Recurrence Models based on Extreme Value Theory: Impacts on Probabilistic Seismic Hazard Assessments and Comparison with the Standard Approach

Anne Dutfoy (anne.dutfoy@edf.fr)
EDF Lab Saclay https://orcid.org/0000-0002-5139-106X

Gloria Senfaute
EDF Lab Saclay

Research Article

Keywords: Recurrence model, Probabilistic Seismic Hazard Analysis, Gutenberg-Richter, Extreme Value Theory, Generalized Pareto Distribution

DOI: https://doi.org/10.21203/rs.3.rs-712783/v1

License: ☝️ This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License
Recurrence Models based on Extreme Value Theory: Impacts on Probabilistic Seismic Hazard Assessments and Comparison with the Standard Approach

Anne DUTFOY, Gloria SENFAUTE

Abstract Probabilistic Seismic Hazard Analyses (PSHA) require that at least the mean activity rate be known, as well as the distribution of magnitudes. Within the Gutenberg-Richter assumption, the magnitudes follow an exponential distribution which is upperly truncated to a maximum possible magnitude denoted $m_{\text{max}}$. This parameter is often fixed from expert judgement under tectonics considerations, due to a lack of universal method.

In this paper, we present two alternatives to the exponential distribution of the magnitudes, based on the extreme value theory and that don’t require to fix a priori the value of $m_{\text{max}}$: the first model is based on a generalized Pareto distribution (GPD) to model the tail distribution of the magnitudes; the second model, the Randomized Gutenberg-Richter model, is a variation on the usual exponential distribution where $m_{\text{max}}$ is randomized and follows a distribution defined from an extreme value analysis.

We use the maximum likelihood estimates taking into account the time varying level of completeness of the catalog and the uncertainty in the magnitude value itself. We apply both modelizations to the Alps region.

Then, we integrate the resulting magnitude-frequency relations into a probabilistic seismic hazard calculation to evaluate their impacts on the seismic hazard levels. These extreme-value-based recurrence models introduce a reduction of the seismic hazard level compared to the common Gutenberg-Richter model conventionally used for PSHA calculations. This decrease is significant mainly for long periods.

Keywords Recurrence model · Probabilistic Seismic Hazard Analysis · Gutenberg-Richter · Extreme Value Theory · Generalized Pareto Distribution

1 Introduction

Probabilistic Seismic Hazard Analysis (PSHA) has become a fundamental tool in assessing seismic hazards and seismic input motions to evaluate critical facilities ([1], [2]) and in developing codes at national or regional scale ([16], [19]). Over the last 15 years, the PSHA practice has greatly improved, mainly motivated by research programs launched around the world ([9], [31], [36], [37]).

One important task in PSHA consists in calculating the activity rate for any seismic source and the earthquake magnitude-frequency distribution. The most common model is based on the empirical Gutenberg-Richter law ([20], [21]), defined by:

$$\log \mathbb{E} \left[ N_m \right] = a - bm$$

(1)

where $\mathbb{E} \left[ N_m \right]$ is the mean number of earthquakes with a magnitude larger than $m$. Under the assumption that earthquakes are generated by a stationary point Poisson process, this relation implies that magnitudes are exponentially distributed. As the total energy released by
There is no widely accepted method to estimate $m_{\text{max}}$, the maximum possible magnitude that is the largest possible earthquake of seismic and tectonic setting. For stable continental regions with poor identification of seismogenic structures and low earthquake rates, the determination of $m_{\text{max}}$ is a tricky matter. Currently, there is no widely accepted method to estimate $m_{\text{max}}$ and the literature is vast on estimating the maximum possible magnitude. Some methods are based on empirical relationships between the magnitude and various tectonic and fault parameters, such as the fault length or the rupture dimension ([3], [43]). Other methods rely on records of the largest historic or paleo earthquakes applied especially in areas of low seismicity, where large events have long return periods [28]. In the absence of any tectonic information, the maximum possible earthquake magnitude is assumed to be equal to the largest magnitude of the catalog, or the largest one plus an increment defined by experts judgement [44]. Many scientific papers propose statistical estimators of the maximum possible magnitude: for a complete review of different methods applicable for the assessment of the maximum possible magnitude, see [8], [22], [23], [24], [25], [26] or [41].

In this paper, we aim to provide two alternatives to the truncated Gutenberg-Richter model where the maximum possible magnitude $m_{\text{max}}$ is fixed by expert judgements. These models are based on a generalized Pareto distribution (GPD) without requiring any assumption on $m_{\text{max}}$:

- the first model uses a GPD to model the tail of the magnitude distribution;
- the second model is a variation on the usual truncated Gutenberg-Richter model: the magnitudes follow an exponential distribution which upper bound is randomized. The randomized maximum magnitude follows a distribution that results from an extreme value analysis.

Furthermore, we go further than just estimating the magnitude-frequency relations associated to both models: we want to quantify their impacts on the seismic hazard levels mainly for long return periods of interest in the nuclear facilities. That’s why we implemented both of them into a complete probabilistic seismic hazard assessment. At last, in Section 4, we apply the models to the Alps region in France, and we integrate them into a complete probabilistic seismic hazard calculation to evaluate their impacts on the seismic hazard levels.

## 2 Earthquake Recurrence Models

We briefly present in this section the magnitude-frequency models that will be used afterwards.

We assume that the temporal distribution of event numbers follow a point Poisson distribution. Consequently, the seismic events follow a point Poisson process.

### 2.1 The truncated Gutenberg-Richter model (GRt)

This model relies on the Gutenberg-Richter law (1), where magnitudes are exponentially distributed, with a maximum magnitude $m_{\text{max}}$ fixed by experts judgement under geophysical considerations. In [17], the author details the model as well as its estimate from data that follow the previous features. The author also exhibits the normal asymptotic properties of the joint estimator of $(a, b)$.

