
Supplementary materials of "Complex or Parsimonious Site-based stochastic ground motion models: which one fits GM datasets better?"

Maijia Su¹ | Mayssa Dabaghi² | Marco Broccardo³

¹Department of Civil, Environmental and Mechanical Engineering, University of Trento, Trento, Italy

²Department of Civil and Environmental Engineering, American University of Beirut, Beirut, Lebanon

³Department of Civil, Environmental and Mechanical Engineering, University of Trento, Trento, Italy

This document supplements our manuscript titled "Complex or Parsimonious Site-Based Stochastic Ground Motion Models: Which One Fits GM Datasets Better?" by providing additional data and figures supporting our validation and analysis. It includes six supplementary sections:

- Section A displays the distribution of scenario parameters for the selected large ground motion dataset. (Supplementary to Section 4.1)
- Section B compares using spline-modulating function and Gamma modulating functions to fit the large dataset. (Supplementary to Section 4)
- Section C shows fitted results of using the baseline model to predict statistics of nonlinear response spectra at various constant ductility levels. (Supplementary to Section 4.3)
- Section D shows other ranking results in the optimal model comparison, including 2%- and 20%-damped linear repose spectra, as well as constant-ductility ($\mu = 1.5, 4$) nonlinear response spectra. (Supplementary to Section 5.3)
- Section E present details of the fitted joint probabilistic model for the GMM parameters, including marginal distributions and two dependence models. (Supplementary to Section 6)
- Section F shows the fitted results of using the site-based GMM to predict nonlinear response spectra. (Supplementary to Section 7.2)

1 List of Figures

2 A1	Distribution of moment magnitude M , source-to-site distance R , and shear-wave velocity V_{S30} for the the 1,001 records.	3
4 B1	Comparison of using two time-modulating functions to fit the large dataset w.r.t the linear response spectra $S_a(T)$	4
6 B2	Comparison of using two time-modulating functions to fit the large dataset w.r.t the nonlinear response spectra $S_a^{NL}(T)$	5
8 C1	Baseline GMM validation w.r.t. three quantiles of nonlinear-response spectra across different given ductility μ	5
10 C2	Baseline GMM validation w.r.t. logarithmic standard deviation of nonlinear-response spectra across different given ductility μ	6
12 C3	Baseline GMM validation w.r.t. nonlinear spectral correlation $\rho(T_1, T_2)$ across different given ductility μ	6
13 D1	Comparion metrics ϵ_{QoI} of 2%-damped $S_a(T)$ for the eight considered parametric GMMs. (a) and (b): Error bars of different metrics ϵ_{QoI} , with the whiskers representing $\pm 1\sigma$ confidence bounds; (c): Biases of $q_n\%$ vary with the quantile level n , with $\pm 1\sigma$ confidence bounds.	7
16 D2	Comparion metrics ϵ_{QoI} of 20%-damped $S_a(T)$ for the eight considered parametric GMMs. (a) and (b): Error bars of different metrics ϵ_{QoI} , with the whiskers representing $\pm 1\sigma$ confidence bounds; (c): Biases of $q_n\%$ vary with the quantile level n , with $\pm 1\sigma$ confidence bounds.	7
19 D3	Comparion metrics ϵ_{QoI} of constant-ductility ($\mu = 1.5$) spectra for the eight considered parametric GMMs. (a) and (b): Error bars of different metrics ϵ_{QoI} , with the whiskers representing $\pm 1\sigma$ confidence bounds; (c): Biases of $q_n\%$ vary with the quantile level n , with $\pm 1\sigma$ confidence bounds.	8
22 D4	Comparion metrics ϵ_{QoI} of constant-ductility ($\mu = 4$) spectra for the eight considered parametric GMMs. (a) and (b): Error bars of different metrics ϵ_{QoI} , with the whiskers representing $\pm 1\sigma$ confidence bounds; (c): Biases of $q_n\%$ vary with the quantile level n , with $\pm 1\sigma$ confidence bounds.	8
25 E1	Gaussian copula (lower triangle) and Vine Copula (upper triangle) fitted to the parameters extracted from the 1,001-GM dataset.	10
27 F1	Validation of the site-based GMM using two different copula models w.r.t. three quantiles of constant- ductility ($\mu = 2$) nonlinear-response spectra.	11
29 F2	Validation of the site-based GMMs using two different copula models w.r.t. logarithmic standard devi- ation of constant-ductility ($\mu = 2$) nonlinear-response spectra.	11
31 F3	Validation of the site-based GMMs using two different copula models w.r.t. spectral correlation of constant-ductility ($\mu = 2$) nonlinear-response spectra.	11

33 List of Tables

34 E1	Fitted marginal models of random GM Parameters.	9
-------	---	---

³⁵ **A | DISTRIBUTION OF SCENARIOS PARAMETERS FOR THE SELECTED DATASET**

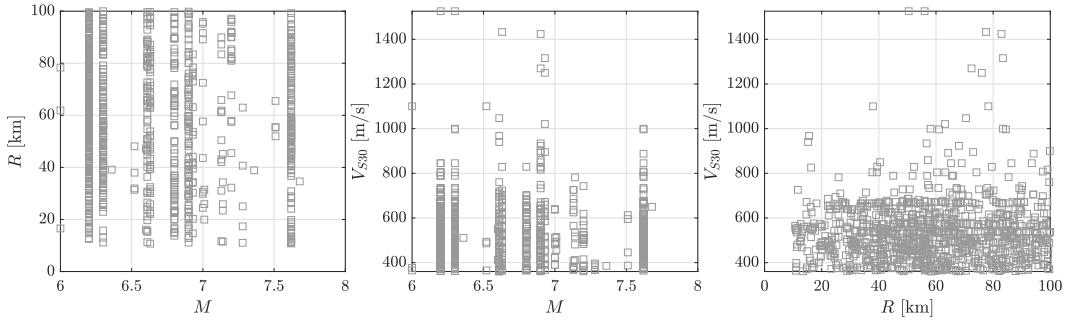


FIGURE A1 Distribution of moment magnitude M , source-to-site distance R , and shear-wave velocity V_{S30} for the 1,001 records.

