Hazard Compatible Stochastic Ground Motion Modelling

Alexandra Tsioulou[a], Alexandros A. Taflanidis[b] and Carmine Galasso[a]

[a]Department of Civil, Environmental and Geomatic Engineering, University College London, UK
[b]Department of Civil and Environmental Engineering and Earth Sciences, University of Notre Dame, IN, USA

Abstract: This paper discusses a computationally efficient framework for the hazard-compatible tuning of existing stochastic ground motion models. The tuning pertains to the modification of the probabilistic predictive relationships that relate the ground motion model parameters to seismicity characteristics, whereas the seismic hazard is quantified through ground motion prediction equations (GMPEs), which for a specified earthquake scenario and period range provide information for both the conditional mean and the dispersion (variability) of the resultant spectral accelerations. The proposed modification is defined as an optimization problem with a dual objective. The first objective corresponds to comparison for a chosen earthquake scenario between the regional conditional hazard and the predictions established through the stochastic ground motion model. The second objective corresponds to comparison of the new predictive relationships with the pre-existing predictive relationships, developed considering regional data. This second objective guarantees that the resultant ground motions not only match the regional hazard (objective one) but are also compatible with observed trends. The relative entropy is adopted to quantify both objectives since they are both related to comparison between probability distributions, and a computational framework relying on Kriging surrogate modeling is established for an efficient optimization.

1 Introduction

The growing popularity the past decade in simulation-based probabilistic seismic risk assessment [9, 10] has increased the importance of ground motion modeling. Though undoubtedly scaling of ground motions [12] is the most popular methodology to do so, an approach that has been gaining increasing interest within the structural engineering community [5, 6] is the use of stochastic ground motion models [2, 3, 8, 13, 18]. These models are based on modulation of a stochastic sequence, through functions (filters) that address spectral and temporal characteristics of the excitation. The parameters of these filters are related to seismicity and site characteristics (i.e., seismicity scenarios) through predictive relationships. Sample ground motions for any desired seismicity scenario can be generated by determining the parameters of the stochastic ground motion model through these predictive relationships and by utilizing a sample stochastic sequence. This approach may ultimately support a comprehensive description of the seismic hazard [9]. Various methodologies have been established for developing the aforementioned predictive relationships, with main representatives being record-based and physics-based models. Record-based models (also known as site-based) are developed by fitting a preselected "waveform" to a suite of recorded regional ground motions [13, 18]. On the other hand, physics-based models rely on physical modeling of the rupture and wave propagation mechanisms [2, 3]. Emphasis in this study will be on the former models.

An important concern related to the use of stochastic ground motion models for structural engineering applications is the fact that through current approaches in selecting their predictive relationships compatibility to the seismic hazard for specific structures and sites is not necessarily obtained. This realization has motivated researchers to investigate the selection of predictive relationships for stochastic ground motion models so that compatibility with Ground Motion Prediction Equations (GMPEs), the established approach for describing seismic hazard, is explicitly achieved, an idea first introduced in [15]. Formulation considers an explicit optimization for matching the median predictions of the ground motion model to the spectral acceleration estimates of GMPEs for a range of seismicity scenarios, while maintaining physics-based principles or the matching to trends from real ground motions as an optimization constraint, in an attempt to preserve desired ground motion characteristics. Vetter et al. [17] recently extended the work of [15] by providing a versatile and computationally efficient approach, relying on surrogate modeling principles, for tuning stochastic ground motion models to establish compatibility with the median GMPE predictions for a range of structural periods, seismicity scenarios and site conditions of interest. Two significant drawbacks of this tuning approach, though, is that (i) the physical characteristics of the resulting acceleration time-series are incorporated in the optimization merely as constraints, something that requires significant experience in ground motion characterization for proper definition of the optimization problem and (ii) match only to the median hazard is targeted without considering the associated variability of the GMPE predictions. This study addresses these shortcomings by offering a computationally efficient framework to modify stochastic ground motion models for specific seismicity scenarios with a dual goal of (i) matching the prescribed hazard for a specific site and structure while (ii) preserving desired trends and correlations in the physical characteristics of the resultant ground acceleration time-series. The modification of the ground motion model is formulated as a multi-objective optimization problem and a surrogate modeling approach is adopted, similar to [17], to facilitate an efficient optimization.

2 Hazard-Compatible Ground Motion Modelling Framework

2.1 Preliminaries and Baseline Predictive Equation Formulation

Consider a stochastic ground motion model that provides acceleration time-histories $\ddot{a}(t\,|\,\boldsymbol{\theta},\mathbf{w})$ by modulating a Gaussian white-noise sequence, $\mathbf{w}$, through appropriate time/frequency functions that are parameterized through the n_θ-dimensional model parameter vector $\boldsymbol{\theta} \in \mathbb{R}^{n_\theta}$. This vector defines the model and is typically composed of various excitation properties such as Arias intensity, strong ground motion duration or parameters related to frequency characteristics of the ground motion. Section 3 presents the specific model chosen for this study. Synthetic time-histories can be created by relating $\boldsymbol{\theta}$ to seismicity and local site properties through predictive relationships. The vector of these properties, referred to as seismological parameters, is denoted as $\mathbf{z}$. Common characteristics used for $\mathbf{z}$ [3, 13] include the fault type, the moment magnitude and rupture distance of seismic events and the local soil profile. For record-based models the standard approach for development of these predictive relationships [13] relies on first matching the waveform characteristics to recorded ground motions (i.e., identify first $\boldsymbol{\theta}$ for each of the recorded ground motions in a given database) and then performing a regression to relate $\boldsymbol{\theta}$ to $\mathbf{z}$. Typically this is performed by first transforming the problem to the standard Gaussian space through a nonlinear mapping for each component θ_i. The transformed Gaussian vector is denoted $\mathbf{v}(\boldsymbol{\theta})$ herein. This ultimately leads to a Gaussian probability model $\mathbf{v}\sim N(\boldsymbol{\mu}_p(\mathbf{z}),\boldsymbol{\Sigma}_p)$ with mean $\boldsymbol{\mu}_p(\mathbf{z})$ that is dependent on $\mathbf{z}$ and covariance matrix $\boldsymbol{\Sigma}_p$, identified by the residuals of the regression, that is independent of $\mathbf{z}$. The resultant probability model for $\boldsymbol{\theta}$ is denoted $p(\boldsymbol{\theta}|\ \boldsymbol{\mu}_p(\mathbf{z}),\boldsymbol{\Sigma}_p)$.

