Addendum

The authors recommend the replacement of Sections 3.5–3.6 and Table 3.15 with the content of this addendum. Consequently, the recommendation is to replace the 13 models and their weights with the 17 models provided in Electronic Appendix A.

The development of weights (Table 3.15) follows the same procedure that was described in the original Section 3.7 of the report, with the final numbers updated for the 17 models (Electronic Appendix A).

Updated Sections 3.5 and 3.6

3.5 DISCRETIZATION OF GROUND-MOTION SPACE

3.5.1 Definition of Range in Ground-Motion Space

In the previous section, the visualization of the sampled models from the continuous ground-motion distribution $P(\mathbf{Y})$ was presented. This results in Sammon's maps, one for each frequency, which are a representation of the high-dimensional ground-motion model space: each map is an approximation of $P(\mathbf{Y})$. Since $P(Y)$ describes epistemic uncertainty, and the map is an approximation of $P(Y)$, then the center, body, and range (CBR) on the map is an approximation to CBR of $P(Y)$. Thus, defining the CBR of the models in the two-dimensional space (on the map) is an approximation to the center, body, and range of $P(\mathbf{Y})$ The definition of the range is done similar to a one-dimensional distribution; in that case, often a range of $\pm 2\sigma$ is chosen (in the case of a normal distribution). Such a range covers 95.45% of the total probability of a normal distribution:

$$\int_{-2}^{2} N(x;\mu=0,\sigma=1)\,dx = 0.9545 \tag{3.17}$$

However, the Sammon's maps represent a two-dimensional distribution. Based on the definition of a two-dimensional normal distribution and the distribution of the sampled models (gray points in Figure 3.49), an ellipse was selected to represent the range on the map. The half-axes of the ellipses are determined by the standard deviations of the distribution of points in the *x*- and *y*-direction. These are calculated and then scaled by a factor α such that the resulting ellipse covers 95% of the total probability of a two-dimensional normal distribution. The factor α is calculated in the following way: the probability density function of the two-dimensional normal distribution is converted to polar coordinates *r*, θ, and the angle θ is marginalized out. This results in a Rayleigh distribution, and the scaling factor α can be calculated from

$$p = F(s) = 1 - e^{\left(-s^2/\sigma^2\right)} \tag{3.18}$$

where *F* is the cumulative distribution of the Rayleigh distribution, and *p* is the amount of probability that should be enclosed by the ellipse (p = 0.95). The resulting scale factor is α = 2.45. This is equivalent to using a second order chi-squared distribution at the 0.95 bilateral level.

The ellipse is centered on the point {0, 0}, which corresponds to the mean of the seed GMMs (see Section 8.3). Figure 3.51 shows the ellipse defining the range for two frequencies.

This range covers large portions of the map, and thus also includes the ground-motion model space.

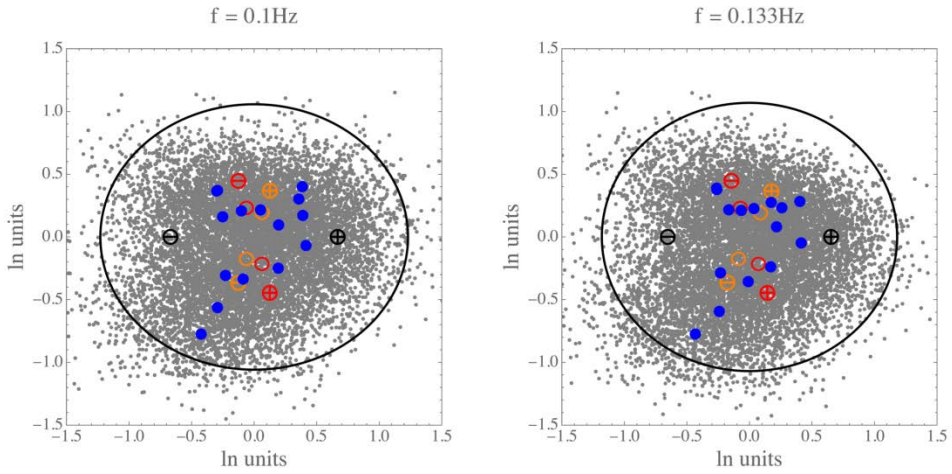

Figure 3.51 Sammon's maps for two different frequencies and 10,000 sampled models (gray points). The range defined by the TI team is a black ellipse.