³⁶ **B | COMPARISON OF TWO TIME-MODULATING FUNCTIONS**

³⁷ This section presents the results of replacing the spline-modulating function in the baseline model with the gamma-
³⁸ modulating function, highlighting the reasons for preferring the spline-modulating function in our study. The new
³⁹ stochastic GMM consists of seven parameters: three for frequency nonstationarity, three for the gamma-modulating
⁴⁰ function, and one for the corner frequency.

⁴¹ The gamma modulating function is defined as [1]:

$$q(t|\theta_T) = \alpha_1 t^{\alpha_2-1} e^{-\alpha_3 t}. \quad (1)$$

⁴² Here $\theta_T = [\alpha_1, \alpha_2, \alpha_3]$, where α_1 controls the intensity of the process, α_2 controls the shape, and α_3 controls the
⁴³ duration of the motion. This modulating function is fitted by matching partial features of simulated cumulative energy
⁴⁴ $\hat{I}_a(t|\theta_T)$ to those of recorded cumulative energy $I_a(t)$ [1]. Specifically, this method first optimizes α_2 and α_3 by
⁴⁵ minimizing two error metrics simultaneously: $|\hat{D}_{5-95} - D_{5-95}|$ and $|\hat{t}_{mid} - t_{mid}|$, where $D_{5-95} = t_{95} - t_5$ and $t_{mid} = t_{45}$
⁴⁶ are defined on the recorded GM and the corresponding hatted parameters are computed based on $\hat{I}_a(t|\theta_T)$. Once
⁴⁷ α_2 and α_3 are fitted, α_1 is analytically computed to satisfy $\hat{I}_a(t_f|\theta_T) = I_a(t_f)$:

$$\alpha_1 = \sqrt{I_a(t_f) \frac{(2\alpha_3)^{2\alpha_2-1}}{\Gamma(2\alpha_2-1)}}, \quad (2)$$

⁴⁸ where $\Gamma(\cdot)$ is the gamma function. The gamma modulating function is truncated between $[0, 2D_{5-95}]$ [1], i.e., the
⁴⁹ simulated GM starts at $t_0 = 0$ and ends at $t_f = 2D_{5-95}$.

⁵⁰ The remaining parameters of the new stochastic GMMs are fitted following the procedure in Section 3. The
⁵¹ fitting results for the large dataset (Section 4.1) are compared with the baseline GMM. The comparisons include
⁵² various quantile levels of linear response spectra (Figure B1) and nonlinear response spectra (Figure B2). Figure B1
⁵³ evaluates the ability of the two time-modulating functions to capture linear spectral trends across different damping
⁵⁴ ratios ($\zeta = 0.02, 0.05$, and 0.2). The results show that both functions provide similar approximations for the low-level

quantile $q_{1\%}(S_a)$ and for other quantile curves at long periods (around $T > 1$ s). However, discrepancies emerge at short periods, particularly for higher quantiles ($q_{n\%}(S_a)$ with $n > 50$). The gamma-modulating function generally underestimates spectral values at lower periods, whereas the spline-modulating function more closely aligns with the real dataset. Figure B2 presents similar findings for nonlinear response spectra across different constant ductility levels ($\mu = 1.5, 2$, and 4). These results suggest that the spline-modulating function more accurately predicts both linear and nonlinear response spectra, particularly at short periods.