2.2 Modification of Predictive Equations to Match a Target Hazard

This formulation for the predictive relationships of stochastic ground motion models, prioritizing a match to regional trends, provides synthetic ground motions whose statistics (mean and dispersion) of output IMs do not necessarily match hazard-compatible IMs (in terms of their mean and dispersion) as derived from GMPEs. To achieve this, a modification of the existing probabilistic regression model for $\boldsymbol{\theta}$ is proposed for specific seismicity scenarios defined by $\mathbf{z}$ with objective to get a suite of acceleration time-series (i) whose mean and dispersion match a target IM mean and dispersion vectors, while (ii) maintaining similarity to the predictive relationships already established for the model. The IM vector may include different response quantities of interest, for example characteristics of the ground motion, such as Peak Ground Acceleration (PGA), Velocity (PGV) and Displacement (PGD) or elastic and inelastic spectral responses for different periods of a Single Degree of Freedom (SDOF) oscillator. The target for these IMs can be described through a GMPE [4,16]. Note that if match to spectral responses is of interest, then a range of structural periods for which the match is established needs to be determined as well.

To formalize these concepts, let Y_i, i=1,…,n_y denote the response quantities of interest. The target hazard for there response quantities is provided by GMPEs, which yield a probabilistic description $\ln(Y_i) \sim N(\log \overline{Y}_i(\mathbf{z}), \sigma_i^2(\mathbf{z}))$ with $\log \overline{Y}_i(\mathbf{z})$ and $\sigma_i^2(\mathbf{z})$ corresponding to the mean and variance, respectively, of the logarithmic IM. Let $Y_i^g(\boldsymbol{\theta}, \mathbf{w})$ denote the estimate for Y_i established through the stochastic ground motion model for specific values of the model parameter vector $\boldsymbol{\theta}$ and a specific white noise sequence $\mathbf{w}$ [i.e. for a specific ground motion time-history $\ddot{a}(t \mid \boldsymbol{\theta}, \mathbf{w})$]. Considering the variability in both $\boldsymbol{\theta}$ and $\mathbf{w}$, the mean and variance for $\log(Y_i^g)$ are

$$\log(\overline{Y}_i^g(\boldsymbol{\mu}, \boldsymbol{\Sigma})) = E\left[\log(Y_i^g)\right] = \int\int \log(Y_i^g(\boldsymbol{\theta}, \mathbf{w})) p(\boldsymbol{\theta} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma})\, p(\mathbf{w})\, d\boldsymbol{\theta}\, d\mathbf{w} \tag{1}$$

$$(\sigma_i^g(\boldsymbol{\mu}, \boldsymbol{\Sigma}))^2 = Var\left[\log(Y_i^g)\right] = \int\int \left[\log(Y_i^g(\boldsymbol{\theta}, \mathbf{w})) - \log(\overline{Y}_i^g(\boldsymbol{\mu}, \boldsymbol{\Sigma}))\right]^2 p(\boldsymbol{\theta} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma})\, p(\mathbf{w})\, d\boldsymbol{\theta}\, d\mathbf{w} \tag{2}$$

where $E[.]$, $Var[.]$ denote the expectation and variance operators, respectively, $p(\mathbf{w})$ is the probability distribution for the stochastic sequence $\mathbf{w}$ and $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ represent the mean and covariance respectively for $\mathbf{v}$. Note that both these quantities may be different from the ones given by the predictive relationships $\boldsymbol{\mu}_p(\mathbf{z})$,$\boldsymbol{\Sigma}_p$. Similar to the GMPE predictions we will assume a lognormal distribution for Y_i^g, $\ln(Y_i^g(\boldsymbol{\mu}(\mathbf{z}), \boldsymbol{\Sigma})) \sim N(\log(\overline{Y}_i^g(\boldsymbol{\mu}(\mathbf{z}), \boldsymbol{\Sigma})), (\sigma_i^g(\boldsymbol{\mu}(\mathbf{z}), \boldsymbol{\Sigma}))^2)$. This simplifies the comparison between target and underlying (by ground motion model) hazard since they both correspond to normal distributions for the logarithmic IM.

