

RESEARCH ARTICLE

Modification of stochastic ground motion models for matching target intensity measures

Alexandra Tsioulou¹ | Alexandros A. Taflanidis²  | Carmine Galasso¹

¹Department of Civil, Environmental and Geomatic Engineering and Institute for Risk and Disaster Reduction, 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

Stochastic ground motion models produce synthetic time-histories by modulating a white noise sequence through functions that address spectral and temporal properties of the excitation. The resultant ground motions can be then used in simulation-based seismic risk assessment applications. This is established by relating the parameters of the aforementioned functions to earthquake and site characteristics through predictive relationships. An important concern related to the use of these models is the fact that through current approaches in selecting these predictive relationships, compatibility to the seismic hazard is not guaranteed. This work offers a computationally efficient framework for the modification of stochastic ground motion models to match target intensity measures (IMs) for a specific site and structure of interest. This is set as an optimization problem with a dual objective. The first objective minimizes the discrepancy between the target IMs and the predictions established through the stochastic ground motion model for a chosen earthquake scenario. The second objective constrains the deviation from the model characteristics suggested by existing predictive relationships, guaranteeing that the resultant ground motions not only match the target IMs but are also compatible with regional trends. A framework leveraging kriging surrogate modeling is formulated for performing the resultant multi-objective optimization, and different computational aspects related to this optimization are discussed in detail. The illustrative implementation shows that the proposed framework can provide ground motions with high compatibility to target IMs with small only deviation from existing predictive relationships and discusses approaches for selecting a final compromise between these two competing objectives.

KEYWORDS

hazard-compatible tuning, kriging surrogate modeling, seismic risk, stochastic ground motion model

1 | INTRODUCTION

The growing interest in performance-based earthquake engineering^{1,2} and in simulation-based, seismic risk assessment approaches^{3,4} has increased in the past decades the relevance of ground motion modeling techniques. These techniques describe the entire time-series of seismic excitations, providing a characterization appropriate for dynamic time-history analysis. Undoubtedly

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.

© 2017 The Authors. Earthquake Engineering & Structural Dynamics Published by John Wiley & Sons Ltd.

the most popular methodology for performing this task for seismic risk assessment (or seismic design) applications is the selection and modification of real (ie, recorded from past events) ground motions based on a target intensity measure (IM) level,^{5,6} eg, an elastic pseudo-acceleration response spectrum. For seismic risk assessment, such modification is performed for specific seismicity scenarios (typically defined through moment magnitude and source-to-site distance) contributing to the seismic hazard for the chosen site, with the target IM commonly⁷ derived through ground motion predictive equations (GMPEs).^{8,9}

An alternative philosophy for describing seismic excitations is to use simulated ground motions.^{10,11} A specific modeling approach for the latter, which has been steadily gaining increasing attention by the structural engineering community,¹²⁻¹⁴ is the use of stochastic ground motion models.¹⁵⁻²⁰ These models are based on modulation of a stochastic sequence, through functions (filters) that address spectral and temporal characteristics of the excitation. The parameters of these filters are related to seismicity (eg, moment magnitude and rupture distance) and site characteristics (eg, shear wave velocity for soil profile) through predictive relationships.^{15,19} Sample ground motions for any desired seismicity scenario can be generated by determining the parameters of the stochastic ground motion model through these predictive relationships and by using a sample stochastic sequence. This approach may ultimately support a comprehensive description of the seismic hazard.²¹

The essential component of stochastic ground motion models is the development of the associated predictive relationships, and various approaches have been established to accomplish this, with main representatives being record-based and physics-based models. Record-based models (also known as site-based) are developed by fitting a preselected “waveform” to a suite of recorded regional ground motions.^{15,18,22} On the other hand, stochastic physics-based models rely on physical modeling of the rupture and wave propagation mechanisms.^{19,20} Emphasis in this study will be on the former models, although the techniques discussed can be extended to any type of stochastic ground motion model.

An important concern related to the use of stochastic ground motion models for structural engineering applications is the fact that through current approaches in selecting their predictive relationships compatibility to the seismic hazard for specific structures and sites is not necessarily obtained¹⁵ (this is also shown in the illustrative example considered later). Although validation of these models is frequently performed by comparison of their spectral acceleration outputs to GMPEs,^{15,17} the match to GMPEs is not explicitly incorporated in the predictive relationships development. Such a match to some desired GMPE (or target IMs in general) is though important for subsequent use of the stochastic ground motion models to describe the seismic hazard. Take for example the recent FEMA P-58²³ guidelines for seismic performance assessment of structures; the scenario-based description of the seismic hazard requires match of the median response to the one described by a GMPE (for the specific seismicity scenario examined). Similarly, the intensity- and time-based descriptions in FEMA P-58 require compatibility with seismic hazard curves, which are ultimately defined through use of GMPEs.²⁴

This realization has motivated researchers to investigate the selection of predictive relationships for stochastic ground motion models so that compatibility with GMPEs is explicitly established.²⁵ The formulation introduces an explicit optimization for matching the median predictions of the ground motion model to the spectral acceleration estimates of GMPEs, while maintaining physics-based principles or the matching to trends from real ground motions as an optimization constraint, in an attempt to preserve desired ground motion characteristics. Vetter et al²⁶ recently extended the work of²⁵ by providing a versatile and computationally efficient approach, leveraging surrogate modeling principles, for tuning stochastic ground motion models to establish compatibility with the median GMPE predictions for a range of structural periods and seismicity scenarios of interest. One of the main drawbacks of this tuning approach, though, is that the physical characteristics of the resulting acceleration time-series are incorporated in the optimization merely as constraints, something that requires significant experience in ground motion characterization for proper definition of the optimization problem and can furthermore lead to synthetic time histories with unrealistic properties for some seismicity scenarios.

The current study addresses this critical shortcoming and looks at the modification of stochastic ground motion models for specific seismicity scenarios with a dual goal of (i) matching a target IM for a specific structure (or range of structures) while (ii) preserving desired trends and correlations in the physical characteristics of the resultant ground acceleration time series. This is ultimately formulated as a multi-objective optimization problem. The first objective is to minimize the discrepancy between the median ground motion output and the target IM for a given seismicity scenario. Any desired IM can be used for this purpose with only requirement to have a corresponding seismicity scenario. For instance, if the target IM is derived through probabilistic seismic hazard analysis, a corresponding seismicity scenario (or ‘design earthquake’) can be derived through the disaggregation of seismic hazard⁷ for a given hazard level. The second objective is to establish the smallest deviation from the model characteristics suggested by existing predictive relationships. This second objective aims at maintaining regional physical characteristics and parameter correlations with respect to existing predictive relationships. Approach differs significantly from Vetter et al²⁶; rather than tuning the ground motion for hazard compatibility ignoring any existing predictive relationships, goal here is the minimum modification of the existing relationships that will yield the desired compatibility. This is ultimately posed as a multi-objective problem to better investigate the compromise between

the two different objectives, and for efficiently solving it a surrogate modeling approach is adopted, similar to that of Vetter et al.²⁶ A surrogate model (ie, metamodel) is trained based on an initial database of ground motion simulations, and ultimately provides a highly efficient approximation for the spectral acceleration predictions of the stochastic ground motion model. The surrogate model is then leveraged to solve the optimization problem. Emphasis is also placed here on the selection of the database to inform the metamodel development, which constitutes a significant advancement over the approach by Vetter et al.²⁶ Blind search and gradient-based approaches are considered for the multi-objective optimization, and the relative computational benefits of each are explored.

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

2 | PROBLEM FORMULATION

Consider a stochastic ground motion model that provides acceleration time-histories $\ddot{a}(t|\boldsymbol{\theta}, \mathbf{W})$ by modulating a Gaussian white noise sequence, \mathbf{W} , through appropriate time/frequency functions that are parameterized through the n_θ -dimensional model parameter vector $\boldsymbol{\theta} = [\theta_1 \ \theta_2 \ \dots \ \theta_{n_\theta}] \in \mathfrak{R}^{n_\theta}$. This vector completely defines the model and is typically composed of various excitation properties such as Arias intensity, strong ground motion duration or parameters related to frequency characteristics of the ground motion. A specific example for such a model, the one used in the illustrative example later, is provided in Appendix A. This particular record-based model efficiently addressed both temporal and spectral nonstationarities. The former is established through a time-domain modulating envelope function, whereas the latter is achieved by filtering a white noise process by a filter with characteristics that vary in time.

Synthetic time-histories can be created by relating $\boldsymbol{\theta}$ to seismicity and local site properties through predictive relationships. The vector of these properties, called seismological parameters, is denoted as \mathbf{z} . Common characteristics used for \mathbf{z} ^{15,19} include the fault type F , the moment magnitude, M , the rupture distance, R , and the shear wave velocity in the upper 30 meters of soil, V_{s30} . For record-based models, the standard approach for development of these predictive relationships^{15,27} relies on first matching the waveform characteristics to recorded ground motions (ie, identify first $\boldsymbol{\theta}$ for each of the recorded ground motions in a given database) and then performing a regression to relate $\boldsymbol{\theta}$ to \mathbf{z} . This leads ultimately to a probabilistic regression model for $\boldsymbol{\theta}$ with mean predictive relationship $\boldsymbol{\mu}_\theta(\mathbf{z})$ that is dependent on \mathbf{z} and some associated uncertainty characterization U , identified from the residuals of the regression, that is independent of \mathbf{z} . Typically, this is performed by first transforming the problem to the standard Gaussian space through a nonlinear mapping for each component θ_i . The transformed Gaussian vector is denoted $\mathbf{v}(\boldsymbol{\theta})$ herein. Approach ultimately leads to a Gaussian probability model $\mathbf{v} \sim N(\boldsymbol{\mu}(\mathbf{z}), \boldsymbol{\Sigma})$ with mean $\boldsymbol{\mu}(\mathbf{z})$ and covariance matrix $\boldsymbol{\Sigma}$. In this case, the uncertainty characterization U corresponds to the covariance matrix $\boldsymbol{\Sigma}$ and to the fact that probability model for \mathbf{v} is identified as Gaussian. Appendix A includes more details for a specific ground motion model.¹⁵ Note that a similar description can be established for physics-based models. In this case, the predictive relationships $\boldsymbol{\mu}_\theta(\mathbf{z})$ are obtained through rupture and wave propagation principles,^{19,28} whereas the uncertainty characterization U can be established by assigning probability models for $\boldsymbol{\theta}$ through an epistemic uncertainty treatment.^{12,29}

As discussed in Section 1, this formulation for the predictive relationships of stochastic ground motion models, prioritizing a match to regional trends, provides synthetic ground motions whose output IMs do not necessarily match hazard-compatible IMs (eg, as derived from GMPEs). For this purpose, a modification of the model parameter vector $\boldsymbol{\theta}$ is proposed for specific seismicity scenarios defined by \mathbf{z} with objective to (i) match a target IM vector, while (ii) maintaining similarity to the predictive relationships already established for the model. Equivalently, this can be viewed as identifying the model characteristics $\boldsymbol{\theta}$ that are closest to the established model $\boldsymbol{\mu}_\theta(\mathbf{z})$ (considering, when available, any additional information provided by U) and also match the intended hazard (described through some IM). The IM vector may include different response quantities of interest, for example, (i) direct characteristics of the ground motion, such as peak ground acceleration, velocity, and displacement; (ii) elastic and inelastic spectral responses for different periods of a single degree of freedom (SDoF) oscillator; or (iii) more complex spectral or ground motion related quantities proposed by different researchers.³⁰⁻³² The target for most of these IMs can be described through a GMPE.^{8,33} However, this is not necessary; any IM description can be used, with only requirement that a corresponding seismicity scenario is defined. Note that if match to spectral responses (ie, a spectral plot) is of interest, then a range of structural periods for which the match is established needs to be determined.

To formalize these concepts mathematically, let, $Y_i(\mathbf{z}); i = 1, \dots, n_y$, denote the target response quantities of interest and $Y_i^m(\boldsymbol{\theta})$ the median predictions for the same quantities provided through the stochastic ground motion model. The median predictions are obtained through the following process:

- Step 1. Generate n_w sample acceleration time-histories for different white noise sequences $\ddot{a}_\kappa(t|\boldsymbol{\theta}, \mathbf{W}_\kappa)$; $\kappa = 1, \dots, n_w$.
- Step 2. For each sample evaluate the responses of interest. For spectral quantities this will entail numerical simulation of SDoF responses.
- Step 3. Estimate the statistics (median) over the established sample-set.

