

Uncertainties in GMPEs, Effect of Non-Ergodic Models

Seismic Hazard in California using Non-Ergodic GMPEs



AUTHORS		REVIEW		APPROVAL	
Name	Date	Name	Date	Name	Date
<i>N. Abrahamson, N. Kuehn, M. Walling, and N. Landwehr</i>	<i>2018/05/25</i>		<i>YYYY/MM/DD</i>	<i>Albert Koeche</i>	<i>YYYY/MM/DD</i> <i>2020/12/21</i>

DISSEMINATION: This document must not be distributed to any person, institution or company other than members of SIGMA-2 steering and scientific committees, except under written formal permission of SIGMA-2 steering committee.

Document history

DATE	VERSION	COMMENTS
2018/05/25	0	<i>Initial document.</i> <i>Based on draft of Journal paper (probabilistic seismic hazard analysis in California using non-ergodic ground-motion prediction equations, Abrahamson, Kuehn, Walling, and Landwehr)</i>

Preliminary version

	Research and Development Program on Seismic Ground Motion	Ref : SIGMA2-2018-D5-006/1
		Page 3/25

Executive Summary

The large increase in ground-motion data sets shows that ergodic GMPEs significantly overestimate the aleatory variability. Multiple studies have shown that after removing the repeatable source, path, site effects, the aleatory variability is about 0.4 In units. This value of the nonergodic aleatory variability is stable for different regions of the world. The issue is that to use this lower value of the aleatory variability in seismic hazard studies, the site/source specific source, path, and site terms need to be estimated, including the epistemic uncertainty in the estimated values.

A methodology for developing a complete non-ergodic ground-motion model for California, including estimates of the epistemic uncertainties is described. Three different concepts are combined into this method: (1) modeling the epistemic uncertainty in the base scaling of the regional GMPE as a continuous distribution used for Diablo Canyon and described in Geopentech (2015), (2) modeling the spatial variability of the GMPE within a region using the spatially varying coefficient model (VCM) of Landwehr et al (2016), and (3) modeling the path-specific terms using concept of cell-specific attenuation for the large distance scaling as describe by Dawood and Rodriquez-Marek (2013). For the last two, a method for modeling the epistemic uncertainty in the VCM and cell-specific attenuation is described.

The PHSA code HAZ45 was modified to be able to use fully non-ergodic GMPEs. Example hazard calculations for three sites in California are used to compare the mean hazard and the epistemic uncertainty range of the hazard for ergodic and non-ergodic GMPEs. This comparison shows the large increase in epistemic uncertainty for sites site with little to no data (but no change in the mean hazard). For sites with some ground-motion data to constrain the non-ergodic GMPE, the mean hazard can increase or decrease depending on the location of the site. The width of the epistemic uncertainty range depends on the amount of ground-motion available in the site region. To reduce the epistemic uncertainties requires additional ground-motion data from dense arrays of stations or from numerical simulations using 3-D crustal models.

The data collected over the past decade show that ground motions do not follow an ergodic model. The non-ergodic approach should be used in both regions with large amounts of data and in regions with sparse data. The key is that lack of data does not mean certainty; lack of data implies large uncertainty. For regions with sparse data, the aleatory variability will still be the non-ergodic value of about 0.4, but there will be large epistemic uncertainty in the site/source specific source, path, and site terms. For regions without data, the covariance structure of the source, path, and site terms from the regions with sufficient data can be used. That is, an ergodic assumption can be made on the covariance, but not on the median values.

1 Introduction

Probabilistic seismic hazard analysis (PSHA) requires a ground-motion model to describe the range of ground motions that can occur for a given earthquake scenario. Most often, the ground-motion model is given by empirical ground-motion prediction equations (GMPE) which describe the median and standard deviation of the ground-motion parameter for the given scenario. In PSHA, at long return periods, the hazard is strongly dependent on the value of the standard deviation as it controls the slope of the hazard curve (Bommer and Abrahamson, 2006). Because of the limited number of ground-motion recordings from large magnitude earthquakes at short distances, GMPEs have traditionally been developed using global data sets that combine the ground-motion data from similar tectonic regions around the world. Typically, three broad tectonic categories are used for GMPEs: active crustal regions, stable continental regions, and subduction zones. Within each broad tectonic category, the ground motion is assumed to follow the same scaling with magnitude, distance, and site condition (i.e. for a given scenario, the median and aleatory variability of the GMPE are assumed to be applicable to any location within the broad tectonic category). This is called the ergodic assumption (Anderson and Brune, 1999).

In the past decade, ground-motion data sets have grown rapidly resulting in data sets with recordings from multiple earthquakes at specific recording stations. Many researchers have found that for a specific site and for earthquakes with located within a small region, the aleatory variability of the ground motion is much smaller than the aleatory variability of global models based on the ergodic assumption (e.g. Atkinson, 2006; Hiemer et al., 2011; Lin et al., 2011; Morikawa et al., 2008; Yagoda-Biran et al., 2015). These studies have shown that for a specific site and earthquake pair, the variance of the aleatory variability is only 30-40% of the ergodic variance. This means that most of the variability treated as randomness in the ergodic approach is actually due to systematic source, path, and site effects. For example, using the NGA-W2 data set, Landwehr et al. (2016) found that the magnitude-independent aleatory variability has a standard deviation in the range of 0.68 to 0.87 natural log units for an ergodic model for California; whereas, the standard deviation is in the range of 0.46 to 0.54 natural log units for a non-ergodic model.

In addition to the empirical data, numerical simulations with 3-D crustal models have been

used to estimate the path and site terms. In Southern California, Cybershake has been used to compute the non-ergodic hazard (Graves et al., 2011). An analysis of the variance components from the Cybershake simulations showed that the single-path variability was about 30% of the variance measured over all sites and paths (Wang and Jordan, 2014), consistent with the empirical results.

This difference in the ergodic and non-ergodic aleatory variability has a large effect on the seismic hazard as shown by the following simplified example for a single earthquake scenario. Consider a site in which the hazard is controlled by a single nearby fault with a single representative scenario: M7 earthquake at a distance of 10 km with a recurrence interval of 200 years. The ground-motion distribution for an ergodic ground-motion model is shown together with the distributions from three alternative non-ergodic models (Figure 1). The three non-ergodic distributions have different values of the non-ergodic median for this scenario (source+path+site terms) and exhibit a lower aleatory variability than the ergodic model. These three estimates of the non-ergodic term represent three possible logic tree branches. Assuming a recurrence interval of 200 years for the scenario, Figure 1 shows the resulting hazard curves for the ergodic model and the three non-ergodic branches. Each of the non-ergodic hazard curves are steeper compared to the ergodic one due to the lower aleatory variability and they span a wide range of hazard due to the uncertainty in the median ground motion for the scenario at the specific location. This large uncertainty range is typical in the initial phase of moving from ergodic to non-ergodic hazard analyses: with limited data to constrain the location-specific median ground motion, there is an increase in the epistemic uncertainty of the hazard (Walling and Abrahamson, 2012).

With the large increase in ground-motion data sets, we now know that the standard deviation of the systematic location-specific effects on the ground motion is large compared to the remaining aleatory variability. We know that the ground motions at a specific site from a specific earthquake will have a narrow distribution as shown by the non-ergodic models, but the median will, in general, be different from the global model median. If there is no site-specific data available to estimate the source- and site-specific median, then there will be a large epistemic uncertainty in the value of the median ground motion. As new data are collected and the site/source specific terms become better constrained, the epistemic uncertainty in the median

can be reduced . Walling and Abrahamson (2012) computed fully non-ergodic hazard showing the uncertainty for example cases in which there was no site-specific information available and for which there was some limited site-specific information available. This increased epistemic uncertainty is not captured in the current ergodic models and it highlights a short-coming in traditional hazard analyses based on ergodic ground-motion models.

