



Seismic risk analysis with reliability methods, part I: Models

M. Mahsuli, T. Haukaas*

Dept. of Civil Engineering, Sharif University of Technology, Azadi Ave., Tehran, Iran

Dept. of Civil Engineering, University of British Columbia, 6250 Applied Science Lane, Vancouver, BC, Canada V6T 1Z4

ARTICLE INFO

Article history:

Received 12 March 2012

Received in revised form 27 December 2012

Accepted 16 January 2013

Available online 28 February 2013

Keywords:

Reliability methods

Probabilistic models

Seismic risk

Software

ABSTRACT

A library of probabilistic models for prediction of seismic risk is presented. The models are specifically intended for use with reliability methods to compute event probabilities, such as seismic loss probabilities. Several models are presented here for the first time. In particular, new and generic models are proposed for earthquake location, regional loss, building response, building damage, and building loss. Each model is presented with an explanation of its development and a discussion of its predictions. In addition, models from the literature are “smoothed” to make them amenable to reliability analysis. The models are implemented in a new computer program that is tailored for reliability and optimization analysis with many probabilistic models. The models and the computer program are employed in the companion paper to assess the seismic risk to the Vancouver metropolitan region in Canada.

© 2013 Elsevier Ltd. All rights reserved.

1. Introduction

The overarching objective in this paper is to improve the prediction of seismic risk for civil infrastructure. Risk in this context refers to loss probabilities, where the losses are due to repair of damaged structural and non-structural components. Loss probabilities are often presented in the form of a “loss curve”, which displays the probability of exceeding any loss value. Loss curves are commonly employed on a regional scale in the insurance industry [1] and for individual buildings in modern performance-based engineering [2]. However, the definition of risk in terms of loss curves contrasts the classical definition from the field of structural reliability. Risk is there defined as expected loss, *i.e.*, the product of probability and cost of a predefined failure event. In classical reliability analysis, the failure event is specified by a limit-state function, the uncertainty is characterized by random variables, and reliability methods are employed to estimate the failure probability [3]. This approach has mostly been applied to capacity-demand-type limit-state functions. However, two observations are made: (1) Reliability methods are tailored to estimate the probability of rare events, *i.e.*, small probabilities; (2) The tail of the loss curve, *i.e.*, where the probability is low but the loss is high, is particularly important for seismic mitigation decisions. These observations motivate the present effort to explore the use of reliability methods in risk analysis and loss estimation, with emphasis on accuracy in the tail of the loss curve.

This paper is part of a two-part contribution. New probabilistic models are presented here, while analysis methods, loss curves,

and other results for the Vancouver metropolitan region in Canada are presented in Part II. Although the details of the analysis are presented in Part II, it is immediately clear that the use of reliability methods influences the form of the sought models. To appreciate this fact, recall that reliability methods, such as the first- and second-order reliability methods (FORM and SORM) and sampling methods address problems with two ingredients: random variables and limit-state functions. For example, this paper employs limit-state functions defined in terms of seismic loss. The reliability analysis progresses by repeatedly evaluating the limit-state function for new realizations of the random variables. Subsequently, the probability that the limit-state function take on negative outcomes is obtained. Because no probabilities enter into the limit-state function, the models used in this type of reliability analysis must return measurable responses, not probabilities. In short, reliability methods require models that take realizations of random variables as input and return the value of one or more responses, which are not probabilities. As described in greater detail in Part II, this contrasts with contemporary approaches that employ conditional probability models. To this end, the following “rules” are identified for each probabilistic model intended for use in reliability analysis:

- Rule 1: It is an equation or algorithm
- Rule 2: It simulates possible scenarios of physical phenomena without conservative bias
- Rule 3: It discretizes all uncertainty in terms of random variables
- Rule 4: It takes as input the realization of continuous random variables
- Rule 5: Ideally, it has random model parameters whose probability distribution is updated when new data emerge; *i.e.*, it includes epistemic uncertainty

* Corresponding author. Tel.: +1 6048275557.

E-mail addresses: terje@civil.ubc.ca (T. Haukaas), mahsuli@sharif.edu (M. Mahsuli).

- Rule 6: It may take constants and decision variables as input; decision variables are not uncertain, rather, their value is at the discretion of the engineer
- Rule 7: It returns one or more physical responses, not probabilities
- Rule 8: It returns a unique response for each unique realization of the random variables
- Rule 9: It returns a response that is “continuously differentiable,” *i.e.*, smooth, with respect to the random variables; this is necessary for gradient-based reliability methods, such as FORM
- Rule 10: It permits the simulation of all possible realizations of the outcome space
- Rule 11: It is modularized in the sense that it can take input from “upstream” models and/or return output to “downstream” models.

Some models in the literature conform to these definitions and are directly amenable to reliability analysis. One example is the following ground shaking intensity model, which predicts the value of the site-specific acceleration response spectrum, S_a :

$$\ln(S_a) = h(m, R, V_{S30}, T_n) + \varepsilon, \tag{1}$$

where h = function or algorithm, m = earthquake moment magnitude, R = site-to-hypocenter distance, V_{S30} = shear-wave velocity of the top 30 m of the ground at the site, T_n = first natural period of vibration of the structure, and ε = model uncertainty. Intensity models of the format in Eq. (1) include those presented by Atkinson and Boore [4] and Boore and Atkinson [5], which are employed later in this study. Notably, these intensity models yield a unique realization of S_a for given realizations of m, R, V_{S30}, T_n , and ε . In a reliability analysis, the parameters m, R , and V_{S30} are typically random variables. Thus, this model satisfies all 11 rules, except Rule 5 if the regression parameters in the attenuation relationship are constants instead of random variables.

A contrasting example of models that are intractable in reliability analysis is the damage fragility model for a concrete shear wall high-rise building in Fig. 1 presented by FEMA–NIBS [6]. The solid thick line in Fig. 1 is the probability that the damage to the building exceeds Damage State 3 for a given value of the maximum lateral displacement, Δ . The probability that the building is in Damage State 3 is the distance between the solid and dashed thick curves. This model is here referred to as a conditional probability model because it has a probability as output. Thus, it violates Rule 7 and cannot be directly employed in a reliability analysis. However, this model is employed later to calibrate models that obey this rule. As it stands, the model in Fig. 1 limits the analysis formats to analytical integration by the total probability theorem or sampling. It is also noted that with models like the one in Fig. 1, each building configuration requires a unique model. In contrast, this paper advocates the development of probabilistic models that take an array of material, geometry, and model parameters as input, thus covering a range of configurations.

