Baker, J. W. (2015). "Efficient analytical fragility function fitting using dynamic structural analysis." *Earthquake Spectra*, 31(1), 579-599.

TECHNICAL NOTE

Efficient analytical fragility function fitting using dynamic structural analysis

Jack W. Baker[a)] **M.EERI**

Estimation of fragility functions using dynamic structural analysis is an important step in a number of seismic assessment procedures. This paper discusses the applicability of statistical inference concepts for fragility function estimation, describes appropriate fitting approaches for use with various structural analysis strategies, and studies how to fit fragility functions while minimizing the required number of structural analyses. Illustrative results show that multiple stripe analysis produces more efficient fragility estimates than incremental dynamic analysis for a given number of structural analyses, provided that some knowledge of the building's capacity is available prior to analysis so that relevant portions of the fragility curve can be approximately identified. This finding has other benefits, as the multiple stripe analysis approach allows for different ground motions to be used for analyses at varying intensity levels, to represent the differing characteristics of low intensity and high intensity shaking. The proposed assessment approach also provides a framework for evaluating alternate analysis procedures that may arise in the future.

INTRODUCTION

This paper describes statistical procedures for estimating parameters of fragility functions using nonlinear dynamic structural analysis results, and uses those procedures to evaluate various strategies for performing dynamic structural analysis to estimate fragility functions. A fragility function specifies the probability of collapse, or some other limit state of interest, of a structure as a function of some ground motion intensity measure, *IM*. The parameter *IM* is often quantified by spectral acceleration with a specified period and damping, though any measure of ground motion intensity can be used with the procedures below. Collapse fragility functions obtained from structural analysis results are increasingly popular in structural assessment

[a)] Stanford University, 473 Via Ortega, MC 4020, Stanford, CA 94305

procedures (Applied Technology Council 2012, Chapter 6; Federal Emergency Management Agency 2009). An estimated fragility function can also be combined with a ground motion hazard curve to compute the mean annual rate of structural collapse (e.g., Shome 1999; Ibarra and Krawinkler 2005; Haselton and Deierlein 2007; Liel and Deierlein 2008).

For a given ground motion and dynamic structural analysis result, the occurrence or non-occurrence of collapse can be defined in a number of ways (Zareian and Krawinkler 2007). In this paper it is assumed that this definition is established and that for a given analysis there is a criterion to determine whether or not the ground motion caused collapse. The results below are not limited to any specific collapse definition, and in fact the performance state of interest need not be related to collapse (e.g., it could be exceedance of a some drift threshold), but the term "collapse" will be used below given the common use of these procedures with collapse assessment.

There are a number of procedures for performing nonlinear dynamic structural analyses to collect the data for estimating a fragility function. One common approach is incremental dynamic analysis (IDA), where a suite of ground motions are repeatedly scaled in order to find the *IM* level at which each ground motion causes collapse (Vamvatsikos and Cornell 2002; Federal Emergency Management Agency 2009). A second common approach is multiple stripes analysis, where analysis is performed at a specified set of *IM* levels, each of which has a unique ground motion set (Jalayer 2003). As the type of data collected in these two cases differs, the appropriate approach for estimating fragility functions from the data also differs. This paper presents appropriate methods for fitting fragility functions to data from these and other related approaches. Given these methods for fragility function fitting, the paper then discusses optimal strategies for performing structural analysis in order to obtain an accurate fragility estimate with a minimal number of structural analyses.

Fragility functions are in general derived using a variety of approaches such as field observations of damage, static structural analyses, or judgment (e.g., Kennedy and Ravindra 1984; Kim and Shinozuka 2004; Calvi et al. 2006; Villaverde 2007; Porter et al. 2007; Shafei et al. 2011), but here the focus is on so-called analytical fragility functions developed from dynamic structural analysis. Unlike some other methods, in the case of analytical fragility functions the analyst has control over the data collected, by means of choosing the *IM* levels at which analysis is performed and the number of analyses performed at each level. This motivates the investigation below of effective ways to perform that data collection.

A lognormal cumulative distribution function is often used to define a fragility function

$$P(C \mid IM = x) = \Phi\left(\frac{\ln(x/\theta)}{\beta}\right) \quad (1)$$

where $P(C \mid IM = x)$ is the probability that a ground motion with $IM = x$ will cause the structure to collapse, $\Phi(\ )$ is the standard normal cumulative distribution function (CDF), θ is the median of the fragility function (the *IM* level with 50% probability of collapse) and β is the standard deviation of ln*IM* (sometimes referred to as the dispersion of *IM*). Equation 1 implies that the *IM* values of ground motions causing collapse of a given structure are lognormally distributed; this is a common assumption has been confirmed as reasonable in a number of cases (e.g., Ibarra and Krawinkler 2005; Porter et al. 2007; Bradley and Dhakal 2008; Ghafory-Ashtiany et al. 2010; Eads et al. 2013), but it is not required and alternate assumptions can be used with the procedures described below. Calibrating equation 1 for a given structure requires estimating θ and β from structural analysis results. We will denote estimates of those parameters as $\hat{\theta}$ and $\hat{\beta}$.

Parameter estimation is the field of statistics associated with estimating values of model parameters based on observed data that has a random component (e.g., Rice 1995). In this case, our parameters of interest are θ and β, and we have randomness because record-to-record variability causes ground motions with the same *IM* level to produce different demands on a given structure. There are two common statistical approaches for estimating parameters from data. The method of moments finds parameters such that the resulting distribution has the same moments (e.g., mean and standard deviation) as the sample moments of the observed data. The maximum likelihood method finds the parameters such that the resulting distribution has the highest likelihood of having produced the observed data. Common desirable properties of the estimators are that they be unbiased (the estimators do not systematically overestimate or underestimate the true parameter's value), efficient (the estimators have as small variance) and consistent (as the number of data goes to infinity, the estimator converges to the true parameter value).

The use of parameter estimation procedures to estimate fragility functions is discussed in the following section. Appropriate estimation procedures are defined and then used to investigate approaches for performing structural analysis, to identify approaches that produce

accurate fragility function estimates with a minimum number of required structural analyses. The findings are then summarized to offer recommendations on structural analysis and parameter estimation strategies that produce efficient estimates of collapse fragility functions.

ESTIMATING FRAGILITY FUNCTION PARAMETERS

There are a number of ways to estimate parameter values for a fragility function that are consistent with observed data, depending upon the procedure used to obtain structural analysis data, as discussed in this section.

INCREMENTAL DYNAMIC ANALYSIS

Incremental dynamic analysis involves scaling each ground motion in a suite until it causes collapse of the structure (Vamvatsikos and Cornell 2002). This process produces a set of *IM* values associated with the onset of collapse for each ground motion, as illustrated in Figure 1a. The probability of collapse at a given *IM* level, *x*, can then be estimated as the fraction of records for which collapse occurs at a level lower than *x*. A plot of these probabilities is shown in Figure 1b, and is referred to as an empirical cumulative distribution function. Fragility function parameters can be estimated from this data by taking logarithms of each ground motion's *IM* value associated with onset of collapse, and computing their mean and standard deviation (e.g., Ibarra and Krawinkler 2005).

$$\ln \hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} \ln IM_i \tag{2}$$

$$\hat{\beta} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(\ln\left(IM_i / \hat{\theta}\right)\right)^2} \tag{3}$$

where n is the number of ground motions considered, and IM_i is the *IM* value associated with onset of collapse for the ith ground motion. This is a method of moments estimator, as $\ln\theta$ and β are the mean and standard deviation, respectively, of the normal distribution representing the ln*IM* values. Note that the mean of ln*IM* is equal to the median of *IM* in the case that *IM* is lognormally distributed, which is why using the sample mean in this manner produces an estimate of θ. The mean and standard deviation, or moments, of the distribution are estimated using the sample moments from a set of data. A fragility function fitted using this approach is shown in Figure 1b.

This fragility fitting approach has been used widely, and is denoted "Method A" by Porter et al. (2007). It has been used to calibrate fragility functions for data other than structural collapse (e.g., Aslani and Miranda 2005). A related alternative is to use counted fractiles of the IM_i values, rather than their moments, to estimate θ and β (Vamvatsikos and Cornell 2004).