The fully non-ergodic model considers the systematic source, path, and site effects. A partial non-ergodic model that only removes the systematic site effects has been used in hazard applications over the past five years (e.g. PRP, South Africa, SWUS, BCHydro, Hanford). If only the systematic site terms are addressed, then about 30% of the variance of the ergodic model is removed as compared to 60-70% for the fully non-ergodic case. The reduced aleatory variability for this partially non-ergodic approach is called the single-station sigma. Because site-specific site response calculations are commonly conducted, the partially non-ergodic approach is straightforward to implement (e.g. Rodriguez-Marek et al., 2014). For this type of partially non-ergodic approach, the hazard is computed for a reference rock condition using the ergodic median and the single-station sigma for the aleatory variability. The site-specific site amplification and its epistemic uncertainty are then computed using traditional site response studies methods. Combining the single-station sigma reference rock hazard with the site amplification (including the epistemic uncertainty in the site amplification) leads to a partially non-ergodic hazard curve.

As ground-motion data sets have grown over the past decade, there has been a trend of moving from ergodic to non-ergodic GMPEs. For example, the 2008 NGA-W1 GMPEs (Power et al., 2008) were fully ergodic models with the same models being applied to all regions within the same broad tectonic category. The 2014 NGA-W2 GMPEs (Bozorgnia et al., 2014) had access to much larger data sets. In these expanded data sets, it became clear that there were strong regional differences to the ground-motion scaling. For example, the large distance scaling between Japan and California was different. The NGA-W2 models included regional differences in up to four terms of the GMPEs: constant term, large distance scaling, the VS30 scaling, and the basin depth scaling. Other studies have included regional terms in which the regional terms are modeled as random effects (Kotha et al., 2016; Kuehn and Scherbaum, 2016; Stafford, 2014).

In the studies mentioned above, regional differences have been demonstrated for broad regions. This raises the question as to the scale for regionalization. For example, if Japan and California have different scaling, are there differences in the scaling within these broad regions? Landwehr et al. (2016) used the NGA-W2 Californian data subset used by Abrahamson et al. (2014) to regionalize the GMPE using the spatially-varying coefficient model in which the coefficients of the GMPE depend on the longitude and latitude of the site and the earthquake (Bussas et al., 2017). A large part of the NGA-W2 data in California is from small and moderate earthquakes which are useful for constraining path and site effects and average source effects, but not for large magnitude scaling. Therefore, Landwehr et al. (2016) regionalized the constant, distance, and site scaling terms but maintained the global magnitude scaling everywhere. They found that the regional differences in the source and path terms had correlation lengths of about 20 km, whereas, the regional differences in the site terms had correlation lengths of just a few km.

The varying coefficient model of Landwehr et al. (2016) models path effects as an average over all directions for an event, which means that the path effects are not directional. Dawood and Rodriguez-Marek (2013) proposed a method to include source-to-site specific path effects, which models the attenuation along a specific path as the sum of attenuation in small cells, each of which has its own cell-specific attenuation coefficient.

In this paper, we develop a fully non-ergodic GMPE combining the varying coefficient model with the large distance scaling from the cell-specific approach. To demonstrate the approach, we use the model to compute non-ergodic hazard for three sites in California.

2 Non-Ergodic PSHA

The goal in PSHA is to calculate the expected rate of exceedance for a ground-motion level A

$$\nu(Y > A) = \sum_i \nu_i \int_M \int_R f_{m_i}(m) f_{r_i}(m, r) P(Y > A|m, r, \dots) dr dm \quad (1)$$

where Y is the ground-motion parameter of interest, and $P(Y > A|m, r, \dots)$ is the probability that Y is larger than A for a given magnitude, distance and other relevant parameters. The functions $f_{m_i}(m)$ and $f_{r_i}(m, r)$ describe the densities of magnitudes and distances for the i^{th}

source. The exceedance probability can be calculated from the cumulative distribution function (CDF) of a standard normal distribution, $\Phi(x)$.

$$P(Y > A|m, r, \dots) = 1 - \Phi\left(\frac{\ln(A) - \mu(m, r, \dots)}{\sigma(m, r)}\right) \quad (2)$$

where the distribution of $\ln(Y)$ is typically modeled as a normal distribution with median μ and standard deviation σ . The median and standard deviation are functions of the source and site related parameters

$$\mu = g(\vec{\theta}; m, r, \dots) \quad (3)$$

$$\sigma = h(\vec{\beta}; m, r, \dots) \quad (4)$$

where $\vec{\theta}$ and $\vec{\beta}$ are coefficients of the models for the μ and σ given by f and g . Traditionally, the coefficients of the empirical GMPEs are estimated by a regression analysis using the ergodic assumption (Anderson and Brune, 1999), which means that the $P(Y|m, r, \dots)$ is the same for all sites with the same site conditions.

An ergodic GMPE is an average over all sources, paths, and sites of the underlying data set. This leads to an inflated estimate of aleatory variability, because systematic effects are not modeled (Lin et al., 2011). To move from an ergodic to a non-ergodic PSHA, the repeatable systematic source, path and site effects need to be incorporated in the GMPE. This means that $P(Y|m, r, \dots)$ becomes explicitly dependent on source and site coordinates, which is achieved by making the median ground motion dependent on the coordinates of the earthquake and the site. Compared to an ergodic GMPE, there are source and site specific adjustment terms that are added to the ergodic median prediction.

$$\mu_{nonerg}(m, r, \dots, \vec{x}_{src}, \vec{x}_{site}) = \mu_{erg}(m, r, \dots) + \delta S(\vec{x}_{site}) + \delta L(\vec{x}_{src}) + \delta P(\vec{x}_{src}, \vec{x}_{site}) \quad (5)$$

where \vec{x}_{src} and \vec{x}_{site} are the source and site coordinates, respectively. δS is a location-specific site effect, δL is a location-specific source effect, and δP is a location-specific path effect.

The inclusion of the systematic, location-specific effects in eq 5 leads to a reduction in the value of aleatory variability (Al-Atik et al., 2010; Lin et al., 2011). This is an example of how

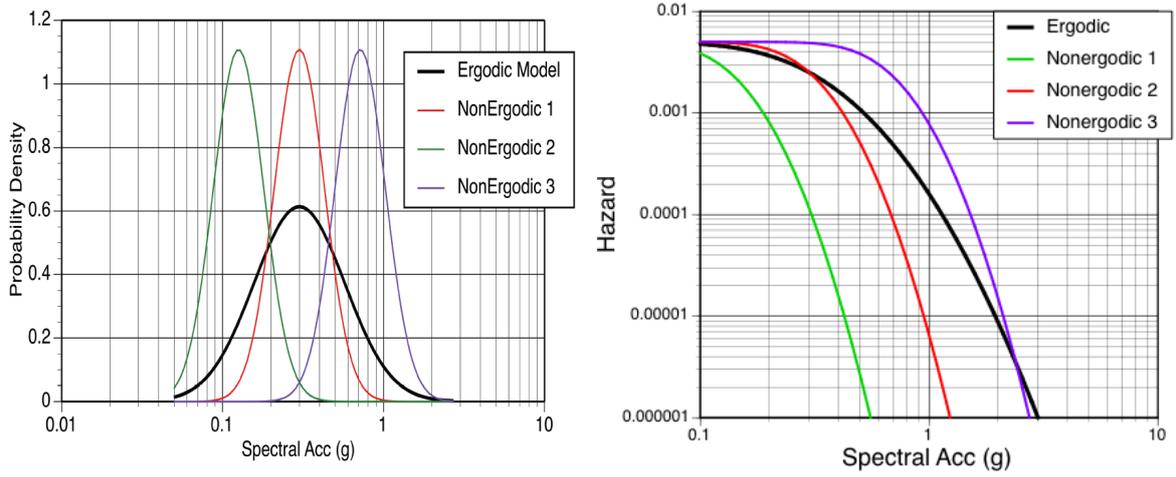


Figure 1: Non-ergodic PSHA

the separation of aleatory variability and epistemic uncertainty is not absolute and depends on the model parameterization. The effects of parameters not included in the model are treated as aleatory variability. Therefore, as additional significant predictive parameters are added to the model, there will be reduced aleatory variability

Because the non-ergodic GMPE includes location-specific site, path, and source effects as part of the median model, the epistemic uncertainties in these location-specific effects need to be considered. If the available data sets are sparse for the specific source/site pair, then the systematic source, path, and site effects can have large epistemic uncertainties that are important to include in the PSHA calculations. This can be done in a logic tree framework in which different values of the adjustment terms for each source/site pair occupy the logic tree branches. For non-ergodic ground-motion models, each logic tree branch is a map of the adjustment terms.

