Empirical vulnerability curves for Icelandic low-rise buildings based on zero-inflated beta regression model

Bjarni Bessason – University of Iceland, Faculty of Civil and Environmental Engineering, Reykjavik, Iceland, bb@hi.is

Rajesh Rupakhety – University of Iceland, Faculty of Civil and Environmental Engineering, Reykjavik, Iceland, rajesh@hi.is

Jón Örvar Bjarnason – Natural Catastrophe Insurance of Iceland, Hlídarsmári 14, 201 Kópavogur, Iceland, jon@nti.is

Abstract: In June 2000, two earthquakes of ~Mw6.5 struck in South Iceland, and in May 2008 the same region was hit again further west, with Mw6.3 event. Almost 5000 residential buildings were affected in each of these two seismic events. To fulfil insurance claims, detailed, and complete loss data were collected in each case, and the *2000 dataset* and *2008 dataset* were established. Having access to two high quality loss datasets from different size earthquakes, affecting the same building typologies in the same region, is rare to find in the literature. An advanced empirical vulnerability model based on zero-inflated beta regression was fitted to five building typologies, classified according to the GEM taxonomy system, independently for the 2000 dataset and the 2008 dataset. Status of seismic codes was considered when defining the building typologies. PGA was used as intensity measure. For all the five building typologies, the calibrated vulnerability functions and the fragility curves are substantially different from these two datasets. This indicates that PGA is not alone an adequate intensity measure to predict losses. The results also show that status of seismic code affects the performance of the buildings as one would like to see.

Keywords: Zero-inflated beta regression, vulnerability, GEM taxonomy, seismic codes.

1. Introduction

Seismicity in Iceland is moderate to high and comparable to activity in Southern Europe (Einarsson, 2008). On 17 and 21 of June 2000, two earthquakes of similar size, Mw6.52 and Mw6.44 (Jónasson et al. 2021), struck in the eastern part of the South Iceland Seismic Zone (SISZ) and affected nearly 5000 low-rise residential buildings. Eight years later, on 29 May 2008, a Mw6.31 earthquake struck further west in the zone (Halldórsson et al, 2009; Jónasson et al. 2021), and again nearly 5000 buildings were affected. The epicentres and fault ruptures of all the three events were close to small towns and farms (Fig.1). Despite, substantial damage, no residential building collapsed and there were no fatalities. Mandated by law, all properties in Iceland are insured against natural hazards, like earthquakes, at the Natural Catastrophe Insurance of Iceland (2022). Therefore, after the two 2000 earthquakes and again after the 2008 event, monitory losses were estimated for each damaged building to address insurance claims. Registers Iceland (2022) maintains a detailed property database for all building units in Iceland. Combination of the loss data and the property database were used to build two loss dataset hereafters referred to as the *2000 dataset* and the *2008 dataset*, respectively. The two datasets include loss estimates for every building exposed to estimated PGA of 0.05g or more. Buildings that had no losses are also a part of the datasets, so they are complete. The two databases give unique possibilities to study different aspects of seismic vulnerability. This has been done to some extent. The first vulnerability model using only the 2008 dataset was presented by Bessason et al. (2012). As mentioned before the loss data is detailed, especially the 2008 dataset, which is classified into different subcategories and sub-subcategories of structural and non-structural losses. An overview of how the 2008 loss data was distributed in subclasses was reported in Bessason et al. (2014). In Bessason et al. (2016) fragility functions and vulnerability curves, using the lognormal distribution assumption, were calibrated using the 2000 and 2008 datasets combined into one dataset. New developments were done in a work by Ioannou et al. (2018) when an advanced beta-regression (Ferrari et al. 2004) was used to model the losses caused by the two June 2000 earthquakes. Bessason et al. (2020) improved this loss model by applying a zero-inflated beta regression model (ZIBRM) (Ospina et al. 2012). Finally, Bessason et a al. (2022), used the ZIBRM again to model separately, the 2000 dataset and 2008 dataset, as well as to study the effect of status of seismic codes (Crowley et al. 2021) when defining the building typologies. Furthermore, the GEM taxonomy was used to define the building typologies (Brzev et al. 2013).

To our best knowledge, the use of ZIBRM to model post-earthquake loss data has only by done twice, that is by the authors of this study (Bessason et al. 2020; 2022). The most important novelty of the ZIBRM is that it treats the no-loss buildings specially and helps to bend the vulnerability curves down towards zero loss at low intensity. By this the models better reflect the dominance of no-loss data in the low intensity range.

The objective of this conference paper is to present and demonstrate the capability of using ZIBRM to model empirical loss data, using the 2000 and the 2008 loss datasets from Iceland. The model is believed to be an important contribution in modelling of empirical loss data. The ZIBRM model is fitted to building typologies based on the GEM building taxonomy (Brzev et al. 2013) and the status of seismic codes refer to classification given by Crowley et al. (2021) and can therefore be compared to other global models.

