IMPORTANCE OF INTENSITY MEASURE SUFFICIENCY FOR STRUCTURAL SEISMIC DEMAND HAZARD ANALYSIS

S.L.N. Dhulipala¹, A. Rodriguez-Marek², and M.M. Flint³

ABSTRACT

Seismic Intensity Measures (IM) are regularly used in demand hazard estimation of structures to mediate between seismological parameters such as magnitude, distance and the Engineering Demand Parameter (EDP). Conventionally, the quality of an IM is described by the IM’s capability to predict an EDP (termed as efficiency) and render the EDP independent from seismological parameters (termed as sufficiency). While efficiency has received much attention in the past, sufficiency albeit being known as a sine qua non requirement has been overlooked. In this paper, we quantify sufficiency of IMs using a Total Information Gain metric ($\overline{TIG}$) which represents the additional knowledge about the EDP from all the seismological parameters under consideration. We apply the $\overline{TIG}$ to a four-story steel moment frame building for several combinations of EDPs – IMs – ground motion record sets and find that record selection can have a considerable impact on IM sufficiency. Next, we show that the $\overline{TIG}$ generally represents changes in demand hazard curves when seismological parameters are included in computations. Finally, we explore the relation between efficiency and sufficiency of IMs and propose a unified metric for both sufficiency and efficiency of an IM.

¹Graduate Student, Dept. of Civil and Env. Engg., Virginia Tech, Blacksburg, VA 24061 (email: lakshd5@vt.edu)
²Professor, Dept. of Civil and Env. Engg., Virginia Tech, Blacksburg, VA 24061
³Assistant Professor, Dept. of Civil and Env. Engg., Virginia Tech, Blacksburg, VA 24061

Importance of Intensity Measure Sufficiency for Structural Seismic Demand Hazard Analysis

S.L.N. Dhulipala¹, A. Rodriguez-Marek², and M.M. Flint³

ABSTRACT

Seismic Intensity Measures (IM) are regularly used in demand hazard estimation of structures to mediate between seismological parameters such as magnitude, distance and the Engineering Demand Parameter (EDP). Conventionally, the quality of an IM is described by the IM’s capability to predict an EDP (termed as efficiency) and render the EDP independent from seismological parameters (termed as sufficiency). While efficiency has received much attention in the past, sufficiency albeit being known as a sine qua non requirement has been overlooked. In this paper, we quantify sufficiency of IMs using a Total Information Gain metric (\(\overline{TIG}\)) which represents the additional knowledge about the EDP from all the seismological parameters under consideration. We apply the \(\overline{TIG}\) to a four-story steel moment frame building for several combinations of EDPs – IMs – ground motion record sets and find that record selection can have a considerable impact on IM sufficiency. Next, we show that the \(\overline{TIG}\) generally represents changes in demand hazard curves when seismological parameters are included in computations. Finally, we explore the relation between efficiency and sufficiency of IMs and propose a unified metric for both sufficiency and efficiency of an IM.

Introduction

Performance Based Methods relying on the PEER framework for Performance Based Earthquake Engineering often assume that a scalar Intensity Measure (IM) is adequate to characterize the complex seismic features at a site. To represent the adequacy of an IM in representing seismological parameters such as magnitude and distance, a desirable quality of an IM termed

¹Graduate Student, Dept. of Civil and Env. Engg., Virginia Tech, Blacksburg, VA 24061 (email: lakshd5@vt.edu)
²Professor, Dept. of Civil and Env. Engg., Virginia Tech, Blacksburg, VA 24061
³Assistant Professor, Dept. of Civil and Env. Engg., Virginia Tech, Blacksburg, VA 24061

Sufficiency was introduced [1]. Sufficiency has been frequently evaluated through a pass/fail basis using p-values. If the Engineering Demand Parameter (EDP) residuals given an IM show significant dependence on a seismological parameter, p-value will be close to zero and such an IM will be deemed insufficient. Otherwise, the p-value will be greater than a significance level (say 0.05) and the IM is rendered sufficient. An obvious question that arises is: what happens when multiple IMs pass or fail the p-value test across a suite of seismological parameters considered? In such cases p-values only demarcate between sufficient and insufficient IMs and do not provide a unified measure for degree of sufficiency from all the seismological parameters of interest.

