12th CANADIAN CONFERENCE ON EARTHQUAKE ENGINEERING
JUNE 17-20, 2019
CHÂTEAU FRONTENAC
QUÉBEC, QC.
12ème CONFÉRENCE CANADIENNE DU GÉNIE PARASISMIQUE
17-20 JUIN 2019
CCEE 2019

Uncertainty spread in the 5th Generation seismic hazard results used in NBCC2015

John Adams[1], Stephen Halchuk[1]
[1]Seismologist, Natural Resources Canada, Ottawa, Canada

ABSTRACT

The 2015 National Building Code of Canada (NBCC) uses hazard results calculated using the GSC's 5th Generation hazard model, but uses just a single value, the mean. However, the mean hazard value has uncertainty associated with it, and the amount of uncertainty has implications for the reliability of engineering designs. Uncertainty arises from aleatory and epistemic uncertainty in the model. The uncertainty spread is illustrated through the use of percentiles (5, 16, 50 (median), 84, 95, etc; also called fractiles) of the distribution. The mean hazard values, as used in NBCC2015, correspond to the 60-85th percentiles of the distribution. Calculated distributions are provided for La Malbaie, Montreal, Toronto, Tofino, Victoria, Vancouver, and Kelowna to compare their spreads. The spreads are large, and for most localities can be approximated to a lognormal distribution. The slope of the cumulative distribution (proportional to standard deviation) can be similar (or different) for all periods at a single site, but differs greatly between sites. It is largest for low-seismicity eastern sites and smallest for sites above the Cascadia subduction zone. The main contributors to the spread is the uncertainty in the Ground Motion Models (for most places in Canada) and the uncertainty in the earthquake rates for low-hazard regions.

Keywords: uncertainty, seismic hazard, Canada, National Building Code, percentiles

INTRODUCTION

Aleatory uncertainty has been included in National Building Code of Canada (NBCC) hazard estimates since the 3rd Generation model (NBCC1985). Treatment of epistemic uncertainty has improved from the 3rd Generation, which was a "best estimate" with no epistemic uncertainty [1], to the 4th Generation (NBCC2005) which was not a fully probabilistic model since it used the "robust" combination of earthquake sources but included epistemic uncertainty for the first time [2]. The 5th Generation is a fully probabilistic model including both aleatory and epistemic uncertainty [3,4]. Epistemic uncertainty is sampled by using 3-branch representations of the uncertainty in: Ground Motion Model, magnitude-recurrence, maximum magnitude, earthquake depth, and alternative regional source models. While any percentile of the distribution can be provided, the mean is the only parameter from the national seismic hazard model that is generally used.

This paper investigates the hazard distributions about the mean (i.e. the percentiles or fractiles or quantiles of the distribution) revealed by computing the distributions of 5%-damped spectral acceleration, Sa(T) (where T = period in seconds), using the GSCFRISK and OpenQuake (OQ; [5]) computational codes. The goal is (i) to represent the distribution about the mean as a way of quantifying the uncertainty of the 5th Generation model used for NBCC2015, and (ii) to confirm that the OQ code produces similar results to our legacy code, GSCFRISK, hence further validating OQ's use for the 6th Generation model proposed for NBCC2020 [6].

HAZARD CODES USED

GSCFRISK is a derivation of the FRISK88 program (a proprietary product of Risk Engineering Inc) that was modified to perform deaggregations of the mean and to output percentiles of the distribution. The program determines a smoothed representation of the percentile values by determining a weighted average of branch results that fall in the vicinity of the desired percentile. The averaging is performed for a specified range of percentiles about the desired percentile (in the tests done for this document the range is ± 5%). The branches immediately above and below this range are also considered at half weight if they fall within a "tolerance" interval above or below the range (± 0.5% for this study).

The replacement code, OQ, uses a set of values for the quantile_hazard_curves parameter to produces quantile values not only in the form of hazard curves but also UHS values for individual quantiles (quantiles are considered to be the same as percentiles, at least in the manner that the terms are used in describing Canadian hazard distributions). However, OQ does not treat depth uncertainty in an epistemic sense, and this may cause some of the differences seen below.

Most of the hazard results in this paper comprise Sa(0.2) values for the 2%/50 year probability on Site Class C. Because of the linear F(T) amplification factors used in NBCC2015 [7], the lognormal spread of the hazard (see below) will be the same for all Site Classes under the NBCC2015. That is, the F(T) are simple scalars, albeit ones that change with ground motion intensity. Explicitly, no extra uncertainty is added into the hazard used for soils other than Site Class C.

EASTERN RESULTS

The 5th Generation model was run with GSCFRISK for Montreal, La Malbaie and Toronto. Montreal is a typical eastern moderate-hazard site, La Malbaie is in Charlevoix, one of the highest hazard regions in eastern Canada, and Toronto represents the low-moderate hazard regions. The approximate position of the mean is indicated by the heavy weighted dots on each curve on Figures 1 and 3.

The spread for each locality in Figure 1A (linear-scale for ground motion) looks rather different, but in Figure 1B (using log-scale for ground motion), they are seen to be rather similar. Therefore the spread is approximately lognormal for all three localities. Notice that the equivalent percentile for each mean value decreases from Toronto to Montreal to La Malbaie.

