Engineering Structures 209 (2020) 109969

Contents lists available at ScienceDirect

Engineering Structures

journal homepage: www.elsevier.com/locate/engstruct

Statistical modelling of seismic vulnerability of RC, timber and masonry buildings from complete empirical loss data

Bjarni Bessason[a,*], Jón Örvar Bjarnason[b], Rajesh Rupakhety[a]

[a] Faculty of Civil and Environmental Eng., University of Iceland, Hjardarhagi 2, 107 Reykjavik, Iceland
[b] Natural Catastrophe Insurance of Iceland, Hlídarsmári 14, 201 Kópavogur, Iceland

ARTICLE INFO

Keywords:
Seismic vulnerability
Zero-inflated beta model
Beta regression
RC buildings
Timber buildings
Masonry buildings

ABSTRACT

In June 2000 two shallow, strike slip, Mw6.5 earthquakes occurred in the middle of the largest agricultural region in Iceland. The epicentres were close to small towns and villages and almost 5000 residential buildings were affected. A great deal of damage occurred but no residential buildings collapsed and there was no loss of life. Insurance against natural disasters is compulsory for all buildings in Iceland and they are all registered in a comprehensive official property database. Therefore, to fulfil insurance claims, a field survey was carried out after the two earthquakes where repair cost was estimated for every damaged building. By combing the loss data with the property database it was possible to establish a complete loss database, where all residential buildings in the affected area were included, both buildings with loss as well as buildings with no-loss. The main aim of the study was to fit a statistical vulnerability model to the data. Due to the high proportion of no-loss buildings in the database (~84%) a new and novel vulnerability model was used based on a zero-inflated beta regression model. The model was fitted to the three main building typologies in the affected region, i.e. low-rise structural wall RC, timber, and masonry buildings. The proposed model can be used to predict the mean and desired prediction limits of the losses for a given intensity level as well as to create fragility functions. All the typologies showed outstanding performance in the two destructive earthquakes, which is important to report, model and learn from.

1. Introduction

Knowledge of seismic hazard, structural vulnerability and exposed structures is fundamental to estimate and mitigate seismic risk. The characteristics of earthquake ground motion depend on the tectonic environment and geology in a given seismic region, and seismic vulnerability of buildings are affected by building traditions, materials and structural system, workmanship and quality of construction in the same region. It is therefore important to “learn from earthquakes” every time a destructive earthquake occurs in a specific seismic zone.

Seismic hazard in Iceland is the highest in North Europe and is comparable to that in South Europe. Within Iceland, most of the damaging earthquakes are strike-slip events at a shallow depth (< 10 km) occurring in two complex fracture zones. One of them is the South Iceland Seismic Zone (SISZ) which lies in the middle of the largest agricultural region in the country. The other one is the Tjörnes Fracture Zone (TFZ) which is mainly off the north shore (Fig. 1) [1,2]. Since 1700, about 25 earthquakes in the magnitude range of 6.0–7.0 have occurred in these zones [3]. Based on historical records, geological evidence, fault mechanism, and crustal strength and thickness, it has been estimated that earthquakes as large as 7.0 (Mw) can be expected in both the SISZ and the TFZ [1,3].

In June 2000, two Mw6.5 earthquakes struck in the middle of the SISZ. High peak ground accelerations ($PGA > 0.6$ g) were registered at some locations in both of these events [4]. Nearly 5000 residential buildings were affected (estimated $PGA > 0.05$ g). A great deal of damage occurred but no residential buildings collapsed, and there was no loss of life.

All buildings in Iceland are registered in a comprehensive official property database [5]. Insurance against earthquakes is compulsory for all buildings in Iceland [6]. Therefore, to evaluate insurance claims, field surveys were carried out after these quakes to estimate damage and repair costs of all affected structures. By combining the official property database and the insurance loss database a comprehensive and complete (all dwellings covered) dwelling-to-dwelling loss database was established, which is rare to find in the literature [7]. In May 2008, a Mw6.3 earthquake struck again in the SISZ and caused significant losses. The same procedure as in 2000 was carried out in order to map and register all losses on a dwelling-to-dwelling basis.

The focus in this study was on the loss data from the two June 2000

---

[*] Corresponding author.

https://doi.org/10.1016/j.engstruct.2019.109969
Received 25 February 2019; Received in revised form 15 October 2019; Accepted 20 November 2019
Available online 27 November 2019
0141-0296/ © 2019 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license
(http://creativecommons.org/licenses/BY-NC-ND/4.0/).

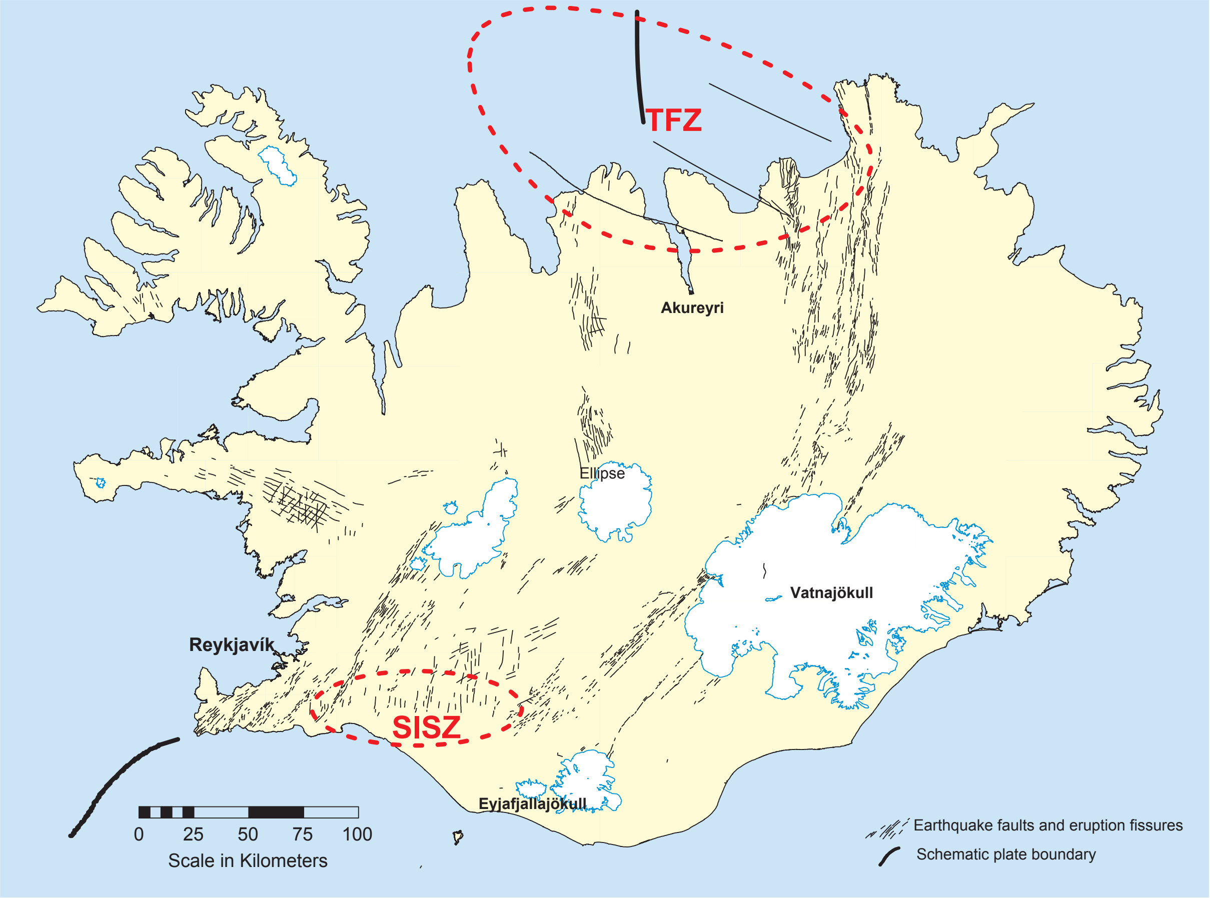

Fig. 1. Map of Iceland showing the South Iceland Seismic Zone (SISZ) and the Tjörnes Fracture Zone (TFZ). The map is based on data from the National Land Survey of Iceland.

events (Fig. 2) covering all residential buildings where the estimated *PGA* was greater than 0.05 g. There were at least four challenges in modelling the data. First there was the uneven spatial distribution of the buildings due to scattered farms spread over the entire area, as well as the denser towns and villages (Fig. 2). Second, a high proportion of buildings had no-loss (84%). This can partly be explained by the fact that most of the buildings were at sites with low *PGA*. Third, some buildings were located between the two faults and may therefore have been affected by both the 17 June and the 21 June earthquakes. Fourth, there was high scatter in the data with both no-loss and total-loss in the same acceleration bins.