In order to overcome the above limitation, we discuss a unified metric for sufficiency of IMs proposed by the authors in [2] (termed as Average Total Information Gain, $\overline{TIG}$). This unified metric relies on Bayes’ rule and information theory to formulate a method that gauges sufficiency in a neutral manner across all EDP-IM combinations. The metric for sufficiency is applied to a steel moment frame building located in Los Angeles, CA across 192 combinations of EDPs, IMs and record sets. Three important seismological parameters are taken into consideration. The degree to which $\overline{TIG}$ represents changes in demand hazard when seismological parameters are considered is explored. And the quality of ground motion set selected in terms of rendering IM sufficiency is also investigated. Finally, we explore the relation between efficiency (another important quality of a desirable IM) and sufficiency of IMs and find that these metrics which describe an IM’s quality are Bivariate Normal distributed. We leverage this finding to propose a unified metric for both sufficiency and efficiency of an IM.

A Measure for Sufficiency of an Intensity Measure

Let $\Phi = (\phi_1, \phi_2, ...)$ be a vector of ground motion or seismological parameters against which conditional independence of response is to be assessed. Let $IM_i$ be the $i$th ground motion intensity measure in a suite of alternative intensity measures. Let $EDP$ be the structural response quantity under consideration. We use a linear model to relate $EDP$ and $IM_i$ ($lnEDP = a_0 + a_1lnIM_i$) and a bilinear model to relate $EDP$, $IM_i$ and $\phi_j$, the $j$th seismological parameter ($lnEDP = b_0 + b_1lnIM_i + b_2 \phi_j$). Let $Pr(EDP > y|IM_i)$ and $\overline{Pr}(EDP > y|IM_i)$ be the probability of exceedances of a response level given $IM_i$. The latter exceedance probability ($\overline{Pr}$) denotes that the $j$th seismological parameter has been considered in the empirical relation and has been marginalized as shown below,

$$\overline{Pr}(EDP > y|IM_i) = \int_{\phi_j} Pr(EDP > y|IM_i, \phi_j) f(\phi_j|IM_i) d\phi_j$$

In this equation $f(\phi_j|IM_i)$ denotes the density distribution of $\phi_j$ given an IM level. For example, if $\phi_j$ is the magnitude of earthquake, the distribution $f(\phi_j|IM_i)$ can be obtained by deaggregation. The probability density function of $IM_i$ given a response level $EDP > y$ can be obtained using Bayes’ rule [3] as shown below,

$$f(IM_i|EDP > y) \propto Pr(EDP > y|IM_i) \left| \frac{d\lambda(IM_i)}{dIM_i} \right|$$

where, $\lambda(IM_i)$ denotes the seismic hazard curve for the $i$th IM in the suite; it is noted that the
normalizing constant which is an integration of the numerator across all IM values has been omitted for brevity. When considering the \( j \)th seismological parameter, a density distribution can be obtained in a similar fashion and will be denoted by \( f_j(IM_i|EDP > y) \).

In this paper, we employ principles of information theory to assess the degree of conditional independence of scalar IMs. For more details on information theory the reader is referred to [4]. In particular, we gauge the influence of a seismological parameter (\( \phi_j \)) by computing the Kullback-Liebler Divergence (KLD) between the density distributions \( f(IM_i|EDP > y) \) and \( f_j(IM_i|EDP > y) \). The KLD, also known as information gain, utilized to measure IM sufficiency is given by [2],

\[
IG^i_j(y) = \int_{IM_i} f_j(IM_i|EDP > y) \log_2 \frac{f_j(IM_i|EDP > y)}{f(IM_i|EDP > y)} dIM_i
\]

where, \( IG^i_j(y) \) is the information gain due to consideration of the \( j \)th seismological parameter given the \( i \)th IM and EDP level \( y \). A schematic illustrating the method of computing the information gain due to \( j \)th seismological parameter at a particular response level (\( EDP > y \)) for the \( i \)th IM is shown in Figure 1.

![Figure 1](image1.png)

**Figure 1.** Schematic illustrating the convolution of hazard curve and fragility function using Bayes’ rule to give density distribution of IM given response level. \( \phi_j \) is \( j \)th the seismological parameter.

