
USING GAUSSIAN PROCESS MODELS FOR DYNAMIC POST-EARTHQUAKE IMPACT ESTIMATION WITH REGIONAL RISK PREDICTORS

 **Lukas Bodenmann***

 **Yves Reuland***

 **Bozidar Stojadinovic***

March 1, 2022

ABSTRACT

The widespread damage to the built environment that may be caused by earthquakes, not only leads to direct consequences and losses, but also disrupts safe functionality of buildings and thus, potentially induces severe long-term societal consequences. Better community resilience may be achieved through well-organized recovery, which is complicated by the conditions, in which decisions need to be taken, characterized by intense time pressure and limited information on the severity and the spatial distribution of building damage. Approximate initial damage estimates are produced using regional risk models, which convolute early ground-motion data with indirect mapping schemes of buildings to typological classes and with generic building fragility functions. This domain-knowledge is integrated into Gaussian process models, a probabilistic machine-learning tool, to leverage the continuous data inflow from a building inspection campaign. Initial post-earthquake damage estimates are dynamically improved by simultaneously updating the distributions of ground-motion intensity, typological attribution, and building damage. Hence, the inspection information becoming available in the first days following an earthquake helps constraining uncertainties and provides reliable estimates of the geographical distribution of damaged buildings after a fraction of the time required to inspect all the buildings. The performance of the proposed methodology is demonstrated on a fictitious earthquake scenario and two real damage datasets from historic earthquakes and a comparison with purely data-driven methods shows that the underlying risk model reduces the number of building inspections that is required to provide reliable and precise predictions.

Keywords Post-earthquake damage assessment · Gaussian Process Models · Regional Earthquake Risk Models · Uncertainty reduction

1 Introduction

Damaging earthquakes carry the potential to cause catastrophic consequences. In addition to direct financial losses, injuries, and fatalities caused by damages to the built environment, the disruption in the capacity of buildings to shelter social and economic functionalities gives rise to negative long-term consequences, such as a decrease in available low-cost rental houses and increase in house-purchase prices (Potter et al., 2015). The severity and duration of such disruptions are not only affected by the earthquake-resistance of affected buildings, but also by the efficiency and speed, with which public and private stakeholders respond to the event. Information of the regional severity of the damage plays a central role in supporting and organising response and recovery efforts (Amin and Goldstein, 2008). However, such information is scarce, limited, and imprecise in the early aftermath of an earthquake, when, under intense time pressure, stakeholders and agencies need to take crucial decisions that affect disaster assistance and recovery. Thus, post-earthquake inspection of buildings, in addition to enabling safe building reoccupancy, leads to a systematic collection of building-specific data (Mcentire and Cope, 2004; Lallemand et al., 2017). However, the visual inspection of all buildings within the affected region is time-intensive and may take several weeks (Marquis et al., 2017; Kusunoki,

*Dept. of Civil, Environmental and Geomatic Engineering, ETH Zurich, Switzerland

2021). Therefore, decision-makers typically rely on damage estimates that are deduced from incomplete and uncertain data.

Earthquake risk models may provide initial, quantitative, and spatially exhaustive damage estimates. Such risk models aim to quantify the probability of earthquake-induced consequences on the built environment within a region and, in a broader sense, on its users and inhabitants. While this study focuses on the use of risk models for rapid post-earthquake damage estimation (Erdik et al., 2011; Guérin-Marthe et al., 2021), they are also widely employed in pre-event settings to analyze long-term seismic risk and to conduct scenario-based analyses (Silva et al., 2020; Dolce et al., 2021). Risk models require information on the exposed buildings and on the susceptibility of those buildings to ground-motions together with estimates of the spatial distribution of ground-motion intensity. In post-event cases, the latter can be constrained with ground-motion recordings (Wald et al., 1999). The precision and accuracy of the resulting damage estimates depends, amongst other, on the density of the seismic network, the level of detail of the available exposure information and, finally, how well the vulnerability estimates reflect local conditions. While these model-based estimates are particularly helpful in the immediate aftermath, when damage observations are not available yet, the multiple sources of uncertainty underline the need for a methodological framework to confirm or correct early damage estimates using empirical evidence once available.

Permanently monitoring buildings with sensors may help to detect and quantify earthquake-induced damage. This near-real-time data source has gained increasing attention at building (Noh et al., 2012; Reuland et al., 2019; Giordano and Limongelli, 2020) and regional levels (Goulet et al., 2015). Yet, translating data-driven damage indicators into robust decision-making remains an open research topic. In addition, extensive regional instrumentation of buildings is not realistic in the short-to-medium term and therefore, robust methodologies to link damage of single buildings with regional impact assessment, such as proposed in this paper, will continue to be of interest. Remote-sensing data, such as satellite, aerial, and drone imagery, consist another data source. However, most automated image-based applications either focus on binary assessment of building collapses or involve damage proxies that are difficult to link with overall building damage (Dell’Acqua and Gamba, 2012; Booth et al., 2011; Paal et al., 2014). In addition, remotely taken images are limited to the building exterior and suffer from occlusion, complex facade ornaments, and changing shadow patterns (Kerle et al., 2019).

Expert-conducted building inspection in the aftermath of an earthquake guarantees a continuous, yet slow, data inflow that qualifies the induced damage to individual buildings and, as a byproduct, provides important building information that is typically missing in an exposure database (e.g. the lateral load-resisting system). Given the data inflow, machine-learning and statistical tools may be leveraged to estimate damage sustained by buildings that have not been inspected yet. Kovačević et al. (2018) have applied random forests (RFs) to predict the damage level for individual buildings due to the 2010 Kraljevo (Serbia) earthquake. Stojadinovic et al. (2021) have extended the use of RFs and formulated an operational methodology for rapid regional repair cost estimation that requires careful prioritization of inspections. Using inspection data from the 2015 Nepal earthquake, Loos et al. (2020) employed geo-statistical techniques to predict the mean damage level aggregated to equally spaced grid cells and, assuming normally-distributed mean damage levels, produced variance-based uncertainty estimates for their predictions. However, the proposed method only applies to continuous data, requiring aggregation and averaging of categorical inspection results over grid cells. Apart from the risk of information loss, this also imposes operational constraints on the inspection process, because all buildings in a grid cell have to be inspected before the corresponding results can be used. Sheibani and Ou (2020) explored Gaussian Process (GP) models by focusing on quantities that either require sensors within buildings, such as peak floor acceleration, or become available much later in the recovery process (such as repair costs). The GP model is applied only to continuous quantities, even though application to continuous quantities is not a limitation of GP models in general.

We propose the use of GP models to fuse available inspection data with a pre-existing earthquake risk model, in order to dynamically update regional post-earthquake damage estimates. The proposed method allows the processing of observed damage, taking the form of multiple ordinal categories (i.e. damage states), in order to reduce the uncertainty of the geographical distribution of the shaking intensity, and involves the simultaneous updating of the components governing building damage. Thus, instead of learning a new, entirely data-driven, model after the event, individual risk-model components are updated, which allows for increasingly constrained simulation-based predictions of building damage that can be consistently aggregated at different spatial scales. The present work builds upon the theoretical study by Pozzi and Wang (2018), who proposed GP models to predict the failure probability of individual components in spatially distributed infrastructure systems based on binary component state observations.

The structure of this paper is as follows: First, we present the mathematical background of GP models in a general context (Section 2), followed by an introduction to earthquake risk models in Section 3. Then, Section 4 describes the GM model inference and updating process using post-earthquake data from visual inspections. Finally, in Section 5, the

methodology is applied to real data from the 1998 Pollino (Italy) and the 2010 Kraljevo (Serbia) earthquakes, as well as to simulated data for a fictitious event in Zurich (Switzerland).

2 Statistical Learning with Gaussian Process Models

A Gaussian Process (GP) model is a versatile tool that performs regression and classification under the Bayesian paradigm (Rasmussen and Williams, 2006; Kuss, 2006). GP models assume a directional dependency between a d -dimensional input vector (or covariate) $\mathbf{x}$ from some domain, $\mathcal{X}$, and the corresponding observable scalar output (or response), y . Assuming this dependency to be separable into a systematic and a random component, the systematic dependency is given by a latent function, $f : \mathcal{X} \rightarrow \mathbb{R}$, such that the likelihood of the output takes the form $p(y|f(\mathbf{x}), \boldsymbol{\vartheta})$. We use $\boldsymbol{\vartheta}$ to denote additional parameters of the likelihood. The objective of GP regression and classification is to infer knowledge about the function f from training data and prior beliefs. After performing inference, the updated function produces probabilistic predictions on the unobserved output at the target inputs.

Consider a training set of empirical pairwise observations $\mathcal{D} = \{\mathbf{x}_i, y_i\}_{i=1}^m = (\mathbf{X}, \mathbf{y})$, where $\mathbf{y}$ is a vector of $m \times 1$ outputs and $\mathbf{X}$ collects the d -dimensional inputs in a matrix of size $m \times d$. Besides the training data, inference about f also involves formalizing the prior belief about the latent function. Instead of assuming a fixed parametric form of f with a finite number of parameters $\boldsymbol{\psi}$ and making inference about $\boldsymbol{\psi}$ only², a GP serves as prior on the latent function

$$f \sim \mathcal{GP}(m(\mathbf{x}, \boldsymbol{\theta}_m), k(\mathbf{x}, \mathbf{x}', \boldsymbol{\theta}_k)) , \quad (1)$$

where $m(\cdot)$ and $k(\cdot)$ denote the mean and positive definite covariance function with parameters $\boldsymbol{\theta}_m$ and $\boldsymbol{\theta}_k$, respectively. Many potential mean and covariance functions exist and the reader is referred to Rasmussen and Williams (2006) for details. The introductory example, introduced in Section 2.1, uses two popular examples, whereas Section 3 discusses $m(\cdot)$ and $k(\cdot)$ in the context of earthquake risk models.

Technically, a GP is a collection of random variables, any finite subset of which follows a joint multivariate normal distribution. A GP prior on f means that each input position $\mathbf{x} \in \mathcal{X}$ has an associated random variable $f(\mathbf{x})$, and that the prior joint distribution for a collection of function values $\mathbf{f} = [f(\mathbf{x}_1), \dots, f(\mathbf{x}_k)]^\top$ associated with an arbitrary set of inputs $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_k]^\top \subset \mathcal{X}$ is multivariate normal $p(\mathbf{f}|\mathbf{X}, \boldsymbol{\theta}) = \mathcal{N}(\mathbf{m}, \mathbf{K})$. The mean vector $\mathbf{m} \in \mathbb{R}^{k \times 1}$ has entries $[\mathbf{m}]_i = m(\mathbf{x}_i, \boldsymbol{\theta}_m)$ and covariance matrix $\mathbf{K} \in \mathbb{R}^{n \times n}$ has entries $[\mathbf{K}]_{ij} = k(\mathbf{x}_i, \mathbf{x}_j, \boldsymbol{\theta}_k)$.

For the inference process, the likelihood of y is required to depend on f only through $f(\mathbf{x})$. As a result the likelihood of f given $\mathcal{D}$ factorises $p(\mathbf{y}|f, \boldsymbol{\vartheta}) = \prod_{i=1}^n p(y_i|f(\mathbf{x}_i), \boldsymbol{\vartheta}) = p(\mathbf{y}|\mathbf{f}, \boldsymbol{\vartheta})$. The posterior distribution of the function values $\mathbf{f}$ is then computed through Bayes' rule:

$$p(\mathbf{f}|\mathcal{D}, \boldsymbol{\theta}, \boldsymbol{\vartheta}) = \frac{p(\mathbf{y}|\mathbf{f}, \boldsymbol{\vartheta})p(\mathbf{f}|\mathbf{X}, \boldsymbol{\theta})}{p(\mathcal{D}|\boldsymbol{\theta}, \boldsymbol{\vartheta})} , \quad (2)$$

where $p(\mathcal{D}|\boldsymbol{\theta}, \boldsymbol{\vartheta}) = \int p(\mathbf{y}|\mathbf{f}, \boldsymbol{\vartheta})p(\mathbf{f}|\boldsymbol{\theta})d\mathbf{f}$ is the marginal likelihood or model evidence. Then, the posterior predictive distribution of function values $\mathbf{f}_*$ at desired target positions $\mathbf{X}_*$ is derived by marginalizing the function variables over the available training inputs:

$$p(\mathbf{f}_*|\mathcal{D}, \boldsymbol{\theta}, \boldsymbol{\vartheta}) = \int p(\mathbf{f}_*|\mathbf{f}, \boldsymbol{\theta})p(\mathbf{f}|\mathcal{D}, \boldsymbol{\theta}, \boldsymbol{\vartheta})d\mathbf{f} . \quad (3)$$

The desired probabilistic predictions for the target outputs is derived by taking the expectation $p(\mathbf{y}_*|\mathcal{D}, \boldsymbol{\theta}, \boldsymbol{\vartheta}) = \int p(\mathbf{y}_*|\mathbf{f}_*, \boldsymbol{\vartheta})p(\mathbf{f}_*|\mathcal{D}, \boldsymbol{\theta}, \boldsymbol{\vartheta})d\mathbf{f}_*$. Note that we omit the dependence on inputs, $\mathbf{X}$, in the definitions above to lighten the notation. The analytical tractability of posterior distributions depends on the type of the output quantity of interest and assumptions about the likelihood $p(y|f(\mathbf{x}), \boldsymbol{\vartheta})$. The following section 2.1 serves as an introductory example, where we use a Gaussian likelihood. Section 2.2 specifies the likelihood for non-Gaussian cases, i.e. ordinal and nominal categorical data. Those are of special interest for earthquake-risk applications, because they describe building damage categories and typological building classes (see Section 3). Subsequently, we present approximate inference in cases with such non-Gaussian mappings.

2.1 Introductory Example: Gaussian Likelihood

For continuous outputs $y \in \mathbb{R}$ and under the assumption of independent normally distributed noise we can write the likelihood as $p(y|f(\mathbf{x}), \boldsymbol{\vartheta}) = \mathcal{N}(f(\mathbf{x}), \sigma_0^2)$. The noise variance σ_0^2 is thus an additional parameter of the likelihood $\boldsymbol{\vartheta} = \sigma_0^2$. Bayesian inference of the latent function f in this model is analytically tractable and the posterior process

²An example is the linear model $f(\mathbf{x}, \boldsymbol{\psi}) = \mathbf{x}^\top \boldsymbol{\psi}$, where prior uncertainty about f is usually expressed in terms of a prior distribution on $\boldsymbol{\psi}$.

$f|\mathcal{D}$ is again a GP. Thus, the posterior of the function values at the input points of the training set $\mathbf{X}$ also follows a multivariate normal distribution $p(\mathbf{f}|\mathcal{D}, \boldsymbol{\theta}, \boldsymbol{\vartheta}) = \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where the parameters are computed as

$$\boldsymbol{\mu} = \mathbf{m} + \mathbf{K}(\mathbf{K} + \sigma_0^2 \mathbf{I})^{-1}(\mathbf{y} - \mathbf{m}) \quad (4a)$$

$$\boldsymbol{\Sigma} = (\mathbf{K}^{-1} + \sigma_0^{-2} \mathbf{I})^{-1}, \quad (4b)$$

where $\mathbf{I}$ is the identity matrix with identical dimensions to $\mathbf{K}$. The posterior predictive distribution for an arbitrary set of target locations $\mathbf{X}_*$ also follows a multivariate normal distribution $p(\mathbf{f}_*|\mathcal{D}, \boldsymbol{\theta}, \boldsymbol{\vartheta}) = \mathcal{N}(\boldsymbol{\mu}_*, \boldsymbol{\Sigma}_{**})$. By evaluating the covariance matrix between target and training inputs, $\mathbf{K}_*$, and between the target inputs themselves, $\mathbf{K}_{**}$, we can compute the posterior mean and covariance as

$$\boldsymbol{\mu}_* = \mathbf{m}_* + \mathbf{K}_*(\mathbf{K} + \sigma_0^2 \mathbf{I})^{-1}(\mathbf{y} - \mathbf{m}) \quad (5a)$$

$$\boldsymbol{\Sigma}_{**} = \mathbf{K}_{**} - \mathbf{K}_*(\mathbf{K} + \sigma_0^2 \mathbf{I})^{-1} \mathbf{K}_*^\top. \quad (5b)$$

Finally, the predictive distribution of outputs $\mathbf{y}_*$ is also multivariate normal with identical mean but with the addition of the noise variance σ_0^2 to the diagonal elements of $\boldsymbol{\Sigma}_{**}$, e.g. $p(\mathbf{y}_*|\mathcal{D}, \boldsymbol{\theta}, \boldsymbol{\vartheta}) = \mathcal{N}(\boldsymbol{\mu}_*, \boldsymbol{\Sigma}_{**} + \sigma_0^2 \mathbf{I})$.