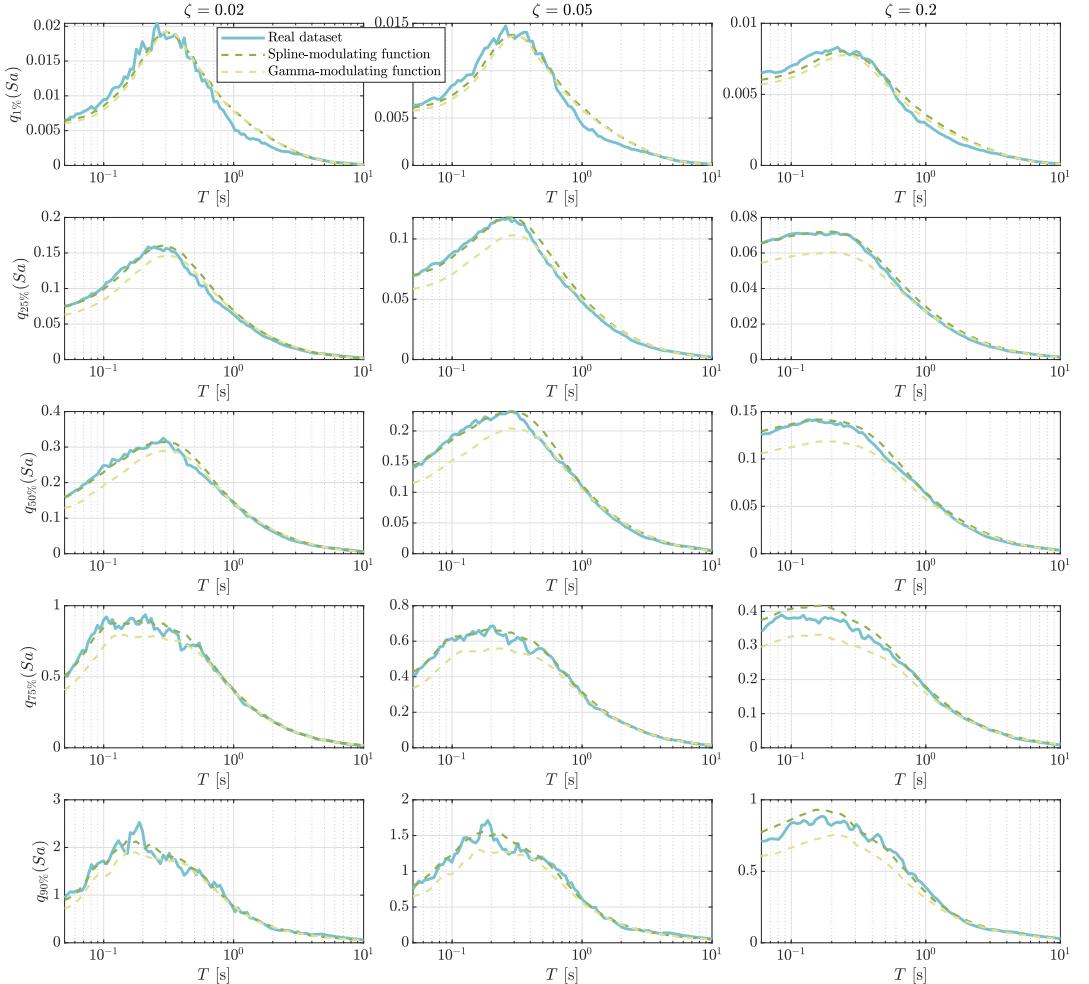


FIGURE B1 Comparison of using two time-modulating functions to fit the large dataset w.r.t the linear response spectra $Sa(T)$

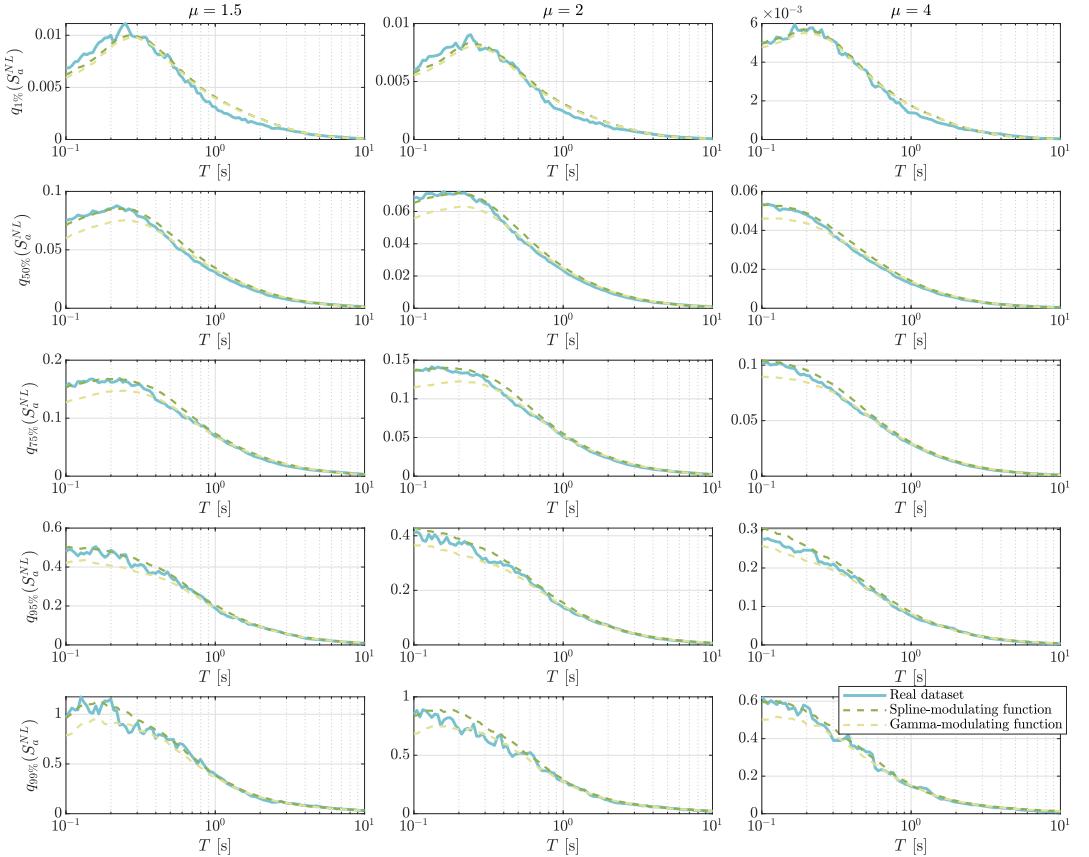


FIGURE B2 Comparison of using two time-modulating functions to fit the large dataset w.r.t the nonlinear response spectra $S_a^{NL}(T)$

