ORIGINAL PAPER

Exploring the impact of epistemic uncertainty on a regional probabilistic seismic risk assessment model

Petros Kalakonas[1] · Vitor Silva[2] · Amaryllis Mouyiannou[3] · Anirudh Rao[2]

Abstract

Probabilistic earthquake loss models are widely used in the (re)insurance industry to assess the seismic risk of portfolios of assets and to inform pricing mechanisms for (re)insurance contracts, as well as by international and national organizations with the remit to assess and reduce disaster risk. Such models include components characterizing the seismicity of the region, the ground motion intensity, the building inventory, and the vulnerability of the assets exposed to ground shaking. Each component is characterized by a large uncertainty, which can be classified as aleatory or epistemic. Modern seismic risk assessment models often neglect some sources of uncertainty, which can lead to biased loss estimates or to an underestimation of the existing uncertainty. This study focuses on exploring and quantifying the impact of a number of sources of uncertainties from each component of an earthquake loss model to the loss estimates. To this end, the residential exposure of Guatemala and Guatemala City were used as case studies. Moreover, a comparison of the predicted losses for an insured portfolio in the country between OpenQuake-engine and a vendor catastrophe platform was performed, assessing the potential application of OpenQuake in the (re)insurance industry. The findings from this study suggest that the uncertainty in the hazard component has the most significant effect on the loss estimates.

1 Introduction

Probabilistic earthquake loss models, similar to any other catastrophe model (CAT), are structured by four main components: hazard, exposure, vulnerability and a financial model. Such models are characterized by a large variability, due to the uncertainties associated with each component. These sources of uncertainty originate from the input of numerous

✉ Petros Kalakonas
petros.kalakonas@iusspavia.it

1 University School for Advanced Studies (IUSS), Pavia, Italy

2 Global Earthquake Model Foundation (GEM), Pavia, Italy

3 Partner Re, Zurich, Switzerland

parameters that define the seismicity, the ground motion intensity, the seismic vulnerability and the exposure characteristics of a building inventory. Inevitably, the loss estimates carry a high degree of variability, as different assumptions during the modelling process can lead to considerably different results (e.g. Silva 2018). Such uncertainties can be classified into *aleatory* and *epistemic*, depending on their nature and physical basis. Epistemic uncertainty arises from incomplete scientific knowledge and can be reduced in principle, such as through increased data or advanced scientific principles.

These uncertainties can also be classified into primary and secondary, according to their source and order within the model. The former represents both the epistemic and aleatory uncertainty associated with the event generation process and occurrence of earthquakes, while the latter represents the uncertainties involved in loss estimation given that an event has occurred. Although the treatment of uncertainties in risk analysis has been the topic of a wide discussion (e.g. Paté-Cornell 1996; Faber 2005; Spiegelhalter and Riesch 2011), modern catastrophe loss models tend to oversimplify risk estimates, leading to an insufficient treatment and quantification of the uncertainties (e.g. Taylor 2015). Such simplifications are often due to the need to reduce the computational demand required by each model. As demonstrated by Bazzurro and Luco (2007), the results of a loss analysis that does not properly incorporate all the sources of uncertainty can be misleading and lead to an underestimation of the actual risk. The work presented herein focuses on the impact of epistemic uncertainty in loss estimation of spatially distributed building portfolios.

The hazard component reflects the extent (spatial distribution) and intensity of potential earthquakes. Several platforms for probabilistic seismic hazard analysis (PSHA, e.g. Cornell 1968) are currently available (e.g. OpenSHA—Field et al. 2003; CRISIS—Ordaz et al. 2013; OpenQuake-engine—Pagani et al. 2014). In this process, the seismic source model is used to generate an earthquake rupture forecast, which represents all the possible ruptures that can occur in the region of interest and its associated rate of occurrence. The development of a seismic source model for a region is a complex process involving numerous tasks, such as the derivation of geological and tectonic maps, the development of an earthquake catalogue and the identification of crustal faults and faults systems. The majority of the assumptions and decisions during the development of a source model are subjected to epistemic uncertainties. For instance, the absence of information related to the date, size, focal mechanism, and hypocentral locations of past earthquakes, or the existence of unmapped faults (e.g. Hayes et al. 2010) may significantly affect the geometry and parameters of a modelled seismic zonation. The maximum magnitude of each source is also affected by epistemic uncertainty, as it is defined from geologic data and the historical catalogue of a region. Furthermore, the selection of ground motion prediction equations (GMPEs), which represent the intensity of shaking at a given site due to a rupture scenario, has been indicated as one of the largest epistemic uncertainties in earthquake loss estimation (e.g. Crowley et al. 2005).

An exposure model contains the information regarding the assets potentially at risk. One important source of uncertainty arises from the grouping of the building inventory in a certain number of building classes (e.g. Pittore et al. 2018), which depends on the detail of exposure information (primary and secondary modifiers), but also on the availability of distinct vulnerability functions. Typically, exposure modellers apply a mapping scheme to the available exposure data to relate each asset with a vulnerability function, which is inevitably subjected to epistemic uncertainty. Another essential source of uncertainty is related to the spatial resolution of an exposure model, which originates in the aggregation of the buildings within each exposure unit. For example, regional or national seismic risk assessments (e.g. Yepes-Estrada et al. 2017) use information from remote sensing and

national housing databases, where information is typically provided at a coarse resolution (e.g. district or municipality level). On the other hand, exposure models of portfolios of buildings in the (re)insurance industry can be developed at a high resolution, as the location of each asset is documented. In these cases, the spatial aggregation of the assets is performed in order to minimize the computational effort. This aggregation and relocation of buildings results in a misrepresentation of the distance between the assets and the seismic sources, and the implicit (full) correlation in the ground motion for all assets at a given location (Bazzurro and Park 2007). Bal et al. (2010) conducted a study on the impact of urban exposure resolution, and one important conclusion is that the loss estimates become accurate and stable beyond a certain (fine) spatial resolution. A potential way to reduce this type of uncertainty is by improving the detail of information concerning the location of the building inventory. However, this process can be time and resource-demanding, and in many cases it is simply impractical (e.g. risk analysis at the national level). Nevertheless, there are effective alternatives which may involve the disaggregation of the exposure in each unit using night-time lights, satellite imagery or the location of roads (e.g. Dabbeek and Silva 2020).

Vulnerability models are typically a set of damage (or vulnerability) functions, which provide mean damage ratio (also known as loss ratio) conditioned on a set of intensity measure levels (IM). Depending on the derivation methodology, vulnerability functions are categorized into empirical (e.g. Colombi et al. 2008; Rossetto et al. 2015), analytical (e.g. D'Ayala et al. 2014; Yepes-Estrada et al. 2016), or hybrid (e.g. Kappos et al. 2006). In general, the advantage of the analytical approach is that it facilitates the modelling and propagation of the effects of epistemic and aleatory uncertainty in ground motion, selection of structural response parameters, characteristics of the buildings, and correlation in damage or loss (e.g. Sousa et al. 2016). Regardless of the derivation methodology, vulnerability functions are characterized by a large uncertainty (e.g. Porter 2010), principally due to four main sources (Silva 2019): (1) record-to-record variability, (2) building-to-building variability, (3) uncertainty in the damage criteria and (4) uncertainty in the damage to loss model. Crowley et al. (2005) investigated the impact of epistemic uncertainties in the seismic demand and capacity on an earthquake loss model for Istanbul. One of the key conclusions is that the capacity parameters, and generally the epistemic uncertainty in the derivation of vulnerability functions, have the most significant impact on the loss estimates. Moreover, the choice of the engineering demand parameter (EDP) used to estimate the expected damage is affected by considerable uncertainty as the same values of an EDP could lead to different damage levels, and consequently to distinct loss ratios (e.g. Martins et al. 2016).

The CAT vendor platforms used in the (re)insurance industry treat, incorporate, and propagate uncertainty to the loss metrics in various manners. The latter task is considered to be one of the most challenging parts of a CAT model, and it is where significant differences amongst models and platforms are located (e.g. Mitchell-Wallace et al. 2017). Even though there is a recognized need for more transparent models, common vendor models often do not enable the quantification of the impact of uncertainty in some of the components. In this study the sources of epistemic uncertainty previously described are explored using an open platform for seismic hazard and risk analysis (OpenQuake-engine).

This study was carried out during the development of the global seismic risk model (Silva et al. 2020) by the Global Earthquake Model Foundation (GEM). This global model is comprised by almost 200 exposure datasets, 30 seismic hazard models (Pagani et al. 2018) and more than 600 vulnerability functions (Martins and Silva 2020). In order to evaluate the impact of several sources of epistemic uncertainty in earthquake risk analysis, the loss model

for the country of Guatemala was used as a case study. The quality and detail of the components for Guatemala are comparable with the vast majority of the models in the global seismic risk model. Moreover, the resolution and quality of many models used in the (re)insurance industry are similar to that of Guatemala. Finally, sensitivity studies using such models can inform decision makers responsible for risk management in regions with high seismic risk and a relatively low insurance penetration. The authors acknowledge, however, that the findings from this study are contingent on the characteristics of the case study, as further discussed in the final remarks.

2 Case study: Guatemala

2.1 Seismic hazard, exposure and vulnerability models

Central America is one of the most active seismic regions in the world (e.g. Alvarado et al. 2017). The tectonic setting of the region is complex due to the interaction of the boundaries of five tectonic plates: North American, Caribbean, Coco's, Nazca and South American (e.g. Benito et al. 2012; Alvarado et al. 2017). Some examples of past destructive events include the 1972 Managua (M_w 6.3) and 1976 Guatemala (M_w 7.5) earthquakes, which caused more than 10,000 and 20,000 fatalities, respectively. Earthquake risk is particularly important in Guatemala, as it is the most populated country in the Central America with 17.25 million inhabitants, and its capital Guatemala City (1 million citizens) is located in an area characterized by high seismic hazard (e.g. Villagran et al. 1996) due to its proximity to both the subduction zone and crustal active faults (e.g. Motagua fault). Furthermore, the first seismic design regulation was implemented only in 1996, and consequently a large portion of the building stock does not have adequate seismic provisions.