The total information gain \( TIG^i(y) \) under the \( i \)th IM due to all the seismological parameters considered (\( \Phi \)) is simply the sum of information gains attributable to the individual seismological parameters (\( \phi_j \)). This final equation for the conditional independence metric of a given IM and response level is shown below,

\[
TIG^i(y) = \sum_{j=1}^{N_{\phi}} IG^i_j(y)
\]

When the above metric is averaged across all the response levels considered, we obtain a quantity for sufficiency comparison across different IMs and EDPs. The \( TIG \) averaged across the response levels will be denoted by \( \overline{TIG} \).

**Case Study Description**
We apply the proposed metric for sufficiency measurement to a four-story steel moment frame building designed for seismic hazard at a site in Los Angeles, CA [33.996°N, 118.162°W]. A 2D model for the moment frame in EW direction was developed by [5] in OpenSees considering material and geometric nonlinearities utilizing sophisticated models. Eight IMs are considered in the present study: $Sa(T_1 = 1.33s)$, $Sa(T_2 = 0.43s)$, $Sa(T_3 = 0.22s)$, $Sa(1s)$, $Sa(1.5s)$, $Sa(2s)$, $PGA$ and $PGV$. Six EDPs are investigated: Roof Drift ($RD$), Inter-story Drift Ratios at the first and fourth stories ($IDR1$ and $IDR4$), Joint Rotation ($JR$), Peak Floor Accelerations at the first and fourth stories ($PFA1$ and $PFA4$). Sufficiency from three seismological parameters are studied: Magnitude ($M$), Distance ($R$) and epsilon ($\varepsilon$: normalized difference between observed and predicted log ground motions).

To obtain the conditional deaggregation information to compute equation (1), OpenSHA software [6] was utilized. For a simplified method to approximate the deaggregation probabilities across the IM space, the reader is referred to [2].

Four ground motion sets were used to simulate the EDPs considered and construct demand fragility functions. The first set is FEMA P695 far-field set comprising of forty-four records [7]. The second set is Medina-Krawinkler LMSR-N set [8]. The third and fourth sets at this site were constructed using the Conditional Spectrum approach [9] where the former set comprises of far-field records and the latter one only constitutes near-field motions. Considering multiple record sets enables a study of the influence of record selection on IM sufficiency.

**Impact of IM Sufficiency on Demand Hazard Analysis**

The metric for sufficiency ($\overline{TIG}$) obtained by averaging equation (4) across the EDP levels of interest was computed across 192 the combinations of EDPs, IMs and record sets. The degree to which $\overline{TIG}$ represents changes in the demand hazard when seismological parameters are included in the calculations will be explored in this section. When considering the $j$th seismological parameter ($\phi_j$), the demand hazard is calculated using,

$$\bar{\lambda}(EDP > y) = \int_{IM_i} \bar{Pr}(EDP > y|IM_i) \left[ \frac{d\lambda(IM_i)}{dIM_i} \right] dIM_i$$  \hspace{1cm} (5)

where $\bar{Pr}(EDP > y|IM_i)$ is the demand fragility obtained from equation (1) and $\left[ \frac{d\lambda(IM_i)}{dIM_i} \right]$ is the slope of the seismic hazard curve. It is noted from the above equation that each seismological parameter is treated separately for demand hazard computation, although, there is a combined effect of seismological parameters in the EDP-$IM_i$-$\phi_j$ relationship. This treatment can be attributed to the finite datasets used for generating the fragility functions. However, if the datasets used for fragility estimation are satisfactorily large, then, equations (1) and (5), and the metric for sufficiency can be conveniently updated to consider multiple seismological parameters at once.

Figures 2 and 3 provide demand hazard curves for a select combinations of EDPs, IMs and records sets among the plethora considered in this study. Also provided in these figures are the sufficiency ($\overline{TIG}$) and efficiency ($\beta_{EDP|IM}$) values. It can be readily noticed from these figures that under the same EDP and without including the seismological parameters, different combinations of IMs and record sets result in different estimates of the demand hazard (examples: Figures 2a and 2b; Figures 3a and 3b). This alleviates the importance of proper record and IM selection which has been emphasized in other studies too. Including the seismological parameters in demand hazard
computations results in a variable effect. In some cases the seismological parameters have no effect
at all (examples: Figures 2d; 3d); while in other cases seismological parameters significantly
influence demand hazard estimates (example: Figure 3c). In some interesting cases, the
seismological parameters seem to have a mixed effect on the demand hazard—incrementing the
hazard at some EDP levels and reducing it at other levels (example: Figure 2a).

