COMPDYN 2013
4th ECCOMAS Thematic Conference on
Computational Methods in Structural Dynamics and Earthquake Engineering
M. Papadrakakis, V. Papadopoulos, V. Plevris (eds.)
Kos Island, Greece, 12–14 June 2013

NEW DAMAGE MODELS FOR PERFORMANCE-BASED EARTHQUAKE ENGINEERING

Terje Haukaas[1] **and Abbas Javaherian Yazdi**[1]

[1] Department of Civil Engineering, University of British Columbia
6250 Applied Science Lane, Vancouver, BC, V6T 1Z4, Canada
{terje, abbas}@civil.ubc.ca

Keywords: Damage Models, Performance-based Earthquake Engineering, Logistic Regression, Probabilistic Methods.

Abstract. *This paper suggests two improvements to the modeling of damage in contemporary performance-based earthquake engineering. Prior to describing these developments, a review of existing performance-based engineering approaches is provided. The review describes three approaches and is accompanied by worked examples to expose the detailed steps. Particular attention is given to the modeling of damage. Subsequently, logistic regression is described for the modeling of discrete damage, followed by a description of Bayesian linear regression for the modeling of continuous visual damage. These two improvements of the state-of-the-art in damage modeling are demonstrated by application to reinforced concrete shear walls.*

1 INTRODUCTION

This paper recommends two improvements to the damage models that are presently employed in performance-based earthquake engineering (PBEE). The work is motivated by the central role that damage modeling plays in PBEE. The prediction of damage and ensuing costs is the key novelty in PBEE compared with traditional force-based and displacement-based seismic design approaches. As such, PBEE represents a crucial step forward because it allows design decisions to be based on cost-benefit considerations that weigh the cost of damage against the cost of construction to prevent such damage. However, cost predictions require a host of models for hazards and structural performance. Furthermore, significant shortcomings remain in our understanding of structural behavior and damage. As a result, this paper stresses the importance of probabilistic modeling and the gathering of data for steady improvement of the models over time.

The point of departure for the two improvements suggested in this paper is the groundbreaking work conducted by the Pacific Earthquake Engineering Research Center (PEER) and the Applied Technology Council 58 project (ATC-58). Conditional probability models form the basis for these efforts, as exemplified by the utilization of fragility functions as damage models in the ATC-58 project. The use of fragility functions has important implications for the damage modeling. Although these functions fit into an appealing analytical framework, they impose several restrictions on the damage modeling that are discussed and relaxed in this paper. The first improvement is to employ logistic regression to create multinomial and multivariate damage models that fit directly into the ATC-58 analysis procedure.

In contrast, the second improvement represents a departure from the ATC-58 analysis procedure. Instead of conditional probability models, which yield probability values, this paper advocates the development of models that predict visual damage, such as crack widths. This approach retains the advantages of the first improvement, namely multivariate modeling, and provides additional advantages in the downstream use of the damage predictions. Specifically, the prediction of visual damage facilitates the prediction of repair actions, which in turn facilitates better predictions of the cost and time of repair.

A challenge associated with the two improvements suggested in this paper is the lack of data to calibrate the new models. In traditional testing of structural components it is common to record force-deformation results, but not visual damage that can be linked to repair actions. This lack of data is a general phenomenon that hampers all damage modeling and this highlights a pressing research need. This paper argues that the solution is not to revert to simple fragility functions based on expert opinion. Rather, the solution is a commitment by the research community to increase the focus on data gathering and steady improvement of probabilistic models. The PEER and ATC-58 projects offer a substantial step in the right direction, and this paper suggests two additional steps.

2 COMPARISON OF METHODOLOGIES

The approach selected for damage modeling is naturally influenced by the methodology selected for the overall PBEE analysis. At present, several PBEE approaches are under development, and all aim to predict the monetary costs associated with earthquakes. To provide context for the damage models put forward in this paper, three PBEE approaches are reviewed. The first is the total-probability-based framing equation established by PEER [1]; the second is the sampling-based analysis approach adopted by ATC-58 [2]; the third is the reliability-based approach employed in the authors' research group [3]. The following exposure stresses the different approaches for damage modeling, and employs worked examples with artificial data to explain the details of each step in the different PBEE approaches.

2.1 PEER framing equation

To understand the PEER approach, consider the four continuous random variables *IM*, *EDP*, *DM*, and *DV*. These represent, respectively, the earthquake intensity measure, the engineering demand parameter that measures the structural response, the damage measure, and the decision variable that measures the repair cost. In the following, random variables are denoted by uppercase letters, while the same lowercase letters represent their realizations. The PEER framing equation employs the theorem of total probability to express the sought probability distribution for DV [1, 4]

$$\begin{aligned} G(dv) &= \int_0^\infty \int_0^\infty \int_0^\infty G(dv \mid dm) \cdot \left| \frac{\mathrm{d}G(dm \mid edp)}{\mathrm{d}dm} \right| \cdot \left| \frac{\mathrm{d}G(edp \mid im)}{\mathrm{d}edp} \right| \cdot \left| \frac{\mathrm{d}G(im)}{\mathrm{d}im} \right| \cdot \mathrm{d}dm \cdot \mathrm{d}edp \cdot \mathrm{d}im \\ &= \int_0^\infty \int_0^\infty \int_0^\infty G(dv \mid dm) \cdot f(dm \mid edp) \cdot f(edp \mid im) \cdot f(im) \cdot \mathrm{d}dm \cdot \mathrm{d}edp \cdot \mathrm{d}im \end{aligned} \tag{1}$$

where G denotes the complementary cumulative distribution function (complementary CDF) and f denotes the probability density function (PDF). $G(\ |\ )$ and $f(\ |\ )$ denote conditional probability distributions. In the original formulation [1] the exceedance probabilities $G(dv)$ and $G(im)$ were replaced by the annual exceedance rates $\lambda(im)$ and $\lambda(dv)$. This is not a principal difference, but to understand the meaning of G it is helpful to note that

$$G(dv) \equiv P(DV \geq dv) \tag{2}$$

and

$$G(dm \mid edp) \equiv P(DM \geq dm \mid EDP = edp) \tag{3}$$

An important purpose of Eq. (1) is its compartmentalization of intensity modeling, structural modeling, damage modeling, and cost modeling. For example, researchers charged with the task of developing damage models for the PEER formulation are asked to develop conditional probability models of the type in Eq. (3). To understand the details of the PEER approach, which is the foundation for the ATC-58 approach described later, Eq. (1) is evaluated in the following for a set of artificial data and models. Figure 1 displays the assumed hazard function, $G(im)$, which is selected to be the lognormal distribution with mean 0.8 and 20% coefficient of variation. For the sake of clarity, the variable *IM* can here be thought of as a unidirectional horizontal pseudo-spectral acceleration, measured in units of acceleration of gravity.

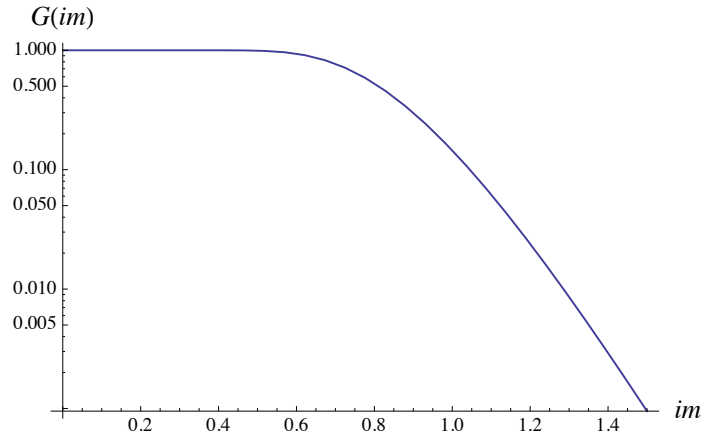

Figure 1. Hazard function.

Figure 2 displays the structural model, $G(edp|im)$ namely the conditional probability distribution for *EDP* given *IM*. Assuming that *EDP* is an inter-storey drift in percent, i.e., the ratio of peak displacement to storey height multiplied by 100, the distribution for *EDP* is here selected to be the lognormal distribution with mean $\mu_{EDP}=2im$ and 20% coefficient of variation. This means that drift realizations will be greater than zero, while drift realizations greater than 5% are unlikely. Naturally, Figure 2 shows that the probability for a high *edp*-value increases with increasing *im*-value.

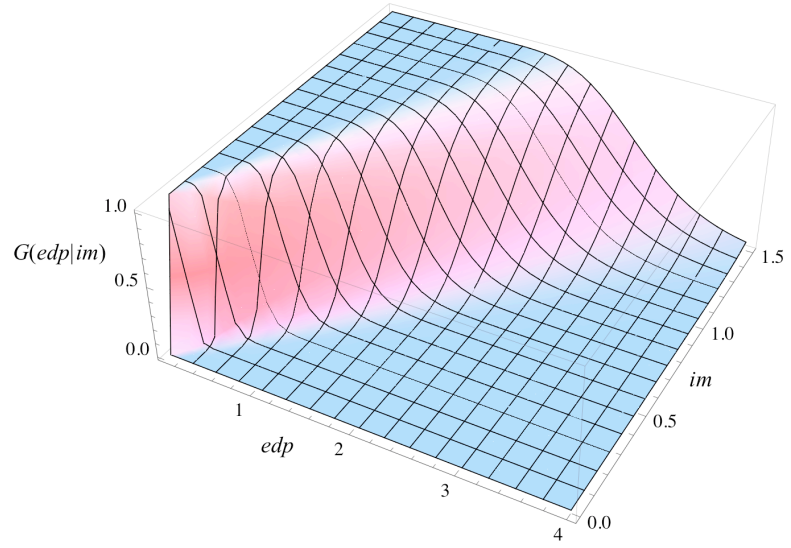

Figure 2. Structural model.

Figure 3 displays the damage model, $G(dm|edp)$, i.e., the conditional probability distribution for *DM* given *EDP*. Assuming that *DM* is a damage index that takes on values in the range from zero to around unity, the distribution for *DM* is here selected to be the lognormal distribution with mean $\mu_{DM}=0.25edp$ and 20% coefficient of variation. Figure 3 shows that the probability for high damage increases with increasing *edp*-value.

