Aalto University
MS-E2177: Seminar on Case Studies in Operations Research

Probabilistic Seismic Hazard Analysis for a Nuclear Power Plant

Final Report

Project Manager: Tran Quang Anh Tuan

Team Members: Anh Dao, Eeli Asikainen, Einari Stenberg, Waldemar Sorjonen

Date: June 3, 2025

Contents

Symbols and Abbreviations

Symbols

m_i	Magnitude of earthquake i
$m_{\min}$	Minimum magnitude considered in the PSHA calculations
$m_{\max}^{\text{obs}}$	Maximum observed magnitude
$m_{\max}$	(True) Maximum magnitude
m^u	Candidate value for $m_{\max}$ in the Bayesian formulation
$\mathrm{M_w}$	Moment magnitude
N_{obs}	Number of earthquakes in the earthquake catalog
N	Completeness corrected number of earthquakes
$N_{\geq \min}$	Number of earthquakes with magnitudes equal to or larger than $m_{\min}$
$\hat{\mu}$	Sample mean
$\hat{\sigma}$	Sample standard deviation
$\mathcal{N}(\mu, \sigma)$	Normal distribution with mean μ and standard deviation σ
a, α	Intercept in the Gutenberg-Richter law, $\alpha = a \cdot \ln(10)$
b, β	Slope in the Gutenberg-Richter law, $\beta = b \cdot \ln(10)$
W_{area}	Spatially weighted magnitude average
W_{SSZ}	Weight distribution of magnitudes within a given seismic source zone
n_{SSZ}	Number of earthquakes in a given seismic source zone
$\exp\{\cdot\}$	Exponential function
$\ln(\cdot)$	Natural logarithm
$F[m_{\max}^{\text{obs}}]$	CDF of the observed maximum magnitude
$F_M(m)$	CDF of earthquake magnitudes up to m in the Kijko formulation
$L[m^u]$	Likelihood function for m^u
$Pr(m_{\max} < z)$	Probability that the (true) maximum magnitude is less than z

Abbreviations

NPP	Nuclear Power Plant
EPRI	Electric Power Research Institute
PSHA	Probabilistic Seismic Hazard Analysis
PRA	Probabilistic Risk Assessment
UH	University of Helsinki
SSZ	Seismic Source Zone
CEUS	Central and Eastern United States
SCR	Stable Continental Region
DN	Domain
SD	Superdomain
GR	Gutenberg-Richter
CDF	Cumulative Distribution Function

1 Introduction

1.1 Background

Given the push for renewable energy sources around the world, nuclear energy has been heralded by many as an important addition to the array of clean solutions. However, a major challenge to nuclear acceptance comes from its potential risks. Unlike other energy sources, nuclear energy incidents can push public sentiment against it for years, which happened with the accidents at Chernobyl in 1986 and Fukushima in 2011. This means that nuclear power plants are designed with extremely strict safety standards.

A major standard involves safety against seismic activity, giving rise to Seismic Hazard and Risk Analysis as a key research field. At its core, Probabilistic Seismic Hazard Analysis (PSHA) models the likelihood and consequences of earthquake activity using recorded seismic data. PSHA is also a critical component of Probabilistic Risk Analysis (PRA), which evaluates the overall risk to a nuclear power plant by integrating multiple hazard and system failure scenarios.

At Fortum's Loviisa nuclear power plant, PRA models have been updated annually for decades. However, Finland's low seismic activity presents challenges for accurate risk estimation. The scarcity of relevant seismic data further complicates the modeling process, necessitating the adaptation of new methods to suit Loviisa's specific conditions.

This project aims to enhance Loviisa's PSHA by addressing these limitations. We focus on developing improved models by identifying more suitable methodologies and refining existing approaches. A key element of this effort is the assessment of the maximum possible earthquake magnitude ($m_{\max}$), which is critical in defining the upper bounds of seismic hazard. Since the underlying risk models do not impose an inherent limit on how large $m_{\max}$ can be, a carefully justified and site-appropriate estimation is essential. Accurate determination of $m_{\max}$ ensures that the risk assessment accounts for the most extreme, yet practically plausible, seismic scenarios, contributing to the overall resilience and safety of the nuclear facility.

1.2 Objectives

The primary objective of this project is to refine and adjust existing statistical models for estimating the maximum magnitudes of earthquakes of Loviisa's PSHA. As current models are generally not tailored for application in low-seismicity regions such as Loviisa, targeted modifications are required to enhance their relevance and reliability. In addition, the methodologies underlying these models are often not

sufficiently detailed in the existing literature, necessitating the development of practical tools for their reconstruction, calibration, and validation.

To further support transparency and reproducibility, Fortum requested that the project also document the use of the Bayesian method described in Section 4.1, which had been employed in Fortum's PSHA 2021 by an external consulting firm. This supplementary task complements the model refinement effort by addressing gaps in documentation and clarifying previous methodological choices.

Although the core focus of the project remains on the development of robust statistical models, particular attention is paid to the systematic documentation and rigorous justification of the research process and outcomes. Given the inherent complexity of the matter, it is anticipated that a fully definitive solution may not be attainable. Therefore, a critical aspect of this study will be to elucidate the limitations of various approaches, analyze the reasons behind their shortcomings, and assess the strategies employed to address these challenges. By documenting unsuccessful approaches alongside successful ones, this research aims to contribute to a broader understanding of earthquake modeling in low-seismicity regions, providing a foundation for future improvements.

2 Literature Review

2.1 Probabilistic Seismic Hazard Analysis

This literature review critically examines the methods commonly applied in PSHA, particularly focusing on their suitability and limitations within low-seismicity regions like Finland. The scarcity of seismic activity in these low activity regions make it challenging to get reliable datasets for reliable models. We focus especially on model selection and on analyses of the prior usage of the chosen methods: the Bayesian approach and Kijko's method. We also aim to suggest alternative approaches to conducting PSHA based on the existing literature.

Probabilistic Seismic Hazard Analysis is important in evaluating earthquake risks for nuclear power plants. Nuclear energy has a growing importance as part of the global shift toward renewable and cleaner energy sources. As a part of this shift the energy companies have even more responsibility in creating and maintaining the safety and reliability of these facilities. Earthquakes remain a key concern due to their potentially large consequences, and this is why PSHA is still getting a lot of attention.

PSHA was first introduced in the literature in 1986 [1]. Cornell [1] modeled the earthquake hazard using four elements: source characterization, recurrence rate, maximum magnitude, and ground motion prediction. This conventional Cornell source-based approach involves using all earthquake magnitudes and sources to get the annual probability of exceeding a certain ground motion at a location of interest. This method has been widely implemented and several alternative approaches have been developed.

The application of PSHA to low-to-moderate seismicity regions posed new unique problems. This led to development of many new approaches. For example Historical method [2] uses past earthquake data to estimate future hazard with alternative modeling techniques, Frankel method [3] employs smoothed historical seismicity to calculate probabilistic amplitudes, Kijko and Graham method [4] uses "deductive" and "historic" procedures, and the direct Amplitude-Based approach [5] ground-motion attenuation model to refine PSHA to be more accurate.

It has been shown that using PSHA in areas like Finland can be very challenging. Fülöp et al. [6] investigated how PSHA results change if some parameters were changed for a model developed for low-seismicity areas. The study shows that parameters such as the b-value, seismicity rates, the largest possible magnitude ($m_{\max}$), and the median ground motion prediction equation significantly influence PSHA results in low seismicity areas. They also note that this result does not restrict itself

to Finland but is applicable to the whole Europe. The result also applies to updating parameter values, and the current model that Fortum is using might experience significant changes if the updated $m_{\max}$ values are implemented.

2.2 Estimation of Maximum Magnitude

As mentioned, PSHA has difficulties with robustness in regions of low-seismicity. This arises mostly because the area lacks sufficient earthquake catalogs, meaning there is not enough data to achieve significant results. Furthermore, it has been shown that PSHA becomes somewhat sensitive to parameter changes when there is not enough data backing the model. Thus, updating and refining the model parameters becomes very important. Furthermore, we aim to analyze the sensitivity of these parameters to understand how methodological choices influence the results.

Many quantitative approaches have been developed to estimate the $m_{\max}$ from seismicity catalogs. As reviewed by Pisarenko & Robin [7], these methods usually rely on statistical analysis of observed earthquake magnitudes. Pisarenko and Robin reviewed different methods and compared them with each other. They looked at Moment method, Bayesian method, extreme-value theory, order-statistics estimators and Kijko method.

The moment method matches the sample mean of observed magnitudes to the same moments of truncated Gutenberg-Richter law. This method works as a good baseline but has a large sensitivity to large events when the catalog used is small. The Bayesian method has a similar issue: if the dataset is small, the prior can dominate the outcome, giving too much weight to prior beliefs and assumptions.[7]

Extreme value theory only looks at magnitudes above a certain threshold. Then, these "peaks" are fitted with a generalized pareto distribution. This way an upper limit or an upper quantile can be calculated. This method is also sensitive if there is not enough data, but can be reliable if there is at least half a dozen data points.[7]

The Kijko method was introduced in 1998 by Kijko and Graham [8]. The method fits a Gutenberg-Richter model to a combined catalog with different completeness levels. This way, datasets collected with different accuracies can be combined. For areas with sparse data, this method is useful as it can be used for incomplete catalogs. This suggests that the Kijko method is useful for Loviisa as well.

Overall, the estimation of maximum magnitudes is a difficult task for low-seismicity areas. Thus, multiple methods should be combined and weighted to achieve the best possible result. However, even with multiple methods there should be thorough sensitivity analysis to assess the validity of the results.

3 Data

3.1 Stable Continental Region Earthquake Catalog

The Stable Continental Region (SCR) Earthquake catalog for this project is a global database of earthquakes from the historic era up to 2003. It includes two components: the earthquake catalog itself and the catalog of crustal domains. The catalog has also been declustered, meaning that dependent events, such as aftershocks, have been removed, leaving only independent earthquakes in the dataset.

A core parameter of the catalog is the uniform moment magnitude (M_w), estimated based on correlations with other size measures such as magnitude scales of instrumentally recorded earthquakes or shaking intensity measures for pre-instrumental earthquakes. While the catalog was originally developed in 1994 (see [9]), it has been expanded with earthquake data recorded from 1991 to 2003 which are utilized in the updated ERPI study in 2012 [10].

3.1.1 Completeness Correction

To account for potential incompleteness of the catalog regarding sample size (N_{obs}) of earthquakes with magnitude $M_w \geq 4.5$, a process of completeness correction was introduced in [9] to get the corrected sample size. The process estimates catalog completeness periods for portions of the SCR. These periods were then utilized to create a completeness-corrected equivalent earthquake count for each domain based on the catalog completeness period for magnitudes equal to $m_{\text{max}}^{\text{obs}}$. For example, if the observed magnitude is smaller than $m_{\text{max}}^{\text{obs}}$, the completeness period is shorter than the period for $m_{\text{max}}^{\text{obs}}$.

