent [13, 20].

2.1.3 Ground motion prediction equations

The relationship of magnitude and distance to ground movement is described by GMPEs. A large number of GMPEs have been developed to match the different geologies and available seismic data around the world. The sparsity of strong ground movement data brings challenges to choosing the form of the function and

its parameters. Developing GMPEs for Finland is a continuing task that has seen much recent advancements [30–33].

For PSHA the ground motion is captured by a probability of exceedance at the site

$$P_\omega(C > c \mid m, r)\,, \tag{16}$$

with the given earthquake magnitude m and distance r. The parameter ω denotes the dependency on frequency. Commonly a log-normal distribution is used, so that the GMPE function is the mean of the logarithmic characteristic $\log c$ [3, 6]. While this uncertainty is treated as aleatory and integrated over in the hazard calculations, GMPEs are also an important source of epistemic uncertainty [5]. The common approach is to include multiple GMPEs in the logic tree approach [3, 34].

2.1.4 Integration of hazard

The integration step of PSHA can be performed in multitudes of ways using law of total probability. Generally one seeks to add up the hazard of all sources, integrating over the source volume and magnitude.

The cumulative magnitude distribution is conventionally defined as the proportion of magnitudes below the value m, which is the reverse of the convention with the GR law. The truncated GR equation (15) implies Poisson distribution of magnitudes with probability density

$$f_M(m) = \beta\, \frac{e^{-\beta(m-m_{\min})}}{1 - e^{-\beta(m_{\max}-m_{\min})}}\,. \tag{17}$$

The cumulative form and the truncated GR equation become:

$$F(M \le m) = \int_{m_{\min}}^{m} f_M(M)\, \mathrm{d}M \tag{18}$$

$$n(m) = n_{\min}\Big(1 - F(M \le m)\Big)\,. \tag{19}$$

The seismic sources were defined at the first step of PSHA. Each source will be integrated over with the appropriate dimensionality depending on the type of the

source. For SSAs, the integration should consider the depth distribution in addition to the surface area. Given the minimum depth $d_{\min}$ and the threshold value d, the CDF of depth will be

$$F(D \leq d) = \int_{d_{\min}}^{d} f_D(D)\ \mathrm{d}D\,. \tag{20}$$

Options for the depth density function $p_D(d)$ include triangular, normal and uniform distributions. The source area integral is

$$A = \int_{x_1}^{x_2} \int_{y_1}^{y_2} \mathrm{d}y\,\mathrm{d}x\,. \tag{21}$$

Let (x_0, y_0, d_0) denote the coordinates of the NPP or other location of interest, and let r denote the distance

$$r = \sqrt{(x - x_0)^2 + (y - y_0)^2 + (d - d_0)^2}\,.$$

Consider a set of sources S, which consists solely of SSAs. The frequency of exceedance for the location is

$$\nu(c) = \sum_{S} \int_{m_{\min}}^{m_{\max}} \int_{x_1}^{x_2} \int_{y_1}^{y_2} \int_{d_{\min}}^{d_{\max}} \frac{n_{\min}}{A}\ f_D(d)\ f_M(m)\ P_\omega(C > c \mid m, r)\ \mathrm{d}d\,\mathrm{d}y\,\mathrm{d}x\,\mathrm{d}m\,. \tag{22}$$

The integration bounds, the recurrence per area $\frac{n_{\min}}{A}$ and the density functions $f_D(d)$ and $f_M(m)$ are assessed separately for each source. In real studies, the integrals in eq. (22) are calculated numerically and continuous definitions may not even be considered. The total output will be the $\nu(c)$ calculated for a number of threshold values c and frequencies.

Each parameter in equation (22) is assessed as a part of the PSHA, therefore carrying epistemic uncertainty. A logic tree approach was introduced by Kulkarni et al. in 1984 to take the epistemic uncertainty into consideration [5, 35]. Each node of the tree corresponds to an uncertain parameter. Multiple values of the uncertain parameters are picked and a branch is made for each. The branches are associated with weights that commonly sum to unity. If a parameter is dependent on another, the dependent parameter is placed later in the tree.

Hazard is calculated for every leaf node of the tree, picking the parameters from the path to the leaf. Total weights of the paths are calculated as the product of the

associated weights. Rather than summing over the sources like in equation (22), the sources are dealt with separately.

The mean hazard and fractiles can be calculated by either calculating the hazard through each path or by random sampling of the tree with a Monte Carlo method. In the former method, the mean is attained by weighting the results with the corresponding total weights and summing them together. Fractiles can be calculated by sorting the total hazard results and finding the fractile from the cumulative sum of the total weights. When adding up the contribution of each source, every combination of the results of different sources has to be considered. The Monte Carlo method becomes useful for large trees, but may cause slow convergence as a new problem [12].

The weighting scheme makes it possible to consider expert opinions and epistemic uncertainty without specific requirements on the distribution. However it is recognized by some that if the weights represent subjective merits of the different models rather than valid probabilities, the calculated mean result does not necessarily correspond to the expected value [36]. When a distribution of error can be assumed it can be considered with a mathematically founded approximation of the associated distribution. For instance, the epistemic uncertainty of the GR parameter estimation can be reduced into a one-dimensional normal distribution [20]. The branch weights are chosen to describe the shape of the distribution e.g. with the method introduced by Miller and Rice [37].

3 Setup for the sensitivity analysis

3.1 Loviisa PSHA

This thesis focuses on a 2021 PSHA by Slate geotechnological consultants [38], which was developed following some remarks on the previous 2018 PSHA by AFRY (former ÅF) and Fortum [39–41]. The assessments are founded on a comprehensive seismotectonic study by Korja et al. (2016), which compiles works by the University of Helsinki, Uppsala University, Geological Survey of Finland, Geological Survey of Sweden and Geological Survey of Estonia [8]. The logic tree used in the 2021 PSHA is depicted in Figure 1. The parameters and weights are coded directly into a HAZ input file. It was found however that changing the parameters seismogenic thickness and fault plane dip in the HAZ input do not affect the results. A separate tree was used to assess various GMPEs producing a single median estimate.

The SSAs come directly from the report by Korja et al.. Only those that reside within a 300 km radius of the NPP are considered in the calculations. Two models were adopted as branches to the logic tree, with the only difference of a subdivision of area 6 into three smaller areas 6a, 6b and 6c with individually assessed earthquake recurrence. The Loviisa NPP itself situates on area 10, while 6 is in close proximity westward from the location. Areas 1–5 and 9 are not included in Loviisa PSHA. Areas 7 and 11 are within the study area but have had virtually no seismicity. This has lead to area 6 recurrence being used as a conservative estimate for the two.

The earthquake catalog used in the report was compiled from the previously used by Korja et al., with additional earthquakes from FENCAT to cover years 2012–2014. The catalog was processed further to contain moment magnitudes for all events, to remove clusters and to consider completeness. Earthquake cluster removal was done with the Reasenberg algoritm [42]. The original catalog already sought to remove events that do not result from seismicity, such as ice quakes and explosions. Completeness was evaluated for magnitude intervals with a visual inspection of the occurrences. For this work, the starting point is chosen to be the first day of the year of the first observed event. The earthquake counts and completeness periods are given for each SSA in Table 1.

Mean GR parameter estimates are obtained via the ordinary LS method, fitting a line to the cumulative earthquake rates located at the centers of the magnitude intervals. Covariance matrix elements of the two estimates are calculated with usual methods. The epistemic uncertainty is assessed with a single node in the logic tree with pairs of GR parameters. The two-sided 90 % confidence bounds used in the 2021 PSHA are developed by assuming a normal distribution of error of the parameter a. Denoting the mean estimate with $\hat{a}$ and the estimated variance with $\mathrm{var}(a)$, the 90 % confidence interval is

$$\left(\hat{a} - 1.65\,\sqrt{\mathrm{var}(a)},\ \hat{a} + 1.65\,\sqrt{\mathrm{var}(a)}\right). \tag{23}$$

The b-values are then picked by considering its covariance with a. Denoting the mean estimate with $\hat{b}$ and the covariance with $\mathrm{cov}(a,b)$, the values corresponding to the parameter a confidence bounds are given by

$$\left(\hat{b} - 1.65\,\frac{\mathrm{cov}(a,b)}{\sqrt{\mathrm{var}(a)}},\ \hat{b} + 1.65\,\frac{\mathrm{cov}(a,b)}{\sqrt{\mathrm{var}(a)}}\right). \tag{24}$$

The recurrence values were recalculated for this thesis because precise information on the dataset completeness times used in the 2021 PSHA was unavailable. The earthquake events were also recounted from the catalog. The estimated GR parameters are given in appendix B. As a novel modelling decision to the 2021 PSHA, the estimated b-values are only used up to magnitude 5.75. Beyond this the mean b-value is set to 2 and the confidence bounds are repicked from (24).

Mean hazard curves for each SSA and the total are presented in Figure 2. The most significant part of the hazard can be seen to come from areas 6 and 10. The large contribution of area 6 is explained by the comparatively large seismic activity within area 6c and the close proximity of area 6b to Loviisa.

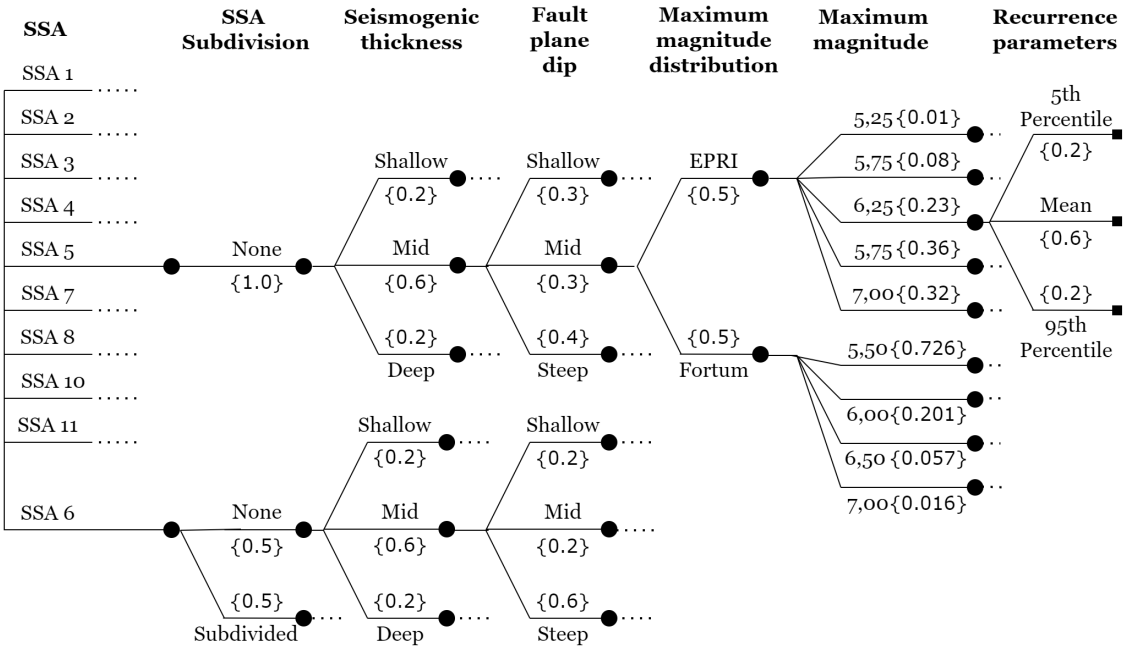


Figure 1: The logic tree used in the study. Branch weights are included in braces. Note that Loviisa PSHA excludes the areas 1–5 from the study.

Bin	**1–1.5**	**1.5–2**	**2–2.5**	**2.5–3**	**3–3,5**	**3.5–4**	**4–4.5**	**4.5–5**
End	31.12.2014							
Start	1.1.2013	1.1.2012	1.1.1994		1.1.1944		1.1.1909	
SSA 1	3	1	8	2	2	2	0	0
SSA 2	9	8	16	1	1	0	0	0
SSA 3	3	6	16	1	0	0	0	0
SSA 4	0	6	7	1	2	0	0	0
SSA 5	37	22	22	5	6	0	1	0
SSA 6a	1	1	2	0	0	0	0	0
SSA 6b	3	1	1	0	0	0	0	0
SSA 6c	1	1	1	0	0	3	1	0
SSA 8	9	3	6	1	2	0	1	0
SSA 10	34	9	1	0	1	0	0	0

Table 1: Earthquake counts for each SSA for each magnitude bin. Completeness period end and start for each magnitude interval. Areas 7 and 11 are excluded since their GR parameters were not determined from historical seismicity. Area 9 was not included in the study.

3.2 Olkiluoto PSHA

The methods and data of the Loviisa 2021 PSHA were also used to produce a PSHA for Olkiluoto NPP. For this thesis, the sensitivity studies selected for Loviisa are duplicated for Olkiluoto. The Olkiluoto results are given as relative changes only, and the absolute values of the hazard estimates are omitted. The locational difference means that Olkiluoto model uses all of the SSAs 1–11 excluding 9. Situated in area 6a, the contribution of the different areas changes. Area 10 no longer has a high contribution to the hazard and area 6 becomes dominant.

3.3 Choice of parameters and presentation

The use of a different b-value for high magnitudes in the 2021 PSHA is a novel approach, hence its effect on the estimated hazard is chosen to be presented in this thesis. This is achieved by comparing the calculated hazard curve to one with the basic truncated GR equation in use with parameters estimated from the data. Another point of interest is the contribution of the two SSA models, other with the undivided area 6 and the other with the subdivided area 6. The two models were included in the 2021 PSHA with equal weighting. Their contribution to the estimated hazard is studied by fully weighting each one in separate hazard calculations.

The GR parameters were estimated with an LS method instead of the appropriate MLE. The recurrence parameters are re-estimated from the original data by using the MLE method presented in this thesis. While the accurate confidence interval estimation produces recurrence bounds for each magnitude, HAZ accounts for uncertainty only by considering pairs of b-value and recurrence at minimum magnitude $n_{\min}$. The 90 % confidence bounds are included by inserting the accurate recurrence bounds at minimum magnitude and the unconditional estimates for b-value:

$$\left(\hat{b} + 1.65\,\sqrt{\mathrm{var}(b)},\ \hat{b} - 1.65\,\sqrt{\mathrm{var}(b)}\right), \tag{25}$$

where $\hat{b}$ is the mean estimate and $\mathrm{var}(b)$ the estimated variance.

Some additional parameter variations are explored with the MLE method to form a

picture of its behaviour compared to the LS. Finally, the datapoints used to estimate the GR parameters were set at the centers of the magnitude intervals. The points are moved to the lower bounds of the intervals and GR estimates are recalculated with the LS method.

The the mean hazard curves plotted for each SSA showed that a significant portion of the total hazard of Loviisa NPP comes from areas 6 and 10. For the sensitivity analyses that involve the recurrence parameters, these two areas are focused on. As discussed previously, changes to the recurrence of area 6 also applies to areas 7 and 11, making it even more important to the total hazard.

One of the most significant differences to the previous PSHA was the completeness evaluation of the earthquake dataset. As a result, a large part of the available seismic history was omitted. The effect of the completeness evaluation for high magnitudes is studied by adding back the removed history for magnitudes 4–4.5. The data is retrieved from FENCAT and individual earthquakes are picked according to the dates of omitted events in the Loviisa PSHA report. The observation periods of magnitude bins 4–4.5 and higher were increased down to the first observed event.

The earthquake data for the lowest magnitudes spanned two years only, while new data is constantly being updated to the FENCAT catalog. Despite being far from the magnitudes relevant to structural integrity, it may considerably affect the hazard through GR parameter estimation when the higher magnitude data is sparse. This motivates the study of the sensitivity to the addition of new low-magnitude data. A Monte Carlo method was implemented to assess the variation of the GR parameters with the addition of data. The new data was acquired from FENCAT search tool [7] for years 2015–2021. The involved methods are discussed in section 3.4.

