16th World Conference on Earthquake, 16WCEE 2017

Santiago Chile, January 9th to 13th 2017

Paper N° 2902

Registration Code: S-K1463177514

EMPIRICAL FRAGILITY CURVES OF RC BUIDLINGS IN CHILE USING A CUMULATIVE LINK MODEL

P. Favier(1,*), F. Quintana(2), C. Magna-Verdugo(3,*), H. Santa María(4,*), M. Hube(5,*), J.C. de la Llera(6,*)

(1) *Postdoctoral researcher, Pontificia Universidad Católica de Chile, philomene.favier@cigiden.cl*
(2) *Professor, Department of Statistics, Faculty of mathematics, Pontificia Universidad Católica de Chile*
(3) *Assistant professor, Faculty of Engineering, Universidad Diego Portales, carolina.magna@udp.cl*
(4) *Associate professor, Department of structural and geotechnical engineering, Faculty of engineering, Pontificia Universidad Católica de Chile*
(5) *Associate Professor, Department of structural and geotechnical engineering, Faculty of engineering, Pontificia Universidad Católica de Chile*
(6) *Professor, Faculty of engineering, Pontificia Universidad Católica de Chile, jcllera@ing.puc.cl*
(*) *Cigiden, National Research Center for Integrated Natural Disaster Management CONICYT/FONDAP/15110017*

Abstract

Chile is known as one of the most seismic countries in the world showing approximately one event above magnitude 8 every ten years. Unexpected damage occurred in reinforced concrete (RC) buildings during the 2010 M_w 8.8 Maule earthquake. Calculating the risk for RC buildings to exceed a given damage state during potential future seismic events is thus of paramount importance. It is assumed that the risk can be calculated from a combination of an exposure model, an intensity hazard distribution, and the vulnerability definition of RC buildings. Construction of fragility curves is part of the process for vulnerability definition. We propose here in an innovative methodology to build empirical fragility curves with the collected data of Chilean RC buildings. After the 2010 Maule earthquake, a database of the damage states of Chilean RC buildings was built by the assessment of structural engineers from companies or municipalities. Damage was classified in five states: no damage, slight, moderate, severe and complete damage states. The database is composed of observations made in six Chilean cities exposed to different seismic intensity. An interpolation of a 2010 shakemap provides the intensity measure occurring at the location of each building. Different generalized linear models were tested and it was shown that the cumulative logit link model provides the best fit of fragility curves. Cumulative link models for ordinal data ensure that fragility curves do not cross. The number of buildings floors and the year of construction are considered as covariates of our model. A Bayesian approach is adopted to obtain estimates and distribution of the parameters describing the statistical model. The resulting fragility curves for RC buildings are compared with results available in the literature. The provided fragility curves for Chilean RC buildings is the major contribution of this study. Finally, the usability of the curves is investigated from a risk perspective, and the need of more information to improve the fitting procedure is discussed.

Keywords: seismic fragility curves; empirical data; Bayesian statistic; RC buildings; earthquake vulnerability.

1. Introduction

In recent years, fragility curves derivation is gaining popularity due to their crucial roles for assessing risk and vulnerability in natural hazards such as earthquakes. For instance, the vulnerability is deduced from the fragility [1] and can be expressed in terms of the cost of the direct damage, casualties or downtime. Fragility curves provide the probability of exceeding a given physical damage knowing the seismic intensity measure, e.g. structural fragility is expressed as the probability of exceeding a given structural damage. Fragility curves are deduced in a probabilistic framework. They are defined as increasing curves with values in [0, 1], following similar characteristics as a cumulative density function. The most commonly used seismic intensity measures are: intensity scales (e.g. the Modified Mercalli Intensity, the Medvedev-Sponheuer-Karnik intensity), peak ground characteristics (e.g. Peak Ground Acceleration (PGA), Peak Ground Velocity, Peak Ground Displacement) or spectral description at a given period (e.g., spectral acceleration, spectral velocity, spectral displacement). The fundamental period of the studied structure is usually the period used to determine the seismic spectral intensity measure of interest.

Fragility curves can be classified into four groups based on how they are derived: empirical, judgmental, analytical, and hybrid [2]. Empirical assessment of fragility curves is developed from post-disaster observations. On one hand, empirical assessment is very sensitive to the quality and quantity of the available data. On the other hand, it is often considered as the most realistic assessment of fragility, i.e. the data used for assessing the empirical fragility curves incorporates all the real uncertainty/variability, such as the variability of the structural response of the studied stock, the variability of the soil-structure interaction and site effects, the uncertainty on the path and source characteristics of the earthquake [2]. Such variability is difficult to take into account in analytical or judgmental fragility curves, e.g. the Hazus project suggests to consider the uncertainty in the damage state threshold, variability in the response properties of the model building type, and the uncertainty in response due to the spatial variability of ground motion demand [3]. Empirical approaches can be based on direct discrete damage probability matrix (DPM) assessment (i.e., providing the ratio of the number of damaged to the total number of buildings, usually of the same structural type), statistical fitting with well-chosen statistical distributions (e.g. log-normal distribution, normal distribution) or statistical regression [4].

