Towards Vector Hazard Analysis, Earthquake Loss Estimation and Record Selection

Peter J. Stafford
Willis Research Fellow & RCUK Fellow
p.stafford@imperial.ac.uk
Acknowledgements

The vast majority of the work that is presented in the following has been conducted over the past couple of years in close collaboration with Prof. Julian Bommer.

The work also reflects significant contributions from colleagues:

- Fleur Strasser
- Ellen Rathje
- Ahmed Elghazouli
- Helen Crowley
- Sinan Akkar
- Elkhayam Dorra
- Rui Pinho
- Nicola Buratti
- Alice Modica
- John Berrill
- Jon Hancock
- John Alarcón
- Jarg Pettinga
- Sara Sgobba
- Beppe Marano
Topics to be Addressed

General research themes relate to the characterisation of ground-motion for seismic hazard and risk applications

• Vector hazard analysis
  – Development of empirical predictive models
  – Development of correlation models

• Earthquake loss estimation
  – A few brief comments on current research themes

• Record selection and modification
  – A summary of recent work
Basis for Vector Hazard

- Vector-valued probabilistic seismic hazard analysis (VPSHA) was proposed over a decade ago (Bazzurro, 1998; Bazzurro & Cornell, 2002)
- Whereas common scalar PSHA provides the mean annual rate at which a particular value of a single intensity measure would occur, VPSHA simultaneously provides the mean annual rate at which combinations of values of two or more intensity measures would occur
- We know that earthquake ground-motion cannot be adequately described by a single parameter so VPSHA makes perfect intuitive sense when more than one aspect of the ground motion is important
Why has VPSHA not become widespread?

**Soft barriers to implementation**
- General understanding of scalar PSHA is not very good and the idea of an extension to VPSHA turns many people away
- It is not implemented in accessible software routines

**Real barriers to implementation**
- Correlation models between intensity measures are required and these are not readily available (with the notable, significant exception of Baker’s work considering spectral ordinates and also Arias intensity)
- It has been shown that the additional complexity may not be warranted in many cases (Rajeev et al., 2008)
- It is not always clear what combinations of intensity measures are useful – different combinations are important for different problems
- It is more difficult to present the results in a tractable manner, particularly in current code formats
Mathematical basis

- Scalar PSHA may be written as follows:

\[
MRD_{IM}(im) = \sum_{i=1}^{N} \nu_i \left\{ \int \int f_{IM}(im|m,r) f_{M,R}(m,r) \, dm \, dr \right\}_i
\]

Marginal PDF in scalar case

JPDF in VPSHA

Seismicity analysis is the same in both cases

- Using this formulation, the extension to VPSHA is straightforward and one can appreciate that there should be no real conceptual barrier to understanding VPSHA (provided one understands the scalar case)

\[
MRD_{IM_1,IM_2}(im_1,im_2) = \sum_{i=1}^{N} \nu_i \left\{ \int \int f_{IM_1,IM_2}(im_1,im_2|m,r) f_{M,R}(m,r) \, dm \, dr \right\}_i
\]
JPDF of intensity measures

- The marginal PDF that is required in the scalar case is precisely what is provided by empirical ground-motion models.
- It can be shown that the JPDF that is required can be obtained directly from empirical ground-motion models for each intensity measure and an estimation of the correlation between these measures.
- If one wishes to consider an intensity measure within VPSHA then a decent empirical ground-motion model must first be available.
- Models for peak ground acceleration and spectral acceleration have dominated and options are therefore limited.
New empirical ground-motion models

- As a first step towards enabling correlations to be developed and for VPSHA to be implemented, a series of empirical ground-motion models have been developed
  - Fourier amplitude spectrum
  - Arias intensity
  - Husid parameters
  - Bracketed, Uniform and Significant Durations
  - Equivalent numbers of cycles
  - Modification factors for over-damped spectral ordinates
- Current attention is being placed on the revision of models for Arias intensity (including a model for spatial correlations) and the Mean Period
- Preliminary correlations have been obtained so far...
Preliminary Intra-event Correlations
Degrees of Correlation

• Thus far, the correlations that have received the most attention are those among spectral ordinates (and to a lesser extent Arias intensity)
• In both of these cases, the general level of correlation is quite high
• The additional intensity measure may be providing largely the same information
• Need to investigate if there are combinations of parameters that can ultimately lead to reduced variability in response estimates. Possibly these will have intermediate correlations
• Work is currently underway on this topic (with the aim of developing vector fragility functions for earthquake loss estimation)
Bracketed and Uniform Duration

