


RESEARCH ARTICLE

Hazard-compatible modification of stochastic ground motion models

Alexandra Tsioulou¹ | Alexandros A. Taflanidis²  | Carmine Galasso¹ 

¹Department of Civil, Environmental and Geomatic Engineering, University College London, London, UK

²Department of Civil and Environmental Engineering and Earth Sciences, University of Notre Dame, Notre Dame, IN, USA

Correspondence

Alexandros A. Taflanidis, Associate Professor, Department of Civil and Environmental Engineering and Earth Sciences, University of Notre Dame, 156 Fitzpatrick Hall, Notre Dame, IN 46556, USA.

Email: a.taflanidis@nd.edu

Summary

A computationally efficient framework is presented for modification of stochastic ground motion models to establish compatibility with the seismic hazard for specific seismicity scenarios and a given structure/site. The modification pertains to the probabilistic predictive models that relate the parameters of the ground motion model to seismicity/site characteristics. These predictive models are defined through a mean prediction and an associated variance, and both these properties are modified in the proposed framework. For a given seismicity scenario, defined for example by the moment magnitude and source-to-site distance, the conditional hazard is described through the mean and the dispersion of some structure-specific intensity measure(s). Therefore, for both the predictive models and the seismic hazard, a probabilistic description is considered, extending previous work of the authors that had examined description only through mean value characteristics. The proposed modification is defined as a bi-objective optimization. The first objective corresponds to comparison for a chosen seismicity scenario between the target hazard and the predictions established through the stochastic ground motion model. The second objective corresponds to comparison of the modified predictive relationships to the pre-existing ones that were developed considering regional data, and guarantees that the resultant ground motions will have features compatible with observed trends. The relative entropy is adopted to quantify both objectives, and a computational framework relying on kriging surrogate modeling is established for an efficient optimization. Computational discussions focus on the estimation of the various statistics of the stochastic ground motion model output needed for the entropy calculation.

KEYWORDS

hazard compatibility, kriging surrogate modeling, Monte Carlo simulation, seismic risk, stochastic ground motion model

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2018 The Authors. Earthquake Engineering & Structural Dynamics Published by John Wiley & Sons Ltd

1 | INTRODUCTION

The relevance of techniques that model acceleration time-series of seismic events has increased during the past decades due to the growing popularity of simulation-based probabilistic seismic risk assessment¹⁻³ and performance-based earthquake engineering.⁴⁻⁶ Though the most popular methodology for performing this task is the selection of real (ie, recorded from past events) ground motions,⁷⁻¹⁰ potentially scaled based on a target intensity measure (IM), an alternative philosophy is the use of simulated ground motions.^{11,12} A specific modeling approach for the latter which has been steadily gaining increasing attention by the structural engineering community,¹³⁻¹⁵ is the use of stochastic ground motion models.¹⁶⁻²² These models are based on a parametric description of the spectral and temporal characteristics of the excitation, with synthetic time-histories obtained by filtering a stochastic sequence through the resultant frequency and time domain modulating functions. The parameters involved in this description are related to seismicity (for example moment magnitude and rupture distance) and site characteristics (for example shear wave velocity for soil profile) through predictive models/relationships.^{16,20} Sample ground motions for a specific seismicity scenario and site 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,^{6,23} and its essential component is the predictive models relating seismicity/site characteristics to ground motion model parameters.

The two main methodologies for establishing such stochastic ground motion models are record-based and physics-based approaches. Record-based models (also known as site-based) are developed by fitting a preselected “waveform” to a suite of recorded regional ground motions. Regression analysis is used for establishing the predictive models, which leads to a probabilistic characterization described by mean predictions along with an associated variability.^{16,19} On the other hand, stochastic physics-based models rely on physical modeling of the rupture and wave propagation mechanisms.^{20,21} The predictive relationships in this case are typically described by deterministic models that represent the underlying mean physical properties, though approaches exist for addressing variability in these properties.^{13,24} Emphasis in this study will be on record-based models, though the techniques discussed can be extended to any type of stochastic ground motion model, with the assumption that the corresponding predictive models are characterized by both a mean prediction and an associated variance.

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 models, compatibility to the seismic hazard for specific structures and sites is not necessarily obtained. This hazard is typically characterized through a Probabilistic Seismic Hazard Analysis,²⁵ for example through de-aggregation that identifies the seismicity scenarios, described through relevant seismicity characteristics, mainly the moment magnitude and rupture distance, that have the largest contribution to the hazard for a specific structure. Essential part of Probabilistic Seismic Hazard Analysis are ground motion prediction equations (GMPEs). GMPEs provide predictions, as function of seismicity characteristics, for both the median and the dispersion of IMs, determining ultimately the conditional hazard for seismic events corresponding to these characteristics.

Recognizing the importance of matching stochastic ground motion models to some target IM the modification of the predictive relationships for accommodating such a match was recently examined.^{26,27} This was posed as an optimization problem with objective to minimize the discrepancy between the median predictions of the ground motion model and the spectral acceleration estimates of GMPEs for different period ranges (or, more broadly, a range of target IMs). Appropriate constraints were set in this optimization to preserve desired ground motion physical characteristics, so that unrealistic time-histories are not obtained. This formulation was first introduced in Scherbaum et al,²⁶ with Vetter et al²⁷ offering a computationally efficient approach for performing the associated optimization, leveraging surrogate modeling principles. Recent work by the authors²⁸ extended this framework by addressing a critical shortcoming, the fact that physical descriptors of the resulting acceleration time-series were incorporated in the optimization merely as constraints, something that required significant experience in ground motion characterization for proper definition of the optimization problem. The shortcoming was addressed²⁸ by introducing a bi-objective optimization problem, transforming the aforementioned constraint for the physical characteristics of the resultant ground acceleration time-series to an explicit objective. The proposed tuning was performed for specific seismicity scenarios and identified the ground motion model that achieves the minimum modification of the existing predictive relationships that will yield the desired compatibility with the target IM.

All the aforementioned studies focused, however, on the mean model characteristics and associated hazard. Optimization utilized only the mean of the predictive relationships of the stochastic ground motion model, whereas, more importantly, match only to a target IM vector (taken to represent the mean hazard) was investigated, ignoring any variability in the IM predictions. The latter is an important constraint because for seismic risk assessment applications

hazard compatibility is expressed in terms of both the mean and dispersion of the target IMs.²⁵ From a practical standpoint, capturing the actual variability of target IMs is essential to capture extreme structural response values and, therefore, in properly assessing the likelihood of consequences due to such extreme seismic demand values.

The current study extends approach²⁸ to (1) match the prescribed conditional hazard (not simply mean IMs) for a specific site and structure (or range of structures) while (2) preserving desired trends and correlations in the physical characteristics of the resultant ground acceleration time-series, including consideration of the variability of these characteristics. This is again formulated as a bi-objective optimization problem. The first objective is to minimize the discrepancy between the statistics (mean and dispersion) of the outputs for a suite of acceleration time-histories obtained from the ground motion model and the target IM statistics for a given seismicity scenario. The second objective is to establish the smallest deviation from the model characteristics suggested by existing predictive models, examining both the mean and the variance of the model parameters. Both objectives are expressed as comparison between probability distributions, and the relative entropy is adopted to quantify them. This setting creates a fundamental difference to the previous study²⁸ with respect to both the optimization characteristics (alter both mean and variance of predictive models) as well as the goal (match to mean and dispersion for conditional hazard). In Tsioulou et al,²⁸ the ground motion model was tuned so that outputs from a single parametric description of the ground motion model match a target IM vector for each seismicity scenario. The goal of the current study is to produce an ensemble of ground motion models whose output statistics yield the desired compatibility with the hazard (IM mean and dispersion) for that scenario. For efficiently solving the multi-objective optimization problem, a surrogate modeling approach is adopted^{27,28} for approximating the desired IMs for specific values of the ground motion model parameters. Emphasis is placed here on the efficient estimation of the response statistics for the modified ground motion model output, leveraging Monte Carlo techniques. This requires further extension of the surrogate modeling framework, compared with the approach adopted in Tsioulou et al,²⁸ for facilitating this estimation. Different assumptions are also examined for the evaluation of the entropy for the first objective. The corresponding bi-objective optimization is finally solved using a random search approach. The novelty of the current work stems from both the fundamentally different theoretical framework for formulating the stochastic ground motion modification as well as the computational advances required for efficiently calculating the response statistics for the first objective and performing the associated optimization.

In the next section, the general problem of developing simulated ground motions compatible with target IM distributions is defined, and then specific aspects of the framework are discussed in detail.

2 | PROBLEM FORMULATION

2.1 | Preliminaries and baseline predictive relationships formulation

The foundation of the problem formulation is the same as in Tsioulou et al.²⁸ A stochastic ground motion model is considered that provides acceleration time-histories $\ddot{a}(t|\boldsymbol{\theta}, \mathbf{w})$ by modulating a discretized Gaussian white-noise sequence, \mathbf{w} , through appropriate time/frequency functions that are parameterized through the n_θ -dimensional *model parameter* vector $\boldsymbol{\theta} = [\theta_1 \ \theta_2 \ \dots \ \theta_{n_\theta}] \in \mathcal{R}^{n_\theta}$. $\boldsymbol{\theta}$ completely defines the parametric description of the model, ie, along with \mathbf{w} facilitates the generation of a time-history $\ddot{a}(t|\boldsymbol{\theta}, \mathbf{w})$, 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. It should be noted that $\boldsymbol{\theta}$ corresponds typically to a low-dimensional vector and \mathbf{w} to a high-dimensional sequence.

Predictive models/relationships are utilized to relate $\boldsymbol{\theta}$ to seismicity and local site characteristics, such as the fault type F , the moment magnitude, M , the rupture distance, R , and the shear wave velocity in the upper 30 m of soil, V_{s30} .^{16,20} The vector of these characteristics is denoted as \mathbf{z} . For record-based models, the standard approach for development of these predictive relationships^{16,29} relies on first matching the waveform of recorded ground motions (ie, identify first $\boldsymbol{\theta}$ for each of the recorded ground motions in a given database) and then carrying out a regression to relate $\boldsymbol{\theta}$ to \mathbf{z} . Typically, this is performed by first transforming problem to the standard Gaussian space through a nonlinear mapping for each component θ_i .¹⁶ The transformed Gaussian vector is denoted \mathbf{v} . Approach ultimately leads to a Gaussian probability model $\mathbf{v} \sim \mathcal{N}(\boldsymbol{\mu}_r(\mathbf{z}), \boldsymbol{\Sigma}_r)$ with mean $\boldsymbol{\mu}_r(\mathbf{z})$ and covariance matrix $\boldsymbol{\Sigma}_r$. Note that the latter is independent of \mathbf{z} . The notation $\mathcal{N}(a, b)$ stands for Normal distribution with mean a and covariance b whereas notation $c \sim d$ stands for random variable c following distribution d . The resultant probability model for $\boldsymbol{\theta}$ is denoted $p(\boldsymbol{\theta}|\boldsymbol{\mu}_r(\mathbf{z}), \boldsymbol{\Sigma}_r)$ and determines the predictive model for the ground motion model parameters. Note that a similar description can be readily established

for physics-based models,²⁸ with the uncertainty characterization stemming from an explicit treatment of the epistemic uncertainties associated with the physics-based formulation.^{13,24} The predictive model for θ will be denoted herein as $p(\theta|\boldsymbol{\mu}_r(\mathbf{z}), \boldsymbol{\Sigma}_r)$, with the understanding that it is not necessarily constrained to models established through regression analysis, rather simply can be parameterized by quantities $\boldsymbol{\mu}_r(\mathbf{z})$ and $\boldsymbol{\Sigma}_r$, with $\boldsymbol{\mu}_r(\mathbf{z})$ representing the mean predictions, and $\boldsymbol{\Sigma}_r$ the variability of these predictions.

2.2 | Modification of predictive models for hazard matching

As discussed in the introduction, the formulation of the predictive model for θ provides synthetic ground motions whose statistics (mean and dispersion) of output IMs do not necessarily match the intended hazard for specific structures and sites. For accommodating such a match, a modification of the existing predictive model for θ is proposed for specific seismicity scenarios defined through \mathbf{z} , with objective to get a suite of acceleration time-series for that scenario whose (1) mean and dispersion match a target IM mean and dispersion vectors, while (2) maintaining similarity to the predictive relationships already established for the model. Equivalently, this can be viewed as identifying the updated probabilistic model $p(\theta|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ that is closest to the established model $p(\theta|\boldsymbol{\mu}_r(\mathbf{z}), \boldsymbol{\Sigma}_r)$ and also matches the intended conditional hazard. $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ represent the updated parametric description for the probability model of θ . In the context of record-based models, these correspond, respectively, to the mean vector and covariance matrix for \mathbf{v} . The IM vector may include different response quantities of interest,²⁸ for example direct characteristics of the ground motion, such as peak ground acceleration or elastic and inelastic spectral responses for different periods of a single degree of freedom oscillator. The conditional hazard for most of these IMs may be described through a GMPE.^{30,31} Because the proposed modification refers to the conditional hazard, for simplifying terminology for the remainder of the paper, the description as “conditional” will be removed: term hazard corresponds to conditional hazard.

To formalize these concepts mathematically, let, $Y_i(\mathbf{z}); i = 1, \dots, n_y$, denote the IMs of interest. The target hazard for them is quantified through a probabilistic description, denoted by $p_t(\ln(Y_i)|\mathbf{z})$. As is common in earthquake engineering and without loss of generality, the statistics are assumed here to be determined for the logarithm of the IM. To better align approach with current GMPE standards a lognormal underlying model is assumed, leading to $\ln(Y_i(\mathbf{z}))^t \sim \mathcal{N}(\ln(\bar{Y}_i(\mathbf{z})), \sigma_i^2(\mathbf{z}))$ with $\ln(\bar{Y}_i(\mathbf{z}))$ and $\sigma_i^2(\mathbf{z})$ corresponding to the mean and variance, respectively, of the logarithmic IM. Note, however, that the computational framework can support any probabilistic IM description $p_t(\ln(Y_i)|\mathbf{z})$, not constrained to one provided by GMPEs or defined through a lognormal probabilistic model. Also, the superscripts/subscripts t and g (the latter defined in the next paragraph) are used herein to distinguish between target prediction and prediction facilitated through the stochastic ground motion model.

To quantify the hazard predictions through the stochastic ground motion model, let $Y_i^g(\theta, \mathbf{w})$ denote the estimate for Y_i established through this model for specific values of the model parameter vector θ and a specific white noise sequence \mathbf{w} (ie for a specific ground motion time-history $\ddot{a}(t|\theta, \mathbf{w})$). $Y_i^g(\theta, \mathbf{w})$ will be referenced herein as response output of the ground motion model. The statistical characterization for $\ln(Y_i)$ through the stochastic ground motion model for the updated parametric description is denoted by $p_g(\ln(Y_i)|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and using the total probability theorem is equal to:

$$p_g(\ln(Y_i)|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \int \int p(\ln(Y_i^g)|\theta, \mathbf{w}, \boldsymbol{\mu}, \boldsymbol{\Sigma})p(\theta, \mathbf{w}|\boldsymbol{\mu}, \boldsymbol{\Sigma})d\theta d\mathbf{w} = \int \int p(\ln(Y_i^g)|\theta, \mathbf{w})p(\theta|\boldsymbol{\mu}, \boldsymbol{\Sigma})p(\mathbf{w})d\theta d\mathbf{w} \quad (1)$$

where $p(\mathbf{w})$ is the probability distribution for the stochastic sequence \mathbf{w} , which by definition is independent of $p(\theta|\boldsymbol{\mu}, \boldsymbol{\Sigma})$. In deriving the second equality in Equation 1 the fact that the stochastic ground motion model is completely defined by pair θ, \mathbf{w} was also used. This simplifies $p(\ln(Y_i^g)|\theta, \mathbf{w}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = p(\ln(Y_i^g)|\theta, \mathbf{w})$ because knowledge of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ is redundant if θ is also known. Note that Y_i^g is itself a random variable with randomness in its description stemming from both \mathbf{w} and θ . In this context, $Y_i^g(\theta, \mathbf{w})$ may be viewed as a realization of the random variable.

