13th World Conference on Earthquake Engineering
Vancouver, B.C., Canada
August 1-6, 2004
Paper No. 3384

CHOICE OF A VECTOR OF GROUND MOTION INTENSITY MEASURES FOR SEISMIC DEMAND HAZARD ANALYSIS

Jack W. BAKER[1], C. Allin CORNELL[2]

SUMMARY

Among the challenges of performance-based earthquake engineering is estimation of the demand on a structure from a given earthquake. This is often done by calculating an "Intensity Measure" (*IM*) for a given earthquake ground motion, and then computing the probability that the earthquake will cause a given level of demand in the structure as a function of this *IM*. Traditional *IM*s include peak ground acceleration and spectral acceleration at the first-mode period of vibration. These *IM*s consist of a single parameter. In this paper, *IM*s consisting of two parameters are proposed: spectral acceleration at the first-mode period of vibration along with a measure of spectral shape (the ratio of spectral acceleration at a second period to the original spectral acceleration value). A method for predicting the probability distribution of demand using a vector *IM* is presented. This method accounts for the effect of collapses on the distribution of demand. Two complimentary methods for determining the optimum second period at a given intensity level are described, and an improvement in the accuracy of demand predictions is shown.

INTRODUCTION

In this paper we present a method for determining an optimal vector "Intensity Measure" for predicting the demand in a structure. The demand on a structure can be defined as a measure of the structural response (e.g. the maximum inter-story drift angle seen in the structure). To predict this response effectively, one needs an Intensity Measure (*IM*) for the given earthquake. In the past, the Peak Ground Acceleration (*PGA*) of the earthquake was commonly used as an *IM*. More recently, spectral response values (i.e. spectral acceleration at the first-mode period of vibration – $S_a(T_1)$) have been used as *IM*s. These *IM*s are generally a single parameter. In this paper, we propose intensity measures containing two parameters. These intensity measures are called "Vector *IM*s," as opposed to the "Scalar *IM*s" that contain only a single parameter. One would expect that a Vector *IM* would contain more information about the ground motion than a Scalar *IM*, and would thus be more effective at predicting the response of a structure. This will be shown, and criteria for finding an optimal set of two parameters will be described.

An example of the need for prediction of structural response is seen in the work of the Pacific Earthquake Engineering Research (PEER) Center (Cornell [1]). Here, the response of a structure is termed an

[1] Graduate Student, Dept. of Civil and Env. Eng., Stanford University, Stanford, USA. E-mail: bakerjw@stanford.edu
[2] Professor, Dept. of Civil and Env. Eng., Stanford University, Stanford, USA.

Engineering Demand Parameter, or *EDP*. The annual frequency of exceeding a given *EDP* is calculated as follows:

$$\lambda_{EDP}(z) = \sum_{\text{all } x_i} P(EDP > z \mid IM = x_i) \cdot \Delta\lambda_{IM}(x_i) \quad (1)$$

Where $\lambda_{EDP}(z)$ is the annual frequency of exceeding a given *EDP* value z, $\lambda_{IM}(x_i)$ is the annual frequency of exceeding a given *IM* value x_i (this is commonly referred to as a ground motion hazard curve), and $\Delta\lambda_{IM}(x_i) = \lambda_{IM}(x_i) - \lambda_{IM}(x_{i+1})$ is approximately the annual frequency of $IM = x_i$. The final element of this equation is $P(EDP>z|IM=x_i)$, the probability of exceeding a specified *EDP* level, given a level of *IM*. This is the value that we are interested in estimating efficiently using a vector *IM*. If the *IM* becomes a vector, Equation 1 must be generalized, as will be discussed below.

In this paper, we will consider only a specific class of candidates as potential vector *IM*s. Given that $S_a(T_1)$ has been verified as an effective predictor of structural response for a wide class of structures, we will always use this as the first element of our vector. For the second element of our vector, we will consider the predictor $R_{T1,T2} = S_a(T_2)/ S_a(T_1)$ (see Figure 1 for an illustration). The predictor $R_{T1,T2}$ is a measure of spectral shape. Together the vector $S_a(T_1)$ and $R_{T1,T2}$ define two points on the spectrum of an accelerogram. We will keep T_1 equal to the first-mode period of vibration of the building, but we will let T_2 vary, and choose the value that optimally predicts the response of the structure. (At an early stage in this research, a range of T_1 values was examined, but other values did not show any significant improvement when compared with the first-mode period of the structure.) The spectral shape predictor $R_{T1,T2}$ has been found by others to be a useful predictor of structural response (e.g. Cordova [2] and Vamvatsikos [3]). In this study, we will limit the size of the vector to two elements, although the same procedure is easily applicable to vectors of a larger size. This study also considers only maximum interstory drift (referred to as θ_{max}) as an *EDP*, although an identical study could be performed on any other *EDP* value.

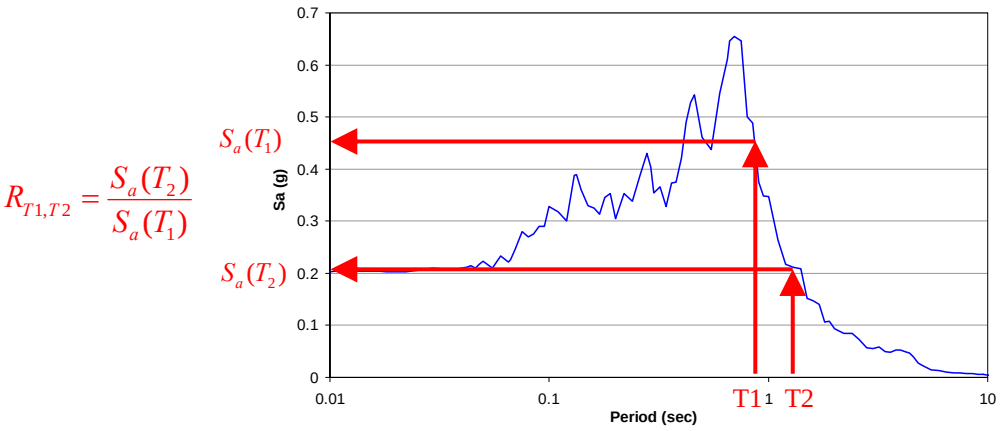

Figure 1: Illustration of the calculation of $R_{T1,T2}$ using a response spectrum.

PREDICTION OF BUILDING RESPONSE USING A SCALAR IM

The prediction of building response requires estimation of the term $P(EDP>z|IM=x_i)$ in Equation 1. This is true for both the scalar and vector *IM* cases. In this section, an estimation method is presented for the scalar *IM* case, and in the next section it is modified for use with vector *IM*s. The scalar *IM* $S_a(T_1)$ is used in this section, both because of its wide use elsewhere, and because it will be easily generalized to our vector case, $IM = \{S_a(T_1), R_{T1,T2}\}$.

