Accounting for ground motion uncertainty in empirical seismic fragility modeling

Lukas Bodenmann, M.EERI*[1], Jack W. Baker, M.EERI[2], and Božidar Stojadinović, M.EERI[1]

[1]Department of Civil, Environmental and Geomatic Engineering, ETH Zurich
[2]Department of Civil and Environmental Engineering, Stanford University

July 29, 2024

ABSTRACT

Seismic fragility models provide a probabilistic relation between ground motion intensity and damage, making them a crucial component of many regional risk assessments. Estimating such models from damage data gathered after past earthquakes is challenging because of uncertainty in the ground motion intensity the structures were subjected to. Here, we develop a Bayesian estimation procedure that performs joint inference over ground motion intensity and fragility model parameters. When applied to simulated damage data, the proposed method can recover the data-generating fragility functions, while the traditionally used method, employing fixed, best-estimate, intensity values, fails to do so. Analyses using synthetic data with known properties show that the traditional method results in flatter fragility functions that overestimate damage probabilities for low-intensity values and underestimate probabilities for large values. Similar trends are observed when comparing both methods on real damage data. The results suggest that neglecting ground motion uncertainty manifests in apparent dispersion in the estimated fragility functions. This undesirable feature can be mitigated through the proposed Bayesian procedure.

*Corresponding author: bodenmann (at) ibk.baug.ethz.ch

INTRODUCTION

Procedures for earthquake risk assessments often rely on probabilistic relations between ground shaking intensity and physical damage to elements of the built environment. A fragility function plots the probability of reaching or exceeding a certain damage limit state for increasing ground motion intensity measure (IM) levels. The fragility function parameters are derived from expert opinion (e.g., Rojahn and Sharpe, 1985; Hadlos et al., 2023), or estimated from data that is either generated through structural analyses of building archetypes, the so-called mechanical or analytical approach (e.g., Mosalam et al., 1997; Martins and Silva, 2020), or gathered after an earthquake event, the so-called empirical approach (e.g., Basöz and Kiremidjian, 1998; Rossetto et al., 2014).

Besides challenges in quality and coverage of post-earthquake damage surveys, empirical fragility modeling also requires assumptions on the IM values that caused the observed damage (Rossetto and Ioannou, 2018). These values are only known at locations of seismic network stations, and are uncertain at other locations. Thus, if a group of damaged buildings is observed, we do not know *a-priori* whether they were damaged because they were particularly vulnerable, because the ground shaking was particularly strong, or because of both. This study aims to provide further insights into this causality dilemma and proposes a new methodology to estimate fragility function parameters.

One common approach to address this causality dilemma assumes fixed, best-estimate, IM values for parameter estimation (Rossetto et al., 2014). The fixed values represent best-estimates that are typically chosen as the median of the IM distribution conditional on available data from seismic network stations (e.g., from a ShakeMap system). This fixed IM approach served as the basis for the vast majority of available empirical fragility functions, including early (e.g., Basöz and Kiremidjian, 1998) and recent studies (e.g., Rosti et al., 2021). By considering fixed IM values, such an approach does not account for potential deviations of the actual IM from this best-estimate. The sensitivity of estimated fragility function parameters with respect to IM uncertainty was examined by Ioannou et al. (2015) through simulated case studies. They found that unrealistically dense seismic networks are required to recover the "true" fragility parameters used to generate the data. For realistic network configurations, however, the approach based on fixed IM values led to substantial differences between estimated and "true" parameters. Approaches to account for IM uncertainty vary in the existing literature. For example, King et al. (2005) focused on damage survey data collected within a limited distance to a seismic network station. Lallemant et al. (2015) suggested weighting the observations according to the inverse of the remaining IM variance, thus assigning higher weight to data collected close to stations. Ioannou et al. (2015) simulated numerous IM values at the survey locations and performed parameter estimation for each simulation run. The resulting set of estimated parameters was then assumed to reflect IM uncertainty.

While aiming to account for IM uncertainty, the above approaches miss a crucial component of the causality dilemma, which is best explained by considering the inverse problem, where the true fragility function parameters are available. In this hypothetical case, studied by Pozzi and Wang (2018), the damage observations serve as noisy measurements of the uncertain IM values. Consequently, these observations can be used to obtain a constrained estimate of the IM values that caused the damage.

In cases of practical interest, both the fragility parameters and the IM values are uncertain. This study presents a Bayesian approach to estimate their joint posterior distribution, which can account for uncertainty in both the IM values and the fragility parameters. We start with an overview of fragility function definitions, followed by mathematical descriptions of the proposed Bayesian parameter estimation approach and the traditional, fixed IM, approach. Both methods are first explained on a simplified example and then compared using data from the 2009 L'Aquila (Italy) earthquake.

FRAGILITY FUNCTIONS

Empirical fragility modeling aims to estimate the fragility function parameters using observed or estimated values of IMs and building damage states. Damage states are obtained from post-earthquake building inspections, where experts describe the building characteristics and the sustained damage based on pre-specified guidelines and data-recording forms. Information on the building characteristics pertain to the material and the type of the load resisting system, as well as geometrical attributes such as the number of storeys. Buildings with similar characteristics are grouped into a 'building class' with a corresponding set of fragility functions. Damage descriptions are commonly reduced to ordered, collectively exhaustive and mutually exclusive damage states (e.g., no, slight, moderate, heavy and complete damage), which are numerically encoded as $ds \in \{0, 1, ..., c\}$. To ease the notation in the following definitions, we consider a single building class and assume that all damage observations for buildings of that class are treated equally.

