Risk-informed representative earthquake scenarios for Valparaíso and Viña del Mar, Chile

Hugo Rosero-Velásquez[1], Mauricio Monsalve[4,5], Juan Camilo Gómez Zapata[2,3], Elisa Ferrario[4,5,6], Alan Poulos[7], Juan Carlos de la Llera[5,8], and Daniel Straub[1]

[1]Engineering Risk Analysis Group, TU Munich, Arcisstr. 21, 80333 Munich, Germany
[2]Seismic Hazard and Risk Dynamics, GFZ German Research Centre for Geosciences, 14473 Potsdam, Germany
[3]Institute for Geosciences, University of Potsdam, Karl-Liebknecht-Str. 24–25, 14476 Potsdam, Germany
[4]School of Engineering, Pontificia Universidad Católica de Chile, Santiago, Chile
[5]Research Center for Integrated Disaster Risk Management (CIGIDEN), ANID/FONDAP/1522A0005, Santiago, Chile
[6]Ricerca sul Sistema Energetico - RSE S.p.A., Milano, Italy
[7]Department of Civil and Environmental Engineering, Stanford University, Stanford, California, U.S.A
[8]Department of Structural and Geotechnical Engineering, Pontificia Universidad Católica de Chile, Santiago, Chile

Correspondence: Hugo Rosero-Velásquez (hugo.rosero@tum.de)

Abstract. Different risk management tasks, such as land-use planning, preparedness and emergency response, utilize scenarios of earthquake events. A systematic selection of such scenarios should aim at finding those that are representative of a certain severity, which can be measured by its consequences to the exposed assets. For this reason, it has been proposed to define a representative scenario as the most likely one leading to a loss with a specific return period, e.g., the 100-year loss. We adopt
5   this definition and develop enhanced algorithms for determining such scenarios for multiple return periods, based on a seismic catalog. With this approach, we identify representative earthquake scenarios for the Valparaíso and Viña del Mar communes in Chile. Because the earthquake scenarios are defined in terms of the loss exceedance, the scenarios vary in function of the exposed system. In this contribution, we consider separately the residential building stock and the electrical power network, and identify and compare earthquake scenarios that are representative for these systems.

10   1   Introduction

Due to the complexity of earthquake events and the response of infrastructure and society to these events, risk managers analyze potential impacts of strong seismic events through representative earthquake scenarios (e.g., Salgado-Gálvez et al., 2018; Aguirre et al., 2018). Such earthquake scenarios are also the starting point for more detailed risk assessment (e.g., Jayaram and Baker, 2010; Miller and Baker, 2015), and recommendations for improving risk management (e.g., Chatelain
15   et al., 1995).

Representative scenarios are commonly selected based on expert knowledge (e.g., Aguirre et al., 2018) and past events (e.g., Indirli et al., 2011). In Probabilistic Seismic Hazard Analysis (PSHA) (Cornell, 1968), seismic catalogs generated from source models have also been used for the selection of representative scenarios (McGuire, 1995; Miller and Baker, 2015).

A particular approach for scenario selection is based on hazard disaggregation (Bazzurro and Cornell, 1999). It utilizes the
20   conditional probability of different hazard scenarios given that an intensity measure (e.g., peak ground acceleration, PGA)
at a specific site of interest either equals or exceeds a threshold (Fox et al., 2016). Classic hazard disaggregation does not
explicitly consider the losses of the affected engineering systems, which are often a function of the intensity measures at
multiple locations and which are subject to uncertainty. However, the method has been extended to take into account direct
losses from building stocks, considering scenarios in terms of magnitude and hypocentral distance (Goda and Hong, 2009).
25   Because the accumulated loss can be defined for any stock of spatially distributed buildings, loss disaggregation implicitly
considers spatially distributed intensity measures.

     Rosero-Velásquez and Straub (2022) proposed a definition of a representative hazard scenario, associated with a return period
t. It is defined as the most likely earthquake scenario that leads to the loss value associated with this return period t. They also
presented a numerical procedure for selecting the representative hazard scenario in a continuous space of source parameters
30   with a surrogate model and active learning, thus considering the uncertainty in the conditional losses given a hazard scenario.
The definition formalizes the loss disaggregation presented by Goda and Hong (2009) and generalizes it for any natural hazard
and spatially distributed systems, including infrastructure networks. In the context of seismic risk analysis, that definition tells
that representative earthquake scenarios can be different for different engineering systems, even if they are located in the same
area.

35   Considerable work has been devoted to the study of the seismic hazard, vulnerability, and risk in the Valparaíso coastal
area of Chile due to its high population density and economic importance in combination with strong seismic activity. Recent
earthquakes that led to significant damages occurred in 1971 with $M_w = 7.8$, in 1985 with $M_w = 8.0$ (Indirli et al., 2011),
and in 2010 with $M_w = 8.8$ (de la Llera et al., 2017). Recent studies on the Valparaíso area deal with seismic characterization
(e.g., Carvajal et al., 2017; Candia et al., 2020), source models (e.g., Poulos et al., 2019; Pagani et al., 2021), ground motion
40   models (Montalva et al., 2017), building exposure models (e.g., Yepes-Estrada et al., 2017; Jiménez et al., 2018; Gómez-Zapata
et al., 2022b, b), damage analysis on individual buildings (e.g., Indirli et al., 2011; Jünemann et al., 2015), socio-economic
impact (Jiménez Martínez et al., 2020), and seismic risk analysis of the electric power network (Ferrario et al., 2022) and road
network (Allen et al., 2022). Additionally, Indirli et al. (2011) identified representative earthquake scenarios using historical
events and expert knowledge, for generating representative ground motion time series, but solely from the hazard point of view
45   and disregarding the risk component.

     This paper determines representative earthquake scenarios for different return periods for the residential building stock and
the power supply in Valparaíso and Viña del Mar communes. We adapt and extend the methodology described in (Rosero-
Velásquez and Straub, 2022) for identifying scenarios associated with different return periods from a seismic catalog. The
representative scenario is found directly by solving a stochastic optimization problem; namely the identification of the mode of
50   the conditional distribution of the source parameters given the t-year loss among the scenarios in the catalog. Kernel-density
estimation (KDE) is used to approximate this conditional density. The stochastic optimization problem is solved with an active
learning strategy, whereby the uncertainty in the objective function is estimated by bootstrapping.

We introduce the definition of representative earthquake scenario more formally in Section 2. Then we present the methodology for computing the scenarios on a seismic catalog in Section 3, and illustrate it with idealized examples in Section 4. The description of the study area, the utilized hazard and system models, are presented in Section 5. The results are given in Section 6 and discussed in Section 7.

2 Definition of representative earthquake scenario

An earthquake scenario can be typically described by a vector θ of source parameters, including the magnitude, hypocentral distance, source longitude, latitude and depth. In a stochastic model, the scenario is a single realization of a random vector Θ, with joint probability density function (PDF) $f_\Theta(\theta)$. The PDF of Θ is obtained from a seismic source model (e.g., Poulos et al., 2019) and is conditioned on the occurrence of a seismic event, whose frequency (occurrence rate) is λ_H.

A seismic catalog is a set of n earthquake scenarios $\theta^{(1)}, ..., \theta^{(n)}$, which are independent realizations of the PDF of Θ. The catalog can be obtained by random sampling from $f_\Theta(\theta)$. Alternatively, the seismic catalog can be obtained from past ground motion events, by declustering to ensure statistical independence of these events (e.g., Poulos et al., 2019). Seismic catalogs are commonly used in PSHA (e.g., Salgado-Gálvez et al., 2018; Ferrario et al., 2022; Allen et al., 2022).

For a given earthquake scenario, ground motion prediction equations (GMPEs) result in spatially distributed intensity measures, e.g., PGA and spectral accelerations, which are input to assess the vulnerabilities to earthquakes of the exposed systems. In the general case, these predictions are stochastic. Thereafter, the model of the engineering system considers the physical and functional vulnerability and results in a loss value L.

Because of the randomness and uncertainty in the earthquake scenario, GMPE, vulnerabilities and exposure, L is a random variable with cumulative distribution function (CDF) $F_L(l)$. By combining this CDF with the earthquake occurrence rate λ_H, one obtains the loss exceedance function $\lambda_L(l)$:

$$\lambda_L(l) = \lambda_H \left(1 - F_L(l)\right). \tag{1}$$

Based on the loss exceedance function, the losses l_t with a specific return period t can be found as

$$l_t = \lambda_L^{-1}\left(\frac{1}{t}\right) = F_L^{-1}\left(1 - \frac{1}{\lambda_H t}\right), \tag{2}$$

which is defined only for $t \geq \frac{1}{\lambda_H}$. The loss l_t is also called the $t-$year loss.

Following Rosero-Velásquez and Straub (2022), the representative earthquake scenario θ_t, associated with a return period t, is defined as the most likely scenario among those causing the t-year loss l_t. In other words, θ_t is the mode of the conditional PDF of Θ given the loss $L = l_t$, also called the loss disaggregation of Θ given $L = l_t$:

$$\theta_t = \arg \max_\theta f_{\Theta|L}(\theta|l_t). \tag{3}$$

Figure 1 illustrates this conditional distribution $f_{\Theta|L}(\theta|l_t)$ for two source parameters $\Theta = (\Theta_1, \Theta_2)$, e.g., representing the magnitude M_w and the hypocentral distance R with respect to a location of interest.
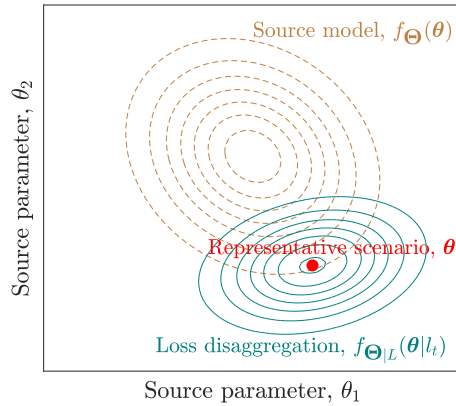
Figure 1. Illustration of representative scenario in the source parameter space, modified after Rosero-Velásquez and Straub (2022)

By Bayes' rule, Eq. (3) can be expressed in terms of $f_\Theta(\theta)$,

$$\theta_t = \arg \max_\theta f_{L|\Theta}(l_t|\theta) f_\Theta(\theta), \tag{4}$$

wherein $f_{L|\Theta}(l|\theta)$ is the conditional PDF of the losses given the hazard scenario θ. Eq. (4) illustrates that the scenario selection criterion balances the probability of the earthquake scenario (quantified by $f_\Theta(\theta)$) and the probability of the t-year losses (quantified by $f_{L|\Theta}(l_t|\theta)$). To ease the notation in the following section, we let $z_t(\theta)$ denote the objective function of Eq. (4):

$$z_t(\theta) = f_{L|\Theta}(l_t|\theta) f_\Theta(\theta). \tag{5}$$

3 Method for using scenario selection based on a seismic catalog

We consider the case where the randomness of earthquake events is represented through a seismic catalog. Specifically, we aim at identifying the earthquake scenarios in the catalog that maximize the objective function in Eq. (4) for different return periods t.