3 Non-Ergodic GMPE

In this section, we briefly describe the non-ergodic GMPE used to perform the hazard calculations. The non-ergodic GMPE follows the principles described in Equation (5) and consists of an ergodic base GMPE that is adjusted by source, path, and site terms that explicitly depend

on the source and site locations. The components of the non-ergodic GMPE are

$$\mu_{erg} = f_{base}(M, R_{RUP}, F, V_{S30}, Z_{TOR}) + f_{HW}(M, R_{RUP}) + f_{NL-site}(V_{S30}, PSA_{1100}) \quad (6)$$

$$\delta S(\vec{x}_{site}) = f_{site}(V_{S30}; \vec{x}_{site}) + \delta\theta_{0A}(\vec{x}_{site}) \quad (7)$$

$$\delta L(\vec{x}_{src}) = f_{geom}(R_{RUP}; \vec{x}_{src}) + \delta\theta_{0B}(\vec{x}_{src}) \quad (8)$$

$$\delta P(\vec{x}_{cls}, \vec{x}_{site}) = f_{attn}(R_{RUP}; \vec{x}_{cls}, \vec{x}_{site}) \quad (9)$$

In this non-ergodic GMPE, there are five spatially dependent terms: the geometrical spreading $f_{geom}(R_{RUP}; \vec{x}_{src})$, the linear V_{S30} -scaling $f_{site}(V_{S30}; \vec{x}_{site})$, one site/source dependent constant each, $\delta\theta_{0A}(\vec{x}_{site})$ and $\delta\theta_{0B}(\vec{x}_{src})$, and the anelastic attenuation $f_{attn}(R_{RUP}; \vec{x}_{cls})$.

There are two coordinates used for the source in Equation (9). For the source term, \vec{x}_{src} is the coordinate of the center of the rupture. For the path term, \vec{x}_{cls} is the coordinate of the closest point on the rupture to the site.

A suite of alternative ergodic base GMPE (Equation (6)) is derived based on the methodology described in the Southwestern United States (SWUS) ground-motion model report (Geopentech 2015). The suite captures the epistemic uncertainty in the base scaling of the ergodic GMPEs. The adjustment terms of Equation (7) and (8) are taken from Landwehr et al. (2016), while the anelastic attenuation term (Equation (9)) is calculated similar to Dawood and Rodriguez-Marek (2013). These three parts of the ground-motion model are described in the subsequent sections.

3.1 Ergodic Base GMPE

The suite of ergodic base GMPEs are derived based on the methodology used in the SWUS project (Geopentech 2015). The goal of the method is to obtain a good representation of the epistemic uncertainty in the main components of the scaling of the response spectrum with the important predictor variables. The basic assumption is that there is a continuous distribution of GMPEs, which can be derived from the range of published models that are considered applicable to the southwestern US region.

The continuous distribution of GMPEs is approximated by the joint distribution of coeffi-

coefficients of a GMPE function which is estimated by fitting a common functional form to a set of published GMPEs, and then estimating the variances and correlations between the obtained sets of coefficients. This is done in the following way. First, we generate response spectral values for a suite of scenarios with different magnitude, distance, V_{S30} , Z_{TOR} and focal mechanism values F using the five NGA West 2 GMPEs (Abrahamson et al., 2014; Boore et al., 2014; Campbell and Bozorgnia, 2014; Chiou and Youngs, 2014; Idriss, 2014). For each GMPE, this gives a data set $D_i = \{(M_1, R_1, V_{S30,1}, Z_{TOR,1}, F_1, y_{i,1}), \dots, (M_N, R_N, V_{S30,N}, Z_{TOR,N}, F_N, y_{i,N})\}$, where N is the number of scenarios and y_i is the vector of median ground motions computed using GMPE i . Second, each synthetic data set from the underlying GMPEs is fit to the following functional form

$$\ln y = \theta_0 + g_M(M) + (\theta_4 + \theta_5(M - 5)) \ln \sqrt{R_{RUP}^2 + \theta_6^2} - \theta_7^2 R_{RUP} + \theta_8^2 Z_{TOR} + \theta_9^2 F_R - \theta_{10}^2 F_{NO} - \theta_{11}^2 \ln \frac{V_{S20}}{760} \quad (10)$$

$$g_M(M) = \begin{cases} -\theta_1 + \theta_2(M - 5.5) & M < 5.5 \\ \theta_1(M - 6.5) & 5.5 \leq M \leq 6.5 \\ \theta_3(M - 6.5) & M > 6.5 \end{cases}$$

where $F_R = 1$ for reverse faulting and $F_{NO} = 1$ for normal faulting. Some of the coefficients are squared to constrain the sign of the coefficient to be consistent with physical constraints.

For each GMPE i , there is a vector $\theta_i = \{\theta_0, \dots, \theta_{11}\}$ of coefficients. These coefficient vectors are assumed to be distributed according to a multivariate normal distribution

$$\begin{pmatrix} \theta_0 \\ \vdots \\ \theta_{11} \end{pmatrix} \sim N \left(\begin{pmatrix} \mu_{\theta_1} \\ \vdots \\ \mu_{\theta_{11}} \end{pmatrix}, \begin{pmatrix} \Sigma_{1,1} & \dots & \Sigma_{1,11} \\ \vdots & \ddots & \vdots \\ \Sigma_{11,1} & \dots & \Sigma_{11,11} \end{pmatrix} \right) \quad (11)$$

where the means and covariances of the coefficients can be estimated from the fitted sets of coefficients. Equation (11) describes a continuous distribution over coefficients for a GMPE parameterized by Equation (10). Therefore, it describes a continuous distribution of GMPEs. For the hazard calculation, we sampled 100 sets of coefficients θ from their distribution to cover the epistemic uncertainty in the base scaling of the GMPE. The magnitude and distance scaling of the 100 sampled base GMPEs, as well as the underlying five NGA West2 models, is shown

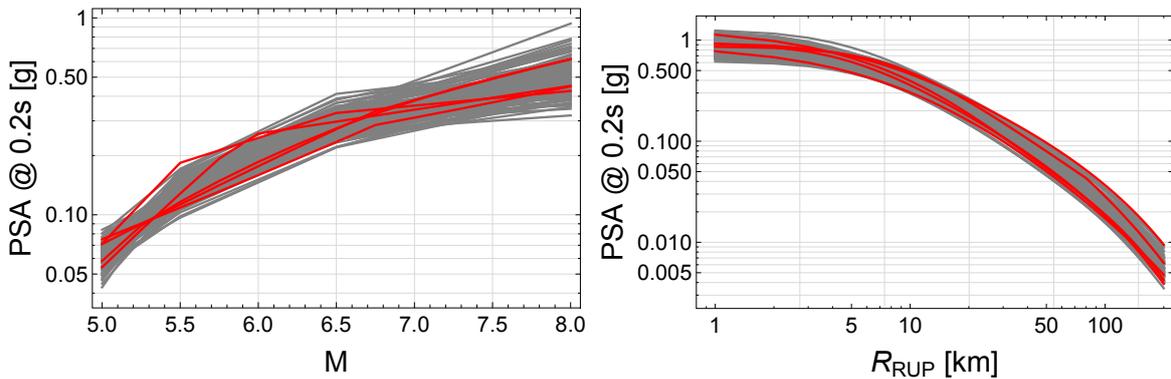


Figure 2: Scaling of base GMPEs, for $R_{RUP} = 20$ and $M = 6$.

in Figure 2.