The seismic hazard model used herein is from a PSHA study covering Central America, the RESIS II project (Molina et al. 2008; Benito et al. 2012,). In that study, three seismogenic models were considered and associated with focal depths: *crustal seismicity* ($h \leq 25$ km), *subduction interface* (25 km $< h \leq 60$ km) and *subduction intra-slab* ($h > 60$ km). Two seismic zonations (i.e. seismic source models) for Central America were proposed, with different degrees of detail: *Regional* and *National*. The former distinguishes large seismogenic zones that encompass the main seismic tectonic units of the region, while the latter (Benito et al. 2012) has a greater number of seismogenic zones within each country, avoiding discontinuities at the national boundaries. It should be noted that both seismic source models include the three tectonic environments, but differ in the geometry and parameters of the seismic area sources, such as the magnitude frequency distribution and maximum magnitude. The hazard models consider the epistemic uncertainty in the selection of the GMPEs through the implementation of a logic tree, as presented in Table 1.

In the present study, a site model for Guatemala was implemented using Vs_{30} proxy values obtained from the USGS Vs_{30} server, following the slope topography methodology proposed

Table 1 GMPEs and the associated weights per tectonic region used by Benito et al. (2012)

Active shallow crust	Subduction interface	Subduction intra-slab
Climent et al. (1994), 0.5	Youngs et al. (1997), 1.0	Youngs et al. (1997), 0.5
Zhao et al. (2006), 0.5		Zhao et al. (2006), 0.5

by Wald and Allen (2007). Classical PSHA analyses were carried out using the OpenQuake-engine (Pagani et al. 2014), and hazard maps of Guatemala, in terms of peak ground acceleration (PGA), for the 500-year return period (i.e. 9.5% probability of exceedance in 50 years) for both zonations were obtained, as presented in Fig. 1.

The exposure model was exported from Calderon et al. (2020), in which a regional exposure model for the Central American countries was developed based on the national census databases, World Housing Encyclopedia reports, information available in the literature and local expert judgement. Calderon et al. (2020) used the GEM Building Taxonomy (Brzev et al. 2013) to classify the building stock in a uniform manner. In the case of Guatemala, the data from the most recent national census survey developed in 2002 were utilized. In the current study, only residential buildings were considered and the structural characteristics of the buildings stock are described in Table 2.

The distribution of construction material of the residential building stock at the national level and in Guatemala City (i.e. 12–13% of the total building stock) are depicted in Fig. 2. This distribution of the different building classes across the country or its capital is mostly defined based on a combination of data from the housing census and the opinion of local engineers. Consequently, it is also a component of the risk model that is affected by epistemic uncertainty. Another source of uncertainty in the exposure model is related with the spatial distribution of the assets in the region. The data for Guatemala is publicly available at the municipality level, and therefore a spatial disaggregation of the assets within each administrative area may be required to avoid an over-aggregation of the assets (e.g. Bazzurro and Park 2007), and consequently the introduction of a full correlation in both the ground shaking and the vulnerability. In this study, the exposure model was spatially disaggregated into evenly spaced grids to assess the impact of employing different spatial resolutions.

The vulnerability model for Guatemala was adopted from Martins and Silva (2020), which consists of 34 distinct vulnerability functions for the aforementioned building classes. The analytical derivation methodology incorporates building-to-building and record-to-record variability, through the generation of a large set of single-degree-of-freedom systems and the consideration of a large set of ground motion records (GMRs), respectively. In particular, Martins and Silva (2020) generated 150 capacity curves for each building class using a Monte Carlo simulation (Silva et al. 2014b), and selected over 300 GMRs associated to earthquakes from subduction and active shallow tectonic regimes. The records were scaled using factors below 2, and for each intensity measure type 10 levels of acceleration were considered in the range between 0.05 and 2.0 g. The fragility and corresponding vulnerability functions were derived using nonlinear dynamic analyses.

The definition of the loss ratios per intensity measure can follow a deterministic (i.e. only mean values are considered) or probabilistic (i.e. a probability model is used to define the distribution of loss ratio) approach. For the latter, either a lognormal (e.g. Porter 2010; Rodrigues et al. 2018) or a beta (e.g. Lallemant and Kiremidjian 2014; Silva 2019) distribution can be adopted. These modelling options are also explored in this study.

2.2 Selection of epistemic uncertainties

The selection of epistemic uncertainties for investigation was influenced by the level of sophistication of the adopted models, the capabilities and limitations of the OpenQuake-engine, and the availability of additional information (e.g. alternative fragility functions).
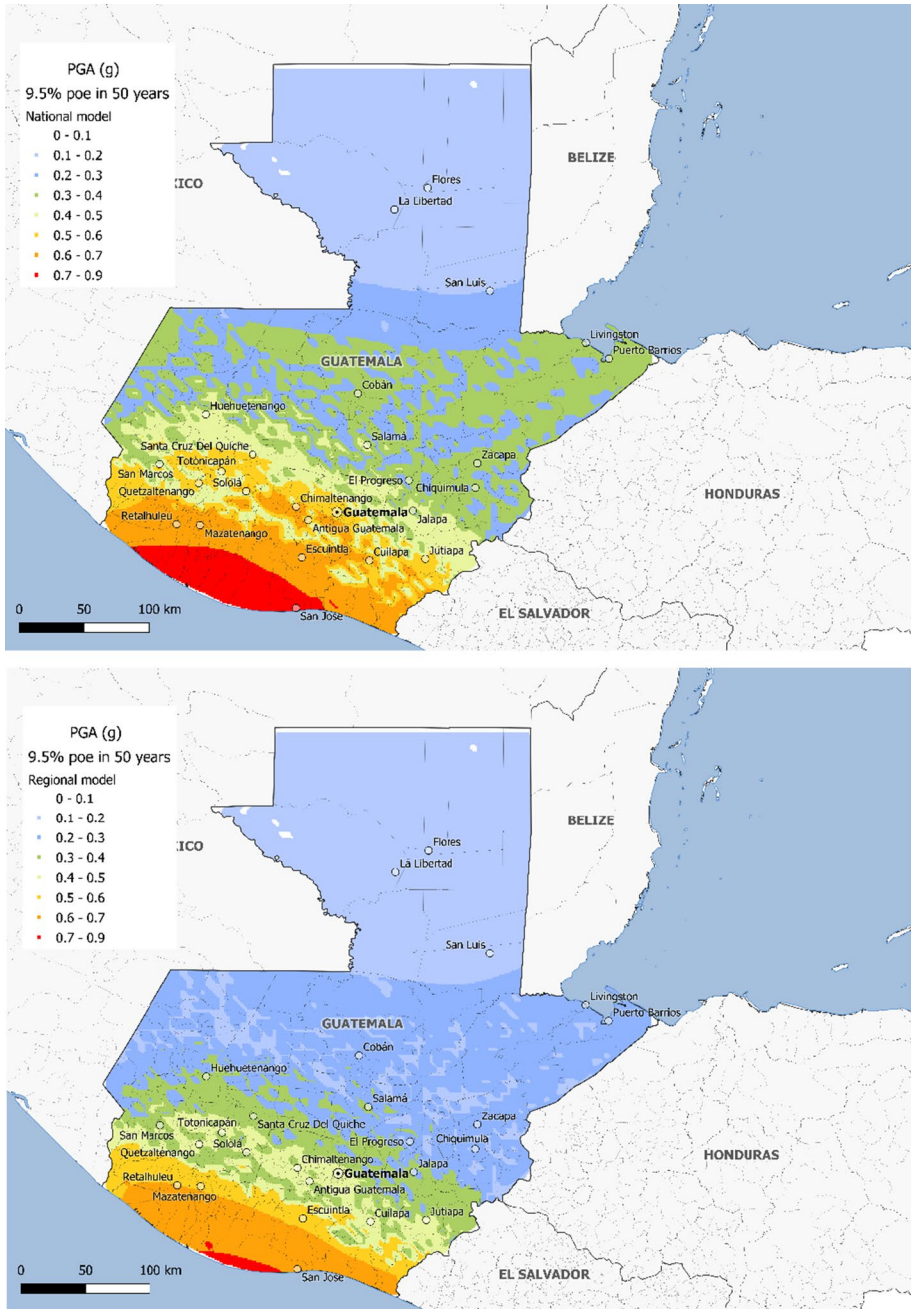
Fig. 1 Hazard maps of Guatemala for 500-years return period on soil conditions, using National (*top*) and Regional (*bottom*) zonation

Table 2 Building classes of the residential building stock of Guatemala

| Building class	Material	LLRS	Ductility	Storeys
W + WLI/LWAL + DNO	Wood	Panels	Non-ductile	1 and 2
W + WBB/LPB + DNO	Wood	Post and beams	Non-ductile	1
W + WWD/LWAL + DNO	Wood and earth	Wall	Non-ductile	1
W + WS/LPB + DNO	Wood	Post and beams	Non-ductile	1
MUR + ADO/LWAL + DNO	Adobe masonry	Wall	Non-ductile	1 and 2
MUR/LWAL + DNO	Unreinf. masonry	Wall	Non-ductile	1 and 2
MR/LWAL + DNO	Reinf. masonry	Wall	Non-ductile	1–3
MR/LWAL + DUM	Reinf. masonry	Wall	Moderate	1–3
MCF/LWAL + DNO	Conf. masonry	Wall	Non-ductile	1–3
MCF/LWAL + DUM	Conf. masonry	Wall	Moderate	1–3
CR/LFM + DUL	Reinf. concrete	Bare frame	Low	1–2
CR/LFINF + DUL	Reinf. concrete	Infilled frame	Low	2–5
CR/LFINF + DUM	Reinf. concrete	Infilled frame	Moderate	3–6
CR/LDUAL + DUH	Reinf. concrete	Wall and frame	High	7–12
UNK	Unknown	Unknown	Low	1