To illustrate GP regression, we use a one-dimensional toy example, where the prior GP is specified using the popular squared exponential covariance function defined as $k(\mathbf{x}, \mathbf{x}', \boldsymbol{\theta}_k) = \sigma^2 \exp(-\|\mathbf{x} - \mathbf{x}'\|_2^2 / 2\ell^2)$, where σ^2 is the variance, ℓ the lengthscale, and $\|\cdot\|_2$ is the euclidean distance measure. The variance controls the amplitude and the lengthscale controls the range of correlation: lower lengthscale values imply a correlation that decreases faster with the distance between input points. Considering the introductory character of this example, a zero mean function ($m(\mathbf{x}) = 0$) is chosen and thus, the hyper-parameters of the prior GP are limited to $\boldsymbol{\theta} = (\sigma^2, \ell)$.

The dataset $\mathcal{D}$ consists of $n = 40$ datapoints, at which y is sampled from the function $\sin(0.5x) + 0.5\cos(0.2x)$ polluted with a white noise $\epsilon \sim \mathcal{N}(0, \sqrt{0.1})$. Input positions, x , are uniformly sampled in two separate intervals. A first model mimics the situation where the hyper-parameters are fixed to 1, 7 and 0.15 for the GP variance σ^2 , lengthscale ℓ , and noise variance σ_0^2 , respectively. Figure 1a shows the posterior predictive distribution over the input space $x_* \in [-20, 20]$. The resulting posterior function, whose density is shown as shaded area, gives a poor fit to the generating function, represented as solid line. The fit is especially poor in regions that are far from available data points. Figure 1c plots the posterior predictive in the output space, e.g. $p(y_*|\mathcal{D}, \boldsymbol{\theta}, \boldsymbol{\vartheta})$, with the mean function identical to Figure 1a above, but with the variance being increased by the noise σ_0^2 .

The choice of appropriate hyper-parameters, $\boldsymbol{\theta}$ and $\boldsymbol{\vartheta}$, prior to the inference process is often challenging, especially in situations with limited prior domain knowledge. In a full Bayesian setting hyper-prior distributions $p(\boldsymbol{\theta})$ and $p(\boldsymbol{\vartheta})$ are assigned and inference is performed over these parameters to update the joint posterior of the latent function and the hyper-parameters simultaneously. This approach is analytically intractable and thus, requires an approximation via a sampling approach, such as Markov-chain Monte Carlo. In this study, we use the maximum a-posteriori (MAP) estimates of the hyper-parameters, e.g. the mode of their posterior distributions $p(\boldsymbol{\theta}, \boldsymbol{\vartheta}|\mathcal{D})$, via solving the optimization problem

$$(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\vartheta}}) = \underset{\boldsymbol{\theta}, \boldsymbol{\vartheta}}{\operatorname{argmax}} \log p(\mathcal{D}|\boldsymbol{\theta}, \boldsymbol{\vartheta}) + \log p(\boldsymbol{\theta}) + \log p(\boldsymbol{\vartheta}), \quad (6)$$

which is analytically tractable for a Gaussian likelihood and factorizing hyper-prior distributions. In this example we chose log-normal hyper-priors, where the median values are identical to the initial values from above and the standard deviation is 0.8 for ℓ and σ^2 and 0.2 for σ_0^2 .

When using the MAP estimated hyper-parameters the resulting posterior function provides a better fit to the true data generating function (see Figure 1b and d). The improved fit results from the reduction of the estimated lengthscale, to a value of 3, from the initially set value of 7. The smaller lengthscale, ℓ , reduces the overconfidence in regions away from known data points, evidenced by the larger spread in the posterior density of the function f .

GP regression, as presented above, is also widely applied in the geo-statistics field, where it is known as kriging (Matheron, 1973) and naturally focused mostly on two-dimensional input spaces. The method for rapid post-earthquake damage mapping proposed by Loos et al. (2020), first uses the non-spatial inputs in least-squares regression to remove a linear trend in the data and secondly uses the geo-coordinates to update GP models with the remaining residuals. The method proposed by Sheibani and Ou (2020) also relies on GP regression but applies it to the entire input space and not only on the geo-coordinates. Both methods, however, have in common that they rely on a Gaussian likelihood, which explains why they are not directly applicable to categorical data, such as the damage levels assigned during post-earthquake building inspections.

2.2 Non-Gaussian Likelihoods for Categorical Data

If the output quantity of interest is categorical, meaning it only takes values from a finite set of discrete categories, likelihood functions that differ from the Gaussian likelihood, used in Section 2.1, are required. Although the inference

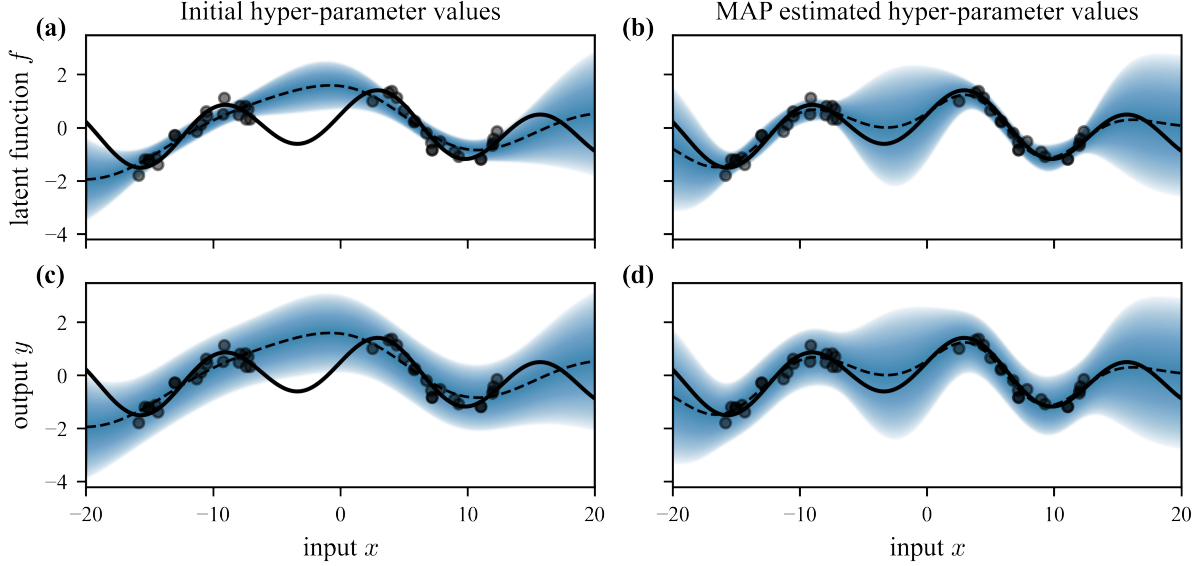


Figure 1: GP regression with independent normal noise. The input data $\mathcal{D}$, indicated with dots, was generated from the function shown as a solid line with additive normal noise. The top row shows the posterior of the latent functions, where the shaded area shows the density $p(f|\mathcal{D}, \theta, \vartheta)$ and the dashed lines indicate the corresponding mean. The bottom row shows the posterior in the output space $p(y|\mathcal{D}, \theta, \vartheta)$. In the left column the parameters θ and ϑ are fixed, whereas for the right column parameters are estimated as their MAP values.

process loses its analytical tractability for categorical data, GP models also apply in these contexts (Hensman et al., 2015; Chu and Ghahramani, 2005). The next two paragraphs contain commonly used likelihoods for ordered categorical output data (ordinal GP regression) and nominal categorical outputs (multi-class GP classification). Finally, a method for approximate Bayesian inference in these non-conjugate cases is presented.

Ordinal GP regression Ordinal data is defined by scalar outputs y_i being elements of a finite set $\mathcal{Y}$ of $c + 1$ ordered categories, which are denoted as integers, $y_i \in \mathcal{Y} = \{0, 1, \dots, c\}$, with preserved ascending ordering information. According to Chu and Ghahramani (2005), GP regression can be rewritten for such ordinal data by employing c parameters $\eta_1 < \dots < \eta_c$ to define the class membership probabilities of label y_i conditional on (latent) function value $f_i = f(\mathbf{x}_i)$:

$$p(y_i|f_i, \vartheta) = \begin{cases} 1 - \Phi[(f_i - \eta_{y_i+1})/\beta], & \text{if } y_i = 0 \\ \Phi[(f_i - \eta_{y_i})/\beta], & \text{if } y_i = c \\ \Phi[(f_i - \eta_{y_i})/\beta] - \Phi[(f_i - \eta_{y_i+1})/\beta], & \text{otherwise} \end{cases} \quad (7)$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function (CDF). The threshold parameters, $\boldsymbol{\eta} = (\eta_1, \dots, \eta_c)$, partition the real line into contiguous intervals, mapping the continuous function f_i into the discrete variable y_i . The dispersion of these threshold parameters is denoted as β . The parameters of the likelihood function thus become $\vartheta = (\boldsymbol{\eta}, \beta)$.

Multi-class GP classification Nominal categorical data, in a similar way to ordinal data, is discrete and finite with possible class outputs, $y_i \in \mathcal{C} = \{0, 1, \dots, c\}$, where c is the number of classes. Unlike ordinal labels, nominal class labels do not provide any ranking information and, instead of having one function f as is the case for ordinal data, we have c independent latent GP functions, i.e. one for each class. The function values for data point $\mathbf{x}_i$ are denoted as $\mathbf{g}_i = (g_i^0, \dots, g_i^c)$, where we denote functions as g to avoid confusion with preceding sections. The conditional class membership probabilities of label y_i are calculated using the softmax function as (Rasmussen and Williams, 2006):

$$p(y_i|\mathbf{g}_i) = \frac{\exp g_i^{y_i}}{\sum_{k=1}^c \exp g_i^k}. \quad (8)$$

If for class k , we write $\mathbf{g}^k = (g_1^k, \dots, g_n^k)$, then $p(\mathbf{g}^k|\mathbf{X})$ follows a multivariate normal distribution and, because the c latent functions are independent, so is $p(\mathbf{g}|\mathbf{X})$, where $\mathbf{g} = (\mathbf{g}^1, \dots, \mathbf{g}^c)$. In this case the likelihood function has no specific parameters $\vartheta = \emptyset$.

Variational Gaussian Approximation Because of the non-Gaussian likelihoods, defined in Eq. 7 and Eq. 8, the posterior distribution, defined in Eq. 2, is not analytically tractable and thus, approximated using the variational Gaussian approximation scheme (Oppor and Archambeau, 2009). The true, but intractable, posterior is approximated with a multivariate normal distribution $q(\mathbf{f})$ that is close to the true posterior $q(\mathbf{f}) = \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \approx p(\mathbf{f}|\mathcal{D}, \boldsymbol{\theta}, \boldsymbol{\vartheta})$. Specifically, we search parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ that minimize the Kullback-Leibler (KL) divergence between the variational distribution $q(\mathbf{f})$ and the true posterior. We can write this KL divergence as

$$\text{KL}(q(\mathbf{f}) \parallel p(\mathbf{f}|\mathcal{D}, \boldsymbol{\theta}, \boldsymbol{\vartheta})) = \log p(\mathcal{D}|\boldsymbol{\theta}, \boldsymbol{\vartheta}) + \mathcal{F}_{\text{vfe}}(q, \boldsymbol{\theta}, \boldsymbol{\vartheta}), \quad (9)$$

where $\mathcal{F}_{\text{vfe}}(\cdot)$ denotes the variational free energy. Because of the non-Gaussian likelihood, the evidence $p(\mathcal{D}|\boldsymbol{\theta}, \boldsymbol{\vartheta})$ is not analytically tractable undermining a direct evaluation of Eq. 9. However, the evidence being a constant term, minimizing Eq. 9 is equivalent to minimizing the variational free energy. The latter is expressed as

$$\mathcal{F}_{\text{vfe}}(q, \boldsymbol{\theta}, \boldsymbol{\vartheta}) = \text{KL}(q(\mathbf{f}) \parallel p(\mathbf{f}|\boldsymbol{\theta})) - \sum_{i=1}^n \left[\int \log(p(y_i|f_i, \boldsymbol{\vartheta})) q(f_i) df_i \right]. \quad (10)$$

The first term in Eq. 10 is analytically tractable because it is the KL-divergence between two multivariate normal distributions, the variational distribution and the prior of f . The second term, also called the expected log likelihood of the data with respect to $q(\mathbf{f})$, and the one-dimensional integrals can, for example, be approximated by means of Gauss-Hermite quadrature. In order to keep the variational free energy at small values, any model should provide a good explanation of the data (via large values for the second term) while not deviating too far from the prior (by keeping the first term small). Negative variational free energy is also called the evidence lower bound (ELBO), because it provides a lower bound to the log evidence. This follows from Eq. 9 and the fact that the KL-divergence measure is always positive. Therefore, $\mathcal{F}_{\text{vfe}}(\cdot)$ is used to learn the parameters of the variational distribution, $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, while in the same time providing approximate MAP-estimates for the hyper-parameters, $\boldsymbol{\theta}$ and $\boldsymbol{\vartheta}$, using the following objective function:

$$(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\vartheta}}) = \underset{\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\theta}, \boldsymbol{\vartheta}}{\text{argmax}} - \mathcal{F}_{\text{vfe}}(q, \boldsymbol{\theta}, \boldsymbol{\vartheta}) + \log p(\boldsymbol{\theta}) + \log p(\boldsymbol{\vartheta}). \quad (11)$$

The application of this objective function to derive MAP estimates for the parameters $\boldsymbol{\vartheta}$ of the ordinal likelihood require a two-fold re-parameterization: First, to enable a parallel estimation of the threshold parameters and their dispersion, we perform an up-front scaling $\tilde{\boldsymbol{\eta}} = \boldsymbol{\eta}/\beta$. Second, to fix the order of the threshold parameters, positive parameters $\Delta_y = \tilde{\eta}_y - \tilde{\eta}_{y-1}$ are introduced for categories $y > 0$. Thus, MAP-estimation is performed for parameters $\boldsymbol{\vartheta} = \{\beta, \tilde{\eta}_0, \Delta_1, \dots, \Delta_c\}$.

Because the approximated posterior distribution, $q(\mathbf{f})$, is multivariate normal, the approximate posterior predictive $q(\mathbf{f}_*|\hat{\boldsymbol{\theta}}) = \int p(\mathbf{f}_*|\mathbf{f}, \hat{\boldsymbol{\theta}})q(\mathbf{f})d\mathbf{f}$ is also multivariate normal and we can evaluate its parameters as

$$\boldsymbol{\mu}_* = \mathbf{m}_* + \mathbf{A}\boldsymbol{\mu} \quad (12a)$$

$$\boldsymbol{\Sigma}_{**} = \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top + \mathbf{B}, \quad (12b)$$

where $\mathbf{A} = \mathbf{K}_*\mathbf{K}^{-1}$ and $\mathbf{B} = \mathbf{K}_{**} - \mathbf{K}_*\mathbf{K}^{-1}\mathbf{K}_*^\top$. In the following section we link the outlined GP theory to regional earthquake risk models and we specify well-suited mean and covariance functions of the different GPs described above.

3 Regional Earthquake Risk Models

Regional risk models, when employed in a post-earthquake context, provide rapid, yet uncertain, damage predictions at regional scale. However, the components of the risk model contain uncertainties, as the geographical distribution of shaking-intensity measures remains uncertain, despite seismic network stations, detailed information about the buildings composing a region are often not available, and the simulation of structural models for all buildings within a region is impossible. Thus, buildings are clustered into several predefined types, for which an average vulnerability is established that often represents buildings at national, if not continental, scale (Martins and Silva, 2020). Typological attribution models, or exposure mapping models, correlate building types with socio-economic indicators and building attributes from public databases, such as building height, value and year of construction (Diana et al., 2019; Crowley et al., 2020), in order to overcome lack of building-specific information. These typological attribution models may suffer from limited applicability for the region hit by an earthquake, further adding to the uncertainties arising from the two other components of loss assessment.

As input to the risk model we assume official and public databases to provide basic building-specific information. For any building, this information is gathered in a d -dimensional vector $\mathbf{x}$, where the individual entries are, for example, the geographic coordinates, the construction year, and the number of stories. Building damage is then modelled as a

consequence of earthquake ground-motion intensity at its location, with fragility functions expressing the probability of reaching a given damage category as a function of ground-motion intensity. Because economic and technical constraints often hinder the pre-earthquake collection of detailed information and the simulation of structural models for all buildings within a region, we employ fragility functions that are defined for building classes expected to behave similarly under earthquake loading. Then, we use typological attribution models to map individual buildings to these pre-defined classes.