The loss data from the May 2008 Mw6.3 earthquake have been studied earlier [8,9]. A subset of the loss data from the two June 2000 earthquakes, along with the loss data from the May 2008 quake, have also been used to evaluate fragility curves based on classical methods [10].

Advanced statistical models are more and more used when working with empirical loss data [11,12]. One such model was recently fitted by Ioannou et al. [13] to the loss data from the two June 2000 Mw6.5 events, and at the same time an effort was made to model accumulated damage from these two nearby events, which occurred 4 days apart on two parallel faults 16 km apart. In Ioannou et al. [13] the loss data were aggregated using an adaptive spatial grid based on certain assumptions and then fitted using a beta regression model [14]. A beta regression model has also been used to fit Australian loss data [15].

The main aim of the present study was to develop further and improve the beta regression model proposed by Ioannou et al. [13]. The new model is based on using building-by-building loss data, and no spatial aggregation was used as in the previous model. Furthermore, an independent model was constructed for each of the three building typologies covered in the study, which gives more flexibility when fitting the data. Finally, and the most important novelty, was that the new model treats the no-loss buildings specially. Such models, based on beta regression, are called zero-inflated beta regression models [16]. They help to bend the vulnerability curves down towards zero loss at low intensity and by that better reflect the dominance of no-loss data in that intensity range, which is more difficult when using standard beta regression. The model is a combination of two regression models, a logistical regression model that makes it possible to predict the probability of incurring loss for a given intensity level, and a conditional beta regression model that can predict the loss distribution given that there was a loss. The combined proposed statistical model allows prediction of loss and desired prediction limits and can also be used to construct fragility curves.

2. Background data and presumptions

2.1. The South Iceland earthquakes of June 2000

In June 2000 two shallow earthquakes of Mw6.5 struck in the SISZ (Fig. 2). The first occurred on June 17, 2000, at 15:41 (GMT) in the eastern part of the zone. It was a right-lateral strike-slip quake, with the fault striking in the north-south direction and had a focal depth of 6.3 km. The second earthquake, also M_w6.5, struck further west on June 21, 2000, at 00:52 (GMT). Like the first one, it was also a right-lateral strike-slip earthquake, with the fault striking in the north-south direction and with a focal depth of 5.3 km. The highest recorded *PGA* in these two events was 0.84 g [4]. The largest aftershock was M_L5.0, and all the other events were of a magnitude less than M_L4.5. Time histories and response spectra of ground motion recorded during the two Mw6.5

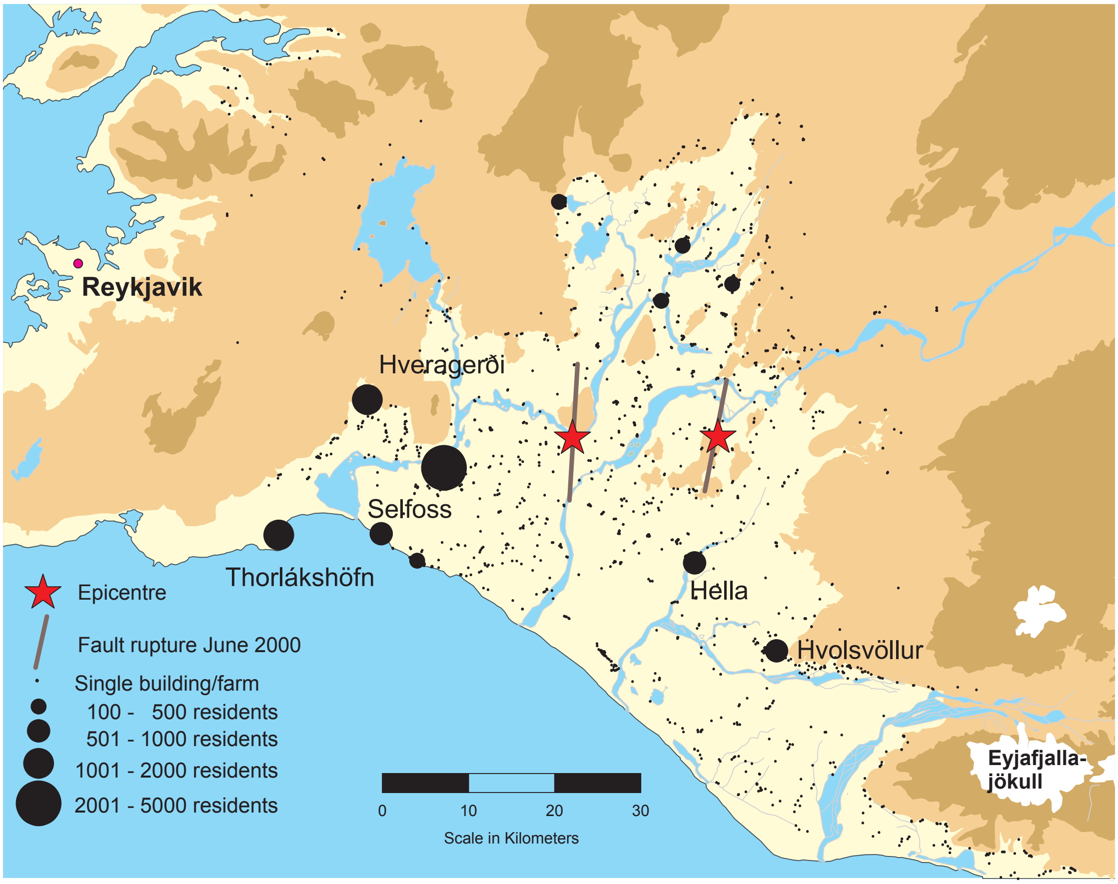

Fig. 2. Epicentre and fault rupture of the two June 2000 South Iceland Mw6.5 earthquakes. The residential buildings are marked with black dots. Larger circles indicate small towns and villages. The data for residents are for the year 2000. (Map data are from the National Land Survey of Iceland and Register Iceland.)

events can be found in the ISESD database [17]. More details of these events can be found in [18–20].

2.2. Ground motion prediction model

In a vulnerability assessment, it is necessary to have an intensity measure (IM) that can be correlated with the observed damage. It is most practical to use single parameter IM that can be obtained by available ground motion prediction equations (GMPE), like *PGA*, *PGV*, or spectral acceleration or spectral displacement at representative structural periods [12,21]. It is, however, also known that damage of structural elements and non-structural components depends on combination of high amplitude impacts and repeated reversals of significant ground and floor motion amplitudes (see for instance [22]). High-frequency peak values of accelerations and short period spectral acceleration can be observed in a wide range of earthquake magnitudes, whereas duration of significant intensive ground motion is more correlated to magnitude.

In this study the affected buildings were low-rise, stiff, and with low natural periods and therefore, for simplicity, the intensity measure was expressed in terms of the *PGA,* which is representative for the short period part of a response spectrum. The GMPE of Rupakhety and Sigbjörnsson [23] was adopted. It is given as:

$$\log_{10}(PGA) = -1.038 + 0.387\cdot \mathrm{M_w} - 1.159\cdot\log_{10}\left(\sqrt{\mathrm{H}^2 + 2.6^2}\right) + 0.123\cdot S + \varepsilon\cdot 0.287 \tag{1}$$

where *H* is the distance to surface trace of the fault in km, *S* is a site factor which takes the value 0 for rock sites and 1 for stiff soil sites. The last term is an error/scatter term where ε follows a standard normal distribution, i.e. $\varepsilon \sim \mathrm{N}(0,1)$. The unit of *PGA* is in $\mathrm{m/s^2}$. Following common practice, the *PGA* level at a given site was estimated as the median *PGA* from Eq. (1), ignoring the error term. A geological map of South Iceland was used to determine the soil conditions at each building site [24]. The adopted GMPE was calibrated to the larger horizontal component from each station. Most of the strong motion recordings were from Icelandic earthquakes but the data were augmented by records from continental Europe and the Middle East. The main characteristic of GMPE given by Eq. (1) is that it predicts a relatively high *PGA* in the near fault area whilst the attenuation with distance is more than generally found in well-known GMPE of similar form from other regions. The high *PGA*s can be explained with shallow earthquakes ($<$ 10 km) in the SISZ, and the high attenuation with distance in Iceland has been explained by the existence of young, fractured and relatively weak rock in the seismic source area that dampens the propagating seismic waves faster than in more solid rock [25,26].

