

THE CHICKEN AND EGG PROBLEM IN EMPIRICAL FRAGILITY MODELING AND THE APPROACHES TO SOLVE IT

L. Bodenmann¹, J.W. Baker² & B. Stojadinović³

¹ Department of Civil, Environmental and Geomatic Engineering, ETH Zurich, Zurich, Switzerland, bodenmann@ibk.baug.ethz.ch

² Department of Civil and Environmental Engineering, Stanford University, Stanford, USA

³ Department of Civil, Environmental and Geomatic Engineering, ETH Zurich, Zurich, Switzerland

Abstract: *Regional earthquake risk analyses employ fragility models to estimate shaking-induced damage to the built environment. Empirical fragility models are estimated from damage data gathered after past earthquakes and present a valuable resource to calibrate and validate models derived from physics-based computer simulations. Yet, empirical fragility modelling is challenging because it requires knowledge of the ground motion intensity the buildings were subjected to. The latter, however, is only known at locations of seismic network stations while its uncertainty increases for locations further away. Thus, if a group of damaged buildings is observed, we do not know, a-priori, whether they were damaged because they were particularly vulnerable or because the ground shaking was particularly strong or because of both. We refer to this as the chicken and egg problem in empirical fragility modelling. In this work, we present a Bayesian approach to quantify the joint posterior distribution of the fragility model parameters and the ground motion intensity. Using a simulated damage data set, we compare the Bayesian approach to the traditionally used approach that relies on fixed ground motion intensity estimates.*

1 Introduction

Regional risk assessments often rely on probabilistic relations between ground shaking intensity and physical damage to the exposed structures. A fragility function plots the probability of reaching or exceeding a certain damage limit state for increasing ground motion intensity measure (IM) levels. Empirical fragility modeling aims to estimate such functions from damage survey data collected after past earthquakes. Besides the need for high-quality and high-coverage damage surveys, this empirical approach also requires information on the IM values at the survey locations (Rossetto and Ioannou, 2018). Unless the surveyed structures were equipped with seismic recording stations, these IM values are unknown. Yet, the observed damage is caused by both: the structures' fragilities and the IM values they were subjected to. This leads to the chicken and egg problem in empirical fragility modelling: How can one estimate the structures' fragilities if the actual IM values are unknown?

One approach to tackle this causality dilemma estimates fragility function parameters using fixed, deterministic, IM values. 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 shake map). The fixed IM approach served as the basis for most available empirical fragility functions (e.g., Rosti et al., 2021). In simulated case studies, Ioannou et al. (2015) showed that this approach has difficulties in recovering the "true"

data-generating fragility functions. This may be explained by the fact that this approach does not account for potential deviations of the actual IM values from the fixed, best-estimate, values.

To address this shortcoming, the authors recently proposed a novel estimation approach that explicitly accounts for IM uncertainty and uses Bayesian inference to estimate the joint posterior distribution of IM values and fragility function parameters (Bodenmann *et al.*, 2023). This study illustrates the new approach: it starts with an overview of fragility function definitions, followed by mathematical descriptions of the traditional, fixed IM, estimation approach and the new Bayesian approach. Both methods are then compared using a simulated damage data set.

2 Fragility function parameterization

This study considers fragility functions for ordered, collectively exhaustive and mutually exclusive damage states (e.g., no, slight, moderate, heavy and complete damage), which are numerically encoded as $ds \in \{0, 1, \dots, c\}$. A fragility function then plots the probability of reaching or exceeding a certain damage state, ds , for increasing levels of ground motion intensity, im . The most common functional form for fragility functions is based on the log-normal cumulative distribution and is parametrized as

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

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 logarithmic IM, here referred to as the dispersion parameter. To avoid a crossing of the fragility functions, we choose an identical dispersion parameter β for all damage states. This is a standard assumption taken in fragility modeling for ordered damage states (e.g., Nguyen and Lallemand, 2022). For $c + 1$ damage states, we thus have c increasing parameters $\theta_1 < \theta_2 < \dots < \theta_c$ and one common dispersion parameter β , which are collectively denoted with vector $\boldsymbol{\vartheta}$.

