Dynamic Updating of Building Loss Predictions Using Regional Risk Models and Conventional Post-Earthquake Data Sources

Conference Paper

Author(s):

Bodenmann, Lukas ; Reuland, Yves ; Stojadinovic, Bozidar

Publication date:

2021-09

Permanent link:

<https://doi.org/10.3929/ethz-b-000507866>

Rights / license:

In Copyright - Non-Commercial Use Permitted

Originally published in:

https://doi.org/10.3850/978-981-18-2016-8_559-cd

Funding acknowledgement:

821115 - Real-time Earthquake Risk Reduction for Europe (EC)

Dynamic Updating of Building Loss Predictions using Regional Risk Models and Conventional Post-Earthquake Data Sources

Lukas Bodenmann

*Dept. of Civil, Environmental and Geomatic Engineering, ETH Zurich, Switzerland.
E-mail: bodenmann@ibk.baug.ethz.ch*

Yves Reuland

*Dept. of Civil, Environmental and Geomatic Engineering, ETH Zurich, Switzerland.
E-mail: reuland@ibk.baug.ethz.ch*

Bozidar Stojadinovic

*Dept. of Civil, Environmental and Geomatic Engineering, ETH Zurich, Switzerland.
E-mail: stojadinovic@ibk.baug.ethz.ch*

Earthquakes can cause widespread damage to the built environment, disrupt the function of many residential buildings to provide safe housing capacities and thus, potentially induce severe long-term societal consequences. Rapid recovery significantly improves the short-term resilience of communities after an earthquake. However, time pressure and scarce information on the severity and the spatial distribution of damage complicate the decision-making. Therefore, early damage estimates are produced using regional earthquake risk models with rapid earthquake intensity data and typological building vulnerability functions. While the precision of the former depends, amongst other issues, on the density of seismic network stations and the region-specific geological knowledge, the typological classification of buildings often involves attribution models correlating exposure data, such as building height and age, with typological seismic vulnerability classes. Typological attribution models are approximate and locally add to the uncertainties resulting from the average representation of buildings forming one building class. Employing probabilistic machine-learning tools, the continuous inspection data inflow is leveraged to dynamically update initial regional earthquake risk predictions by updating simultaneously the functions that govern typological attribution and building damage. Hence, while completing inspection of all affected buildings may take several weeks, the limited information becoming available in the first days following an earthquake helps constraining underlying uncertainties. This leads to more reliable rapid estimates of losses of building functions and their respective spatial distribution. The framework is demonstrated on a region in Switzerland subjected to a fictitious earthquake scenario.

Keywords: Rapid Loss Assessment, Earthquake Recovery, Spatial Damage Distribution, Typological Building Classification, Community Resilience, Gaussian Process.

1. Introduction

Earthquake events carry the potential to cause widespread damage to the built environment. In addition to direct financial losses, the loss of functionality of residential buildings to provide safe housing capacities, gives rise to negative long-term consequences. While performance-based seismic design of new buildings and risk-informed seismic retrofitting of vulnerable buildings form the backbone of long-term risk mitigation strategies, adequate post-earthquake decision-making is a crucial enabler of rapid post-earthquake recovery. However, in the early aftermath of an earthquake, decisions, which may shape the recovery trajectory, need to be taken with limited and imprecise data and under intense time pressure.

Post-earthquake inspection is crucial in enabling safe building reoccupancy and functional recovery. Still, visual inspection of all buildings affected from the earthquake may take several weeks. Therefore, regional seismic risk models of exposure and vulnerability are typically combined with ground-motion estimates that are possibly constrained by seismic recordings to provide rapid loss assessment (Erdik et al., 2011). Such regional risk models provide rapid, yet uncertain, damage predictions at regional scale. However, economic and technical constraints hinder the pre-earthquake collection of detailed information and the simulation of structural models for all buildings within a region. Instead, buildings are clustered into several predefined classes, for which an average vulnerability is established. Ideally,

definition of representative classes would start with a detailed building database with relevant attributes (Brzev et al., 2013), many of which are only available in rare cases. Therefore, typological attribution models are employed, which link building types with socio-economic indicators and building attributes that may be found in public databases, such as building height, value and year of construction. These typological attribution models are mostly valid for exposure models that are aggregated at very large scales (Crowley et al., 2020) and thus, may suffer from limited applicability within the region hit by an earthquake. Hence, typological attribution adds to other uncertainties present in ground-motion models and vulnerability functions.