Given the periods covered in the catalog, and assuming earthquake occurrence rates are stationary, the maximum likelihood estimate of the rate of earthquakes m_i depends on the observed number of earthquakes, $N_{\text{obs}}(m_i)$ in the completeness period for m_i divided by the length of the completeness period, $T_C(m_i)$. The total number of earthquakes that would have been recorded in the completeness period for $m_{\text{max}}^{\text{obs}}$ can then be estimated using the estimated rate of occurrence for magnitude m_i multiplied by $T_C(m_{\text{max}}^{\text{obs}})$:

$$N_{\text{Completeness Corrected}}(m_i) = N_{\text{In Period } T_C(m_i)} \times \frac{T_C(m_{\text{max}}^{\text{obs}})}{T_C(m_i)} \tag{1}$$

The values of $N_{\text{Completeness Corrected}}(m_i)$ are summed up to get the completeness-corrected sample size for each domain. We refer to this result with the simple notation N.

3.1.2 Division of Domains and Superdomains

As described in Chapter 5 of [9] and Chapter 5.2.1 of [10], the catalog has been divided into a combination of domains and superdomains. SCR tectonic crusts have been divided into 255 domains by Johnston et al. in [9] which were reviewed and updated in terms of priors in the 2012 EPRI report [10]. These domains have then been grouped together based on shared characteristics to form superdomains. One of the key classifications used was whether the domain lies in an extended crust or non-extended crust. Afterwards, properties like crustal age, stress state were used to further divide the domains into specific superdomains. The result is a set of superdomains for calculating the priors specific to certain regions. Figure 1 shows a list of earthquakes recorded in the SCR catalog according to their respective superdomains.

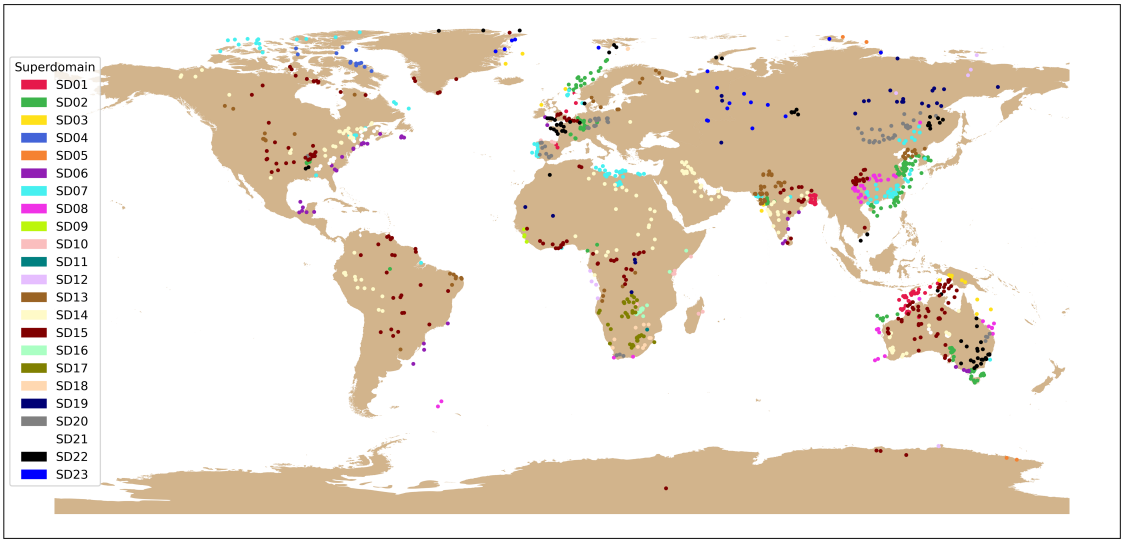


Figure 1: Earthquakes by superdomain in the SCR catalog

3.2 University of Helsinki Earthquake Catalog

The catalog [11], developed by the University of Helsinki (UH) in 2023, is a declustered update of the UH 2016 catalog [12], with declustering also carried out by UH. It contains 22,123 seismic observations spanning the period from 1467 to 2021. This updated catalog was assembled in a manner similar to the 2016 version, with magnitudes homogenized and converted to moment magnitudes, which will be used throughout this report. The processing methods are detailed in the 2016 study by Korja et al. [12] and briefly summarized in the 2024 Master's thesis by Vehmas [13]. The maximum observed magnitude ($m_{\max}^{\text{obs}}$) in the entire catalog is 4.5, while within the seismic source zones (SSZ) located within a 300 km radius of the Loviisa Nuclear Power Plant (NPP), the maximum observed magnitude is 4.4.

Vehmas also conducted completeness analysis for the catalog, resulting in two versions, with and without SSZ groups. In this context, a SSZ is a region where seismic activity is assumed to be uniformly distributed based on geological and seismic properties. In this report, we will be using the completeness analysis data without SSZ groups. The SSZs and earthquake observations used in [14] are shown in Figure 2.

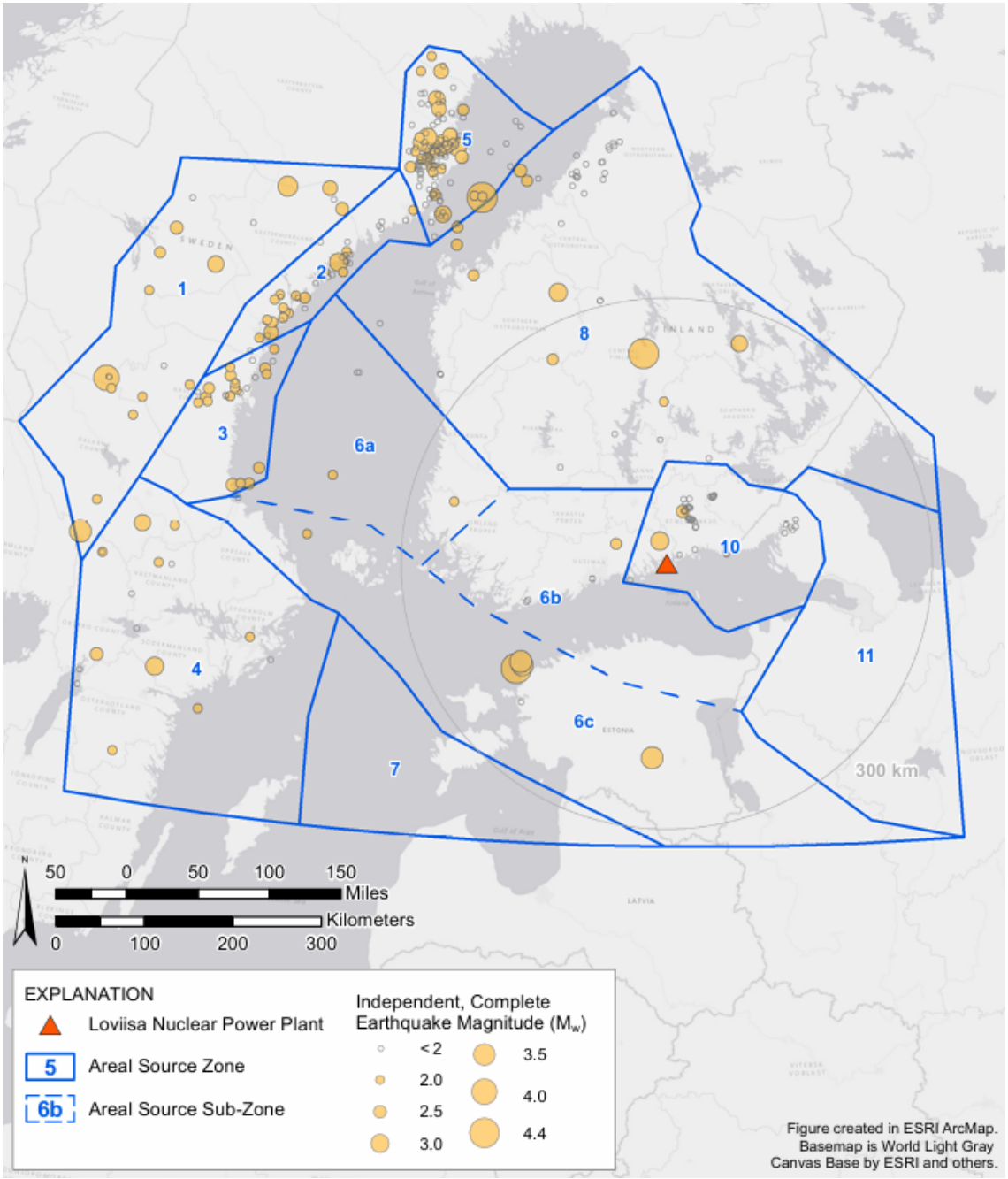


Figure 2: SSZs and considered earthquake observations in Loviisa's PSHA 2021 [14]

4 Methodology

This project seeks to estimate the maximum magnitude ($m_{\max}$) by evaluating and refining two of the most widely applied approaches: the Bayesian method developed in the 1994 EPRI study [9] and the method proposed by Kijko [8]. These methods are assessed in the context of low-seismicity regions, with particular attention to their applicability and limitations in Loviisa's PSHA.

4.1 Bayesian Method

The Bayesian method is grounded in Bayes' theorem [15]

$$p(\theta|y) = \frac{p(\theta)p(y|\theta)}{p(y)}, \tag{2}$$

where $p(\theta|y)$ denotes the posterior distribution of the unknown parameter θ given the observed data y, $p(\theta)$ is the prior distribution of θ, $p(y|\theta)$ is the likelihood function, and $p(y) = \int p(\theta)p(y|\theta)d\theta$ is the marginal distribution of the data. In essence, Bayes' theorem provides a mechanism for updating prior beliefs about a parameter in light of new evidence. One of the main advantages of the Bayesian framework is its capacity to incorporate existing knowledge, which is particularly valuable when dealing with limited data—a common challenge in not just low-seismicity regions but all around the world.

For estimating $m_{\max}$, the Bayesian method was first introduced in the 1994 EPRI study [9] by Johnston et al., who developed a probabilistic framework for Central and Eastern United States (CEUS). The method involved constructing prior distributions based on analogies with seismically comparable regions, followed by updating these priors using regional earthquake data for events of magnitude $\mathrm{M_w} \geq 4.5$.

Johnston et al. also proposed a set of globally applicable prior distributions, intended to be updated with local data from regions outside the CEUS. However, in the case of Loviisa, the scarcity of high-magnitude seismic events means that the posterior distribution of $m_{\max}$ would be largely influenced, if not entirely dominated, by the global prior. Since these global priors are based predominantly on high-seismicity regions, their direct application may misrepresent the seismic characteristics of Loviisa, thereby limiting the method's relevance in this context.

To evaluate and address these challenges, we adopt a multi-step approach. First, we replicate the original EPRI computations to reconstruct the prior distributions. We then explore alternative data subsets for the development of priors that may better reflect the characteristics of low-seismicity regions. Finally, we assess the potential for updating these priors using available data from the vicinity of the Loviisa NPP.

In response to Fortum's request for greater methodological transparency, we also provide a detailed account of how the Bayesian method was applied in Fortum's PSHA 2021.