2.3. Accumulation effect from the two events

The fault-to-fault distance of the two June 2000 earthquakes was about 16 km. Some buildings, especially those located between the two faults were, to some degree, affected by both events and must have experienced some accumulated damage (Fig. 3). Approximately 100 buildings were located in a rectangle bounded by the two faults in the west and east, and by a line drawn between the two end points of the fault rupture in the north and similar line in the south (blue dotted lines in Fig. 3). This corresponded to about 2% of all the affected buildings. A computed shake map based on Eq. (1) shows the high attenuation, and

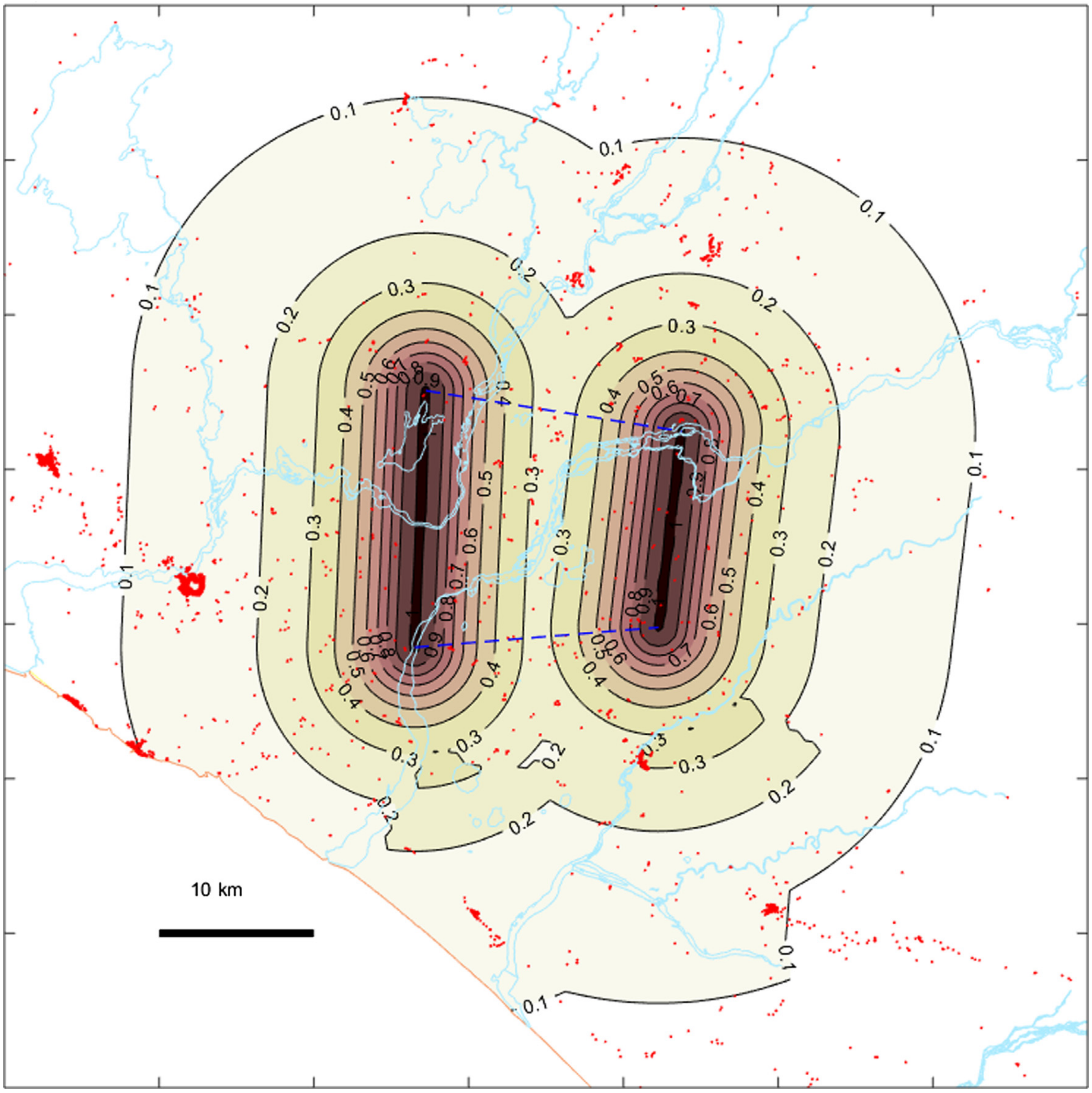

Fig. 3. Shake map for the two June 2000 earthquakes in South Iceland (see also Fig. 2). The map shows computed *PGA* contours (g) based on Eq. (1) and Eq. **(2)**. Red dots show building locations. About 100 buildings were located within the box limited by the two faults and the dotted blue lines. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

for the majority of the buildings located between the two faults the most intense shaking must have been dominated by the nearest fault (Fig. 3). Only a very few buildings were exposed to the same/similar ground shaking from both events. For simplicity in this study, all the losses were treated as if they were caused by a single event with ground motion intensity defined by the *PGA*, corresponding to the larger one of the two events, i.e.:

$$PGA_{\max} = \max(PGA_{EQ17June},\ PGA_{EQ21June}) \tag{2}$$

Table 1
Classification of dwellings in Iceland in both urban and rural areas based on data from Icelandic Property Registers.

Building material	Lateral structural System	No. of storeys	Percentage of dwellings belonging to each class
Reinforced concrete	Structural walls	1–3	53.7%
		4–7	26.4%
		> 7	4.6%
Timber	Structural walls	1–3	12.5%
Masonry	Structural walls	1–3	2.8%
Total:			**100%**

2.4. Property database and main characteristics of the buildings

The Property Registry in Iceland contains information on construction year, main construction material, number of storeys, geographical coordinates and occupancy for all properties. It also contains results of valuation, both for taxation and reconstruction insurance value (replacement value) [5]. Based on this database, all dwellings in Iceland are classified in Table 1 by main construction material and number of storeys. It should be noted that although the official database does not list structural bearing systems, almost all residential buildings have structural walls for resisting lateral seismic forces. This is valid for reinforced concrete (RC), timber, and masonry buildings. In contrast, South European building forms, consisting of moment-frames with or without masonry or brick infills hardly exist in Iceland.

The study area consists mainly of agricultural land with many farms and few small towns, villages and service centres (Fig. 2). The residential buildings were low-rise single-family buildings, but there were also two-family duplexes, town houses and apartment blocks. In total

Fig. 4. Distribution of the loss data for RC, timber and masonry buildings. PGA_{max} was computed with Eq. (2) and DF with Eq. (3).

4776 residential buildings were affected by the two June 2000 earthquakes. Here the affected area was defined as where the computed PGA_{max} from Eqs. (1) and (2) was more than 0.05 g and corresponded to a 35 km distance from the fault rupture (Fig. 2). Of the 4776 buildings, 54% were built of reinforced concrete (RC), 36% of timber, 9.3% were masonry buildings, and the rest, only 0.3%, were built of other building material. The RC buildings constructed before 1976 had a limited amount of reinforcement, typically only around openings in structural walls (low code). Furthermore, 68% of the buildings were one storey, 23% two storeys, 7.9% three storeys, 0.3% four storeys, and no buildings higher. The building stock was quite young in an international context. No building was constructed before 1870 and 92% of them were constructed after 1940. These numbers are different from those shown in Table 1, which reflect numbers for the whole country.

Most of the RC buildings were in-situ cast, although there were a few prefabricated buildings. Prescribed wind loads are very high in Iceland, among the highest in Europe. The *fundamental value of the base wind velocity* is $v_{b,0} = 36$ m/s according to the National Annexes of Eurocode 1-1-4 [27,28] and is the same for the whole country. Based on old tradition and craftsmanship, the Icelandic timber houses are therefore strongly built and well suited to withstand earthquake forces. The bottom floor slab and the foundations are usually made of reinforced concrete, as in the concrete houses. The masonry buildings were built of unreinforced manufactured hollow pumice (*high porosity volcanic rock*) blocks in walls and tied together with rigid RC floors. The weight density of the pumice blocks is low, typically around 14 kN/m^3, and consequently the inertia forces are lower than in ordinary Southern Europe stone or clay brick masonry buildings. The masonry buildings were mainly built before 1980 and are no longer constructed [9].

2.5. Loss data

The present study was restricted to residential buildings. Repair cost was assigned for each dwelling where the owner reported damage after the earthquakes. The loss was estimated by trained technical people, who made a detailed report for each property. Since all dwellings have compulsory catastrophe insurance provided by the Natural Catastrophe Insurance of Iceland [6], it can be assumed that all damaged dwellings were covered as it financially benefits the owners to file a claim. By combining the loss information with the official property database, a complete loss data record (all dwellings included) for all the affected area was created. Such complete loss data for a given destructive earthquake is exceptional to find in post-earthquake damage studies [7].