The objective function of Eq. (4) consists of the PDF $f_\Theta(\theta)$, which is known from the earthquake source model, and $f_{L|\Theta}(l_t|\theta)$, which can be approximated with conditional samples of losses. To account for the aleatory uncertainty in the modeled ground motions, it has been typically advised to draw thousands of random samples (Silva, 2016) and propagate it to the loss metrics. However, performing this amount of loss evaluations for an entire seismic catalog (normally containing dozens of thousands of events) is computationally (too) expensive. Therefore, we propose to first perform only one loss evaluation for each scenario in the catalog and use these to approximate the loss-exceedance function and l_t. The same samples are used for an initial approximation of $f_{L|\Theta}(l_t|\theta)$, the second part of the objective function. This approximation is improved by the use of active learning (AL) to identify earthquake scenarios in the catalog for which additional loss evaluations are to be performed. This methodology is an adaptation of the one proposed in Rosero-Velásquez and Straub (2022).

Figure 2 illustrates the main steps of the methodology for selecting representative earthquake scenarios for n_t return periods $t_1 > t_2 ... > t_{n_t}$. The earthquake model in this example is a single seismic source within a bounding volume, and the system is a single building.
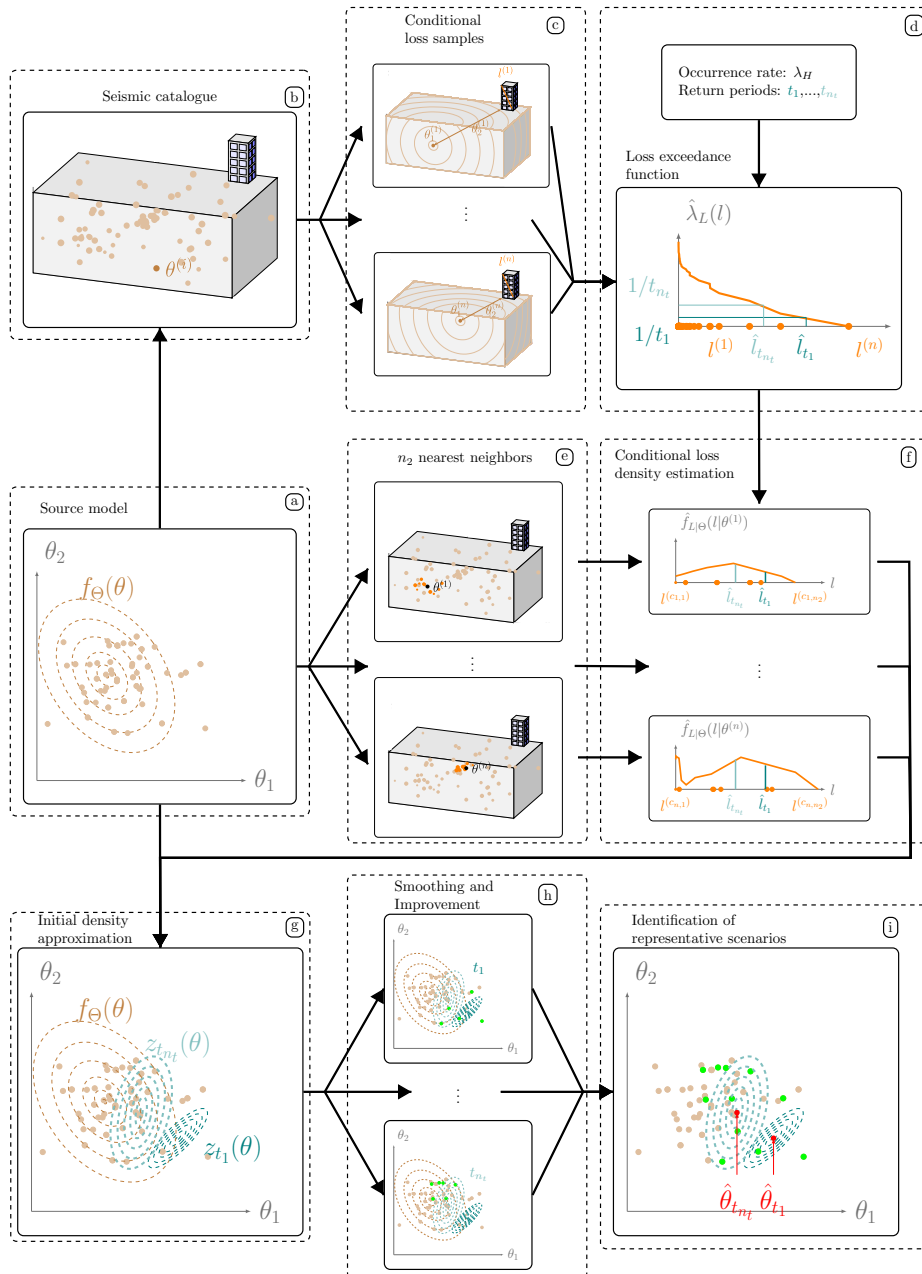
Figure 2. General procedure for selecting representative earthquake scenarios with a stochastic seismic catalog. Subsection 3.1 explains with more detail panels e and f, and Subsections 3.2 and 3.3 explain panel h. The remaining panels are referred in Section 3

The starting point is a seismic source model (panel a in Fig. 2), which consists of the occurrence rate λ_H and the PDF of the source parameters, $f_\Theta(\theta)$, together with an associated stochastic seismic catalog (panel b in Fig. 2). The catalog consists of a set of n random and independent earthquake scenarios $\theta^{(1)}, ..., \theta^{(n)}$ generated from $f_\Theta(\theta)$, possibly associated with weights. The generation of such catalogs for the study site is described in Section 5.2.

For each scenario, one simulates the ground motion fields in terms of the intensity measure (e.g., the peak ground acceleration PGA) through the GMPE. These intensity measures are the input to assess the performance of the system components, by combining them with vulnerability functions. Based on the component performances, the total losses in the system, l, are evaluated (panel c in Fig. 2). Details on the simulation of the ground motion, the system response and loss calculation for the study site are given in Section 5.

From these samples of the system losses, one obtains an estimate of the loss exceedance curve $\hat{\lambda}_L(l)$:

$$\hat{\lambda}_L(l) = \frac{\lambda_H}{n} \sum_{i=1}^{n} \mathbf{1}(l^{(i)} > l), \tag{6}$$

where $\mathbf{1}(\cdot)$ is the indicator function. In addition, one obtains estimates of the t-year losses $\hat{l}_t$ for all return periods of interest $t_1, ..., t_{n_t}$ following Eq. (2) (panel d in Fig. 2).

Since the conditional density of the losses, $f_{L|\Theta}(l|\theta)$, is not available in analytical form, we propose to approximate it with $\hat{f}_{L|\Theta}(l|\theta)$ (panels e and f in Fig. 2), as detailed in Section 3.1. We utilize this approximation in the objective function of Eq. (4) (panel g in Fig. 2) to obtain initial estimates of the objective function $z_t(\theta)$ at each scenario of the catalog, which we denote as $z_t^{(1)}, ..., z_t^{(n)}$. To reduce the scatter in the estimates of z_t, we add a smoothing step (panel h in Fig. 2), which is described in Section 3.2.

At this, and any later stage of the algorithm, we approximate the solution of Eq. (4) by (panel i in Fig. 2):

$$i_t = \underset{i=1,...,n}{\arg\max} \, z_t^{(i)} \tag{7}$$

$$\theta_t \approx \theta^{(i_t)} \tag{8}$$

The initial approximation based on a single loss evaluation per scenario in the catalog is typically poor. To enhance the accuracy, we use an active learning strategy (panel h in Fig. 2). It intelligently selects earthquake scenarios from the catalog, for which additional loss evaluations are performed. This is presented in Section 3.3.

3.1 Approximation of the objective function $z_t(\theta)$ with kernel density estimation

We approximate the conditional density $f_{L|\Theta}(l|\theta)$ using weighted kernel density estimation (KDE) (Gisbert, 2003). The KDEs at each scenario $\theta^{(i)}$ are evaluated with n_2 loss evaluations, which come from the closest scenarios $\theta^{(c_{i,1})}, ..., \theta^{(c_{i,n_2})}$ and have associated weights $w_{i,1}, ..., w_{i,n_2}$ which sum up to 1, i.e., $\sum_{j=1}^{n_2} w_{i,j} = 1$:

$$\hat{f}_{L|\Theta}(l|\theta^{(i)}) = \sum_{j=1}^{n_2} w_{i,j} \kappa\left(l, l^{(c_{i,j})}, \gamma\right) \tag{9}$$

where κ is a kernel function and γ is the bandwidth. We define the weights as $w_{i,j} = \exp\left(-d_{i,j}\right) / \sum_{k=1}^{n_2} \exp\left(-d_{i,k}\right)$, where $d_{i,j}$ is the Mahalanobis distance between $\theta^{(i)}$ and $\theta^{(c_{i,j})}$. This ensures that the loss values from scenarios similar to $\theta^{(i)}$ are given more weight in the KDE.

A common choice for κ is the Gaussian kernel function , which employs the standard Gaussian PDF $\phi(\cdot)$:

$$\kappa\left(l, l^{(c_{i,j})}, \gamma\right) = \frac{1}{\gamma}\phi\left(\frac{l - l^{(c_{i,j})}}{\gamma}\right) \tag{10}$$

and γ computed as suggested in (Silverman, 1986). Alternatively, one can employ a lognormal kernel, excluding the zero loss values (if any), whose probability $p_0^{(i)}$ is estimated from the conditional loss samples $l^{(c_{i,1})}, ..., l^{(c_{i,n_2})}$. That is,

$$\kappa\left(l, l^{(c_{i,j})}, \gamma\right) = \mathbf{1}\left(l^{(c_{i,j})} > 0\right)\left(1 - p_0^{(i)}\right)\frac{1}{l\gamma}\phi\left(\frac{\ln l - \ln l^{(c_{i,j})}}{\gamma}\right) \tag{11}$$

wherein the bandwidth γ is computed as suggested in (Silverman, 1986) but only using the logarithm of the nonzero loss samples. In consequence, the weights $w_{i,j}$ have to be adjusted excluding the zero loss samples, i.e.,

$$w_{i,j} = \mathbf{1}\left(l^{(c_{i,j})} > 0\right)\frac{\exp\left(-d_{i,j}\right)}{\sum_{k=1}^{n_2} \mathbf{1}\left(l^{(c_{i,k})} > 0\right)\exp\left(-d_{i,k}\right)} \tag{12}$$

In this case, for scenarios where all the n_2 conditional loss samples are zero, the density at l_t equals zero.