- One obstacle that was encountered when developing empirical models for bracketed and uniform duration was how to deal with the possibility of observing zero values.
Biased observations

Higher than average accelerations may generate non-zero durations

Must therefore account for the probability, or the likelihood, that the acceleration associated with an observation is above the threshold

Use conditional likelihood within MLE
Treatment of Zero values

- This problem is not straightforward to deal with, but due account must be taken for zero-valued duration if robust correlations are to be derived.
- The problem is also relevant for other applications such as prediction of slope displacements, landslides, onset of liquefaction etc.
- Solution found using a two-part correlated random-effects model where the inter- and intra-event correlation is derived as part of the empirical model for duration.

\[ a_{ij} = \mu_A(X_{ij} | \gamma) + \eta_{A,i} + \varepsilon_{A,ij} \]

\[ d_{ij} = \mu_D(Z_{ij} | \theta) + \eta_{D,i} + \varepsilon_{D,ij} \]

\[ \Sigma_\eta = \begin{bmatrix} \tau_A^2 & \rho_{\eta} \tau_A \tau_D \\ \rho_{\eta} \tau_D \tau_A & \tau_D^2 \end{bmatrix} \]

\[ \Sigma_\varepsilon = \begin{bmatrix} \sigma_A^2 & \rho_{\varepsilon} \sigma_A \sigma_D \\ \rho_{\varepsilon} \sigma_D \sigma_A & \sigma_D^2 \end{bmatrix} \]
Model Comparison

![Graph showing model comparison](image)

- Uniform Duration, $D_{UA}(0.05g)$ (s)
- Rupture Distance, $R_{rup}$ (km)

Lines indicate:
- Nonlinear Least Squares
- Random Effects
- Censored Conditional Random Effects
Arias intensity and Husid parameters

- As another aside, the consideration of Arias intensity has led to the development of an energy-based envelope for stochastic simulation of ground motions.
- A complaint that is often made against stochastically simulated ground-motions is that they do not possess the correct energy content.
- We desired a non-stationary stochastic model that captured the energy content correctly.
Arias intensity and Husid parameters

- I realised that the shape of the Husid plot is very similar to a lognormal CDF
- If we can relate the moments of this distribution to seismological characteristics then we capture the temporal distribution of energy
- The probability distribution must be scaled by some intensity factor – for which we use the Arias intensity

\[
E(t) = \sqrt{\frac{4gI_a}{t\sigma^2\pi \sqrt{2\pi}}} \exp\left[-\frac{(\ln(t) - \mu)^2}{2\sigma^2}\right]
\]

- How could such a model actually fit into a seismic design scenario?
## Implementing Stochastic Models

<table>
<thead>
<tr>
<th>Scenario-based analysis</th>
<th>PSHA in terms of Arias intensity</th>
<th>PSHA in terms of spectral acceleration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scenario specified directly $M, R$</td>
<td>$I_a$ from PSHA for a given $\lambda$ $I_a \mid \lambda$</td>
<td>$S_a(T)$ from PSHA for a given $\lambda$ $S_a(T) \mid \lambda$</td>
</tr>
<tr>
<td>Envelope parameters are all median values $\hat{I}_a(M, R)$ $\hat{\mu}(M, R)$ $\hat{\sigma}(M, R)$</td>
<td>Scenario from disaggregation $M^<em>, R^</em>, \varepsilon_{lnI_a}^*$</td>
<td>Scenario from disaggregation $M^<em>, R^</em>, \varepsilon_{lnS_a(T)}^*$</td>
</tr>
<tr>
<td>Envelope function and standard deviation $\hat{F}(t \mid M, R)$ $\hat{\sigma}_{lnF}(t \mid M, R)$</td>
<td>Envelope parameters conditioned on $S_a(T)$ $I_a(M^<em>, R^</em>, \varepsilon_{lnI_a}^<em>)$ $\mu(M^</em>, R^<em>) \mid \varepsilon_{lnI_a}^</em>$ $\sigma(M^<em>, R^</em>) \mid \varepsilon_{lnI_a}^*$</td>
<td>Envelope function and standard deviation $E(t \mid M^<em>, R^</em>, \varepsilon_{lnI_a}^<em>)$ $\sigma_{lnE}(t \mid M^</em>, R^<em>, \varepsilon_{lnI_a}^</em>)$</td>
</tr>
</tbody>
</table>

Correlations are derived among the envelope parameters and Arias intensity.