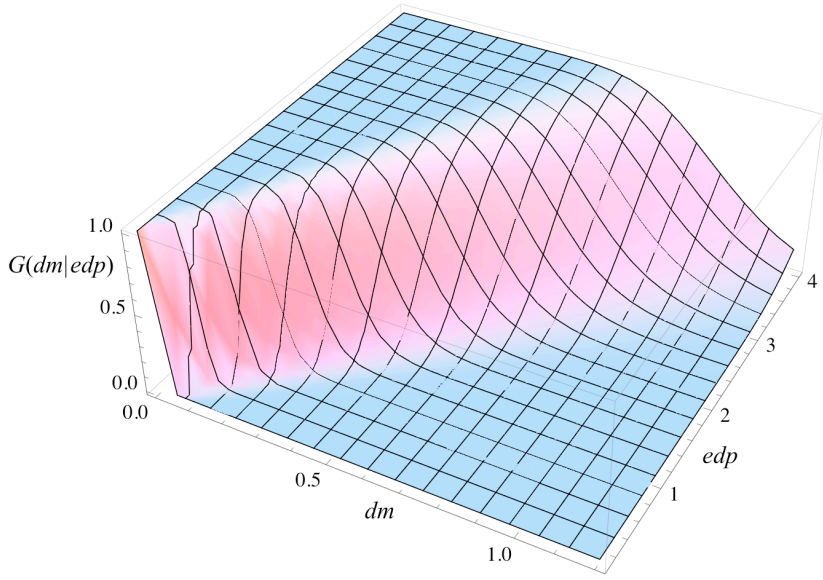

Figure 3. Damage model.

Under the assumption that DV represents the cost of repair of damage, $G(dv|dm)$, is selected to be a lognormal distribution with mean μ_{DV}=\$10,000+\$20,000dm and 20% coefficient of variation. Figure 4 shows how the probability for high cost increases with increasing damage.

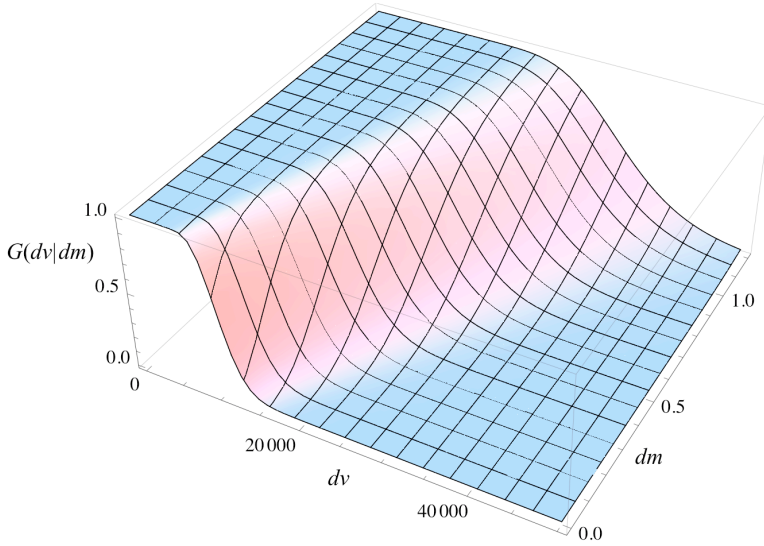

Figure 4. Cost model.

The analytical evaluation of the triple-integral in Eq. (1) is cumbersome. A robust and appealing alternative is to evaluate it numerically as a matrix-vector product. This is achieved by evaluating the probability distributions $G(dv|dm)$, $G(dm|edp)$, $G(edp|im)$, and $f(im)$ at a grid of points along the axes dv, dm, edp, and im. Employing index notation, these points are denoted by dv_l, dm_k, edp_j, im_i. Figures 1 through 4 display $f(im_i)$, $G(edp_j|im_i)$, $G(dm_k|edp_j)$, $G(dv_l|dm_k)$ evaluated at those points. This discretization strategy means that $f(im_i)$ is a vector and that the conditional PDFs $f(edp_j|im_i)$, $f(dm_k|edp_j)$, and $f(dv_l|dm_k)$ in Eq. (1) are matrices. As a result, Eq. (1) is evaluated by the matrix-vector product

$$G(dv_l) = G(dv_l \mid dm_k) \cdot f(dm_k \mid edp_j) \cdot f(edp_j \mid im_i) \cdot f(im_i) \tag{4}$$

where summation over repeated indices is implied. From the index notation it is understood that the left-hand side is a vector, and evaluating Eq. (4) with the values displayed in Figures 1 through 4 yields the function $G(dv_l)$ that is plotted in Figure 5. In other words, Figure 5 displays the probability of exceedance of a range of cost values. It is observed that the mean is around \$20,000 and that it is likely that the cost will be greater than \$10,000. Conversely, it is unlikely that the cost will be greater than \$40,000.

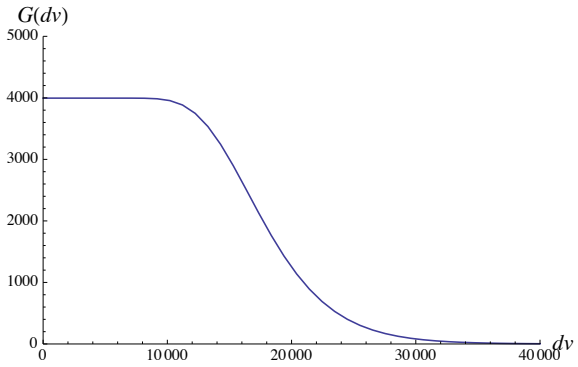

Figure 5. Exceedance probability curve for the decision variable.

The models and results presented above are intended to illustrate the key concepts of the PEER approach. In particular, it is observed that damage is modeled by a conditional probability distribution, as the one visualized in Figure 3. This is a model with significant limitations. Most importantly, neither the model nor the evaluation of Eq. (1) maintains the above simplicity when *EDP* and *DM* are vector-valued. This has been recognized in PEER because Eq. (1) is primarily employed to coordinate the research among researchers who address different components of the PBEE problem.

2.2 ATC-58 approach

The ATC-58 project originated with efforts to extend the PEER approach to comprehensive and realistic building analysis [5]. A key aspect of the framing equation is retained in the ATC-58 formulation, namely the use of conditional probability functions. In particular, damage is modeled by "fragility functions." However, these fragility functions should not be confused with the conditional probability model $G(dm|edp)$ that appears in the PEER framing equation. As illustrated shortly, the ATC-58 project employs discrete random variables to characterize damage, while *DM*, introduced earlier, is a continuous random variable.

An ATC-58 analysis commences with the creation of a detailed finite element model of the structure under consideration. This model is subjected to a suite of recorded ground motions that are scaled to specific hazard levels. For example, the ground motions may be scaled so that they match the pseudo-spectral acceleration that has 50%, 10%, and 5% probability of being exceeded in 50 years. For the sake of this demonstration, suppose Table 1 is obtained by recording the peak inter-storey drifts of a three-storey building at those three hazard levels. Typically, finite element analyses are conducted to obtain these results, but Table 1 is here created with artificial data. The first-, second-, and third-storey drift realizations are denoted edp_1, edp_2, and edp_3, respectively and these three variables are collected in the vector **edp**. A comprehensive analysis would of course include additional EDPs that may affect the damage, such as peak inter-storey acceleration values.

	50% in 50 years			10% in 50 years			5% in 50 years		
	edp_1	edp_2	edp_3	edp_1	edp_2	edp_3	edp_1	edp_2	edp_3
GM#1	1.10	0.76	1.19	1.76	2.14	2.46	2.71	2.93	2.96
GM#2	1.29	0.91	0.91	1.87	1.23	2.38	2.98	3.23	3.61
GM#3	1.11	0.97	0.74	2.15	1.85	2.09	3.55	2.60	3.37
GM#4	0.80	1.01	1.20	1.57	1.68	1.79	2.89	2.86	3.83
GM#5	0.89	0.75	0.94	1.68	1.48	1.58	2.71	3.16	3.07
GM#6	0.94	0.66	0.96	1.56	2.38	1.68	3.98	3.14	3.41
GM#7	1.09	1.01	0.99	2.29	3.16	1.55	2.77	3.42	3.00
GM#8	0.86	0.95	1.03	2.16	2.40	2.20	2.36	2.90	2.88
GM#9	0.87	0.76	1.36	1.40	2.33	2.26	3.76	2.42	3.81
GM#10	0.69	0.91	0.90	1.99	1.90	1.54	3.11	2.54	3.04

Table 1: EDP values, drift in percent (GM=ground motion).

The ATC-58 approach proceeds by fitting a joint lognormal distribution to the EDPs to facilitate computationally efficient generation of many more realizations of the EDPs. In accordance with [5] this is accomplished by first taking the natural logarithm of all values in Table 1. The result is shown in Table 2, where the notation $\mathbf{Y}=\{Y_1, Y_2, Y_3\}$ is employed to de-

note these variables, whose realizations are denoted $\mathbf{y}=\{y_1, y_2, y_3\}$. According to the lognormality assumption for **edp**, the variables **Y** have the joint normal distribution.

50% in 50 years			10% in 50 years			5% in 50 years		
y_1	y_2	y_3	y_1	y_2	y_3	y_1	y_2	y_3
0.09	-0.28	0.18	0.56	0.76	0.90	1.00	1.08	1.09
0.25	-0.09	-0.09	0.62	0.21	0.87	1.09	1.17	1.28
0.10	-0.03	-0.31	0.76	0.62	0.74	1.27	0.95	1.21
-0.22	0.01	0.18	0.45	0.52	0.58	1.06	1.05	1.34
-0.12	-0.28	-0.06	0.52	0.39	0.46	1.00	1.15	1.12
-0.06	-0.42	-0.04	0.45	0.87	0.52	1.38	1.14	1.23
0.09	0.01	-0.01	0.83	1.15	0.44	1.02	1.23	1.10
-0.16	-0.05	0.03	0.77	0.87	0.79	0.86	1.06	1.06
-0.14	-0.27	0.31	0.34	0.85	0.81	1.33	0.88	1.34
-0.37	-0.09	-0.10	0.69	0.64	0.43	1.14	0.93	1.11

Table 2: The natural logarithm of the values in Table 1.