Most of statistical procedures for estimating fragility functions are developed with frequentist approaches [5]. Recent researches focused on developing fragility curves in a Bayesian statistical framework. For instance, by combining multiple experts' estimates as priors and field-based observations as likelihood, fragility curves for global building types were already built in a Bayesian setting [6]. In other study, the data on building damages are used to estimate the likelihood functions and the parameters of analytical fragility curves are used to estimate the prior distributions of damage in the Bayesian analysis [7]. The clustered effort of the GEM group (Global Earthquake Model group) lead to some courses and publications on how to use the Bayesian framework for fragility curves determination aiming to better take into consideration uncertainty propagation [4]. Bayesian approaches were also developed to propose a decision model to select the good fragility curves estimation, englobing from theoretical approaches to more practical ones [8].

A database of Chilean reinforced concrete (RC) buildings is used in this research. It consists of 587 RC buildings surveyed after the 2010 Chilean earthquake, where the number of floors, the year of construction, and the seismic damages were known. The intensity measure of the earthquake at each buildings location is calculated using the shakemap provided by the United States Geological Survey (USGS) scientific agency [9]. Generalized linear models to obtain fragility curves are tested on the data. Frequentist and Bayesian frameworks are set and used for obtaining the parameters of the fragility curves. The determination of the most suitable model is based on the estimation of three statistical indexes. A comparison with existing fragility curves of the literature helped identifying the consistency and weaknesses of the fragility curves. Finally, the conclusion highlights the main scientific and operational findings and puts forward several perspectives for future research and improvement of the proposed fragility functions.

2. Data

2.1 Damage database

Two sets of data describing RC building damage from the 2010 Chilean earthquake were used for this study. The data sets were constructed by Chilean structural engineers. The largest set was obtained from the Municipality of Viña del Mar and was done by randomly inspecting RC buildings in the city. The shorter set was done upon request by a Chilean structural engineering company. The request was done by buildings owners whatever the type of damages observed, e.g. no damage, few damages, risk of collapse. The latter set is composed of data from the cities of Constitución, Concepción, Talca, Curicó, Santiago, and Viña del Mar in Chile. Overlapping of data have been carefully avoided. For each building, the address and thus the location in the longitude and latitude coordinates were known. Both databases consist mostly in residential buildings, where shear-wall RC buildings is the common lateral resistant system used for residential buildings type in Chile [10, 11]. The number of floor and the year of construction for each building were also collected in both databases.

The database of buildings was categorized according to the number of floors (4 categories) and the year of construction (3 categories). The sorting according to the number of floors was done following the division: 1-3, 3-9, 10-24 and 25+ stories buildings previously proposed [12, 13]. The year of construction is related to the design code provisions ruling at their time of construction. Three time periods are defined based on the evolution of the official Chilean seismic design codes for buildings: before 1972, between 1973 and 1996, and since 1997. The 1972 date corresponds to the publication of the first official Chilean code NCh433Of 72. The 1996 date corresponds to the Chilean seismic code which followed the 1985 Center Chile earthquake (NCh433Of.96 official code [14]). It is important to notice that the new version of this code (NCh433Of.96mod2009) is not used in this study, because its provisions were mandatory for a short period of time before the 2010 earthquake stroke. A time line from the year 1960 summarizing the main seismic events and relevant induced changes in the Chilean seismic design code for RC buildings is presented in Table 1.

From the set of data of Viña del Mar municipality, five damage categories are defined: (0) No Damage, (1) Very Light, (2) Light, (3) Moderate and (4) Severe. From the other set of data, four damage categories are defined: (0) Green - Habitable, (1) Yellow - Restricted use, (2) Orange - No habitable, (3) Red - Danger of collapse. The (0) Green - Habitable level is allocated to buildings having either no damage or low qualified damages. The (1) Yellow - Restricted use, (2) Orange - No habitable and (3) Red - Danger of collapse levels are respectively allocated to buildings having moderate, strong and severe qualified damages. To treat the data on the same scale, i.e. a four-damage state scale, assumptions had to be made. In this work, it was assumed that the damage categories (0) No Damage and (1) Very Light of the first set were equivalent to the category (0) Green - Habitable level of the second set.

Consulting the engineers in charge of collecting the data, it can be assumed that the damage states of the final database follows the FEMA [15] center definitions of the four damage state for concrete shear walls:

- Slight Structural Damage: Diagonal hairline cracks on most concrete shear wall surfaces; minor concrete spalling at few locations.
- Moderate Structural Damage: Most shear wall surfaces show diagonal cracks; some shear walls exceeded their yield capacity by presenting larger diagonal cracks and concrete spalling at wall ends.
- Extensive Structural Damage: The majority of concrete shear walls have exceeded their yield capacities; some walls reached their ultimate capacities which its exhibit by large, through-the-wall diagonal cracks, extensive spalling around the cracks and visibly buckled wall reinforcement or rotation of thin walls with inadequate foundations. Partial collapse may occur due to failure of non-ductile columns not designed to resist lateral loads.
- Complete Structural Damage: Structure has collapsed or is in imminent danger of collapse due to failure of most of the shear walls and failure of some critical beams or columns.