The method used in this paper requires a suite of earthquake of records, all at the same *IM* value, $S_a(T_1)=x$ (e.g., in this study, 40 records are used at each *IM* level). We scale a suite of historical earthquake records to the given $S_a(T_1)$ value (e.g. Shome [4]). (We use the same suite of records for different $S_a(T_1)$ levels, although one could use different record suites at different levels if PSHA disaggregation suggested that, for example, the representative magnitude level was changing.) This suite of records is used to perform Nonlinear Dynamic Analysis (NDA) on a model of the structure. Now we have *n* records, all with $IM = x$, and *n* corresponding values of *EDP* (see an illustration in Figure 2). If we had a perfect *IM*, then all records with $IM = x$ would have an identical value of *EDP*. However, with our less-than-perfect *IM*s, there will be a distribution of *EDP*'s. So in fact, *EDP* given $IM = x$ is a random variable with an unknown distribution, and our *n* values of *EDP* are a sample from this distribution. We then need to estimate this distribution in order to calculate the probability that *EDP* is greater than a given value. Our *EDP*, maximum interstory drift, has been found to be well represented by a lognormal distribution (e.g. Shome [5], Aslani [6]). Because of this we work with the natural logarithm of *EDP*, which then has the normal distribution. We can estimate the parameters for this normal distribution using the method of moments (Benjamin [7]). For each *IM* level, estimate the mean of ln*EDP* as the sample average of the ln*EDP*s, and denote this $\hat{\mu}_{\ln EDP|IM=x}$. Estimate the standard deviation of ln*EDP*, which we call the "dispersion" and denote $\hat{\beta}_{\ln EDP|IM=x}$, as equal to the sample standard deviation. These two parameters fully define the normal distribution. The probability that *EDP* exceeds *z* given $IM = x$ can now be calculated using the Gaussian Complimentary Cumulative Distribution Function:

$$P(EDP > z \mid IM = x) = 1 - \Phi\left(\frac{\ln z - \hat{\mu}_{\ln EDP|IM=x}}{\hat{\beta}_{\ln EDP|IM=x}}\right) \tag{2}$$

where $\Phi(\bullet)$ denotes the standard normal distribution.

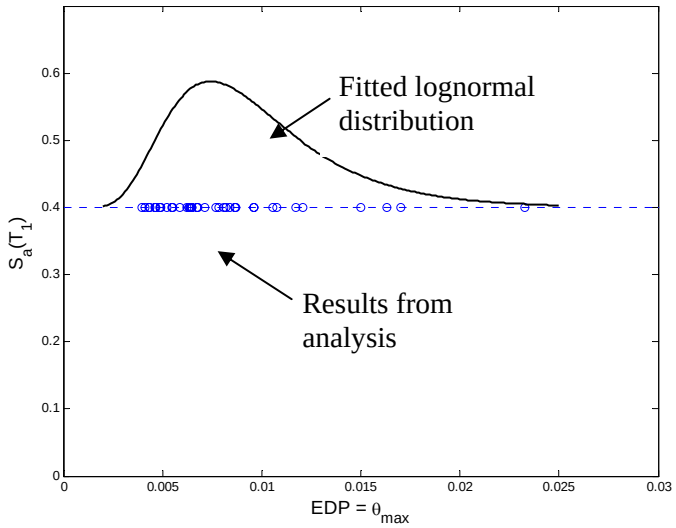

Figure 2: Forty records scaled to $S_a(T_1) = 0.4g$, and a superimposed lognormal probability distribution function, generated using the method of moments.

Accounting for Collapses

In the previous section, it was stated that *EDP* given $IM = x$ has a lognormal distribution. While this may be a valid assumption at lower ground motion levels, it does not explicitly account for the possibility that some records may result in a collapse of the structure at higher levels of *IM*. For these records, we would say that EDP_i is greater than *z* for any value of *z* (or equivalently, $EDP_i = \infty$). This causes two problems:

the probability that $EDP_i = \infty$ is zero in the lognormal distribution, and collapses cause our estimates of the lognormal mean and standard deviation to be infinite. To address this issue, we need to make a modification to our procedure. First, we separate our realizations of *EDP* into collapsed and non-collapsed data. We then estimate the probability of collapse at the given *IM* level as:

$$P(C \mid IM = x) = \frac{\text{number of records collapsed}}{n} \tag{3}$$

We then use the method of moments to estimate $\mu_{\ln EDP|IM=x}$ and $\beta_{\ln EDP|IM=x}$ using *only* the non-collapsed records. Combining the two possibilities, our estimate of the probability that *EDP* exceeds *z* given *IM* = *x* is now:

$$P(EDP > z \mid IM = x) = P(C \mid IM = x) + \left(1 - P(C \mid IM = x)\right)\left(1 - \Phi\left(\frac{\ln z - \hat{\mu}_{\ln EDP|IM=x}}{\hat{\beta}_{\ln EDP|IM=x}}\right)\right) \tag{4}$$

We can now proceed with this estimate.

PREDICTION OF BUILDING RESPONSE USING A VECTOR IM

We now adapt the procedure of the preceding section for use with a vector *IM*. If we label the two elements of our vector *IM* as $IM_1 = S_a(T_1)$ and $IM_2 = R_{T1,T2}$, then we are trying to estimate $P(EDP > z \mid IM_1 = x_1, IM_2 = x_2)$. Ideally, we would like to scale our records to both $IM_1 = x_1$ and $IM_2 = x_2$. However, when we scale our record by a factor *y*, each spectral acceleration value is changed by the same factor *y*. Therefore, $R_{T1,T2} = S_a(T_2)/ S_a(T_1)$ is unchanged by scaling. (In general, because there is only one factor that we are scaling by—a uniform scale factor on the entire record—we cannot match two *IM*s simultaneously, even if we consider other classes of IM_2 than $R_{T1,T2}$.) So we need a supplement to scaling.

The solution we adopt is to scale on IM_1 ($S_a(T_1)$) as before, and then apply regression analysis to estimate *EDP* versus IM_2 ($R_{T1,T2}$) for each IM_1 level (Neter [8]). In the Results section of this paper, we will explain the reasoning behind scaling to IM_1, rather than using regression analysis on both variables. Our approach will parallel the scalar *IM* case, in that we separate out the collapsing records first, and then deal with the remaining non-collapsed records.

Accounting for Collapses with the Vector IM

As with the scalar *IM*, there is a possibility that some records may cause collapse of the structure. However, instead of taking the probability of collapse to be simply the fraction of records that collapse, we would like to take advantage of our vector *IM* to predict the probability of collapse more accurately. We do this using logistic regression, which is commonly used to generate predictions for binary data (Neter [8]). Each record has a value of $R_{T1,T2}$, which we denote more simply as R_i and use as our predictor variable. We then build an indicator variable for collapse (call this Y_i and set it to 1 if the record causes collapse and 0 otherwise). We then use the logit transform to predict collapse:

$$Y_i = \frac{\exp(\beta_0 + \beta_1 R_i)}{1 + \exp(\beta_0 + \beta_1 R_i)} \tag{5}$$

where β_0 and β_1 are parameters to be estimated using our set of data points $\{Y_i, R_i\}$. Logistic regression is available as a built-in function in many mathematical and statistical software packages. We then use this function (and our estimated parameters) to predict the probability of collapse, given $S_a(T_1)$ and $R_{T1,T2}$:

$$P\left(C \mid S_a(T_1) = x_1, R_{T1,T2} = x_2\right) = \frac{\exp(\hat{\beta}_0 + \hat{\beta}_1 x_2)}{1 + \exp(\hat{\beta}_0 + \hat{\beta}_1 x_2)} \tag{6}$$

where $\hat{\beta}_0$ and $\hat{\beta}_1$ are used to denote the estimates of β_0 and β_1 obtained from regression on a dataset that has been scaled to $S_a(T_1) = x_1$ (i.e. $\hat{\beta}_0$ and $\hat{\beta}_1$ will be different for different values of $S_a(T_1)$). We now have a probability of collapse prediction that varies with $R_{T1,T2}$, rather than being constant as in the scalar case. An example of this data and a fitted logistic regression curve is presented in Figure 3.