In addition to the base model scaling, nonlinear site effects and hanging-wall effects are included in the suite of ergodic GMPEs. The nonlinear site amplification term is taken from the model of [Abrahamson et al. \(2014\)](#) and is applied without epistemic uncertainty. The hanging-wall term is taken from the SWUS project ([Geopentech, 2015](#)) which consists of five equally probable HW models that capture the epistemic uncertainty in the hanging-wall effects. Each sampled base set of coefficients is randomly paired with one of the five hanging-wall models.

3.2 Local Adjustment Terms from [Landwehr et al. 2016](#)

In this section, we describe the local adjustment terms (dependent on source or site coordinates) of Equation (7) and (8). These comprise an adjustment to the geometrical spreading, the linear site scaling, and the constant term. The constant terms has two parts: one term for the source coordinate and one term for the site coordinate:

$$f_{geom}(R_{RUP}; \vec{x}_{source}) = \delta\theta_4(\vec{x}_{src}) \sqrt{R_{RUP}^2 + \theta_6^2} \quad (12)$$

$$f_{site}(\vec{x}_{site}; V_{S30}) = \delta\theta_{11}(\vec{x}_{site}) \ln \frac{V_{S30}}{760} \quad (13)$$

$$f_{const} = \delta\theta_{0A}(\vec{x}_{site}) + \delta\theta_{0B}(\vec{x}_{src}) \quad (14)$$

We assume that the the coordinate-dependent adjustment coefficients $\delta\theta(\vec{x})$ from the spatially varying coefficient model (VCM) of [Landwehr et al. \(2016\)](#) are applicable to each of the 100 base GMPEs. In the VCM, each model coefficient is a continuous function of either source or site coordinate (cf. Figure 4 of [Landwehr et al. \(2016\)](#)). The VCM contains spatially varying

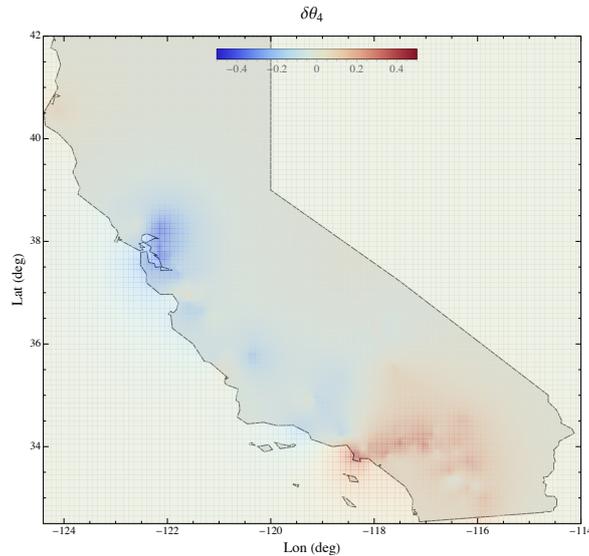


Figure 3: Spatially varying adjustment coefficients $\delta\theta_4(\vec{x}_{source})$, controlling geometrical spreading.

coefficients controlling the geometrical spreading and linear site scaling, as well as two constants describing source and site effects. To generate Figure 4 of Landwehr et al. (2016), these coefficients were calculated at a spatial grid over California of size 2×2 km. At each grid point, we calculate the difference between the coefficients at that grid point and the mean of the respective coefficient—this results in local adjustments for each of the four spatially varying coefficients. These adjustment terms are zero in regions where data is sparse, and potentially large close to observed events/stations.

To simplify PSHA calculations, we use a slightly larger grid than Figure 6 of Landwehr et al. (2016), dividing California into cells of size 5×5 km. For each cell, we calculate the mean adjustment coefficients for the four coefficients from all grid points inside that cell. As an example, the mean local adjustment term for the geometrical spreading, $\delta\theta_4$ is shown in Figure 3. For the rest of the spatially varying coefficients, we refer to Figure 6 of Landwehr et al. (2016).

It is important to capture the epistemic uncertainty in the median adjustments for the non-ergodic PSHA. The VCM of Landwehr et al. (2016) does not provide direct estimates of the uncertainty of the individual coefficients, but it does provide the total epistemic uncertainty, ψ , associated with ground-motion median predictions for different locations (cf. Figure 6 (b) and Equation (11) of Landwehr et al. (2016)). We use the spatially-varying values of ψ as estimates for the total epistemic uncertainty of all adjustment terms discussed in this section. Hence,

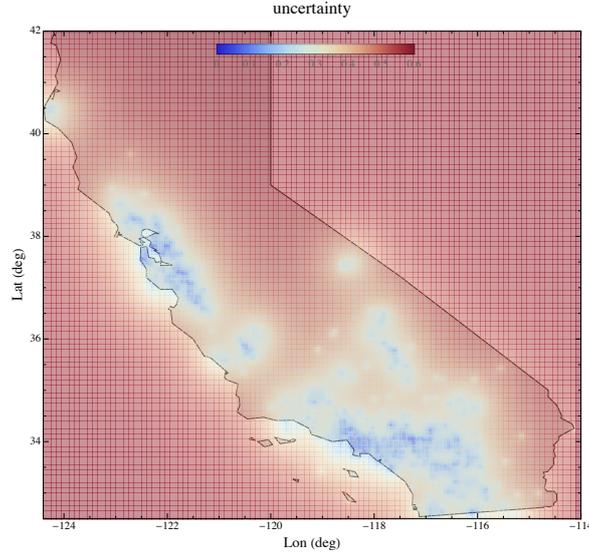


Figure 4: Epistemic uncertainty ψ associated with median predictions of [Landwehr et al. \(2016\)](#)

the epistemic uncertainty associated with the adjustment terms $\delta L(\vec{x}_{src})$ and $\delta S(\vec{x}_{site})$ is fully modeled as uncertainty in the constants $\delta\theta_{0B}$ and $\delta\theta_{0A}$, while the coefficients controlling the source-location dependent geometrical spreading and site-location dependent V_{S30} -scaling are modeled as error-free.

As an example, the total ψ -values for PGA are shown in Figure 4. We partition ψ into two parts, one accounting for source adjustments (σ_E) and one accounting for site adjustments (σ_S):

$$\psi^2 = \sigma_S^2 + \sigma_E^2 \quad (15)$$

$$\sigma_S^2 = \rho\psi^2 \quad (16)$$

$$\sigma_E^2 = (1 - \rho)\psi^2 \quad (17)$$

where the partitioning factor ρ depends on the number of events/station in each 5×5 km cell:

$$\rho = 0.2 + 0.6 \frac{1}{1 + \frac{n_S}{n_E}} \quad (18)$$

where n_S and n_E are the number of stations and events in a cell, respectively. Hence, if the number of stations in a cell is large compared to the number of events, most of the epistemic uncertainty is associated with the source adjustment and vice versa. We do not want to assign the full epistemic variance to either the event or station term, so the partitioning factor ρ is

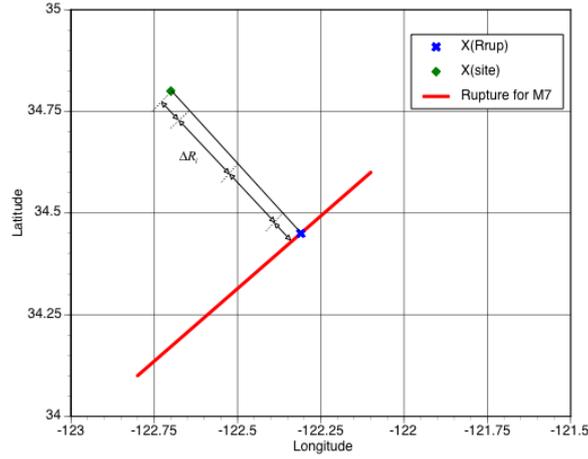


Figure 5: Calculation of cell distances

constrained to be between 0.2 and 0.8. In the future, we plan to use directly the uncertainties of the adjustment coefficients, which will avoid the problem of partitioning the variance ψ^2 , but those are currently not available.