Table 1: The history of the adoption of new seismic regulations in Chile [16]. The grey cells show the dates corresponding to the two main Chilean seismic codes for buildings.

Earthquake	**Year**	**Chilean seismic regulations**
The Valdivia Earthquake M_w=9.5	1960	
	1968	433p69 provisory code
	1972	NCh433Of 72 official code: first official Chilean code providing a design spectrum
The Center Chile Earthquake M_w=8.0	1985	
	1993	NCh433Of 93 official code
The Antofagasta Earthquake M_w=8.0	1995	
	1996	NCh433Of 96 official code: adding of a soil type classification, incorporating modification factor for static and dynamic answer, new design spectrum
	2003	NCh2745Of 2003 official code: Analysis and design of buildings with seismic isolation, and NCh2369Of 2003: Earthquake resistant design of industrial structures and facilities.
	2009	NCh433Of 96 official code with 2009 modifications
The Maule Earthquake M_w=8.8	2010	
	2011	DS 117, DS 118, DS 60, DS 61 supreme decrees
The Iquique Earthquake M_w=8.2	2014	

2.2 Earthquake intensity

The peak ground acceleration is the earthquake intensity measure used for this study. For the 2010 Chilean Maule earthquake, the USGS provides spectral acceleration maps at 0.3s, 1.0s and 3.0s, and a PGA shakemap [9]. The PGA shakemap and the locations of the buildings are shown in Fig. 1. A nearest neighbor interpolation was done on the USGS shakemap to calculate the PGA for each building location. A study [17] provided the 31 strong motion recording in the region through an instrumented network of accelerometers, but due to the low number of accelerometers, the USGS shakemap was preferred to estimate the PGA at each building site.

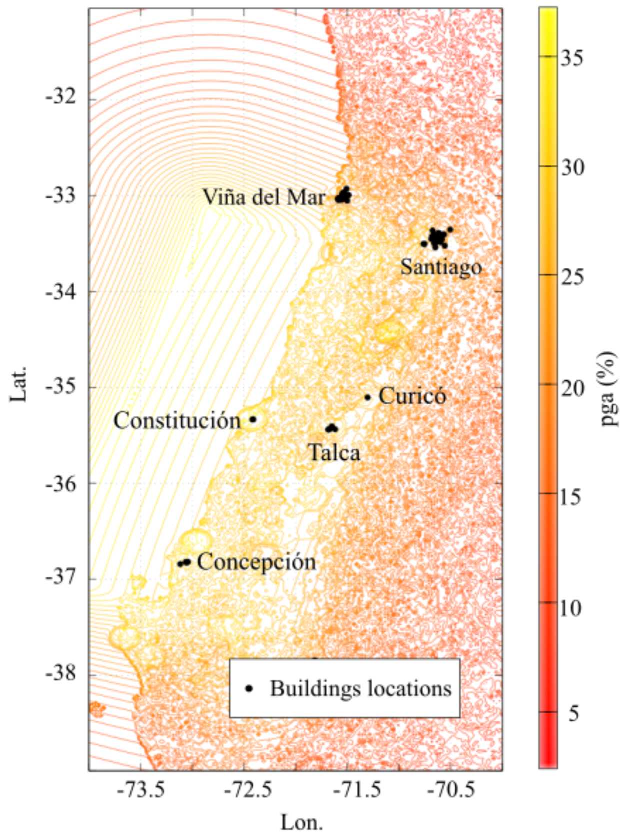

Fig. 1 - Location of the RC buildings with the shakemap from the USGS.

3. Statistical procedures

3.1 Statistical distributions

For each of the 587 buildings, the number of stories, the year of construction and the PGA seismic intensity measure approximation at the location of the building are known. The probability for a building to be in a damage state superior or equal to the damage state i under the intensity measure im is $P(DS \geq ds_i|IM = im)$ where DS is the random variable representing the damage and ds_i is the i^{th} damage state. Standard statistical procedures for developing earthquake damage fragility curves [5] were first tested. Lognormal fitting and classical generalized linear model lead to fragility curves that crossed, i.e. for a range of seismic intensity measures of values im, we observed $P(DS \geq ds_{i+1}|IM = im) > P(DS \geq ds_i|IM = im)$ for $i \in [0,3]$ violating the ordinal properties of the damage states. In other words, the fragility curve associated to the damage state i has to stochastically dominate at the order one the fragility curve associated to the damage state $i + 1$.