LLRS, lateral load resisting system
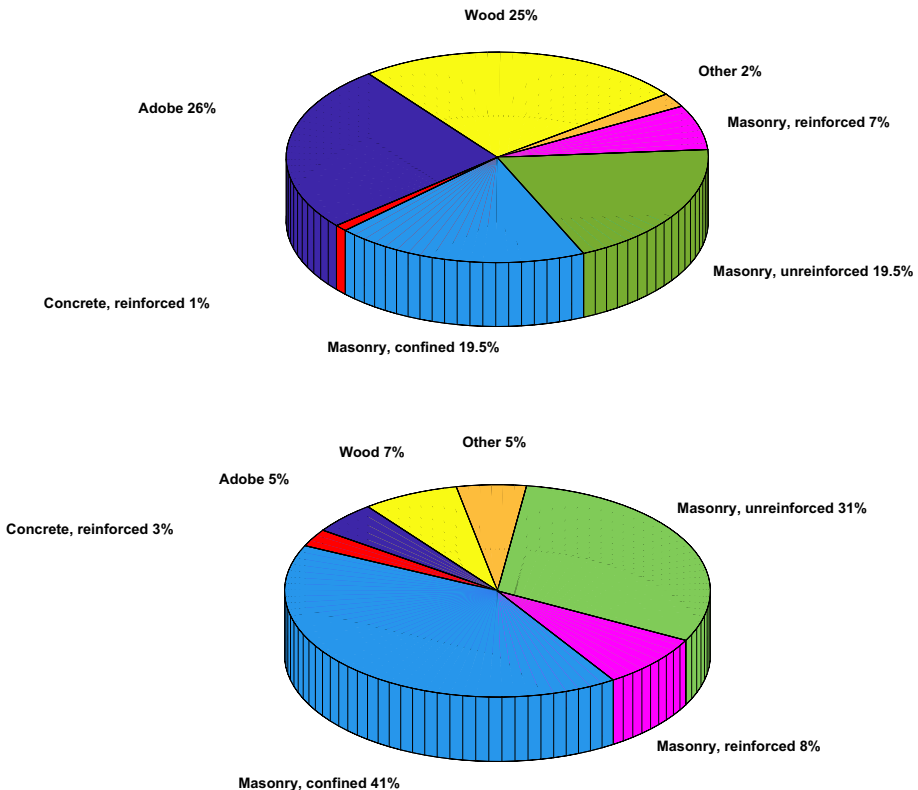
Fig. 2 Construction material distribution of the residential building inventory of Guatemala (*top*) and Guatemala City (*bottom*)
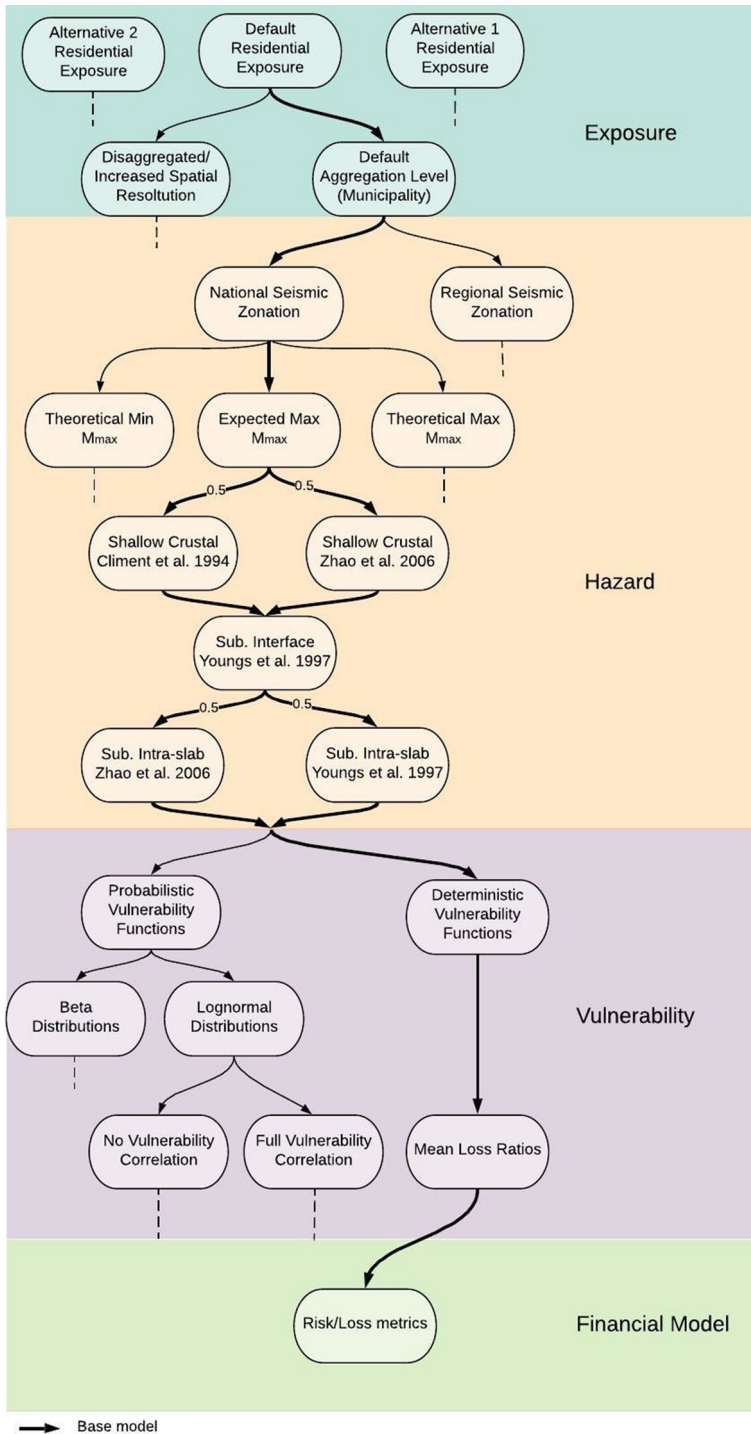
Fig. 3 Schematic representation of the considered epistemic uncertainties

The list of uncertainties that are explored in this study are represented in Fig. 3, and a description of each uncertainty is provided below.

Seismic hazard

1. *Seismic zonation*: the national and regional zonation were considered as two alternative modelling realizations of the seismicity of the region.
2. *Maximum M_w*: in addition to the default option (expected M_{max}), the theoretical maximum and minimum M_{max} for each seismic zone/source of the national zonation from Benito et al. (2012) were considered.
3. *GMPEs selection*: the influence of the selected GMPEs for each tectonic region in the aforementioned logic tree was assessed.

Exposure

1. *Spatial resolution*: the default exposure model was disaggregated in five resolutions ranging from 480 to 30 arc-seconds.

The methodology proposed by Dabbeek and Silva (2020) was used to disaggregate the exposure model within each municipality. In this methodology, night-time lights are used to re-allocate the assets. The night-time lights data show the reflected lights from different sources such as human settlements, industrial parks, roads and railways. Figure 4 depicts the night-time lights for Guatemala (Román et al. 2018), where Guatemala City is located in the area with the highest concentration of light. It is worth mentioning that the base exposure model consists of 334 exposure points (municipalities), while the 480 and 30 arc-second resolution models resulted into 314 and 30,842 locations, respectively. In the case of Guatemala City, the base model consists of 4 exposure locations, while the 480 and 30 arc-second resolution models resulted in 6 and 448 locations, respectively.

2. *Distribution of construction material*: Alternative versions of the residential exposure in terms of the distribution of construction material were tested, based on the judgment of local engineers.

Fig. 4 Night-time lights of Guatemala

During the development of the residential exposure model for Central America, Calderon et al. (2020) carried out surveys with local experts and engineers regarding the most common types of construction in urban and rural areas in the country. For the case of Guatemala, two alternative variations of the exposure model in terms of construction material distribution at the national level were proposed, as presented in Fig. 5. These distributions were used to derive two alternative exposure models of the residential building stock for Guatemala.

Vulnerability

1. *Loss ratio distribution*: The uncertainty in the loss ratio (LR) conditioned on different intensity levels, arising from the propagation of the building-to-building and record-to-record variability was modelled using the mean values, beta and lognormal distributions.

The consideration of the uncertainty in the LR was performed following the procedure proposed by Silva (2019), where a standard deviation (σ_{LR}) is calculated based on the expected LR. The impact in the consideration of correlation in the vulnerability was also assessed. However, due to limitations in the loss modelling approach, correlation was only considered for the vulnerability functions using a lognormal distribution (and not for the beta distribution).
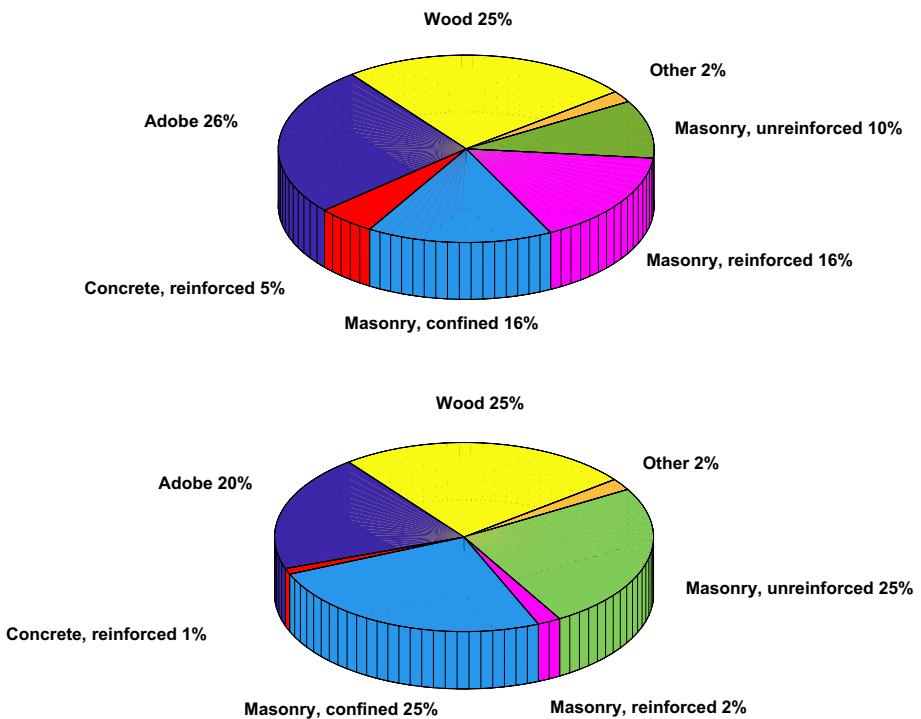