Several modeling techniques are available to develop probabilistic models suitable for reliability analysis. One approach is the development of linear models by Bayesian inference, as described in [7]. This approach was applied in [8] to reinforced concrete members in the context of seismic risk. This approach is appealing because the resulting models conform to all the abovementioned rules. In fact, model uncertainty is explicitly included by means of random model parameters in accordance with Rule 5. The extension to nonlinear models, *i.e.*, models that are nonlinear in terms of the model parameters, is described in [9].

The models presented in this paper are implemented in the computer program Rt. Rt is an object-oriented program developed by the authors for reliability and optimization analysis with many interacting probabilistic models. The software architecture to support this

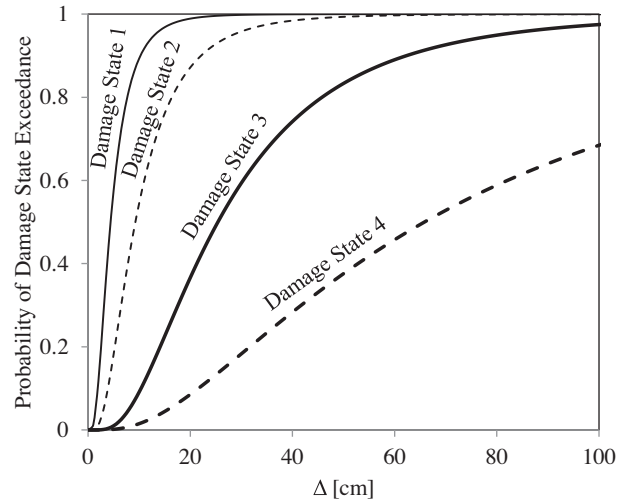


Fig. 1. Damage fragility model for a concrete shear wall high-rise building from FEMA–NIBS [6].

	Region Level	Building Level	Component Level	
Loss	Regional Loss Model	Building Loss Model		\$
Damage	Regional Damage Model	Building Damage Model		Icons of buildings
Response	Regional Damage Model	Building Response Model		Icons of buildings
Ground Motion	Intensity Model			Icon of seismic waves
Earthquake	Occurrence Model	Magnitude Model	Location Model	Icon of earthquake location

Fig. 2. Overview of models at the region, building, and component levels.

type of analysis is presented by Mahsuli and Haukaas [10], but the first comprehensive utilization of Rt is presented here. Rt and its object-oriented library of models are freely available for download at www.inrisk.ub.ca, together with examples and video tutorials.

An overview of the models addressed in this paper is provided in Fig. 2. Each of the following sections presents one of the models, or collection of models, identified in Fig. 2. The exception is the occurrence model, which is not described because the well-known Poisson point process is employed without modifications. Conversely, a model that discounts future losses to present value is not shown in Fig. 2, but it is briefly described in a section below. It is also emphasized that several other models are available in Rt. Fig. 2 emphasizes that the buildings are modeled at three different levels in Rt, *i.e.*, region, building, and component levels. A regional model covers many buildings and provides rough estimates. This contrasts with building models, which utilize characteristics such as building height and structural system to estimate response, damage, and loss. The most detailed approach is component models, *i.e.*, finite element analysis with element models for each structural and non-structural member. However, in this paper the attention is centered on regional loss estimation, thus justifying the use of simplified models to estimate the damage for individual buildings.

2. Location models

Contemporary risk analysis approaches often take a “hazard curve” as a starting point; see *e.g.*, [11]. In the context of seismic

risk, a hazard curve displays the probability of exceeding values of a site-specific ground shaking intensity, such as S_a . This study aims at circumventing the hazard curve to expose the underlying models. This is necessary for the comprehensive and continuously improving modeling of uncertainties in seismic risk analysis. Behind each hazard curve are models for the location and magnitude of earthquakes, and for the propagation of rupture energy. This section addresses the location models, which aim at predicting in probabilistic terms the earthquake location.

The literature provides probability distributions for the epicenter-to-site distance, R , for various source geometries [12]. Technically, these probability distributions can be employed directly in the reliability analysis, with the premise that the epicenter-to-site distance is input as a random variable. However, this approach has several disadvantages in the present regional analysis. To address all buildings in a region, a different random variable with different distribution would be necessary for each site. This is cumbersome and computationally inefficient, and the stipulation of correlation between the ground motions at different sites poses an additional challenge. Therefore, an alternative is proposed here, whereby location models are developed that take random variables as input and give earthquake location as output. Each realization of the random variable(s) is associated with one earthquake location and once the coordinates of this location are known, the distance to any and many building sites is readily computed. The proposed approach is also advantageous if aftershocks are modeled, because their locations are dependent on the location of the main shock. The dedicated location models presented here also make it easier to model non-uniform probability of occurrence within a source, although this is not relevant in this paper.

Models are developed here for area sources and line sources. The line source model is straightforward. It takes two random variables as input, together with the longitude and latitude of the end points, as well as the longitude and latitude of potential intermediate points for multi-segment lines. One of the random variables represents the epicenter location along the line, which is usually a uniform random variable. The other represents the earthquake depth.

The formulation of the area source model is new in this paper. It models arbitrarily shaped area sources and takes three random variables, x_1 , x_2 , and x_3 , as input. The output is the earthquake depth, d , and the longitude, L_o , and latitude, L_a , of the epicenter. The random variable x_3 directly represents the depth, while x_1 and x_2 are uniformly distributed random variables that indirectly represent L_o and L_a . The key objective of the model is to transform realizations of x_1 and x_2 into realizations of L_o and L_a so that the probability for the epicenter location is uniformly distributed within the arbitrary area.

To understand how the objective is accomplished, consider Fig. 3. The shape identified by thick solid lines is an arbitrary polygon that represents an earthquake source. The user specifies the area by defining any number of longitude–latitude pairs. This is conveniently done in Rt’s map interface. Next, it is recognized that if L_o and L_a have the probability density functions (PDFs) shown in Fig. 3 and denoted by $f(L_o)$ and $f(L_a)$, then the sought model is obtained. Specifically, if $f(L_o)$ is uniform and $f(L_a)$ is proportional to the width of the area at that latitude, then the joint PDF $f(L_o, L_a) = f(L_o)f(L_a)$ is uniform over the area. Therefore, the model transforms the uniform distributions $f(x_1)$ and $f(x_2)$ to the distributions $f(L_o)$ and $f(L_a)$, where $f(L_a)$ is automatically computed based on the width of the area source at every latitude L_a .