Distribution $p_g(\ln(Y_i)|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ can be approximated through kernel density estimation (KDE) as discussed in Appendix A. To further simplify framework, the same hypothesis as for the hazard can be adopted, assuming a lognormal distribution, leading to $\ln(Y_i)^g \sim \mathcal{N}(\ln(\bar{Y}_i^g(\boldsymbol{\mu}, \boldsymbol{\Sigma})), (\sigma_i^g(\boldsymbol{\mu}, \boldsymbol{\Sigma}))^2)$, where the mean and variance for $\ln(Y_i)^g$, considering the variability in both the low-dimensional θ and high-dimensional \mathbf{w} , are:

$$\ln(\bar{Y}_i^g(\boldsymbol{\mu}, \boldsymbol{\Sigma})) = E[\ln(Y_i^g)] = \int \int \ln(Y_i^g(\theta, \mathbf{w}))p(\theta|\boldsymbol{\mu}, \boldsymbol{\Sigma})p(\mathbf{w})d\theta d\mathbf{w} \quad (2)$$

$$(\sigma_i^g(\boldsymbol{\mu}, \boldsymbol{\Sigma}))^2 = \text{Var}[\ln(Y_i^g)] = \int \int [\ln(Y_i^g(\theta, \mathbf{w})) - \ln(\bar{Y}_i^g(\boldsymbol{\mu}, \boldsymbol{\Sigma}))]^2 p(\theta|\boldsymbol{\mu}, \boldsymbol{\Sigma})p(\mathbf{w})d\theta d\mathbf{w} \quad (3)$$

with $E[\cdot]$, $Var[\cdot]$ denoting the expectation and variance operators, respectively. The functional dependence of the statistics of $\ln(Y_i(\mathbf{z}))^g$ on $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ is explicitly noted herein to facilitate an easier understanding of the ground motion model modification framework. The log-normal assumption will be primarily used for $p_g(\ln(Y_i)|\boldsymbol{\mu}, \boldsymbol{\Sigma})$, and its validity will be examined in the illustrative example later by comparing to the KDE-based estimation.

The hazard compatible modeling corresponds to modification of the probability model for $\boldsymbol{\theta}$, ultimately of the parametric description defined through $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, and is formulated as multi-objective optimization problem with two competing objectives

$$[\boldsymbol{\mu}, \boldsymbol{\Sigma}]^* = \arg \min \{F_{p1}(\boldsymbol{\mu}, \boldsymbol{\Sigma}|\mathbf{z}), F_{p2}(\boldsymbol{\mu}, \boldsymbol{\Sigma}|\mathbf{z})\} \quad (4)$$

The first objective F_{p1} corresponds to the weighted discrepancy of the target seismic hazard to the hazard predicted through the ground motion model, ie, to a comparison between $p_g(\ln(Y_i(\mathbf{z})|\boldsymbol{\mu}(\mathbf{z}), \boldsymbol{\Sigma}))$ and $p_t(\ln(Y_i)|\mathbf{z})$. The relative entropy, a popular measure to quantify differences between distributions,³² is utilized as metric for this discrepancy, given by

$$D\left[p_g(\ln(Y_i)|\boldsymbol{\mu}, \boldsymbol{\Sigma})\|p_t(\ln(Y_i)|\mathbf{z})\right] = \int_{\mathfrak{R}} p_g(\ln(Y_i)|\boldsymbol{\mu}, \boldsymbol{\Sigma}) \log \left[\frac{p_g(\ln(Y_i)|\boldsymbol{\mu}, \boldsymbol{\Sigma})}{p_t(\ln(Y_i)|\mathbf{z})} \right] d \ln(Y_i). \quad (5)$$

This leads to definition of F_{p1} as

$$\begin{aligned} F_{p1}(\boldsymbol{\mu}, \boldsymbol{\Sigma}|\mathbf{z}) &= \frac{1}{\sum_{i=1}^{n_y} \gamma_i} \sum_{i=1}^{n_y} \gamma_i D\left[p_g(\ln(Y_i)|\boldsymbol{\mu}, \boldsymbol{\Sigma})\|p_t(\ln(Y_i)|\mathbf{z})\right] \\ &= \frac{1}{\sum_{i=1}^{n_y} \gamma_i} \sum_{i=1}^{n_y} \gamma_i D\left[p_g(\ln(Y_i)|\boldsymbol{\mu}, \boldsymbol{\Sigma})\|N(\ln(\bar{Y}_i(\mathbf{z})), \sigma_i^2(\mathbf{z}))\right] \end{aligned} \quad (6)$$

with γ_i corresponding to the weights prioritizing the match to different IM components (eg, spectral accelerations at different structural periods). For the assumption of lognormal distribution for Y_i^g , this simplifies to

$$F_{p1}(\boldsymbol{\mu}, \boldsymbol{\Sigma}|\mathbf{z}) = \frac{1}{\sum_{i=1}^{n_y} \gamma_i} \sum_{i=1}^{n_y} \gamma_i D\left[N(\ln(\bar{Y}_i^g(\boldsymbol{\mu}, \boldsymbol{\Sigma})), (\sigma_i^g(\boldsymbol{\mu}, \boldsymbol{\Sigma}))^2)\|N(\ln(\bar{Y}_i(\mathbf{z})), \sigma_i^2(\mathbf{z}))\right] \quad (7)$$

with closed-form solution³²:

$$\begin{aligned} D\left[N(\ln(\bar{Y}_i^g(\boldsymbol{\mu}, \boldsymbol{\Sigma})), (\sigma_i^g(\boldsymbol{\mu}, \boldsymbol{\Sigma}))^2)\|N(\ln(\bar{Y}_i(\mathbf{z})), \sigma_i^2(\mathbf{z}))\right] &= \\ \frac{(\ln(\bar{Y}_i^g(\boldsymbol{\mu}, \boldsymbol{\Sigma})) - \ln(\bar{Y}_i(\mathbf{z})))^2}{2\sigma_i^2(\mathbf{z})} + \frac{1}{2} \left[\frac{(\sigma_i^g(\boldsymbol{\mu}, \boldsymbol{\Sigma}))^2}{\sigma_i^2(\mathbf{z})} - 1 - \ln \left(\frac{(\sigma_i^g(\boldsymbol{\mu}, \boldsymbol{\Sigma}))^2}{\sigma_i^2(\mathbf{z})} \right) \right]. \end{aligned} \quad (8)$$

Objective F_{p2} measures that discrepancy between the initial predictive model for $\boldsymbol{\theta}$, $p(\boldsymbol{\theta}|\boldsymbol{\mu}_r(\mathbf{z}), \boldsymbol{\Sigma}_r)$, and the modified one, $p(\boldsymbol{\theta}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$. The relative entropy is utilized again as measure to quantify differences:

$$F_{p2}(\boldsymbol{\mu}, \boldsymbol{\Sigma}|\mathbf{z}) = D[p(\boldsymbol{\theta}|\boldsymbol{\mu}, \boldsymbol{\Sigma})\|p(\boldsymbol{\theta}|\boldsymbol{\mu}_r(\mathbf{z}), \boldsymbol{\Sigma}_r)] = \int_{\mathfrak{R}^{n_\theta}} p(\boldsymbol{\theta}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) \log \left[\frac{p(\boldsymbol{\theta}|\boldsymbol{\mu}, \boldsymbol{\Sigma})}{p(\boldsymbol{\theta}|\boldsymbol{\mu}_r(\mathbf{z}), \boldsymbol{\Sigma}_r)} \right] d\boldsymbol{\theta}. \quad (9)$$

Because the relative entropy is invariant under a coordinate transformation for $\boldsymbol{\theta}$,³² the comparison can be established in the transformed Gaussian space, leading to

$$F_{p2}(\boldsymbol{\mu}, \boldsymbol{\Sigma}|\mathbf{z}) = D[N(\boldsymbol{\mu}, \boldsymbol{\Sigma})\|N(\boldsymbol{\mu}_r(\mathbf{z}), \boldsymbol{\Sigma}_r)] \quad (10)$$

with the latter expression readily evaluated³² as

$$D[N(\boldsymbol{\mu}, \boldsymbol{\Sigma})\|N(\boldsymbol{\mu}_r(\mathbf{z}), \boldsymbol{\Sigma}_r)] = 1/2 \left[\text{tr}[\boldsymbol{\Sigma}\boldsymbol{\Sigma}_r^{-1}] + (\boldsymbol{\mu}_r(\mathbf{z}) - \boldsymbol{\mu})^T \boldsymbol{\Sigma}_r^{-1} (\boldsymbol{\mu}_r(\mathbf{z}) - \boldsymbol{\mu}) - n_\theta - \ln(\det[\boldsymbol{\Sigma}\boldsymbol{\Sigma}_r^{-1}]) \right] \quad (11)$$

where $\text{tr}[\cdot]$ and $\det[\cdot]$ stand for trace and determinant, respectively.

Objective F_{p1} enforces hazard compatibility (mean and dispersion), whereas objective F_{p2} guarantees compatibility of the physical characteristics of the resultant ground motions with the regional trends observed in recorded ground

motions. Solution of the multi-objective optimization of Equation 4 ultimately leads to a Pareto set of dominant solutions $\{(\boldsymbol{\mu}_p, \boldsymbol{\Sigma}_p); p = 1, \dots, n_p\}$ that expresses a different compromise between the competing objectives F_{p1} and F_{p2} . A solution is characterized as dominant (or Pareto optimal) and belongs in the Pareto set if there is no other solution that simultaneously improves both objectives F_{p1} and F_{p2} . The representation of the Pareto set in the performance objective $[F_{p1}, F_{p2}]$ space, $\{[F_{p1}(\boldsymbol{\mu}_p, \boldsymbol{\Sigma}_p), F_{p2}(\boldsymbol{\mu}_p, \boldsymbol{\Sigma}_p)]; p = 1, \dots, n_p\}$ is termed as the Pareto front. Illustrations of such Pareto fronts are included in the example discussed later. One can eventually select a model configuration from the identified Pareto set that yields the desired hazard compatibility without deviating significantly from regional ground motion characteristics. This will be further discussed in the illustrative implementation. Note that a Pareto set for optimization of Equation 4 always exist apart from the extreme case that the original stochastic ground motion model provides an exact match to the hazard, ie, $F_{p1}(\boldsymbol{\mu}_r(\mathbf{z}), \boldsymbol{\Sigma}_r | \mathbf{z}) = 0$. In that case, the entire set corresponds to a single point $(\boldsymbol{\mu}_r(\mathbf{z}), \boldsymbol{\Sigma}_r)$.

Identifying the Pareto set is, however, challenging because the computational burden in evaluation of objective F_{p1} is significant, requiring calculation of the multidimensional integrals of Equations 2 and 3 (or samples needed for the KDE approximation discussed in Appendix A) which can be only performed numerically and entails hundreds estimates of the response output of the ground motion model. To facilitate an efficient optimization that can be repeated for any desired seismicity scenario \mathbf{z} , a surrogate modeling (ie, metamodeling) approach is adopted, specifically selecting kriging as metamodel due to its proven capability to approximate well even complex functions.³³ As input for the metamodel, the low-dimensional vector $\boldsymbol{\theta}$ is chosen. This choice corresponds to the smallest possible dimension for the metamodel input, something that can greatly enhance accuracy,³³ a very important consideration because metamodel will be eventually used for optimization, and larger metamodel errors can lead to identification of suboptimal solutions. Alternative choices were to additionally include \mathbf{w} in the metamodel input definition which is completely impractical (due to high-dimensionality of \mathbf{w}), or to use directly pair $\{\boldsymbol{\mu}, \boldsymbol{\Sigma}\}$ as the metamodel inputs (and statistics of Equations 2 and 3 as outputs) which significantly increases, however, the input dimension. Under this selection of $\boldsymbol{\theta}$ as metamodel input, the metamodel output corresponds to the response output statistics considering variability with respect to \mathbf{w} only (ie, conditional on $\boldsymbol{\theta}$ statistics). Here, statistics refers to all quantities needed to eventually calculate F_{p1} . The variability with respect to $\boldsymbol{\theta}$ in quantifying the hazard (and ultimately objective F_{p1}) will be addressed through the approach discussed in Section 4.

This creates a similar setting as in Tsioulou et al²⁸ for the metamodel formulation, with the additional requirement, however, to eventually incorporate variability stemming from $\boldsymbol{\theta}$ in the overall framework, so that variability in both $\boldsymbol{\theta}$ and \mathbf{w} is explicitly considered. The kriging metamodel development is briefly reviewed in the next section, and the details of the kriging-aided optimization problem are discussed in Section 4. A schematic of the overall optimization approach is provided in Figure 1 on the next page.

3 | KRIGING METAMODEL DEVELOPMENT

As discussed in the previous section, $\boldsymbol{\theta}$ is chosen as metamodel input. A modification of this input should be further adopted if the relationship between some components of $\boldsymbol{\theta}$ and the response output $Y_i^g(\boldsymbol{\theta}, \mathbf{w})$ is explicitly known. This is true for stochastic ground motion models that include a scaling parameter, denoted θ_s herein, that directly controls the amplitude of the excitation. For example, for the model that will be used in the illustrative example¹⁶ θ_s corresponds to the Arias intensity. This means that $Y_i^g(\boldsymbol{\theta}, \mathbf{w}) = f(\theta_s) s_i(\mathbf{x}, \mathbf{w})$ with \mathbf{x} corresponding to the remaining model parameters excluding θ_s , $s_i(\mathbf{x}, \mathbf{w})$ representing the output $Y_i^g(\boldsymbol{\theta}, \mathbf{w})$ for $\theta_s = 1$ and $f(\theta_s)$ being a simple function of θ_s . Without loss of generality, we will adopt here this assumption, ie, representation $Y_i^g(\boldsymbol{\theta}, \mathbf{w}) = \sqrt{\theta_s} s_i(\mathbf{x}, \mathbf{w})$. The choice of square root for function $f(\cdot)$ is made here simply to match the model used in the illustrative example. Numerical approach discussed herein can accommodate any other function. This setting leads to modification of the metamodel input to \mathbf{x} , and, respectively, of metamodel output to the statistics of $s_i(\mathbf{x}, \mathbf{w})$ conditional on \mathbf{x} . Once the latter statistics are known the statistics for $Y_i^g(\boldsymbol{\theta}, \mathbf{w})$ conditional on $\boldsymbol{\theta}$ can be easily obtained as will be shown in Section 4, because the relationship to the remaining component of vector $\boldsymbol{\theta}$, θ_s , is explicitly known. This modification ultimately reduces the dimension of the metamodel input, which as discussed earlier facilitates improved accuracy. Note that similar to Y_i^g , s_i is also a random variable with randomness in its description stemming from \mathbf{w} and \mathbf{x} .