In this study, the data were simplified by aggregating losses from all dwellings within the same building. Here, "building" was defined by the street address. As an example, a building (block) with 12 apartments divided into three staircases, each with its own street address, was classified as three buildings (for more details see [13]). The loss measure used in this study was determined by normalising the estimated repair cost with the replacement value taken from the official property database. This is called the damage factor *DF*:

$$DF = \frac{\text{Estimated repair cost}}{\text{Replacement value}} \tag{3}$$

The replacement value is the same as the fire insurance value of a building and is the depreciated replacement value plus the cost of removing the destroyed building. The depreciation is based on age, building material and general condition. On the other hand, the repair cost was in general not depreciated. The loss cannot be greater than 1 (100%) and in practice the expression "total loss" was assigned to most of the residential buildings that suffered an estimated repair cost of more than 70% of their replacement value. Assignment of total loss to these buildings was in some cases due to administrative procedures involved in loss estimation. That is, when there is a disagreement about insurance compensation between house owner and the insurance company, it sometimes is more economical for the insurer to approve a higher *DF* than "true" *DF* to avoid expensive litigation.

The three main typologies, i.e. RC buildings, timber buildings and masonry buildings, were analysed in the study as they cover more than 99% of all residential buildings in the affected area.

3. Exploration of the data

Most of the affected buildings ($PGA_{max} > 0.05$ g) were exposed to low *PGA* values (Fig. 4). This was due to two main reasons, that is, areas experiencing lower intensity shaking were larger than those experiencing higher shaking, and the main towns and villages including most of the building stock were located at some distance from the faults (Fig. 2). The scatter in the data was quite wide and there were buildings with both no-loss ($DF = 0$) and total loss ($DF = 1$) which were exposed to the same computed intensity. The scatter can partly be explained by the inherent variability in the ground motion, site-specific soil conditions, and the variability in structural and non-structural performance of the buildings due to construction age, irregularities, maintenance, workmanship, etc. It should be noticed that even at low intensity there were examples of total loss. To get a better overview the data were classified in four acceleration bins (0.05–0.10 g, 0.10–0.20 g, 0.20–0.40 g > 0.40 g) and then the number of all buildings within each bin as well as the proportions of no-loss, loss and total-loss buildings were computed (Fig. 5). The majority of the buildings as well as the highest proportions of no-loss were in the two first acceleration bins. In fact, the no-loss buildings dominated in these two bins for all the three building typologies.

4. Statistical vulnerability model

4.1. Beta regression models

The two-parameter beta distribution is useful to model random variables which are bounded in the range $x \in (0, 1)$, i.e. where the occurrence of $x = 0$ and $x = 1$ is not possible. When the distribution of these bounded random variables systematically change with one or more explanatory variables, then beta regression models may be preferable. The model parameters then variate and are then associated with the explanatory variables through a linear predictor with unknown coefficients and a link function [14]. In cases when the random variables can also take a value of zero or one (or both), extended beta regression models exist which allow these values. Such a model is called a zero-or-one-inflated beta model [16].

The damage factor, *DF* (see Eq. (3)), can be considered as a random variable, which mean value changes with increased ground motion intensity, that is, more losses can be expected with increased intensity. The *DF* can also reflect no-loss ($DF = 0$) and total loss buildings ($DF = 1$). In our dataset no-loss buildings dominated the two lowest acceleration bins and were also substantial in the other two bins in Fig. 5. In fact, nearly 85% of residential buildings had no-loss. On the other hand, total loss buildings ($DF = 1$) were very few in all the cases, less than 1%, and these cases can be treated by assigning them a value of less than a unit. Hence, a zero-inflated beta regression model, for $x \in [0,1)$, where the beta distribution parameters are related to linear predictors and link function can be expected to be sufficient and preferable to model the data and deal with all the no-loss buildings.

For insurance purposes as well as mitigation planning it is useful to have models that can predict the probability of incurring loss ($DF > 0$) for a given intensity or assumed scenario event. Therefore, in this study two-step regression analysis was used to build the zero-inflated beta regression model, which was shown to be practical methodology.

First, a logistical regression model was computed, followed by conditional beta regression given that there was a loss. The two regression models were then combined to create the final vulnerability model. The model than can be used to predict mean loss and desired prediction limits.

4.2. Logistical regression model

The first step was to use a logistic regression to model the probability of incurring loss ($DF > 0$) as a function of PGA_{max}. The model for each building typology was given as:

$$\log\left(\frac{p_j}{1-p_j}\right) = \beta_{0,j} + \beta_{1,j}\cdot PGA_{\max} \tag{4}$$

where p_j is the probability that $DF > 0$ for given PGA_{max}, i.e. the probability of getting *True* value in the binary process; $\beta_{0,j}$ and $\beta_{1,j}$ are the regression parameters; PGA_{max} is the larger *PGA* during the two events, and *j* refers to building typology, i.e. $j = \{\text{RC, timber, masonry}\}$.

4.3. Conditional beta regression model

The logistical regression model only gives information about the probability of incurring loss, but it contains no information on the extent or distribution of loss. A conditional probability model for the loss expressed by *DF*, given the occurrence of a loss ($DF > 0$), is modelled by a beta distribution which is bounded in the unit interval (0, 1). In the case of total loss ($DF = 1$) of a building, only occurring in a very few cases, *DF* was replaced with a value less than a unit. The probability density function (PDF), expected value and variance of the model, are respectively given as:

$$\begin{aligned} f(x\,|DF>0) &= \frac{\Gamma(\varphi)}{\Gamma(\mu\varphi)\Gamma(1-\mu)\varphi}x^{\mu\varphi-1}(1-x)^{(1-\mu)\varphi-1} && 0<x<1 \\ E[X\,|DF>0] &= \mu && 0<\mu<1 \\ \mathrm{Var}[X\,|DF>0] &= \frac{\mu(1-\mu)}{1+\varphi} && \varphi>0 \end{aligned} \tag{5}$$

In Eq. (5), μ is the mean value and φ is the precision [14]. It should be noted that in many references and software, the beta PDF is expressed differently than in Eq. (5) and the two factors are then called shape parameters (see for instance Wikipedia.org; Python (SciPy.org), Matlab (Mathworks®)). An appropriate transformation of the parameters is then needed. A beta regression model links μ and possibly φ with a systematic component that is a function of a vector of explanatory variables. The mean value, μ, is related to the explanatory variables through a link function, $g_1(\cdot)$:

$$\mu = g_1^{-1}(\eta_1) \tag{6}$$

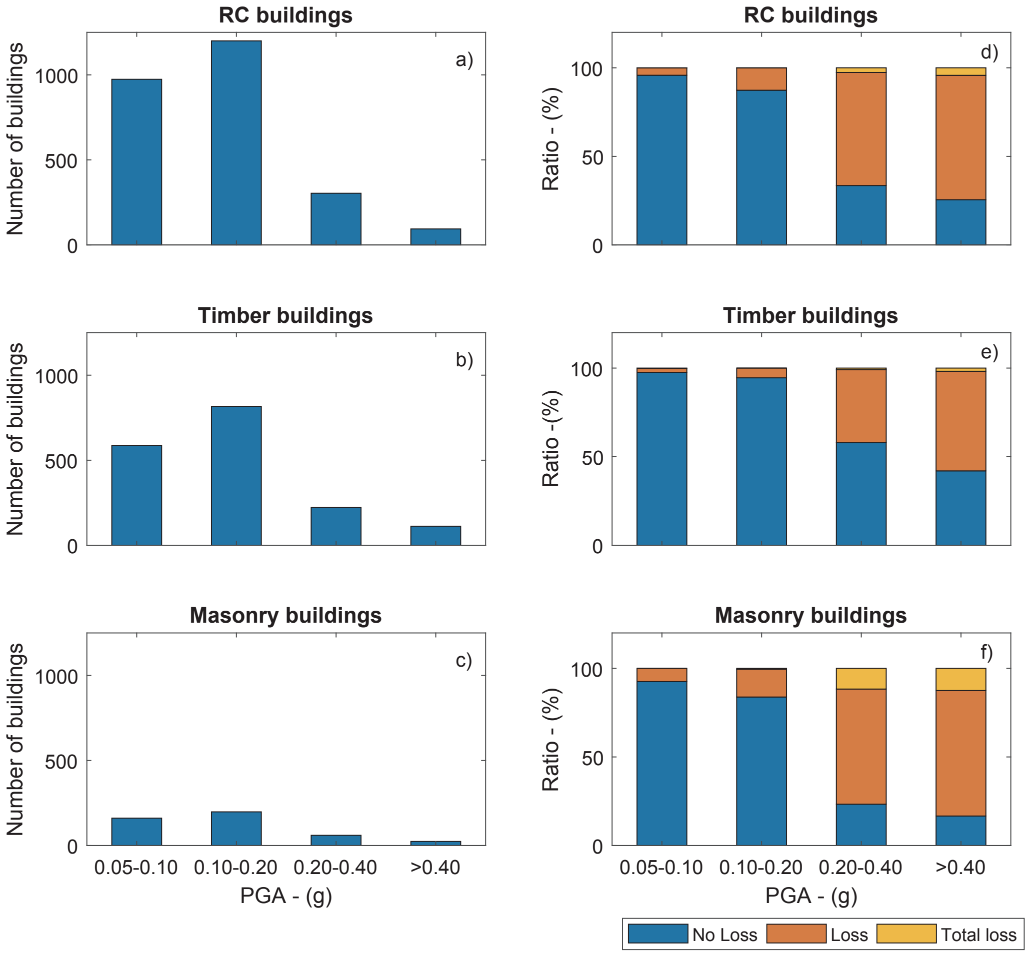

Fig. 5. *Left bars:* Number of buildings as a function of acceleration bins. *Right bars:* Distribution between "No loss", "Loss" and "Total Loss" as a function of acceleration bins.