Next, considering one hazard level at a time, the second-moment statistical information for the variables **Y** is computed. For the values shown in Table 2, the second-moment results are shown in Table 3. Following common statistical notation, the second-moment statistics for **Y** at each hazard level are collected in the mean vector, $\mathbf{M_Y}$, the diagonal matrix of standard deviations, $\mathbf{D_Y}$, and the correlation matrix $\mathbf{R_{YY}}$. This information, shown in Table 3, essentially defines the probability distribution for the inter-storey drifts and represents the uncertainty in the peak structural responses due to the varying characteristics of the ground motions. It will shortly be demonstrated how inter-storey drift realizations are generated from the information in Table 3.

50% in 50 years			10% in 50 years			5% in 50 years		
y_1	y_2	y_3	y_1	y_2	y_3	y_1	y_2	y_3
-0.0533	-0.1485	0.0081	0.5989	0.6871	0.6534	1.1139	1.0653	1.1881
0.1859	0.1507	0.1749	0.1632	0.2716	0.1866	0.1648	0.1128	0.1069
1	0.0398	-0.2548	1	0.2215	-0.1162	1	-0.3891	0.5990
	1	-0.2538		1	-0.1844		1	-0.2742
		1			1			1

Table 3: Second-moment information for the random variables **Y**. First row of numbers are mean values. Second row of numbers are standard deviations. The three last rows are symmetric correlation matrices.

Having established the probability distribution for the structural responses, attention is turned to the damage and repair-cost models. As mentioned earlier, the ATC-58 approach stands apart from the PEER formulation by adopting a discrete variable, *DS*, to denote the damage state of a component. This random variable typically has two to four damage states, denoted by ds_1, ds_2, etc. In ATC-58 parlance, models that describe the probability that a component is in a particular damage state or greater are referred to as fragility functions [2]. This terminology is adopted in this paper although, formally, a fragility function should represent the probability of failure for a component or building for given values of an intensity measure, including uncertainty in both demand and capacity. In the ATC-58 methodology, the generic fragility function is written [2]

$$\mathrm{P}(DS \geq ds_j \mid EDP = edp) = \Phi\left(\frac{1}{\sigma_j} \cdot \ln\left(\frac{edp}{\mu_j}\right)\right) \tag{5}$$

where j is the damage state number, Φ is the standard normal cumulative distribution function, and μ_j and σ_j are parameters of the distribution function. However, it is strongly emphasized that Φ does *not* represent a probability distribution in this equation. It is reiterated that the function Φ has no connection with the function $G(dm|edp)$ employed earlier because DM is continuous while DS is discrete. Rather, the distribution function Φ is employed because it is an appealing two-parameter function that varies smoothly between zero and unity as edp increases. As a result, it is understood that there is no principal reason to prefer Φ in place of, say, a one-parameter trigonometric function or a multi-parameter polynomial function. Φ is simply a well-behaved two-parameter function that is employed in the ATC-58 project to model the probability that a component is in a specific damage state or greater.

Figure 6 shows the fragility functions that are employed in this demonstration. Each curve displays the probability that the damage state is equal to or greater than the associated state. Hence, this example includes four damage states. The left-most curve in the figure displays the probability that the component is in damage state 1 or greater; the next curve displays the probability that the component is in damage state 2 or greater, and so forth. As a result, the probability that a component is in damage state number j for a given EDP value is:

$$\mathrm{P}(DS = ds_j) = \mathrm{P}(DS \geq ds_j) - \mathrm{P}(DS \geq ds_{j+1}) \tag{6}$$

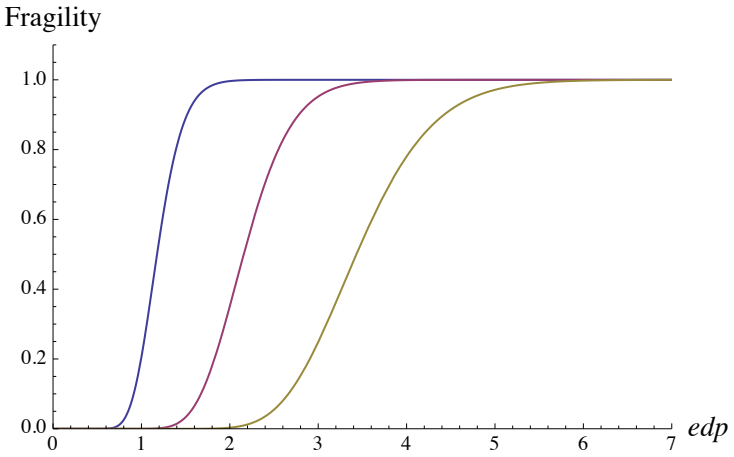

Figure 6. Fragility functions.

In the ATC-58 approach, the damage state determines the surface area to be repaired, which directly determines the repair cost. Economy-of-scale is included by reducing the repair cost for large repair jobs. In this simple demonstration example it is assumed that the repair cost, dv, per storey is \$0, \$10,000, \$20,000, and \$40,000 for damage states 1, 2, 3, and 4, respectively. Having established the damage and repair-cost models, the sampling analysis that is at the heart of the ATC-58 is conducted by executing the following algorithm, at each hazard level:

1. For each sample:
    a. Generate the realization of three independent standard normal random variables, collected in the vector **u**
    b. Transform the realization **u** into a realization of **y** by the well-known probability transformation [6]

$$\mathbf{y} = \mathbf{D}_{\mathbf{Y}}\mathbf{L}_{\mathbf{Y}}\mathbf{u} + \mathbf{M}_{\mathbf{Y}} \tag{7}$$

    where $\mathbf{L}_{\mathbf{Y}}$ is the lower-triangular Cholesky decomposition of $\mathbf{R}_{\mathbf{YY}}$.

    c. Transform the realization of **y** into a realization of **edp** by taking the exponent of each element of the vector **y**
    d. For each edp_i:
        i. Generate a realization, q, of a random variable that is uniformly distributed between 0 and 1
        ii. If $q>\mathrm{P}(DS\geq ds_1|EDP=edp_i)$, that is, if q lies above the left-most fragility function at the given edp, then the damage state is 1
        iii. Otherwise, if $q>\mathrm{P}(DS\geq ds_2|EDP=edp_i)$, then the damage state is 2
        iv. Otherwise, if $q>\mathrm{P}(DS\geq ds_3|EDP=edp_i)$, then the damage state is 3
        v. Otherwise the damage state is 4
        vi. Compute the cost of repair, dv, which is directly determined by the damage state
    e. Append the total cost, i.e., the sum of the repair cost in each storey, to an array that contains the total cost value for each sample

Figure 7 shows the complementary cumulative frequency diagram for DV obtained by 10,000 samples at each hazard level. Because the damage and ensuing loss is discrete, the diagrams that are produced from the sampling results are jagged. The smooth functions in Figure 7 are lognormal distributions that are fitted to the jagged distributions.

Each curve in Figure 7 is a conditional distribution $G(dv|im)$, where the cost is denoted dv. The final loss curve is obtained, in theory, by the integral

$$G(dv) = \int_0^\infty G(dv|im)\cdot f(im)\,\mathrm{d}\,im \tag{8}$$

In practice, this integral must be evaluated by summation instead of integration because $G(dv|im)$ is evaluated only at three hazard levels. The rule of total probability yields

$$G(dv) = \sum_{i=1}^{3} G(dv|im_i)\cdot p_i \tag{9}$$

where p_i represents the relative probability of occurrence of hazard level i, obtained by differentiation of the hazard curve and scaled to satisfy the constraint $p_1+p_2+p_3=1$. The values $p_1=0.2$, $p_2=0.6$, and $p_3=0.2$ yield the combined loss curve, $G(dv)$, in Figure 8.

In this simple demonstration example, one repair cost is associated with each EDP. This is a substantial simplification of the ATC-58 approach, although the difference is not conceptual.

The complete methodology developed by ATC-58 divides the structural and non-structural components into a host of "performance groups." For example, one performance group may be reinforced concrete columns of a certain type, while another may be internal partition walls of a specific type. Each performance group is associated with an *EDP*, such as inter-storey drift or inter-storey acceleration.