The width of the area source at each latitude is denoted by $W(L_a)$, as shown in Fig. 3. Several geometrical considerations are made in the computation of $W(L_a)$. First, note that L_o and L_a are measured in radians, and that L_a is zero at the Equator. Second, let the radius of the Earth be denoted by r , under the approximation that the Earth is spherical. The radius of a circle around the

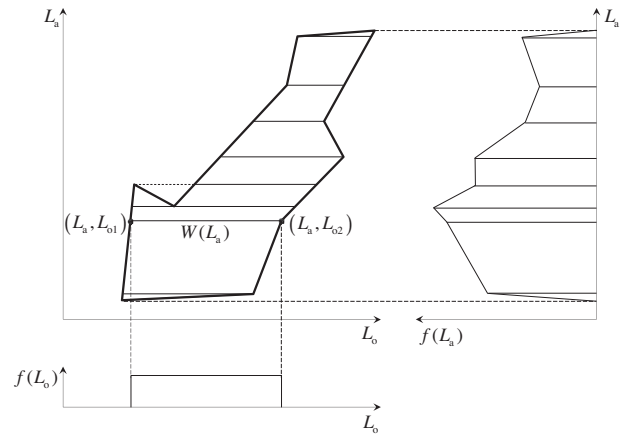


Fig. 3. Earthquake location model for an arbitrary area source.

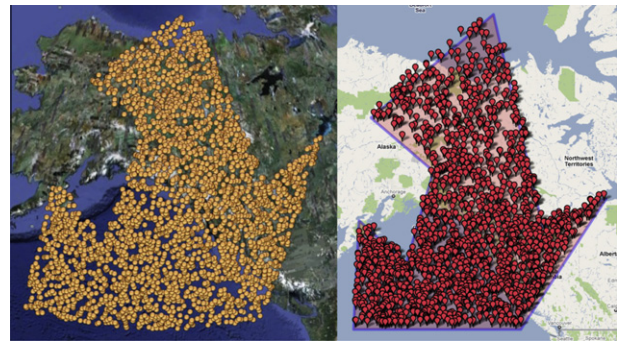


Fig. 4. Uniformly distributed epicenter realizations shown in the spherical Google Earth (left) and the flattened Google Maps (right).

Earth at a specific latitude is $r \cdot \cos(L_a)$. As a result, the length of a segment on this circle, delimited by the longitudes L_{o1} and L_{o2} , is

$$W(L_a) = |L_{o2} - L_{o1}| \cdot r \cdot \cos(L_a). \quad (2)$$

Furthermore, the “great circle effect” is included when determining longitude and latitude values on the straight lines between the user-given corner points, shown by thick solid lines in Fig. 3. This means that the slight curvature of these lines, due to the curvature of the Earth, is accounted for.

To demonstrate the effectiveness of the area source model, Fig. 4 shows 2000 locations generated by random samples of x_1 and x_2 . An arbitrary area in the northern hemisphere is selected for this demonstration because it highlights the ability of the model to produce uniformly distributed earthquake locations. It is observed in Fig. 4 that the locations correctly appear uniformly distributed in Google Earth®, which gives a correct spherical view of the Earth, while they appear non-uniform in Google Maps®, which gives an incorrect flattened view of the Earth. Fig. 4 also shows that it is possible to model concave polygons with this new probabilistic model. All the location models presented here are available online in Rt.

3. Magnitude model

The magnitude of earthquakes is commonly represented by a bounded exponential random variable [11]. The probability distribution of this random variable is based on the Gutenberg–Richter law [13] and the PDF is

$$f(m) = \frac{b' \cdot \exp[-b' \cdot (m - M_{\min})]}{1 - \exp[-b' \cdot (M_{\max} - M_{\min})]} \text{ for } M_{\min} \leq m \leq M_{\max}, \quad (3)$$

where m = moment magnitude, b' = parameter that depends on the relative occurrence of different magnitudes, M_{\min} = magnitude lower bound, and M_{\max} = magnitude upper bound. This study addresses the uncertainty in b' and M_{\max} by modeling them as lognormal random variables. Furthermore, because Eq. (3) produces a probability density, thus violating Rule 7, it is necessary to transform it into a model that yields a magnitude realization as output. For this purpose, a standard normal random variable, x , is introduced as an input variable. Its cumulative distribution function (CDF) is denoted by $\Phi(x)$ and it is employed as a surrogate measure of the earthquake magnitude. Specifically, a relationship between m and x is introduced by the well-known probability-preserving transformation [14]

$$\Phi(x) = F(m), \tag{4}$$

where F is the CDF that corresponds to the PDF in Eq. (3). As a result, magnitude realizations are produced for given realizations of x by the formula:

$$m = F^{-1}(\Phi(x)), \tag{5}$$

where F^{-1} is the inverse CDF of x . In summary, the magnitude model takes the realization of b' , M_{\max} , and x as input and produces the corresponding realization of the magnitude, m , as output. This model is also available online in Rt.

4. Intensity models

An earthquake intensity model employs characteristics of the earthquake and the path of shock wave propagation, and perhaps features of the structure, to predict characteristics of the site-specific ground shaking. A variety of models and intensity measures exist in the literature. While time-series of the ground acceleration are sought in the component-level approach in Fig. 2, the region- and building-level analysis in this paper requires models that produce a scalar intensity measure. Specifically, models are sought that take earthquake location and magnitude as input and return the elastic 5%-damped Sa for given periods T_n at specific sites. The models employed in this study are of the generic form of Eq. (1) and proposed by Atkinson and Boore [4] and Boore and Atkinson [5], with ε as a random variable in the reliability analysis. In this study, a particular effort is made to model V_{S30} . The authors were given access to a comprehensive database of V_{S30} -measurements for the Vancouver metropolitan region in Canada. As a result, V_{S30} is modeled as a lognormal random variable with probability distribution inferred from the data as described in the companion paper for the Vancouver metropolitan region.

The contribution to intensity modeling in this paper is “smoothing” to satisfy Rule 9. In particular, it is required that the model response, Sa , is continuously differentiable with respect to m , R , and V_{S30} . The models presented by Atkinson and Boore [4] and Boore and Atkinson [5] violate Rule 9 due to the presence of discrete variables that depend on V_{S30} and if-statements that introduce different model forms for different ranges of V_{S30} , m , T_n , and depth. To make the intensity models continuously differentiable, the variation of Sa with respect to its parameters is given a third-order polynomial shape in the vicinity of the parameter values where the model changes form. For brevity, detailed equations are omitted here, but Figs. 5 and 6 illustrate the effect of smoothing for the two utilized intensity models. It is arbitrarily selected to plot the model response against magnitude and shear wave velocity. As a result, the solid lines in Figs. 5 and 6 demonstrate that kinks and discontinuities in the original model are replaced by smooth transitions. This makes the models amenable to gradient-based reliability analysis, and the smoothed models are available online in Rt.