From the fragility functions, one obtains the probability mass function $P(DS = ds|im) = p(ds|im)$ as

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

We denote the fragility model in terms of the probability mass function $p(ds|im, \boldsymbol{\vartheta})$, where we explicitly condition on parameters $\boldsymbol{\vartheta}$.

In accordance with the current state-of-the-art in empirical fragility modelling, we group buildings with similar characteristics in building classes, here denoted as bc . Then, we estimate fragility function parameters for each class $bc \in \mathcal{BC}$, with \mathcal{BC} being the set of all considered classes. Thus, the total parameter set is $\boldsymbol{\vartheta} = \{\boldsymbol{\vartheta}_{bc} | bc \in \mathcal{BC}\}$.

3 Estimation of empirical seismic fragility functions

Figure 1 shows the relations between 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 \mathbf{rup} . The fragility model of Eq. (2) relates the observed damage states, $\mathbf{ds} = [ds_1, \dots, ds_n]^T$, to the IM values at the locations of the surveyed buildings, $\mathbf{im} = [im_1, \dots, im_n]^T$. 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 $\mathbf{im}_s = [im_{n+1}, \dots, im_{n+m}]^T$ for the case of m stations, while their counterparts at the survey locations are uncertain.

To constrain the above-mentioned uncertainty, we follow the approach proposed by Engler *et al.* (2022) and implemented in the most recent version of the United States Geology Survey's shake map system. The distribution of \mathbf{im} conditional on station data \mathbf{im}_s and rupture characteristics \mathbf{rup} is a multivariate log-normal

$$p(\mathbf{im}|\mathbf{im}_s, \mathbf{rup}) = \mathcal{LN}(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \quad (3)$$

with the distributional parameters, $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, being computed via Equations 4 and 5 in Engler et al. (2022) and using an empirical ground motion model (GMM) and a spatial correlation model. In the remainder of this study, we use the term prior IM distribution for the distribution specified in Eq. (3).

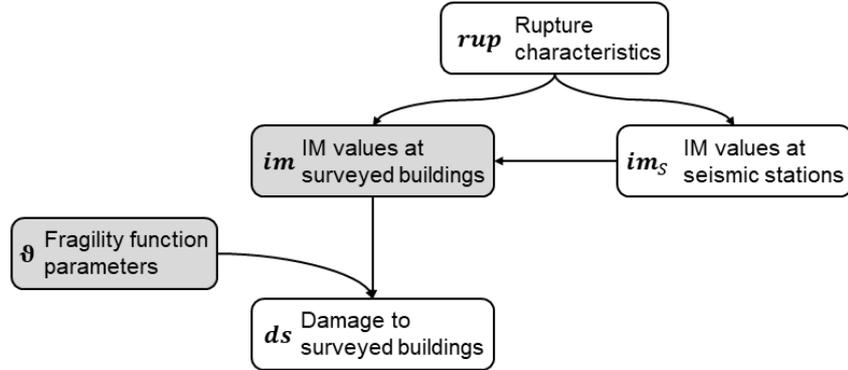


Figure 1. Relations between observed (white) and unobserved (grey) variables involved in the fragility function estimation from damage survey data and seismic network station data.

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

3.1 Maximum likelihood estimation with fixed IM values

This estimation method replaces uncertain IM values at the survey sites with fixed, best-estimate, values derived from the prior IM distribution $p(\boldsymbol{im}|\boldsymbol{im}_s, \boldsymbol{rup})$. For site i , the best-estimate value corresponds to $\overline{im}_i = \exp[\boldsymbol{\mu}]_i$, which is the median IM conditional on station data \boldsymbol{im}_s . The joint distribution of building damage conditional on the IM values at the building locations is assumed to factorize, i.e., $p(\boldsymbol{ds}|\boldsymbol{im}, \boldsymbol{\vartheta}) = \prod_{i=1}^n p(ds_i|im_i, \boldsymbol{\vartheta})$. The parameters are then estimated by maximizing the log-likelihood

$$\hat{\boldsymbol{\vartheta}} = \underset{\boldsymbol{\vartheta}}{\operatorname{argmax}} \sum_{i=1}^n \ln p(ds_i|\overline{im}_i, \boldsymbol{\vartheta}), \quad (4)$$

such that the obtained parameter estimates maximize the probability of observing the surveyed damage under the chosen best-estimate IM values. In the following, we refer to this estimation method as the fixed IM approach.