Fig. 5 First (*top*) and second (*bottom*) alternative version of the residential exposure as per construction material distribution

3 The impact of epistemic uncertainties in loss assessment

The impact of the described uncertainties in the loss metrics of Guatemala and Guatemala City was quantified through sensitivity analysis. A "base model" was defined in each case (see thicker branches in Fig. 3), by selecting the "default" modelling option for each component. More specifically, the base country and city models were defined considering: the building classes distribution and spatial resolution (municipality level) developed by Calderon et al. (2020), the national seismic zonation, the expected M_{max}, the mean seismic hazard among the four GMPEs combinations, and the deterministic vulnerability functions (i.e. mean LRs). In all cases, event-based risk analyses were carried out using the Open-Quake-engine (Silva et al. 2014a). A stochastic event set with a length of 250,000 years was considered, and loss exceedance curves are presented until the 2500-year return period, following the recommendations of Silva (2018). The loss metrics are presented in terms of loss ratio (as a percentage) rather than absolute monetary values, as the purpose of this study is to investigate the relative impact of the selected epistemic uncertainties and not to focus on the actual monetary loss.

The loss exceedance curves for varying the hazard and exposure parameters are illustrated in Figs. 6 and 7, respectively. Each curve represents the results of the risk assessment by altering only the denoted modelling option in the legend, in respect to the base model. Furthermore, the tornado (sensitivity) plots for two selected return periods (100 and 2500 years) and the average annualized losses (AAL) are presented in Figs. 8, 9 and 10. Regarding the tornado plots, the variation from the base model is shown by altering one parameter at a time (y-axis), while the base value is indicated in each figure by the vertical black line. It should be noted that the abbreviation used for the GMPEs selection follows the notation: C→Climent et al. (1994), Y→Youngs et al. (1997), and Z→Zhao et al. (2006), while the order of the letters denotes the GMPE for the active shallow, subduction interface and subduction intra-slab, respectively.

Utilizing the regional zonation source model resulted in the lowest losses amongst all parameters. The impact of regional zonation appears to be constant through the entire range of return periods, for both the City's and country's losses. Such outcome was expected since significant differences in the seismic hazard estimates between the two source models was observed in Fig. 1, especially around Guatemala City and the south west part of the country. This finding highlights the importance of the seismic zonation (source model) in probabilistic seismic risk assessment.

The maximum M_{max} parameter led to the highest losses compared to all other parameters, and especially for high return periods where the losses are about 60% higher than the expected M_{max} prediction. This trend is associated with the appearance of large rare events in such return periods. The impact of the maximum M_{max} on the city's losses is also significant, although it is not magnified for longer return periods due to the smaller contribution of the large rare events from the subduction zone compared to the local shallow crustal events. This is further demonstrated below. On the other hand, the effect of the minimum M_{max} is constant for all the return periods (and AAL), leading to around 20% lower values. Overall, the significant sensitivity of the loss metrics to M_{max} demonstrates the importance of including this source of epistemic uncertainty in the assessment of earthquake losses.

The loss exceedance curves corresponding to each branch of the GMPE logic tree indicate that for low return periods (i.e. < 300 years), the country's losses are controlled by the active shallow crustal earthquakes, while for longer return periods the subduction intra-slab seismicity has a stronger influence. This can be verified from the tornado plots, in
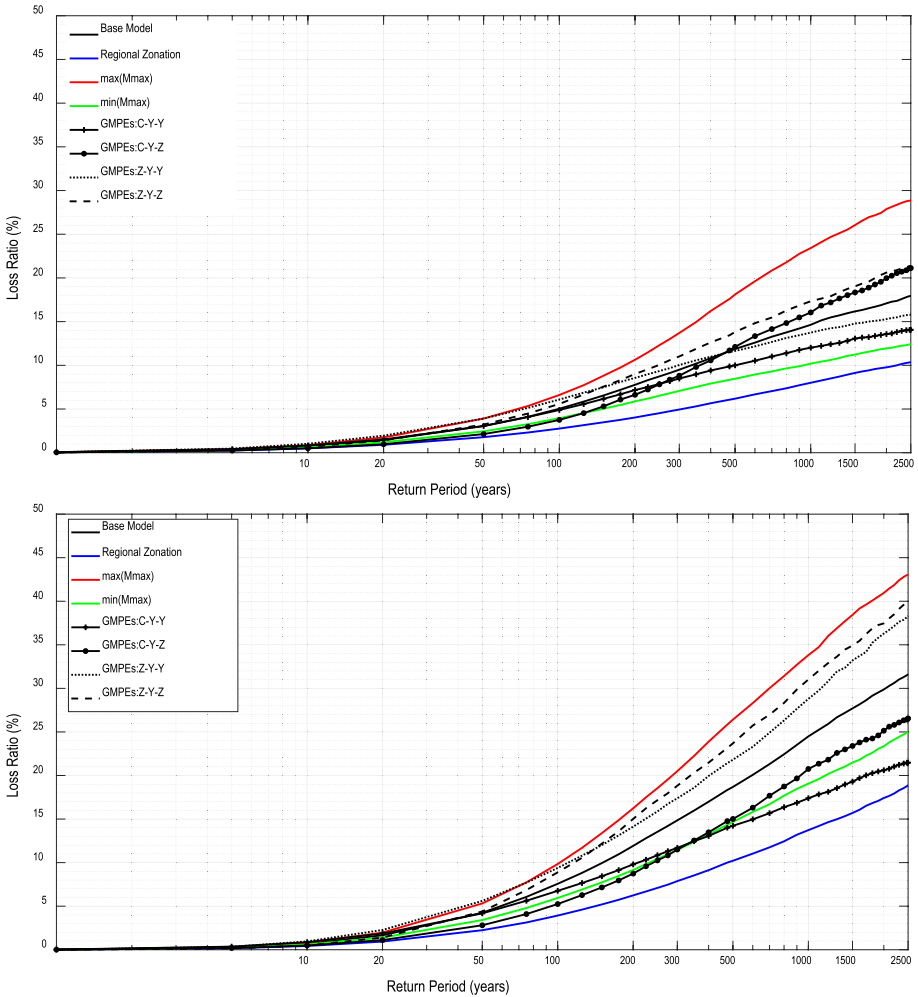
Fig. 6 Country (*top*) and City (*bottom*) loss exceedance curves conditional on the epistemic uncertainties from the hazard component

which the selection of the GMPE for the associated tectonic type governs the predicted losses from each branch. Hence, the contribution from the subduction zone in long return periods is greater than the crustal zones and vice versa. Furthermore, the significance of the epistemic uncertainty in the selection of the GMPEs and the associated weights is demonstrated, with the relative difference in the predicted AAL between two branches reaching up to 50%.

The variations in the AAL for the country indicate that the selection of the GMPE for the subduction intra-slab zones strongly affects the results. This was expected, since a seismic event from the subduction zone will affect a broader region of the country compared to a shallow crustal event. On the contrary, it is quite evident that the economic losses for Guatemala City for all return periods are dominated by the shallow crustal seismicity, as the selection of the GMPE for this tectonic regime governs the variability of predicted
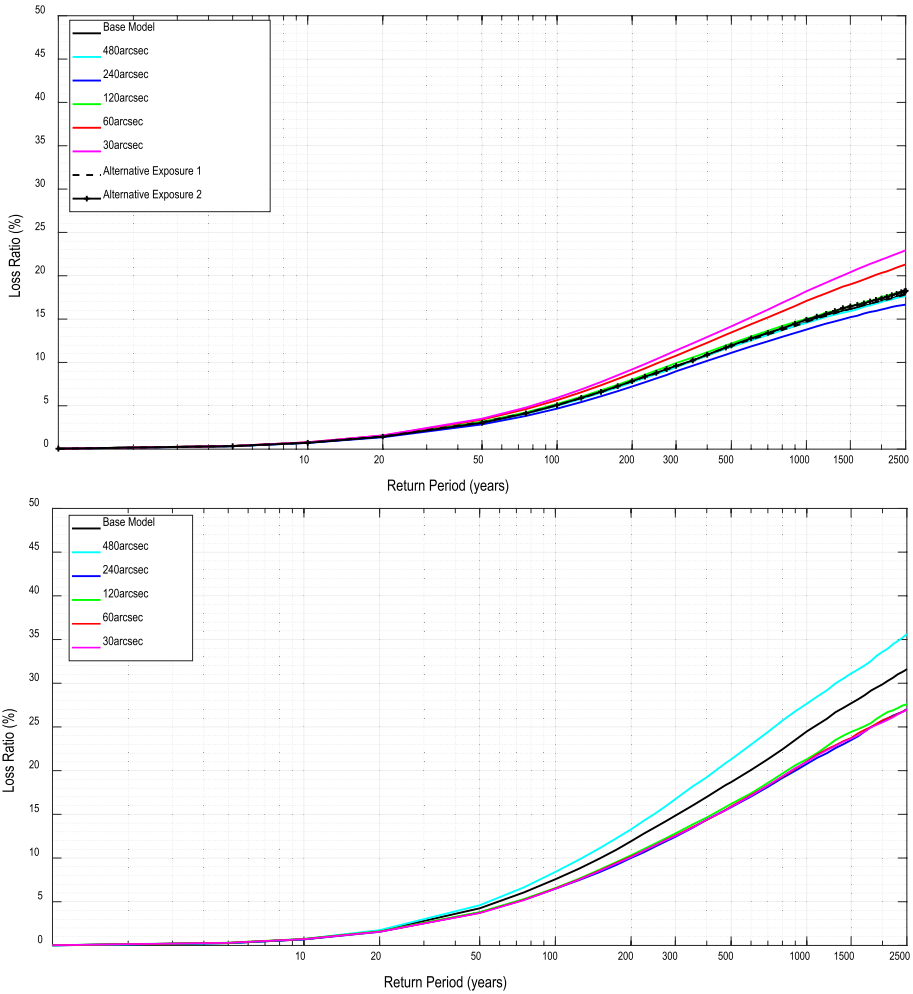
Fig. 7 Country (*top*) and City (*bottom*) loss exceedance curves conditional on the epistemic uncertainties from the exposure parameters

losses from individual branches. Particularly, Zhao et al. (2006) led to higher losses, while Climent et al. (1994) led to lower losses in all cases. Similar findings were also reported by Benito et al. (2012), regarding the predominant seismic event in Guatemala City, using seismic hazard disaggregation.