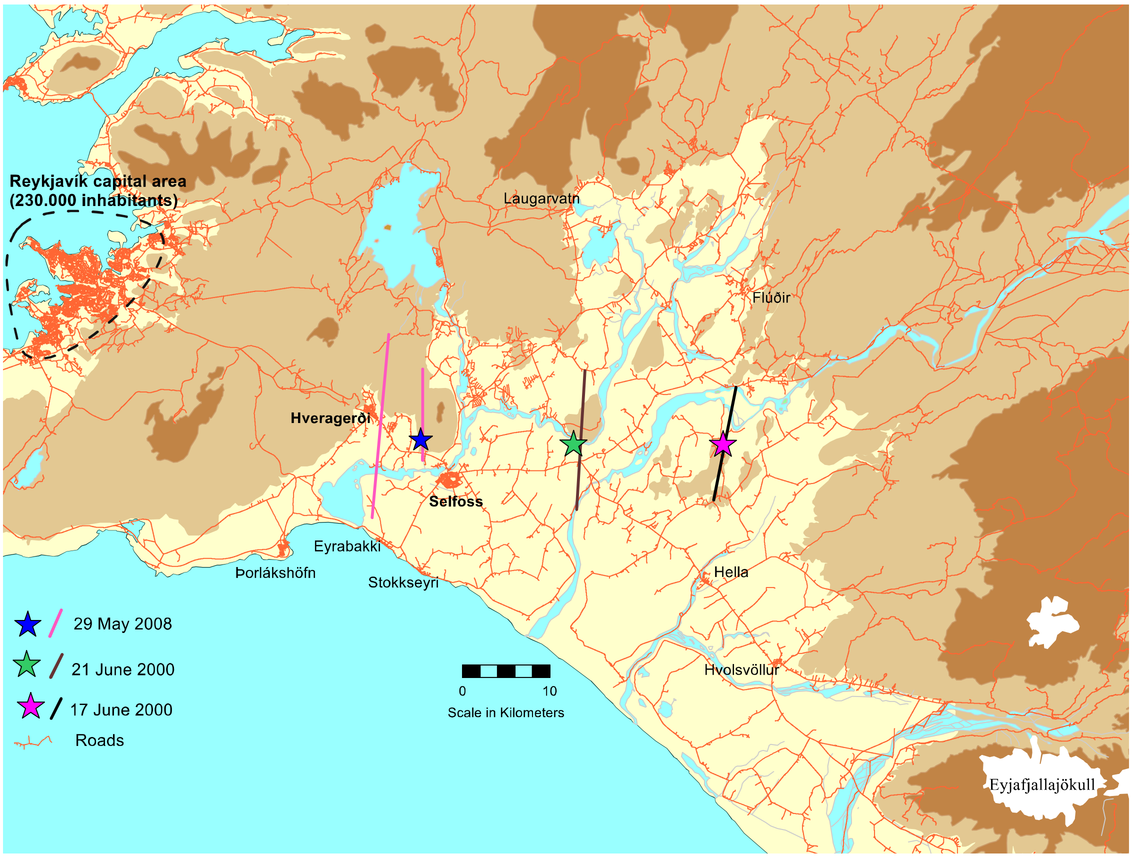

Fig. 1 - Epicentre and fault ruptures of the South Iceland earthquakes in June 2000 and May 2008. The main towns and villages in the affected area are marked with their names on the map. Buildings in the Reykjavik capital area were not affected by these events (PGA < 0.05g).

2. The empirical loss data and building typologies

2.1. Loss data

It is mandated by law in Iceland to have insurance against destructive natural events, like earthquakes, avalanches, floods etc. Therefore, after the June 2000 earthquakes, which occurred with four-day interval, losses were estimated by trained assessors in all damaged buildings in the affected area to fulfil insurance claims. Since only four days were between the two June 2000 earthquakes, the estimated losses include accumulated losses from both events. The shortest distance between the two fault ruptures is 15.5 km (Fig. 1). The attenuation of seismic waves is high Iceland due to geologically young, soft and cracked bedrock, and it has been argued that the great majority of the buildings in June 2000 were only affected by the earthquake closest to the building in question (Bessason et al. 2020). In addition, only relatively few buildings were located between the two faults and at similar distance from each of them, and thereby affected by both events (Fig.1). Similar loss assessment procedure was carried out after the May 2008 earthquake to establish the 2008 dataset. Both, in 2000 and 2008, the insurance deductibles were low (650 euros per dwelling) so it believed that all owners announce damage of their properties in order to get insurance benefits. Therefore, it can be assumed the both the 2000 and the 2008 datasets are complete in the sense that they include all affected buildings in the region (PGA > 0.05g) (see Rossetto et al. 2014).