The approximated statistics, the ones required for calculating F_{p1} , are the conditional on \mathbf{x} mean and variance of $\ln(s_i)$, given by

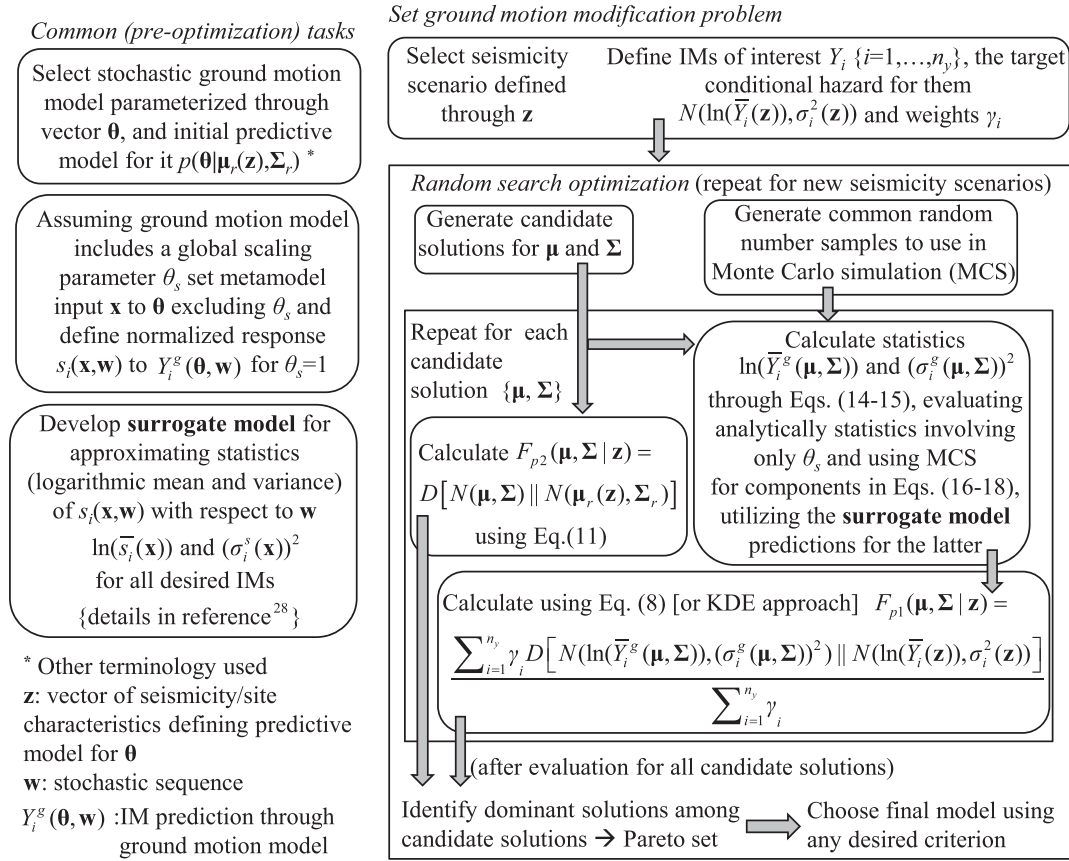


FIGURE 1 Overview schematic of the hazard compatible stochastic ground motion modeling approach

$$\ln(\bar{s}_i(\mathbf{x})) = E[\ln(s_i)|\mathbf{x}] = \int \ln(s_i(\mathbf{x}, \mathbf{w}))p(\mathbf{w})d\mathbf{w} \quad (12)$$

$$(\sigma_i^g(\mathbf{x}))^2 = Var[\ln(s_i)|\mathbf{x}] = \int [\ln(s_i(\mathbf{x}, \mathbf{w})) - \ln(\bar{s}_i(\mathbf{x}))]^2 p(\mathbf{w})d\mathbf{w} \quad (13)$$

where independence of \mathbf{w} from \mathbf{x} was used in deriving the last equalities in both expressions. These statistics define the metamodel output. It should be pointed out that this output is independent of the predictive relationships; rather is simply functions of \mathbf{x} (reason for introducing the functional dependence on \mathbf{x} notation). This is what contributes to the efficiency of the approach: the surrogate model is established with respect to the low dimensional vector \mathbf{x} and can be then leveraged to evaluate the required statistics for different selections of the predictive models (ie, different μ and Σ) as detailed in the next section.

For developing the metamodel, a database with n observations is initially obtained that provides information for the $\mathbf{x}-\ln(\bar{s}_i(\mathbf{x}))$ and $\mathbf{x}-\sigma_i^g(\mathbf{x})$ input/output pairs. For this purpose, n samples for $\{\mathbf{x}^j, j = 1, \dots, n\}$ also known as *support points* or *experiments* are obtained over the domain of interest for \mathbf{x} . This domain, denoted X , should encompass the anticipated range that the metamodel will be implemented in to avoid extrapolations, ie, domain covered by $p(\theta|\mu, \Sigma)$, approximated in this case as domain for $p(\theta|\mu_r(\mathbf{z}), \Sigma_r)$ (for further details for definition of X please refer to Tsioulou et al²⁸). The predictions provided through the ground motion model for each \mathbf{x}^j are then established through the following process: (Step 1) generate n_w sample acceleration time-histories for different white-noise sequences $\{\mathbf{w}^k, k = 1, \dots, n_w\}$ using $\theta_s = 1$ for all samples; (Step 2) for each sample evaluate the response outputs of interest $\{s_i(\mathbf{x}^j, \mathbf{w}^k); i = 1, \dots, n_y\}$ using response-history analysis for spectral IMs; (Step 3) estimate the statistics (logarithmic mean and variance) over the sample-set to obtain $\ln(\bar{s}_i(\mathbf{x}^j))$ and $\sigma_i^g(\mathbf{x}^j)$.

Using this database, the surrogate model can be formulated following the approach presented in Tsioulou et al.²⁸ Only difference is that output for metamodel also includes $\sigma_i^g(\mathbf{x})$. The computationally intensive aspect of the formulation is the development of the database which requires response-history analysis. This needs to be performed, though, only once. As soon as the metamodel is established using this database, it can predict $\ln(\bar{s}_i(\mathbf{x}))$ and $\sigma_i^g(\mathbf{x})$ for any other \mathbf{x} desired. Metamodel predictions can be also vectorized,³⁴ something that will be leveraged in the numerical

optimization discussed in the next section. The accuracy of the metamodel depends on the number of experiments n as well as the exact selection of these experiments. Details for both these tasks, including a sequential, adaptive metamodel formulation that gradually increases n , stopping when pre-specified accuracy criteria are satisfied, are provided in Tsioulou et al.²⁸

4 | MULTI-OBJECTIVE OPTIMIZATION TO MATCH CONDITIONAL HAZARD SUPPORTED BY KRIGING METAMODELING

4.1 | Calculation of statistics of interest

Evaluation of objective F_{p1} boils down to estimation of Equations 2 and 3. Using representation $Y_i^g(\boldsymbol{\theta}, \mathbf{w}) = \sqrt{\theta_s} s_i(\mathbf{x}, \mathbf{w})$, this simplifies to

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

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

where $\text{Cov}[a,b]$ stands for the covariance between random variables a and b . The statistics with respect to θ_s , that is the mean $E[\ln(\theta_s)]$ and variance $\text{Var}[\ln(\theta_s)]$ can be readily calculated using the marginal distribution $p(\theta_s|\boldsymbol{\mu}, \boldsymbol{\Sigma})$. The statistics that involve s_i may be calculated using the metamodel to approximate variability with respect to \mathbf{w} . Using the laws of total expectation and variance we have for these statistics

$$E[\ln(s_i)] = E[E[\ln(s_i)|\mathbf{x}]] = E[\ln(\bar{s}_i(\mathbf{x}))] = \int \ln(\bar{s}_i(\mathbf{x})) p(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) d\mathbf{x} \quad (16)$$

$$\begin{aligned} \text{Var}[\ln(s_i)] &= E[\text{Var}[\ln(s_i)|\mathbf{x}]] + \text{Var}[E[\ln(s_i)|\mathbf{x}]] = E\left[(\sigma_i^s(\mathbf{x}))^2\right] + \text{Var}[\ln(\bar{s}_i(\mathbf{x}))] \\ &= \int (\sigma_i^s(\mathbf{x}))^2 p(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) d\mathbf{x} + \int (\ln(\bar{s}_i(\mathbf{x})) - E[\ln(\bar{s}_i(\mathbf{x}))])^2 p(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) d\mathbf{x} \end{aligned} \quad (17)$$

$$\begin{aligned} \text{Cov}[\ln(\theta_s), \ln(s_i)] &= E[\ln(\theta_s) \ln(s_i)] - E[\ln(\theta_s)]E[\ln(s_i)] = E[E[\ln(\theta_s) \ln(s_i)|\mathbf{x}]] - E[\ln(\theta_s)]E[\ln(s_i)] \\ &= E[\ln(\theta_s)E[\ln(s_i)|\mathbf{x}]] - E[\ln(\theta_s)]E[\ln(s_i)] = E[\ln(\theta_s) \ln(\bar{s}_i(\mathbf{x}))] - E[\ln(\theta_s)]E[\ln(s_i)] \\ &= \int \ln(\theta_s) \ln(\bar{s}_i(\mathbf{x})) p(\boldsymbol{\theta}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) d\boldsymbol{\theta} - E[\ln(\theta_s)]E[\ln(s_i)] \end{aligned} \quad (18)$$

where the expectation (integrals) with respect to \mathbf{x} or $\boldsymbol{\theta}$ have been explicitly expressed in all these equations. These integrals address the variability stemming from the predictive relationships and need to be calculated through Monte Carlo simulation (MCS). Rather than performing two different MCS: one for the integrals involved in the expectation $E[\ln(s_i)]$ and variance $\text{Var}[\ln(s_i)]$, which require samples from the marginal distribution $p(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$, and one for the covariance $\text{Cov}[\ln(\theta_s), \ln(s_i)]$, which requires samples from joint distribution $p(\boldsymbol{\theta}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$, a single MCS is performed, utilizing a common set of samples for all these statistics. This leads to the following approximations for the quantities needed for Equations 14 and 15

$$E[\ln(s_i)] \approx \frac{1}{N_s} \sum_{j=1}^{N_s} \ln(\bar{s}_i(\mathbf{x}^j)) \quad (19)$$

$$\text{Var}[\ln(s_i)] \approx \frac{1}{N_s} \sum_{j=1}^{N_s} (\sigma_i^s(\mathbf{x}^j))^2 + \frac{1}{N_s} \sum_{j=1}^{N_s} (\ln(\bar{s}_i(\mathbf{x}^j)) - E[\ln(s_i)])^2 \quad (20)$$

$$\text{Cov}[\ln(\theta_s), \ln(s_i)] \approx \frac{1}{N_s} \sum_{j=1}^{N_s} \ln(\theta_s^j) \ln(\bar{s}_i(\mathbf{x}^j)) - E[\ln(\theta_s)]E[\ln(s_i)] \quad (21)$$

where $[\theta_s^j, \mathbf{x}^j]$ correspond to samples from $p(\theta|\boldsymbol{\mu}, \boldsymbol{\Sigma})$, N_s is the total number of samples used, and $\ln(\bar{s}_i(\mathbf{x}))$ and $\sigma_i^s(\mathbf{x})$ are approximated through the kriging metamodel for each one of these samples. Utilizing vectorized manipulations for the metamodel predictions, both these quantities can be calculated with very small computational effort, meaning that the MCS-based estimation of Equations 19 to 21 can be performed very efficiently. Further numerical details (computational times) are provided in the illustrative example.

For reducing the relative importance of the MCS estimation error within the multi-objective problem of Equation 4, common random numbers (CRN) are utilized.³⁵ This is facilitated by getting first samples in the standard Gaussian space (which is independent of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$) and then transforming them to the desired samples for θ . Approach is equivalent to transforming integrals in Equations 16 to 18 to standard Gaussian space and for record-based model is easily performed, as these models typically entail¹⁶ a Gaussian predictive model for \mathbf{v} . The CRN is implemented by using the same sample set for the standard Gaussian samples across the entire optimization as also shown earlier in Figure 1. Adoption of CRN creates a consistent estimation error in the MCS application for the different examined $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ values and therefore improves the optimization accuracy for identifying the correct Pareto front.³⁵ In other words, it allows use of smaller value for N_s because it reduces the relative importance of the MCS estimation error within the optimization.

4.2 | Multi objective optimization

Calculation of statistics given by Equations 14 and 15, utilizing MCS estimates of Equations 19 to 21, facilitates an efficient approximation for performance objective F_{p1} given by Equation 8. If the lognormal assumption for the distribution of Y_i^g is not used, then F_{p1} can be estimated through the approach discussed in Appendix A, leveraging again MCS principles. Note that calculation of objective F_{p2} is computationally trivial. Therefore, through the introduction of the metamodel, an efficient estimation of both objectives involved in the optimization described by Equation 4 can be established. It should be noted that in Tsioulou et al²⁸ the explicit incorporation in the optimization of the metamodel error was also considered, established through appropriate modification of the equivalent objective function F_{p1} . It was shown, however, that this reduces computational efficiency and is not necessary if the underlying surrogate model has high accuracy. As calculation of objective F_{p1} requires N_s evaluations of the metamodel in a MCS setting (only 1 evaluation was needed in Tsioulou et al²⁸) the inclusion of the metamodel error is not advocated here as the computational burden is expected to be higher. Additionally, the benefits of explicitly including this error in the current formulation are expected to be smaller because objective F_{p1} represents a statistical quantity over θ . In evaluating the necessary statistic, the response, and therefore associated metamodel errors, over different θ values are averaged. Potential large errors that may exist for specific θ values end up averaged with smaller errors from other θ 's and therefore do not impact F_{p1} estimates as much as they would if these θ values were the only contribution to F_{p1} (as was the case in Tsioulou et al²⁸).

For solving the resultant multi-objective optimization, a variety of numerical approaches can be utilized.³⁶ In Tsioulou et al,²⁸ two such approaches were examined, one gradient-free and one gradient-based, and preference was ultimately given to the former because it can support higher robustness and computational efficiency for identifying the entire Pareto front. This recommendation is adopted here, and a gradient-free, random search approach is implemented as follows. A large number of n_{bc} candidate solutions for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are generated that are close to $\boldsymbol{\mu}_r(\mathbf{z})$ and $\boldsymbol{\Sigma}_r$. This is established by creating uniform random samples centered around $\boldsymbol{\mu}_r(\mathbf{z})$ and $\boldsymbol{\Sigma}_r$. The range for these samples is chosen so that the value of F_{p2} does not become excessively large, because the latter indicates significant departure of the modified model from the initial one, which might produce ground motions with unrealistic characteristics. Because estimation of F_{p2} is simple, a pre-screening can be implemented, rejecting any candidate solution with (large) value of F_{p2} over some desired threshold. Value close to 10 is appropriate for the latter threshold. This choice corresponds to modification of the standard deviation up to 50% and modification of $\boldsymbol{\mu}$ within hypersphere 5 standard deviations away from $\boldsymbol{\mu}_r(\mathbf{z})$. Once all n_{bc} candidate solutions are obtained, objective functions F_{p1} and F_{p2} are calculated for each of them. Estimation of objective F_{p1} in this case leverages the computational efficiency of the metamodel in performing vectorized predictions: the calculations are simultaneously performed for all n_{bc} candidate solutions, or using subsets with a lower number of members depending on the available computational resources (memory can be a problem for vectorized operations depending on n and n_{bc} ²⁸). The dominant solutions representing the Pareto front can be then readily identified by comparing the values for the two objectives over all candidate solutions. The only challenge is that the value of n_{bc} needs to be large in order to obtain an adequate representation of the Pareto front. Proper selection for

n_{bs} is examined in the illustrative example. Once the Pareto front has been identified, a solution (among the Pareto set) can be adopted using any desired criterion. This will be further discussed in the illustrative implementation.

5 | ILLUSTRATIVE IMPLEMENTATION

The illustrative implementation considers the stochastic ground motion model developed by Rezaeian and Der Kiureghian¹⁶ with model parameter vector including the arias intensity I_a , the significant duration D_{5-95} , the time corresponding to 50% of the intensity t_{mid} and the associated spectral frequency ω_{mid} , the rate of change for that frequency $\dot{\omega}$, and the damping ratio ζ_f for the excitation spectrum. The arias intensity simply scales the ground motion, so it corresponds to parameter θ_s in representation $Y_i^g(\boldsymbol{\theta}, \mathbf{w}) = \sqrt{\theta_s} s_i(\mathbf{x}, \mathbf{w})$, with the 5 remaining parameters corresponding to \mathbf{x} . For the target hazard, GMPEs used in the Western US are selected here,³⁷⁻⁴⁰ whereas the suggestions in Kaklamanos et al⁴¹ are adopted to estimate unknown inputs for some of the GMPEs. As target, IM predictions of the logarithmic mean and variance from individual GMPEs as well as the average of their predictions are used. Note that the latter provides single target hazard for each structural period examined, simply that hazard is obtained by averaging information from multiple GMPEs. All computations are performed in a quad-core 3.0-GHz Xeon processor with 16 Gb of RAM. Fault and site characteristics are taken, respectively, as strike-slip fault and shear wave velocity in upper 30 m of soil $V_{s,30} = 800$ m/s. For moment magnitude M and rupture distance R , different values will be examined. As IM for the seismic hazard description, the peak pseudo-acceleration (PSA) (S_a) for a single degree of freedom system with 5% damping ratio is utilized. Different ranges of structural periods will be examined for S_a and, unless otherwise specified, the objective function F_{p1} is estimated as the weighted average of the entropies for each scalar IM as in Equation 7, with the weights chosen as $\gamma_i = 1$, so that no specific structural period is prioritized.