The recurrence model is given for all $m \geq 0$ by:

$$\log E[N_m] = a + \log_{10} \frac{10^{-bm} - 10^{-bm_{\text{max}}}}{1 - 10^{-bm_{\text{max}}}}$$

### 2.2 The Generalized Pareto Distribution based model (GPD)

This model is an alternative to the truncated Gutenberg-Richter law (1). We model extreme magnitudes using a generalized Pareto distribution rather than an exponential distribution. As a matter of fact, under some general conditions, the tail of a distribution considered beyond a high threshold may be approximated ever more closely by a GPD. Extreme magnitudes are those that exceed a high threshold $u$ to be determined. In [18], the author details the model as well as its estimate from data that follow the previous features. The author also exhibits the normal asymptotic properties of the joint estimator of the parameters.

A GPD($\sigma, \xi$) is parametered by a scale parameter $\sigma$ and a shape parameter $\xi$. Given that the strain energy released in a region is necessarily bounded, the magnitude distribution must be bounded too. Thus, the magnitude distribution belongs to the Weibull attraction domain where $\xi < 0$. If the random variable $X$ follows
a GPD($\sigma, \xi$), then its cumulated distribution function (cdf) where $\xi < 0$ is defined for all $x \in [0, -\frac{\sigma}{\xi}]$ by:

$$H_{\sigma, \xi}(x) = \mathbb{P}[X \leq x] = 1 - \left(1 + \frac{x}{\sigma}\right)^{-1/\xi}$$

Then, when $\xi < 0$, X is upperly bounded by $-\frac{\sigma}{\xi}$.

We denote by $M_0 = M|M \geq m_0$ the magnitudes larger than $m_0$ and by $F_0$ its cdf. If $u$ is large enough, then, according to the Tippett-Gnedenko-de Haan theorem ([32]), the tail distribution of $M_0$ can be written for all $m \geq u$ as:

$$F_0(m) = F_0(u)H_{\sigma, \xi}(m - u)$$

where $F_0 = 1 - F_0$ and $H_{\sigma, \xi} = 1 - H_{\sigma, \xi}$.

The use of a GPD to model the tail distribution of the magnitudes is not new. In [7], [13], [34], [35] and [41], the authors propose estimators of the GPD parameters ($\sigma, \xi$) using the maximum likelihood. But the proposed estimators are adapted to catalogs which level of completeness is constant in time. The time-varying level of completeness entails a correlation between the magnitude and the time and requires to use the likelihood of the Poisson process rather than the likelihood of the magnitude distribution only.

In [5] and [8], the authors propose to truncate the GPD tail distribution of the magnitudes to the endpoint $T_M$: their numerical experiences led them to think that even when using a negative extreme value index $\xi$, extreme value methods are not able to capture truncation at high levels. They propose an estimator for $T_M$, using estimators of extreme quantiles, more stable than those proposed in [6].

A vast literature is focused on the Pareto distribution because they consider the energy $E$ released by earthquakes rather than the magnitude. Both variables are linked by the relation $M = (1/1.5 \ln 10)\ln(E/2) + 1$, where $M$ is expressed on the Richter scale and $E$ in megajoule. If the magnitudes follow an exponential distribution, then the energy $E$ follows a Pareto distribution. The Pareto distribution belongs to the Frechet distribution, then the energy $E$ follows the Frechet distribution. The Exponential distribution follows the so-called Truncated GR or Truncated Pareto model; a soft truncation to the so-called Tapered GR or Tapered Pareto model ([22] or [8]).

In this paper, we focus on the magnitude distribution rather than on the energy distribution. The GPD is only used to model the tail of the magnitude distribution. Furthermore, we do not use any additional truncation since the estimated GPD is already upperly bounded as $\xi < 0$. We check the good fitting of the GPD by comparing its right tail function with non-parametric right tail estimates (using histograms and kernel smoothing techniques).

At last, we have adapted the formulas of the likelihood of the point Poisson process to the time-non constant level of completeness of the seismic catalog and to the uncertainty on the magnitude data.

The recurrence model is defined for all $m \geq u$ by:

$$\log_{10} \mathbb{E}[N_m] = \log_{10} \mu_u - \frac{1}{\xi} \log_{10} \left(1 + \frac{m - u}{\sigma}\right)$$

where $\mu_u$ is the mean annual number of earthquakes with magnitude larger than $u$.

2.3 The Randomized Gutenberg-Richter model (GRt-rd)

The Randomized Gutenberg-Richter model is a new variation on the common truncated Gutenberg-Richter model: magnitudes are still supposed to follow an upperly truncated exponential distribution. But the upper bound $M_{\text{max}}$ is randomized and follows a distribution denoted by $\mathcal{L}_{M_{\text{max}}}$. The model is formally written as:

$$\begin{cases}
M | M_{\text{max}} \sim \text{Exp}(b \log 10) \text{ truncated to } [0, M_{\text{max}}] \\
M_{\text{max}} \sim \mathcal{L}_{M_{\text{max}}}
\end{cases}$$

where $b$ is the usual Gutenberg-Richter parameter, estimated in Section 2.1. It is important to note that the estimate of $b$ is independent of the upper bound of the exponential distribution, which allows to estimate it in a separate way.

This model is a way to avoid an abrupt truncation: Kagan in [23] has argued that a sharp upper threshold does not agree with the known behavior of dissipative physical dynamic systems and fundamental seismological scaling principles. The sharp truncation is then replaced by a soft normal-like roll-off. Such a taper enables to take into account uncertainties on the right endpoint of the magnitude distribution. The resulting magnitude-frequency law is characterized by its gradual rather than steep decrease in frequency at the upper cutoff magnitude, and provides better agreement with both fundamental seismological principles and earthquake catalogs.