The choice of n_2 for the KDEs is associated with a trade-off: On the one hand, a small n_2 leads to a poor density estimation, but one that is based on loss samples coming from similar scenarios. On the other hand, a large n_2 produces a biased KDE from the true conditional density, since it incorporates loss samples of more dissimilar scenarios. However, the bias can be reduced by additional model evaluations. In fact, n_2 or more model evaluations at a scenario θ provide a more accurate KDE than a KDE based on model evaluations coming from the n_2 closest scenarios to θ.

For each return period, we obtain an estimate of $f_{L|\Theta}(l_t|\theta^{(i)})$ by evaluating Eq. (9) with argument $\hat{l}_t$. This process is illustrated in panels e and f of Fig. 2. By multiplication with the prior, an estimate of the objective function at all scenarios in the catalog is obtained:

$$z_t^{(i)} = \hat{f}_{L|\Theta}(\hat{l}_t|\theta^{(i)})f_\Theta(\theta^{(i)}) \tag{13}$$

To reduce the significant noise associated with these estimates, we update them with an additional smoothing step described in Sec. 3.2.

Even after the additional smoothing step, the estimates $z_t^{(i)}$ remain subject to uncertainty, due to the limited number of noisy loss evaluations and the need to pool the evaluations from multiple scenarios. For the purpose of the active learning procedure presented in Sec. 3.3, we approximate the uncertainty associated with the objective function values by modeling the estimates $z_t^{(i)}$ as Gaussian random variables $Z_t^{(i)}$. We denote their mean values as $\mu_{Z_t}^{(i)}$ and their standard deviations as $\sigma_{Z_t}^{(i)}$. The mean values are set to $z_t^{(i)}$. We estimate the standard deviations $\sigma_{Z_t}^{(i)}$ for $i = 1, ..., n$ via bootstrapping (Efron and Tibshirani, 1993) on their n_2 nearest neighbors n_b times, wherein conditional losses are resampled according to the weights $w_{i,j}$.

3.2 Smoothed estimation of the objective function $z_t(\theta)$ with Gaussian process regression

To reduce the noise in the estimates $z_t^{(i)}$, we perform an additional smoothing step via Gaussian process regression (GPR)

165 (Rasmussen and Williams, 2006). For each return period t and in each step of the active learning algorithm described in Section 3.3, we perform a separate GPR.

A drawback of GPR is that the computational cost escalates with the size of the training set n_{train}. Fitting and estimating the objective function using standard GPR is an $\mathcal{O}(n^3)$ task (Rasmussen and Williams, 2006). Therefore, we perform GPR smoothing only for estimates $\theta^{(i)}$ near the current solution of Eq. (7), and the GPR hyperparameters are learned only once in

170 the first step. Specifically, we identify the $n_{train} = 1500$ nearest scenarios using the Mahalanobis distance, train the GPR and replace the estimate of $z_t^{(i)}$ only for the training set. The other estimates of $z_t^{(i)}$ are left unaltered.

3.3 Active learning

An accurate estimation of the objective function is only important near the solution. We exploit this by employing an active learning (AL) strategy to identify scenarios for which further model evaluations are performed.

175 AL selects scenarios to evaluate through the acquisition function. Here we use the augmented expected improvement (AEI) as an acquisition function (Huang et al., 2006). It approximates for each scenario the expected value of the improvement of the objective function over the current maximum.

We modify the AEI of Huang et al. (2006) with a correction factor, which assesses the quality of the KDE at each scenario. The resulting AEI at scenario $\theta^{(i)}$ is

$$180 \quad AEI(\theta^{(i)}) = \frac{1}{c_{neigh}^{(i)}} \mathbb{E}\left[\max\left(Z_t^{(i)} - z_t^*, 0\right)\right] \tag{14}$$

wherein z_t^* is the estimate of the objective function at the current best solution θ^*, which is defined as (Huang et al., 2006):

$$\theta^* = \underset{i=1,\dots,n}{\operatorname{argmax}}\left(z_t^{(i)} + c\sigma_{Z_t}^{(i)}\right) \tag{15}$$

with $c = 1$.

The factor $c_{neigh}^{(i)}$ considers the KDE estimation quality at $\theta^{(i)}$. We define it as follows:

$$185 \quad c_{neigh}^{(i)} = \max\left(\sum_{j=1}^{n_2} \exp\left(-d_{i,j}\right), n^{(i)}\right) \tag{16}$$

wherein $n^{(i)}$ is the sample size of conditional loss values simulated at $\theta^{(i)}$ and $d_{i,j}$ is the Mahalanobis distance between $\theta^{(i)}$ and $\theta^{(j)}$.

The expected value in Eq. (14) is computed in terms of the standard normal PDF $\phi(\cdot)$ and CDF $\Phi(\cdot)$ (Huang et al., 2006):

$$\mathbb{E}\left[\max\left(Z_t^{(i)} - z_t^*, 0\right)\right] = \left(\mu_{Z_t}^{(i)} - z_t^*\right)\Phi\left(\frac{\mu_{Z_t}^{(i)} - z_t^*}{\sigma_{Z_t}^{(i)}}\right) + \sigma_{Z_t}^{(i)}\phi\left(\frac{\mu_{Z_t}^{(i)} - z_t^*}{\sigma_{Z_t}^{(i)}}\right) \tag{17}$$

190 For each return period t, we perform n_l loss evaluations at the n_s scenarios with the largest AEI. Taking into account the $n_s \times n_l \times n_t$

new model evaluations, we update the KDEs, the density observations $z_t^{(i)}$ and the bootstrap standard deviations $\sigma_{Z_t}^{(i)}$. At scenarios where more than n_2 loss evaluations have been computed, we deviate from Eq. (9) and evaluate the KDE with all these evaluations (instead of only n_2 evaluations).

195 The AL steps are repeated until convergence is achieved or the maximum number of AL iterations n_3 is exceeded. Convergence is achieved when the AEI of all scenarios is below a threshold ϵ for at least n_d consecutive AL iterations, which prevents premature stopping. A suggested value for n_d is $d+1$, wherein d is the dimensionality of the source parameter random vector Θ (Huang et al., 2006). We also choose $n_3 = 1000$ for encouraging the AL procedure to stop by convergence. The threshold ϵ is chosen as (Huang et al., 2006):

200
$$\epsilon = r \times \left(\max_{i=1,\cdots,n} \hat{f}_{L|\Theta}^0(\hat{l}_t|\theta^{(i)}) f_\Theta(\theta^{(i)}) - \min_{i=1,\cdots,n} \hat{f}_{L|\Theta}^0(\hat{l}_t|\theta^{(i)}) f_\Theta(\theta^{(i)}) \right) \tag{18}$$

$\hat{f}_{L|\Theta}^0(\hat{l}_t|\theta^{(i)})$ is the initial KDE, which is computed before the AL.

4 Illustrative examples

In this section, we present two simple examples to illustrate the methodology. The first one is a one-dimensional example, where we focus on the performance of AL and the approximations of the objective function obtained with the noisy KDE 205 estimations and GPR. The second one is a two-dimensional example, where we show the variability of the scenario selection. In both examples, the exact solution is known.

4.1 Single building subjected to a single seismic source with variable magnitude

Figure 3 illustrates the performance of the acquisition function on an one-dimensional example. In the example, the only source parameter is the magnitude M_w, which is beta distributed distributed with shape parameters $\alpha = 1$, $\beta = 3$, and scaled between 210 0 and 10. The conditional distribution of the logarithm of the losses given $M_w = m_w$ is a normal random variable with mean $\mu_{\ln L}(m_w) = -\frac{1}{2}\sin(\frac{5}{2}(m_w - 5)) + \exp(\frac{m_w - 5}{7}) + 7$ and standard deviation $\sigma_{\ln L} = 0.7$. We set $l_t = 10$ for this example.

We generate a seismic catalog of size $n = 100$. The KDEs are computed with Gaussian kernel and $n_2 = 70$ loss samples, or more if the scenario has more than n_2 conditional loss evaluations, and the bootstrap standard deviation is computed based on $n_b = 100$ samples. We perform the GPR on the whole catalog and learn the hyperparameters at every AL step, since the catalog 215 size in this example is not restrictive. For the AL stage, we select $n_s = 5$ scenarios per AL iteration for computing $n_l = 10$ loss evaluations at each scenario (i.e., $n_s \times n_l = 50$ damage evaluations per AL iteration). The acquisition function is the AEI, as introduced in Eq. (14). We also let the algorithm to achieve convergence with a maximum of $n_3 = 1000$ AL iterations, with the convergence criterion in Eq. (18) and $r = 0.001$.

Figure 3 compares intermediate and final results for this example to the true results. After the initial loss evaluations at the 220 100 scenarios, the estimate of the objective function is poor. However, the acquisition function is able to select scenarios near