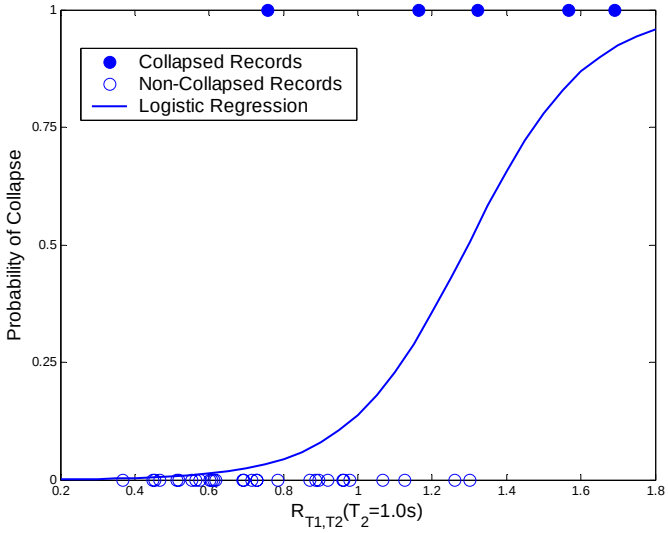

Figure 3: An example of prediction of the probability of collapse using logistic regression applied to binary collapse/non-collapse results ($S_a(T_1) = 0.9g$).

Accounting for Non-Collapses with the Vector IM

Once the probability of collapse has been estimated, it is necessary to quantify the behavior of the non-collapse records. We now separate out the non-collapsed records. Note that each of these records has been scaled to $S_a(T_1) = x_1$. Each of the records has a value of $R_{T1,T2}$, which we again refer to as R_i, and it has a value of EDP, which we refer to as EDP_i. We have found that there tends to be a relationship between $R_{T1,T2}$ and EDP of the form $EDP \mid S_a(T_1), NC \approx a(R_{T1,T2})^b \tilde{\varepsilon}$, where a and b are constant coefficients, and $\tilde{\varepsilon}$ is a random variable representing the randomness in the relationship. This becomes a linear relationship after we take logarithms of both sides: $\ln EDP \mid S_a(T_1), NC \approx \ln a + b \ln R_{T1,T2} + \varepsilon$ (we have defined a new random variable, $\varepsilon = \ln \tilde{\varepsilon}$). So we now use linear least-squares regression (Neter [8]) to estimate the our two regression coefficients, $\beta_2 = \ln a$ and $\beta_3 = b$:

$$\ln EDP_i = \beta_2 + \beta_3 \ln R_i + \varepsilon_i \tag{7}$$

Linear regression is available in many mathematical and statistical software packages, and can be used to obtain estimates of the coefficients, $\hat{\beta}_2$ and $\hat{\beta}_3$ (again, these values will vary for different $S_a(T_1)$ levels). A graphical example of this data and the regression fit is shown in Figure 4.

When using linear least-squares regression on a dataset, several assumptions are implicitly made, and the accuracy of the results depends on the validity of these assumptions. If we denote our predicted value of $\ln EDP$ for record i as $\ln \hat{EDP}_i$, then the prediction error for this record is called the "residual" from record i:

$$\varepsilon_i = \ln EDP_i - \ln \hat{EDP}_i \tag{8}$$

These residuals are assumed to be mutually independent. In addition, when estimating the distribution of ln*EDP* below, we will assume the residuals to be normally distributed with constant variance (this condition is termed homoscedasticity). The assumptions of independent normal residuals with constant variance have been examined for the data in this study, and found to be reasonable. An estimate of the variance of the residuals is also available from the analysis software, and we will denote it $\hat{Var}[\varepsilon] \equiv \hat{\sigma}_{\varepsilon}^2$. This variance in the residuals is displayed graphically in Figure 4, by superimposing the estimated normal distribution of the residuals over the data.

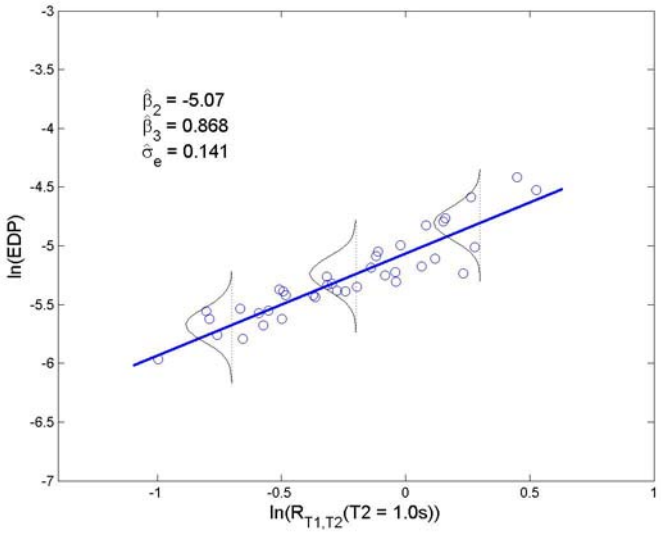

Figure 4: An example of non-collapse data, and a fit to the data using linear regression. The data comes from records scaled to $S_a(T_1) = 0.3g$. The estimated distribution of the residuals has been superimposed over the data.

From regression, we now know that given $S_a(T_1) = x_1$ and $R_{T1,T2} = x_2$, and given no collapse, the mean value of $\ln EDP$ is:

$$\ln EDP = \hat{\beta}_2 + \hat{\beta}_3 \ln x_2 \tag{9}$$

where $\hat{\beta}_2$ and $\hat{\beta}_3$ have been obtained by regressing on records scaled to $S_a(T_1) = x_1$. We also know that $\ln EDP$ is normally distributed, and that it has a variance equal to the $\hat{\sigma}_{\varepsilon}^2$. So the probability that $\ln EDP$ is greater than z, given $S_a(T_1) = x_1$, $R_{T1,T2} = x_2$, and no collapse can be expressed:

$$P\left(EDP > z \mid S_a(T_1) = x_1, R_{T1,T2} = x_2, \text{no col.}\right) = 1 - \Phi\left(\frac{\ln z - \left(\hat{\beta}_2 + \hat{\beta}_3 \ln x_2\right)}{\hat{\sigma}_{\varepsilon}}\right) \tag{10}$$

This equation is very similar to Equation 2 used in the scalar case. We previously estimated the mean of the normal distribution by the average response of all records, but now we use a result from regression on $R_{T1,T2}$. We have also replaced the standard deviation of the records by the standard deviation of the regression residual. But otherwise, the equation is the same.