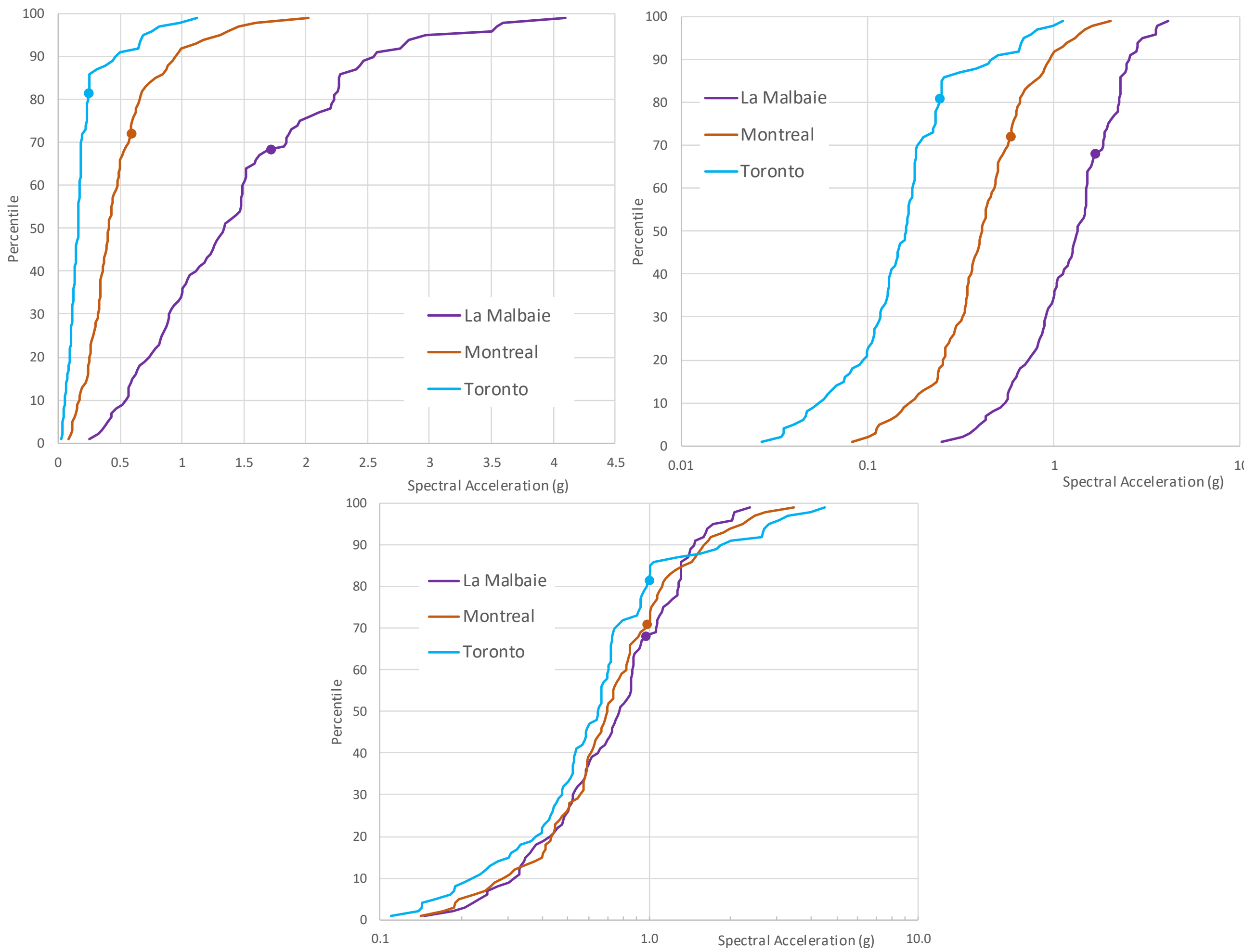


Figure 1. A: (upper left) Percentiles determined for La Malbaie, Montreal and Toronto using GSCFRISK and the 5th Generation model. B: (upper right) As for Figure 1A but using logarithmic spectral acceleration scale. C: (bottom) Same data as 1B, but normalized to a mean value of 1 g to emphasize the similar shape of the curves.

To emphasize the similarities, Figure 1C normalizes the three curves results by the mean hazard. The curves are very similar for all three cities, despite their hazard-level differences. There are small differences in total spread, with Toronto>Montreal>La Malbaie, as expected because of the larger uncertainty in the earthquake rates for the lower-seismicity regions. However, the close similarity of the curves suggests that much of the spread comes from the Ground Motion Models (GMMs), which are common to all eastern cities [8].

Figure 2A shows the suite of Montreal results for various Sa(T) for Class C. There is clearly less variability within the suite than between the eastern cities. The Sa(0.2) curve is the same as for Figure 1, but in Figure 2A the scale of the vertical axis has been adjusted from simple percentile to Z-score [9] assuming a lognormal distribution. This turns the curves from being sigmoidal to nearly straight lines, confirming the lognormal assumption. Although there are small deviations near the tails, the overall distributions for Sa(T) are very close to straight lines and have similar slopes. The similarity of the slopes is reinforced by Figure 2B, which normalizes the curves results by the mean hazard of each. The results from Figures 1C and 2B suggest that a representative curve, adjusted for each Sa(T)'s mean hazard, could be used for all periods at a site in eastern Canada, but that no single curve will capture the differences between sites.