the true solution. In the final step, one can observe that estimates of the objective function values have high noise, but the GPR is effective in reducing this noise. The resulting estimate of the objective function is close to the true value around the optimum.
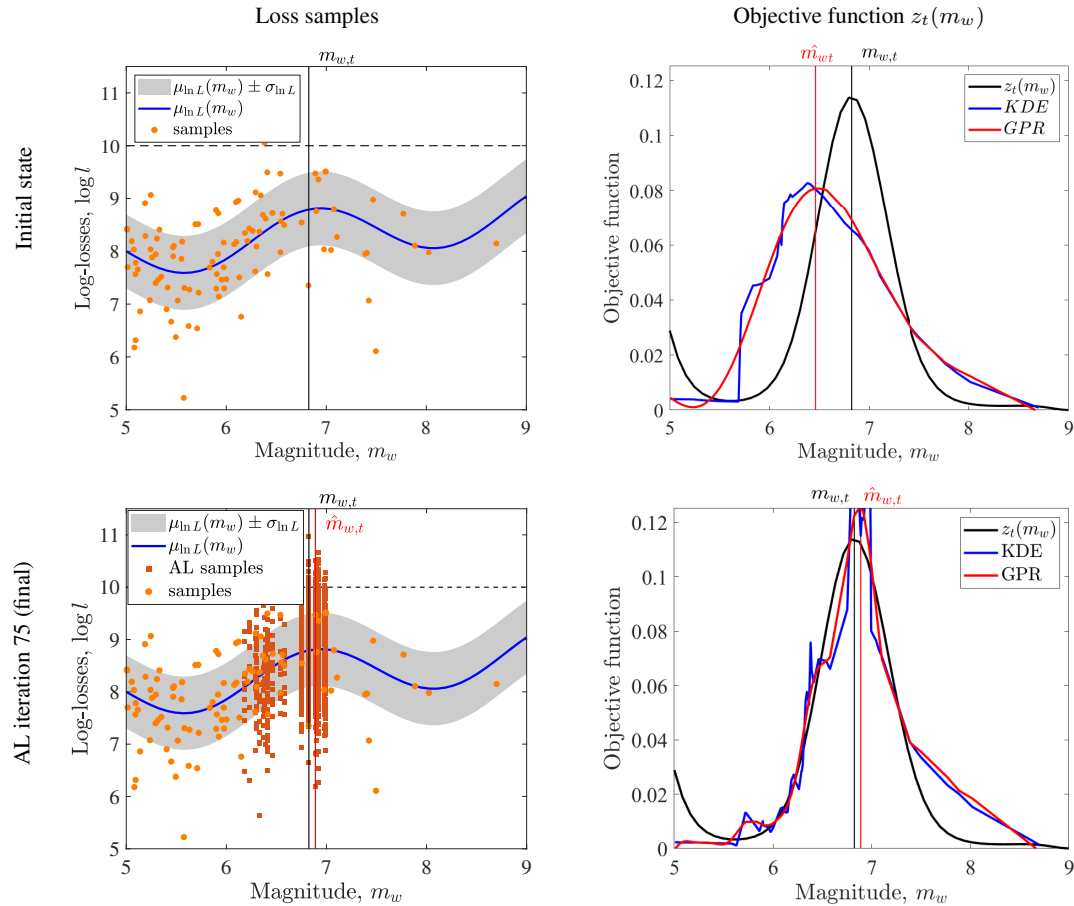
Figure 3. Illustration of AL for maximizing the objective function $z_t(m_w)$ with fixed value l_t, marked with dashed line in the plots in the left panel. The solution $m_{w,t}$ is approximated with the sample point $\hat{m}_{wt}$. The loss samples before and during the AL steps are shown on the left panel. The approximations of the objective function, either with KDE or GPR, are shown in the right panel.

4.2 Single building subjected to an earthquake with unknown magnitude and location

This simple example is adapted from Rosero-Velásquez and Straub (2022). It considers a hypothetical fault, where strong
225   earthquakes occur with a rate of $\lambda_H = 0.3\,\mathrm{yr}^{-1}$. We consider the damages that earthquakes cause to a single building. The source parameters $\Theta = [M_w, \ln R]^{\mathsf{T}}$ are the magnitude M_w and the hypocentral distance R from the earthquake source to the building. The source model for Θ, $f_\Theta(\theta)$, is a normal distribution with mean vector μ_Θ and covariance matrix Σ_Θ given as

follows:

$$\mu_\Theta = \begin{bmatrix} 7.00 \\ 4.38 \end{bmatrix}, \quad \Sigma_\Theta = \begin{bmatrix} 0.36 & -0.08 \\ -0.08 & 0.49 \end{bmatrix}$$

230 A standard deviation σ represents the uncertainty in the ground motion, damage measure, and losses. The losses L are a log-normal random variable with parameters $\mu_{\ln L} = -3.16$, $\sigma_{\ln L} = \sqrt{2.46 + \sigma^2}$. With these choices, the conditional density $f_{\Theta|L}(\theta|l_t)$ can be evaluated analytically. It is a normal distribution, whose mean vector is the representative earthquake scenario for a return period t.

We set $\sigma = 0.5$, and use return periods of 50, 100, 500 and 1000 years. The resulting exact representative earthquake sce-
235 narios are: $\theta_{50} = [7.42, 3.42]^\intercal$, $\theta_{100} = [7.51, 3.21]^\intercal$, $\theta_{500} = [7.69, 2.81]^\intercal$, and $\theta_{1000} = [7.75, 2.65]^\intercal$. We use them to verify the proposed sampling-based algorithm.

To estimate θ_t with the proposed methodology, a stochastic seismic catalog with $n = 2 \times 10^4$ random scenarios is employed. We simulate the losses once at each scenario, approximate l_t, and compute the KDEs at each scenario. The KDEs are based on $n_2 = 200$ loss values, computed with Gaussian kernel, and the bootstrap variance with $n_b = 100$ repetitions. We perform GPR
240 with a training set of size $n_{train} = 1500$, which is constructed as described in Section 3.2.

The maximum number of AL iterations is $n_3 = 1000$, where at every step the losses are evaluated at $n_s = 2$ scenarios $n_l = 10$ times, for each return period. The procedure stops after the maximum AEI is below ϵ, with $r = 0.001$, for at least $n_d = 5$ consecutive AL iterations. For analyzing the uncertainty in the estimation of θ_t, we repeat the experiment 20 times. For the $50, 100$ and 500-year representative scenarios, all experiments converged in less than 10 AL iterations, whereas for the
245 1000-year return period at most 30 AL iterations were required.

Figure 4 shows the spread of the resulting representative hazard scenarios for each return period, which is mainly caused by the numerical approximation of the objective function with limited number of samples.

5 Case study: Valparaíso and Viña del Mar communes

5.1 Context of the study area

250 We apply the proposed methodology to determine representative earthquake scenarios for the communes of Valparaíso and Viña del Mar, which are located in the Valparaíso Region of Chile on the Pacific coast. The study area is the second-largest Chilean urban centre; based on the latest Chilean Census (INE, 2017), it is home to 630 903 inhabitants. It hosts the Port of Valparaíso, which is an important container and the main passenger port in Chile. The area shows a heterogeneous building inventory, ranging from apartment buildings to informal settlements, and a historic district declared a World Heritage Site by
255 UNESCO in 2003 (Indirli et al., 2011; Jiménez et al., 2018).

The National Electric System (SEN) provides the area's power supply. The SEN is the largest Chilean transmission grid, and covers most of the national territory. The SEN connects power plants and substations with the consumer areas through
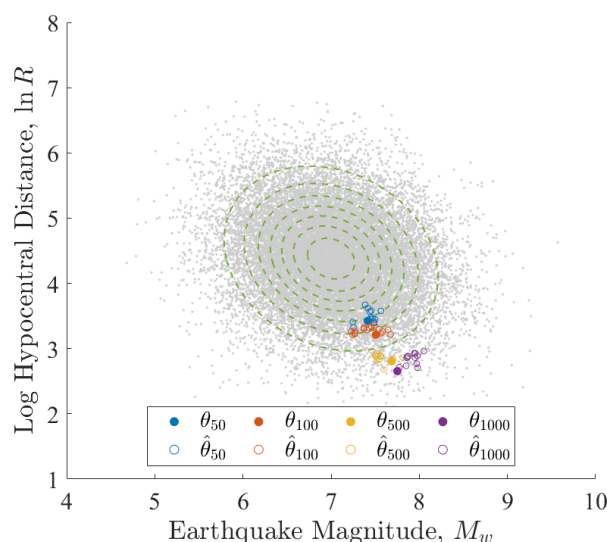
Figure 4. Numerical approximation of the representative earthquake scenarios $\hat{\theta}_t$ of four different return periods $t = 50, 100, 500, 1000$ years, based on a Monte Carlo sample of scenarios. Each return period is represented by a different color. For each return period, the 20 approximations $\hat{\theta}_t$, corresponding to 20 experiments, are the colored empty circles, and the corresponding exact solutions are depicted by filled circles. The grey points are the scenarios of the catalog, and the dashed contours represent the PDF of the source parameters.

high-voltage lines. The topology of the SEN is characterized as a single-scale network with a fast decaying tail, and most of the load substations are close to a generation unit, with a median distance of 9 km (Ferrario et al., 2022).

260   Powerful earthquakes have hit the area in the past, such as the 1730 earthquake, with inferred magnitude M_w in the range 9.1—9.3, and the 1906 event, with inferred moment magnitude M_w 8.0—8.2 (Carvajal et al., 2017). More recently, the 1985 M_w 8.0 event affected around 230 000 dwellings, 1 million people, and caused losses of about USD 1.4 billion (ONEMI, 1985). The most recent M_w 8.8 Maule earthquake (2010) caused severe structural damage in buildings in Viña del Mar, including in buildings retrofitted in 1985 (Jünemann et al., 2015).

265   A hazard evaluation of the Valparaíso urban area presented in (Indirli et al., 2011) selected representative earthquake scenarios based on historical events, considering the seismicity around the study area, to specify an average regional seismic input and to generate synthetic seismograms. The scenarios are summarized in Table 1; their magnitudes range from $M_w = 6.0$ to $M_w = 8.2$.

5.2   Earthquake model and seismic catalog

270 We employ the earthquake model presented in (Poulos et al., 2019) to generate a catalog of earthquake scenarios. The catalog has 2×10^4 scenarios with magnitude larger than or equal to $M_w = 5.0$, which is the minimum magnitude defined by Poulos et al. (2019) for performing the declustering on the historical seismic catalogs on which the earthquake model is based on. The

Table 1. Representative earthquake scenarios for the urban area of Valparaíso selected by Indirli et al. (2011) from historic events. The epicenter location is reported with a map in (Indirli et al., 2011), hence their numeric values, as well as moment magnitude and depth, are here reproduced from the earthquake records of (USGS, 2023). The 1985 M_w 8.0 earthquake event (lon = $-71.850°$, lat = $-33.240°$, depth = 33km) has similar source parameters as the 1906 event, while the 2010 earthquake event occurred after the study of Indirli et al. (2011) was submitted for publication. The event dates are in local time.

Source param., θ	Event date			
	16/08/1906	28/03/1965	06/07/1979	16/10/1981
Longitude [°]	-72.400	-71.233	-71.321	-73.074
Latitude [°]	-32.400	-32.522	-32.148	-33.134
Depth [km]	35	70	45	33
Magnitude, M_w	8.2	7.4	5.7	7.2

catalog covers the whole country of Chile, consists of scenarios at the subduction interface and subduction intraslab zones. The earthquake model utilizes the slab geometry proposed in (Hayes et al., 2012) for the depth contours and trench geometry, and
275 divides the Chilean subduction zone into three subduction interface and four intra-slab zones, whose combined occurrence rate equals $\lambda_H = 43.3 \text{ yr}^{-1}$.

The epicentral locations of the nation-wide catalog are generated randomly, based on the occurrence rate associated with the seismic zones defined by the occurrence model. The magnitude is sampled with an importance sampling (IS) approach, following the Gutenberg-Richter relations at each seismic zone according to (Poulos et al., 2019). Using IS for the magnitude
280 allows obtaining a catalog with similar proportions of scenarios at all magnitudes. We employ a uniform distribution with minimum and maximum values defined by the magnitude range of each seismic zone, as IS density. The corresponding IS weights are considered when determining the loss-exceedance function. The resulting catalog is depicted in Figure 5a.