where η_1 is a function of the explanatory variables. The logit link function was adopted in this study:

$$g_1(\mu) = \text{logit}(\mu) = \log\left(\frac{\mu}{1-\mu}\right) \tag{7}$$

Similarly, φ can be considered as a constant intercept or a function of the explanatory variables, η_2, through a link function, g_2:

$$\varphi = g_2^{-1}(\eta_2) \tag{8}$$

where η_2 is determined by the plot of residuals against η_1. If the residuals appear to be randomly distributed the use of a constant precision is adequate. In this study, the link function of the precision was expressed in the form:

$$g_2(\varphi) = \log(\varphi) \tag{9}$$

Having determined the main properties of the statistical model, η_1 and η_2 need to be expressed as functions of the explanatory variables. Instead of fitting one model to all the three building typologies using categorical variables, three models were fitted independently for each of the typologies.

In Ioannou et al. [13] different combinations of explanatory variables were tested to find a "best function" for η_1 to use in Eq. (6). Among the tested explanatory variables were construction year class (categorical variable), building typology (categorical variable), and ground motion intensity defined by *PGA*. The model that gave the best results included PGA_{max} (see Eqs. (1) and (3)) and building typology. Furthermore, $\log(PGA_{max})$ gave a better fit than PGA_{max}. Based on these observations, the functions of explanatory variables were taken in this study as:

$$\eta_{1j} = \theta_{0j} + \theta_{1j} \times \log(PGA_{\max})$$
$$\eta_{2j} = \theta'_{0j} \tag{10}$$

where θ_{0j}, θ_{1j} and θ'_{0j} are the regression parameters of the conditional beta regression model and j = {RC, timber, masonry}. In this model all three parameters, θ_{0j}, θ_{1j} and θ'_{0j} are related to building typology, whilst in [13] only θ_0 was related to building typology but the other two parameters were common. Separate treatment of these parameters made the model more flexible.

4.4. Combination of the logistical and the conditional beta regression model

To determine the expected value of loss or desired prediction limit, the logistical regression model and the conditional beta model were combined. If X is defined as a random variable representing building *DF* then, based on the given model, there are two mutually exclusive events (E_1 and E_2) that can occur for a given building typology j and for given PGA_{max}.

$$E_1 - \text{No damage},\quad X=0,\quad P[X=0\ |E_1]=1 \qquad P[E_1]=1-p_j$$
$$E_2 \text{ - Damage},\quad 0<X<1,\quad P[X\leqslant x\ |E_2]=F_X(x,\mu_j,\varphi_j) \quad P[E_2]=p_j \tag{11}$$

Here P[·] stands for probability, p_j is the probability of loss ($DF > 0$) for given PGA_{max} (see Eq. (4)) and $F_X(x,\mu_j,\varphi_j)$ is the conditional beta cumulative distribution function for a given building typology j with parameters μ_j, φ_j, both of which are functions of PGA_{max} (see Eqs. (7)–(10)). The total probability theorem can now be used to obtain the expected *DF*:

$$E[DF_j] = E[X|E_1]\cdot P[E_1] + E[X|E_2]\cdot P[E_2] = 0\cdot(1-p_j) + \mu_j\cdot p_j = \mu_j\cdot p_j \tag{12}$$

The total probability theorem can also be used to compute the limits of the desired prediction interval as:

$$\begin{aligned} \mathrm{P}[X<x] &= P[X<x|E_1]\cdot P[E_1] + \mathrm{P}[X\leqslant x|E_2]\cdot P[E_2] \\ &= 1\cdot(1-p_j) + F_X(x,\mu_j,\varphi_j)\cdot p_j \\ &= 1 + p_j(F_X(x,\mu_j,\varphi_j)-1) \end{aligned} \tag{13}$$

By, for instance, putting $\mathrm{P}[X < x] = 0.90$ and solving Eq. (13) it is possible to find the 90% upper bound for *DF*, i.e. 90% of the losses (*DF*) will be less than this value.

5. Results and discussion

The logistic regression was performed using the *glm()* function in the R package [29]. To prepare the data for regression all buildings with loss, $DF > 0$, were replaced with $DF = 1$, whilst buildings with $DF = 0$ were left unchanged. The β-parameters in Eq. (4) were then estimated individually for the three building typologies (Table 5). The null-hypotheses ($\beta_0 = 0$ or $\beta_1 = 0$) could be rejected with high z-values ($\mathrm{P}[>|z|] < 2 \times 10^{-16}$) for both the regression coefficients in all three cases. The models are compared in Fig. 6. Results of the logistical regression alone are helpful in finding the probability of incurring loss for a given ground motion intensity, which is useful for post-event emergency planning and risk mitigation.

A simplification and presumption in the regression was to assume that buildings located between the two faults were only affected by one event, i.e. either the 17 June or the 21 June earthquake, depending on which gave a higher *PGA* at the site in question (Eq. (2)). The fault-to-fault distance of the two earthquakes was about 16 km. The GMPE given in Eq. (1) predicts a median *PGA* value of 0.12 g for $H = 16$ km and Mw = 6.5. For $PGA = 0.12$ g the probability of incurring loss ($DF > 0$) for the three typologies was in the range of 6 to 13% based on the logistical regression (Fig. 6). At a distance of 10 km ($PGA = 0.20$ g), > 65% of the buildings were undamaged for all three building typologies (Fig. 6). Therefore, the main damaging effect of buildings located in the near-fault area, say less than 5 km from either fault rupture, were by the closest event, although some accumulated effect may have occurred. In addition, only a few buildings were located in the central zone, where there was a similar computed intensity from both events (Figs. 2 and 3). This supports the assumption that the losses were mainly controlled by the nearest fault and the effect of accumulated damage was limited.

The scatter in the data, i.e., a wide range of *DF* for a given ground motion intensity, is a well-known challenge in empirical vulnerability modelling. The *DF* for a great majority of the buildings exposed to low intensity shaking was low, but there were some buildings which were assigned total loss, i.e. $DF = 1.0$ (see also Section 2.5). Incurring total loss at a low PGA_{max} was usually related to old buildings in bad condition with low replacement value prior to the earthquake. Anyway, these high values were a kind of outlier in the loss data and affected the beta regression which was carried out by the *betareg()* function in the R package [29]. In beta regression all *DF* values must be in the range $DF \in (0, 1)$. As mentioned before, all data points with $DF = 0$ were removed from the database in the logistical regression whilst $DF = 1$ cases were not. A common practice in beta regression is to replace unit values with high values, like $DF_{max} = 0.99$, but such values were nevertheless outliers as most of the data had low *DF* values. It turned out that the regression was sensitive to what values were assigned to the unit values in the database. Since buildings with a loss exceeding 0.7 were in most cases assigned a 1.0 loss it may seem fair to assign all data points with $DF > 0.85$ a max value $DF_{max} = 0.85$, which is midway between 0.7 and 1.0 (mean value).

To test the quality of the proposed model for using different DF_{max} values three measures were used. First, residuals were computed based on the *sweight2* formula/method as recommended by Espinheira et al. [30], see also the Betareg model in R [14,29]. Secondly, two ratios, R_{DF} and R_{Loss}, were computed where predicted values for every data point were compared to real values of the data points computed from the database, i.e.:

$$R_{DF} = \frac{\text{Predicted mean DF}}{\text{Mean DF from loss data}} \tag{14}$$

$$R_{Loss} = \frac{\text{Predicted accumlated loss}}{\text{Accumulated loss from data}} \tag{15}$$

Plots of residuals against the linear predictor η_1 are shown in Fig. 7a–c for the $DF_{max} = 0.99$ and in Fig. 7d–f for $DF_{max} = 0.85$. The ratios R_{DF} and R_{Loss} for each case are shown in Table 2.