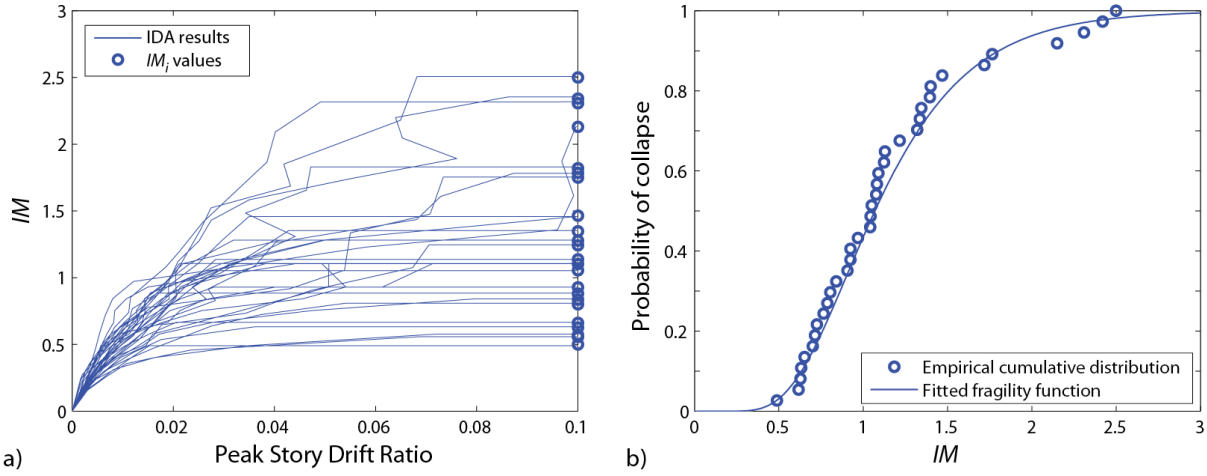

Figure 1. a) Example incremental dynamic analyses results, used to identify *IM* values associated with collapse for each ground motion. b) Observed fractions of collapse as a function of *IM*, and a fragility function estimated using equations 2 and 3.

TRUNCATED INCREMENTAL DYNAMIC ANALYSIS

With incremental dynamic analysis, some ground motions may need to be scaled to large *IM* values in order to produce collapse, which raises several concerns. First, it is computationally expensive, as it requires many structural analyses to be performed with increasing *IM* levels, in order to finally observe a collapse. Second, the large-*IM* results are less practically relevant, as the fragility function values at large *IM* levels are of less interest than values at small *IM* levels, as will be discussed further below. Finally, it is questionable whether scaling typical moderate-*IM* ground motions up to extreme *IM* levels is an accurate way to represent shaking associated with real occurrences of such large *IM* levels (Baker and Cornell 2005a). One potential strategy to address these concerns is to perform incremental dynamic analysis only up to some level, IM_{max}, above which no further analyses are performed. Illustrative results from this type of analysis are shown in Figure 2a. If *n* ground motions are used in the analysis, there will in general be *m* ground motions that caused collapse at *IM* levels less than IM_{max}, and *n*-*m* ground motions that did not cause collapse prior to the analyses being stopped.

This type of data cannot be used to estimate fragility functions using equations 2 and 3. Instead we use the maximum likelihood method to compute the likelihood of observing the data that was observed, given a candidate fragility function.

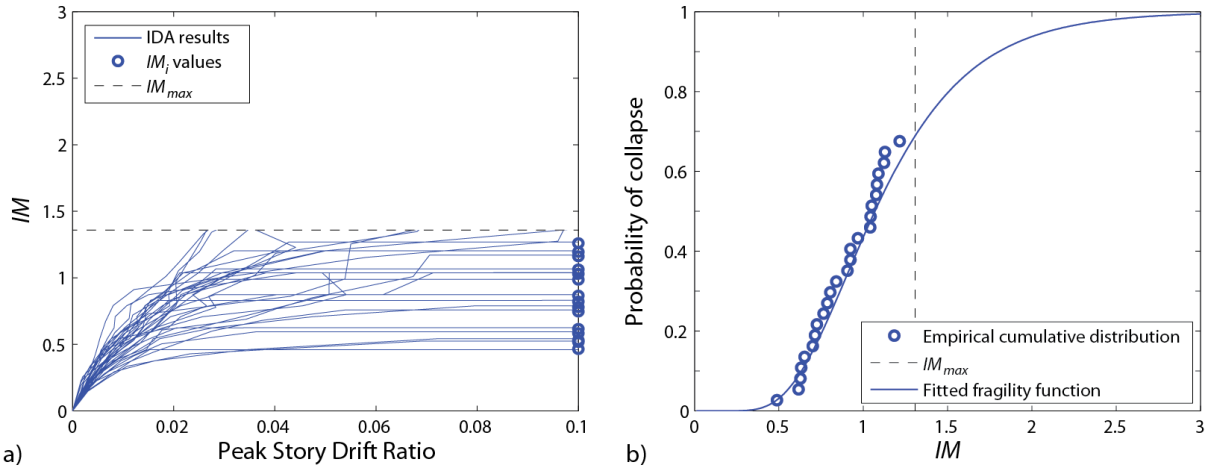

Figure 2. a) Example truncated IDA analysis results. b) Observed fractions of collapse as a function of *IM*, and a fragility function estimated using equation 7.

For the *m* ground motions that were observed to cause collapse, their *IM* values at collapse (IM_i) are known. The likelihood that an arbitrary ground motion causes collapse at IM_i, given a fragility function defined by equation 1, is the normal distribution probability density function (PDF)

$$\text{Likelihood} = \phi\left(\frac{\ln(IM_i / \theta)}{\beta}\right) \tag{4}$$

where $\phi(\ )$ denotes the standard normal distribution PDF. The *n-m* ground motions that did not cause collapse at IM_{max} are called censored data, as we only know that IM_i is greater than IM_{max} (e.g., Klugman et al. 2012, section 15.2.4). The likelihood that a given ground motion can be scaled to IM_{max} without causing collapse is the probability that IM_i is greater than IM_{max}

$$\text{Likelihood} = 1 - \Phi\left(\frac{\ln(IM_{\max} / \theta)}{\beta}\right) \tag{5}$$

Making the reasonable assumption that the IM_i value for each ground motion is independent, the likelihood of the entire data set being observed is the product of the individual likelihoods

$$\text{Likelihood} = \left(\prod_{i=1}^{m} \phi\left(\frac{\ln(IM_i / \theta)}{\beta}\right)\right)\left(1 - \Phi\left(\frac{\ln(IM_{\max} / \theta)}{\beta}\right)\right)^{n-m} \tag{6}$$

where Π denotes a product over *i* values from 1 to *m* (corresponding to the *m* ground motions that caused collapse at *IM* levels less than IM_{max}). Using this equation, the fragility function parameters are then obtained by varying the parameters until the likelihood function is maximized. It is mathematically equivalent and numerically easier to maximize the logarithm of the likelihood function, so in general we do that.

$$\left\{\hat{\theta},\hat{\beta}\right\} = \arg\max_{\theta,\beta} \sum_{i=1}^{m}\left\{\ln\phi\left(\frac{\ln(IM_i/\theta)}{\beta}\right)\right\} + (n-m)\ln\left(1-\Phi\left(\frac{\ln(IM_{\max}/\theta)}{\beta}\right)\right) \quad (7)$$

It is simple to perform this maximization using a spreadsheet or simple computer program, as will be discussed later. A fragility function obtained using this equation is shown in Figure 2b, as obtained using the IM_i and IM_{max} values shown in that figure.

Note that in the special case where all *n* ground motions cause collapse at *IM* values less than IM_{max}, equation 7 has an analytical solution for the values of $\hat{\theta}$ and $\hat{\beta}$ that maximize the equation, and the solution is equivalent to that of equations 2 and 3, except that the "*n*-1" in equation 3 is an "*n*" in this solution (the difference being due to the common choice in equation 3 to use an unbiased variance estimator rather than the maximum likelihood estimator). Note also that the normal distribution PDF and CDF in equation 7 can be replaced with the PDF and CDF of another distribution type, in order to fit a fragility function for some other distribution.