The earthquake model of (Poulos et al., 2019) only considers the subduction zone, hence the independent source parameters are the moment magnitude M_w, longitude X and latitude Y of the epicenter. Other parameters, such as the depth H, strike,
285 dip and rake angles, are determined by the geometry derived from (Hayes et al., 2012), depending of the epicenter location. Therefore, The PDF of the source parameters $f_\Theta(\theta)$ is represented, respectively, by the conditional PDF $f_{M_w|X,Y}(m_w|x,y)$, and the location-dependent occurrence rate $\lambda(x,y)$:

$$f_\Theta(\theta) \propto \lambda(x,y) f_{M_w|X,Y}(m_w|x,y) \tag{19}$$

For verification purposes, we also employ a seismic catalog generated from the occurrence model developed in the SARA
290 Project (Pagani et al., 2021), simulating an investigation time of 10^5 years, and disaggregated with the event-based PSHA calculator of OpenQuake (Pagani et al., 2014), based on the exceedance function of the spectral acceleration at a representative geographic location of the study area. The catalog is part of an application example of the web service Quakeledger, provided by the GFZ-German Research Center of Geosciences (Pittore et al., 2021b). The catalog consists of 23 337 scenarios, with magnitude larger or equal than $M_w = 6.0$, which is considered the minimum moment magnitude at which exposed buildings in

295   the study area can exhibit serious damage. We refer to this catalog as the SARA catalog. It is shown in Figure 5b. We employ
the SARA catalog only with the residential building stock, and not the power network, as the catalog only includes earthquakes
close to the study site. They are representative for the building stock, but not the power network, which spans a larger area;
compare also Figures 5 and 6.

Figure 5 shows that with both catalogs the scenarios with the largest magnitudes are located in the subduction interface zone.
300   The SARA catalog has more scenarios closer to the study area than the nation-wide catalog. However, due to the IS scheme,
the nation-wide catalog has more scenarios with large magnitude ($M_w \geq 8.0$) close to the study area.
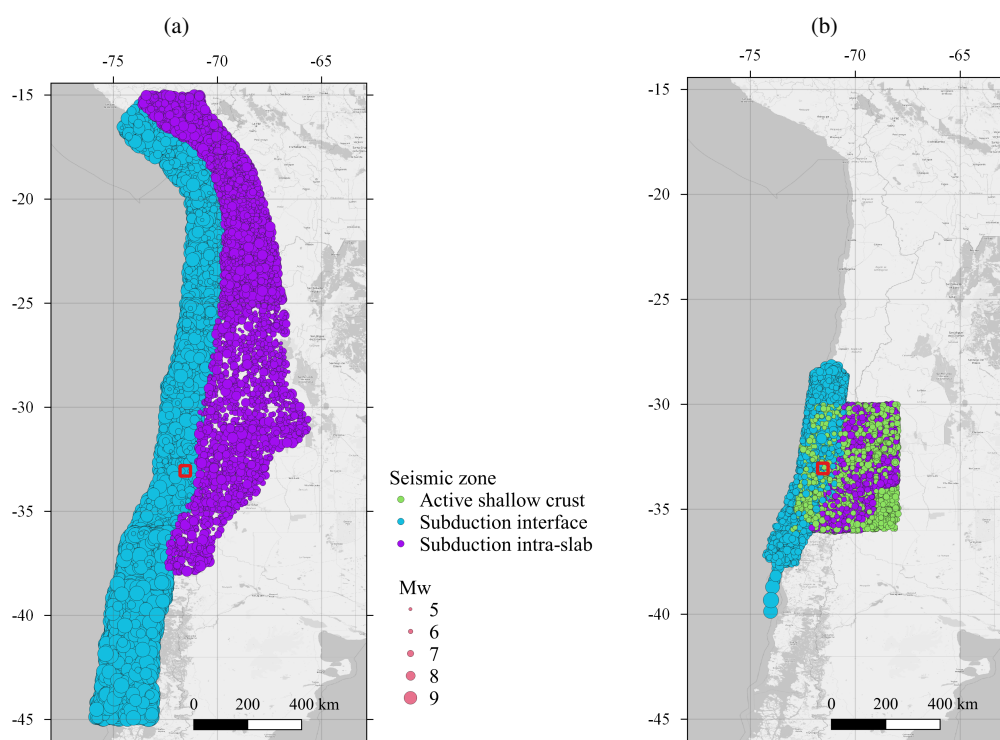
Figure 5. Stochastic catalogs: (a) nation-wide (Poulos et al., 2019), and (b) SARA (Pagani et al., 2021). The circle size corresponds to the
scenario magnitude. The red square contains the study area. Basemap from ©OpenStreetMap contributors 2023. Distributed under the Open
Data Commons Open Database License (ODbL) v1.0.

5.2.1   Ground motion models

For the residential building stock, we evaluate the PGA and spectral accelerations at $0.3\,s$ and $1.0\,s$ with the Ground Motion
Prediction Equation (GMPE) presented in (Montalva et al., 2017). The uncertainty in the median prediction is modelled with
305   a Gaussian random field, with the spatial correlation model of Jayaram and Baker (2009). The choice of these models is based

on the epistemic uncertainty analysis of different ground motion and correlation models by Gómez-Zapata et al. (2022a), who analyzed the same study area. For the scope of this study, we do not consider cross-correlated ground motion fields.

For the Chilean power network, we employ the GMPE of Abrahamson et al. (2016) and the spatial correlation model developed by Goda and Atkinson (2010) for predicting the PGA. This is the same ground motion model as the one utilized by Ferrario et al. (2022).

The functional form of both GMPEs is similar, and therefore, their predictions do not differ significantly, as observed in previous studies (e.g., Hussain et al., 2020; Gómez-Zapata et al., 2022a). In particular, Hussain et al. (2020) found negligible differences in direct loss estimates for the residential building stock of Santiago de Chile after using these two GMPEs to simulate the associated ground motion from subduction earthquake scenarios.

5.3 Model for the building stock in the communes of Valparaíso and Viña del Mar

We employ the Bayesian exposure model of the building stock with the building classes described in (Gómez-Zapata et al., 2022a) and available in (Pittore et al., 2021a). The model was constructed by taking the OpenStreetMap footprint of the buildings in the two communes, and assigning to each footprint the most likely building class. The buildings are counted within a regular 500×500 m resolution grid in the urban areas, as shown in Fig. 6. Detailed building counts for each class are presented in (Gómez-Zapata et al., 2022a)

The model considers 16 building classes (Gómez-Zapata et al., 2022a), which correspond to the building classes proposed in the SARA project (Yepes-Estrada et al., 2017), and have an associated replacement cost. Furthermore, each building class has an associated fragility model with five damage states (Villar-Vega et al., 2017). The fragility model for each building associates an intensity measure (spectral acceleration at $0.3\ s$, $1.0\ s$, or the PGA) with the probabilities of achieving a damage state. We assume the following relative replacement cost for each damage level: 0% for no damage, 2% for slight damage, 10% for moderate damage, 50% for extensive, and 100% for complete damage.

We utilize the model to evaluate the ground motion and simulate the building damage. Given an intensity level of the ground motion, the damage is simulated randomly at each building with a discrete distribution with probabilities defined by the fragility functions. The losses for each scenario in the catalog are computed as the accumulated reconstruction cost of the damaged residential buildings, based on the simulated damage.

5.4 Model for the Chilean National Electric System, SEN

We model the SEN and its components, following (Ferrario et al., 2022). The network model consists of 1494 nodes, representing 500 generation units and 994 substations, and the transmission lines connecting them, with a total power generation capacity of 21.9 GW. The model considers seismic interaction and system performance subjected to component failures. The interaction with the earthquake scenarios is modeled with fragility functions and recovery times at the nodes, according to FEMA (2003). Given a scenario with a ground motion field (in this case, the PGA), each node is randomly associated with a damage state, by means of the fragility function, and a recovery time associated with their damage state.
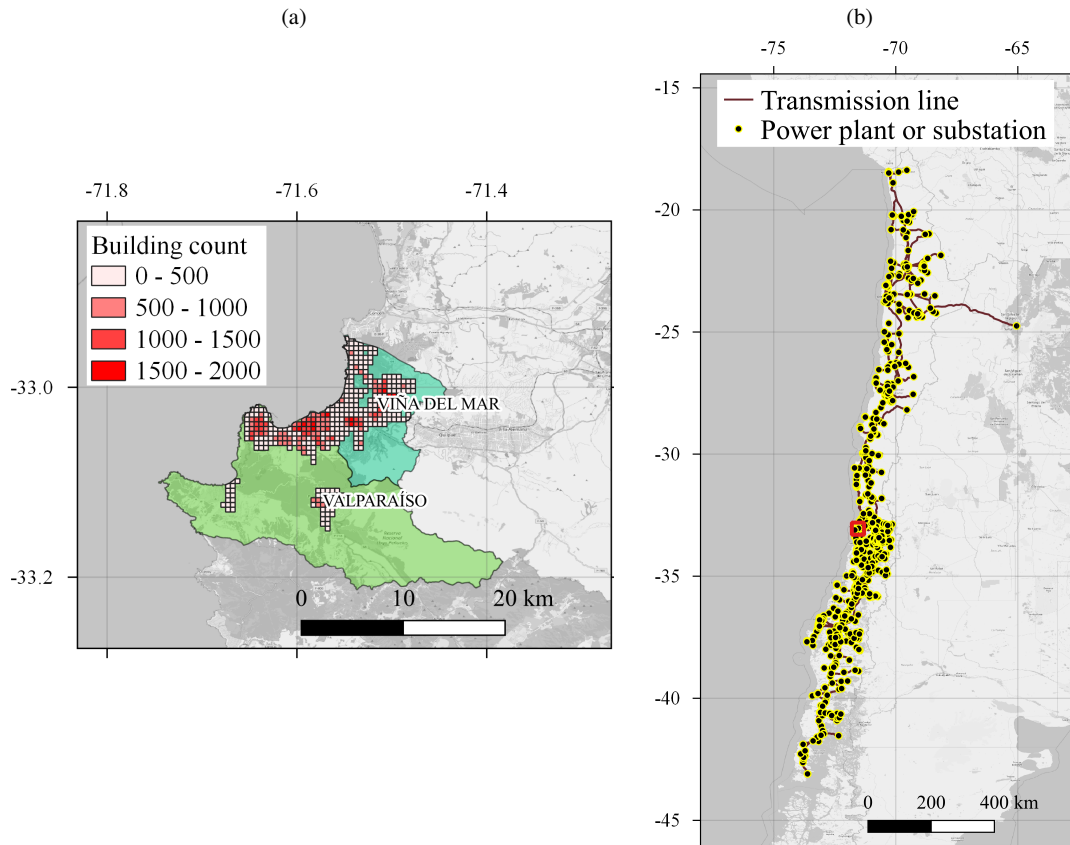
Figure 6. (a) Exposure model of the residential building stock. Each rectangular cell shows the total count of residential buildings, indicating the most dense areas. Source: Pittore et al. (2021a). (b) Geographic location of the SEN. The network follows the narrow shape of the country, and the communes of Valparaíso and Viña del Mar (inside red square) are in a central location within the network. Source: Coordinador Eléctrico Nacional (2019). Basemap from ©OpenStreetMap contributors 2023. Distributed under the Open Data Commons Open Database License (ODbL) v1.0.