We can combine Equations 6 and 10 to compute the probability that *EDP* exceeds *z*:

$$P(EDP > z \mid S_a(T_1) = x_1, R_{T1,T2} = x_2) = P(C) + \left(1 - P(C)\right)\left(1 - \Phi\left(\frac{\ln z - \left(\hat{\beta}_2 + \hat{\beta}_3 \ln x_2\right)}{\hat{\sigma}_\varepsilon}\right)\right) \quad (11)$$

$$\text{where } P(C) = \frac{\exp(\hat{\beta}_0 + \hat{\beta}_1 x_2)}{1 + \exp(\hat{\beta}_0 + \hat{\beta}_1 x_2)}$$

Although x_1 does not appear in Equation 11, our estimate is implicitly a function of x_1, because the data used to estimate $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3$ and $\hat{\sigma}_\varepsilon$ all comes from records scaled to $S_a(T_1) = x_1$. This gives us a response prediction that is similar to the original prediction of Equation 4, but that now incorporates a two-element vector. It is a simple matter to generalize this to a larger vector, by using regression on multiple variables in Equations 5 and 7, but results are not included in this report.

CHOICE OF A VECTOR

Once the above method has been established for predicting drift, we can proceed to choose an "optimal" vector for use in prediction. The general goals to be considered when choosing an *IM* are efficiency (minimum variance in *EDP* for records with the same *IM* value), sufficiency (see Luco [9]), and ease of calculation (i.e., determining the value of *IM* for a given record should not be too difficult). As noted earlier, in this study we are limiting the choices to $S_a(T_1)$ and $R_{T1,T2}$, and letting T_2 vary over all possible values. Spectral acceleration values are easy to calculate, and familiar to many engineers, so the ease of calculation criterion is met. The scalar $IM = S_a(T_1)$ has been found to be a sufficient predictor, and the addition of a second element has been found to not render the *IM* insufficient. So in this paper we focus on finding the *IM* that maximizes efficiency.

The question of efficiency arises when we examine, for example, the estimated mean of ln*EDP*. We are trying to estimate the mean of a population based on a sample from that population. A basic result in statistics tells us that the standard deviation of that estimate is equal to $\beta_{\ln EDP|IM=x} / \sqrt{n}$. So if we want to gain accuracy in our estimate of the mean, we either need to increase *n* or decrease $\beta_{\ln EDP|IM=x}$. Increasing *n* requires analyzing more records, which can be expensive. Therefore, we would like to decrease $\beta_{\ln EDP|IM=x}$ (i.e., we would like greater efficiency). This same principle holds in vector case—to increase the accuracy of our regression estimate, we need to decrease the standard deviation of the regression residuals. For example, in Figure 5 an IM_2 with a large residual standard deviation is shown, whereas in Figure 6 an IM_2 with a small residual standard deviation is displayed. If we are trying to estimate the trend in the data, clearly we will be able to do so more accurately using the IM_2 from Figure 6.

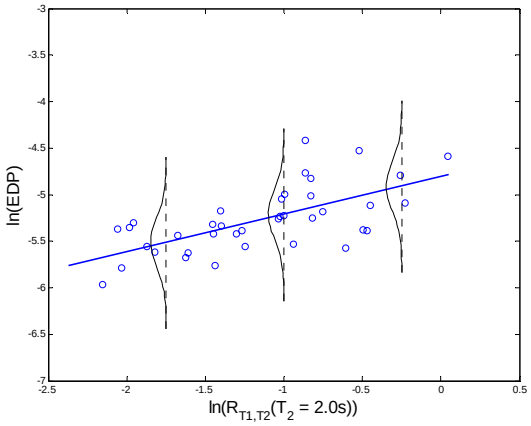

Figure 5: An IM_2 with low efficiency

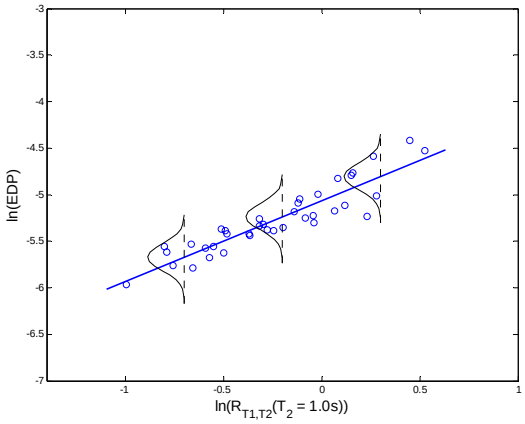

Figure 6: An IM_2 with high efficiency

To choose the optimal T_2 value at a given $S_a(T_1)$ level, we regress on $R_{T1,T2}$ for a range of T_2 values. We will then compute the standard deviation of the residuals for each of these regressions. To normalize these results, we will compute the fractional reduction in standard deviation relative to the standard deviation from the scalar *IM* case with $S_a(T_1)$ alone. If the fractional reduction is zero, then we have gained no efficiency by including the given IM_2. If the fractional reduction is one, then we have a perfect relation between IM_2 and *EDP*, and there is no remaining randomness. We will choose the T_2 that has the largest fractional reduction among all possible T_2 values.

BUILDING MODEL AND GROUND MOTION SELECTION

To perform the analysis as described above, one needs both a structural model and suite of earthquake records. The structure used in this study is a reinforced concrete moment frame building. The building has 1960's era construction and is serving as a test-bed for PEER research activities [10]. A 2D model of the transverse frame created by Jalayer [11] is the model used here. This model contains nonlinear elements that degrade in strength and stiffness, in both shear and bending (Pincheira [12]). The frame has seven stories and three bays. The first mode of the model has a period of 0.8 seconds, and the second mode has a period of 0.28 seconds.

Forty historical earthquake ground motions are used for the analysis. All of the records come from California, and the events range in Magnitude from 5.7 to 7.3. The distances vary from 6.5 km to 56 km. Attempts were made to avoid directivity effects by choosing records with small distances only when the rupture and site geometry suggested that near-fault effects would be unlikely, and velocity histories were not observed to contain pulse-like intervals. This set of 40 records was then scaled to 12 levels of $S_a(T_1)$ between 0.02*g* and 1.0*g*.