The impact of the spatial resolution of the exposure model also reveals interesting findings. The smallest spatial resolution (i.e. 30 arc seconds—which is assumed to reduce the most the error in the computation of the site-to-source distances) leads to the highest probable maximum losses or average annualised losses. This result indicates that the aggregation of the assets at the municipality level (i.e. base model) or at other spatial resolutions causes a shift of the economic value to regions of lower seismic hazard (i.e. further away from active faults). Despite the clear importance of improving the spatial resolution of exposure models, it is important to note that these findings cannot be generalized to other regions. As demonstrated by Dabbeek (2017), the impact of spatial resolution of an
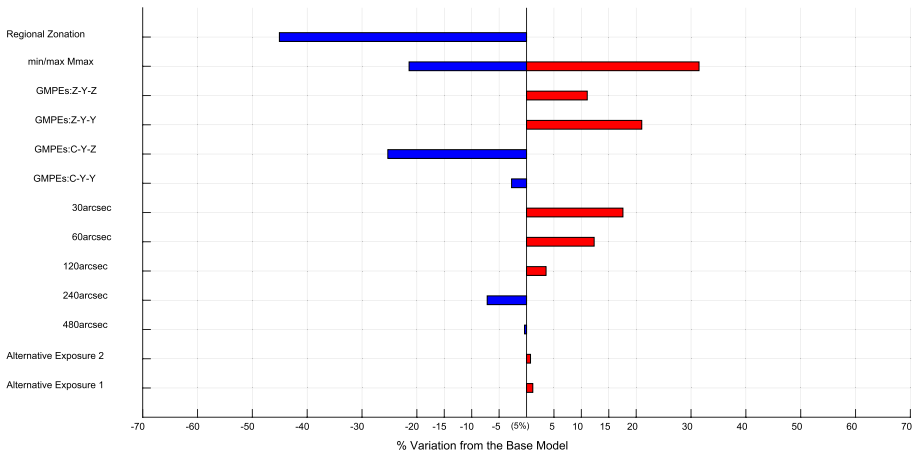
Fig. 8 Sensitivity plot of country loss estimates for the 100-year return period
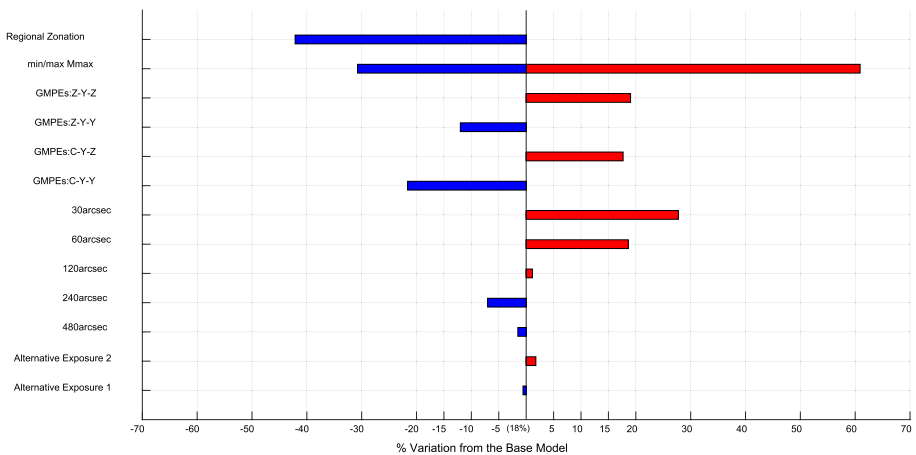

Fig. 9 Sensitivity plot of country loss estimates for the 2500-year return period

exposure model on the predicted losses is strongly dependent on the spatial distribution of the seismic hazard and soil conditions. Moreover, other studies (e.g. Bal et al. 2010) have indicated that adopting a very high resolution might not improve significantly the accuracy of the risk estimates to justify the additional computational effort.

The exposure models for Guatemala City with resolutions equal or smaller than 240 arc-seconds illustrate almost identical losses in the entire range of return periods. A law of diminishing returns is observed in going to very high resolutions since sufficient loss convergence can be achieved at lower resolutions, a trend which was also observed and documented by Bal et al. (2010). In that study, the authors suggest that: "Working at the equivalent of the postcode level of the case study (Istanbul) constitutes a reasonable trade-off between precision and stability in the loss estimates". It is also mentioned that moving to higher resolutions is unlikely to be justified for most applications. Considering the
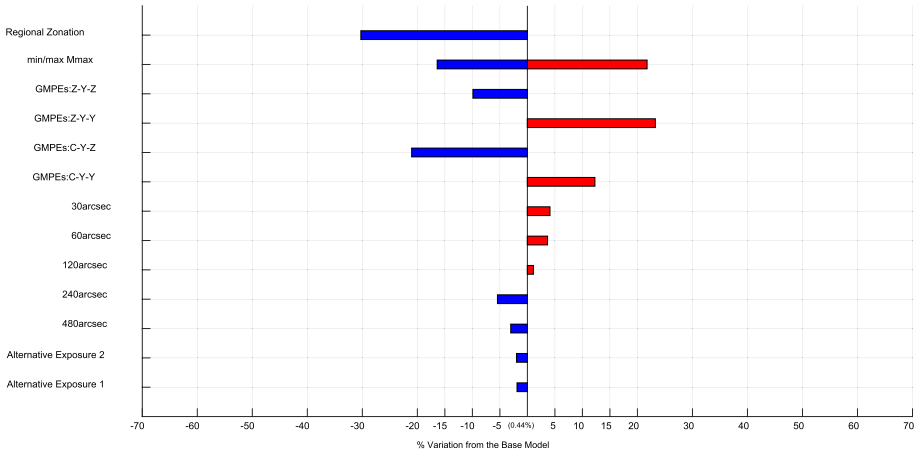
Fig. 10 Sensitivity plot for the country AAL

above observation, we can conclude that the default exposure model aggregation level (i.e. municipality) and the 480 arc-seconds resolution model over-aggregate the building port-folio, which introduces a bias in the results.

Regarding the loss estimates using the different options to model the uncertainty in the vulnerability, Fig. 11 presents the histogram of the annual loss ratios (ALR) considering the three modelling approaches. For the particular case using a lognormal distribution to model the uncertainty in the vulnerability loss ratios, Fig. 12 also presents the results with and without full correlation between the vulnerability of assets of the same building class.

The histogram of annual loss ratios indicates similar average annual loss ratio (AALR), regardless of the modelling approach. While such outcome is expected (as the mean loss ratio in the vulnerability functions across the three modelling options is the same),
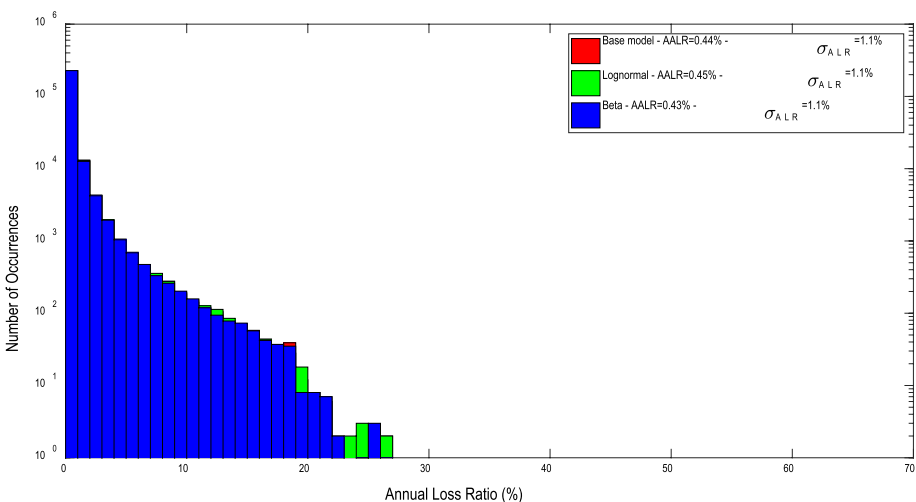