The hazard compatible modelling corresponds to modification of the probability model for $\boldsymbol{\theta}$, ultimately of $\boldsymbol{\mu}(\mathbf{z})$ and $\boldsymbol{\Sigma}$ and is formulated as multi-objective optimization problem with two competing objectives

$$\left[\boldsymbol{\mu}, \boldsymbol{\Sigma}\right]^* = \arg\min\{F_{p1}(\boldsymbol{\mu}, \boldsymbol{\Sigma} \mid \mathbf{z}), F_{p2}(\boldsymbol{\mu}, \boldsymbol{\Sigma} \mid \mathbf{z})\} \tag{3}$$

The first objective corresponds to the discrepancy of the target seismic hazard to the hazard established through the ground motion model, i.e. to a comparison between the GMPE-based probabilistic hazard description $\ln(Y_i) \sim N(\log \overline{Y}_i(\mathbf{z}), \sigma_i^2(\mathbf{z}))$ and the probability model for the response output $\ln(Y_i^g(\boldsymbol{\mu}, \boldsymbol{\Sigma})) \sim N(\log(\overline{Y}_i^g(\boldsymbol{\mu}, \boldsymbol{\Sigma})), (\sigma_i^g(\boldsymbol{\mu}, \boldsymbol{\Sigma}))^2)$. The relative entropy $D[a||b]$ is utilized as measure for the difference between probability distributions a and b, a popular measure to quantify discrepancies between distributions, leading to

$$F_{p1}(\boldsymbol{\mu}, \boldsymbol{\Sigma} \mid \mathbf{z}) = \sum_{i=1}^{n_y} D\left[N(\log(\overline{Y}_i^g(\boldsymbol{\mu}, \boldsymbol{\Sigma})), (\sigma_i^g(\boldsymbol{\mu}, \boldsymbol{\Sigma}))^2) \,||\, N(\log \overline{Y}_i(\mathbf{z}), \sigma_i^2(\mathbf{z}))\right] \tag{4}$$

where the summation of the entropy terms for each IM component for definition of the entropy of the IM vector assumes that target hazard for each of them is independently determined. Since the compared distributions are Gaussians the relative entropy can be readily evaluated as

$$D\Big[N(\log(\overline{Y}_i^g(\boldsymbol{\mu},\boldsymbol{\Sigma})),(\sigma_i^g(\boldsymbol{\mu},\boldsymbol{\Sigma}))^2)\,\|\,N(\log\overline{Y}_i(\mathbf{z}),\sigma_i^2(\mathbf{z}))\Big]= \\ \frac{\left(\log(\overline{Y}_i^g(\boldsymbol{\mu},\boldsymbol{\Sigma})-\log\overline{Y}_i(\mathbf{z})\right)^2}{2\sigma_i^2(\mathbf{z})}+\frac{1}{2}\left[\frac{(\sigma_i^g(\boldsymbol{\mu},\boldsymbol{\Sigma}))^2}{\sigma_i^2(\mathbf{z})}-1-\ln\left(\frac{(\sigma_i^g(\boldsymbol{\mu},\boldsymbol{\Sigma}))^2}{\sigma_i^2(\mathbf{z})}\right)\right] \tag{5}$$

Objective F_{p2} measures that discrepancy between the initial probability model for the predictive relationship $p(\boldsymbol{\theta}|\ \boldsymbol{\mu}_p(\mathbf{z}),\boldsymbol{\Sigma}_p)$ and the modified one $p(\boldsymbol{\theta}\,|\,\boldsymbol{\mu},\boldsymbol{\Sigma})$. The relative entropy is utilized again as measure to quantify differences and since this entropy is invariant under a coordinate transformation, the comparison can be established in the standard Gaussian space, leading to

$$F_{p2}(\boldsymbol{\mu},\boldsymbol{\Sigma}\,|\,\mathbf{z})=D\Big[N(\boldsymbol{\mu},\boldsymbol{\Sigma})\,\|\,N(\boldsymbol{\mu}_p(\mathbf{z}),\boldsymbol{\Sigma}_p)\Big] \tag{6}$$

with the entropy readily evaluated as

$$D\Big[N(\boldsymbol{\mu},\boldsymbol{\Sigma})\,\|\,N(\boldsymbol{\mu}_p(\mathbf{z}),\boldsymbol{\Sigma}_p)\Big]= \\ =1/2\left[tr\Big[\boldsymbol{\Sigma}\boldsymbol{\Sigma}_p^{-1}\Big]+\left(\boldsymbol{\mu}_p(\mathbf{z})-\boldsymbol{\mu}\right)^T\boldsymbol{\Sigma}_p^{-1}\left(\boldsymbol{\mu}_p(\mathbf{z})-\boldsymbol{\mu}\right)-n_\theta-\ln\left(\det\Big[\boldsymbol{\Sigma}\boldsymbol{\Sigma}_p^{-1}\Big]\right)\right] \tag{7}$$

where *tr*[.] and det[.] stand for trace and determinant, respectively.

Solution of the multi-objective optimization of Eq. (3) ultimately leads to a Pareto-front of dominant solutions that express a different compromise between the competing objectives F_{p1} and F_{p2}. A solution is characterized as dominant (and belongs in the Pareto-set) if there is no other solution that simultaneously improves both objectives F_{p1} and F_{p2}. One can eventually select a model configuration from this Pareto-set that yields the desired IM-compatibility without deviating significantly from regional ground motion characteristics. Identifying the Pareto front for this problem is, though, challenging since the computational challenge in evaluation of objective F_{p1} is significant, requiring evaluation of the multidimensional integrals of Eqs. (1) and (2). To facilitate an efficient optimization that can be repeated for any desired seismicity scenario **z** a surrogate modeling approach is adopted here, similar to the one utilized in [17]. An overview of the surrogate modeling is presented in Section 4 with the resultant surrogate model –aided optimization discussed in detail in Section 5. Before doing so, the stochastic ground motion model utilized in the illustrative example is discussed.