3.2 Bayesian estimation with uncertain IM values

To account for ground motion uncertainty, Bodenmann et al. (2023) take a Bayesian approach treating both the fragility function parameters, $\boldsymbol{\vartheta}$, and the IM values at the survey sites, \boldsymbol{im} , as realizations of random variables. The joint posterior distribution of \boldsymbol{im} and $\boldsymbol{\vartheta}$ conditional on the triplet of damage survey data, \boldsymbol{ds} , station data, \boldsymbol{im}_s , and rupture characteristics, \boldsymbol{rup} , is expressed as

$$p(\boldsymbol{\vartheta}, \boldsymbol{im} | \boldsymbol{ds}, \boldsymbol{im}_s, \boldsymbol{rup}) = \left(\prod_{i=1}^n p(ds_i|im_i, \boldsymbol{\vartheta}) \right) \frac{p(\boldsymbol{im}|\boldsymbol{im}_s, \boldsymbol{rup}) p(\boldsymbol{\vartheta})}{p(\boldsymbol{ds}|\boldsymbol{im}_s, \boldsymbol{rup})} \quad (5)$$

where the denominator, $p(\boldsymbol{ds}|\boldsymbol{im}_s, \boldsymbol{rup})$, is the marginal likelihood, i.e., the probability that the prior model assigns to the observed damage data conditional on station data and rupture characteristics, $p(\boldsymbol{im}|\boldsymbol{im}_s, \boldsymbol{rup})$ is the prior IM distribution specified in Eq. (3), $p(ds_i|im_i, \boldsymbol{\vartheta})$ follows from the fragility model defined in Eq. (2), and $p(\boldsymbol{\vartheta})$ is the prior distribution of the fragility function parameters. The latter consists of weakly informative priors for each parameter that provide sufficient adaptability to cover building classes with varying susceptibility to earthquake damage. Then, Markov Chain Monte Carlo (MCMC) is used to draw samples $(\boldsymbol{\vartheta}, \boldsymbol{im})$ from the target posterior specified in Eq. (5). The MCMC implementation and the prior distributions of the fragility function parameters are described in Bodenmann et al. (2023).

Finally, we note that the use of the entire prior IM distribution in Eq. (5) requires access and storage of the potentially high-dimensional covariance matrix Σ . This makes the Bayesian approach computationally more expensive than the traditionally used, fixed IM approach.

4 Case study

This case study estimates fragility function parameters using simulated damage data from buildings located in the L'Aquila (Italy) region. The damage data was simulated by considering the rupture characteristics and station data from the 2009 M6 L'Aquila earthquake as obtained from the engineering strong motion database (Luzi *et al.*, 2020). With this analysis, we aim to compare the performance of the two previously described estimation procedures in recovering the “true” fragility functions that were used to simulate the damage data set. Figure 2 shows the surface projection of the rupture together with the locations of the seismic network stations and of the surveyed buildings.