The two loss datasets include estimated repair cost of both structural and non-structural elements. Here, non-structural damage includes damage to all fixtures, as well as technical systems (plumbing, electrical installations etc.) but does not include damage of loose household items like furniture, TVs, computers, etc. In the 2000 dataset the damage (estimated repair cost) in each building was classified into five subcategories, two covering structural damage and three covering non-structural damage. In the 2008 dataset ten subcategories were used. For more details of the subcategories see (Bessason et al. 2014; 2022). The subcategories in the 2008 dataset can effortlessly be combined to obtain identical categories as in the 2000 dataset. In this study the focus is however on total loss, i.e. the sum of loss in all subcategories (structural and non-structural losses). A damage factor*,* DF, is computed for each building, which is defined as:

$$DF = \frac{\text{Estimated loss}}{\text{Replacement value}} = \frac{EL}{RV} \tag{1}$$

where EL cover the total estimated repair cost and RV is the fire insurance value obtained from the official property database. The RV is estimated as the depreciated replacement value plus the cost of dismantling and transporting the debris. Depreciation is based on age, main construction material and general condition. On the other hand, the repair cost was not depreciated.

The DF can reach 1.0 (100%) for buildings with total damage and correspond to payment of full replacement values to the building owners. In practice damage equivalent to “total loss” was assigned to most of the buildings that suffered an estimated repair cost of more than 70% of their RV value in the 2000 dataset. In 2008, the level was lower and estimated on a case-by-case basis for each building with estimated loss in the range 50–70%.

The great majority of the buildings are low-rise with short natural periods, hence the PGA was adopted as intensity parameter. It was computed for each building site based on local ground motion prediction model given by Rupakhety et al. (2009).

2.2. Building taxonomy

The GEM taxonomy (Brzev et al. 2013) was used to classify all the buildings in the loss datasets. In the 2000 dataset, 54% of the buildings were built of reinforced concrete (RC), almost all cast-in-place (CIP). Furthermore, 37% were timber buildings (W+WLI), 9.3% masonry buildings (MR+CBH+MOC), and the rest (only 0.3%) used other main construction material. Moreover, 68% of the buildings were one-storey, 23% two-storey, 7.9% three-storey, 0.3% four-storey, and no building was higher. In the 2008 dataset, 45% of the dwellings were built of RC, 48% timber and 7.6% masonry. Furthermore, 74% were one-storey buildings, 19% two-storey, 5.9% three-storey, and 0.5% four-storey. Since only a low fraction of the affected buildings are three-storey or higher in both datasets, all the buildings are placed in HBET:1,2 class. The lateral load resisting system for the great majority of the affected buildings in the two datasets are structural walls, hence LWAL is used to identify the structural system for the three building typologies. For more details of the building characteristic see Bessason et al. (2022).

Crowley et al. (2021) classified status of seismic design codes in different European countries, including Iceland, in four categories based on construction period (Table 1).

Table 1. Status of seismic design codes in Icelandic for different construction periods (Crowley et al. 2021)

Status	Description	Comment	Period
CDN	*No Code*	No seismic design code	< 1958
CDL	*Low code*	First generation of seismic codes	1958 – 1975
CDM	*Moderate code*	Second generation of seismic codes	1976– 2001
CDH	*High code*	Latest generation of seismic codes	≥ 2002

In Bessason et al. (2022) it was argued that the no-code and the low-code (CDN,CDL) buildings could be combined in one class, and also the moderate-code and the high code (CDM,CDH) buildings. Since, 98% of the masonry buildings were built before 1976 all these buildings are in the CDN,CDL class with respect to status of seismic code.

In summary, Table 2 gives the distribution of the buildings in the two loss datasets according to the GEM taxonomy and the status of seismic codes.

Table2. Classification of residential buildings affected by the June 2000 and May 2008 earthquakes.

GEM building taxonomy	June 2000 earthquakes		May 2008 earthquake	
	Number	(%)	Number	(%)
CR+CIP / LWAL / HBET:2,1 / CDN,CDL	1665	64.7	1112	52.6
CR+CIP / LWAL / HBET:2,1 / CDM,CDH	907	35.3	1003	47,4
Sum:	*2572*	*100*	*2115*	*100*
W+WLI / LWAL / HBET:2,1 / CDN,CDL	692	39.8	649	28,6
W+WLI / LWAL / HBET:2,1 / CDM,CDH	1047	60.2	1623	71.5
Sum:	*1739*	*100*	*2272*	*100*
MR+CBH+MOC / LWAL / HBET:2,1 / CDN,CDL	443	100	359	100
Total sum	**4754**		**4746**	