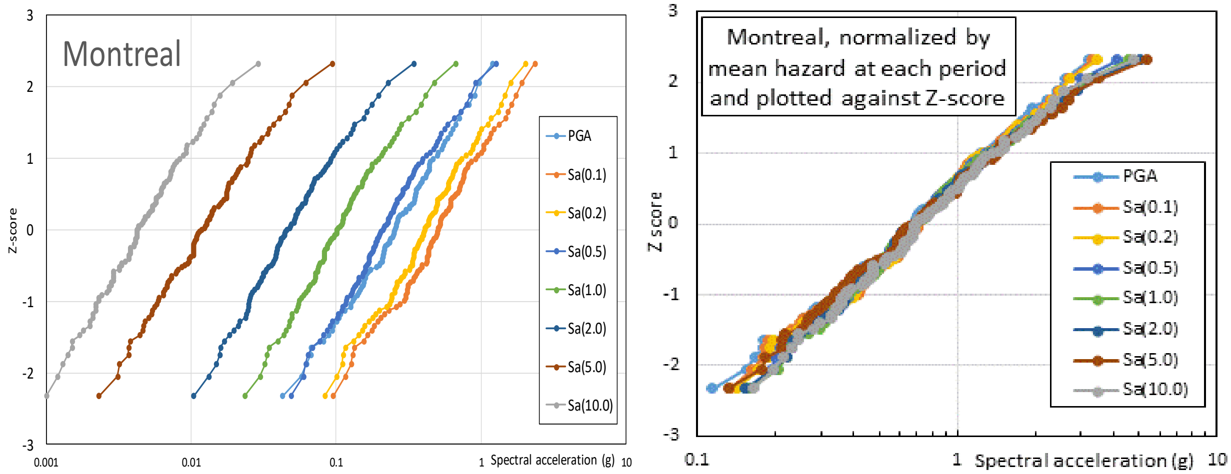


Figure 2. A: (left) Percentiles of 2%/50 year seismic hazard at Montreal for Site Class C with the percentiles plotted as the Z-score of a normal distribution. B: (right) Same data as Figure 2, but normalized to a mean value of 1 g to emphasize the similar slope of the curves.

CONFIRMATION THAT GSCFRISK DISTRIBUTIONS CAN BE MATCHED BY OQ

A suite of values for the same eastern cities were generated in OQ using NRCan's implementation [10] of the 5th Generation model. Note that our implementation differs from the open-source one used to generate the GEM global hazard map [11]. Equivalent models were run with GSCFRISK for Montreal, La Malbaie and (not shown) Toronto.

In Figure 3A the OQ (grey) curve does a good job of matching the GSCFRISK curve (orange) for Montreal. The approximate location of the mean is again indicated by the heavy weighted dots on each curve. The spread of the results closely matches GSCFRISK, though the OQ curve is still more stepped than the GSCFRISK curve. Figure 3A also shows the result (blue curve) of collapsing the 3-branch representation of the magnitude-frequency distribution to a single representation. Such a simplified representation was required to reduce the computer run times to practicable durations for the early calculation of the 6th Generation hazard results [6] proposed for NBCC2020. It can be seen that the mean values are replicated, but as expected the spread about the mean is considerably reduced. Using the collapsed model proved to be an acceptable simplification for the practical use of OQ for calculating mean hazard values from the 6th Generation model, where run times were exceeding days. However, while the simplification reduced the run time considerably, it is unsatisfactory for generating percentiles of the distribution.

For La Malbaie (Fig. 3B) the agreement was not as good as for Montreal. To improve the agreement the "area source discretization parameter" in OQ was reduced from 20 to 5 km; this increased the run time from ~18 hours to ~10 days. The spread of the distribution curve was similar, though it remained offset from the GSCFRISK curve. Further investigations showed that the hazard values, and the spread of the distribution, can be very sensitive to the "area source discretization parameter", especially for high-hazard regions such as Charlevoix that have spatially-concentrated frequent earthquakes. Figure 4 displays the different distributions for La Pocatière, another site near Charlevoix. It can be seen that the overall spread of the percentiles is slightly smaller for OQ and that the differences are largest near the upper tail. For three other Charlevoix sites (not shown), the OQ percentiles have higher hazard than the GSCFRISK ones, but the over-estimate is reduced when 5-km discretization is used instead of 20-km discretization. For La Pocatière the situation is different: the hazard jumps from lower-than to higher-than GSCFRISK when going from 20-km to 5-km discretization (Fig. 4). There is also a reduction from 20% to 3% in the discrepancy to the GSCFRISK mean values, the latter being at our accepted tolerance from computational engines. As a result of these tests, an area source discretization of 10 km was adopted for the 6th Generation model; this is a pragmatic compromise for calculating mean values on a national scale.