It is conceivable that the model is sensitive to new events in the high-magnitude end, especially in areas 10 and 6. This is studied for area 10 by adding an event to the highest magnitude interval with a nonzero event count. It is of particular interest to see how the GR parameter estimation behaves when an event is added to the empty magnitude interval 2.5–3. Two studies are made, one with an event added to this interval and one with an additional event in the interval 3–3.5 where the largest event was observed.

To evaluate the importance of parameter variations, point estimates of hazard (hereinafter called hazard estimates) were chosen. Following the formulation of PSHA, AFE was used as the target variable at fixed values of frequency and spectral acceleration. Loviisa seismic risk analysis and DBE use the mean hazard curves, so the target AFE is also picked from the mean curve for the purpose at hand.

Regarding the choice of frequencies two things were considered. The most important frequencies for hazard are those near the acceleration peak in the different GRS. For Loviisa, this is around 10 Hz–15 Hz. Regulatory guides are also interested in the ground movement where the oscillator is rigid, known as peak ground acceleration [1]. In the previous Loviisa PSHA this has been defined as the 100 Hz curve. Two values of frequency were chosen for presentation: 10 Hz and 100 Hz. After choosing the frequency, the spectral acceleration which corresponds to AFE closest to the value 10^{-5} 1/a was picked from the mean hazard curve. The chosen frequencies and acceleration levels with the corresponding mean AFE are given in Table 2.

In addition to the hazard estimates, hazard curves including fractals are presented for some parameter variations. For this study, the fractile calculation was performed in a simplified way, where the seismic model is assumed to be the same for every SSA. This means that instead of calculating combinations of the different SSAs results, they were summed and the total weights were calculated as for a single SSA. The results for the fractals, namely the median and the 90 % confidence interval, should be taken with some reservation.

Table 2: Two point estimates are picked from hazard curves to make comparisons between the parameter variations. The chosen frequency and spectral acceleration with corresponding annual frequency of exceedance are tabulated here for both estimates. The Olkiluoto AFE results are omitted.

	Frequency	**Acceleration**	**AFE**
Loviisa	10 Hz	0.1 g	$2.4 \cdot 10^{-5}$ 1/a
	100 Hz	0.05 g	$1.32 \cdot 10^{-5}$ 1/a
Olkiluoto	10 Hz	0.25 g	
	100 Hz	0.1 g	

3.4 New earthquake data generation

The yearly numbers of earthquakes of the acquired FENCAT dataset are given in Table 3 for areas 6 and 10 and for 6 subareas in Table 4. Original data is presented down to year 2012 for comparison. The events included in the study were considered up to magnitude 2.5, as the return periods for higher magnitudes become much larger than the timeframe. One event with magnitude 2.5–3 was found in the new data and no larger events. A mean yearly occurrence rate was calculated for each bin taking into account both the acquired data and the original data included in the study. These are also given in the forementioned tables.

There were uncertainties in the new data that were deemed too large for the scope of this thesis. The acquired data is not declustered or checked for independence of events in any way, which may lead to unusually high event counts and increased variance. The magnitude scales used in the dataset vary between events, and in many cases a local magnitude measure is used. Given the small magnitudes of this study, the differences between the magnitude scales may not be large [13]. The verification of the data remains a subject for further study.

For the purpose of sensitivity analysis of the GR parameters, it made sense to generate a distribution of new data, using the full dataset to assess the distribution parameters. It was decided to force the generated yearly counts to follow the Poisson distribution, which provides the amount of variance that is expected for ideal data. Furthermore, the effect of nonindependent event occurrence within the data calls for a separate study. Using a random generator, 300 Poisson distributed earthquake counts were generated for each magnitude bin for each SSA. The area 6 counts were calculated as a sum from the subareas counts.

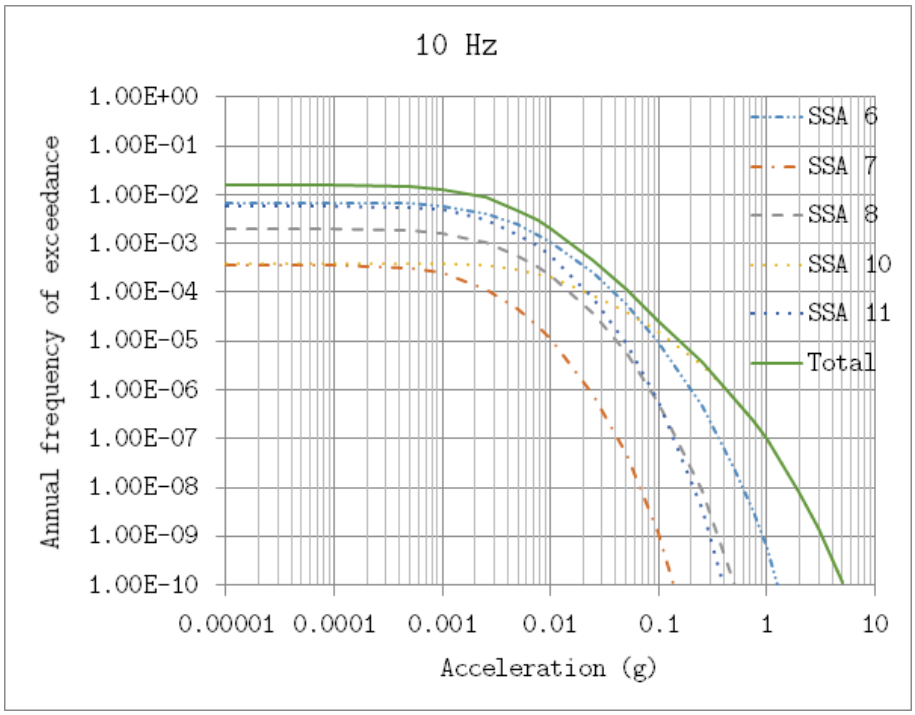
Three studies were made with one, three and six years of the generated data added to the GR parameter calculations. Magnitude bins 0.5–1 and lower were omitted to match the original study. The parameter pairs were plotted in scatter plots for each SSA. Hazard calculations were performed for six pairs from each study to sample the change in hazard by the new data.

Table 3: Earthquake event counts between year 2012 and end of 2021. Counts of years 2012–2014 are from the original dataset, and are presented here for comparison. Mean rates calculated from the full dataset are given after the data.

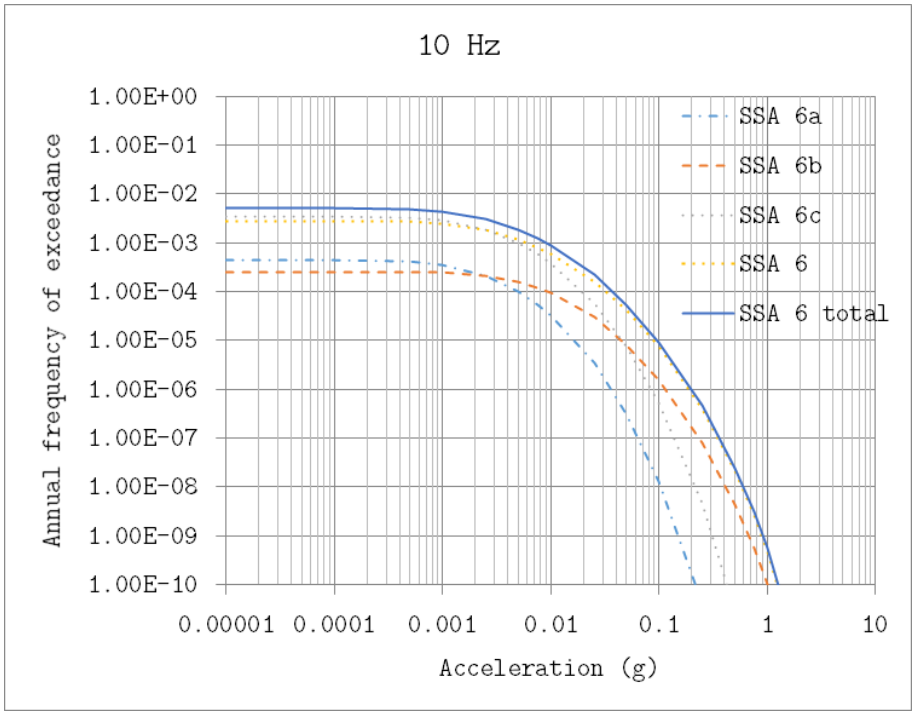
	Area 10				**Area 6**			
	0.5–1	**1–1.5**	**1.5–2**	**2–2.5**	**0.5–1**	**1–1.5**	**1.5–2**	**2–2.5**
2021	23	11	2	1	2	5	4	0
2020	6	4	1	0	2	5	0	0
2019	4	5	1	1	6	5	1	0
2018	4	6	1	0	1	4	2	2
2017	3	2	1	0	0	3	2	2
2016	2	6	2	1	3	5	1	0
2015	12	11	3	1	1	1	1	0
2014	18	28	5	0	0	3	2	0
2013	1	6	2	0	0	2	1	0
2012			2	0			0	0
Mean (1/a)	8.778		2	0.1786		3.667	1.4	0.319

Table 4: Earthquake event counts between year 2012 and end of 2021. Counts of years 2012–2014 are from the original dataset, and are presented here for comparison. Mean rates calculated from the full dataset are given after the data.

	Area 6a			**Area 6b**			**Area 6c**		
	1–1.5	**1.5–2**	**2–2.5**	**1–1.5**	**1.5–2**	**2–2.5**	**1–1.5**	**1.5–2**	**2–2.5**
2021	1	0	0	1	2	0	3	2	0
2020	4	0	0	1	0	0	0	0	0
2019	2	0	0	3	0	0	0	1	0
2018	2	0	2	1	0	0	1	2	0
2017	0	0	0	1	1	0	2	1	2
2016	4	0	0	1	0	0	0	1	0
2015	0	0	0	1	0	0	0	1	0
2014	1	1	0	2	0	0	0	1	0
2013	0	0	0	1	1	0	1	0	0
2012		0	0		0	0		0	0
Mean (1/a)	1.556	0.1	0.143	1.333	0.4	0.069	0.778	0.9	0.107



(a)



(b)

Figure 2: **(a)** Loviisa 10 Hz mean hazard curves from each SSA and the total corresponding to the sum of the individual curves. **(b)** Individual weighted mean hazard curves of the area 6 subareas and the undivided area. Total hazard curve of area 6.

4 Results

The changes in the Loviisa hazard estimate for the different parameter variations are given in Table 5. The relative changes of the 10 Hz hazard estimates are shown in a tornado plot in Figure 3 for Loviisa. The Loviisa 100 Hz results can be found in Appendix D. Figure 4 shows the 10 Hz tornado plot for Olkiluoto, other results are gathered in Appendix E. Some individual results are presented in following subsections.

Table 5: Absolute hazard estimate changes ($\Delta\nu$) and the relative changes for Loviisa for 10 Hz, 0.1 g and 100 Hz, 0.05 g.

	10 Hz		**100 Hz**	
Parameter change	**$\Delta\nu$ (1/a)**	**Relative**	**$\Delta\nu$ (1/a)**	**Relative**
Area 10 additional 2.5-3 event	$-6.05\cdot10^{-6}$	-22.92 %	$-3.97\cdot10^{-6}$	-30.07 %
Area 10 additional 3-3.5 event	$2.32\cdot10^{-5}$	88.03 %	$1.53\cdot10^{-5}$	115.66 %
High-magnitude completeness	$-4.12\cdot10^{-6}$	-15.61 %	$-1.18\cdot10^{-6}$	-8.92 %
Without high-magnitude b split	$1.40\cdot10^{-5}$	53.10 %	$4.21\cdot10^{-6}$	31.93 %
Minimum magnitude 4	$5.04\cdot10^{-6}$	19.09 %	$3.71\cdot10^{-6}$	28.14 %
Minimum magnitude 5	$-4.74\cdot10^{-6}$	-17.97 %	$-3.57\cdot10^{-6}$	-27.06 %
MLE recurrence estimation	$-2.25\cdot10^{-5}$	-85.20 %	$-1.22\cdot10^{-5}$	-92.67 %
Lower-bound magnitude datapoints	$-1.15\cdot10^{-5}$	-43.45 %	$-6.10\cdot10^{-6}$	-46.22 %
Area 6 subdivision 100%	$-5.50\cdot10^{-6}$	-20.84 %	$-1.84\cdot10^{-6}$	-13.94 %
Area 6 subdivision 0%	$5.50\cdot10^{-6}$	20.84 %	$1.84\cdot10^{-6}$	13.94 %

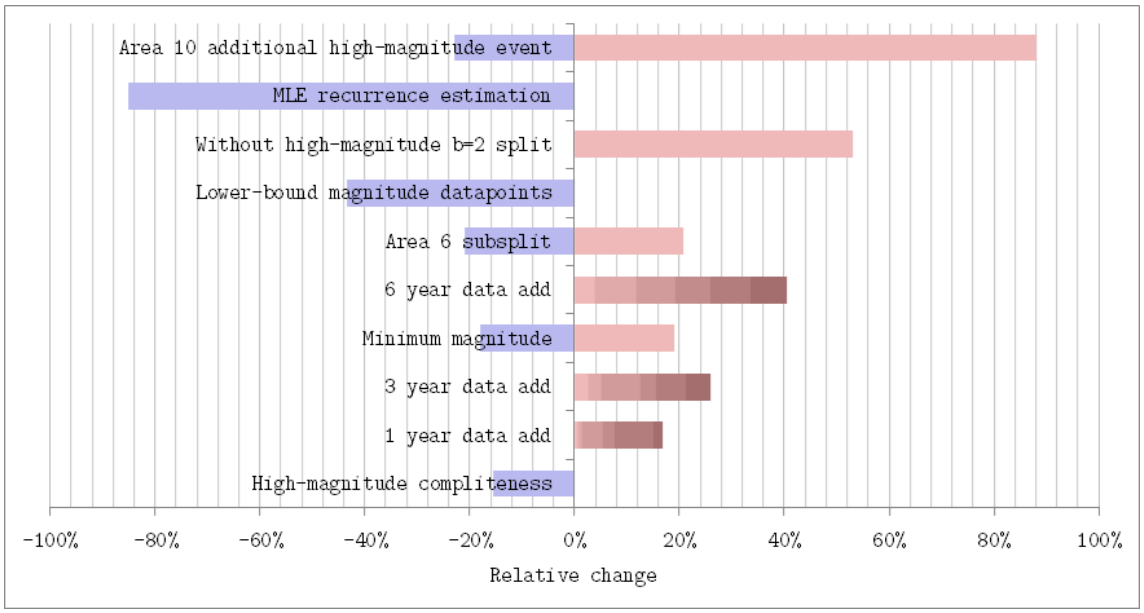


Figure 3: Relative Loviisa hazard estimate changes at 10 Hz for each parameter variation. Each color and shade denotes a single result. A positive 100 % change means that the hazard estimate is doubled, while a negative 100 % means that the estimate becomes zero.