To overcome this issue, we focused the study on generalized linear models for ordinal data. As an overview and a reminder, the generalized linear model can be defined specifying three elements [18]: the probability γ (1st element) we are interested in predicting, using covariates $x_1, x_2, \ldots x_p$ through a linear predictor (2nd element) $\eta = \sum_{k=1}^{p} x_k\beta_k$, such as $\eta = g(\gamma)$, where g is the so called link function (3rd element). In our case, the probability $\gamma_i = P(DS \geq ds_i|x)$, where x is the vector of covariates and the vector is defined from combinations of the covariates such as the seismic intensity measure, the number of floors in the building and the year of construction. The linear predictor is expressed as:

$$\eta = \alpha_i + \beta_1 x_1 + \sum_{l=1}^{3} \beta_{2l} 1_{x_2=l} + \sum_{k=1}^{4} \beta_{3k} 1_{x_3=k} \ , \tag{1}$$

where α_i is the intercept term which depends on the category of damage state i. The effects of the other explanatory variables, i.e. the seismic intensity measures, the number of floors of the buildings and the year of construction, are the same for all categories. The coefficient β_1 is the coefficient of the seismic intensity measure x_1. The coefficient β_{2l} is the coefficient of the year of construction and $1_{x_{2=l}}$ is the indicator function of the year of construction category. The $1_{x_{2=l}}$ indicator is equal to one when the building belong to the category l, which can be one over the three year of construction categories previously set. The coefficient β_{3k} is the coefficient of the number of floors and $1_{x_{3=k}}$ is the indicator function of the number of stories category. To avoid identifiability issues, we impose $\beta_{21} = \beta_{31} = 0$. Three link functions were implemented, the logit function, the probit function, and the complementary log-log function. The logit function is:

$$g_1(\gamma) = \log(\gamma/(1-\gamma)), \tag{2}$$

the probit or inverse normal function is

$$g_2(\gamma) = \phi^{-1}(\gamma), \tag{3}$$

where ϕ is the Normal cumulative distribution function, and the complementary log-log function is

$$g_3(\gamma) = \log(-\log(1-\gamma)). \tag{4}$$

Table 2 shows the models which were tested considering the three link functions g_1, g_2, g_3 and a linear combination of the three above covariates. In the frequentist approach, the estimators from the maximum likelihood estimation method were calculated.

Table 2: Combination of models depending on the link function and the covariates.

covariate / link function	IM, year and floor	IM and floor	IM	IM and year
complementary log-log	M1	M2	M3	M4
logit	M5	M6	M7	M8
probit	M9	M10	M11	M12

3.2 Bayesian framework

For ordered categories data, the theoretical framework for Bayesian analysis for multinomial responses has already been presented [19]. In some conditions and particularly in the perspective of a Bayesian analysis, a continuous latent variable z is preferred to represent the damage state. Cut points C_i reproduce the ordinal response such as:

$$p(Z \geq C_i | x) = p(DS \geq ds_i | x), \tag{5}$$

where the illustration is depicted in Fig. 2. As $\gamma_i = P(DS \geq ds_i|x)$, one can write $g_j(\gamma_i) = \beta^T x - C_i$ for $j = 1,2,3$ and for instance the cumulative logit model is:

$$g_1(\gamma_i) = \log\left(\frac{\gamma_i}{1-\gamma_i}\right) = \beta^T x - C_i \, . \tag{6}$$

For Bayesian methods, we can use uniform prior distributions [20] to ensure that γ_i and so C_i are ordered:

$$\begin{aligned} &C_1 = 1, \\ &C_i \sim U[C_{i-1}, C_{i+1}], \quad j = 2,\ldots,I-2, \\ &C_{I-1} \sim U[C_{I-2}, 10[. \end{aligned} \tag{7}$$

where I is the number of categories of the model. The numerical calculation was done by adapting the numerical code of Dobson et al, 2008 [20] examining the use of the continuous latent variables for ordinal regression. The WinBUGS [21] software was used. It is a free software package to run Bayesian analysis to obtain the Monte Carlo Markov Chain (MCMC) of the parameters.

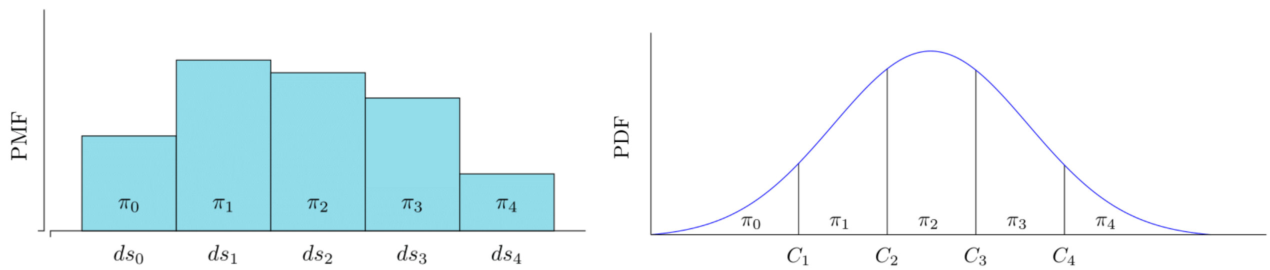

Fig. 2 - Probability mass function of the damage state for a discrete damage states and the probability density function with a latent variable, where $\gamma_i = \sum_{k=i}^{4} \pi_k$.