Expert-conducted building inspection in the aftermath of an earthquake guarantees a continuous data inflow regarding the induced damage and, as a byproduct, the building materials as well as the load-resisting and floor systems. In this paper, machine-learning tools are leveraged to update the post-earthquake risk assessment in a restricted geographic area. Previous attempts have been made to apply machine-learning tools to regional seismic risk prediction, assuming that inspected building subsets are representative of a much larger building stock (Diana et al., 2019). Kovačević et al. (2018) proposed random forests to classify damaged buildings based on inspection results for a subset of buildings. However, these entirely data-driven methods require either highly correlated inputs, for example the type of the lateral load resisting system, large amounts of inspected buildings, and/or careful prioritization of inspections to provide stable predictions.

We propose the use of Gaussian processes to fuse available inspection data with an underlying pre-event earthquake risk model in order to dynamically update regional estimates on the loss of housing capacity. Pozzi and Wang (2018) employed Gaussian processes to predict the failure probability of individual components of a known type in spatially distributed infrastructure systems based on binary component state observations. The method outlined in the present paper allows to process observations in terms of multiple ordinal damage categories and involves a simultaneous updating of the functions governing building damage and typological attribution.

This paper starts with a general description of Gaussian processes for classification of ordinal and general categorical data in Section 2 and their application to post-earthquake impact assessments in Section 3. In Section 4 the methodology is applied to a simulated case study in the greater Zurich area showing the potential for a rapid reduction in uncertainty ranges of regional impact assessments.

2. Gaussian Processes for Supervised Learning

Consider a training set $\{\mathbf{x}_i, y_i\}_{i=1}^n = (\mathbf{X}, \mathbf{y})$, where available input vectors $\mathbf{x}_i$ and scalar outputs y_i are collected in matrix $\mathbf{X}$ and vector $\mathbf{y}$, respectively. The objective is to produce probabilistic predictions for outputs corresponding to unseen inputs $\mathbf{X}_*$. A non-linear function f is employed for the probabilistic mapping between inputs and outputs $p(y_i | f_i, \theta)$, where $f_i = f(\mathbf{x}_i)$ and θ are hyperparameters. Instead of explicitly parametrizing the function f , a Gaussian Process (GP) prior is imposed on the latter

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

where $m(\cdot)$ and $k(\cdot)$ denote the mean and positive definite covariance function, respectively. As defined in Rasmussen and Williams (2006), a GP is a collection of random variables, any finite subset of which follows a joint multivariate Gaussian distribution. Thus, when denoting a vector of random variables as $\mathbf{f} = (f_1, \dots, f_n)'$, $p(\mathbf{f}|\mathbf{X})$ is a multivariate normal distribution with mean vector $\boldsymbol{\mu}$ with entries $(\boldsymbol{\mu})_i = m(\mathbf{x}_i)$ and covariance matrix $\boldsymbol{\Sigma}$ with entries $(\boldsymbol{\Sigma})_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$.

In a supervised learning setting, we aim to infer knowledge about the function f provided by the training set via the calculation of the posterior distribution $p(\mathbf{f}|\mathbf{y})$ by applying Bayes' theorem:

$$p(\mathbf{f}|\mathbf{y}) = \frac{1}{Z} \prod_{i=1}^n p(y_i | f_i) p(\mathbf{f}) , \quad (2)$$

where $Z = \int p(\mathbf{y}|\mathbf{f})p(\mathbf{f})d\mathbf{f}$ is the marginal likelihood. Then the posterior predictive distribution $p(f_*|\mathbf{y})$ is derived by marginalizing the function variables over the training set. Note that the dependence on hyper-parameters and inputs is omitted to lighten the notation. The next two sections specify the mapping $p(y_i | f_i)$ for cases with ordinal and general categorical data. Those are of special interest, because they are linked to building damage categories and typological building classes as will be explained in Section 3.