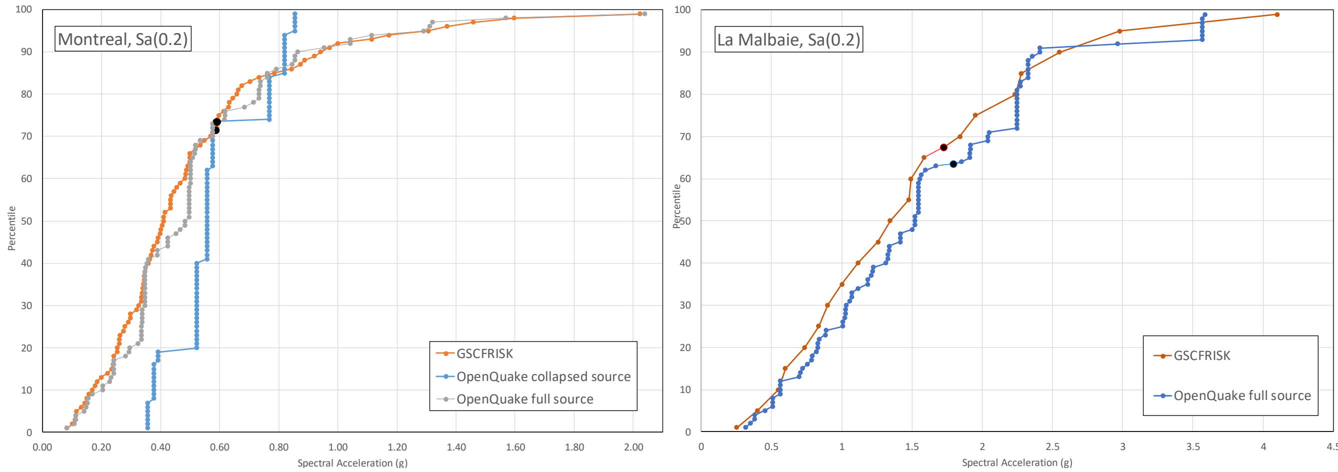


Figure 3. A: (left) Percentiles for Montreal – curves from Fig. 1 plus OpenQuake full-source model (grey curve) and collapsed-rate curve (blue). B: (right) Percentiles for La Malbaie - curves from Fig. 1 plus 5-km discretized OpenQuake model (dark blue).

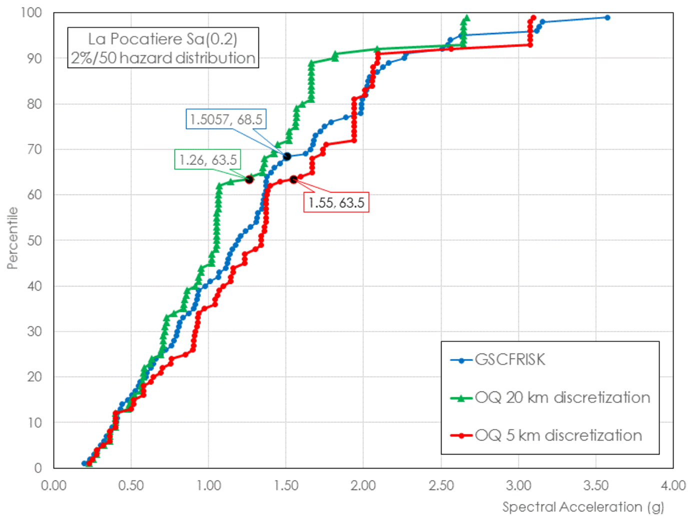


Figure 4. Open Quake curves for La Pocatière, Charlevoix run with 20-km and 5-km discretization, compared to GSCFRISK curve. Note sensitivity of OQ percentiles to the discretization parameter.