The partitioning factor ρ is modeled as spatially correlated using an exponential correlation function (as in Landwehr et al. (2016)). Thus, we can partition the overall uncertainty ψ at every location into a source-related part and a site-related part.

3.3 Anelastic Attenuation Term

The calculation of the non-ergodic anelastic attenuation term (cf. Equation (9)) is based on the methodology proposed by Dawood and Rodriguez-Marek (2013). The region under study, in this case California, is divided into rectangular cells of size 28×30 km. For each recording in the data set, the length of the ray path, ΔR_i , within each cell i is calculated, based on a straight line from the site (\vec{x}_{site}) to the closest point on the rupture (\vec{x}_{cls}). This is conceptually shown in Figure 5. For each record, we have that $\sum_{i=1}^{N_C} \Delta R_i = R_{RUP}$, where N_C is the number of cells.

The anelastic attenuation is modeled by a cell-specific coefficient $\theta_{7,i}$, so that

$$f_{attn}(R_{RUP}; \vec{x}_{cls}, \vec{x}_{site}) = \sum_{i=1}^{N_C} \Delta R_i(\vec{x}_{site}, \vec{x}_{cls}) \theta_{7,i} \quad (19)$$

This makes it possible to include path-specific attenuation in the GMPE.

The cell-specific attenuation terms are estimated from the Californian data set of Abraham-

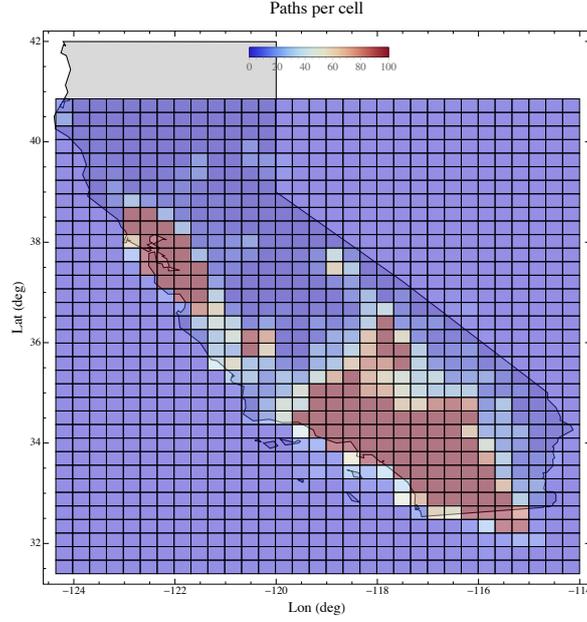


Figure 6: Number of paths

son et al. (2014). Figure 6 shows events, stations, paths and cell coverage. The ground-motion data are first corrected to remove the nonlinear site effects and hanging-wall effects using the model of Abrahamson et al. (2014). Using the corrected data, the residuals are computed with respect to the central ergodic base GMPE (cf. Equation (10)), where all coefficients are fixed to the values of the central model and the ergodic anelastic attenuation term $\theta_7 R_{RUP}$ is replaced with the non-ergodic one of Equation (19).

The model is cast as a Bayesian hierarchical model/multi-level model, where the individual cell-specific attenuation terms $\theta_{7,i}$ are assumed to be distributed according to a (truncated) normal distribution:

$$\theta_7 \sim N(\mu_{\theta,7}, \sigma_{\theta,7}) T(0,) \quad (20)$$

The prior distributions for the mean and standard deviation of the cell-specific attenuation parameters are a normal and a half-Cauchy distribution:

$$\mu_{\theta,7} \sim N(0, 1) T(0,) \quad (21)$$

$$\sigma_{\theta,7} \sim HC(0, 1) \quad (22)$$

where $\mu_{\theta,7}$ is a Californian mean attenuation parameter, and $\sigma_{\theta,7}$ describes how much the

individual coefficients can change across cells. To avoid a nonphysical attenuation term, the $\theta_{7,i}$ terms are required to be positive for each cell. This constraint is captured by the $T(0,)$ terms which means that the distribution is truncated to be greater than 0. The prior for the standard deviation is a half-cauchy distribution as recommended by [Gelman \(2006\)](#), denoted by the HC term.

The parameters are estimated via Bayesian inference using the program Stan [Carpenter et al. \(2016\)](#); [Team \(2015\)](#). The results are shown in Figure 7 in terms of the difference between the cell-specific attenuation and the mean attenuation, $\delta\theta_{7,i} = \theta_{7,i} - \mu_{\theta,7}$. The left panel shows the mean values of the $\delta\theta_{7,i}$ and the right panel shows the epistemic uncertainty on the value of $\delta\theta_{7,i}$ given in terms of the standard deviation of the $\delta\theta_{7,i}$.

The hierarchical nature of the model causes the mean of the cell-specific attenuation coefficients, $\mu_{\theta_{7,i}}$, to automatically move towards the mean over all cells ($\delta\theta_{7,i}$ goes to zero) for cells that do not have a lot of coverage. This can be seen in the left panel of Figure 7 which shows that cells with low path coverage (Figure 6) have mean cell-specific attenuation coefficients that are close to the Californian mean, while their epistemic uncertainty is large (as shown in the right panel). In these regions without data, the epistemic uncertainty in $\delta\theta_{7,i}$ becomes standard deviation of the posterior distribution of the $\theta_{7,i}$. If earthquakes at large distances contribute significantly to the hazard, then it is important to take both the path effects and the epistemic uncertainty in the path effects into account during hazard calculations.

4 Implementation for PSHA

In this section, we describe the steps necessary to perform a site-specific PSHA with the previously developed non-ergodic GMPE. Generally, the non-ergodic PSHA can be calculated like an ergodic one, except that one has to add the necessary adjustment terms for each source, as well as an adjustment term for the site.