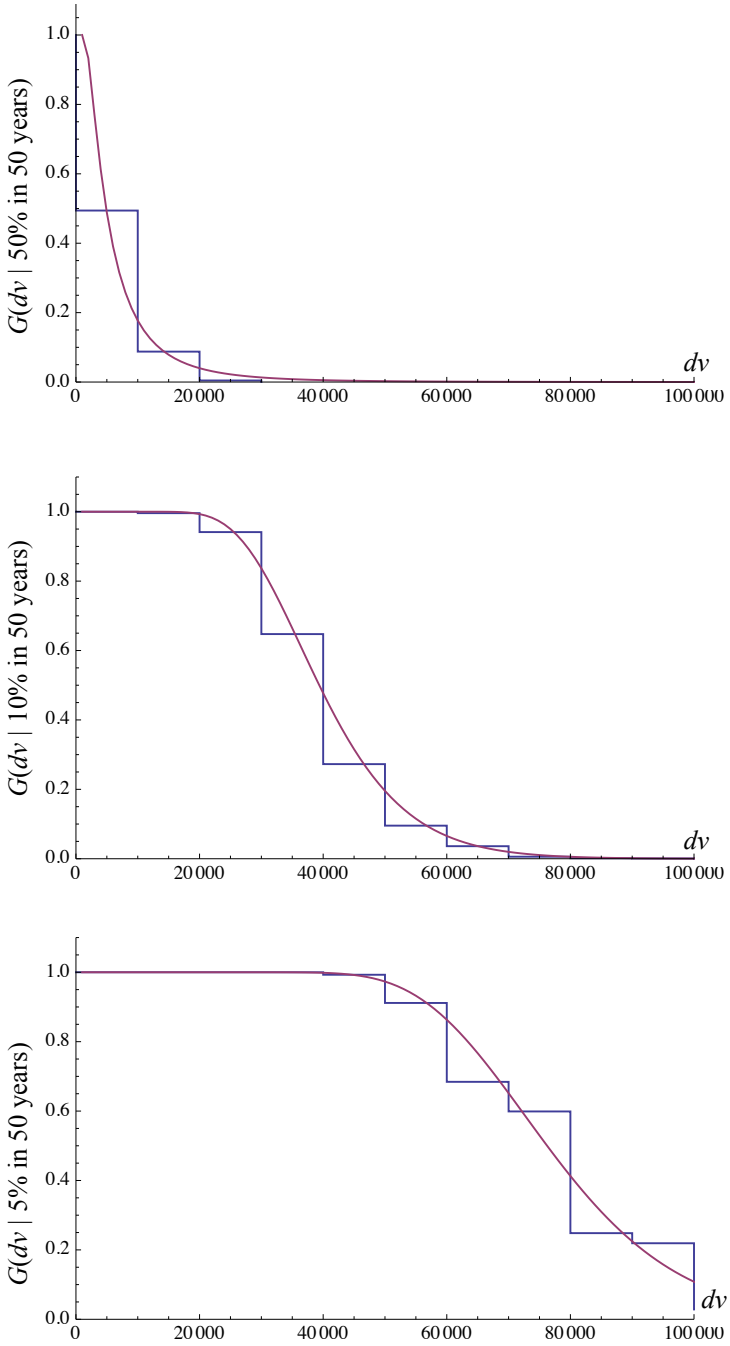

Figure 7. Loss curves at the three hazard levels.

The damage within each performance group is governed by fragility functions of the type in Eq. (5). This type of damage modeling is omnipresent in contemporary performance-based earthquake engineering. The reason for this is, in part, the appealing analytical form, which renders possible total probability integration. Another reason is that damage modeling is in-

herently challenging, and it is often difficult to justify more complex models. On the other hand, several disadvantages are associated with the use of Eq. (5). First, it may be argued that a better approach would be to model the actual visual damage rather than the probability of damage. It is also observed that the damage in Eq. (5) is solely determined by the value of one structural response. Neither the influence of material and geometry parameters, nor the influence of boundary conditions and other structural responses are accounted for. As a result, an entire "phonebook" of fragility functions are needed to cover the range of building components encountered in practice. Therefore, although the ATC-58 project represents a comprehensive and commendable effort to create a library of univariate fragility functions, the potential for improvements exists.

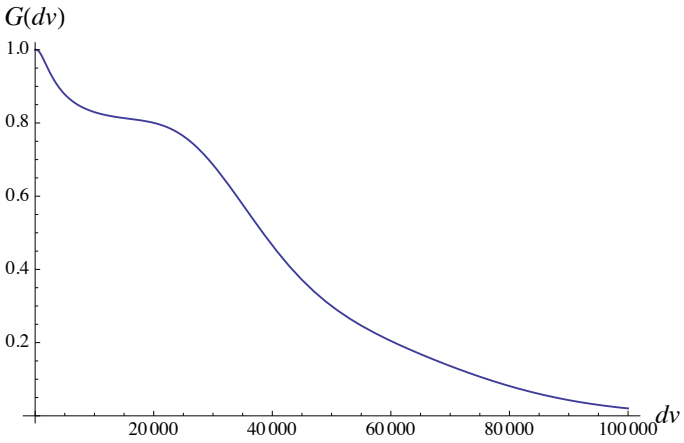

Figure 8. Combined loss curve.

2.3 Reliability-based approach

It is understood from the previous examples that the key result of a PBEE analysis is the probability distribution for the repair cost, DV. Figure 5 shows this result obtained by the PEER approach and Figure 8 shows the result obtained by the ATC-58 approach. An alternative analysis approach, based on reliability methods, is now demonstrated. This approach has been referred to as unified reliability analysis [3] because of the flexible but unified format of the models employed in this approach.

To understand the reliability-based methodology it is helpful to recall that reliability problems have two ingredients: random variables and limit-state functions. The random variables describe the uncertainty and the limit-state functions define the response events for which the probability is sought. In the following, one limit-state function, $g(\mathbf{x})$, is employed, where $\mathbf{x}$ is the vector of random variables. Following classical reliability theory [7], g is defined so that the event of interest yields negative realizations of g. Conversely; the complement of that event yields positive realizations of g. In the context of PBEE, the limit-state function is

$$g(\mathbf{x}) = dv_o - dv(\mathbf{x}) \tag{10}$$

where dv_o is a cost value selected by the analyst. A reliability analysis with the limit-state function in Eq. (10) yields the probability that dv exceeds the threshold dv_o. In other words, it yields one point on the function $G(dv)$.

A key strength of the reliability-based approach is the flexibility it offers in the choice of models. In a plug-and-play fashion, advanced finite-element-based damage models can be replaced with a simple regression-based damage model. This flexibility is available for all mod-

els, from earthquake characteristics to repair cost. In fact, it is straightforward to add costs related to construction and cost of environmental impacts, without any changes in the analysis framework. This is a substantial advantage in the quest to continuously improve probabilistic models in PBEE.

To illustrate the reliability-based approach, consider the following simple example. Suppose the ground motion intensity, *IM*, is modeled by the same random variable as the one visualized in Figure 1, namely a lognormal random variable with mean 0.8 and 20% coefficient of variation. It is here stressed that a reliability-based analysis would often aim to model the physical phenomena behind the ground motion intensity, i.e., the magnitude, location, and attenuation of the earthquake [8]. However, this is not pursued in this paper.

Next, the ground motion intensity is translated into forces applied to the structural model shown in Figure 9. This structural model is created in the computer program St, which is freely available from the website that belongs to the authors' research group, www.inrisk.ubc.ca. The loads are applied as three point loads with value F_1 at the first storey and two point loads with value F_2 at the second storey. These loads are modeled as $F_1=\alpha_1 \cdot IM$ and $F_2=\alpha_2 \cdot IM$ where α_1 and α_2 are lognormal random variables with means 500kN and 600kN, respectively, and 20% coefficient of variation. These models are selected simply to demonstrate the flexibility in introducing random variables in each model. The structural model in Figure 9 has linear material properties and consists of frame, truss, and plane continuum elements. It is unnecessary to describe the structural model in detail here, but it is noted that the model takes many material and geometry random variables as input and gives displacements and other responses as output. Of course the use of a nonlinear structural model instead of a linear model poses no conceptual problems for the reliability-based methodology. The results employed in the following are the inter-storey displacement of the two storeys.

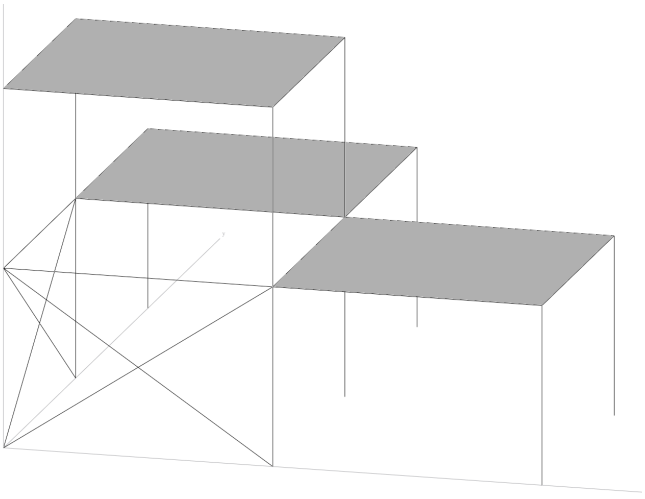

Figure 9. Structural model.

Denoting inter-storey drift by δ, the structural responses are input to the damage model

$$\eta = \begin{cases} 0 & \delta < \delta_{max} \\ \frac{1}{2} \cdot \left(\sin\left(\frac{\pi}{\varepsilon \cdot \delta_{max}} \cdot \delta - \frac{\pi}{2} \right) + 1 \right) & 0 < \delta < \delta_{max} \\ 1 & \delta > \delta_{max} \end{cases} \tag{11}$$

where η is a damage indicator, here referred to as a damage ratio, ε is a normally distributed random variable with unit mean and standard deviation 0.1, and δ_{max} is set equal to 6% drift. As shown in Figure 10, this model is a function that describes a smooth increase in damage as the drift increases.

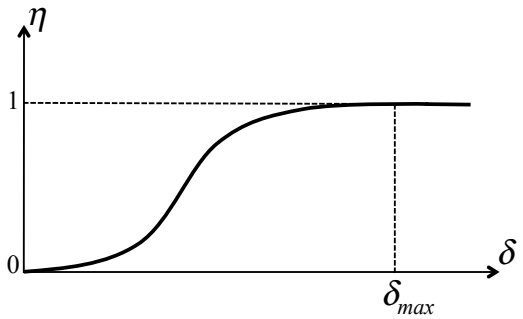

Figure 10. Visualization of the damage model in Eq. (11).

Finally, the cost is computed with the model

$$dv = \left(\eta_1 \cdot A_1 \cdot C \cdot \varepsilon\right) + \left(\eta_2 \cdot A_2 \cdot C \cdot \varepsilon\right) \tag{12}$$

where dv is the repair cost, η_1 and η_2 are the damage ratios from the damage model. A_1=108m^2 and A_2=36m^2 are floor areas of the first and second storeys, respectively. C is a random variable that represents the cost of repair per unit area, which is modeled as a lognormal random variable with mean \$2,000/m^2 and standard deviation \$400/m^2. ε is a normal random variable with unit mean and standard deviation 0.3.