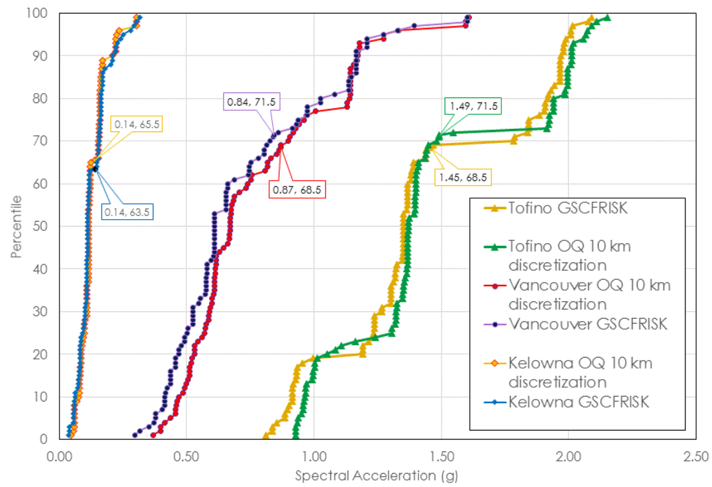
WESTERN RESULTS

Equivalent OQ results were generated for western cities Tofino, Vancouver, and Kelowna (Fig. 5) to compare with the results from GSCFRISK. Tofino represents a high-hazard locality immediately above the Cascadia subduction interface and the hazard is almost exclusively from great earthquakes on the interface. Victoria and Vancouver have high hazard, with the short-period hazard coming chiefly from the inslab earthquakes and the longer period from the great Cascadia earthquakes. Kelowna is a low-hazard region in the interior of British Columbia where the short-period hazard is dominated by local active crustal sources. The agreement of GSCFRISK and OQ curves is a little unsatisfactory for Vancouver, as OQ gives higher hazard for most of the central percentiles (+3.3% at the mean relative to GSCFRISK, which is slightly above our desired tolerance of 3%). At Tofino and Kelowna the agreement is acceptable (despite a 20-km discretization). The Vancouver disagreement may be due to a mismatch between the OQ discretization and the finer GSCFRISK slicing parameter (as for Charlevoix). A feature of these western curves is their step-like nature, especially evident for Tofino (see discussion below); the steps are more pronounced than for the eastern curves (Fig. 1C). It appears that the 3-branch representations (specifically, of the GMMs – see below) within the logic tree are too coarse to provide a smooth curve for some sites.

Overlooking the small differences in Figure 5, the combined western plot (Fig. 6A) and the normalized curves (Fig. 6B) show a clear pattern. Note the smaller variation in the distribution (i.e. steeper slope) for sites closer to the subduction zone (Tofino << Victoria ≈ Vancouver < Kelowna) indicating less uncertainty close to the well-defined active subduction interface (especially the low uncertainty in rates) versus higher uncertainty in the low-seismicity regions. However, this does not translate into the mean value being a higher percentile, as Kelowna (63.5th) is smaller than Tofino (68.5th). The explanation likely lies in the relative asymmetry of the distributions, with Kelowna having more of a low-tail than Tofino. We looked at the spread of the Sa(T) percentiles for Vancouver and found that there is greater variability in their slope than for the Montreal ones.

Long-period Sa(T) have a smaller spread than for short periods. In logarithmic standard deviation terms (see later) the spread for Montreal is ~0.31 ($\log_{10}$ units) for all periods but for Vancouver varies from 0.169 at T=0.2 s to 0.144 at T= 10 s. The Vancouver variation probably arises from different amounts of epistemic uncertainty between inslab and interface GMMs of Atkinson and Adams [8] (which contribute chiefly short-period and long-period hazards respectively), and within each suite of GMMs for different periods.

Figure 5. Comparison of GSCFRISK and OQ percentiles for three western sites.



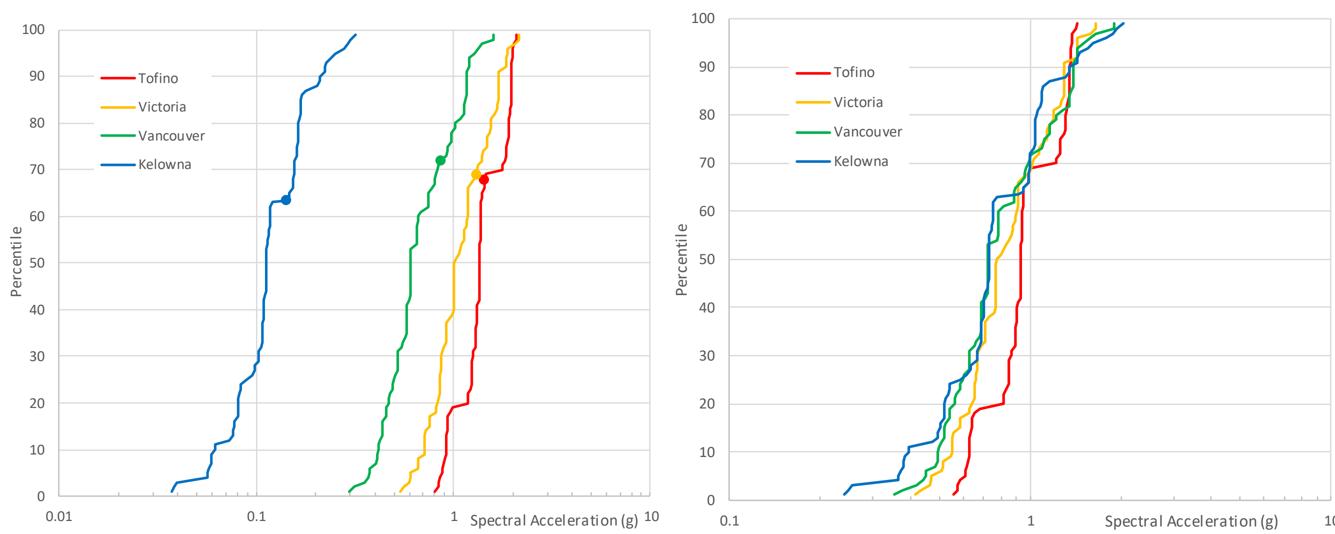


Figure 6 A: (left) Percentiles determined for Tofino, Victoria, Vancouver, and Kelowna hazard (Sa(0.2) for probability 2%/50 yr on Site Class C) using GSCFRISK 5th Generation model. Dots identify the mean values. B: (right) normalized to a mean value of 1 g to emphasize the generally similar shape of the curves but differences in hazard uncertainty (see text).