2.1. Ordinal categorical data

In case of ordinal data, the scalar outputs y_i are elements of a finite set $\mathcal{O}$ of c ordered categories, which are denoted as integers e.g. $y_i \in \mathcal{O} = \{1, 2, \dots, c\}$ with preserved ascending ordering information. We follow Chu and Ghahramani (2005) for this brief explanation of ordinal GP regression. By employing $c + 1$ parameters $\eta^1 < \dots < \eta^c < \eta^{c+1}$, where $\eta^1 = -\infty$ and $\eta^{c+1} = \infty$, the class membership probability of label y_i is:

$$p(y_i | f_i) = \Phi[(f_i - \eta^{y_i}) / \beta] - \Phi[(f_i - \eta^{y_i+1}) / \beta] , \quad (3)$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function. The threshold parameters, η , partition the real line into contiguous intervals, mapping the real function f_i into the discrete variable y_i . Parameter β denotes a common dispersion of the threshold parameters. Section 3 provides additional insights into these parameters in the earthquake-risk model context.

2.2. General categorical data

Categorical data, in a similar way to ordinal data, is discrete and finite with possible class outputs, $y_i \in \mathcal{C} = \{1, 2, \dots, c\}$, where $c > 2$ is the number of classes. However, categorical class labels do not provide any ranking information and, instead of having one function f as is the case for ordinal data (see Section 2.1), we have c independent latent functions, i.e. one for each class. The latent function values for data point $\mathbf{x}_i$ are denoted as $\mathbf{g}_i = (g_i^1, \dots, g_i^c)$, where we denote functions as g to avoid confusion with preceding sections. Following the representation outlined in Rasmussen and Williams (2006), the softmax function is used to describe the class membership probability of label y_i :

$$p(y_i | \mathbf{g}_i) = \frac{\exp g_i^{y_i}}{\sum_{k=1}^c \exp g_i^k}. \quad (4)$$

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 the following section we link the outlined theory to regional earthquake risk models where we will also specify the mean and covariance functions of the different Gaussian processes described above.

3. Regional Earthquake Risk Model

Regional earthquake risk models are used to predict the impact of an earthquake scenario on n residential buildings. The impact quantity of interest is the total number of buildings for which the severity of damage prevents immediate re-occupancy, either until emergency structural measures or additional detailed assessments are carried out (yellow damage tag), or because major repair and reconstruction works are required (red damage tag). Apart from the damage tag, the final decision taken by inspectors on the habitability state may depend on external hazards, such as damage to neighboring buildings, and cascading hazards, such as the potential for rockfall or slope instability.

In the present study we assume basic building-specific information to be available from official databases. For one building this information is gathered in a vector $\mathbf{x} \in \mathbb{R}^4$, where the first two entries, $x^{(1:2)}$, contain the geographic coordinates

(in km), $x^{(3)}$ the construction year and $x^{(4)}$ the number of stories.

This section starts with a description of the general workflow to predict damage tags for an earthquake scenario if all buildings are of an identical class, before a framework to account for typological attribution is outlined in Section 3.2.