5.1 | Details for metamodel development

For the characterization of domain X and the selection of the support points for the metamodel development, the same approach as in Tsioulou et al²⁸ is adopted, the only difference being that a larger domain is considered here with a relative increase of 70% compared with Tsioulou et al.²⁸ The reason for the latter is to support higher accuracy in the MCS implementation: in Tsioulou et al²⁸ evaluation only at the new (modified) predictive relationships was required so proximity to the initial predictive relationships needed to account only for that modification, whereas here the MCS estimation will need to utilize the metamodel predictions for parameters even further away from the modified (and therefore from the initial) mean predictive relationships.

Logarithmic mean and variance for S_a for different periods, the ones used by the aforementioned GMPEs, is adopted as the response output for the metamodel development whereas the white noise samples are chosen as $n_w = 100$. Three different accuracy criteria with associated coefficient of determination 0.92, 0.94, and 0.96 are selected for the adaptive metamodel development.²⁸ This leads to metamodels with 1500, 3000, and 4500 number of support points, respectively. Note that the accuracy of the established metamodels is much higher for prediction of the logarithmic mean rather than the logarithmic variance. For example, for the metamodel with 4500 support points the average coefficient of determination is 0.99 for the logarithmic mean and 0.94 for the logarithmic variance. This trend agrees with past studies⁴² that have examined the use of metamodels in approximating seismic hazard when stochastic ground motion models are utilized for the description of the latter. This should be also viewed as a positive attribute of the metamodel approximation for the purposes of this study: the logarithmic mean has a higher importance towards the hazard description for the typical earthquake engineering applications that have logarithmic variance smaller than 1 (the scaling of difference of logarithmic means by $1/\sigma_i^2(\mathbf{z}) > 1$ in Equation 8 when estimating discrepancy from target also reveals this), which means that a higher degree of confidence exists for properly approximating this hazard than the average coefficient of determination (averaged over both logarithmic mean and variance) indicates, because the accuracy for the more important component (logarithmic mean) is higher.

Estimation of metamodel response for 10 000 samples requires 3.6, 7.5, and 10.7 seconds for the metamodels with 1500, 3000, and 4500 support points, respectively. Note that adoption of larger value of samples prohibits efficient vectorization of operations for the $n = 4500$ points due to memory restrictions. Calculation of objective function F_{p1} for $N_s = 100$ MCS samples requires 0.005, 0.10, and 0.16 seconds for the metamodels with 1500, 3000, and 4500 support points, respectively. It should be stressed that, as advocated earlier, evaluation of the metamodel across multiple candidate solutions is simultaneously performed to better leverage the computational efficiency allotted by the vectorized

metamodel evaluation. For example, for $N_s = 100$, 100 different candidate solutions are examined; this leads to a total of $100 \times 100 = 10\,000$ samples for the metamodel evaluation, which avoids memory problems even for the metamodel with 4500 support points.

5.2 | Validation of lognormal distribution assumption

First the validation of the lognormal assumption for the response output Y_i^g distribution is examined. This is performed separately for the response output from the exact stochastic ground motion model as well as for the metamodel approximation. Distribution $p_g(\ln(Y_i) | \boldsymbol{\mu}, \boldsymbol{\Sigma})$ is approximated through KDE (Equation A1) as discussed in Appendix A using $n_s = 1000$ samples of Y_i^g . Comparison is primarily expressed through objective F_{p1} because this is the critical quantity utilized in the proposed framework. The target hazard corresponds to the average of the aforementioned four GMPEs. The predictive model proposed in Rezaeian and Der Kiureghian,¹⁶ referenced herein as unmodified model, is used for defining $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, but comparison extends over a wide range of seismicity scenarios, M in range [6–8] and R in range [10–100] km. The latter guarantees that validation extends over the wider range that the optimization is going to examine with respect to the predictive relationships. Figure 2 shows F_{p1} evaluated either through the lognormal assumption or through the KDE approximation for three periods $T_s = [0.01\ 0.5\ 2]$ seconds (columns of the figure). In this instance F_{p1} is evaluated for each period separately to assess impact over specific structural characteristics, rather than averaged over the different IMs. Top row corresponds to comparison with respect to the actual stochastic ground motion model and bottom row to comparison with respect to the metamodel approximation. The comparisons in this figure show exceptionally close agreement for objective F_{p1} between the lognormal assumption and the KDE approximation (compare the two curves in each subplot) over the entire seismicity range and for all structural periods. This agreement holds for both the actual model and the metamodel. The results in the figure allow for some additional observations. First there is a very good agreement between the metamodel and the actual model (compare the results across the rows). This provides a preliminary validation, examined further in the following sections, of the proposed, metamodel-based framework. Secondly, for certain seismicity ranges, the unmodified model has large discrepancy (large F_{p1} values) from the target hazard. This validates the claim that motivated this study, that existing approaches for selecting predictive relationships do not necessarily provide a close match to the desired hazard for some structures or seismicity scenarios. For

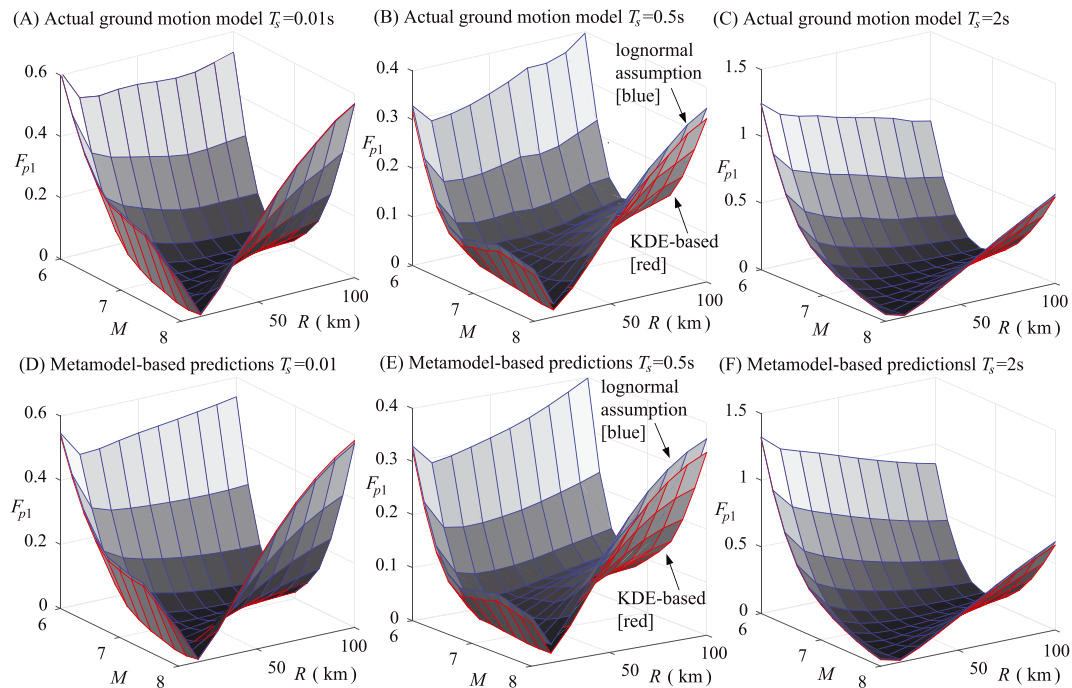


FIGURE 2 Comparison for F_{p1} evaluated either through lognormal assumption for Y_i^g or KDE approximation utilizing samples for actual ground motion model (top row) and metamodel predictions (bottom row). Columns correspond to three different Y_i , representing, respectively, PSA for 0.01, 0.5, or 2 seconds (s) [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com)]

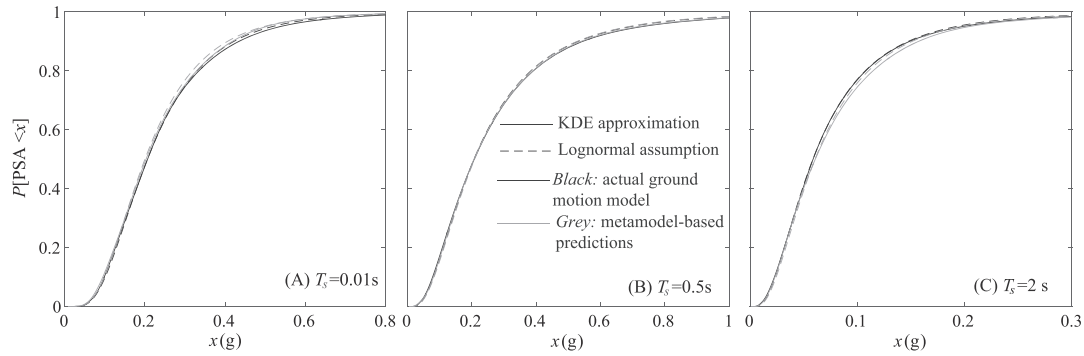


FIGURE 3 Comparison for the CDF $Y_i^g = \text{PSA} (P[\text{PSA} < x])$ based on lognormal assumption or KDE approximation utilizing samples for actual ground motion model and metamodel predictions. Cases correspond to PSA for (A), 0.01, (B), 0.5, or (C), 2 seconds (s)

assessing what constitutes large value of F_{p1} based on Equation 8, values close to 0.005 should be considered as small, values close to 0.02 as moderate, and values over 0.05 as large.

The compatibility with respect to the lognormal assumption is further examined in Figure 3, which shows the cumulative distribution functions (CDF) for Y_i^g for a particular seismicity scenario, corresponding to $M = 6$ and $R = 20$ km. Similarly to Figure 2, each column corresponds to a different period T_s . The results in Figure 3 show, again, a very good match not just between the CDFs using the KDE approximation or the lognormal assumption but also between the CDFs using the actual model and the metamodel. Overall, the comparisons in Figures 2 and 3 justify the use of a lognormal assumption for the distribution of Y_i^g . Additionally, they provide a first validation for the accuracy of the metamodel approximation, when compared with the actual ground motion model, with respect to both the entropy (Figure 2) but also the exact distribution of Y_i^g (Figure 3).

5.3 | Optimization details and metamodel accuracy

This section examines details of the numerical solution of the optimization problem. For the MCS of Equations 19 to 21, Latin hypercube sampling was adopted, and for the sample size N_s different values in range [20 150] were examined. The selection of the exact value of N_s is a compromise between numerical efficiency (it proportionally impacts the computational burden) and robustness of the multi-objective optimization (identification of correct Pareto front). After sensitivity analysis, a value of $N_s = 70$ was chosen. The coefficient of variation for F_{p1} using MCS estimates of Equations 19 to 21 with $N_s = 70$ was found to be in range 3% to 4% which should be deemed sufficient considering the fact that CRN are further utilized³⁵ to reduce the importance of the associated estimation error in the comparisons established across the optimization.

The discussion moves next to the impact of the number of points used in the random search n_{bc} . The considerations are similar to the selection of N_s (efficiency versus robustness) although in this case the selection is not as straightforward because there are no simple statistics like the coefficient of variation to compare. The sensitivity analysis is performed instead by solving the optimization problem for different values of n_{bc} . The target used in this section corresponds to structural periods $T_s = [0.4 \ 0.5 \ 0.75 \ 1.0 \ 1.5 \ 2.0]$ seconds and hazard described by the average of the considered GMPEs. Three different seismicity scenarios $M = 6-R = 20$ km, $M = 7-R = 30$ km, and $M = 8-R = 50$ km are considered. The first scenario represents a case that the unmodified model provides a poor match to the target hazard, the second one achieves a good match, and the last case lies in between the other two. Figure 4 presents the Pareto fronts identified through random search for different number of points n_{bc} . The metamodel with 4500 support points is used in this case. Only a few representative solutions and not the entire front are shown for clarity. Results are reported in this figure and in the remaining of the manuscript with respect to $\sqrt{F_{p1}}$ and $\sqrt{F_{p2}}$ to facilitate an easier comparison (differences of extreme values easier to discern). Figure 4 shows that the differences between the identified fronts for different n_{bc} values are only minor, and they occur primarily for small $\sqrt{F_{p1}}$ - large $\sqrt{F_{p2}}$ combinations. Note that the objective function $\sqrt{F_{p1}}$ values for the $M = 7-R = 30$ km scenario are very low (due to an already good match of the unmodified model to the target hazard); therefore, any identified differences between the Pareto fronts for different number of points n_{bc} are of smaller relevance. This comparison shows that a value of n_{bc} around 100 000 to 250 000

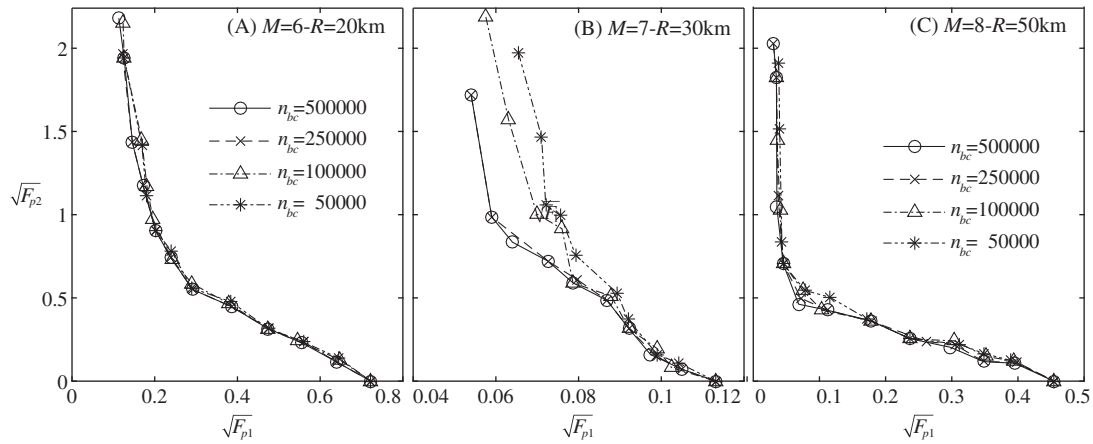


FIGURE 4 Pareto fronts identified by random search for different number of candidate solutions n_{bc} for different seismicity scenarios: (A), $M = 6-R = 20$ km, (B), $M = 7-R = 30$ km, and (C), $M = 8-R = 50$ km

should be considered as sufficient for identifying an adequate representation of the Pareto front. An n_{bc} value equal to 150 000 is adopted for the random search results presented in the remaining of the manuscript.

The impact of the metamodel accuracy is examined next. This is established by considering additionally the results obtained by using the exact stochastic ground motion model, which represents the measure for evaluating the actual hazard compatibility of the identified ground motion model. Comparison is performed across different Pareto optimal solutions. The Pareto fronts identified by using the metamodels with the three different number of support points are presented in Figure 5 for seismicity scenario $M = 6-R = 20$ km and $M = 8-R = 50$ km. Note that for seismicity scenario $M = 7-R = 30$ km which was also presented in Figure 4, the results are of limited interest because the unmodified ground motion model provides a good compatibility to the target hazard. This is the reason that this seismicity scenario is not presented here. In all cases, the random search is implemented with the same candidate solutions, to facilitate a consistency in the corresponding comparisons. Then, Figure 6 presents spectral plot comparisons for the solution (among the Pareto set identified in each case) corresponding to the minimum of F_{p1} for the seismicity scenario $M = 6-R = 20$ km. Similar trends hold for the $M = 8-R = 50$ km seismicity scenario (not reported here due to space constraints). In all plots, the predictions using the metamodel and the actual ground motion model are reported. Figure 6 offers comparisons in context of both average response (top row) using curves corresponding to median and 14th to 86th percentiles (denoted as median $\pm \sigma_{log}$ herein), and logarithmic standard deviation (bottom row). The former assesses the hazard compatibility with respect to different IM statistics and the latter explicitly with respect to the IM dispersion.