There exist different ways to fix the distribution $\mathcal{L}_{M_{\text{max}}}$. For example, it may be the asymptotic distribution of one of the estimators of $M_{\text{max}}$ proposed in the literature. In this paper, we suggest that the distribution of the roll-off is fixed from an extreme value analysis that enlights possible values for the maximum possible magnitude: then, in this paper, we consider
the asymptotic distribution of the estimator of a well-chosen quantile of large order of the GPD-tail distribution of the magnitudes. Furthermore, as the upper bound is randomized, the model does not fix the maximum magnitude to one value in particular but lets it take several possible values.

The quantile \( q_p \) of order \( p \) is defined by \( F_0(q_p) = p \), and from (3), with \( \zeta_u = F_0(u) \) and \( \xi < 0 \), we deduce that:

\[
q_p = u + \frac{\sigma}{\xi} \left( \left( \frac{1 - p}{\zeta_u} \right)^{-\xi} - 1 \right)
\]  

We denote by \( g_p \) the function such that \( q_p = g_p(\sigma, \xi) \) and we consider the estimator \( \hat{q}_p \) of \( q_p \):

\[
\hat{q}_p = g_p(\hat{\sigma}, \hat{\xi})
\]

We know that \((\hat{\sigma}, \hat{\xi})\) is asymptotically normal, with a covariance matrix denoted \( \Sigma \). From regularity properties of \( g_p \), we deduce that \( \hat{q}_p \) is also a maximum likelihood estimator, asymptotically normal, with covariance matrix \( \Gamma \) that is written as:

\[
\Gamma = \nabla g_p(\hat{\sigma}, \hat{\xi}) \Sigma \nabla g_p(\hat{\sigma}, \hat{\xi})^T
\]

where \( \nabla g_p(\hat{\sigma}, \hat{\xi}) = \left( \frac{\partial g_p}{\partial \sigma}(\hat{\sigma}, \hat{\xi}), \frac{\partial g_p}{\partial \xi}(\hat{\sigma}, \hat{\xi}) \right) \).

To fix the distribution of \( M_{\text{max}} \), we must take into account the maximum magnitude \( m_{\text{max,cat}} \) of the catalog: it would be nonsense that \( M_{\text{max}} \) be lesser than \( m_{\text{max,cat}} \). Thus, the distribution \( \mathcal{L}_{M_{\text{max}}} \) has to be truncated to the lower bound \( m_{\text{max,cat}} \) to ensure that \( M_{\text{max}} \) be larger than \( m_{\text{max,cat}} \) with a probability equal to 1.

It is worth noticing that the cdf \( F \) of the magnitudes is related to \( F_0 \) through the relation \( F(m) = \zeta_u F_0(m) \) for all \( m \geq u \). Then, we have \( F(q_p) \geq F_0(q_p) \), which entails that \( q_p \) is indeed a \( \hat{p} \)-quantile of order \( \hat{p} \geq p \). In other words, \( q_p \) is a more extreme value with respect to the distribution \( F \) than with respect to the \( F_0 \) distribution.

We want to emphasize that assigning to \( M_{\text{max}} \) the distribution of the GPD(\( \sigma, \xi \)) upper bound: \( u - \sigma/\xi \) deduced from the normal asymptotic distribution of \((\hat{\sigma}, \hat{\xi})\) is not the best option, due to the irregularity of the function \((\sigma, \xi) \to u - \sigma/\xi\) at any point of kind \((\sigma, 0)\): any small variation on \( \xi \) around the zero value leads to a large variation on the upper bound. Thus, estimate uncertainty on \( \xi \) induces a huge variance in the distribution of \( u - \sigma/\xi \). This instability had already been noticed in [34]. On the contrary, the estimate of any quantile of large order is much more stable as the function \( q_p \) is infinitely regular at any point of kind \((\sigma, 0)\) (for example, the derivability at \((\sigma, 0)\) is ensured by the equivalent \( \partial g/\partial \xi(\sigma, \xi) \sim (1/2)\sigma \log^2 \alpha \) with \( \alpha = \zeta_u(1 - p) \)).

Generally speaking, the upper bound of a distribution gives poor information on its extreme values: the decreasing speed of its tail is more informative. Consider for example the standard normal and log-normal distributions: both ones have an infinite upper bound whereas the probability to be greater than 4 standard deviations from the mean is about \( 10^{-5} \) for the Normal distribution and \( 10^{-2} \) for the log-normal one. Furthermore, the study of large order quantiles is more significant: the quantile of order 99\% is 2.33 standard deviations from the mean for the normal distribution, while it is 4.9 standard deviations for the lognormal one. In conclusion, even if both distributions have an infinite upper bound, the behavior in extremes is completely different: the tail of the log-normal distribution is heavier than the tail of the normal distribution, letting extreme values appear much more frequently than the normal distribution does.

The principle of the Randomized Gutenberg-Richter model is near the principle of the Gamma distribution proposed by Kagan in [22]: here, an normal taper is applied to the density of the exponential distribution of the magnitudes, while Kagan applies an exponential taper to the Pareto density of the seismic moment. Indeed, the final distribution of the magnitudes is defined by its probability density (pdf):

\[
p_M(m) = \int p_{M|M_{\text{max}}=m_{\text{max}}}(m)p_{M_{\text{max}}}(m_{\text{max}}) \, dm_{\text{max}}
\]

where \( M \) is the magnitude, \( p_{M|M_{\text{max}}=m_{\text{max}}} \) the pdf of the conditional magnitude \( M|M_{\text{max}} = m_{\text{max}} \) and \( p_{M_{\text{max}}} \) the pdf of \( M_{\text{max}} \).

Determining that final distribution is not easy to do. We used the open source software OpenTURNS [4] that provides advanced probabilistic modelizations and statistical functionalities. The integral (8) is calculated thanks to quadrature methods that prove to be very efficient in that case.

2.4 Recurrence models evaluation

This section evaluates each recurrence model.