3.1. Damage tag prediction

We denote the damage tag of building $i \in \{1, \dots, n\}$ as a discrete random variable DT_i , with sequential, collectively exhaustive and mutually exclusive states $dt \in \{1, 2, 3\}$, corresponding to green, yellow and red damage tags, and the total number of buildings with a specific damage tag dt is N_{dt} . For the damage tag prediction on a regional scale, we take the following assumptions, which comply with the state of the art: First, conditional on a specific scenario event, the logarithm of the ground-motion intensity measure, IM, at different building sites follow a multivariate normal distribution. Second, conditional on IM, damage to individual buildings is independent. Third, building capacities are measured in terms of the value of IM at which the building exceeds yellow and red damage tag thresholds following a log-normal distribution with parameter sets $(\eta^{(2)}, \beta^{(2)})$ and $(\eta^{(3)}, \beta^{(3)})$. For the purpose of the present study, we assume the IM to be the peak ground acceleration (PGA).

The analogy to the ordinal GP classification, outlined in Section 2.1, motivates the use of the logarithmic PGA as function $\mathbf{f} = f(\mathbf{X})$, whose prior mean and covariance function are derived with the ground-motion prediction equation (GMPE) of Akkar and Bommer (2010) and the spatial correlation model of Esposito and Iervolino (2011). Specifically, the covariance function is a combination of a constant and exponential kernel:

$$k_f(\tau) = \sigma_W^2 \exp\left(\frac{-\tau}{l}\right) + \sigma_B^2, \quad (5)$$

where $\tau = \|\mathbf{x}^{(1:2)} - \mathbf{x}^{(1:2)'}\|_2$ is the euclidean distance between geographic coordinates. σ_W , and σ_B are the within-event and between-event standard deviations and l denotes the length scale.

Then, similar to conventional fragility functions, the probabilities of exceeding yellow and red damage tag thresholds, for a building of class b_i conditional on the logarithm of PGA at the building site, f_i , are:

$$P(DT_i \geq 2 | f_i, b_i) = \Phi\left(\frac{f_i - \eta^{(2)}(b_i)}{\beta^{(2)}(b_i)}\right) \quad (6a)$$

$$P(DT_i \geq 3 | f_i, b_i) = \Phi\left(\frac{f_i - \eta^{(3)}(b_i)}{\beta^{(3)}(b_i)}\right) \quad (6b)$$

The evaluation of the probability masses for the green and yellow damage tags follow from basic probability calculus. Observe that in contrast to Eq. 3, the dispersion parameter β is not necessarily identical for both cases and thus the condition $\eta^{(2)} < \eta^{(3)}$ might not be sufficient to prevent negative probability masses or, in other words, the fragility functions might cross if parameters are not chosen carefully.

The above formulation can be extended to cases involving multiple IMs. In this 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. Typological attribution model

Although some building attributes can be found in municipal databases, information on the lateral load-resisting system (LLRS) is typically unavailable. As a result, the class definition often relies on expert judgement. Here we assume that 16 representative attribute combinations have been identified, for which capacity parameter sets (η, β) exist.

Because of the available information in $\mathbf{x}_i$, the randomness pertains to the taxonomic attributes related to the LLRS and the material of the floor system, for which we consider four possible combinations $b \in \{1, 2, 3, 4\}$ provided in Table 1 and denoted as building classes in the following. Once the random class B_i is known, we can categorize the building into a building class and use the corresponding damage tag thresholds, as discussed in Section 3.1. In the following we discuss possible models to derive building class membership probabilities as a function of the available inputs $\mathbf{x}_i$.

Table 1. Combinations of taxonomic attributes forming the four considered building classes.

Attribute	$b = 1$	$b = 2$	$b = 3$	$b = 4$
LLRS Material	MUR	MUR	CR	CR
LLRS Type	LWAL	LWAL	LWAL	LFINF
Floor Material	Wood	CR	CR	CR

Note: MUR: Unreinforced Masonry; CR: Reinforced Concrete; LWAL: Wall ; LFINF: Infilled frame

In the absence of region-specific data, an attribution model that is based on expert-opinion presents the most likely alternative. In the present case, experts may estimate the proportions of building classes characterizing different construction periods, denoted as $p_{EO}(B_i = b|\mathbf{x}_i)$ and illustrated in Fig. 1. Within a risk model, a class

sample could be attributed to each building independently. However, whereas p_{EO} might be a valid average over an entire country, it is reasonable to assume that construction practices vary from region to region, thus leading to systematic local deviations from the country-wide average. In addition, such an independent sampling approach neglects probable correlations of classes for geographically close buildings of similar age and geometry. As explained below, we employ GP latent functions to model such variability and to incorporate field observations that are gathered during rapid safety screening after a damaging earthquake.