The results show that good agreement is established between the metamodel predictions and the actual model predictions along the Pareto front for all metamodel cases: 1500, 3000, and 4500 support points. This is evident in both the Pareto fronts (Figure 5) and the corresponding spectral plots (Figure 6), the latter indicating a good match in terms of both the mean and variability of the hazard. For the lower accuracy metamodel utilizing only 1500 support points

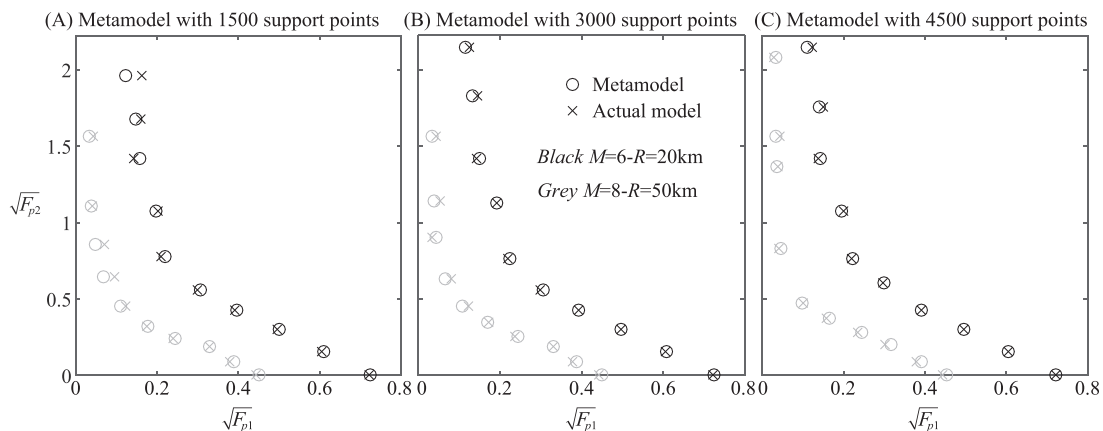


FIGURE 5 Pareto fronts identified using metamodels with (A), 1500, (B), 3000, or (C), 4500 support points and comparison to predictions by exact stochastic ground motion model. Two seismicity scenarios examined $M = 6-R = 20$ km [black] and $M = 8-R = 50$ km [gray]

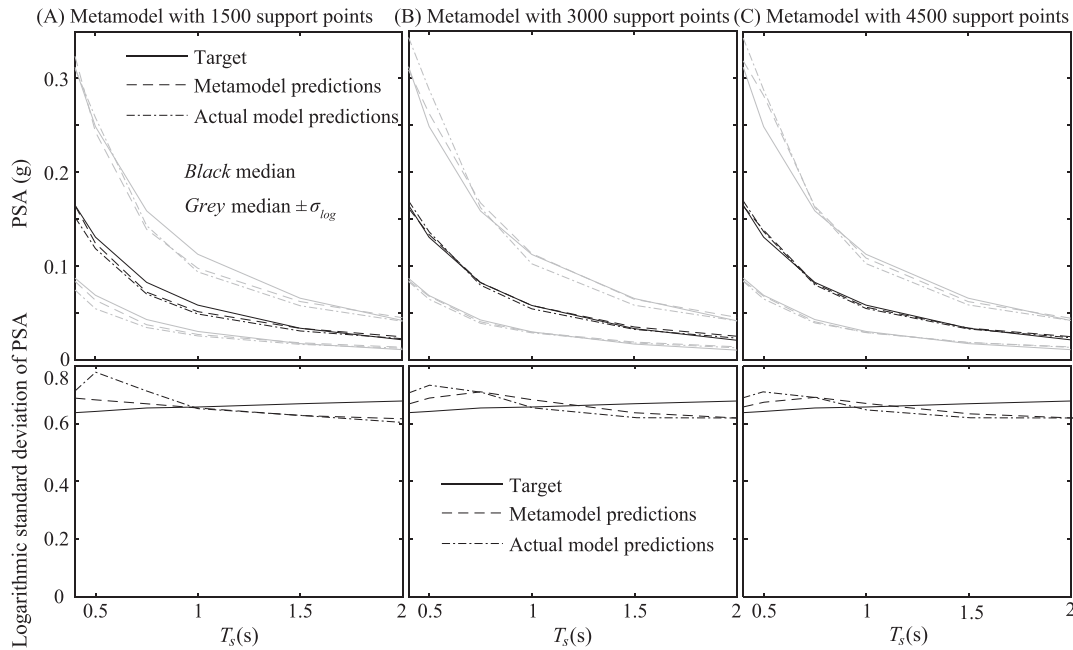


FIGURE 6 Spectral plots for the solutions corresponding to minimum of F_{p1} in the Pareto fronts identified in Figure 5 for the $M = 6$, $R = 20$ km seismicity scenario. Top row shows curves corresponding to median and median $\pm \sigma_{\log}$ for the response. Bottom row shows logarithmic standard deviation (σ_{\log})

greater differences exist, especially for smaller F_{p1} values, but the discrepancies even for it are overall quite small, significantly smaller than the ones reported in Tsioulou et al²⁸ where this metamodel was shown to lead to erroneous results, identifying suboptimal solutions when the metamodel error was not explicitly considered in the optimization. This should be attributed to the fact, discussed also in Section 4.2, that objective F_{p1} represents a statistical quantity over θ , and averaging over different θ values for evaluating F_{p1} reduces the potential influence of larger errors for specific θ values. This discussion and the good agreement reported in Figures 5 and 6 also further validate the choice to avoid the explicit consideration of the metamodel prediction error in the problem formulation (see discussion in Section 4.2). This consideration would increase the computational burden while providing negligible benefits in terms of the quality of the identified Pareto fronts. Overall, all examined here metamodels provide adequate accuracy in the identification of the Pareto front, with no need to explicitly consider the metamodel prediction error, with preference towards the metamodel with 3000 or 4500 support points. The latter will be utilized in all remaining comparisons in this manuscript.

Some final remarks are warranted with respect to the overall computational cost. The primary computational burden of optimization of Equation 4 stems from the $N_s \cdot n_{bc}$ evaluations of the metamodel required for the MCS of Equations 19 to 21 across the search for the Pareto front. For the scenario advocated earlier with $N_s = 70$, $n_{bc} = 150\,000$ and use of metamodel with 4500 support points the time is 180 minutes per seismicity scenario. If the metamodel with 3000 support points is used instead the required time reduces to 120 minutes. Compared with the study Tsioulou et al,²⁸ this represents an important increase of the computational cost: there is ultimately a N_s -fold increase of this cost for same n_{bc} value, stemming from the MCS step.

5.4 | Implementation for different seismicity scenarios

With the computational details ironed out, the discussion moves next to the hazard compatibility established by the proposed modification of the ground motion model. Figure 7 shows results for 6 seismicity scenarios targeting seismic hazard given by the average of the aforementioned GMPEs for two different ranges of T_s : $T_s = [0.4\ 0.5\ 0.75\ 1.0\ 1.5\ 2.0]$ seconds and $T_s = [0.4\ 0.5\ 0.75]$ seconds. These two different cases are referenced herein as long and short, respectively, period ranges. The proposed framework identifies in each case a Pareto front that clearly demonstrates the compromise between the two objectives, with different characteristics in each case, depending on how close the unmodified ground motion model was to the target hazard. Choosing a shorter period range for the target IMs facilitates an overall better

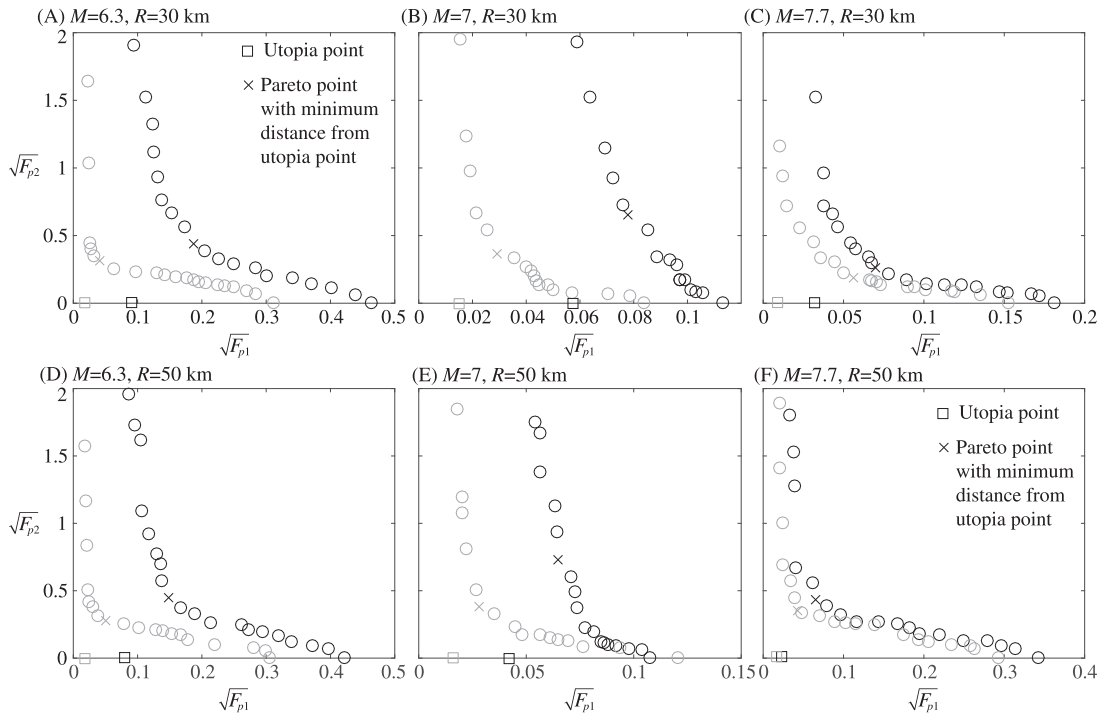


FIGURE 7 Pareto fronts for different seismicity scenarios considering match to long (black) or short (gray) period range IMs for defining the seismic hazard. In each plot, the Pareto point with minimum distance from utopia point is shown with x and the utopia point with a □

match (smaller F_{p1} values); this is anticipated because objective F_{p1} imposes less strict requirements (fewer number of components to match) in this case.

To select a solution within the identified Pareto set, one of the most common approaches within the multi-objective optimization literature⁴³ is to choose the one that has the smallest normalized distance from the utopia point, defined as the point in the Pareto front that corresponds to the idealized (unachievable) minimum of the two objectives across the front. Following the guidelines in Tsioulou et al²⁸ for choosing the normalization, the following distance metric is chosen

$$D_p(\theta) = \sqrt{\sum_{i=1,2} \left(\frac{\sqrt{F_{pi}(\theta|z)} - \min \sqrt{F_{pi}}}{\max \sqrt{F_{pi}} - \min \sqrt{F_{pi}}} \right)^2}; \max \sqrt{F_{pi}} = \max_{\{\theta_p; i=1, \dots, n_p\}} (\sqrt{F_{pi}}), \min \sqrt{F_{pi}} = \min_{\{\theta_p; i=1, \dots, n_p\}} (\sqrt{F_{pi}}). \quad (22)$$

The corresponding point is identified in all cases in Figure 7. Another option would have been to choose the solution that satisfies a pre-determined threshold for the match to the targeted IMs.

These selections are finally demonstrated for a wide range of seismicity scenarios, M in range [6 8] and R in range [10 100] km, in Figures 8 to 13. For each scenario, three different Pareto points are selected, the one with smallest distance $D_p(\theta)$ from the utopia point and the ones with objective $\sqrt{F_{p1}}$ smaller than 0.15 or 0.075. These three cases are denoted, U_t , C_l , and C_s , respectively. The thresholds for C_l and C_s modifications were chosen so that to reflect medium and small, respectively, incompatibility to the target hazard. In addition, results for the unmodified model are presented, denoted U_n . Figure 8 shows plots for (i—first row) $\sqrt{F_{p1}}$ for U_t and U_n ($\sqrt{F_{p1}}$ is constrained for the other two cases) and for (ii—second and third rows) $\sqrt{F_{p2}}$ for U_t , C_l , and C_s ($\sqrt{F_{p2}}$ is zero for U_n). To better demonstrate the differences, results are presented separately for U_t (second row) and for the pair C_l and C_s (third row) in the latter case. The three different columns in the figure correspond to three different implementation cases: target hazard given by the average of the aforementioned four GMPEs for both the (A) long and (B) short period ranges as well as (C) target hazard given only by GMPE³⁸ for the long period range. These scenarios are denoted herein as SC_1 , SC_2 , and SC_3 , respectively. Figures 9 to 11 show spectral plots for a selection of seismicity scenarios, defined by combinations of M [6.2, 6.8, 7.4, 8] and R [30, 60, 90] km, for SC_1 . For each of the 12 M - R combinations spectral curves corresponding to the target hazard, the unmodified model and the predictions by the three aforementioned model modifications are shown to facilitate comparisons. More specifically, Figure 9 shows curves corresponding to different statistics of the response

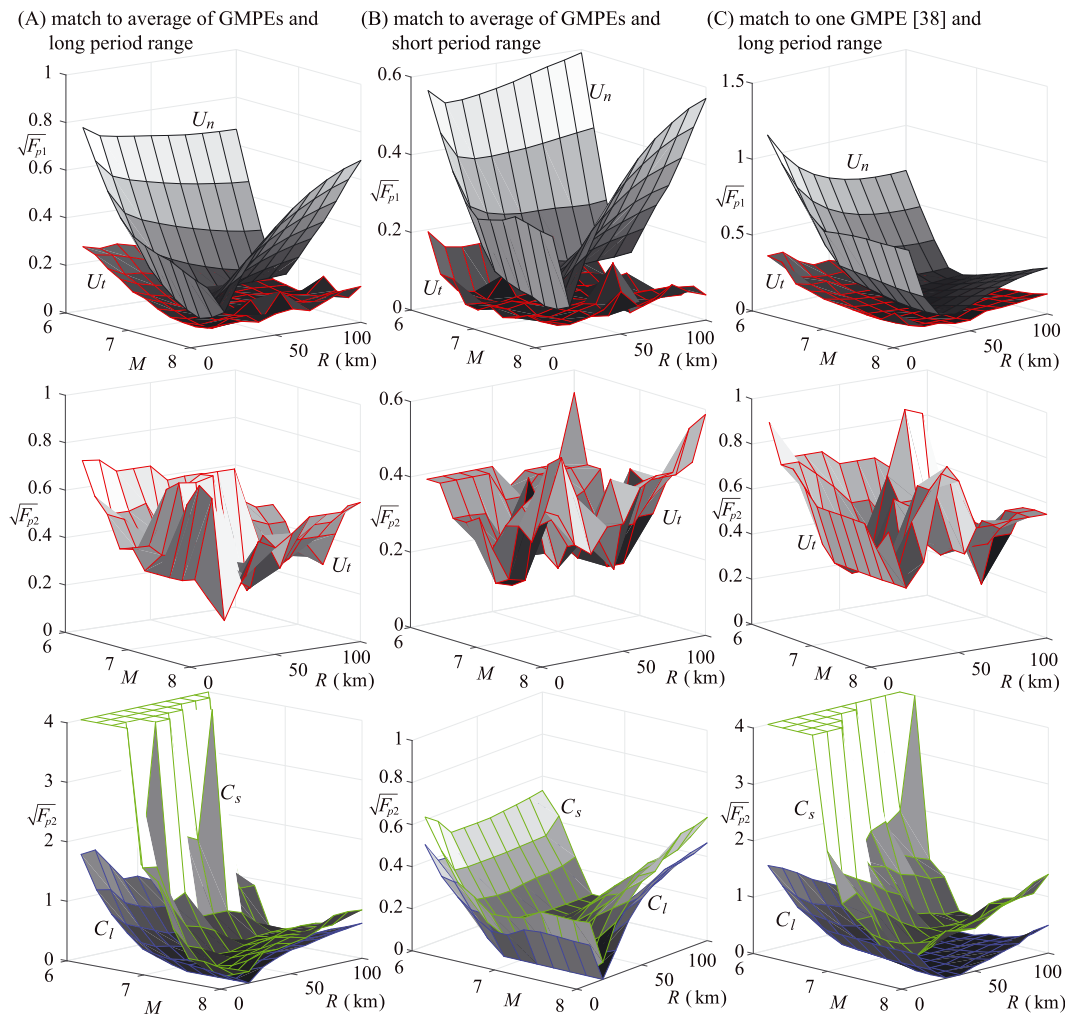


FIGURE 8 Results for $\sqrt{F_{p1}}$ and $\sqrt{F_{p2}}$ for unmodified ground motion model (U_n) and modified ground motion model corresponding to three different selections criteria: Minimum distance from utopia point (U_i) and value $\sqrt{F_{p1}}$ smaller than 0.15 (C_i) or 0.075 (C_s). Implementations in the different columns correspond to (A) long and (B) short period ranges for matching to the average considered GMPEs and (C) long period ranges for match to GMPE³⁸ [Colour figure can be viewed at wileyonlinelibrary.com]

(median and median $\pm \sigma_{log}$) for the unmodified model and the model corresponding to the Pareto point with smallest distance from the utopia point (U_i case). Figure 10 presents the same curves for the models corresponding to the Pareto points with average relative entropy thresholds C_i and C_s . Figure 11 presents same comparison directly in terms of logarithmic standard deviation (ie, IM dispersion). In all these figures, the curves corresponding to the target are also shown.