The losses associated with the SEN due to an earthquake scenario are quantified in terms of the energy not supplied (ENS). The ENS is evaluated at each substation by solving the power in normal steady state operation through the Direct Current

340 - Optimal Power Flow (DCOPF) model (Wood et al., 2013), and comparing it with the power in a damaged state operation caused by the earthquake scenario. To quantify the loss in the power supply in the communes of Valparaíso and Viña del Mar, we calculate the total ENS with the sum of the ENS of all substations located in the two communes (14 in total).

DCOPF is typically adopted in practice for transmission networks (Frank and Rebennack, 2016). It optimizes the power generation cost, taking into account the capacity of the power plants and transmission lines connected to the power grid, the

345 generation cost associated with each power plant, and the demand from the clients. For modeling the system response to an earthquake, the DCOPF considers the reduced capacity of components affected by the earthquake. For the SEN, the capacities

are known from the specifications of its components (Coordinador Eléctrico Nacional, 2019; Comisión Nacional de Energía, 2019), and the demand is modelled based on the database of Coordinador Eléctrico Nacional (2018). A detailed description and validation of the network model of the SEN can be found in (Ferrario et al., 2022).

350 ## 6 Evaluation of representative earthquake scenarios for Valparaíso and Viña del Mar communes

6.1 Results for the residential building stock

Figure 7 shows the annual exceedance rate of the losses, computed with the two seismic catalogs. The loss estimations are similar from 10^8 USD on, corresponding to events with 50-year return period and beyond. For more frequent events with smaller losses, the estimations differ.

355 USGS (2023) records that between 1960 and 2020 there were 12 seismic events that produced a macroseismic intensity greater or equal than VI on the Mercalli scale in the two communes. This corresponds to an occurrence rate of 0.20 yr^{-1}. It is reasonable to assume that events with macroseismic intensity of VI or higher lead to losses of 10^6 USD and higher. This data therefore seems to confirm the results obtained with the nation-wide catalog. Overall, we conclude that the nation-wide catalog seems to result in reasonable estimates of the losses.
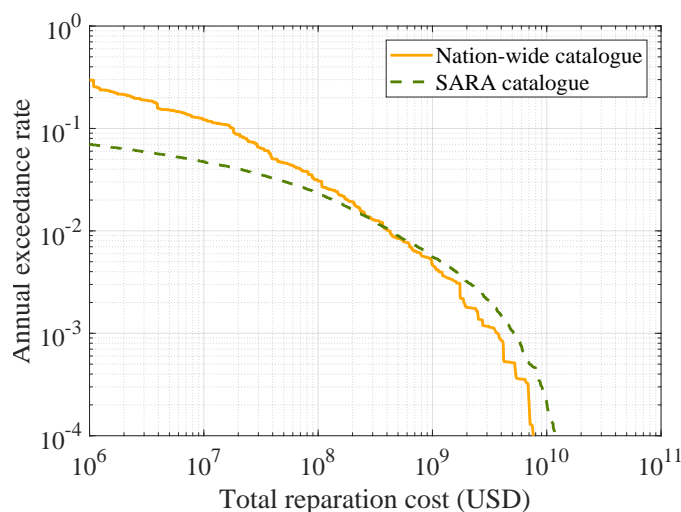
Figure 7. Loss-exceedance function of the reconstruction costs associated with the building portfolio in Valparaíso y Viña del Mar communes, computed with two seismic catalogs. On the range between 50 and 1000 year return period, the two curves show similar values.

360 As in Section 4.2, we evaluate the representative scenarios in 20 independent runs of the algorithm, to check the robustness of the results. In all evaluations, we found a limited spread of the identified representative scenarios, similar to that of Figure 4. This spread is larger for higher return periods, but is deemed acceptable. In the following, we only present the modes, i.e., the representative scenarios that were identified the most frequently in the 20 repetitions.

Figure 8 shows the representative earthquake scenarios for the analyzed return periods. One can observe that large return periods are associated with scenarios that have larger magnitude. The fact that the magnitude for the 1000-year scenario equals only $M_w = 7.01$ is a consequence of the size of the study area. On the one hand, among the 36 seismic events with $M_w \geq 8.0$ registered along the Chilean coast between 1570 and 2023 (CSN, 2023), only 4 events had epicenter near to the two communes, i.e., within a radius of approximately 70km. For comparison, the identified 1000yr scenario is at a distance of around 20km from Valparaiso and Viña del Mar. On the other hand, the spatial correlation of the ground motion within a small study area leads to an increased likelihood of extreme losses in a scenario with lower earthquake magnitude and extreme ground motion residuals. This tendency was also found by Goda and Hong (2009).
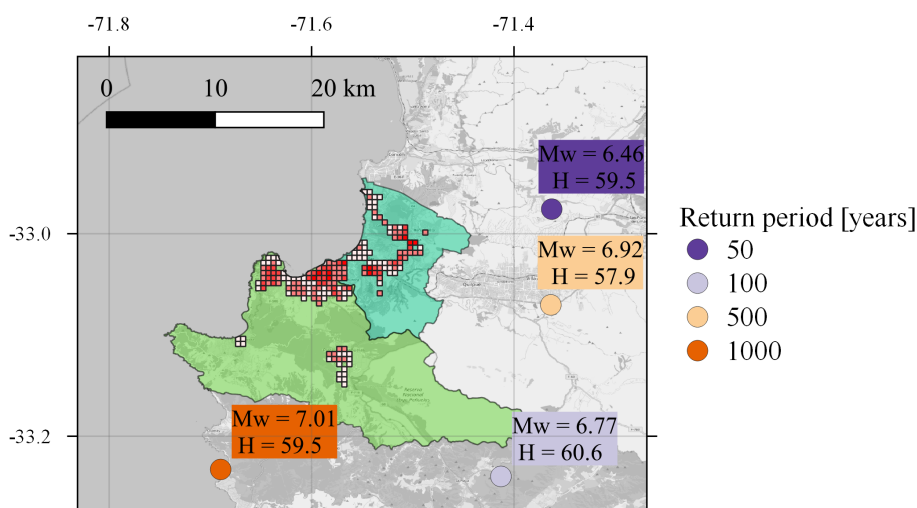
Figure 8. Representative earthquake scenarios for the residential building portfolio in Valparaíso and Viña del Mar communes. The hypocentral depth H is displayed in km. Source of the exposure morel: Pittore et al. (2021a). Basemap from ©OpenStreetMap contributors 2023. Distributed under the Open Data Commons Open Database License (ODbL) v1.0.

6.2   Results for the power network

Figure 9 shows the loss exceedance function in terms of the ENS obtained with the nation-wide catalog. The largest sampled ENS value is around 2×10^5MWh, which is around 20% of the annual energy demand of the two communes.

Figure 10 shows the resulting representative earthquake scenarios for the analyzed return periods. One can observe that the scenarios are close to the two communes but less concentrated than those of the residential building stock, and have a different magnitude range. This reflects the fact that the total ENS, although computed only at the substations located within the two communes, depends on the damage state of the components of the rest of network.

Even though the power supply network is spread out over a larger area, the representative earthquake scenarios are close to the study area. They reflect that the most important components of the network for the two communes are in their proximity.
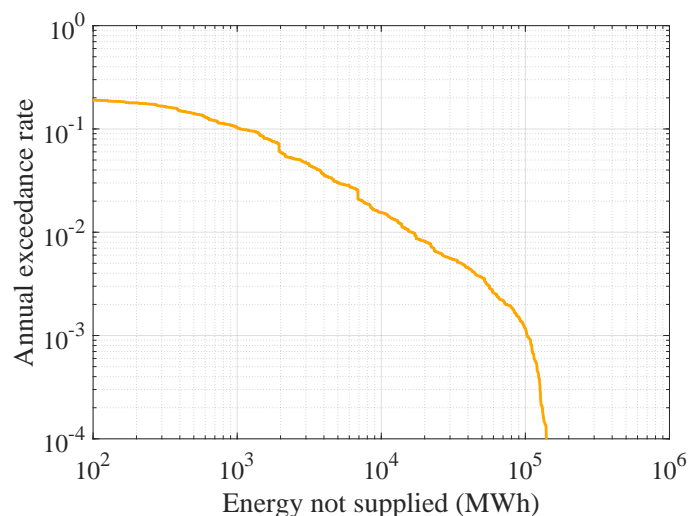
Figure 9. Loss-exceedance function of the total energy not supplied in communes of Valparaíso and Viña del Mar.
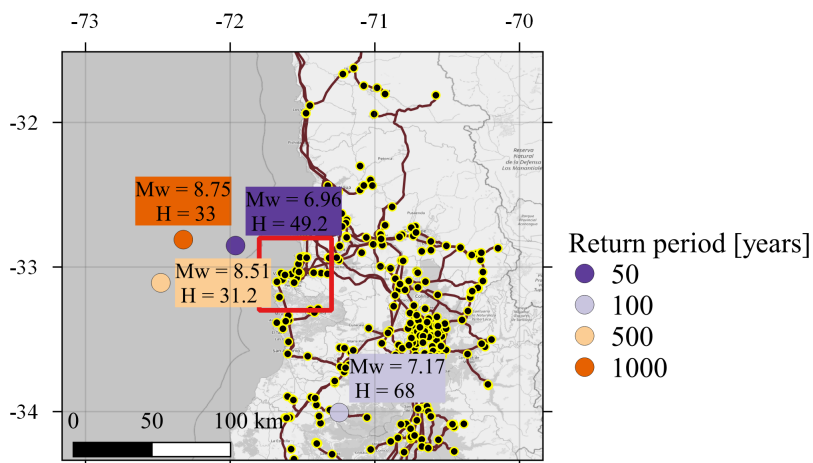
Figure 10. Representative earthquake scenarios for the power supply considering the total ENS of Valparaíso and Viña del Mar communes. The red square indicates the location of the two communes, and the hypocentral depth H is displayed in km. Source of the power network: Coordinador Eléctrico Nacional (2019). Basemap from ©OpenStreetMap contributors 2023. Distributed under the Open Data Commons Open Database License (ODbL) v1.0.