MULTIPLE STRIPES ANALYSIS

Rather than using incremental dynamic analysis, structural analyses are sometimes performed at a discrete set of *IM* levels, and different ground motions are used at each *IM* level. This multiple stripes analysis (MSA) approach is common when using the Conditional Spectrum or other approaches to select ground motions representative of a specific site and *IM* level, because the target properties of the ground motions change at each *IM* level and thus so do the representative ground motions (e.g., Bradley 2010; Iervolino et al. 2010; Lin et al. 2013). With this approach, the analysis need not be performed up to *IM* amplitudes where all ground motions cause collapse. Data of this type is illustrated in Figure 3. Due to the differing ground motions used at each *IM* level, the analyst may not observe strictly increasing fractions of collapse with increasing *IM*, even though it is expected that the true probability of collapse is increasing with *IM*.

With data of this type, we cannot use the estimation approaches described earlier, because we do not have the IM_i values associated with the onset of collapse for a given ground motion.

Instead, the structural analysis results provide the fraction of ground motions at each *IM* level that cause collapse. The appropriate fitting technique for this type of data is to use the method of maximum likelihood, as has been noted by a number of authors (Shinozuka et al. 2000; Baker and Cornell 2005b; Straub and Der Kiureghian 2008). The approach is described briefly here.

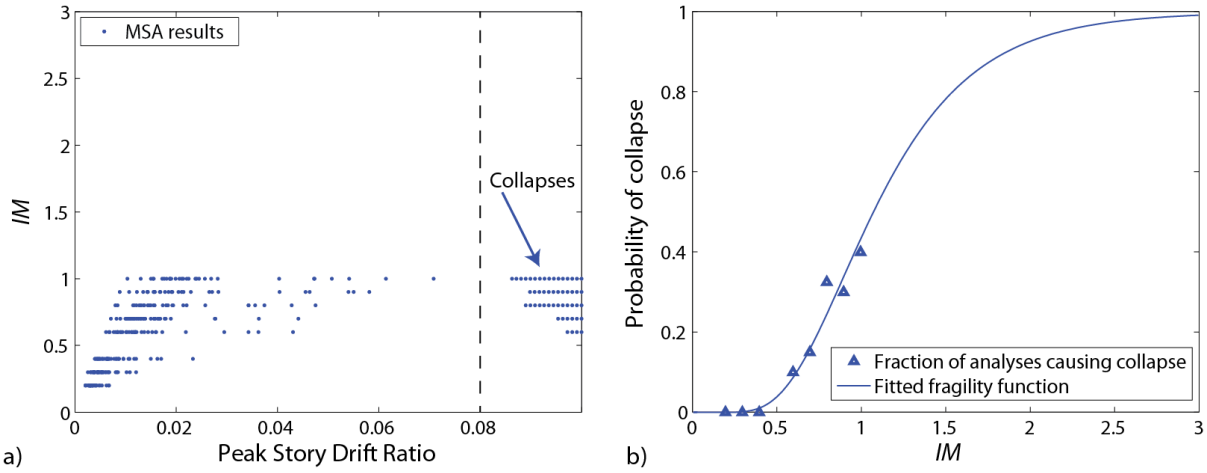

Figure 3. a) Example MSA analysis results. Analyses causing collapse are plotted at Peak Story Drift Ratios of greater than 0.08, and are offset from each other to aid in visualizing the number of collapses. b) Observed fractions of collapse as a function of *IM*, and a fragility function estimated using equation 11.

At each intensity level $IM = x_j$, the structural analyses produce some number of collapses out of a total number of ground motions. Assuming that observation of collapse or no-collapse from each ground motion is independent of the observations from other ground motions, the probability of observing z_j collapses out of n_j ground motions with $IM = x_j$ is given by the binomial distribution

$$P(z_j \text{ collapses in } n_j \text{ ground motions}) = \binom{n_j}{z_j} p_j^{z_j} (1-p_j)^{n_j - z_j} \tag{8}$$

where p_j is the probability that a ground motion with $IM = x_j$ will cause collapse of the structure. Our goal is to identify the fragility function that will predict p_j, and the maximum likelihood approach identifies the fragility function that gives the highest probability of having observed the collapse data that was obtained from structural analysis. When analysis data is obtained at multiple *IM* levels, we take the product of the binomial probabilities (from equation 8) at each *IM* level to get the likelihood for the entire data set

$$\text{Likelihood} = \prod_{j=1}^{m} \binom{n_j}{z_j} p_j^{z_j} (1-p_j)^{n_j - z_j} \tag{9}$$

where m is the number of *IM* levels, and Π denotes a product over all levels. We then substitute equation 1 for p_j, so the fragility parameters are explicit in the likelihood function.

$$\text{Likelihood} = \prod_{j=1}^{m} \binom{n_j}{z_j} \Phi\left(\frac{\ln(x_j/\theta)}{\beta}\right)^{z_j} \left(1-\Phi\left(\frac{\ln(x_j/\theta)}{\beta}\right)\right)^{n_j-z_j} \quad (10)$$

Estimates of the fragility function parameters are obtained by maximizing this likelihood function. It is equivalent and numerically easier to maximize the logarithm of the likelihood function, so we do that

$$\left\{\hat{\theta},\hat{\beta}\right\} = \arg\max_{\theta,\beta} \sum_{j=1}^{m} \left\{ \ln\binom{n_j}{z_j} + z_j \ln\Phi\left(\frac{\ln(x_j/\theta)}{\beta}\right) + \left(n_j - z_j\right)\ln\left(1-\Phi\left(\frac{\ln(x_j/\theta)}{\beta}\right)\right)\right\} \quad (11)$$

A fragility function obtained using this approach is displayed in Figure 3.

A few comments can now be made regarding this formulation. First, equation 11 is written using a lognormal cumulative distribution function for the fragility function, but other functions can be substituted without changing the fitting approach. Second, this formulation does not require multiple observations at each *IM* level of interest (i.e., n_i can equal 1). This makes it useful, for example, when fitting a fragility function using unscaled ground motions, each having unique *IM* amplitudes. Third, this formulation assumes independence of observations, so that the overall likelihood is product of likelihoods at each *IM* level. This independence may not be strictly true if the same ground motion is used for structural analysis at multiple x_i levels, although quantifying this dependence may be somewhat challenging and anecdotal evidence suggests that relaxing the assumption typically makes little numerical difference in the estimated parameters. In fact, example analyses appear to indicate that the approach produces effective fragility estimates even with IDA data, where identical ground motions are used at all *IM* levels. Straub and Der Kiureghian (2008) discuss a generalization of the above approach that allows consideration of dependent samples, though the formulation is more complex to implement.

A numerically equivalent alternative to equation 11 is to use generalized linear regression with a Probit link function to predict the probability of collapse as a function of ln*IM* (Agresti 2012). Generalized linear regression uses maximum likelihood for estimation and the Probit link function is equivalent to using the normal cumulative distribution function as the fragility function. Generalized linear regression is available in many statistical software packages (e.g., R Team 2012; The Mathworks 2012). Logistic regression has also been used for fragility

function calibration using this type of data, and is consistent with maximum likelihood principles, though it fits a logistic rather than lognormal distribution for the fragility (e.g., Basöz and Kiremidjian 1998; Baker and Cornell 2005a).

POTENTIAL ALTERNATIVE METHODS FOR MULTIPLE STRIPES ANALYSIS

One alternative to the use of equation 11 to estimate a fragility function from multiple stripes analysis data would be to minimize the sum of squared errors (SSE) between the observed fractions of collapse and probabilities of collapse predicted by the fragility function. Mathematically, this would be stated

$$\left\{\hat{\theta},\hat{\beta}\right\} = \arg\min_{\theta,\beta} \sum_{j=1}^{m} \left(\frac{z_j}{n_j} - \Phi\left(\frac{\ln(x_j/\theta)}{\beta} \right) \right)^2 \tag{12}$$

where all variables are defined earlier. Example results from this approach are shown in Figure 4. Another alternative would be to fit the function using the "Method B" approach of Porter et al. (2007). This method transforms the observed fractions of collapse so that linear regression can be used to estimate the fragility function parameters. This approach also minimizes an error metric between observations and a fitted function—in this case the sum of squared errors in the transformed space.