3. Statistical model.

In the June 2000 and May 2008 earthquakes high proportion of the buildings suffered no losses (DF=0). Total losses (DF=1) were on the other hand very rare. Since the loss data includes both “zero” values and “one” values and is bounded in the range [0,1], it is preferable to use a mixed continuous-discrete regression to model the data. That is, discrete models to cover the “zeros” and “ones”, and then continuous regression for data in the range (0,1). In our case where the data includes high fraction of zero loss incidents but negligible number of total losses, a zero-inflated beta regression model is well-suited (Ospina et al. 2012). The discrete modelling of the total loss buildings (DF=1) is omitted, but instead the DF for these buildings is assigned a value less than 1. A two-step regression process is used to construct the vulnerability model. This approach is explained schematically in Fig.2 but more details, and mathematical formulations, are available in Ferrari et al. (2004); Ospine et al. (2012); Ioannou et al. (2018); and Bessason et al. (2020; 2022).

In Fig.2a the loss data transformed to DF using Eq.(1) is shown. Loss data after the June 2000 earthquake for RC buildings built before 1976 (CND,CDL) is used as an example (1665 datapoints, see Table 2). In Fig.2b the data is transformed to logistical data, i.e Y=0 if DF=0 and Y=1 if DF>0. To better show the distribution of the data it is jittered in the range -0.05-0.05 for Y=0 and in the range 0.95-1.05 for Y=1. A logistical regression model (LM) is used to evaluate the probability of obtaining loss (Y=1) for a given PGA. Then, all data points with DF=0 are filtered out, and a beta regression is applied to model the continuous loss distribution conditioned on loss (Fig.2c). The logistical model and the condition beta model are then combined to obtain the vulnerability model (Fig.2d). From this model the probability density function for the DF for any given PGA can be found. An example of this is shown in Fig.2e for PGA=0.4g. Finally, by defining bins for different damage stages (see Table 3), desired fragility curves can be constructed (see Fig.2e and Fig.2f). All the above regression is carried out in R (R Core Team, 2022).

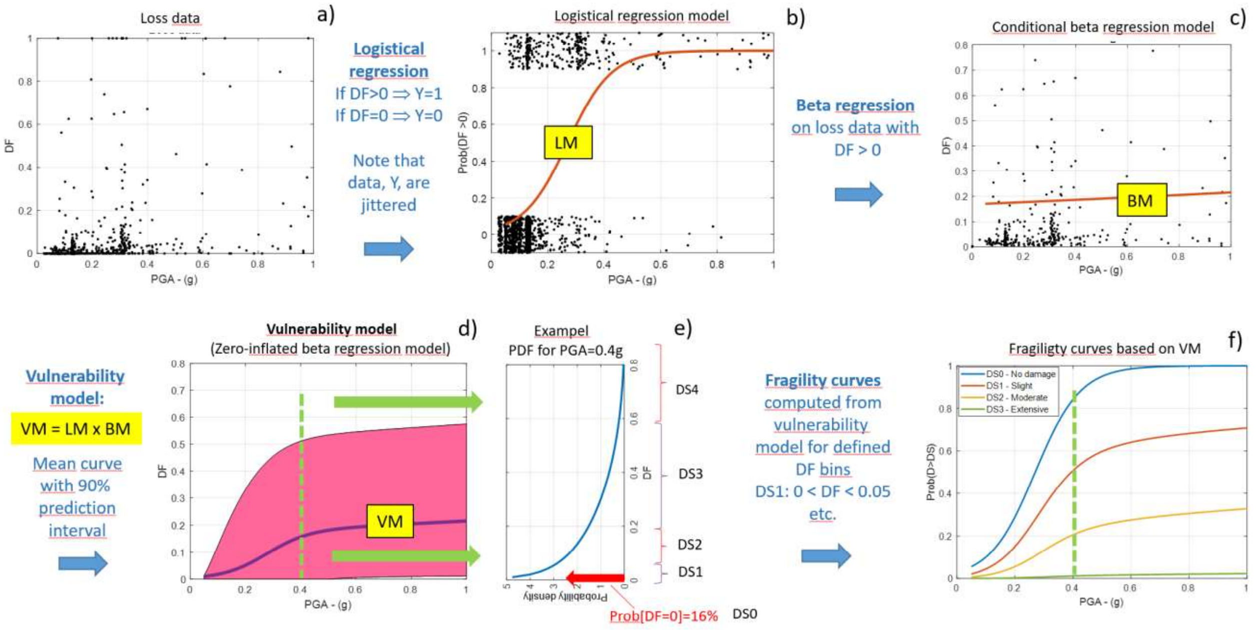

Fig. 2 - Flowchart that schematically explains the main steps in the regression process to obtain vulnerability model and thereafter fragility curves. Real loss data for RC buildings built before 1976 (CR+CIP / LWAL / HBET:2,1 / CDN,CDL) for the 2000 loss dataset are used to describe the method (see Table 2). The damage stages shown in Fig.2e are defined int Table 3.