DISCUSSION

Comparison of percentile distributions

Figure 7 gives the relative distributions for all 7 localities studied, normalized to 1 g and plotted against the Z-score to display their lognormal distribution. There is a considerable range in slope represented, from a steep slope for Tofino to the flattest slope for Toronto. This range reflects the smaller amount of uncertainty in Tofino relative to the larger amount in Toronto. It is quite possible that the range represented covers most (but not all) sites in Canada.

Origin of the spread in hazard

We have looked at the origin of the spread in hazard for just Tofino, as it has the lowest spread. Figure 8 shows the total hazard spread in black compared to the hazard spread that results when each of its contributors is successively allowed to be the sole contributor. Most of the epistemic uncertainty is captured by 3-branch alternatives within the 5th Generation model [3], and the contributions can be discerned by the small or large horizontal "treads" in the upward-stepping slopes. The spreads arising from uncertainty in the magnitude-recurrence parameters (used to represent earthquake rates), in the largest considered magnitude, Mmax, and in the depth range assigned are very small (near-vertical curves), while the uncertainty in GMMs makes up almost the entire contribution to the total hazard. Two things are apparent from Figure 8: a) that most of the uncertainty arises from epistemic uncertainty in the GMMs, and b) to obtain a smoother, less-stepped spread curve would involve a N-branch representation of the GMMs, and perhaps of all other variables in the model, where N>>3. However, such a refinement would greatly increase the computation time, which might be acceptable for a site-specific study but is not for national hazard mapping. We are not surprised that the GMMs are a major contributor to the hazard spread at Tofino, and expect that they would be major contributors to the hazard spread across Canada. This may be especially true in eastern Canada where a single, 3-branched GMM is used for the stable crust (instead of three, 3-branch GMM models in southwest BC). Indeed the broad similarity in the slopes for the eastern sites (Fig. 1C) is an indication that the GMMs make a large contribution to the spread, probably chiefly supplemented by uncertainty in the magnitude-recurrence relations.

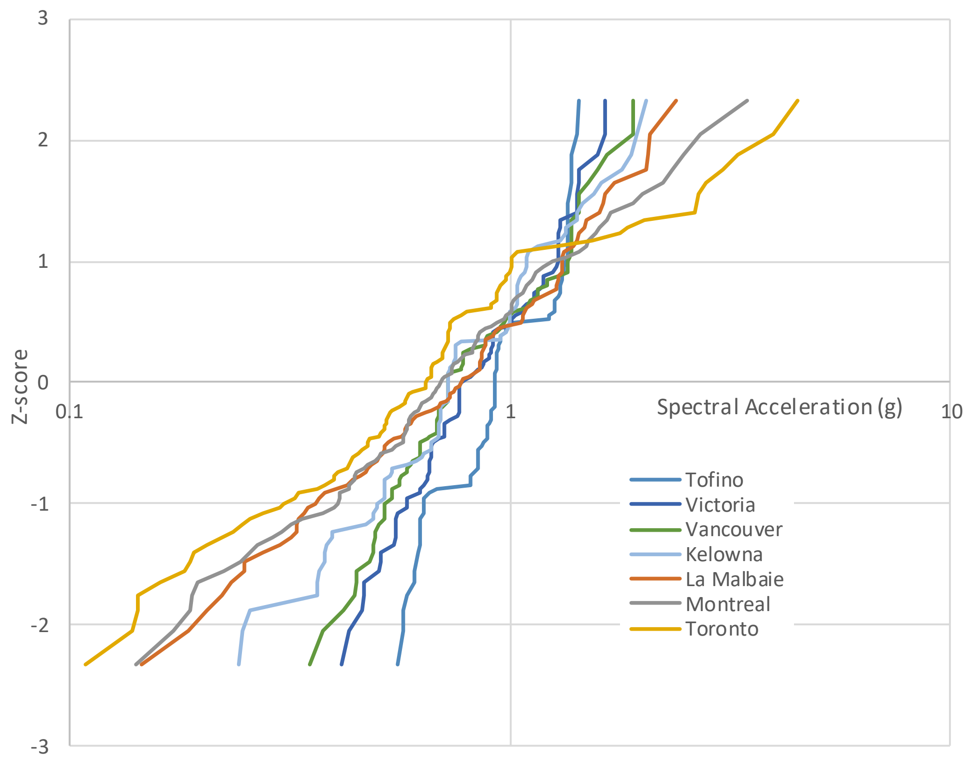


Figure 7. Percentile distributions for the 7 localities considered, normalized to 1 g and plotted against the Z-score. Note that the order of the curves in the legend applies to their order from left to right across the top of the figure.

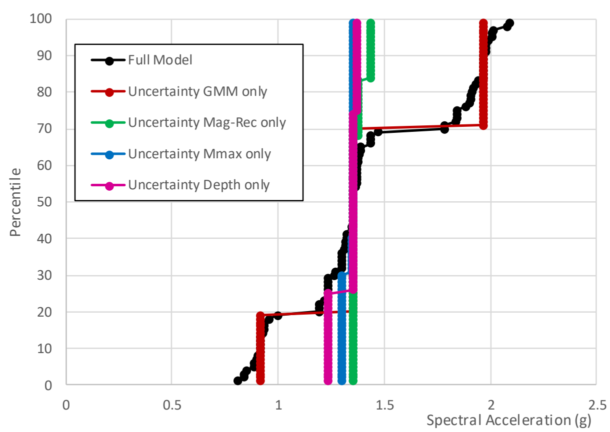


Figure 8. Unbundling of the hazard contributions to the total spread at Tofino (Sa(0.2) for probability 2%/50yr on Site Class C). See text for discussion.