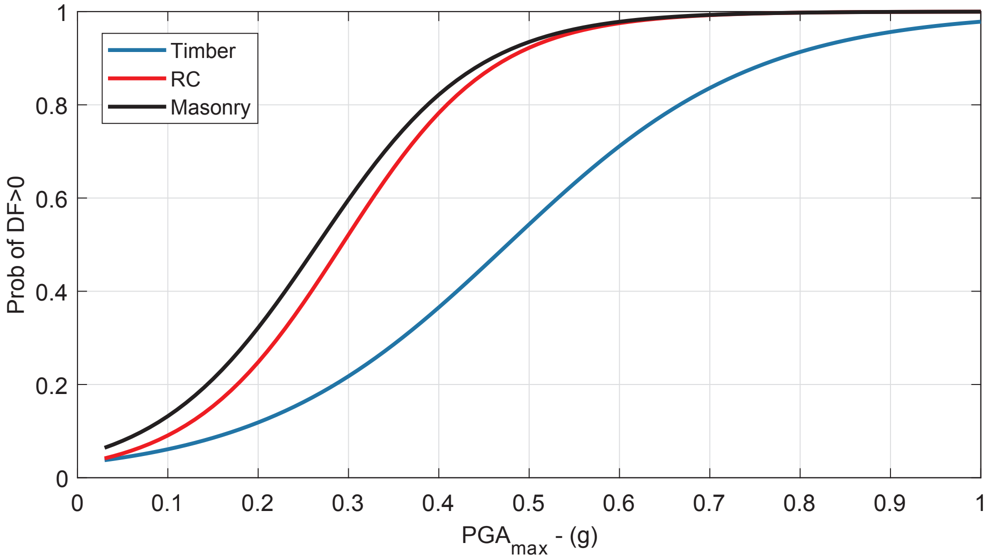

Fig. 6. Probability of getting $DF > 0$ as a function of PGA_{max} for three different building typologies in South Iceland based on logistic regression model. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

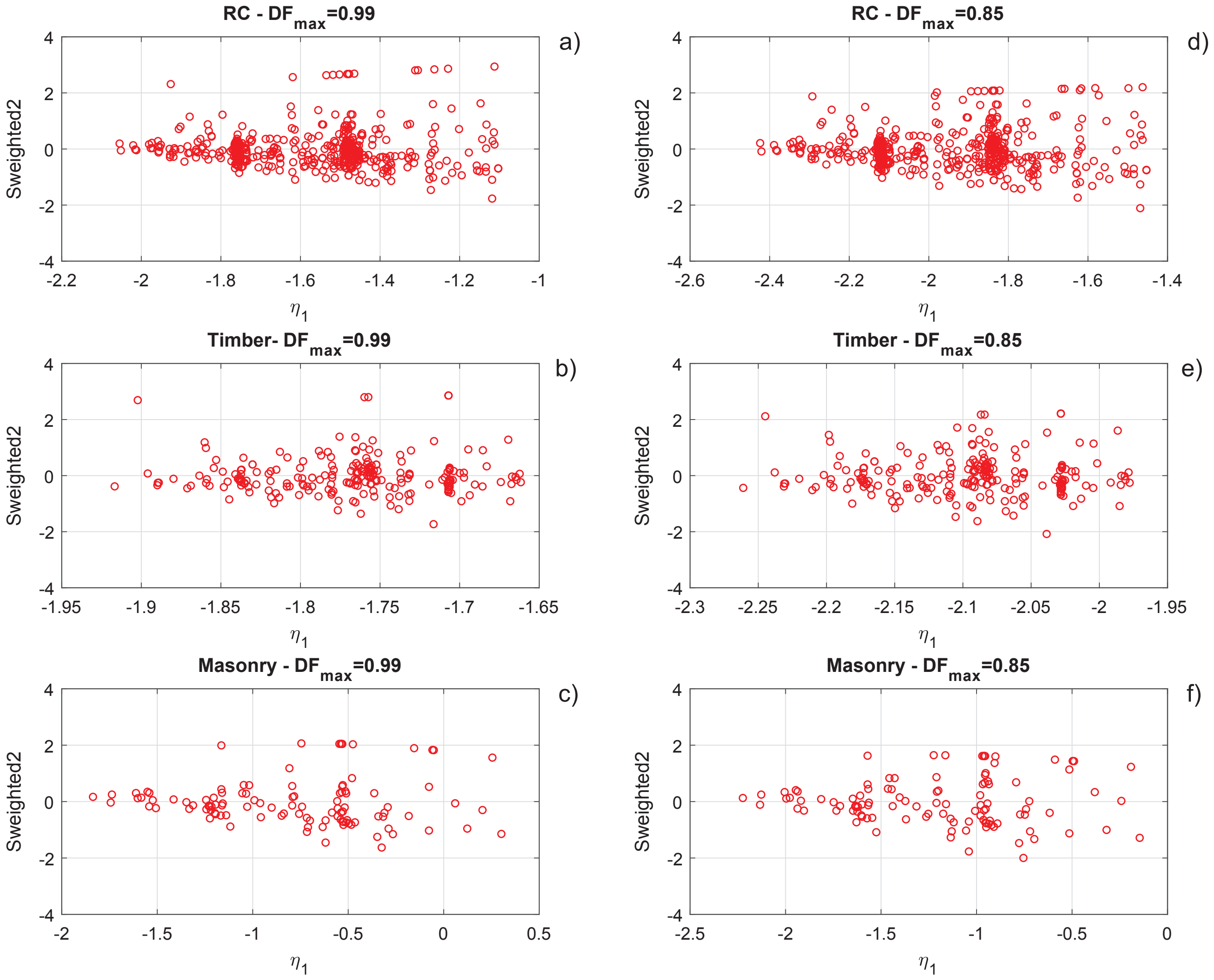

Fig. 7. Residuals for all the three building typologies for two different DF_{max} values, i.e. 0.99 and 0.85.

Table 2
Ratio of predicted mean *DF* to mean *DF* from loss data (R_{DF}) and the ratio of the predicted accumulated loss to real accumulated loss (R_{Loss}) for different assignments of DF_{max}. No weighting used.

Assigning of DF_{max}	RC buildings		Timber buildings		Masonry buildings	
	R_{DF}	R_{Loss}	R_{DF}	R_{Loss}	R_{DF}	R_{Loss}
0.99	1.59	2.08	1.52	2.08	1.38	1.76
0.85	1.17	1.54	1.14	1.56	1.04	1.32

By using $DF_{max} = 0.85$ the residuals are more evenly spread around the zero line and are in most cases between -2 and 2, which is preferable. Both R_{DF} and R_{Loss} are closer to unit, as preferable, if $DF_{max} = 0.85$ is used instead of $DF_{max} = 0.99$. Although the predicted mean *DF* may be at an acceptable level (1.04 to 1.17) the model overpredicts the losses (1.32 to 1.56).

The main aim of the study was to develop a reliable statistical vulnerability model that can be used to predict financial losses. New buildings have higher replacement value than older buildings, both because they are newer but also because of higher standards (more luxurious). Older buildings are generally more likely to be assigned a higher DF than the new ones, due to low-code design and more age degeneration of the building itself. These effects are ignored in the normalised *DF*, where low-cost old buildings and more luxurious (and newer) buildings have the same weight in the regression. This affects the prediction of accumulated loss (Table 3). To account for this, a factor, w_i, was used to weight the *DF* for each building, *i*, before the regression was carried out:

Table 3
Ratio of predicted mean *DF* to mean *DF* from loss data (R_{DF}) and the ratio of predicted accumulated loss to actual accumulated loss (R_{Loss}) for different assignments of DF_{max} and where weighting based on Eq. (16) was used in the regression.

Assigning of DF_{max}	RC buildings		Timber buildings		Masonry buildings	
	R_{DF}	R_{Loss}	R_{DF}	R_{Loss}	R_{DF}	R_{Loss}
0.99	1.15	1.52	1.02	1.40	1.14	1.45
0.85	0.91	1.19	0.83	1.14	0.87	1.11

$$w_i = \frac{\text{Replacment cost of building } i}{\min(\text{Replacment value})} \tag{16}$$

By this the building with the lowest replacement value was assigned a weight of "1″ whilst more expensive ones get higher weights. The weighting, which was done independently for each building typology, helped to improve the statistical model and by this it was possible to

Table 4
Total number of buildings; number of buildings with $DF > 0$; and number buildings with $DF > 0.85$ that were assigned a max value of $DF_{max} = 0.85$.

	N_{total}	$N_{DF>0}$	$N_{DF=0.85}$
RC	2572	465	15
Timber	1739	218	5
Masonry	443	110	13

predict accumulated loss closer to actual loss from the loss data. The effect of weighting is shown in Table 3 for two cases as $DF_{max} = 0.99$ and $DF_{max} = 0.85$.

After this, the accumulated loss was over-predicted in the range of 11 to 19%. The conclusion was therefore to use the weighting based on Eq. (16) and assigning of $DF_{max} = 0.85$ to all DF values greater than 0.85 to estimate the modal parameters. Overview of the loss data is given in Table 4 and the estimated model parameters in Table 5.