The modification problem is ultimately formulated as bi-objective optimization problem:

$$\boldsymbol{\theta} = \arg \min \{F_1(\boldsymbol{\theta}|\mathbf{z}), F_2(\boldsymbol{\theta}|\mathbf{z})\}. \quad (1)$$

The first objective F_1 corresponds to a measure of the discrepancy from the chosen target. One choice for this measure is the average weighted square error given by

$$F_1(\boldsymbol{\theta}|\mathbf{z}) = \frac{1}{n_y} \sum_{i=1}^{n_y} \gamma_i^2 (Y_i(\mathbf{z}) - Y_i^m(\boldsymbol{\theta}))^2 \quad (2)$$

with γ_i corresponding to the weights prioritizing the match to different IM components. A typical selection for γ_i is $1/Y_i(\mathbf{z})$, then the quantity in Equation 2 corresponds to the average squared relative error. Alternative formulations for this first objective, facilitating perhaps a better physical intuition, are the average absolute error or the maximum absolute error over the different IMs, given, respectively by,

$$F_{1r}(\boldsymbol{\theta}|\mathbf{z}) = \frac{1}{n_y} \sum_{i=1}^{n_y} \gamma_i |Y_i(\mathbf{z}) - Y_i^m(\boldsymbol{\theta})|, \quad (3)$$

$$F_{1m}(\boldsymbol{\theta}|\mathbf{z}) = \max_{i=1, \dots, n_y} \gamma_i |Y_i(\mathbf{z}) - Y_i^m(\boldsymbol{\theta})| \quad (4)$$

Both these objectives lead, though, to more challenging optimization problems as they correspond to nondifferentiable functions. Thus, preference will be here for the objective given by Equation 2. The alternative measures will be used to evaluate the suitability of different solutions.

The second objective F_2 measures the discrepancy of $\boldsymbol{\theta}$ from the established predictive relationships. One choice for $F_2(\boldsymbol{\theta}|\mathbf{z})$ could be $[\boldsymbol{\theta} - \boldsymbol{\mu}_\theta(\mathbf{z})]^T [\boldsymbol{\theta} - \boldsymbol{\mu}_\theta(\mathbf{z})]$, ie, the discrepancy from the mean predictive relationships. If there is no available uncertainty characterization U , then this is the only option that can be made. If this characterization is available, as is the case typically with record-based models,¹⁵ correlation and variability information for the model parameters can be additionally incorporated. This is established by considering the maximization of the likelihood of $\boldsymbol{\theta}$ based on the probability model described through U . For the stochastic ground motion model described in Appendix A, this corresponds to the maximization of the probability density for $\mathbf{v}(\boldsymbol{\theta})$, leading to

$$F_2(\boldsymbol{\theta}|\mathbf{z}) = [\mathbf{v}(\boldsymbol{\theta}) - \boldsymbol{\mu}(\mathbf{z})]^T \boldsymbol{\Sigma}^{-1} [\mathbf{v}(\boldsymbol{\theta}) - \boldsymbol{\mu}(\mathbf{z})] \quad (5)$$

The covariance matrix $\boldsymbol{\Sigma}$ incorporates in the formulation the correlation between the model characteristics as well as the fact that variability for each of these characteristics is different.

Objective function F_1 enforces the match to the target IMs. Objective F_2 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 1 ultimately leads to a Pareto set of dominant solutions $\{\boldsymbol{\theta}_p; p = 1, \dots, n_p\}$ that express a different compromise between the competing objectives F_1 and F_2 . A solution is characterized as dominant (and belongs in the Pareto set) if there is no other solution that simultaneously improves both objectives F_1 and F_2 . The representation of the Pareto set in the performance objective $[F_1, F_2]$ space, $\{[F_1(\boldsymbol{\theta}_p|\mathbf{z}), F_2(\boldsymbol{\theta}_p|\mathbf{z})]; p = 1, \dots, n_p\}$ is termed as the Pareto front. Illustrations of such Pareto fronts are included in the example discussed later. One extreme point of this front will always correspond to the unmodified model with $\boldsymbol{\theta} = \boldsymbol{\mu}_\theta(\mathbf{z})$, representing the minimum of objective $F_2 = 0$. Unless this point also yields a match to the targeted hazard (ie, corresponds to $F_1 = 0$), optimization of Equation 1 will identify points that improve upon $F_1(\boldsymbol{\mu}_\theta(\mathbf{z})|\mathbf{z})$ while

deviating from the unmodified model ($F_2 > 0$). One can eventually select a model configuration from the identified Pareto set that yields the desired IM-compatibility without deviating significantly from regional ground motion characteristics. This will be further discussed in the illustrative implementation in Section 5.4.

Identifying the Pareto set for this problem is challenging because the computational burden in evaluation of objective F_1 is significant, requiring $n_w \cdot n_y$ time-history analyses for each objective function estimation. To facilitate an efficient optimization that can be repeated for any desired seismicity scenario \mathbf{z} , a surrogate modeling approach is adopted here, similar to the one used by Vetter et al.²⁶ Specifically, kriging is used as metamodel since it has a proven capability to approximate highly complex functions,³⁴ while simultaneously providing gradient information that will be directly exploited in the optimization and allowing to explicitly consider the local metamodel approximation error within the optimization formulation. These aspects of the optimization problem will be discussed in Section 4. The details of the kriging metamodel development are discussed first in the next section.

3 | KRIGING METAMODEL DEVELOPMENT

The kriging metamodel is developed to provide an efficient approximation to the input–output relationship $\boldsymbol{\theta} - Y_i^m(\boldsymbol{\theta})$ considering every potential response quantity of interest that can be eventually used for the definition of objective F_1 . A further simplification can be established if the relationship between some components of $\boldsymbol{\theta}$, and the response output $Y_i^m(\boldsymbol{\theta})$ is explicitly known. This is true for stochastic ground motion models that include a parameter, denoted θ_s herein, that directly controls the amplitude of the excitation. This means that $Y_i^m(\boldsymbol{\theta}) = \theta_s s_i^m(\mathbf{x})$ with \mathbf{x} corresponding to the remaining model parameters excluding θ_s and $s_i^m(\mathbf{x})$ representing the output $Y_i^m(\boldsymbol{\theta})$ for $\theta_s = 1$. For the model described in Appendix A $\theta_s = \sqrt{I_a}$ and $\mathbf{x} = \{D_{5-95}, t_{\text{mid}}, \omega_{\text{mid}}, \omega', \zeta_f\}$. Without loss of generality, we will adopt here this assumption, ie, representation $Y_i^m(\boldsymbol{\theta}) = \theta_s s_i^m(\mathbf{x})$. In this case, the metamodel needs to be established to approximate only relationship $\mathbf{x} - s_i^m(\mathbf{x})$.

For developing the metamodel, a database with n observations is initially obtained that provides information for the $\mathbf{x} - s_i^m(\mathbf{x})$ pair. 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 and its determination is discussed later in this section. The median predictions provided through the ground motion model are then established through the 3-step process discussed in Section 2 considering n_w white noise samples. Using the data set $\{\mathbf{x}^j - s_i^m(\mathbf{x}^j); j = 1, \dots, n\}$, the kriging model can be formulated. Details for this formulation may be found in the studies of Sacks et al.³⁵ or Vetter et al.,²⁶ with the latter reference focusing on a similar application as the one considered here, looking at approximating the predictions of stochastic ground motion models.

This approach ultimately leads to a kriging predictor that has a Gaussian nature with predictive mean $\hat{s}_i^m(\mathbf{x})$ and local prediction variance, which is also a function of \mathbf{x} , $\sigma_i^2(\mathbf{x})$.³⁵ Each response output can be approximated through this predictor leading to

$$Y_i^m(\boldsymbol{\theta}) = \theta_s \hat{s}_i^m(\mathbf{x}) + \varepsilon_i \theta_s \sigma_i(\mathbf{x}), \quad (6)$$

where ε_i is a standard Gaussian variable. This facilitates a computationally efficient approximation to $Y_i^m(\boldsymbol{\theta})$ for each $\boldsymbol{\theta}$. This efficiency can be improved by ignoring the metamodel prediction error (ie, setting $\varepsilon_i = 0$) since calculation of predictive variance $\sigma_i^2(\mathbf{x})$ entails a significant higher computational cost than estimation of predictive mean. The computationally intensive aspect of the entire formulation is the development of the database $\{\mathbf{x}^j - s_i^m(\mathbf{x}^j); j = 1, \dots, n\}$, which requires response-history analysis for a large number of model parameters to populate X and a sufficient number of white noise samples to address the resultant variability in the response. However, this needs to be performed only once. As soon as the kriging metamodel is established based on this database, it can be then used to efficiently predict the responses for any other $\boldsymbol{\theta}$ desired. Calculation of $\hat{s}_i^m(\mathbf{x})$ and $\sigma_i^2(\mathbf{x})$ can be also vectorized,³⁶ something that will be leveraged in the numerical optimization discussed in the next section.

The accuracy of this metamodel depends on the number of experiments n used as well as the exact selection of these experiments. A larger value of n can improve accuracy but at the same time can reduce significantly computational efficiency. Because the latter is important for solving the challenging multi-objective optimization problem discussed here, the value of n needs to be kept moderately low. Therefore, the improvement of metamodel accuracy is primarily sought after through an adaptive design of experiments (DoE). The adaptive DoE strategy gradually increases the number of support points, leveraging the metamodel developed through the existing support points to guide the selection of the new experiments. This leads to an iterative identification of support points, whereas the specific strategy adopted here for selecting the support points in each of the iterations corresponds to a sample-based DoE.^{37,38} In the first iteration, since no metamodel is yet available, the initial

n_1 experiments are obtained using Latin hypercube sampling in X , ie, a space-filling DoE. Subsequent iterations adopt the adaptive DoE strategy. At the k th iteration, a surrogate model is developed using the available $n_{av}^{(k)}$ support points. The prediction error of this metamodel is then leveraged to identify new experiments in regions with low metamodel accuracy. This is accomplished through a sample-based implementation: a large number of candidate experiments is first sampled within X that are distributed proportional to $\sigma_i^2(\mathbf{x})$ (for example, through rejection sampling³⁹), and these experiments are then clustered (for example, using k-means clustering⁴⁰) to the desired number n_a of additional experiments, which are added to the existing experiments for a total of $n_{av}^{(k+1)} = n_{av}^{(k)} + n_a$ support points. The clustering is important for avoiding close-proximity support points that ultimately provide information over the same domain in X . A new metamodel is then developed using the $n_{av}^{(k+1)}$ experiments, and its accuracy is assessed, for example, by calculating error statistics through cross-validation.³⁸ If sufficient accuracy is achieved, then adaptive DoE is terminated and $n = n_{av}^{(k+1)}$, else algorithm proceeds to the $k + 1$ st iteration.

Another important feature of the metamodel development is the definition of domain X . This domain needs to (i) cover the entire range of values that the metamodel will be eventually used for (to avoid extrapolations that unavoidably have reduced accuracy) while (ii) avoiding unnecessarily broad definitions that lead to computational effort spent in subdomains within X of no practical interest. Here the definition of X is established by ignoring subdomains that correspond to high values for $F_2(\boldsymbol{\theta}|\mathbf{z})$ for any potential seismicity scenario \mathbf{z} , ie, \mathbf{x} values that are away from the current predictive relationships. Solutions within such subdomains will not be selected in the multi-objective optimization because they correspond to large values for one of the objectives. The X definition is established through these steps:

- Step 1. Create a range of seismicity scenarios $\{\mathbf{z}_c^l; l = 1, \dots, n_c\}$ that are representative of the scenarios that could be eventually considered in the ground motion model tuning.
- Step 2. Define a box-bounded domain X^d that is expected to be a superset of X , and create a large number of samples for $\{\mathbf{x}_d^j; j = 1, \dots, n_s\}$ uniformly distributed in X^d . Range defined through the scenarios in step 1, $\{\boldsymbol{\mu}_0(\mathbf{z}_c^l); l = 1, \dots, n_c\}$, can be also used to guide selection of X^d .
- Step 3. For each candidate sample \mathbf{x}_d^j , calculate value $F_2(\boldsymbol{\theta}_d^j|\mathbf{z}_c^l)$ for each seismicity scenario \mathbf{z}_c^l where $\boldsymbol{\theta}_d^j$ is the sample corresponding to \mathbf{x}_d^j and the mean value for θ_s (for the given \mathbf{z}_c^l). Evaluate

$$D^j = \min_{l=1, \dots, n_c} F_2(\boldsymbol{\theta}_d^j|\mathbf{z}_c^l), \quad (7)$$

which corresponds to the smallest distance for \mathbf{x}_d^j from the current predictive relationships for any potential seismicity scenario.

- Step 4. Set a threshold δ_d and then classify each sample \mathbf{x}_d^j as belonging in set X if D^j is smaller than δ_d (sample is given membership classification 1) or not otherwise (sample is given membership classification 0). Using this classification information, domain X can be characterized through support vector machine (SVM).⁴¹

This approach ultimately leads to an SVM characterization of X . Figure 1 demonstrates some of the steps of this process. Finally, samples within X , as needed for the adaptive DoE, can be generated with negligible computational effort by creating first samples within box-bounded domain X^d and then using the SVM classifier to maintain only the samples belonging in X .