4.1.1 Prior Distribution Computation

Based on the 1994 and 2012 EPRI studies [9, 10], the Bayesian prior distribution for $m_{\max}$ is estimated using data from the SCR catalog. The sample mean ($\hat{\mu}$) and sample standard deviation ($\hat{\sigma}$) of $m_{\max}^{\text{obs}}$ across selected superdomains are used to parameterize a normal distribution for $m_{\max}^{\text{obs}}$. This distribution, in turn, informs the prior for $m_{\max}$. For instance, in regions characterized by non-extended crustal types, the prior distribution is based on the mean and standard deviation of $m_{\max}^{\text{obs}}$ for SDs classified as non-extended.

However, because the observed maximum magnitudes tend to underestimate the true maximum magnitude, this approach introduces bias. To address this, Johnston et al. [9] propose a correction based on the cumulative distribution function (CDF) of $m_{\max}^{\text{obs}}$, defined as

$$F[m_{\max}^{\text{obs}}] = \left[\frac{1-\exp\{-b\ln(10)(m_{\max}^{\text{obs}}-m_{\min})\}}{1-\exp\{-b\ln(10)(m^u-m_{\min})\}}\right]^N, \text{ for } m_{\min} \le m_{\max}^{\text{obs}} \le m^u, \quad (3)$$

where $b\ln(10)$ represents the slope of the Gutenberg-Richter law computed for the selected subset of SDs, $m_{\min}$ is the minimum magnitude recorded in the SCR catalog, m^u is a candidate value for $m_{\max}$, and N denotes the completeness-corrected number of earthquakes. This formulation assumes that the size distribution of earthquakes within a source region follows a truncated exponential distribution bounded between $m_{\min}$ and m^u.

Solving Equation (3) for m^u yields

$$m^u = m_{\min} - \frac{1}{b\ln(10)}\ln\left(1-\frac{1-\exp\{-b\ln(10)(m_{\max}^{\text{obs}}-m_{\min})\}}{F[m_{\max}^{\text{obs}}]^{1/N}}\right), \quad \text{for } m_{\min} \le m_{\max}^{\text{obs}} \le m^u. \quad (4)$$

Since $m_{\max}^{\text{obs}}$ is assumed to be normally distributed, setting $F[m_{\max}^{\text{obs}}] = 0.5$, which corresponds to the median, is equivalent to obtaining $\hat{\mu}$. This median is then bias-corrected using (4) to estimate the mean of m^u which gives the mean of the $m_{\max}$ distribution. The standard deviation of the $m_{\max}$ prior is taken to be the same as the original $\hat{\sigma}$.

Furthermore, the sample size N should be sufficiently large (preferably greater than 10) to ensure the bias-corrected estimate remains finite and statistically reliable. All in all, as all required inputs for Equation (4) can be derived from the SCR catalog, this method offers a self-contained framework for constructing the Bayesian prior distribution for $m_{\max}$.

4.1.2 Prior with Superdomains

Given the significant influence of the prior distribution in the Bayesian model, it is essential to construct a prior that appropriately reflects the seismic characteristics of the Loviisa region while still incorporating as much global data as possible. One viable solution is to use of superdomains, as introduced in the 1994 EPRI study [9].

In this project, we base the prior distribution on catalog data from superdomain 13 (SD13), which includes Loviisa. As illustrated in Table 1 and Figure 3, SD13 encompasses 12 seismic domains (DNs) distributed across six global regions. For each region, GR b-values were obtained from the 1994 EPRI report. The overall b-value for SD13 is calculated as the arithmetic mean of the regional values, resulting in an approximate value of 1.03.

To estimate the mean of the $m_{\max}$ prior, we first compute the average observed maximum magnitude ($m_{\max}^{\text{obs}}$) across the superdomain. This value is then bias-corrected following the procedure described in Section 4.1.1. The resulting corrected mean, along with the standard deviation derived from the superdomain data, forms the basis for the Bayesian prior distribution.

Table 1: Domain-wise completeness-corrected number of earthquakes, observed maximum magnitudes, and regional GR b-values in SD13

DN	N	$m_{\max}^{\text{obs}}$ (M_w)	Region	Regional GR b
113	71.7	6.7	China	1.029
133	107.7	6.7	India	0.966
236	2	6.0	North America	0.790
42	1	4.9	South America	1.212
56	8.4	5.4	South America	1.212
72	13.6	6.2	Africa	0.982
171	2	4.7	Europe	1.156
173	3	5.3	Europe	1.156
175	2	4.8	Europe	1.156
177	2	5.9	Europe	1.156
238	9.3	5.3	North America	0.790
242	2.5	5.2	North America	0.790
Mean	18.8	5.6		1.033

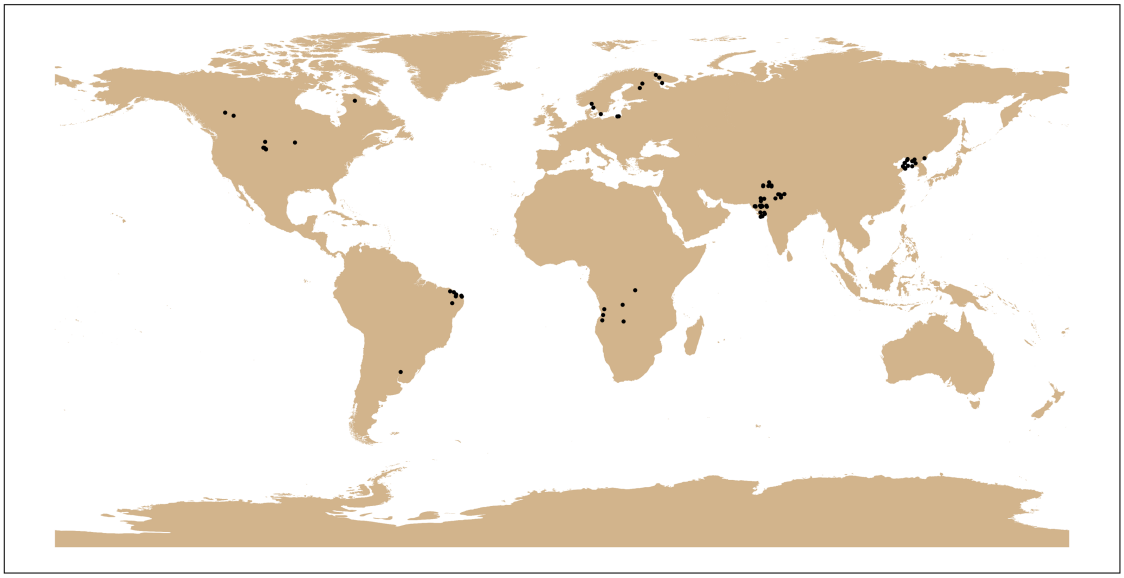


Figure 3: Earthquakes of SD13 in the SCR catalog

While this method enables the incorporation of geographically broad data into the prior, it has its limitations. As noted in the EPRI reports [9, 10], the statistical robustness of the criteria used to define superdomains is relatively weak. This observation is supported by our analysis: a one-way ANOVA conducted on SDs comprising more than one domain yields a p-value of approximately 0.14, indicating that the observed variation in $m_{\max}^{\text{obs}}$ across SDs may be attributable to random chance rather than systematic differences.

Nonetheless, given current data constraints—particularly the global scarcity of high-resolution explanatory variables, such as rift lengths—superdomains likely remain a practical and justifiable framework for grouping seismically similar regions in a global context.

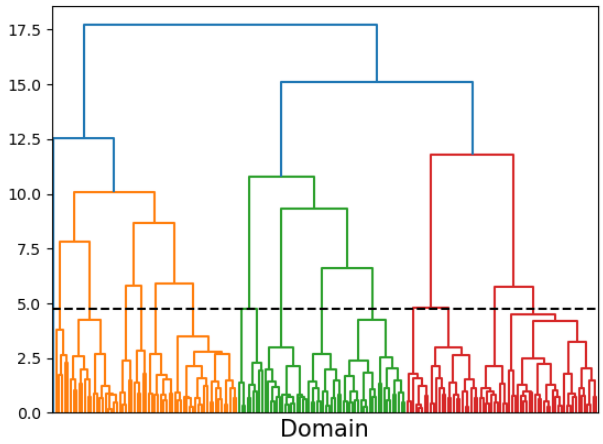
4.1.3 Prior with Clustering

To address the issues associated with superdomains, clustering is introduced as an alternative method to group domains based on their seismic characteristics, aiming to identify a set of domains representative of the Loviisa region. In cluster analysis, a set of objects are grouped such that similar objects, i.e. objects close to each other in the feature space, are in the same group. In this case, the objects are the domains of the SCR catalog, and the feature space consists of continuous variables describing tectonic, seismic, or spatial characteristics of those domains. Constructing the prior using domains with these characteristics helps ensure that it is both justifiable and representative of the Loviisa region.

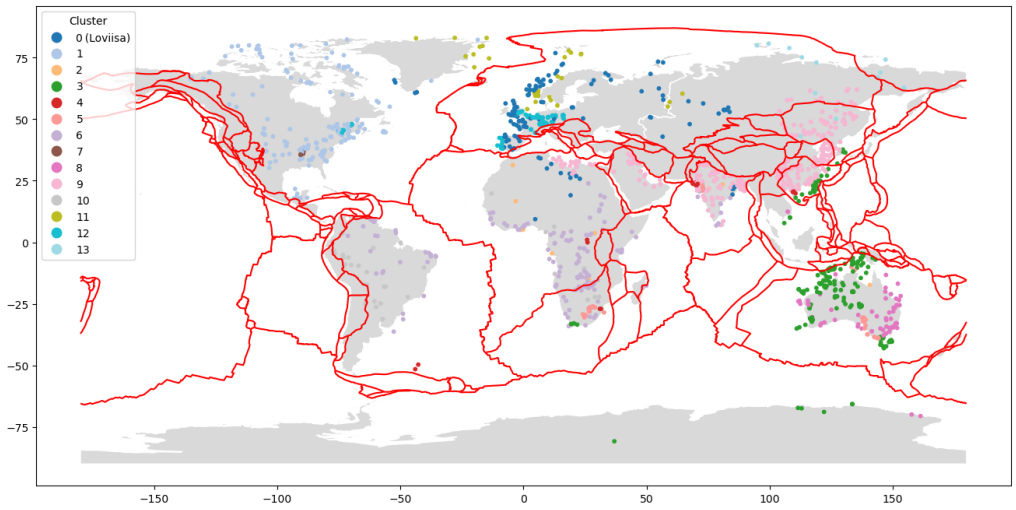
To implement the clustering, both k-means clustering and agglomerative hierarchical clustering were considered. Both methods are widely popular, trusted, and well documented. Hierarchical clustering was selected as the primary method since it is fully deterministic and does not require predefining the number of clusters. Euclidean distance and ward linkage were chosen due to their popularity in similar tasks. All variables are normalized because the method is affected by the scale of the variables. The clustering is carried out in two approaches: one based on seismic variables and the other on spatial variables.