To capture the epistemic uncertainty associated with the GMPE, 100 realizations are run—this corresponds to a logic tree with 100 branches for the ground-motion model. Each branch is a combination of an ergodic base GMPE (Equation (6)), a hanging-wall model, a site-specific adjustment term, source-specific adjustment terms, and cell-specific attenuation coefficients. The nonlinear site amplification and the source and site specific adjustment terms for geomet-

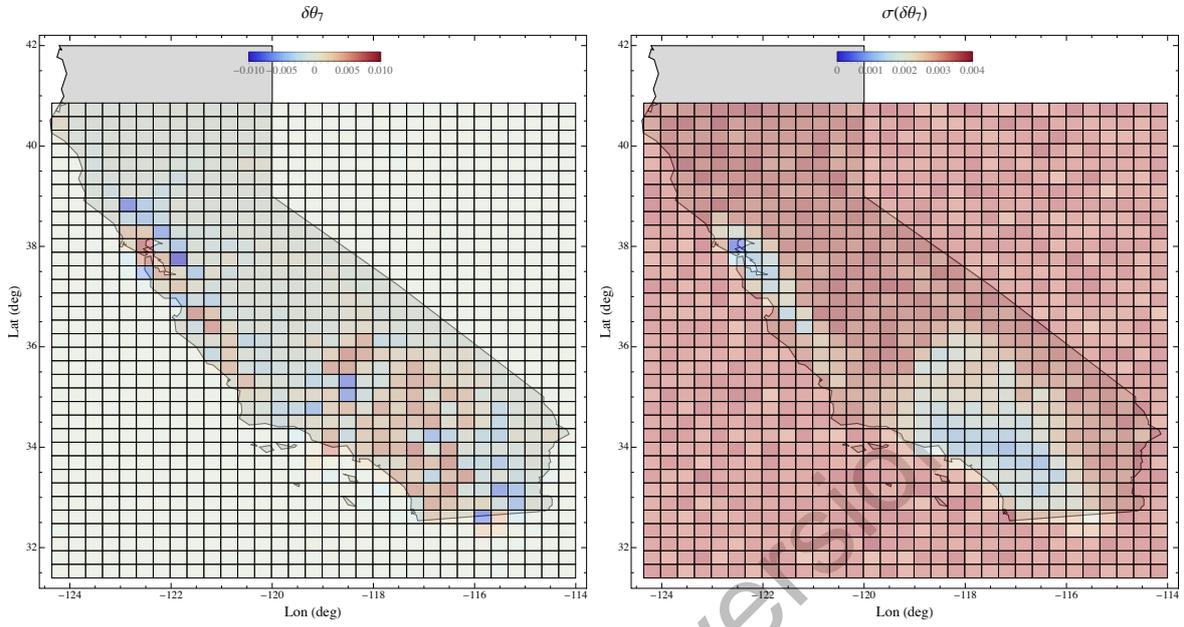


Figure 7: $\Delta\theta_7$ (left panel) and $\sigma\Delta\theta_7$ (right panel)

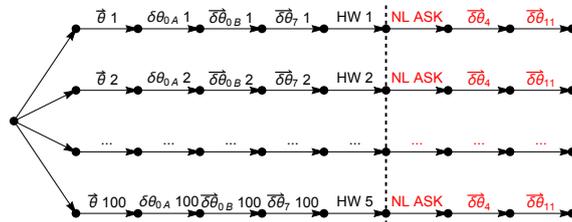


Figure 8: Nonergodic GMM logic tree

rical spreading and V_{S30} -scaling are the same for all 100 GMPE realizations. The GMPE logic tree is shown conceptually in Figure 8.

Before running the hazard calculation, we randomly sample 100 sets of the base coefficients from the distribution of coefficients estimated in Section [Ergodic Base GMPE](#), and pair them randomly with one of the five SWUS hanging-wall models. We also randomly sample 100 sets of constant source adjustments $\vec{\delta}\theta_{0B}(\vec{x}_{source})$ from a multivariate normal distribution with mean zero and a covariance matrix whose elements describe the spatial correlation of the source terms. The diagonal elements of the covariance matrix are σ_E^2 , the partitioned epistemic uncertainty from [Landwehr et al. \(2016\)](#) (see [Local Adjustment Terms from Landwehr et al. 2016](#) and Equation (15)), while the non-diagonal elements are calculated via an exponential covariance function whose length scale is taken from [Landwehr et al. \(2016\)](#). Next, we sample 100 sets of the cell-specific coefficients $\vec{\delta}\theta_7$ from their posterior distribution. Finally, we sample 100 site-specific adjustments $\delta\theta_{0A}$, from a normal distribution whose mean and standard deviation are determined by the cell in which the site is located. The mean corresponds to the value from [Landwehr et al. \(2016\)](#), while the standard deviation is σ_S (cf. Equation (15)).

All of these sampled sets of coefficients are generated outside of the hazard code and are read as tables. The adjustment terms for the geometrical spreading and the V_{S30} -scaling are the same across all 100 logic tree branches—their uncertainty is mapped into the constant terms.

The calculation of the distances within each cell for the cell-specific distance attenuation can be computationally intensive in a hazard calculation for a large number of scenarios. Therefore, we calculate the cell-specific distances to the site from each center point of each cell. Then, we calculate the adjustment term $f_{attn}(R_{RUP}; \vec{x}_{cell}, \vec{x}_{site}) = \sum_{i=1}^{N_C} \Theta_7^i \Delta R_i(\vec{x}_{site}, \vec{x}_{cell})$ for the center of each cell, where \vec{x}_{cell} is the center point of the cell. The adjustment term for the rupture coordinate (closest point on the rupture) is then calculated as an interpolation of the four closest cells.

With these steps, the hazard calculations can be done in a normal way. As with any hazard code, there is a loop over the alternative GMPEs. Here, we are using non-ergodic GMPEs in place of ergodic GMPEs. For the implementation of the non-ergodic GMPEs, there are some additional calculations that need to be done within the hazard code: one needs to find the index of the grids containing the site (\vec{x}_{site}) and the center of the rupture (\vec{x}_{src}) and the four

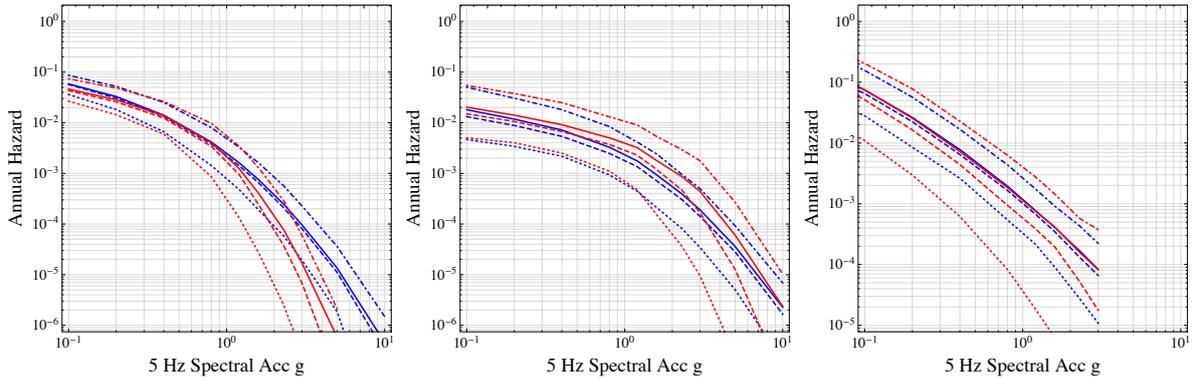


Figure 9: Hazard results: Left San Jose, middle SLO, right NE CA

cell around the closest point on the rupture plane (\vec{x}_{cls}). With these indexes, it is a simple lookup table for the non-ergodic terms.

5 Example PSHA Calculations

We calculate non-ergodic hazard for three different sites: San Jose, San Luis Obispo, and a north-eastern California site. For San Jose, there exists a significant amount of ground-motion data, for San Luis Obispo there exists some ground-motion data, whereas the north-eastern California site lacks data in its vicinity. The source model is taken from the PG&Es source models for these three regions. For comparison, we also calculate ergodic hazard for the three sites, using the 100 ergodic base models. For simplicity, the aleatory standard deviation is assumed to be magnitude independent value of 0.6 (ln units) for the ergodic case and 0.4 for the non-ergodic case. Magnitude-dependent sigma models can be used, but here we have only applied a constant sigma model.

The results are shown in Figure 9. For the north-eastern California site, the mean hazard does not change compared to the ergodic model. This is expected because there is no ground-motion data, so the average non-ergodic adjustment terms are zero and the reduction in the aleatory variability is offset by the increase in the epistemic uncertainty in the non-ergodic terms.