4 | MULTI-OBJECTIVE OPTIMIZATION TO MATCH TARGET IMS SUPPORTED BY KRIGING METAMODELING

The multi-objective optimization of Equation 1 can be efficiently performed by using the kriging approximation given by Equation 6 when evaluating performance objective F_1 . Note that calculating objective F_2 is computationally trivial. In addition, the approximation error of the metamodel can be incorporated in the objective function definition, leading to the following modification²⁶:

$$F_1(\boldsymbol{\theta}|\mathbf{z}) = \frac{1}{n_y} \sum_{i=1}^{n_y} \gamma_i^2 \left[(Y_i(\mathbf{z}) - \theta_s \widehat{s}_i^m(\mathbf{x}))^2 + \theta_s^2 \sigma_i^2(\mathbf{x}) \right] \quad (8)$$

The motivation for incorporating this error is to improve the robustness of the optimization and to avoid convergence to erroneous solutions due to poor quality of the metamodel. This feature will be further explored in the illustrative implementation.

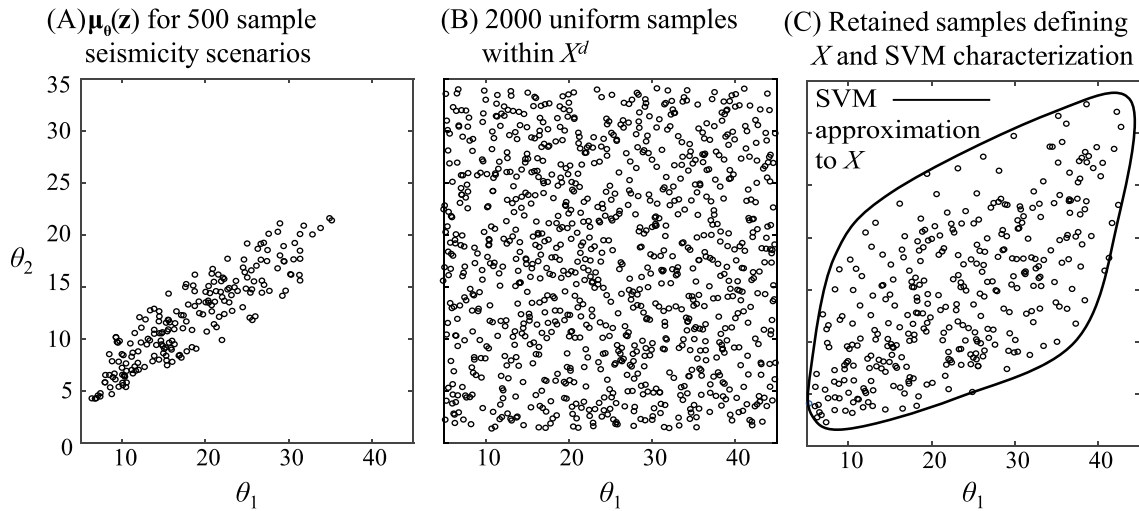


FIGURE 1 Steps for SVM classification of X : A, samples $\mu_0(\mathbf{z})$ for a range of seismicity scenarios; B, uniform samples created within X^d ; C, samples as belonging to X along with developed SVM (curve) [Colour figure can be viewed at wileyonlinelibrary.com]

For solving the multi-objective optimization, a variety of numerical approaches can be used.⁴² Here two are considered, one gradient-free and one gradient-based, offering different advantages. The first approach adopts an exhaustive search.⁴³ A very large number of n_{bc} samples for θ are generated that are close to $\mu_0(\mathbf{z})$, and objective functions F_1 and F_2 are calculated. Estimation of objective F_1 in this case leverages the computational efficiency of the metamodel in performing vectorized predictions: the calculations are simultaneously performed for all n_{bc} samples, or using subsets with a lower number of samples depending on the available computational resources (memory can be a problem for vectorizing operation). This greatly reduces computational time for estimating F_1 . The dominant solutions representing the Pareto front can be then readily identified by comparing the values for the two objectives. The challenge in this case is that the value of n_{bc} needs to be large to obtain an adequate representation of the Pareto front. The advantage is that vectorized calculations can be used for the metamodel predictions. For the stochastic ground motion model described in Appendix A, the samples for θ can be generated by obtaining samples for \mathbf{v} from Gaussian distribution $N(\mu(\mathbf{z}), \Sigma)$ and then transforming these to samples for θ through the inverse of Equation A6. This guarantees that samples will correspond to lower values for F_2 [ie, are close to $\mu_0(\mathbf{z})$] and can therefore emerge as dominant solutions. Note that since this approach does not use gradients, it can seamlessly accommodate the alternative objective functions for F_1 given by Equations 3 and 4.

The second optimization approach is a gradient-based one. The epsilon constraint approach⁴⁴ is specifically adopted because of its ability to explicitly define the value for one of the objectives. This method converts the multi-objective optimization problem to a set of single-objective constraint optimization problems with different constraint bounds ϵ^r . Through systematic variations of this constrain, different Pareto optimums can be obtained. Here objective function F_2 is adopted as objective and F_1 as constraint. This allows identification of the stochastic ground motion model that provides a specific compatibility with the target hazard (the prescribed constraint). The range of values ϵ^r of interest is determined by identifying first the anchor points of the Pareto front, corresponding to the minimum of objective functions $F_1(\theta | \mathbf{z})$ and $F_2(\theta | \mathbf{z})$ (unconstrained single-objective optimizations). Evidently, optimization for $F_2(\theta | \mathbf{z})$ yields solution $\mu_0(\mathbf{z})$. The range for ϵ^r corresponds then to $[\min F_1, F_1(\mu_0(\mathbf{z}) | \mathbf{z})]$, and a number of different constraint values can be considered, with the exact number depending on the desired resolution of the front. For each such value, the single-objective, constrained optimization is solved

$$\begin{aligned} \theta &= \arg \min F_2(\theta | \mathbf{z}) \\ &\text{subject to } F_1(\theta | \mathbf{z}) \leq \epsilon^r. \end{aligned} \quad (9)$$

This optimization problem is not convex, and a gradient-based approach appropriate for constrained global optimization problems needs to be adopted. This is accomplished through an approximate multistart approach that addressed both the potential existence of multiple local minima as well as challenges associated with identifying feasible starting points for the gradient-based approach. Initially, a large value n_{init} of trial solutions for θ is examined, then the solutions corresponding to lower values of $F_2(\theta | \mathbf{z})$ while satisfying constraint ϵ^r for $F_1(\theta | \mathbf{z})$ are taken as initial points for a gradient-based optimization. The latter is achieved through a SOL solver implemented through the TOMLAB optimization environment.⁴⁵ The same candidate solutions

are used for all values of ϵ^r (no need to repeat this step). Evaluation of $F_1(\boldsymbol{\theta} | \mathbf{z})$ over the n_{init} candidate solutions is vectorized, so has small burden, whereas the efficiency of the gradient-based optimization is improved by obtaining analytically the gradients for both objectives. For objective F_1 , the components of the gradient vector are

$$\frac{\partial F_1}{\partial \theta_i} = \frac{1}{n_y} \sum_{i=1}^{n_y} \gamma_i^2 \left[2(Y_i(\mathbf{z}) - \theta_s \hat{s}_i^m(\mathbf{x})) \frac{\partial \theta_s \hat{s}_i^m(\mathbf{x})}{\partial \theta_i} + \frac{\partial \theta_s^2 \sigma_i^2(\mathbf{x})}{\partial \theta_i} \right], \quad (10)$$

where the partial derivatives inside the brackets can be readily obtained through the metamodel.²⁶ For the specific objective function F_2 that will be used in the illustrative example later, given by Equation 5, the components of the gradient vector are calculated by the chain rule as

$$\frac{\partial F_2}{\partial \theta_i} = \frac{\partial F_2}{\partial v_i} \frac{\partial v_i}{\partial \theta_i} = \frac{\partial F_2}{\partial v_i} \frac{F_{\theta_i}'(\theta_i)}{\varphi(\Phi^{-1}(F_{\theta_i}(\theta_i)))}, \quad (11)$$

where $\varphi(\cdot)$ stands for the standard Gaussian pdf, the partial derivative $\partial v_i / \partial \theta_i$ was calculated by differentiating Equation A6, and the partial derivatives $\partial F_2 / \partial v_i$ correspond to the components of the gradient row vector

$$\nabla_v F_2 = 2[\mathbf{v}(\boldsymbol{\theta}) - \boldsymbol{\mu}(\mathbf{z})]^T \boldsymbol{\Sigma}^{-1}. \quad (12)$$

The challenge for this optimization approach is that a gradient-based step needs to be repeated multiple times (for each different value of ϵ^r) and cannot leverage vectorized calculations for the metamodel because the metamodel is separately utilized for each sequential objective function evaluation. The advantage is that gradient information can improve computational efficiency and that the optimization can be performed for specific values of ϵ^r . This allows the identification of a specific part for the Pareto front if desired, for example, the front that corresponds to specific levels of compatibility to the chosen IMs.

Once the Pareto front has been identified, a dominant solution can be adopted using any desired criterion, for example, the solution that provides a specific compromise between the two objective functions. 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 Kiureghian¹⁵ and reviewed in Appendix A. For the target IMs, GMPEs used in the Western US are considered here,^{9,46-48} whereas the suggestions by Kaklamanos et al⁴⁹ were adopted to estimate unknown inputs for some of the GMPEs. As target, IM predictions from individual GMPEs as well as the average of their predictions will be adopted later. Note that the latter still provides a single target IM for each structural period examined. All computations are performed in a quad-core 3.0 GHz Xeon processor with 16 GB of RAM.

5.1 | Details for metamodel development

The box-bounded domain X^d was determined based on the ranges reported by Rezaeian and Der Kiureghian¹⁵ as [5 45] s for D_{5-95} , [0.5 40] s for t_{mid} , [0.1 30] Hz for $\omega_{\text{mid}}/2\pi$, [-2, 0.5] Hz/s for $\omega'/2\pi$, and [0.02 0.99] for ζ_f . For the response output, the peak pseudo-acceleration ($Y = \text{PSA}$) for a SDoF system with 5% damping ratio and for 22 different periods, the ones used by the aforementioned GMPEs, is adopted. The white noise samples are chosen as $n_w = 200$. For the domain X characterization $n_c = 1000$ seismicity scenarios are considered in range [6 8] for M , [10 100] km for R , [300 1600] m/s for $V_{s,30}$ and discrete {0,1} for F . These ranges correspond to the ones for which the initial predictive relationships¹⁵ were developed. The samples for the SVM-based characterization of X are chosen as $n_s = 10000$ with $\delta_d = 9$ (latter corresponding to radius of 3 standard deviations away from mean). The adaptive DoE discussed in Section 3 is implemented with $n_1 = 600$ and $n_a = 300$. Three different accuracy criteria are selected, with associated coefficient of determination (averaged over all outputs) 0.95, 0.97, and 0.985. This leads to number of support points 1500, 3000, and 4500, respectively.

For generating a total of 900,000 time histories and performing the required 19,800,000 simulations to develop the database for the metamodel for the elastic responses, close to 600 CPU hours were required. Although this computational burden is significant, it should be stressed that it corresponds to an initial only overhead of the approach. Once the metamodel is developed, it can be then used for any required predictions since the established accuracy is high. This large burden should be also attributed to the large number of white noise sequences (200), periods (22), and the wide seismicity range examined. The former provides high accuracy for the calculation of the relevant statistics, whereas the latter two support a wide applicability of the developed metamodel, as it can provide accurate predictions for all the responses of interest for the considered stochastic ground motion model (covers its range of applicability) and GMPEs (covers all the periods addressed). The burden for a metamodel that considered smaller number of stochastic sequences or constrained seismicity or period ranges would be drastically decreased.

Estimation of metamodel response for 10000 samples requires 4.9 s, 8.8 s, and 14.05 s for the metamodels with 1500, 3000, and 4500 support points, respectively. When the metamodel prediction error variance is not computed, the corresponding times are 3.6 s, 7.5 s, and 10.7 s, respectively. Note that adoption of larger value of samples prohibits efficient vectorization of operations for the $n = 4500$ points because of memory restrictions. The calculation of objective function F_1 along with its gradient requires 0.031 s, 0.052 s, and 0.091 s for the metamodels with 1500, 3000, and 4500 support points, respectively. When the metamodel prediction error variance is not included in the calculation, the corresponding times are 0.008 s, 0.0107 s, and 0.015 s, respectively. Comparison of these computational times shows that (i) increase of the support points has a considerable effect on computational efficiency, (ii) vectorization of calculations provides significant benefits, and (iii) inclusion of the prediction error variance in the calculations increases the computational burden, especially when gradient information needs to be obtained. All these aspects should be taken into account when choosing computational details for the optimization problem.