The log-normal cumulative distribution is the most common functional form for fragility functions

$$\mathrm{P}(DS \geq ds|im) = \Phi\left(\frac{\ln(im/\theta_{ds})}{\beta_{ds}}\right) = \Phi\left(\frac{\ln im - \ln \theta_{ds}}{\beta_{ds}}\right) \ , \tag{1}$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function, θ_{ds} denotes the median IM which causes the structure to reach or exceed ds, and β_{ds} is the standard deviation of the $\ln IM$, here referred to as the dispersion parameter. From the fragility functions, one obtains the probability mass function, $\mathrm{P}(DS = ds|im) = p(ds|im)$, as

$$p(ds|im) = \begin{cases} 1 - \mathrm{P}(DS \geq ds + 1|im), & \text{if } ds = 0 \\ \mathrm{P}(DS \geq ds|im), & \text{if } ds = c \\ \mathrm{P}(DS \geq ds|im) - \mathrm{P}(DS \geq ds + 1|im), & \text{otherwise .} \end{cases} \tag{2}$$

To prevent negative probabilities in Eq. (2), we assume an identical dispersion parameter β for all damage states. This avoids a crossing of the fragility functions and is a standard assumption taken in analytical and empirical fragility modeling for ordered damage states (e.g., Martins and Silva, 2020; Rosti et al., 2021). For $c+1$ damage states, we thus have c increasing parameters $\theta_1 < \theta_2 < \cdots < \theta_c$ and one common dispersion parameter β.

A second fragility function definition, equivalent to Eq. (1), is obtained from the cumulative probit model (Lallemant et al., 2015). By introducing the threshold parameters $\eta_{ds} = \beta^{-1} \ln \theta_{ds}$, this alternative definition reads as

$$\mathrm{P}(DS \geq ds|im) = \Phi\left(\beta^{-1} \ln im - \eta_{ds}\right) \ . \tag{3}$$

Following Nguyen and Lallemant (2022), we use the re-parametrized definition of Eq. (3) with the transformed parameters η_{ds} still being of increasing order. To account for this ordering in the estimation algorithms, we introduce positive parameters $\delta_{ds} = \eta_{ds} - \eta_{ds-1}$ for states $ds > 1$. The parameters of interest are collectively denoted as vector $\boldsymbol{\vartheta} = [\beta, \eta_1, \delta_2, \ldots, \delta_c]^\top$. Finally, we denote the fragility model in terms of the probability mass function $p(ds|im, \boldsymbol{\vartheta})$, where we explicitly condition on parameter vector $\boldsymbol{\vartheta}$.

The use of log-normal fragility functions – leading to the cumulative probit model – dates back to earthquake risk assessments of critical infrastructure components (e.g., Kennedy et al., 1980) and is well-established in disaster risk analysis (e.g., Kircher et al., 1997; Dolce et al., 2021) as well as in performance-based earthquake engineering (e.g., Porter et al., 2007; Baker, 2015). Yet, the herein presented estimation algorithms can also be used for alternative models, such as the cumulative logistic model employed by Basöz and Kiremidjian (1998).

ESTIMATING FRAGILITY FUNCTION PARAMETERS

Figure 1 depicts the probabilistic model relating the variables involved in the estimation of $\boldsymbol{\vartheta}$ by considering a damage data set from n buildings gathered after an earthquake with rupture characteristics **rup**. The fragility model of Eq. (2) relates the observed damage states, $\mathbf{ds} = [ds_i, \ldots, ds_n]^\top$, to the IM values at the locations of the surveyed buildings, $\mathbf{im}_\mathcal{B} = [im_1, \ldots, im_n]^\top$. Knowing the latter would allow estimating $\boldsymbol{\vartheta}$ with similar techniques used in analytical fragility modeling (e.g., Baker, 2015). In most cases, however, we only know the IM values at the locations of seismic stations, here denoted as $\mathbf{im}_\mathcal{S} = [im_{n+1}, \ldots, im_{n+m}]^\top$ for the case of m stations, while their counterparts at the survey locations are uncertain.

To constrain the above-mentioned uncertainty, we evaluate the distribution of $\mathbf{im}_\mathcal{B}$ conditional on station data $\mathbf{im}_\mathcal{S}$ and rupture characteristics **rup** by following Engler et al. (2022), who present the mathematical background for the most recent version of the United States Geological Survey's ShakeMap system. Conditional on **rup** only, the joint distribution of $\mathbf{im}_\mathcal{B}$ and $\mathbf{im}_\mathcal{S}$ is assumed to be multivariate log-normal (Jayaram and Baker, 2008), with parameters derived from empirical ground motion models (GMMs) and spatial correlation models. Thus, the conditional distribution of interest also follows a multivariate log-normal distribution, i.e., $p(\mathbf{im}_\mathcal{B}|\mathbf{im}_\mathcal{S}, \mathbf{rup}) = \mathcal{LN}(\boldsymbol{\mu}_{\mathcal{B}|\mathcal{S}}, \boldsymbol{\Sigma}_{\mathcal{BB}|\mathcal{S}})$. The distributional parameters, $\boldsymbol{\mu}_{\mathcal{B}|\mathcal{S}}$ and $\boldsymbol{\Sigma}_{\mathcal{BB}|\mathcal{S}}$ are computed via Equations 4 and 5 in Engler et al. (2022), also explained in Appendix A of the electronic supplement.

We next present two methodologies to estimate $\boldsymbol{\vartheta}$. The first, traditional, method assumes fixed – or deterministic – values for $\mathbf{im}_\mathcal{B}$. The second, proposed, method, treats both, $\boldsymbol{\vartheta}$ and $\mathbf{im}_\mathcal{B}$, as uncertain and follows a Bayesian approach for inference.