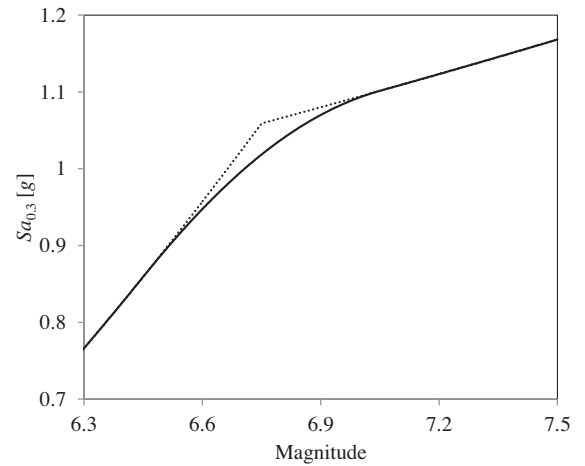


Fig. 5. Smoothed ground motion relationship by Boore and Atkinson [5]: spectral acceleration at 0.3 s plotted against the magnitude.

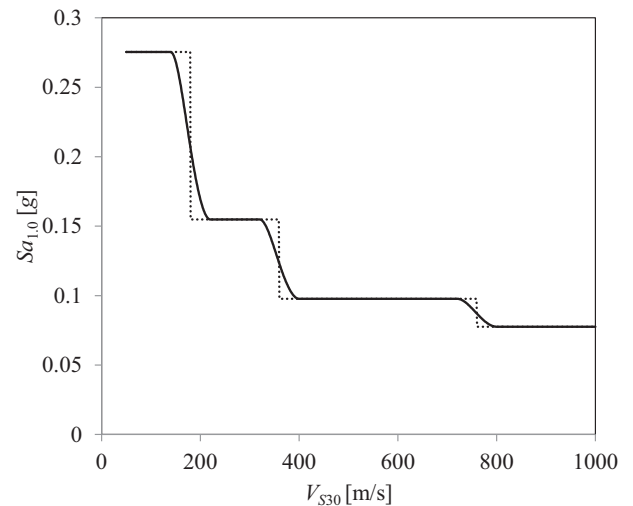


Fig. 6. Smoothed ground motion relationship by Atkinson and Boore [4]: spectral acceleration at 1.0 s plotted against the shear wave velocity.

5. Regional damage model

Three levels of modeling were identified in Fig. 2. Modeling at the regional scale, which is described in this section, requires less data and less computational resources compared to the building-by-building modeling that is described later. As the name suggests, a regional model predicts damage and loss within an entire region. Naturally, the savings in modeling and computer efforts are counteracted by high model uncertainty. The characterization of this uncertainty is the primary objective in this section.

In general, a region encompasses different land uses, here categorized as “zones.” These zones are analogous to typical city zoning and—in the implementation in Rt—each region has K zones. Damage to the building stock within a zone is measured by a damage ratio, η_k , ($k = 1, 2, \dots, K$) where $\eta_k = 0$ indicates no damage and $\eta_k = 1$ indicates complete damage. Complete damage requires replacement of the entire building stock. In this study $K = 5$ because five zone types are considered: single-residential (detached family housing); multi-residential (apartments); commercial; industrial; and comprehensive development (high-rise construction).

In the following, regression models for η_k for the different zone types are developed, with spectral acceleration at different periods

as regressors, as explained shortly. For this purpose it is necessary to have observed earthquake damage in specific regions. Ideally, data from past earthquakes are utilized, such as insurance claims data. In lieu of such information, detailed analysis of a large building inventory is conducted. The authors were given access to a database for 8330 buildings in Vancouver, for which key characteristics are known. One of these characteristics is the occupancy type, described shortly, which is employed to categorize the buildings into different zones.

To generate the data, first a range of earthquake magnitudes and distances are randomly generated. For each magnitude and distance pair, Sa at $T_n = 0.3$ s and $T_n = 1.0$ s, denoted by $Sa_{0.3}$ and $Sa_{1.0}$, at the location of each building is computed. $Sa_{0.3}$ and $Sa_{1.0}$ are not independently generated because that would miss the correlation between these two quantities. Sa at these two periods is commonly employed in the literature to construct a complete response spectrum, see, e.g., NEHRP [15] and FEMA–NIBS [6]. $Sa_{0.3}$ and $Sa_{1.0}$ for each building are input to the capacity spectrum method [16], which is implemented in Rt as a separate model. Furthermore, the peak displacement and acceleration response from the capacity spectrum method are input to the HAZUS fragility and loss functions [6] to compute the loss for each of the 8330 buildings. Thereafter, the damage ratio, η_k , for each zone is computed by dividing the total loss of the buildings in that zone by their total replacement cost. The latter is obtained for each building by multiplying the cost of replacement per unit floor area, which is obtained from HAZUS, by its floor area of construction from the database. Finally, η_k is recorded together with $Sa_{0.3}$ and $Sa_{1.0}$ at the centroid of the region. Although the correlation coefficient between $Sa_{0.3}$ and $Sa_{1.0}$ is substantial at 0.85, both $Sa_{0.3}$ and $Sa_{1.0}$ were considered as regressors in the model development.

A number of model forms for η_k as a function of $Sa_{0.3}$ and $Sa_{1.0}$ are explored. To model the transition from no damage ($\eta_k = 0$) for low values of $Sa_{0.3}$ and $Sa_{1.0}$ to complete damage ($\eta_k = 1$) for high values of $Sa_{0.3}$ and $Sa_{1.0}$, several functions are explored. Rather than the linear and the sine function, the most successful turned out to be the standard normal CDF, $\Phi(\cdot)$, which transitions smoothly from zero to unity as the argument increases. It is first attempted to let the argument of Φ be a linear function of $Sa_{0.3}$ and $Sa_{1.0}$, i.e., with the following model form for each zone type:

$$\eta = \Phi(\theta_1 + \theta_2 \cdot Sa_{0.3} + \theta_3 \cdot Sa_{1.0}) + \varepsilon, \quad (6)$$

where θ_i = model parameters and ε = model error. However, Eq. (6) turns out to provide a poor fit to the data; it tends to over-predict low damage and under-predict high damage. As a result, the following nonlinear form of the argument is introduced:

$$\ln(\eta) = \ln \left[\theta_1 \cdot \Phi \left(\theta_2 + \theta_3 \cdot (Sa_{0.3})^{\theta_5} + \theta_4 \cdot (Sa_{1.0})^{\theta_5} \right) \right] + \varepsilon, \quad (7)$$

where it is noted that the powers of $Sa_{0.3}$ and $Sa_{1.0}$ are equal, and that the logarithmic transformation is introduced to obtain a homoscedastic model, confirmed by diagnostics plots that are omitted here for brevity. It is reiterated that $Sa_{0.3}$ and $Sa_{1.0}$ are evaluated at the centroid of the region.