The truncated Gutenberg-Richter model is an improvement with respect to the Gutenberg-Richter model as the frequency-magnitude curve falls off from the straight line when the magnitude increases. The difficulty remains in the choice of \( m_{\text{max}} \) on which the recurrence model strongly depends. This value raises debates among the seismologist community.

The GPD-based model is an improvement with respect to the truncated Gutenberg-Richter model as the
frequency-magnitude curve falls off from the straight line when the magnitude increases without any additional hypothesis on the maximum possible magnitude: its upper bound results from the GPD estimate and is equal to \( m_{\text{max}} = u - \alpha / \xi \) if \( \xi < 0 \).

However, for areas with low seismicity, the estimate of the GPD-based model might be tricky, due to the eventual low number of earthquakes in stable continental regions and particularly earthquakes with large magnitude.

The Randomized Gutenberg-Richter model is based on the following advantages: 1. Its estimate is based on all the data and not only the largest ones; 2. It is able to model the uncertainty of the upper truncature thanks to an extreme value analysis. In domains where the seismicity is low, that feature proves to be essential.

To conclude, it is worth considering the following good practices: for domains of high seismicity, if no value \( m_{\text{max}} \) achieves universal agreement, the GPD-based model is likely to help to model the tail distribution without any additional hypothesis. For domains of low seismicity, the Randomized Gutenberg-Richter model should be used after the random modelization of \( m_{\text{max}} \) has been done on a larger domain thanks to the GPD-based model. In general, the simultaneous use of all the models help to better understand their respective limitations and their impacts on hazard assessments.

### 3 Probabilistic Seismic Hazard Assessment

This section presents the integration of the three recurrence models previously defined into a probabilistic seismic hazard calculation. The aim is to check their consistency with the truncated Gutenberg-Richter model, commonly used in PSHA, as well as to assess their impacts on seismic hazard assessments.

The PSHA computational model used in this study corresponds to Cornell’s approach [14]. This approach calculates the probability of exceeding a target value of ground motion, during a fixed period of time. A probabilistic seismic hazard calculation takes into account the influence of all seismic sources able to contribute to the hazard, as well as the spatial and temporal distribution of the seismicity of each source. Four main steps are necessary to perform a PSHA calculation:

1. Identification of the area seismic sources that might affect the site of interest. Area source models are often used when one can not identify a specific fault. Usually, in PSHA, a uniform seismicity distribution is assigned to each area source, implying that earthquakes are equally likely to occur at any point within the source zone.

2. Specification of temporal and magnitude seismicity distributions for each source. Cornell’s approach assumes that earthquakes follow a Poisson point process. The most commonly used frequency-magnitude model is the Gutenberg-Richter relationship (1).

3. Calculation of ground motions and their uncertainty. Ground Motion Prediction Equations (GMPE) are used to predict the ground motion at the site itself. The parameters of interest include different intensity measures (peak ground acceleration, peak ground velocity, peak ground displacement, spectral acceleration, intensity, strong ground motion duration). Most GMPE available today are empirical and depend on the earthquake magnitude, source-to-site distance, type of faulting and local site conditions.

4. Integration of uncertainties on earthquake location, magnitude and GMPE.

The ultimate result of a PSHA calculation is a seismic hazard curve giving the annual probability of exceeding a specified ground motion parameter. Conceptually, the computation of the seismic hazard is relatively simple, as the calculation is carried out with the following equation:

\[
\lambda(IM > x) = \lambda(M > m_{\text{min}}) \times \int \int P(IM > x | (m, r, \theta)) f_M(M > m_{\text{min}}) f_R(r) \, dm \, dr \tag{9}
\]

where \( IM \) is the Intensity Measure of the ground motion parameter; \( x \) is the target intensity; \( \lambda(IM > x) \) is the annual rate of earthquakes leading to an excess of the target; \( \lambda(M > m_{\text{min}}) \) is the annual rate of earthquakes of magnitude \( M \) larger than the minimum magnitude \( m_{\text{min}} \) selected for the source zone; \( f_M(M > m_{\text{min}}) \) and \( f_R \) are the pdf of the magnitude and the distance from the source zone; \( P(IM > x | (m, r, \theta)) \) is the probability that an earthquake of magnitude \( m \) at the distance \( r \) from the site generates an intensity larger than \( x \); \( \theta \) regroups all the other fixed parameters regarding the site, the type of faults, etc. This probability is determined using a model of GMPE that takes into account the model errors due to amplifications and misunderstandings of geological phenomena.

The PSHA model used in this study is a simplified but realistic model composed of only one area source. It uses real data and a conceptual calculation model respecting the international rules. The tool used for the calculations is OpenQuake [29].

The PSHA calculations take into account the following uncertainties:

- Maximum magnitude uncertainty: both models GPD and GRT-Rd can model this epistemic uncertainty,
indirectly through the asymptotic distribution of the parameter estimators for the GPD, or directly with the distribution of the random upper bound for the $GRt$-rd model.

- Recurrence parameters uncertainty: this epistemic uncertainty is taken into account thanks to the asymptotic distribution of the parameter estimators that enables to generate several possible sets of parameters: $(a, b)$ for the Gutenberg-Richter based models and $(\sigma, \xi, \mu_u)$ for the GPD model. Each set of parameters thus constitutes an alternative branch of the logic tree.

The final logic-tree of PSHA calculations is composed of 100 branches sum up on Figure 1 illustrated in the Gutenberg-Richter case, each one associated to some specific probabilistic hypotheses. In this study, all the branches have the same weight.

The GMPE selected for the calculations is based on records in the Pan-European region [10] with the following main characteristics: Magnitudes range: $[4.0, 7.6]$, the distances: $R_{jb} = 0 - 300$ km, the spectral periods: (PGA; 0.01-4 s)

4 Application to the Alps Region, France

4.1 The Alps seismic catalog

In this section, we apply the whole approach to the mountainous region Alps which is one of the most active regions in France. A homogenous seismotectonic zone was delimited based on geological, structural, geophysical, neotectonic and seismological data [16].

