

Supplemental Material

Fast Probabilistic Seismic Hazard Analysis through Adaptive Importance Sampling

Soung Eil Houng and Luis Ceferino

List of Supplemental Material Captions

This supplementary document includes one Text, one Algorithm, one Table, and 17 Figures.

- **Text S1** Explanation on Algorithm S1
- **Algorithm S1** VEGAS adaptive importance sampling PSHA pseudocode.
- **Table S1** Seismic source information for comprehensive numerical examples to test our AIS PSHA framework
- **Figure S1** K-S D statistics examples.
- **Figure S2** Areal Source Example: COV as a function of computation time
- **Figure S3** Areal Source Example: Box plot showing the distribution of AIS MC estimates.
- **Figure S4** Areal Source Example: Convergence of iterated IS density to the hazard disaggregation
- **Figure S5** Areal Source Example: K-S D statistic and Mean difference between the hazard disaggregation and the proposed IS density.
- **Figure S6** Fault Source Example: COV as a function of computation time
- **Figure S7** Fault Source Example: Box plot showing the distribution of AIS MC estimates.
- **Figure S8** Fault Source Example: Convergence of iterated IS density to the hazard disaggregation.
- **Figure S9** Fault Source Example: K-S D statistic and mean difference between the hazard disaggregation and the proposed IS density.
- **Figure S10** Combined Sources Example: COV as a function of computation time

- **Figure S11** Combined Sources Example: Box plot showing the distribution of Full AIS MC estimates.
- **Figure S12** Combined Sources Example: Box plot showing the distribution of Partial AIS MC estimates.
- **Figure S13** Combined Sources Example: Convergence of iterated IS distribution to the hazard disaggregation
- **Figure S14** Dipping Fault Example: COV as a function of computation time
- **Figure S15** Dipping Fault Example: Convergence of iterated IS density to the hazard disaggregation
- **Figure S16** Smart AIS versus Smart Riemann Summation Example: COV as a function of computation time
- **Figure S17** The contribution of each part of VEGAS AIS PSHA algorithm to the total computation time.

[**Text S1:** Explanation on algorithm S1] From lines 1 to 5, the generation index t is set to be zero, and the sampling density q is set to be uniform. The main algorithm loop is from lines 6 to 24. The loop is continued until there is no improvement in the coefficient of variation or there is no previous generation (line 6). At line 7, X_m, X_r, X_ε are sampled from the distribution $q^{(t)}$. Then, the probability $\hat{\lambda}$ is estimated in lines 8-10, and its coefficient of variation is also calculated in line 11. From lines 12 to 23, the sampling function $q^{(t)}$ is updated. The steps are repeated over $m, r,$ and ε (lines 12-15). From lines 16 to 20, the contribution of each grid is calculated, at line 21, it is smoothed and dampened, and $q^{(t)}$ is updated by subdivision depending on d_i and restore the number of grids to the original number, N , in line 22. Updated $q^{(t+1)}$ is obtained by multiplying $q_m^{(t+1)}, q_r^{(t+1)},$ and $q_\varepsilon^{(t+1)}$ in line 23, and the while loop is ended by increasing the generation index t . When the main algorithm loop is terminated, it returns the hazard estimate $\hat{\lambda}$ and proposed optimal density q in line 25.

Algorithm S1 VEGAS adaptive importance sampling PSHA pseudocode

	[Parameters and Functions]
a	Ground motion intensity of interest (e.g., $a = 0.1$ g)
N	Number of grids (e.g., $N = 50$)
N_{subgrid}	Number of sub-grids (e.g., $N_{\text{subgrid}} = 10,000$)
N_s	Number of MC samples (e.g., $N_s = 1,000$)
\vec{X}_m	Magnitude sample vector ($\vec{X}_{m,i} = i$ th element of \vec{X}_m)
\vec{X}_r	Distance sample vector
\vec{X}_ε	Ground motion random variable sample vector
\vec{X}	$[\vec{X}_m; \vec{X}_r; \vec{X}_\varepsilon]$
$G(\cdot)$	Ground Motion Model
$I(\cdot)$	Indicator function
$f_X(\cdot)$	Original sampling distribution
$q^{(t)}(\cdot)$	Proposed sampling distribution at t th iteration step
$\hat{\lambda}^{(t)}$	Hazard estimate at t th iteration step
$\text{COV}^{(t)}$	COV of the hazard estimate at t th iteration step
ϵ	pre-defined iteration stopping criteria (e.g. 10%)
i	Index for partitioned grids ($i = 1, 2, \dots, N$)
j	Index for samples ($j = 1, 2, \dots, N_s$)

[Algorithm]

```

1:  $t = 0$ 
2:  $q_m^{(t)} = \text{U}([m_{\min}, m_{\max}])$  ( $\text{U}(a,b) =$  uniform distribution from  $a$  to  $b$ )
3:  $q_r^{(t)} = \text{U}([r_{\min}, r_{\max}])$ 
4:  $q_\varepsilon^{(t)} = \text{U}([\varepsilon_{\min}, \varepsilon_{\max}])$ 
5:  $q^{(t)} = q_m^{(t)} q_r^{(t)} q_\varepsilon^{(t)}$ 

6: while  $|\sum d_i^{(t)} - \sum d_i^{(t-1)}| / \sum d_i^{(t-1)} < \epsilon$  and  $t > 1$ :
7:    $\vec{X} = \{X_m, X_r, X_\varepsilon\}^{N_s}$ , ( $X_m \sim q_m^{(t)}$ ,  $X_r \sim q_r^{(t)}$ ,  $X_\varepsilon \sim q_\varepsilon^{(t)}$ )
8:    $\vec{H}^{(t)} = \{ I(G(x) > a|x) f_X(x) \mid x \in \vec{X} \}$ 
9:    $\vec{\lambda}^{(t)} = \vec{H}^{(t)} / q^{(t)}(\vec{X})$ 
10:   $\hat{\lambda}^{(t)} = \sum_1^{N_s} \lambda_i^{(t)} / N_s$ 
11:   $\text{COV}^{(t)} \leftarrow \text{Eq. (20)}$ 
12:  for  $u$  in  $[m, r, \varepsilon]$  :
13:    if  $u = m$  :  $(v, w) \leftarrow (r, \varepsilon)$ 
14:    if  $u = r$  :  $(v, w) \leftarrow (\varepsilon, m)$ 
15:    if  $u = \varepsilon$  :  $(v, w) \leftarrow (m, r)$ 
16:    for  $i$  in  $\{1, 2, \dots, N\}$  :
17:      for  $j$  in  $\{1, 2, \dots, N_s\}$  :
18:        if  $u_{i-1} < X_{u,j} < u_i$  :  $\bar{H}_i \leftarrow \bar{H}_i + (H_j^2 / (q_v(X_{v,j}) \cdot q_w(X_{w,j})))$ 
19:         $\bar{H}_i \leftarrow \sqrt{\bar{H}_i}$ 
20:         $d_i \leftarrow \left( \frac{\bar{H}_i \Delta x_i}{\sum_i \bar{H}_i \Delta x_i} \right)$ 
21:         $d_i \leftarrow d_d$  (Eq. (23) and (24))
22:         $q_u^{(t+1)} \leftarrow$  Subdivision and Restoration
23:         $q^{(t+1)} \leftarrow q_m^{(t+1)} q_r^{(t+1)} q_\varepsilon^{(t+1)}$ 
24:         $t \leftarrow t + 1$ 
25: return  $\hat{\lambda}^{(t-1)}$ ,  $q^{(t-1)}$ 

```

Table S1: Seismic source information for comprehensive numerical examples

	Area1	FaultA	FaultB	FaultC	Area2
Source Type	Area	Vertical Fault	Vertical Fault	Dipping Fault	Area
Source Geometry	Circle (R=100 km)	Line (L = 50 km)	Line (L = 85 km)	Line (L = 100 km)	Circle (R=100km)
EQ Occurrence Model	Exponential	Characteristic	Characteristic	Exponential	Exponential
b-value	0.9	0.9	0.9	0.9	0.9
M_{\min}	5.0	5.0	5.0	5.0	5.0
M_{\max}	6.5	6.75	7.0	6.5	6.5
M_{char}	-	6.5	6.75	-	-
$\nu(M > M_{\min})$ (yr^{-1})	0.0395	-	-	0.0395	0.0395
slip rate (mm/yr)	-	1	2	-	-
Seismogenic depth (km)	5-10	0-12	0-12	0-15	5
Rupture type	Point	Floating rupture	Floating rupture	Floating rupture	Point
Magnitude Scaling	-	$\log_{10}(A) = M - 4$ ($L/W = 2$)	$\log_{10}(A) = M - 4$ ($L/W = 2$)	WC94*	
GMM	Sea97*	Sea97	Sea97	ASK14*	Sea97