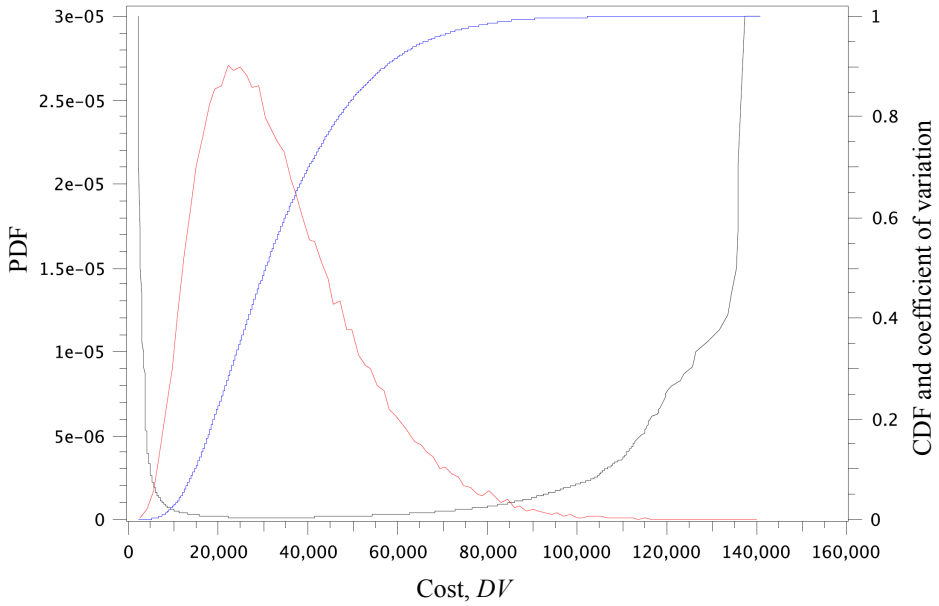

Figure 11. Visualization of $f(dv)$ and $F(dv)$ by sampling.

The reliability analysis with the above models is conduced with the computer program Rt [9], which is freely available from the website www.inrisk.ubc.ca. Rt has many analysis options and two of them are employed in the following. It is noted that the models described above take the realization of an array of random variables as input and ultimately give a realization of dv as output. For this reason, Monte Carlo sampling can be utilized to get a rough

sense of the probability distribution for *DV*. The red line in Figure 11 shows the relative frequency diagram, which is comparable to a PDF, and the blue line shows the cumulative frequency diagram, which is comparable to a CDF. It is emphasized that Monte Carlo sampling is a computationally costly approach that does not yield high accuracy in the probability estimates in the tails of the distribution. In fact, the black line in Figure 11 shows the coefficient of variation of the cumulative frequency diagram at various cost thresholds. The black line shows that the uncertainty in the sampling result is highest in the distribution tails.

To obtain the loss curve with better accuracy in the tails of the distribution, FORM analysis is carried out with the limit-state function in Eq. (10) at different cost thresholds. The accuracy of these results can be further improved by computationally effective importance sampling about the design point from FORM, but this is not done here. Figure 12 shows the resulting loss curve, $G(dv)$, obtained by FORM analysis at ten cost thresholds.

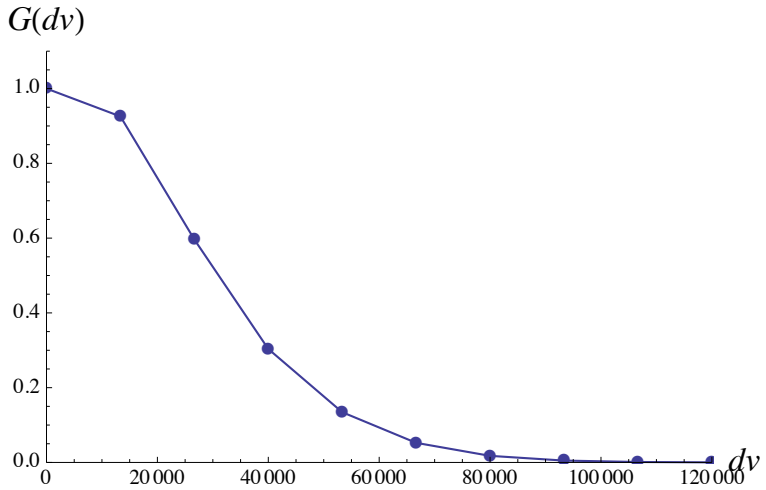

Figure 12. Points on the $G(dv)$ curve obtained by FORM and sampling reliability analysis.

It is not meaningful to compare the loss curves in Figure 5, Figure 8, and Figure 12, because the models employed in the three examples are different. Instead, the purpose has been to expose the steps of the analysis methods, and to compare the types of models that can be employed. In the reliability-based approach demonstrated in this section, none of the models are conditional probability models. Instead, all the models take random variable realizations as input and give corresponding response realizations as output. Ultimately, the response of the last model in the chain is input to the limit-state function. This modeling concept is sometimes not well understood, and the following overview of characteristics is provided to emphasize the format of models employed in the reliability-based approach.

A model, in the reliability-based PBEE approach, is an equation or algorithm, that essentially simulates possible scenarios of physical phenomena without conservative bias. This type of model discretizes all uncertainty in terms of random variables, and it takes as input the realization of continuous random variables. Ideally, it has random model parameters whose probability distribution is updated when new data emerge; i.e., it includes epistemic uncertainty. It may also take constants and design variables as input. The model returns one or more physical responses, not probabilities, and it returns a unique response for each unique realization of the random variables. It returns a response that is "continuously differentiable," i.e., smooth, with respect to the random variables, which is necessary for gradient-based reliability methods, such as FORM. It permits the simulation of all possible realizations of the outcome

space, and finally, it is modularized in the sense that it can take input from "upstream" models and/or return output to "downstream" models.

In the introduction to the seminal textbook on structural reliability methods [7], Ditlevsen and Madsen argue that probabilistic analysis is an extension of deterministic analysis because deterministic quantities can be interpreted as particularly trivial random variables. This philosophy is adopted in the above definition of a probabilistic model; the starting point is always a deterministic model in which random variables are substituted for the deterministic ones. This is the format for the collection of models that are implemented in the computer program Rt.

3 MODELING DISCRETE DAMAGE BY LOGISTIC REGRESSION

In contemporary PBEE, the ATC-58 project is an important initiative that provides context for this paper. As seen above, the ATC-58 employs fragility functions to model damage. In particular, the fragility curve represents P(*DS*≥*ds*|*EDP*=*edp*).The idea in this paper is to utilize logistic regression to develop multivariate models for P(*DS*≥*ds*), where any number of EDPs and material and geometry parameters influence the damage. For readers unfamiliar with logistic regression, an introduction is provided in the following. Additional details are provided in many textbooks, including [10, 11]. As the starting point, a generic regression model is written

$$r = h(\mathbf{x}, \boldsymbol{\theta}) \tag{13}$$

where r is a continuous response variable, h is a linear or nonlinear function, $\mathbf{x}$ is a vector of explanatory variables, such as material, geometry, and structural response parameters, and $\boldsymbol{\theta}$ is a vector of model parameters. The arctypical example of Eq. (13) is the linear regression model

$$r = \theta_1 + \theta_2 \cdot x_2 + \cdots + \theta_m \cdot x_m + \varepsilon \tag{14}$$

where ε is the model error and the model is said to be linear because it is linear in the model parameters. Conversely, nonlinear regression employs a right-hand side that is nonlinear in the model parameters. Regression models of the form in Eq. (14) have been employed in a variety of engineering disciplines. For example, [12] employed models of the form

$$r = \theta_1 + \theta_2 \cdot h_2(\mathbf{x}) + \cdots + \theta_m \cdot h_m(\mathbf{x}) + \varepsilon \tag{15}$$

where h_i are explanatory functions that consist of individual or combined explanatory variables, to model the capacity of reinforced concrete columns.

Logistic regression differs from ordinary regression in two aspects: The model error, ε, is removed from the right-hand side and the left-hand side is reformulated. Specifically, in logistic regression the right-hand side is the logit function:

$$\underbrace{\ln\left(\frac{p}{1-p}\right)}_{r} = \theta_1 + \theta_2 \cdot h_2(\mathbf{x}) + \cdots + \theta_m \cdot h_m(\mathbf{x}) \tag{16}$$

where p=P(*DS*≤*ds*)=probability of a certain damage *or less*. That is, the probability in Eq. (16) is the *complement* of the fragility function presented in Eq. (5). The logit function is sketched in Figure 13 to emphasize that p varies from zero to unity while r varies from minus to plus infinity. As a result, a regression model of the form in Eq. (16) can be employed to model

probability values, which vary between 0 and 1, such as the probability that a structural component is in a specific damage state. More precisely, Eq. (16) is employed in this paper to model the probability that a component is in damage state *j* or less. The two key advantages of logistic regression in this context are that it is readily implemented in the ATC-58 analysis procedure, and that it facilitates the development of multivariate damage models.

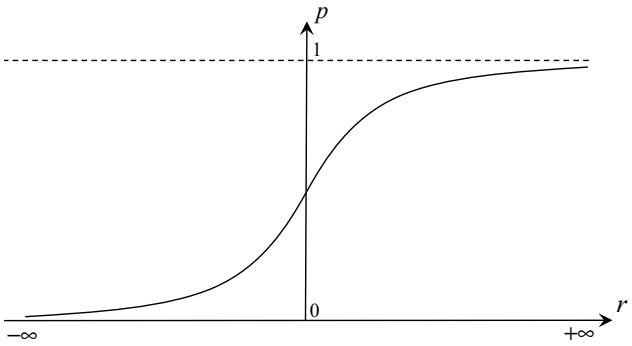

Figure 13. The logit function.

3.1 Binomial logistic regression

In order to explain how to develop multivariate models for P(*DS*≥*ds*) by means of Eq. (16) it is useful to start with binomial logistic regression, where only two damage states are considered. Suppose *n* observations are made, and that regressor values and the corresponding damage state have been recorded in each experiment. Furthermore, let the binary variable *b*={0,1} denote the damage state.