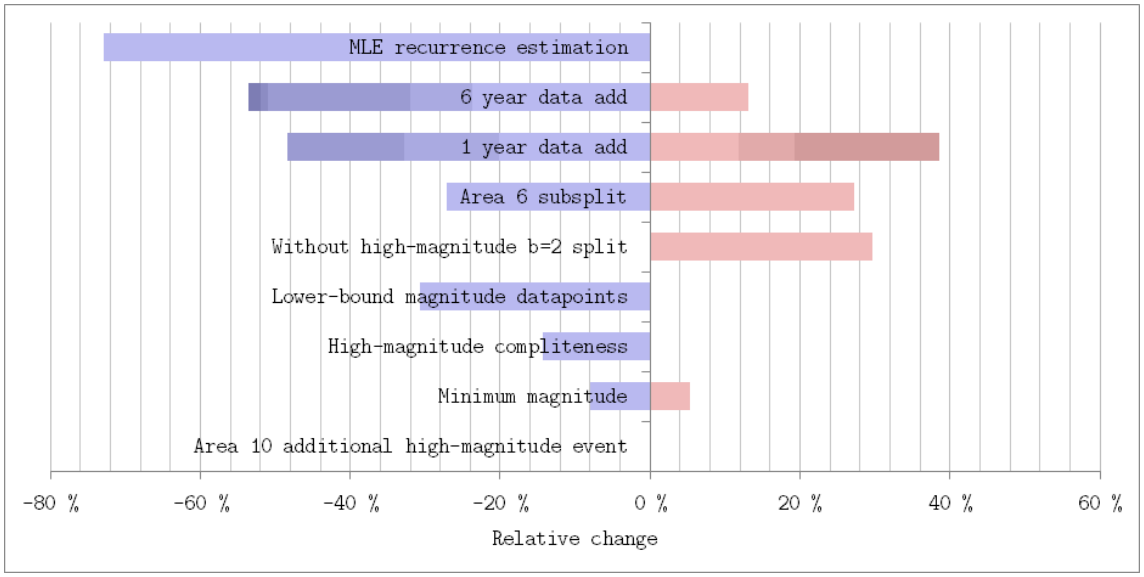


Figure 4: Olkiluoto relative hazard estimate changes at 10 Hz for each parameter variation. Each color and shade denotes a single result. A positive 100 % change means that the hazard estimate is doubled, while a negative 100 % means that the estimate becomes zero.

4.1 Source area 6 subdivision

Figure 5 shows the 10 Hz hazard curves with fully weighted undivided area 6 model and Figure 6 shows the fully subdivided area 6 hazard curves. The subsdivided model has the lower hazard of the two at important levels of acceleration. The subdivision can also be seen widen the confidence interval.

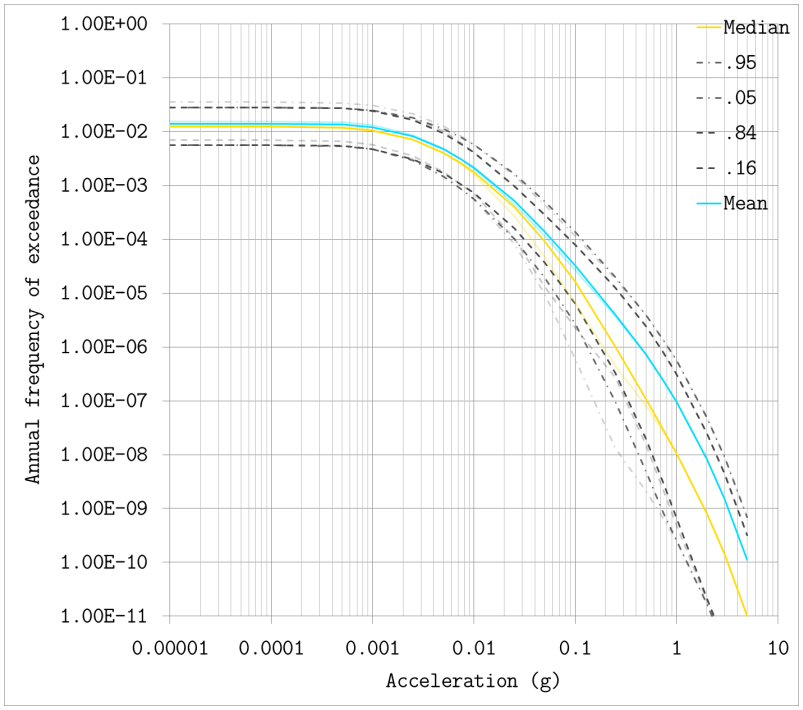


Figure 5: Loviisa 10 Hz hazard curves after fully weighting undivided area 6. The original curve is plotted as a reference with lighter colour.

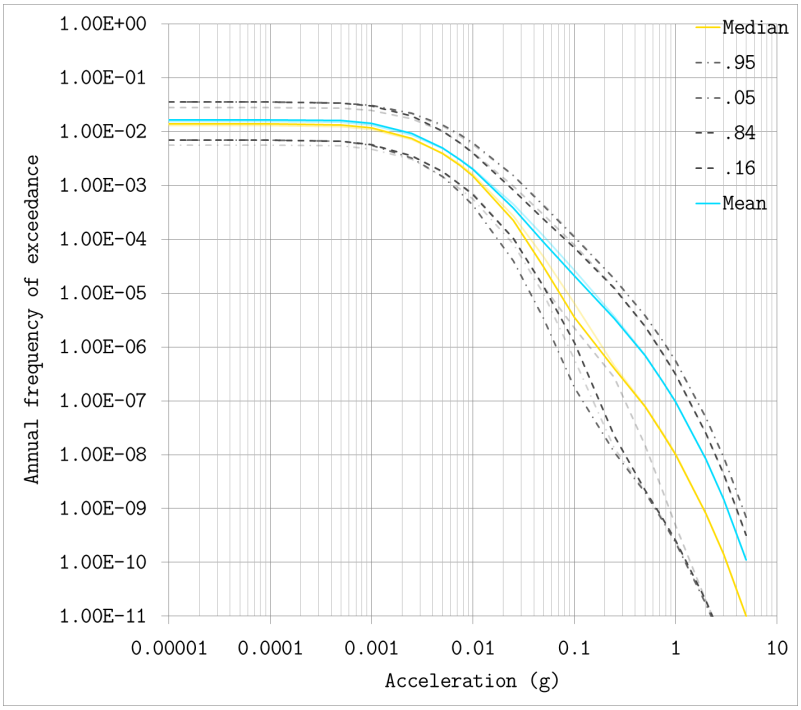


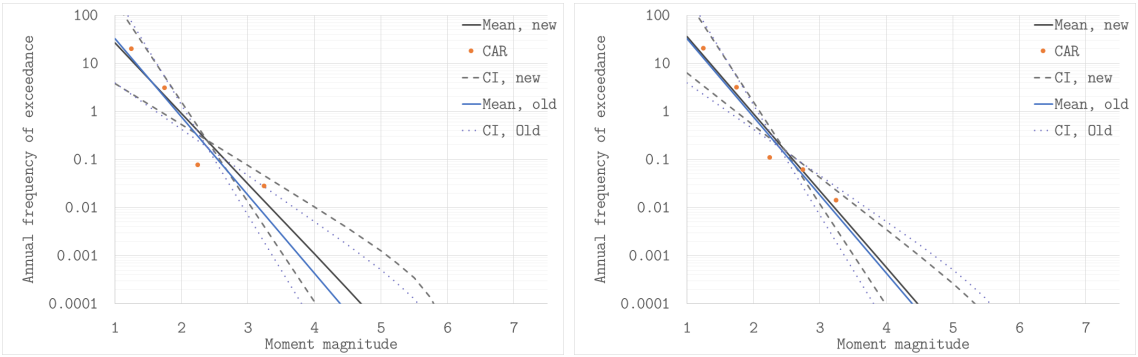
Figure 6: Loviisa 10 Hz hazard curves after fully subdivided areas 6a, 6b and 6c. The original curve is plotted as a reference with lighter colour.

4.2 High-magnitude events in source area 10

The recalculated GR parameters after the addition of an event to area 10 high-magnitude bins are in Table 6. The mean b-values decrease after the additions as could be expected, since the event magnitudes were higher than the mean of the previously included magnitudes. Figure 7 shows the recurrence curves after the modification. In contrast to the increase in the mean recurrence, the hazard estimates decrease with the event added to the second highest magnitude interval. The corresponding 10 Hz hazard curve (Figure 8) shows that this applies to the entire mean curve. The decrease in mean can only be attributed to the significant decrease in the upper confidence bound of the recurrence. The simultaneous increase in median hazard supports this, as the median is less sensitive to changes in confidence bounds of the model parameters.

Table 6: Sensitivity of area 10 b-value and recurrence parameter to additional high-magnitude events. Two variations were made, one with an added event in magnitude interval 2.5–3 and the other with an additional event in 3–3.5. Results are given for the new value, absolute and relative difference to the original.

	b			$n_{\min}$		
	mean	**0.05**	**0.95**	**mean**	**0.05**	**0.95**
Magnitude 3–3.5 add	1.466	2.079	0.854	$1.98{\cdot}10^{-4}$	$1.00{\cdot}10^{-5}$	$3.93{\cdot}10^{-3}$
Change	-0.159	-0.215	-0.104	$1.32{\cdot}10^{-4}$	$7.47{\cdot}10^{-6}$	$2.21{\cdot}10^{-3}$
Relative	-9.81 %	-9.36 %	-10.88 %	199.82 %	292.42 %	129.06 %
Magnitude 2.5–3 add	1.603	2.121	1.085	$8.89{\cdot}10^{-5}$	$7.91{\cdot}10^{-6}$	$9.98{\cdot}10^{-4}$
Change	-0.023	-0.173	0.126	$2.27{\cdot}10^{-5}$	$5.36{\cdot}10^{-6}$	$-7.16{\cdot}10^{-4}$
Relative	-1.42 %	-7.53 %	13.18 %	34.32 %	209.87 %	-41.78 %



(a) (b)

Figure 7: Area 10 recurrence curves and the confidence intervals (CI) after adding an earthquake event to the GR parameter estimation to magnitude interval **(a)** 3–3.5 and **(b)** 2.5–3. The original recurrence curves are plotted as a reference. Curves are drawn with maximum magnitude 6.

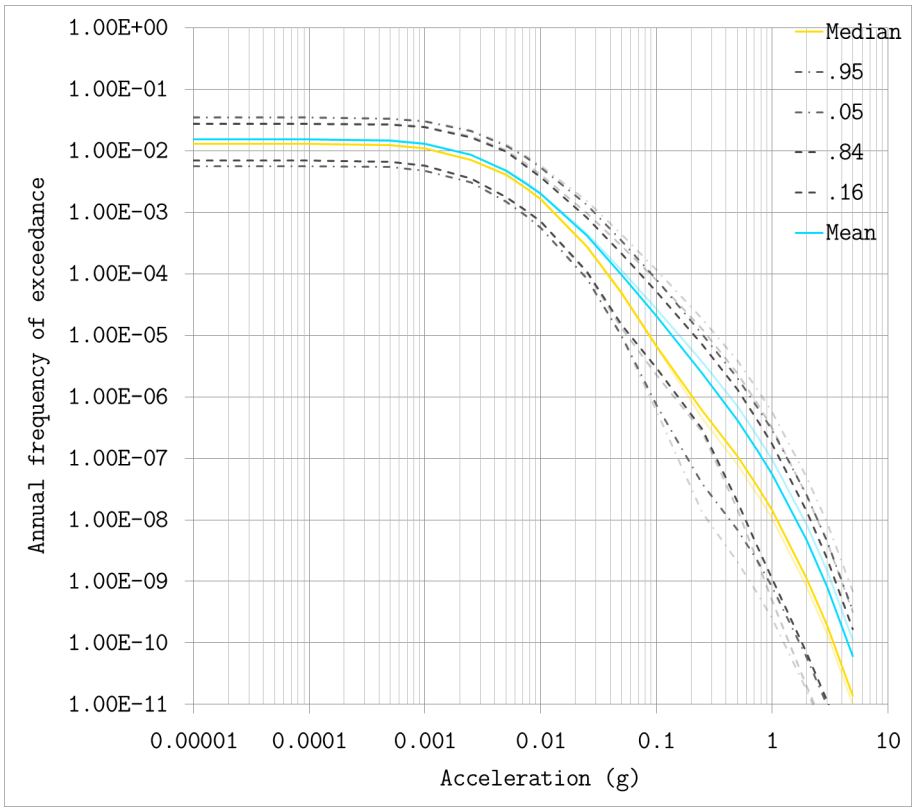


Figure 8: Loviisa 10 Hz hazard curve after adding an event to area 10 magnitude interval 2.5–3. The original curve is plotted as a reference with lighter colour.

4.3 High-magnitude completeness

The earthquake events with magnitudes 4 and higher included in the 2021 PSHA dataset are listed in Table 7. One event from year 1497 with magnitude 4.5 was omitted from this thesis. For the completeness study, the earthquake counts were updated to include all events listed in the table. Observation times of the affected magnitude intervals were increased down to the year of the first included event 1540. Due to the LS estimation method, only those areas with nonzero counts for the magnitudes above 4 saw change in the GR parameters. The b-values of these areas are given in Table 8. It is interesting to note that the upper recurrence bound of area 4 saw the b-value change sign, shown by the larger-than 100 % relative drop. This is not allowed by the GR law, and the result should not be used for hazard calculations.

Table 7: Events of magnitudes 4–4.5 that were included in the 2021 PSHA before completeness evaluation. Provided are the date, SSA, geographic coordinates and inclusion after completeness evaluation for each event.

Year	Month	Day	SSA	Lat.	Long.	Included
1540			4	57.70	18.70	False
1723	8		2	62.60	18.00	False
1751	11	14	2	63.50	19.50	False
1827	9	28	6c	59.00	23.50	False
1879	2	2	4	58.80	16.20	False
1888	7	28	2	63.30	19.00	False
1909	3	9	5	64.00	22.00	True
1931	11	16	8	62.50	25.80	True
1976	10	25	6c	59.26	23.39	True

Table 8: Relative change of b-values and their confidence bounds after high-magnitude completeness increase. The reference are the original values. A higher b-value means a steeper slope and reduced hazard at high magnitudes. Red color highlights a decrease in the b-value, green color an increase.

Source area	b-value	0.05	0.95
2	-65.15 %	-61.21 %	-72.78 %
4	-85.99 %	-76.23 %	-109.07 %
5	16.22 %	16.20 %	16.26 %
6	11.56 %	11.70 %	11.29 %
6c	15.47 %	15.73 %	14.96 %
8	21.03 %	20.68 %	21.62 %

4.4 Sensitivity to recurrence estimation method

The recurrence estimation with the presented MLE method demonstrated all-around reduced hazard with significant changes to the upper confidence bound and mean curves. The formulas without assumption of small magnitude intervals were used. The changes of b-values are shown in Table 9. Most SSAs saw considerable changes to the slope, with largest differences in the upper confidence bound. Figure 9 shows the mean hazard curves for each SSA with 10 Hz frequency. A massive reduction in hazard from area 10 is seen, and consequently area 6 almost completely dominates the mean hazard. The GR estimates of each SSA calculated with MLE are given in appendix B.

Table 9: Relative change of b-values and their confidence bounds after re-estimating recurrence with the MLE method. A higher b-value means a steeper slope and reduced hazard at high magnitudes. Red color highlights a decrease in the b-value, green color an increase.