* WC94: Wells and Coppersmith (1994); Sea97: Sadigh et al. (1997); ASK14: Abrahamson et al. (2014)

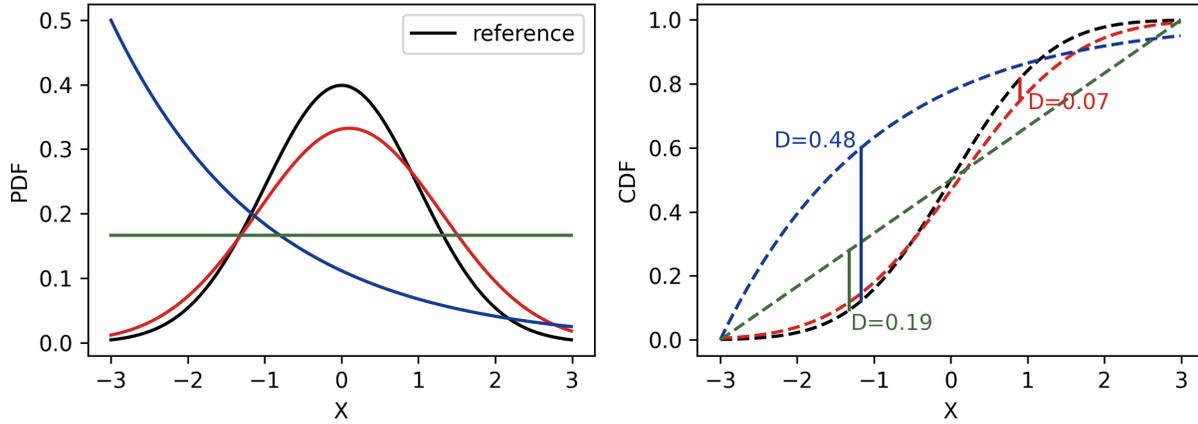


Figure S1. K-S D statistics examples. (left) Probability density functions of the reference ($\mathcal{N}(0, 1)$, black) and three distributions to compare: Red - Normal distribution, $\mathcal{N}(0.1, 1.2)$; Green - Uniform distribution $U(-3, 3)$; Blue - Exponential distribution, $\text{Exp}(\lambda = 0.5)$. (right) The corresponding cumulative distribution functions (CDF) and K-S D statistics. The maximum differences between the example distribution CDFs (dotted lines) and the reference CDF (black dotted line) are shown as vertical solid lines with their D values. The corresponding K-S D statistics for the normal, uniform, and exponential distributions are 0.07, 0.19, and 0.48, respectively.

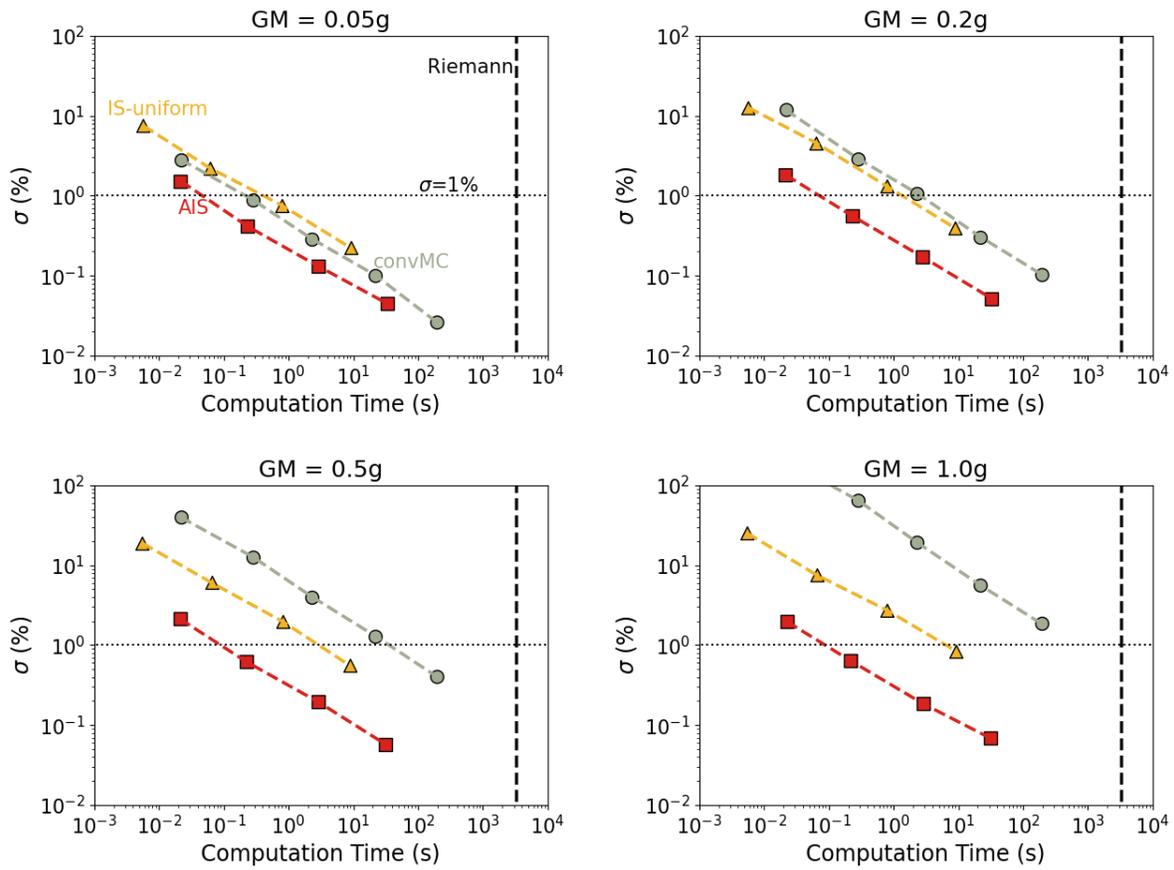


Figure S2. Areal Source Example: COV of the conventional MC (gray), IS (orange), and AIS (red) estimates as functions of computation time at ground motion of 0.05 g, 0.2 g, 0.5 g, and 1.0 g. The computation time required for Riemann summation is presented as black vertical dashed lines. At low ground motion, conventional MC outperforms IS and AIS, however, as the target ground motion increases, the performance of AIS makes a dramatic improvement in terms of both accuracy and computational cost.