The mean vulnerability curves for RC, timber and masonry buildings are compared in Fig. 8 using the proposed model and the regression parameters from Table 5. In Fig. 9 the curves are shown for each building catalogue with 90% prediction limits (5% below and 5% above). For instance, for $PGA_{max} = 0.8$ g, the predicted mean loss was 8% for the timber buildings and 95% of them incurred a loss less than 29% of their replacement value. In all cases, the mean curve as well as the upper 95% limit exhibited a sharp decline at low PGA_{max}, which is in good agreement with the loss data (Fig. 5).

In Fig. 10 the proposed statistical model in this study is compared to two models presented by Ioannou et al.[13], which were also fitted to the same loss data but based on beta regression model instead of zero-inflated beta regression. In the study by Ioannou et al. the loss data was split in two groups, i.e. buildings that were believed to be affected by 'both' the 2000 events (accumulated loss) and then buildings only affect by 'single' event (see also Section 2.3). The mean DF within each acceleration bin in Fig. 5 is also shown for comparison in Fig. 10, where the x-coordinate is the central value of each bin. At low PGA_{max} both the 'single' model and the 'both' model curves are above the upper 95% prediction limits of the proposed model in this study for all the three building typologies, whilst the mean value of the loss data within each accelertion bin (the yellow dots) fairly well fit the model in this study at this PGA_{max} level. This indicates that the proposed model in this study is indeed an improved model of the previous model (Ioannou et al. [13]) as was one of the main objective of this study.

From the statistical model and Eq. (13) ($P[X > x] = 1 - P[X < x]$) it is also possible to construct fragility curves which can be used for comparison with other studies. Four damage states were defined using loss bins to define the damage states:

DS0 − *minor*, loss less than 1% of replacement value;
DS1 − *slight*, loss in the range 1–5% of replacement value;
DS2 − *moderate*, loss in the range 5–20% of replacement value;
DS3 − *substantial*, loss in the range 20–50% of replacement value;

The fragility curves are shown in Fig. 11. The probability of exceeding DS3 ($DF > 0.50$) is very low for both RC and timber buildings at all intensity levels.

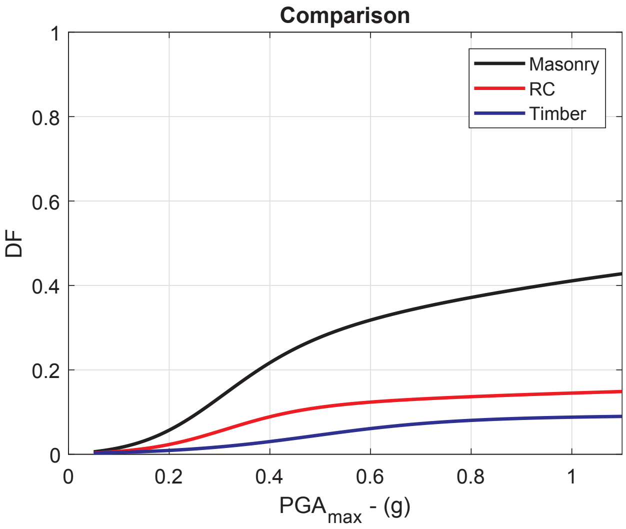

Fig. 8. Mean vulnerability curves for RC, timber and masonry buildings.

6. Conclusions

This paper describes a new and novel statistical vulnerability model which can be used to predict seismic losses for low-rise Icelandic buildings. The proposed model, a zero-inflated beta regression model, was based on a two-step regression of detailed and complete building-by-building loss data which were recorded in the aftermath of two June 2000 Mw6.5 earthquakes in South Iceland. The database covers almost 5000 residential buildings. The first regression provides a logistical model representing the probability of experiencing damage for a given ground motion intensity level. The second regression provides a beta distribution representing the extent of loss conditioned on occurrence of loss. The predicted losses include both structural and non-structural losses (interior finishing work, partition walls, cladding, interior fixtures, paintwork, flooring, plumbing, electrical installations, etc.), excluding the loss to household contents. The ground motion intensity measure is expressed in terms of *PGA*. Three independent, five-parameter models are presented, i.e. one for RC, one for timber and one for masonry buildings. The statistical model was used to compute vulnerability functions with 90% prediction bounds. The model was also used to construct fragility curves for four damage states which were defined by loss bins.

It is of interest to note that despite very strong ground shaking ($Mw = 6.5$) the mean loss was less than 15% for RC buildings and less than 9% for timber buildings in the near-fault area at high *PGA*s. For masonry buildings, however, the mean loss was above 40% in the vicinity of the faults. When the earthquakes struck in June 2000 less than 10% of the residential buildings were masonry buildings, and they are no longer constructed.

It is believed that the models give conservative predictions for magnitudes lower than $Mw = 6.5$, but they should be used with caution to predict losses in larger magnitude events.

Table 5
Estimated model parameters, Mean and Standard Error (SE) based on two-step regression.

	β_0		β_1		θ_0		θ_1		θ_0'	
	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE
RC	−3.503	0.123	11.953	0.632	−1.774	0.031	0.305	0.023	1.645	0.024
Timber	−3.457	0.143	7.267	0.517	−2.315	0.025	0.103	0.025	1.894	0.027
Masonry	−3.025	0.259	11.370	1.357	−0.360	0.031	0.725	0.019	1.012	0.017

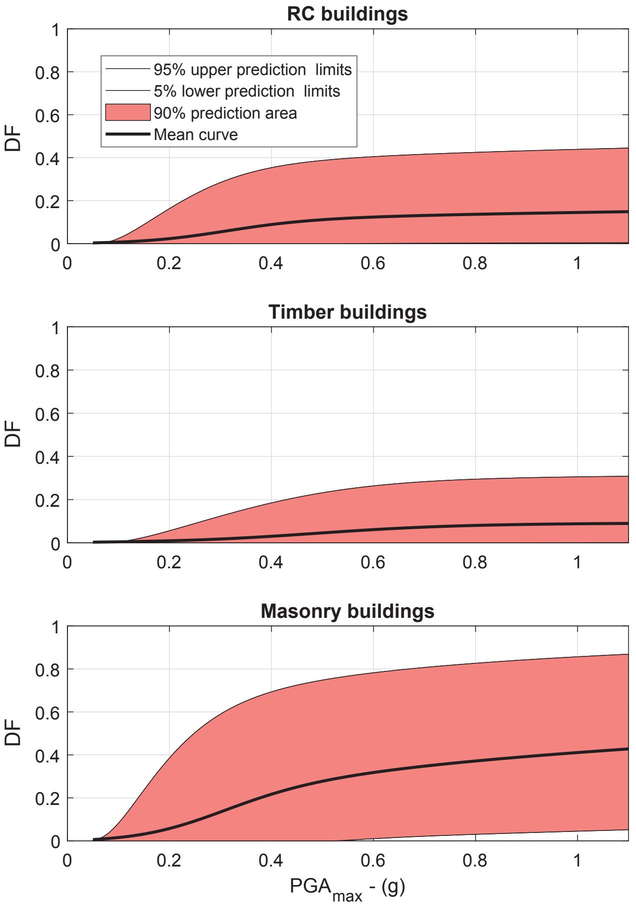

Fig. 9. Mean vulnerability curve (solid black line) with 95% upper and 5% lower prediction limits for RC, timber and masonry buildings.

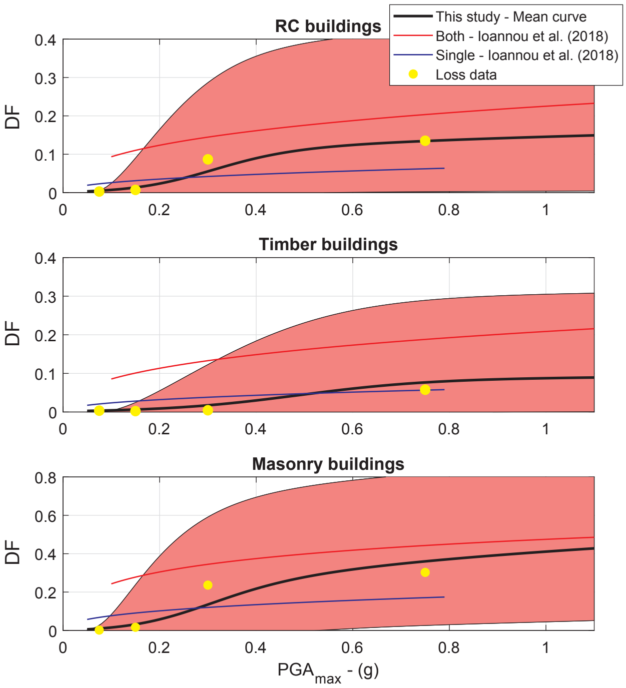

Fig. 10. Comparison of proposed vulnerability model in this study to the ‘Single’ and ‘Both’ models from Ioannou et al. (2018) [13] based on same data. The dots show the mean DF within each acceleration bin in Fig. 5.