Source area	b-value	0.05	0.95
1	6.46 %	4.29 %	10.18 %
2	-19.63 %	-30.03 %	0.53 %
3	-16.57 %	-31.08 %	19.26 %
4	-27.93 %	-36.69 %	-7.19 %
5	4.85 %	-9.07 %	28.44 %
6	16.45 %	10.12 %	28.42 %
6a	22.28 %	13.85 %	50.87 %
6b	2.51 %	-5.83 %	30.22 %
6c	-8.60 %	7.53 %	-39.63 %
8	18.68 %	11.92 %	30.23 %
10	21.72 %	-0.43 %	74.74 %

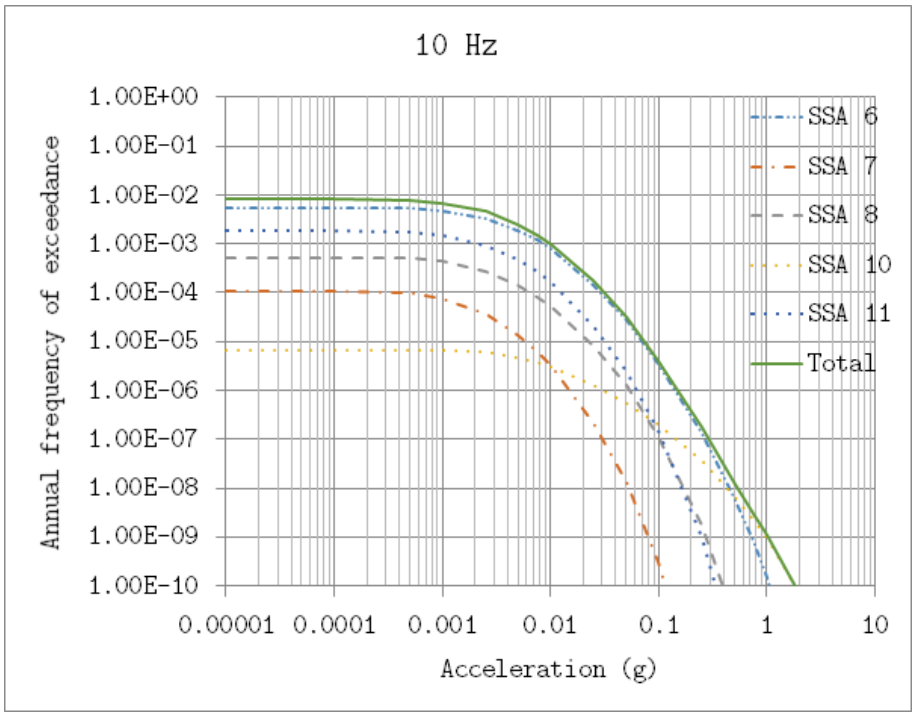


Figure 9: Loviisa mean hazard curves with the MLE recurrence estimation at 10 Hz for each SSA and the total corresponding to the sum of the individual curves.

The area 6 recurrence curves with the three estimation methods are plotted in Figure 10. The figure shows that the regression line fitted on the center magnitude intervals overestimates the intercept a compared to MLE. Recurrence curves for the other

areas are given in Appendix B. Note that these curves do not include the change in b-value at high magnitudes.

The recurrence confidence interval extrapolation from accurate bounds at magnitude 4.5 is also demonstrated in Figure 10. Visual inspection shows a good match between the used method and accurate bounds for area 6. Alternative method of linear interpolation of the slope between minimum magnitude and magnitude 7 showed a relative difference less than 1 % for all other areas but 6c for which it was less than 5 %.

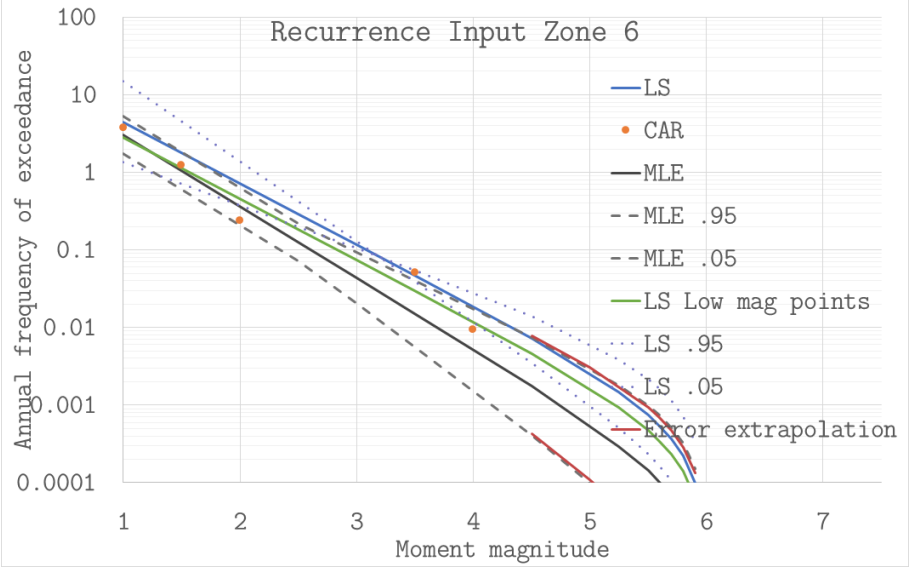


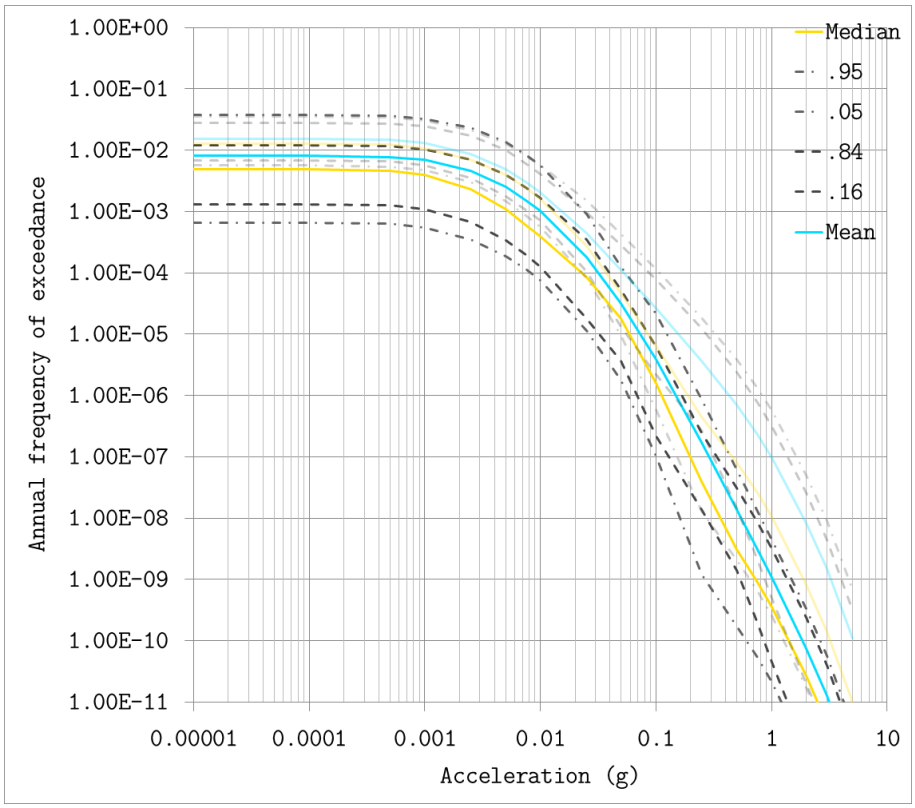
Figure 10: Area 6 recurrence curves calculated with LS with center magnitude interval datapoints (blue), with the points at lower bounds of the intervals (green) and MLE estimate (gray). The cumulative annual earthquake rates (CAR) set at lower bounds of the intervals (orange points). Ten percent confidence intervals for center magnitude LS (dots) and MLE (dashes). Confidence interval extrapolation from point estimates at magnitude 4.5 with unconditional 90 % estimates for b-value (red). Curves are drawn with maximum magnitude 6.

Hazard curves using the MLE method and LS with datapoints shifted to lower bounds of the magnitude intervals are plotted in Figure 11 for 10 Hz frequency. The 100 Hz plots are in Appendix D. With the use of the MLE method, a substantial drop in AFE is evident at high accelerations. This is consistent with the increase of b-values of most areas. The shifted datapoint LS hazard curve looks much like the original but scaled down, which is anticipated as the shift leaves the b-values unchaged. The effect is much smaller than that of the MLE, suggesting the position of the datapoints

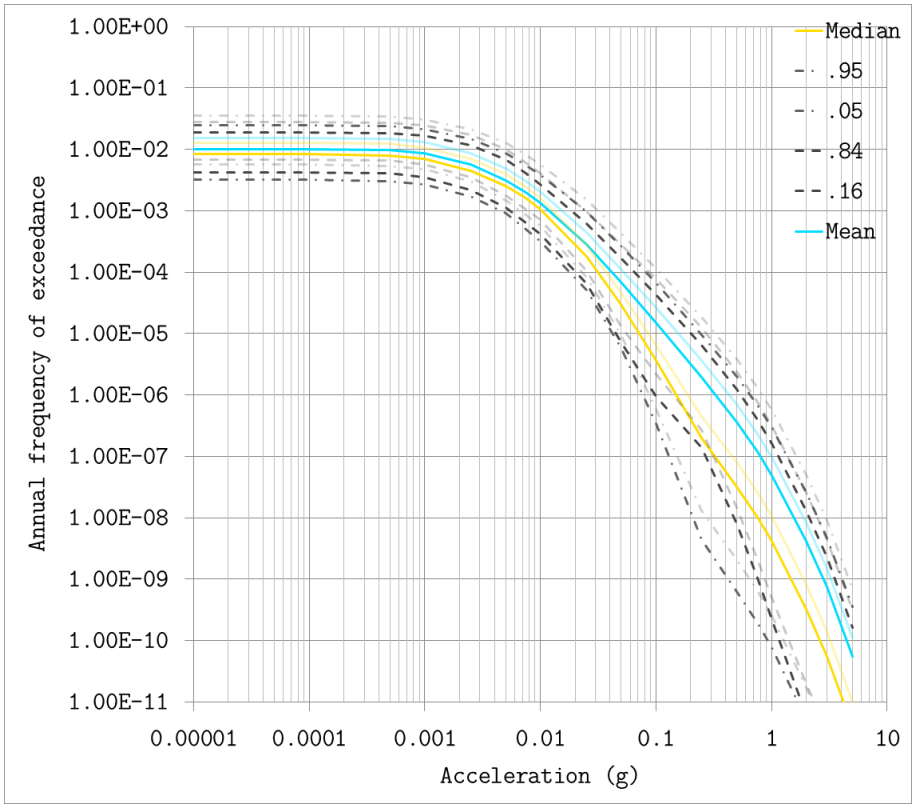
is not the predominant difference leading to the smaller hazard with the MLE.

The addition of high-magnitude event to area 6 and the high-magnitude completeness increase were performed together with MLE method. The GR parameter changes after new event addition are given in Table 10. Unlike with LS, adding an event to the magnitude bin with zero previous events does not lead to a major decrease in the upper confidence bound. The changes to hazard estimates are almost neglible: the 10 Hz hazard estimates at 0.05 g only increased 2.0 % when the event was added to magnitude interval 3–3.5 and 1.3 % when it was added to the interval 2.5–3.

The completeness increase affected all SSAs, as the method was able to include high-magnitude bins without events in the calculation. The changes in b-values are presented in Table 11. While all SSAs were affected, the individual differences were small except for area 6c. The importance of area 6 is highlighted by the hazard estimate change: −53.61 % for 10 Hz frequency.



(a)



(b)

Figure 11: Loviisa 10 Hz hazard curves with GR parameters estimated by **(a)** MLE method and **(b)** LS with points shifted to lower bounds of the magnitude intervals. Original curves are plotted as a reference with lighter colour.

Table 10: Sensitivity of area 10 b-value and recurrence parameter calculated with MLE to additional high-magnitude events. Two variations were made, one with an added event in magnitude interval 2.5–3 and the other with an additional event in 3–3.5. Results are given for the new value, absolute and relative difference to the original.

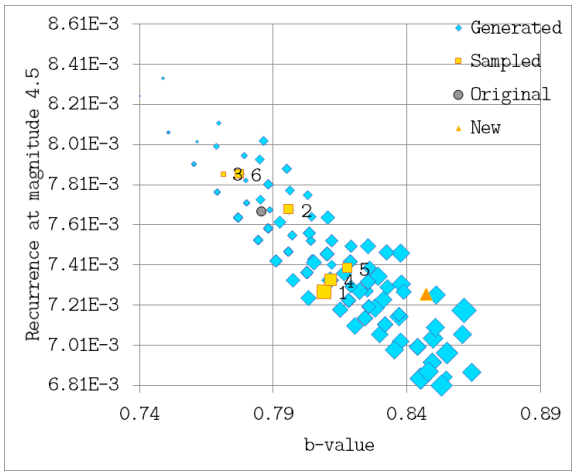
	b			$n_{\min}$		
	mean	**0.05**	**0.95**	**mean**	**0.05**	**0.95**
Magnitude 3–3.5 add	1.854	2.123	1.584	$6.292 \cdot 10^{-6}$	$7.727 \cdot 10^{-7}$	$5.123 \cdot 10^{-5}$
Change	-0.125	-0.161	-0.090	$3.969 \cdot 10^{-6}$	$5.596 \cdot 10^{-7}$	$2.591 \cdot 10^{-5}$
Relative	-6.33 %	-7.03 %	-5.38 %	170.87 %	262.65 %	102.32 %
Magnitude 2.5–3 add	1.885	2.162	1.608	$4.929 \cdot 10^{-6}$	$5.677 \cdot 10^{-7}$	$4.279 \cdot 10^{-5}$
Change	-0.094	-0.121	-0.066	$2.606 \cdot 10^{-6}$	$3.547 \cdot 10^{-7}$	$1.747 \cdot 10^{-5}$
Relative	-4.74 %	-5.32 %	-3.96 %	112.20 %	166.46 %	68.99 %

Table 11: Relative change of b-values and their confidence bounds after high-magnitude completeness increase. The reference are the original MLE calculated values.

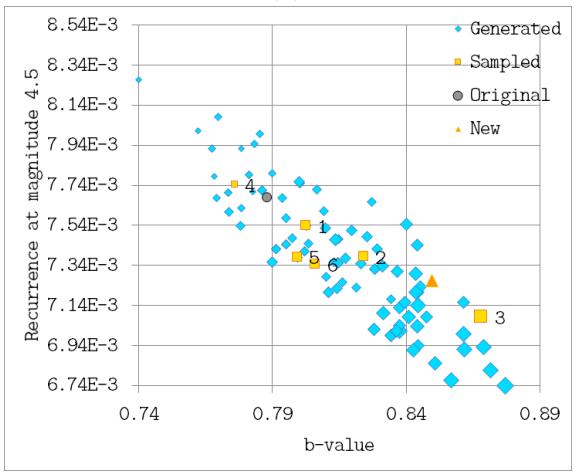
Source area	b-value	0.05	0.95
1	13.84 %	8.78 %	22.03 %
2	-6.17 %	-8.25 %	-3.35 %
3	6.18 %	3.84 %	9.52 %
4	-1.57 %	-5.76 %	5.20 %
5	3.24 %	2.41 %	4.23 %
6	4.32 %	-0.03 %	11.37 %
6a	3.62 %	0.32 %	12.04 %
6b	0.53 %	-0.17 %	2.23 %
6c	25.70 %	9.49 %	81.29 %
8	5.15 %	2.95 %	8.37 %
10	0.09 %	36.38 %	0.19 %

4.5 Sensitivity to new data

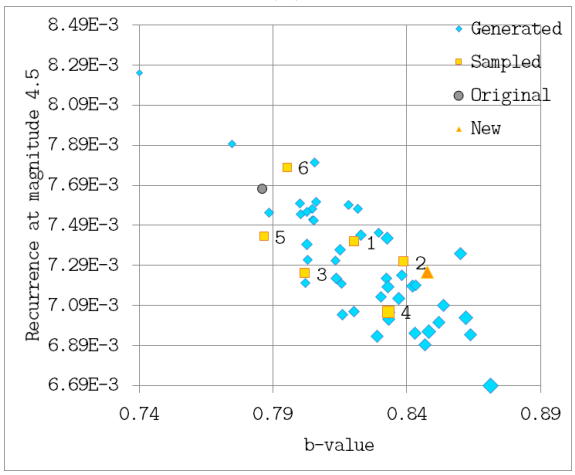
The recalculated recurrence parameters after adding generated earthquake data are in Figure 12 for area 6 and Figure 13 for area 10. Figures for 6 subareas are in Appendix C. The figures also show the parameters after adding the full 7-year FENCAT dataset covering magnitudes 1–2.5.