3.5.2 Discretization of the Ground-Motion Space into Cells

With the range as defined in the previous section, the ellipse encloses the subset of ground-motion model space (Figure 3.51) that the TI team intends to capture. As described above, this range covers 95% of the total probability on the map. The range needs to be discretized into a manageable number of GMMs. Therefore, the ellipse defining the range is partitioned into several cells, and a representative model for each cell is developed.

The space inside the ellipse is further partitioned via two ellipses (forming rings) with different (smaller) scale factors. Thus, the considered range is partitioned into a central ellipse and two outer rings. These rings are further partitioned into eight cells, based on equal angular distances (45°), to capture epistemic uncertainty in magnitude and distance scaling. Thus, in total there are 17 cells: one central cell (ellipse), and eight cells each on the central and outer ring.

The scale factors to calculate the semi-axes of the inner ellipses are again based on a two-dimensional Gaussian distribution. The center cell represents the center, the middle ring the body, and the outer ring the range of the distribution. Based on a two-dimensional normal distribution, the center should correspond to 10% of the density, the body should capture 75% (including the center), and the full range should capture 95% of the distribution (as stated in the previous section). Thus, the body corresponds to 65% (75–10), and the range corresponds to 20%. Hence, the scale factors for the inner ellipses are calculated such that the cumulative distribution function of a two-dimensional normal distribution equals 0.1 and 0.75. The scale factors are calculated from the cumulative distribution function of the Rayleigh distribution according to Equation (3.18). The resulting scale factors are α = 0.46, 1.65 and p = 0.1, 0.75.

The TI team tested a different discretization, using the same outer ellipse and three rings to define a total of 13 and 29 models. The 29 models scheme was the original approach presented at the SSHAC workshops. The TI team concluded that there is not enough information to defend a more complex discretization, and that the 17 models scheme was appropriate (and practical) to capture a range of ground-motion values. The number of models is sufficient to capture alternate magnitude and distance scaling behaviors represented in the space away from the center. An example of the discretized ground-motion space is shown in Figure 3.52, for all frequencies. Figure 3.53 shows the fraction of areas of the different cells with respect to the cell in the center.

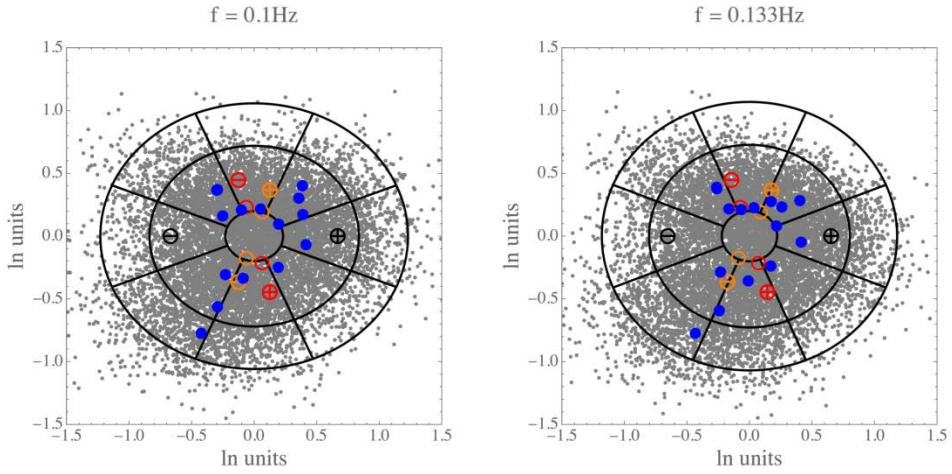

Figure 3.52 **Sammon's maps for two different frequencies and 10,000 sampled models (gray points). The partition of the ground-motion space is shown as black cells.**