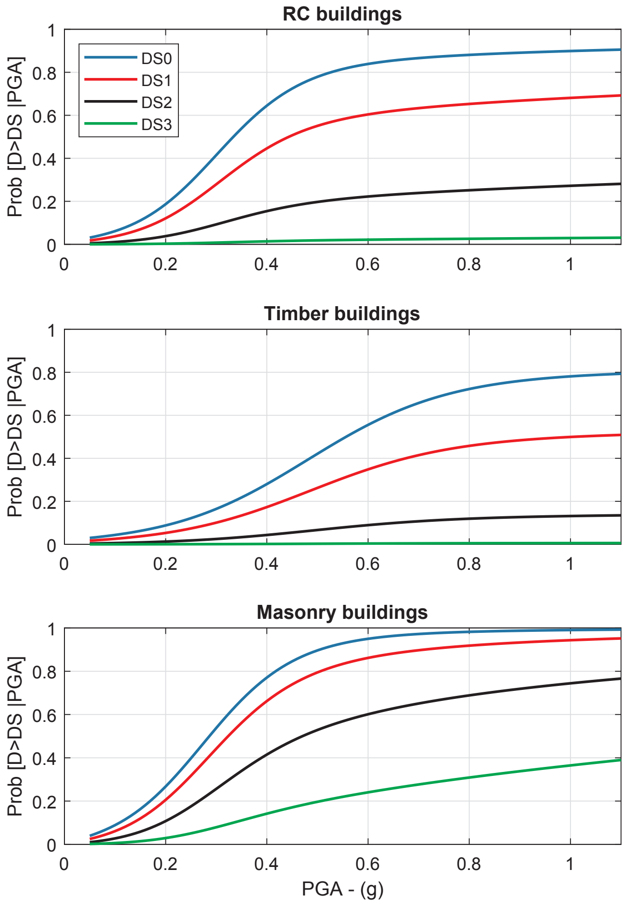

Fig. 11. Fragility curves based on the statistical model for RC, timber and masonry buildings. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Declaration of Competing Interest

Author declares that there is no conflict of interest

Acknowledgements

The authors thank the Natural Catastrophe Insurance of Iceland for placing the earthquake damage database and other relevant information at their disposal, and the University of Iceland Research Fund for financial support (Grant no. RSJ-2017).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.engstruct.2019.109969.

References

[1] Einarsson P. Earthquakes and present-day tectonism in Iceland. Tectonophysics 1991;189(1):261–79.

[2] Einarsson P. Plate boundaries, rifts and transforms in Iceland. Jökull 2008;58:35–58.

[3] Halldórsson P, Björnsson S, Brandsdóttir B, Sólnes J, Stefánsson R, Bessason B. Earthquakes in Iceland. In: Sólnes J, Sigmundsson F, Bessason B, editors. Natural Hazard in Iceland, Volcanic Eruptions and Earthquakes. University of Iceland Press and Natural Catastrophe Insurance of Iceland (in Icelandic); 2013.

[4] Thórarinsson Ó, Bessason B, Snæbjörnsson J, Ólafsson S, Sigbjörnsson R, Baldvinsson G. The South Iceland earthquakes 2000: Strong motion measurements. Proceedings of the 12th European Conference on Earthquake Engineering, Paper no. 321, Paris, France. 2002.

[5] Icelandic Property Registers (2018). https:/www.skra.is/english/individuals/ Accessed 5/10, 2018.

[6] Natural Catastrophe Insurance of Iceland (2018). https://nti.is/en/ Accessed 5/10, 2018.

[7] Pitilakis K, Crowley H, Kaynia AM. Introduction. In: Pitilakis K, Crowley H, Kaynia AM, editors. SYNER-G: Typology definition and fragility curves for physical elements at seismic risk. Springer. 2014.

[8] Bessason B, Bjarnason JÖ, Gudmundsson A, Sólnes J, Steedman S. Probabilistic earthquake damage curves for low-rise buildings based on field data. Earthq Spectra 2012;28(4):1353–78.

[9] Bessason B, Bjarnason JÖ, Guðmundsson A, Sólnes J, Steedman S. Analysis of damage data of low-rise buildings subjected to a shallow Mw6.3 earthquake. Soil Dyn Earthquake Eng 2014;66:89–101.

[10] Bessason B, Bjarnason JÖ. Seismic vulnerability of low-rise residential buildings based on damage data from three earthquakes (Mw6.5, 6.5 and 6.3). Eng Struct 2016;111:64–79.

[11] Rossetto T, Ioannou I, Grant D. Existing empirical fragility and vulnerability relationships: compendium and guide for selection. Pavia: GEM Foundation; 2013.

[12] Rossetto T, Ioannou I, Grant D, Maqsood T. Guidelines for empirical vulnerability assessment. Pavia: GEM Foundation; 2014.

[13] Ioannou I, Bessason B, Kosmidis I, Bjarnason JÖ, Rossetto T. Empirical seismic vulnerability assessment of Icelandic buildings affected by the 2000 sequence of earthquakes. Bull Earthq Eng 2018;16(12):5875–903.

[14] Ferrari SLP, Cribari-Neto F. Beta regression for modelling rates and proportions. J Appl Statist 2004;31(7):799–815.

[15] Maqsood T, Edwards M, Ioannou I, Kosmidis I, Rossetto T, Corby N. Seismic vulnerability functions for Australian buildings by using GEM empirical vulnerability assessment guidelines. Nat Hazards 2015;80:1625–50. https://doi.org/10.1007/s1106 9-015-2042-x.

[16] Ospina R, Ferrari SLP. A general class of zero-or-one inflated beta regression models. Comput Stat Data Anal 2012;56:1609–23.

[17] Ambraseys N, Smit P, Sigbjörnsson R, Suhadolc P, Margaris B. Internet-Site for European Strong-Motion Data (ISESD). European Commission, Research-Directorate General, Environment and Climate Programme. 2002.

[18] Sigbjörnsson R, Ólafsson S. On the South Iceland earthquakes in June 2000: Strong-motion effects and damage. *Bollettino* di *Geofisica teorica ed applicate* 2004;45(3):131–52.

[19] Sigbjörnsson R, Ólafsson S, Snæbjörnsson JT. Macroseismic effects related to strong ground motion: a study of the South Iceland earthquakes in June 2000. Bull Earthq Eng 2007;5(4):591–608.

[20] Vogfjörd K, Sigbjörnsson R, Snæbjörnsson JTh, Halldórsson B, Sólnes J, Stefánsson R. The South Iceland earthquakes 2000 and 2008. In: Sólnes J, Sigmundsson F, Bessason B, editors. Natural Hazard in Iceland, Volcanic Eruptions and Earthquakes, University of Iceland Press and Natural Catastrophe Insurance of Iceland (in Icelandic). 2013.

[21] D'Ayala DA, Meslem A, Vamvatsikos D, Porter K, Rossetto T, Silva V. Guidelines for Analytical Vulnerability Assessment of Low/Mid-Rise Buildings. Pavia: GEM Foundation; 2014.

[22] Park YJ, Ang AH-S. Mechanistic seismic damage model for reinforced concrete. J Struct Eng, ASCE 1984;111(ST4):722–39.

[23] Rupakhety R, Sigbjörnsson R. Ground-motion prediction equations (GMPEs) for inelastic response and structural behaviour factors. Bull Earthq Eng 2009;7(3):637–59.

[24] Jóhannesson H, Jakobsson SP, Sæmundsson K. Geological map of Iceland, sheet 6, S-Iceland. Reykjavik, Iceland: Icelandic Museum of Natural history and Iceland Geodetic Survey; 1982.

[25] Sigbjörnsson R, Snæbjörnsson JT, Higgins SM, Halldórsson B, Ólafsson S. A note on the Mw6.3 earthquake in Iceland on 29 May 2008 at 15: 45 UTC. Bull Earthq Eng 2009;7(1):113–26.

[26] Ólafsson S. Attenuation of earthquake waves. In: Sólnes J, Sigmundsson F, Bessason B, editors. Natural Hazard in Iceland, Volcanic Eruptions and Earthquakes. University of Iceland Press and Natural Catastrophe Insurance of Iceland (in Icelandic); 2013.

[27] CEN (2005). EN 1991-1, Eurocode 1: Actions on structures - Part 1-4: General actions - Wind actions. European Committee for standardization (CEN).

[28] Icelandic Standards. Icelandic National Annexes to Eurocodes, Staðlaráð Íslands; 2010.

[29] R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/; 2018.

[30] Espinheira PL, Ferrari SLP, Cribari-Neto F. On beta regression residuals. J Appl Statist 2008;35(4):407–19.