This study focuses on the physical damage to residential buildings as prediction output. Two facts motivate this choice: First, most model approaches for other earthquake consequences, such as recovery time and repair costs, use building damage as starting point for subsequent calculations (Burton et al., 2019). Second, empirical observations of building damage provide one of the first building-specific data sources to constrain estimates for other impacts after an earthquake occurs.

3.1 Ground-Motion Intensity Measures

Predicting earthquake-induced damage to a spatially distributed set of buildings, involves the following state-of-the-art assumptions: (i) Conditional on a specific scenario event, the logarithm of the ground-motion intensity measure (IM) at dispersed building sites follow a multivariate normal distribution; (ii) Conditional on IM, damage to individual buildings is independent; (iii) Building capacities are given in terms of the value of IM at which the building exceeds pre-defined damage thresholds; (iv) For the purpose of the present study, we assume that the capacities of all buildings are expressed with respect to the same IM.

Let im_i denote the ground-motion IM at site location $\mathbf{x}_i$. The characteristics of the corresponding event, such as its magnitude, source geometry and style of faulting, are collected in vector $\mathbf{e}$. To estimate im_i , we employ empirical and ergodic ground-motion models (GMMs) of the form

$$\ln im_i = m(\mathbf{x}_i, \mathbf{e}, \boldsymbol{\theta}_m) + \delta_{W,i} + \delta_B, \quad (13)$$

where $m(\cdot)$ denotes the GMM trend function with parameters $\boldsymbol{\theta}_m$ that predicts the median IM conditional on the magnitude, the style of faulting, the source-to-site distance, and local soil conditions. The terms $\delta_{W,i}$ and δ_B denote the within-event and between-event residuals that are both normally distributed with zero mean and variances σ_W^2 and σ_B^2 , respectively. Standard GMMs provide the parameters of Eq. 13 based on mixed-effects regression on ground-motion recordings from historic earthquakes. The between-event residual indicates a systematic deviation from the trend function for a given event, whereas the remaining deviations are summarized in the within-event residual. Thus, for a specific event, δ_B is identical for two sites $\mathbf{x}_i$ and $\mathbf{x}_j$, whereas within-event residuals, $\delta_{W,i}$ and $\delta_{W,j}$ differ.

Past studies suggest that the latter are spatially dependent and that their joint distribution is multivariate normal (Boore et al., 2003; Wang and Takada, 2005; Goda and Hong, 2008; Jayaram and Baker, 2008). Most empirical evaluation studies suggest that the spatial correlation, ρ , decreases exponentially with the euclidean distance between two sites:

$$\rho(\mathbf{x}_i, \mathbf{x}_j, \ell) = \exp \left(\frac{-\|\mathbf{x}_i^{(1:2)} - \mathbf{x}_j^{(1:2)}\|_2}{\ell} \right), \quad (14)$$

where ℓ denotes the lengthscale. For illustration, the distance at which ρ is lower than 5% is approximately 3ℓ . The superscript (1 : 2) indicates that the correlation function acts on the geographical coordinates (i.e. Easting and Northing) of the input vectors, here assumed to be stored in dimensions one and two.

All random variables in Eq. 13 are Gaussian and the correlation function in Eq. 14 leads to positive definite covariances. Therefore, the random field of logarithmic ground-motion IMs, $f = \ln im$, follows a GP as defined in Eq. 1. The mean function is identical with the trend function of the GMM $m(\mathbf{x}_i, \mathbf{e}, \boldsymbol{\theta}_m)$ and the covariance function is $k(\mathbf{x}, \mathbf{x}', \boldsymbol{\theta}_k) = \sigma_W^2 \rho(\mathbf{x}, \mathbf{x}', \ell) + \sigma_B^2$. For a finite set of buildings with inputs $\mathbf{X}$ and event characteristics $\mathbf{e}$ the joint distribution $p(\mathbf{f}|\mathbf{X}, \mathbf{e}, \boldsymbol{\theta}_f)$ is multivariate normal. The hyper-parameters $\boldsymbol{\theta}_f$ of this GP model are the GMM parameters ($\boldsymbol{\theta}_m, \sigma_W^2, \sigma_B^2$) and the lengthscale ℓ of the spatial correlation model.

The above formulation can be extended to cases involving multiple IMs. In such a case, each IM is represented as a separate GP with different mean and covariance functions. The correlation between the GPs can then be modelled using, for example, a linear model of co-regionalization.

3.2 Building Damage

We denote the damage to a building i as a discrete random variable Y_i , with sequential, collectively exhaustive and mutually exclusive categories $y \in \mathcal{Y} = \{0, 1, \dots, c_y\}$, where $y = 0$ indicates no damage. A building enters a certain

damage category $y > 0$, if the ground-motion IM at the building location exceeds its corresponding capacity threshold. In compliance with the state of the art, we assume these thresholds to be log-normally distributed with location parameters $\eta_1 < \eta_2 < \dots < \eta_{c_y}$ and a common dispersion parameter β . Based on the logarithm of IM, here denoted as f_i , the probability of entering a damage state $y > 0$ is derived as

$$P(Y_i \geq y | f_i, \boldsymbol{\vartheta}) = \Phi \left(\frac{f_i - \eta_y}{\beta} \right), \quad (15)$$

where $\boldsymbol{\vartheta}$ collects the parameters of the log-normal distributions. The expression in Eq. 15 is called a fragility function. The dispersion parameter β is chosen to be the same for all damage categories to prevent negative probability masses or, in other words, a crossing of the fragility functions. Given the similarity with the likelihood employed in case of ordinal GP regression, as defined in Eq. 7, inference using damage data from a small number of inspected buildings can be performed.

The approaches to estimate the parameters $\boldsymbol{\vartheta}$ for existing building stocks differ in whether they are calibrated with numerical models of buildings, damage data from past earthquake events or a combination of both (D'Ayala et al., 2015; Lagomarsino and Giovinazzi, 2006; Rosti et al., 2021a). Here we assume that $\boldsymbol{\vartheta}$ have been derived for a finite set of fragility classes $\mathcal{B} = \{1, 2, \dots, c_b\}$, such as low-rise unreinforced masonry (URM) buildings with flexible floors and mid-rise reinforced concrete (RC) shear wall buildings. Ideally, the inputs $\mathbf{x}_i$ would contain the corresponding fragility class $b_i \in \mathcal{B}$ for each building. In reality, however, this is rarely the case and some attributes are uncertain, if not unknown, for large parts of the buildings within the region. For example, the number of floors could be available from municipal databases or estimated from digital elevation models to separate low- and mid-rise buildings. Acquiring input data on the lateral load resisting system often requires detailed and time-consuming field surveys. To overcome missing data, typological attribution models are used to estimate unknown attributes from known inputs.

3.3 Typological Attribution Models

Typological attribution, in the present context, describes the process of deriving unknown taxonomic attributes of buildings from known information. The models, discussed in this section, assume that the input data $\mathbf{x}_i$ contains for each building i the construction period or year and the number of floors or an approximate height indication. The lateral load resisting system (LLRS) is unavailable, as in most real-world scenarios. As discussed in Section 3.2, the required attributes depend on the fragility classes $\mathcal{B}$.

Given the information in $\mathbf{x}_i$, the randomness pertains to the LLRS, which we model with a discrete random variable A_i with c_a possible attribute combinations $\mathcal{A} = \{1, 2, \dots, c_a\}$. For example $A_i = 1$ indicates an URM building with flexible floors and $A_i = 2$ indicates an URM building with stiff floors. A typological attribution model provides class membership probabilities for A_i as a function of known input data $\mathbf{x}_i$, which allows to generate samples a_i for all buildings and we can attribute a fragility class $b = h(\mathbf{x}, a)$ using a deterministic function $h : \mathcal{X} \cup \mathcal{A} \rightarrow \mathcal{B}$, and consequently gather the corresponding parameters of the fragility function as described above. As an example, the deterministic function $h(\cdot)$ may simply add the height class (i.e. low-rise) of a building, stored in $\mathbf{x}_i$, to the sampled unknown attribute combination a_i .

In the absence of region-specific data, an attribution model that is based on expert-opinion presents the best alternative. Experts may provide insights into the proportions of building classes characterizing different construction periods for instance in the form of a table saying that $x\%$ of buildings built after 1988 are expected to be reinforced concrete shear wall buildings. This might be a valid average over entire countries or states, however, construction practices vary from region to region, thus leading to systematic local deviations from the average. In addition, unknown attribute combinations of geographically close buildings of similar age and geometry might be correlated.

In Bodenmann et al. (2021), we proposed the use of latent functions to account for such variability and to incorporate typological information gathered during post-earthquake inspections. In analogy to the multi-class GP classification scheme presented in Section 2.2, we employ c_a independent latent functions $\mathbf{g} = [g^{(1)}, \dots, g^{(c_a)}]$ and impose a GP prior on them. The class membership probabilities conditional on these functions are given by Eq. 8. The prior mean function $u(\cdot)$ is calibrated with an expert-opinion based proportion estimate. For the covariance function, we combine multiple squared-exponential covariance functions to account for spatially correlated deviations over longer scales, due to regional differences in construction practices, v_{LS} , and short-scale correlated deviations for classes of close buildings that share similar age and number of stories v_{SS} and the final covariance function is denoted as

$$v(\cdot) = v_{LS} + v_{SS,1} \cdot v_{SS,2} \cdot v_{SS,3}, \quad (16)$$

where the individual functions of $v_{SS}(\cdot)$ measure similarity with respect to geo-coordinates, number of floors and construction year. The multiplicative structure limits high correlation to buildings with all three individual functions showing high similarity. Detailed examples of typological attribution models are shown in the three case-studies, presented in Section 5 and the corresponding appendices.

3.4 Simulation workflow

The damage to residential buildings $\mathbf{X}$ resulting from a specified earthquake scenario $\mathbf{e}$ is estimated using the following framework: First, we draw r samples from the multivariate normal distributions $p(\mathbf{f}|\mathbf{e}, \mathbf{X})$ (ground-motion intensity) and $p(\mathbf{g}|\mathbf{X})$ (typological attribution). Then, for each sample and for each building i , we first sample unknown attributes a_i from the categorical distribution $p(A_i|\mathbf{g}_i)$, with probabilities given by Eq. 8 and subsequently sample a damage category y_i from the categorical distribution $p(Y_i|f_i, a_i)$, with probabilities given by Eq. 7. Finally, for each sample j buildings with identical level of damage are counted and the values stored in a vector $\mathbf{n}_j = \{n_{yj}|y \in \mathcal{Y}\}$. Thus, we obtain r samples from $p(\mathbf{N}|\mathbf{e})$, which denotes the joint predictive distribution of how the exposed buildings are partitioned amongst the different damage levels. The marginal predictive distributions $p(N_y|\mathbf{e})$, i.e. of the entries of random vector $\mathbf{N}$, denote the predicted number of buildings in a certain damage level y .

In case of early post-earthquake assessments, the available information that is measured by seismic networks or gathered from field surveys, should be incorporated to allow for more constrained estimates.

4 Model Updating with Post-Earthquake Data

The following sections explain the inference process for ground-motion measurement data from seismic stations, and for damage and typological data from building inspections. Figure 2 schematically illustrates the proposed framework. The first rapid damage estimates are obtained with the regional risk model, where estimates of ground-motion IMs are constrained by seismic recordings in analogy to current shake map systems. Then, we use data from the first inspected buildings to further constrain the shake map and, in parallel, update the fragility function parameters, as well as the typological attribution model.

We use the subscripts $\mathcal{S}, \mathcal{I}$ and $\mathcal{T}$ to denote index sets related to the seismic stations, the inspected buildings and the target buildings for which we aim to predict the damage category. In the immediate aftermath of an event, the model predicts the damage to buildings with indices $\mathcal{T} = \{1, 2, \dots, n\}$. Indices of buildings inspected after t time steps are denoted as $\mathcal{I}_t \subset \mathcal{T}$ and the indices of the remaining not (yet) inspected buildings are denoted as $\mathcal{T}_t = \mathcal{T} \setminus \mathcal{I}_t$.

4.1 Seismic recordings

We denote the dataset of recorded ground-motion intensity measures from a seismic network as $\mathcal{D}_\mathcal{S} = \{\mathbf{x}_i, z_i | i \in \mathcal{S}\}$, where z_i are noise free measurements of the ground-motion intensity ($\ln z_i = f_i$). This section explains how to constrain the predictions of ground-motion intensity f using seismic recordings, where we closely follow Worden et al. (2018). In a first step we condition the distribution of the (constant) inter-event residual on the seismic recordings by deriving its posterior distribution $p(\delta_\mathbf{B}|\mathcal{D}_\mathcal{S}) = \mathcal{N}(\xi_\mathbf{B}, \psi_\mathbf{B})$, where the parameters are calculated as

$$\psi_\mathbf{B}^2 = \left(\frac{1}{\sigma_\mathbf{B}^2} + \frac{\mathbf{1}^\top \mathbf{C}_{\mathcal{S}\mathcal{S}}^{-1} \mathbf{1}}{\sigma_\mathbf{W}^2} \right)^{-1}, \quad (17a)$$

$$\xi_\mathbf{B} = \frac{\psi_\mathbf{B}^2}{\sigma_\mathbf{W}^2} (\mathbf{1}^\top \mathbf{C}_{\mathcal{S}\mathcal{S}}^{-1} (\ln \mathbf{z} - \mathbf{m}_\mathcal{S})) , \quad (17b)$$

with $\mathbf{C}_{\mathcal{S}\mathcal{S}}$ denoting the correlation matrix of the within-event residuals with entries $[\mathbf{C}_{\mathcal{S}\mathcal{S}}]_{ij} = \rho(\mathbf{x}_{\mathcal{S}i}, \mathbf{x}_{\mathcal{S}j}, \ell)$. Then we evaluate the posterior predictive distribution of f evaluated at the target inputs $\mathbf{X}_\mathcal{T} = \{\mathbf{x}_i | i \in \mathcal{T}\}$, which is a multivariate normal distribution, e.g. $p(\mathbf{f}_\mathcal{T}|\mathcal{D}_\mathcal{S}) = \mathcal{N}(\boldsymbol{\nu}_\mathcal{T}, \boldsymbol{\Psi}_{\mathcal{T}\mathcal{T}})$. For this calculation the posterior mean of the between-event residual $\xi_\mathbf{B}$ is added to the prior GP mean function as a constant term, and the remaining variance of the between-event residual $\psi_\mathbf{B}^2$ is added to the prior variance of the within-event residual $\sigma_\mathbf{W}^2$. The mean vector $\boldsymbol{\nu}_\mathcal{T}$ and covariance matrix $\boldsymbol{\Psi}_{\mathcal{T}\mathcal{T}}$ of the posterior predictive distribution are then computed as

$$\boldsymbol{\nu}_\mathcal{T} = \mathbf{m}_\mathcal{T} + \xi_\mathbf{B} + \mathbf{C}_{\mathcal{T}\mathcal{S}} \mathbf{C}_{\mathcal{S}\mathcal{S}}^{-1} (\mathbf{z} - \mathbf{m}_\mathcal{S} - \xi_\mathbf{B}), \quad (18a)$$

$$\boldsymbol{\Psi}_{\mathcal{T}\mathcal{T}} = (\sigma_\mathbf{W}^2 + \psi_\mathbf{B}^2) (\mathbf{C}_{\mathcal{T}\mathcal{T}} - \mathbf{C}_{\mathcal{T}\mathcal{S}} \mathbf{C}_{\mathcal{S}\mathcal{S}}^{-1} \mathbf{C}_{\mathcal{S}\mathcal{T}}) , \quad (18b)$$

where matrices $\mathbf{C}_{\mathcal{T}\mathcal{S}}$ and $\mathbf{C}_{\mathcal{T}\mathcal{T}}$ denote the correlation matrices evaluated between the target points and seismic stations and between the target points themselves. To generate correlated realizations of a damage level for the target buildings $\mathbf{X}_\mathcal{T}$, we follow the workflow described in Section 3.4, where we sample from the posterior $p(\mathbf{f}_\mathcal{T}|\mathcal{D}_\mathcal{S})$ instead of the prior $p(\mathbf{f}_\mathcal{T})$.