In seismic clustering, domains are clustered based on seismic structure of historical earthquakes. The variables used are mean and standard deviation of earthquake magnitudes, frequency of earthquakes per area, latitude and longitude. The idea is that domains with similar earthquake structure would also be similar in terms of maximum magnitudes. However, there are notable limitations when relying on historical seismic variables to predict similarities in maximum magnitudes. The period during which humans have documented seismic events is relatively brief, resulting in a limited dataset. Consequently, similarities observed in this small sample over such a short time frame may not indicate future patterns reliably. This approach biases the results towards historical realizations, which may not be accurate due to the low sample size. The dendrogram and earthquakes of the chosen domains are in Figure 4.

In spatial clustering, variables used are longitude, latitude and distance to nearest tectonic plate. Since tectonic plate boundary data was not readily available from standard sources, it was obtained from Hasterok et al. [16]. While this approach overcomes the problems of seismic clustering, it has low explanatory power on $m_{\text{max}}^{\text{obs}}$, i.e. domains similar in this way may not be similar in terms of maximum magnitude. The dendrogram and earthquakes of the chosen domains are shown in Figure 5.

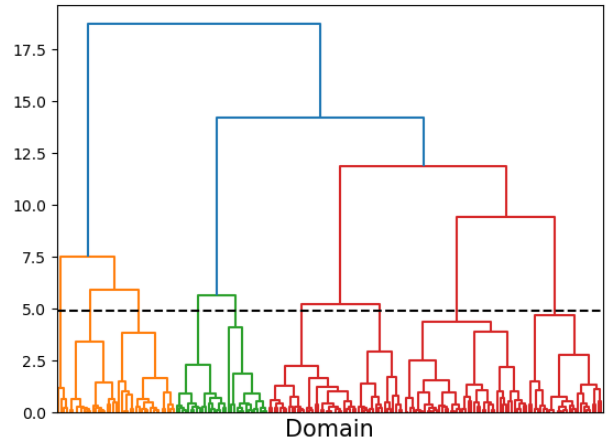


(a) Dendogram of seismic clustering

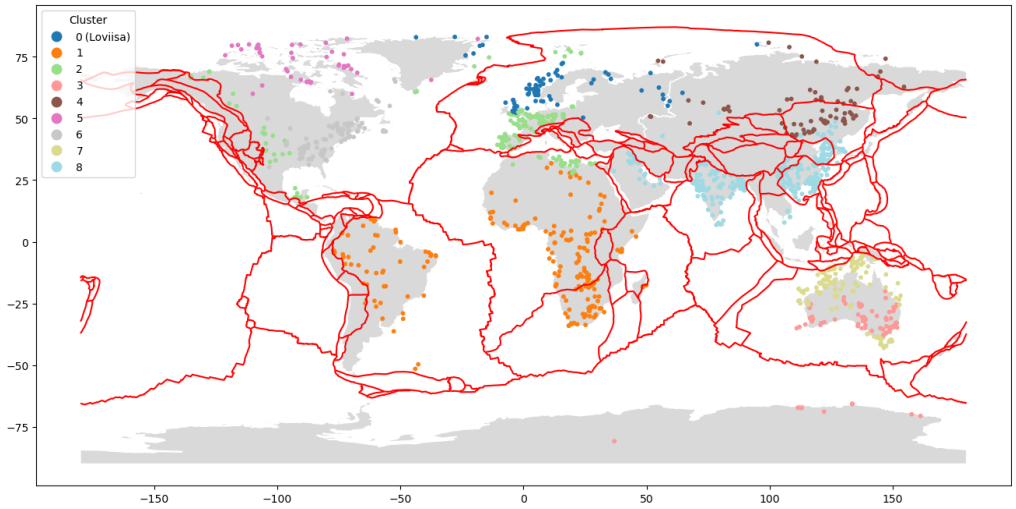


(b) Earthquakes color-coded by seismic cluster, 14 clusters

Figure 4: Dendrogram (a) and map of earthquakes (b) for seismic clustering. Cluster of Loviisa is 0.



(a) Dendogram of spatial clustering



(b) Earthquakes color-coded by spatial cluster, 9 clusters

Figure 5: Dendrogram (a) and map of earthquakes (b) for spatial clustering. Cluster of Loviisa is 0.

Important hyperparameter regarding the clustering is the number of clusters, k. It essentially defines the size of the Loviisa cluster, which then defines the prior distribution. Number of clusters was chosen to be 14 for seismic clustering and 9 for spatial clustering. The decision was made based on a suitable sample size, the position of the clusters on the map and the one-way ANOVA test. Different number of clusters are included as a way of sensitivity analysis. This is further discussed in section 5.1.2.

Since polygonal boundaries of the domains were not readily available, domain lo-

cations were approximated using the centroid of earthquakes within each domain. Following this procedure, position of Loviisa domain ends up at Central Europe rather than South-Finland. To mitigate this issue within the spatial clustering framework, a proxy domain is assigned to the geographic coordinates of Loviisa. Those domains ending up in the same cluster as the proxy domain are then used to create the prior.

4.1.4 Bayesian Updating of the Prior

According to the EPRI 1994 and 2012 studies [9, 10], the posterior probability distribution is obtained by multiplying the prior distribution with a likelihood function

$$L[m^u] = \begin{cases} 0, & \text{for } m^u < m_{\max}^{\text{obs}} \\ \left[1 - \exp\left\{-b\ln(10)(m^u - m_{\min})\right\}\right]^{N_{\geq \min}}, & \text{for } m^u \geq m_{\max}^{\text{obs}} \end{cases} \tag{5}$$

Here, $b\ln(10)$ represents the slope of the GR law estimated from the local catalog, $m_{\max}^{\text{obs}}$ is the maximum observed magnitude in the local dataset, and $N_{\geq \min}$ is the number of earthquakes with magnitudes equal to or greater than the defined threshold $m_{\min} = 4.5$.

Following multiplication, the posterior distribution must be normalized to ensure that it integrates to one. This is achieved by dividing the product of the prior and the likelihood by its integral over all possible values of m^u, which corresponds to the marginal distribution in Bayes' theorem (Equation (2)).

The likelihood function has two key effects on the prior distribution. First, it truncates the distribution at the observed maximum magnitude, setting the probability of lower values to zero. Second, it sharpens or narrows the posterior around $m_{\max}^{\text{obs}}$, with the intensity of this effect depending on $N_{\geq \min}$. A higher number of qualifying events increases the influence of the likelihood, whereas if no events in the local catalog exceed $m_{\min}$, the likelihood function becomes ineffective in updating the shape of the prior. In such cases, the only effect of the Bayesian updating is the truncation of the prior below $m_{\max}^{\text{obs}}$.

4.1.5 Synthetic Data Generation for Bayesian Updating

Since one of the main challenges behind updating the priors involves the lack of a significant seismic catalog in the Loviisa area, a potential solution lies in the generation of additional data. Although methods vary between research, it is usually accomplished by drawing out samples from a pre-defined distribution which is supposedly representative of the seismic activity seen in the region [17]. This has been suggested in the literature, see eg. [18] for utilizing synthetic catalogs to perform

statistical analysis like clustering quality test. The literature also focuses on the creation of simulation models to simulate ground motions based on the characteristics of a specific region [19]. However, since the project focuses on PSHA and not PRA, these aspects are not considered in the final results.

One of the key challenges regarding synthetic data is overfitting. If the true parameters are not known, any additional synthetic data will not reflect the existing bias in the catalog. A considered method involved utilizing the University of Helsinki Earthquake Catalog to generate additional data points that better reflect the superdomain Loviisa is located in before combining it with the SCR catalog to update the Bayesian prior. However, as noted in previous section, there is a notable issue with using the University of Helsinki Earthquake catalog. The maximum observed magnitude in the catalog is 4.5 while the SCR catalog consists of only earthquakes with magnitude of 4.5 and above. Generated data points from the university catalog would not be relevant to the SCR superdomain prior. However, generating earthquake data using the superdomain prior would only create additional bias that goes against the goals of the project.

Upon realizing the ineffectiveness of this method, a further literary review was conducted regarding the use of synthetic data generation. The review indicates that synthetic catalogs are more widely used for validating specific parameters within the Gutenberg-Richter distribution. For example, to test the validity of a specific region's b-value, synthetic catalogs can be generated at different parameters following different models to estimate overall bias of the provided b-value. As mentioned above, other use cases include testing the clustering of earthquake dataset. As both are outside the scope of this project, this method is not considered in the final results.

4.1.6 Application in Fortum's PSHA 2021

Fortum's PSHA from 2021 was conducted by Slate Geotechnical Consultants. According to their report [14], a slightly modified version of a prior originally proposed in the 1994 EPRI study [9], referred to as *the undifferentiated SCR $M_{\max}$ prior distribution*, was used. The authors report that the prior included $m_{\max}$ values on the M_w interval $[5.25, 7.75]$.

As the local seismic catalog for Loviisa did not include any events exceeding M_w 5.25, no Bayesian updating was carried out. Consequently, the selected prior served directly as the posterior distribution for $m_{\max}$. However, at Fortum's request, the upper bound of the distribution was later truncated to M_w 7.00, to better reflect the low-seismicity nature of the region. This adjustment resulted in the final discrete distribution for $m_{\max}$ shown in Table 2.

Table 2: Maximum magnitude distribution in Fortum's PSHA 2021 (Bayesian method)

$m_{\max}$ (M_w)	Weight
5.25	0.01
5.75	0.08
6.25	0.23
6.75	0.36
7.00	0.32

Although the table provides the discrete weights for each magnitude bin, the full form and parameters of the underlying distribution are not disclosed in the report. To approximate the continuous distribution used, we assume that the posterior follows a truncated normal distribution. Specifically, we model the distribution as a normal distribution with zero probability beyond M_w 7.00, scaled so that the CDF is bounded within the interval $[0, 1]$.

To estimate the parameters μ and σ of this truncated normal distribution, we use the Solver tool in Microsoft Excel. The goal is to find parameter values such that the probability density values at each discrete magnitude in Table 2 are approximately proportional to the corresponding reported weights. This procedure yields an estimated distribution of $\mathcal{N}(6.74, 0.56)$ for both the prior and the posterior. The resulting distribution is illustrated in Figure 6.