(a)

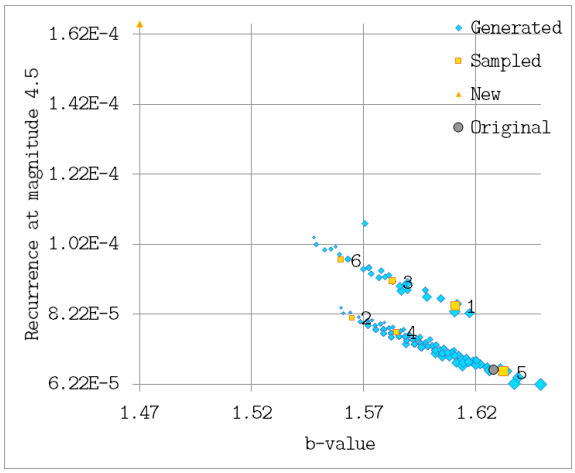


(b)

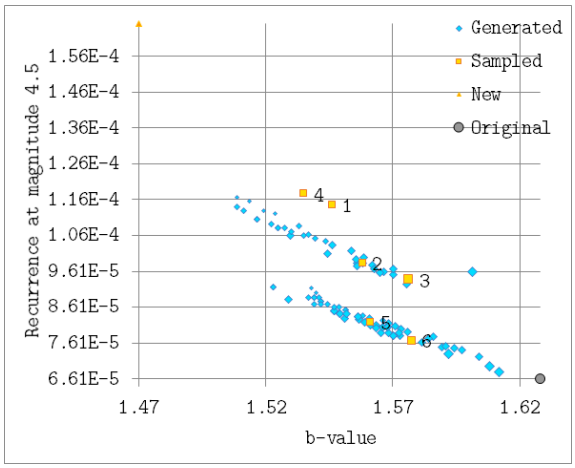


(c)

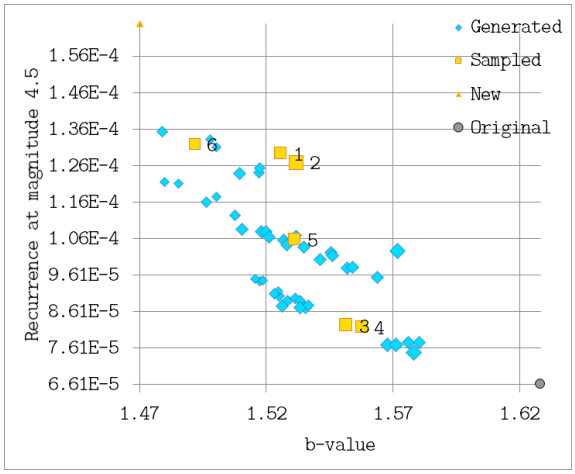
Figure 12: Scatter plot of b-values and $n_{\min}$ at magnitude 4.5 after adding generated earthquake data for area 6. **(a)** 1 year of added data. **(b)** 3 years of added data. **(c)** 6 years of added data. The numbered squares are the samples included for hazard calculation. Also shown are the original values (gray sphere) and values after adding the whole acquired dataset (orange triangle). The sizes of the sample points are scaled by the total number of added events and divided by the original number of events.



(a)



(b)



(c)

Figure 13: Scatter plot of b-values and $n_{\min}$ at magnitude 4.5 after adding generated earthquake data for area 10. **(a)** 1 year of added data. **(b)** 3 years of added data. **(c)** 6 years of added data. The numbered squares are the samples included for hazard calculation. Also shown are the original values (gray sphere) and values after adding the whole acquired dataset (orange triangle). The sizes of the sample points are scaled by the total number of added events and divided by the original number of events.

5 Discussion and conclusions

The 2021 PSHA demonstrated high sensitivity to some modelling decisions. The largest variation was related to the GR parameter estimation method. Both Loviisa and Olkiluoto saw the hazard estimates drop significantly after using MLE method for estimating the GR parameters. The difference of the GR parameters estimated with LS and MLE was mixed, however. LS often showed a shallower slope, as expected from it not taking into account empty magnitude intervals at the high-magnitude end. This was not the case for the SSAs 2, 3, 4 and 6c for undetermined reasons. The use of magnitude interval centers for the LS regression results in an upwards shift to the recurrence curves. While the rationale behind its use is unclear, it is to be noted that important questions regarding the handling of magnitude scales were not considered in this thesis.

With area 10, the high earthquake count at low magnitudes makes the MLE recurrence confidence interval tiny compared to LS estimate. This may suggest that the MLE method is more sensitive to the assumption of Poisson distribution of event counts. On the other hand, area 6c shows a wider confidence interval with the MLE method. Both methods rely on a large enough earthquake count to make the confidence intervals accurate. For areas with low earthquake counts, some suggest the use of bootstrapping as an alternative method.

The MLE method has its own intricacies that the implementer has to consider. The theory permits using arbitrarily large magnitude intervals with zero events, but omitting the maximum magnitude truncation makes the variance estimates less accurate if magnitudes close to it are considered. In this thesis, bins were considered up to magnitude 5, which should remain consistent with the maximum magnitudes as well as the use of a different b-value for magnitudes higher than 5.75. Regarding the choice of making the small magnitude interval approximation, it is hard to see where this approximation would be necessary as the accurate equations are equally easy to solve. Without the approximation the magnitude interval widths can be chosen arbitrarily, but the effect on the estimates calls for another study.

The high-magnitude completeness increase showed small changes compared to other parameters, although this would change if the MLE method was used. The MLE and

LS methods seem to have a very different behaviour regarding the magnitude bin observation times. Naturally increasing the admitted history should make the GR paramteter estimates less sensitive to addition of new data, bias nothwithstanding. The LS method only shows this change for the areas where high-magnitude events had occurred. The study left out an event with magnitude 4.5-5 within area 4 that happened before 1540, though the changes were also omitted for this SSA due to the change of sign of the b-value confidence bound. While a negative b-value does not make sense for the GR law, the result could be used if the correct normalization of the magnitude distribution within the calculation software is verified.

Sensitivity to the addition of low-magnitude earthquake data was investigated. A sample of Poisson distributed yearly event counts were created with rate parameters assessed from the combined original data and new data acquired from FENCAT. The GR parameters proved out to be quite sensitive to the variation of the yearly event counts, as could be expected from the short completeness times of the low-magnitude intervals. The study did not directly show how the sensitivity changes after adding the data. For area 6 where the generated distribution remains similar after the different numbers of years added, the estimates could converge quickly. The GR parameter confidence estimation should accurately take into account the variation in the yearly event counts, but this depends on the assumption of Poisson process.

With area 10 the original and the fully updated datapoints appear at extremes of the generated data. It seems unlikely that the original and the new yearly counts come from a Poisson distribution with a common rate parameter. Comparing the original yearly event counts to the new data suggests that year 2014 is an outlier with its high count between magnitudes 0.5–1.5, although high variation can be seen occurring other times as well. Further studies should be made to verify the new data and to make sure it is treated identically to the original data.

The datapoints after updating with the new data in full showed a substantial change for both NPPs near the important levels of spectral acceleration. Even after generating noise to the event counts a consistent increase to the Loviisa hazard estimate was observed. Olkiluoto saw a lot more variation with inclination towards lower hazard. In interpreting the results, it is to be kept in mind that the dataset was used without rigorous verification and that the events were added only up to magnitude 2.5. The study was only performed with LS GR estimation. The results could look very

different with MLE due to the dissimilarities discussed previously.

The newly used GMPE in the 2021 PSHA was known to be a major contributor to the changes in the estimated hazard compared to the previous PSHA. This important part was left out of this thesis for technical reasons. The GMPEs are coded directly into HAZ and modifying them requires access to the source code. In addition, the new PSHA adopted novel methods to the GMPE assessment that were out of the scope of this thesis. Ground motion prediction for Finnish soil likely remains a subject of much discussion in the near future.

The results of the chosen 10 Hz and 100 Hz hazard estimates showed some disparity. This can be understood against the background that the hazard curves themselves remained apart at important levels of spectral acceleration and the estimates were picked at different points. It would be desirable to be able to report the sensitivity also in terms of the acceleration to deduce effects on the design basis earthquake. However, this can not be directly obtained from the presented point estimates. With some additional analysis on the hazard curves, changes in acceleration for a set frequency of exceedance could be produced.

References

[1] Radiation and nuclear safety authority (STUK), "Provisions for internal and external hazards at a nuclear facility (YVL B.7)," 2019. [Online]. Available: http://www.finlex.fi/data/normit/41791-YVL_B.7e.pdf (visited on 03/01/2022).

[2] N. C. Nigam and P. C. Jennings, "Calculation of response spectra from strong-motion earthquake records," *Bulletin of the Seismological Society of America*, vol. 59, no. 2, pp. 909–922, Apr. 1969, ISSN: 0037-1106. DOI: 10.1785/BSSA0590020909.

[3] D. Wakefield, M. Ravindra, K. Merz, and G. Hardy, *Seismic Probabilistic Risk Assessment Implementation Guide*, 1002989. EPRI, Palo Alto, CA, 2003.

[4] *Seismic Hazards in Site Evaluation for Nuclear Installations*, ser. Specific Safety Guides SSG-9 (Rev. 1). Vienna: INTERNATIONAL ATOMIC ENERGY AGENCY, 2022, ISBN: 9789201178213. eprint: https://www.iaea.org/publications/14665/seismic-hazards-in-site-evaluation-for-nuclear-installations.

[5] J. J. Bommer, F. Scherbaum, H. Bungum, F. Cotton, F. Sabetta, and N. A. Abrahamson, "On the Use of Logic Trees for Ground-Motion Prediction Equations in Seismic-Hazard Analysis," *Bulletin of the Seismological Society of America*, vol. 95, no. 2, pp. 377–389, Apr. 2005, ISSN: 0037-1106. DOI: 10.1785/0120040073.

[6] R. Budnitz, G. Apostolakis, D. Boore, L. Cluff, K. Coppersmith, C. Cornell, and P. Morris, *Recommendations for Probabilistic Seismic Hazard Analysis: Guidance on Uncertainty and Use of Experts: Main Report*. Jan. 1997, vol. 1.

[7] U. o. H. Institute of Seismology. (2022). "FENCAT earthquake search tool," [Online]. Available: https://www.seismo.helsinki.fi/EQ-search/ (visited on 03/01/2022).

[8] P. Mäntyniemi, M. Uski, T. Vuorinen, T. Huotari-Halkosaari, K. Korhonen, N. Mikko, M. Pajunen, J.-P. Palmu, H. Virkki, J. Virtasalo, E. Kosonen, H. Soosalu, K. Suuroja, S. Suuroja, A. Veski, S. Grigull, C. Smith, C.-H. Wahlgren, K. Högdahl, and B. Lund, "Seismic source areas in central Fennoscandia," English, Institute of Seismology, Report S, no. 64, A. Korja, S. Kihlman, and K. Oinonen, Eds., Nov. 2016.

[9] P. Mäntyniemi, M. Malm, S. Burck, O. Okko, P. Välikangas, and L. Fülöp, "Sensitivity of seismic hazard analysis in finland: Overview of the SENSEI project," English, *ATS Ydintekniikka*, vol. 50, no. 2, pp. 19–23, Jul. 2021, ISSN: 0356-0473.

[10] N. Abrahamson, *HAZ*, version 45.2, GitHub repository, 2017. [Online]. Available: https://github.com/abrahamson/HAZ (visited on 03/01/2022).

[11] C. A. Cornell, "Engineering seismic risk analysis," *Bulletin of the Seismological Society of America*, vol. 58, no. 5, pp. 1583–1606, Oct. 1968, ISSN: 0037-1106. DOI: 10.1785/BSSA0580051583.

[12] T. C. Hanks, N. A. Abrahamson, D. M. Boore, K. J. Coppersmith, and N. E. Knepprath, "Implementation of the SSHAC guidelines for level 3 and 4 PSHAs - experience gained from actual applications," 2009. eprint: https://pubs.usgs.gov/of/2009/1093/.

[13] R. K. McGuire, *Seismic Hazard and Risk Analysis*, ser. Engineering monographs on miscellaneous earthquake engineering topics. Earthquake Engineering Research Institute, 2004, ISBN: 9780943198019.

[14] M. Stirling, "The continued utility of probabilistic seismic-hazard assessment," *Earthquake Hazard, Risk and Disasters*, pp. 359–376, Dec. 2014. DOI: 10.1016/B978-0-12-394848-9.00013-4.

[15] C. Godano, E. Lippiello, and L. de Arcangelis, "Variability of the b value in the Gutenberg-Richter distribution," *Geophysical Journal International*, vol. 199, no. 3, pp. 1765–1771, Oct. 2014, ISSN: 0956-540X. DOI: 10.1093/gji/ggu359.

[16] M. W and L. Sandri, "A review and new insights on the estimation of the b-value and its uncertainty," *Annals of Geophysics*, vol. 46, Dec. 2009. DOI: 10.4401/ag-3472.

[17] C. Frohlich and S. D. Davis, "Teleseismic b values; or, much ado about 1.0," *Journal of Geophysical Research: Solid Earth*, vol. 98, no. B1, pp. 631–644, 1993. DOI: 10.1029/92JB01891.

[18] K. Aki, "Maximum likelihood estimate of b in the formula logn=a-bm and its confidence limits," *Bulletin of Earthquake Research Institute*, vol. 43, pp. 237–239, 1965.

[19] D. Weichert, "Estimation of the earthquake recurrence parameters for unequal observation periods for different magnitudes," *Bulletin of the Seismological Society of America*, vol. 70, pp. 1337–1346, 1980.

[20] D. Stromeyer and G. Grünthal, "Capturing the uncertainty of seismic activity rates in probabilistic seismic-hazard assessments," *Bulletin of the Seismological Society of America*, vol. 105, pp. 580–589, Mar. 2015. DOI: 10.1785/0120140185.

[21] A. Kijko and M. Sellevoll, "Estimation of earthquake hazard parameters from incomplete data files. Part I. Utilization of extreme and complete catalogs with different threshold magnitudes," *Bulletin of the Seismological Society of America*, vol. 79, pp. 645–654, Jun. 1989. DOI: 10.1785/BSSA0790030645.

[22] A. C. Johnston, L. R. Kanter, K. J. Coppersmith, and C. A. Cornell, "The earthquakes of stable continental regions. volume 1, assessment of large earthquake potential, final report," Dec. 1994. eprint: https://www.osti.gov/biblio/34452.

[23] M. G. Kendall and A. Stuart, *The Advanced Theory of Statistics vol 2 (London: Charles Griffin)*. 1961.

[24] 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," *Annals of Geophysics*, vol. 50, pp. 329–339, Jun. 2007. DOI: 10.4401/ag-4432.

[25] H. A. Merz and C. A. Cornell, "Seismic risk analysis based on a quadratic magnitude-frequency law," *Bulletin of the Seismological Society of America*, vol. 63, no. 6-1, pp. 1999–2006, Dec. 1973, ISSN: 0037-1106. DOI: 10.1016/0148-9062(74)90731-1.

[26] J. Bommer and H. Crowley, "The purpose and definition of the minimum magnitude limit in PSHA calculations," *Seismological Research Letters*, vol. 88, pp. 1097–1106, Jul. 2017. DOI: 10.1785/0220170015.

[27] A. Kijko and G. Graham, "Parametric-historic procedure for probabilistic seismic hazard analysis Part I: Estimation of maximum regional magnitude m max," *Pure and Applied Geophysics*, vol. 152, pp. 413–442, Oct. 1998. DOI: 10.1007/s000240050161.

[28] P. Koskinen, "Orientation of faults and their potential for reactivation in the present stress field in Finland," Master's thesis, University of Helsinki, 2013.

[29] A. Kijko, "Estimation of the maximum earthquake magnitude, m max," *Pure and Applied Geophysics*, vol. 161, no. 8, pp. 1655–1681, 2004.

[30] L. Fülöp, V. Jussila, R. M. Aapasuo, T. A. T. Vuorinen, and P. B. Mäntyniemi, "A ground-motion prediction equation for Fennoscandian nuclear installations," English, *Bulletin of the Seismological Society of America*, vol. 110, no. 3, pp. 1211–1230, 2020, ISSN: 0037-1106. DOI: 10.1785/0120190230.