In the case of the other two sites, the mean hazard changes, and the epistemic uncertainty is reduced compared to the case of no data. For the San Jose site, the non-ergodic hazard at high exceedance rates is close to the ergodic hazard, but deviates strongly at low exceedance

rates.

6 Discussion and Conclusions

The large increase in ground-motion data sets shows that ergodic GMPEs significantly overestimate the aleatory variability. We know that after removing the repeatable source, path, site effects, the aleatory variability is about 0.4 ln units. This value of the non-ergodic aleatory variability is stable for different regions of the world. The issue is that to use this lower value of the aleatory variability in seismic hazard studies, the site/source specific source, path, and site terms need to be estimated, including the epistemic uncertainty in the estimated values. In the development of GMPE over the last decade, there is a clear move to non-ergodic models.

Site-specific site effects can be estimated using observed ground motions at the site or using analytical site response methods with epistemic uncertainties on the inputs. This is standard practice for critical facilities and it is used with the partially non-ergodic single-station sigma method. Because path effects in the crust are in the linear range, the site-specific path effects can be estimated empirically using recordings from small earthquakes. In addition, numerical simulations with 3-D crustal models can provide constraints on the path effects. This approach is being applied by SCEC as part of the Central California Seismic Project. Site-specific source effects are more difficult to estimate empirically than path effects because it is not yet been shown that source effects from small earthquakes in a region are correlated with source effects for large earthquakes

The aleatory variability has a strong effect on the seismic hazard at low probability levels typically used for design of structures. The move to non-ergodic ground-motion models will lead to the largest changes in seismic hazard estimates since the change to including the aleatory variability for ground-motion models in the 1980s. The non-ergodic approach should be used in both regions with large amounts of data and in regions with sparse data. The key is that lack of data does not mean certainty; lack of data implies large uncertainty. For regions with sparse data, the aleatory variability is the non-ergodic value of about 0.4, but there will be large epistemic uncertainty in the hazard due to the epistemic uncertainty in the site/source specific source, path, and site terms. The mean hazard will be unchanged from the ergodic model, but the non-ergodic approach will show the limitations of the available data which is

masked with the ergodic approach.

Using the non-ergodic approach will show how large are the current uncertainties in the hazard estimates. In many cases, the uncertainties will be large. To reduce the uncertainties requires additional ground-motion data: from dense arrays of stations or from numerical simulations using 3-D crustal models. Understanding the current limitations of the seismic hazard information due to the uncertainty in the non-ergodic ground-motion models will help to make the business case for seismic instrumentation and region-specific numerical simulations.

7 Acknowledgments

We would like to thank Brian Chiou for providing the points on the rupture closest to the stations for the NGA West 2 data set, used to calculate the rupture distances within the cells. This work was partially supported by the PG&E Long-Term Seismic Program and it is part of the PG&E contribution to the SIMGA2 project.

References

- Abrahamson, N. A., W. J. Silva, and R. Kamai (2014, 8). Summary of the ASK14 Ground Motion Relation for Active Crustal Regions. *Earthquake Spectra* 30(3), 1025–1055.
- Al-Atik, L., N. Abrahamson, J. J. Bommer, F. Scherbaum, F. Cotton, and N. Kuehn (2010, 8). The Variability of Ground-Motion Prediction Models and Its Components. *Seismological Research Letters* 81(5), 794–801.
- Anderson, J. G. and J. N. Brune (1999, 1). Probabilistic Seismic Hazard Analysis without the Ergodic Assumption. *Seismological Research Letters* 70(1), 19–28.
- Atkinson, G. M. (2006, 4). Single-Station Sigma. *Bulletin of the Seismological Society of America* 96(2), 446–455.
- Bommer, J. J. and N. A. Abrahamson (2006). Why Do Modern Probabilistic Seismic-Hazard Analyses Often Lead to Increased Hazard Estimates? *Bulletin of the Seismological Society of America* 96(6), 1967–1977.

- Boore, D. M., J. P. Stewart, E. Seyhan, and G. M. Atkinson (2014, 8). NGA-West2 Equations for Predicting PGA, PGV, and 5% Damped PSA for Shallow Crustal Earthquakes. *Earthquake Spectra* 30(3), 1057–1085.
- Bozorgnia, Y., N. A. Abrahamson, L. A. Atik, T. D. Ancheta, G. M. Atkinson, J. W. Baker, A. Baltay, D. M. Boore, K. W. Campbell, B. S.-J. Chiou, R. Darragh, S. Day, J. Donahue, R. W. Graves, N. Gregor, T. Hanks, I. M. Idriss, R. Kamai, T. Kishida, A. Kottke, S. A. Mahin, S. Rezaeian, B. Rowshandel, E. Seyhan, S. Shahi, T. Shantz, W. Silva, P. Spudich, J. P. Stewart, J. Watson-Lamprey, K. Wooddell, and R. Youngs (2014, 8). NGA-West2 Research Project. *Earthquake Spectra* 30(3), 973–987.
- Bussas, M., C. Sawade, N. Kühn, T. Scheffer, and N. Landwehr (2017, 4). Varying-coefficient models for geospatial transfer learning. *Machine Learning*.
- Campbell, K. W. and Y. Bozorgnia (2014, 8). NGA-West2 Ground Motion Model for the Average Horizontal Components of PGA, PGV, and 5% Damped Linear Acceleration Response Spectra. *Earthquake Spectra* 30(3), 1087–1115.
- Carpenter, B., A. Gelman, M. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. A. Brubaker, P. Li, and A. Riddell (2016). Stan: A Probabilistic Programming Language. *Journal of Statistical Software* VV(Ii).
- Chiou, B. S.-J. and R. R. Youngs (2014, 8). Update of the Chiou and Youngs NGA Model for the Average Horizontal Component of Peak Ground Motion and Response Spectra. *Earthquake Spectra* 30(3), 1117–1153.
- Dawood, H. M. and A. Rodriguez-Marek (2013, 5). A Method for Including Path Effects in Ground-Motion Prediction Equations: An Example Using the Mw 9.0 Tohoku Earthquake Aftershocks. *Bulletin of the Seismological Society of America* 103(2B), 1360–1372.
- Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian analysis* (3), 515–533.
- Graves, R., T. H. Jordan, S. Callaghan, E. Deelman, E. Field, G. Juve, C. Kesselman, P. Maechling, G. Mehta, K. Milner, D. Okaya, P. Small, and K. Vahi (2011). CyberShake: A Physics-

- Based Seismic Hazard Model for Southern California. *Pure and Applied Geophysics* 168(3-4), 367–381.
- Hiemer, S., F. Scherbaum, D. Roessler, and N. Kuehn (2011, 5). Determination of θ and Rock Site from Records of the 2008/2009 Earthquake Swarm in Western Bohemia. *Seismological Research Letters* 82(3), 387–393.
- Idriss, I. M. (2014, 8). An NGA-West2 Empirical Model for Estimating the Horizontal Spectral Values Generated by Shallow Crustal Earthquakes. *Earthquake Spectra* 30(3), 1155–1177.
- Kotha, S. R., D. Bindi, and F. Cotton (2016, 4). Partially non-ergodic region specific GMPE for Europe and Middle-East. *Bulletin of Earthquake Engineering* 14(4), 1245–1263.
- Kuehn, N. M. and F. Scherbaum (2016, 10). A partially non-ergodic ground-motion prediction equation for Europe and the Middle East. *Bulletin of Earthquake Engineering* 14(10), 2629–2642.
- Landwehr, N., N. M. Kuehn, T. Scheffer, and N. Abrahamson (2016). A Nonergodic Ground-Motion Model for California with Spatially Varying Coefficients. *Bulletin of the Seismological Society of America* 106(6), 2574–2583.
- Lin, P.-S., B. Chiou, N. Abrahamson, M. Walling, C.-T. Lee, and C.-T. Cheng (2011, 9). Repeatable Source, Site, and Path Effects on the Standard Deviation for Empirical Ground-Motion Prediction Models. *Bulletin of the Seismological Society of America* 101(5), 2281–2295.
- Morikawa, N., T. Kanno, A. Narita, H. Fujiwara, T. Okumura, Y. Fukushima, and A. Guerpinar (2008). Strong motion uncertainty determined from observed records by dense network in Japan. *Journal of Seismology* 12(4), 529–546.
- Power, M., B. Chiou, N. Abrahamson, Y. Bozorgnia, T. Shantz, and C. Roblee (2008). An overview of the NGA project. *Earthquake Spectra* 24, 3.
- Rodriguez-Marek, A., E. M. Rathje, J. J. Bommer, F. Scherbaum, and P. J. Stafford (2014). Application of single-station sigma and site-response characterization in a probabilistic seismic-

hazard analysis for a new nuclear site. *Bulletin of the Seismological Society of America* 104(4), 1601–1619.