3 Stochastic Ground Motion Model Utilized in the Study

The specific ground motion model used in this study is the record-based model proposed by Rezaeian and Der Kiureghian [13] that efficiently addresses both temporal and spectral non-stationarities. The former is established through a time-domain modulating envelope function, while the latter is achieved by using a frequency-domain modulation with characteristics that vary in time. The baseline discretized time-history of the ground motion according to this model is expressed as

$$\ddot{a}(t\,|\,\boldsymbol{\theta},\mathbf{w})=q(t,\boldsymbol{\theta})\sum_{i=1}^{k}\frac{h[t-t_i,\boldsymbol{\theta}(t_i))]}{\sqrt{\sum_{j=1}^{k}h[t-t_j,\boldsymbol{\theta}(t_j))]^2}}w(i\Delta t)\quad k\Delta t<t<(k+1)\Delta t \tag{8}$$

where $\mathbf{w}=[w(i\Delta t)\colon i=1,2,\ldots,N_T]$ is a Gaussian white noise sequence, Δt is the chosen discretization interval (assumed constant and equal to 0.005s in this study), $q(t,\boldsymbol{\theta})$ is the time-

modulating function, and $h[t\text{-}\tau,\boldsymbol{\theta}(\tau)]$ an impulse response function corresponding to the pseudo-acceleration response of a SDOF linear oscillator with time varying frequency $\omega_f(\tau)$ and damping ratio $\zeta_f(\tau)$, in which τ denotes the time of the pulse

$$\begin{aligned} h[t-\tau,\boldsymbol{\theta}(\tau))] &= \frac{\omega_f(\tau)}{\sqrt{1-\zeta_f^2(\tau)}}\exp\left[-\omega_f(\tau)\zeta_f(\tau)(t-\tau)\right]\sin\left[\omega_f(\tau)\sqrt{1-\zeta_f^2(\tau)}(t-\tau)\right]; \quad \tau \le t \\ &= 0; \quad \text{otherwise} \end{aligned} \tag{9}$$

For the time varying characteristics a linear function has been proposed for the frequency and a constant for the damping

$$\omega_f(\tau) = \omega_{mid} + \omega'(\tau - t_{mid}) \qquad \zeta_f(\tau) = \zeta_f \tag{10}$$

with ω_{mid} (central frequency), ω' (frequency variation) and ζ_f ultimately corresponding to model parameters for the filter and t_{mid} corresponding to the mid-time of the strong motion duration (defined next). The time envelope $q(t,\boldsymbol{\theta})$ is given by [13]

$$q(t, I_a, \alpha_2, \alpha_3) = \sqrt{I_a}\left[\sqrt{\frac{2}{\pi}\frac{(2\alpha_3)^{2\alpha_2-1}}{\Gamma(2\alpha_2-1)}}\right] t^{\alpha_2-1}\exp(-\alpha_3 t) \tag{11}$$

where $\Gamma(.)$ is the gamma function, I_a is the Arias intensity expressed in terms of g, and $\{\alpha_2, \alpha_3\}$ are additional parameters controlling the shape and total duration of the envelope that can be related to the strong motion duration, $D_{5\text{-}95}$ (defined as the duration for the Arias intensity to increase from 5% to 95% of its final value), and t_{mid}, the time Arias intensity corresponds to 45% of its final value. The pair $\{\alpha_2, \alpha_3\}$ can be easily determined based on the values of $\{D_{5\text{-}95}, t_{mid}\}$ [13]. To assure zero residual velocity, the simulated process is eventually high-pass filtered, yielding the final excitaton $\ddot{a}(t\,|\,\boldsymbol{\theta},\mathbf{w})$. The filter corresponds to a critically damped oscillator, and has minimal impact on the response beyond chosen corner frequency, ω_c [13].

Ultimately, the ground motion model has as parameters $\boldsymbol{\theta}=\{I_a, D_{5\text{-}95}, t_{mid}, \omega_{mid}, \omega', \zeta_f\}$ with the first one directly impacting (scaling) the output and the remaining five, denoted by $\mathbf{x}$ herein, having a complex nonlinear relationship to that output, so that $\boldsymbol{\theta}=\{I_\alpha, \mathbf{x}\}$. The normalized responses will be denoted s_i so that $Y_i^g(\boldsymbol{\theta},\mathbf{w}) = \sqrt{I_a}\,s_i(\mathbf{x},\mathbf{w})$.

Predictive relationships have been established for $\boldsymbol{\theta}$ by fitting the stochastic model to a subset of the next generation attenuation (NGA) relationships strong motion database [13]. These predictive relationships relate $\boldsymbol{\theta}$ to the following earthquake and site characteristics, defining seismicity vector $\mathbf{z}$: the moment magnitude, M, the rupture distance, r_{rup}, the type of fault F [F=0 denoting strike slip and F=1 reverse fault] and the shear wave velocity of the top 30m of the site soil, V_{s30}. Details for these relationships, ultimately defining probability model denoted $p(\boldsymbol{\theta}|\boldsymbol{\mu}_p(\mathbf{z}),\boldsymbol{\Sigma}_p)$ may be found in [13].