4.2 Inspection data

During inspection of a certain building i experts usually provide an estimate of the inflicted damage y_i via attributing a certain damage category. Here we assume that those are consistent with the damage categories $\mathcal{Y}$ employed in the

fragility functions of the risk model, e.g. $y_i \in \mathcal{Y}$. Besides this damage description, experts are also asked to provide some basic building attributes, e.g. the dominating material of the lateral load-resisting system and its type. Again we take the assumption that this information allows the attribution of a typological combination $a_i \in \mathcal{A}$ employed in the risk model. We denote the dataset from building inspections conducted up to a given time step t after the event as $\mathcal{D}_{\mathcal{I}t} = \{\mathbf{x}_i, y_i, a_i | i \in \mathcal{I}_t\}$, where we omit the subscript t in the following for better readability. This data is used to perform two separate inference steps: First, we perform ordinal GP regression using the function f , where we take also into account the seismic recordings $\mathcal{D}_{\mathcal{S}}$. Second, we perform multi-class GP classification using the functions $\mathbf{g}$, related to the unknown typological attributes.

First, we use a_i to allocate a certain fragility class $b_i = h(\mathbf{x}_i, a_i)$ for each of the inspected buildings. Then we use the data y_i and the likelihood for ordinal data (see Eq. 7) to perform variational inference on f . Specifically, we approximate the true posterior $p(\mathbf{f}_{\mathcal{I}} | \mathcal{D}_{\mathcal{S}}, \mathcal{D}_{\mathcal{I}}) \propto p(\mathbf{y} | \mathbf{f}_{\mathcal{I}}) p(\mathbf{f}_{\mathcal{I}} | \mathcal{D}_{\mathcal{S}})$ with a variational multivariate normal distribution $q(\mathbf{f}_{\mathcal{I}})$. Note that the posterior is conditioned on both, the inspection data $\mathcal{D}_{\mathcal{I}}$ and seismic recordings $\mathcal{D}_{\mathcal{S}}$. To evaluate the variational free energy in Eq. 10, we replace the prior $p(\mathbf{f}_{\mathcal{I}})$ with the posterior predictive $p(\mathbf{f}_{\mathcal{I}} | \mathcal{D}_{\mathcal{S}})$ from Section 4.1. As outlined in Section 2.2, the objective function of this variational inference step (given by Eq. 11) allows for a simultaneous MAP estimation of the model hyper-parameters. We exploit this capacity in order to update the fragility functions. Specifically, we calculate MAP estimates of the damage threshold parameters $\hat{\eta}_b$ for all fragility classes $b \in \mathcal{B}$. Given the limited amount of available building inspections in the early aftermath of an earthquake event, we keep the dispersion parameters (β in Eq. 15), and other hyper-parameters, such as the empirical GMM parameters, fixed at the values specified in the risk model.

In a second step we perform approximate inference on the latent functions $\mathbf{g}$ using the data a_i via standard variational inference as described in Section 2.2, e.g. the true posterior $p(\mathbf{g}_{\mathcal{I}} | \mathcal{D}_{\mathcal{I}})$ is approximated by a multivariate normal distribution $q(\mathbf{g}_{\mathcal{I}})$. We keep the parameters of the mean functions fixed to the initial expert-judgment based estimates and perform MAP estimation for the variances and lengthscales of the covariance function specified in Eq. 16.

To predict damage for a set of not (yet) inspected target buildings $\mathbf{X}_{\mathcal{T}t}$, we follow the procedure outlined in Section 3.4 with two exceptions: First, we sample from the (approximate) posterior predictive distributions $q(\mathbf{f}_{\mathcal{T}})$ and $q(\mathbf{g}_{\mathcal{T}})$. The parameters of these multivariate normal distributions are calculated using Eq. 12. To take into account the seismic recordings in case of function f , we use the mean vector $\boldsymbol{\nu}_{\mathcal{T}}$ and covariance matrices $\boldsymbol{\Psi}_{\mathcal{T}\mathcal{T}}$ and $\boldsymbol{\Psi}_{\mathcal{T}\mathcal{I}}$ to evaluate matrices $\mathbf{A}$ and $\mathbf{B}$ in Eq. 12. Second, to sample damage categories we employ the updated fragility functions $p(Y | f, a, \boldsymbol{\vartheta})$, using the inferred MAP-estimates of the threshold parameters. The inference schema described above is numerically implemented in Python using the software library GPFlow (Matthews et al., 2017)³.

5 Application to three case-studies

Three case studies, one simulated earthquake scenario in the Swiss Canton of Zurich and two real-world case-studies, namely the 1998 Pollino (Italy) and the 2010 Kraljevo (Serbia) earthquake events, are used to apply and evaluate the proposed framework. Specifically, we seek to predict the spatial pattern of earthquake damage as the number of buildings within sub-regions (i.e. zip-codes or municipalities) being attributed to a given damage level. Table 1 summarizes key characteristics of the three examples, while Figure 3 shows the location of the epicenter, the boundaries to the considered sub-regions, and the locations of seismic network stations, if available. Whereas the earthquake events share similar magnitudes, the three case studies strongly differ in terms of the size of the considered region, the quantity of exposed buildings and the coverage of the seismic network. The Zurich case represents a densely populated region that is equipped with four seismic network stations. The Pollino case-study covers the largest area but has a relatively low building density. Finally, the Kraljevo case-study covers a small region with few buildings and no seismic network stations.

Table 1: Characteristics of the earthquake event and the considered region of the three case studies.

	Zurich (simulated)	Pollino (1998)	Kraljevo (2010)
Magnitude M_w	5.8	5.6	5.5
Number of buildings	33594	20528	1959
Area size [km ²]	97.1	931.1	25.4
Number of subregions	22	14	3

Four levels indicate the severity of earthquake damage to residential buildings: No, slight, moderate, and extensive damage. The four damage levels are encoded as increasing integers $\mathcal{Y} = \{0, 1, 2, 3\}$. We mimic a post-event

³The code will be shared on gitlab once ready for submission

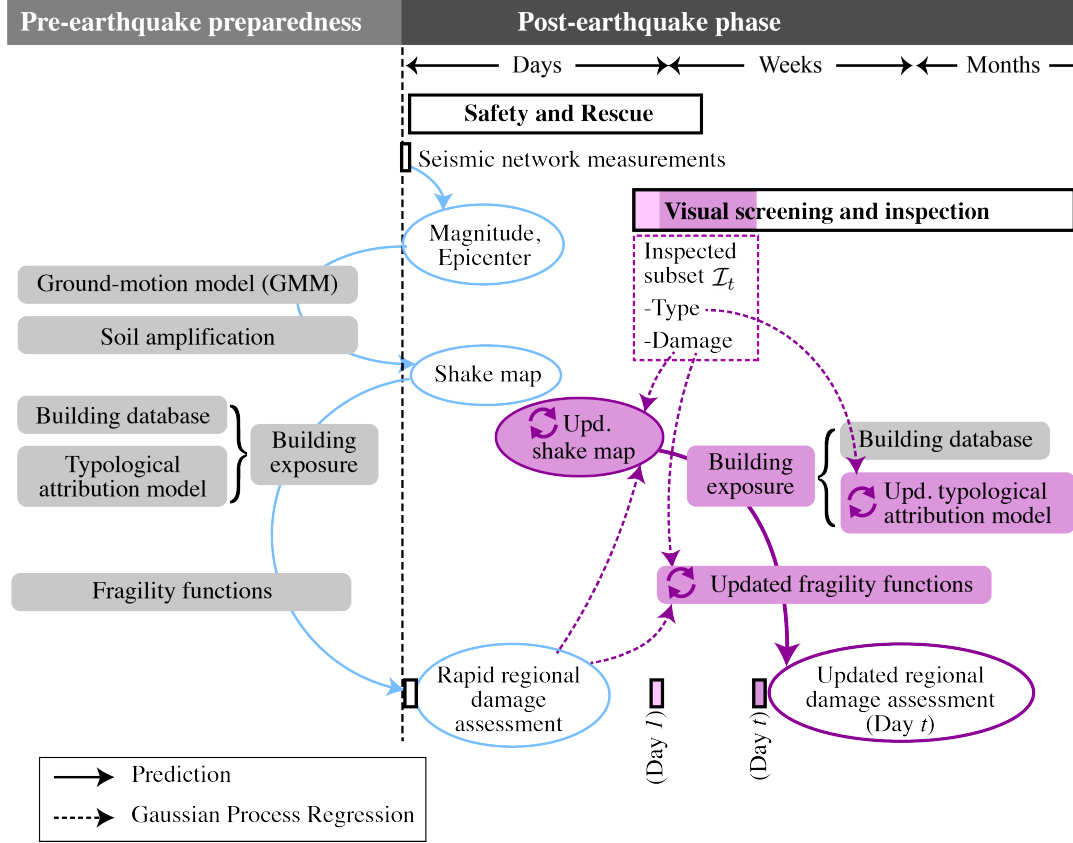


Figure 2: Schematic overview of the post-earthquake implementation of regional risk models and their updating using inspection data.

application by providing first predictions in the immediate aftermath ($t = 0$) with shake maps, obtained following the process described in Section 4.1, that are only constrained using ground-motion recordings $\mathcal{D}_S$. Subsequently, gradually increasing amounts of data from building inspections $\mathcal{D}_{\mathcal{I}_t}$ enable updating of the risk model components⁴ (see Section 4.2) and thereby, lead to improved impact predictions at later time steps $t > 0$. To assess the performance of the proposed risk-model-informed Gaussian process model (RMGP), we compare, at each time step t and for each subregion k , the predictions $\mathbf{N}_{tk}$ with the actual number of buildings at each damage level on one hand and with predictions obtained using two entirely data-driven methods (presented in Section 5.2) on the other hand.

5.1 Data and Models

For all three case-studies, the available building-specific input data, $\mathbf{X}_{\mathcal{T}}$, consists of the geographical coordinates, the Eurocode 8 soil class (Eurocode 8, 2004), the construction year (or period), and the number of stories (either exact or as a range). The event characteristics $\mathbf{e}$ consist of the earthquake (moment-) magnitude, the coordinates of the epicenter and the style of faulting.

The historic inspection data of the Pollino and Kraljevo events do not contain information regarding the time when a building was inspected. Therefore, the inspection process is simulated, in the same manner than for the simulated Zurich case study, by randomly attributing an initial building to each available inspection team at the beginning of each time step (inspection day). Each inspection team then continues by inspecting the four geographically closest (non-inspected) buildings during this time step before being assigned to the next randomly chosen initial building at the following time step. Whereas this procedure successfully produces a spatial clustering of inspection data, the time trajectory of the entire inspection process is simplified. Therefore, the results are presented as a function of the number of inspected buildings available for inference rather than stating a specific time after the event.

⁴The inference process at each time step, starts from the initial risk model and shake map, and considers data from all buildings that were inspected until this time step.

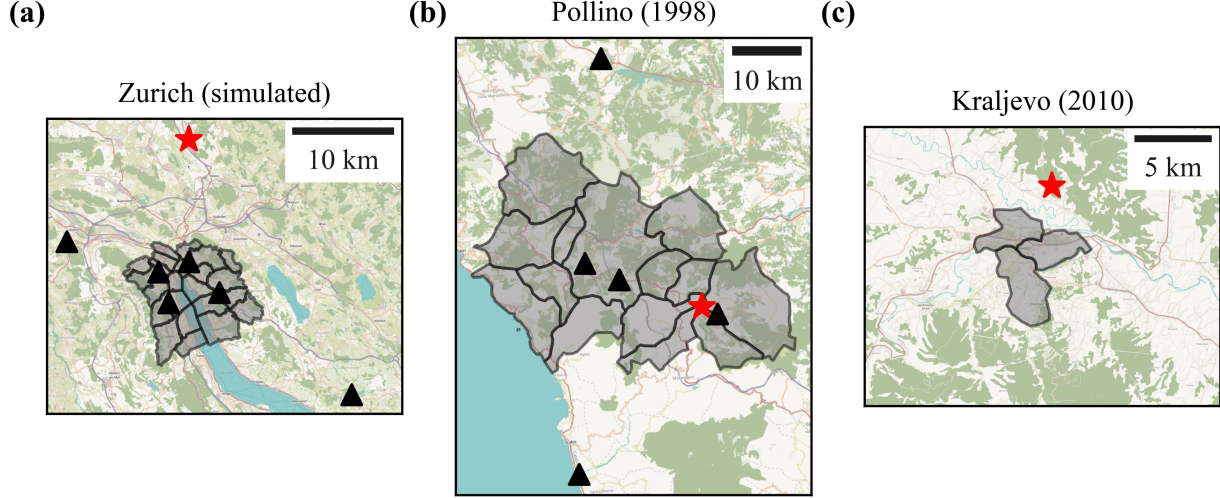


Figure 3: Overview of the three case studies: (a) Simulated M6 scenario in the canton of Zurich (Switzerland), (b) Pollino (Italy) earthquake from 1998 and (c) Kraljevo (Serbia) earthquake from 2010. The grey shaded area delimit the sub-regions of interest, the star indicates the epicenter location, and the triangles indicate stations of a seismic network, if available.

5.1.1 Zurich

This case-study examines damage to approximately 34'000 residential buildings within the Canton of Zurich (Switzerland) inflicted by a scenario earthquake of $M_w = 5.8$ (see Figure 3). The following paragraphs introduce the available input data, the risk model components and the models used to generate the ‘true’ building damage. The electronically available supplementary material A1 provides more details to interested readers.

Available input data Building-specific information $\mathbf{X}_{\mathcal{T}}$ is retrieved from the office for spatial development of the Canton of Zurich (2020). Approximately 75% of the building stock was built before 1980 with 65% of all buildings having between 3 and 5 stories. The locations of the seismic network stations $\mathbf{X}_{\mathcal{T}}$ are fictitious and deliberately chosen to present a dense network within the region of interest. We simplify the entire region to be located on soil class B.

Risk model components The earthquake-risk model, assumed to be available before the scenario earthquake, builds upon the empirical ground-motion model of Akkar and Bommer (2010) and the spatial correlation model of Esposito and Iervolino (2011). The employed fragility functions are defined for the peak ground acceleration (PGA) and specified for 12 classes $b \in \mathcal{B}$, encompassing buildings from low-rise (1-2 stories) URM with flexible floors to high-rise (6 stories or higher) RC shear wall buildings designed according to modern seismic guidelines. Because of the available input information $\mathbf{x}$, the randomness pertains to four unknown typological attribute combinations $a \in \mathcal{A}$, which are URM with either flexible or stiff floors, and RC buildings with either shear walls or infill frames forming as structural system.

Data-generating models The same input information, $\mathbf{X}_{\mathcal{T}}$, as specified above is used to generate the ‘true’ data. Four cross-correlated random fields of PGA and elastic, 5%-damped spectral acceleration at periods of 0.3, 0.6 and 1.0 seconds are generated using the GMM of Chiou and Youngs (2014). The spatial cross-correlation of the within-event residuals is accounted for using the model of Markhvida et al. (2018) and the cross-correlation of the inter-event residuals follows the relations provided by Baker and Cornell (2006). Finally, the fragility functions developed by Martins and Silva (2020) are used to simulate damage for each building. These functions are defined for one of the four aforementioned IMs, depending on the building type and the number of stories. The data-generating model is deliberately chosen to differ from the risk model used for updating to reflect real conditions in which, models never fully capture the real processes leading to regional damage distributions.

While the available building data (coordinates, construction year, number of stories) is deemed realistic, the fragility functions and soil conditions are not. Therefore, neither the ‘true’ simulated results nor the prior risk model predictions reflect real earthquake risk conditions in Zurich.

5.1.2 Pollino

On September 9, 1998, the Pollino region in Italy was struck by an $M_w = 5.6$ earthquake causing damage to many buildings in the surrounding municipalities. This case-study examines approximately 20'000 residential buildings spread over 14 subregions (see Figure 3).

Available input data The event-characteristics and seismic recordings are taken from the engineering strong-motion (ESM) database (Luzi et al., 2020). The main data source for this case study is the publicly available building-damage database Da.D.O. (Dolce et al., 2019), consisting of inspection results for about 13'000 residential buildings in the affected region. This database provides the geographical coordinates, the construction period (from pre-1919 to post-1981) and the number of stories, with the latter categorized into three height classes: Low-rise (1-2 stories), mid-rise (3-4 stories) and high-rise (>5 stories). In cases where multiple entries were lumped at the same coordinates, the European Settlement Map (Corbane and Sabo, 2019) serves as reference to attribute unique coordinates to each building.