Fig. 11 Histogram of the annual loss ratios for Guatemala following three different options to model the uncertainty in the vulnerability
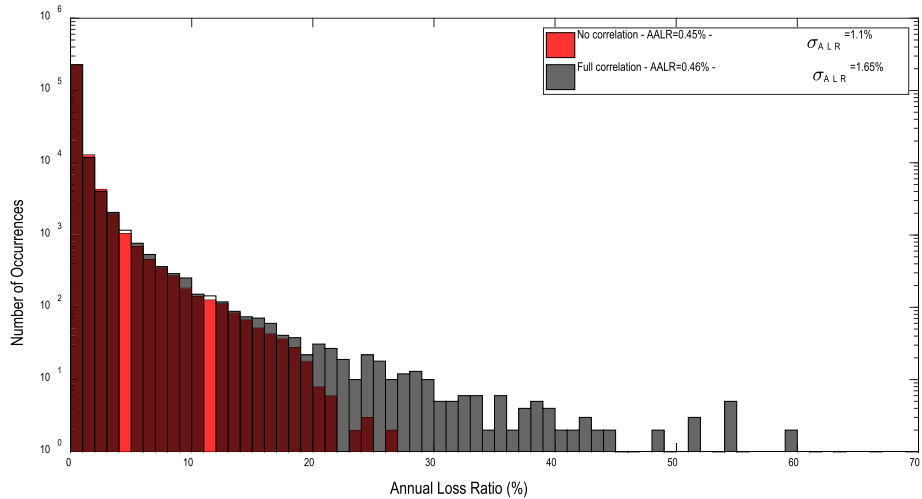
Fig. 12 Histogram of the annual loss ratios for Guatemala assuming a lognormal distribution to model the uncertainty in the vulnerability, with and without full correlation

one could anticipate a higher standard deviation (σ_{ALR}) when modelling the vulnerability uncertainty using a beta or lognormal distribution. However, as also reported in Silva (2019), in earthquake risk analyses covering large regions there is an averaging effect, where the inclusion of vulnerability uncertainty will lead to losses below the mean in some areas, and likewise above the mean in other areas. These results tend to compensate for each other, leading to aggregated losses (i.e. sum of the losses across the entire portfolio) equivalent to the case when only the mean vulnerability functions are used. The exception occurs when vulnerability correlation is considered. In this case, either all the losses will be above the mean, or below the mean, thus leading to scenarios where very low or very high aggregated losses were obtained. This is visible in the histogram in Fig. 12, where ALR above 30% only occur for the full correlation case, leading to a standard deviation 50% higher.

It is recognized that both assumptions of full or no vulnerability correlation are unrealistic, as they describe two extreme cases. The first assumes that buildings within the same building class share identical characteristics (e.g. plan irregularities, construction technique, etc.) and therefore will have exactly the same loss when subjected to the same IM (i.e. same sample LR), while the second assumes that their experienced loss at the same IM is totally uncorrelated. It is reasonable to assume that the "true" loss exceedance curve lies between the curves with no correlation and full correlation, presented in Fig. 13.

4 Loss estimation for insured portfolio

In this section, an insured (cedant) portfolio in Guatemala is presented along with the developed alternative versions in OpenQuake, using various assumptions and techniques to match the vendor model's building classes and corresponding vulnerability functions to the ones used for the residential exposure. In order to map the assets of a portfolio to representative building classes and associated vulnerability functions, the vendor model considers
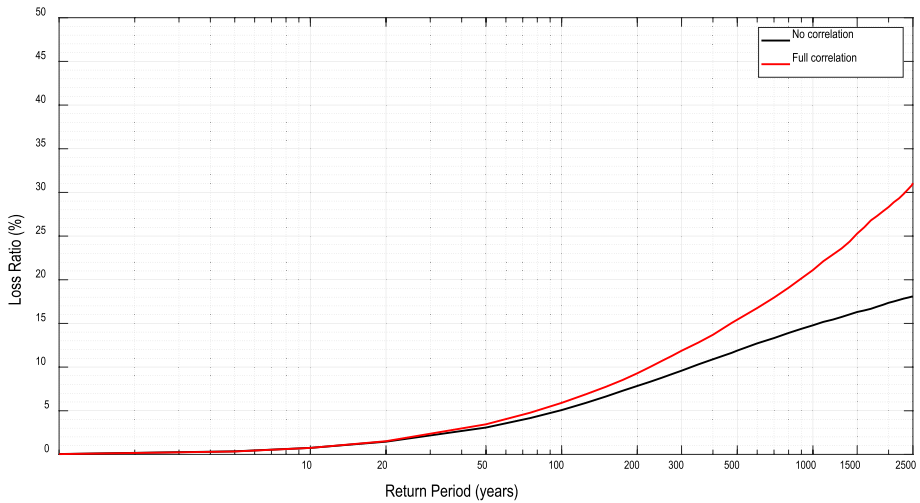
Fig. 13 Country loss exceedance curves using lognormal distributions with and without full vulnerability correlation

the following primary building characteristics: occupancy class, construction type, year of construction, and building height (number of storeys).

There are various occupancy classes included in the portfolio such as residential, commercial and industrial, although the vulnerability functions of the vendor model for Guatemala are independent of the occupancy class. The insured portfolio is also organized according to different coverages. The assets are insured as buildings, contents and business interruption. For the purposes of this study, only the value of the buildings was considered, which represents 90% of the total insured value (TIV). Furthermore, 98% of the buildings of the insured portfolio are modelled as reinforced concrete (RC) in the vendor model. Figure 14 presents the distribution of buildings per year of construction and number of storeys, along with the percentage of the TIV.

Considering this information, the number of storeys and the age (year of construction) of the buildings are the main criteria to link the exposure characteristics of the insured portfolio to the GEM building taxonomy. As previously mentioned, the first seismic design code in Guatemala was implemented in 1996. Therefore, it was assumed that buildings constructed before 1996 are non-ductile or have a low ductility level (DNO or DUL), while buildings constructed after 1996 have a medium or high ductility level (DUM or DUH), according to the available building classes and corresponding vulnerability functions described in Table 2. In order to consider the epistemic uncertainty in this assumption since some of the buildings constructed between 1987 and 1996 are ductile, the buildings with an unknown age were considered to have medium ductility level. For the unknown number of storeys category, the average number of storeys across the entire insured portfolio (i.e. 3 storeys) was assigned to these buildings.

Even though nearly the entire portfolio is modelled as RC buildings in the vendor model, 47% of the TIV consists of low-rise buildings (1–3 storeys). In order to further explore possible discrepancies between the vendor and the OpenQuake-engine models, six alternative versions of the insured portfolio were considered for the low-rise buildings. In the first two cases this subset was assumed to be RC, while in the other four cases it was
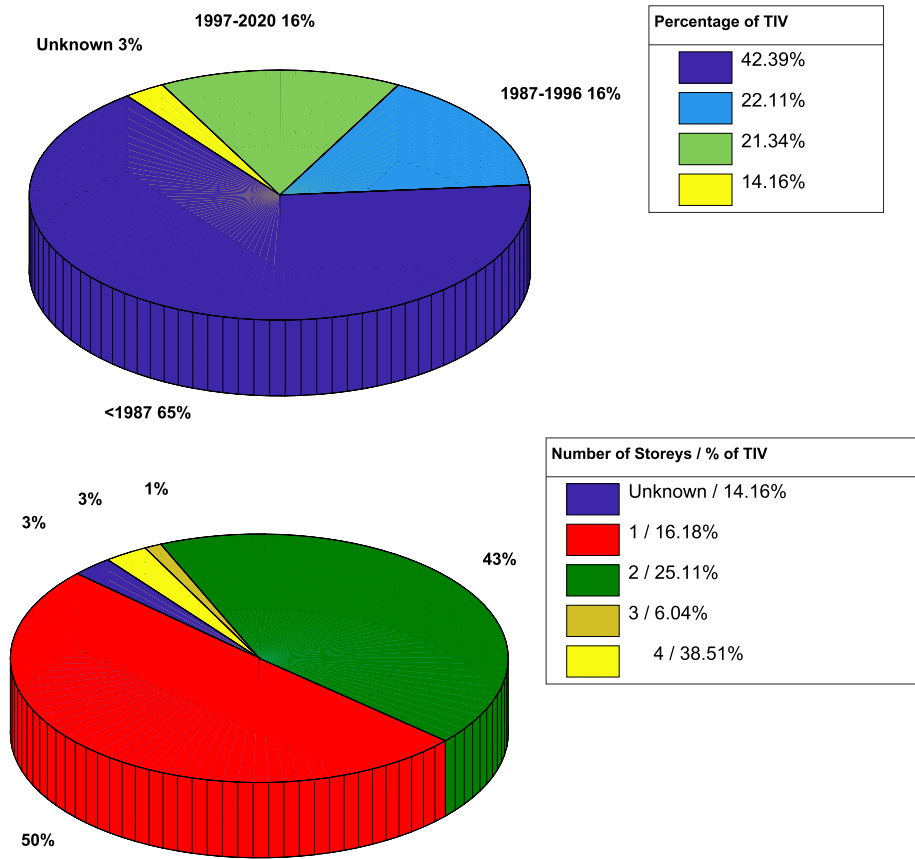
Fig. 14 Distribution of buildings per age of construction (*top*) and number of storeys (*bottom*)

assumed to be composed by unreinforced, confined or reinforced masonry. The alternative versions of the portfolio are summarized in Table 3. Note that the total portfolio (100% of TIV) was considered in cases RC1 and RC2.

It is worth noting that in the vendor model, all the high-rise buildings are characterized by two vulnerability functions, one for 8–15 storeys and the other for more than 15-storey buildings. Finally, the vulnerability functions for confined and reinforced masonry are the same in the vendor model, which limits the loss estimation of the low-rise masonry versions to LM1 and LM3.

As for the seismic hazard component, the seismic source model between the two platforms is the same (Benito et al. 2012). Moreover, in order to ensure that the seismic hazard results were the same, the GMPE logic tree proposed by the vendor model was also used within the OpenQuake-engine. For the sake of clarity, the GMPEs and associated weights proposed by the vendor model are described in Table 4.

The ground up loss exceedance curves and the AAL obtained from both platforms for the alternative versions of the insured portfolio are presented in Figs. 15, 16, 17.