Table 3 – Definitions of damage states in this study.

Damage state	**Description**	DF bins
DS0	No damage	DF=0
DS1	Slight damage	0 < DF ≤ 0.05
DS2	Moderate damage	5 < DF ≤ 0.20
DS3	Extensive damage	0.20 < DF ≤ 0.60
DS4	Complete damage	DF > 0.6

4. Results and discussion

The Zero-Inflated beta regression model described in section 3 was used to fit the five building typologies shown in Table 1 independently for both the 2000 and the 2008 dataset. The mean vulnerability curves for RC buildings and Timber buildings are shown in Figure 3 and 4. The scatter in the loss data is quite wide although only data points in the DF range 0 to 0.4 are shown on the plot (some datapoints are located higher). As mentioned before when losses were in the range of 50-70% or higher of the replacement value for a given property, full replacement value was in most cases paid to the owner, making the effective DF equal to 1.0 in the loss database.

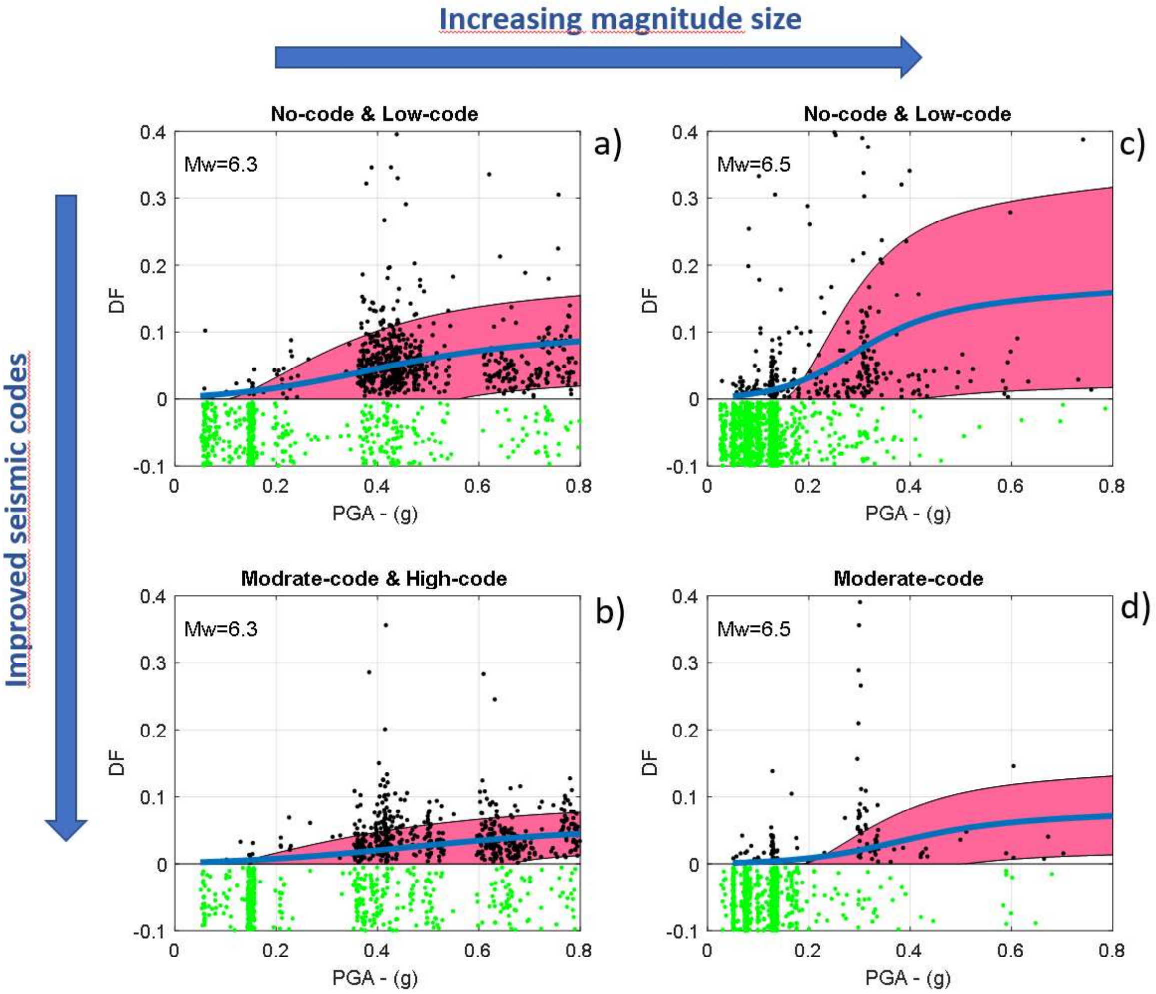

Fig. 3 - Vulnerability model for RC buildings. Mean loss (blue curve) with 16% and 84% prediction limits (pink area). Black dots show loss data with DF>0 and green dots show no-loss data (DF=0) jittered: a) 2008 dataset and CDN & CDL buildings, b) 2008 dataset and CDM & CDH buildings c) 2000 dataset and CDN & CDL buildings, d) 2000 dataset and CDM buildings.