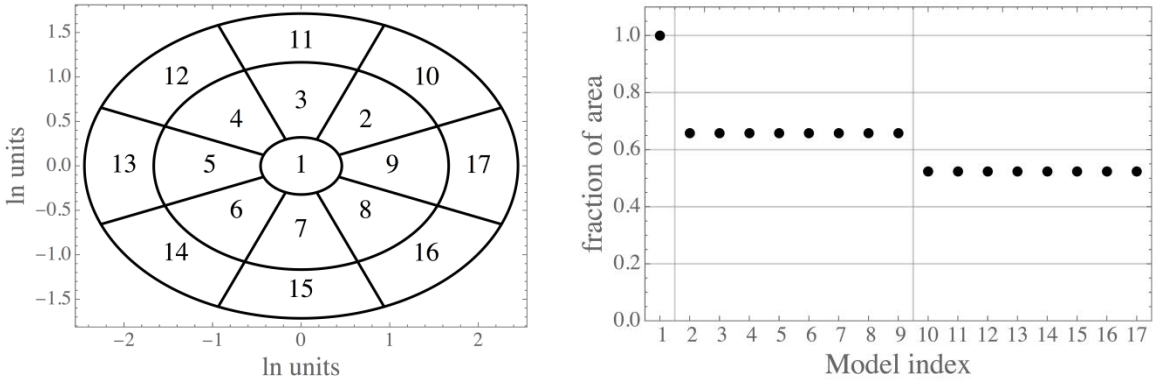

Figure 3.53 **Model indices of different cells (left) and fraction of area of different cells with respect to the center cell (right).**

3.5.3 Selection of Representative GMM for Each Cell

Each of the 17 cells defined above covers a fraction of the area on the map, which is the full representation of $P(\mathbf{Y})$. The next step is to define a representative GMM for each cell. Various candidate representative models were considered by the TI team, with preference given to an approximation to the expectation of $\boldsymbol{Y}$ over each cell. Since $P(\mathbf{Y})$ is a distribution over vectors of ground-motion estimates (an approximation of a continuous GMM), this results in a valid GMM. The approximation to the expectation is calculated by averaging over all models inside one cell:

$$y_k = \frac{1}{N_k}\sum_{i=1}^{N_k} y_i \tag{3.19}$$

where k indexes the cell, and N_k is the number of samples inside a cell. This is an approximation of:

$$E\left[Y\right]_{A_k} = \int_{A_k} y\,P\left(y\right)dA \tag{3.20}$$

The representative model for each cell is therefore an average of samples from $P(\mathbf{Y})$, which allows the representation of the systematic trends in that cell. For example, if the representative model was randomly selected from any model in the cell, it could lead to extreme model realizations and larger variations from frequency to frequency (and very jagged spectra). Because the maps are all oriented the same way, the average metric has the advantage of producing smoother spectra.

All the samples passed the criteria of physicality established by the TI team, ensuring that the selected models also pass the physicality constraints. Figure 3.54 shows the scaling of the 17 selected models against magnitude and distance.

The spectra of the selected models are shown in Figure 3.55, for a single ($\mathbf{M}$, R_{RUP}) values. For each frequency, the models with the same model index (see right-hand panel of Figure 3.53) are combined into one GMM. Considering that the process is performed for each frequency independently, the spectra are reasonably smooth. The smoothness is ensured by the fact that the Sammon's maps, on which the selected models are based, are all rotated and flipped in the same way. Figure 3.56 shows an example 3D plot for two different frequencies. The three axes correspond to ground motions from three different ($\mathbf{M}$, R_{RUP}) scenarios. The seeds and samples align into plane-like cloud structures, one for each frequency. These two planes cover different ground-motion values, but the location of the seeds and samples remain in similar positions relative to each plane. This is especially true for close-by frequencies and can explain the relative smoothness of the spectra. Once the high-dimensional space is mapped in two dimensions, the structure is preserved across frequencies by the rotation and reflection of the Sammon's maps in a consistent way. Hence, the regions on the Sammon's maps for the different frequencies correspond roughly to the same scaling properties. However, smoother individual models may be desired for specific applications. The smoothing of the 17 models is described in the next section.

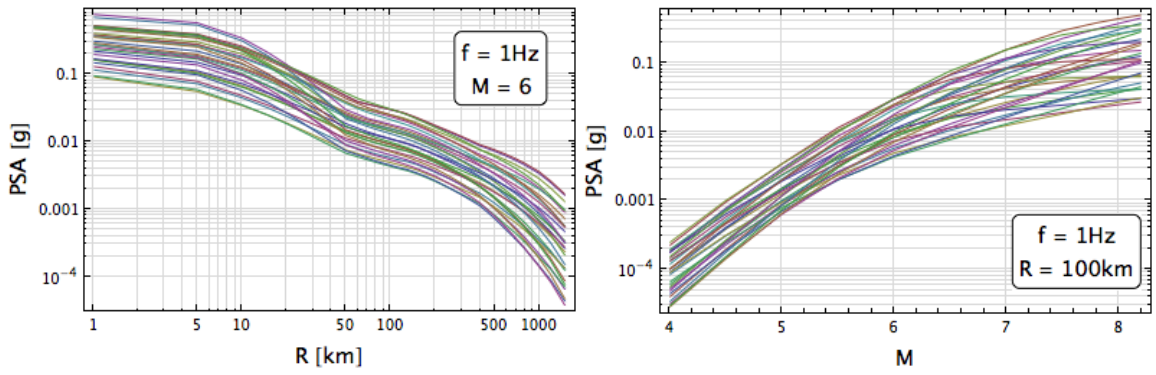

Figure 3.54 Distance and magnitude scaling of selected models for f = 1 Hz.