Given that the same probabilistic seismic hazard model and exposure dataset were used, the differences between the various loss exceedance curves are mostly due to the

Table 3 Alternative portfolio versions and the corresponding building characteristics

Modelling cases	Construction type/version	Number of storeys	Lateral load resisting system
All buildings are modelled as RC (100% TIV)	RC1	1–2	MRF
		3	MRFinf
		4–6	MRFinf
		7+	LDUAL
	RC2	1	MRF
		2–3	MRFinf
		4–6	MRFinf
		7+	LDUAL
Low-rise (1–3 storeys) buildings are modelled as RC (47% of TIV)	LRC1	1–2	MRF
		3	MRFinf
	LRC2	1	MRF
		2–3	MRFinf
Low-rise (1–3 storeys) buildings are modelled as masonry (47% of TIV)	LM1	1–2	URM
		3	CF
	LM2	1–2	URM
		3	RM
	LM3	1–3	CF
	LM4	1–3	RM

MRF, moment resisting frame; URM, unreinforced masonry; Inf, with URM infills; CF, confined masonry; RM, reinforced masonry; LDUAL, moment resisting frame with shear walls

Table 4 GMPEs and the associated weights per tectonic region used in the vendor model

Active shallow crust	Subduction interface	Subduction intra-slab
Boore and Atkinson (2008), 0.33	Zhao et al. (2006), 0.5	Zhao et al. (2006), 0.2
Campbell and Bozorgnia (2008), 0.33	Atkinson and Boore (2003), 0.25	Atkinson and Boore (2003), 0.2
Chiou and Youngs (2008), 0.33	Youngs et al. (1997), 0.25	Youngs et al. (1997), 0.6

vulnerability component. For the RC1 and RC2 versions, where the total portfolio was considered, the OpenQuake-engine predicted higher losses than the vendor model for return periods longer than 200 years, with a clear increasing trend at higher return periods. Nevertheless, due to the significantly lower loss estimates at shorter return periods, OpenQuake estimates lower AAL values than the vendor model. As for the low-rise RC variations, we can comment that LRC1 follows the trends of RC1 and RC2, but LRC2 shows higher losses estimated with OpenQuake than the vendor model for almost the entire range of return periods. On the contrary, a significant divergence is observed in the loss estimates for the low-rise masonry buildings. The vendor model appears to overestimate the losses for the unreinforced masonry buildings, since the predicted losses of LM1 (47% of TIV) are higher than the loss estimates of the entire portfolio cases RC1 and RC2 (100% of TIV).
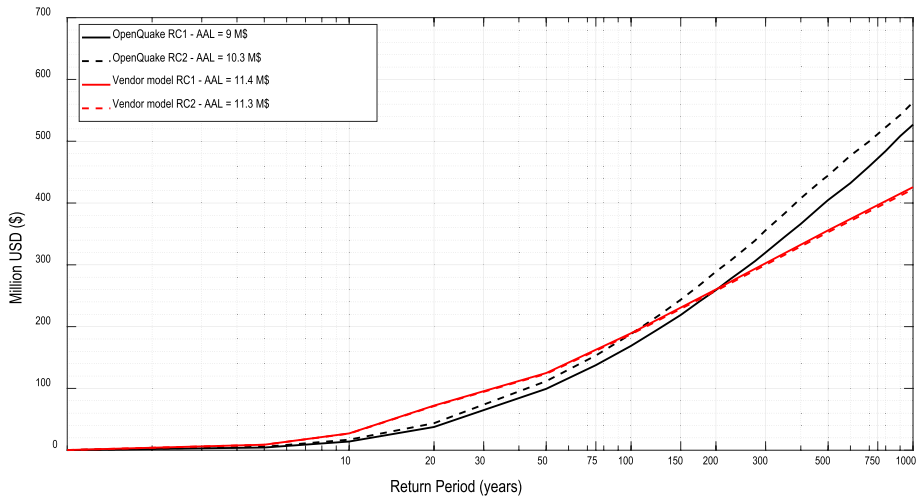
Fig. 15 Loss exceedance curves for RC1 and RC2 portfolio versions obtained from OpenQuake and the vendor model
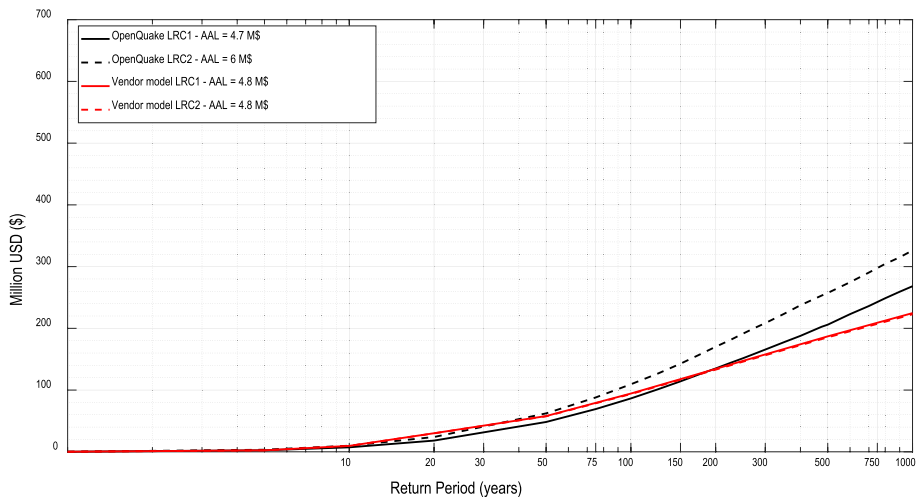

Fig. 16 Loss exceedance curves for the low-rise buildings, LRC1 and LRC2 portfolio versions obtained from OpenQuake and the vendor model

Reasons for these differences in the predicted losses should be due to the differences in the assumptions and methodology adopted for the derivation of the vulnerability functions. The vendor model's functions are derived using a different analytical methodology (i.e. Capacity Spectrum Method, e.g. Freeman 1998), but more importantly, the aleatory uncertainty of the GMPEs along with the total secondary uncertainty are incorporated and represented in the vulnerability functions (and not on the hazard footprints).
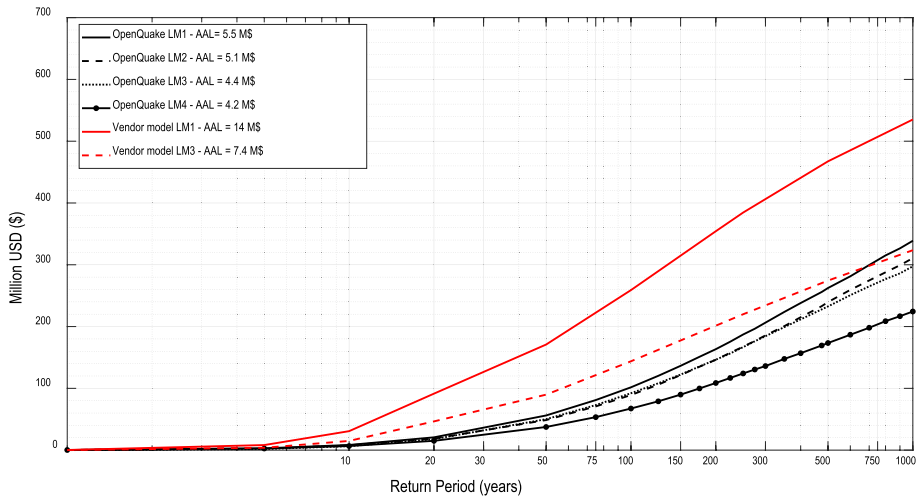
Fig. 17 Loss exceedance curves for the low-rise buildings, LM1, LM2, LM3 and LM4 portfolio versions obtained from OpenQuake and the vendor model

Further comparisons between the two platforms were not possible as most of the components are not accessible.

5 Conclusions

The aim of this study was to explore and quantify the impact of epistemic uncertainty in a probabilistic seismic risk assessment model. To this end, the residential exposure of Guatemala and Guatemala City were selected as the case studies. Overall, the results of the study indicate that the epistemic uncertainty in the hazard component is the most significant among the three components. In particular, the loss estimates illustrated a high sensitivity on the source model (seismic zonation), the maximum magnitude assumed for each seismic source, and the selection of the GMPEs and associated weights. The latter results are also in agreement with the findings from Crowley et al. (2005) and Silva (2018).

For the evaluation of the impact of the exposure component, alternative versions of the residential exposure model were considered, and negligible differences in the loss estimates were observed. Moreover, various spatial resolutions of the exposure model were tested, and the results suggest that for the risk assessment at a city (urban) level, coarse resolutions should be refined in order to avoid biased loss estimates. For national assessments, the predicted losses for increasing resolutions presented a strong dependency on the spatial distribution of seismic hazard. Finally, in order to assess the impact of the vulnerability component, probabilistic vulnerability functions were used to incorporate the building-to-building and record-to-record variability. It was shown that the vulnerability correlation between assets of the same class dominates the predicted losses.

An insured portfolio in Guatemala was translated into the OpenQuake format using the building classes developed by Calderon et al. (2020) and the fragility functions of Martins and Silva (2020). In general, OpenQuake predicted higher losses than the vendor model for long return periods, while the opposite trend was observed in the short return periods.

Significant differences in vulnerability modelling of the low-rise masonry buildings were observed. It is relevant to mention that the two platforms incorporate and propagate the secondary uncertainty in different manners, which did not allow further comparison.

The results from this study indicate that different sources of epistemic uncertainty can influence the risk estimates significantly, and consequently any decisions that may be informed by those risk results. These can include the pricing of (re)insurance products, but also the investment in disaster risk management measures. The employment of an open-source platform for the assessment of the losses allows a better understanding of the impact of these sources of uncertainty, as well as to identify which components are driving the uncertainty, and consequently where additional efforts should be devoted to limit the variability. The OpenQuake-engine is a suitable tool for such analyses, but it still lacks of a robust financial module to translate the ground-up losses into results that can be directly used in the insurance industry.

The authors would like to highlight that the main conclusions from this study are valid for the case of Guatemala, and their application to other countries or cities calls for due care. However, we note that the levels of sophistication of this model is similar to that of other models covering other regions in the world (e.g. Silva et al. 2020), and therefore these findings can also support modellers in understanding which modelling assumptions might affect their risk estimates.