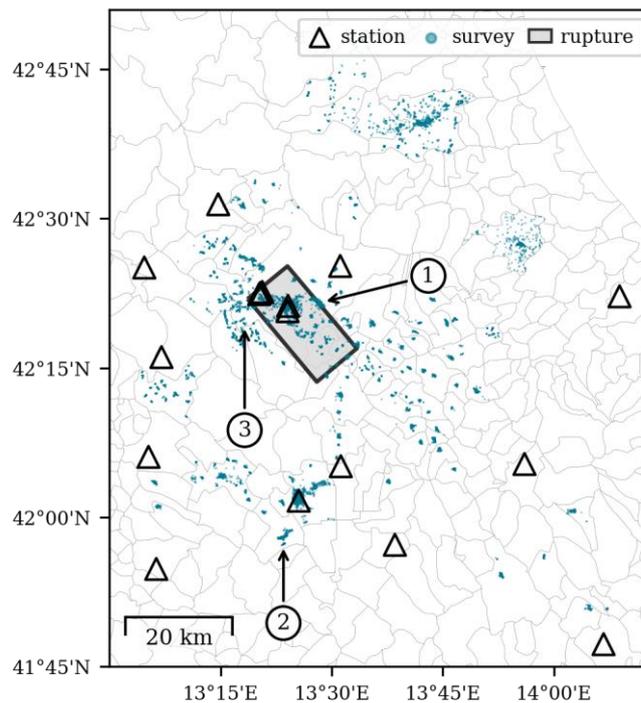


Figure 2. Overview of the L'Aquila case study with the surface projection of the rupture from the 2009 M6 earthquake, locations of seismic network stations and the considered buildings in the simulated damage survey data set. The numbered circles refer to example locations that are analyzed in Section 4.2.

The considered data set contains data from 12,000 buildings that are categorized into three building classes, A, B and C, with decreasing susceptibility to earthquake-induced damage. The damage states consist of six categories: no, negligible to slight, moderate, severe, very heavy damage and collapse, numerically encoded as $ds \in \{0,1,2,3,4,5\}$. Thus, we estimate six parameters for each building class, leading to a total of 18 parameters.

We estimate fragility functions using the elastic, 5%-damped spectral acceleration at 0.3 second, $SA(0.3s)$, as the IM of interest. Thus, the vectors \mathbf{im} and \mathbf{im}_S are spectral accelerations at sites of surveyed buildings and network stations, respectively. To compute the prior IM distribution, $p(\mathbf{im}|\mathbf{im}_S, \mathbf{rup})$, we use the Bindi *et al.* (2011) GMM and the spatial correlation model of Esposito and Iervolino (2012). Figure 3a plots the median spectral accelerations conditional on station data \mathbf{im}_S , while Figure 3b illustrates the logarithmic standard deviation. The latter shows the reduced IM uncertainty in the vicinity of stations but also highlights how this uncertainty increases with distance to the stations. As shown in Figure 2, many buildings of the considered survey data set are located far away from the recording stations with corresponding IM values being subject to increased uncertainty.

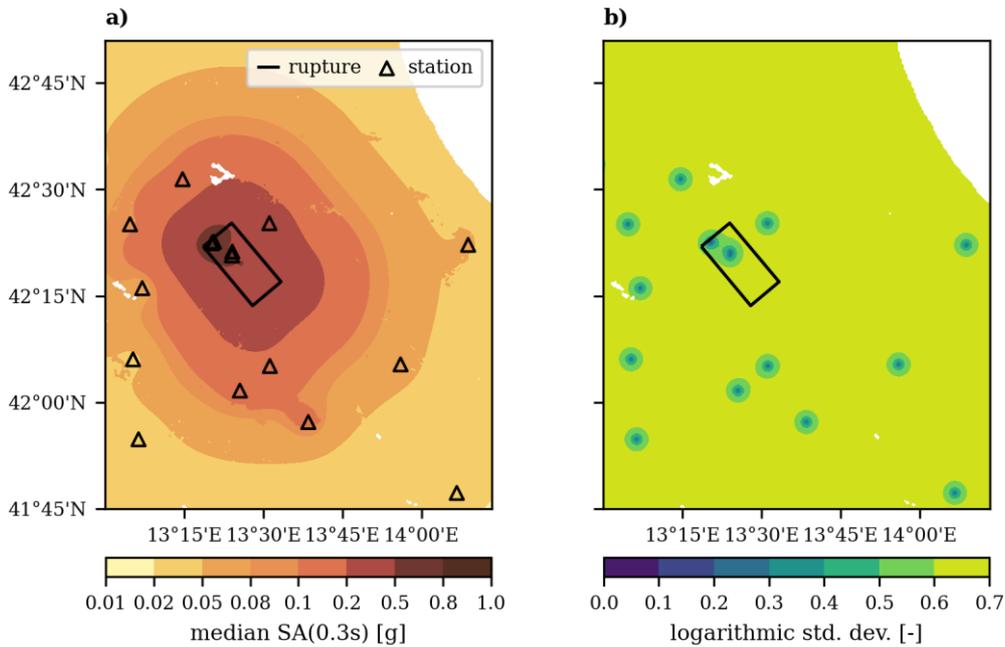


Figure 3. The spatial distribution of median IM values (a) and logarithmic standard deviation (b) conditional on seismic network recordings from the stations indicated in (a).