4 Metamodel Details

The metamodel is developed to provide an efficient approximation to the stochastic ground motion model output for specific model characteristics $\boldsymbol{\theta}$, i.e. to address the variability of the response with respect to $\mathbf{w}$. Ultimately this is established for the normalized response s_i and pertaint to both the logarithimc mean and variance with respect to $\mathbf{w}$, since both these statistics are needed for the estimation of the objective function of Eq. (4). These statistics are

$$\log(\overline{s}_i(\mathbf{x})) = E\left[\log(s_i(\mathbf{x}))\right] = \int \log(s_i(\mathbf{x},\mathbf{w}))p(\mathbf{w})d\mathbf{w} \tag{12}$$

$$(\sigma_i^s(\mathbf{x}))^2 = Var\left[\log(s_i(\mathbf{x}))\right] = \int \left[\log(s_i(\mathbf{x},\mathbf{w})) - \log(\overline{s}_i(\mathbf{x}))\right]^2 p(\mathbf{w})d\mathbf{w} \tag{13}$$

and note that they are both independent of the characteristics of the predictive relationships; rather are simply functions of **x** directly. This is what makes the surrogate modelling approach efficient: the surrogate model needs to be simply established within the domain of interest for the model parameters of the ground motion model (i.e. the domain covered by the initial predictive relationships), and can be then leveraged, as detailed in the next Section to evaluate the required statistics for different selections of the predictive relationships **μ**(**z**), **Σ**. In other words the input to the metamodel is the low-dimensional vector **x**, and the formulation is independent of **z** or the probability distribution characteristics **μ**(**z**) and **Σ**. The metamodel provides predictions (metamodel output) for both $\log(\overline{s}_i(\mathbf{x}))$ and $\sigma_i^s(\mathbf{x})$.

For developing the metamodel a database with n observations is initially obtained that provides information for the the $\mathbf{x}-\log(\overline{s}_i(\mathbf{x}))$ and $\mathbf{x}-\sigma_i^s(\mathbf{x})$ pairs. For this purpose n samples for $\{\mathbf{x}^j, j=1,...,n\}$, also known as *support points or experiments*, are obtained over the domain of interest for **x**. This domain, denoted *X*, should encompass the anticipated range that the metamodel will be implemented in [i.e domain covered by $p(\boldsymbol{\theta}|\,\boldsymbol{\mu}(\mathbf{z}),\boldsymbol{\Sigma})$]. The mean predictions provided through the ground motion model are then established through the folowing process considering n_w white-noise samples

Step 1. Generate n_w sample acceleration time-histories for different white-noise sequences $\ddot{a}_\kappa(t\,|\,\boldsymbol{\theta},\mathbf{w}^k);\kappa=1,...,n_w$. Set model parameter I_a equal to 1.
Step 2. For each sample evaluate the responses of interest. For spectral quantities this will entail numerical simulation of SDOF responses.
Step 3. Estimate the statistics over the established sample-set to obtain $\log(\overline{s}_i(\mathbf{x}))$ and $\sigma_i^s(\mathbf{x})$.

Using this database the surrogate model can be formulated. Here Kriging is selected as surrogate model due to its ability to accurately describe complex functions [14]. Details for the Kriging metamodel formulation may be found in [14] or [17], with the latter reference focusing on a similar application as the one considered here, i.e. explicitly looking at metamodel development for approximating the predictions of stochastic ground motion models. The metamodel ultimately provides approximations for $\log(\overline{s}_i(\mathbf{x}))$ and $\sigma_i^s(\mathbf{x})$.

The computationally intensive aspect of the entire formulation is the development of the database which requires response-history analysis for a large number of model parameters to populate *X*, and a sufficient number of white-noise samples to address the resultant variability in the response. This needs to be performed, though, only once. As soon as the Kriging metamodel is established based on this database, it can be then used to efficiently predict the responses for any other **x** desired. Calculation of $\log(\overline{s}_i(\mathbf{x}))$ and $\sigma_i^s(\mathbf{x})$ can be also vectorized, something that will be leveraged in the numerical optimization discussed in the next section.

5 Kriging-aided Multi-Objective Optimization

5.1 Calculation of Statistics of Interest

Calculation of the statistics of interest, given by Eqs. (1) and (2), simplifies to

$$\log(\overline{Y}_i^g(\boldsymbol{\mu},\boldsymbol{\Sigma})) = E\left[\log(Y_i^g)\right] = E\left[\log(\sqrt{I_a}s_i)\right] = 1/2E\left[\log(I_a)\right] + E\left[\log(s_i)\right] \quad (14)$$

$$(\sigma_i^g(\boldsymbol{\mu},\boldsymbol{\Sigma}))^2 = Var\left[\log(\sqrt{I_a}s_i)\right] = 1/4Var\left[\log(I_a)\right] + Var\left[\log(s_i)\right] + Cov\left[\log(I_a),\log(s_i)\right] \quad (15)$$

where Cov[a,*b*] stands for the covariance between random variables *a* and *b*. The statistics for I_a, that is the mean $E[\log(I_a)]$ and variance $Var[\log(I_a)]$ can be easily calculated using the marginal distribution $p(I_a|\boldsymbol{\mu},\boldsymbol{\Sigma})$. The statistics that involve s_i require use of the metamodel to

approximate the influence of **w** and Monte Carlo Simulation (MCS) to approximate the influence of the predictive relationship variability. Rather than performing two different MCS: one for the expectation $E[\log(s_i)]$ and variance $Var[\log(s_i)]$ [which require samples from the marginal distribution $p(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\Sigma})$], and one for the covariance $Cov[\log(I_a),\log(s_i)]$ [which requires samples from joint distribution $p(\boldsymbol{\theta}|\boldsymbol{\mu},\boldsymbol{\Sigma})$], a single MCS simulation is performed, utilizing a common set of samples for all these statistics. This leads to the following approximations