5.2 | Comparison of optimization approaches

The focus is first placed on the numerical solution of the optimization problem. The target used in this subsection, and also in the next one 5.3, corresponds to structural periods $T_s = [0.4 \ 0.5 \ 0.75 \ 1.0 \ 1.5 \ 2.0]$ s and IM described by the average of the considered GMPEs.^{9,46-48} Weights are chosen as $\gamma_i = 1/Y_i(\mathbf{z})$, so that objective function is expressed in terms of the relative error. In other words, no specific structural period is prioritized in evaluating the match to the target IMs. Three reference seismicity scenarios are examined in this section, corresponding to $M = 6-R = 20$ km, $M = 7.8-R = 30$ km, and $M = 7-R = 40$ km, for a strike-slip fault ($F = 0$) and $V_{s,30} = 800$ m/s. The first two scenarios correspond to cases where the unmodified stochastic ground motion model does not provide an adequate match to the target GMPEs¹⁵ (significantly over predicts for the former, moderately under predicts for the latter) and the last to a case where the unmodified model facilitates a good match. Because of space limitations, discussion will focus around specific cases of interest. The metamodel with 4500 support points is used in this section. Comparison across metamodels with different number of support points (and therefore accuracy) will be discussed in the next section.

Figure 2 presents illustrative swarms of candidate solutions in the objective space from the exhaustive search using $n_{bc} = 50000$ samples for seismicity $M = 6-R = 20$ km and $M = 7-R = 40$ km, in the former case examining the case with and without metamodel error in the calculation of objective F_1 . For the second objective, results are reported with respect to $\sqrt{F_2}$, which corresponds to the distance between the initial and the modified predictive relationships (not the squared distance)

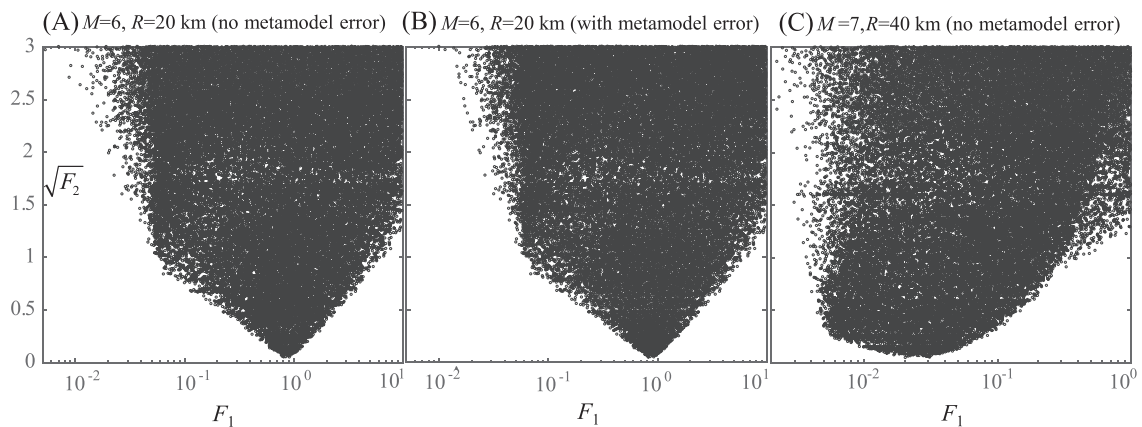


FIGURE 2 Projection to the objective space of swarm of 50000 candidate solutions obtained using an exhaustive search for A,B, $M = 6-R = 20$ km and C, $M = 7-R = 40$ km. Objective F_1 is calculated without the metamodel error for A, and C, and with the metamodel error for B

and offers a better normalization for the results in the comparison. The solutions located at the left boundary of the swarms in Figure 2 correspond ultimately to Pareto optimal solutions since there is no other solution that can simultaneously improve both performance objectives. It is evident from these swarms that modification of the predictive relationships can indeed facilitate a difference in the IM match (check the range of F_1 values obtained), whereas for smaller F_1 values, the candidate solutions deviate more from the rest of the swarm. This means, ultimately, that there are fewer model configurations that can provide a good match to the target IMs (small F_1 values). When the unmodified model is closer to the target IMs ($M = 7-R = 40$ km seismicity scenario), higher compatibility to these IMs can be obtained through the modification of the predictive relationships [compare case (C) to the other two], whereas addition of the metamodel error [compare case (B) to (A)] increases function F_1 , especially for smaller F_1 values.

Figure 3 examines different approaches for the solution of the multi-objective problem for the reference seismicity case of $M = 6-R = 20$ km (similar trends hold for the other cases). Part (A) shows the Pareto fronts identified through the exhaustive search using three different n_{bc} values. Results are reported herein with respect to $\sqrt{F_1}$ and $\sqrt{F_2}$ because this facilitates an easier comparison (differences of extreme values easier to discern). Only 10 representative solutions are shown, and not the entire front. It is evident that minor differences only exist between the identified fronts for different n_{bc} values, and such differences occur primarily for small F_1 -large F_2 combinations. This comparison shows that a value of n_{bc} around 200,000 to 400,000 should be considered as sufficient for efficiently identifying the front. For the remaining of the manuscript, results for the exhaustive search will be presented for a value of n_{bc} equal to 400,000. Figure 3B then compares the Pareto fronts obtained by the exhaustive search of the gradient-based (epsilon constraint) approaches for the case that the metamodel error is considered or not in the objective function F_1 . For the epsilon constraint, n_{init} is taken as 10000, whereas the gradient-based optimization is performed using a couple only different initial points. For small F_1 constraints, some challenges were encountered in converging to an admissible solution. This should be attributed to the trend identified in Figure 2; for such values, a smaller number of candidate solutions exist that can satisfy the constraint, and a potential increase in n_{init} might be needed to identify feasible initial points for the gradient-based optimization. The comparison in Figure 3B shows that the two approaches identify similar Pareto fronts, with the gradient-based optimization converging to suboptimal solutions for lower F_1 values, evidently because of the existence of local minima with greater differences in achieved performance. For such performance ranges (ie, with an already good match to the target IM, as indicated by the lower F_1 value), these differences might be unimportant. With respect to the computational burden, the exhaustive search requires 15 s per 10000 candidate solutions examined (11 s if metamodel error is disregarded). For the epsilon constraint, the computational cost is 15 s for the initial 10000 trials and 20 s (2 s if metamodel error is disregarded) for each different constraint examined. These comparisons show that both optimization approaches may be considered as adequate and preferred, depending on the application context. Overall, some preference exists for the exhaustive search due to the fact that epsilon constraint method needs identification of an appropriate starting point to avoid convergence to suboptimal local minima. The epsilon constraint approach, though, might be beneficial when a single solution is sought after, the one that satisfies a desired match to the target IM, rather than the entire front. Consideration or not of the metamodel error has no effect on the differences between the optimization approaches. The computational efficiency for the gradient-based search is reduced, as discussed in the previous section, when this error is included. Table 1 provides summary of these results, including some details discussed in the next section.

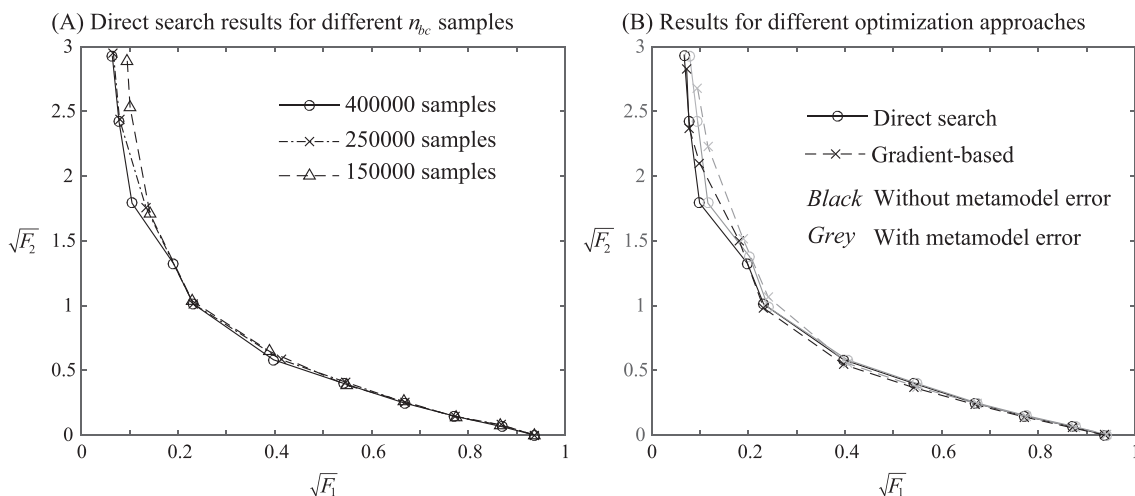
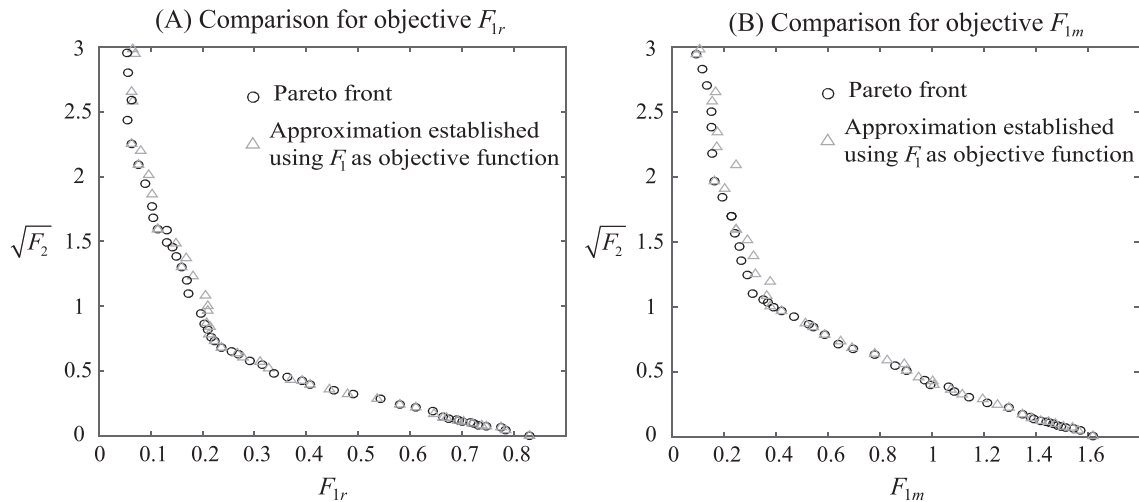


FIGURE 3 A, Pareto fronts identified by exhaustive search for different n_{bc} values. B, Comparison of Pareto fronts obtained by exhaustive search and epsilon constraint approach, considering or not the metamodel error. Case presented corresponds to seismicity scenario $M = 6-R = 20$ km

TABLE 1 Summary of pros/cons characteristics for the different implementation cases

	Pros	Cons
Optimization algorithms		
Gradient-free exhaustive search	Can leverage vectorized metamodel predictions for increased computational efficiency	Large number of evaluations required and will identify always the entire front. Number of samples used in exhaustive search has (small) effect on solutions
Gradient-based	Very efficient when a single only solution is sought, rather than entire front. Allows identification of a specific part of the Pareto front	Requires identification of an appropriate starting point to avoid convergence to local minima. Greater computational burden when metamodel error is considered
Metamodel characteristics		
Larger number of support points	Higher accuracy in predictions established. Reduces necessity to include metamodel error for obtaining high quality solutions	Increased computational burden. Remedied by the fact that metamodel error does not need to be considered
Inclusion of metamodel error	Facilitates greater robustness, avoiding identification of erroneous points	Considerably increased computational burden especially when combined with gradient-based optimization algorithms

**FIGURE 4** Comparison between exact and approximate Pareto fronts for objective functions A, F_{1r} , and B, F_{1m} describing the discrepancy from the target IMs. Case presented corresponds to seismicity scenario $M = 6-R = 20$ km

Finally, Figure 4 presents a comparison between the alternative objective function selections for quantifying the discrepancy from the target IM, ie, comparison between F_1 , F_{1r} , and F_{1m} . Figure 4A discusses F_{1r} and Figure 4B F_{1m} . The exact Pareto front, ie, optimization using F_{1r} (or F_{1m}) and F_2 as objectives, as well as an approximate front are compared. The approximate front is obtained by using F_1 and F_2 as objectives, identifying the Pareto set for θ and then evaluating objective F_{1r} (or F_{1m}) over that Pareto set. As discussed earlier, this optimization is less challenging but evidently identifies a sub-optimal solution. The results in Figure 4 show that there is overall very good correlation between the different objective functions and that the approximate solutions have only small deviations from the optimal front and those only for small values of the first objective. Therefore, use of F_1 as objective may be considered as an adequate surrogate even when the interest is in objectives F_{1r} or F_{1m} for quantifying discrepancy from the target IMs.