To determine the mean, standard deviation, and correlation for the model parameters θ_i in Eq. (7), the nonlinear regression described in [9] is employed. The regression analysis is carried out in MATLAB[®] with the Levenberg–Marquardt algorithm [17] for solving the nonlinear least squares problem. According to Seber and Wild [9], the model parameters, θ_i , are jointly t -distributed. As an approximation, in the application of Eq. (7) for regional risk analysis, the model parameters are assumed to have the normal distribution. Table 1 shows the mean and coefficient of variation that is obtained for θ_i and ε for the five zone-types. The table shows that for the single-residential, the mean of θ_3 is several times larger than the mean of θ_4 . This correctly suggests that $Sa_{0.3}$ is a more

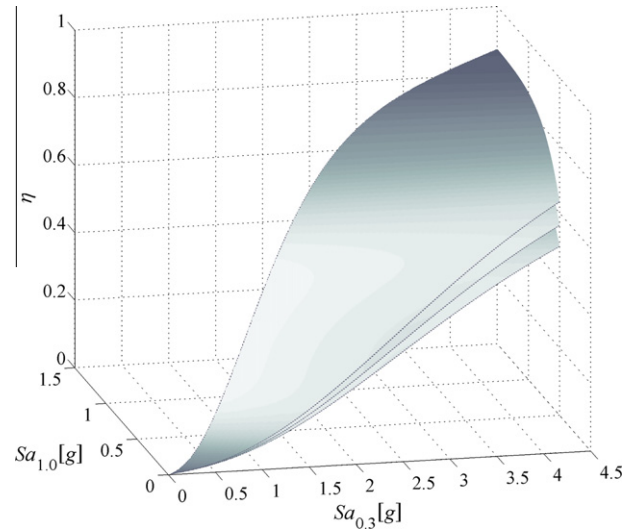


Fig. 7. Damage ratio predicted by the regional damage model versus $Sa_{0.3}$ and $Sa_{1.0}$.

important regressor than $Sa_{1.0}$ for single-residential buildings, which are mostly low-rise and thus have lower periods. Conversely, the Table 1 shows that $Sa_{1.0}$ is more influential in a comprehensive development zone, which mostly consists of high-rise buildings with high periods.

To demonstrate the predictions made by the regional damage model in Eq. (7), Fig. 7 shows the median damage plus/minus one standard deviation for a single-residential zone. The plot displays damage for $Sa_{1.0} \leq Sa_{0.3} \leq 3 \cdot Sa_{1.0}$ because points outside this range are uncommon. The desired smooth S-shaped increase in damage for increasing ground shaking intensity is observed. It is also observed in Fig. 7 that the uncertainty in the damage is significant. This is reasonable because it covers a range of possible building types within a zone. In fact, it is stressed that the uncertainty in Eq. (7) has four contributions: (1) Variability in the ground shaking intensity, which is included in $Sa_{0.3}$ and $Sa_{1.0}$; (2) Variability in the spatial distribution of ground shaking, which is included in the model parameters because each of the 8330 buildings have a different intensity according to their exact location; (3) Variability in the amount of damage to each building, which is included in the FEMA–NIBS fragility curves that are employed; and (4) Variability in the building characteristics within a zone.

6. Regional loss model

This paper addresses seismic loss due to the cost of repair of damaged buildings. Once the zone-specific damage ratio, η_k , is computed by Eq. (7), the total regional loss is

$$l = \sum_{k=1}^K (\eta_k \cdot A_k \cdot C_k), \quad (8)$$

where A_k = area of zone k and C_k = replacement cost of the buildings in zone k measured per unit area. Two interpretations are possible for A_k and C_k . One is that A_k is the floor area of construction, so that C_k is the replacement cost per unit floor area. Another interpretation is that A_k is the area of the entire zone, so that C_k is the replacement cost per unit area of land. In the analysis presented in the companion paper, the former interpretation is adopted, but in Rt the analyst is free to adopt either one. In Rt, the analyst defines any number of corner points of the region by clicking on an interactive map. Thereafter, zone percentages within the region are specified. For example, of the total area of a region, 10% may be commercial and 40%

Table 1Second moments of model parameters for regional damage models. CoV = coefficient of variation and σ_ε = standard deviation of the model error, ε .

		Single-residential	Multi-residential	Commercial	Industrial	Comprehensive development
θ_1	Mean	0.779	1.415	0.580	0.304	0.663
	CoV	0.021	0.101	0.020	0.013	0.030
θ_2	Mean	-3.981	-11.397	-6.333	-5.677	-9.461
	CoV	0.016	0.109	0.030	0.034	0.049
θ_3	Mean	3.126	3.667	2.282	3.401	1.215
	CoV	0.018	0.119	0.039	0.035	0.082
θ_4	Mean	0.698	6.641	4.126	2.916	7.919
	CoV	0.051	0.115	0.030	0.036	0.049
θ_5	Mean	0.436	0.081	0.209	0.249	0.117
	CoV	0.028	0.145	0.042	0.046	0.065
σ_ε	Mean	0.117	0.095	0.095	0.100	0.086
	CoV	0.023	0.023	0.023	0.023	0.023

residential. Because of the different possible interpretations of A_k , the zone percentages need not add up to 100%. A_k is automatically computed in Rt once the analyst has specified the zone-percentages, and the analyst inputs the replacement cost, C_k , as a parameter, which may be given as a constant or a random variable. Values for C_k for different zones are provided in the companion paper.

7. Building response models

With reference to Fig. 2, it is straightforward in Rt to replace regional models with many models for individual buildings. This refinement leads to more accurate results at the cost of more modeling and computer efforts. The characteristics of the building models are obtained by walk-down surveys, satellite imagery, and municipal databases. In this and the next sections, the following information is employed: (1) Load bearing system and construction material, which places the building in one of 13 “prototype” categories, listed in Table 2; (2) Occupancy type, which places the building in one of, coincidentally, 13 occupancy classes; (3) State of seismic retrofit and time of construction, which determines the parameter α for construction quality in accordance with Table 3; this parameter is analogues to the building code level in FEMA-NIBS [6]; (4) Number of stories, N , which determines the building height, $H = N \cdot 3$ m; (5) Footprint area, A ; (5) Plan irregularity, I_{pi} , which is unity if the building has a non-rectangular plan and zero otherwise; (6) Vertical irregularity, I_{vi} , which is unity if there is a change in the building plan along its height and zero otherwise; (7) Soft story, I_{ss} , which is unity if the building has a story with significantly less stiffness than other stories, e.g., parking in the first floor, and zero otherwise; (8) Short column, I_{sc} , which is unity if there are columns in the building with high ratio of width to height and zero otherwise; and (9) Pounding, I_p , which is unity if there is insufficient separation with adjacent buildings and zero otherwise.