$$E\left[\log(s_i)\right] = \int \log(\overline{s}_i(\mathbf{x}))\, p(\mathbf{x}\,|\,\boldsymbol{\mu},\boldsymbol{\Sigma}) d\mathbf{x} \approx 1/N_s \sum_{j=1}^{N_s} \log(\overline{s}_i(\mathbf{x}^j)) \tag{16}$$

$$\begin{aligned} Var\left[\log(s_i)\right] &= \int \left(\sigma_i^s(\mathbf{x})\right)^2 p(\mathbf{x}\,|\,\boldsymbol{\mu},\boldsymbol{\Sigma}) d\mathbf{x} + \int \left(\log(\overline{s}_i(\mathbf{x})) - E\left[\log(s_i)\right]\right)^2 p(\mathbf{x}\,|\,\boldsymbol{\mu},\boldsymbol{\Sigma}) d\mathbf{x} \\ &\approx 1/N_s \sum_{j=1}^{N_s} \left(\sigma_i^s(\mathbf{x}^j)\right)^2 + 1/N_s \sum_{j=1}^{N_s} \left(\log(\overline{s}_i(\mathbf{x}^j)) - E\left[\log(s_i)\right]\right)^2 \end{aligned} \tag{17}$$

$$\begin{aligned} Cov\left[\log(I_a),\log(s_i)\right] &= \int \log(I_a)\log(\overline{s}_i(\mathbf{x}))\, p(\boldsymbol{\theta}\,|\,\boldsymbol{\mu},\boldsymbol{\Sigma}) d\boldsymbol{\theta} - E\left[\log(I_a)\right] E\left[\log(s_i)\right] \\ &\approx 1/N_s \sum_{j=1}^{N_s} \log(I_a^j)\log(\overline{s}_i(\mathbf{x}^j) - E\left[\log(I_a)\right] E\left[\log(s_i)\right] \end{aligned} \tag{18}$$

where $[I_a^j, \mathbf{x}^j]$ correspond to samples from $p(\boldsymbol{\theta}|\boldsymbol{\mu},\boldsymbol{\Sigma})$, N_s is the total number of samples used and $\log(\overline{s}_i(\mathbf{x}))$ and $\sigma_i^s(\mathbf{x})$ are approximated through the Kriging metamodel for each one of these samples. Utilizing vectorized manipulations for the metamodel both these quantities can be calculated with very small computational effort, meaning that the MCS based estimation of Eqs. (16), (17) and (18) can be efficiently performed.

5.2 Multi-objective Optimization

Calculation of statistics given by Eq. (14) and (15) [utilizing estimates of Eqs. (16), (17) and (18)] facilitates an efficient approximation for objective F_{p1}. The multi-objective optimization problem of Eq. (3) is then solved adopting a gradient-free exhaustive search [7]. A very large number of n_{bc} samples for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are generated that are close to $\boldsymbol{\mu}_p(\mathbf{z})$ and $\boldsymbol{\Sigma}_p$, and objective functions F_{p1} and F_{p2} are calculated. Estimation of objective F_{p1} in this case leverages the computational efficiency of the metamodel in performing vectorized predictions by simultaneously performing calculations for all n_{bc} samples, or using subsets with a lower number of samples depending on the available computational resources (memory can be a problem for vectorized operations depending on the number of support points n). Note that for each of the samples of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ the metamodel needs to be utilized to provide prediction for the N_s samples utilized within the MCS. The dominant solutions representing the Pareto front can be then readily identified by comparing the values for the two objectives. The challenge in this case is that the value of n_{bc} needs to be large in order to obtain an adequate representation of the Pareto front. The advantage is that vectorized calculations can be utilized for the metamodel predictions.

6 Illustrative Implementation

The box-bounded domain X for development of the metamodel is determined based on the ranges reported in [13] as [5 45] s for D_{5-95}, [0.5 40]s for t_{mid}, [-2, 0.5] Hz/s for $\omega^{'}/2\pi$, [0.1 30] Hz for $\omega_{mid}/2\pi$ and [0.02 0.99] for ζ_f. For the response output the peak pseudo-acceleration (Y=PSA) for a SDOF system with 5% damping ratio and for 22 different periods T_s=[0.01 0.02 0.03 0.04 0.05 0.075 0.1 0.15 0.2 0.25 0.3 0.4 0.5 0.75 1.0 1.5 2.0 3.0 4.0 5.0 7.5 10.0] s is adopted. The white-noise samples are chosen as $n_w = 100$. A total of 4500 support points is utilized for the metamodel, leading to high coefficient of determination, over 98% for mean predictions and 95% for variance predictions. For generating the total of 4500000 time-histories and performing the required simulations to develop the output database, close to 600

CPU hours were required. Though this computational burden is significant, it should be stressed that it corresponds to an initial only overhead of the approach. Once the metamodel is developed, it can be then used for any required predictions since the established accuracy is high and can support the multi-objective optimization as discussed in Section 5.2.

Here for illustration purposes three different events are examined, corresponding to seicmicity scenarios {M=6, r_{rup}=20 km}, {M=7, r_{rup}=50 km} and {M=8, r_{rup}=20 km}. For both cases the match is provided for local site condition corresponding to shear wave velocity in the upper 30 m V_{s30}=600 m/s and a strike-slip fault. The target PSA values were estimated as the average of four different GMPEs developed for crustal earthquakes in the Western U.S. [1] for a range of periods T chosen as [0.3 0.4 0.5 0.75 1] s. The specific GMPES correspond to the 2008 models [1] developed by Abrahamson and Silva, Boore and Atkinson, Campbell and Bozorgnia, and Chiou and Youngs. Since many of the chosen GMPEs require additional seismicity inputs, beyond M and r_{rup}, the nominal relationships suggested in [11] are used to derive these inputs, starting from the M and r_{rup} values.