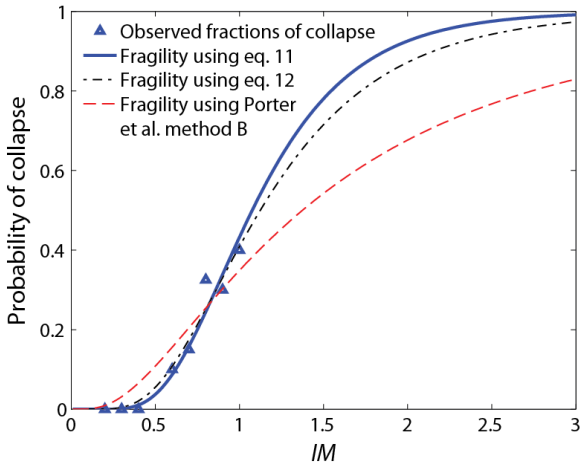

Figure 4. Observed fractions of collapse from Figure 3, and fitted fragility functions obtained using the maximum likelihood approach and two potential alternative methods.

The fragility functions obtained from equations 11, 12 and Porter et al. Method B will in general differ, because the least-squares method ignores a fundamental property of the data: the variance of the observed fractions of collapse is non-constant, in violation of the requirements of least squares fitting (Agresti 2012). For example, if zero collapses are observed at a given *IM* level and the fitted probability of collapse is 0.1, then this error is much larger

than fitting a probability of collapse of 0.6 at an *IM* level where 50% of motions are observed to cause collapse.

The results from these approaches are shown in Figure 4, where the alternate approaches predict relatively higher probabilities of collapse at low *IM* levels that the data suggests are very unlikely to cause collapse. Additionally, the parameter estimates obtained from these approaches will be biased, as can be determined using the assessment procedure proposed in the following section. These alternate approaches are thus not recommended for use in fitting this type of data. They are only mentioned briefly here to illustrate the potential problems with seemingly-reasonable alternative approaches.

EFFICIENT STRATEGIES TO PERFORM STRUCTURAL ANALYSIS FOR FRAGILITY FUNCTION FITTING

For a given set of collapse data obtained from multiple stripes analysis, equation 11 provides the statistically appropriate approach for fragility estimation. But an engineer also needs to choose how to collect structural analysis data for the fragility estimation. There have been approaches proposed for efficiently estimating a fragility function from a small number of analyses (Bradley 2013; Eads et al. 2013), but none that evaluated the proposed approach by studying statistics of the estimated parameters. This section proposes an approach to study how to effectively collect structural analysis data for fragility function fitting.

ANALYSIS APPROACH

To study potential structural analysis strategies, the following Monte Carlo approach is proposed. We first simulate hypothetical structural analysis data from a known fragility function, and then re-estimate that fragility function from the data, using the following steps:

1. Assume a fragility function (i.e., numerical values of θ and β for equation 1).
2. "Collect data" via Monte Carlo simulation by generating collapses or non-collapses at each *IM* level of interest according to the collapse probability given by the fragility function from step 1. After this data is collected, the assumed fragility function is treated as unknown.
3. Using the data from step 2, estimate a fragility function using equation 11.
4. Repeat steps 2 and 3 many times, to measure how similar the estimated fragility functions are to the original assumed fragility function.

These steps are illustrated graphically in Figure 5, for an assumed fragility function with θ = 1 and β = 0.4.

With this approach, the collapse and non-collapse data in step 2 are coming from an assumed fragility function rather than from structural analysis results. This has several advantages for the purpose of this study. First, because the fragility function producing the data is known, we can easily quantify how close the later fragility estimates are to this correct answer. Second, because the collapse and non-collapse data is simulated using a random number generator, it is very fast to produce and thus facilitates studies of a large number of analysis strategies without the computational expense of real structural analysis. Finally, an essentially infinite number of data can be produced, while the finite number of available ground motions would limit the amount of real data that could be obtained from structural analysis. For related reasons, a similar procedure was used recently by Gehl et al. (2013) to quantify the accuracy of structural response estimation strategies.

Once the data have been collected, we must then determine how well the fragility function is being estimated. The following metrics are considered here: the θ and β parameter estimates, and rate-of-collapse estimates, given a ground motion hazard curve and fitted fragility function (as defined below). Desirable properties of our estimates of these metrics are that they be unbiased and efficient. We measure bias by comparing the mean of repeated parameter estimates to the true parameter value being estimated. For example, in Figure 5a, the true median of the fragility curve is 1, and the mean of the θ estimates in Figure 5d is essentially 1, indicating that the estimated θ values are correct on average (i.e., unbiased). When considering analysis strategies below, any potential case that appears to produce a bias is rejected (such cases arise, for example, when one uses only two closely spaced *IM* levels and the resulting data sometimes produces unstable fragility estimates); note that this problem can be mitigated by use of larger numbers of ground motions at each IM level, though the 40 ground motions considered here is already near the upper limit of most industry or research efforts to fit fragility functions. We measure efficiency by computing the standard deviation of repeated parameter estimates. An efficient data collection strategy will lead to a small standard deviation, indicating that parameter estimates will be relatively stable despite the expected record-to-record variation in analysis results. For example, the fragility estimates in Figure 5 were obtained using 40 hypothetical ground motions each at *IM* = {0.6, 1, 1.5}, and resulted in a standard deviation of θ of 0.056. If only 20 ground motions are used at the same three *IM*

levels, the standard deviation of θ increases to 0.078, indicating the tradeoff between the computational expense of doing more analyses and the reduced estimation uncertainty that the additional analyses provide.