One technique for determining point estimates for the model parameters, $\boldsymbol{\theta}$, is the maximum likelihood method. The likelihood function represents the probability of observing the observed data. Provided independent observations, the likelihood function is the following product:

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} p_i^{b_i} \cdot (1 - p_i)^{1-b_i} \tag{17}$$

The validity of Eq. (17) as a likelihood function is understood by first considering the case with one observation with *b*=1, in which case the likelihood function correctly evaluates to *p*, namely the probability of observing *b*=1. Conversely, for an observation with *b*=0 the likelihood function correctly evaluates to 1–*p*, namely the probability of observing *b*=0. The maximization of Eq. (17), or its logarithm, implies that the following equation must be solved:

$$\frac{\partial \ln\left(L(\boldsymbol{\theta})\right)}{\partial \boldsymbol{\theta}} = \mathbf{0} \tag{18}$$

This is a popular and straightforward approach for the determination of point estimates for the model parameters. Additional techniques are available in the literature for the determination of the variances and covariances of the model parameters [13]. Once the statistics for the model parameters, $\boldsymbol{\theta}$, in Eq. (16) are determined, "fragilities" of the type employed in the ATC-58 project are determined from the formula

$$\mathrm{P}(DS \geq ds_j) = 1 - p = 1 - \frac{e^r}{1 + e^r} \tag{19}$$

3.2 Multinomial logistic regression

The difference between binomial and multinomial regression is that the latter allows more than two damage states. The following notation is employed:

$$p_j = P(DS \le ds_j) \tag{20}$$

where j=1, 2,…, k–1, where k is the number of damage states. This implies that the sought fragility model for the ATC-58 analysis format is

$$P(DS \ge ds_{j+1}) = 1 - p_j \tag{21}$$

where the above definition of j is maintained because caution must be exercised to ensure consistent numbering of the damage states. The generic format of the multinomial regression model that are adopted in this paper is [14]

$$\ln\left(\frac{p_j}{1-p_j}\right) = \theta_{ds_j} + \theta_2 \cdot h_2(\mathbf{x}) + \theta_3 \cdot h_3(\mathbf{x}) + \cdots + \theta_m \cdot h_m(\mathbf{x}) \tag{22}$$

Importantly, the first model parameter in the right-hand-side, namely the intercept parameter, is the only model parameter that changes from the model for one damage state to the model for another. As a result, the following constraint is employed:

$$\theta_{ds_1} \le \theta_{ds_2} \le \cdots \le \theta_{ds_{k-1}} \tag{23}$$

In short, each model is a shifted version of the model for the first damage state, and the shift is incorporated in the intercept parameter. As a result, for a situation with *four* damage states the complete set of model parameters for the *three* logistic regression models is:

$$\boldsymbol{\theta} = \left(\theta_{ds_1}, \theta_{ds_2}, \theta_{ds_3}, \theta_2, \theta_3, \ldots, \theta_m\right) \tag{24}$$

In accordance with the maximum likelihood approach, point estimates for the model parameters are obtained by solving Eq. (18), where

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n}\left(\prod_{j=1}^{k} p_{ij}^{b_{ij}}\right) \tag{25}$$

where b_{ij}=1 if observation i is in damage state j, otherwise b_{ij}=0. Analogous to Eq. (6), the probability of a particular damage state is now

$$\mathrm{P}(DS = ds_j) = p_{j+1} - p_j \tag{26}$$

where the above definition of j and k are maintained. An application of logistic regression to damage modeling for reinforced concrete shear walls is provided later in this paper.

4 MODELING VISUAL DAMAGE BY BAYESIAN LINEAR REGRESSION

The second suggested improvement to damage modeling is modeling of measurable damage and use of the reliability-based analysis procedure. This contrasts with the use of fragility curves, which gives the probability of a damage state as output instead of a physical measurable quantity. The advantage of predicting the physical damage, hereafter called visual damage, is twofold. First, it allows direct use of recorded damage data, such as crack widths, without the need for an intermediate and subjective step to characterize the damage as being in a cer-

tain damage state. Second, the prediction of visual damage makes it easier to predict the time and cost of repair. This is because the contractor will invariably use the visual damage as the primary basis for selection of repair actions, which in turn determines the time and cost of the repair.

Several techniques are available for developing models for visual damage. One approach is detailed finite element modeling of each structural component. Another approach is suggested here, namely the use of Bayesian linear regression. This is an appealing approach for two reasons. First, the Bayesian approach implies that epistemic uncertainty, i.e., model uncertainty, is explicitly addressed by probability distributions for the model parameters. Second, the use of a multivariate regression technique means that any number of EDPs, material properties, and geometry parameters can be included in the model. In the past, this type of linear regression has been employed in a variety of applications, including the development of capacity models for structural components [12].

Echoing the previous description of logistic regression, a brief description of Bayesian linear regression is provided in the following. The reader is referred to standard statistics textbooks for further details, including [15]. The generic model form is that of Eq. (15), where the model parameters are collected in the vector $\boldsymbol{\theta}=\{\theta_1, \theta_2, \ldots\}$ and the model error, ε, is assumed to be normally distributed. The observations that are employed to calibrate the model in Eq. (15) are organized in a vector $\mathbf{r}$ and a matrix $\mathbf{X}$. The matrix contains observed values of the explanatory functions, $\mathbf{x}$. One observation occupies one component of $\mathbf{r}$ and the correspond row of $\mathbf{X}$. The mean of the model parameters is the point estimate from ordinary least squares regression:

$$\mathbf{M}_{\boldsymbol{\theta}} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{r} \tag{27}$$

The standard error from classical linear regression

$$s^2 = \frac{1}{n-m}\cdot\left(\mathbf{r}-\mathbf{X}\mathbf{M}_{\boldsymbol{\theta}}\right)^T\left(\mathbf{r}-\mathbf{X}\mathbf{M}_{\boldsymbol{\theta}}\right) \tag{28}$$

also plays a role in Bayesian regression, where n is the number of observations and m is the number of explanatory variables. Assuming non-informative priors, the model parameters have the multivariate t-distribution and the standard deviation of the model error has the inverse chi-squared distribution with ν is the degrees of freedom, where $\nu=n-m$. In summary, when ν is sufficiently large, the covariance matrix for the model parameters is

$$\boldsymbol{\Sigma}_{\boldsymbol{\theta}\boldsymbol{\theta}} = s^2\left(\mathbf{X}^T\mathbf{X}\right)^{-1} \tag{29}$$

The mean of the standard deviation of the model error is

$$\mu_\sigma = s \tag{30}$$

The variance of the standard deviation of the model error is

$$\sigma_\sigma^2 = \frac{s^2}{2(\nu-4)} \tag{31}$$

Although the above formulas for the second-moment statistics of the model parameters appear simple, it is critical that sufficient efforts are devoted to model diagnostics and potential remedial actions. A host of potential issues exist, including collinearity among the explanato-

ry variables, heteroskedasticity in the model error, correlation between observed errors, non-normality in the observed errors, outliers, and model nonlinearity.

5 APPLICATIONS TO REINFORCED CONCRETE SHEAR WALLS

The two proposed improvements to damage modeling are demonstrated in the following for concrete walls. In the first application, logistic regression is employed to develop models based on actual observations. Thereafter, Bayesian regression is employed to model the amount of repairable cracks, based on data generated by an advanced numerical simulation software.

5.1 Logistic regression models for shear wall damage

The authors have been given access to data for reinforced concrete walls of the type illustrated in Figure 14. A number of researchers have conducted laboratory experiments for such shear walls, and the data have been merged into a unified database. The shear walls were subjected to cyclic testing, and load-displacement curves like those shown in Figure 15 were recorded. From such hysteresis curves, the "backbone curve," shown as solid straight lines in Figure 15, was extracted for each test specimen. In this paper, the lateral displacement is measured by the dimensionless quantity drift ratio rather than displacement, where the drift ratio is defined as

$$\delta = \frac{u}{h_w} \tag{32}$$

where u and h_w are defined in Figure 14. The data that is available for this study have the format shown in Table 4, and includes the axial load ratio, $P/A_g f'_c$, the wall height, h_w, horizontal wall length, l_w, concrete compressive strength, f'_c,yield strength of flexural reinforcement in web, f_{yl}, yield strength of flexural reinforcement in boundary zone, f_{ylb}, flexural reinforcement ratio in web zone, ρ_{lw}, flexural reinforcement ratio in boundary zone, ρ_{lb}, transverse reinforcement ratio in both web, ρ_{hw}, transverse reinforcement ratio in boundary zone, ρ_{hb}, and the drift ratio at three key points on the backbone curve.

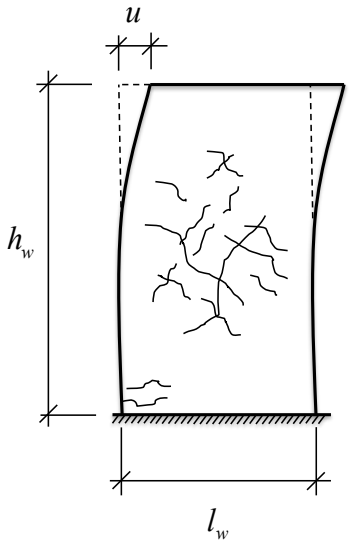

Figure 14. Reinforced concrete shear wall.

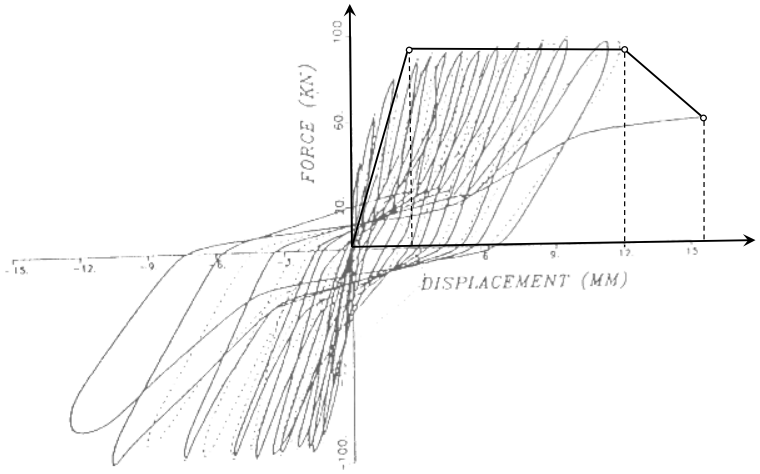

Figure 15. Backbone curve from cyclic testing.