Stafford, P. J. (2014, 3). Crossed and Nested Mixed-Effects Approaches for Enhanced Model Development and Removal of the Ergodic Assumption in Empirical Ground-Motion Models. *Bulletin of the Seismological Society of America* 104(2), 702–719.

Team, S. D. (2015). Stan: A C++ Library for Probability and Sampling, Version 2.5.0.

Walling, M. and N. A. Abrahamson (2012). Non-Ergodic Probabilistic Seismic Hazard Analyses. *15th World Conference on Earthquake Engineering (15WCEE)* (2008).

Wang, F. and T. H. Jordan (2014). Comparison of probabilistic seismic-hazard models using averaging-based factorization. *Bulletin of the Seismological Society of America* 104(3), 1230–1257.

Yagoda-Biran, G., J. G. Anderson, H. Miyake, and K. Koketsu (2015). Between-Event Variance for Large Repeating Earthquakes. *Bulletin of the Seismological Society of America* 105(4), 2023–2040.

Preliminary version

Review of “Uncertainties in GMPEs, Effect of Non-Ergodic Models: Seismic Hazard in California using Non-Ergodic GMPEs” by N. Abrahamson, N. Kuehn, M. Walling and N. Landwehr (SIGMA2-2018-D5-005/1, version 1)

The authors present a procedure for the development of non-ergodic ground motion prediction equations by combining three previous approaches. Next they demonstrate the procedure for California. Finally, they compare hazard curves from ergodic and non-ergodic approaches for three sites in California. The report is interesting and significantly moves on the state-of-the-art in this important field. The approach appears correct and promising for application elsewhere.

There are a number of editorial and technical issues that should be addressed in the next version of this deliverable (see below, where **bold** indicates the more important comments). Because the topic of this work is novel, there is little guidance on this topic in the public literature and the topic is poorly understood by the community, the authors are encouraged to spend some time expanding the explanations (perhaps by adding some more diagrams and examples) and being more verbose, particularly in figure captions. Finally, the resulting maps and graphs are interesting and warrant more care being spent on their production so that this work has the highest impact and encourages additional work in this field.

1. Title, “GMPE”: It would be best to avoid an abbreviation in the report title.
2. Front page: The reference number given in the top right-hand corner is “SIGMA2-2018-D5-005/1” whereas elsewhere it is given as “SIGMA2-2018-D5-006/1”. This needs to be made consistent (I think that the front page is correct).
3. Page 2: This report is stated to be based on a draft journal article but it is not clear if this article has been submitted nor to which journal. This should be clarified. The authors are encouraged to submit the manuscript.
4. Page 3, ”sites site”: Delete “site”.
5. Page 3: This Executive Summary would flow better with the last paragraph moved up to become the second paragraph.
6. Page 4, “it controls the slope of the hazard curve”: It would be useful to clarify that this is true if all other parameters are kept constant because the slope of the hazard curve is affected by things other than the standard deviation of the GMPE. It would be useful for non-experts to state in which direction the slope changes with changes in the standard deviation: higher standard deviations lower slope.
7. Page 6, “PRP, South Africa, SWUS, BCHydro, Hanford”): It would be useful to give the exact references for these studies as some of them will likely be unfamiliar to the average reader.
8. Page 9, Figure 1: It would be useful to give an expanded caption to better explain this calculation.
9. Page 10: It would be useful to define the various parameters in Equations 6 to 9 for non-experts.
10. Page 11, Equation 10: “ V_{S20} ” should be “ V_{S30} ” I assume.
11. **Pages 11-12, Figure 2: It is stated that 100 sets of coefficients are used but Figure 2 appears to indicate that predictions of one of the actual NGA West2 models is not captured within these 100 sets (one of the red model higher than the grey ones about M 5.5). Does this suggest the need to increase the number of models? Apart from increased computational time is there a reason for not generating many more sets?**
12. Figure 2: It would be useful to indicate which base GMPE is which by using different colours. Again, an expanded caption could be useful.
13. **Page 12, Equation 12: It would be useful to add a brief discussion of the assumption that the adjustment to the geometric spreading term only depends on the earthquake**

- location and not the site location. Can this assumption be relaxed?** On the other hand, the assumption that the site adjustment on depends on the site location is reasonable.
14. Figure 3: This figure should be made larger to improve legibility.
 15. Figure 4: Similarly this figure should be made larger.
 16. **Page 13, Equation 18: What is the source of this equation? It would be useful to add a couple of examples for different values of n_s and n_e to show the implications for this equation (and for Equations 16 and 17).**
 17. Figure 5: This figure should also be made larger. It is currently very difficult to read.
 18. **Page 15, “28 x 30km”: Why not “30 x 30km”? Or, in fact, another size entirely, e.g. 10 x 10km or 100 x 100km? Is the cell size connected with the potential variations in geology? Or to the strong-motion network/earthquake density?**
 19. Figure 6: Increase the size of this figure.
 20. Page 16, “Bayesian hierarchical model/multi-level model”: It would be useful to briefly explain what this means.
 21. Page 17, “half-cauchy”: It should be “half-Cauchy”.
 22. Page 17: The citations of Carpenter et al. (2016) and “Team (2015)” should be in brackets (and is “Team” the correct author name?).
 23. **Section 3.3: Is there a risk of double-counting the spatial variability in the cells using this approach and the approach in Section 3.2 simultaneously? In some sense they appear to be capturing similar things (e.g. geometric spreading and anelastic attenuation are often difficult to constrain independently).**
 24. Page 17, “100 realizations”: Is this sufficient to capture to full epistemic uncertainty in the results?
 25. **Figure 7, left panel: Is the sometimes large changes in anelastic attenuation between two cells (e.g. in the middle of the State there are neighbouring point mid-blue and mid-red) justified given knowledge of the crustal properties? For example, does this map match with maps for $L_g Q$ values from other studies?**
 26. Figure 8: A slightly larger figure would be useful here.
 27. Page 19, paragraph “Before running ... cf. Equation (15))”: This is a dense paragraph. It could be useful to expand the explanation slightly.
 28. Figure 9: The meaning of the red, blue, solid, dashed and dotted lines needs to be given in the caption or in a legend. An expanded caption would help too. These graphs are not particularly easy to read because of the large number of lines and their small sizes. As these are key figures it would be useful to spend some time making them more understandable.
 29. Page 23, Bussas et al. (2017): Add the volume and page numbers for this article.
 30. Page 23, Carpenter et al. (2016): Add the page numbers or DOI for this article.
 31. Page 24, Hiemer al. (2011): There seems to be something wrong in the title of this article (“Determination of θ and rock site ...”) – I think that “kappa” is missing.
 32. Page 24, Power et al. (2008): Add the page numbers for this article.

John Douglas

29th May 2018