61 **C | VALIDATION OF NONLINEAR-RESPONSE SPECTRUM FOR THE BASE-
62 LINE GMM**

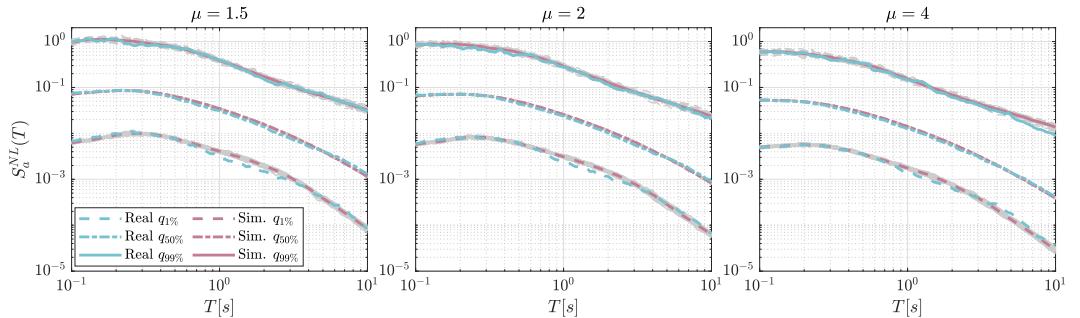


FIGURE C1 Baseline GMM validation w.r.t. three quantiles of nonlinear-response spectra across different given ductility μ .

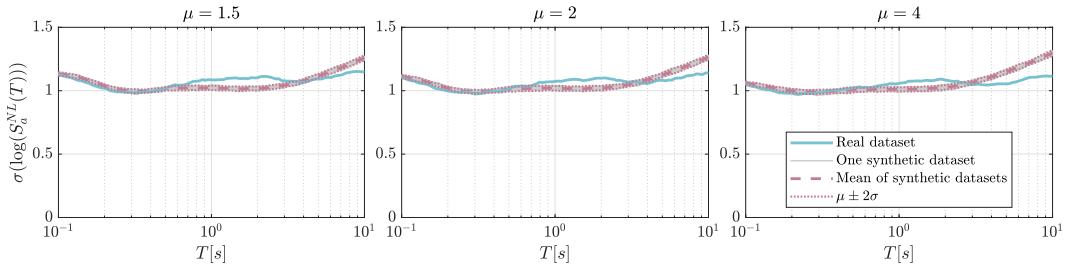


FIGURE C2 Baseline GMM validation w.r.t. logarithmic standard deviation of nonlinear-response spectra across different given ductility μ .

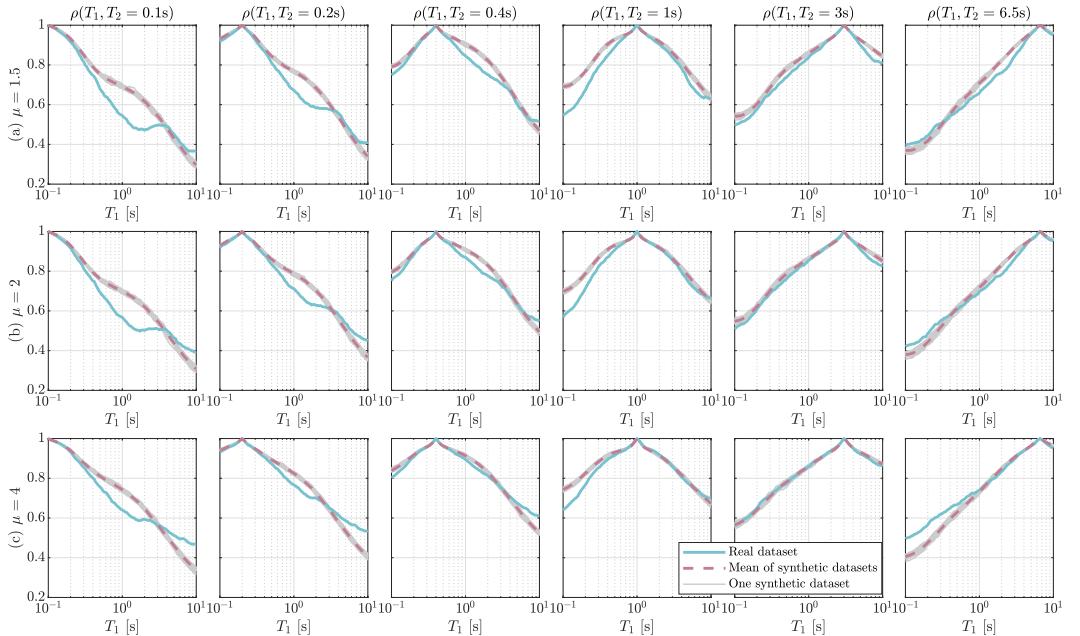


FIGURE C3 Baseline GMM validation w.r.t. nonlinear spectral correlation $\rho(T_1, T_2)$ across different given ductility μ .