The optimization is performed utilizing a total number of n_{bc}=50,000 samples, yielding a Pareto front of dominant solutions for each event, reported in Figure 1. To validate the accuracy of the Kriging approximation, the estimates of the objective function F_{p1} established through the metamodel as well as through use of the actual ground motion model, i.e. utilizing directly Eq. (1) and (2) [statistics are calculated again through Monte Carlo simulation, in this case simultaneously extending over both $\boldsymbol{\theta}$ and $\mathbf{w}$], are reported in this figure for each Pareto front point. It should be noted that the resultant Pareto fronts include a larger number of points, but only 10 points are reported in Figure 1 for clearer presentation. The results for the match between the target (i.e, GMPE-based) PSA probability model (mean and dispersion) estimated through GMPEs and the resultant probability model for the ground motion outputs for all cases are presented in Figure 2. In particular, Figure 2 shows spectral plots (for all T considered) for the mean and mean ± one standard deviation estimated by the GMPEs for each seismicity scenario and the corresponding spectral plots (estimated by the metamodel) for the original (F_{p2}=0) and the modified predictive relationships for the ground motion model parameters. The modified predictive relationship shown in Figure 2 corresponds to the point on the Pareto front with the smallest discrepancy from the probability model of the target response output (smallest F_{p1}).

As shown in Figure 1, the predictions of the objective function F_{p1} by the metamodel are in good agreement with the ones provided through the stochastic ground motion model, validating the accuracy of the metamodel approximation for the response output PSA. The Pareto front points shown in Figure 1 correspond to ground motion models that represent a different compromise between the two objective functions, and ultimately one can be chosen based on any desired criterion; for example the solution that remains below a desired threshold for F_{p1} (i.e provides sufficient hazard compatibility) or the solution that corresponds to the best comparative improvement of F_{p1} without a significant compromise of F_{p2} (this is facilitated by evaluating the slope of the Pareto front). Figure 2 verifies the hazard match in terms of both mean and dispersion. Differences exist between the GMPE target and the original model in terms of both the desired mean and desired standard deviation. Note that the reported trends (overprediction by the ground motion model for small M and underprediction for large M) agree with the ones reported in [13]. The proposed modification of the predictive relationships yields in all cases examined, i.e. irrespective of the over or under prediction trends, a model that matches both the target mean and the target dispersion (Figure 2). The compromise over the second objective F_{p2} is larger for magnitude 6.

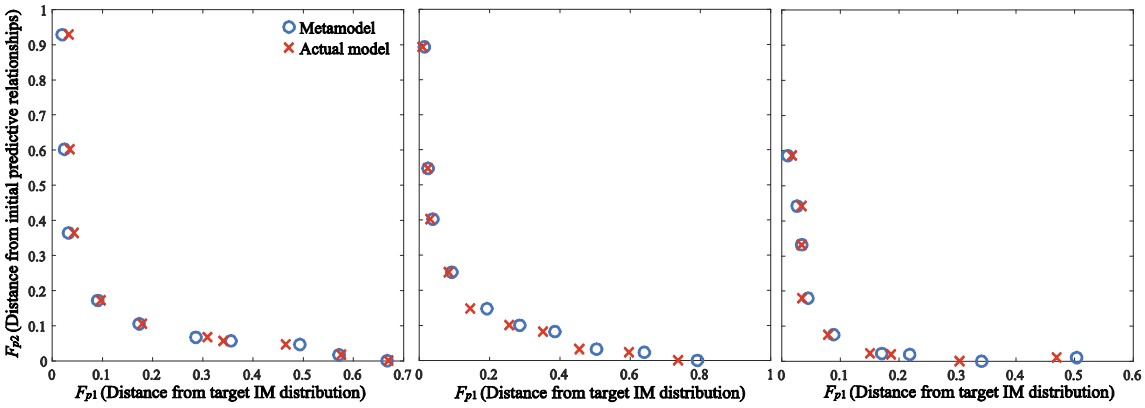

Figure 1: Pareto fronts for M=6, r_{rup}=20km (left), M=7, r_{rup}=50km (middle) and M=8, r_{rup}=20km (right). Predictions through the metamodel as well as the actual ground motion model are shown.

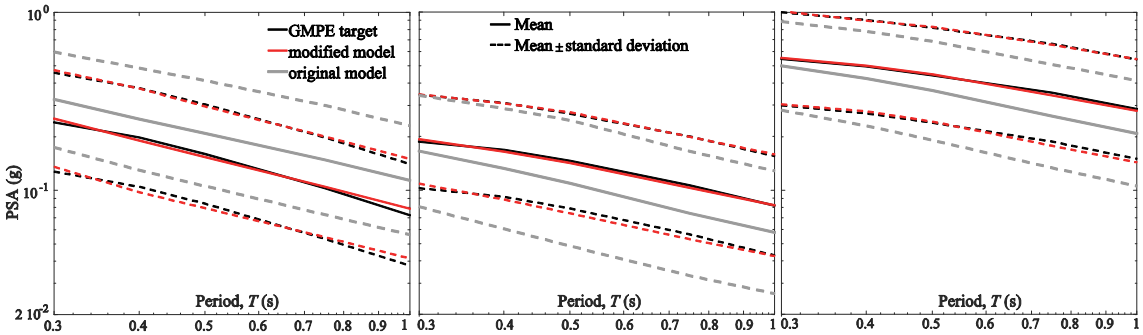

Figure 2: Comparison of PSA mean (solid line) and mean ± dispersion (dashed lines) predictions by the original stochastic ground motion model to the ones by the modified stochastic ground motion model as well as the targeted GMPE estimates for M=6, r_{rup}=20km (left), M=7, r_{rup}=50km (middle) and M=8, r_{rup}=20km (right).