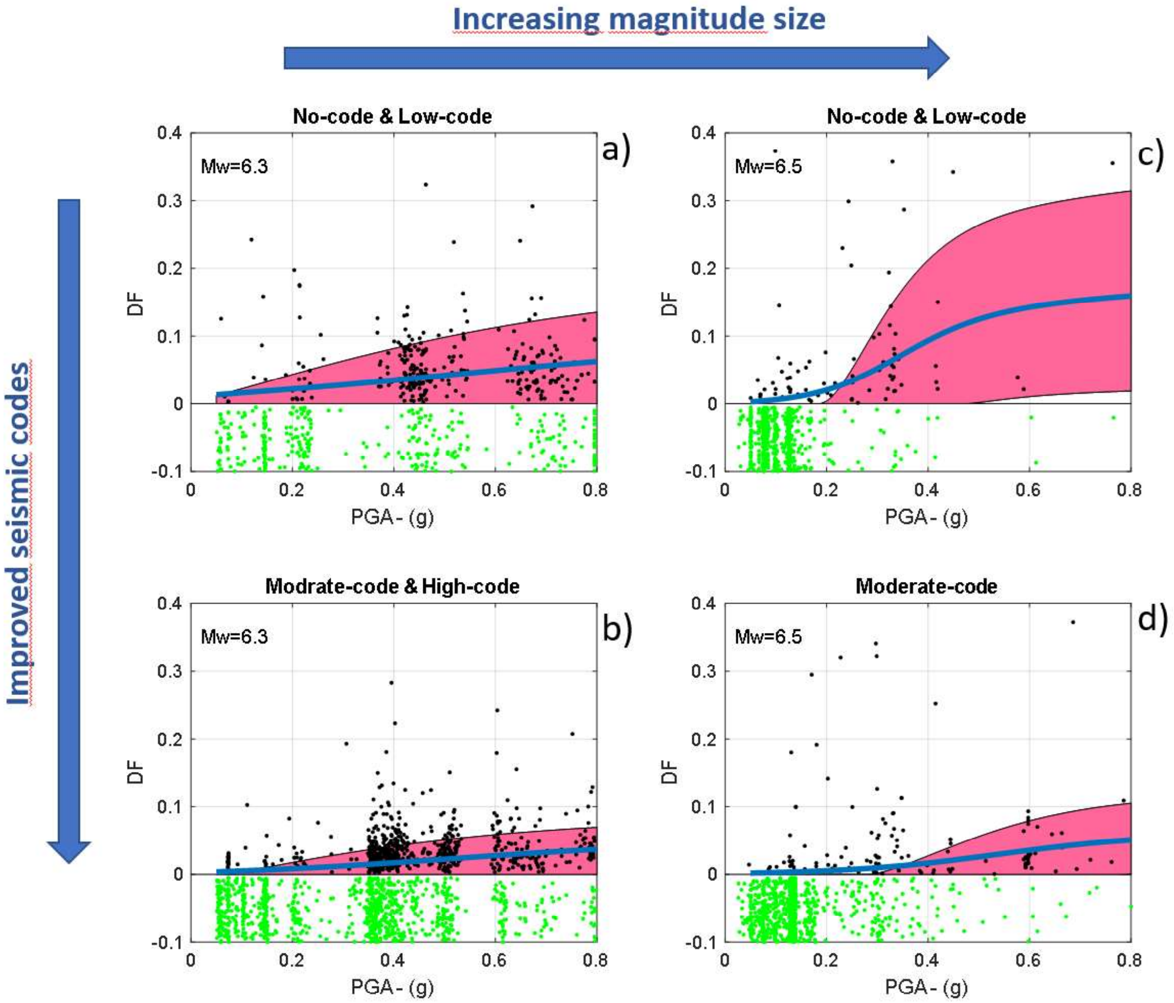

Fig. 4 - Vulnerability model for Timber buildings. Mean loss (blue curve) with 16% and 84% prediction limits (pink area). Black dots show loss data with DF>0 and green dots show no-loss data (DF=0) jittered: a) 2008 dataset and CDN & CDL buildings, b) 2008 dataset and CDM & CDH buildings c) 2000 dataset and CDN & CDL buildings, d) 2000 dataset and CDM buildings.