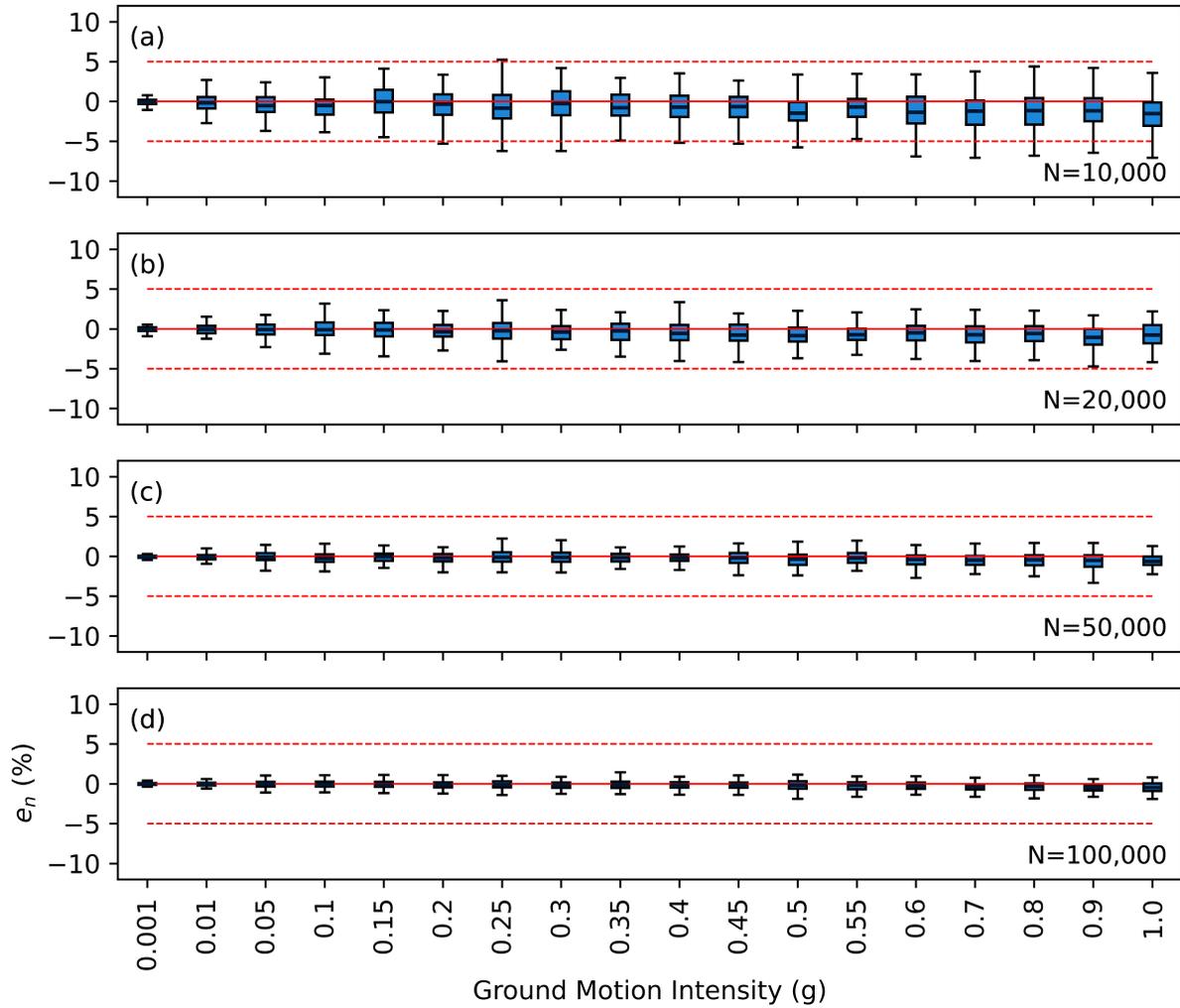


Figure S3. Areal Source Example: Box plot showing the distribution of AIS MC estimates with different numbers of samples. Each box is generated using the statistics of a hundred MC estimates. The x-axis is the ground motion, and the y-axis is the error to the benchmark PSHA curve, the relative difference with respect to the “true hazard” computed using the Riemann sum. The boxes represent the first and third quartile boundaries, and the whiskers indicate the 1.5 times interquartile ranges. The red dotted horizontal lines indicate the 5% error ranges.

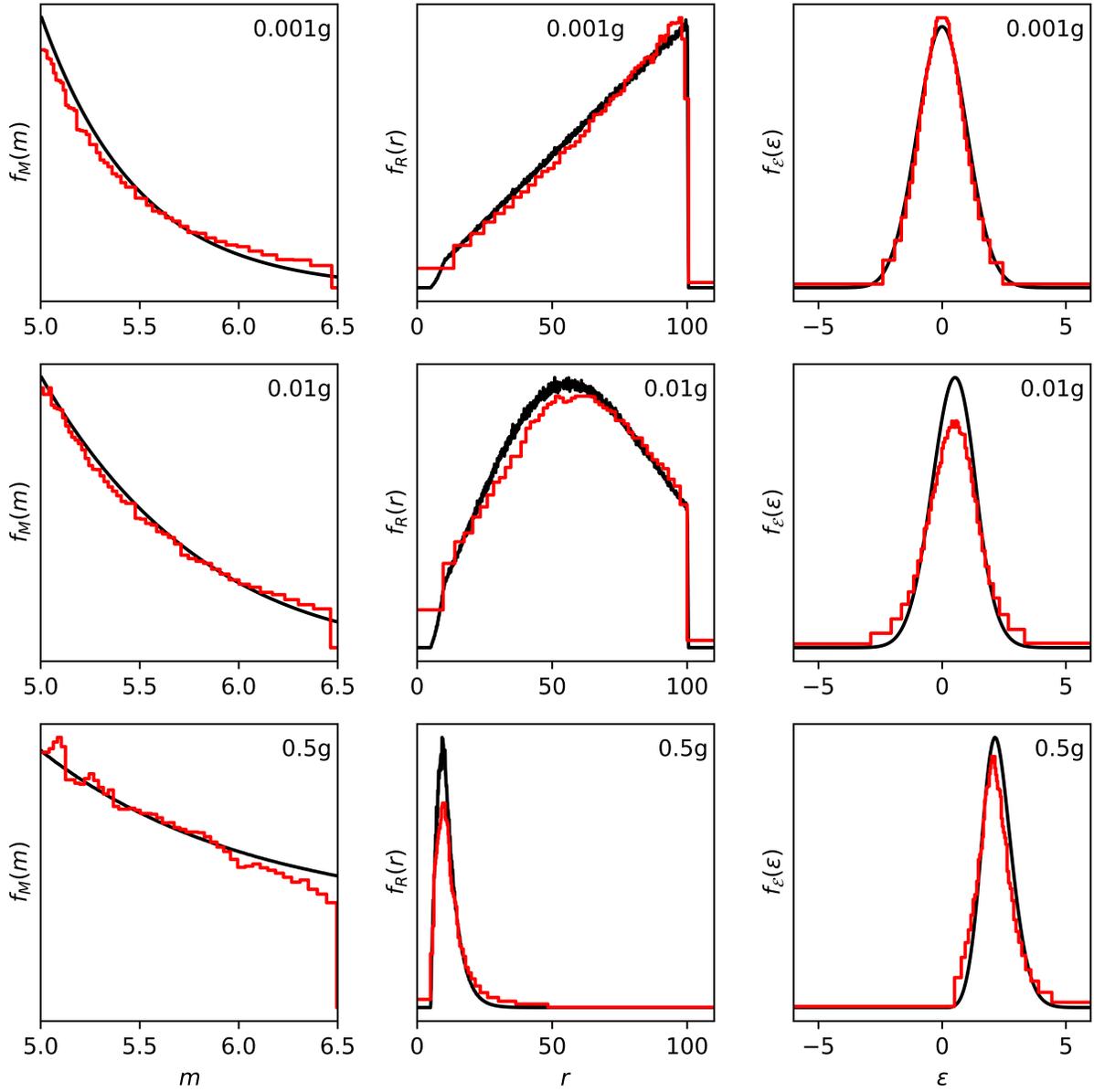


Figure S4. Areal Source Example: The convergence of m , r , and ε iterated IS densities density (red) derived in AIS algorithm ($N = 100,000$) to marginal distributions of hazard disaggregation (black) at ground motion intensities of 0.001 g, 0.01 g, and 0.5 g.

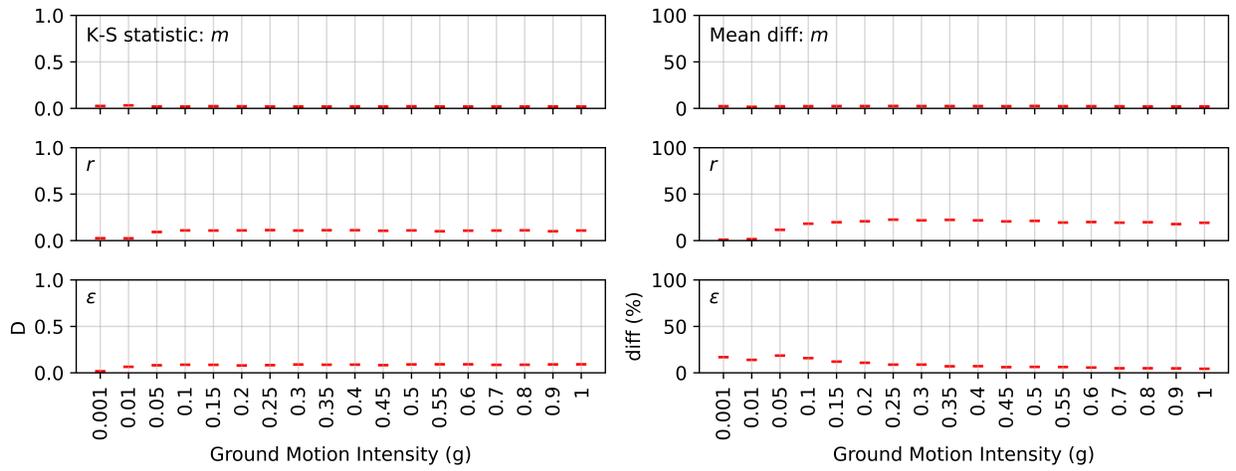


Figure S5. Areal Source Example: Kolmogorov–Smirnov D statistic and Mean difference between the hazard disaggregation and the proposed optimal density.