5.3 | Impact of metamodel accuracy

The discussion moves next to the examination of the effect of the metamodel accuracy. This is established by considering additionally the results obtained by using the exact stochastic ground motion model (ie, not relying on metamodel predictions), which represents the measure for evaluating the actual hazard compatibility of the identified ground motion model. The

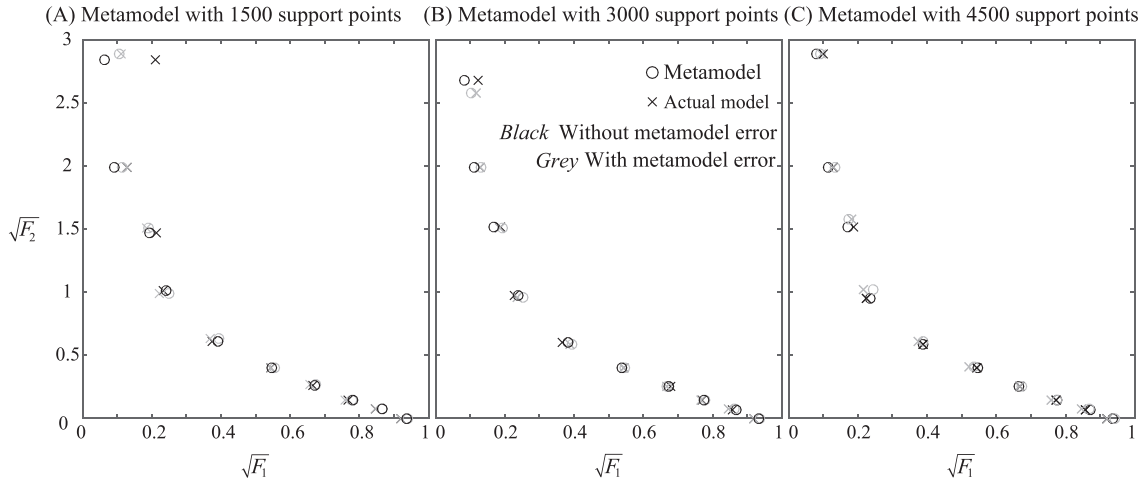


FIGURE 5 Pareto fronts identified using metamodels with A, 1500, B, 3000, or C, 4500 support points and comparison to predictions by exact stochastic ground motion model. Case presented corresponds to seismicity scenario $M = 6-R = 20$ km

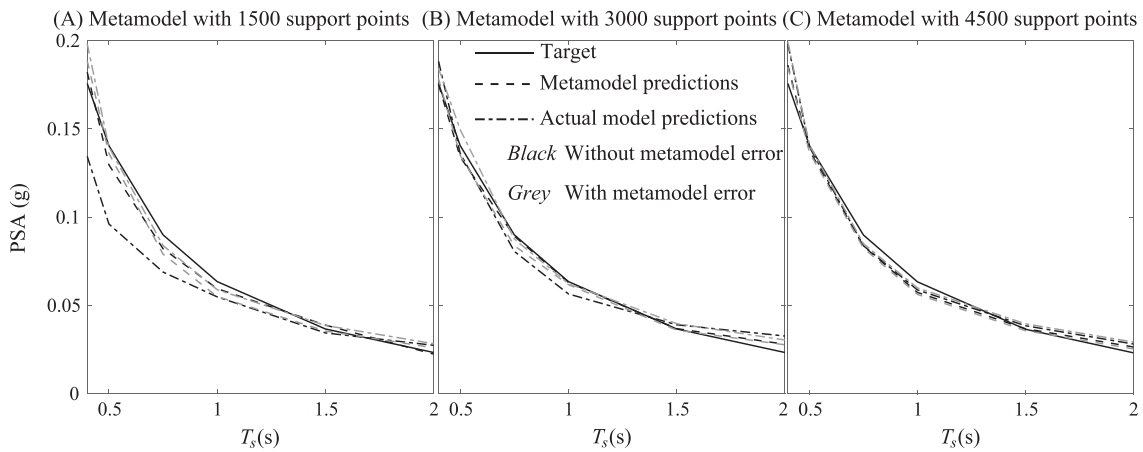


FIGURE 6 Spectral plots for the solutions corresponding to minimum of F_1 in the Pareto fronts identified in Figure 5

details for the study are the ones used in the previous section. Results for seismicity scenario $M = 6-R = 20$ km are presented in Figures 5 and 6. Figure 5 shows the Pareto fronts identified by using the metamodels with the three different number of support points. Cases with or without the metamodel error in the estimation of objective function F_1 are separately shown. This leads to two different Pareto sets, one without error and one with error, and for each two different fronts are reported, one corresponding to metamodel predictions and one to the use of the exact stochastic ground motion model. Then Figure 6 shows the spectral plot comparisons for the solution (among the Pareto set identified in each case) corresponding to the minimum of F_1 . The period range used in Figure 6 corresponds to the target structural periods T_s . In all cases, the exhaustive search is implemented with the same candidate solutions to facilitate a consistency in the corresponding comparisons. Figure 7 then shows Pareto front results for different seismicity scenario, $M = 7.8-R = 30$ km. Note that for seismicity scenario, $M = 7-R = 40$ km results (another case discussed in the previous section) are of limited interest since the unmodified ground motion model provides a good compatibility to target IMs.

The results show that for the higher accuracy metamodel (4500 support points), good agreement is established between the metamodel predictions and the actual model predictions along the Pareto front, whereas the inclusion of the metamodel error has only a small effect on the identified Pareto front. This is observed in both Pareto fronts (Figures 5 and 7) as well as in the corresponding spectral plots (Figure 6). Different trends are observed, though, for the lower accuracy metamodel (1500 support points). For lower F_1 values, and therefore large F_2 values, the Pareto optimal solutions identified when metamodel error is not included in the problem formulation lead to erroneously modified ground motion models. Based on the metamodel predictions, these models provide a very good match to the target IMs (small predicted F_1 value), but when response is evaluated with the exact model, larger differences are observed from the target IMs. This is particularly evident in the spectral plots shown in Figure 6. Evidently, the respective solutions identified correspond to parameters θ for which the metamodel accuracy is low.

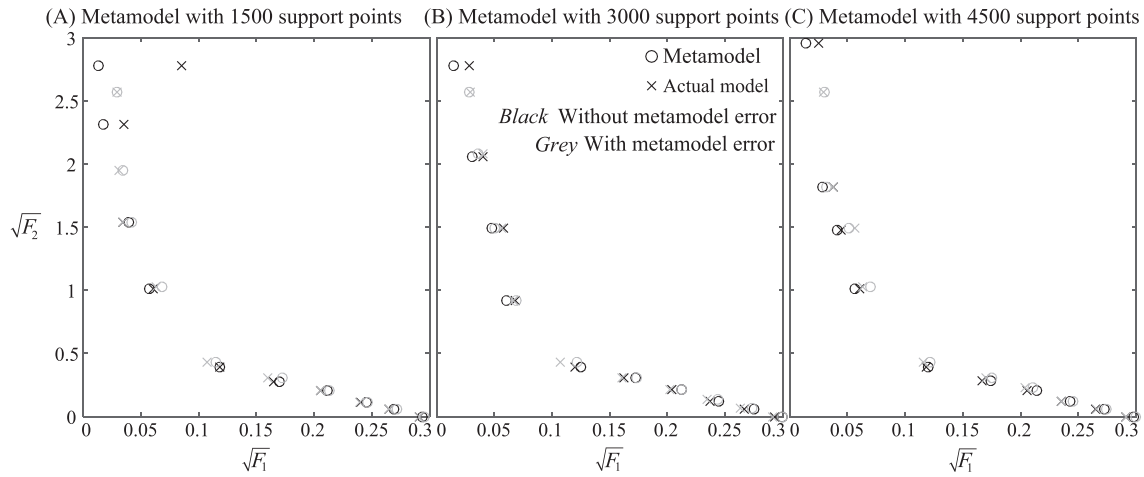


FIGURE 7 Pareto fronts identified using metamodels with A, 1500, B, 3000, or C, 4500 support points and comparison to predictions by exact stochastic ground motion model. Case presented corresponds to seismicity scenario $M = 7.8-R = 30$ km

The moderate accuracy metamodel (3000 support points) falls in between the two aforementioned cases, with characteristics that resemble more closely the ones for the high accuracy metamodel.

Note that the seismicity case examined here is ideal for exploring vulnerabilities in the optimization associated with lower metamodel accuracy, as it corresponds to a case at the boundary of the scenarios used to define domain X . Therefore, for larger values of F_2 , the corresponding parameters θ are expected to be close to the boundary of X , where metamodel accuracy is lower. However, even for this challenging case, metamodels with higher accuracy (3000 or 4500 support points) face small challenges, whereas the inclusion of the metamodel error for the lower accuracy metamodel (1500 support points) greatly improves the robustness of the optimization, leading to identified solutions with good agreement between metamodel and actual model. Figure 8 sheds further light into this topic. It focuses on the case examined in Figure 6A, but besides the mean metamodel predictions, it includes the predictions that are 1.5 standard deviations (σ) from the mean based on the estimated error variance. The solution identified when metamodel error is not included in the evaluation of F_1 is associated with a larger anticipated error (Figure 8A). Note that the actual model is actually even further away than the plotted 1.5σ . When the metamodel error is included in the evaluation of F_1 , such θ values with large associated σ are avoided because the large σ contributes to larger values for the objective function. This ultimately contributes to identification of solutions with greater robustness, ie, better agreement between metamodel and actual model (Figure 8B).

Overall, the above discussion shows that metamodels with higher accuracy (coefficient of determination 98%) can be considered as a good surrogate for the proposed optimization, whereas the inclusion of the prediction error greatly improves the

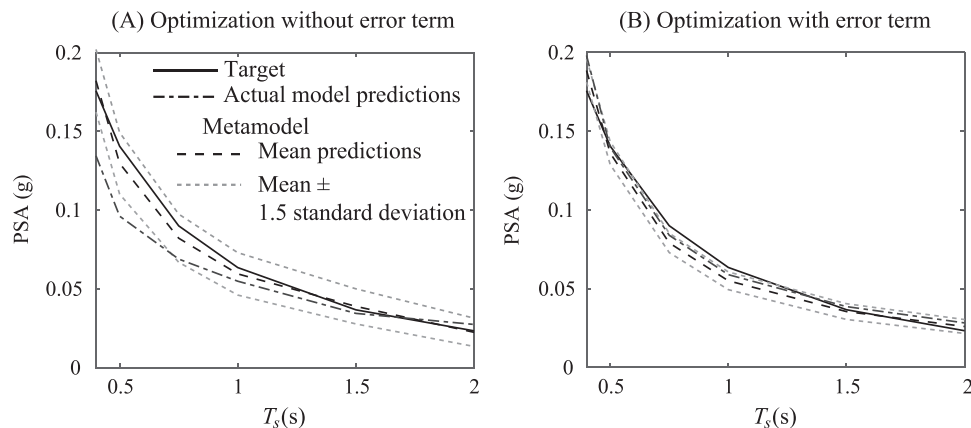


FIGURE 8 Spectral plots for the solutions corresponding to minimum of F_1 in the Pareto fronts identified in Figure 5 for the metamodel with 1500 support points. For the metamodel predictions, the mean predictions and the predictions within 1.5 standard deviations from the mean are shown

robustness of this optimization, avoiding identification of erroneous solutions, even when metamodells with lower accuracy are adopted. Consideration of this error is not necessary, though, when higher accuracy metamodells are used.

5.4 | Implementation for different seismicity scenarios

With the computational details ironed out, the discussion moves finally to the IM compatibility established by the proposed modification of the ground motion model. Figure 9 shows results for three seismicity scenarios targeting PSA given by the average of the aforementioned GMPEs for two different ranges for T_s : $T_s = [0.4 \ 0.5 \ 0.75 \ 1.0 \ 1.5 \ 2.0]$ s and $T_s = [0.4 \ 0.5 \ 0.75]$ s. These two different cases are referenced herein as long and short, respectively, period ranges. The proposed approach identified 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 IM. Choosing a shorter period range for this target facilitates an overall better match; this is anticipated because objective F_1 imposes less strict requirements in terms of IM compatibility (fewer number of components to match).

The question finally arises which point should be selected within the identified Pareto set. Various approaches have been proposed in the greater multi-objective optimization literature for making this choice.⁵⁰ Perhaps the most common one is to select the solution that has the smallest normalized distance from the utopia point, defined as the point in the Pareto front (ie, objective function space) that corresponds to the minimum of the two objectives across the front. This utopia point represents the best but unachievable performance $[\min F_1 \ \min F_2]$. Normalization is typically established with respect to the maximum of each performance objective across the Pareto front, leading to distance metric

$$D_p(\boldsymbol{\theta}) = \sqrt{\sum_{i=1,2} \left(\frac{F_i(\boldsymbol{\theta}|\mathbf{z}) - \min F_i}{\max F_i - \min F_i} \right)^2}; \max F_i = \max_{\{\boldsymbol{\theta}_p; i=1, \dots, n_p\}} (F_i), \min F_i = \min_{\{\boldsymbol{\theta}_p; i=1, \dots, n_p\}} (F_i). \quad (13)$$

Instead of F_1 and F_2 , this can be implemented for $\sqrt{F_1}$ and $\sqrt{F_2}$ due to the better normalization properties. This point is identified in all cases in Figure 9. Another choice would have been to choose the solution that satisfies a predetermined threshold for the match to the targeted IMs. Perhaps this is better set with respect to objectives F_{1r} or F_{1m} rather than F_1 .