This means that the DF in some of these cases does not reflect the actual damage and this also creates outliers that affect the regression. It can also be underlined that no buildings suffered full or partial collapse. To account for these outliers all data points with DF>0.85 were replaced with a max value of DF_{max}=0.85. In the 2000 dataset this was done for 33 buildings, and in the 2008 dataset for 23 buildings. To better show the high proportion of buildings with no damage (DF=0) these data points are randomly jittered and plotted in the range -0.1 to 1 (green dots). The horizontal blue arrow in Fig.3 and Fig.4 underlines the effect of increased magnitude size between the May 2008 event (Mw6.3) and the June 2000 events (~Mw6.5). The mean vulnerability curve is considerably higher for the larger 2000 events for alle the four building typologies. Furthermore, the vertical blue arrow in the figures show the effect of improved seismic codes. In all cases moderate-code & high-code (CDM,CDH) buildings (grouped together) perform better than no-code & low-code buildings (CDN,CDL). The vulnerability model, calibrated for each building typology, consists of five parameters which are given in Bessason et al. (2022) along with all formulas needed to construct the curves in Fig.3 and Fig.4.

One simple measure of the quality of the vulnerability model is to use it to predict losses in a scenario event which is the same (size and location) as the event as the empirical loss data was obtained from. The predicted accumulated loss for given building typology can then be normalized with the actual observed accumulated losses. The results of this are shown in Table 4 for both the 2000 and the 2008 datasets. For RC and timber buildings and both datasets the ratio is within 10% from the actual losses except for the no-code & low-code RC buildings ($RC_{CDN,CDL}$) for the 2000 dataset which has also the largest spread of the data (see Fig.3). For the Masonry buildings the difference is up to 13% in the 2008 dataset.

Table 4. Ratio of predicted accumulated loss to actual accumulated loss (R_{Loss}) for the five building typologies and the two datasets. Weighting based on Eq. (10) was used in the regression and DF_{max}=0.85.

Dataset	$RC_{CDN,CDL}$	$RC_{CDM,CDH}$	$W_{CDM,CDH}$	$W_{CDM,CDH}$	$MR_{CDN \& CDL}$
2000	1.19	1.04	1.07	1.09	1.11
2008	1.09	0.99	0.99	1.01	1.13

Finally, the model can easily be used to compute fragility curves (see Fig.2f, and Bessason et al. 2022) for desired damage states which must be defined by DF bins (see Table 3).

5. Conclusions

In June 2000 two earthquakes (Mw6.52 and Mw6.44) struck in the South Iceland with four-day interval. Then in May 2008, the western part of the same region was hit again by single earthquake (Mw6.31). These events affected and damaged similar building typologies. This means that the same workmanship, building traditions, material properties, seismic codes, regulations etc., were used for all the buildings in the region, but of course some of these factors change with time. After the 2000 and 2008 events detailed and complete empirical loss data were collected and two independent loss datasets were established, that is the 2000 and the 2008 loss dataset. Each dataset includes nearly 5000 residential buildings/dwellings.

This study explores and compares the losses suffered by residential buildings during these events from different angles. To facilitate vulnerability modelling and comparison with similar structures, the GEM taxonomy was used to classify the affected buildings and in addition the status of seismic codes (no-, low-, moderate- and high-code) was used to subclassify them loss dataset. A novel statistical vulnerability models (VM) based on zero-inflated beta regression were calibrated from the 2000 and the 2008 datasets. These models are denoted as ICE2000VM and ICE2008VM.

One of the main findings of the study is the detected differences in loss patterns, and resulting vulnerability models, of buildings affected by the June 2000 (Mw~6.5) and the May 2008 events (Mw6.3). The losses caused by the 2000 events (larger magnitude) are substantially higher. The results show that two earthquakes that produce similar PGA at an given site can have very different impact on structures. The study demonstrates the limitation of PGA as a ground motion intensity measure and highlights the pitfalls of combining loss data from different-sized earthquakes in vulnerability modelling with simple intensity measures such as PGA. This all points towards a need for better ground motion intensity measures that can capture the event size effect better.

The calibrated vulnerability models showed consistently that moderate-code & high-code buildings, which were grouped together in study, showed better performance than no-code & low-code buildings, which is expected, but important to report.

Owing to the differences in the two models, for seismic risk assessment in Iceland, it is recommended that ICE2000VM should be used for earthquake scenarios with Mw in the range 6.4-6.6 and the ICE2008VM for those with Mw range 6.2-6.4. Caution is need when extrapolating empirical vulnerability models calibrated from given destructive earthquake to other magnitude sizes.

Acknowledgements

The authors thank the Natural Catastrophe Insurance of Iceland for placing the earthquake damage database and other relevant information at their disposal. This work was partly financed by the SERICE project funded by a Grant of Excellence from the Icelandic Centre for Research (RANNIS), Grant Number: 218149-051. We also acknowledge support from the University of Iceland Research Fund.