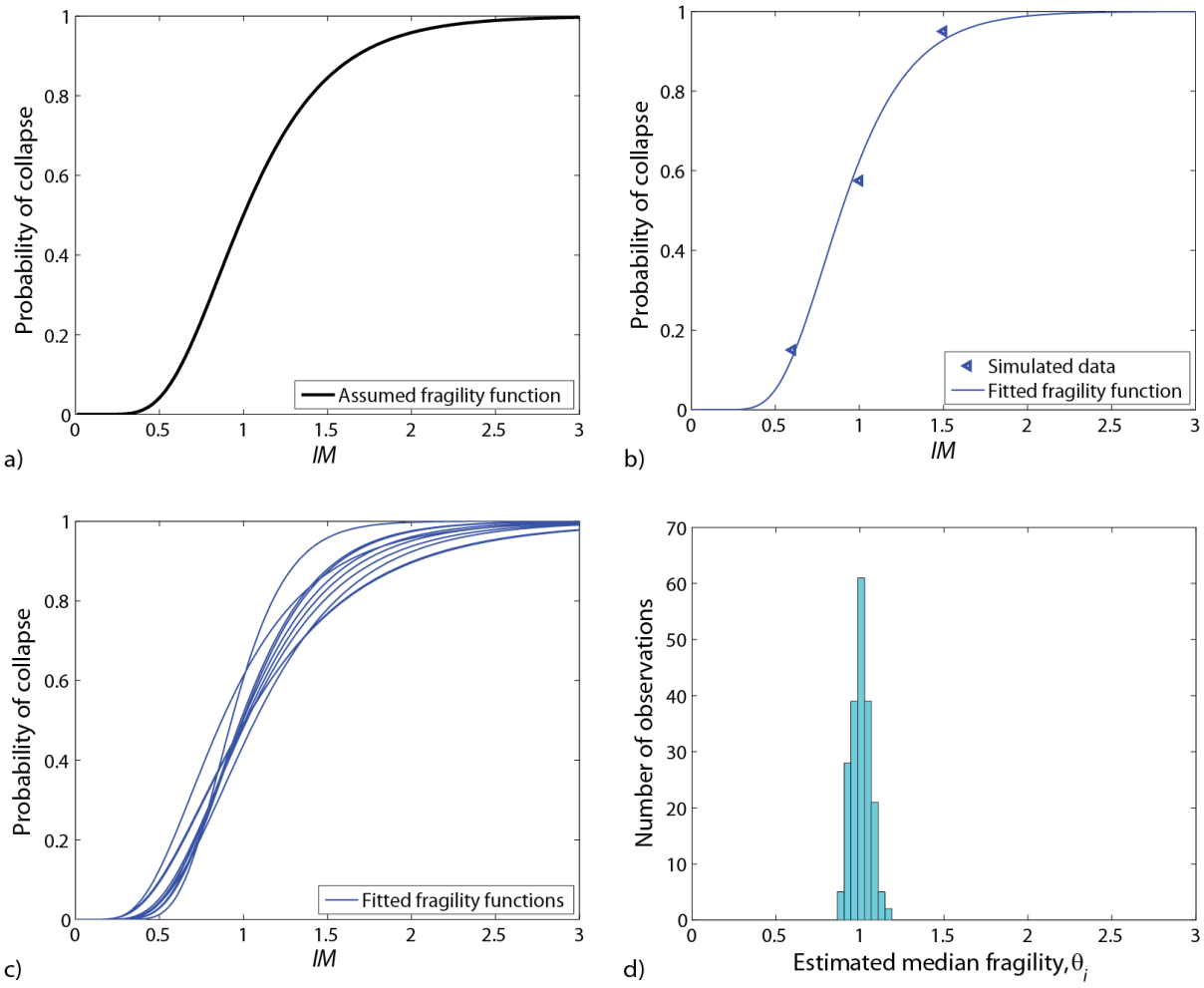

Figure 5. Illustration of procedure used to evaluate structural analysis strategies. (a) Assumed fragility function from step 1. (b) One realization of collapse data at three *IM* levels from step 2, and corresponding estimated fragility function from step 3. (c) Multiple realizations of fitted fragility functions, from step 4. (d) Histogram of estimated θ values from each fitted fragility function.

In the example analyses below (and the analyses shown in Figure 5) the assumed fragility function has $\theta = 1$ and $\beta = 0.4$. The x-axis label on figures in this section is *IM*, but the axis could also be interpreted as IM/θ, or "fraction of median *IM*," for the purposes of translating the results to fragility functions with other medians, so long as β is assumed equal to 0.4 (a common value for collapse fragility functions).

For a given fragility curve, we can also compute the annual rate of collapse of the structure ($\lambda_{collapse}$) using

$$\lambda_{collapse} = \int_x P(C \mid IM = x) \left| d\lambda_{IM}(x) \right| = \int_x \Phi\left(\frac{\ln(x/\theta)}{\beta} \right) \left| d\lambda_{IM}(x) \right| \quad (13)$$

where $\lambda_{IM}(x)$ is a ground motion hazard curve, specifying the annual rate of ground motions with $IM>x$, and $\left| d\lambda_{IM}(x) \right|$ is the absolute value of the derivative of the hazard curve. This calculation requires a hazard curve to be specified. One real hazard curve for downtown Los Angeles is considered here, and is obtained from the U.S. Geological Survey for IM = one-second spectral acceleration on a site with V_{s30} = 760 m/s (Petersen et al. 2008). To generalize the results, idealized power-law hazard curves of the following form are also considered

$$\lambda_{IM}(x) = k_0 \, IM^{-k} \quad (14)$$

The exponent k represents the steepness of the hazard curve. Values of k=2 and k=3 are used to represent typical shapes of spectral acceleration hazard curves observed in seismically active parts of the United States (Yun et al. 2002). The coefficient k_0 scales the overall rate of ground motions and thus the rate of collapse. Here k_0 = 0.0002 and 0.00012 for the k=2 and 3 cases, respectively, are chosen to approximately match the Los Angeles hazard curve, but k_0 does not affect the relative effectiveness of various structural analysis strategies. The three considered hazard curves are shown in Figure 6. For the assumed fragility function considered here, the k=2, k=3 and Los Angeles hazard curves produce probabilities of collapse of 0.013, 0.012 and 0.011 in 50 years, respectively.

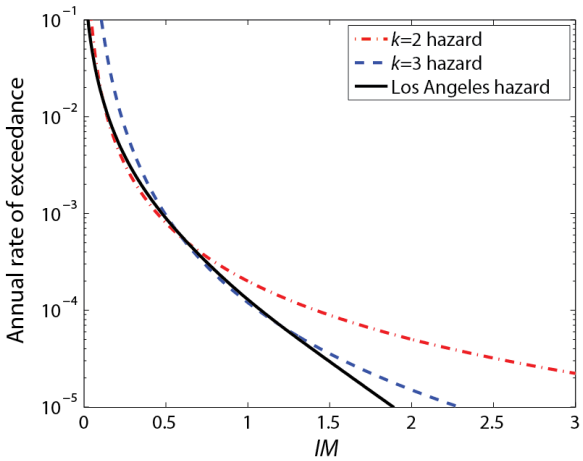

Figure 6. Example ground motion hazard curves considered for $\lambda_{collapse}$ calculations.

ANALYSIS RESULTS

Given the above assumed fragility curves and evaluation metrics, potential structural analysis strategies were studied. Forty analyses at each of three *IM* levels were performed (i.e., it was assumed that 40 ground motions at three *IM* levels would be used when analyzing a real structure). The values of the lower and upper *IM* levels were varied, with the third *IM* level being chosen as the midpoint of the lower and upper value. For each set of potential *IM* levels, 1000 sets of data (i.e., 40 collapse or non-collapse realizations at each of the three *IM* levels) were simulated from the assumed fragility curve, and 1000 corresponding fragility functions were estimated. The 1000 values of θ, β and $\lambda_{collapse}$ were then studied, to see which choices of *IM* levels produced estimates of those parameters with the smallest standard deviations. If a set of candidate *IM* levels produced biased estimates of any of those metrics, they were discarded. Bias was judged to be present if the mean value of the 1000 simulations' metrics differed from the true value by more than 10%. Such cases arose when the IM levels were too closely spaced together or extended far into the ranges where 0 or 100% collapses were predicted, leading to numerically unstable parameter estimates arising in a significant fraction of the 1000 simulations. Of the non-biased *IM* levels, the levels producing the smallest standard deviations for each of the five metrics are shown in Table 1. Figure 7a shows the same values from Table 1, but superimposed on the assumed fragility function to aid in illustrating where the effective *IM* levels fall relative to the fragility function. Additionally, to aid interpretation, the probabilities of collapse at the optimal upper and lower *IM* levels are given in Table 1.

Table 1. Optimal upper and lower *IM* levels for collecting data using three *IM* stripes to estimate five fragility function-related parameters.

	Lower *IM*	P(*C*|lower *IM*)	Upper *IM*	P(*C*|upper *IM*)
θ	0.7	0.2	1.3	0.7
β	0.7	0.2	2.0	0.96
$\lambda_{collapse}$ ($k = 2$)	0.4	0.02	0.9	0.4
$\lambda_{collapse}$ ($k = 3$)	0.4	0.02	1.0	0.5
$\lambda_{collapse}$ (L.A.)	0.4	0.02	0.9	0.4

While these results are specifically relevant only for the number of ground motions and IM levels considered here, and to the case where a lognormal fragility function is assumed, a few general observations can be made. When estimating θ and β, the effective *IM* values are approximately centered around the median of the fragility function. The β parameter is most efficiently estimated by considering a relatively wider range of *IM*, which makes intuitive sense

given that this parameter quantifies the slope of the fragility function and slopes are well estimated using widely spaced data points. The *IM* levels producing optimal estimates of the $\lambda_{collapse}$ values are located towards the left side ("lower tail") of the fragility function. This is because ground motions with low *IM* levels occur much more frequently than ground motions with high *IM* levels, and so when quantifying the rate of collapse it is more important to accurately estimate the left side of the fragility function than the right side.

Table 2. Optimal upper and lower *IM* levels for collecting data using two *IM* stripes to estimate five fragility function-related parameters.