For example, the source location of the 100-year return period scenario lies near a main connection between the substations in the two communes and the rest of the SEN.
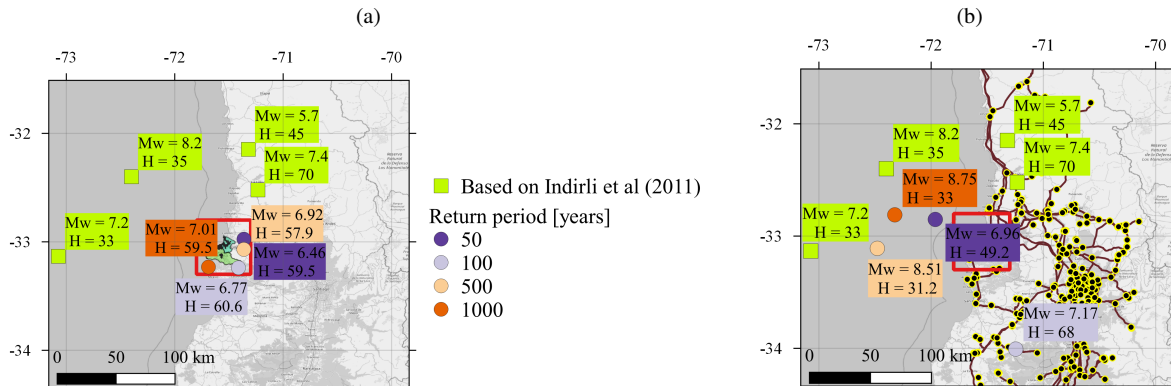
Figure 11. Representative earthquake scenarios for the (a) building stock and (b) power supply, compared with past earthquake events selected in (Indirli et al., 2011). The location of the two communes is within the red square, and the SEN network is also displayed. Source of the exposure model: Pittore et al. (2021a). Source of the power network: Coordinador Eléctrico Nacional (2019). Basemap from ©OpenStreetMap contributors 2023. Distributed under the Open Data Commons Open Database License (ODbL) v1.0.

6.3 Comparison with past earthquake events

Figure 11 compares the results with the historical events selected in (Indirli et al., 2011). Although the representative earthquake scenarios and the selected historical events target the same area of interest, they have different purposes. The historical events presented by Indirli et al. (2011) aim at representing the seismicity of the most important seismic zones affecting the study area. In contrast, the representative earthquake scenarios, as defined in (Rosero-Velásquez and Straub, 2022), take into account the performance and the losses caused by damage and failures in the analyzed engineering system. In addition, the scenarios are selected based on different loss levels, which are attached to return periods.

7 Discussion

The representative earthquake scenarios summarized in Fig. 11 can provide important input to risk assessment and risk management tasks. The fact that the scenarios identified with the proposed approach differ widely from the historic events should not be surprising, as the latter are in some sense just "random samples" of earthquake events. Nevertheless, the historic events can provide a useful validation of the identified scenarios. In this regard, the scenarios identified as representative for the power supply network appear to be in line with the historic events, as they cover a similar range of magnitudes. The identified 500 and 1000-year scenarios are larger than the observed events, which would also be expected. By contrast, the representative scenarios identified for the building stock have significantly smaller magnitudes. According to the employed model, the extreme losses are more likely to occur by a combination of a less strong earthquake with larger-than-average ground motions (i.e., a large value of the inter-event term in the GMPE). This effect occurs for the residential building stock, due to its spatial concentration, and not (or to a much smaller extend) for the power supply network, which is spatially distributed.

The above observations lead to some important conclusions: Firstly, the scenarios, rightly, are different for different assets. Secondly, the scenarios depend on model assumptions beyond the seismic source models. In the application here, the model of the ground motion variability has a distinct effect on the scenarios for the residential building stock. Given that the employed state-of-the-art model might overestimate the event-to-event variability (Bodenmann et al., 2023), the results for the building
405 portfolio should be utilized carefully. Thirdly, because the earthquake scenario in the building case is representative for certain loss return periods only in combination with high ground motions, it should be investigated if and how the representative scenario should also provide ground motion fields together with earthquake event.

Overall, for practical risk management tasks it is recommended to use the historic events jointly with the identified scenarios, in particular for the residential building stock case.

410 The presented results do not consider the potential impact caused by further earthquake-triggered hazards, such as tsunamis and landslides. The tsunami impact during offshore earthquake scenarios with large magnitude should be considered in a complete loss estimation, and may affect the scenario selection associated with large return periods. Similarly, the scenario selection may change if one considers the landslides potential in the study area.

The dependence of the results to the engineering system and the model assumptions implies that the representative scenarios
415 should be regularly updated, depending on how much the analyzed system changes and the models improve. According to demographic projections of the INE, based on censal data (INE, 2017), the population in the two communes by 2035 will increase by around 13% with respect to the population in 2017. This demographical change will likely cause changes in the residential building stock and the power demand. Depending on how these changes develop, the representative earthquake scenarios may change as well.

420 Cascading effects have not been considered in the power network model. Although the topology of the SEN and the redundancy of generators along the network reduce the probability of large blackouts, hub nodes far from the two communes may affect them through a sequence of failures triggered by earthquake scenarios impacting them. This may shift the representative scenarios to further locations from the study area.

Further applications of this methodology include catalogs with multiple hazards (e.g., seismic scenarios with tsunami),
425 loss calculations considering indirect consequences, and high-dimensional scenarios (e.g., including the damage states of the individual components, either buildings or power network components). For the later, however, the dimensionality of the damage states has to be reduced (Rosero-Velásquez and Straub, 2019).

8 Conclusion

We present a methodology and algorithm to determine representative earthquake scenarios from seismic catalogs and apply
430 it to theValparaíso region of Chile. Because the identified scenarios should be representative for extreme losses, they differ depending on the exposed assets. In this contribution, we consider the building portfolio and the electrical supply network. The application shows that the the methodology can work and allows to identify scenarios more systematically than by selection or extrapolation from past events. However, the results for the building portfolio also show that resulting scenarios cannot

be considered independently of the resulting ground motions. Therefore, future work should investigate scenarios that also
435 include the ground motions. Because the description of ground motion fields requires a large number of parameters, the existing
methodology will need to be extended to be able to cope with such scenarios.

Code and data availability. The code for scenario selection is available from the authors upon reasonable request.

Author contributions. This paper was conceptualized by HRV and DS. The methodology was developed by HRV and DS. The investigation
for the hazard modeling (i.e., source model, seismic catalog) was conducted by AP, JCG, JCL, HRV and DS. The investigation for the
440 power supply application (i.e., network model and numerical simulations) was conducted by MM, EF, JCL and HRV. The investigation
for the residential building stock (i.e., exposure and vulnerability model and numerical simulations) was conducted by JCG and HRV. The
visualization was done by HRV, and the interpretation by HRV and DS. The original draft preparation was done by HRV, MM and DS.
Review and editing were done by HRV, AP, JCG, JCL, MM, EF and DS. Funding acquisition for this work was done by DS and JCL.

Competing interests. The contact author has declared that none of the authors has any competing interests.

445 *Acknowledgements.* This work has been sponsored by the research and development project RIESGOS 2.0 (Grant No. 03G0905A-J), funded
by the German Federal Ministry of Education and Research (BMBF) as part of the funding programme "BMBF CLIENT II - International
Partnerships for Sustainable Innovations", and by the Chilean government through the Research Center for Integrated Disaster Risk Man-
agement (CIGIDEN), ANID/FONDAP/1522A0005, the research project Multiscale earthquake risk mitigation of healthcare networks using
seismic isolation, ANID/FONDECYT/1220292. We also thank Prof. Fabrice Cotton (GFZ) for his support during the elaboration of this
450 study.

References

Abrahamson, N., Gregor, N., and Addo, K.: BC Hydro Ground Motion Prediction Equations for Subduction Earthquakes, Earthquake Spectra, 32, 23–44, https://doi.org/10.1193/051712EQS188MR, 2016.

Aguirre, P., Vásquez, J., de la Llera, J. C., González, J., and González, G.: Earthquake damage assessment for deterministic scenarios in Iquique, Chile, Nat. Hazards, 92, 1433–1461, https://doi.org/10.1007/s11069-018-3258-3, 2018.

Allen, E., Chamorro, A., Poulos, A., Castro, S., de la Llera, J. C., and Echaveguren, T.: Sensitivity analysis and uncertainty quantification of a seismic risk model for road networks, Comput. -Aided Civ. Infrastruct. Eng., 37, 516–530, https://doi.org/https://doi.org/10.1111/mice.12748, 2022.

Bazzurro, P. and Cornell, C. A.: Disaggregation of seismic hazard, Bulletin of the Seismological Society of America, 89, 501–520, https://doi.org/10.1785/BSSA0890020501, 1999.

Bodenmann, L., Baker, J. W., and Stojadinović, B.: Accounting for path and site effects in spatial ground-motion correlation models using Bayesian inference, Natural Hazards and Earth System Science, 23, 2387–2402, https://doi.org/10.5194/nhess-23-2387-2023, 2023.

Candia, G., Poulos, A., de la Llera, J. C., Crempien, J. G., and Macedo, J.: Correlations of spectral accelerations in the Chilean subduction zone, Earthquake Spectra, 36, 788–805, https://doi.org/10.1177/8755293019891723, 2020.

Carvajal, M., Cisternas, M., and Catalán, P. A.: Source of the 1730 Chilean earthquake from historical records: Implications for the future tsunami hazard on the coast of Metropolitan Chile, Journal of Geophysical Research: Solid Earth, 122, 3648–3660, https://doi.org/https://doi.org/10.1002/2017JB014063, 2017.

Chatelain, J.-L., Yepes, H., Bustamante, G., Fernández, J., Valverde, J., Kaneko, F., Villacis, C., Yamada, T., and Tucker, B.: Proyecto para manejo del riesgo sísmico de Quito, Escuela Politécnica Nacional, Quito, Ecuador, 1995.

Comisión Nacional de Energía: Energía Maps, https://energiamaps.cne.cl/, 2019.

Coordinador Eléctrico Nacional: Database provided by email, 2018.

Coordinador Eléctrico Nacional: Infotécnica Sistema Eléctrico Nacional, https://infotecnica.coordinador.cl/, 2019.

Cornell, C. A.: Engineering seismic risk analysis, Bulletin of the Seismological Society of America, 58, 1583–1606, 1968.