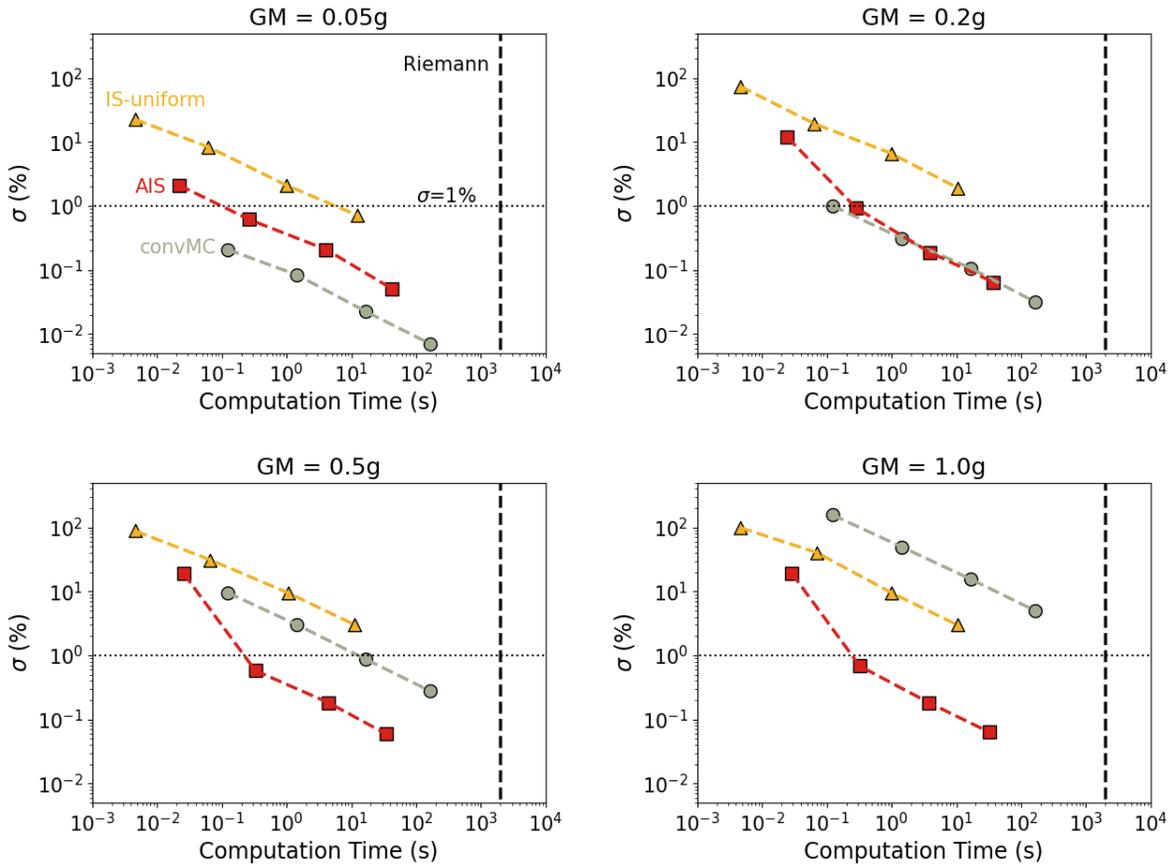


Figure S6. Fault Source Example: COV of the conventional MC (gray), IS (orange), AIS (red) estimates as functions of computation time at ground motions of 0.05 g, 0.2 g, 0.5 g, and 1.0 g.

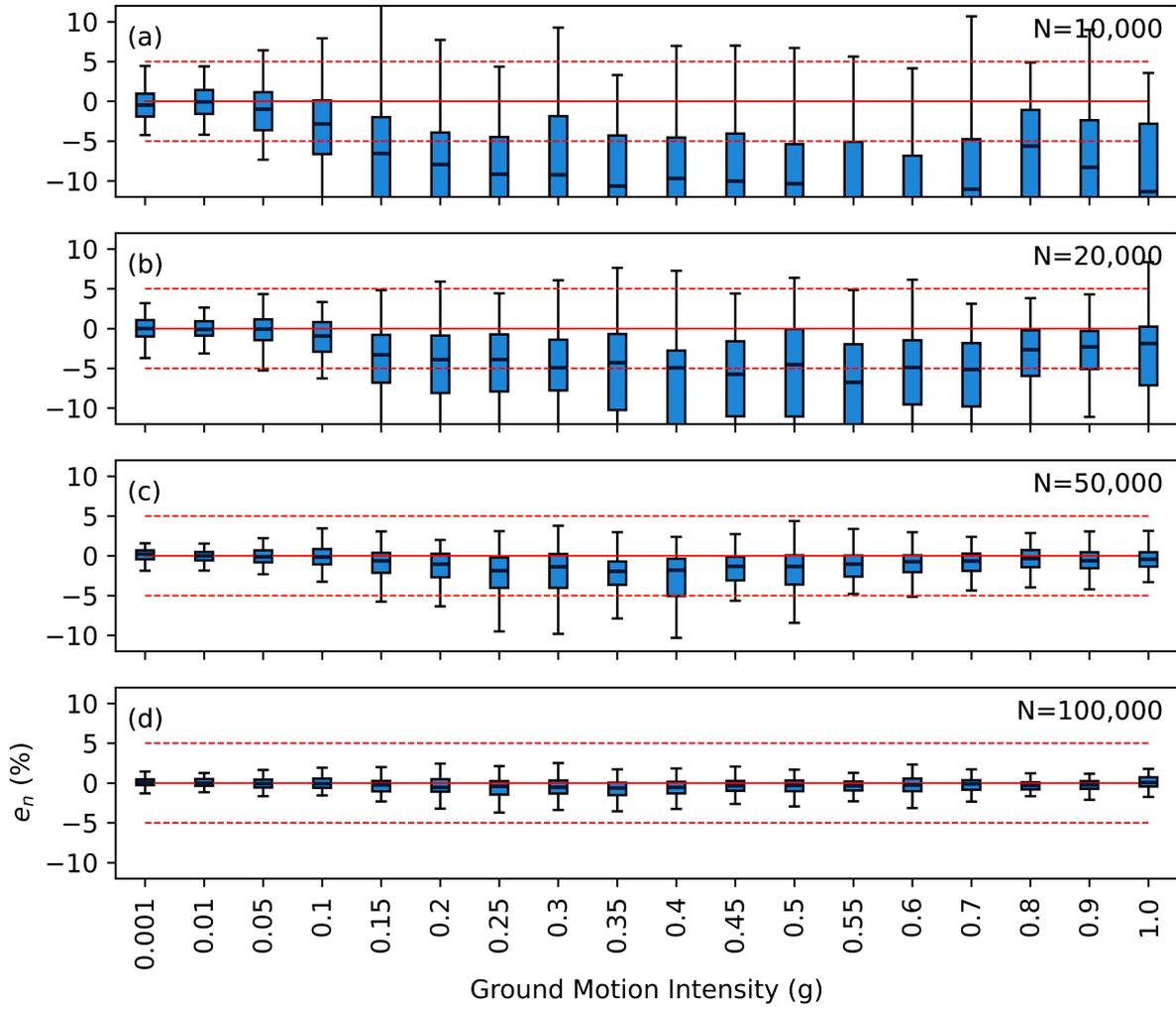


Figure S7. Fault Source Example: Box plot showing the distribution of AIS estimates with different numbers of samples.