The available data does not cover the entire building stock. The Italian National Institute of Statistics (ISTAT, 2001) census data from 2001 provides information on the number of residential buildings from a certain construction period and with a certain number of stories, at municipality level. Based on this information, we augment the damage database with the missing buildings, for which spatial coordinates are generated using the European Settlement Map. The 2001 census data is the temporarily closest information to the 1998 event, yet, changes that occurred in the three years may induce errors that are deemed small and thus, acceptable. Finally, soil classes are attributed according to Forte et al. (2019). The pre-processing steps are described in detail in appendix A2.

Risk model components The earthquake risk model, assumed to be available prior to the event, builds on the empirical GMM of Bindi et al. (2011) and the spatial correlation model of Esposito and Iervolino (2011). In accordance with the Italian risk model (?), buildings are first classified according to their material: URM or RC. For URM buildings, Rosti et al. (2021b) empirically derived fragility functions for three classes (A, B, and C1, with decreasing susceptibility to earthquake-induced damage) further separated into low-rise and mid-/high-rise buildings. For RC buildings, we employ fragility functions from Rosti et al. (2021a) that differentiate between two classes (C2, D) for buildings built before and after the introduction of seismic design guidelines. RC buildings are further separated into low-, mid- and high-rise buildings. Thus, a total of 12 classes $b \in \mathcal{B}$ are defined. Appendix A3 contains further details.

Processing of inspection results The available inspection results provide damage descriptions for individual groups of elements, such as the vertical and horizontal elements of the lateral load resisting system. Following the approach by Dolce et al. (2019), we transform these descriptions into overall levels of building damage and link them to one of the four typological attribute combinations $a \in \mathcal{A}$. Given the absence of documented damage, the buildings, which are added to augment the database, are assumed to be undamaged. This assumption, despite not reflecting small undocumented damage, is assumed reasonable.

5.1.3 Kraljevo

On November 3, 2010, an earthquake of $M_w = 5.4$ struck the region of Kraljevo (Serbia), claiming two fatalities and leading to 16'000 structures for which damage has been reported. This case-study examines damage to approximately 2'000 residential buildings, for which detailed information are publicly available (Stojadinovic et al., 2021). We group the 2'000 buildings artificially into three subregions (see Figure3).

Available input data The event-characteristics are taken from the earthquake catalogue published by the Seismological Survey of Serbia (2017). No seismic recordings from stations sufficiently close to the Kraljevo region are available. The aforementioned building dataset covers single-family houses and 90% of them have no more than two stories. No additional pre-processing of this dataset was performed and we consider following building information available before the event: geographical coordinates, number of stories, construction year, and soil class.

Risk model components The earthquake risk model, builds on the empirical ground-motion model of Akkar and Bommer (2010) and the spatial correlation model of Esposito and Iervolino (2012). Stojadinovic et al. (2021) have classified buildings into four main categories, Adobe, URM with two different types of brick and URM with tie-beams. The latter are further divided into three subcategories based on the construction year. Due to the lack of region-specific fragility functions, we employ functions provided in Martins and Silva (2020). In our prior risk model we consider the four main categories as the random typological attribute combinations $a \in \mathcal{A}$. Using the available input information, we establish a deterministic link to the 12 fragility classes $b \in \mathcal{B}$. Further details are provided in the electronic appendix A4.

5.2 Data-Driven Methods for Comparative analysis

To assess the quality of the predictions obtained with the proposed framework and the additional information gain provided by the use of risk models, two purely data-driven regression methods are used. A simple linear model is selected as basic option and a random forest reflects a common regression approach from previous literature studies (Roeslin et al., 2020; Kovačević et al., 2018). The following building-specific features are used for the data-driven regression: Geo-coordinates, epicentral distance, construction year, number of stories, and soil classes.

Ordered Linear Probit (OLP) The ordered linear probit regression is similar to the ordinal GP regression outlined in Section 2, with the exception of the latent function f not being a GP but a linear combination of the features $\mathbf{x}^\top \boldsymbol{\psi}$. The form of the likelihood is identical to Eq7 and the threshold parameters $\boldsymbol{\eta}$, together with the feature weights $\boldsymbol{\psi}$, are estimated by minimizing the negative log-likelihood.

Random Forest (RF) RF classifiers are ensembles of decision trees that are trained using bootstrap samples of the data (Breiman, 2001). For each set of bootstrap samples, a decision tree is grown by recursively partitioning the input space until all terminal nodes contain a pre-defined minimum number of samples. During training, a random subset of the d input variables is selected at each node and the best split is chosen as the split that leads to the largest decrease in Gini-impurity (Breiman et al., 1984). For prediction, the class membership probabilities are derived via averaging the probabilities predicted by the individual trees⁵. The number of trees in the forest, the minimal required number of samples in a node and the size of the feature subset to split the nodes are hyperparameters of this model. We fixed the number of trees in the forest to 1'000 and perform a grid-search amongst pre-specified values for the other two hyper-parameters, keeping the combination that returns the highest out-of-bag performance.

A GP Regression method, following Sheibani and Ou (2020) for the model specification, was also implemented. However, because the damage data is ordinal (and not continuous), this regression method showed the worse performance and thus, has been omitted from the presentation of the results. Indeed, the application of a regression model to ordinal (or categorical) data would require an additional training step that includes a "calibration" model to transform the continuous outputs to the discrete categories.

5.3 Results

Probabilistic predictions $p(\mathbf{N}_k)$ of the distribution of damage levels $y \in \mathcal{Y}$ sustained by exposed buildings in subregion k are performed. With the aim to assess the performance of the proposed framework and other methods (see 5.2) we introduce two error metrics: i) the marginal prediction error (MPE) that measures the performance on predicting the number of buildings in any single damage level and (ii) the joint prediction error (JPE) that measures performance jointly over all damage levels.

The predictive multivariate distribution $p(\mathbf{N}_k)$ is approximated via r Monte-Carlo samples $\mathbf{n}_{kr}$, as described in Section 3.4. The vector $\tilde{\mathbf{n}}_k$ denotes for a subregion k the 'true' number of buildings in each damage level. We employ the continuous ranked probability score (CRPS) as the error metric to quantify the MPE. The CRPS metric is often used to compare the performance of probabilistic forecasts, for instance from weather forecasting systems (Hersbach, 2000; Gneiting and Raftery, 2007; Gneiting and Katzfuss, 2014). Thus, for the amount of buildings in any subregion k being in damage state y , the MPE for predictive distribution $p(N_{ky})$ with respect to the true value $\tilde{n}_{ky}$ is approximated via r samples as

$$\text{MPE}_{ky} \hat{=} \text{CRPS}(p(N_{ky}); \tilde{n}_{ky}) = \frac{1}{r} \sum_{j=1}^r \left(\|n_{kyj} - \tilde{n}_{ky}\|_2 - \frac{1}{2r} \sum_{i=1}^r \|n_{kyj} - n_{k yi}\|_2 \right). \quad (19)$$

The first term of the MPE formulation measures the bias - or lack of accuracy - of the predictions, while the second term of the MPE accounts for the dispersion - or lack of precision. For deterministic predictions, the MPE simplifies to the absolute error metric, which renders the interpretation of the MPE more intuitive. Also, using the definition of Eq. 19, the MPE is given in the same units than the predicted quantity, which is, in this case, the number of buildings. Finally, the CRPS can be readily extended to the multivariate case, as discussed in the following.

The energy score (ES) is used to quantify the JPE. The ES generalizes the CRPS towards multivariate predictions. The ES is calculated in a similar way to Eq. 19, by replacing the scalar quantities n_{kyj} and $\tilde{n}_{ky}$ with vectors $\mathbf{n}_{kr}$ and $\tilde{\mathbf{n}}_k$, which include all the damage levels. Hence, for the ES, the first term in Eq. 19 measures the Euclidean distance between the sampled damage level partitioning and the true partitioning of the exposed buildings in a four-dimensional space

⁵Based on the features of a given test input, a trained decision tree assigns the corresponding terminal node and calculates the class membership probability as the proportion of class labels in this terminal node that were seen during training.

(because we consider four damage levels in the case studies). Again, the ES as JPE has the same units as the predicted quantity.

5.3.1 Zurich

The Zurich case study, consisting a simulated earthquake scenario, contains the most exhaustive information of the geographical distribution of the ground-motion intensity and the ‘true’ damage. Therefore, the local performance of impact estimation is reported for the Zurich case study in Figure 4, which contains the predicted number of moderately damaged buildings in a single subregion after the M5.8 scenario earthquake (see Figure 3 for the earthquake details). When increasing amounts of buildings are inspected and thus, available to train the GP model, the kernel density of the predictions made with the proposed RMGP increases, meaning that the prediction uncertainty decreases. The initial prediction - relying exclusively on the risk model and the shake-map that is established with seismic-network measurements - suffers from large uncertainties: the amount of buildings with moderate damage ranges from below 100 to over 800 buildings, which corresponds to a range from 7% to 54% of the total amount of 1471 buildings in this subregion. Such large uncertainties undermine good decision making and the organisation of recovery activities. The predictions stemming from the RF, which do not have a physics-based component, require a minimum amount of initial building inspections to enable predictions and therefore, are represented only from the second timestep, $t = 1$. When comparing the RMGP and the RF, the latter returns more precise yet less accurate results as the ‘true’ value is not contained within the 90%-confidence interval. The over-confident predictions with almost no uncertainty produced by the RF stem from the implicit model assumptions. Conditional on the input values x of some buildings, the RF considers damage to these buildings as independent, regardless of how close the buildings are. The RMGP, on the other hand, generates spatially correlated samples of ground-motion intensity, conditional on which building damage is then assumed to be independent.

In addition to the probabilistic distribution of the predicted number of moderately damaged buildings, the MPE values, provided in Figure 4 and derived using Equation 19, quantify the reduction of the predictive uncertainty and the convergence of the median value towards the ‘true’ amount of moderately damaged buildings. Comparing the kernel density distributions with the MPE values highlights the joint contribution of precision and accuracy to the MPE. Providing the MPE for all four damage levels (from no damage, $y = 0$, to extensive damage, $y = 3$) of the considered subregion, Table 2 complements Figure 4. The four time steps coincide with the time-steps reported in Figure 4, where inspection data is available for 0, 0.5, 1 and 1.5 % of all 34’000 buildings. Finally, Table 2 also provides the JPE, which measures the error in the joint distribution of buildings in this subregion belonging the four damage levels.

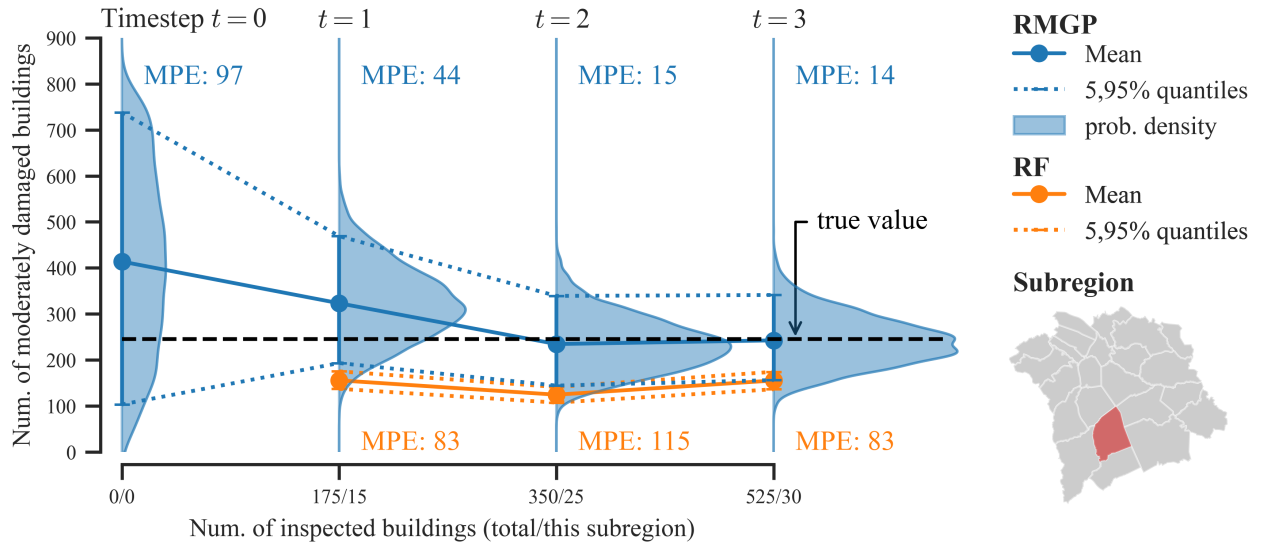


Figure 4: Zurich case-study: Evolution of the predictions on the number of moderately damaged buildings ($y = 2$) in the indicated subregion over four time steps with increasing amounts of inspection data. Predictions made with the proposed RMGP method (blue) and an entirely data-driven random forest (RF, orange) are compared with the true value. The numerical values indicate the marginal prediction error (MPE) at the four time steps for both methods.

While Figure 4 represents the results for a single subregion, impact estimates are generally required to cover the entire affected region. The error in damage predictions for each subregion, represented using the JPE, is therefore reported in Figure 5. The geographical distribution of prediction errors confirms that the RMGP rapidly converges for all

Table 2: Marginal and joint prediction error for damage level counts without inspection data ($t = 0$) and using data gathered until three timesteps for the subregion indicated in Figure 4.

			RMGP				RF		
			$t = 0$	$t = 1$	$t = 2$	$t = 3$	$t = 1$	$t = 2$	$t = 3$
Marginal Prediction Error ¹	MPE_{ky}	$y = 0$	118	65	22	26	36	87	27
		$y = 1$	363	15	9	26	58	36	61
		$y = 2$	97	44	15	14	83	115	83
		$y = 3$	14	4	8	4	23	21	16
Joint Prediction Error	JPE_k		365	80	31	41	113	149	107

¹ The true damage level counts in this subregion are [519, 676, 245, 31] for damage levels $y \in [0, 1, 2, 3]$.

subregions, while the JPE remains high for the RF, especially in some subregions. Also, the cumulated joint prediction error (CJPE), obtained by summing the JPE over all subregions, shows the convergence of the prediction results towards the real building damage. Note that the results shown so far consider the outcome of a single inspection process, which leads to the singularity that the predictions at timestep $t = 2$ are slightly more accurate than at timestep $t = 3$. This singularity may originate from a particularly efficient distribution of buildings inspected until timestep $t = 2$ or a slight overfit to a badly balanced inspection set at timestep $t = 3$. Given the influence of a single random inspection process, multiple inspection runs are considered in the following.

The CJPE is subsequently used to compare the prediction outcomes of RMGP, RF, and OLP in Figure 6. The violin outlines, presented in Figure 6, provide a visual representation of the distribution of the kernel probability density, i.e. the width of the shaded area corresponds to the proportion of the data at this vertical ordinate. The probability distributions is based on 50 random inspection runs, each leading to one instance of a cumulated JPE. The dashed and dotted lines within the violins indicate the median and the inter-quartile range. Statistically, all three methods converge with additional data. For the RMGP, the information from the first 175 inspected buildings provides in all cases a significant improvement in the prediction results, with the median CJPE decreased by over 50% with respect to the initial risk model predictions.

The RMGP produces the most reliable predictions, especially with few available building inspections. The median CJPE of the RMGP with 175 inspected buildings is lower than the median CJPE of the RF after 525 buildings were inspected. Hence, the additional information provided by the underlying risk model reduces the need for building inspections in absence of optimised inspection schemes. The influence of the inspection scheme is larger for the purely data-driven methods, RF and OLP, than for the RMGP, which includes physics-based information. Finally, it is recalled that all buildings are modelled to be founded on soil class B, meaning that additional influence of varying soil classes is likely to further increase the number of inspected buildings that would be required for data-driven methods, such as RF and OLP, to converge.

In real case studies, the shake map cannot be known with precision except at the seismic network stations, which are used to derive the shake maps. Therefore, we use this simulated case study to analyze the performance of the proposed RMGP model in updating the function f , which is the logarithmic ground-motion IM, in this case expressed as PGA. The top of Figure 7 shows the evolution of the inferred median PGA over the entire region when increasing amount of buildings are inspected⁶, while the bottom of the figure illustrates the probability density of the posterior distributions at three specific locations. The uncertainty at specific locations is reduced, especially when some buildings in the vicinity of the location get inspected, as evidenced by the uncertainty reduction after 175 inspections for locations S1 and S2, while at location S3 the variance only reduced after 525 buildings are inspected, some of which are in the same subregion than location S3. This comparison is particularly interesting remembering the fact that the "true", simulated, damage for the vast majority of the inspected buildings originates from IMs other than PGA, namely spectral accelerations at periods of 0.3, 0.6 and 1.0 seconds. The physical meaning of the latent function provides information on the shaking intensity that may prove useful in assessing other earthquake consequences or the risk of cascading events.