RESULTS

Using the given structure and ground motions, we can now use the Vector *IM* methodology to test candidate T_2 values for $R_{T1,T2}$. A plot of the fractional reduction in residual standard deviation is shown in Figure 7, where all records have been scaled to $S_a(T_1) = 0.3g$. We see that the optimal T_2 is one second (note, this optimal T_2 at this spectral acceleration level was shown in Figure 6 above, and a non-optimal T_2 was shown in Figure 5). We see that the reduction in dispersion was approximately 60%. We can make a comparison of the standard errors of estimation:

$$\frac{\sigma_{scalar}}{\sqrt{n_{scalar}}} = \frac{\sigma_{vector}}{\sqrt{n_{vector}}} \tag{12}$$

If we can reduce our dispersion by 60%, ($\sigma_{vector} = 0.4\sigma_{scalar}$), then $n_{vector} = 0.16 n_{scalar}$. That is, by adopting the most efficient vector, we could potentially reduce the number of records used by a factor of approximately six and still maintain the same accuracy in our estimate of the mean response of the structure. This could lead to a great reduction in computational expense.

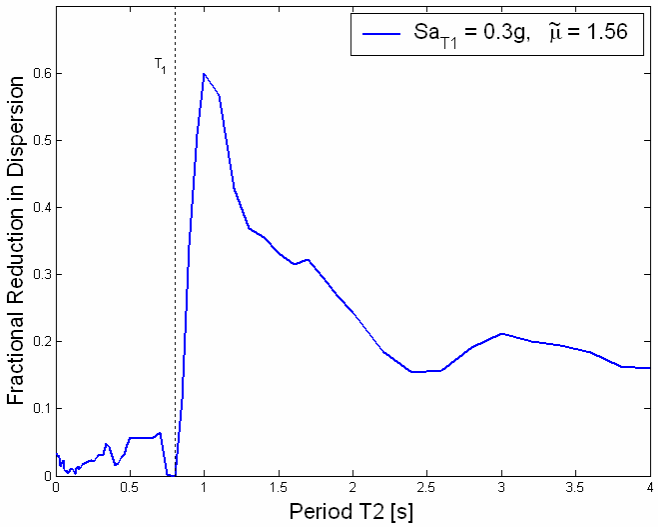

Figure 7: Fractional reduction in dispersion vs. T_2 for T_2 between 0 and 4 seconds for $S_a(T_1)$ = 0.3g.

This optimal T_2 value is only relevant for a single level of $S_a(T_1)$. We repeat this same calculation for two additional levels of $S_a(T_1)$ and show the results in Figure 8. It is apparent that the optimal T_2 value varies depending on the level of $S_a(T_1)$. An additional value, $\tilde{\mu}$ is given in the legend. This is the ratio of the average *EDP* among the 40 records to the approximate *EDP* at yielding (as determined from a pushover analysis of the structure). This value is analogous to a ductility level for the structure.

For $S_a(T_1)$ = 0.1*g*, we find that the optimal T_2 is 0.36 seconds. This is near 0.28 seconds: the second-mode period of the structure. In fact, 0.28 seconds shows a reduction in dispersion that is nearly as large as the reduction at 0.36 seconds. If 0.28 seconds were indeed the best T_2, then this vector *IM* would be somewhat analogous to the modal analysis method of estimating linear response. The modal analysis method (with two modes) would take the spectral acceleration at the first two modes of the building and use them to estimate the response of the structure. Note that at this level of spectral acceleration, our estimate of ductility is 0.533, implying that for most of the records, the structure stays linear.

For $S_a(T_1)$ = 0.3*g*, we find that the optimal T_2 is 1.0 seconds. This is the level that was discussed previously. It is noted now that $\tilde{\mu} = 1.56$, suggesting that most of the records cause some level of nonlinear behavior in the structure. For $S_a(T_1)$ = 0.7*g*, we find that the optimal T_2 is 1.5 seconds. At this level of spectral acceleration, $\tilde{\mu} = 4.66$. This suggests that most of the records experience very large levels of nonlinearity.

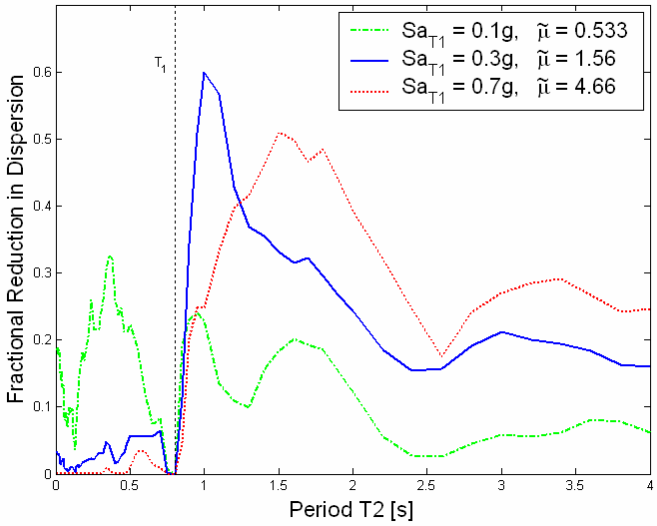

Figure 8: Fractional reduction in dispersion vs. T_2 for three levels of $S_a(T_1)$.

If one were to combine engineering intuition with the results of Figure 8, the following conclusion might be drawn: if $S_a(T_1)$ is low enough that few or no records cause nonlinearity, then the optimal second period to incorporate would be near the second-mode period of the structure, and if $S_a(T_1)$ is large enough that most records cause nonlinearities, then the optimal T_2 will be larger than T_1. The plot of Figure 9 below shows this pattern as well. In this plot, the level of $S_a(T_1)$ is given on the x-axis. On the y-axis is T_2, plotted as a ratio of T_2 to T_1. A circle is placed in the plot at the location of the optimal T_2 for a given $S_a(T_1)$. In addition, a line is plotted over the range of T_2/T_1 where the reduction in dispersion is at least 75% of the reduction seen at the optimal T_2. This line is shown to indicate the breadth of the optimal solution (i.e., are there only a few effective T_2's, or is there a large range of T_2 that reduces dispersion comparably?). Other authors (Cordova [2] and Vamvatsikos [3]) have examined choices of T_2, and have also recognized the dependence on the level of nonlinearity.

The average $\tilde{\mu}$ is less than one for $S_a(T_1)$ between 0 and 0.1g, and this is where we see an optimal T2 that is near to the second-mode period of the building. For $S_a(T_1) \geq 0.2g$ (and $\tilde{\mu} > 1$), the optimal T2 is larger than T_1, and shows an increasing trend as $S_a(T_1)$ increases (and as levels of nonlinearity increase). This increase in T2 with increasing levels of nonlinearity may be related to the idea of an equivalent linear system. The theory is that a nonlinear Single-Degree-Of-Freedom (SDOF) system may be represented by an "equivalent" linear system with a longer period. As the level of nonlinearity increases, the period of the equivalent linear system increases. A force-deformation diagram of a nonlinear system and its equivalent linear system is shown in Figure 10. The original SDOF has elastic stiffness k_e, and the equivalent linear system has a reduced stiffness k^* that is dependant on the ductility demand (μ_{max}) of the nonlinear system. Methods for determining an equivalent nonlinear system based on level of ductility have been proposed by Iwan [13] and Kennedy [14]. The suggested equivalent periods from these two papers are plotted on Figure 9, and they show trends similar to the ideal T_2's seen in this study. However, it should be noted that there is a difference between the equivalent linear system, and the IM_2 being selected for a vector. The equivalent linear system is used to *replace* $S_a(T_1)$, while the IM_2 is used to *supplement* $S_a(T_1)$. The discrepancy between these two goals is most apparent in Figure 9 when $S_a(T_1)$ is 0.3 or 0.4g. Here, the equivalent nonlinear system has a period almost identical to the elastic period of structure. However, the optimal T_2 from this study is at a longer period. This is because if T_2 is very near to T_1, then $S_a(T_1)$ is very highly correlated with $S_a(T_2)$ (see Inoue [15]). Thus, $R_{T1,T2} = S_a(T_2)/ S_a(T_1)$ is almost always very near to one and so there is little or no predictive ability provided by $R_{T1,T2}$. The