[31] L. Fülöp, V. Jussila, R. Aapasuo, T. Vuorinen, and P. Mäntyniemi, "Evolving the Fennoscandian GMPEs (EVOGY)," English, in *SAFIR2018 - The Finnish Research Programme on Nuclear power Plant Safety 2015-2018*, J. Hämäläinen and V. Suolanen, Eds., 349. Finland: VTT Technical Research Centre of Finland, 2019, vol. T349, pp. 422–433. DOI: 10.32040/2242-122X.2019.T349.

[32] J. Tang, "Sensitivity analysis of the probabilistic seismic hazard assessment on the Hanhikivi site," Master's thesis, Aalto University, 2019.

[33] L. Rinne, "Seismic wave attenuation and the spectral decay parameter κ (kappa) in crystalline bedrock at Olkiluoto, SW Finland," Master's thesis, University of Helsinki, 2021.

[34] J. Douglas, "Capturing geographically-varying uncertainty in earthquake ground motion models or what we think we know may change," in *Recent Advances in Earthquake Engineering in Europe: 16th European Conference on Earthquake Engineering-Thessaloniki 2018*, K. Pitilakis, Ed. Springer International Publishing, 2018, pp. 153–181, ISBN: 978-3-319-75741-4. DOI: 10.1007/978-3-319-75741-4_6.

[35] R. Kulkarni, R. Youngs, and K. Coppersmith, "Assessment of confidence intervals for results of seismic hazard analysis," in *Proceedings of the Eighth World Conference on Earthquake Engineering*, vol. 1, 1984, pp. 263–270.

[36] N. A. Abrahamson and J. J. Bommer, "Probability and uncertainty in seismic hazard analysis," *Earthquake Spectra*, vol. 21, no. 2, pp. 603–607, 2005. DOI: 10.1193/1.1899158.

[37] A. C. Miller and T. R. Rice, "Discrete approximations of probability distributions," *Management Science*, vol. 29, pp. 352–362, 1983.

[38] "Finland probabilistic seismic hazard analysis, Loviisa Nuclear Power Plant, Finland," Aug. 2021, Slate Geotechnical Consultants Inc. LO1-T84252-00016Liite1.

[39] T. Leppänen, "Re-evaluation of seismic hazard spectra in Loviisa NPP site," Dec. 2018, Fortum Power and Heat Oy. LO1-T84252-00007.

[40] M. Malm and O. Kaisko, "Teollisuuden voima Oyj / Fortum Oyj, Re-evaluation of seismic hazard spectra in Olkiluoto and Loviisa," Dec. 2017, ÅF-Consult Ltd. LO1-T84252-00006.

[41] J. Saari and M. Malm, "Teollisuuden voima Oyj / Fortum Oyj, Re-evaluation of seismic hazard in Olkiluoto and Loviisa," Dec. 2016, ÅF-Consult Ltd. LO1-T84252-00004.

[42] P. Reasenberg, "Second-order moment of central California seismicity, 1969—1982," *Journal of Geophysical Research: Solid Earth*, vol. 90, no. B7, pp. 5479–5495, 1985. DOI: 10.1029/JB090iB07p05479.

A Derivation of maximum likelihood estimation for GR–parameters

This appendix presents a derivation of the maximum likelihood estimation method for double truncated GR–equation parameters, without the assumption of small magnitude intervals. Minimum magnitude shift is omitted from the equations, because it leaves the MLE estimates without change. This is the case also for the inclusion of maximum magnitude truncation, although the variance of α and the covariance will be affected.

The truncated GR-relation is

$$n = e^{\alpha} \frac{e^{-\beta m} - e^{-\beta m_{\max}}}{1 - e^{-\beta m_{\max}}} \,. \tag{A1}$$

Equation A1 implies that the number of events between magnitudes $m_i - \frac{1}{2}\delta m$ and $m_i + \frac{1}{2}\delta m$

$$\delta n_i = \Big[\frac{e^{\alpha-\beta m} - e^{-\beta m_{\max}}}{1 - e^{-\beta m_{\max}}} \Big] \Big|_{m=m_i+\frac{1}{2}\delta m}^{m_i-\frac{1}{2}\delta m} \tag{A2}$$

$$= \frac{e^{\alpha-\beta m_i}}{1 - e^{-\beta m_{\max}}} \Big(e^{\beta \frac{1}{2}\delta m} - e^{-\beta \frac{1}{2}\delta m} \Big) \tag{A3}$$

$$= 2 \sinh{(\beta \frac{1}{2} \delta m)} \frac{e^{\alpha-\beta m_i}}{1 - e^{-\beta m_{\max}}} \,. \tag{A4}$$

Let us assume that earthquakes recur following a Poisson process with rate parameter $\delta n_i t_i$. The probability of observing n_i events within magnitude interval $m_i - \frac{1}{2}\delta m$ to $m_i + \frac{1}{2}\delta m$ in time period t_i is

$$P_i(n) = \frac{(\delta n_i t_i)^{n_i}}{n_i!} e^{-\delta n_i t_i} \tag{A5}$$

Assuming that the occurrences between magnitude intervals are independent, the likelihood of observing n_i observations for each magnitude interval $i \in [1,I]$ with completeness periods t_i is

$$\mathcal{L} = \prod_{i=1}^{I} P_i(n_i) \,. \tag{A6}$$

The logarithm of its likelihood is

$$\ln \mathcal{L} = \sum_{i=1}^{I} \Big[n_i \ln(\delta n_i t_i) - \ln(n_i!) - \delta n_i t_i \Big] . \tag{A7}$$

To maximize this, it suffices to find the roots of the derivatives

$$\frac{\partial \ln \mathcal{L}}{\partial \alpha} = \sum_{i=1}^{I} \Big[n_i - \delta n_i t_i \Big] = N - \sum_{i=1}^{I} \delta n_i t_i = 0 \tag{A8}$$

$$\frac{\partial \ln \mathcal{L}}{\partial \beta} = \sum_{i=1}^{I} \Big[n_i \frac{\delta n_i'}{\delta n_i} - \delta n_i' t_i \Big] = 0 . \tag{A9}$$

The parameter α can be evaluated from (A8)

$$\frac{e^{\alpha}}{1 - e^{-\beta m_{\max}}} = \frac{N}{\sum_{i=1}^{I} 2 \sinh\left(\beta \frac{\delta m_i}{2}\right) e^{-\beta m_i} t_i} . \tag{A10}$$

The partial derivative of the occurrence δn_i with respect to β has the following expression:

$$\begin{aligned}
\delta n_i' &= 2 e^{\alpha - \beta m_i} \Big(\frac{\delta m_i}{2} \cosh\left(\beta \frac{\delta m_i}{2}\right) \frac{1}{1 - e^{-\beta m_{\max}}} - m_i \sinh\left(\beta \frac{\delta m_i}{2}\right) \frac{1}{1 - e^{-\beta m_{\max}}} \\
&\quad - \sinh\left(\beta \frac{\delta m_i}{2}\right) \frac{m_{\max} e^{-\beta m_{\max}}}{(1 - e^{-\beta m_{\max}})^2} \Big) \\
&= \delta n_i \Big(\frac{\delta m_i}{2} \coth\left(\beta \frac{\delta m_i}{2}\right) - m_i - \frac{m_{\max}}{1 - e^{-\beta m_{\max}}} \Big) \\
&= \delta n_i \phi_i .
\end{aligned}$$

The multiplier in parenthesis is marked with ϕ_i to shorten the notation. Using this, equation (A9) can be written as

$$\sum_{i=1}^{I} \Big[(n_i - \delta n_i t_i) \phi_i \Big] = 0 . \tag{A11}$$

This can be solved for β numerically by using the definition of δn_i (A4) and equation (A10) to obtain

$$\delta n_i = \frac{N \sinh\left(\beta \frac{\delta m}{2}\right) e^{-\beta m_i}}{\sum_{i=1}^{I} \sinh\left(\beta \frac{\delta m_i}{2}\right) e^{-\beta m_i} t_i} . \tag{A12}$$

Note that equation (A8) implies that only those terms of ϕ_i which depend on i contribute to the sum (A11), hence the term involving $m_{\max}$ can be omitted.

To assess uncertainty, the covariance matrix $C_{\alpha\beta}$ can be approximated by inverting the Hessian of negative of the log-likelihood

$$C_{\alpha\beta} \approx \begin{bmatrix} -\frac{\partial^2 \ln \mathcal{L}}{\partial \alpha^2} & -\frac{\partial \ln \mathcal{L}}{\partial \alpha \partial \beta} \\ -\frac{\partial^2 \ln \mathcal{L}}{\partial \alpha \partial \beta} & -\frac{\partial^2 \ln \mathcal{L}}{\partial \beta^2} \end{bmatrix}^{-1} . \tag{A13}$$

Here α and β are assumed to be optimal, hence A8 and A9 can be used to simplify the expressions for the derivatives. The covariance matrix elements are:

$$\mathrm{var}(\alpha) \approx \frac{-1}{D} \sum_{i=1}^{I} \Big[(n_i - \delta n_i t_i)(\frac{\delta m_i}{2})^2 (1 - \coth^2(\beta \frac{\delta m_i}{2})) - \delta n_i t_i \phi_i^2 \Big] \tag{A14}$$

$$\mathrm{var}(\beta) \approx \frac{N}{D} \tag{A15}$$

$$\mathrm{covar}(\alpha, \beta) \approx \frac{-1}{D} \sum_{i=1}^{I} n_i \phi_i \, . \tag{A16}$$

Here D is the determinant of the Hessian, given by

$$\begin{aligned} D =& \Big(\frac{\partial^2 \ln \mathcal{L}}{\partial \alpha^2}\Big)\Big(\frac{\partial^2 \ln \mathcal{L}}{\partial \beta^2}\Big) - \Big(\frac{\partial^2 \ln \mathcal{L}}{\partial \alpha \partial \beta}\Big)^2 \\ =& - N \sum_{i=1}^{I} \Big[(n_i - \delta n_i t_i)(\frac{\delta m_i}{2})^2 (1 - \coth^2(\beta \frac{\delta m_i}{2})) - \delta n_i t_i \phi_{i2}^2 \Big] \\ & - (\sum_{i=1}^{I} n_i \phi_{i2})^2 \, , \end{aligned} \tag{A17}$$

where ϕ_{i2} only contains terms which depend on i:

$$\phi_{i2} = \frac{\delta m_i}{2} \coth(\beta \frac{\delta m_i}{2}) - m_i \, . \tag{A18}$$

A simple and efficient algorithm for solving (A11) is the Newton's method. The update rule for a Newton's method iteration is

$$\hat{\beta}_{i+1} = \hat{\beta}_i - \frac{\sum_{i=1}^{I} \Big[(n_i - \delta n_i t_i) \phi_{i2} \Big]}{\sum_{i=1}^{I} \Big[(n_i - \delta n_i t_i)(\frac{\delta m_i}{2})^2 (1 - \coth^2(\beta \frac{\delta m_i}{2})) - \delta n_i t_i {\phi_{i2}}^2 \Big]} \, . \tag{A19}$$

The terms containing $m_{\max}$ have been omitted as they do not affect the estimate at convergence.

B Recurrence curves

Figure B1 shows the recurrence curves of Loviisa source areas, calculated with LS and MLE methods. The curves do not show the change in slope at high magnitudes that were used in the hazard calculations. The recurrence curve of source area 6 is given in the main text (Figure 10). The estimated b-value and recurrence parameter $n_{\min}$ are given in Tables B1 for LS estimation and B2 for MLE.

Table B1: GR parameters for each source area calculated with LS method. The parameters of area 6 are also used for areas 7 and 11.

LS	**b-value**			$\boldsymbol{n_{\min}}$		
Source area	**mean**	**0.05**	**0.95**	**mean**	**0.05**	**0.95**
1	0.7846	0.9897	0.5795	0.0068210	0.0030440	0.0152845
2	1.4523	1.9163	0.9883	0.0002467	0.0000283	0.0021535
3	1.2823	1.8252	0.7395	0.0005144	0.0000275	0.0096295
4	1.3059	1.8355	0.7762	0.0005234	0.0000532	0.0051522
5	1.1802	1.4844	0.8759	0.0037024	0.0012586	0.0108911
6	0.7907	1.0344	0.5470	0.0076737	0.0036095	0.0163143
6a	0.9895	1.5285	0.4505	0.0006383	0.0000237	0.0172175
6b	1.5973	2.4553	0.7393	0.0000131	0.0000001	0.0024904
6c	0.5800	0.7633	0.3966	0.0096638	0.0054786	0.0170461
8	0.9429	1.1889	0.6969	0.0033003	0.0013794	0.0078966
10	1.6258	2.2933	0.9584	0.0000661	0.0000026	0.0017138

Table B2: GR parameters for each source area calculated with MLE method. The parameters of area 6 are also used for areas 7 and 11.

MLE	**b-value**			$\boldsymbol{n_{\min}}$		
Source area	**mean**	**0.05**	**0.95**	**mean**	**0.05**	**0.95**
1	0.8353	1.0321	0.6385	0.0034951	0.0009804	0.0124609
2	1.1672	1.3408	0.9936	0.0007722	0.0002250	0.0026509
3	1.0699	1.2579	0.8819	0.0011140	0.0003025	0.0041030
4	0.9412	1.1620	0.7203	0.0015970	0.0003650	0.0069880
5	1.2374	1.3497	1.1251	0.0012567	0.0005591	0.0028249
6	0.9208	1.1391	0.7025	0.0018193	0.0004269	0.0077531
6a	1.2100	1.7403	0.6797	0.0000655	0.0000015	0.0029376
6b	1.6374	2.3120	0.9627	0.0000036	0.00000002	0.0006265
6c	0.5301	0.8208	0.2394	0.0079353	0.0014340	0.0439103
8	1.1191	1.3306	0.9076	0.0006756	0.0001530	0.0029836
10	1.9790	1.6746	1.6746	0.0000023	0.0000002	0.0000253