Parameters to quantify the distributions

We examined the consistency of the slopes from Figure 7, and concluded that an acceptable parameter to quantify the slope of the line is the difference between the 2nd and 98th percentile hazard values. This represents the range Z=-2 to Z= +2, and so equals 4 times the standard deviation. Because of the lognormal nature of the distribution, the standard deviation is reported in $\log_{10}$ units. Table 1 gives the parameters summarizing the distribution of Sa(0.2) hazard, including the logarithms of the standard deviation. As seen on Figure 7, the logarithmic standard deviations increase from Tofino to Toronto. As we have four estimates of the standard deviation (Z=-2 to Z=-1, Z=-1 to Z=0, etc) we can estimate the uncertainty of using any one of these intervals to estimate the population's value. We express that as the standard deviation of the four estimated standard deviations, and more clearly as the variability of the standard deviation as a percentage of the standard deviation. The variability is highest for Tofino, which is not surprising given the stepped nature of its curve. The variability is relatively low for the rest of the cities (except Toronto, which has a long upper tail). The bottom rows of Table 1 show the consequence of this variability. Here we compare the 84th percentile directly calculated from the model to that estimated from using the median hazard and standard deviation. While most of the estimates are within ±15%, the Toronto estimate is off by 40%. The corollary is that using hazard estimates of just the median and 84th percentile can give a poor estimate of the overall distribution of the hazard percentiles.

Table 1. Parameters summarizing the distribution of Sa(0.2) hazard for 2%/50 yr on Site Class C and estimation of 84th percentile from calculated mean and standard deviation

	Tofino	Victoria	Vancouver	Kelowna	Toronto	Montreal	La Malbaie
mean hazard (g)	**1.459**	**1.304**	**0.842**	**0.142**	**0.248**	**0.587**	**1.724**
percentile for mean	**68.5**	**68.5**	**71.5**	**63.5**	**81.5**	**71.5**	**68.5**
2nd percentile (g)	0.838	0.565	0.315	0.039	0.035	0.101	0.322
16th percentile (g)	0.933	0.762	0.438	0.078	0.076	0.238	0.626
50th percentile (g)	1.351	1.008	0.611	0.113	0.160	0.409	1.341
84th percentile (g)	1.965	1.682	1.137	0.169	0.251	0.734	2.274
98th percentile (g)	2.077	2.156	1.603	0.303	0.987	1.597	3.594
Average SD ($\log_{10}$)	**0.099**	**0.145**	**0.177**	**0.224**	**0.364**	**0.300**	**0.262**
SD of average SD ($\log_{10}$)	0.064	0.045	0.054	0.058	0.145	0.057	0.051
Variability in SD	65%	31%	31%	26%	40%	19%	20%
Estimated 84th percentile	1.695	1.408	0.917	0.189	0.371	0.816	2.452
Calculated 84th percentile	1.965	1.682	1.137	0.169	0.251	0.734	2.274
Percent deviation	-14%	-16%	-19%	12%	48%	11%	8%

Note: SD= Standard Deviation. Average SD (in log units) = ($\log_{10}$(hazard at 98th percentile)-$\log_{10}$(hazard at 2th percentile))/4

We consider that a good estimate of the spread of the hazard values can be obtained by using the 2nd and 98th percentiles, while using some combination of the calculated 50th and 84th percentiles and/or the mean can be misleading. The misleading results arise from the stepped nature of the curve, which is a direct result of the necessary crudeness of the national model for capturing the epistemic uncertainty. The choice of the estimation method is important because the nature of the upper tail of the hazard distribution strongly influences the intended reliability of structural designs. On a national scale an adequate distribution parameter (standard deviation, in log units) could be obtained by contouring one quarter of the difference between the logarithms of the 2nd and the 98th percentiles on a fairly coarse grid and interpolating for a required site. This would probably work for all periods in eastern Canada, but would need some adjustment for different periods in western Canada.

Comparable work on uncertainty in contemporary U.S. national seismic hazard maps

Both Canadian and U.S. seismic hazard maps have the same goal of providing mean hazard at 2%/50 yr for a range of Sa(T), and it is to be hoped that their models reflect a similar amount of epistemic uncertainty. A study by Lee et al. [12] uses an innovative logic tree sampling to replicate the 2014 U.S. NSHMP mean hazard curves and also to sample and quantify their uncertainty. Like our paper, they conclude that the epistemic uncertainty can be best expressed by the lognormal standard deviation, which they obtain by fitting a cumulative probability curve to the hazard percentiles. They tabulate the lognormal standard deviation for a number of cities, probabilities and soil conditions. They do not tabulate Sa(0.2), so their most comparable results to ours are for PGA at 2%/50yr on Class B/C. Those U.S. lognormal standard deviations for PGA are ~0.22 for west coast cities, and 0.4 - 0.54 for eastern sites, both near the New Madrid seismic zone. Comparable locations from Table 1 for Sa(0.2) are 0.1 - 0.18 for west coast cities and 0.26 - 0.36 for east coast cities. For all places it appears that the U.S. model includes more epistemic uncertainty than the Canadian model, though the reason for the difference is still under investigation.