4.1 Maximum likelihood estimation with fixed IM values

The fixed IM approach uses the median IM values, illustrated in Figure 3a, to perform maximum likelihood estimation. For the three considered building classes, Figure 4 compares the estimated fragility functions with the “true” data-generating functions. We observe that the estimated functions are flatter than the “true” ones. Compared to the “true” functions, the fixed IM approach overestimates damage probabilities for low IM values and underestimates probabilities for large IM values. Such behavior results from an overestimation of the dispersion parameter β . The “true” β values are 0.59, 0.80, and 0.81 for the three building classes, while the estimated values are 1.10, 1.37, and 1.19, respectively. This overestimation indicates that IM uncertainty manifests as apparent uncertainty – or dispersion – in the fragility functions.

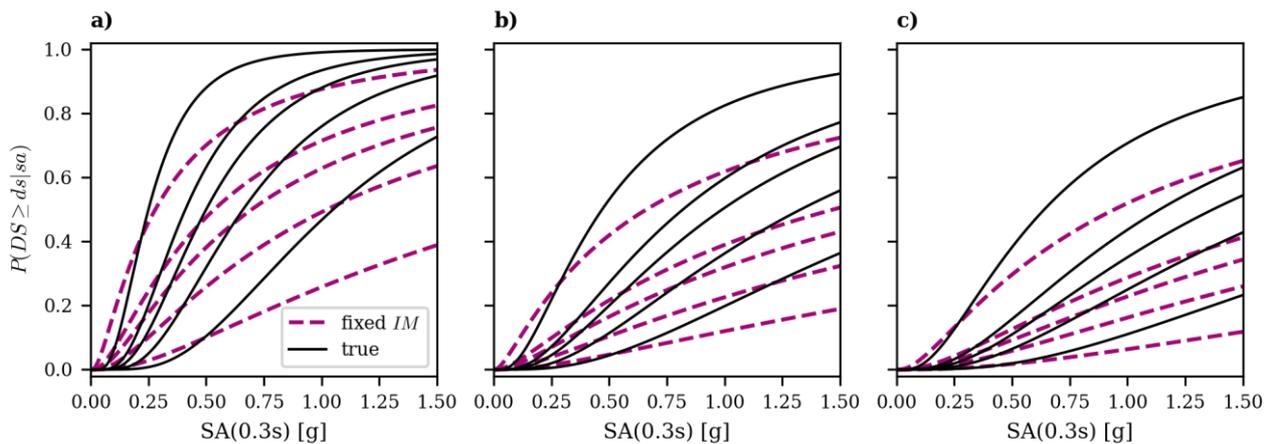


Figure 4. Fragility functions for three building classes (a to c) estimated with the fixed IM approach and compared to the “true” data-generating functions. The five functions correspond to $ds \in \{1,2,3,4,5\}$ from the top left to the bottom right.

4.2 Bayesian estimation with uncertain IM values

The Bayesian approach provides samples from the joint posterior of IM values and fragility parameters. Figure 5 plots the fragility functions obtained from the mean of the posterior parameter samples (dashed lines) while

the shaded areas indicate the 90% credibility interval. The latter is the difference between the 95% and 5% quantiles obtained at each IM level for all functions computed from the posterior parameter samples. In this case, the estimated mean β values are 0.61, 0.80 and 0.81 for the three building classes, which closely match the “true” values (0.59, 0.80, and 0.81). Compared to the fixed IM approach, shown in Figure 4, the Bayesian approach performs better in recovering the “true” functions. By explicitly accounting for IM uncertainty, the Bayesian approach avoids additional apparent fragility function dispersion observed with the fixed IM approach.