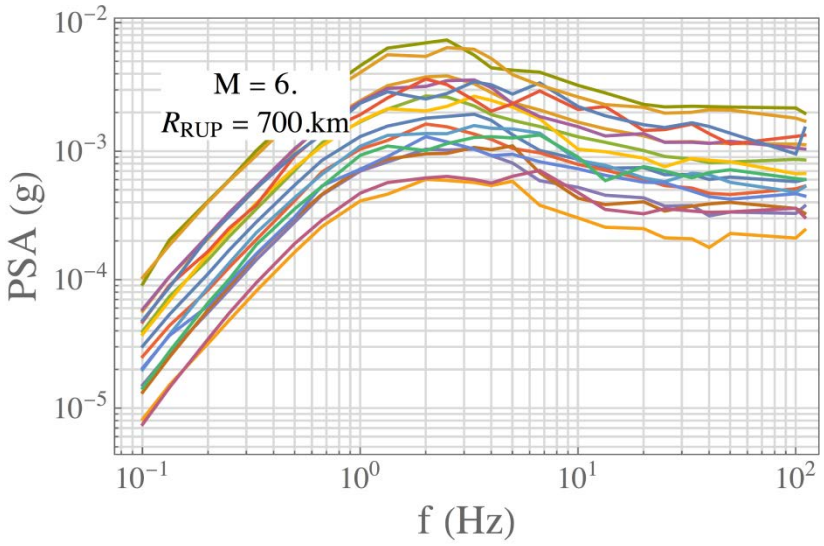

Figure 3.55 Spectra of 17 selected models for a single (M, R) scenario.

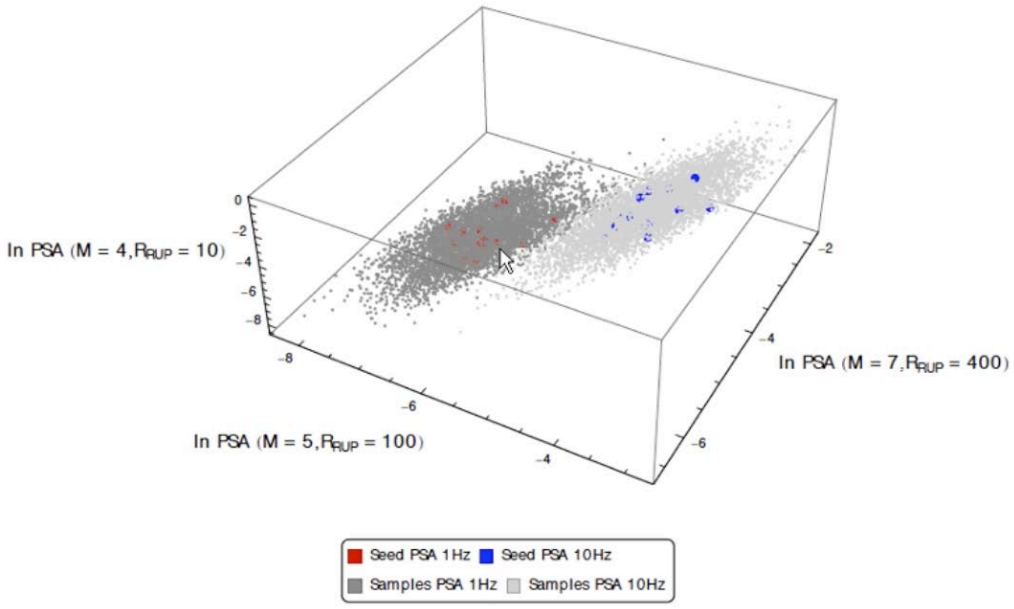

Figure 3.56 Three-dimensional ground-motion space for seeds and samples at two frequencies (1 and 10 Hz).