4. Validation

4.1 Cross-validation

The cross-validation is an internal validation procedure. It consists of splitting the database into two sets. One set is called the training data set D^{train} and is used to fit the model. The second set is called the testing data set D^{test} and is used only to test predictions arising from the model [22]. We used 90% of the data for training and the remaining 10% for testing. The data were randomly divided into 50 groups of 10%-90% combination sets. Models with link functions among logit, probit and complementary log-log adopting one to three covariates (Table 2) were compared using scoring rules such as the Deviance Information Criterion (DIC), the Log Pseudo Marginal Likelihood (LPML) and the test log-likelihood: the DIC and the LPML assess the quality of the fitting using the training data set D^{train} while the test log-likelihood assessed the quality of the probabilistic forecast using the testing data set D^{test}.

4.2 Deviance-based criteria

Deviance-based criteria are often estimated for testing statistical hypothesis and for comparing models. For a given model, one defines the Deviance Information Criterion, i.e. the DIC [23], as:

$$DIC = p_D + \overline{D}, \tag{8}$$

where $\overline{D}$ is the average deviance over all values of θ, the deviance is $D(\theta) = -2\log(p(y|\theta))$, that is to say minus twice the log-likelihood, and $p_d = \overline{D} - D(\bar{\theta})$, where $D(\bar{\theta}) = -2\log\left(p(y|\bar{\theta})\right)$. The most suitable model is the model with the smallest DIC. In the frequentist framework, the Akaike Information Criterion, i.e. the AIC [24], is calculated. The AIC is defined as:

$$AIC = D(\theta) + 2k, \tag{9}$$

where k is the number of parameters of the model. The AIC is a goodness of fit statistic based on the log-likelihood function with adjustment for the number of parameters estimated.

4.3 Log Pseudo Marginal Likelihood (LPML)

The Log Pseudo Marginal Likelihood is calculated as the sum of the log-conditional predictive ordinate (CPO) over the *n* samples of the training data:

$$LPML = \sum_{i=1}^{n} \log\left(CPO_i\right). \tag{10}$$

Chen et al. (2002) [25] proposed an estimator of the CPO_i as:

$$C\hat{P}O_i = \left(\frac{1}{L}\sum_{l=1}^{L}\frac{1}{f\left(y_i \mid \theta_l, x_i\right)}\right)^{-1}, \tag{11}$$

where L is the length of the MCMC, and $\theta_1,...,\theta_L$ are the corresponding parameter samples. The best model has the highest *LPML*. In the frequentist approach, the log-likelihood evaluated at the maximum likelihood estimation (MLE) was calculated instead of the *LPML*.

4.4 Test log-likelihood

The test log-likelihood [26] is calculated as the log-likelihood evaluated on the testing data set D^{test} with N_{test} points and where the model k is fit to the training data D^{train} obtaining parameters $\theta^k(D^{train})$:

$$l_k^{test} = \log p\left(D^{test} \mid \theta^k\left(D^{train}\right)\right). \tag{12}$$

In the Bayesian framework, the log-likelihood is calculated as the mean on the MCMC, i.e. $p\left(D^{test}|\theta\left(D^{train}\right)\right) = \int p\left(D^{test}|\theta^k\right)p\left(\theta^k|D^{train}\right)$. Let us write $i_k = -l_k^{test}/N_{test}$. The final score is calculated as the mean of i_k over the 50 data set combinations:

$$E\left[i_k\right] = E\left[\log p\left(D^{test} \mid \theta^k\left(D^{train}\right)\right)\right]. \tag{13}$$

5. Results

5.1 Model selection

The AIC, the log-likelihood, the test log-likelihood, the DIC and the LPML were calculated for each of the twelve models tested. Table 3 shows these values. In Table 3, the minimum values of the AIC and the DIC are shown in grey cells. The maximum values of the log-likelihood, the test log-likelihood and the LPMLs are also shown in grey cells. In the frequentist framework, model M5 (Table 2) obtains the best values of LL, which means that M5 is better fitting the training data than the other models. Model M6 shows the lowest value of the test log-likelihood, meaning that it is the best model for forecasting the testing data. Model M6 shows the lowest AIC meaning that M6 provides a good compromise between fitting of the training data and the number of parameters needed for fitting the model.

In the Bayesian framework, model M6 obtains the best values of DIC and test log-likelihood. Model M5 obtains the best value of LPML. It has to be noticed that the difference between the DIC and test log-likelihoods of models M5 and M6 is very small. The difference between the DIC of M5 and the DIC of M6 represents 0.13% of the DIC of M6. The difference between the LPML of M5 and the LPML of M6 represents 0.05% of the LPML of M5.