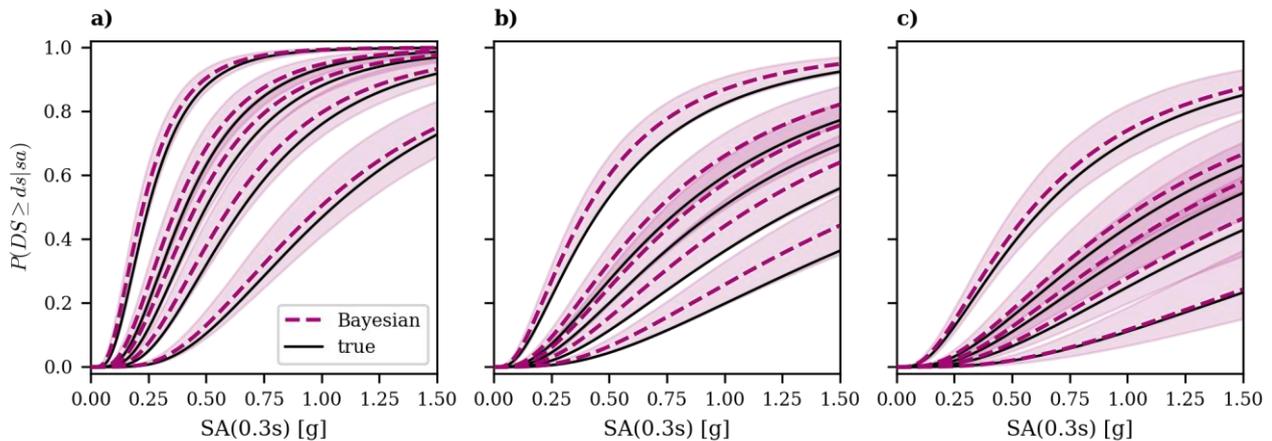


Figure 5. Fragility functions for three building classes (a to c) estimated with the Bayesian approach together with the “true” data-generating functions. Fragility functions derived using the Bayesian approach are illustrated using the mean posterior samples (solid lines) and the 90% credibility interval (shaded area).

To illustrate the posterior samples of the IM values, Figure 6a shows the spatial distribution of the posterior median IM across the study region and Figure 6b illustrates the associated logarithmic standard deviation. Compared to the prior IM distribution shown in Figure 3, the conditioning on damage survey data leads to a more refined spatial pattern of the median IM and a reduced IM uncertainty in the vicinity of surveyed buildings.

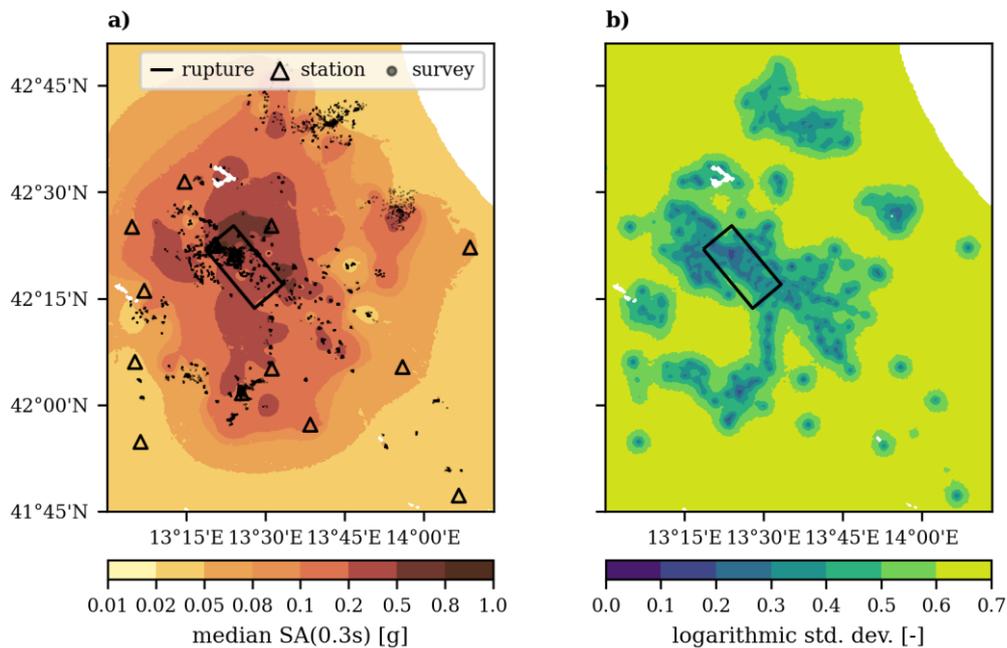


Figure 6. The spatial distribution of median IM values (a) and logarithmic standard deviation (b) estimated from the posterior samples obtained from the Bayesian approach using seismic network recordings and damage data from the stations and the surveyed buildings indicated in (a).