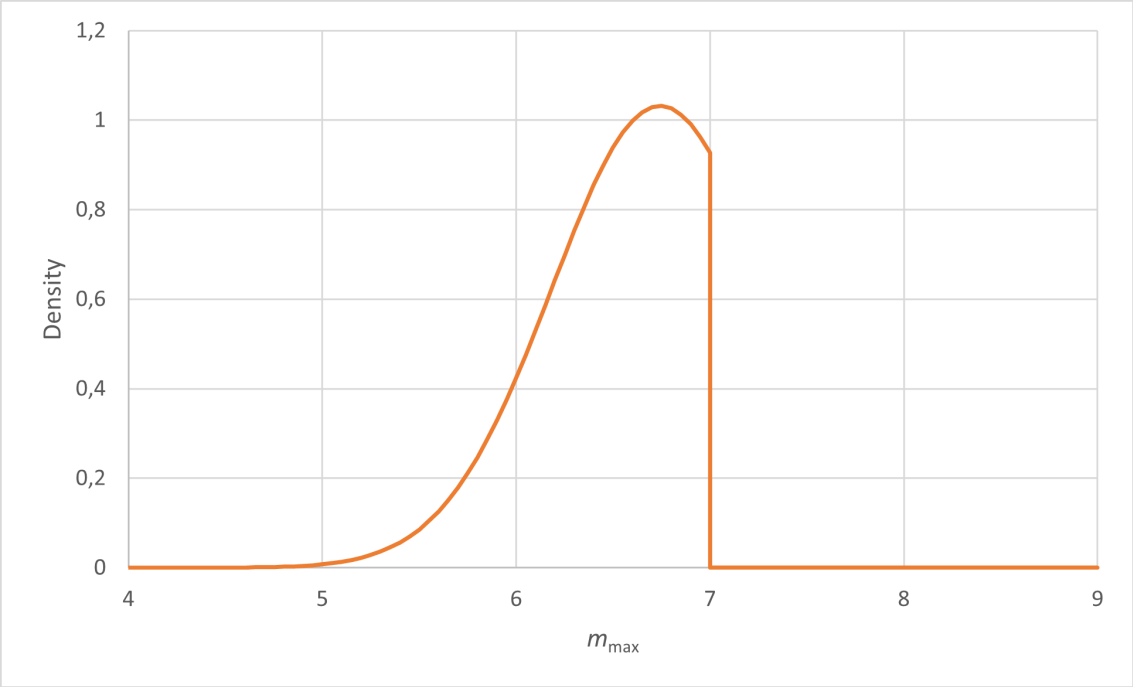


Figure 6: Estimated posterior probability distribution for $m_{\max}$ in Fortum's PSHA 2021

4.2 Kijko Method

The Kijko method [8] is a statistical approach used in seismic hazard analysis to estimate the seismicity parameters of a region, particularly when dealing with incomplete or uncertain earthquake catalogs. According to [8], the CDF which is bounded from above by $m_{\max}$ is

$$F_M(m) = \begin{cases} 0 & \text{for } m < m_{\min} \\ \frac{1-\exp\{-\beta(m-m_{\min})\}}{1-\exp\{-\beta(m_{\max}-m_{\min})\}} & \text{for } m_{\min} \leq m \leq m_{\max} \\ 1 & \text{for } m > m_{\min}, \end{cases} \tag{6}$$

where β is the slope in Gutenberg-Richter equation.

Based on the CDF, the confidence limit for estimated maximum earthquake magnitude $m_{\max}$ is

$$Pr(m_{\max} < z) = 1 - [F_M(m_{\max}^{\text{obs}}; z)]^n, \tag{7}$$

where $m_{\max}^{\text{obs}}$ is the maximum observed magnitude in the region, and n is the number of events.

4.2.1 Application in Fortum's PSHA 2018

In Fortum's 2018 report [20], the weights are reported for each magnitudes 5.5, 6.0, 6.5, 7 with a strip of 0.5 magnitude units. These weights were calculated by Dr. Jouni Saari from ÅF-Consult Ltd in [21] and [22]. Seismicity parameters calculated in University of Helsinki 2016 report [12] for area within 300 km of Loviisa are used for the calculation. These include time span T of 245 years, $m_{\min} = 1.4$, $m = m_{\max}^{\text{obs}} = 4.5$ according to [12]. The reported weights are in Table 3.

Table 3: Maximum magnitude weights for 300 km area around Loviisa in 2018 report

$m_{\max}$ (M_w)	Weight
5.50	0.726
6.00	0.20
6.50	0.057
7.00	0.016

It is also noted that there is a concern with Kijko's method applicability in Finland. In this report, the value $P(m_{\max} < z)$ is below 0.5 with increasing z, where $P(m_{\max} > 7) = 1 - P(m_{\max} < 7)$, which represents the truncated probability, is above 0.5. According to [8], when $z \to +\infty$, the probability in Equation 7 approaches a number less than 1. According to Chapter 5 of 2012 EPRI report [10], it is recommended that the truncated distribution should be below 0.5, which is not achieved in [20]. This recommendation is built upon the fact that the weights for different magnitudes should be determined based on the majority (over 50%) of the distribution. With a higher value of truncated distribution, the weight are calculated based on a minor part of the distribution, thus making it less justified. The high truncated probability obtained in 2018 report by Fortum is caused by the small catalog for Finland, with lower the reliability of estimating $m_{\max}$ using Kijko's method.

4.2.2 Implementation Using Updated Parameters

In 2024, Fortum conducted a probabilistic seismic hazard analysis [23] recalculating seisimic parameters for region around Loviisa, using catalog from [11], with completeness analysis conducted in Juhana Vehmas's thesis [13]. In this report, we calculate new weights for the different magnitudes according to the method in Fortum's 2018 report [20] using these updated parameter values.

The parameters are calculated separately for each SSZ instead of for the whole area within 300 km of Loviisa. Therefore, in our analysis, we will use the updated b-values to determine new maximum magnitude weights for each SSZ within 300 km of Loviisa separately. This includes SSZs 6, 8, 10, and 11. Due to low earthquake activity, SSZ 11 is excluded from the analysis. Therefore, only SSZ 6, 8, and 10 are

included in the analysis.

In determining the weights for each SSZ, we compute an average set of weights between the SSZs, weighted by number of earthquakes in each SSZ, as

$$W_{area} = \sum_{SSZ} \frac{n_{SSZ}}{\sum n_{SSZ}} \cdot W_{SSZ}, \tag{8}$$

where $\sum_{SSZ}$ represents summation over all SSZs, n_{SSZ} is the number of earthquakes in a given seismic source zone, and W_{SSZ} represents the corresponding magnitude weight distribution for that zone. This method represents each SSZ's contribution to the overall magnitude distribution by weighting zones based on their seismic activity, ensuring that more active zones have a greater influence on the final calculation.

In addtion, we also provide sensitivity analysis of the truncated probability $P(m_{\max} > z)$ with regards to different radius around Loviisa for a further discussion on Kijko's applicability to the region. For this analysis, an average b-values of all included SSZs is used for simplicity. We experiment with different values of radius at 100 km, 300 km, and 500 km. The SSZs included with regards to different radius were obtained using geographic information system application QGIS [24]. Figure 7 shows these different radius and included SSZs. For a radius of 100 km, SSZ 6 and 10 are included. For a radius of 300 km, SSZ 6, 8, and 10, 11 are included. For a radius of 500 km, SSZs 2, 3, 4, 5, 6, 7, 8, 10, and 11 are included. Similar to the weights calculation, SSZ 7 and 11 are excluded due to low earthquake activity.

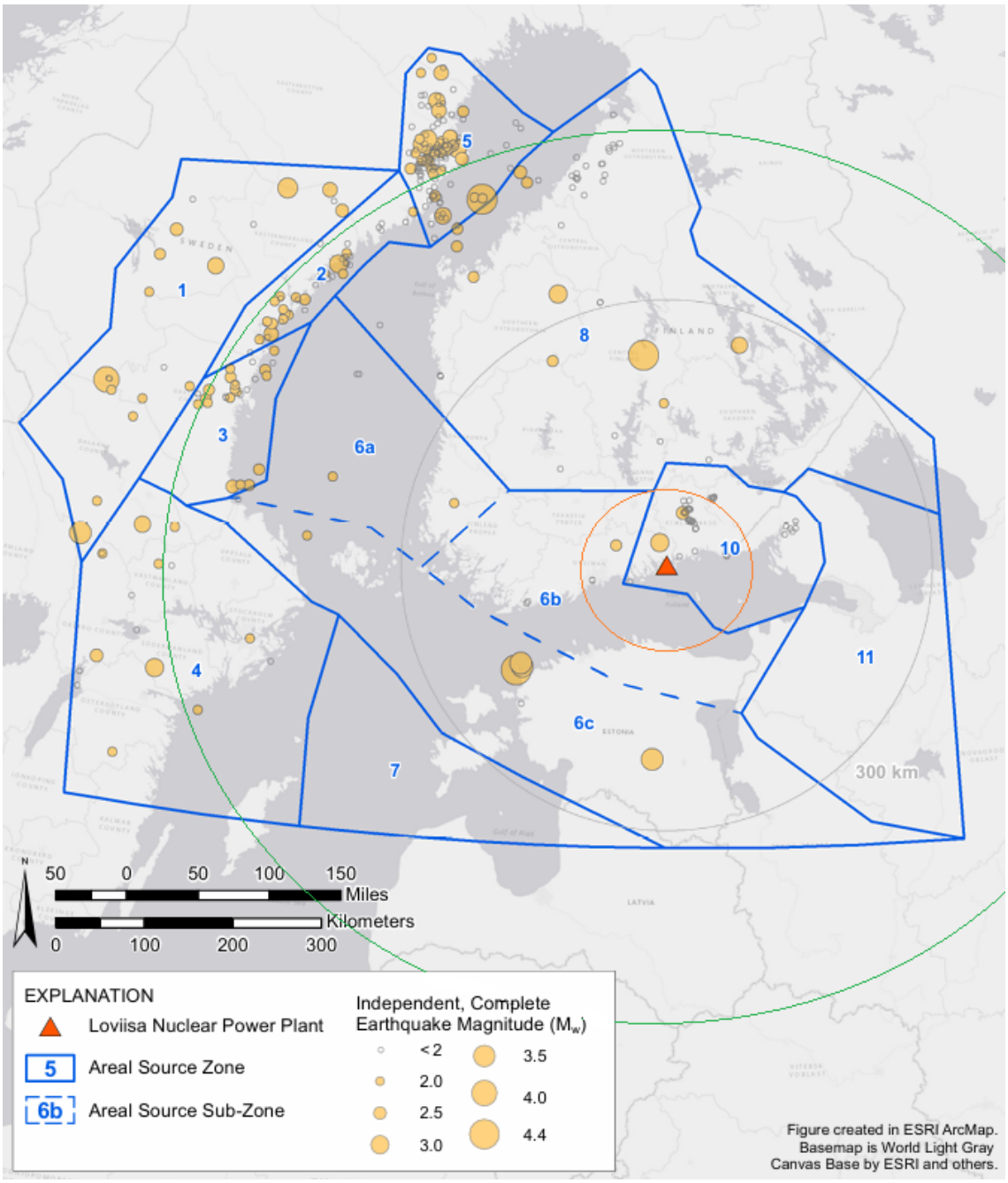


Figure 7: SSZs and the radius 100 km, 300 km, and 500 km around Loviisa

4.3 Integration of Methods

Ultimately, we chose to combine the outcomes of the Bayesian approach and Kijko's method to obtain a more balanced and robust estimate of $m_{\max}$. Each method offers distinct advantages: the Bayesian framework incorporates prior knowledge and is particularly valuable in low-data environments, whereas Kijko's method leverages statistical inference based directly on the observed earthquake record and is well-suited for capturing empirical patterns in the data.

Given the lack of definitive evidence favoring one method over the other in the context of this study, we opted to assign equal weights (50 % each) to the two methods. This weighting scheme reflects a neutral stance, avoiding the imposition of potentially unwarranted assumptions about the relative accuracy or reliability of either approach. It also ensures that both prior knowledge and empirical observations are given due consideration in the final estimate.