	Lower *IM*	P(*C*|lower *IM*)	Upper *IM*	P(*C*|upper *IM*)
θ	0.7	0.2	1.3	0.7
β	0.6	0.1	1.8	0.9
$\lambda_{collapse}$ ($k = 2$)	0.6	0.1	1.3	0.7
$\lambda_{collapse}$ ($k = 3$)	0.6	0.1	1.6	0.9
$\lambda_{collapse}$ (L.A.)	0.6	0.1	1.3	0.7

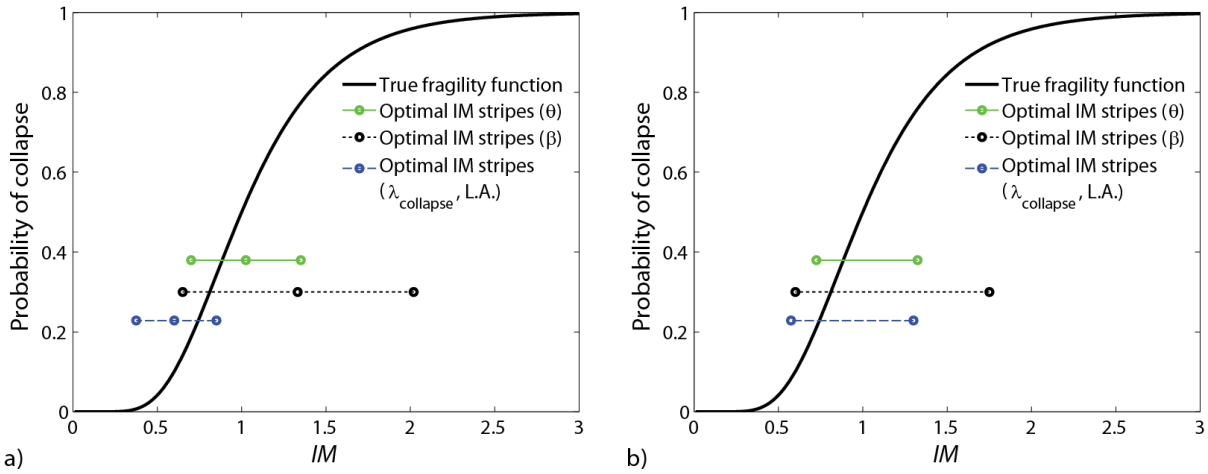

Figure 7. *IM* levels for collecting data that most efficiently estimates fragility function-related parameters, superimposed on fragility function a) with three *IM* stripes, b) with two *IM* stripes. The y-axis values associated with the optimal IM stripes are arbitrary, and have been offset for each case to aid in viewing.

To study an alternate analysis strategy, the procedure was performed again using only two *IM* levels for data collection rather than three. Again 40 analyses at each *IM* level were considered, and the optimal *IM* levels with respect to the various parameter values were determined. These optimal *IM* levels are summarized in Table 2 and Figure 7b. Relative to the comparable results with three *IM* levels, several observations can be made. When estimating θ and β, the effective *IM* upper and lower values are nearly unchanged from the case then

three *IM* levels are used. For the $\lambda_{collapse}$ estimates, the optimal upper and lower *IM* levels shift to the right relative to the three-*IM* case. This is presumably because in the previous case the lower *IM* level produced almost zero probability of collapse; observing zero or very few collapses at a given *IM* level is a good constraint on a fragility function, as long as there are two or more other *IM* levels for which collapses are observed. When only two *IM* levels are considered, however, then if one *IM* level produces no collapses then there is no unique solution for the fragility curve.

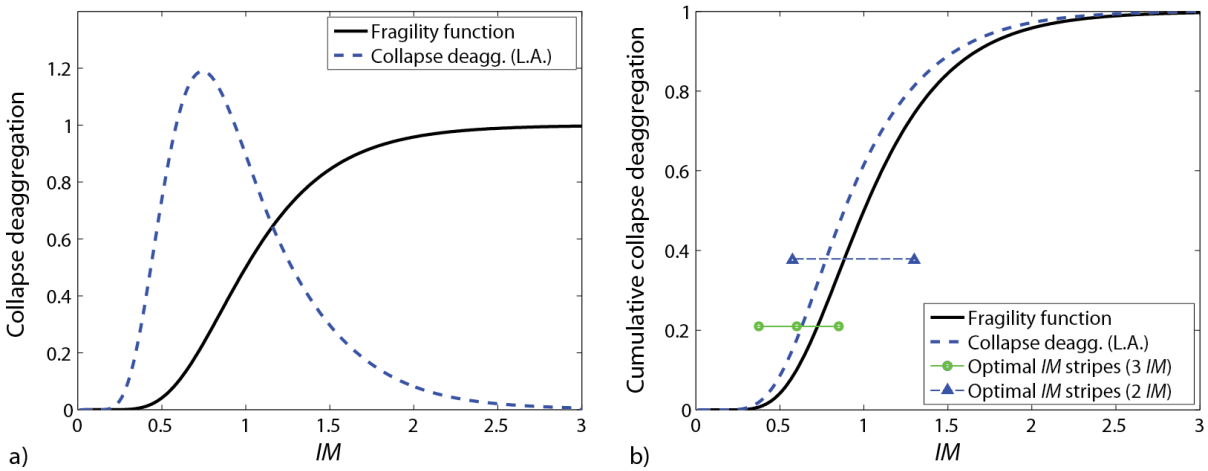

Figure 8. a) Collapse deaggregation plot for the Los Angeles hazard curve considered, with the fragility function shown for reference. b) Cumulative collapse deaggregation plot, with the fragility function shown for reference and the optimal *IM* levels from above noted on the deaggregation plot. The y-axis values associated with the optimal IM stripes are arbitrary, and have been offset for each case to aid in viewing.

Another way to evaluate the results from the $\lambda_{collapse}$ estimation is to study the collapse deaggregation plots, as illustrated in Figure 8. The collapse deaggregation is the probability distribution of *IM* levels contributing to collapse, as computed by

$$P(x \le IM < x + dx \mid C) / dx = \frac{P(C \mid IM = x)\left|d\lambda_{IM}(x)\right|}{\lambda_{collapse}} \tag{15}$$

A cumulative distribution of these deaggregated values can also be obtained by integrating this result from 0 to the *IM* level of interest, and results of this type are shown in Figure 8b. Deaggregation is a simple way to identify important *IM* levels, as the numerator in equation 15 is large for *IM* levels where both the probability of collapse and probability of IM occurrence is large, and the denominator is simply a normalizing constant. This calculation requires the fragility function and a hazard curve, so in Figure 8 results are shown for the Los Angeles hazard curve. We see that the highest values of the deaggregation plot are to the left of the median, indicating that the most important *IM* levels (and thus the most important portion of

the fragility curve to estimate accurately) fall below the median of the fragility curve. This is expected, because the probability of observing a ground motion with a given *IM* level drops quickly as the *IM* level increases. By comparing these deaggregation plots to the results of Table 1 and Table 2, we see that the optimal *IM* levels for effectively estimating $\lambda_{collapse}$ correspond to *IM* levels for which the collapse deaggregation is high. This comparison is motivated by the work of Eads et al. (2013), who proposed a strategy for choosing *IM* levels for data collection and fragility fitting relative to collapse deaggregation levels. That study proposed using two *IM* levels for collapse deaggregation, corresponding to the estimated 35% and 90% points on the cumulative collapse deaggregation plots. Figure 8b indicates that the optimal two *IM* levels for the L.A. hazard curve case fall at approximately the 15% and 80% points on the cumulative collapse deaggregation. Similar percentages were observed for the other hazard curves as well—10% to 70% and 20% to 90% for the *k*=2 and *k*=3 hazard curve cases, respectively. When three or more *IM* levels are considered, the optimal percentage levels decrease significantly, as indicated in Figure 8b.

EVALUATION OF IDA AND MSA ANALYSIS STRATEGIES

The proposed evaluation approach can also be used to compare the effectiveness of incremental dynamic analysis and multiple stripes analysis approaches for collecting data. We will consider five strategies, again using an assumed fragility function with $\theta = 1$ and $\beta = 0.4$. For each strategy, we perform 1000 Monte Carlo simulations of synthetic structural analysis results to obtain 1000 sets of estimated fragility parameters, and then study the variability of the metrics resulting from the fitted fragilities.