optimal T_2 must be significantly different than T_1 in order to decrease the correlation between $S_a(T_1)$ and $S_a(T_2)$. This can be seen in Figure 8, where the fractional reduction in dispersion is zero or nearly zero for T_2 values that are very close to T_1. So, although the equivalent linear systems somewhat match the optimal T_2 values at moderate to large levels of ductility, there is a discrepancy between the two at low levels of nonlinearity.

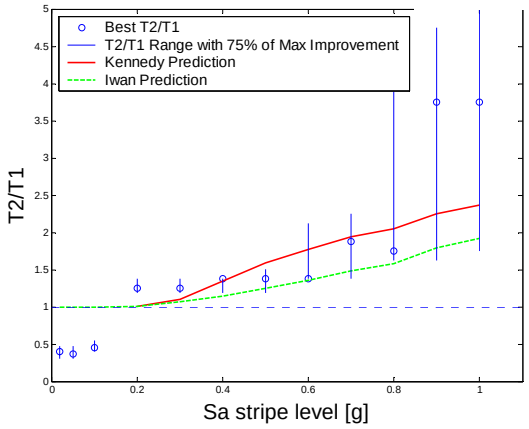

Figure 9: The optimum second period T_2, versus the level of $S_a(T_1)$.

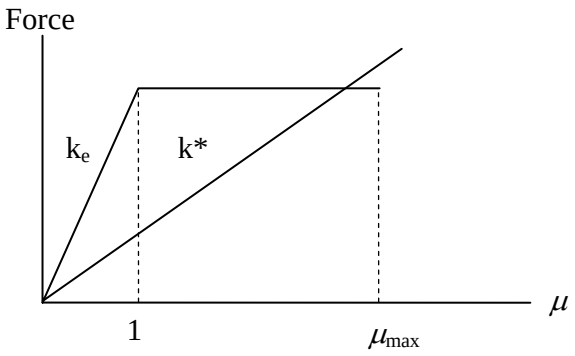

Figure 10: Illustration of a non-linear SDOF, and its equivalent linear system based upon the maximum ductility observed from a given earthquake ground motion.

With these results in mind, it is valuable to reconsider our original scheme for determining the optimal IM_2. We chose to scale records to $S_a(T_1)$ first, and then regress on $R_{T1,T2}$. Therefore, all of our regression coefficients were allowed to vary at each $S_a(T_1)$ level. This allows for full interaction between $S_a(T_1)$ and $R_{T1,T2}$. If we had not scaled the records, and simply regressed on both $S_a(T_1)$ and $R_{T1,T2}$ simultaneously, then there would be a need to parameterize the interactions, and this could be a difficult task to do well. In addition, had we regressed on both $S_a(T_1)$ and $R_{T1,T2}$ simultaneously without scaling, we would only find a single optimal T_2 for all $S_a(T_1)$ levels, rather than an optimal T_2 that varies with the level. These two issues were why scaling to $S_a(T_1)$ and then regressing on $R_{T1,T2}$ was chosen as the methodology. It should be noted however, that regressing on two *IM*s simultaneously has been done by Shome [5].

Optimization of the Choice of IM_2 Using the Bootstrap and the Drift Hazard Curve

Another issue with the proposed methodology that should be mentioned is the criterion for choosing an optimal $R_{T1,T2}$. Recall that the quantity being optimized in Figure 7 is the reduction in residual dispersion from regression on the *non-collapse* records only. This quantity does not measure the ability of $R_{T1,T2}$ to predict the probability of collapse. Perhaps if the collapse capacity of the structure is of primary interest, the quantity to optimize would be the reduction in residual dispersion from Equation 6. Ideally, the quantity to optimize should incorporate the improvements made in prediction of both collapse *and* non-collapse records. Because of this issue, the plot of Figure 9 is limited to $S_a(T_1) \leq 1g$. For $S_a(T_1) > 1g$, more than one-quarter of the records cause the structure to collapse, so it would not be appropriate to neglect the collapse records as has been done here.

It is possible to address this issue by computing a weighted sum of the reduction in dispersion from the collapse prediction (Equation 6) and the reduction in dispersion from regression on the non-collapse records (Equation 7). But the weights used should be a function of the relative importance of collapse and non-collapse behavior, and it is not clear what that function should be. The best resolution of this issue is

to carry the prediction of *EDP* all the way through to the computation of the mean rate of exceeding an *EDP* value *z*, as discussed in the introduction. The vector version of Equation 1 is given below:

$$\lambda_{EDP}(z) = \sum_{\text{all } x_{1,i}} \sum_{\text{all } x_{2,i}} P(EDP > z \mid S_a(T_1) = x_{1,i}, R_{T1,T2} = x_{2,i}) \cdot \Delta\lambda_{IM}(x_{1,i}, x_{2,i}) \quad (13)$$

This rate is the final result we desire, and this calculation incorporates both the collapse prediction and the non-collapse response prediction. Our ultimate goal is to calculate this accurately, so a reduction in variability here would be a good criteria to use in selecting $R_{T1,T2}$. To measure the statistical variability of the estimate of $\lambda_{EDP}(z)$, one would employ the bootstrap (Efron [16]). The procedure is as follows: select *n* records *with replacement* from the original set of *n* records (some records will be duplicated and others will not be present at all). With this new record set and a candidate $R_{T1,T2}$, use the vector *IM* to predict the building response as outlined above. Using this new estimate, re-compute $\lambda_{EDP}(z)$ using Equation 13. Repeat this process *N* times (no new structural analysis is performed, so *N*=1000 may not be too expensive). The standard deviation of these *N* values is an estimate of the standard error of estimation of $\lambda_{EDP}(z)$. A good T_2 value for $R_{T1,T2}$ will result in a $\lambda_{EDP}(z)$ that has significantly reduced variability as compared to the result of Equation 1. There are two advantages introduced by this method. The first is that it incorporates gains in efficiency from both the collapse and non-collapse predictions, but rather than using a potentially arbitrary weighting scheme, it incorporates them naturally into the final computation where the results will be used. The second advantage is that the probability of exceeding a given *EDP* value does not come from a single $S_a(T_1)$ level, but from a range of levels. Thus, a good predictor according to this criterion will show efficiency gains over the range of $S_a(T_1)$ values that contribute significantly at the given *EDP* level. Note that this calculation requires knowledge of vector ground motion hazard (the rate of jointly exceeding both $S_a(T_1)$ and $R_{T1,T2}$). This result is available (Bazzurro [17] and Somerville [18]), but not yet in widespread use in engineering practice.