While this simple averaging approach is pragmatic, we acknowledge its limitations. In particular, it does not account for differences in uncertainty, sample size, or methodological bias between the two approaches. Future work could conduct more thorough literature review on this area and explore more refined weighting strategies, potentially based on uncertainty quantification, cross-validation, or expert elicitation, to further improve the integration of these methods.

For the purposes of this study, however, the equal weighting approach provides a transparent and justifiable means of synthesizing the available information, producing an estimate that benefits from the complementary strengths of both the Bayesian and Kijko methodologies.

5 Results

5.1 Results of the Bayesian Method

We developed three alternative prior distribution proposals using different subsets of the SCR catalog. The first prior was constructed based on superdomain 13, which includes the Loviisa region. The other two priors were derived using clustering techniques applied to varying catalog parameters to explore alternative groupings of seismically comparable regions.

In the updated UH 2023 catalog [11], $m_{\max}^{\text{obs}} = 4.4$ within a 300 km radius of the Loviisa NPP. As this value falls below the minimum magnitude threshold ($m_{\min} = 4.5$) defined for Bayesian Mmax weighting, Bayesian updating had no effect on the shape of the priors. Consequently, each posterior distribution remained identical to its corresponding prior, with the exception that it was truncated from below at $m_{\max} = 4.4$ and rescaled to ensure a total probability of one. The weights assigned to the discretized $m_{\max}$ bins were determined by integrating the posterior probability density over the range defined by each bin and subsequently normalized to ensure that the sum of all weights equals one.

5.1.1 Results Using SD13 Prior

As shown in Table 1 in Section 4.1.2, the mean observed maximum magnitude ($m_{\max}^{\text{obs}}$) within SD13 is 5.59, with a standard deviation of 0.70. Applying the bias correction procedure described in Section 4.1.1 yields a corrected mean of 5.86, resulting in a prior distribution approximated by $\mathcal{N}(5.86, 0.70)$.

Compared to the original prior for non-extended domains proposed in the 1994 EPRI study [9], which follows $\mathcal{N}(6.3, 0.5)$, this prior places significantly greater weight on lower $m_{\max}$ values. This adjustment is motivated by the relatively low observed seismicity and geological constraints in regions analogous to Loviisa, which suggest that lower maximum magnitudes may be more plausible, thus supporting the use of a less conservative prior in such low-seismicity contexts.

After truncating the distribution at the lower bound of $m_{\max} = 4.4$ and rescaling the density to ensure proper normalization, we obtained the posterior probability distribution shown in Figure 8.

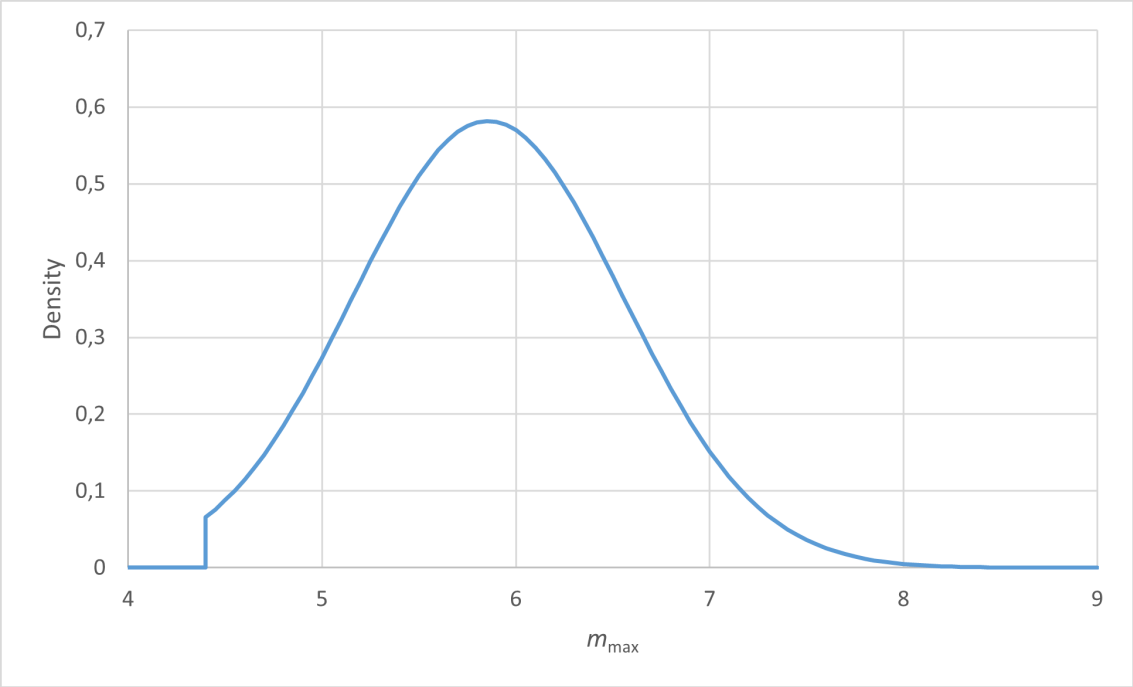


Figure 8: Posterior probability distribution for $m_{\max}$ using the SD13 prior

The weight distribution for the discretized magnitude bins within the M_w interval $[5.00, 7.00]$ is presented in Table 4. Each bin represents a symmetric interval of width 0.5 units; for example, the bin centered at 5.00 corresponds to the interval $[4.75, 5.25]$. The weight assigned to each bin is proportional to the integral of the posterior probability density function over the corresponding interval.

Table 4: Maximum magnitude weight distribution with the SD13 prior.

$m_{\max}$ (M_w)	Weight
5.00	0.14
5.50	0.28
6.00	0.30
6.50	0.20
7.00	0.08

5.1.2 Results Using Prior with Clustering

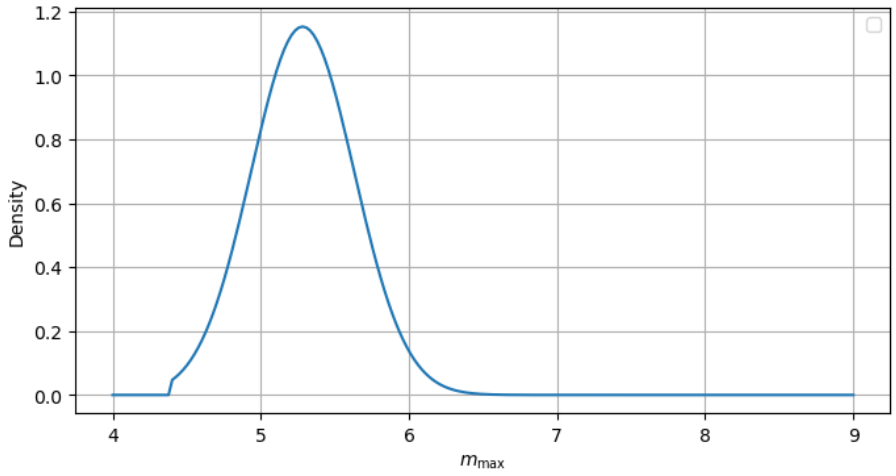
Results regarding both seismic and spatial clustering are displayed in Table 5. Following the procedure in Sections 4.1.2 and 4.1.3, with domains listed in Table 5, priors for the clusters were obtained. For seismic and spatial clustering, the corresponding priors become $\mathcal{N}(5.28, 0.35)$ with $k = 14$ and $\mathcal{N}(6.07, 0.58)$ with $k = 9$,

respectively. P-values of one-way ANOVA test for the mean of $m_{\max}^{obs}$ in these clusters is statistically significant, indicating that at least for some clusters, $m_{\max}^{\text{obs}}$ does not come from the same distribution. Varying the number of clusters does seem to impact mean $m_{\max}^{\text{obs}}$, but the magnitude of the change is acceptable, i.e. the result does not change completely as k changes.

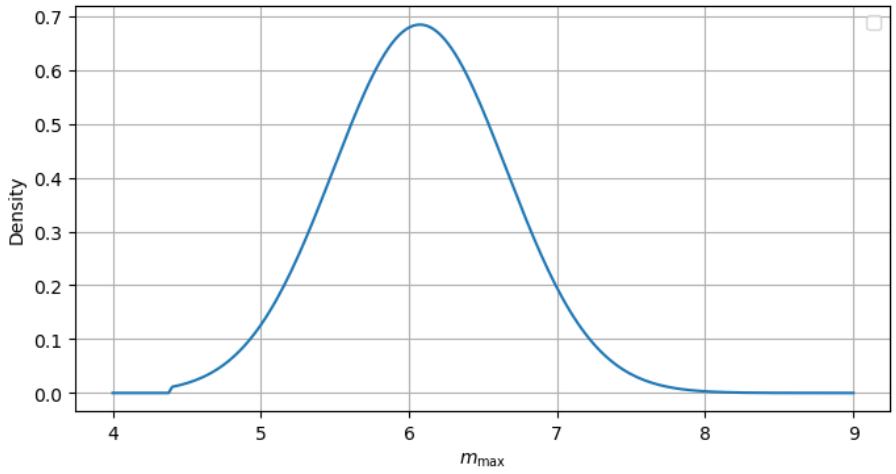
Table 5: Results for seismic and spatial clustering in stable continental regions.

	Seismic clustering			Spatial clustering		
	k = 9	k = 14	k = 17	k = 6	k = 9	k = 14
Chosen domain numbers	89, 93, 108, 109, 134, 138, 140, 141, 144, 146, 148, 152, 154, 156, 157, 158, 160, 161, 166, 167, 168, 170, 171, 173, 175, 182, 183, 184, 185, 186, 188, 189, 190, 193, 196, 197, 198, 200, 201, 244, 246	89, 93, 108, 109, 156, 157, 158, 160, 161, 166, 167, 168, 171, 173, 175, 182, 183, 184, 185, 186, 188, 189, 190, 193, 196, 197, 198, 200, 201, 244, 246	89, 93, 108, 156, 157, 158, 160, 161, 166, 167, 168, 171, 173, 175, 182, 183, 184, 185, 186, 188, 190, 196, 197, 198, 200, 201, 244, 246	154, 159, 166, 167, 168, 171, 173, 177, 179, 183, 184, 185, 186, 188, 189, 190, 194, 217, 218, 221, 222, 223, 224, 225, 226, 227, 228, 229, 230, 235, 236, 237, 239, 246, 247, 248, 249, 250, 251, 253	154, 159, 166, 167, 168, 171, 173, 177, 179, 183, 184, 185, 186, 188, 189, 190, 194, 251, 253	167, 177, 179, 183, 184, 185, 186, 188, 189, 190, 194, 251, 253
Mean m_{max}^{obs}	5.05	5.06	5.08	5.57	5.32	5.42
STD m_{max}^{obs}	0.31	0.35	0.35	0.80	0.58	0.65
Mean N	3.34	4.03	4.29	5.44	4.05	4.92
One-way ANOVA p-value	0	0	0	0.60	0.02	0.06

After truncating and normalizing the prior distributions, the posterior distributions for the clustering approach were obtained. The posterior distributions are displayed in Figure 9. Table 6 displays weights obtained from these posteriors.