3.6 FINAL MODELS AND SMOOTHING PROCESS

Figure 3.55 shows that although the spectra are relatively smooth, they may be too jagged for certain applications. This is especially true for models near the tail ends of the distributions (the outer cells in ellipse), which are averaged over a smaller number of models in each cell. Therefore, the TI team decided to smooth the 17 selected models to ensure a reasonable expected shape of all spectra for all (**M**, R_{RUP}) scenarios.

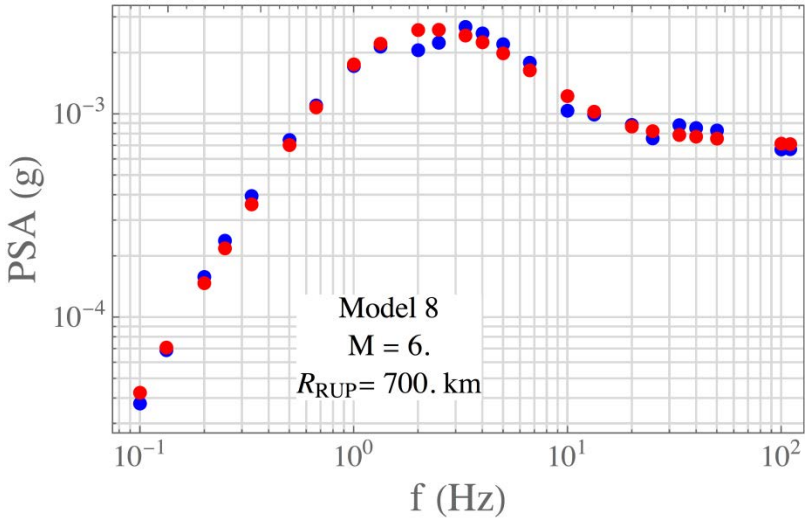

Figure 3.57 Example of spectrum smoothing (blue is raw, red is smoothed) for an individual model and a single (M, R) scenario.

The smoothing is done by fitting each of the 17 models to a function that depends on frequency, magnitude and distance.

$$Y_k = g_k\left(f, \mathbf{M}, R_{\text{RUP}}\right) \tag{3.21}$$

where Y is the logarithmic PSA value at one of the 24 NGA East frequencies (f = 0.1 to 100 Hz plus PGA as described above), and k indexes the 17 models. The function $g\left(f, \mathbf{M}, R_{\text{RUP}}\right)$ has the following form (modified from McGuire et al. [2001]):

$$g(f, \mathbf{M}, R_{\text{RUP}}) = a_0 + \frac{a_1}{\cosh[\exp(a_2) f^{a_3}]} + \frac{\exp(a_5 f)}{f^{a_6}} \tag{3.22}$$

where some of the coefficients a* depend on magnitude and distance. The dependence of the coefficients on **M** and R_{RUP} is as follows

$$\begin{aligned}
a_4 &= c_{40} + \mathbf{M}\left[c_{41} + c_{42}\,\mathbf{M} + c_{43}\ln(c_{44}R_{RUP} + 1)\right] \\
a_6 &= c_{60} + c_{61}R_{\text{RUP}} \\
a_7 &= \exp\left[c_{70} + \mathbf{M}(c_{71} + c_{72}\,\mathbf{M} + c_{73}R_{\text{RUP}})\right] \\
a_8 &= c_{80} + +\mathbf{M}(c_{81} + c_{82}R_{RUP}) \\
a_9 &= c_{90} + c_{91}\,\mathbf{M}
\end{aligned} \tag{3.23}$$

where the coefficients c* were estimated. The other coefficients (a_1, a_2, a_3, a_5) were held constant across all magnitudes and distances. The coefficient a_0 corresponds to the PGA value. It is different for each of the 374 **M**-R_{RUP} scenarios.

All coefficients (a*,c*) were estimated via Bayesian inference using the program Stan [Carpenter et al. 2017; Stan Development Team 2017]. The prior distributions for the coefficients θ (save a_0) were set to be weakly informative; they are normal distributions with mean zero and standard deviation 10, or $\theta \sim N(0,10)$.

The prior distribution for a_0 is a normal distribution whose mean is the PGA value of the unsmoothed model, with a standard deviation of 0.1. This ensures that the estimated coefficient a_0 is similar to the unsmoothed PGA value.

The parameters are estimated by maximum-a-posteriori (MAP) optimization. Since the inference is sensitive to the starting values, four different starting values were used, and the final model with the highest log-probability was used.