In this section, a new building response model is developed for each of the 13 prototypes. The responses are peak inter-story drift ratio, δ_p , and peak acceleration response, A_p . These two responses are selected because, according to the definition that is adopted in this paper, they are directly related to the structural and non-structural damage. In fact, corresponding damage models are developed in the next section, while a later section presents loss models for the 13 occupancy classes. The motivation for developing new models instead of employing those implemented in HAZUS is threefold: (1) To explicitly describe the model uncertainty by random variables; (2) To have continuously differentiable models, which is not the case for, e.g., the response model in HAZUS due to presence of “kinks” in the demand curve; and (3) To create models that are updated in a Bayesian fashion once new observations become available.

Seismograph readings of δ_p and A_p from past earthquakes are limited. Therefore, the capacity spectrum method [16] is employed to generate data for the regression analysis. In particular, δ_p and A_p are computed and H , α , and Sa at T_n are recorded as regressors. In the initial modeling efforts, these regressors were directly employed in a variety of model forms that unfortunately were incapable of providing good predictions of δ_p and A_p . Therefore, a stronger emphasis on the mechanics of the problem was introduced. As a result, several “explanatory functions” that define the structural dynamics of the building are considered as regressors. These parameters include T_n , as well as the strength-to-weight ratio, V , yield drift ratio, δ_y , ductility capacity, μ , ultimate drift ratio, δ_u , and degradation factor, κ . Sub-models are established for these parameters, in which T_n is modeled as a function of the building height:

$$T_n = \theta_1 \cdot H^{\theta_2}. \quad (9)$$

V and μ are modeled as functions of code level and height:

$$V = \theta_1 \cdot \exp(-\theta_2 \cdot H) \cdot \frac{2 + (\alpha - 2) \cdot (\alpha - 1)}{8}, \quad (10)$$

$$\mu = \theta_1 \cdot H^{-\theta_2} \cdot \frac{10 + (\alpha - 2) \cdot (\alpha - 1)}{16}, \quad (11)$$

because the data indicate that these two parameters typically increase with code level and decrease with height. As stated earlier, α is a parameter that reflects the construction quality, with values given in Table 3. The α -dependent terms in Eqs. (10) and (11) are introduced in this paper based on judgment to achieve the desired variation of V and μ with construction quality. For instance, the α -dependent factor in the right-hand side of Eq. (10) equals 1.0 for the high code level, while it reduces V by factors of 0.5 and 0.25 for moderate and low code levels, respectively. It is noted that V represents a means by which damage can be introduced into the model, for example in aftershock predictions, i.e., in a damaged building, the strength is reduced to the residual capacity of the structure, as pointed out in [18]. Next, the yield drift ratio equals the strength, $V \cdot g$, where g is the acceleration of gravity, divided by the elastic stiffness, $4\pi^2/T_n^2$, and normalized by H :

$$\delta_y = \frac{T_n^2}{4\pi^2} \cdot \frac{V \cdot g}{H}. \quad (12)$$

The ultimate drift ratio is

$$\delta_u = \mu \cdot \delta_y. \quad (13)$$

Finally, the degradation factor is modeled as a function of the spectral acceleration demand and the building code level:

$$\kappa = \exp(-\theta_1 \cdot Sa \cdot \alpha^{-\theta_2}), \quad (14)$$

where the negative sign in the exponent appears because the structure degrades more as the demand increases. When Sa increases,

Table 2
Building prototypes.

Material	Load bearing system				
	Shear wall	Moment frame	Braced frame	Light frame	Frame/masonry wall
Reinforced concrete	Concrete shear wall	Concrete moment frame	–	–	Concrete frame with masonry infill wall
Precast concrete	Precast frame with concrete shear wall	–	–	–	–
Steel	Steel frame with concrete shear wall	Steel moment frame	Steel braced frame	Steel light frame	Steel frame with masonry infill wall
Wood	–	Wood large frame	–	Wood light frame	–
Reinforced masonry	–	–	–	–	Reinforced masonry bearing wall
Unreinforced masonry	–	–	–	–	Unreinforced masonry bearing wall

Table 3
Building code levels, i.e., construction quality.

Time of construction	Unretrofitted	Retrofitted
Before 1940	$\alpha = 1$: Pre-code	$\alpha = 3$: Moderate-code
From 1940 to 1975	$\alpha = 2$: Low-code	$\alpha = 3$: Moderate-code
After 1975	$\alpha = 3$: Moderate-code	$\alpha = 4$: High-code

then κ tends to zero, i.e., full degradation. Conversely, $\kappa = 1$ when $Sa = 0$, i.e., no degradation.

Eqs. (9)–(14) are established separately for each of the 13 prototypes. In turn, the six parameters $T_n, V, \delta_y, \mu, \delta_u$, and κ are utilized as regressors to model δ_p and A_p for all prototypes. A number of different model forms are tried. Each model is assessed by plotting the model predictions against the data and the model residuals against the regressors. In this process, some model forms exhibited inadequate predictions and some suffered from heteroscedasticity. In conclusion, the models that best predict δ_p and A_p are:

$$\ln(\delta_p) = \theta_1 + \theta_2 \cdot \ln(\delta_y) + \theta_3 \cdot \ln(\delta_u) - \theta_4 \cdot \ln(V) - \theta_5 \cdot \ln(\kappa) + \theta_6 \cdot \ln(Sa) + \theta_7 \cdot Sa + \varepsilon, \tag{15}$$

$$\ln(A_p) = \theta_1 - \theta_2 \cdot \ln(\delta_y) + \theta_3 \cdot \ln(V) - \theta_4 \cdot \ln(\mu) + \theta_5 \cdot \ln(\kappa) + \theta_6 \cdot \ln(Sa) + \varepsilon. \tag{16}$$

The total number of parameters for all models is high, which prevents the presentation of second-moment information for the model parameters θ_i here. However, this information is implemented in Rt and is available in the Ph.D. thesis of the first author [19]. It is observed that the model form in Eqs. (15) and (16) is essentially multiplicative due to the natural logarithm on the left-hand side. A model without the natural logarithm on the left-hand side suffers from heteroscedasticity and non-normality of the residuals. The latter is illustrated in Fig. 8, where the residual quantiles for such a model are plotted against normal theoretical quantiles. The points in this plot significantly deviate from the 45° line, which indicates the non-normality of errors. This is remedied by the natural logarithm on the left-hand side. The natural logarithms on the right-hand side improve the model prediction and homoscedasticity of the model.

Fig. 9 shows the median model predictions against the data for the drift model in Eq. (15). The data points are relatively close to the solid line, which is one indication that the model provides reasonable predictions. The models in Eqs. (15) and (16) originally included damping and overstrength as regressors, but these were omitted in a stepwise modeling process, similar to the one described in [8]. In conclusion, $T_n, V, \delta_y, \mu, \delta_u$, and κ appear to be the most important building characteristics.