63 D | RANKING RESULTS IN THE OPTIMAL GMM STUDY

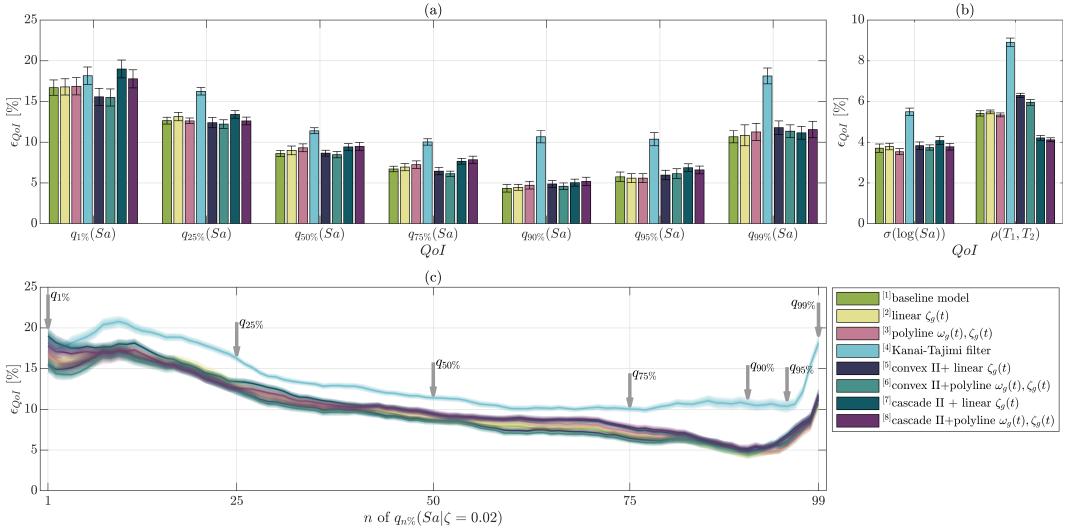


FIGURE D1 Comparison metrics ϵ_{QoI} of 2%-damped $Sa(T)$ for the eight considered parametric GMMs. (a) and (b): Error bars of different metrics ϵ_{QoI} , with the whiskers representing $\pm 1\sigma$ confidence bounds; (c): Biases of $q_n\%$ vary with the quantile level n , with $\pm 1\sigma$ confidence bounds.

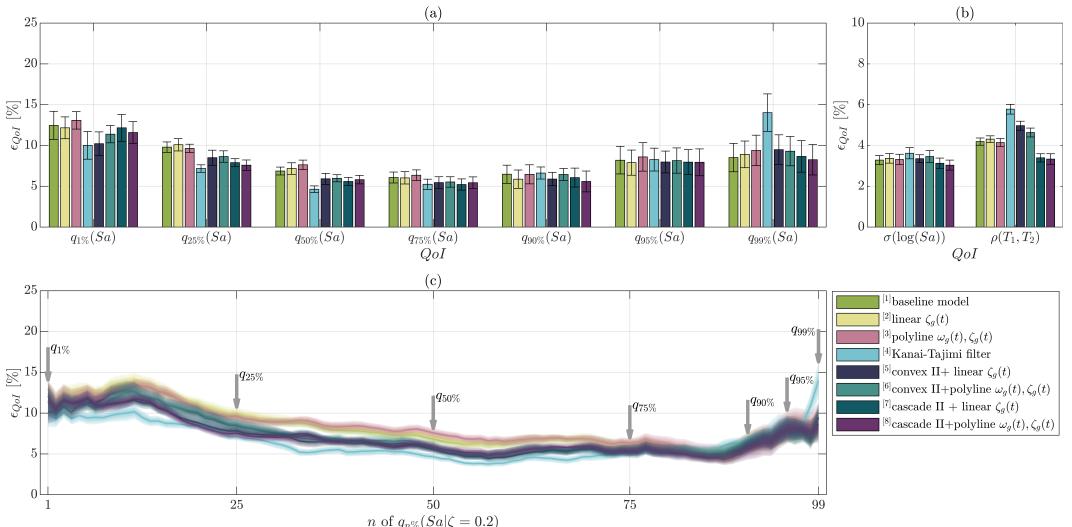


FIGURE D2 Comparison metrics ϵ_{QoI} of 20%-damped $Sa(T)$ for the eight considered parametric GMMs. (a) and (b): Error bars of different metrics ϵ_{QoI} , with the whiskers representing $\pm 1\sigma$ confidence bounds; (c): Biases of $q_n\%$ vary with the quantile level n , with $\pm 1\sigma$ confidence bounds.