Using Baker’s correlation between $S_a$ and Arias intensity we can incorporate the model into common frameworks.
Conditioned upon the actual Arias intensity for the recording.
Peak Factor Distribution

- The peak factor relating the maximum of the envelope to the peak acceleration is well modelled by a lognormal distribution with an average value of around 2.5.
- The parameters of this distribution are effectively uncorrelated with those of the envelope function.
Multi-purpose model

- A remarkable feature of the model is its ability to be used to also predict PGA – obtain results that are very consistent with NGA both for median and standard deviation.
Regional Earthquake Loss Estimation

• Work closely with the reinsurance industry
• Currently addressing three key themes associated with regional loss estimation
  – The degree of epistemic uncertainty associated with ELE
  – Improved characterisation of ground motions and damage using vector-based approaches
  – Simulation of stochastic event sets

• Briefly outline aspects of the above themes that relate to discussions yesterday
Epistemic uncertainty in ELE

• Yesterday the issue of model comparisons was raised and a good point was made that the best model is not necessarily the most complex, or theoretically robust, but may instead be the model that we are actually able to apply
• We are currently building an earthquake loss model for Tunisia (for commercial reasons)
  – Very little data exists for any part of the problem (hazard, vulnerability and exposure)
  – Apply multiple approaches where each approach is necessarily missing certain key pieces of information
  – Ultimately develop a framework with multiple modelling approaches implemented within a logic-tree framework whereby weights are applied to approaches (as well as to branches within a given approach)
  – Allow these weights to be tuned/updated given the available constraints
  – These issues are very relevant for GEM
Improved characterisation of ground motion

• Epistemic uncertainty and aleatory variability of ground motions overwhelmingly dominates the uncertainty in loss estimates - how accurate do we need a priori distributions to be for regional ELE? Does it matter?

• Regional differences among ground-motions may not be as large as had been thought
  – Robust models, like the NGA models, may be applied in Europe and elsewhere
  – There is little justification for using poorly constrained national models

• The macro-spatial correlation of ground motion should be accounted for, and should also be coupled with period-to-period correlations

• The issues associated with the ergodic assumption must be resolved – by how much can we reduce sigma?
Stochastic event sets for Loss Estimation

• Yesterday we saw discussion regarding time dependence and multiple events with respect to aftershock hazard and multi-hazard approaches
• The reinsurance industry is extremely interested in these sorts of problems but they work over a very short time-frame – usually on the basis of one year
• Early loss estimation was just built on top of PSHA
  – We know that you can’t do this for portfolios due to the lack of consideration of spatial correlation and the requirement to consider inter- and intra-event variance explicitly in loss estimation
  – Ground-motion modelling has changed, but the seismic source models have not (with some exceptions of time-dependent hazard and stress transfer models)
• Stochastic event set generation from source models derived using a Poisson assumption is probably not appropriate
How many time-history analyses are required?

• It is well known that code requirements for the numbers of time-history analyses that must be made are not based upon any robust statistical analysis
• A question that commonly arises is ‘how many time histories do I need to perform?’
• The answer depends largely upon what you’re trying to achieve with an important distinction being made between trying to obtain the median structural response or some estimate of the distribution of the response
• The work regarding the number of records required to estimate the median response is based upon the doctoral work of Jon Hancock
Establishing a point of reference

- The PEER GMSM group recently completed a comprehensive analysis in which the point-of-reference (true drift) is based upon ~100 natural records that were consistent with some M-R design scenario.
- The approach that we’d previously taken was similar in concept, but rather than only consider a particular scenario, we consider a very large number of scenarios.
- Run 1666 time history analyses for single frames using unscaled records:
  - 8 storey concrete wall-frame (median) [Seismostruct]
  - 6 storey concrete frame (distribution) [OpenSees]
- Derive empirical regression equations for Sa and the EDPs.
Median Case

- Consider multiple EDPs
  - Peak roof drift
  - Peak inter-storey drift, for all storeys
  - Peak member end rotation
  - Member rotational fatigue
  - Absorbed hysteretic energy
  - Park and Ang damage index

- Consider multiple selection and modification strategies (all consisting of restricting the magnitude range and considering spectral shape)
Selection & Modification Approaches

Scaled to $\text{Sa}(T_1)$

Scaled to $0.5\text{Sa}(T_1)$ to $2\text{Sa}(T_1)$

Spectrum Matched to 5%

Spectrum Matched to 1%, 5%, 10% & 20%

Earthquake Engineering by the Beach, Capri 2009
Effect of Multiple Damping Ratios

- The influence of duration on structural response has differing degrees of significance depending upon the EDP that is considered.