(a) Posterior probability distribution for $m_{\max}$ using the seismic prior.



(b) Posterior probability distribution for $m_{\max}$ using the spatial prior.

Figure 9: Posterior distributions for priors constructed by seismic and spatial clustering.

Table 6: Maximum magnitude weight distributions with clustering

(a) Seismic weights

$m_{\max}$ (M_w)	Weight
5.00	0.42
5.50	0.49
6.00	0.09
6.50	0.00
7.00	0.00

(b) Spatial weights

$m_{\max}$ (M_w)	Weight
5.00	0.07
5.50	0.22
6.00	0.34
6.50	0.27
7.00	0.10

All in all, while the clustering is an interesting concept, it is limited by the quality and amount of the variables available. Among the clustering strategies, seismic clustering appears to yield promising results. However, as noted in the literature [9], these results are likely to be biased. This bias arises from the fact that clustering based on observed magnitudes inherently produces priors that predominantly reflect the sparse local data available for the region of interest, rather than broader seismic characteristics. At the same time, spatial variables used have little correlation with $m_{\max}^{\text{obs}}$. We therefore conclude that constructing the prior based on SD13 — effectively clustering by tectonic characteristics — is the most rational and well-justified approach, supported by existing literature.

One way to improve the clustering approach could be to include categorical tectonic attributes as one-hot-encoded binary variables. In addition, the position of the domain could be used. This would be similar to using SD13, but also taking into consideration the geographic positions of the domains.

5.2 Results of the Kijko Method

The Kijko's final weighted average weight is in Table 7. Details of the parameters and calculations are in Table A1, A2, and A3. Noticeably, $m_{\min} = 1$ is used instead of 1.4 as in 2018 Fortum report [20] for consistency with Vehmas' complete analysis [13], which also uses $m_{\min} = 1$ as the minimum threshold. In addition, the number of earthquakes N is derived directly from the number of earthquakes in each area instead of being calculated from the rate of earthquake λ as in the 2018 report.

Table 7: Maximum magnitude weights for 300 km area around Loviisa.

$m_{\max}$ (M_w)	Weight
5.50	0.721
6.00	0.204
6.50	0.058
7.00	0.016

Table 8 shows the truncated probability $P(m_{\max} > 7)$ for SSZ 6, 8, and 10. It can be seen that all the values exceeded the recommended value 0.5. Furthermore, there is a significant difference between SSZ 10 and the other SSZs. This is likely due to the lower $m_{\max}^{\text{obs}} = 2.9$ reported in this area, compared to $m_{\max}^{\text{obs}} = 4$ in SSZ 6 and $m_{\max}^{\text{obs}} = 4.4$ in SSZ 8.

Table 8: Truncated probability for SSZ 6, 8, and 10.

SSZ	6	8	10
$P(m_{\max} > 7)$	0.988	0.904	0.573

Finally, Table 9 shows the truncated probability at different radius value around Loviisa. It can be seen that the truncated probability decreases with increasing radius.

Table 9: Sensitivity of the truncated probability with regards to radius.

Radius	100 km	300 km	500 km
$P(m_{\max} > 7)$	0.978	0.95	0.723

6 Discussion and Conclusion

6.1 Final Maximum Magnitude Distribution

For the Bayesian method, the main focus was on the construction of the prior distribution, i.e. choosing which domains are the most representative of the Loviisa region. Three different ways of constructing the prior distribution were demonstrated, one of which was using the SD13 as defined in the literature. The other two ways were related to clustering of the domains based on historical seismic activity and spatial features. Using SD13 to construct the prior is theoretically sound and pragmatic approach. Nevertheless, there is a possibility that the superdomains exhibit no meaningful differences in their $m_{\text{max}}^{\text{obs}}$ distributions. This would imply that clustering by superdomains may not be that much better than using a global prior.

The clustering method aimed to address these limitations. The main challenge of the clustering method lies in the limited number of explanatory variables that correlate with $m_{\text{max}}^{\text{obs}}$. For the clustering method to be improved, tectonic characteristics of the domains should be included one way or another. Furthermore, there is always a bit of unavoidable ambiguity in choosing the number of clusters: how close is close enough for a point in the feature space to be considered representative of the Loviisa region? As is, the use of clustering priors would not be advised due to inherent bias with the seismic clustering and low explanatory power of the spatial clustering with respect to $m_{\text{max}}^{\text{obs}}$.

In essence, we recommend adopting the SD13-based prior due to its stronger methodological justification and lower degree of ambiguity compared to alternative clusters. The resulting distribution also lies between those obtained from the seismic and spatial clustering methods, suggesting it offers a balanced representation of the underlying data. Nonetheless, we acknowledge that the current superdomain grouping could be further refined through clustering, particularly if extended datasets incorporating additional tectonic characteristics become available.

For Kijko's method, it can be seen that the weights are heavily distributed in the lower magnitude region between 5 to 6 M_w, with the strip 5-5.5 taking up a value at approximately 0.72. Regarding the truncated probability, our analysis yields truncated probability much higher than the recommended value of 0.5. This is attributable to the limited size of the Finnish earthquake catalog, and the even smaller subset of data for each SSZ.

The sensitivity analysis shows that increasing radius around Loviisa lower the truncated portion of the distribution, due to the increasing number of earthquakes. However, at the highest radius value of 500 km, the truncated probability $P(m_{\text{max}} > 7)$

approaches 0.723, which indicates that 72% of the distribution is truncated, surpassing the 50% value recommended in [10]. Therefore, it may be appropriate to find ways to increase the data samples to be used with Kijko's method. Furthermore, other methods could also be considered for the analysis.

6.2 Comparison with Previous Studies

As illustrated in Figure 10, all Bayesian posterior distributions proposed in this study assign significantly greater probability mass to lower $m_{\max}$ values compared to Fortum's previously applied implementation. This suggests that the seismic potential of the Loviisa region may be lower than previously reported. Additionally, the relatively broad spread of the distributions reflects the heightened uncertainty associated with limited data availability.

Among the proposed approaches, the SD13 and spatial clustering posteriors yield broadly consistent results, while the seismic clustering posterior deviates more noticeably. This divergence is likely due to the inherent bias discussed in Section 5.1.2, wherein clustering based on observed magnitudes tends to exaggerate lower $m_{\max}$ values in regions characterized by sparse seismic activity.

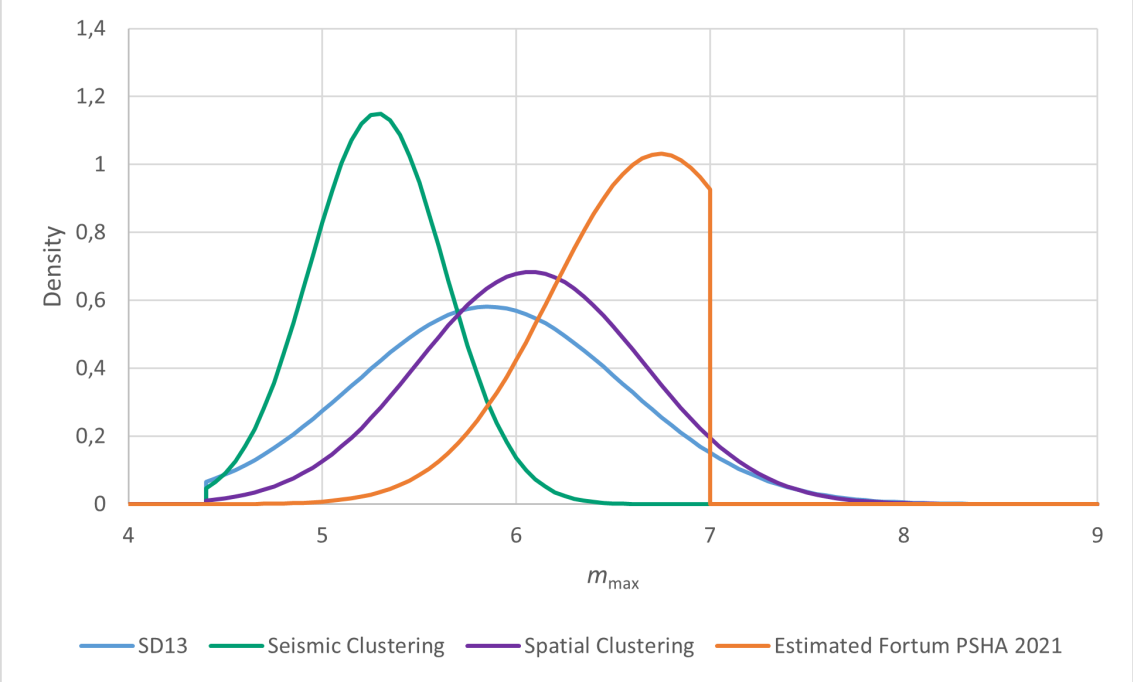


Figure 10: Proposed Bayesian posteriors for $m_{\max}$ compared to the estimated posterior probability distribution in Fortum's PSHA 2021

Since in all cases Bayesian updating only truncated the prior distributions without reshaping them, the observed differences among the posteriors are entirely attributable to the underlying priors. This highlights the critical importance of selecting a well-justified and representative prior, particularly in regions characterized by low seismicity.

Regarding the weights calculated using Kijko's method [8], our final results are rather similar to those obtained in [20]. However, it is difficult to compare the calculation process. This is because the weights are calculated separately for each SSZ in our report. While in 2018 Fortum report, the weights are calculated at once for the whole 300 km area around Loviisa.

6.3 Conclusion

The primary goal of this study was to improve the existing probability models that estimate the maximum magnitudes of a Fortum nuclear site in Loviisa. Two approaches were used: Bayesian modeling [9, 10] and the Kijko method [8]. Both of these methods were implemented successfully. Furthermore we explored the use of clustering methods to extend the feasible catalog that can be used. Despite the very different approaches of these methods, they both converge to broadly similar central values. For both, the most probable $m_{\max}$ is between 5.5 and 6.0. However, we see significant divergence on larger values of $m_{\max}$. Using the clustered superdomain, the Bayesian posterior assigns roughly probability of 0.07 to $m_{\max}$ event of 7, but for Kijko method the similar $m_{\max}$ probability is at 0.016.

A clustering of superdomains was also performed in order to group domains to achieve a complete dataset with larger amount of data. K-means and agglomerative hierarchical clustering were used and the clustering was based on seismic structure of historical earthquakes. The method showed promising results but due to limited quality of data no justified reason wad found to use the clustered dataset for prior estimation.

A comparison of these results with the original estimate from 2021 points to significantly differences. The study from 2021 [14] resulted in a most probable $m_{\max}$ of 6.7. However, it is difficult to say whether the new estimate is more accurate. More research is needed to ascertain that the new estimates reflect the environmental factors of Loviisa more accurately.

We recommend treating the two distributions together rather than selecting either one alone. A simple 50–50 mixture captures the tighter lower-range confidence of the Kijko result while keeping the heavier upper tail from the Bayesian model. This mixed probability distribution reduces the median hazard relative to the 2021 study,

but maintains a conservative look for even more rare high-magnitude events. With this in mind, a deeper look of the parameter weights should be performed as the 50-50 mix may not be optimal for Loviisa site.