![Figure 2. Demand hazard curves computed without and with considering the seismological
parameters (M, R, ε) for the EDPs Roof drift (a and b), IDR1 (c) and IDR4 (d). The combination
of record set and IM is shown in each subfigure. The values of average TIG (\(\overline{TIG}\)) and standard
deviation in predicting \(\ln EDP\) given \(\ln IM\) are also shown.](image)

The \(\overline{TIG}\) metric is seen to represent changes in the demand hazard when seismological parameters
are included, thereby, serving as a metric for sufficiency. Low values of \(\overline{TIG}\) indicate consistency
(examples: Figures 2b; 3d), moderate values of $\overline{TIG}$ indicate noticeable deviation from the “only IM” curve (examples: Figures 2a; 3b), and high values of $\overline{TIG}$ suggest considerable influence of seismological parameters (examples: Figures 3c). Finally, it is noted that if other seismological parameters such as $M^2, M \cdot \ln R$ or $\ln V_{S30}$ in a GMPM are also considered, the computed $\overline{TIG}$ is bound to increase due to the positivity of KLD [4].

![Diagram](image)

Figure 3. Demand hazard curves computed without and with considering the seismological parameters ($M$, $R$, $\epsilon$) for the EDPs Joint Rotation (a and b), PFA1 (c) and PFA4 (d). The combination of record set and IM is shown in each subfigure. The values of average TIG ($\overline{TIG}$) and standard deviation in predicting lnEDP given lnIM are also shown.

**Quality of the Ground Motion Set in Rendering IM sufficiency**

To compare the capability of individual ground motion record sets in rendering IM sufficiency, Figure 4 shows the average information gains ($\overline{TIG}$) for the most sufficient IM across different
EDP-ground motion record set combinations. It can be observed that the FEMA P695 and conditional spectrum matched pulse-like sets are better in terms of IM sufficiency across all EDPs as compared to the other two records sets adopted in this study. The conditional spectrum matched non pulse-like ground motion set on the other hand, has the least capability to render IM sufficiency among the ground motion record sets adopted in this study. So, ground motion selection procedures such as the CS approach which try to enforce hazard consistency need not always render IMs to be sufficient. This observation is in line with that of [10]. However, it is noted that a limited number of IMs are considered in this study.

![Graph](image)

Figure 4. Quality of ground motion records in rendering IM sufficiency (for the IMs considered in this study): Average information gains of the most sufficient IMs for various EDP-record set combinations.

**A Unified Metric for Assessing the Sufficiency and the Efficiency of an Intensity Measure**

A unified metric for efficiency and sufficiency of an IM will be derived in this section by studying the relationship between $\beta_{EDP|IM}$, which is a measure for efficiency of an IM, and $\overline{TIG}$, which is measure for sufficiency of an IM. A scatter plot of $ln(\overline{TIG})$ versus $ln(\beta_{EDP|IM})$ considering all EDP-IM-ground motion combinations adopted in this study is shown in Figure 5 [2]. The median prediction, standard deviation of this prediction, and Pearson correlation coefficient are also shown. It can be observed from this figure that there is considerable scatter around the median prediction, indicating that efficiency and sufficiency of an IM are weakly correlated. This is also implied by the Pearson correlation coefficient, which is 0.29. However, there is a positive correlation between efficiency and sufficiency indicating as an IM becomes more efficient the same IM also tends to become more sufficient on an average. To test the bi-variate normality of $ln(\overline{TIG})$ and $ln(\beta_{EDP|IM})$, the Henze-Zirkler test and the Mardia’s test for skewness and kurtosis [11] were performed. All three tests fail to reject the null hypothesis that $ln(\overline{TIG})$ and $ln(\beta_{EDP|IM})$ come from a bi-variate normal distribution at a significance level of 0.05. The p value for the Henze-Zirkler test is 0.19 and the p-values for the Mardia’s test are 0.17 and 0.18 for skewness.
and kurtosis respectively. Utilizing this conclusion, a unified metric for both efficiency and sufficiency of an IM is proposed.