We used the FCAT17 catalog [27] that gathers recent instrumental magnitudes measured by seismometers ([11], [12]) and magnitudes attributed to historical events [38]. Magnitudes were derived from historical documents ([39], [37]) in a 2-step process: 1. The historical information was translated into macroseismic intensities; 2. Magnitudes were derived from these macroseismic intensities. All the magnitudes are given in $M_w$.

Figure 2 presents the location of earthquakes inside the delimited seismotectonic domain.

A declustering method was applied to separate the time-dependent part of seismicity from the seismic activity related to aftershocks, foreshocks or clusters of earthquakes [16].

Seismic events of this catalog exhibit the following main features: given the limited sensitivity of seismographic networks, small earthquakes are not completely sampled in the earthquake catalog. This makes it necessary to introduce a general catalog completeness threshold equal to $m_0 = 3.5$.

Furthermore, the catalog level of completeness is not constant in time: the further we go in the past, the higher the threshold is. Complete periods associated to each class of magnitudes are defined in Table 1. We assume that the resulting restricted catalog provides a complete statistical sample. Our database finally contains $N = 583$ earthquake events, from year 1524 until year 2016, whose magnitudes range from 3.5 until 6.69. The database is large enough to achieve robust results with all the models presented.

At last, the accuracy of the magnitudes can sometimes be no better than 0.5 unit: this can be due to imprecise measurement methods or to bias introduced when converting from various local magnitude scales. In this case, we can not account for the precise value of the magnitude to fit our models: we only consider that it belongs to a class of length 0.5. This way of taking into
account uncertainties on the magnitudes was first proposed by [42] and then generalized by [17] where the joint estimate of the parameters is introduced. In [18], the author has extended it to the GPD-based model.

For all these reasons, we can not separately analyse the two distributions: the temporal sequence of earthquake numbers and the magnitudes. That is why the estimators are built from the likelihood of the point Poisson process rather than from the separate likelihoods (temporal sequence and magnitude).

4.2 Recurrence models estimations

4.2.1 The Gutenberg-Richter model (GRt)

The maximum magnitude selected for the Alps region is $m_{\text{max}} = 7.5$, based on tectonic information and expert opinion. The $(a, b)$ estimate with its 90%-confidence intervals is:

\[
\begin{align*}
\hat{a} &= 3.87 & P\{a \in [3.68, 4.06]\} &= 0.9 \\
\hat{b} &= 0.965 & P\{b \in [0.912, 1.02]\} &= 0.9
\end{align*}
\]

Generating a sample of $(a, b)$ from its asymptotic normal distribution allows to draw all the corresponding recurrence models and thus visualize the impact of the estimate uncertainty of $(a, b)$ on the recurrence model. Figure 3 shows 100 statistically possible models. Table 2 gives the 90% confidence intervals of $\mathbb{E}[N_m]$ for several magnitudes: we note that the mean annual number of earthquakes whose magnitude is larger than 7 is less than $1.11 \times 10^{-3}$ with a 95% probability.

Among several tests to validate the fitted Poisson point process, we focus here on the good fitting of the estimated distribution of magnitudes on data: it consists of drawing the estimated distribution of $M | M \in C_k$, where $C_k$ is a class of magnitudes, and the empirical cumulative rates in the same plot: Figure 4 shows that the fitting is correct.

### Table 1: Complete periods for each class of magnitude, in years [16].

| Magnitude class ($M_w$) | Complete period (years) |
|------------------------|-------------------------|
| [3.5, 4]               | [1850, 2016]            |
| [4, 4.5]               | [1850, 2016]            |
| [4.5, 5]               | [1700, 2016]            |
| [5, 5.5]               | [1600, 2016]            |
| $\geq 5.5$             | [1500, 2016]            |

### Table 2: GRt model: Uncertainties on $\mathbb{E}[N_m]$ induced by uncertainties on $(a, b)$: 5% quantile, 95% quantile and median of $\mathbb{E}[N_m]$.

| $m (M_w)$ | 5% quant. | 50% quant. | 95% quant. |
|-----------|-----------|------------|------------|
| 3.5       | 3.25      | 3.49       | 3.75       |
| 4         | 8.46 $10^{-1}$ | 9.22 $10^{-1}$ | 1.00       |
| 4.5       | 2.99 $10^{-1}$ | 3.39 $10^{-1}$ | 3.84 $10^{-1}$ |
| 5         | 1.05 $10^{-1}$ | 1.25 $10^{-1}$ | 1.48 $10^{-1}$ |
| 5.5       | 2.56 $10^{-2}$ | 2.32 $10^{-2}$ | 4.14 $10^{-2}$ |
| 6         | 8.78 $10^{-3}$ | 1.17 $10^{-2}$ | 1.56 $10^{-2}$ |
| 6.5       | 2.89 $10^{-3}$ | 4.05 $10^{-3}$ | 5.66 $10^{-3}$ |
| 7.0       | 5.10 $10^{-4}$ | 7.54 $10^{-4}$ | 1.11 $10^{-3}$ |

### Fig. 3
GRt: Uncertainties on the recurrence model due to uncertainties on $(a, b)$: 100 statistically possible models.