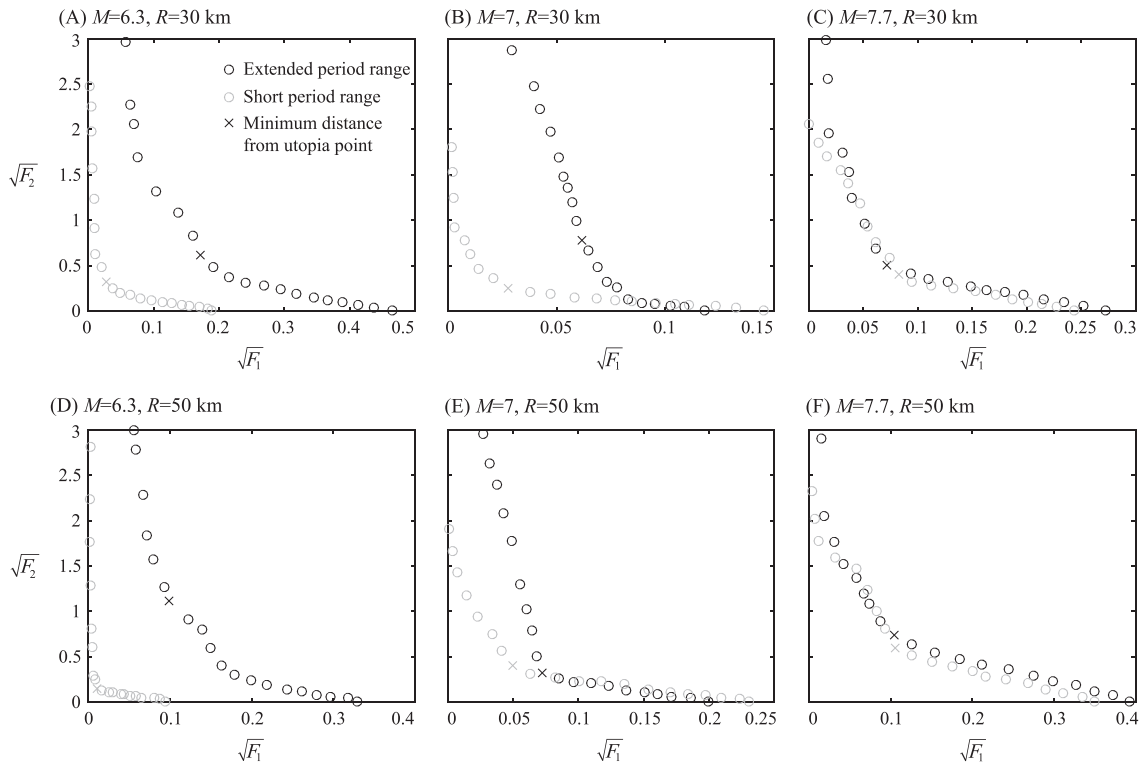


FIGURE 9 Pareto fronts for different seismicity cases considering match to long (black) or short (gray) period range IMs. In each plot the Pareto point with minimum distance from utopia point is shown with x

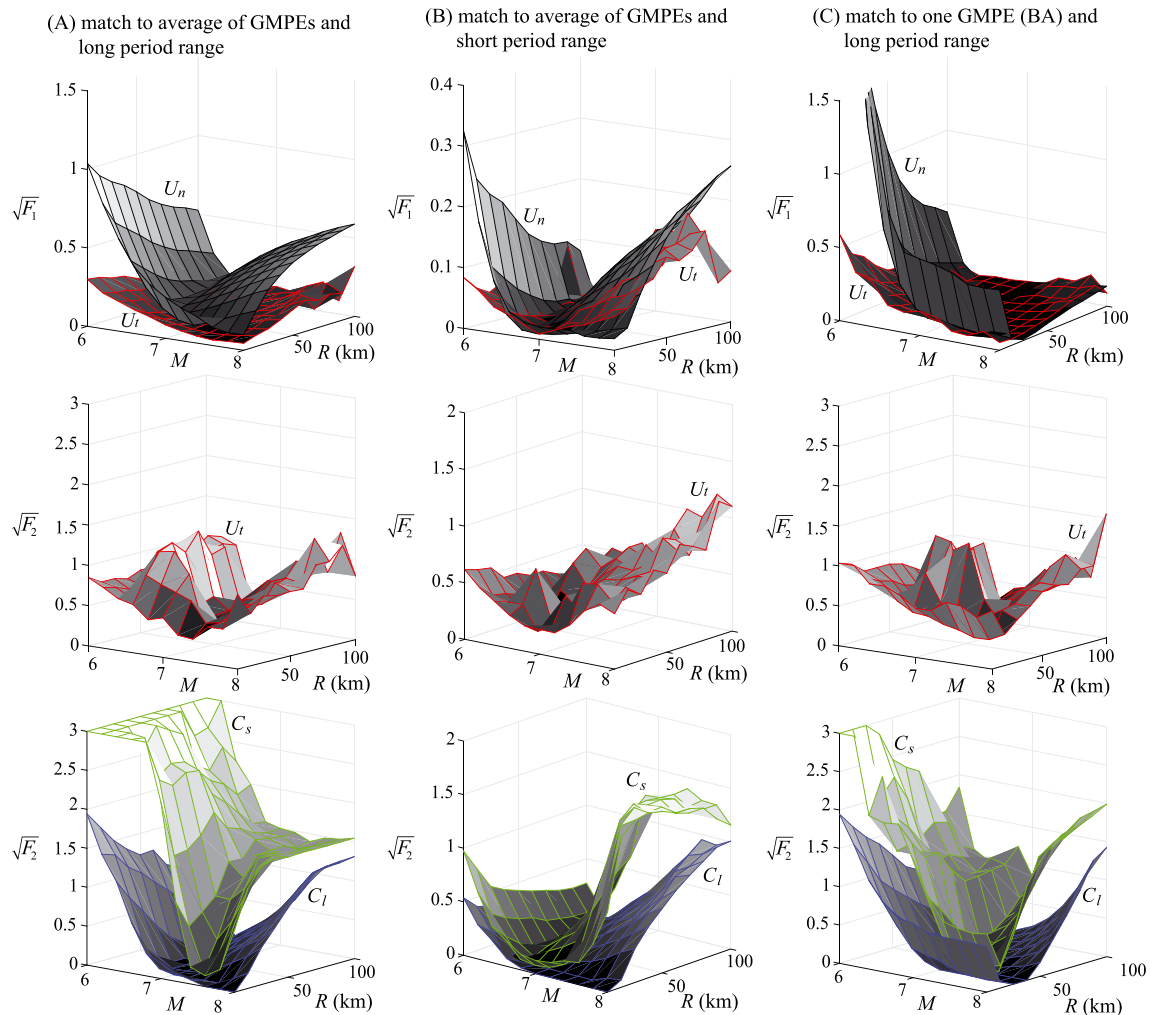


FIGURE 10 Results for $\sqrt{F_1}$ and $\sqrt{F_2}$ for unmodified ground motion model (U_n) and modified ground motion model corresponding to three different selection criteria: Minimum distance form utopia point (U_t) and value $\sqrt{F_1}$ smaller than 0.15 (C_t) or 0.05 (C_s). Implementations in 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 GMPE⁴⁷ [Colour figure can be viewed at wileyonlinelibrary.com]

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 10–12. 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 average absolute relative error $\sqrt{F_1}$ smaller than 0.15 or 0.05. These three cases are denoted, U_t , C_t , and C_s , respectively. The thresholds for C_t 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 10 shows plots for (i, first row) $\sqrt{F_1}$ for U_t and U_n ($\sqrt{F_1}$ is constrained for the other two cases) and for (ii, second and third rows) $\sqrt{F_2}$ for U_t , C_t , and C_s ($\sqrt{F_2}$ is zero for U_n). To better demonstrate the differences, results are presented separately for U_t (second row) and for the pair C_t and C_s (third row) in the latter case. The three different columns in the figure correspond to three different implementation cases: target IM given by the average of the aforementioned 4 GMPEs for both (A) the long and (B) the short period ranges as well as (C) target IM given only by one only GMPE⁴⁷ for the long period range. These scenarios are denoted herein as SC_1 , SC_2 , and SC_3 , respectively. Figure 11 shows spectral plots for a smaller (due to space limitations) 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, the curves corresponding to the target IM, the unmodified model, and the predictions by the three aforementioned model modifications are shown to facilitate comparisons. Finally, Figure 12 shows for all examined seismicity scenarios the model parameters θ for the unmodified ground motion model as well as for the modified model corresponding to the Pareto point with smallest distance from the utopia point (U_t case) for SC_1 . Note that some of the curves shown in these figures have nonsmooth characteristics. This should be attributed to multiple facts: (i) a discrete representation of the Pareto front is identified, rather than the actual Pareto front; (ii) problem has multiple local minima as discussed earlier, especially for larger F_2

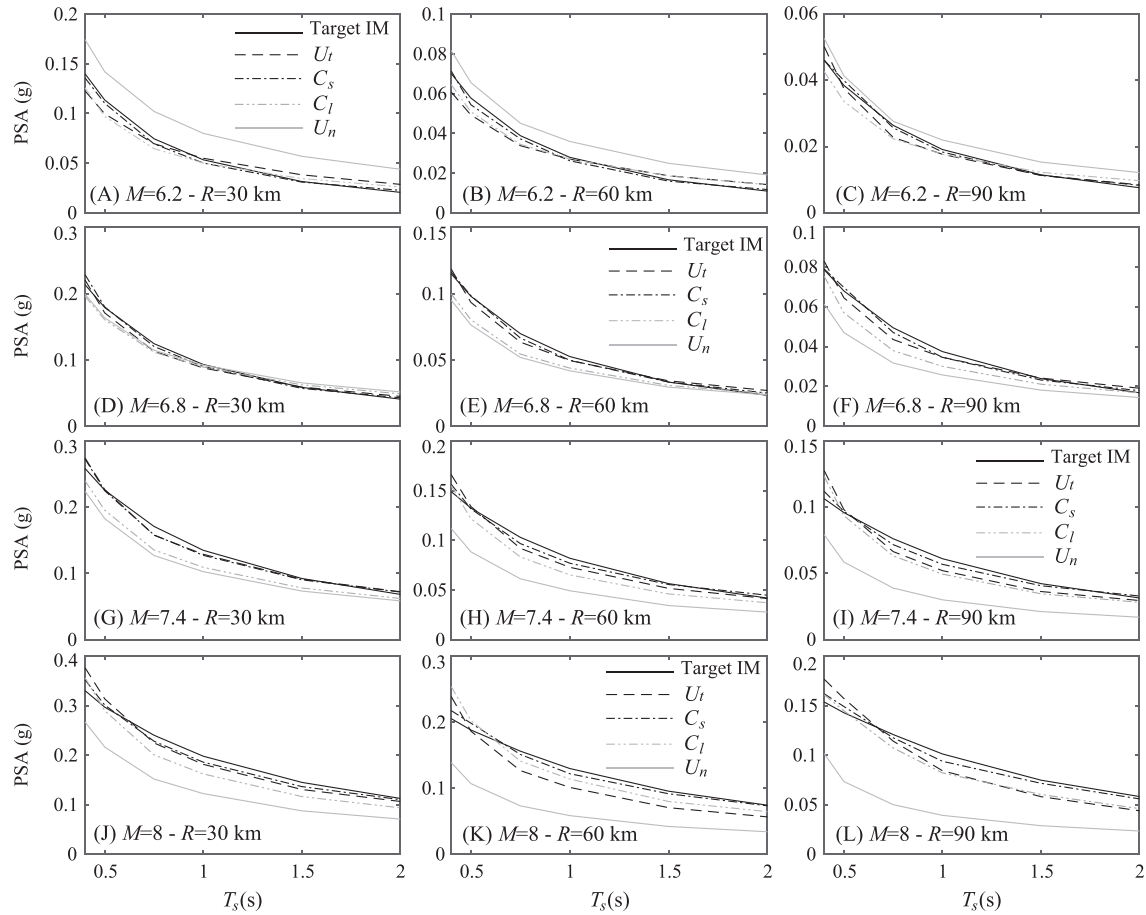


FIGURE 11 Spectral plots for seismicity scenarios (different subplots) corresponding to combinations of M [6.2, 6.8, 7.4, 8] and R [30, 60, 90] km, for the target IM (target), the unmodified ground motion model (U_n) and modified ground motion model corresponding to three different selection criteria: minimum distance from utopia point (U_t) and value $\sqrt{F_1}$ smaller than 0.15 (C_l) or 0.05 (C_s). Implementation scenario shown corresponds to matching to the average considered GMPEs and long period range

values; and (iii) algorithms used for the optimization have characteristics of stochastic search, well known to lead to discontinuous results.