8. Building damage models

Given the responses δ_p and A_p , this section addresses the ensuing damage. Damage is here expressed as the ratio of the repair

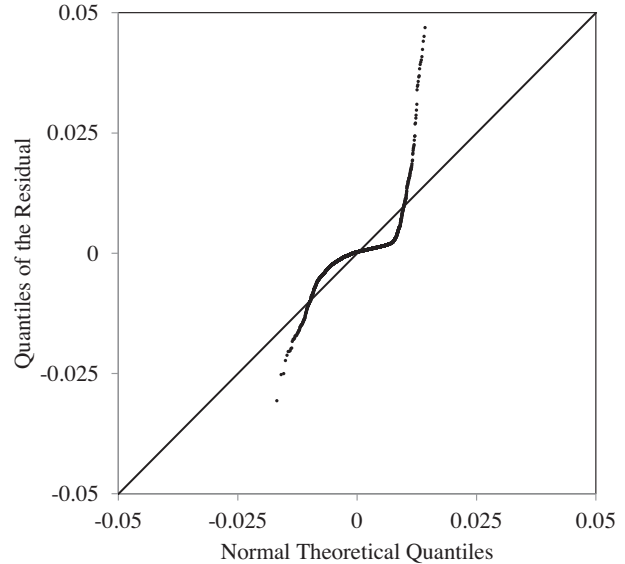


Fig. 8. Quantile–quantile plot to assess the normality of residuals of a drift response model without the natural logarithm on the left-hand side.

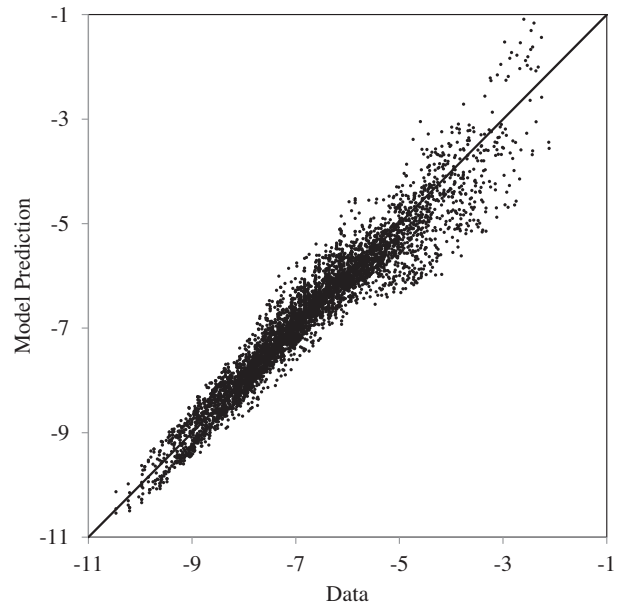


Fig. 9. Predictions of the drift model in Eq. (15) versus data.

cost to the replacement cost of the building. Four damage ratios are developed: (1) Structural damage, η_S ; (2) Non-structural drift-sensitive damage, η_{ND} ; (3) Non-structural acceleration-sensi-

tive damage, η_{NA} ; and (4) Content damage, η_C . The first two factors depend on δ_p , while the last two depend on A_p .

Data for the regression analysis is generated by substitution of realizations of δ_p and A_p into the fragility curves from FEMA–NIBS [6]. To account for the increased damage due to building irregularities, the information from the building scoring system in ATC-21 [20] is employed. ATC-21 assigns a base score to each building, which is the negative \log_{10} of the probability of exceeding a 60% damage ratio. In addition, different irregularities modify the base score. For example, ATC-21 proposes a decrease of 0.5 to the base score for a wood frame building if it has a vertical irregularity. Consequently, to generate data for such a building, the probability of the highest damage state is increased by $10^{0.5} = 3.2$. A similar method is employed in [21] to incorporate irregularities in the risk assessment of a city. Using the updated damage state probabilities and the damage ratios associated with each damage state, the expected damage ratio is computed, with δ_p, A_p, H, α , and irregularity indices, I_i , recorded as regressors.

For all models, a smooth increase in damage from 0 to 1 due to increasing building responses is sought. As in the regional modeling, polynomial, trigonometric, and even logit functions were tested, but the standard normal CDF, Φ , again turned out to provide the best fit. The building irregularities are included in the structural damage model by means of the exponential function. This function produces a factor to increase the damage if irregularities exist. This yields the following model for structural damage:

$$\eta_s = \left(\frac{\Phi(\theta_1 + \theta_2 \cdot \ln(\delta_p) + \theta_3 \cdot \ln(H) - \theta_4 \cdot \alpha)}{\exp(\theta_5 \cdot I_{VI} + \theta_6 \cdot I_{PI} + \theta_7 \cdot I_{SS} + \theta_8 \cdot I_{SC} + \theta_9 \cdot I_P)} \right) + \varepsilon. \quad (17)$$

The second moment information for the model parameters θ_i for each of the 13 building prototypes is available in Rt and in [19]. The other damage models are considered independent from the building prototype and building irregularities:

$$\eta_{ND} = \Phi(\theta_1 + \theta_2 \cdot \ln(\delta_p)) + \varepsilon, \quad (18)$$

$$\eta_{NA} = \Phi(\theta_1 \cdot \ln(A_p) - \theta_2 \cdot \alpha) + \varepsilon, \quad (19)$$

$$\eta_C = \Phi(\theta_1 \cdot \ln(A_p) - \theta_2 \cdot \alpha) + \varepsilon. \quad (20)$$

Fig. 10 shows median predictions of the structural damage according to Eq. (17) for four common prototypes. The figure shows that at the same drift ratio, unreinforced masonry buildings incur the most damage, while steel frame and concrete shear wall buildings experience the least damage. Light wood frame buildings fall in between. The negative sign of the θ_i parameters associated with α in Eqs. (17), (19) and (20) correctly indicates that the damage decreases as the quality of construction increases. It is also noted that the structural damage model in Eq. (17) suggests that taller buildings incur more damage at the same level of drift ratio. Furthermore, amongst the θ_i parameters that correspond to irregularities ($\theta_5 - \theta_9$) in Eq. (17), regression yields the highest mean for θ_7 for most prototypes. This implies that soft-story irregularity is the most detrimental type of irregularity. Conversely, θ_9 has the lowest mean for most prototypes, which indicates that pounding imposes least damage compared with other irregularities.