CONCLUSIONS

We have quantified the spread of hazard values about the mean, as calculated from the 5th Generation model used for NBCC2015. We find that it follows a lognormal distribution between its extreme percentiles. Calculated distributions show a large range in ground motion values between adjacent percentiles, displayed as steps in the cumulative curves. These arise from the crudeness of the model (for example using a 3-branch representation of the GMM instead of a 11-branch one), but such crudeness is currently required for efficient computation of national hazard maps. We find that OQ gives a similar distribution for the hazard, providing some of its parameters are chosen appropriately. Therefore OQ's use for generating the distribution of hazard from the 6th Generation model appears possible, though the practicalities of the very long run times makes its actual use challenging for our Canadian model.

The shapes of the Sa(T) lognormal distributions are very similar for all periods at a given site in eastern Canada, but are not as similar for a site in western Canada. For a given period it appears that the spread can differ greatly from site to site, so that no single "spread" parameter can represent all hazard distributions across Canada. For the localities studied we find that the spread provided by the model is smallest at Tofino and largest at Toronto. Much of the spread arises from epistemic uncertainty in the GMMs, but in low-hazard regions the uncertainty in earthquake rates is also important.

Due to the stepped nature of the curves, we consider that the actual spread in the hazard values is better represented by the separation between the 2nd and 98th percentiles, rather than by using the values calculated for intermediate percentiles. This conclusion has implications for the reliability of designs that use either the mean or the 84th percentile to make assumptions about values in the upper tail.

ACKNOWLEDGMENTS

The deaggregation and percentile version of GSCFRISK is a legacy of Frank Anglin (1935-2016). Trevor Allen created the 5th Generation model for OQ. We thank our colleague Michal Kolaj and Marco Pagani of GEM for their assistance in running OQ. Michal Kolaj provided a useful internal review.

REFERENCES

1. Basham, P.W., Weichert, D.H., Anglin, F.M., and Berry, M.J. (1985): New probabilistic seismic ground motion maps of Canada, *Bulletin Seismological Society America*, **75** (2), 563–595.
2. Adams, J. and Halchuk, S. (2003). Fourth generation seismic hazard maps of Canada: values for over 650 Canadian localities intended for the 2005 National Building Code of Canada, *Geological Survey of Canada Open File* 4459, 155 pp, doi:10.4095/214223.
3. Adams, J., Halchuk, S., Allen, T., Rogers, G. (2015). Canada's 5th Generation seismic hazard model, as prepared for the 2015 National Building Code of Canada, *11th Canadian Conference on Earthquake Engineering*, Victoria, Canada, Paper 93775, 11pp.
4. Halchuk, S., Allen, T.I., Adams, J., and Rogers, G. C. (2014). Fifth generation seismic hazard model input files as proposed to produce values for the 2015 National Building Code of Canada, *Geological Survey of Canada Open File* 7576, 15 pp, https://doi.org/10.4095/293907 .
5. Pagani, M, Monelli, D, Weatherill, G, Danciu, L, Crowley, H, Silva, V, Henshaw, P, Butler, R, Nastasi, M, Panzeri, L, Simionato, M, Vigano, D (2014): OpenQuake Engine: An open hazard (and risk) software for the Global Earthquake Model, *Seismological Research Letters*, **85** (3), 692–702, doi: 10.1785/0220130087.
6. Adams, J., Allen, T., Halchuk, S., Kolaj, M. (2019). Canada's 6th Generation Seismic Hazard Model, as Prepared for the 2020 National Building Code of Canada. *12th Can. Conf. Earthquake Engineering*, Quebec City, paper 192-WGm8-169, 8 pp.
7. NBCC2015. (2015). National Building Code of Canada 2015. National Research Council of Canada, Ottawa, NRCC 56190 Vol 1 & 2.
8. Atkinson, G.M., Adams, J. (2013). Ground motion prediction equations for application to the 2015 Canadian national seismic hazard maps, *Can. J. Civ. Eng.*, **40**, 988–998, doi: 10.1139/cjce-2012-0544.
9. Z-score percentile table used from https://math.tutorvista.com/statistics/z-score-table.html last accessed 20190130.
10. Allen, T.I., Halchuk, S., Adams, J., Rogers, G.C. (2017). Canada's 5th generation seismic hazard model: 2015 hazard values and future model updates, *16th World Conference on Earthquake Engineering*, Santiago, Chile, Paper 3494.
11. GEM Global hazard map https://hazard.openquake.org/gem/models/mosaic/ last accessed 20190130.
12. Lee, Y., Graf, W., and Hu, Z. (2018). Characterizing the epistemic uncertainty in the USGS 2014 National Seismic Hazard Mapping Project (NSHMP). *Bulletin Seismological Society America*, 108(3A) 1465-1480.