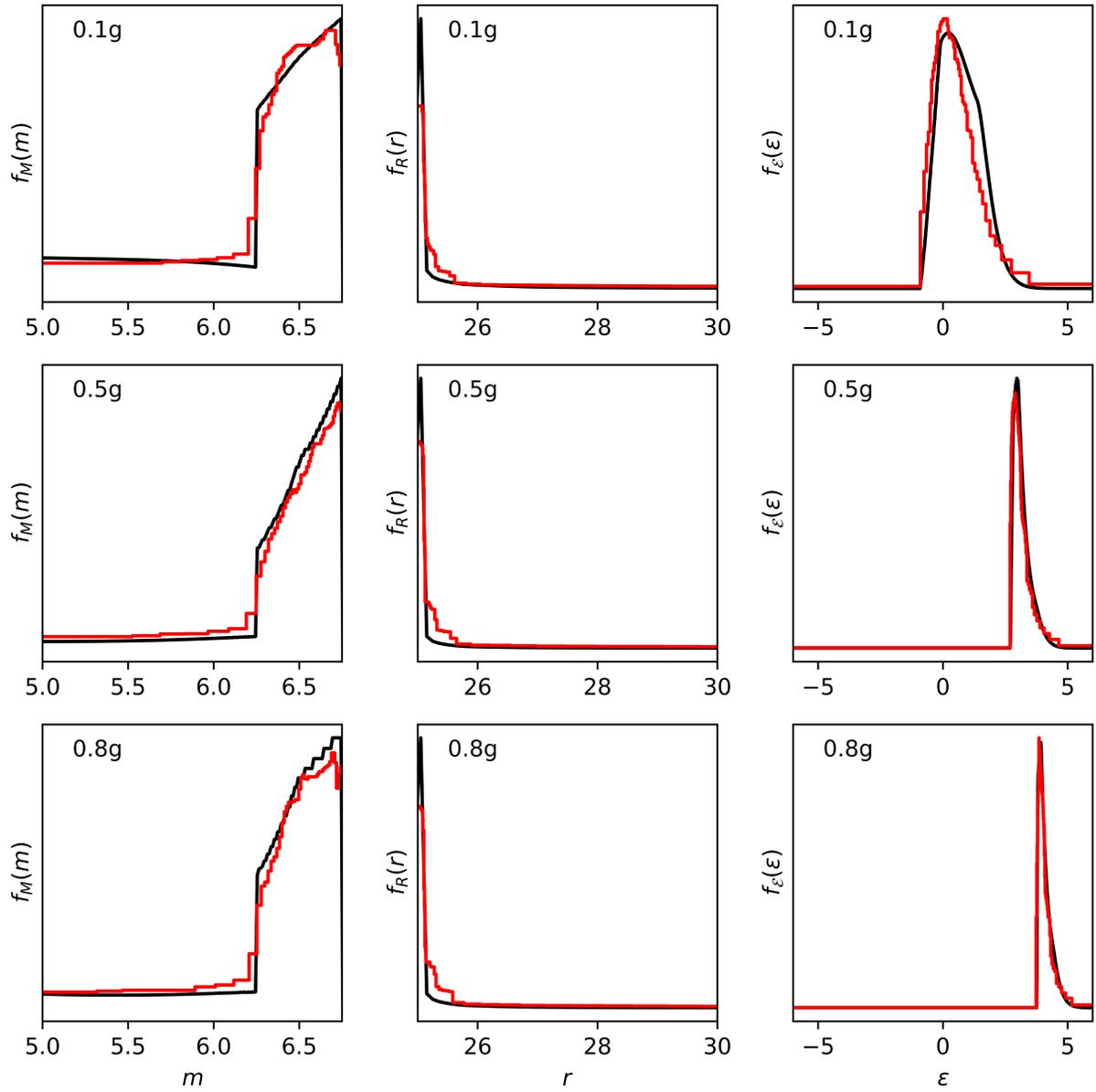


Figure S8. Fault Source Example: The convergence of m , r , and ϵ iterated IS densities (red) derived in AIS algorithm ($N = 1,000,000$) to marginal distributions of hazard disaggregation (black) at ground motion intensities of 0.1 g, 0.5 g, and 0.8 g.

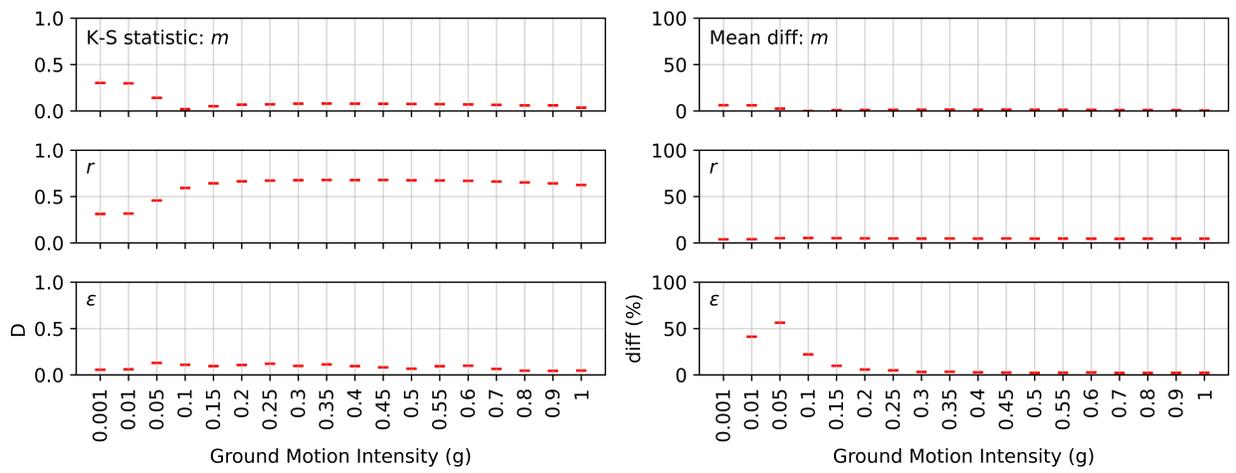


Figure S9. Fault Source Example: Kolmogorov–Smirnov D statistic and the mean differences between the hazard disaggregation and the proposed IS density.

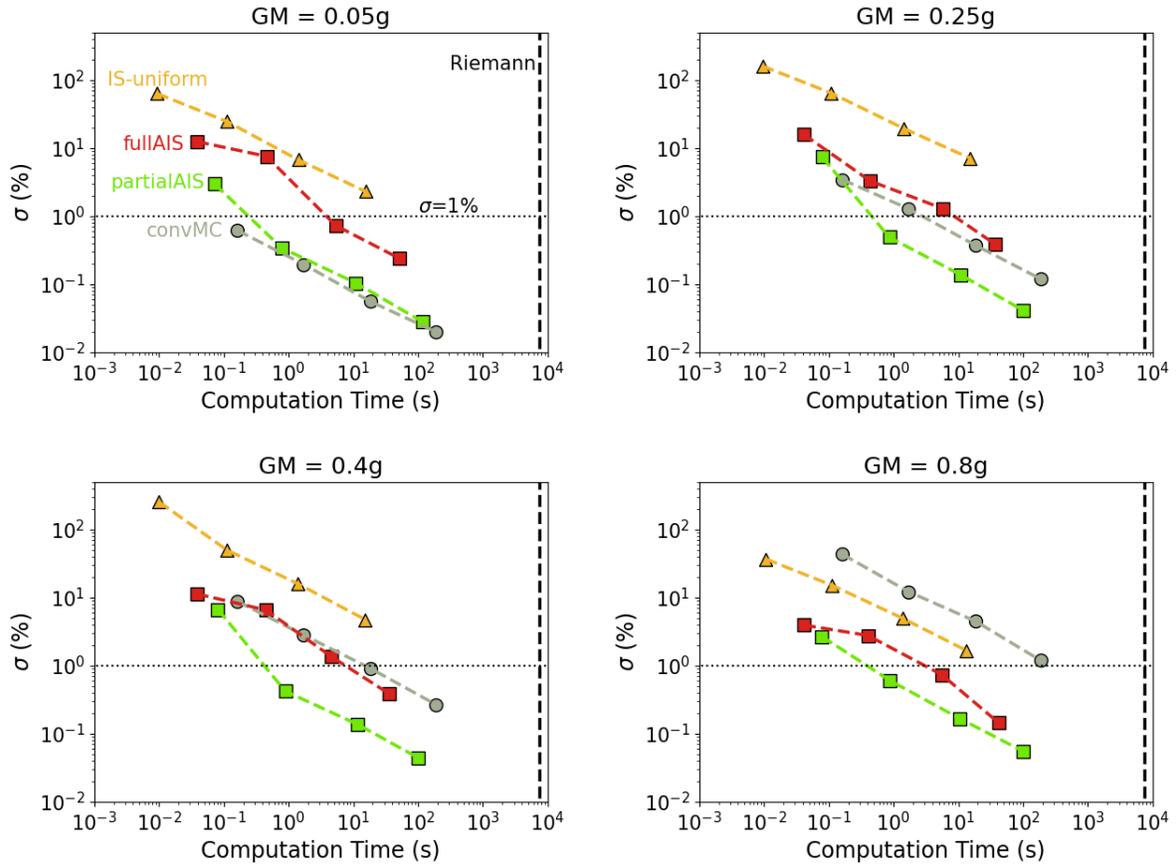


Figure S10. Combined Sources Example: COV of the conventional MC (gray), IS (orange), full AIS (red), and partial AIS (green) estimates as functions of computation time at the ground motion of 0.05 g, 0.25 g, 0.4 g, and 0.8 g