As an illustration, the complete drift hazard curve $\lambda_{EDP}(z)$ is computed using the scalar and vector procedures (Equations 1 and 13 respectively). The ground motion hazard is calculated for the actual location of the structure being studied (a soil site in the Los Angeles area). The vector *IM* used is $\{S_a(T_1), R_{T1,T2}(T_2=1.0\text{s})\}$. The scalar and vector-based curves are shown in Figure 11. Note that the flattening of the curve towards the right occurs because the exceedance of these *EDP* values is dominated by collapses (P(collapse) $\cong 3*10^{-3}$). There is not a large difference between the two curves, but what is not apparent from this picture is that there is much less variability in the curve estimated using the vector *IM*. To measure this variability, we now use the bootstrap. Four example bootstrap replicates of the recordset have been generated, and their corresponding vector-*IM*-based drift hazard curves are shown in Figure 12. To display the variability in our estimates, we compute histograms of the $\lambda_{EDP}(z)$ values for a given *z*, using both the vector *IM* and scalar *IM* methods. These curves use maximum interstory drift = *z* as the *EDP* of interest.

We now examine histograms at a value of *z* of 0.01, to compare our scalar and vector *IM*s (see Figure 13). We see that the results from the vector-based drift hazard curve are much more tightly bunched around their central value than the results from the scalar-based drift hazard curve. This means that we can be more confident about the value obtained from the vector-based drift hazard calculation. Or equivalently, if we adopt the vector-based drift hazard calculation, we do not need to use as many records in the analysis to obtain an accuracy comparable to the scalar method. (In this case, we cut the standard deviation of the bootstrapped replicates by a factor of about two, so in principle we could reduce the number of records used by a factor of about four. However, the method does demand a minimum of approximately five to ten records per stripe.)

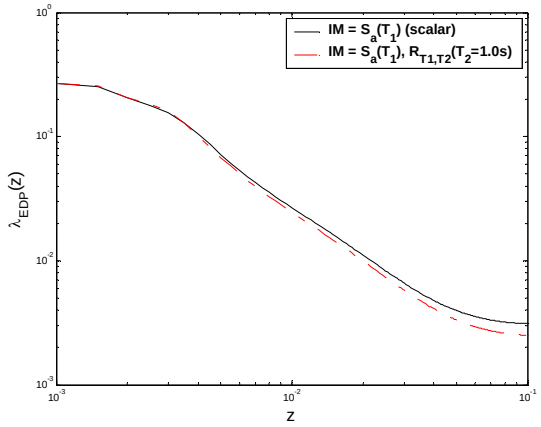

Figure 11: Maximum interstory drift hazard curves computed using a scalar *IM* and a vector *IM*

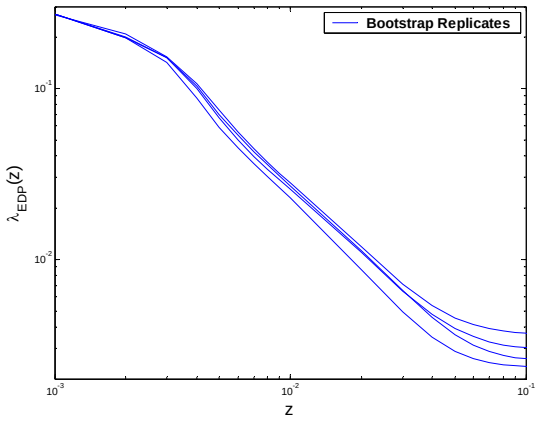

Figure 12: Bootstrap replicates of the vector *IM* drift hazard curve

The histogram of Figure 13 only shows results for one vector *IM* at a single *EDP* value, but we can extend this calculation to additional candidate vector *IM*s, and additional values of *EDP*. Consideration of candidate vector *IM*s is slightly more complicated now, because the vector ground motion hazard $\lambda_{IM}(x_1,x_2)$ needs to be recomputed for each candidate *IM*. For this reason, the analysis here is limited to three vectors that appeared promising in Figure 9: T_2 = 0.28s, 1.0s and 2.0s. The coefficient of variation of the replicate results is used as a measure of dispersion of the bootstrap replicates (because the mean values of $\lambda_{EDP}(z)$ vary by two orders of magnitude depending on z, standard deviations of $\lambda_{EDP}(z)$ are much more variable than the coefficient of variation). A plot of the coefficient of variation versus *EDP* level is shown for the three candidate vector *IM*s in Figure 14, along with the coefficient of variation using the scalar *IM*. We see that the vector $\{S_a(T_1), R_{T1,T2}(T_2=1.0\text{s})\}$ produces a significant reduction in coefficient of variation for nearly all levels of *IM*, while the other two vectors do not show a significant improvement. This result fits with results seen earlier. We saw in Figure 9 that the vector with T_2 = 0.28s was only helpful for very small levels of $S_a(T_1)$ (and in fact we see that for very small levels of *EDP*, this *IM* produces a small improvement). The vector with T_2 = 2.0s was helpful as $S_a(T_1)$ levels got very large, but these large-intensity events are rare enough that they do not significantly affect the *EDP* hazard curve except at large levels of *EDP* (and here we see a slight improvement). The vector with T_2 = 1.0s showed a significant improvement over a large range of important $S_a(T_1)$ levels, and thus it is the most useful. These results are consistent with earlier results, but perhaps this method reveals more information about the overall usefulness of a candidate vector than the previous method did. Note that to select a vector, we still need to specify a value of z to consider. But this choice of z may not be too hard if we are concerned with specific limit states (e.g. collapse, or EDP = 1% drift).

This bootstrap procedure requires more computation than the regression procedure. However, it has distinct advantages in that it directly measures uncertainty in the value of interest ($\lambda_{EDP}(z)$), and it incorporates estimates from both collapse and non-collapse prediction at many *IM* levels simultaneously. We saw from the example calculation above that the results from the two procedures appears consistent. Research to date seems to show that the two alternative techniques agree at levels of *EDP* where collapses are not frequent.

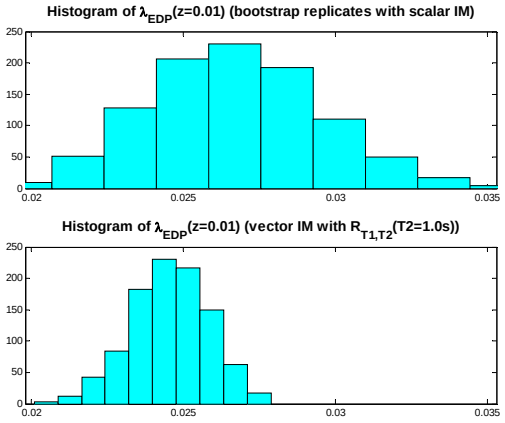

Figure 13: Histograms of the scalar and vector drift hazard curves, for z = 0.01.