Table 4. Excerpt from the shear wall database.

Test No.	$P/A_g f'_c$	h_w/l_w	$f'c$ [MPa]	f_{yl} [MPa]	f_{ylb} [MPa]	ρ_{lw} [%]	ρ_{lb} [%]	ρ_{hw} [%]	ρ_{hb} [%]	δ [%]
										0.61
1	0.07	2.0	47	400	471	0.27	3.17	0.26	1.5	2.17
										2.87
										0.73
2	0.07	2.0	48	448	477	0.61	7.14	0.61	1.5	3.03
										3.03
										0.49
3	0.07	1.5	48.7	399	471	0.32	3.17	0.36	1.5	3.06
										3.94
										.
.	.	.	.	.	.	.	.	.	.	.
										.

In lieu of more information about damage from the tests, it is hypothesized in this study that the characteristic points on the backbone curve are indicative of damage. Specifically, it is assumed that the drift-values δ_y, δ_u, and δ_c identify the transition from one damage state to another. Figure 16 illustrates this assumption by showing the drift-ranges where the wall is assumed to be in Damage State 1 (ds_1), Damage State 2 (ds_2), etc. Even after identifying damage states for each shear wall, another step is necessary before the data can be employed in logistic regression. This is because the observed values for δ_y, δ_u, and δ_c cannot serve as observations for developing logistic regression model. The reason is that δ_y, δ_u, and δ_c identify the *transition* between damage states, rather than regressor values within damage states. Figure 16 illustrates one way of making the data amenable to logistic regression. Instead of directly using the values for δ_y, δ_u, and δ_c, drift-values that belong to specific damage states are recorded. In Figure 16, the points on the abscissa axis that are marked by asterisk show a strategy where two drift-values are employed for each damage state. In contrast, the drift values marked my lines show an alternative strategy, where five drift-values are employed in

each damage state. Results presented shortly show that the results are not highly sensitive to the number of selected drift-values.

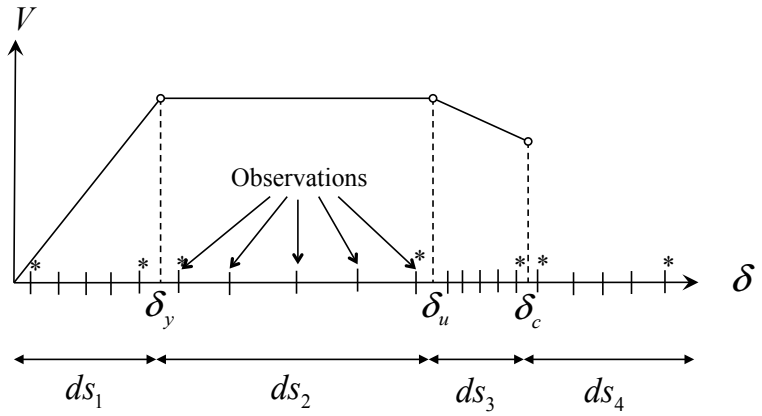

Figure 16. Damage states relative to the backbone curve.

Table 5 enumerates the explanatory functions that are used in the multinomial logistic regression models. It is noted that the natural logarithm of δ is employed rather than δ itself, to ensure zero probability of damage at zero drift.

Table 5. List of explanatory functions, $h_i(\mathbf{x})$.

h_2	Axial load ratio	$P/A_g f'_c$
h_3	Aspect ratio	h_w/l_w
h_4	Concrete compressive strength	f'_c
h_5	Yield strength of flexural reinforcement (web)	f_{yl}
h_6	Yield strength of flexural reinforcement (boundary)	f_{yb}
h_7	Flexural reinforcement (web)	ρ_{lw}
h_8	Flexural reinforcement (boundary)	ρ_{lb}
h_9	Horizontal reinforcement (web)	ρ_{hw}
h_{10}	Horizontal reinforcement (boundary)	ρ_{hb}
h_{11}	Drift ratio	$\ln(\delta)$

For the sake of simplicity of notation, let **h** denote an n-by-10 dimensional matrix whose rows contain the value of the 10 explanatory functions shown in Table 5 for each of the n observations. The n-dimensional vector **y** is also defined, to collect the damage state for each observation. Following the methodology described earlier, the MATLAB® subroutine "mnrfit" for multinomial logistic regression is utilized to estimate model parameters. Specifically, the model parameters are computed by the command $\boldsymbol{\theta}$=mnrfit(**h**,**y**,'model','ordinal'), where $\boldsymbol{\theta}$ are point estimates for the model parameters and the options 'model' and 'ordinal" are employed to obtain multinomial logistic regression models of the form described earlier. For the considered shear wall data, which includes four damage states, the models read

$$\ln\left(\frac{p_1}{1-p_1}\right) = \theta_{ds_1} + \theta_2 \cdot h_2(\mathbf{x}) + \theta_3 \cdot h_3(\mathbf{x}) + \cdots + \theta_{11} \cdot h_{11}(\mathbf{x}) \tag{33}$$

$$\ln\left(\frac{p_2}{1-p_2}\right) = \theta_{ds_2} + \theta_2 \cdot h_2(\mathbf{x}) + \theta_3 \cdot h_3(\mathbf{x}) + \cdots + \theta_{11} \cdot h_{11}(\mathbf{x}) \tag{34}$$

$$\ln\left(\frac{p_3}{1-p_3}\right) = \theta_{ds_3} + \theta_2 \cdot h_2(\mathbf{x}) + \theta_3 \cdot h_3(\mathbf{x}) + \cdots + \theta_{11} \cdot h_{11}(\mathbf{x}) \tag{35}$$

where Table 6 provides the point estimates for the model parameters, given two or five drift-observations in each damage state, as explained earlier.

Table 6. Model parameters for the models in Eqs. (33) to (35).

Model parameter	2 recorded observations in each damage state	5 recorded observations in each damage state
θ_{ds1}	-6.2815	-7.8127
θ_{ds2}	-3.5362	-4.4351
θ_{ds3}	-1.5746	-2.2594
θ_2	-2.8518	-3.818
θ_3	0.8837	1.1815
θ_4	0.0021	0.003
θ_5	0.0013	0.0016
θ_6	0.0011	0.0015
θ_7	-0.2861	-0.3224
θ_8	0.0078	0.0039
θ_9	1.5821	2.0884
θ_{10}	-0.1604	-0.2148
θ_{11}	-2.7209	-3.5176

It is observed in Table 6 that the results for two drift-observations in each damage state are correlated with the results for with five observations. In other words, the results are not highly dependent on the choice of drift-observations. This point will be further discussed shortly. Table 6 also provides information about the relative influence of each explanatory function. However, caution must be exercised, because the θ-values in Table 6 have different units. Nonetheless, by accounting for the practical range of variation of each explanatory function, it can be concluded the yield strengths of reinforcement and the compressive strength of concrete are relatively insignificant in comparion with h_w/l_w and P/Agf'_c.

To demonstrate that the models in Eqs. (33) to (35) can be employed to produce the same results as a fragility function, Figure 17 shows the fragility functions for a wall with $P/A_g f'_c$=0.5, h_w/l_w=2, f'_c=30MPa, f_{yl}=450MPA, f_{ylb}=500MPA, ρ_{lw}=0.25% ρ_{lb}=4%, ρ_{hw}=2%, and ρ_{hb}=1.5%. It is observed in Figure 17 that the models developed with two drift-observations are quite similar to those with five dift-observations. Furthermore, from the slope of the fragility curves in Figure 17 it is concluded that the drift at which the wall transitions into damage state 4 is far more uncertain than the drift at which the wall enters into damage state 1. It is also stressed that the models in Eqs. (33) to (35) are far more versatile than the impression given in Figure 17. In fact, Figure 17 is provided only to show that results similar to the ATC-58 fragility functions can be obtained when values are given for all the explanatory functions that influence damage.

At present, the models in Eqs. (33) to (35) are intended primarily for demonstration purposes. In ongoing research, more parsimonious models are developed by various model selection procedures [16].

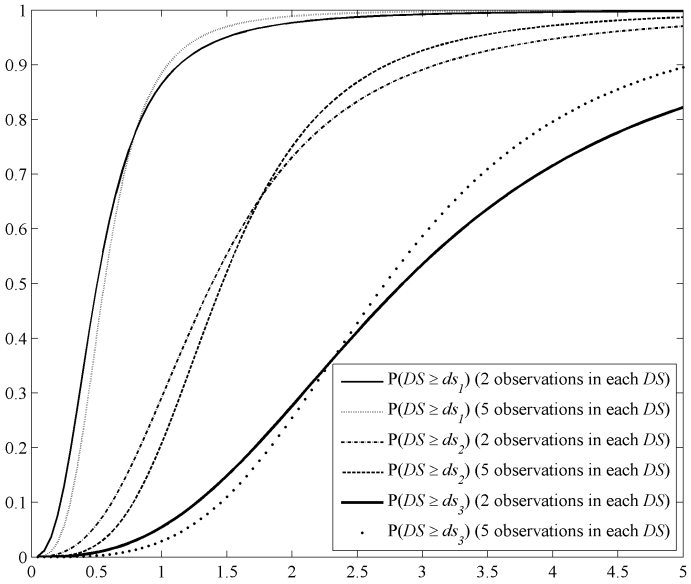

Figure 17. Variation of fragility probabilities with δ for a specific wall from multinomial logistic regression.

5.2 Linear regression models for shear wall damage

A probabilistic model for visual damage is now presented, based on work in [17]. This model aims at predicting a physical measure of damage, rather than the probability of different damage states. In particular, wall damage is formulated as a continuous variable. As mentioned earlier, once the visual damage is determined, the repair action and the associated direct and indirect costs can be estimated. Here, the physical quantity of damage is the length of cracks in the wall. In lieu of data from experiments, data is generated by detailed finite element analysis, as described shortly. To this end, 32 reinforced concrete shear walls are designed based on the Canadian Standard Association code requirements. Table 7 shows an excerpt of the properties of the walls; further details are provided in [17].