Figure 3.57 shows the fit of one particular spectrum, and Figure 3.57 shows the smoothed version of the spectra from Figure 3.55. Figure 3.59 compares the hazard curve distribution for f = 1 Hz, calculated for the smoothed and the unsmoothed models. Figure 3.60 shows a sample uniform hazard spectrum (UHS), calculated for both the smoothed and unsmoothed models. Although there are some differences at some frequencies, there is no defensible argument to maintain the “kinks,” which are relatively small. The benefit of smooth spectra was deemed to outweigh benefits of preserving the actual spectral shapes.

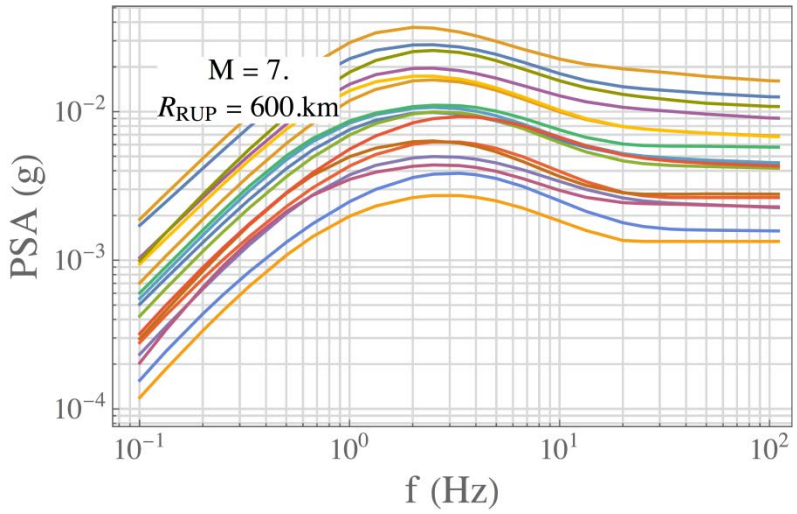

Figure 3.58 **Smoothed spectra of 17 selected models for a single (M, R) scenario (same models as in Figure 3.43).**