References

Alvarado GE, Benito B, Staller A, Climent Á, Camacho E, Rojas W, Marroquín G et al (2017) The new Central American seismic hazard zonation: mutual consensus based on up to day seismotectonic framework. Tectonophysics 721:462–476

Atkinson Gail, Boore David (2003) Empirical Ground-Motion Relations for Subduction-Zone Earthquakes and Their Application to Cascadia and Other Regions. Bulletin of the Seismological Society of America. 93:1703–1729. https://doi.org/10.1785/0120020156

Bal IE, Bommer JJ, Stafford PJ, Crowley H, Pinho R (2010) The influence of geographical resolution of urban exposure data in an earthquake loss model for Istanbul. Earthq Spectra 26(3):619–634

Bazzurro P, Luco N (2007) Effects of different sources of uncertainty and correlation on earthquake-generated losses. Aust J Civ Eng. https://doi.org/10.1080/14488353.2007.11463924

Bazzurro P, Park J (2007) The effects of portfolio manipulation on earthquake portfolio loss estimates. AIR Worldwide, Boston

Benito MB, Lindholm C, Camacho E, Climent Á, Marroquín G, Molina E, Rojas W et al (2012) A new evaluation of seismic hazard for the Central America Region. Bull Seismol Soc Am 102(2):504–523

Boore DM, Atkinson GM (2008) Ground-motion prediction equations for the average horizontal component of Pga, Pgv, and 5%-damped Psa at spectral periods between 0.01 S and 10.0 S. Earthq Spectra 24(1):99–138

Brzev S, Scawthorn C, Charleson LA, Greene M, Jaiswal K, Silva V (2013) Gem building taxonomy version 2.0. Vol. GEM technical report 2013-02 V1.0.0, Pavia, IT: GEM Foundation. https://doi.org/10.13117/gem.exp-mod.tr2013.02

Calderon A, Silva V, Avilez M, Mendez R, Castillo R, Gil JC, Lopez A (2020) Towards a uniform earthquake loss model across Central America. Earthquake Spectra. Under review.

Campbell KW, Bozorgnia Y (2008) Nga ground motion model for the geometric mean horizontal component of Pga, Pgv, Pgd and 5% damped linear elastic response spectra for periods ranging from 0.01 to 10 S. Earthq Spectra 24(1):139–171

Chiou BJ, Youngs RR (2008) An Nga model for the average horizontal component of peak ground motion and response spectra. Earthq Spectra 24(1):173–215

Climent Á, Taylor W, Real MC, Strauch W, Villagran M, Dahle A, Bungum H (1994) Spectral strong motion attenuation in Central America. NORSAR Technical Report 2-17, p 46

Colombi M, Borzi B, Crowley H, Onida M, Meroni F, Pinho R (2008) Deriving vulnerability curves using Italian earthquake damage data. Bull Earthquake Eng 6:485–504. https://doi.org/10.1007/s10518-008-9073-6

Cornell CCA (1968) Engineering seismic risk analysis. Bull Seismol Soc Am 58:1583–1606

Crowley H, Bommer J, Pinho R, Bird J (2005) The impact of epistemic uncertainty on an earthquake loss model. Earthq Eng Struct Dyn 34:1653–1685

Dabbeek J (2017) The effect of spatial resolution in exposure models for seismic loss estimation. Instituto Universitario di Studi Superiori, Pavia

Dabbeek J, Silva V (2020) Modeling the residential building stock in the Middle East for multi-hazard risk assessment. Nat Hazards 100(2):781–810. https://doi.org/10.1007/s11069-019-03842-7

D'Ayala D, Meslem A, Vamvatsikos D, Porter K, Rossetto T (2014) Guidelines for analytical vulnerability assessment of low/mid-rise buildings. GEM Technical Report 2014-12. GEM Foundation, Pavia, Italy. https://doi.org/10.13117/GEM.VULN-MOD.TR2014.12

Faber MH (2005) On the treatment of uncertainties and probabilities in engineering decision analysis. J Offshore Mech Arct Eng 127(3):243–248

Field E, Jordan T, Cornell C (2003) Opensha: a developing community-modeling environment for seismic hazard analysis. Seismol Res Lett 74:406–419

Freeman S (1998) Development and use of capacity spectrum method. In Proceedings of the IEEE—PIEEE

Hayes GP, Briggs RW, Sladen A, Fielding EJ, Prentice C, Hudnut K, Mann P et al (2010) Complex rupture during the 12 January 2010 Haiti earthquake. Nat Geosci 3(11):800–805

Kappos A, Panagopoulos G, Panagiotopoulos C, Penelis G (2006) A hybrid method for the vulnerability assessment of R/C and Urm buildings. Bull Earthq Eng 4(4):391–413

Lallemant D, Kiremidjian A (2014) A beta distribution model for characterizing earthquake damage state distribution. Earthq Spectra 31:140514111412006

Martins L, Silva V (2020) Development of fragility and vulnerability model for global seismic risk analyses. Bull Earthq Eng. https://doi.org/10.1007/s10518-020-00885-1

Martins L, Silva V, Marques M, Crowley H, Delgado R (2016) Development and assessment of damage-to-loss models for moment-frame reinforced concrete buildings. Earthq Eng Struct Dyn 45(5):797–817

Mitchell-Wallace K, Jones M, Hillier J, Foote M (2017) Natural catastrophe risk management and modelling: a practitioner's guide. Wiley, Hoboken

Molina E, Marroquín G, Escobar J, Talavera E, Climent A, Rojas W, Camacho E, Benito B, Lindhom C (2008) Proyecto Resis Ii Evaluación De La Amenaza Sísmica En Centro América. NORSAR

Ordaz M, Martinelli F, D'Amico V, Meletti C (2013) Crisis 2008: a flexible tool to perform probabilistic seismic hazard assessment. Seismol Res Lett 84:495–504

Pagani M, Monelli D, Weatherill G, Danciu L, Crowley H, Silva V, Henshaw P et al (2014) OpenQuake engine: an open hazard (and risk) software for the global earthquake model. Seismol Res Lett 85:692–702

Pagani M, Garcia-Pelaez J, Gee R, Johnson K, Poggi V, Styron R, Weatherill G, Simionato M, Viganò D, Danciu L, Monelli D (2018) Global earthquake model (Gem) seismic hazard map (Version 2018.1—December 2018)

Paté-Cornell ME (1996) Uncertainties in risk analysis: six levels of treatment. Reliab Eng Syst Saf 54(2–3):95–111

Pittore M, Haas M, Megalooikonomou KG (2018) Risk-oriented, bottom-up modeling of building portfolios with faceted taxonomies. Front Built Environ 4:41

Porter K (2010) Cracking an open safe: uncertainty in Hazus-based seismic vulnerability functions. Earthq Spectra 26(3):893–900

Rodrigues D, Crowley H, Silva V (2018) Earthquake loss assessment of precast RC industrial structures in Tuscany (Italy). Bull Earthquake Eng 16:203–228. https://doi.org/10.1007/s10518-017-0195-6

Román MO, Wang Z, Sun Q, Kalb V, Miller SD, Molthan A, Schultz L et al (2018) Nasa's black marble nighttime lights product suite. Remote Sens Environ 210:113–143

Rossetto T, Ioannou I, Grant D (2015) Existing empirical fragility and vulnerability functions: compendium and guide for selection. https://doi.org/10.13117/gem.vulnsmod.tr2015.01

Silva V (2018) Critical issues on probabilistic earthquake loss assessment. J Earthq Eng 22(9):1683–1709

Silva V (2019) Uncertainty and correlation in seismic vulnerability functions of building classes. Earthq Spectra 35(4):1515–1539

Silva V, Crowley H, Pagani M, Monelli D, Pinho R (2014a) Development of the OpenQuake engine, the global earthquake model's open-source software for seismic risk assessment. Nat Hazards 72(3):1409–1427

Silva V, Crowley H, Varum H, Pinho R, Sousa L (2014b) Investigation of the characteristics of Portuguese regular moment-frame RC buildings and development of a vulnerability model. Bull Earthquake Eng 13:1455–1490 (2015). https://doi.org/10.1007/s10518-014-9669-y

Silva V, Amo-Oduro D, Calderon A, Costa C, Dabbeek J, Despotaki V, Pittore M (2020) Development of a global seismic risk model. Earthqu Spectra. https://doi.org/10.1177/8755293019899953

Sousa L, Silva V, Marques M, Crowley H (2016) On the treatment of uncertainties in the development of fragility functions for earthquake loss estimation of building portfolios. Earthquake Engng Struct Dyn 45:1955–1976. https://doi.org/10.1002/eqe.2734

Spiegelhalter D, Riesch H (2011) Don't know, can't know: embracing deeper uncertainties when analysing risks. Philos Trans Ser A Math Phys Eng Sci 369:4730–4750

Taylor P (2015) Calculating financial loss from catastrophes. Paper presented at the SECED, Cambridge, United Kingdom

Villagran M, Lindholm C, Dahle A, Cowan H, Bungum H (1996) Seismic hazard assessment for Guatemala city. Nat Hazards 14(2):189–205

Wald D, Allen T (2007) Topographic slope as a proxy for seismic site conditions and amplification. Bull Seismol Soc Am 97:1379–1395

Yepes-Estrada C, Silva V, Rossetto T, D'Ayala D, Ioannou I, Meslem A, Crowley H (2016) The global earthquake model physical vulnerability database. Earthq Spectra 32(4):2567–2585

Yepes-Estrada C, Silva V, Valcárcel J, Acevedo AB, Tarque N, Hube M, Coronel G, Santa-Maria H (2017) Modeling the Residential Building Inventory in South America for Seismic Risk Assessment. Earthquake Spectra, 33(1), 299–322. https://doi.org/10.1193/101915eqs155dp

Youngs RR, Chiou SJ, Silva WJ, Humphrey JR (1997) Strong ground motion attenuation relationships for subduction zone earthquakes. Seismol Res Lett 68(1):58–73

Zhao JX, Zhang J, Asano A, Ohno Y, Oouchi T, Takahashi T, Ogawa H et al (2006) Attenuation relations of strong ground motion in Japan using site classification based on predominant period. Bull Seismol Soc Am 96(3):898–913