The unmodified model U_n does not provide a good match to the target IMs for the entire seismicity range, with trends observed in first row of Figure 10 and in the spectral plots in Figure 11 being similar to the ones reported by Rezaeian and Der Kiureghian,¹⁵ ie, greater challenges in lower moment magnitudes and combination of higher magnitudes and larger distances. Figure 10 shows that the proposed modification (cases U_t , C_l and C_s) improves this match, establishing a balance between F_1 and F_2 , with the characteristics of the balance depending on the criteria for selection of the final model among the Pareto optimal solutions (ie, which specific case is chosen). When the unmodified model has larger discrepancies from the target IMs, then the modifications lead to larger values for F_2 , but still successfully identify models, independent of the implementation case, that provide an improved match to the IM target. The U_t modification identifies a model with moderate discrepancy from the unmodified one, corresponding to values of F_2 in the range of 0.3–1, 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 IMs and therefore modification of it provides limited advantages. This is perhaps better captured by the C_l case. This discussion shows that selection of the Pareto optimal model based on a targeted accuracy to the GMPEs, ie, value for F_1 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. However when that threshold is selected small (C_s case) and the unmodified model has larger discrepancy from the target IM, then the modification leads to identification of a model with big differences from the original one (large F_2 values). This model will typically be far away from the U_t case and will belong in a steep part of the Pareto front, meaning that small improvements in F_1 come at a large increase of F_2 (check the Pareto fronts shown in previous figures). A multilevel selection criterion seems therefore more appropriate: select the Pareto optimal solution that satisfies a certain accuracy threshold

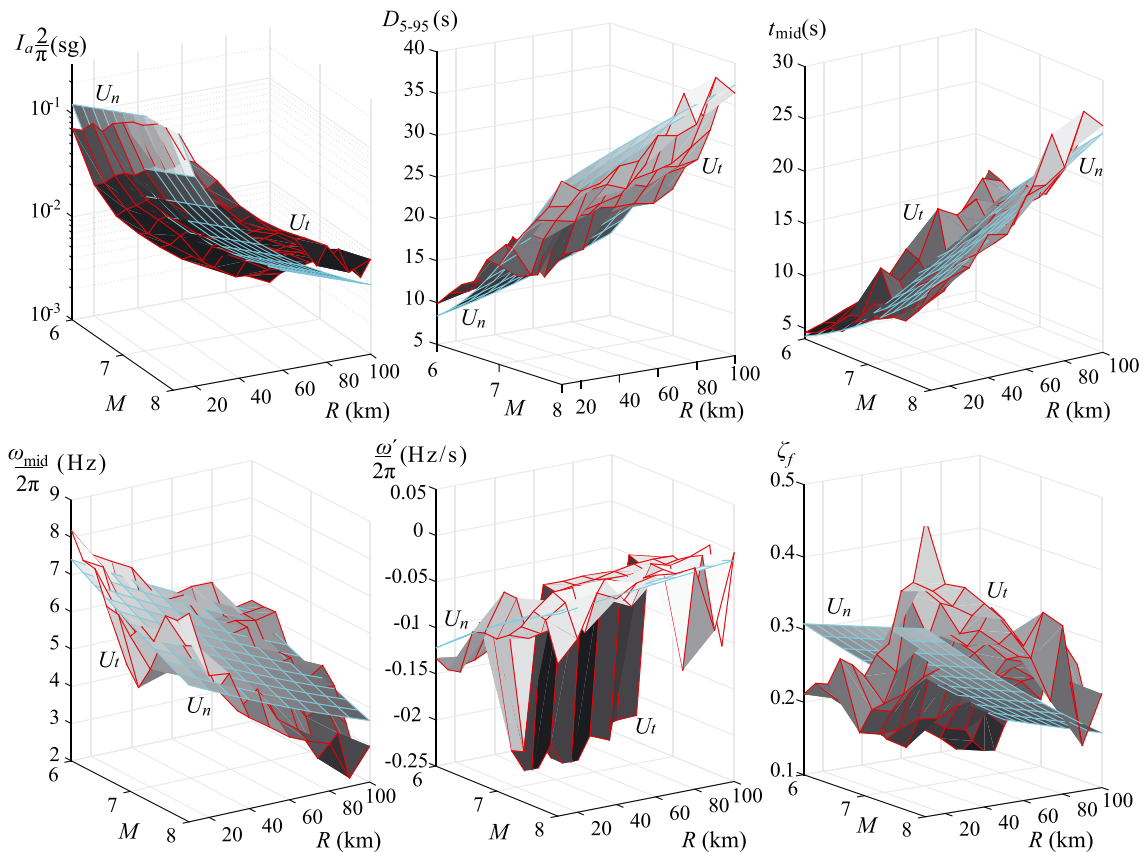


FIGURE 12 Physical ground motion model parameters θ corresponding to unmodified ground motion model (U_n) and modified ground motion model with minimum distance from utopia point (U_t) for implementation scenario of matching to the average considered GMPEs and long period range [Colour figure can be viewed at wileyonlinelibrary.com]

for F_1 unless this solution leads to a greater F_2 value than the Pareto optimal solution with minimum distance from the utopia point. If the latter happens, then select the Pareto optimal solution with minimum distance from the utopia point.

With respect to the different implementation cases shown in Figure 10, selection of a shorter period range [compare cases (B) and (A)] promotes, as also identified earlier, an easier match to the target (smaller overall values for F_1 and F_2). This demonstrates the importance of carefully selecting the target IMs, as this selection affects the ability to match this target, and of the availability of a framework, as the one developed here, that allows you to do so. The exact selection of the target [compare cases (A) and (C)] does not impose any additional constraint in the optimization implementation, although the results evidently change.

The spectral plots in Figure 11 provide further validation for the observed accuracy of the different examined cases, showing the decomposition of the overall F_1 match to the different structural periods. The ground motion model modification greatly improves the match to the target spectral curves, with C_s modification always providing a better match compared to the C_l one. The U_t modification exhibits a varying behavior, frequently in between the C_l and C_s cases, in other instances outside their envelope, with characteristics that ultimately depend as discussed above (based on Figure 10) on the ability of the unmodified model to provide an adequate match to the target IMs (ie, how much is the modification truly needed). Of course the proper evaluation of the proposed modifications comes from examining both F_1 and F_2 values, as detailed above (discussion focusing on Figure 10).

Finally, the results for the ground motion model parameters in Figure 12 show that the model modification leads to similar trends as observed for the unmodified model. This is the direct results of incorporating the difference between these parameters as an objective in the problem formulation (objective F_2) rather than merely as a constraint. Parameters I_a , ζ_f and ω' show bigger variability compared with their initial values. This should be attributed to a greater sensitivity with respect to them of the resultant ground motions.

The overall discussion shows the importance of the established framework: once the initial metamodel is developed, through the adaptive guidelines established here, it can support the efficient identification of ground motion models that (i) match any desired IMs for any chosen period range while (ii) maintaining a small deviation from the initial predictive relationships. This

can be seamlessly repeated for any seismicity scenario. The final ground motion model modification can be chosen based on the criteria discussed earlier.

6 | CONCLUSIONS

The modification of stochastic ground motion models to match target IMs for specific seismicity scenarios was presented in this article. This was formulated as a multi-objective optimization problem with first objective (F_1) quantifying the discrepancy between the ground motion predictions and the target IMs, and the second objective (F_2) corresponding to the discrepancy between the new model characteristics and the model characteristics suggested by existing predictive relationships (ie, the unmodified model). The second objective explicitly incorporates in the modification process physical characteristics and parameter correlations described in the initial ground motion model. The developed framework facilitates a match to any desired IM or to a collection of them, for example, spectral accelerations over a period range, for any chosen seismicity scenario. Repeating process for different seismicity scenarios can then facilitate the development of a suite of models that can support comprehensive seismic risk assessment. Computational efficiency was achieved by adopting a metamodel for approximating the median ground motion model predictions for the targeted IMs. Although the upfront cost for development of this metamodel is significant, once established, it can be used to support a highly efficient multi-objective optimization. Gradient-based and gradient-free approaches were discussed for the latter, whereas to reduce the computational burden for the metamodel development, an adaptive design of experiments was proposed for selecting the database informing the metamodel.

The framework was demonstrated in an illustrative example considering a recently developed record-based stochastic ground motion model and IMs described through ground motion prediction equations. It was shown in the context of this example that the metamodel-aided optimization can support an accurate identification of the Pareto front of dominant solutions, provided that the metamodel accuracy is significant high, and that inclusion of the metamodel error in the optimization formulation greatly improves the robustness of this optimization, avoiding the identification of erroneous solutions. Comparisons between the two optimization approaches showed that the gradient-free one demonstrates overall preferable attributes because the gradient-based one might converge to suboptimal local minima, especially for lower F_1 values. However, the gradient-based one provides greater relative efficiency when identification of a single solution, rather than of the entire front, is desired. Still, an adequate representation of the overall Pareto front can be obtained in as little as 2 minutes using the blind-search, gradient-free optimization, which should be considered as an acceptable computational burden. Different approaches can be then used to select the final ground motion model, eg, the Pareto optimal point that has the minimum distance from the utopia point or the point that satisfies a desired compatibility to the target IMs (ie, value of F_1 smaller than a threshold). Implementation for a range of seismicity scenarios showcased the advantages of the proposed framework: small modifications of the original ground motion model (small to moderate F_2 values) provided significant improvement for the match to the target IM (F_1 value) for seismicity ranges where the unmodified model faces challenges in matching the target IMs. With respect to the selection of the final ground motion model, the following suggestion is provided after carefully examining the various trends observed: select the Pareto optimal solution that satisfies a certain accuracy threshold for F_1 unless this solution leads to a greater F_2 value than the Pareto optimal solution with minimum distance from the utopia point. If the latter happens, then select the Pareto optimal solution with minimum distance from the utopia point.

REFERENCES

1. Bozorgnia Y, Bertero V. *Earthquake Engineering: From Engineering Seismology to Performance-Based Engineering*. Boca Raton, FL: CRC Press; 2004.
2. 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.
3. 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.
4. Jensen HA, Catalan AA. On the effects of non-linear elements in the reliability-based optimal design of stochastic dynamical systems. *Int J Non Linear Mech*. 2007;42:802-816.
5. 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.
6. Katsanos EI, Sextos AG, Manolis GD. Selection of earthquake ground motion records: a state-of-the-art review from a structural engineering perspective. *Soil Dyn Earthq Eng*. 2010;30(4):157-169.
7. McGuire RK. *Seismic Hazard and Risk Analysis*. Oakland, CA: Earthquake Engineering Research Institute; 2004.