5.3.2 Pollino

The rapid impact assessment with the risk model for Pollino, shown in Figure 8 (top left), produces large prediction errors for three subregions. When inspection outcomes are used to update the risk model, the quality of the estimates

⁶The median PGA corresponds to the exponential mean of the posterior distributions of function f , which is, for these maps, evaluated at a regularly spaced grid.

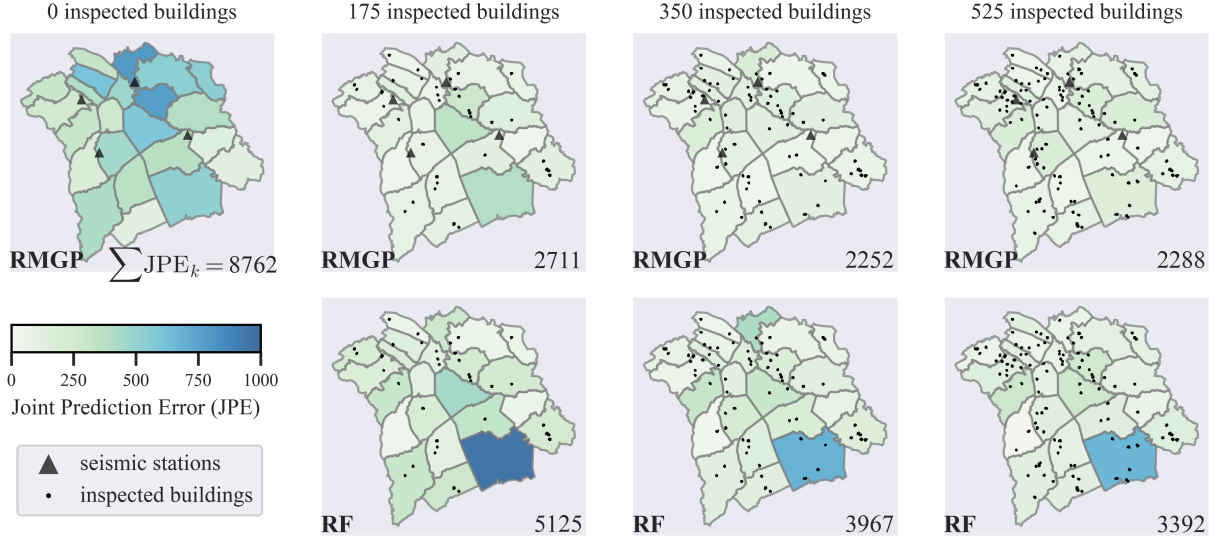


Figure 5: Zurich Case-Study: Joint prediction error (JPE) for the proposed RMGP (top row) and a random forest (RF, bottom row) in all considered subregions at four time steps (columns) with increasing amounts of inspection data. The numerical values indicate the cumulated (over all subregions) JPE achieved by both methods at the different time steps.

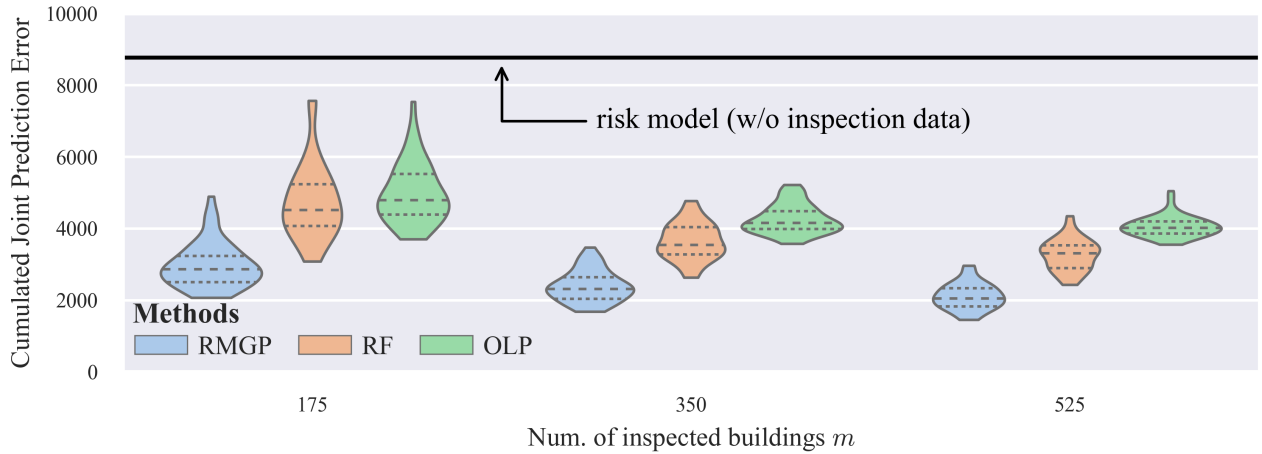


Figure 6: Zurich Case-Study: Violin plots for the cumulated joint prediction error (CJPE) over 50 random inspection processes using the herein proposed method (RMGP), a random forest (RF) and an ordered linear probit regression (OLP) for three time steps corresponding to increasing amounts of inspection data available for training. The black solid line illustrates the cumulated JPE obtained with the prior risk model and recordings from seismic stations.

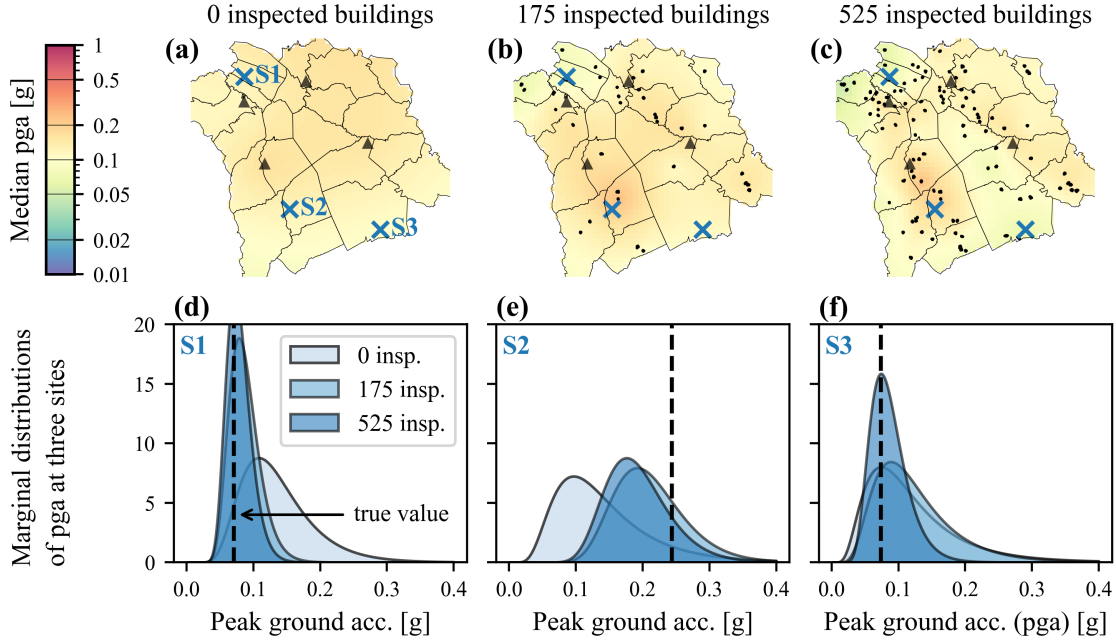


Figure 7: Zurich Case-Study: Inferred estimates of peak ground acceleration (PGA). The top row (a-c) show maps of the median PGA inferred from increasing amounts of inspection data and the bottom row (d-f) illustrates the posterior distributions of PGA for three building sites.

increases rapidly in all subregions. The CJPE, representing the total number of mispredicted buildings at a subregional aggregation level, decreases by over 60%, from 3480 to 1397, by fusing the risk model with inspection data of 600 buildings, corresponding to 3% of the entire building stock in the affected region. For this specific inspection sequence, the RF model (shown in the bottom row of Figure 8) performs better than the initial risk model. In addition, with increasing amounts of data, the prediction performance of the RF, measured through the CJPE, approaches the one of the RMGP. This underlines the importance of the additional information provided by the risk model to the RMGP, especially in the early stages of inspection, when data is scarce.

As outlined in Section 4, the proposed RMGP model leverages inspection data not only to constrain the distribution of the ground-motion IMs, in this case PGA, but also to update fragility function parameters. Figure 9 shows the evolution of the inferred median PGA in the top row and the evolution of the estimated fragility curves, for low-rise URM buildings of class A, in the bottom row. Compared to the Zurich case study, the herein considered region is almost ten times larger and equipped with a much sparser seismic network which results in larger heterogeneity of the inferred median PGAs. Whereas the initial shake map varies from below 0.02g to 0.25g, the incorporation of inspection data reveals a more fine-grained spatial pattern. While the updating does not change the fragility curve for slight damage (delimiting $y = 0$ from $y = 1$), the initial fragility curve for extensive damage shifts towards less conservative predictions with the addition of inspection outcomes. Such updated fragility curves may be useful in case of aftershocks or earthquake events in neighbouring regions.

In a similar manner than for the Zurich case, the violin outlines of kernel densities of the CJPE are reported in Figure 10 and correspond to 50 random simulation outcomes of inspection processes. Compared to the simulated Zurich earthquake, the initial risk model is contained within the range of possible outcomes for the RF and the OLP. Updating the risk model with inspection outcomes leads to a median reduction of approximately 50%, yet, the dispersion of CJPE results due to inspection processes is large for the first 200 inspected buildings. In addition, the median and 75-percentile CJPE values of the RMGP with 200 inspected buildings are similar to the corresponding RF values based on 600 building inspections. Again, this underlines that the fusion of the risk model with inspection data outperforms the risk model without inspection data and the purely data-driven predictions that do not have a physics-based foundation.

5.3.3 Kraljevo

The Kraljevo case study is characterized by a much smaller impacted area (see Figure 3), for which no seismic network stations are available. In addition, the predictions originating from the original risk model have a comparable accuracy

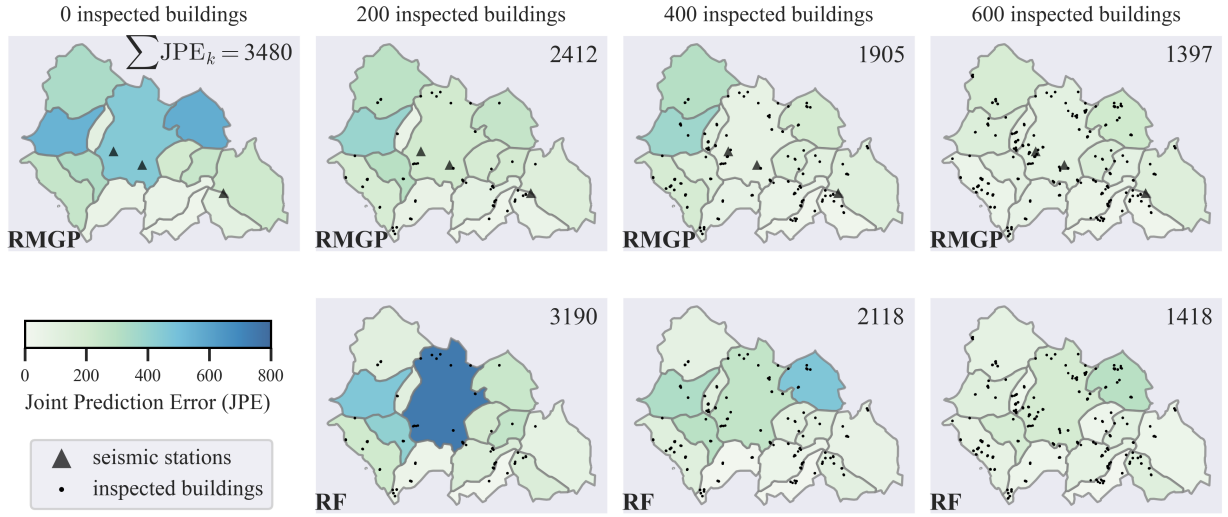


Figure 8: Pollino Case-Study: Joint prediction error (JPE) for the herein proposed method (RMGP, top row) and a random forest (RF, bottom row) in all considered subregions at four time steps (columns) with increasing amounts of inspection data. The numerical values indicate the cumulated (over all subregions) JPE achieved by both methods at the different time steps.

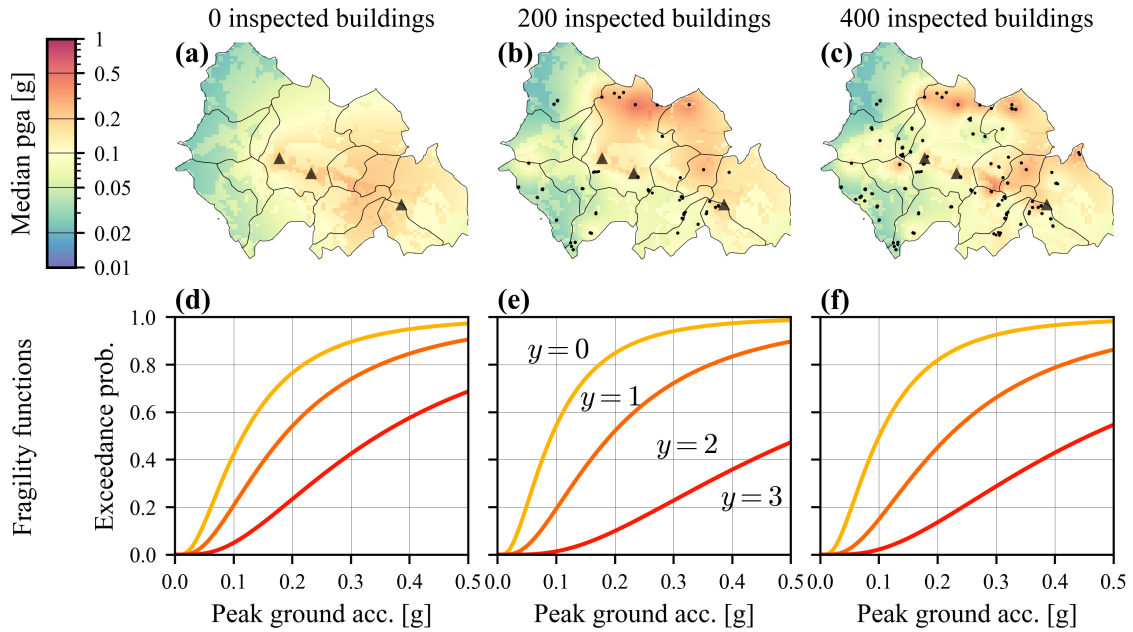


Figure 9: Pollino case-study: Top row shows the predicted median PGA inferred from increasing amounts of inspection data. The bottom row shows the fragility functions for low-rise URM buildings of class A (most vulnerable).

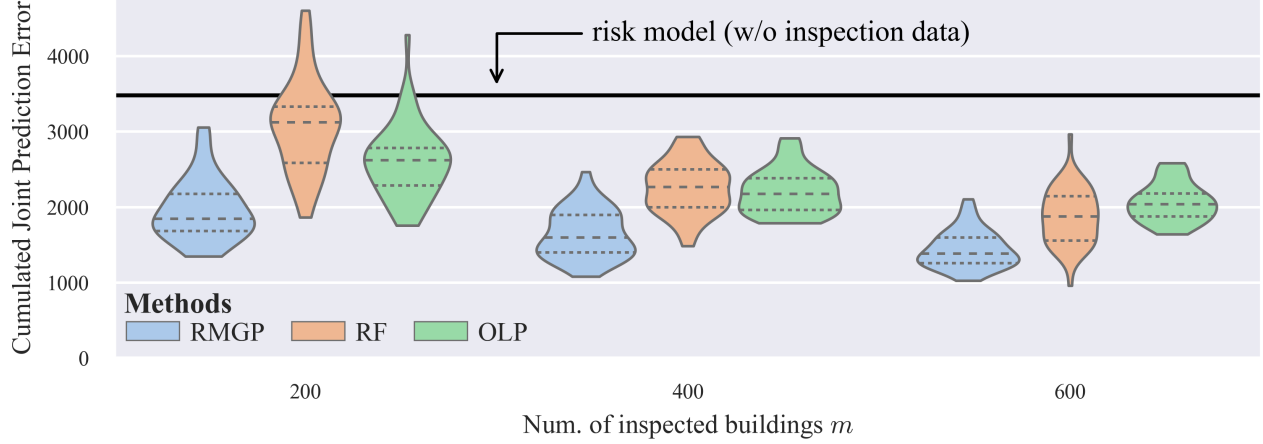


Figure 10: Pollino Case-Study: Violin plots for the cumulated joint prediction error (JPE) over 50 random inspection processes using the herein proposed method (RMGP), a random forest (RF) and an ordered linear probit regression (OLP) for three time steps with increasing amounts of inspection data available for training. The black solid line illustrates the cumulated JPE obtained with the prior risk model and recordings from seismic stations.

to the predictions with the updated model and even performs better than the median outcomes from RF and the OLP, when 80 building inspections are available. Despite the limited geographic range of the case study, 80 buildings are very few considering the amount of information and correlations that needs to be inferred, especially in absence of an underlying (risk) mode.