Fig. 1. Building class proportions as derived from expert-opinion and as a function of the construction year.

In analogy to Section 2.2 we use four independent latent functions, denoted as $g^1, \dots, g^4$ and class membership probabilities are given by Eq. 4. The prior mean function $m_g(\cdot)$ is calibrated such that $p(B_i = b|m_g(\mathbf{x}_i))$ is equal to the expert-based model $p_{EO}(B_i = b|\mathbf{x}_i)$ shown in Fig. 1. For the covariance function, we chose a combination of squared-exponential kernels that are of the following general type:

$$k(\tau) = \sigma^2 \exp \left(-\frac{(\tau)^2}{2l^2} \right), \quad (7)$$

where τ denotes the euclidean distance between input dimension d , and σ^2 and l denote the variance and the length scale. Multiple covariance functions are combined in order to account for long-scale correlated deviations due to regional differences in construction practices, $k_{g,LS}$, and short-scale correlated deviations for classes of close buildings that are similar w.r.t. their age and number of stories $k_{g,SS}$. The latter is a multiplicative combination of three kernels acting on different input dimensions and the final covariance

function is denoted as:

$$k_g(\cdot) = k_{g,LS} + k_{g,SS,1} \cdot k_{g,SS,2} \cdot k_{g,SS,3}, \quad (8)$$

The relevant hyperparameters and input dimensions are further specified in Table 2.

Table 2. Parameters for the different covariance functions composing $k_g(\cdot)$.

	variance σ^2	lengthscale l	active dim. d
$k_{g,LS}(\cdot)$	0.8	5	$\mathbf{x}^{(1:2)}$
$k_{g,SS,1}(\cdot)$	0.6	5	$\mathbf{x}^{(3)}$
$k_{g,SS,2}(\cdot)$	0.6	2	$\mathbf{x}^{(4)}$
$k_{g,SS,3}(\cdot)$	0.6	0.15	$\mathbf{x}^{(1:2)}$

3.3. Simulation workflow

Following workflow is employed to estimate impacts for a specified earthquake scenario using the model described above: First, we generate independently r samples from the multivariate normal distributions $p(\mathbf{f})$ and $p(\mathbf{g})$. Then, for each sample and for each building, we first sample a building class b_i from the categorical distribution $B_i|\mathbf{g}_i$ (Eq.4) and subsequently sample a damage tag dt_i from the categorical distribution $DT_i|f_i, b_i$ (Eq.6). By aggregating buildings with identical damage tag and building class for each sample we obtain r realizations of the total number of buildings with a certain damage tag, N_{dt} , and building class, N_b , respectively.

In case of early post-earthquake assessments the available information, gathered by seismic networks or field surveys, should be incorporated to allow for more constrained estimates. This is discussed in the following section.

3.4. Bayesian inference workflow

This section illustrates how the concepts of GP-based ordinal regression (Section 2.1) and multi-class classification (Section 2.2) can be used to incorporate data gathered in early phases after an event. First we consider seismic network stations at locations $\mathbf{X}_s$ providing, typically noisy, measurements $\mathbf{z}_s \in \mathbb{R}$ of ground-motion intensity measures and thus also of the function f , which is used for a first inference step. Assuming independent and normally distributed noise, the posterior predictive distribution $p(\mathbf{f}|\mathbf{z}_s)$ is analytically tractable and itself a multivariate normal distribution (see Worden et al. (2018) for a detailed treatment in the context of shake maps). In the absence of any further information we can use

this distribution instead of $p(\mathbf{f})$ in the simulation workflow described in Section 3.3 to estimate first rapid impact assessments.