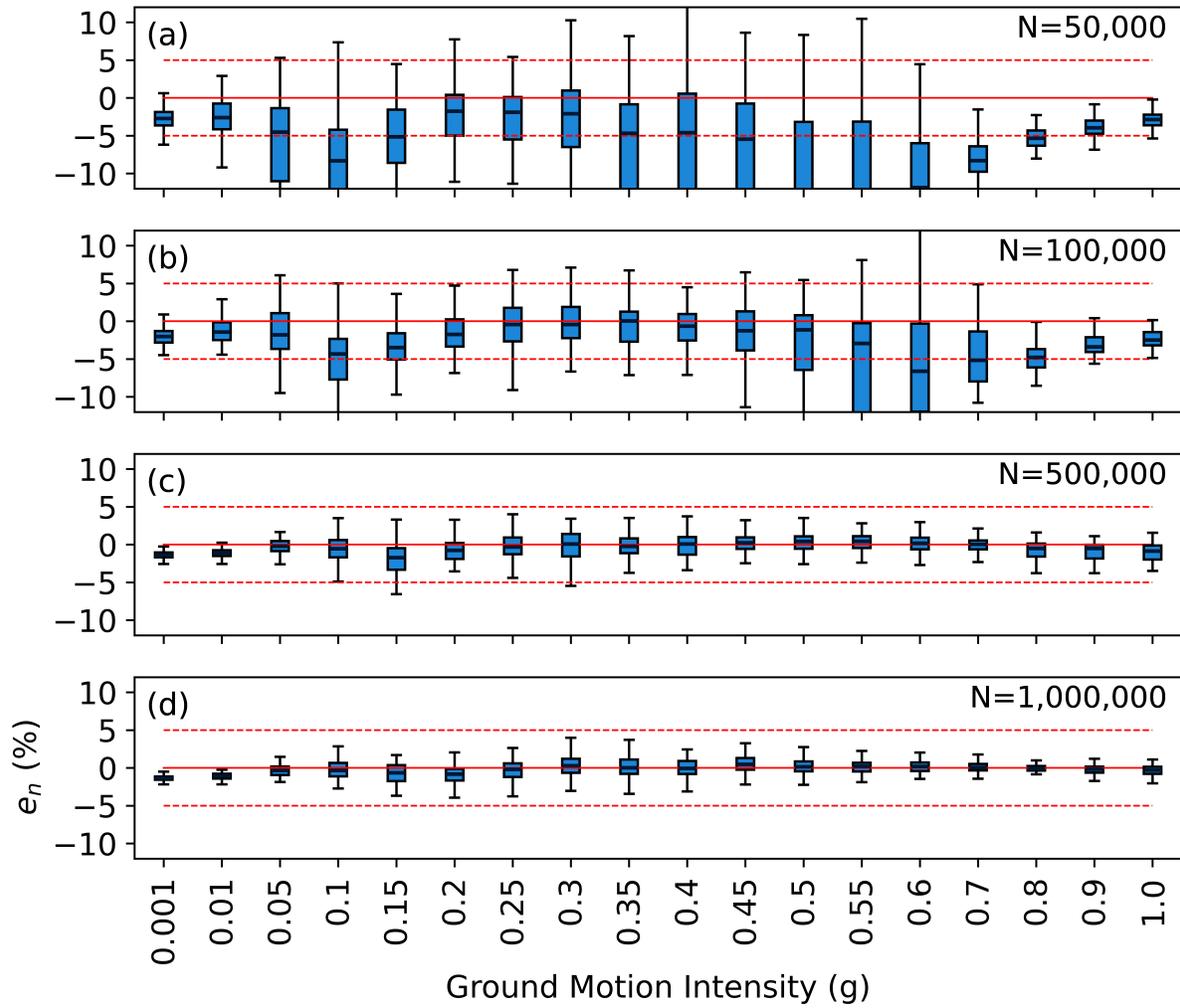


Figure S11. Combined Sources Example: Box plot showing the distribution of Full AIS MC estimates with different numbers of samples.

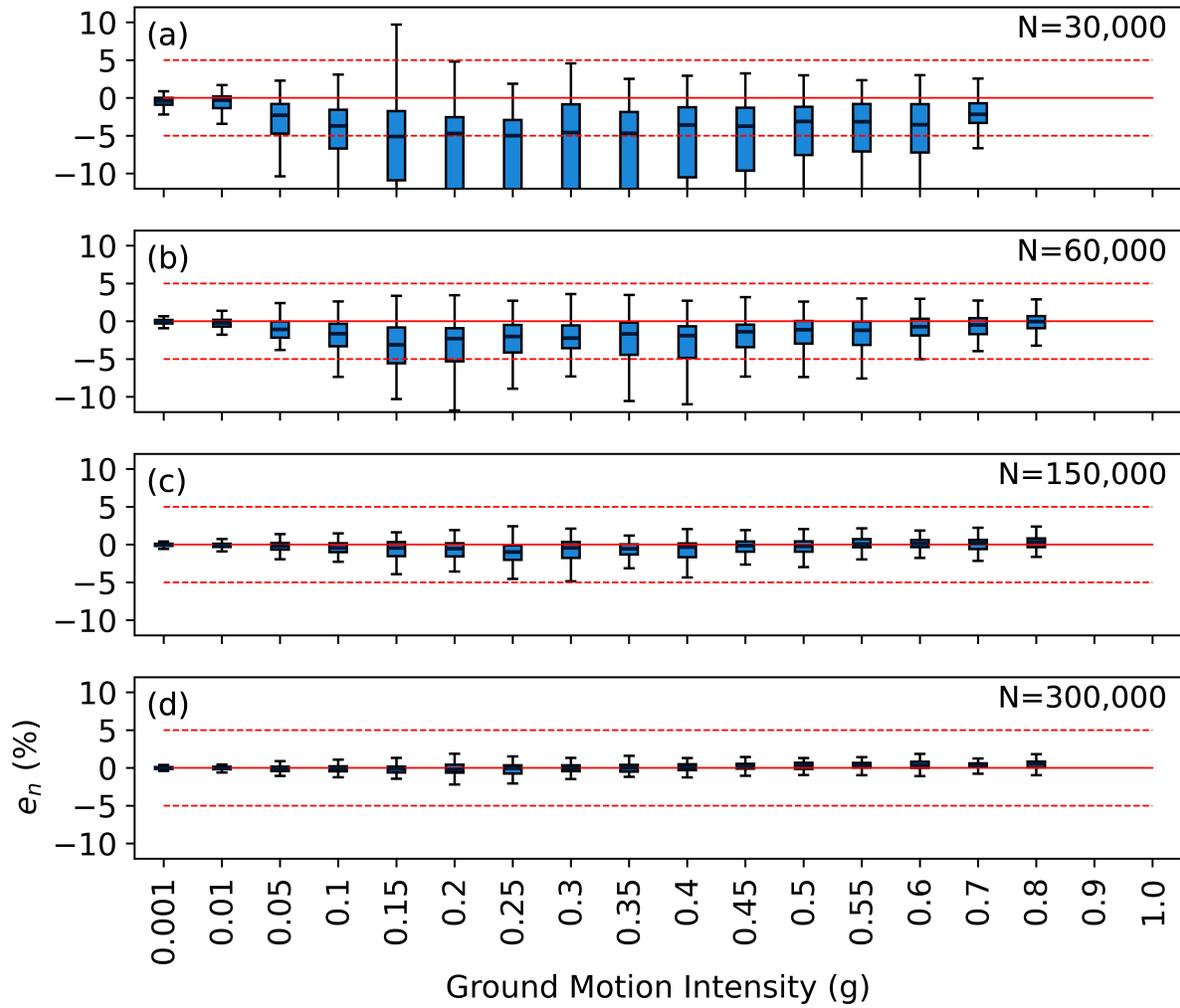


Figure S12. Combined Sources Example: Box plot showing the distribution of Partial AIS MC estimates with different numbers of samples. Ground motions greater than 0.8 g are not shown because the hazard from fault B reaches a numerical instability due to its low exceedance probability ($< 10^{-12}$ /yr).