For a further comparison of the prior and posterior IM values, Figure 7 plots the corresponding distributions at the locations of three surveyed buildings indicated in Figure 2. The dotted vertical lines show the simulated “true”, but unobserved, IM value at these sites, and the labeled points indicate the prior and posterior median IM values. Recall that the fixed IM approach uses these prior median values to estimate the fragility functions, while the Bayesian approach considers the entire prior IM distribution. The posterior distribution, obtained from the Bayesian approach, has reduced uncertainty and the median values are closer to the “true” values.

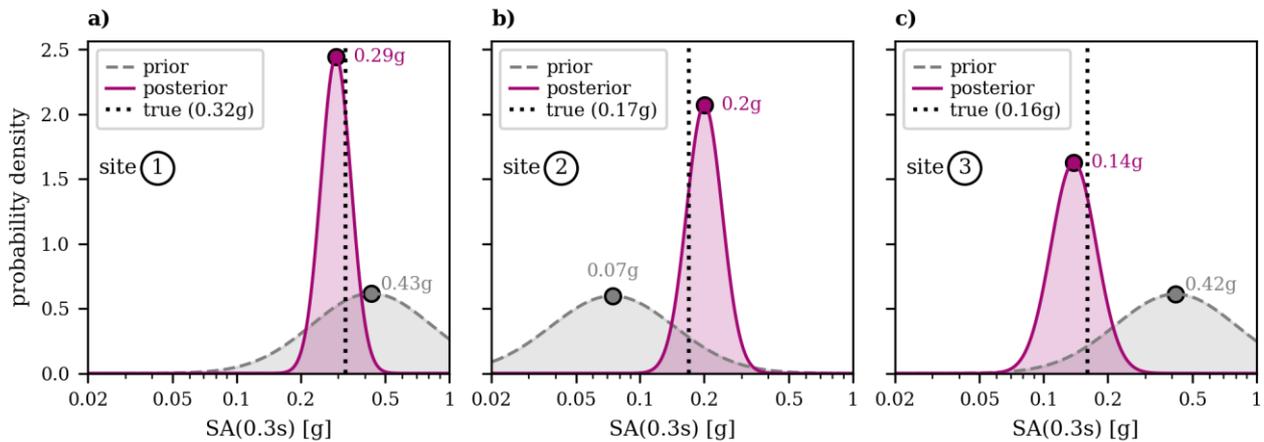


Figure 7. Comparison of the prior and posterior distributions of $SA(0.3s)$ at the three survey sites indicated in Figure 2, with the simulated “true” values at these sites. The prior distribution is conditioned on the rupture information and the station data, while the posterior distribution additionally considers the damage survey data through the proposed Bayesian approach.

We note that the presented case study results are obtained for a simulated damage data set. To establish the prior IM distribution, we used the same ground motion and spatial correlation models as the ones used to generate the data. In Bodenmann *et al.* (2023), the authors compare the methodologies using actual damage data and test their robustness with respect to the employed ground motion and spatial correlation models.

5 Conclusions

This study explored the chicken-and-egg problem in estimating fragility function parameters from post-earthquake damage survey data. This causality dilemma is due to the uncertainty in the IM values the surveyed buildings were subjected to. We compared two parameter estimation approaches that tackle this issue differently: the first, traditional, approach uses fixed, best-estimate, IM values for parameter estimation, while the second, recently proposed, approach considers IM values and fragility function parameters as uncertain variables and uses Bayesian inference to estimate their joint posterior distribution.