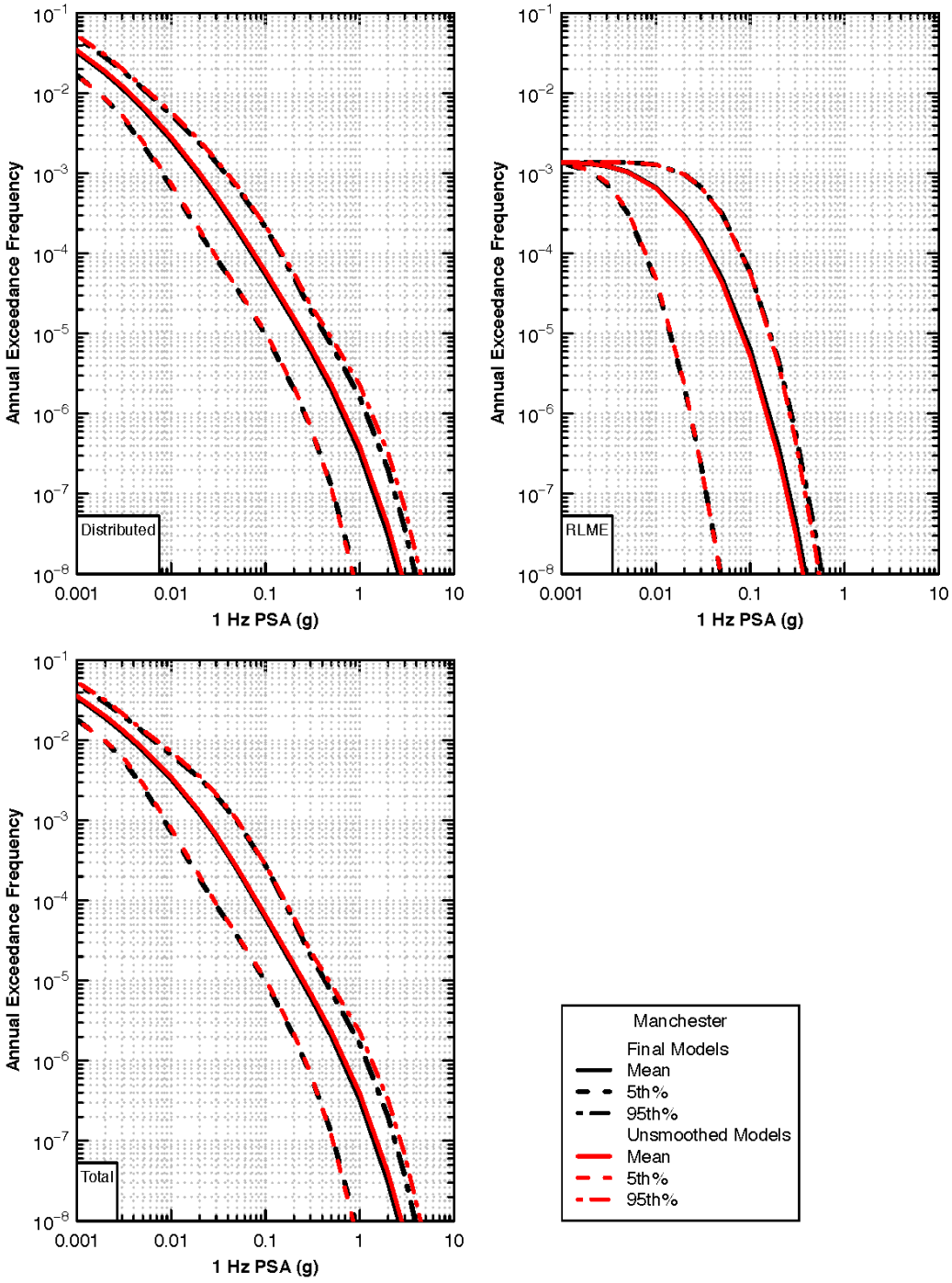

Figure 3.59 **Sample 1 Hz hazard curves comparing results from the as-is and smoothed spectra, Manchester site.**