In the Kraljevo case study, 160 inspections - corresponding to approximately 8% of the building stock - are required to improve upon the predictions from the initial risk-model, in a manner that is robust with respect to the inspection sequence. The inspection sequences are based on random starting points and careful, or even optimal, selection of the buildings to be inspected would help ensuring CJPE values that remain close to the lower bound of the predicted ranges.

Finally, when using data from 8% and 12% of the building stock, the low CJPE values for RF are aligned with the findings of Stojadinovic et al. (2021) who, based on an analysis of the same dataset, proposed the use of RFs for post-earthquake repair cost estimation. Specifically, the similar performance achieved by RMGP and RF indicates a diminishing benefit from an explicit modelling of spatially distributed ground-motion estimates for this small-scale case-study.

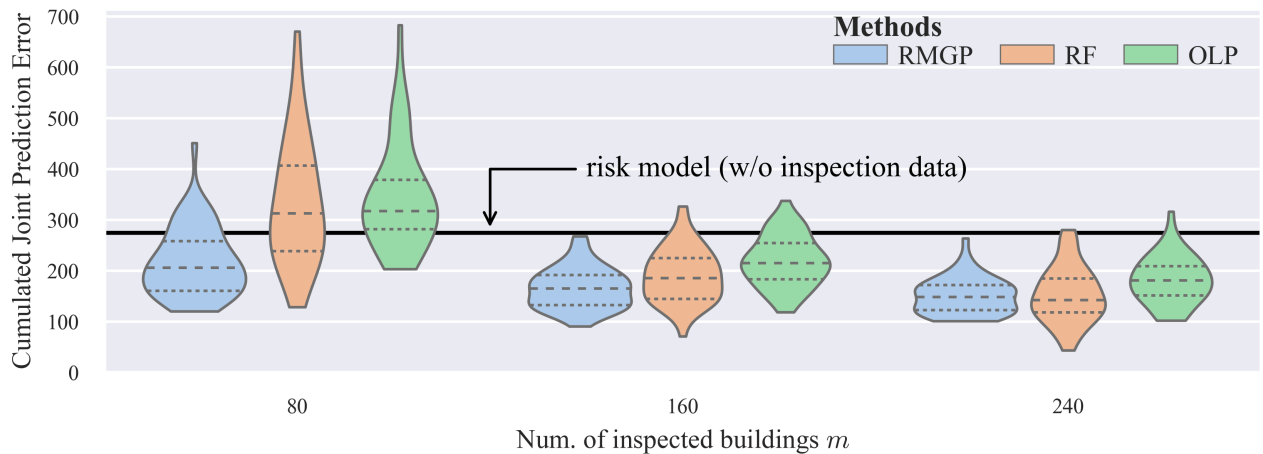


Figure 11: Kraljevo Case-Study: Violin plots for the cumulated joint prediction error (JPE) over 50 random inspection processes using the herein proposed method (RMGP), a random forest (RF) and an ordered linear probit regression (OLP) for three time steps with increasing amounts of inspection data available for training. The black solid line illustrates the cumulated JPE obtained with the prior risk model.

5.4 Summary and Limitations

Comparing the results of all three case studies, Table 3 summarizes Figures 6, 10, and 11 by stating the mean, minimal and maximal CJPE over 50 randomly sampled inspection processes. For a more transparent comparison, the CJPE is normalized by the number of exposed buildings in the respective region and the best-performing methodology is highlighted in bold for each time step. The potential of the risk-model-informed GP framework to reduce errors and uncertainty in the geographical distribution of earthquake-induced building damage is confirmed for all three case studies.

The first, simulated, *Zurich* case study shows that despite a relatively dense seismic network, the initial predictions with the assumed prior risk model may be inaccurate. Updating the components of the risk model rapidly captures the actual trend in the data and thus, provides reliable regional damage estimates after a fraction of the time that would be required to inspect the entire city. This is also confirmed by the case of *Pollino*, where the initial risk-model predictions are more accurate despite the less dense seismic network. For the *Kraljevo* case, the updating after the first timestep leads to a less significant uncertainty reduction. This reduced performance may be attributed to the small case-study region, with the diminished performance indicating reduced efficiency, if inspection results are limited to a small region or small absolute quantities of inspected buildings, which is less than half, when compared with the first two case studies.

Yet, in all case studies, the proposed RMGP method performs better (*Zurich* and *Pollino*) or equally well (*Kraljevo*) than purely data-driven approaches, such as RF. In addition, for the first timestep ($t = 1$), with very limited data, the RMGP performs significantly better than the data-driven approaches. In general, we focus on the early aftermath, with very few inspected buildings (no more than 600), which undermines data-driven methods that typically require large training sets. It should be noted that the RFs, used for comparison in this paper, represent post-earthquake studies from the literature. Considering the rapid evolution of machine-learning tools, there may exist modified RF versions, or other methods such as neural networks, that outperform the herein adopted version.

The indicated performances of the initial rapid damage estimates ($t = 0$) naturally depend on the assumed prior risk models and do not allow for conclusions on how official national risk models might perform in future events. Such models might profit from more detailed information on the exposed building stock, and can incorporate better knowledge from local experts. The damage-estimation method that we propose involves a bottom-up earthquake risk model and thus, may prove to have a limited applicability in low-income countries, where geo-coordinates of buildings are not always complete. However, grass-root exposure modelling Scaini et al. (2022) may allow to overcome this limitation.

The case studies idealize the visual inspection outcome as accurate and unbiased. Thus, subjectivity in visual inspections and damage level definitions is not accounted for. However, the good performance of the framework for the real cases, *Pollino* and *Kraljevo*, underlines its applicability to datasets stemming from human inspectors.

The dataset of buildings, for which inspection data becomes available in the early aftermath, is based on random starting points for inspection teams and subsequent choice of the nearest buildings. However, especially in less affected subregions, inspections may be performed only on demand of building owners and thus, limited to buildings that have a higher a-priori probability of being damaged than randomly chosen buildings. This introduces a sampling bias in the dataset and may undermine the use of machine-learning tools. To overcome this issue, either random buildings need to be inspected in addition to those with an open inspection demand or special training schemes need to be implemented. In the case of RMGP, this could be achieved by conditioning the fragility functions on the event that the owner reported damage. These conditional functions are used for inference and damage predictions for buildings that requested an inspection, whereas the unconditional versions are used to predict damage to buildings for which an inspection is not yet requested. We leave this for future work.

6 Conclusion

This paper contains a framework for post-earthquake damage estimation that dynamically leverages early arriving empirical observations of ground-motion intensity and building damage. The framework relies on GP models to update the individual components of a regional earthquake risk model that is established prior to the event. Based on the application to three case-studies the conclusions are as follows:

- GP models provide a powerful tool to combine inspection evidence with prior estimates of traditional risk models, thus enabling precise and accurate predictions of earthquake-inflicted impacts to residential buildings in a fraction of the time required to inspect the entire building stock.
- The parallel updating of ground-motion intensity estimates, fragility functions, and typological attribution models enable a local calibration of prior risk models with broader geographical scopes. The produced

Table 3: Summary of the Cumulated Joint Prediction Errors for all three case-studies using the herein proposed method (RMGP) and random forests (RF) as well as ordered linear probit (OLP), without inspection data ($t = 0$) and using data gathered up until three time-steps t after the event. The indicated values are the mean (min., max.) over 50 random inspection processes.

Case-study	Model	Cumulated Joint Predictive Error (CJPE) in % of the number of exposed buildings			
		$t = 0$	$t = 1$	$t = 2$	$t = 3$
Zurich ¹	RMGP (ours)	26.1	8.8 (6.1, 14.6)	7.1 (5.0, 10.3)	6.3 (4.3, 8.8)
	RF		14.1 (9.2, 22.5)	10.8 (7.8, 14.2)	9.7 (7.2, 12.9)
	OLP		14.8 (10.6, 22.4)	12.7 (10.6, 15.5)	12.1 (10.6, 15.0)
Pollino ²	RMGP (ours)	17.0	9.5 (6.6, 14.9)	7.9 (5.3, 12.0)	7.0 (5.0, 10.2)
	RF		15.0 (9.1, 22.4)	11.0 (7.2, 14.3)	9.1 (4.7, 14.4)
	OLP		12.7 (8.6, 20.8)	10.7 (8.7, 14.2)	10.0 (8.0, 12.6)
Kraljevo ³	RMGP (ours)	14.0	11.1 (6.1, 23.0)	8.5 (4.6, 13.7)	7.7 (5.2, 13.5)
	RF		17.0 (6.6, 34.2)	9.8 (3.6, 16.7)	7.8 (2.2, 14.3)
	OLP		17.6 (10.4, 34.8)	11.2 (6.0, 17.2)	9.3 (5.2, 16.1)

¹ Available inspection data for the three time steps: 175 (0.5%), 350 (1.0%), 525 (1.5%).

² Available inspection data for the three time steps: 200 (1.0%), 400 (2.0%), 600 (3.0%).

³ Available inspection data for the three time steps: 80 (4.0%), 160 (8.0%), 240 (12.0%).

byproducts, such as the updated fragility functions, provide useful information for risk modellers beyond rapid damage assessment.

- Using GP models to fuse inspection data with prediction models for probabilistic regional risk analysis requires less data compared with approaches that solely rely on data to provide estimates, such as random forests, a common machine-learning technique.
- Compared to purely data-driven approaches, predictions obtained with inspection-informed GP models are more robust with respect to changing inspection sequences and do not require prior optimization or allocation of buildings for priority inspection.

The promising results from the application of the proposed framework to three case-studies highlight its potential for operational use after future earthquakes. Yet, this requires close collaboration between risk modellers and agencies responsible for planning the inspection process. Future work may involve understanding the minimum amount of inspection data that is required to yield stable and reliable updating results, especially in presence of possibly biased inspection results. In addition, a combination of the dynamically improving damage estimates with damage-to-loss and recovery models to examine its effect on financial loss predictions and recovery forecasts. Finally, applying the framework to more complete data, such as available at later stages of the inspection, the proposed framework may offer a robust and automated methodology to derive empirical fragility curves.

Acknowledgements

The work presented in this paper was financially supported by the Real-time Earthquake Risk Reduction for a Resilient Europe 'RISE' project, financed under the European Union's Horizon 2020 research and innovation program, under grant agreement No 821115, as well as the ETH Risk Center project 'DynaRisk', financed under grant agreement ETH-11 18-1.

References

- Akkar, S., Bommer, J.J., 2010. Empirical Equations for the Prediction of PGA, PGV, and Spectral Accelerations in Europe, the Mediterranean Region, and the Middle East. *Seismological Research Letters* 81, 195–206. URL: <https://pubs.geoscienceworld.org/srl/article/81/2/195-206/143661>, doi:doi:10.1785/gssrl.81.2.195.
- Amin, S., Goldstein, M., 2008. Data Against Natural Disasters : Establishing Effective Systems for Relief, Recovery, and Reconstruction. World Bank, Washington D.C. URL: <http://hdl.handle.net/10986/6511>.
- Amt für Raumentwicklung, 2020. Geografisches Informationssystem des Kantons Zürich (GIS-ZH). URL: <http://maps.zh.ch>.

-
- Baker, J.W., Cornell, A., 2006. Correlation of Response Spectral Values for Multicomponent Ground Motions. *Bulletin of the Seismological Society of America* 96, 215–227. URL: https://pubs.geoscienceworld.org/ssa/bssa/article-pdf/96/1/215/2726988/215_961_05060.pdf, doi:doi:10.1785/0120050060.
- 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, 1899–1920. URL: <http://link.springer.com/10.1007/s10518-011-9313-z>, doi:doi:10.1007/s10518-011-9313-z.
- Bodenmann, L., Reuland, Y., Stojadinovic, B., 2021. Dynamic Updating of Building Loss Predictions Using Regional Risk Models and Conventional Post-Earthquake Data Sources, in: *Proceedings of the 31st European Safety and Reliability Conference (ESREL 2021)*, Research Publishing Services, Singapore. pp. 1411–1418. URL: <https://rpsonline.com.sg/proceedings/9789811820168/html/559.xml>, doi:doi:10.3850/978-981-18-2016-8_559-cd.
- Boore, D.M., Gibbs, J.F., Joyner, W.B., Tinsley, J.C., Ponti, D.J., 2003. Estimated Ground Motion From the 1994 Northridge, California, Earthquake at the Site of the Interstate 10 and La Cienega Boulevard Bridge Collapse, West Los Angeles, California. *Bulletin of the Seismological Society of America* 93, 2737–2751. URL: <https://pubs.geoscienceworld.org/bssa/article/93/6/2737-2751/121020>, doi:doi:10.1785/0120020197.
- Booth, E., Saito, K., Spence, R., Madabhushi, G., Eguchi, R.T., 2011. Validating Assessments of Seismic Damage Made from Remote Sensing:. *Earthquake Spectra* 27. URL: <https://journals.sagepub.com/doi/full/10.1193/1.3632109>, doi:doi:10.1193/1.3632109.
- Breiman, L., 2001. Random Forests. *Machine Learning* 45, 5–32. doi:doi:10.1023/A:1010933404324.
- Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 1984. *Classification And Regression Trees*. Routledge. URL: <https://www.taylorfrancis.com/books/9781351460491>, doi:doi:10.1201/9781315139470.
- Burton, H.V., Miles, S.B., Kang, H., 2019. Integrating Performance-Based Engineering and Urban Simulation to Model Post-Earthquake Housing Recovery:. *Earthquake Spectra* 34, 1763–1785. URL: <https://journals.sagepub.com/doi/full/10.1193/041017EQS067M>, doi:doi:10.1193/041017EQS067M.
- Chiou, B.S.J., Youngs, R.R., 2014. Update of the Chiou and Youngs NGA Model for the Average Horizontal Component of Peak Ground Motion and Response Spectra. *Earthquake Spectra* 30, 1117–1153. URL: <http://journals.sagepub.com/doi/10.1193/072813EQS219M>, doi:doi:10.1193/072813EQS219M.
- Chu, W., Ghahramani, Z., 2005. Gaussian Processes for Ordinal Regression. *Journal of Machine Learning Research* 6, 1019–1041. URL: <http://www.jmlr.org/papers/volume6/chu05a/chu05a.pdf>.
- Corbane, C., Sabo, F., 2019. European Settlement Map from Copernicus Very High Resolution data for reference year 2015, Public Release 2019. URL: <https://data.jrc.ec.europa.eu/dataset/8bd2b792-cc33-4c11-afd1-b8dd60b44f3b>, doi:doi:10.2905/8BD2B792-CC33-4C11-AFD1-B8DD60B44F3B.
- Crowley, H., Despotaki, V., Rodrigues, D., Silva, V., Toma-Danila, D., Riga, E., Karatzetou, A., Fotopoulou, S., Zugic, Z., Sousa, L., Ozcebe, S., Gamba, P., 2020. Exposure model for European seismic risk assessment. *Earthquake Spectra* 36, 252–273. URL: <http://journals.sagepub.com/doi/10.1177/8755293020919429>, doi:doi:10.1177/8755293020919429.
- D’Ayala, D., Meslem, A., Vamvatsikos, D., Porter, K., Rossetto, T., Silva, V., 2015. Guidelines for Analytical Vulnerability Assessment - Low/Mid-Rise. Technical Report. Vulnerability Global Component Project. doi:doi:10.13117/GEM.VULN-MOD.TR2014.12.
- Dell’Acqua, F., Gamba, P., 2012. Remote Sensing and Earthquake Damage Assessment: Experiences, Limits, and Perspectives. *Proceedings of the IEEE* 100, 2876–2890. URL: <http://ieeexplore.ieee.org/document/6235981/>, doi:doi:10.1109/JPROC.2012.2196404.
- Diana, L., Thiriot, J., Reuland, Y., Lestuzzi, P., 2019. Application of Association Rules to Determine Building Typological Classes for Seismic Damage Predictions at Regional Scale: The Case Study of Basel. *Frontiers in Built Environment* 5. URL: <https://www.frontiersin.org/article/10.3389/fbuil.2019.00051/full>, doi:doi:10.3389/fbuil.2019.00051.
- Dolce, M., Prota, A., Borzi, B., da Porto, F., Lagomarsino, S., Magenes, G., Moroni, C., Penna, A., Polese, M., Speranza, E., Verderame, G.M., Zuccaro, G., 2021. Seismic risk assessment of residential buildings in Italy. *Bulletin of Earthquake Engineering* 19, 2999–3032. URL: <https://link.springer.com/10.1007/s10518-020-01009-5>, doi:doi:10.1007/s10518-020-01009-5.
- Dolce, M., Speranza, E., Giordano, F., Borzi, B., Bocchi, F., Conte, C., Meo, A.D., Faravelli, M., Pascale, V., 2019. Observed damage database of past Italian earthquakes: The Da.D.O. WebGIS. *Bollettino di Geofisica Teorica ed Applicata* 60, 141–164. doi:doi:10.4430/bgta0254.