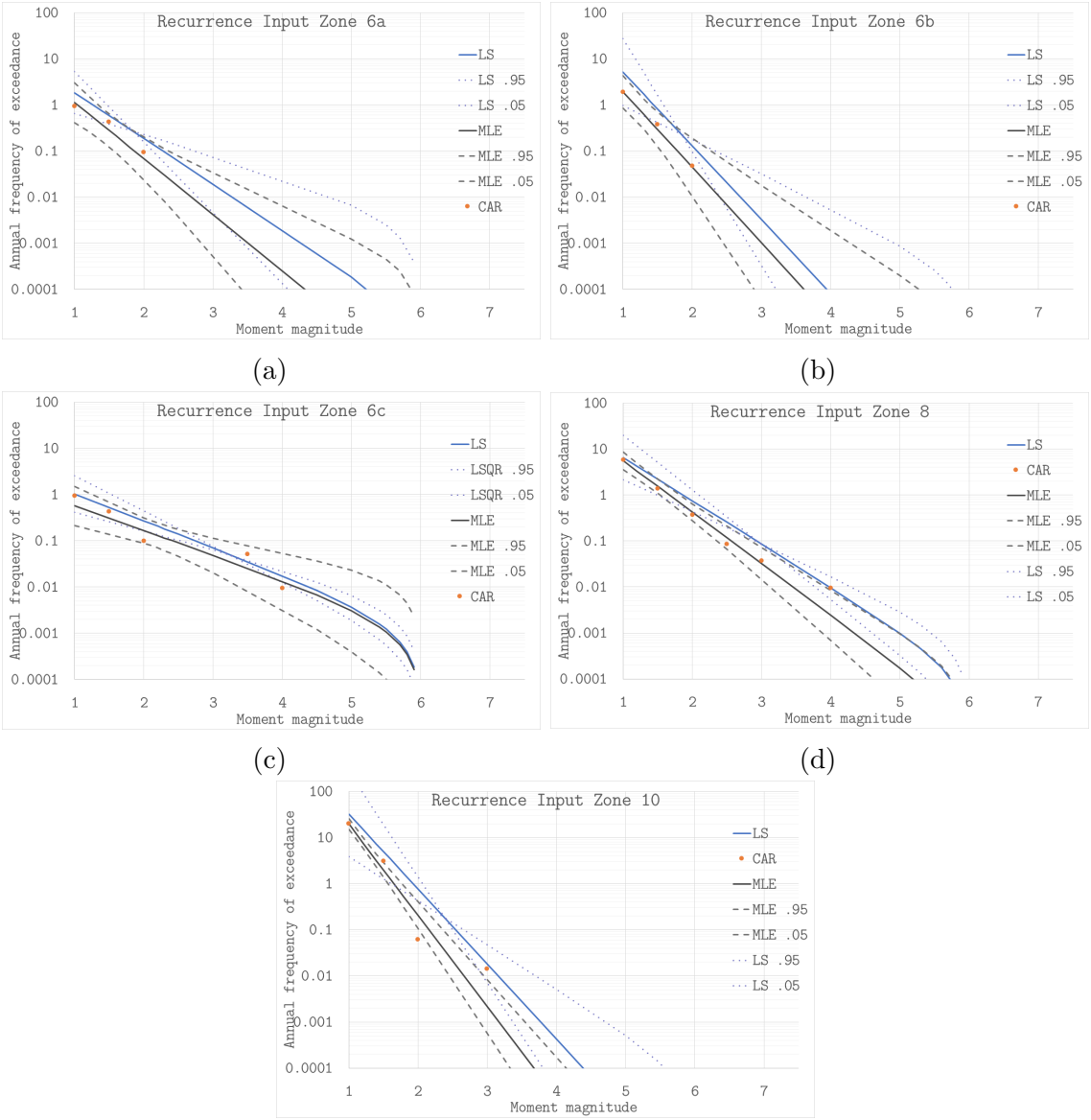


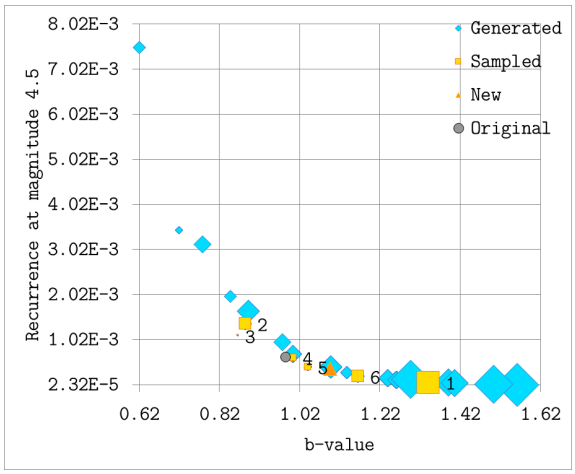
Figure B1: Recurrence curves calculated with LS with center magnitude interval datapoints (blue) and MLE estimate (gray). Cumulative annual earthquake rates (CAR) at lower bounds of the intervals (orange points). Ten percent confidence intervals for LS (dots) and MLE (dashes). Curves are drawn with maximum magnitude 6.

C Results for addition of new data

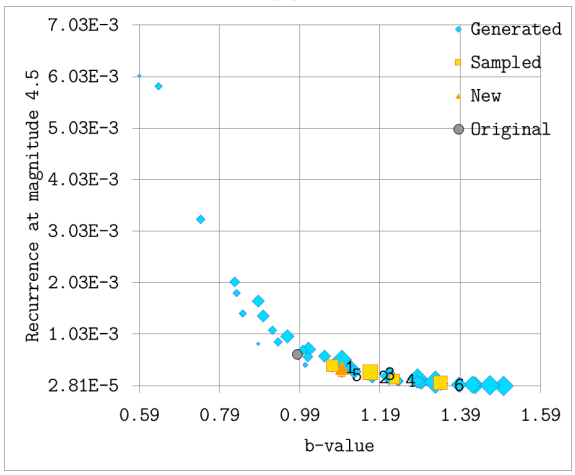
Recurrence parameters b and $n_{\min}$ at magnitude 4.5 are plotted after addition of new data in Figure C1 for area 6a, Figure C2 for area 6b and Figure C3 for area 6c. The Loviisa hazard estimate changes after using the chosen recurrence parameter samples are listed in Table C1. Changes after adding the unmodified new FENCAT data are also presented.

Years added, sample	**10 Hz**		**100 Hz**	
	Absolute (1/a)	**Relative**	**Absolute (1/a)**	**Relative**
6 a, sample 6	$1.07{\cdot}10^{-5}$	40.40 %	$5.95{\cdot}10^{-6}$	45.07 %
6 a, sample 5	$5.02{\cdot}10^{-6}$	19.05 %	$3.28{\cdot}10^{-6}$	24.82 %
6 a, sample 4	$3.09{\cdot}10^{-6}$	11.70 %	$1.77{\cdot}10^{-6}$	13.44 %
6 a, sample 3	$9.82{\cdot}10^{-7}$	3.72 %	$1.04{\cdot}10^{-6}$	7.89 %
6 a, sample 2	$8.80{\cdot}10^{-6}$	33.36 %	$5.43{\cdot}10^{-6}$	41.13 %
6 a, sample 1	$6.77{\cdot}10^{-6}$	25.67 %	$4.85{\cdot}10^{-6}$	36.76 %
3 a, sample 6	$6.90{\cdot}10^{-7}$	2.61 %	$7.10{\cdot}10^{-7}$	5.38 %
3 a, sample 5	$1.33{\cdot}10^{-6}$	5.06 %	$1.14{\cdot}10^{-6}$	8.60 %
3 a, sample 4	$6.88{\cdot}10^{-6}$	26.06 %	$4.31{\cdot}10^{-6}$	32.63 %
3 a, sample 3	$4.04{\cdot}10^{-6}$	15.33 %	$2.60{\cdot}10^{-6}$	19.68 %
3 a, sample 2	$3.23{\cdot}10^{-6}$	12.25 %	$2.44{\cdot}10^{-6}$	18.51 %
3 a, sample 1	$5.53{\cdot}10^{-6}$	20.96 %	$3.83{\cdot}10^{-6}$	29.02 %
1 a, sample 6	$4.43{\cdot}10^{-6}$	16.81 %	$2.66{\cdot}10^{-6}$	20.18 %
1 a, sample 5	$1.14{\cdot}10^{-7}$	0.43 %	$1.49{\cdot}10^{-7}$	1.13 %
1 a, sample 4	$3.60{\cdot}10^{-7}$	1.36 %	$6.37{\cdot}10^{-7}$	4.82 %
1 a, sample 3	$3.92{\cdot}10^{-6}$	14.85 %	$2.33{\cdot}10^{-6}$	17.69 %
1 a, sample 2	$1.35{\cdot}10^{-6}$	5.12 %	$1.03{\cdot}10^{-6}$	7.79 %
1 a, sample 1	$1.97{\cdot}10^{-6}$	7.47 %	$1.51{\cdot}10^{-6}$	11.48 %
FENCAT	$1.02{\cdot}10^{-5}$	38.70 %	$7.24{\cdot}10^{-6}$	54.82 %

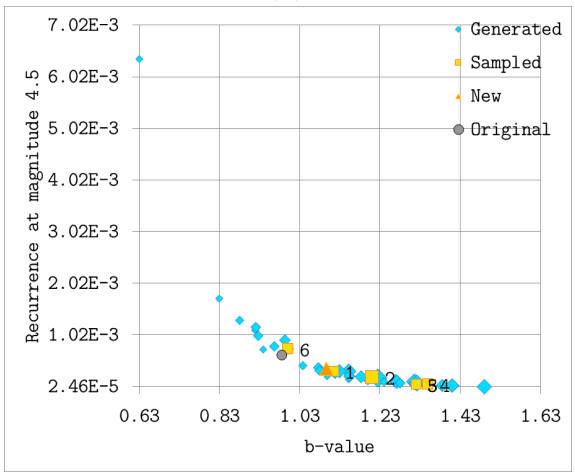
Table C1: Loviisa hazard estimate change at 10 Hz, 0.1 g and 100 Hz, 0.05 g after adding generated data. Also included is the estimates after addition of the acquired 7-year FENCAT dataset.



(a)

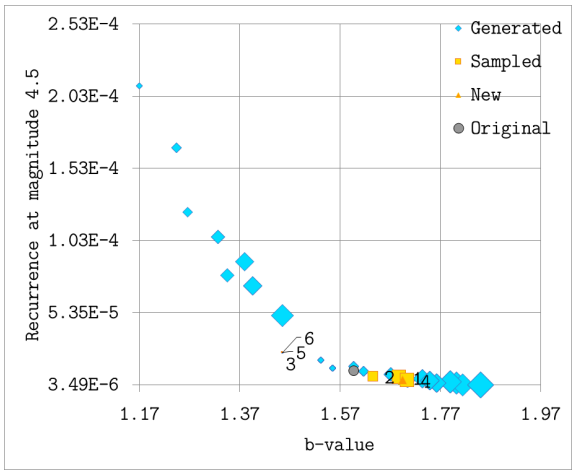


(b)

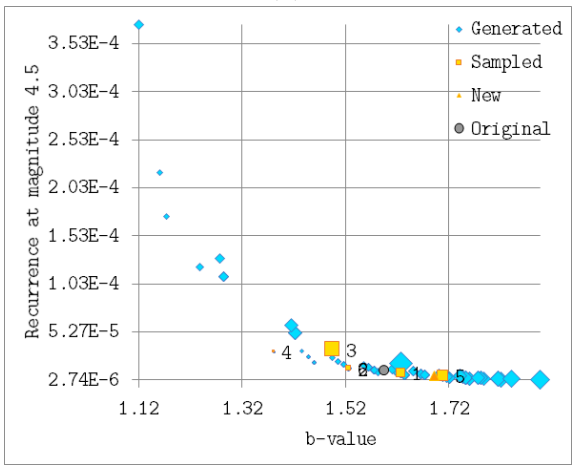


(c)

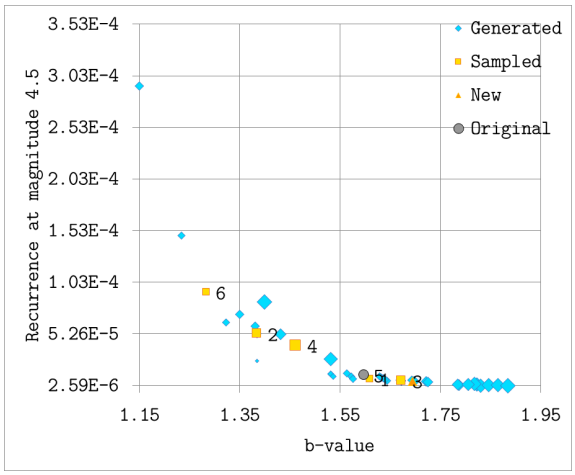
Figure C1: Scatter plot of b-values and $n_{\min}$ at magnitude 4.5 after adding generated earthquake data for source area 6a. **(a)** 1 year of added data. **(b)** 3 years of added data. **(c)** 6 years of added data. Numbered squares are the samples included for hazard calculation. Also shown are the original values (gray sphere) and values after adding the whole acquired dataset (orange triangle). The sizes of the sample points are scaled by the total number of added events and divided by the original number of events.



(a)

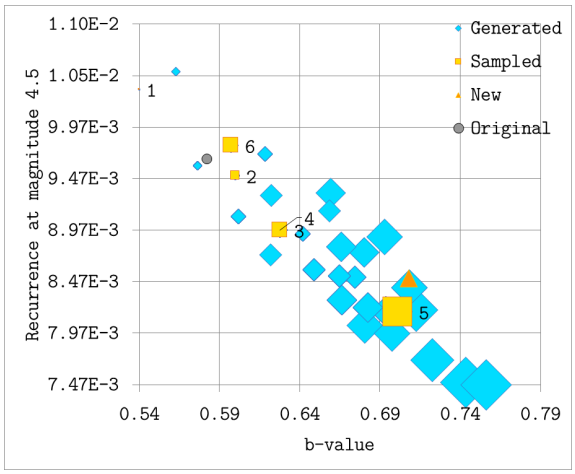


(b)

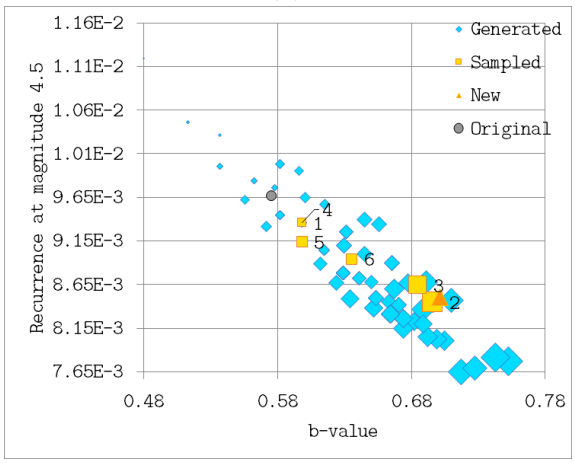


(c)

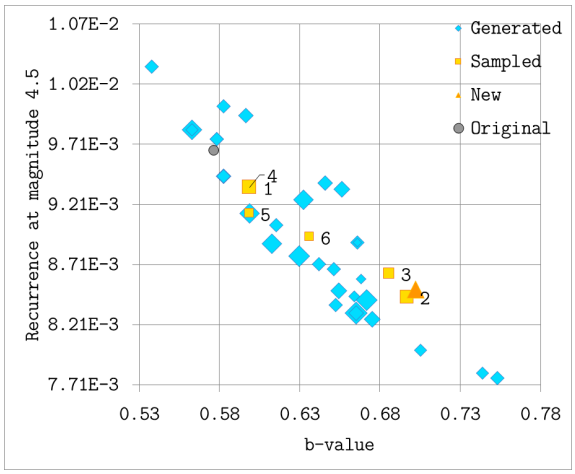
Figure C2: Scatter plot of b-values and $n_{\min}$ at magnitude 4.5 after adding generated earthquake data for source area 6b. **(a)** 1 year of added data. **(b)** 3 years of added data. **(c)** 6 years of added data. Numbered squares are the samples included for hazard calculation. Also shown are the original values (gray sphere) and values after adding the whole acquired dataset (orange triangle). The sizes of the sample points are scaled by the total number of added events and divided by the original number of events.



(a)



(b)



(c)

Figure C3: Scatter plot of b-values and $n_{\min}$ at magnitude 4.5 after adding generated earthquake data for source area 6c. **(a)** 1 year of added data. **(b)** 3 years of added data. **(c)** 6 years of added data. Numbered squares are the samples included for hazard calculation. Also shown are the original values (gray sphere) and values after adding the whole acquired dataset (orange triangle). The sizes of the sample points are scaled by the total number of added events and divided by the original number of events.

D Loviisa 100 Hz figures

All hazard results were acquired for 10 Hz and 100 Hz frequencies. Here is presented some figures for 100 Hz that were left out of the main text. Figure D1 shows the tornado plot of all parameter variations. Figure D2 gives the mean hazard curves for each SSA. Figure D3 contains the hazard curves for MLE and shifted datapoint LS methods.