### Fig. 4
GRt: Validation of the magnitude distribution: empirical and estimated distributions of $M | M \in C_k$ for each class $C_k$ of length 0.5: the blue line corresponds to the model while the empirical cumulative rates are drawn in black.
4.2.2 The Generalized Pareto Distribution based model (GPD)

This model and its application to the Alps region has already been presented in [18]. In this section, we mention the essential relations only and we focus on the additional developments. We first determined the threshold \( u \) thanks to different modelizations of the distribution of \( M_{\theta} = M|M \geq m_0 \); non parametric fitting with normal or uniform kernels. We fixed \( u = 4.0 \) that is a quantile of order larger than 80%. The estimates with their 90%-confidence intervals are:

\[
\begin{align*}
\hat{\sigma} &= 0.586 \quad \mathbb{P}[\sigma \in [0.503, 0.668]] = 0.9 \\
\hat{\xi} &= -0.143 \quad \mathbb{P}[\xi \in [-0.226, -0.0608]] = 0.9 \\
\hat{\mu}_u &= 0.900 \quad \mathbb{P}[\mu_u \in [0.794, 1.00]] = 0.9 \\
\end{align*}
\]

(10)

In [18], the Reader can find 100 statistically possible models due to uncertainties in the estimation of \((\sigma, \xi, \mu_u)\) and the 90%-confidence interval of \(\mathbb{E}[N_m]\). We note that the mean annual number of earthquakes whose magnitude is larger than 7 is less than \(3.37 \times 10^{-4}\) with a probability of 95%.

In addition to many validation tests already detailed in [18], we focus here on an additional test based on the annual maximum magnitude over years where at least one earthquake with magnitude larger than \( u \) happened. If \( M_u = M|M \geq u \) denotes the magnitudes greater than \( u \), and \( N_u \) the Poisson distribution parametrized by \( \mu_u \) that counts the annual number of such earthquakes, then the annual maximum is defined by:

\[ M_{N_u} = \max_{1 \leq i \leq N_u} M^i_u \quad \text{given that} \quad N_u \geq 1 \quad (11) \]

Its cdf can be written as:

\[ G_u(m) = \frac{e^{-\mu_u R_u \mu (m-u)}}{1 - e^{-\mu_u}}, \quad \forall m \geq u \quad (12) \]

The test consists in comparing \( G_u \) built with \((H_u \hat{\xi}, \hat{\mu}_u)\) with non parametric fittings built on empirical maximum annual magnitudes (a kernel smoothing with normal kernels or a histogram). Figure 5 shows that the fitting is correct.

4.2.3 The Randomized Gutenberg-Richter model (GRT-rd)

To elaborate the Randomized Gutenberg-Richter model, we considered the distribution of the quantile of order \(1 - 10^{-4}\) of the GPD model to randomize the upper bound of the exponential distribution. From (6) and (10), we obtained the estimate \( \hat{q}_N = 6.75 \).

Its asymptotic distribution is normal, with mean 6.75 and variance \( 7.63 \times 10^{-2}\) calculated from (7). We truncated that normal distribution to the lower bound equal to the maximum magnitude contained in the catalog \( m_{\text{max, cat}} = 6.69 \), which can be written as:

\[ M_{\text{max}} \sim \mathcal{N}(6.75, 7.63 \times 10^{-2}) \text{ truncated to } [6.69, +\infty[ \]

Figure 6 shows the pdf of the distribution of \( \hat{q}_N \).

Figure 7 shows 100 statistically possible models built from couples \((a, b)\) generated by the asymptotic distribution of \((\hat{a}, \hat{b})\).

4.2.4 Recurrence models comparison

This section compares the results obtained in the Alps region for the three recurrence models. Figure 8 shows the pdf of the conditional magnitude \( M_u = M|M \geq u \)
obtained in each case. We observe that the Randomized Gutenberg-Richter model appears as a compromise between the truncated Gutenberg-Richter model and the GPD-based model: for the magnitudes below magnitude 6.69, the Randomized Gutenberg-Richter model behaves like the truncated Gutenberg-Richter model as the blue and red curves overlap; while for the magnitudes above magnitude 7, the Randomized Gutenberg-Richter model behaves like the GPD-based model as the blue and violet curves overlap. Between magnitudes 6.69 and 7, the Randomized Gutenberg-Richter model gradually moves from the GRt model to the GPD-based model. We can visualise here the taper effect applied on the probability function of the magnitudes, which entails a soft normal-like roll-off truncation.

At last, we note that the truncated Gutenberg-Richter model has the heaviest tail distribution, while the GPD-based model has the slightest one: the pdf of GRt (red line on Figure 8) is largely above the pdf of GPD (violet line). It means that large magnitudes have a higher probability to appear with the truncated Gutenberg-Richter model than with the GPD-based model, even if the upper bound of the GPD-based model is larger than $m_{\text{max}} = 7.5$ fixed to the truncated Gutenberg-Richter model.

Figure 9 shows all the recurrence models: we can notice the soft decrease of the GRt-rd recurrence model compared to the steep decrease at the upper cut-off of the GRt model. Table 3 details some values of the recurrence models at different magnitudes.

| $m$ ($M_w$) | GRt-imp | GPD | GRt-rd |
|-------------|---------|-----|--------|
| 6.0         | $1.16 \times 10^{-2}$ | $8.26 \times 10^{-3}$ | $1.04 \times 10^{-2}$ |
| 6.5         | $3.52 \times 10^{-3}$ | $1.22 \times 10^{-3}$ | $2.36 \times 10^{-3}$ |
| 7.0         | $8.72 \times 10^{-4}$ | $8.67 \times 10^{-5}$ | $1.09 \times 10^{-4}$ |
| 7.5         | $0.0$    | $1.16 \times 10^{-6}$ | $4.06 \times 10^{-7}$ |
Table 4 Summary of the main parameters used to perform the PSHA calculations.

| PSHA parameters                  |                      |
|----------------------------------|----------------------|
| Minimum magnitude                | 4.5 $M_w$            |
| Maximum magnitude for GRt        | 7.5 $M_w$            |
| Earthquake depth                 | 10 km                |
| $V_s(30)$                        | 800 m.s$^{-1}$       |
| Maximum integration distance     | 150 km               |
| Magnitude scale relation for simple fault | WC1994               |
| Rupture aspect ratio             | 1                    |

4.3 Seismic Hazard Assessment

4.3.1 Probabilistic seismic hazard calculations

Seismic hazard curves and Uniform Hazard Spectrum (UHS) were calculated using the previous recurrence models. The calculation was done in the center of the area source delimited on Figure 2. The logic tree used for calculations is drawn on Figure 1 and Table 4 recalls the parameters selected for the PSHA calculations.