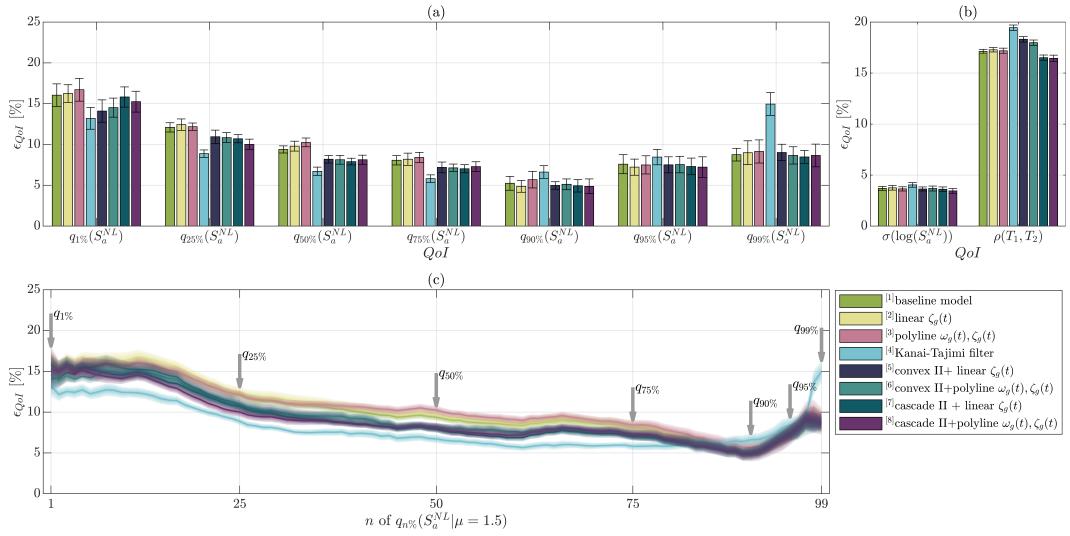


FIGURE D3 Comparison metrics ϵ_{QoI} of constant-ductility ($\mu = 1.5$) spectra for the eight considered parametric GMMs. (a) and (b): Error bars of different metrics ϵ_{QoI} , with the whiskers representing $\pm 1\sigma$ confidence bounds; (c): Biases of $q_n\%$ vary with the quantile level n , with $\pm 1\sigma$ confidence bounds.

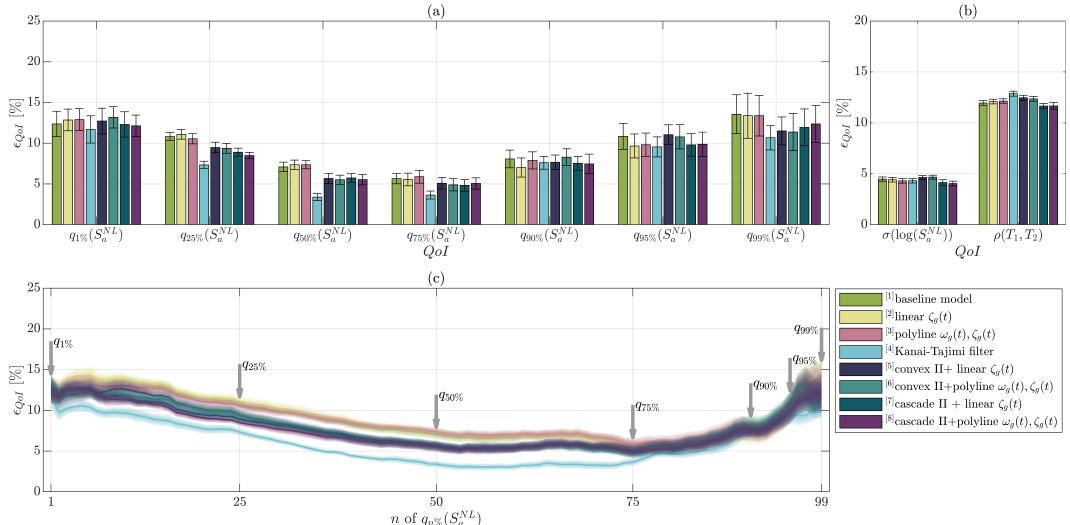


FIGURE D4 Comparison metrics ϵ_{QoI} of constant-ductility ($\mu = 4$) spectra for the eight considered parametric GMMs. (a) and (b): Error bars of different metrics ϵ_{QoI} , with the whiskers representing $\pm 1\sigma$ confidence bounds; (c): Biases of $q_n\%$ vary with the quantile level n , with $\pm 1\sigma$ confidence bounds.

64 E | FITTED MARGINALS AND JOINT DEPENDENCE MODELS

65 Table E1 summarizes the fitted marginal models along with their corresponding distribution parameters, the distribution support, and the first two statistical moments. The distribution supports are determined by visual observations of
 66 corresponding histograms. Moreover, the supports are adjusted to avoid: (1) numerical issues (e.g., $t_f > 200$ s causes
 67 out-of-memory) and (2) unrealistic simulation (e.g., t_f is too small/large or $\omega_g(t) = 0$).
 68

TABLE E1 Fitted marginal models of random GM Parameters.

No.	GM Parameter	Marginal	Parameters		Moments		Distribution support	Unit
			Par#1	Par#2	μ	σ		
1	$\log(I_a)$	Gaussian	-5.557	1.896	-5.557	1.896	(-Inf, Inf)	$\log(g^2 \cdot s)$
2	$\omega(t_{mid})$	Lognormal	3.162	0.610	28.446	19.112	(0, Inf)	$2\pi \cdot \text{Hz}$
3	$\omega'(t_{mid})$	Laplace	-0.227	0.709	-0.227	1.003	(-Inf, Inf)	$2\pi \cdot \text{Hz/s}$
4	$\zeta(t_{mid})$	Weibull	0.505	2.524	0.448	0.190	[0.02, 1]	1
5	D_{0-5}	Gamma	0.595	4.357	7.320	3.507	[0.1, 20]	s
6	D_{5-30}	Weibull	5.398	1.729	4.811	2.869	[0.1, 15]	s
7	D_{30-45}	Gamma	1.167	1.965	1.684	1.202	[0.1, 10]	s
8	D_{45-75}	Gamma	0.637	2.899	4.549	2.672	[0.1, 20]	s
9	D_{75-95}	Gumbel	8.172	3.637	10.271	4.664	[0.1, 40]	s
10	D_{95-100}	Lognormal	3.196	0.960	38.737	47.634	[0.1, 40]	s
11	f_c	Gamma	3.572	0.853	0.239	0.259	[0, 2]	Hz