After the damage event, collection of building-specific damage data is usually done within a rapid safety screening. We denote the input information for buildings screened in the first t days as $\mathbf{X}$ and the gathered damage tag and building class data as $\mathbf{dt}$ and $\mathbf{b}$, respectively. The input information for the remaining, uninspected, buildings is stored in $\mathbf{X}_*$. The next inference steps are to derive posterior distributions $p(\mathbf{f}|\mathbf{dt}, \mathbf{b}, \mathbf{z}_s)$ and $p(\mathbf{g}|\mathbf{b})$. However, because of the non-Gaussian likelihoods in Eq. 3 and Eq. 4 those distributions are not analytically tractable. In this study we resort to the method of variational inference, where the true posterior is approximated by a Gaussian distribution $\tilde{p}(\cdot)$ whose parameters are obtained by minimizing the Kullback-Leibler divergence between $\tilde{p}(\cdot)$ and the exact posterior $p(\cdot)$ (Opper and Archambeau, 2009). Thus, also the approximate posterior predictive distributions $\tilde{p}(\mathbf{f}_*|\mathbf{dt}, \mathbf{b}, \mathbf{z}_s)$ and $\tilde{p}(\mathbf{g}_*|\mathbf{b})$ are multivariate normal distributions and by employing the workflow outlined in Section 3.3 we arrive at increasingly constrained estimates of N_{dt} and N_b , respectively.

The hyperparameters of the latent function g , namely the variances and length scales indicated in Table 2, are optimized during the inference process by means of maximum a-posteriori estimates obtained with variational inference. All other hyperparameters are kept fixed. The inference schema described above is numerically implemented using the software library GPFlow (Matthews et al., 2017).

4. Case-Study

The impact of a scenario earthquake on the residential buildings within the Canton of Zurich (Switzerland) is simulated to illustrate the concepts outlined in the preceding sections. Section 4.1 presents the layout and the simulated impacts as well as the data-generating models. The latter uses hazard, typological attribution and building fragility models that differ from those available in the regional risk model, explained in Section 3. Finally, inference results are presented in Section 4.2.

4.1. Layout and earthquake scenario

The map in Fig. 2 shows the Canton of Zurich, the epicenter of the considered $M_w = 6.0$ scenario earthquake and the locations of the seismic network stations available within 50km from the epicenter. In this study, we focus on impacts in the zip-code highlighted in red. The building specific information $\mathbf{X}_*$, discussed in Section 3, is made publicly available by the office for spatial development of the Canton of Zurich (2020). The considered zip-code contains 1900 residential buildings,

56% of which have been built before 1945 and 79% of which are three to five stories tall.

Fig. 2. Location of the considered zip-code (highlighted in red) within the Canton of Zurich, assumed locations of seismic network stations and of the epicenter of the magnitude six earthquake scenario.

As a result of the scenario event 9% and 27% of the buildings in the zip-code received a red and yellow damage tag, respectively. The rapid safety screening campaign started one day after the event happened and, as illustrated in Fig. 3, after eight days 200 buildings have been screened, corresponding to approx. 10% of the entire building stock. The assigned damage tags and building classes are also indicated.

Fig. 3. Accumulation of screened buildings up to certain days after the event, together with the assigned damage tag dt and building class b .

The models used to produce the simulated dataset are deliberately chosen to differ from the

models used in the prior risk model outlined in Section 3. In addition to PGA we used 5% damped spectral acceleration at periods of 0.3, 0.6 and 1.0 seconds. The building stock is categorized using building classes (and fragility curves) defined in the global risk model (Martins and Silva, 2020). We manually assigned a class to each building by assuming an identical class for all buildings in a street block that share similar construction year and number of stories. The corresponding fragility curves are used to estimate the damage state of each building, which is then translated to green, yellow and red damage tags using a conditional probability table. For the safety screening process we assume a constant number of available inspection teams that are randomly assigned to a building at the beginning of every day. After completing the first inspection of a building, each team continues to inspect the geographically closest uninspected building.