References

- Bessason B, Bjarnason JÖ, Gudmundsson A, Sólnes J, Steedman S (2012). Probabilistic Earthquake Damage Curves for Low-Rise Buildings Based on Field Data, Earthquake Spectra. 28(4):1353–78. https://doi.org/10.1193/1.4000082.
- Bessason B, Bjarnason JÖ, Gudmundsson A, Sólnes J, Steedman S (2014). Analysis of damage data of low-rise buildings subjected to a shallow Mw6.3 earthquake, Soil Dynamics and Earthquake Engineering 66:89–101. https://doi.org/10.1016/j.soildyn.2014.06.025
- Bessason B, Bjarnason JÖ (2016). Seismic vulnerability of low-rise residential buildings based on damage data from three earthquakes (Mw6.5, 6.5 and 6.3), Engineering Structures. 111 64–79. https://doi.org/10.1016/j.engstruct.2015.12.008
- Bessason B, Bjarnason JÖ, Rupakhety R (2020). Statistical modelling of seismic vulnerability of RC, timber and masonry buildings from complete empirical loss data, Engineering Structures, 209:109969. https://doi.org/10.1016/j.engstruct.2019.109969
- Bessason B, Rupakhety R, Bjarnason JÖ (2022). Comparison and modelling of building losses in South Iceland caused by different size earthquakes, Journal of Building Engineering, 46:13806. https://doi.org/10.1016/j.jobe.2021.103806
- Brzev S, Scawthorn C, Charleson AW, Allen L, Greene M, Jaiswal K, Silva V (2013), GEM Building Taxonomy Version 2.0. GEM Technical Report 2013-02.
- Crowley H, Despotaki V, Silva V, Dabbeek J, Romão X, Pereira N, Castro JM, Daniell J, Veliu E, Bilgin H, Adam C, Deyanova M, Ademovic N, Atalic J, Karatzetzou A, Bessason B, Shendova V, Tiganescu A, Toma-Danial D, Zugic Z, Akkar S, Hancilar U (2021). Model of seismic design lateral force levels for the existing reinforced concrete European building stock. Bulletin Earthquake Engineering, 19: 2839–2865. https://doi.org/10.1007/s10518-021-01083-3
- Einarsson P (2008). Plate boundaries, rifts and transforms in Iceland, Jökull. 58 (2008) 35–58.
- Ferrari SLP, Cribari-Neto (2004). Beta Regression for Modelling Rates and Proportions. Journal of Applied Statistics. 31(7):799–815. https://doi.org/10.1080/0266476042000214501
- Halldórsson B, Sigbjörnsson R (2009). The Mw6.3 Ölfus earthquake at 15:45 UTC on 29 May 2008 in South Iceland: ICEARRAY strong-motion recordings. Soil Dynamics and Earthquake Engineering, 29 (2009) 1073–83. https://doi.org/10.1016/j.soildyn.2008.12.006
- Ioannou I, Bessason B, Kosmidis I, Bjarnason JÖ, Rossetto T (2018). Empirical seismic vulnerability assessment of Icelandic buildings affected by the 2000 sequence of earthquakes, Bulletin of Earthquake Engineering. 16(12) (2018) 5875–903. https://doi.org/10.1007/s10518-018-0413-x

- Jónasson K, Bessason B, Helgadóttir Á, Einarsson P, Gudmundsson GB, Brandsdóttir B, Vogfjörd, KS, Jónsdóttir K (2021). A harmonised instrumental earthquake catalogue for Iceland and the northern Mid-Atlantic Ridge, Natural Hazards and Earth System Sciences. 21 (2021) 2197–2214. https://doi.org/10.5194/nhess-21-2197-2021
- Natural Catastrophe Insurance of Iceland (2022). http://www.nti.is/en/ (accessed 1 February 2022).
- Ospina R, Ferrari SLP (2012). A general class of zero-or-one inflated beta regression models, Computational Statistics and Data Analysis. 56:1609–23. https://doi.org/10.1016/j.csda.2011.10.005
- R Core Team (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.
- Registers Iceland (2022). http://www.skra.is/english. 2021 (accessed 1 February 2022).
- Rupakhety R, Sigbjörnsson R (2009). Ground-motion prediction equations (GMPEs) for inelastic response and structural behaviour factors, Bulletin of Earthquake Engineering. 7(3) (2009) 637–59. https://doi.org/10.1007/s10518-009-9105-x
- Rossetto T, D'Ayala D, Ioannou I, Meslem A (2014). Evaluation of existing fragility curves, in: Pitilakis K, Crowley H, Kaynia AM, (Eds). SYNER-G: Typology definition and fragility curves for physical elements at seismic risk, Geotechnical, Geological and Earthquake Engineering. 27 1–28. https://doi.org/10.1007/978-94-007-7872-6