69 Figure E1 visualizes the two dependency models: Gaussian copula and C-vine copula. The Gaussian copula
 70 is the classical choice, with the correlation coefficients displayed in the lower triangle. The Gaussian copula dis-
 71 plays the correlation coefficients, while the Vine Copula plots the contours of the logarithmic copula density (i.e.,
 72 $\log_{10}(c_{i_e,j_e|D_e}(u_{i_e|D_e}, u_{j_e|D_e}))$). The C-vine copula contains a sequence of the sorted nodes, denoted as $S = [8, 4, 7, 11,$
 73 $10, 1, 9, 6, 5]$, where the m -th element s_m of S represents the dominating node of \mathcal{T}_{m+1} . The first tree \mathcal{T}_1 contains 10 fit-
 74 ted bivariate copulas, where each copula links two unconditional random variables. For example, “Frank(θ_8, θ_4)” refers
 75 to a Frank bivariate copula $C_{8,4}(u_8, u_4)$ measuring the dependency between the random variables θ_8 and θ_4 . Note that
 76 independence is abbreviated as “Indep”. In subsequent trees, the bivariate copulas $C_{i_e,j_e|D_e}(\cdot)$ in \mathcal{T}_m ($m \geq 2$) are condi-
 77 tional on the first $(m - 1)$ elements of S , i.e., $D_e = \{s_1, s_2, \dots, s_{m-1}\}$. For example, “Clayton(θ_{11}, θ_{10})” in \mathcal{T}_4 represents a
 78 Clayton copula $C_{11,10|D_e=\{8,4,7\}}(\cdot)$ conditional on three variables θ_8, θ_4 and θ_7 . Specifically, this copula links two condi-
 79 tional CDFs $F(\theta_{11}|\theta_8, \theta_4, \theta_7)$ and $F(\theta_{10}|\theta_8, \theta_4, \theta_7)$, expressed as $C_{11,10|8,4,7}(F(\theta_{11}|\theta_8, \theta_4, \theta_7), F(\theta_{10}|\theta_8, \theta_4, \theta_7))$. These
 80 two conditional CDFs can be derived from the previous trees, such as $F(\theta_{11}|\theta_8, \theta_4, \theta_7) = \frac{\partial C_{7,11}(u_7, u_{11}|u_8, u_4)}{\partial u_7}$, where
 81 $C_{7,11}(\cdot)$ refers to the copula “Frank(θ_7, θ_{11})” in \mathcal{T}_3 . Therefore, the copula $C_{11,10|8,4,7}(\cdot)$ in \mathcal{T}_4 , nested with the two cop-
 82 ulas (i.e., Frank(θ_7, θ_{11}) and Indep(θ_7, θ_{10})) in \mathcal{T}_3 , characterizes the dependence between θ_{11} and θ_{10} conditioned on $\theta_8,$
 83 θ_4 and θ_7 . This nested structure applies to each bivariate copula $C_{i_e,j_e|D_e}(\cdot)$ in \mathcal{T}_m ($m \geq 2$). The product of all bivariate
 84 copulas across the trees fully captures the dependence structure of the random GMM parameters Θ .