Both approaches were applied to a simulated damage survey data set of 12'000 buildings. We assessed their performance in recovering the “true” data-generating fragility functions. The fragility functions estimated with the fixed IM approach are flatter than the “true” functions. In other words, the fixed IM approach overestimates damage probabilities for low IM values and underestimates probabilities for large IM values. By considering fixed, best-estimate, IM values, this approach does not account for the variability of the actual, unobserved, IM around these best-estimates. This causes an overestimation of the fragility function dispersion and results in flatter fragility functions.

The proposed Bayesian approach, on the other hand, explicitly accounts for IM uncertainty, and performs simultaneous inference on the IM values and the fragility function parameters. This procedure avoids an inflation of fragility function dispersion, as observed with the fixed IM approach, and the estimated fragility functions better match the “true” ones used to generate the data. This extra value makes the increased computational cost of the Bayesian approach a good investment for empirical fragility studies.

While the focus of the presented work lies on IM uncertainty, there are several opportunities for further research addressing additional challenges in empirical fragility modelling. The Bayesian approach could be adapted to fragility models that depend on building characteristics in a more complex manner than the commonly used

categorization into pre-defined building classes. It could also be extended to account for uncertainty in the collected building characteristics (or the deduced building classes). Despite these possible extensions of the presented Bayesian approach, its current capability to account for IM uncertainty makes it a valuable tool for analysts interested in improved empirical fragility modeling.

6 Code availability

The presented results were obtained with BayesFrag, an open-source, Python-based, software tool developed by the authors. BayesFrag is available at <https://doi.org/10.5281/zenodo.10074233> and comes with several tutorials that assist analysts interested in applying the Bayesian parameter estimation approach to their data sets.

7 References

- Baker J.W. (2015). Efficient Analytical Fragility Function Fitting Using Dynamic Structural Analysis. *Earthquake Spectra*, 31(1): 579-599, <https://doi.org/10.1193/021113EQS025M>
- Bodenmann L., Baker J.W., Stojadinović B. (2023). Accounting for ground motion uncertainty in empirical seismic fragility modelling, *Engineering Archive* [preprint], <https://doi.org/10.31224/3336>
- Bindi D., Pacor F., Luzi L., Puglia R., Massa M., Ameri G., Paolucci R. (2011). Ground motion prediction equations derived from the Italian strong motion database. *Bulletin of Earthquake Engineering*, 9(6): 1899-1920, <https://doi.org/10.1007/s10518-011-9313-z>
- Engler D.T., Worden C.B., Thompson E.M., Jaiswal K.S. (2022). Partitioning Ground Motion Uncertainty When Conditioned on Station Data. *Bulletin of the Seismological Society of America*, 112(2): 1060-1079, <https://doi.org/10.1785/0120210177>
- Esposito S., Iervolino I. (2012). Spatial correlation of spectral acceleration in European data. *Bulletin of the Seismological Society of America*, 102(6): 2781-2788, <https://doi.org/10.1785/0120120068>
- Ioannou I., Douglas J., Rossetto T. (2015). Assessing the impact of ground-motion variability and uncertainty on empirical fragility curves. *Soil Dynamics and Earthquake Engineering*, 69: 83-92, <https://doi.org/10.1016/j.soildyn.2014.10.024>
- Luzi L., Lanzano G., Felicetta C., D'Amico M., Russo E., Sgobba S., Pacor F., et al. (2020). Engineering Strong Motion Database (ESM) (Version 2.0). <https://doi.org/10.13127/ESM.2>
- Nguyen M., Lallemand D. (2022). Order Matters: The Benefits of Ordinal Fragility Curves for Damage and Loss Estimation. *Risk Analysis* 42(5): 1136-1148. <https://doi.org/10.1111/risa.13815>
- Rossetto T., Ioannou I. (2018). Empirical Fragility and Vulnerability Assessment: Not Just a Regression, In: *Risk Modeling for Hazards and Disasters*. Elsevier, 79-103. <https://doi.org/10.1016/B978-0-12-804071-3.00004-5>
- Rosti A., Rota M., Penna A. (2021). Empirical fragility curves for Italian URM buildings, *Bulletin of Earthquake Engineering*, 19(8): 3057–3076. <https://doi.org/10.1007/s10518-020-00845-9>