The absolute and relative comparisons of the indicators calculated in the frequentist and Bayesian frameworks show that the best model is M6, which is including two covariates, i.e. the seismic intensity measure and the number of floors. It has to be noticed that model M5, and even models M9 and M10, have values very similar to the selected model. Further investigations could focus on a new index to better discriminate the model choice (e.g. test the relevance of the logit model versus the probit model).

Table 3: Values in the frequentist and the Bayesian frameworks of the Akaike information criterion (AIC), the log-likelihood evaluated at the MLE with the training data (LL), the test log-likelihood (TLL), the deviance information criterion (DIC) and the log-pseudo maximum likelihood (LPML) for the twelve models. Values in grey cells are the best indices.

	Frequentist			**Bayesian**		
	AIC	**LL**	**TLL**	**DIC**	**LPML**	**TLL**
M1	1188.23	-583.94	-65.21	1187.53	-588.96	-65.19
M2	1185.00	-584.40	-65.00	1184.82	-588.54	-65.02
M3	1219.29	-604.75	-67.16	1219.33	-607.30	-67.18
M4	1219.33	-602.80	-67.14	1219.30	-606.32	-67.12
M5	1177.47	-579.06	-64.66	1178.12	-584.13	-64.59
M6	1176.07	-580.31	-64.61	1176.62	-584.40	-64.55
M7	1210.62	-600.45	-66.80	1210.78	-602.97	-66.79
M8	1209.49	-597.93	-66.64	1209.74	-601.45	-66.61
M9	1179.97	-580.25	-64.80	1180.21	-585.22	-64.73
M10	1177.06	-580.75	-64.65	1177.17	-584.71	-64.58
M11	1214.00	-602.12	-66.98	1214.04	-604.61	-66.93
M12	1213.15	-599.75	-66.83	1213.26	-603.23	-66.80

5.2 Fragility curves for Chilean RC buildings

The fragility curves are derived in the frequentist analysis and Bayesian approaches, i.e. the parameters of model M6 are calculated as the maximum likelihood estimators and as the mean over the 10,000 MCMC output samples. Fig. 3 shows the fragility curves built following the logit model M6 (i.e. with the seismic intensity measure and the number of floor as covariates) in the frequentist framework (a) and the Bayesian framework (b). Our model shows that the very high-rise buildings with 25 or more floors have the greatest fragility, i.e. the probability for a very high-rise building to reach the collapse state at a given seismic intensity measure is higher than for a high-rise (10-24 floors), medium-rise (4-9 floors) or low-rise (1-3 floors) buildings. The same observation can be said of the three other damage states. On the contrary, the low-rise buildings have the lowest fragility.

Due to the lack of fragility curves developed for Chilean buildings, the comparison with existing fragility curves is done with curves of RC buildings found in the literature. The largest database of fragility curves in America is HAZUS database but fragility curves are expressed as a function of the spectral displacement. Thus, the comparison is hard to do without strong assumptions on the structural behavior of the RC buildings [28]. Pejovic et al. (2016) [28] provided seismic fragility assessment for RC high-rise buildings in Southern Euro-Mediterranean zone, that is to say 20-story, 30-story and 40-story RC high-rise buildings with core wall structural system, which is similar to typical high-rise RC Chilean buildings. The analytically derived fragility curves for the prototype RC high-rise buildings of Pejovic et al. (2016) [28] are defined on a wider range of peak ground acceleration values, i.e. the probability of overpassing the complete damage state at a PGA of 1g is lower than 0.6 whereas we found a probability higher than 0.9. It can be concluded that the curves of Fig. 3 are steeper than the curves developed by Pejovic et al. (2016) [28].