Figure 5. Relation between standard deviation in structural response given IM \((\ln(\beta_{EDP|IM})\), efficiency) and Total Information Gain \((\ln(TIG)\), sufficiency) for the EDPs, IMs, ground motion record sets and structure considered in this study [2].

The natural logarithm of metrics for efficiency and sufficiency, \(\ln(\beta_{EDP|IM})\) and \(\ln(TIG)\), are correlated to some degree and are not on the same scale. First, a Mahalanobis transformation [12] is utilized to de-correlate and transform \(\ln(\beta_{EDP|IM})\) and \(\ln(TIG)\) into a bi-variate standard normal space. This transformation can be performed using,

\[
\tilde{Z}_i = \tilde{S}^{-1/2}(\tilde{X}_i - \tilde{X}_M)
\]

where \(\tilde{Z}\) is a 2 \times 1 vector containing transformed values, \(\tilde{S}\) is the covariance matrix, \(\tilde{X}\) is a 2 \times 1 vector containing original values and \(\tilde{X}_M\) is a 2 \times 1 vector containing mean values of \(\ln(\beta_{EDP|IM})\) and \(\ln(TIG)\) respectively. A scatter plot of the transformed values is shown in Figure 6a. Now, theoretically, the co-ordinates of the “perfect” IM in this transformed space would tend to \((-\infty, -\infty)\). Or in other words, this “perfect” IM would have co-ordinates tending to \((0, 0)\) in the exponent of transformed space. A scatter plot of the exponent of vector \(\tilde{Z}_i\) is shown in Figure 6b. The Euclidean norm of the vector \(exp(\tilde{Z}_i)\) is a measure of how efficient and sufficient an IM is. The definition of this unified metric is mathematically given by,

\[
A_i = \ln(||exp(\tilde{Z}_i)||)
\]

where \(||.||\) represents a Euclidean norm. In the above equation, natural logarithm is used for visualization purposes. The unified metric \(A_i\) is dimensionless and has bounds \((-\infty, +\infty)\). Lesser the value of \(A_i\), better is the IM in terms of efficiency and sufficiency. A histogram of \(A_i\) values considering all combinations of IMs, EDPs and record sets is shown in Figure 6c.
Summary and Conclusions

Sufficiency is an important criterion to avoid a biased evaluation of the seismic demand hazard. In this paper we discuss a unified metric to gauge the sufficiency of IMs from all the seismological parameters. This metric for sufficiency has been formulated utilizing Bayes’ rule and Information Theory principles. A four-story steel moment frame was selected to apply the new sufficiency metric to numerous combinations of EDPs, IMs and ground motion record sets. The influence of IM sufficiency and record selection on demand hazard estimation was explored. Finally, we also propose a unified metric for sufficiency and efficiency (another important quality of a desirable IM). The following conclusions are made:

- Including the seismological parameter in demand hazard computations can have a significant effect depending on EDP-IM-record set combination adopted.
- Seismological parameters have a varied effect when included into the demand hazard calculation. In some cases, these parameters uniformly increase or decrease the hazard. And in other cases, they have a mixed effect incrementing the demand hazard at some EDP levels while decreasing it at other EDP values (with reference to the demand hazard curve computed using only the IM).
- The Average Total Information Gain metric ($\overline{\text{TIG}}$) is generally seen to represent changes in the demand hazard when seismological parameters are included.
- By comparing the $\overline{\text{TIG}}$ value of the most sufficient IM across different EDP and record set combinations, we find that the FEMA P695 and the CS matched pulse sets adopted in this study are better at rendering sufficiency in general.
- Upon plotting the two metrics which gauge an IM’s quality (sufficiency and efficiency) in a log-log space, we find that these measures are Bivariate Normal distributed. We leverage this conclusion to propose a combined metric for both sufficiency and efficiency.
Acknowledgment

This research is supported by the US National Science Foundation, Division of Civil, Mechanical, and Material Innovation, through Award #1455466.

References