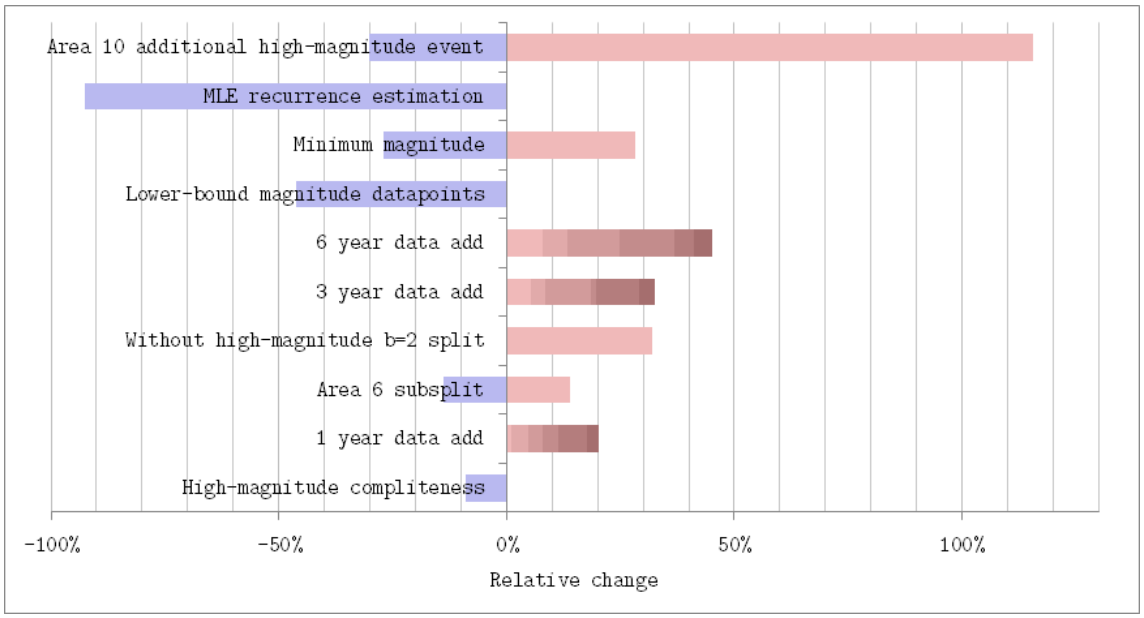
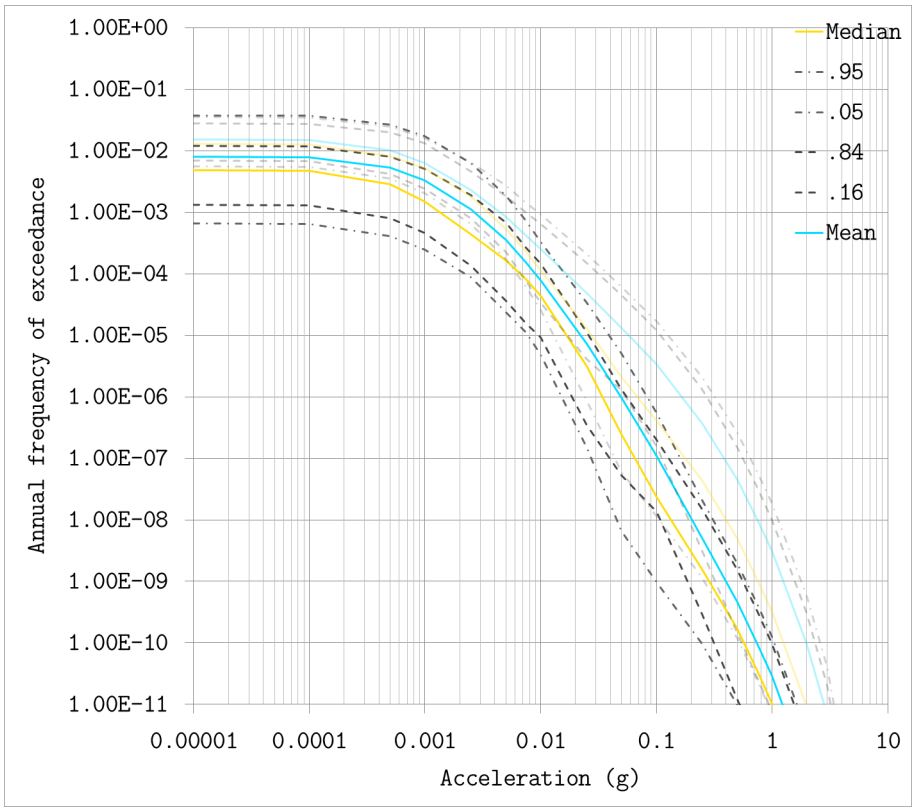
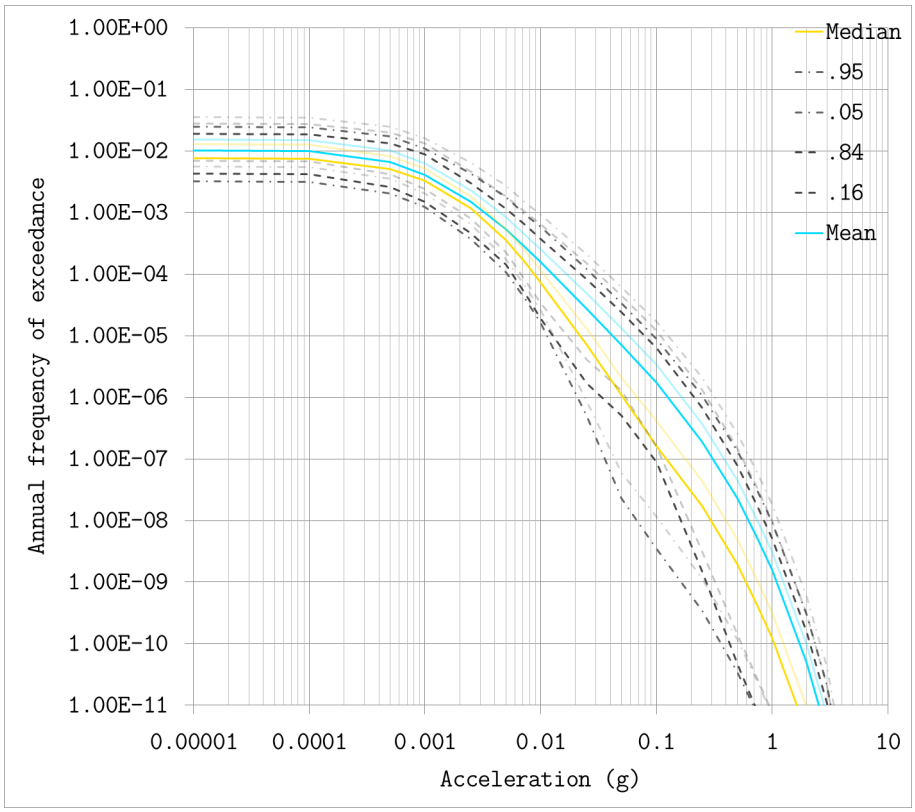


Figure D1: Relative Loviisa hazard estimate changes at 100 Hz for each parameter variation. Each color and shade denotes a single result. A positive 100 % change means that the hazard estimate is doubled, while a negative 100 % means that the estimate becomes zero.

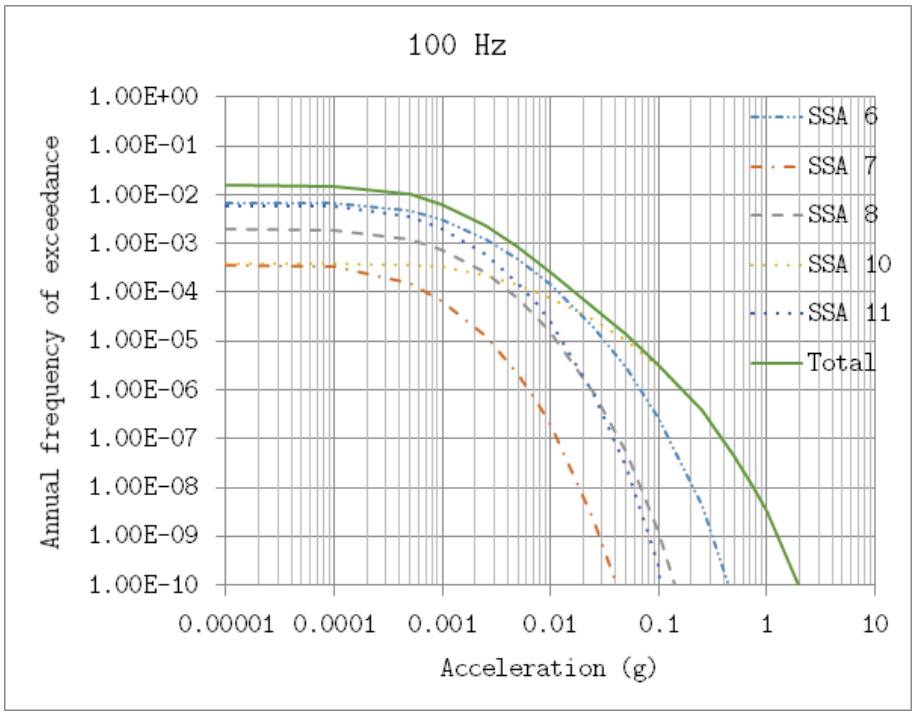


(a)

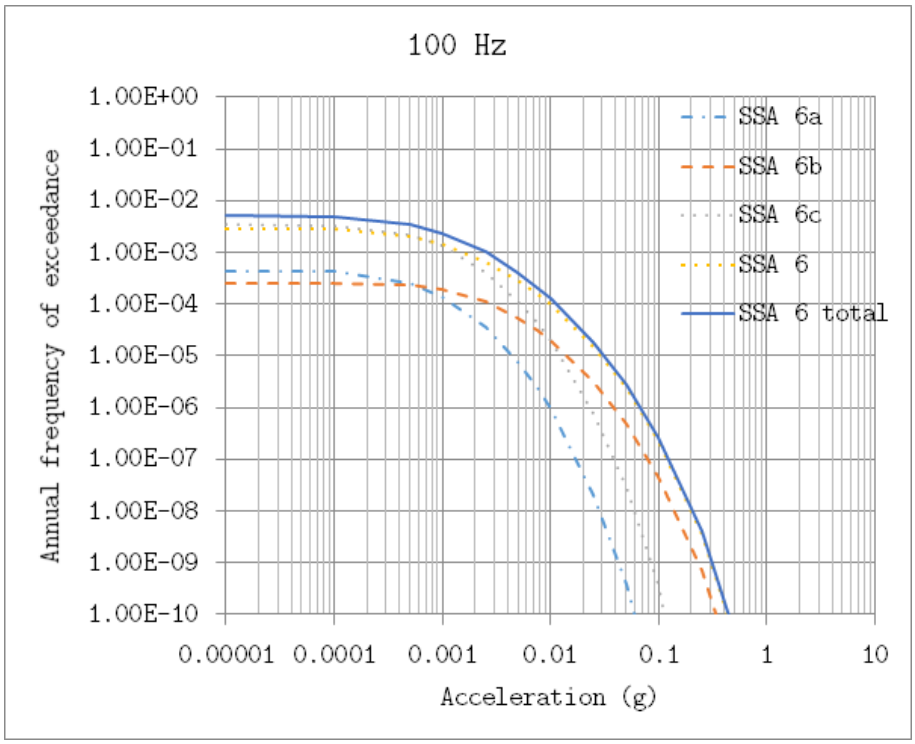


(b)

Figure D3: Loviisa 100 Hz hazard curves with GR-parameters estimated by **(a)** MLE method and **(b)** LS with points shifted to lower bounds of the magnitude intervals. Original curves are plotted as a reference with lighter colour.



(a)



(b)

Figure D2: **(a)** Loviisa 100 Hz mean hazard curves from each SSA and the total corresponding to the sum of the individual curves. **(b)** Individual weighted mean hazard curves of the SSA 6 subareas and the unsplit area. Total hazard curve of SSA 6.

E Results for Olkiluoto powerplant

Hazard estimate changes for different parameter vatiations are given in table E1. The sensitivity analysis of adding generated data was performed for 1 and 6 years of added data. The results are presented in table E2. Changes after adding the unmodified new FENCAT data are also presented. The hazard estimate changes for different parameter vatiations are visualized with a tornado plot Figure E1 for 100 Hz.

Table E1: Absolute and relative hazard estimate changes for Olkiluoto for 10 Hz, 0.1 g and 100 Hz, 0.05 g.

Parameter change	**10 Hz Relative**	**100 Hz Relative**
Area 10 additional 3–3.5 event	0.00 %	0.00 %
High-magnitude compliteness	-14.36 %	-14.64 %
Without high-magnitude b split	42.73 %	31.52 %
Minimum magnitude 4	-2.43 %	3.91 %
Minimum magnitude 5	-1.29 %	-8.31 %
MLE recurrence estimation	-80.90 %	-80.23 %
Lower-bound magnitude datapoints	-30.13 %	-30.25 %
Area 6 subdivision 100%	37.49 %	36.05 %
Area 6 subdivision 0%	-37.49 %	-36.05 %

Years added, sample	**10 Hz Relative**	**100 Hz Relative**
6 a, sample 6	13.05 %	13.05 %
6 a, sample 5	-50.99 %	-49.97 %
6 a, sample 4	-53.65 %	-52.45 %
6 a, sample 3	-51.90 %	-50.87 %
6 a, sample 2	-32.05 %	-30.99 %
6 a, sample 1	-23.66 %	-22.94 %
1 a, sample 6	-32.80 %	-31.96 %
1 a, sample 5	-20.12 %	-19.60 %
1 a, sample 4	11.74 %	11.35 %
1 a, sample 3	19.24 %	18.49 %
1 a, sample 2	38.60 %	37.71 %
1 a, sample 1	-48.35 %	-47.19 %
Full FENCAT	-37.95 %	-36.49 %

Table E2: Olkiluoto hazard estimate change at 10 Hz and 100 Hz after adding generated data. Also included is the estimates after addition of the full acquired FENCAT dataset.

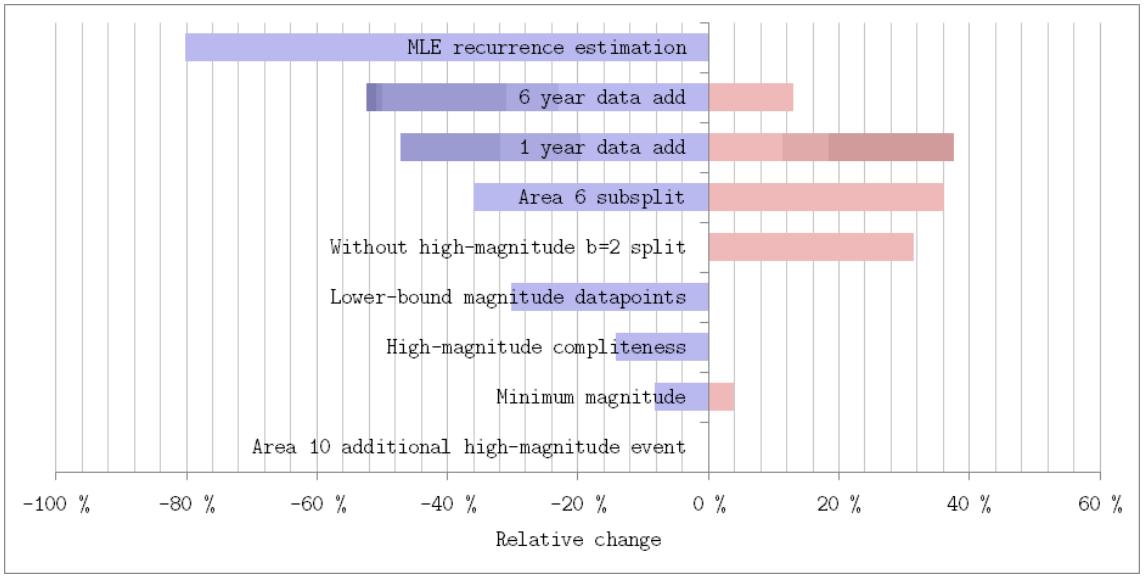


Figure E1: Relative hazard estimate changes at 100 Hz for each parameter variation. Each color and shade denotes a single result. A positive 100 % change means that the hazard estimate is doubled, while a negative 100 % means that the estimate becomes zero.