- Peak responses are insensitive to duration while energy-based responses are sensitive.

- Damping modification factors to alter response spectral ordinates from 5% damping to other damping ratios have been derived and are shown to be strongly dependent upon duration.

- The effect of matching to multiple damped spectra leads to greater consistency with duration.
Numbers required

- Numbers of records required to be within 10% of the ‘true’ median response at 68% confidence level

<table>
<thead>
<tr>
<th>EDP</th>
<th>Element/Location</th>
<th>Scale $Sa(T_1)$</th>
<th>Scale $Sa(T_{ave})$</th>
<th>Match 5%</th>
<th>Match Multiple</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drift</td>
<td>1st Storey</td>
<td>6</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>End Rotation</td>
<td>Base 1st floor column</td>
<td>6</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Fatigue</td>
<td>Base 1st floor column</td>
<td>82</td>
<td>44</td>
<td>28</td>
<td>14</td>
</tr>
<tr>
<td>Hysteretic energy</td>
<td>Base 1st floor column</td>
<td>54</td>
<td>15</td>
<td>11</td>
<td>5</td>
</tr>
<tr>
<td>Park &amp; Ang</td>
<td>Base 1st floor column</td>
<td>7</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>
Distribution of drift response

• Intuitively one would assume that many more records are required in order to estimate the distribution of an EDP

• We have conducted a preliminary study that has yielded some promising results that indicate that the numbers required may not be as high as one might suspect

• The recent work of the PEER GMSM team has confirmed that, when it comes to estimating drift response, spectral shape is fundamentally important

• Propose an approach based upon spectral shape and 3-point approximations to distributions
Degree of nonlinearity

- We are limited in terms of the degree of nonlinearity by the actual records that are available (the PEER NGA database is used).
- However, we are able to observe significant degrees of nonlinear response.

Fundamental period of 0.93s

Target median drift is around 0.4% corresponding to moderate to extensive damage for this frame.
Correlations between residuals of $\text{Sa}(T_1)$ and drift

When $\text{Sa}(T_1)$ is above average, drift is above average
Sa(T₁) and Spectral Shape

• It is well known (thanks to the Stanford group) that Sa(T₁) by itself is not a particular efficient estimator of inelastic response – one needs to correct amplitude and spectral shape.
• The undoubted success of the CMS clearly shows this.
• If you took the same CMS and scaled it up and down you would expect the drift response to also scale up and down.
• The question is whether one can map scaling in terms of spectral ordinates across to scaling of drift response.
• The very strong correlations that are observed among the residuals (for all levels of imposed demand) suggest that such a mapping is possible.
Gauss-Hermite Quadrature

We use three and five point approximations to continuous distributions – Equi-probability and Gauss-Hermite polynomials.
Moment Estimation

- We assume that drift is lognormally distributed and verify this assumption using statistical tests.
- We take the response estimates obtained from the scaled records and combine them using the weights that are associated with the discretisation of the distribution.
- Estimate the first two moments and their associated confidence intervals.
- We find that the Gauss-Hermite approach works extremely well in this case.
- This must be tested for different scenarios, different structures etc, but it looks promising.
3-point approximations

![Graphs showing drift and standard deviation of logarithmic drift for different levels of records per level and approximations. The graphs compare reference levels, 3-point EP without replacement, and 3-point EP with replacement.]
5-point approximations

![Graphs showing 5-point approximations for drift and standard deviation of logarithmic drift vs. number of records per level.](image)

Earthquake Engineering by the Beach, Capri 2009

Stafford
Conclusions

• Accounting for correlations in many aspects of earthquake engineering yield significant benefits

• Significant work is required to develop the empirical models that are then used to develop the correlation models

• Future work is to be focussed upon identifying important combinations of IMs for different applications

• Further work is also required for record selection
  – Orientation dependence
  – Spectrum matching, peak-to-peak variability, etc