7 Conclusions

The tuning of stochastic ground motion models for compatibility with the regional hazard was discussed in this paper. This hazard was described through the mean and dispersion predictions of GMPEs. The tuning was formulated as modification of the probabilistic predictive relationships for the ground motion model parameters (relating these parameters to seismicity and site characteristics) with a dual objective (multi-objective optimization) to match the aforementioned hazard while maintaining small discrepancy with pre-existing predictive relationships, assumed to facilitate similarity to observed regional trends. Computational efficiency was established through a metamodeling approach, which provided an accurate approximation for the response statistics of the ground motion model considering its stochastic variability. Once developed the surrogate model can facilitate an efficient optimization for any desired seismicity scenario providing a Pareto front of optimal solutions. Within the illustrative example the approach was shown to identify ground motion model that provide the desired match to the conditional target hazard.

References

[1] N. Abrahamson, G. Atkinson, D. Boore, Y. Bozorgnia, K. Campbell, B. Chiou, I. M. Idriss, W. Silva, R. Youngs. *Comparisons of the NGA ground-motion relations*. Earthquake Spectra, Vol. 24, pp. 45-66, 2008.

[2] G. M. Atkinson, W. Silva. *Stochastic modeling of California ground motions*. Bulletin of the Seismological Society of America, Vol. 90, pp. 255-274, 2000.
[3] D. M. Boore. *Simulation of ground motion using the stochastic method*. Pure and Applied Geophysics, Vol. 160, pp. 635-676, 2003.
[4] Y. Bozorgnia, M. M. Hachem, K. W. Campbell. *Ground motion prediction equation ("attenuation relationship") for inelastic response spectra*. Earthquake Spectra, Vol. 26, pp. 1-23, 2010.
[5] M. Broccardo, A. Der Kiureghian. *Multicomponent nonlinear stochastic dynamic analysis by tail-equivalent linearization*. Journal of Engineering Mechanics, Vol. 142, p. 04015100, 2015.
[6] J. Chun, J. Song, G. H. Paulino. *Structural topology optimization under constraints on instantaneous failure probability*. Structural and Multidisciplinary Optimization, Vol. 53, pp. 773-799, 2016.
[7] C. A. C. Coello, D. A. Van Veldhuizen, G. B. Lamont. *Evolutionary algorithms for solving multi-objective problems*, Vol. 242: Springer, 2002.
[8] H. P. Gavin, B. W. Dickinson. *Generation of uniform-hazard earthquake ground motions*. Journal of Structural Engineering, ASCE, Vol. 137, pp. 423-432, 2011.
[9] I. Gidaris, A. A. Taflanidis. *Performance assessment and optimization of fluid viscous dampers through life-cycle cost criteria and comparison to alternative design approaches*. Bulletin of Earthquake Engineering, Vol. 13, pp. 1003-1028, 2015.
[10] C. A. Goulet, C. B. Haselton, J. Mitrani-Reiser, J. L. Beck, G. Deierlein, K. A. Porter J. P. Stewart. *Evaluation of the seismic performance of code-conforming reinforced-concrete frame building-From seismic hazard to collapse safety and economic losses*. Earthquake Engineering and Structural Dynamics, Vol. 36, pp. 1973-1997, 2007.
[11] J. Kaklamanos, L. G. Baise, D. M. Boore. *Estimating unknown input parameters when implementing the NGA ground-motion prediction equations in engineering practice.* Earthquake Spectra, Vol. 27, pp. 1219-1235, 2011.
[12] T. Lin, C. B. Haselton, J. W. Baker. *Conditional spectrum-based ground motion selection. Part I: Hazard consistency for risk-based assessments*. Earthquake Engineering and Structural Dynamics, Vol. 42, pp. 1847-1865, 2013.
[13] S. Rezaeian, A. Der Kiureghian. *Simulation of synthetic ground motions for specified earthquake and site characteristics*. Earthquake Engineering & Structural Dynamics, Vol. 39, pp. 1155-1180, 2010.
[14] J. Sacks, W. J. Welch, T. J. Mitchell, H. P. Wynn. *Design and analysis of computer experiments.* Statistical Science, Vol. 4, pp. 409-435, 1989.
[15] F. Scherbaum, F. Cotton, H. Staedtke. *The estimation of minimum-misfit stochastic models from empirical ground-motion prediction equations*. Bulletin of the Seismological Society of America, Vol. 96, pp. 427-445, 2006.
[16] J. P. Stewart, D. M. Boore, E. Seyhan, G. M. Atkinson. *NGA-West2 equations for predicting vertical-component PGA, PGV, and 5%-damped PSA from shallow crustal earthquakes*. Earthquake Spectra, Vol. 32, pp. 1005-1031, 2015.
[17] C. Vetter, A. A. Taflanidis, G. P. Mavroeidis. *Tuning of stochastic ground motion models for compatibility with ground motion prediction equations*. Earthquake Engineering and Structural Dynamics, Vol. 45, pp. 893-912, 2016.
[18] C. Vlachos, K. G. Papakonstantinou, G. Deodatis. *A multi-modal analytical non-stationary spectral model for characterization and stochastic simulation of earthquake ground motions*. Soil Dynamics and Earthquake Engineering, Vol. 80, pp. 177-191, 2016.