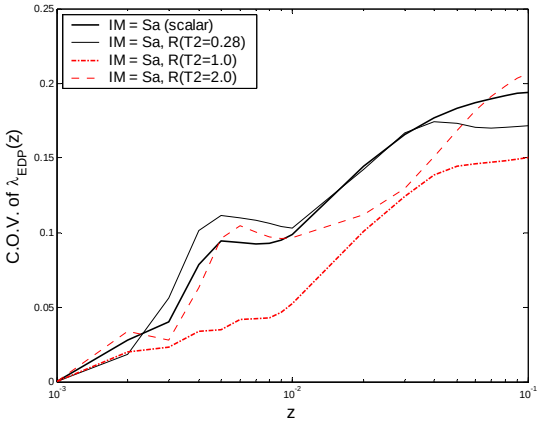

Figure 14: Coefficient of variation vs. *EDP* level for three candidate vector *IM*s and the scalar *IM*.

CONCLUSIONS

A method for selecting an efficient vector intensity measure using regression analysis has been presented. This method is based on scaling to the first *IM* element ($S_a(T_1)$) and then using regression on the second element ($R_{T1,T2}$) to predict the response of the structure. The optimal second period (T_2) for use in the vector was chosen by maximizing the reduction in standard deviation of the prediction errors. It was shown that the proper choice of $R_{T1,T2}$ can significantly reduce prediction error when compared to prediction using $S_a(T_1)$ alone. For the structure considered here, the optimal second period for use in $R_{T1,T2}$ was shown to be approximately equal to the second-mode period of the building when the building response was linear. When $S_a(T_1)$ was large enough to cause nonlinear behavior, the optimal second period was larger than the first-mode period of the building, and increased as the average level of ductility increased. This increase appears to be related to the effective period of an equivalent linear system.

An additional method for evaluating vector *IM*s that utilizes bootstrap replications of the drift hazard curve is also presented. This method has the advantage of directly computing the statistical variability in estimates of the drift hazard curve, and it accounts for the increased prediction accuracy of both collapse and non-collapse cases at many *IM* levels simultaneously. The disadvantage of this method is the requirement of a vector ground motion hazard for each candidate *IM*, and increased overall computational time. Because of this, it is suggested that the regression analysis method be used to narrow down a broad range of potential vector *IM*s to a few promising candidates. The bootstrap method can then be used to examine these few in detail.

A vector intensity measure has the potential to produce a drift hazard curve with much narrower confidence bands than the equivalent curve computed with the scalar intensity measure $S_a(T_1)$ and the same number of nonlinear analyses. Analysis of the structure evaluated in this paper shows the potential for a reduction in the standard deviation of $\lambda_{EDP}(z)$ of as much as a factor of two. This implies that in principle the required number of analyses could be reduced by a factor of as much as four without increasing the standard deviation of the $\lambda_{EDP}(z)$ result. This decrease in computational expense is very appealing, and may justify the use of the vector intensity measure.

ACKNOWLEDGEMENTS

This work was supported in part by the Pacific Earthquake Engineering Research Center through the Earthquake Engineering Research Centers Program of the National Science Foundation under award number EEC-9701568. We thank our co-researchers, Dr. Paul Somerville and Dr. Hong Kie Thio, for providing the vector-valued hazard results used in this study

REFERENCES

1. Cornell C.A., Krawinkler H., "Progress and Challenges in Seismic Performance Assessment". *Peer Center News*, 2000. **3**(2): 4p.
2. Cordova P.P., Deierlein G.G., Mehanny S.S.F., Cornell C.A. "Development of a Two-Parameter Seismic Intensity Measure and Probabilistic Assessment Procedure". *The Second U.S.-Japan Workshop on Performance-Based Earthquake Engineering Methodology for Reinforced Concrete Building Structures*. 2001. Sapporo, Hokkaido. 187-206.
3. Vamvatsikos D., *Seismic Performance, Capacity and Reliability of Structures as Seen through Incremental Dynamic Analysis*. Dept. Of Civil and Environmental Engineering. 2002: Stanford University. 152p.
4. Shome N., Cornell C.A., Bazzurro P., Carballo J.E., "Earthquakes, Records, and Nonlinear Responses". *Earthquake Spectra*, 1998. **14**(3): 32.
5. Shome N., *Probabilistic Seismic Demand Analysis of Nonlinear Structures*. Dept. Of Civil and Environmental Engineering. 1999: Stanford University. 320p.
6. Aslani H., Miranda E., "Probabilistic Response Assessment for Building-Specific Loss Estimation", *PEER 2003-03*. 2003, Pacific Earthquake Engineering Research Center, University of California at Berkeley: Berkeley, California. 49p.
7. Benjamin J.R., Cornell C.A., *Probability, Statistics, and Decision for Civil Engineers*. 1970, New York,: McGraw-Hill. 684p.
8. Neter J., Kutner M.H., Nachtsheim C.J., Wasserman W., *Applied Linear Statistical Models*. 4th ed. 1996, Boston: McGraw-Hill. 1408p.
9. Luco N., Cornell C.A., "Structure-Specific Scalar Intensity Measures for near-Source and Ordinary Earthquake Ground Motions". *Earthquake Spectra*, 2001 (Submitted).
10. *http://www.peertestbeds.net*. PEER Testbeds Project Site.
11. Jalayer F., *Direct Probabilistic Seismic Analysis: Implementing Non-Linear Dynamic Assessments*. Dept. Of Civil and Environmental Engineering. 2003: Stanford University. 244p.
12. Pincheira J.A., Dotiwala F.S., D'Souza J.T., "Seismic Analysis of Older Reinforced Concrete Columns". *Earthquake Spectra*, 1999. **15**(2): 245-272.
13. Iwan W.D., "Estimating Inelastic Response Spectra from Elastic Response Spectra". *Earthquake Engineering & Structural Dynamics*, 1980. **8**: 375-388.
14. Kennedy R.P., Kincaid R.H., Short S.A. "Prediction of Inelastic Response from Elastic Response Spectra Considering Localized Nonlinearities and Soil-Structure Interaction". *8th SMIRT*. 1985. 427-434.
15. Inoue T., Cornell C.A., "Seismic Hazard Analysis of Multi-Degree-of-Freedom Structures", *RMS-8*. 1990, Reliability of Marine Structures: Stanford, CA. 70p.
16. Efron B., Tibshirani R., *An Introduction to the Bootstrap*. Monographs on Statistics and Applied Probabilities; 57. 1993, New York: Chapman & Hall. 436p.
17. Bazzurro P., Cornell C.A. "Vector-Valued Probabilistic Seismic Hazard Analysis". *7th U.S. National Conference on Earthquake Engineering*. 2002. Boston, MA: Earthquake Engineering Research Institute. 1 CD-ROM.
18. Somerville P.G., Thio H.K., "Probabilistic Vector-Valued Ground Motion Intensity Measures and Engineering Demand Measures for the Peer Van Nuys Holiday Inn Pbee Testbed." *SCEC 2003 Project Report*p.