4.3.2 Impact of the recurrence models on the probabilistic seismic hazard

Figure 10 shows the hazard curves for the spectral periods PGA and 1 s. Results show that for all spectral accelerations the curves are similar; we note a decrease in the hazard calculated with the GRt-rd and GPD models compared to the common truncated Gutenberg Richter model GRt. This decrease amplifies for long periods, as it is observed on the hazard curve at 1 s.

Figure 11 illustrates the impact of the three recurrence models on the UHS calculated for the return periods: 475, 5 000, 10 000 and 20 000 years. This impact is particularly important for periods higher than 0.1 s.

Figure 12 quantifies the relative differences of the seismic hazard generated by the GPD and GRt-rd models with respect to the common truncated Gutenberg-Richter model GRt. The curves show a clear truncation at about 0.1 s: below 0.1 s, there is a marked decrease in the seismic hazard with both GPD and GRt-rd recurrence models, as follows:

- for return periods larger than 5 000 years, the decrease in the seismic hazard of the GPD based model is estimated at around 30, 40 and 45% for the periods 1, 2 and 3 s respectively with respect to the common truncated Gutenberg-Richter model GRt;
- for return periods larger than 5 000 years, the decreasing in the seismic hazard of the Randomized Gutenberg-Richter model GRt-rd is estimated at about 15, 18 and 20% for the periods 1, 2 and 3 s respectively with respect to GRt.

- for return periods larger than 5 000 years and spectral periods lower than around 0.1 s, the reduction in seismic hazard is around 20 and 10% for both GPD and GRt-rd models with respect to GRt.

- for return periods of 475 years, the decrease in the hazard for the long periods (1, 2 and 3 s) is almost the same than those observed for higher return periods. However, for shorter periods, lower than 0.1 s, the decrease in the hazard is around 10% for both GPD and GRt-rd models with respect to GRt.

We explain that significant decrease by the different modelizations of extreme magnitudes: the tail distribution of magnitudes of both GPD and GRt-rd is much slighter than the tail of GRt, as drawn on Figure 8. The pdfs of these models drop towards 0 faster than the pdf of the truncated Gutenberg-Richter model. In particular, the probability to exceed a magnitude equal to 7 calculated from GPD and GRt-rd is 8 times lesser.
Fig. 11 Comparison of the Uniform Hazard Spectra (UHS) between the new recurrence models: Randomized Gutenberg-Richter $GRt-rd$, Generalized Pareto Distribution $GPD$ and the common Gutenberg-Richter model $GRt$ for the return periods (RP): 475, 5000, 10 000 and 20 000 years.

Fig. 12 Relative differences of the UHS calculated with the new recurrence models ($GPD$ and $GRt-rd$) with respect to the common Gutenberg Richter model ($GRt$) considered as a reference.
than the probability calculated from $GRt$. That feature is noticeable on the values of the recurrence models: the mean annual number of earthquakes with magnitude larger than 7 is $E[N_7] = 8.72 \times 10^{-4}$ with $GRt$ model, $E[N_7] = 1.09 \times 10^{-4}$ with $GRt$-rd and $E[N_7] = 0.867 \times 10^{-4}$ with $GPD$.

We insist on the fact that even if the maximum possible magnitude estimated by the $GPD$-based model, equal to 8.06, is larger than the $m_{max} = 7.5$ fixed by experts to the truncated Gutenberg-Richter model, the $GPD$-based model gives less weight to large magnitudes than the truncated Gutenberg-Richter model. We have the same feature for the $GRt$-rd model whose maximum possible magnitude is theoretically infinite, but whose probability to exceed the magnitude 7 is $1.47 \times 10^{-8}$.

Thus, these extreme-value-based recurrence models reduce the population of earthquakes with larger magnitudes predicted for the hazard calculations. Such a smaller population of larger magnitudes induces a reduction in the seismic hazard, precisely for longer periods (e.g., 2 and 3 s in this study).

4.3.3 Impact of parameters uncertainty on the probabilistic seismic hazard

In this section, we propagate the uncertainty on the estimation of the recurrence model parameters through PSHA calculations.

We generated 100 statistically possible sets of parameters of each recurrence model, according to the asymptotic distribution of their estimators, which generated 100 statistically possible branches in the PSHA logic-tree (see Figure 1).

For each return period, at each period, we drew on Figure 13 the quantile of level 5%, 50% (median value) and 95% of the UHS. Figure 14 shows the length of the confidence interval of level 90% of the UHS curves, defined as the difference between the 95% quantile and the 5% quantile. We notice that the Randomized Gutenberg-Richter model $GRt$-rd has the lowest uncertainties with respect to the $GPD$ and $GRt$ models for all the periods. For short periods (below 0.1 s), uncertainties are higher than for longer periods for all the models. The higher uncertainties of the $GPD$ model are due to the fact that this model is estimated from less data than for both other models: magnitudes greater than $u = 4$ are considered for the $GPD$ model (around 200 seismic events), while all the magnitudes of the catalog are taken into account for both other models (around 600 seismic events).

![Figure 13](image1.png)

Fig. 13 5%, 50% and 95% quantile curves of the UHS calculated with the new recurrence models ($GPD$, $GRt$-rd) and the common Gutenberg-Richter model $GRt$.