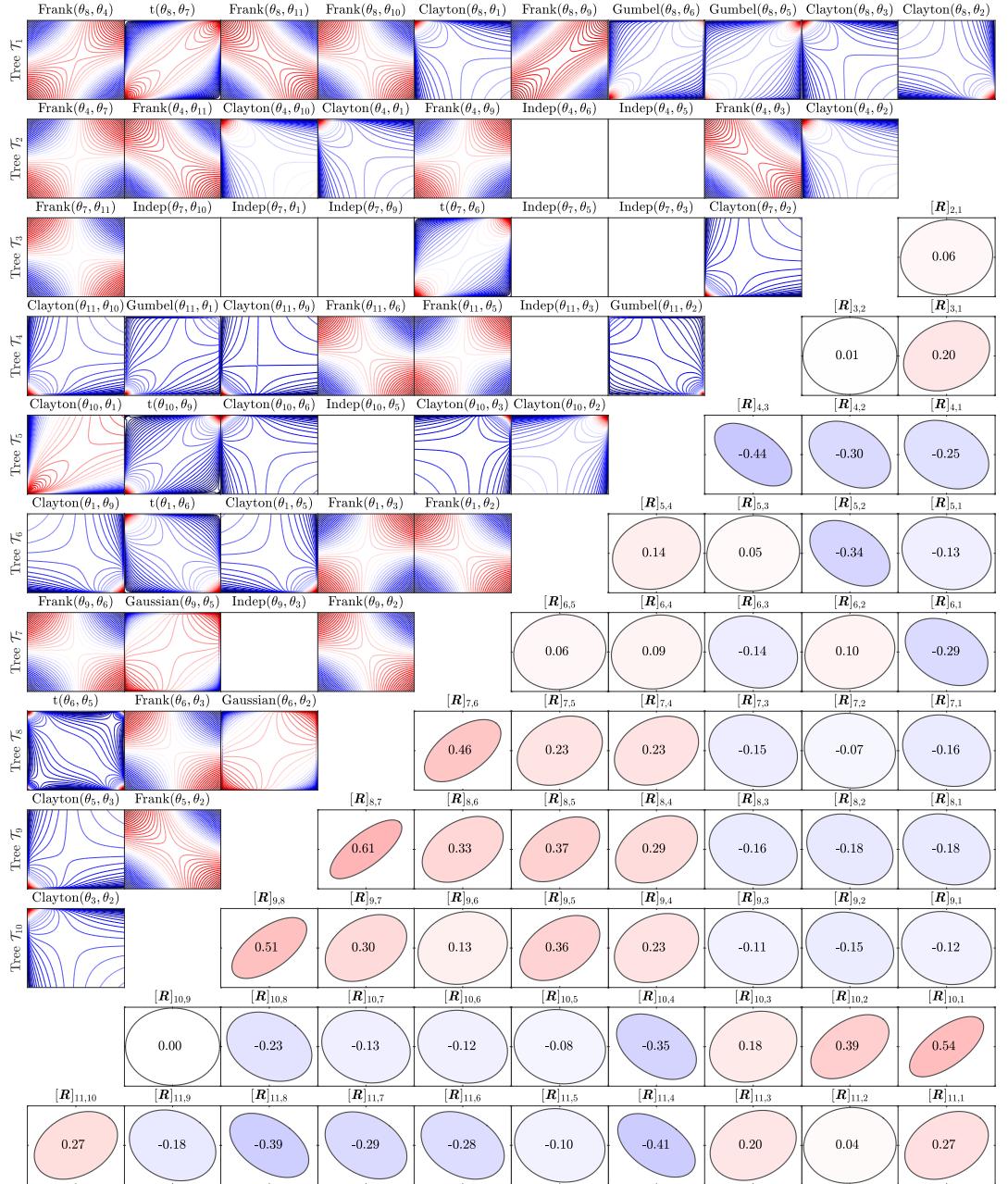


FIGURE E1 Gaussian copula (lower triangle) and Vine Copula (upper triangle) fitted to the parameters extracted from the 1,001-GM dataset.

85 **F | VALIDATION OF NONLINEAR-RESPONSE SPECTRUM FOR THE SITE-BASED
86 GMMS**

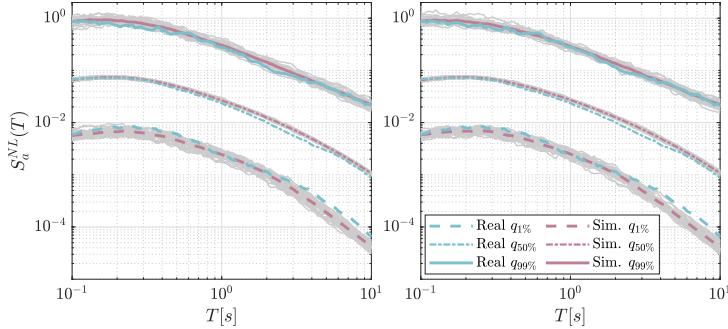


FIGURE F1 Validation of the site-based GMM using two different copula models w.r.t. three quantiles of constant-ductility ($\mu = 2$) nonlinear-response spectra.

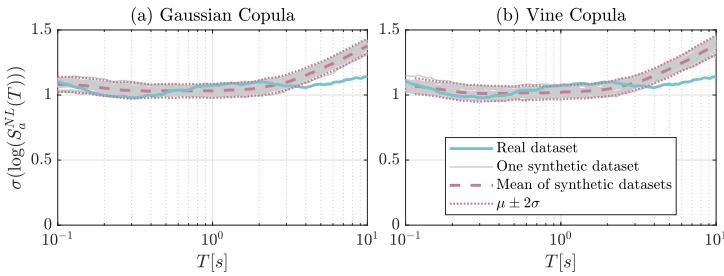


FIGURE F2 Validation of the site-based GMMS using two different copula models w.r.t. logarithmic standard deviation of constant-ductility ($\mu = 2$) nonlinear-response spectra.

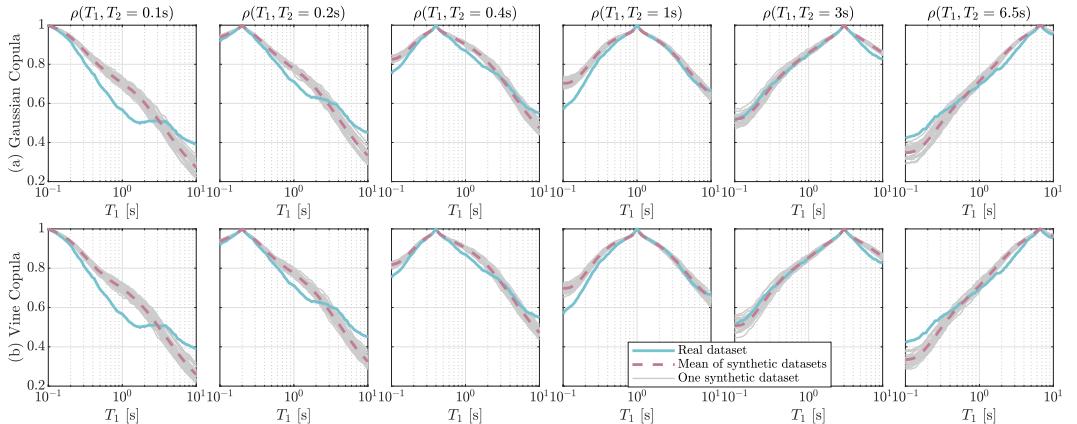


FIGURE F3 Validation of the site-based GMMS using two different copula models w.r.t. spectral correlation of constant-ductility ($\mu = 2$) nonlinear-response spectra.

87 references

- 88 [1] Rezaeian S, Der Kiureghian A. Simulation of orthogonal horizontal ground motion components for specified earthquake
89 and site characteristics. *Earthquake Engineering & Structural Dynamics* 2010;41(2):335–353. <https://onlinelibrary.wiley.com/doi/10.1002/eqe.1132>.
90 _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/eqe.1132>.