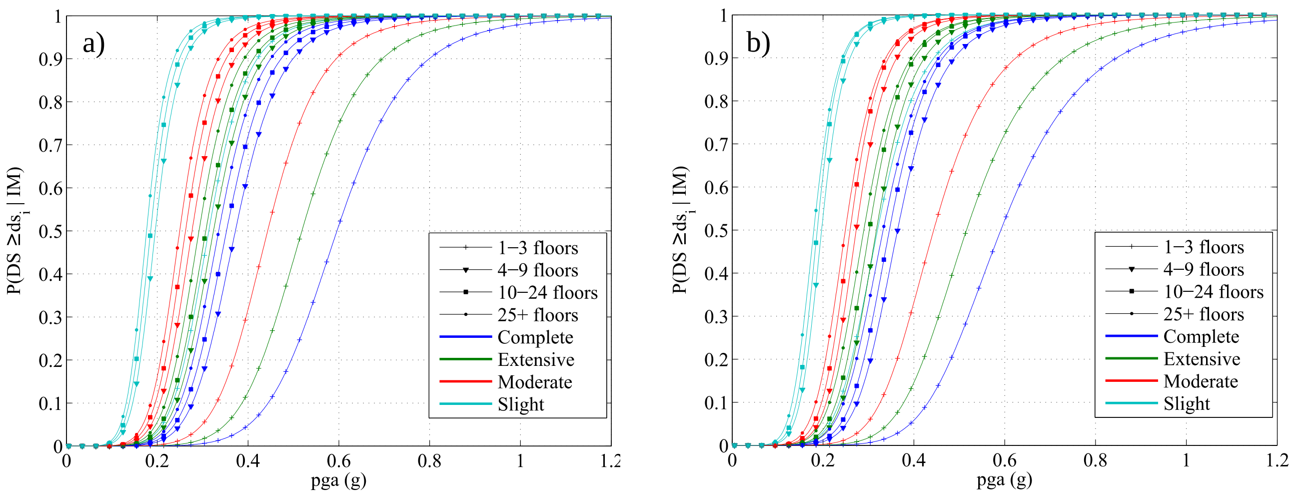

Fig. 3 - Fragility curves of Chilean RC buildings with the logit model with the intensity measure and the number of floor as covariates in the frequentist approach (a) and the Bayesian approach (b).

The study of Haindl et al. (2015) [27] focused on providing fragility curves for a two-story Chilean RC shear wall house using an analytical approach. A comparison is done to the fragility curves for RC buildings of 1-3 floors. In Haindl et al. (2015) [27], it was found that the probability of reaching the collapse state is above 0.9 for a PGA value of 3g. The fragility curves for the 1-3 floors buildings of Fig. 3 reach 0.9 of probability of collapse for a PGA value lower than 1g. The empirical curves developed here show a higher fragility than the fragility curves derived analytically, i.e. a given probability is reached for a lower value of PGA than in the analytical fragility curves of the literature. The two comparisons show that the range of the PGA on which the fragility curves of this paper are derived is lower than for fragility curves found in the literature.

To summarize, the empirical fragility curves of Fig. 3 are steeper than the ones developed analytically in the literature and are defined on a less spread range of PGA. This steepness and short range of PGA definition can be attributed to the limited number of data used to fit the curves. Indeed, the available data for fitting were acquired after one event, which reduced the range of PGA.

6. Discussion and conclusion

This paper presents a set of seismic fragility curves for Chilean RC buildings derived from empirical data. Generalized linear models were compared and different statistical indices calculated. Results show that the logit model with the seismic intensity measure and the number of floors as covariates is the best model among all 12 tested models. The comparison with fragility curves from the existing literature showed that our proposed curves are steeper and hence these fragility curves developed may overestimate the damage probabilities of Chilean RC buildings. The lack of data is the main drawback that biases the resulting fragility curves. Furthermore, more data with high PGA, say greater than 1g, would allow to better characterize the tail of the fragility curves at high PGA. However, the generalized linear model used shows good fitting and enables to include various covariates in the expression of the fragility. Indeed, the model with the number of floors and the year of construction of the buildings as covariates showed the best fit.

The proposed fragility curves are the first to be empirically developed for RC buildings in Chile. The method presented here is the first step to more advanced regression procedures. The addition of damage data from reports of insurance could improve the accuracy of these fragility curves. Second, the development of more complex models such as nonparametric ones or generalized additive models are of interest and could lead to a better fit than the one proposed here. Finally, the Bayesian framework to build fragility curves may account for the uncertainty of covariates such as the PGA value at each building location and/or incorporate expert opinions to greatly improve the quality of the proposed fragility curves.

6. Nomenclature

DS, ds_i	random variable representing the damage state, i^{th} damage state
IM, im	random variable representing the seismic intensity measure, value taken by IM
γ, γ_i	probability of overpassing a damage state with covariates vector x, $P(DS \geq ds_i \mid x)$
π_i	probability of being in a damage state with covariates vector x, $P(DS = ds_i \mid x)$
$x_1, x_2, \dots x_p$	p components of the covariates vector
η, β_k	linear predictor such as $\eta = \sum_{k=1}^{p} x_k \beta_k$, coefficients of the linear prediction
g, g_j	link function, link function of type j
Z	latent variable
C_i	i^{th} cut point
DIC	Deviance Information Criterion
AIC	Akaike Information Criterion
LPML	Log Pseudo Marginal Likelihood
CPO	Conditional Predictive Ordinate
D^{train}, D^{test}	training data and testing data from the cross-validation
l_k^{test}	log-likelihood of the model k calculated with the testing data

7. Acknowledgements

This study has been sponsored by the National Commission for Scientific and Technological Research (CONICYT) under grant Fondecyt 1141187, grant Fondecyt 3160483, and by the National Research Center for Integrated Natural Disaster Management CONICYT/FONDAP/15110017 (CIGIDEN).

8. References

[1] Gueguen P (2013): *Seismic Vulnerability of Structures*. Wiley-ISTE.

[2] Rossetto T, Elnashai A (2003): Derivation of vulnerability functions for European-type RC structures based on observational data. *Eng Struct*, **25**, 1241–1263.