4.2. Results

Applying the method presented in Section 3, we perform approximate Bayesian inference using the data available until the end of each day $t \in \{0, 1, \dots, 8\}$ illustrated in Fig. 3 and use the inferred predictive posterior distributions to estimate the distribution of the total number of buildings of class b , N_b , and damage tag dt , N_{dt} . In the following results are normalized by the total number of buildings n within the considered zip code.

4.2.1. Damage tag proportions

Fig. 4 shows frequency histograms of the proportion of buildings with yellow ($dt=2$) and red ($dt=3$) damage tag, obtained via the workflow described in Sections 3.3 and 3.4 for 10000 realizations. For the first post-earthquake predictions at day 0 ($t=0$) only seismometer measurements $\mathbf{z}_s$ are available, which results in mean predictions of 14% and 4% for proportions of yellow-tagged and red-tagged buildings, respectively. The 90% credibility interval spans from 1.5% to 33% for the yellow and from 0% to 16% for the red damage tag proportions. Recalling the actual proportions of 27% and 9%, we note that the predictions underestimate, on average, the actual results for both damage tags.

In the following days, leveraging the information from rapid safety screening, the uncertainty in the predictions is reduced and the mean predictions approaches the actual results. Specifically, the predictions done at day 8 after the event result in mean estimates of 25% and 8.5%, with credibility intervals of (22,28)% and (6,10)%, respectively.

Fig. 4. Frequency histograms of the proportion of buildings with yellow damage tag ($dt=2$) and red damage tag ($dt=3$), calculated using no inspection data ($t=0$) and data gathered up to day 2, 4 and 8 after the event. The dashed black line plots the actual proportions.

4.2.2. Building class proportions

The results presented in the preceding section reflect simultaneous updating of earthquake intensity (via the function f and the observed damage tag data) and region-specific building class proportions (via the function g and the observed building classes). Fig. 5 illustrates the predictions of the latter as a function of the ratio of buildings that have been inspected and, thus, were available for inference. We observe that for building class $b=1$ the initial predictions, obtained with the prior model outlined in Section 3.2, result in an average proportion of 42% whereas the actual proportion amounts to 63%. By subsequently processing the continuous inflow of inspection data, we note a reduction in the difference.

4.2.3. Effect of using a fixed typological attribution model

The effect of not performing a parallel updating of building class proportions for predicting damage tag proportions is studied by following the workflow described in Sections 3.3 and 3.4, except that building classes are sampled independently from the fixed, expert-based attribution model $p_{EO}(B_i = b | \mathbf{x}_i)$ (see Fig. 1), instead of using the updated estimates shown in Fig. 5. Fig. 6 shows the results in an identical layout as Fig. 4. We observe that the predictions at later time steps (days 4 and 8) fail to capture the actual proportions of buildings with yellow and red damage tags, because the model continuously underestimates

Fig. 5. Predictions of the proportion of buildings being of type b , N_b/n , as a function of the number of inspected buildings available for inference. The gray line and shaded area indicate the mean and the 90% credibility interval of the predictions. The dashed black line plots the actual proportions.

the number of buildings of the most vulnerable class $b=1$.

Fig. 6. Frequency histograms of damage tag proportions calculated using a fixed typological attribution model. See Fig. 4 for additional explanations.

5. Conclusion

Gaussian-process based classification is used to improve knowledge regarding spatial distribution of earthquake intensity and attribution of typological building classes based on building attributes.

Thus, a reduction in the uncertainty ranges of predicted housing capacity loss in a earthquake-hit region is achieved. Within the scope of the simulated scenario earthquake, the following preliminary conclusions are drawn:

- Classification techniques show potential to provide stable and precise predictions of the aggregated loss of housing capacity in a fraction of the time required to inspect the entire building stock.
- Incomplete inspection data from random building sets lead to spatially accurate distributions of unsafe buildings and provide a significant uncertainty reduction with respect to traditional rapid loss assessment based on shake maps.
- Simultaneous updates of building-class attribution models, in addition to earthquake intensity measures, reduces local errors in typological attribution models stemming from high-level aggregation.
- Gaussian Process classification is robust to inevitable discrepancies between the updated regional risk model and the underlying unknown behavior of buildings undergoing an earthquake.