8. 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.
9. 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.
10. 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.
11. Galasso C, Zhong P, Zareian F, Iervolino I, Graves RW. Validation of ground-motion simulations for historical events using MDoF systems. *Earthq Eng Struct Dyn*. 2013;42(9):1395-1412.
12. Vetter C, Taflanidis A. Comparison of alternative stochastic ground motion models for seismic risk characterization. *Soil Dyn Earthq Eng*. 2014;58:48-65.
13. Broccardo M, Der Kiureghian A. Multicomponent nonlinear stochastic dynamic analysis by tail-equivalent linearization. *J Eng Mech*. 2015;142(3): p. 04015100
14. Chun J, Song J, Paulino GH. Structural topology optimization under constraints on instantaneous failure probability. *Earthq Eng Struct Dyn*. 2016;53(4):773-799.
15. Rezaeian S, Der Kiureghian A. Simulation of synthetic ground motions for specified earthquake and site characteristics. *Earthquake Engineering & Structural Dynamics*. 2010;39(10):1155-1180.
16. Gavin HP, Dickinson BW. Generation of uniform-hazard earthquake ground motions. *J Struct Eng ASCE*. 2011;137(3):423-432.
17. Yamamoto Y, Baker JW. Stochastic model for earthquake ground motion using wavelet packet. *Bull Seismol Soc Am*. 2013;103(6):3044-3056.
18. 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.
19. Boore DM. Simulation of ground motion using the stochastic method. *Pure Appl Geophys*. 2003;160:635-676.
20. Atkinson GM, Silva W. Stochastic modeling of California ground motions. *Bull Seismol Soc Am*. 2000;90(2):255-274.
21. 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.
22. Papadimitriou C, Beck J. Nonstationary stochastic characterization of strong-motion accelerograms. in *Proc. of 4th US National Conference on Earthquake Engineering*. 1990.
23. FEMA-P-58, *Seismic performance assessment of buildings*. 2012. American Technology Council: Redwood City, CA.
24. Petersen MD, Frankel AD, Harmsen SC, et al. Documentation for the 2008 update of the United States National Seismic Hazard Maps: U.S Geological Survey Open-File Report 2008-1128. 2008.
25. Scherbaum F, Cotton F, Staedtke H. The estimation of minimum-misfit stochastic models from empirical ground-motion prediction equations. *Bull Seismol Soc Am*. 2006;96(2):427-445.
26. 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.
27. Medel-Vera C, Ji T. A stochastic ground motion accelerogram model for Northwest Europe. *Soil Dyn Earthq Eng*. 2016;82:170-195.
28. Boore DM, Thompson EM. Revisions to some parameters used in stochastic-method simulations of ground motion. *Bull Seismol Soc Am*. 2015;105:1029-1041.
29. 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.
30. Bojórquez E, Iervolino I. Spectral shape proxies and nonlinear structural response. *Soil Dyn Earthq Eng*. 2011;31(7):996-1008.
31. Cosenza E, Manfredi G. The improvement of the seismic-resistant design for existing and new structures using damage criteria. In: *Seismic design methodologies for the next generation of codes*. Rotterdam Netherlands: Balkema; 1997:119-130.
32. Cordova PP, Deierlein GG, Mehanny SS, Cornell CA. Development of a two-parameter seismic intensity measure and probabilistic assessment procedure. in *The Second US-Japan Workshop on Performance-Based Earthquake Engineering Methodology for Reinforced Concrete Building Structures*. 2000.
33. Bozorgnia Y, Hachem MM, Campbell KW. Ground motion prediction equation ("attenuation relationship") for inelastic response spectra. *Earthq Spectra*. 2010;26(1):1-23.
34. Lophaven SN, Nielsen HB, Sondergaard J. DACE-A MATLAB Kriging Toolbox. 2002. Technical University of Denmark.
35. Sacks J, Welch WJ, Mitchell TJ, Wynn HP. Design and analysis of computer experiments. *Stat Sci*. 1989;4(4):409-435.
36. 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.
37. Dubourg V, Sudret B, Bourinet J-M. Reliability-based design optimization using kriging surrogates and subset simulation. *Struct Multidiscip Optim*. 2011;44(5):673-690.

38. 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.
39. Robert CP, Casella G. *Monte Carlo Statistical Methods*. 2nd ed. New York, NY: Springer; 2004.
40. Hartigan JA, Wong MA. Algorithm AS 136: A K-means clustering algorithm. *J R Stat Soc Ser C Appl Stat*. 1979;28(1):100-108.
41. Schölkopf B, Smola AJ. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge, Massachusetts: The MIT Press; 2002.
42. Marler RT, Arora JS. Survey of multi-objective optimization methods for engineering. *Struct Multidiscip Optim*. 2004;26:369-395.
43. Coello CAC, Van Veldhuizen DA, Lamont GB. *Evolutionary Algorithms for Solving Multi-objective Problems*. 2nd ed. New York: Springer; 2002.
44. Mavrotas G. Effective implementation of the ϵ -constraint method in multi-objective mathematical programming problems. *Appl Math Comput*. 2009;213(2):455-465.
45. Holmstrom K, Goran AO, Edvall MM. User's guide for TOMLAB 7. 2009. San Diego, CA: Tomlab Optimization Inc. www.tomopt.com/tomlab/.
46. Abrahamson N, Silva W. Summary of the Abrahamson & Silva NGA ground-motion relations. *Earthq Spectra*. 2008;24(1):67-97.
47. 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.
48. 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.
49. 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.
50. Keeney RL, Raiffa H. *Decisions with multiple Objectives: Preferences and Trade-offs*. Cambridge University Press; 1993.
51. Rezaeian S, Der Kiureghian A. A stochastic ground motion model with separable temporal and spectral nonstationarities. *Earthq Eng Struct Dyn*. 2008;37:1565-1584.

How to cite this article: Tsioulou A, Taflanidis AA, Galasso C. Modification of stochastic ground motion models for matching target intensity measures. *Earthquake Engng Struct Dyn*. 2017;1–22. <https://doi.org/10.1002/eqe.2933>

APPENDIX A

DETAILS FOR STOCHASTIC GROUND MOTION MODEL CONSIDERED IN THE STUDY

The specific ground motion model used in this study is the record-based model proposed by Rezaeian and Der Kiureghian¹⁵ that efficiently addresses both temporal and spectral nonstationarities. The unfiltered discretized time history of the ground motion according to this model is expressed as

$$\ddot{u}(t|\boldsymbol{\theta}, \mathbf{W}) = q(t, \boldsymbol{\theta}) \sum_{i=1}^k \frac{h[t-t_i, \boldsymbol{\theta}(t_i)]}{\sqrt{\sum_{j=1}^k h[t-t_j, \boldsymbol{\theta}(t_j)]^2}} w(i\Delta t) \quad k\Delta t < t < (k+1)\Delta t, \quad (\text{A1})$$

where $\mathbf{W} = [w(i\Delta t): i = 1, 2, \dots, N_T]$ is a Gaussian white noise sequence, Δt is the chosen discretization interval (assumed constant and equal to 0.005 s in this study), $q(t, \boldsymbol{\theta})$ is the time-modulating function, and $h[t-\tau, \boldsymbol{\theta}(\tau)]$ is an impulse response function corresponding to the pseudo-acceleration response of a single-degree-of-freedom linear oscillator with time-varying frequency $\omega_f(\tau)$ and damping ratio $\zeta_f(\tau)$, in which τ denotes the time of the pulse

$$h[t-\tau, \boldsymbol{\theta}(\tau)] = \frac{\omega_f(\tau)}{\sqrt{1-\zeta_f^2(\tau)}} \exp[-\omega_f(\tau)\zeta_f(\tau)(t-\tau)] \sin\left[\omega_f(\tau)\sqrt{1-\zeta_f^2(\tau)}(t-\tau)\right]; \quad \tau \leq t \quad (\text{A2})$$

$$= 0; \quad \text{otherwise.}$$

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

$$\omega_f(\tau) = \omega_{\text{mid}} + \omega'(\tau - t_{\text{mid}}) \quad \zeta_f(\tau) = \zeta_f, \quad (\text{A3})$$

with ω_{mid} (central frequency), ω' (frequency variation), and ζ_f ultimately corresponding to model parameters for the filter and t_{mid} corresponding to the mid-time of the strong motion duration (defined next). The time envelope $q(t, \boldsymbol{\theta})$ is given by¹⁵

$$q(t, I_a, \alpha_2, \alpha_3) = \sqrt{I_a} \left[\frac{\sqrt{(2\alpha_3)^{2\alpha_2-1}}}{\Gamma(2\alpha_2-1)} \right] t^{\alpha_2-1} \exp(-\alpha_3 t), \quad (\text{A4})$$

where $\Gamma(\cdot)$ is the gamma function, I_a is the Arias intensity expressed in terms of $g \cdot \pi/2$ (ie, scaled by that term), and $\{\alpha_2, \alpha_3\}$ are additional parameters controlling the shape and total duration of the envelope that can be related to physical parameters. As advocated by Rezaeian and Der Kiureghian,¹⁵ the strong motion duration D_{5-95} (defined as the duration for the Arias intensity to increase from 5% to 95% of its final value) and t_{mid} corresponding to the time Arias intensity achieves 45% of its final value are used. The pair $\{\alpha_2, \alpha_3\}$ can be then easily determined based on the values of $\{D_{5-95}, t_{\text{mid}}\}$.¹⁵ To assure zero residual velocity, the simulated process is eventually high-pass filtered.⁵¹ This filter corresponds to a critically damped oscillator, and the corrected acceleration record is obtained as the solution of the differential equation

$$\ddot{a}(t|\boldsymbol{\theta}, \mathbf{W}) + 2\omega_c \dot{a}(t|\boldsymbol{\theta}, \mathbf{W}) + \omega_c^2 a(t|\boldsymbol{\theta}, \mathbf{W}) = \ddot{q}(t|\boldsymbol{\theta}, \mathbf{W}); \quad \omega_c = 0.5\pi. \quad (\text{A5})$$

This filter has minimal effect on the response beyond the corner frequency, ω_c .⁵¹

The ground motion model has model parameters $\boldsymbol{\theta} = \{I_a, D_{5-95}, t_{\text{mid}}, \omega_{\text{mid}}, \omega', \zeta_f\}$, with the first one directly affecting (scaling) the output (thus input–output relationship is known) and the remaining 5 having a complex nonlinear relationship to that output. Predictive relationships have been established for $\boldsymbol{\theta}$ by fitting the stochastic model to a subset of the next generation attenuation (NGA) relationships strong motion database.¹⁵ These predictive relationships relate $\boldsymbol{\theta}$ to the following earthquake and site characteristics, defining seismicity vector \mathbf{z} : the moment magnitude, M , the rupture distance, R , the type of fault F [$F = 0$ denoting strike slip and $F = 1$ reverse fault] and the shear wave velocity of the top 30 m of the site soil, V_{s30} . The predictive relationships were established by first transforming the initial parameters in the standard Gaussian space leading to vector \mathbf{v} with components v_i , $i = 1, \dots, 6$, each of which is related to a component θ_i of vector $\boldsymbol{\theta}$ through

$$v_i(\theta_i) = \Phi^{-1}[F_{\theta_i}(\theta_i)]; \quad i = 1, \dots, 6 \quad (\text{A6})$$

where Φ corresponds to the standard Gaussian cumulative distribution function (CDF) and F_{θ_i} corresponds to the CDF for the fitted probability distribution to the i th component of vector $\boldsymbol{\theta}$,¹⁵ given in Table A1 where the two-sided exponential referenced in this table is

$$p_{\omega'/2\pi}(\omega'/2\pi) = \begin{cases} 4.85 \exp(6.77\omega'/2\pi) & -2 < \omega'/2\pi < 0 \\ 4.85 \exp(-17.10\omega'/2\pi) & 0 < \omega'/2\pi < 0.5 \\ 0 & \text{otherwise} \end{cases}. \quad (\text{A7})$$

The predictive relationships for \mathbf{v} are then obtained through regression analysis, providing a probabilistic characterization with mean vector $\boldsymbol{\mu}(\mathbf{z})$ having components¹⁵

$$\begin{aligned} \mu_1(\mathbf{z}) &= c_{1,0} + c_{1,1}F + c_{1,2} \left(\frac{M}{7} \right) + c_{1,3} \ln \left(\frac{R}{25\text{km}} \right) + c_{1,4} \ln \left(\frac{V_s}{750\text{m/s}} \right) \\ \mu_i(\mathbf{z}) &= c_{i,0} + c_{i,1}F + \beta_{i,2} \left(\frac{M}{7} \right) + c_{i,3} \left(\frac{R}{25\text{km}} \right) + c_{i,4} \ln \left(\frac{V_s}{750\text{m/s}} \right); \quad i = 2, \dots, 6 \end{aligned} \quad (\text{A8})$$

where $c_{i,j}$ is the regression coefficient provided in Table A2, and covariance matrix $\boldsymbol{\Sigma}$ corresponding to standard deviation σ_{v_i} and correlation coefficient ρ_{ij} for each component that are also shown in Table A.2.

TABLE A1 Fitted probability distributions

Parameter	Fitted distribution	Distribution bounds	Mean	Standard deviation
I_a (s g)	Lognormal	$(0, \infty)$	0.0468	0.164
D_{5-95} (s)	Beta	$5, 45$	17.3	9.31
t_{mid} (s)	Beta	$[0.5, 40]$	12.4	7.44
$\omega_{\text{mid}}/2\pi$ (Hz)	Gamma	$(0, \infty)$	5.87	3.11
$\omega'/2\pi$ (Hz/s)	Two-sided exponential	$[-2, 0.5]$	-0.0892	0.185
ζ_f	Beta	$[0.02, 1]$	0.213	0.143

TABLE A.2 Regression coefficients for mean predictive relationships and total standard deviation and correlation coefficients for covariance matrix

i	$c_{i,0}$	$c_{i,1}$	$c_{i,2}$	$c_{i,3}$	$c_{i,4}$	σ_{vi}	ρ_{i1}	ρ_{i2}	ρ_{i3}	ρ_{i4}	ρ_{i5}	ρ_{i6}
1	-1.844	-0.071	2.944	-1.356	-0.265	0.654	1	-0.36	0.01	-0.15	0.13	-0.01
2	-6.195	-0.703	6.792	0.219	-0.523	0.730		1	0.67	-0.13	-0.16	-0.20
3	-5.011	-0.345	4.638	0.348	-0.185	0.658			1	-0.28	-0.20	-0.22
4	2.253	-0.081	-1.810	-0.211	0.012	1.001		Symmetric		1	-0.20	0.28
5	-2.489	0.044	2.408	0.065	-0.081	0.962					1	-0.01
6	-0.258	-0.477	0.905	-0.289	0.316	1.021						1