Table 7. Design details of walls.

Wall No.	f'_c [MPa]	$\sqrt{f'_c}$	$P/A_g f'_c$	ρ_{lw}	ρ_{lb}	ρ_h	f_y [MPa]	h_w/l_w
1	35	5.92	5%	0.78%	1.20%	1.00%	455	3.02
2	25	5.00	5%	0.78%	1.20%	1.00%	455	3.02
…	…	…	…	…	…	…	…	…
31	35	5.92	5%	0.79%	1.40%	1.00%	455	3.02
32	35	5.92	5%	0.81%	1.30%	1.00%	455	3.02

Each wall is subjected to a cyclic loading history in the computer program VecTor2 [18]. VecTor2 is a two-dimensional non-linear program for analysis of reinforced concrete elements. A screenshot of a wall modeled in VecTor2 is shown in Figure 18. The load cycles have steadily increasing amplitude, and after each load cycle the repairable residual cracks corresponding to the zero-force position of the wall are determined. Conducting this type of

analysis for all 32 walls yields a database of results that are employed in linear regression analysis. This data is not given in this paper, where focus is on the modeling procedure and the resulting model.

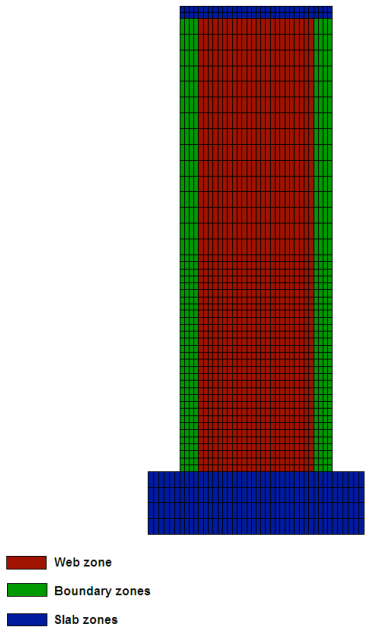

Figure 18. A screenshot of a wall modeled in VecTor2.

Once the data is available, the first step in the modeling procedure is to select a comprehensive set of explanatory functions. These are shown in Table 8. The generic model form is that of Eq. (15) and the model response is l_c/l_w, where l_c is the total crack length and l_w is the horizontal length of the wall.

Table 8. List of explanatory functions.

h_1	Drift ratio	δ
h_2	Concrete tensile strength	$\sqrt{f'_c}$
h_3	Concrete compressive strength	f'_c
h_4	Axial load ratio	$P/A_g f'_c$
h_5	Flexural reinforcement (web)	ρ_{lw}
h_6	Flexural reinforcement (boundary)	ρ_{lb}
h_7	Horizontal reinforcement (web & boundary)	ρ_h
h_8	Yield strength of flexural reinforcement	f_y
h_9	Slenderness ratio	h_w/l_w

The next step is to employ the Bayesian inference presented earlier to assess the model parameters. Furthermore, an iterative procedure is utilized to remove explanatory functions that do not influence the crack length [12]. At each iteration, the second-moment statistics of $\boldsymbol{\theta}$

and σ are calculated and the explanatory function with largest coefficient of variation of the corresponding model parameter is deemed to have the least influence on the crack length. By trial and error, explanatory functions are removed and reinserted while monitoring the standard deviation of the model error, σ. In this manner, the "optimal" model is achieved, which balances parsimoniousness with model accuracy [12]. The first iteration includes all the nine explanatory functions, where the least significant explanatory function is the ratio of flexural reinforcement in the boundary region, ρ_{lb}. In subsequent iterations, the ratio of horizontal reinforcement is removed followed by the compressive and tensile strength of concrete. The final model form, which is achieved in the sixth iteration, after which any further reduction deteriorates the model drastically, is

$$\frac{l_c}{l_w} = \theta_1 \cdot \delta + \theta_4 \cdot \frac{h_w}{l_w} + \theta_5 \cdot \frac{P}{A_g \cdot f'_c} + \theta_6 \cdot \rho_{hw} + \sigma \cdot \varepsilon \tag{36}$$

where the second-moment statistics for the model parameters is shown in Table 9.

Table 9. Statistics of model parameters.

Coefficients	Mean	COV	Correlation coefficient			
			θ_1	θ_4	θ_5	θ_6
θ_1	7.1	0.023	1.00	-0.52	-0.02	-0.16
θ_4	-0.72	0.16	-0.52	1.00	-0.70	-0.18
θ_5	-0.31	0.19	-0.02	-0.70	1.00	-0.20
θ_6	0.86	0.21	-0.16	-0.18	-0.20	1.00
σ	0.95	0.06	-	-	-	-

6 ACKNOWLEDGEMENTS

This work is supported by the first author's Discovery Grant from the National Science and Engineering Research Council of Canada (NSERC). The authors also gratefully acknowledge Ms. Shahrzad Talachian who conducted the VecTor2 analyses during her Master's studies in 2010.

7 CONCLUSIONS

The context for this paper is uncertainty analysis in contemporary performance-based earthquake engineering. In performance-based engineering the prediction of damage is a central ingredient, and this paper suggests two improvements to current damage assessment procedures. These advances are presented with the "ATC-58 approach" as a benchmark, in which fragility functions are employed to model damage probabilities. Specifically, lognormal cumulative distribution functions, which are univariate functions of a demand variable, provide the probability that a component is damaged beyond a pre-defined damage state. Each fragility function is uniquely defined by two parameters, which essentially represent the mean and standard deviation of the lognormal distribution. In the early stages of the ATC-58 project these two parameters are often based on expert opinion, but this is improving as more test data is becoming available.

The first contribution in this paper is the use of logistic regression to develop fragility functions. Logistic regression is similar to ordinary regression in the sense that any number of explanatory variables can be included to describe the phenomena. However, it differs from

ordinary regression in an important way; the output of a logistic regression model is a probability. As a result, logistic regression can be employed to develop damage fragility functions. This is a powerful extension of the ATC-58 approach because it avoids several difficulties in the current approach and it facilitates the inclusion of any number of variables that may influence the damage. In existing models, only one demand variable, for example displacement, is input to the fragility function to obtain the damage probability. With logistic regression it is possible to add any number of geometry and material variables, as well as model uncertainty variables, thus merging many fragility functions into one robust model. This paper provides a detailed description of logistic regression, together with an example and a demonstration of how such models fit within the ATC-58 analysis framework.

The analysis framework adopted by ATC-58 employs Monte Carlo sampling to obtain the key result, namely the loss exceedance probability curve. As an alternative, this paper puts forward a second extension of the ATC-58 approach, in which any reliability method can be used to obtain such results. Importantly, this paper describes how damage is modeled in the alternative approach. While logistic regression models fit directly into the ATC-58 framework, this second extension requires a reformulation of the damage models, whereby the visual damage is explicitly modeled, followed by models that predict repair action(s), in turn followed by models that predict the cost and time of the repair. None of these new models produce a probability as output. Instead, each model produces realizations of physical measurable parameters. On one hand this approach is even more flexible than the use of logistic regression above and points to a future with simulation models, but on the other hand it entails a substantial reformulation of the ATC-58 approach. Several examples are provided in this paper to compare the approaches, in order to highlight advantages and disadvantages.

REFERENCES

[1] C. Cornell and H. Krawinkler, Progress and challenges in seismic performance assessment. *PEER Center News*, http://peer.berkeley.edu/news/2000spring/index.html, 2000.

[2] Applied Technology Council, *Development of next generation performance-based seismic design procedures for new and existing buildings*. Applied Technology Council Report 58, 2008.

[3] T. Haukaas, Unified reliability and design optimization for earthquake engineering, *Probabilistic Engineering Mechanics*, **23**, 471-481, 2008.

[4] A. Der Kiureghian, Non-ergodicity and PEER's framework formula. *Earthquake Engineering and Structural Dynamics*, **34**, 1643-1652, 2005.

[5] T. Y. Yang, J. Moehle, B. Stojadinovic and A. Der Kiureghian, Seismic performance evaluation of facilities: Methodology and implementation. *Journal of Structural Engineering*, **135**, 1146-1154, 2009.

[6] A. H. Ang and W. H. Tang, *Probability Concepts in Engineering, Vol I*. Wiley, 2007.

[7] O. Ditlevson and H. O. Madsen, *Structural Reliability Methods*. Wiley, 1996.

[8] M. Mahsuli and T. Haukaas, Seismic risk analysis with reliability methods, part I: Models. *Structural Safety*, **42**, 54-62, 2013.

[9] M. Mahsuli and T. Haukaas, Computer program for multimodel reliability and optimization analysis. *Journal of Computing in Civil Engineering*, **27**, 87-98, 2013.

[10] T. P. Ryan, *Modern Regression Methods*. Wiley-Interscience, 2008.

[11] A. Agresti, *An Introduction to Categorical Data Analysis*. Wiley, 1996.

[12] P. Gardoni, A. Der Kiureghian and K. M. Mosalam, Probabilistic capacity models and fragility estimates for reinforced concrete columns based on experimental observations. *Journal of Engineering Mechanics,* **128**, 1024-1038, 2002.

[13] D. Hosmer and S. Lemeshow, *Applied Logistic Regression*. Wiley, 2000.

[14] P. McCullagh and J. A. Nelder, *Generalized linear models, Monographs on statistics and applied probability 37*. Chapman Hall, 1989.

[15] G. E. P. Box and G. C. Tiao, *Bayesian Inference in Statistical Analysis*. Addison–Wesley, 1992.

[16] A. Javaherian Yazdi and T. Haukaas, Multivariate probabilistic damage models for performance-based earthquake engineering. *Proceedings of the 11th International Conference on Structural Safety and Reliability, ICOSSAR 2013*.

[17] S. Talachian, *Probabilistic models for damage and repair cost for reinforced concrete structural members*. MASc thesis, Department of Civil Engineering, University of British Columbia, 2010.

[18] P. S. Wong and F. J. Vecchio, *Vector2 and formworks user's manual*. Department of Civil Engineering, University of Toronto, 2002.