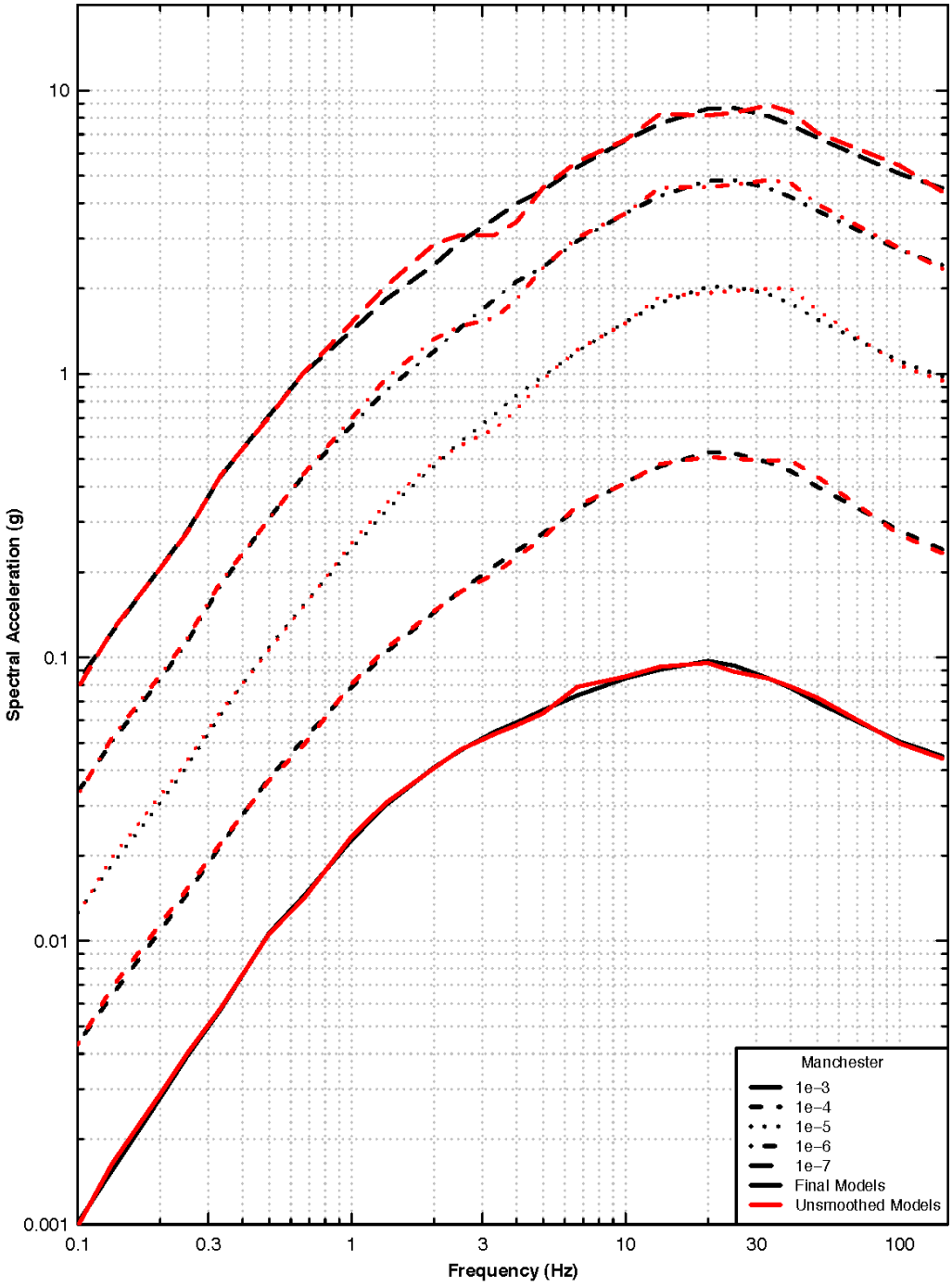

Figure 3.60 **UHRS comparing results from the as-is and smoothed spectra for various hazard levels, Manchester site.**

Table 3.15 Total weights for the 17 models, for all the GMIMs (oscillator frequencies, *f*, in Hz).

	f=4	*f*=5	*f*=6.667	*f*=10	*f*=13.333	*f*=20	*f*=25	*f*=33.333	*f*=40	*f*=50	*f*=100	PGA	PGV
Model 1	0.0921	0.0737	0.0683	0.1047	0.1068	0.0998	0.1069	0.1078	0.0987	0.0949	0.0935	0.1009	0.0976
Model 2	0.0585	0.0994	0.153	0.1175	0.1311	0.1315	0.1256	0.1316	0.1453	0.1176	0.1462	0.1606	0.0678
Model 3	0.0632	0.0892	0.0863	0.0723	0.0697	0.0965	0.088	0.0883	0.0996	0.0985	0.123	0.1151	0.0738
Model 4	0.0739	0.0691	0.0834	0.0676	0.0651	0.0686	0.068	0.0673	0.0653	0.0704	0.0981	0.097	0.0756
Model 5	0.0731	0.0456	0.0342	0.0677	0.0735	0.054	0.0579	0.0512	0.0396	0.0407	0.0472	0.0548	0.0702
Model 6	0.0965	0.1095	0.096	0.0553	0.0519	0.0559	0.06	0.0509	0.062	0.0666	0.033	0.0376	0.0916
Model 7	0.1198	0.102	0.0894	0.0725	0.0917	0.0649	0.0586	0.0627	0.0609	0.0643	0.0522	0.0507	0.098
Model 8	0.1123	0.0876	0.055	0.0642	0.0506	0.0743	0.0784	0.0727	0.0838	0.0984	0.0629	0.0497	0.1054
Model 9	0.0774	0.0859	0.086	0.1075	0.0938	0.1136	0.1221	0.1205	0.1057	0.1064	0.1092	0.0986	0.0956
Model 10	0.0123	0.0281	0.0212	0.0254	0.019	0.0374	0.0298	0.0245	0.0278	0.0246	0.0372	0.0372	0.0108
Model 11	0.0185	0.0214	0.0056	0.0088	0.0008	0.0191	0.0087	0.0016	0.003	0.0147	0.0123	0.01	0.0197
Model 12	0.0344	0.0293	0.0139	0.0175	0.0048	0.0178	0.0139	0.0055	0.0059	0.0174	0.0271	0.0167	0.0274
Model 13	0.0191	0.0176	0.0109	0.0158	0.0126	0.0117	0.0082	0.0057	0.0045	0.0098	0.0076	0.0119	0.0117
Model 14	0.0208	0.0366	0.063	0.082	0.098	0.0406	0.0473	0.0703	0.0648	0.0512	0.0368	0.0436	0.0257
Model 15	0.0418	0.0368	0.0688	0.0649	0.08	0.043	0.0549	0.0758	0.0626	0.0466	0.0418	0.0504	0.0365
Model 16	0.0606	0.0328	0.0196	0.0179	0.0226	0.0268	0.0272	0.0236	0.0345	0.0419	0.0266	0.0282	0.0567
Model 17	0.0257	0.0354	0.0454	0.0384	0.028	0.0445	0.0445	0.04	0.036	0.036	0.0453	0.037	0.0359