[3] Federal Emergency Management Agency Hazus - MH 2.1 Technical Manual, Multi-hazard Loss Estimation Methodology, Earthquake Model, 718.

[4] Rossetto T., Ioannou I., Grant DN, Maqsood T (2014): Guidelines for empirical vulnerability assessment. Pavia, Italy.

[5] Lallemant D., Kiremidjian A., Burton H (2015): Statistical procedures for developing earthquake damage fragility curves. Earthquake Engng Struct. Dyn.

[6] Jaiswal K, Wald D, D'Ayala D (2011): Developing empirical collapse fragility functions for global building types. *Earthquake Spectra*, **27** (3), 775–795.

[7] Singhal A., Kiremidjian AS (1998): Bayesian updating of fragilities with application to RC frames. *Journal of Structural Engineering*, **124**, 922–929.

[8] Damblin G, Keller M, Pasanisi A, Barbillon P, Parent E (2014): A Bayesian decision approach for estimating a fragility curve (in French). *Journal de la Société Française de Statistique*, **155** (3), 78–103.

[9] United States Geological Survey's (USGS) Earthquake Hazards Program Shakemap us2010tfan. http://earthquake.usgs.gov/earthquakes/shakemap/global/shake/2010tfan/. Accessed 30 May 2016

[10] Lagos R, Kupfer M, Lindenberg J, Bonelli P, Saragoni R, Guendelman T, Massone L, Boroschek R, Yañez F (2012): Seismic Performance of High-rise Concrete Buildings in Chile. *International Journal of High-Rise Buildings*, **1** (3), 181–194.

[11] Massone LM, Bonelli P, Lagos R, Lüders C, Moehle J, Wallace JW (2012): Seismic Design and Construction Practices for RC Structural Wall Buildings. *Earthquake Spectra*, **28** (S1), S245–S256.

[12] Jünemann R., Hube M, De La Llera JC, Kausel E (2012): Characteristics of Reinforced Concrete Shear Wall Buildings Damaged During 2010 Chile Earthquake. *15th World Conferences on Earthquake Engineering*, Lisboa.

[13] Alarcon C., Hube M.A., Jünemann R., De La Llera JC (2015): Characteristics and displacement capacity of reinforced concrete walls in damaged buildings during 2010 Chile earthquake. Bull Earthq Eng. doi: 10.1007/s10518-015-9727-0

[14] INN (1996): NCh433.Of96: Earthquake Resistant Design of Buildings (in Spanish), 46.

[15] FEMA (2013): Hazus-MH MR5 Earthquake Model Technical Manual.

[16] Guendelman T., Lindenberg J (2011): Evolucion del diseño sismico de edificios. Procedimientos y normativa. Anales del Instituto de Ingenieros de Chile 123:

[17] Boroschek RL, Contreras V, Kwak DY, Stewart JP (2012): Strong Ground Motion Attributes of the 2010 M w 8.8 Maule, Chile, Earthquake. *Earthquake Spectra*, **28** (S1), S19–S38.

[18] McCullagh P, Nelder JA (1989): *Generalized Linear Models, Second Edition*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability.

[19] Albert JH, Chib S (1993): Bayesian Analysis of Binary and Polychotomous Response Data. *Journal of the American Statistical Association*, **88** (422), 669–679.

[20] Dobson JD, Barnett AG (2008): *An introduction to Generalized Linear Models Third Edition*. Chapman and Hall/CRC.

[21] Lunn DJ, Thomas A, Best N, Spiegelhalter D (2000): WinBUGS-a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*, **10** (4), 325–337.

[22] Steyerberg EW, Harrell FE, Borsboom GJJ., Eijkemans MJ., Vergouwe Y, Habbema JDF (2001): Internal validation of predictive models. *Journal of Clinical Epidemiology*, **54**, 774–781.

[23] Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A (2002): Bayesian Measures of Model Complexity anf Fit. *Journal of the Royal Statistical Society Series B (Statistical Methodology)*, **64**, 583–639.

[24] Akaike H (1973): Information theory and an extension of the maximum likelihood principle. *Second Int. Symp. Inf. Theory*, Akademinai Kiado.

[25] Chen M-H, Shao Q-M, Ibrahim JG (2002): *Monte Carlo Methods in Bayesian Computation*. Springer-Verlag New York.

[26] Smyth P (2000): Model selection for probabilistic clustering using cross-validated likelihood. *Statistics and Computing*, **10** (1), 63–72.

[27] Haindl M., Hube M.A., Arteta C (2015): Seismic performance assessment of a reinforced concrete shear wall house (in Spanish). *VII Congreso Nacional de Ingeniería Sísmica*, Bogotá.

[28] Pejovic J, Jankovic S (2016): Seismic fragility assessment for reinforced concrete high-rise buildings in Southern Euro-Mediterranean zone. *Bulletin of Earthquake Engineering*, **14**, 185–212.