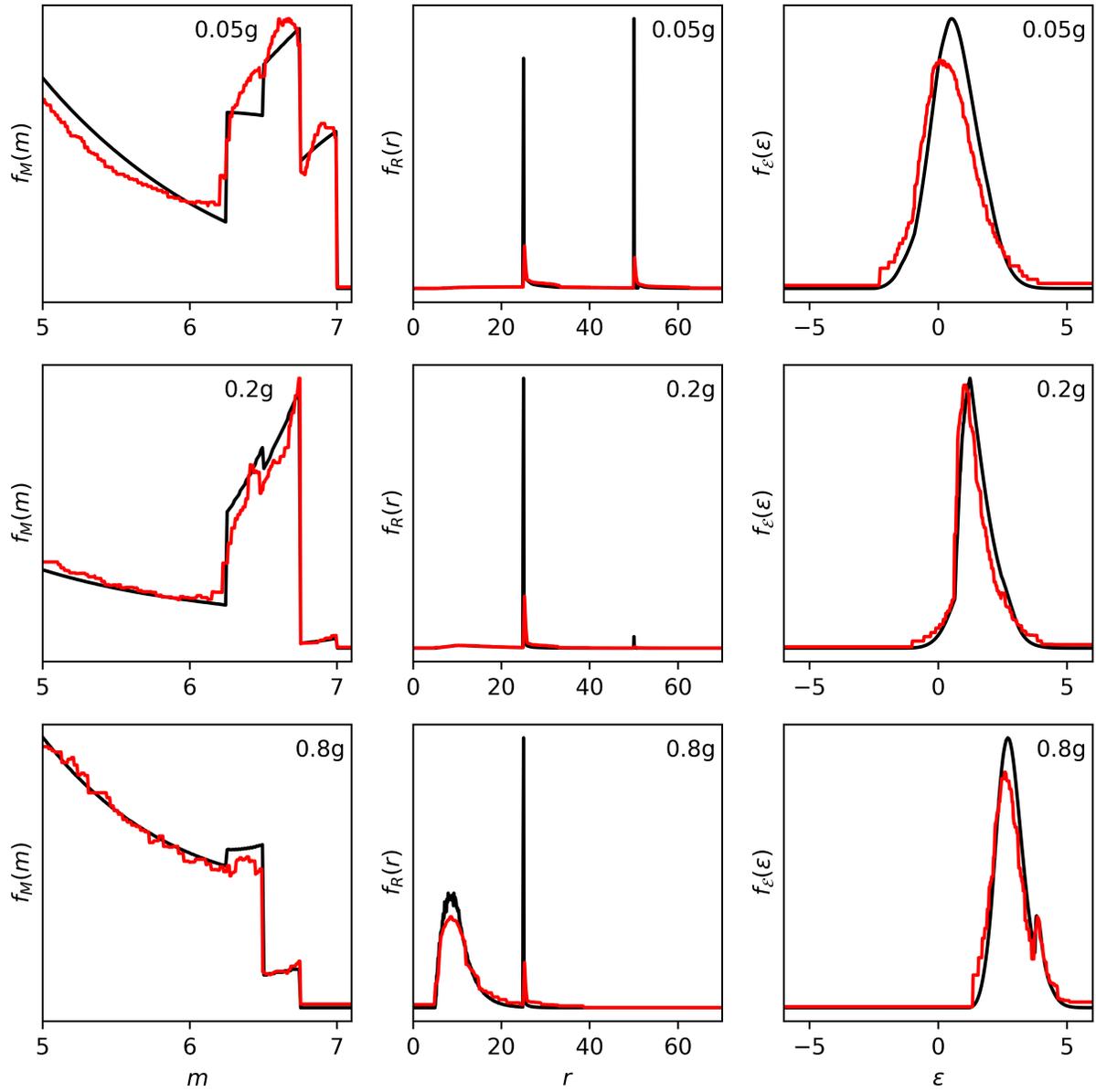


Figure S13. Combined Sources Example: The convergence of m , r , and ε iterated IS densities (red) derived in AIS algorithm ($N_s = 1,000,000$) to marginal distributions of hazard disaggregation (black) at ground motion intensities of 0.05 g, 0.2 g, and 0.8 g.

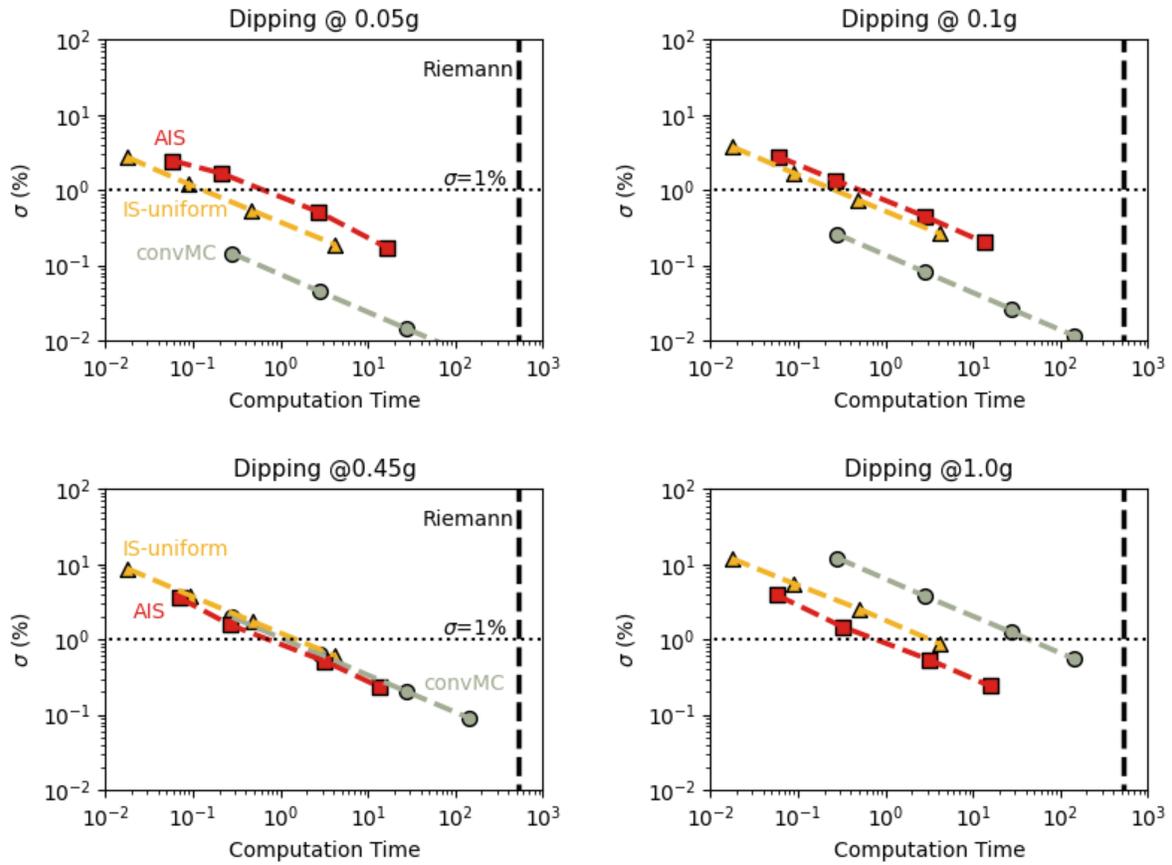


Figure S14. Dipping Source Example: COV of the conventional MC (gray), IS (orange), AIS (red) estimates as functions of computation time at the ground motion of 0.05 g, 0.1 g, 0.45 g, and 1.0 g.