- Erdik, M., Şeşetyan, K., Demircioğlu, M.B., Hancilar, U., Zülfiyar, C., 2011. Rapid earthquake loss assessment after damaging earthquakes. *Soil Dynamics and Earthquake Engineering* 31, 247–266. doi:doi:10.1016/j.soildyn.2010.03.009.
- Esposito, S., Iervolino, I., 2011. PGA and PGV Spatial Correlation Models Based on European Multievent Datasets. *Bulletin of the Seismological Society of America* 101, 2532–2541. URL: <https://pubs.geoscienceworld.org/bssa/article/101/5/2532-2541/326659>, doi:doi:10.1785/0120110117.
- Esposito, S., Iervolino, I., 2012. Spatial correlation of spectral acceleration in European data. doi:doi:10.1785/0120120068.
- Eurocode 8, 2004. Design of structures for earthquake resistance - Part 1: General rules, seismic actions and rules for buildings.
- Forte, G., Chioccarelli, E., De Falco, M., Cito, P., Santo, A., Iervolino, I., 2019. Seismic soil classification of Italy based on surface geology and shear-wave velocity measurements. *Soil Dynamics and Earthquake Engineering* 122, 79–93. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0267726118311114>, doi:doi:10.1016/j.soildyn.2019.04.002.
- Giordano, P.F., Limongelli, M.P., 2020. The value of structural health monitoring in seismic emergency management of bridges. *Structure and Infrastructure Engineering* URL: <https://www.tandfonline.com/doi/abs/10.1080/15732479.2020.1862251>, doi:doi:10.1080/15732479.2020.1862251.
- Gneiting, T., Katzfuss, M., 2014. Probabilistic Forecasting. *Annual Review of Statistics and Its Application* 1, 125–151. URL: <https://www.annualreviews.org/doi/10.1146/annurev-statistics-062713-085831>, doi:doi:10.1146/annurev-statistics-062713-085831.
- Gneiting, T., Raftery, A.E., 2007. Strictly Proper Scoring Rules, Prediction, and Estimation. *Journal of the American Statistical Association* 102, 359–378. URL: <http://www.tandfonline.com/doi/abs/10.1198/016214506000001437>, doi:doi:10.1198/016214506000001437.
- Goda, K., Hong, H.P., 2008. Spatial Correlation of Peak Ground Motions and Response Spectra. *Bulletin of the Seismological Society of America* 98, 354–365. URL: <https://pubs.geoscienceworld.org/bssa/article/98/1/354-365/341898>, doi:doi:10.1785/0120070078.
- Goulet, J.A., Michel, C., Kiureghian, A.D., 2015. Data-driven post-earthquake rapid structural safety assessment. *Earthquake Engineering & Structural Dynamics* 44, 549–562. URL: <https://onlinelibrary.wiley.com/doi/full/10.1002/eqe.2541https://onlinelibrary.wiley.com/doi/abs/10.1002/eqe.2541https://onlinelibrary.wiley.com/doi/10.1002/eqe.2541>, doi:doi:10.1002/EQE.2541.
- Guérin-Marthe, S., Gehl, P., Negulescu, C., Auclair, S., Fayjaloun, R., 2021. Rapid earthquake response: The state-of-the art and recommendations with a focus on European systems. *International Journal of Disaster Risk Reduction* 52, 101958. doi:doi:10.1016/J.IJDRR.2020.101958.
- Hensman, J., Matthews, A., Ghahramani, Z., 2015. Scalable Variational Gaussian Process Classification, in: Lebanon, G., Vishwanathan, S.V.N. (Eds.), *Proceedings of the Eighteenth International Conference on Artificial Intelligence and Statistics, Proceedings of Machine Learning Research*, San Diego, California, USA. pp. 351–360. URL: <https://proceedings.mlr.press/v38/hensman15.html>.
- Hersbach, H., 2000. Decomposition of the Continuous Ranked Probability Score for Ensemble Prediction Systems. *Weather and Forecasting* 15, 559–570. doi:doi:10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2.
- ISTAT, 2001. Censimento generale della popolazione e delle abitazioni. URL: <http://dawinci.istat.it/jsp/MD/>.
- Jayaram, N., Baker, J.W., 2008. Statistical Tests of the Joint Distribution of Spectral Acceleration Values. *Bulletin of the Seismological Society of America* 98, 2231–2243. URL: <https://pubs.geoscienceworld.org/bssa/article/98/5/2231-2243/341955>, doi:doi:10.1785/0120070208.
- Kerle, N., Nex, F., Gerke, M., Duarte, D., Vetrivel, A., 2019. UAV-Based Structural Damage Mapping: A Review. *ISPRS International Journal of Geo-Information* 9, 14. URL: <https://www.mdpi.com/2220-9964/9/1/14>, doi:doi:10.3390/ijgi9010014.
- Kovačević, M., Stojadinović, Z., Marinković, D., Stojadinović, B., 2018. Sampling and machine learning methods for a rapid earthquake loss assessment system, in: 11th U.S. National Conference on Earthquake Engineering, Earthquake Engineering Research Institute, Los Angeles, CA.
- Kuss, M., 2006. Gaussian Process Models for Robust Regression, Classification, and Reinforcement Learning. Ph.D. thesis. Technische Universität Darmstadt, Darmstadt, Germany.
- Kusunoki, K., 2021. Damage Assessment in Japan and Potential Use of New Technologies in Damage Assessment.

- Lagomarsino, S., Giovinazzi, S., 2006. Macroseismic and mechanical models for the vulnerability and damage assessment of current buildings. *Bulletin of Earthquake Engineering* 4, 415–443. doi:doi:10.1007/s10518-006-9024-z.
- Lallemant, D., Soden, R., Rubinyi, S., Loos, S., Barns, K., Bhattacharjee, G., 2017. Post-Disaster Damage Assessments as Catalysts for Recovery: A Look at Assessments Conducted in the Wake of the 2015 Gorkha, Nepal, Earthquake. *Earthquake Spectra* 33, S435–S451. URL: <https://journals.sagepub.com/doi/full/10.1193/120316eqs222m>, doi:doi:10.1193/120316EQS222M.
- Loos, S., Lallemant, D., Baker, J., McCaughey, J., Yun, S.H., Budhathoki, N., Khan, F., Singh, R., 2020. G-DIF: A geospatial data integration framework to rapidly estimate post-earthquake damage. *Earthquake Spectra* 36, 1695–1718. URL: <http://journals.sagepub.com/doi/10.1177/8755293020926190>, doi:doi:10.1177/8755293020926190.
- Luzi, L., Lanzano, G., Felicetta, C., D’Amico, M., Russo, E., Sgobba, S., Pacor, F., ORFEUS Working Group 5, 2020. Engineering Strong Motion Database (ESM) (Version 2.0). doi:doi:10.13127/ESM.2.
- Markhvida, M., Ceferino, L., Baker, J.W., 2018. Modeling spatially correlated spectral accelerations at multiple periods using principal component analysis and geostatistics. *Earthquake Engineering & Structural Dynamics* 47, 1107–1123. URL: <http://doi.wiley.com/10.1002/eqe.3007>, doi:doi:10.1002/eqe.3007.
- Marquis, F., Kim, J.J., Elwood, K.J., Chang, S.E., 2017. Understanding post-earthquake decisions on multi-storey concrete buildings in Christchurch, New Zealand. *Bulletin of Earthquake Engineering* 15, 731–758. URL: <https://link.springer.com/article/10.1007/s10518-015-9772-8>, doi:doi:10.1007/S10518-015-9772-8/FIGURES/5.
- Martins, L., Silva, V., 2020. Development of a fragility and vulnerability model for global seismic risk analyses. *Bulletin of Earthquake Engineering* URL: <http://link.springer.com/10.1007/s10518-020-00885-1>, doi:doi:10.1007/s10518-020-00885-1.
- Matheron, G., 1973. The intrinsic random functions and their applications. *Advances in Applied Probability* 5, 439–468. URL: https://www.cambridge.org/core/product/identifier/S0001867800039379/type/journal_article, doi:doi:10.2307/1425829.
- Matthews, A., van der Wilk, M., Nickson, T., Fujii, K., Boukouvalas, A., León-Villagrà, P., Ghahramani, Z., Hensman, J., 2017. GPflow: A Gaussian Process Library using TensorFlow. *Journal of Machine Learning Research* 18, 1–6. URL: <http://jmlr.org/papers/v18/16-537.html>.
- Mcentire, D.A., Cope, J., 2004. Damage Assessment After the Paso Robles (San Simeon, California) Earthquake: Lessons for Emergency Management.
- Noh, H.Y., Lignos, D.G., Nair, K.K., Kiremidjian, A.S., 2012. Development of fragility functions as a damage classification/prediction method for steel moment-resisting frames using a wavelet-based damage sensitive feature. *Earthquake Engineering & Structural Dynamics* 41, 681–696. URL: <https://onlinelibrary.wiley.com/doi/full/10.1002/eqe.1151><https://onlinelibrary.wiley.com/doi/abs/10.1002/eqe.1151><https://onlinelibrary.wiley.com/doi/10.1002/eqe.1151>, doi:doi:10.1002/EQE.1151.
- Opper, M., Archambeau, C., 2009. The variational gaussian approximation revisited. *Neural Computation* 21, 786–792. doi:doi:10.1162/neco.2008.08-07-592.
- Paal, S.G., Asce, M., Jeon, J.S., Brilakis, I., Asce, A.M., Desroches, R., 2014. Automated Damage Index Estimation of Reinforced Concrete Columns for Post-Earthquake Evaluations. *Journal of Structural Engineering* 141, 4014228. URL: <https://ascelibrary.org/doi/abs/10.1061/%28ASCE%29ST.1943-541X.0001200>, doi:doi:10.1061/(ASCE)ST.1943-541X.0001200.
- Potter, S.H., Becker, J.S., Johnston, D.M., Rossiter, K.P., 2015. An overview of the impacts of the 2010-2011 Canterbury earthquakes. *International Journal of Disaster Risk Reduction* 14, 6–14. doi:doi:10.1016/J.IJDRR.2015.01.014.
- Pozzi, M., Wang, Q., 2018. Gaussian Process Regression and Classification for Probabilistic Damage Assessment of Spatially Distributed Systems. *KSCE Journal of Civil Engineering* 22, 1016–1026. doi:doi:10.1007/s12205-018-0014-x.
- Rasmussen, C., Williams, K., 2006. Gaussian processes for machine learning. The MIT Press, Cambridge, MA.
- Reuland, Y., Lestuzzi, P., Smith, I.F.C., 2019. Measurement-based support for post-earthquake assessment of buildings. *Structure and Infrastructure Engineering* 15, 647–662. URL: <https://www.tandfonline.com/doi/abs/10.1080/15732479.2019.1569071>, doi:doi:10.1080/15732479.2019.1569071.

-
- Roeslin, S., Ma, Q., Juárez-García, H., Gómez-Bernal, A., Wicker, J., Wotherspoon, L., 2020. A machine learning damage prediction model for the 2017 Puebla-Morelos, Mexico, earthquake. *Earthquake Spectra* 36, 875529302093671. URL: <http://journals.sagepub.com/doi/10.1177/8755293020936714>, doi:doi:10.1177/8755293020936714.
- Rosti, A., Del Gaudio, C., Rota, M., Ricci, P., Di Ludovico, M., Penna, A., Verderame, G.M., 2021a. Empirical fragility curves for Italian residential RC buildings. *Bulletin of Earthquake Engineering* 19, 3165–3183. URL: <https://link.springer.com/10.1007/s10518-020-00971-4>, doi:doi:10.1007/s10518-020-00971-4.
- Rosti, A., Rota, M., Penna, A., 2021b. Empirical fragility curves for Italian URM buildings. *Bulletin of Earthquake Engineering* 19, 3057–3076. URL: <https://link.springer.com/10.1007/s10518-020-00845-9>, doi:doi:10.1007/s10518-020-00845-9.
- Scaini, C., Peresan, A., Tamaro, A., Poggi, V., Barnaba, C., 2022. Can high-school students contribute to seismic risk mitigation? Lessons learned from the development of a crowd-sourced exposure database. *International Journal of Disaster Risk Reduction* 69, 102755. URL: <https://linkinghub.elsevier.com/retrieve/pii/S2212420921007160>, doi:doi:10.1016/J.IJDRR.2021.102755.
- Seismological Survey of Serbia, 2017. National Earthquake Catalogue. Technical Report. URL: <http://www.seismo.gov.rs/Seizmichnost/Katalog-zemljotresa.pdf>.
- Sheibani, M., Ou, G., 2020. The development of Gaussian process regression for effective regional post-earthquake building damage inference. *Computer-Aided Civil and Infrastructure Engineering*, mice.12630 URL: <https://onlinelibrary.wiley.com/doi/10.1111/mice.12630>, doi:doi:10.1111/mice.12630.
- Silva, V., Amo-Oduro, D., Calderon, A., Costa, C., Dabbeek, J., Despotaki, V., Martins, L., Pagani, M., Rao, A., Simionato, M., Viganò, D., Yepes-Estrada, C., Acevedo, A., Crowley, H., Horspool, N., Jaiswal, K., Journeay, M., Pittore, M., 2020. Development of a global seismic risk model. *Earthquake Spectra* 36, 372–394. URL: <http://journals.sagepub.com/doi/10.1177/8755293019899953>, doi:doi:10.1177/8755293019899953.
- Stojadinovic, Z., Kovacevic, M., Marinkovic, D., Stojadinovic, B., 2021. Rapid earthquake loss assessment based on machine learning and representative sampling:. *Earthquake Spectra* URL: https://journals.sagepub.com/doi/full/10.1177/87552930211042393?casa_token=VlQpuY3CXwoAAAAA%3AdAA_DlThZRXCU_XaHSe-Z9kkZVi8xnYKwv3hAs9SJYHdNsbC9hQIUoHI4CsrePXhXo0j2b_Z9Mkldow, doi:doi:10.1177/87552930211042393.
- Wald, D.J., Quitoriano, V., Heaton, T.H., Kanamori, H., Scrivner, C.W., Worden, C.B., 1999. TriNet “ShakeMaps”: Rapid Generation of Peak Ground Motion and Intensity Maps for Earthquakes in Southern California. *Earthquake Spectra* 15, 537–555. URL: <http://journals.sagepub.com/doi/10.1193/1.1586057>, doi:doi:10.1193/1.1586057.
- Wang, M., Takada, T., 2005. Macrospatial Correlation Model of Seismic Ground Motions. *Earthquake Spectra* 21, 1137–1156. URL: <http://journals.sagepub.com/doi/10.1193/1.2083887>, doi:doi:10.1193/1.2083887.
- Worden, C., Thompson, E.M., Baker, J.W., Bradley, B.A., Luco, N., Wald, D.J., 2018. Spatial and spectral interpolation of ground-motion intensity measure observations. *Bulletin of the Seismological Society of America* 108, 866–875. URL: <https://pubs.geoscienceworld.org/ssa/bssa/article-pdf/108/2/866/4109765/bssa-2017201.1.pdf>, doi:doi:10.1785/0120170201.

APPENDIX

A1 Zurich case-study: Details on the prior risk model

Follows once ready for submission.

A2 Pollino case-study: Details on data pre-processing

Follows once ready for submission.

A3 Pollino case-study: Details on the prior risk model

Follows once ready for submission.

A4 Kraljevo case-study: Details on prior risk model

Follows once ready for submission.