Furthermore, we suggest that a thorough sensitivity analysis on the results is to be conducted before interpreting the results as given. It is possible that the results derived in this study may be sensitive to parameter changes.

References

[1] C.A. Cornell. Engineering seismic risk analysis. *Bulletin of the Seismological Society of America*, 58(5):1583–1606, 10 1968.

[2] R.K. McGuire. Computations of seismic hazard. *Annali di Geofisica*, 36(3-4):181–200, 1993.

[3] A. Frankel. Mapping seismic hazard in the central and eastern United States. *Seismological Research Letters*, 66(5):8–21, 1995.

[4] A. Kijko and G. Graham. "Parametric-historic" procedure for probabilistic seismic hazard analysis Part II: Assessment of seismic hazard at specified site. *Pure and Applied Geophysics*, 154(1):1–22, 1999.

[5] H. Tsang and A.M. Chandler. Site-specific probabilistic seismic-hazard assessment: direct amplitude-based approach. *Bulletin of the Seismological Society of America*, 96(2):392–403, 2006.

[6] L. Fülöp, P. Mäntyniemi, M. Malm, G. Toro, M.J. Crespo, T. Schmitt, S. Burck, and P. Välikangas. Probabilistic seismic hazard analysis in low-seismicity regions: an investigation of sensitivity with a focus on Finland. *Natural Hazards*, 116:111–132, 2023.

[7] V. F. Pisarenko and M. V. Rodkin. Approaches to Solving the Maximum Possible Earthquake Magnitude ($M_{\max}$) problem. *Surveys in Geophysics*, 43:561–595, 2022.

[8] A. Kijko. Estimation of the maximum earthquake magnitude, $\mathrm{m}_{\max}$. *Pure and Applied Geophysics*, 161:1–27, 2004.

[9] A.C. Johnston, K.J. Coppersmith, L.R. Kanter, and C.A. Cornell. The Earthquakes of Stable Continental Regions: Assessment of Large Earthquake Potential. Technical report, Electric Power Research Institute (EPRI), Palo Alto, California, 1994.

[10] Electric Power Research Institute (EPRI). Central and Eastern United States Seismic Source Characterization for Nuclear Facilities. Technical report, Palo Alto, California, 2012.

[11] P. Mäntyniemi. Earthquake catalog (2023). Institute of Seismology, University of Helsinki, 2023.

[12] A. Korja, S. Kihlman, and K. Oinonen. Seismic source areas in Central Fennoscandia. Technical report, Institute of Seismology, University of Helsinki, 2016.

[13] J. Vehmas. Probabilistic Seismic Hazard Analysis for Nuclear Power Plant Sites in Finland. Master's thesis, Aalto University: School of Science, Espoo, Finland, 2024.

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

[15] A. Gelman, J.B. Carlin, H.S. Stern, D.B. Dunson, A. Vehtari, and D.B. Rubin. *Bayesian Data Analysis*. Chapman & Hall/CRC, 3rd edition, 2013.

[16] D. Hasterok, J. Halpin, M. Hand, A. Collins, C. Kreemer, M. G. Gard, and S. Glorie. New maps of global geologic provinces and tectonic plates. *Earth Science Reviews*, 2021. Preprint available on EarthArXiv.

[17] G.M. Geffers, I.G. Main, and M. Naylor. Biases in estimating b-values from small earthquake catalogues: how high are high b-values? *Geophysical Journal International*, 229(3):1840–1855, 01 2022.

[18] E.G. Daub, D.T. Trugman, and P.A. Johnson. Statistical tests on clustered global earthquake synthetic data sets. *Journal of Geophysical Research: Solid Earth*, 120(8):5693–5716, 2015.

[19] M. Monterrubio-Velasco, S. Callaghan, D. Modesto, J.C. Carrasco, R.M. Badia, P. Pallares, F. Vázquez-Novoa, E.S. Quintana-Ortí, M. Pienkowska, and J. Puente. A machine learning estimator trained on synthetic data for real-time earthquake ground-shaking predictions in Southern California. *Communications Earth & Environment*, 5(1):258, 2024.

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

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

[22] M. Malm and O. Kaisko. Teollisuuden Voima Oyj / Fortum Oyj, Re-evaluation of seismic hazard spectra in Olkiluoto and Loviisa. ÅF-Consult Ltd, 2017. LO1-T84252-00005.

[23] J. Vehmas. Loviisa probabilistic seismic hazard analysis - 2024 update. Fortum Power and Heat Oy, 2024. LO1-T84252-00023.

[24] QGIS Development Team. *QGIS Geographic Information System*. QGIS Association, 2025.

A Kijko Weights for SSZs

Table A1: Kijko weights for SSZ 6

	5	5.5	6	6.5	7
$m_{\min}$	1	1	1	1	1
m	4.4	4.4	4.4	4.4	4.4
b	1.103	1.103	1.103	1.103	1.103
β	2.539	2.539	2.539	2.539	2.539
n	69	69	69	69	69
$m_{\max}$	5	5.5	6	6.5	7
F(m)	1	1	1	1	1
$Pr(m_{\max} < z)$	0.010	0.011	0.012	0.012	0.012
Wn	-	0.724	0.203	0.057	0.016

Table A2: Kijko weights for SSZ 8

	5	5.5	6	6.5	7
$m_{\min}$	1	1	1	1	1
m	4	4	4	4	4
b	1.035	1.035	1.035	1.035	1.035
β	2.383	2.383	2.383	2.383	2.383
n	128	128	128	128	128
$m_{\max}$	5	5.5	6	6.5	7
F(m)	0.999	0.999	0.999	0.999	0.999
$Pr(m_{\max} < z)$	0.087	0.093	0.095	0.095	0.096
Wn	-	0.703	0.213	0.065	0.020

Table A3: Kijko weights for SSZ 10

	5	5.5	6	6.5	7
$m_{\min}$	1	1	1	1	1
m	2.9	2.9	2.9	2.9	2.9
b	1.168	1.168	1.168	1.168	1.168
β	2.689	2.689	2.689	2.689	2.689
n	92	92	92	92	92
$m_{\max}$	5	5.5	6	6.5	7
F(m)	0.994	0.994	0.994	0.994	0.994
$Pr(m_{\max} < z)$	0.426	0.427	0.427	0.427	0.427
Wn	-	0.743	0.193	0.050	0.013

Table A4: Kijko weights for radius 100 km (SSZ 6, 10)

	5	5.5	6	6.5	7
$m_{\min}$	1	1	1	1	1
m	4.4	4.4	4.4	4.4	4.4
b	1.135	1.135	1.135	1.135	1.135
β	2.614	2.614	2.614	2.614	2.614
n	161	161	161	161	161
$m_{\max}$	5	5.5	6	6.5	7
F(m)	1	1	1	1	1
$Pr(m_{\max} < z)$	0.017	0.021	0.022	0.022	0.022
Wn	-	0.734	0.198	0.054	0.014

Table A5: Kijko weights for radius 300 km (SSZ 6, 8, 10)

	5	5.5	6	6.5	7
$m_{\min}$	1	1	1	1	1
m	4.4	4.4	4.4	4.4	4.4
b	1.102	1.102	1.102	1.102	1.102
β	2.537	2.537	2.537	2.537	2.537
n	289	289	289	289	289
$m_{\max}$	5	5.5	6	6.5	7
F(m)	1	1	1	1	1
$Pr(m_{\max} < z)$	0.040	0.047	0.050	0.050	0.050
Wn	-	0.724	0.203	0.057	0.016

Table A6: Kijko weights for radius 500 km (SSZ 2, 3, 4, 5, 6, 8, 10)

	5	5.5	6	6.5	7
$m_{\min}$	1	1	1	1	1
m	4.4	4.4	4.4	4.4	4.4
b	1.063	1.063	1.063	1.063	1.063
β	2.447	2.447	2.447	2.447	2.447
n	1330	1330	1330	1330	1330
$m_{\max}$	5	5.5	6	6.5	7
F(m)	1	1	1	1	1
$Pr(m_{\max} < z)$	0.221	0.261	0.272	0.276	0.277
Wn	-	0.719	0.204	0.060	0.017

B Self-assessment

B.1 How closely did the project follow the plan?

The actual implementation largely adhered to the initial plan, with some deviations that were expected given the exploratory nature of the project. Most deviations stemmed from the need to justify the chosen methodologies. As outlined in the report, concerns arose regarding the applicability of both Kijko's and Bayesian approaches. Consequently, additional analyses were conducted to assess result sensitivity and statistical significance.

Although the project remained on schedule overall, some team members experienced periods of intense workload. Spring break was not factored into the initial timeline, and competing responsibilities from other courses led to some tasks being deferred to the final weeks.

In terms of workload distribution, some team members ended up with more responsibilities than originally planned. Certain aspects of the project proved more time-consuming than anticipated.

B.2 In what regard was the project successful?

The team considers the main analyses and resulting insights a success. We were able to conduct our work with appropriate reasoning and methodological justification, avoiding unfounded conclusions and overly optimistic interpretations. The results are grounded in evidence and reflect a balanced and critical approach to analysis.

The project was also successful in delivering reproducible results and meaningful insights to the client. It addresses several current challenges in the domain, offering a foundation for further internal research at Fortum. The methods and findings can serve as a basis for more targeted studies or development work in the future.

Client communication was effective throughout the project. We received consistent support and valuable guidance from Fortum's representatives, who also granted us considerable autonomy in conducting our research—provided we could justify our choices. The client maintained realistic expectations regarding both the scope of the work and the results, allowing us to focus on depth and quality.

Lastly, the team demonstrated strong internal communication and collaboration. Team members were consistently available for meetings, adhered to deadlines, and contributed actively throughout the project.

B.3 In what regard was the project less successful?

In hindsight, the project manager could have played a more active role in facilitating communication and redistributing tasks to maintain a more balanced workload across the team.

Some of the original analytical directions proved difficult to implement in practice. An earlier and more thorough literature review and preliminary research phase would have helped identify more suitable approaches from the outset. This will also facilitate more balanced work division.

B.4 What could have been done better?

B.4.1 Team

In retrospect, several aspects of the team's workflow could have been improved. More consistent communication regarding individual progress would have helped prevent bottlenecks and periods of intense workload. Additionally, allocating more time to the project during its early phases could have led to more refined analyses and deeper insights in the final results.

B.4.2 Client

Overall, Fortum provided excellent communication and project scoping. We were granted access to high-quality data and internal documentation, which significantly supported our work. However, due to the inherent complexity of the project, some aspects were initially difficult for the team to grasp. These challenges are understandable, given the technical and domain-specific nature of the work, which naturally requires a learning curve.

B.4.3 Teaching staff

We found the guidance from the teaching staff appropriate, especially given the wide variety of projects under supervision. The teaching staff emphasizes collaboration and alignment with the client. That said, we found some inconsistencies between the expectations for the interim report and the interim presentation. In our view, the interim phase could have been simplified by requiring only a presentation, as the written report offered limited added value at that stage.