CSN: Grandes terremotos en Chile, https://www.sismologia.cl/informacion/grandes-terremotos.html, 2023.

de la Llera, J. C., Rivera, F., Mitrani-Reiser, J., Jünemann, R., Fortuño, C., Ríos, M., Hube, M., Santa María, H., and Cienfuegos, R.: Data collection after the 2010 Maule earthquake in Chile, Bulletin of Earthquake Engineering, 15, 555–588, https://doi.org/10.1007/s10518-016-9918-3, 2017.

Efron, B. and Tibshirani, R. J.: An Introduction to the Bootstrap, no. 57 in Monographs on Statistics and Applied Probability, Chapman & Hall/CRC, Boca Raton, Florida, USA, 1993.

FEMA: HAZUS MH MR4 multi-hazard loss estimation methodology - Earthquake model, technical manual, 2003.

Ferrario, E., Poulos, A., Castro, S., de la Llera, J., and Lorca, A.: Predictive capacity of topological measures in evaluating seismic risk and resilience of electric power networks, Reliability Engineering & System Safety, 217, 108 040, https://doi.org/https://doi.org/10.1016/j.ress.2021.108040, 2022.

Fox, M. J., Stafford, P. J., and Sullivan, T. J.: Seismic hazard disaggregation in performance-based earthquake engineering: occurrence or exceedance?, Earthquake Engineering & Structural Dynamics, 45, 835–842, https://doi.org/https://doi.org/10.1002/eqe.2675, 2016.

Frank, S. and Rebennack, S.: An introduction to optimal power flow: Theory, formulation, and examples, IIE Transactions, 48, 1172–1197, https://doi.org/10.1080/0740817X.2016.1189626, 2016.

Gisbert, F. J. G.: Weighted samples, kernel density estimators and convergence, Empirical Economics, 28, 335–351, https://doi.org/10.1007/s001810200134, 2003.

490 Goda, K. and Atkinson, G. M.: Intraevent Spatial Correlation of Ground-Motion Parameters Using SK-net Data, Bulletin of the Seismological Society of America, 100, 3055–3067, https://doi.org/10.1785/0120100031, 2010.

Goda, K. and Hong, H. P.: Deaggregation of seismic loss of spatially distributed buildings, Bulletin of Earthquake Engineering, 7, 255–272, https://doi.org/10.1007/s10518-008-9093-2, 2009.

Gómez-Zapata, J. C., Zafrir, R., Pittore, M., and Merino, Y.: Towards a sensitivity analysis in seismic risk with probabilistic building exposure

495 models: An application in Valparaíso, Chile, using ancillary open-source data and parametric ground motions, ISPRS Int. J. Geo-Inf., 11, https://doi.org/10.3390/ijgi11020113, 2022a.

Gómez-Zapata, J. C., Pittore, M., Cotton, F., Lilienkamp, H., Shinde, S., Aguirre, P., and Santa María, H.: Epistemic uncertainty of probabilistic building exposure compositions in scenario-based earthquake loss models, Bulletin of Earthquake Engineering, https://doi.org/10.1007/s10518-021-01312-9, 2022b.

500 Hayes, G. P., Wald, D. J., and Johnson, R. L.: Slab1.0: A three-dimensional model of global subduction zone geometries, Journal of Geophysical Research: Solid Earth, 117, https://doi.org/https://doi.org/10.1029/2011JB008524, 2012.

Huang, D., Allen, T. T., Notz, W. I., and Miller, R. A.: Sequential Kriging optimization using multiple-fidelity evaluations, Struct. Multidiscip. Optim., 32, 369 – 382, https://doi.org/10.1007/s00158-005-0587-0, 2006.

Hussain, E., Elliott, J. R., Silva, V., Vilar-Vega, M., and Kane, D.: Contrasting seismic risk for Santiago, Chile, from near-field and distant

505 earthquake sources, Natural Hazards and Earth System Sciences, 20, 1533–1555, https://doi.org/10.5194/nhess-20-1533-2020, 2020.

Indirli, M., Razafindrakoto, H., Romanelli, F., Puglisi, C., Lanzoni, L., Milani, E., Munari, M., and Apablaza, S.: Hazard Evaluation in Valparaíso: the MAR VASTO Project, Pure and Applied Geophysics, 168, 543–582, https://doi.org/10.1007/s00024-010-0164-3, 2011.

INE: Servicio de mapas del Censo 2017, Instituto Nacional de Estadísticas (INE). Chile. http://www.censo2017.cl/servicio-de-mapas, 2017.

Jayaram, N. and Baker, J. W.: Correlation model for spatially distributed ground-motion intensities, Earthq. Eng. Struc. D., 38, 1687–1708,

510 https://doi.org/https://doi.org/10.1002/eqe.922, 2009.

Jayaram, N. and Baker, J. W.: Efficient sampling and data reduction techniques for probabilistic seismic lifeline risk assessment, Earthquake Engineering and Structural Dynamics, 39, 1109–1131, https://doi.org/https://doi.org/10.1002/eqe.988, 2010.

Jiménez Martínez, M., Jiménez Martínez, M., and Romero-Jarén, R.: How resilient is the labour market against natural disaster? Evaluating the effects from the 2010 earthquake in Chile, Natural Hazards, 104, 1481–1533, https://doi.org/10.1007/s11069-020-04229-9, 2020.

515 Jiménez, B., Pelà, L., and Hurtado, M.: Building survey forms for heterogeneous urban areas in seismically hazardous zones. Application to the historical center of Valparaíso, Chile, International Journal of Architectural Heritage, 12, 1076–1111, https://doi.org/10.1080/15583058.2018.1503370, 2018.

Jünemann, R., de la Llera, J., Hube, M., Cifuentes, L., and Kausel, E.: A statistical analysis of reinforced concrete wall buildings damaged during the 2010, Chile earthquake, Engineering Structures, 82, 168–185, https://doi.org/https://doi.org/10.1016/j.engstruct.2014.10.014,

520 2015.

McGuire, R. K.: Probabilistic seismic hazard analysis and design earthquakes: Closing the loop, Bulletin of the Seismological Society of America, 85, 1275–1284, https://doi.org/10.1785/BSSA0850051275, 1995.

Miller, M. and Baker, J.: Ground-motion intensity and damage map selection for probabilistic infrastructure network risk assessment using optimization, Earthquake Engineering and Structural Dynamics, 44, 1139–1156, https://doi.org/10.1002/eqe.2506, 2015.

525   Montalva, G. A., Bastías, N., and Rodriguez-Marek, A.: Ground-Motion Prediction Equation for the Chilean Subduction Zone, Bull. Seismol. Soc. Am., 107, 901–911, https://doi.org/10.1785/0120160221, 2017.

ONEMI: Informe sobre el terremoto de marzo 1985 en Chile, 1985.

Pagani, M., Monelli, D., Weatherill, G., Danciu, L., Crowley, H., Silva, V., Henshaw, P., Butler, L., Nastasi, M., Panzeri, L., et al.: OpenQuake engine: An open hazard (and risk) software for the global earthquake model, Seismological Research Letters, 85, 692–702, 2014.

530   Pagani, M., Johnson, K., and Garcia Pelaez, J.: Modelling subduction sources for probabilistic seismic hazard analysis, in: Characterization of Modern and Historical Seismic–Tsunamic Events, and Their Global–Societal Impacts, Geological Society of London, https://doi.org/10.1144/SP501-2019-120, 2021.

Pittore, M., Gómez-Zapata, J., Brinckmann, N., and Rüster, M.: Assetmaster and Modelprop: web services to serve building exposure models and fragility functions for physical vulnerability to natural-hazards, https://doi.org/10.5880/riesgos.2021.005, 2021a.

535   Pittore, M., Haas, M., Gomez-Zapata, J., Brinckmann, N., Rüster, M., and Proß, B.: Quakeledger: a web service to serve earthquake scenarios, https://doi.org/10.5880/riesgos.2021.003, 2021b.

Poulos, A., Monsalve, M., Zamora, N., and de la Llera, J. C.: An Updated Recurrence Model for Chilean Subduction Seismicity and Statistical Validation of Its Poisson Nature, Bulletin of the Seismological Society of America, 109, 66–74, https://doi.org/10.1785/0120170160, 2019.

Rasmussen, C. and Williams, C., eds.: Gaussian Processes for Machine Learning, The MIT Press, 2006.

540   Rosero-Velásquez, H. and Straub, D.: Selecting Representative Scenarios for Contingency Analysis of Infrastructure Systems with Dependent Component Failures, in: Proceedings of the 13th International Conference on Applications of Statistics and Probability in Civil Engineering (ICASP13), pp. 1–8, 2019.

Rosero-Velásquez, H. and Straub, D.: Selection of representative natural hazard scenarios for engineering systems, Earthquake Engineering & Structural Dynamics, 51, 3680–3700, https://doi.org/https://doi.org/10.1002/eqe.3743, 2022.

545   Salgado-Gálvez, M. A., Zuloaga, D., Henao, S., Bernal, G. A., and Cardona, O. D.: Probabilistic assessment of annual repair rates in pipelines and of direct economic losses in water and sewage networks: application to Manizales, Colombia, Natural Hazards, 93, 5–24, https://doi.org/10.1007/s11069-017-2987-z, 2018.

Silva, V.: Critical Issues in Earthquake Scenario Loss Modeling, Journal of Earthquake Engineering, 20, 1322–1341, https://doi.org/10.1080/13632469.2016.1138172, 2016.

550   Silverman, B. W.: Density Estimation for Statistics and Data Analysis, Chapman & Hall/CRC, Boca Raton, FL, 1986.

USGS: Search earthquake catalog, https://earthquake.usgs.gov/earthquakes/search/, 2023.

Villar-Vega, M., Silva, M., Crowley, H., Yepes, C., Tarque, N., Acevedo, A., Hube, M., Gustavo, C., and Santa María, H.: Development of a Fragility Model for the Residential Building Stock in South America, Earthquake Spectra, 33, 581–604, https://doi.org/10.1193/010716EQS005M, 2017.

555   Wood, A., Wollenberg, B., and Sheblé, G.: Power Generation, Operation, and Control, John Wiley & Sons Ltd, https://www.wiley.com/en-ae/Power+Generation,+Operation,+and+Control,+3rd+Edition-p-9780471790556, 2013.

Yepes-Estrada, C., Silva, V., Valcárcel, J., Acevedo, A. B., Tarque, N., Hube, M. A., Coronel, G., and María, H. S.: Modeling the Residential Building Inventory in South America for Seismic Risk Assessment, Earthquake Spectra, 33, 299–322, https://doi.org/10.1193/101915eqs155dp, 2017.