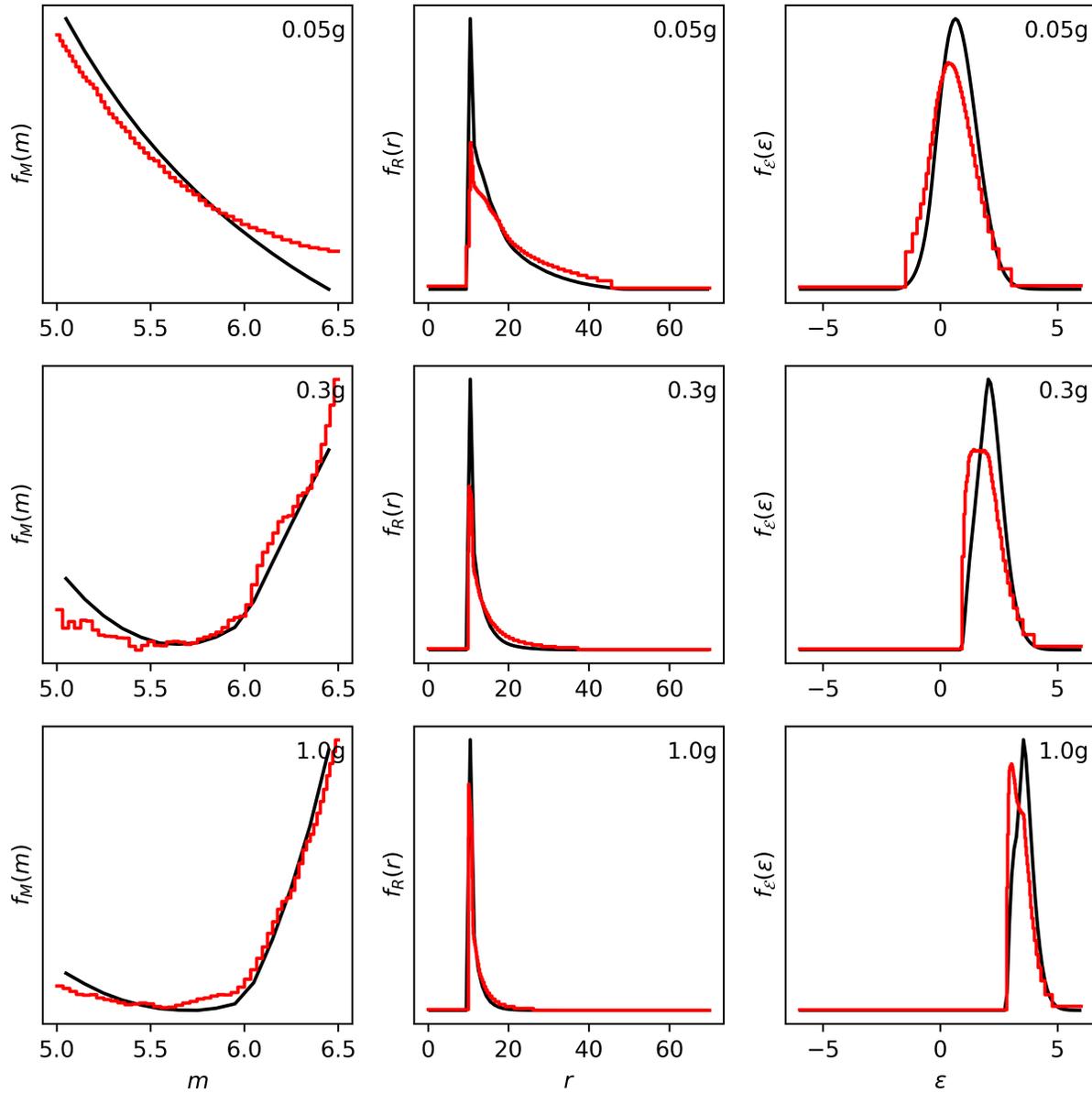


Figure S15. Dipping Source Example: The convergence of m , r , and ϵ iterated IS densities (red) derived in AIS algorithm to marginal distributions of hazard disaggregation (black) at ground motion intensities of 0.05 g, 0.3 g, and 1.0 g.

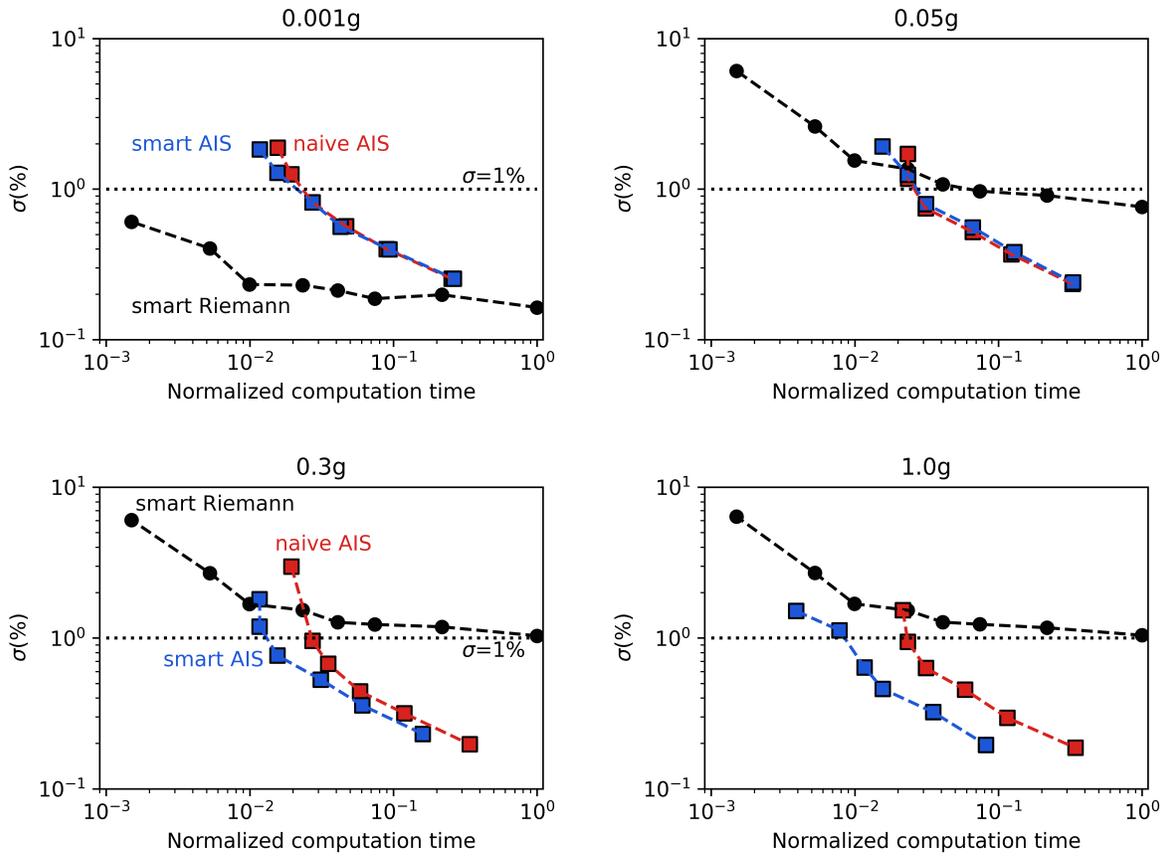


Figure S16. Smart AIS versus Smart Riemann Summation: COV of the smart Riemann Sum (black), naive AIS (red), and smart AIS (blue) estimates as functions of computation time at ground motion of 0.001 g, 0.05 g, 0.3 g, and 1.0 g

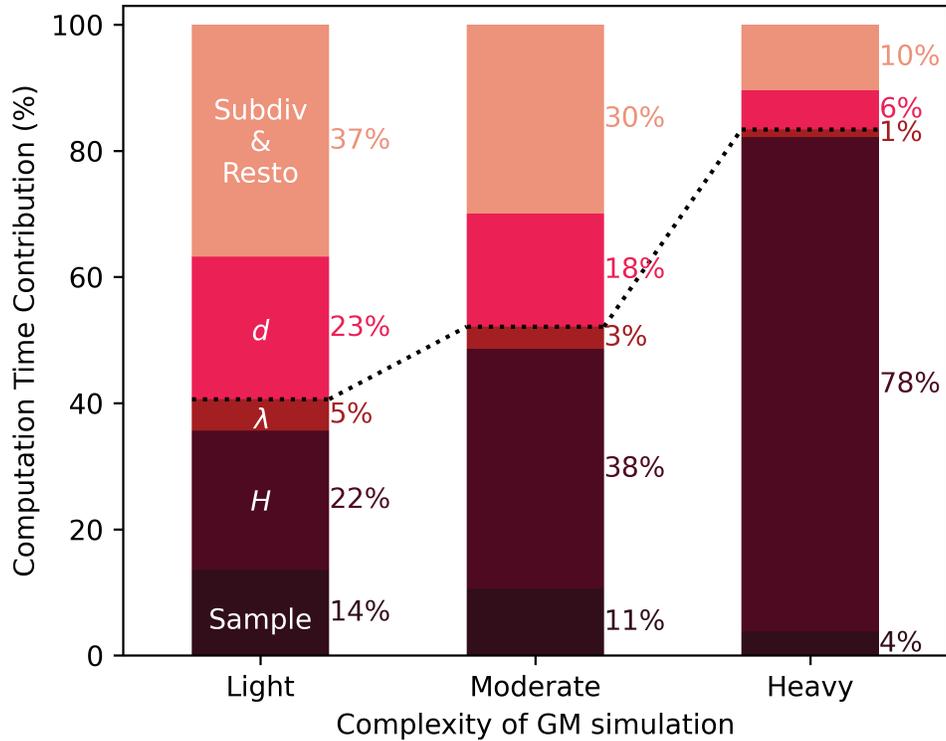


Figure S17. The contribution of each part of VEGAS AIS PSHA algorithm to the total computation time as a function of the computational complexity of ground motion simulation. The areal source example is used with $N_s=100$. “Light” is an as-is case, and for Moderate and Heavy cases, the ground motion simulation part is redundantly repeated 10 and 100 times. The black dotted line indicates the boundary between the “hazard computation” and “update propose density” parts. Sample: random sampling; H : computation of the PSHA integrand each sample, λ : computation of hazard; d : Computation of contribution of each VEGAS grid; Subdiv & Resto: Updating VEGAS grid by subdivision and restoration process. The first three parts correspond to the hazard computation part, which is also present in traditional Riemann and MC methods, and the latter two parts correspond to the updating proposed density part, which is a unique characteristic present in VEGAS AIS PSHA algorithm. As the complexity of the GM simulation increases, the updating portion decreases because the computational burden of the updating process is not a function of the complexity of the ground motion model, but is a function of the number of samples and VEGAS grids(N).