The first strategy is to do incremental dynamic analysis with 20 ground motions. The IDA's are assumed to be performed with an *IM* step size of 0.1 (i.e., "analyses" are performed at *IM* = 0.1, 0.2, etc., until collapse is observed). The true collapse *IM* level for each ground motion is determined using Monte Carlo simulation from a lognormal random variable with $\theta = 1$ and $\beta = 0.4$. The observed collapse *IM* for each IDA analysis is then taken as the midpoint of the steps before and after collapse is observed, and equations 2 and 3 are used to estimate the fragility function parameters. The second strategy is to do truncated IDA with 20 ground motions, and stop performing analyses at the *IM* level for which half of the ground motions have been observed to cause collapse. The true collapse *IM* for each ground motion is obtained using the same Monte Carlo approach, the IDA's are again assumed to have the same 0.1 step size, and equation 7 is used to estimate the fragility function parameters. The final three

strategies use multiple stripes analysis. Two or three *IM* levels are considered, and the *IM* levels are chosen to represent the generally effective *IM* levels identified in Table 1 and Table 2.

For each of these strategies, 1000 sets of fragilities were estimated, the five metrics of interest were computed, and the coefficient of variation (COV) of each metric's 1000 estimates was computed. The results are summarized in Table 3, along with the number of structural analyses that each strategy would require. For the two IDA analysis strategies, the number of required analyses will vary depending upon the particular *IM* levels at which a given ground motion causes collapse, so the mean number of analyses is reported in Table 3.

A few observations can be made from the COV values shown in Table 3. The multiple stripes strategy with 45 ground motions at three *IM* levels is most effective with regard to all five metrics, as it produces the lowest COV values for all five parameters. This is presumably because all of the analyses are focused at *IM* levels where structural analysis results will produce strong constraints on the fragility function, unlike IDA, where a significant number of the analyses are performed for $IM<0.5$, which for this fragility function is very unlikely to produce collapses. The number of analyses can be further reduced, by either reducing the number of stripes or reducing the number of ground motions, without dramatically increasing the COVs of the estimated parameters, as can be seen from the results of the fourth and fifth strategy. Truncating the IDA analyses after 50% of the motions cause collapse is not a particularly effective strategy, as the number of analyses is reduced by only 20% (since all of the ground motions are used for analysis a number of times even if they are not analyzed up to their collapse levels), but the uncertainty in the all of the metrics besides θ increases appreciably.

Table 3: Computation expense of five analysis strategies and associated standard errors of the five metrics considered. For the IDA results, the Number of Analyses is the mean number of required analyses.

	IDA with 20 ground motions	Truncated IDA with 20 motions	45 motions at *IM* = [0.4 0.8 1.2]	30 motions at *IM* = [0.4 0.8 1.2]	45 motions at *IM* = [0.5, 1.2]
Number of analyses	227	184	135	90	90
COV of θ	0.09	0.10	0.06	0.08	0.07
COV of β	0.16	0.26	0.20	0.26	0.40
COV of $\lambda_{collapse}$ (k=2)	0.22	0.39	0.15	0.20	0.24
COV of $\lambda_{collapse}$ (k=3)	0.38	0.57	0.33	0.55	0.51
COV of $\lambda_{collapse}$ (L.A.)	0.29	0.65	0.20	0.27	0.33

From these among these considered options, it appears that an effective strategy for fragility fitting is to focus structural analyses at a limited number of *IM* levels using multiple stripes analysis, rather than doing incremental dynamic analysis. This observation is consistent with the recommendations of Eads et al. (2013). That strategy also has the benefit of allowing the analyst to use different ground motions at each *IM* level, which allows for more accurate representation of ground motion properties (Baker and Cornell 2005a). The results suggest that choosing *IM* levels near the lower tail of the fragility function and up to *IM* levels slightly above the median would be a generally effective strategy.

The obvious limitation of these results is that the *IM* levels that provide good constraints on the fragility curve are defined relative to the fragility curve itself, but for real structural analysis applications the fragility curve is unknown. Nonetheless, an approximate fragility function can be predicted prior to analysis and then used with the above results to estimate effective *IM* levels for analysis (Eads et al. 2013). If we make an accurate prediction of the fragility then the above results are relevant in indicating effective analysis strategies. But if we make a poor prediction of the fragility then we might perform analyses at *IM* levels that are not as informative as other level. It should be noted, however, that multiple stripe analysis strategies using slightly higher or lower *IM* levels than those in Table 3 tend to produce similar results, indicating that there is some margin for error in predicting the fragility curve while still identifying potentially informative *IM* levels for analysis. Another effective way to add robustness to the strategy in the face of an unknown fragility is to do an initial set of analyses at an estimated *IM* level, use those results to update the estimated fragility and identify the next *IM* level for analysis based on that updated estimate, as proposed by Eads et al. (2013). Another effective way to build robustness in the face of an unknown fragility is to spread the analyses over a greater number of *IM* levels; the fourth and fifth cases above indicate that using the same number of analyses spread over three instead of two *IM* levels results in similar statistical estimation efficiency but increases the range of levels at which information is gathered. Use of more than three *IM* levels may also be useful in real applications, in order to ensure useful data at a number of *IM* levels in the case where the capacity of the structure is over- or under-estimated and some *IM* levels produce 0% or 100% collapses.

There are obviously many other data collection approaches that could be considered, some of which may produce more efficient estimates of fragility functions than the ones discussed above, such as using different numbers of ground motions at each *IM* stripe, or using a hunt-

and-seek approach with IDA (Vamvatsikos and Cornell 2002). Identifying effective strategies requires identifying strategies that are optimal from a statistical inference perspective but also practical for an engineer to implement. For example, it is practically easier to select all ground motions prior to beginning structural analysis, which is more difficult when *IM* levels are adaptively updated. A similar implementation issue arises with IDA; while the efficient hunt-and-seek strategy was proposed more than 10 years ago, many users of IDA use the slightly simpler but more expensive approach of scaling ground motions up by a fixed *IM* increment until collapse is observed.

Despite the limitations of the strategies considered here, these results do provide insights regarding the characteristics of effective structural analysis strategies. Further, the proposed approach provides a framework for quantitatively testing the effectiveness of any potential strategy prior to implementing it with a computationally expensive structural analysis model.

CONCLUSIONS

This paper discusses the applicability of statistical inference concepts for fragility function fitting, identifies appropriate fitting approaches for different data collection strategies, and illustrates how one might fit fragility functions using an approach that minimizes the required number of structural analyses. First, incremental dynamic analysis and multiple stripe analysis approaches for data collection were discussed, and corresponding statistically appropriate methods for fragility function fitting were described. Next, an approach was proposed to evaluate the efficiency of those analysis approaches for estimating a fragility function. The proposed approach involved repeatedly simulating synthetic structural analysis data that was consistent with an assumed fragility function, and then studying fragility functions estimated from the simulated data. Unlike most previous studies of fragility function fitting, which used real data from dynamic structural analysis, this study used synthetic data simulated from assumed distributions. This approach is useful, as the true fragility function from which the data came is known, and so strategies can be evaluated to verify that they efficiently produce fragility estimates close to the true answer. The use of synthetic data also allows for much greater numbers of results to be produced, so that more extensive statistical studies can be performed.

This study then compared the efficiency of collapse fragilities obtained from incremental dynamic analysis and multiple stripes analysis. Multiple stripes analysis is seen to be more

efficient than incremental dynamic analysis, because analyses can be targeted at a limited number of important *IM* levels, rather than requiring analysis at high or low *IM* levels associated with collapse for some ground motions but not as critical for constraining the fragility function. This finding is useful from a statistical perspective, but also beneficial from a structural modeling perspective because multiple stripes analysis allows for ground motions selected using the Conditional Spectrum and other approaches that require different ground motions at each *IM* level. The results also indicated that fragility functions may be efficiently estimated by focusing on *IM* levels at which there are lower probabilities of collapse, saving analysis time and focusing the fitting on the region of the fragility most important for risk assessments.

These results provide guidance to efficiently estimate fragility functions from a small number of ground motions and structural analyses. The fitting procedures are easy to implement, and several simple software tools have been provided to facilitate their use.

SOFTWARE TOOLS

To facilitate adoption of the parameter estimation approaches described above, simple software tools have been created to demonstrate the calculations. The user needs only provide observed data, and the software tools will numerically estimate fragility function parameters using the approaches described in this paper. The procedures have been implemented in both Excel and Matlab, and are available at *http://purl.stanford.edu/sw589ts9300*.