Finally, Figures 12 and 13 show the predictive relationships for all examined seismicity scenarios for the modified model corresponding to the Pareto point with smallest distance from the utopia point (U_i case) for SC₁. In particular, Figure 12 shows the mean model parameters θ of the modified predictive model (corresponding to μ for the aforementioned Pareto point). The unmodified ground motion model parameters (corresponding to μ_r) as well as another case that will be discussed in the next section are also included. Then, Figure 13 shows the ratio of standard deviation for the modified and unmodified model (comparison of square root of the diagonal points of Σ for the aforementioned Pareto point and Σ_r). Note that some of the curves shown in these figures have non-smooth characteristics. As discussed in Tsioulou et al,²⁸ this should be attributed to the fact that a discrete representation of the Pareto front is obtained and to the random search characteristics of the adopted optimization algorithm for identifying the Pareto front.

The baseline trends observed in the figures are similar to the ones in Tsioulou et al.²⁸ These trends are enhanced in this study with additional considerations with respect to the variability associated with the seismic hazard. The unmodified model does not provide a good match to the target hazard for the entire seismicity range as observed in first row of Figure 8 and also in the spectral plots in Figures 9 and 11. This is particularly true for the logarithmic mean, ie, median (Figure 9), with logarithmic variance, ie, dispersion (Figure 11), showing better compatibility to start with. The proposed

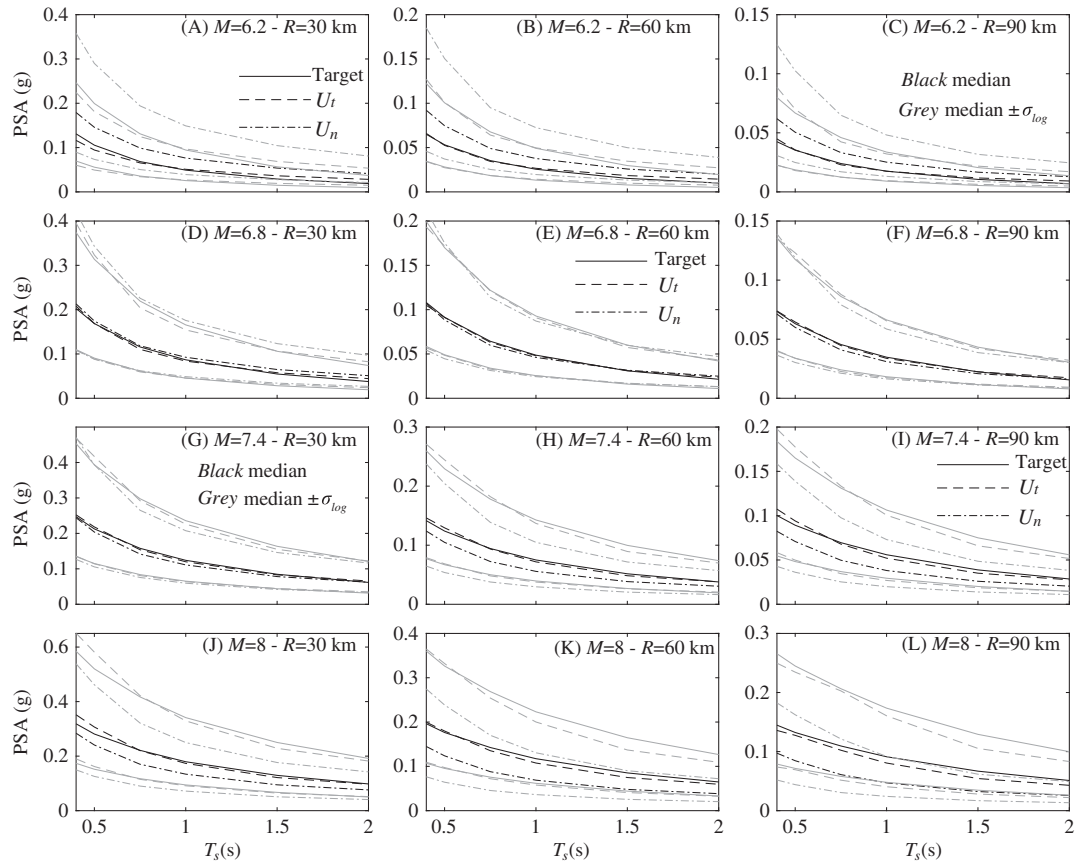


FIGURE 9 Spectral plots of average response for seismicity scenarios (different subplots) corresponding to combinations of M [6.2, 6.8, 7.4, 8] and R [30, 60, 90] km, for the target hazard (target), the unmodified ground motion model (U_n), and the modified ground motion model corresponding to the minimum distance from utopia point (U_t). Implementation scenario shown corresponds to matching to the average considered GMPEs and long period range. Curves corresponding to median and median $\pm \sigma_{log}$ shown

modification (cases U_t , C_l , and C_s) significantly improve the match to the hazard (Figure 8), establishing a balance between F_{p1} and F_{p2} , with the characteristics of the balance depending on the criteria for selection of the final model among the Pareto optimal solutions. Similar to Tsioulou et al²⁸ when the unmodified model has larger discrepancies from the target hazard, then the modifications lead to larger values for F_{p2} , but still successfully identify models, independent of the implementation case, that provide an improved match to the IM target mean and variance. This is clearly observed for C_l and C_s cases; C_s imposes a smaller discrepancy between the modified model and the target hazard and the modification leads to identification of a model with bigger differences from the original one (larger values for F_{p2}). The U_t modification identifies a model with moderate discrepancy from the unmodified one, corresponding to values of $\sqrt{F_{p2}}$ in the range of 0.2 to 0.8, whereas the two other modification approaches, C_l and C_s , identify models with greater variability across the different seismicity scenarios. For scenarios in the range of $M = 7-7.5$, the unmodified model provides a good match to the target hazard, and therefore modification of it provides limited benefits. This is perhaps better captured by the C_l case, which corresponds to low $\sqrt{F_{p2}}$ values for this seismicity range, significantly lower than the U_t modification. This indicates that a satisfactory match to the target hazard, ie, a match satisfying the predefined compatibility threshold, is established with no need to greatly modify the initial predictive models. Thus, selection of the Pareto optimal model based on a targeted accuracy to the GMPEs, ie, value for F_{p1} below a certain threshold as in the C_s and C_l cases, provides a more rational selection for the final model as it allows a more direct recognition of the seismicity ranges where modification is not truly required. Selection of a small threshold (C_s case), however, results in identification of a model with large discrepancies from the unmodified model (large F_{p2} values). This model will typically be far away from the U_t case and belong in a steep part of the Pareto front (check Figure 7 earlier), meaning that small improvements in F_{p1} come at a large increase of F_{p2} . A multi-level selection criterion similar to the one proposed in Tsioulou et al²⁸ may be therefore advocated: select the Pareto optimal solution that satisfies a certain accuracy threshold for F_{p1} unless this solution leads to a greater F_{p2} value than the Pareto optimal solution with minimum distance

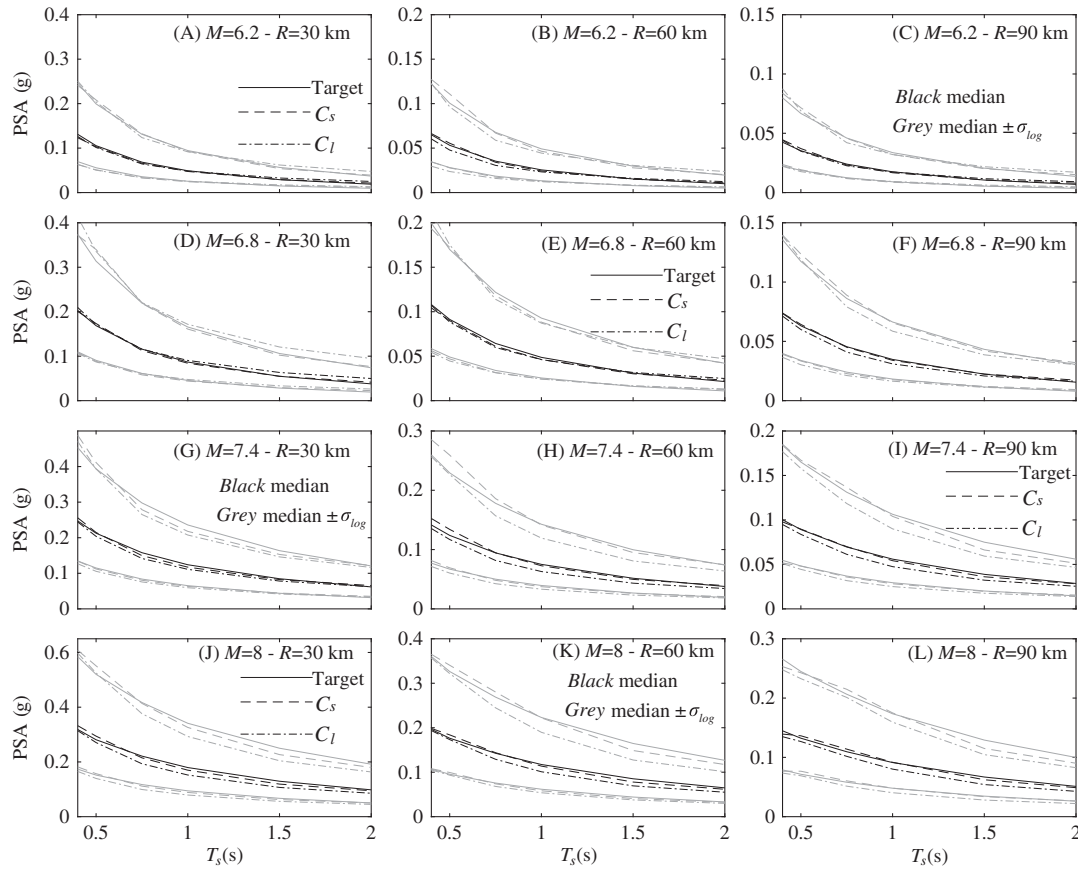


FIGURE 10 Spectral plots of average response for same seismicity scenarios as in Figure 9 (different subplots) for the target hazard (target), and the modified ground motion model corresponding to value $\sqrt{F_{p1}}$ smaller than 0.15 (C_l) or 0.075 (C_s). Implementation scenario shown corresponds to matching to the average considered GMPEs and long period range. Curves corresponding to median and median $\pm \sigma_{log}$ shown

from the utopia point. If the latter happens, then select the Pareto optimal solution with minimum distance from the utopia point.

From the comparison of the different implementation cases shown in Figure 8, it is evident that selection of a shorter period range [compare cases (B) and (A)] establishes an easier match to the target, as it results in smaller overall values for F_{p1} and F_{p2} . The selection of a different target [compare cases (A) and (C)], although not imposing any additional constraint in the optimization implementation, leads to different results. Overall, the discrepancy between the unmodified model and the target hazard follows the same trends in all implementation cases. This demonstrates the versatility of the proposed framework as the modification facilitates an enhanced hazard compatibility for all IMs and hazard targets considered.

The spectral plots in Figure 9 to 11 provide the decomposition of the overall F_{p1} match to the different structural periods and the statistics of the IM distribution (median and dispersion). The comparison between the unmodified and modified cases shows that the match to both the target median and dispersion are separately addressed within the proposed optimization. This is also supported by the trends in Figures 12 and 13; not only the mean (Figure 12) but also the standard deviation (Figure 13) of the predictive model is adjusted. Overall for the ground motion model examined here, the match to the target dispersion offered by the unmodified model, and subsequently by the modified one, is quite good (compare relative discrepancies for U_n in Figures 8 and 11). Although the proposed modifications also impact this dispersion (Figure 11), the impact on the median (Figures 9 and 10) is more substantial. This should be attributed to ability to influence value of objective F_{p1} more by adjustments in $\ln(\bar{Y}_i^g(\boldsymbol{\mu}, \boldsymbol{\Sigma}))$ rather than $\sigma_i^g(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ [formulation of Equation 8 and scaling of differences from the target for the former by $\sigma_i^2(\mathbf{z})$ also reveals that], something automatically leveraged by the optimization. With respect to the different modifications, the C_s always provides a better match compared with C_l one, although the spectral curves of the two modifications are very close to each other especially for lower magnitude ranges $M = 6.2-6.8$. The U_l modification also provides a good

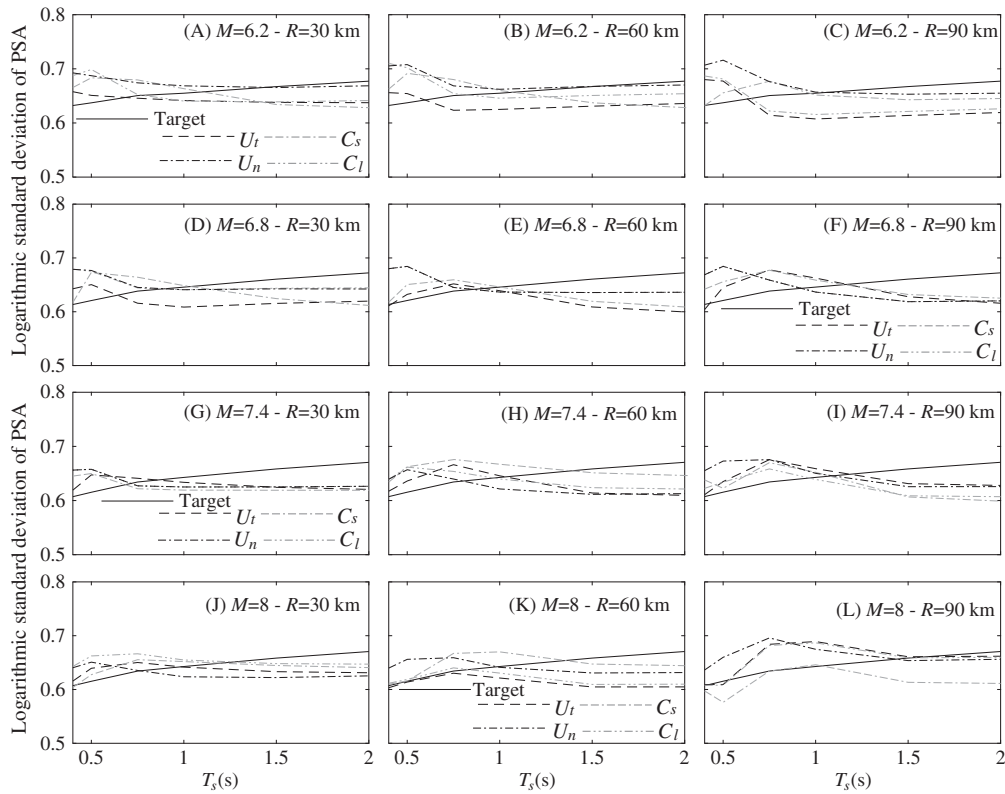


FIGURE 11 Spectral plots for logarithmic standard deviation of response for the same cases examined in Figures 9 and 10

match to the target, frequently very close to the C_l and C_s cases, depending on the original match of the unmodified model (ie, how much is the modification truly needed). This observation further supports the multi-level selection criterion discussed earlier.

The results in Figures 12 and 13 show that the model modification leads to similar characteristics as observed for the unmodified model. This guarantees that the proposed ground motion model modification does not deviate significantly from observed regional trends and was achieved by incorporating this deviation as an objective in the problem formulation (objective F_{p2}). Parameters I_a , ζ_f , and ω show bigger variability compared with their initial mean values (Figure 12). This should be attributed to a greater sensitivity with respect to them of the resultant ground motions and agrees with the trends reported in Tsiouou et al.²⁸ With respect to the adjustment of the variability of the predictive model, the ratio of standard deviations remains close to 1 (Figure 13) with values in the range of 0.8 to 1.05, indicating small (but not negligible) overall adjustment.