9. Building loss model

Provided a damage ratio, the associated repair cost is computed by multiplying it with the building replacement cost per unit floor area and the building floor area. Summation over structural, non-structural, and content yields:

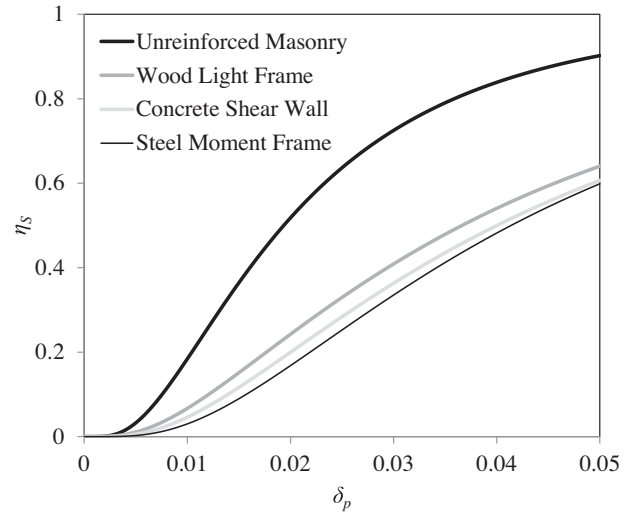


Fig. 10. Structural damage ratio versus drift ratio.

$$l = (\eta_s \cdot C_S + \eta_{ND} \cdot C_{ND} + \eta_{NA} \cdot C_{NA} + \eta_C \cdot C_C) \cdot A \cdot \varepsilon, \quad (21)$$

where η_i = damage ratios from the previous section, C_i = corresponding replacement costs per unit floor area, A = total floor area, and ε = model error variable that is a normal random variable with unit mean and 10% coefficient of variation. Specific information for different occupancy classes is provided in the companion paper.

10. Discounting model

Future seismic losses must be discounted to present value to facilitate comparison of different risk mitigation actions. The discounting model in Rt employs continuous discounting to discount the future loss, l , to present value, l_p . As a result, losses far into the future have less present value than a loss at present time. The discounting model in Rt takes a dedicated Time parameter, t , as input, which identifies the time that the loss occurs. The model reads

$$l_p = l \cdot \exp(-\rho \cdot t), \quad (22)$$

where ρ = effective interest rate. In the reliability analyses conducted in this study, ρ is a normal random variable with mean equal to 3% and 10% coefficient of variation.

11. Conclusions

This study proposes a reliability-based approach for risk analysis. The approach employs a collection of many interacting probabilistic models, and new models are presented in this paper. The models are generic, and they are applied in the companion paper to risk analysis for the Vancouver metropolitan region in Canada. Comprehensive modeling of uncertainty and updating of the models as new information becomes available are key driving forces behind this paper. In fact, an important vision behind this paper is to promote candid modeling of epistemic uncertainty, *i.e.*, reducible uncertainty, which is subsequently reduced by targeted efforts as more observations and better mechanical understanding become available. The models are implemented in Rt, which is a new general-purpose computer program that is tailored for multi-model reliability and optimization analysis. It is freely available at www.inrisk.ubc.ca. An important objective in Rt is to make reliability methods and a library of predictive models available to a broad engineering audience. This is intended to advance the use of probabilistic models and reliability methods in a variety of applications, here with focus on seismic risk.

References

- [1] Grossi P, Patel CC. Catastrophe modeling: a new approach to managing risk. Heidelberg, Germany: Springer Verlag; 2005.
- [2] Yang TY, Moehle J, Stojadinovic B, Der Kiureghian A. Seismic performance evaluation of facilities: methodology and implementation. *J Struct Eng* 2009;135(10):1146–54.
- [3] Ditlevsen O, Madsen HO. Structural reliability methods. Chichester, UK: John Wiley & Sons; 1996.
- [4] Atkinson GM, Boore DM. Empirical ground-motion relations for subduction-zone earthquakes and their application to Cascadia and other regions. *Bull Seismol Soc Am* 2003;93(4):1703–29.
- [5] Boore DM, Atkinson GM. Ground-motion prediction equations for the average horizontal component of PGA, PGV, and 5%-damped PSA at spectral periods between 0.01 s and 10.0 s. *Earthquake Spectra* 2008;24(1):99–138.
- [6] FEMA–NIBS. Earthquake loss estimation methodology – HAZUS technical manual. Washington, DC: Federal Emergency Management Agency and National Institute of Building Sciences; 2003.
- [7] Box GEP, Tiao GC. Bayesian inference in statistical analysis. New Jersey: Wiley-Interscience; 1992.
- [8] Gardoni P, Der Kiureghian A, Mosalam KM. Probabilistic capacity models and fragility estimates for reinforced concrete columns based on experimental observations. *J Eng Mech* 2002;128(10):1024–38.
- [9] Seber GAF, Wild CJ. Nonlinear regression. New York: Wiley-IEEE; 2003.
- [10] Mahsuli M, Haukaas T. Computer program for multimodel reliability and optimization analysis. *J Comput Civ Eng*; in press.
- [11] McGuire RK. Seismic hazard and risk analysis. Berkeley, CA: Earthquake Engineering Research Institute; 2004.
- [12] Der Kiureghian A, Ang AH. A fault-rupture model for seismic risk analysis. *Bull Seismol Soc Am* 1977;67(4):1173–94.
- [13] Gutenberg B, Richter CF. Frequency of earthquakes in California. *Bull Seismol Soc Am* 1944;34(4):185–8.
- [14] Ang AH, Tang WH. Probability concepts in engineering. New York: Wiley; 2007.
- [15] Building Seismic Safety Council. NEHRP guidelines for the seismic rehabilitation of buildings. Washington, DC: Federal Emergency Management Agency; 1997 [report no. FEMA-273].
- [16] Mahaney JA, Paret TF, Kehoe BE, Freeman SA. The capacity spectrum method for evaluating structural response during the Loma Prieta earthquake, Proceedings of the 1993 United States national earthquake conference, May 2–5, Memphis, TN. US Central United States Earthquake Consortium (CUSEC); 1993. p. 501–10.
- [17] Marquardt DW. An algorithm for least-squares estimation of nonlinear parameters. *J Soc Ind Appl Math* 1963;11(2):431–41.
- [18] Luco N, Bazzurro P, Cornell CA. Dynamic versus static computation of the residual capacity of a mainshock-damaged building to withstand an aftershock, Proceedings of the 13th world conference on earthquake engineering, August 1–6, Vancouver, BC, Canada, 2004.
- [19] Mahsuli M. Probabilistic models, methods, and software for evaluating risk to civil infrastructure. Ph.D. dissertation. Vancouver, BC, Canada: The University of British Columbia; 2012.
- [20] ATC. Rapid visual screening of buildings for potential seismic hazards: a handbook. Redwood City, California: Applied Technology Council; 1988 [report no. ATC-21].
- [21] McCormack TC, Rad FN. An earthquake loss estimation methodology for buildings based on ATC-13 and ATC-21. *Earthquake Spectra* 1997;13(4):605–21.