5 Conclusions

In this paper, we presented two extreme-value-based recurrence models and we compared them to the truncated Gutenberg-Richter model that is commonly used in PSHA calculations. We evaluated their impacts on probabilistic seismic hazard assessments, using the Alps data.
Both extreme-value based-models avoid fixing a priori the maximum possible magnitude $m_{\text{max}}$. This parameter results from a statistical analysis adapted to extreme values. This feature is particularly interesting as there is no widely accepted method to estimate $m_{\text{max}}$.

The Randomized Gutenberg-Richter model appears as a mix of the truncated Gutenberg-Richter model and the GPD-based model. It looks promising because:

- the exponential distribution that has been widely accepted by the seismologist community, is still used to model the magnitude distribution,
- the randomness of $M_{\text{max}}$ allows to take into account the uncertainty on the maximum possible magnitude value, whose determination is a tricky matter,
- the distribution of $M_{\text{max}}$ relies on an extreme value analysis that aims at modeling the behavior of the extreme magnitudes,
- the truncation of the magnitude distribution is soft, which provides better agreement with both fundamental seismological principles and earthquake catalogs.

The new recurrence models introduce a reduction of the hazard compared to the truncated Gutenberg-Richter model commonly used in PSHA. This decrease is more significant for long periods (eg 1, 2 and 3 s), it is estimated to between 15 – 20% for the GRT-rd model and to 25 – 40% for the GPD model.

The GPD-based recurrence model as well as the Randomized Gutenberg-Richter model are expected to result in a more realistic assessment of the probabilistic seismic hazard.

The use of these extreme-value-based models for regions of moderate or low seismicity remains difficult, mainly due to the poor number of seismic events. However, the Randomized Gutenberg-Richter model is an encouraging alternative for hazard estimates and could help for such regions.

In perspective, we plan to apply the GPD model in a domain large enough and with sufficient seismicity ($M > 4$) to provide a robust estimate of the tail distribution of the magnitudes in the whole domain. Then, the aim is to use the Randomized Gutenberg-Richter model for each area source inside the tectonic domain: the maximum upper bound follows the tail estimated on the whole domain and both parameters $(a, b)$ are estimated from the seismic events of each area source.

6 Declarations
Not applicable.

References

1. Abrahamson, N.: State of the practice of seismic hazard evaluation. Proc GeoEng, Melbourne, Australia (2000)
2. Abrahamson, N., Birkhauser, P., Koller, M., Mayer-Rosa, D., Smit, P., Sprecher, C., Tinic, S., Graf, R.: A comprehensive probabilistic seismic hazard assessment for nuclear power plant in Switzerland. Proc 12th Eur Conf Earthquake Eng, London, UK, paper 633 (2002)
3. Anderson, J., Wensnousky, G., Stirling, M.: Earthquake size as a function of slip rate. Bull Seism Soc Am 86, 683–690 (1996)
4. Baudin, M., Dutfoy, A., Popelin, A., Iooss, B.: OpenURNS: An Industrial Software for Uncertainty Quantification in Simulation. Handbook of Uncertainty Quantification, Springer (2017). URL www.openURNS.org
5. Beirlant, J., Alves, J.F., Gomes, I.: Tail fitting for truncated and non-truncated pareto-type distributions. Extremes 19(-), 429–462 (2016)
6. Beirlant, J., Alves, I.F., Reynkens, T.: Fitting tail affected by truncation. Elect. J. Stat (-) (2017)
7. Beirlant, J., Goegebeur, Y., Teugels, J., Segers, J.: Statistics of extremes: theory and applications. Wiley, UK (2004)
8. Beirlant, J., Kijko, A., Reynkens, T., Einmahl, J.: Estimating the maximum possible earthquake magnitude using extreme value methodology: the Groningen case. Natural Hazards 98(-), 1091–1113 (2019)
9. Berge-Thierry, C., Voldoire, F., Raguenneau, F., Lopez-Caballero, F., LeMouaflt, A.: Main achieve-ments of the multidisciplinary sinaps@ research project: Towards an integrated approach to per-form seismic safety analysis of nuclear facilities. Pure Appl Geophys 177, 2299–2351 (2020)
10. Bindi, D., Massa, M., Luzi, L., Ameri, G., Pacor, F., Pugliar, R., Augliera, P.: Pan-european ground-motion prediction equations for the average horizontal component of pga, pgv, and 5%-damped psa at spectral periods up to 3.0 s using the resource dataset. Bull Earthquake Eng 12(1), 391–430 (2014)
29. Pagani, M., Monelli, D., G. Weatherill: Openquake: a new catalog of instrumental seismicity for metropolitan france. Bull Soc Geol Fr 186, 3–19 (2015)
30. Page, R.: Aftershock and microaftershocks. Bull Seism. Soc Am. 58(5), 1583–1606 (1968)
31. Pecker, A., Faccioli, E., Gurpinar, A., Martin, C., Renault, P.: An overview of the SIGMA research project. Springer, Berlin (2017)
32. Pickands, J.: Statistical inference using extreme order statistics. Annals of Statistics 3, 119–131 (1975)
33. Pisarenko, V., Sornette, A., Sornette, D., Rodkin, M.: Characterization of the tail of the distribution of earthquake magnitudes by combining the gev and gpd description of extreme value theory. Pure Appl Geophys 171, 1599–1624 (2014)
34. Pisarenko, V., Sornette, A., Sornette, D., Rodkin, M.: Characterization of the tail of the distribution of earthquake magnitudes by combining the gev and gpd description of extreme value theory. Pure Appl Geophys 160, 2343–2364 (2003)
36. Renault, P.: Bollettino di Geofisica Teorica ed Applicata 55(1), 149–164 (2014)
37. Senfaute, G., Pecker, A., Labbé, P., Sis, E., Bergethier, C., Rzepka, J., Conti, P.: Contribution of the SIGMA research program to analyses of uncertainties in seismic hazard assessment. 9iéme Colloque National. AFPS, France (2015)
38. SISFRANCE: URL www.sisfrance.fr.