The overall discussion shows the importance of the established framework: once the initial metamodel is developed, it can support the efficient identification of ground motion models that (1) match conditional hazard for any desired IMs and chosen period range while (2) maintaining a small deviation from the initial predictive models. This can be seamlessly repeated for any seismicity scenario. The final ground motion model modification can be chosen based on the criteria discussed earlier.

5.5 | Comparison to modification of mean value characteristics only

As discussed in Section 5.3, the computational cost of the proposed modification to match the probabilistic seismic hazard is significantly higher than previous efforts to match only the IM predictions corresponding to the mean predictive relationships for the ground motion model,²⁸ ie, completely ignoring the variability stemming from Σ in the predictive model for θ (using $\Sigma = 0$). This increase in computational burden stems ultimately from the need to estimate the statistics of the response when the variability of the predictive model $p(\theta|\mu,\Sigma)$ is considered. It is therefore of interest to examine whether the computationally less demanding problem of matching only the IM predictions corresponding to the mean of the predictive relationships²⁸ can be adopted as a surrogate for the problem of interest here. This is

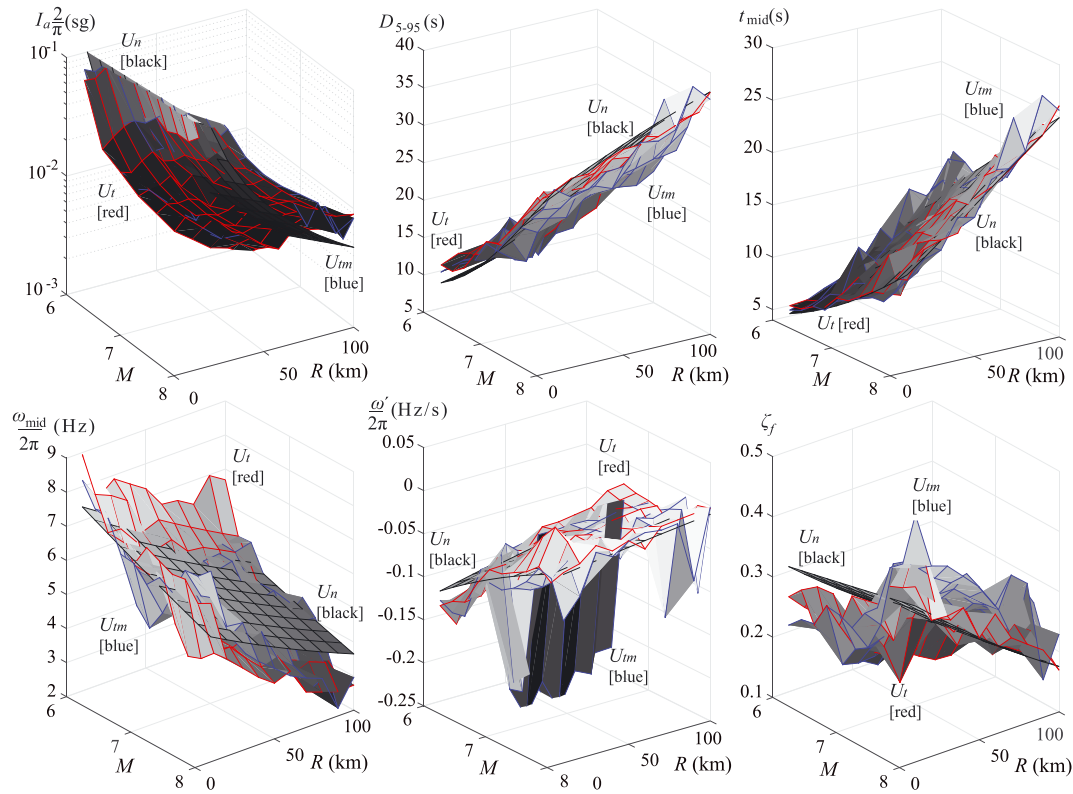


FIGURE 12 Mean for physical ground motion model parameters θ corresponding to unmodified ground motion model (U_n), modified ground motion model with minimum distance from utopia point for matching the complete probabilistic hazard (U_i), or the hazard corresponding to mean predictive relationships (U_{im}). Implementation scenario corresponds to matching to the average considered GMPEs and long period range [Colour figure can be viewed at wileyonlinelibrary.com]

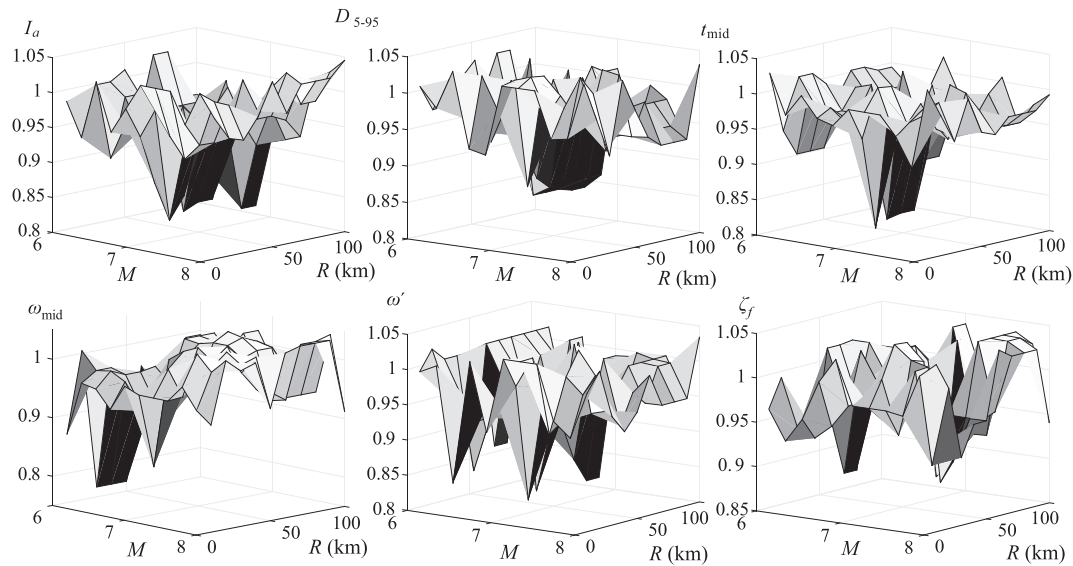


FIGURE 13 Ratio of standard deviations for the ground motion model parameters between modified and unmodified ground motion model (U_n and U_i cases in Figures 12)

equivalent to assuming $\Sigma = 0$ in the optimization of Equation 4 and leads to modification of the mean only predictive relationships, while greatly reduces computational cost of the numerical optimization as it entails no MCS step (because $\Sigma = 0$). This facilitates a N_s fold reduction of the computational burden as discussed earlier. The resultant modification of the predictive relationships will be denoted μ_m herein.

The quality of the solution obtained from this approximate problem may be then assessed by evaluating objectives F_{p1} and F_{p2} assuming distribution $p(\theta|\mu_m, \Sigma_r)$, ie, adopting the initial variability of the predictive models Σ_r , or $p(\theta|\mu_m, \mathbf{0})$, ie, ignoring any variability in the predictive models and utilizing only the variability stemming from the white noise to calculate response statistics. The comparison can be performed with respect to the entire Pareto front for limited number of seismicity scenarios or with respect to a specific solution along the front over a wider range of scenarios. Due to space constraints, the latter is only reported here, with specific solution chosen as the point with minimum distance from the Utopia point. The solution available from Tsioulou et al²⁸ is directly utilized for μ_m . The case corresponding to $p(\theta|\mu_m, \Sigma_r)$ is denoted as U_{tm} and case corresponding to $p(\theta|\mu_m, \mathbf{0})$ as U_{tmn} . Results are shown in Figures 14 and 15, following same guidelines as the study reported in Figures 8 and 9. Comparison with respect to the dispersion only (similar to Figure 10) is not reported due to space constraints. Figure 14 compares U_n , U_t , and U_{tm} across both objectives over a range of seismicity scenarios (ie, adds U_{tm} curve in the results reported in Figure 8), and Figure 15 presents spectral plots for the target seismic hazard, U_{tm} and U_{tmn} . Latter figure should be compared directly to Figure 9 to evaluate the relative advantages of U_t (adding that curve in this plot is avoided to improve clarity of the presentation). Solution for the physical parameters corresponding to μ_m has been also reported earlier in Figure 12. Note that this solution is same for both U_{tm} and U_{tmn} .

With respect to U_{tmn} first (Figure 15), the results show the importance of considering the variability of the predictive relationships. Reliance only on the variability stemming from the white noise provides significantly lower variability for the seismic hazard than the variability prescribed by GMPEs (compare median and median $\pm \sigma_{log}$ curves for target and U_{tm} cases). This ultimately leads to large values for F_{p1} for U_{tmn} which is the reason that results for it are not reported in Figure 14. Comparison now between U_t and U_{tm} in Figure 14 shows that the explicit optimization for the probabilistic hazard provides for some scenarios a noticeably better match (smaller F_{p1} values for U_t) for the same level of modification of the initial probability model (similar F_{p2} values for U_t and U_{tm}). U_{tm} modifications even underperform the unmodified model U_n for some scenarios, corresponding to cases for which U_n provides an adequate match to the target hazard to start with. Same trends are observed in the spectral plots in Figure 15. Overall U_{tm} is shown to provide an improvement over U_n when the initial match to the hazard is not adequate, although it might underperform U_t due to its inability to explicitly accommodate the hazard variability. It should be also pointed out that based on the results in Figure 12, even the mean value vectors for the predictive relationships identified by different modification approaches (U_t and U_{tm}) are different. A final interesting comparison can be established with respect to the mean

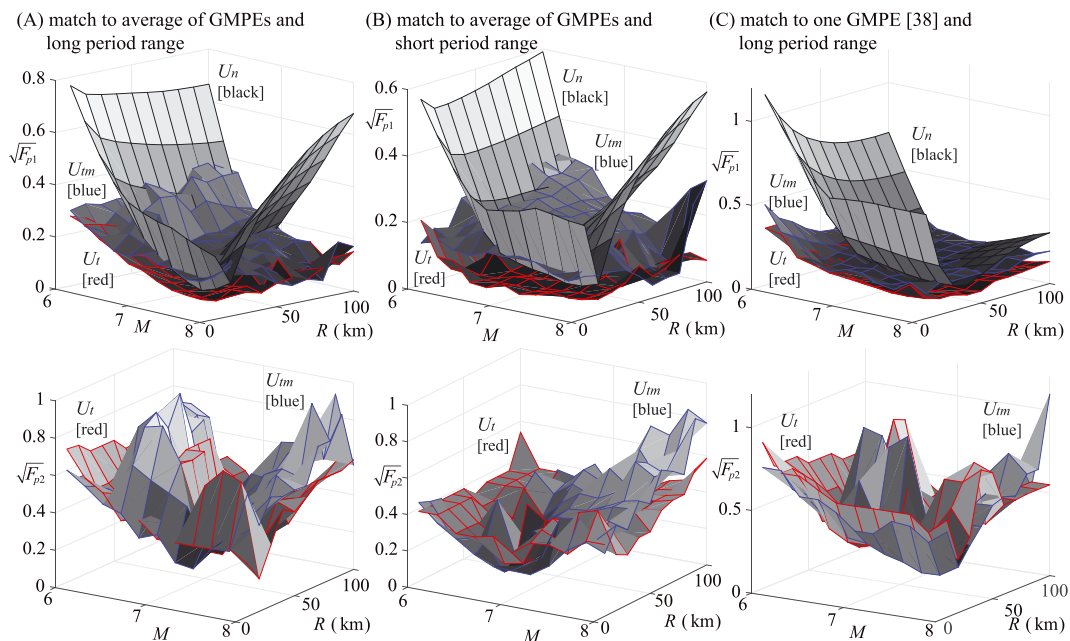


FIGURE 14 Results for $\sqrt{F_{p1}}$ and $\sqrt{F_{p2}}$ for unmodified ground motion model (U_n) and modified ground motion model with minimum distance from utopia point for matching the complete probabilistic hazard (U_t) or the hazard corresponding to mean predictive relationships (U_{tm}). For the latter estimation of objectives adopts variability of the initial (U_n) predictive model but with the corresponding updated predictive mean. Implementations in the different columns correspond to (A) long and (B) short period ranges for matching to the average considered GMPEs and (C) long period ranges for match to GMPE³⁸ [Colour figure can be viewed at wileyonlinelibrary.com]

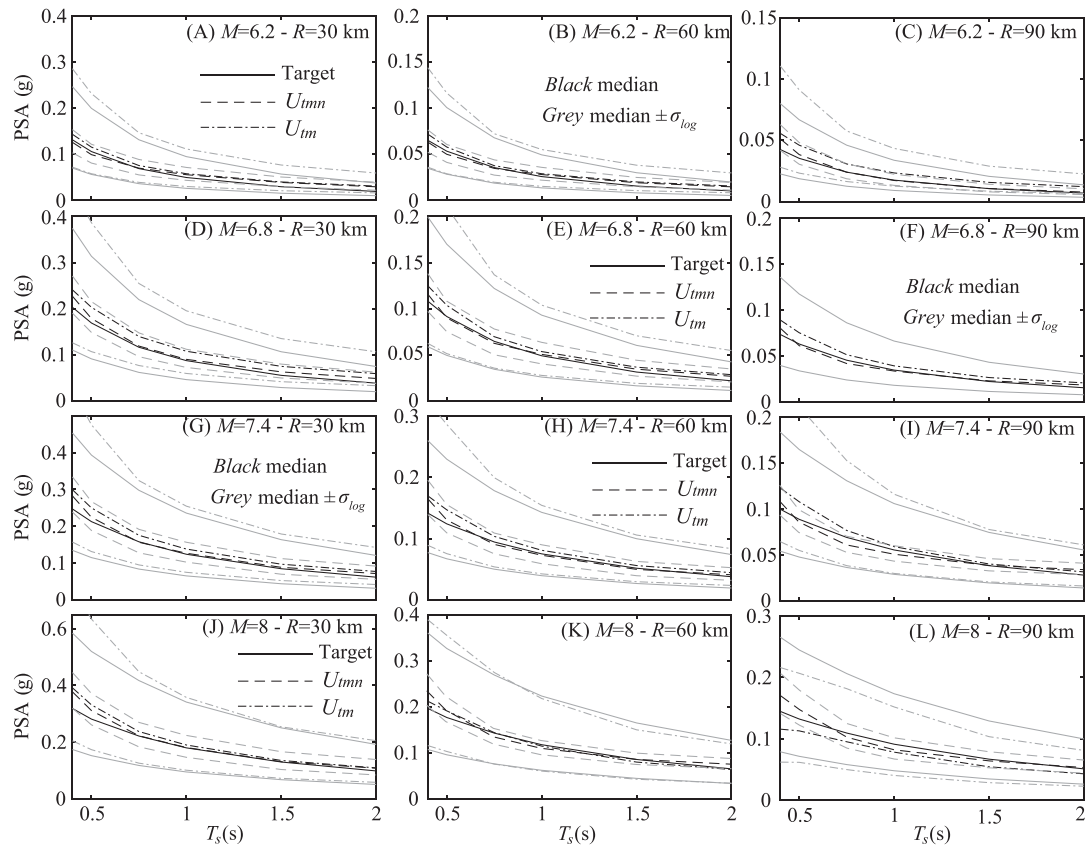


FIGURE 15 Spectral plots for seismicity scenarios (different subplots) corresponding to combinations of M [6.2, 6.8, 7.4, 8] and R [30, 60, 90] km, for the target hazard (target), and modified ground motion model with minimum distance from utopia point for matching the hazard corresponding to mean predictive relationships. For the latter, the cases adopting variability of the initial predictive model (U_{im}) or no variability (U_{imn}) are presented. Implementation scenario shown corresponds to matching to the average considered GMPEs and long period range. Curves corresponding to median and median $\pm \sigma_{log}$ shown

spectral response for U_{im} and U_{imn} (black curves in Figure 15). Consideration of the variability in the predictive model after the optimization (U_{im} case) impacts even the mean of the predicted response, not only the variability of that response, and overall moves that response further away from the desired target, which, recall, was the objective in the underlying modification for the mean predictive relationships μ_m . This discussion shows that even though the surrogate optimization for only μ_m may provide an improvement over the unmodified model for scenarios of high initial discrepancy from the target probabilistic hazard, the post-consideration of variability in the predictive models is problematic. Setting initially $\Sigma = 0$ to identify μ_m and then calculating the hazard for predictive model (μ_m, Σ_r) provides ultimately a lower quality fit to the target hazard. Despite the higher computational burden associated with it, the simultaneous modification of the entire predictive model for θ (both μ and Σ) is therefore advocated.