ACKNOWLEDGEMENTS

Thanks to Greg Deierlein, Iunio Iervolino, Ting Lin, Eduardo Miranda, Beliz Ugurhan and an anonymous reviewer for helpful feedback. This work was supported by the National Science Foundation under NSF grant number CMMI 0726684, and by the NEHRP Consultants Joint Venture (a partnership of the Consortium of Universities for Research in Earthquake Engineering and Applied Technology Council), under Contract SB134107CQ0019, Earthquake Structural and Engineering Research, issued by the National Institute of Standards and Technology, for project ATC-82. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect those of the sponsors.

REFERENCES

Agresti, A. (2012). *Categorical data analysis*. Wiley, New York.

Applied Technology Council. (2012). *ATC-58, Seismic Performance Assessment of Buildings, 100% Draft*. Applied Technology Council, Redwood City, California, 266p.

Aslani, H., and Miranda, E. (2005). "Fragility assessment of slab-column connections in existing non-ductile reinforced concrete buildings." *Journal of Earthquake Engineering*, 9(6), 777–804.

Baker, J. W., and Cornell, C. A. (2005a). "A vector-valued ground motion intensity measure consisting of spectral acceleration and epsilon." *Earthquake Engineering & Structural Dynamics*, 34(10), 1193–1217.

Baker, J. W., and Cornell, C. A. (2005b). *Vector-valued ground motion intensity measures for probabilistic seismic demand analysis*. Report No. 150, John A. Blume Earthquake Engineering Center, Stanford, CA, 321p.

Basöz, N. I., and Kiremidjian, A. S. (1998). "Evaluation of bridge damage data from the Loma Prieta and Northridge, California earthquakes." *Technical Report MCEER*, (98-0004).

Bradley, B. A. (2010). "A generalized conditional intensity measure approach and holistic ground-motion selection." *Earthquake Engineering & Structural Dynamics*, 39(12), 1321–1342.

Bradley, B. A. (2013). "Practice-oriented estimation of the seismic demand hazard using ground motions at few intensity levels." *Earthquake Engineering & Structural Dynamics*, 42(14), 2167–2185.

Bradley, B. A., and Dhakal, R. P. (2008). "Error estimation of closed-form solution for annual rate of structural collapse." *Earthquake Engineering & Structural Dynamics*, 37(15), 1721–1737.

Calvi, G. M., Pinho, R., Magenes, G., Bommer, J. J., Restrepo-Vélez, L. F., and Crowley, H. (2006). "Development of seismic vulnerability assessment methodologies over the past 30 years." *ISET journal of Earthquake Technology*, 43(3), 75–104.

Eads, L., Miranda, E., Krawinkler, H., and Lignos, D. G. (2013). "An efficient method for estimating the collapse risk of structures in seismic regions." *Earthquake Engineering & Structural Dynamics*, 42(1), 25–41.

Federal Emergency Management Agency. (2009). *Quantification of Building Seismic Performance Factors (FEMA P695, ATC-63)*. FEMA P695, prepared by the Applied Technology Council, 421p.

Gehl, P., Douglas, J., and Seyedi, D. (2013). "Influence of the number of dynamic analyses on the accuracy of structural response estimates." *Earthquake Spectra*, (in review).

Ghafory-Ashtiany, M., Mousavi, M., and Azarbakht, A. (2010). "Strong ground motion record selection for the reliable prediction of the mean seismic collapse capacity of a structure group." *Earthquake Engineering & Structural Dynamics*, n/a–n/a.

Haselton, C. B., and Deierlein, G. G. (2007). *Assessing seismic collapse safety of modern reinforced concrete moment frame buildings*. Pacific Earthquake Engineering Research Center Technical Report 2007/08, Berkeley, CA.

Ibarra, L. F., and Krawinkler, H. (2005). *Global collapse of frame structures under seismic excitations*. John A. Blume Earthquake Engineering Center, Stanford, CA, 324.

Iervolino, I., Giorgio, M., Galasso, C., and Manfredi, G. (2010). "Conditional Hazard Maps for Secondary Intensity Measures." *Bulletin of the Seismological Society of America*, 100(6), 3312–3319.

Jalayer, F. (2003). *Direct Probabilistic Seismic Analysis: Implementing Non-Linear Dynamic Assessments*. Ph.D. Thesis, Dept. of Civil and Environmental Engineering, Stanford University, Stanford, CA.

Kennedy, R. P., and Ravindra, M. K. (1984). "Seismic Fragilities for Nuclear Power Plant Risk Studies." *Nuclear engineering and design: an international journal devoted to the thermal, mechanical and structural problems of nuclear energy*, 79, 47–68.

Kim, S.-H., and Shinozuka, M. (2004). "Development of fragility curves of bridges retrofitted by column jacketing." *Probabilistic Engineering Mechanics*, 19(1–2), 105–112.

Klugman, S. A., Panjer, H. H., and Willmot, G. E. (2012). *Loss models: from data to decisions*. John Wiley & Sons.

Liel, A. B., and Deierlein, G. G. (2008). *Assessing the collapse risk of California's existing reinforced concrete frame structures: Metrics for seismic safety decisions*. John A Blume Earthquake Engineering Center Report #166, Stanford University, Stanford, CA.

Lin, T., Haselton, C. B., and Baker, J. W. (2013). "Conditional spectrum-based ground motion selection. Part I: Hazard consistency for risk-based assessments." *Earthquake Engineering & Structural Dynamics*, 42(12), 1847–1865.

Petersen, M. D., Frankel, A. D., Harmsen, S. C., Mueller, C. S., Haller, K. M., Wheeler, R. L., Wesson, R. L., Zeng, Y., Boyd, O. S., Perkins, D. M., Luco, N., Field, E. H., Wills, C. J., and Rukstales, K. S. (2008). "Documentation for the 2008 update of the United States national seismic hazard maps." *US Geological Survey Open-File Report 2008–1128*, Open-File Report 2008–1128, 1128, 61p.

Porter, K., Kennedy, R., and Bachman, R. (2007). "Creating Fragility Functions for Performance-Based Earthquake Engineering." *Earthquake Spectra*, 23(2), 471–489.

R Team. (2012). "The R project for statistical computing." <http://www.r-project.org/>.

Rice, J. A. (1995). *Mathematical statistics and data analysis*. Duxbury Press, Belmont, CA.

Shafei, B., Zareian, F., and Lignos, D. G. (2011). "A simplified method for collapse capacity assessment of moment-resisting frame and shear wall structural systems." *Engineering Structures*, 33(4), 1107–1116.

Shinozuka, M., Feng, M. Q., Lee, J., and Naganuma, T. (2000). "Statistical analysis of fragility curves." *Journal of Engineering Mechanics*, 126(12), 1224–1231.

Shome, N. (1999). *Probabilistic seismic demand analysis of nonlinear structures*. Department of Civil and Environmental Engineering, Stanford University.

Straub, D., and Der Kiureghian, A. (2008). "Improved seismic fragility modeling from empirical data." *Structural Safety*, 30(4), 320–336.

The Mathworks. (2012). "Matlab 2012b Release, Statistics toolbox."

Vamvatsikos, D., and Cornell, C. A. (2002). "Incremental dynamic analysis." *Earthquake Engineering & Structural Dynamics*, 31(3), 491–514.

Vamvatsikos, D., and Cornell, C. A. (2004). "Applied Incremental Dynamic Analysis." *Earthquake Spectra*, 20(2), 523–553.

Villaverde, R. (2007). "Methods to Assess the Seismic Collapse Capacity of Building Structures: State of the Art." *Journal of Structural Engineering*, 133(1), 57–66.

Yun, S.-Y., Hamburger, R. O., Cornell, C. A., and Foutch, D. A. (2002). "Seismic Performance Evaluation for Steel Moment Frames." *Journal of Structural Engineering*, 128(4), 12.

Zareian, F., and Krawinkler, H. (2007). "Assessment of probability of collapse and design for collapse safety." *Earthquake Engineering & Structural Dynamics*, 36(13), 1901–1914.