In this paper, earthquake consequences are limited to the habitability of buildings, governing the loss of regional housing capacity. However, the framework can be applied to other important consequences, such as financial losses or recovery times. Notably, the choice of visual inspection as data source reflects the current state of the art. Yet, the framework can be modified to ingest other post-earthquake damage-information sources, such as remote sensing or structural health monitoring applications.

Acknowledgement

The presented work 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. and J. J. Bommer (2010, 3). Empirical Equations for the Prediction of PGA, PGV, and Spectral Accelerations in Europe, the Mediterranean Region, and the Middle East. *Seismological Research Letters* 81(2), 195–206.

Amt für Raumentwicklung (2020). Geografisches Informationssystem des Kantons Zürich (GIS-ZH).

Brzev, S., C. Scawthorn, A. W. Charleson, L. Allen, M. Greene, K. Jaiswal, and V. Silva (2013). GEM

Building Taxonomy Version 2.0. *GEM Technical Report* 02, 188.

Chu, W. and Z. Ghahramani (2005). Gaussian Processes for Ordinal Regression. *Journal of Machine Learning Research* 6, 1019–1041.

Crowley, H., V. Despotaki, D. Rodrigues, V. Silva, D. Toma-Danila, E. Riga, A. Karatzetou, S. Fotopoulou, Z. Zugic, L. Sousa, S. Ozcebe, and P. Gamba (2020, 10). Exposure model for European seismic risk assessment. *Earthquake Spectra* 36(1_suppl), 252–273.

Diana, L., J. Thiriot, Y. Reuland, and P. Lestuzzi (2019, 4). 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.

Erdik, M., K. Şeşetyan, M. B. Demircioğlu, U. Hancilar, and C. Zülfikar (2011, 2). Rapid earthquake loss assessment after damaging earthquakes. *Soil Dynamics and Earthquake Engineering* 31(2), 247–266.

Esposito, S. and I. Iervolino (2011, 10). PGA and PGV Spatial Correlation Models Based on European Multievent Datasets. *Bulletin of the Seismological Society of America* 101(5), 2532–2541.

Kovačević, M., Z. Stojadinović, D. Marinković, and B. Stojadinović (2018). Sampling and machine learning methods for a rapid earthquake loss assessment system. In *11th U.S. National Conference on Earthquake Engineering*, Los Angeles, CA. Earthquake Engineering Research Institute.

Martins, L. and V. Silva (2020, 6). Development of a fragility and vulnerability model for global seismic risk analyses. *Bulletin of Earthquake Engineering*.

Matthews, A., M. van der Wilk, T. Nickson, K. Fujii, A. Boukouvalas, P. León-Villagrà, Z. Ghahramani, and J. Hensman (2017). GPflow: A Gaussian Process Library using TensorFlow. *Journal of Machine Learning Research* 18, 1–6.

Opper, M. and C. Archambeau (2009). The variational gaussian approximation revisited. *Neural Computation* 21(3), 786–792.

Pozzi, M. and Q. Wang (2018, 3). Gaussian Process Regression and Classification for Probabilistic Damage Assessment of Spatially Distributed Systems. *KSCE Journal of Civil Engineering* 22(3), 1016–1026.

Rasmussen, C. and K. Williams (2006). *Gaussian processes for machine learning*. Cambridge, MA: The MIT Press.

Worden, C., E. M. Thompson, J. W. Baker, B. A. Bradley, N. Luco, and D. J. Wald (2018). Spatial and spectral interpolation of ground-motion intensity measure observations. *Bulletin of the Seismological Society of America* 108(2), 866–875.