6 | CONCLUSIONS

The modification of stochastic ground motion models to establish hazard compatibility for specific seismicity scenarios was discussed in this paper. The hazard for each scenario was described with respect to some IM of interest and a probabilistic description was adopted for it, for example defined through mean and dispersion characteristics. The modification of the ground motion model was defined as an adjustment of the probabilistic predictive models/relationships that relate the parameters of the ground motion model to seismicity characteristics. Both the mean of the predictive model and the associated variance were adjusted. The proposed modification was defined as a bi-objective optimization with dual objective of minimizing the discrepancy between the hazard for a given structure/site and the predictions established through the stochastic ground motion model, while maintaining a small deviation from the original predictive relationships, assumed to facilitate similarity to observed regional trends. This setting extends previous efforts of the

authors²⁸ that examined modification of only the mean predictive relationships with goal to match the corresponding IM predictions, ignoring any variability in either of these two components. The relative entropy was adopted as metric to quantify the objectives considered in the multi-objective optimization problem whereas the same surrogate modeling framework as in Tsioulou et al²⁸ was utilized for an efficient optimization. Emphasis was placed on the estimation of the various response statistics needed for the entropy calculation, and a MCS approach was advocated for it coupled with assumption for lognormal distribution of the response when considering the variability of the predictive models/relationships. The computational approach explicitly considered the fact that most ground motion models involve a separate parameter that impacts their scaling. This parameter was separately treated with respect to both the surrogate model development and the MCS. Different statistical assumption for the distribution of the ground motion model output were also examined for the evaluation of the entropy for the first objective.

In the illustrative example, the proposed framework was applied using a recently developed record-based stochastic ground motion model. It was shown that lognormal distribution assumption for calculating the first objective provides an adequate approximation for the application at hand, whereas the metamodel-aided optimization can facilitate an accurate identification of the Pareto front even when lower accuracy metamodels are utilized, a feature that did not hold in study.²⁸ The necessity to calculate response statistics through MCS increases, though, the computational burden of the approach. Application to wide range of seismicity scenarios and different approaches for determining the seismic hazard (different IMs or sources for the target values) were examined, illustrating the advantages for the proposed framework: it allows significant improvements to the target hazard match, especially for seismicity scenarios for which the unmodified model provides a poor initial fit, with minor only modifications to the original predictive model, something that can guarantee a good agreement with observed regional trends. With respect to selection of the final model across the identified Pareto front, same recommendation as in Tsioulou et al²⁸ is made: select the Pareto optimal solution that satisfies a certain accuracy threshold for match to the target hazard (F_{p1} constraint) unless this solution leads to a greater modification for the predictive model (F_{p2} value) than the Pareto optimal solution with minimum distance from the Utopia point. Finally, the approach presented in this paper was compared with the modification of only the mean predictive relationships to match the corresponding hazard for these mean predictions.²⁸ It was shown in this case that the latter approach may provide an adequate surrogate for seismicity scenarios with high initial discrepancy to the target hazard, though overall it is better to explicitly consider the impact of the variability in the predictive models (ie, calculate response statistics) and simultaneously modify the entire predictive model (ie, not only focus on the mean of the predictive relationships).

The main limitation of the approach is the significant computational burden for performing the multi-objective optimization to identify the Pareto front, a burden stemming from the MCS step. Considering the fact that the proposed modification needs to be repeated for each seismicity scenario of interest, further reduction of this burden, which will have to come from a more computational efficient implementation of the surrogate model predictions, is an important extension of this work, currently under investigation by the authors.

ORCID

Alexandros A. Taflanidis  <http://orcid.org/0000-0002-9784-7480>

Carmine Galasso  <http://orcid.org/0000-0001-5445-4911>

REFERENCES

1. Au SK, Beck JL. Subset simulation and its applications to seismic risk based on dynamic analysis. *J Eng Mech ASCE*. 2003;129(8):901-917.
2. Gidaris I, Taflanidis AA, Lopez-Garcia D, Mavroeidis GP. Multi-objective risk-informed design of floor isolation systems. *Earthq Eng Struct Dyn*. 2016;45(8):1293-1313.
3. Jensen HA, Kusanovic DS. On the effect of near-field excitations on the reliability-based performance and design of base-isolated structures. *Probab Eng Mech*. 2014;36:28-44.
4. Goulet CA, Haselton CB, Mitrani-Reiser J, et al. Evaluation of the seismic performance of code-conforming reinforced-concrete frame building—from seismic hazard to collapse safety and economic losses. *Earthq Eng Struct Dyn*. 2007;36(13):1973-1997.
5. Bobby S, Spence SM, Bernardini E, Kareem A. Performance-Based Topology Optimization for Buildings under Wind and Seismic Hazards. In Proc. of the 2015 Structured Congress. 2015.

6. Gidaris I, Taflanidis AA. Performance assessment and optimization of fluid viscous dampers through life-cycle cost criteria and comparison to alternative design approaches. *Bull Earthq Eng*. 2015;13(4):1003-1028.
7. Lin T, Haselton CB, Baker JW. Conditional spectrum-based ground motion selection. Part I: hazard consistency for risk-based assessments. *Earthq Eng Struct Dyn*. 2013;42(12):1847-1865.
8. Smerzini C, Galasso C, Iervolino I, Paolucci R. Ground motion record selection based on broadband spectral compatibility. *Earthq Spectra*. 2014;30(4):1427-1448.
9. Jalayer F, Ebrahimi H, Miano A, Manfredi G, Sezen H. Analytical fragility assessment using unscaled ground motion records. *Earthq Eng Struct Dyn*. 2017;46(15):2639-2663.
10. Kohrangi M, Bazzurro P, Vamvatsikos D, Spillatura A. Conditional spectrum-based ground motion record selection using average spectral acceleration. *Earthq Eng Struct Dyn*. 2017;46(10):1667-1685.
11. Jalayer F, Beck JL. Effects of two alternative representations of ground-motion uncertainty in probabilistic seismic demand assessment of structures. *Earthq Eng Struct Dyn*. 2008;37(1):61-79.
12. Galasso C, Zareian F, Iervolino I, Graves RW. Validation of ground-motion simulations for historical events using SDoF systems. *Bull Seismol Soc Am*. 2012;102(6):2727-2740.
13. Vetter C, Taflanidis A. Comparison of alternative stochastic ground motion models for seismic risk characterization. *Soil Dyn Earthq Eng*. 2014;58:48-65.
14. Broccardo M, Der Kiureghian A. Multicomponent nonlinear stochastic dynamic analysis by tail-equivalent linearization. *J Eng Mech*. 2015;142(3):04015100.
15. Chun J, Song J, Paulino GH. Structural topology optimization under constraints on instantaneous failure probability. *Struct Multidiscip Optim*. 2016;53(4):773-799.
16. Rezaeian S, Der Kiureghian A. Simulation of synthetic ground motions for specified earthquake and site characteristics. *Earthq Eng Struct Dyn*. 2010;39(10):1155-1180.
17. Gavin HP, Dickinson BW. Generation of uniform-hazard earthquake ground motions. *J Struct Eng ASCE*. 2011;137(3):423-432.
18. Yamamoto Y, Baker JW. Stochastic model for earthquake ground motion using wavelet packet. *Bull Seismol Soc Am*. 2013;103(6):3044-3056.
19. Vlachos C, Papakonstantinou KG, Deodatis G. A multi-modal analytical non-stationary spectral model for characterization and stochastic simulation of earthquake ground motions. *Soil Dyn Earthq Eng*. 2016;80:177-191.
20. Boore DM. Simulation of ground motion using the stochastic method. *Pure Appl Geophys*. 2003;160(3):635-676.
21. Atkinson GM, Silva W. Stochastic modeling of California ground motions. *Bull Seismol Soc Am*. 2000;90(2):255-274.
22. Papadimitriou, C. and J. Beck. Nonstationary stochastic characterization of strong-motion accelerograms. In Proc. of 4th US National Conference on Earthquake Engineering 1990.
23. Grossi, P., Catastrophe modeling: a new approach to managing risk. Vol. 25. 2005: Springer Science & Business Media.
24. Atkinson GM. Ground-motion prediction equations for eastern North America from a referenced empirical approach: implications for epistemic uncertainty. *Bull Seismol Soc Am*. 2008;98(3):1304-1318.
25. McGuire R.K., Seismic hazard and risk analysis. 2004: Earthquake Engineering Research Institute.
26. Scherbaum F, Cotton F, Staedte H. The estimation of minimum-misfit stochastic models from empirical ground-motion prediction equations. *Bull Seismol Soc Am*. 2006;96(2):427-445.
27. Vetter C, Taflanidis AA, Mavroeidis GP. Tuning of stochastic ground motion models for compatibility with ground motion prediction equations. *Earthq Eng Struct Dyn*. 2016;45(6):893-912.
28. Tsioulou A, Taflanidis AA, Galasso C. Modification of stochastic ground motion models for matching target intensity measures. *Earthq Eng Struct Dyn*. 2017;47(1):3-24. <https://doi.org/10.1002/eqe.2933>
29. Medel-Vera C, Ji T. A stochastic ground motion accelerogram model for Northwest Europe. *Soil Dyn Earthq Eng*. 2016;82:170-195.
30. Bozorgnia Y, Hachem MM, Campbell KW. Ground motion prediction equation ("attenuation relationship") for inelastic response spectra. *Earthq Spectra*. 2010;26(1):1-23.
31. Stewart JP, Boore DM, Seyhan E, Atkinson GM. NGA-West2 equations for predicting vertical-component PGA, PGV, and 5%-damped PSA from shallow crustal earthquakes. *Earthq Spectra*. 2015;32(2):1005-1031.
32. Gibbs AL, Su FE. On choosing and bounding probability metrics. *Int Stat Rev*. 2002;70(3):419-435.
33. Lophaven, S.N., H.B. Nielsen, and J. Sondergaard, DACE-A MATLAB Kriging Toolbox. 2002, Technical University of Denmark
34. Jia G, Taflanidis AA. Kriging metamodeling for approximation of high-dimensional wave and surge responses in real-time storm/hurricane risk assessment. *Comput Methods Appl Mech Eng*. 2013;261-262:24-38.
35. Spall JC. *Introduction to Stochastic Search and Optimization*. New York: Wiley-Interscience; 2003.
36. Marler RT, Arora JS. Survey of multi-objective optimization methods for engineering. *Struct Multidiscip Optim*. 2004;26(6):369-395.

37. Abrahamson N, Silva W. Summary of the Abrahamson & Silva NGA ground-motion relations. *Earthq Spectra*. 2008;24(1):67-97.
38. Boore DM, Atkinson GM. Ground-motion prediction equations for the average horizontal component of PGA, PGV, and 5%-damped PSA at spectral periods between 0.01 s and 10.0 s. *Earthq Spectra*. 2008;24(1):99-138.
39. Chiou BSJ, Youngs RR. An NGA model for the average horizontal component of peak ground motion and response spectra. *Earthq Spectra*. 2008;24(1):173-215.
40. Campbell KW, Bozorgnia Y. NGA ground motion model for the geometric mean horizontal component of PGA, PGV, PGD and 5% damped linear elastic response spectra for periods ranging from 0.01 to 10 s. *Earthq Spectra*. 2008;24(1):139-171.
41. Kakkamanos J, Baise LG, Boore DM. Estimating unknown input parameters when implementing the NGA ground-motion prediction equations in engineering practice. *Earthq Spectra*. 2011;27(4):1219-1235.
42. Gidaris I, Taflanidis AA, Mavroeidis GP. Kriging metamodeling in seismic risk assessment based on stochastic ground motion models. *Earthq Eng Struct Dyn*. 2015;44(14):2377-2399.
43. Keeney RL, Raiffa H. *Decisions with Multiple Objectives: Preferences and Trade-offs*. Cambridge University Press; 1993.
44. Scott DW. *Multivariate Density Estimation: Theory, Practise and Visualization*. New York, N.Y: Wiley-Interscience; 1992.

How to cite this article: Tsioulou A, Taflanidis AA, Galasso C. Hazard-compatible modification of stochastic ground motion models. *Earthquake Engng Struct Dyn*. 2018;47:1774–1798. <https://doi.org/10.1002/eqe.3044>

APPENDIX A

KERNEL-BASED APPROXIMATION OF THE PROBABILITY DISTRIBUTION OF THE LN(IM) FOR THE GROUND MOTION MODEL AND ENTROPY ESTIMATION

This Appendix considers the estimation of $p_g(\ln(Y_i)|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and of objective F_{p1} when the lognormal assumption is not invoked for the distribution of Y_i^g . Approach relies on obtaining samples from $p_g(\ln(Y_i)|\boldsymbol{\mu}, \boldsymbol{\Sigma})$, which when using the actual ground motion model is established through the following process. First generate n_d samples for $\boldsymbol{\theta}$ from $p(\boldsymbol{\theta}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $\{\boldsymbol{\theta}^d; d = 1, \dots, n_d\}$, and n_d sample white noise sequences, $\{\mathbf{w}^d; d = 1, \dots, n_d\}$, and obtain the corresponding acceleration time-histories $\ddot{a}(t|\boldsymbol{\theta}^d, \mathbf{w}^d); d = 1, \dots, n_d$. For each sample, the response output of interest is estimated providing samples of $\{\ln(Y_i^{g,d}); d = 1, \dots, n_d\}$ from $p_g(\ln(Y_i)|\mathbf{z})$. The latter can be approximated using KDE based on these samples as:

$$\tilde{p}_g(\ln(Y_i)|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{n_d} \sum_{d=1}^{n_d} \frac{1}{h} K\left(\frac{\ln(Y_i) - \ln(Y_i^{g,d})}{h}\right); K(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \quad (\text{A1})$$

where $K(\cdot)$ is the chosen kernel and h is the Kernel bandwidth. In this study, the widely used Gaussian kernel is employed, shown also in Equation A1 with bandwidth chosen as⁴⁴ $h = 1.06 \cdot n_d^{-1/5} \sigma_d$ where σ_d is the standard deviation of the samples $\{\ln(Y_i^{g,d}); d = 1, \dots, n_d\}$. The entropy in Equation 6 can be then approximated using the KDE estimate in Equation A1, with the scalar integral calculated through numerical integration (trapezoidal rule).

When the IM is approximated through use of the metamodel $Y_i^g(\boldsymbol{\theta}, \mathbf{w}) = \sqrt{\bar{\theta}_s} s_i(\mathbf{x}, \mathbf{w})$, then the approach for obtaining the samples $\{\ln(Y_i^{g,d}); d = 1, \dots, n_d\}$ changes and requires, additionally, another statistical assumption for the distribution of $Y_i^g(\boldsymbol{\theta}, \mathbf{w})$ under the influence of \mathbf{w} . Note that this is a different setting than invoking a specific distribution for Y_i^g , because in this case the assumption pertains only to the influence of the white noise. The standard statistical approximation, with proven accuracy in a number of studies,⁴² is lognormal assumption. In this case, the desired samples are obtained through the following process. First generate n_d samples for $\boldsymbol{\theta}$ from $p(\boldsymbol{\theta}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $\{\boldsymbol{\theta}^d; d = 1, \dots, n_d\}$, and estimate through the metamodel outputs $\ln(\bar{s}_i(\mathbf{x}^d))$ and $\sigma_i^s(\mathbf{x}^d)$ for $d = 1, \dots, n_d$. Generate n_d samples $\{e^d; d = 1, \dots, n_d\}$ from a standard Gaussian distribution and obtain each of the desired samples as

$$\ln(Y_i^{g,d}) = \ln(\bar{s}_i(\mathbf{x}^d)) + \sigma_i^s(\mathbf{x}^d) e^d + \frac{1}{2} \ln \theta_s^d. \quad (\text{A2})$$

Once these samples are obtained, then the KDE estimation of $p_g(\ln(Y_i)|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and the calculation of F_{p1} follow same approach as in the case that actual ground motion model was utilized.