INSIGHTS INTO SEISMIC HAZARD FROM BIG DATA ANALYSIS OF GROUND MOTION SIMULATIONS

KRISTY F. TIAMPO[1], JAVAD KAZEMIAN[2], HADI GHOFRANI[2],
YELENA KROPIVNITSKAYA[2] & GERO MICHEL[3]
[1] Department of Geological Sciences and Cooperative Institute for Research in Environmental Sciences, University of Colorado Boulder, USA.
[2] Department of Earth Sciences, Western University, Canada.
[3] Chaucer Copenhagen, Denmark.

ABSTRACT

Despite the fact that most seismic activity occurs along plate boundaries, large earthquakes do occur within the plates and result in significant human and economic losses. Unlike seismicity along the plate boundaries, this intraplate seismicity is distributed unevenly, is generally rare, and often poorly understood. For example, the most seismically active region in eastern Canada is located in a stable continental region within the North American plate. While it has a relatively low rate of earthquake activity, this largely urban region has experienced six earthquakes of approximately magnitude six since 1663, with large earthquakes concentrated in regions of crustal weakness. Montréal, located in the seismically active region of the western Quebec seismic zone, is one the largest and most populated cities in the eastern Canada. Here we present the results of a high-performance computing application for ground motion simulations for Montréal. The simulations are performed based on the pipelining implementation of the EqHaz program suite. These high-resolution scenario shakemaps produces estimates of ground shaking that are not tied to particular scenario earthquakes, but are probabilistic in nature, and that can be combined with exposure values to estimate most likely risk and loss scenarios. These results can be used not only to estimate potential losses but to mitigate against risks and losses due to large events through better estimation of stability and damage and the directed implementation of earthquake resistant construction standards.

Keywords: Big data, earthquake ground motions, seismic hazard, seismic risk

1 INTRODUCTION

In recent years, technological advances have resulted in an avalanche of data from a wide range of areas, including remote sensing satellites, the health care industry, and internet and financial network data. Although the increasing number of datasets presents a significant challenge to the acquisition and analysis of that data, most definitions of these growing data sets emphasize the difficulties associated with their large volumes [1–4]. In addition, user-generated data are increasing that volume at previously unforeseen rates because, as noted in the Harvard Business Review, "each of us is now a walking data generator" [5]. Today, this phenomenon is designated as 'big data'. However, it is important to note that big data is not only about large volumes. Complications such as the requirement for real-time processing, specific end-user requirements, and the structure of big data, require innovative methodologies for its procurement, transmission, analysis and conservation [6].

U.S. Department of Energy (DOE) described this new era of data-intensive computing in two recent report [7, 8]. However, for publicly-funded science to provide relevant and credible results and information to industry and policy makers, it is necessary to understand the constraints on the use of that data and to identify those factors and opportunities that improve data usability. Data usability is best characterized in terms of its likely usage and requires interaction between the producers and users of that data [9]. In that capacity, big data, whether collected via satellite or generated in large computer simulations, is critical to modern studies of natural and anthropogenic hazards. These studies range from forecasting, monitoring, and

ISSN: 2041-9031 (paper format), ISSN: 2041-904X (online), http://www.witpress.com/journals
DOI: 10.2495/SAFE-V9-N1-1-12

response to high wind and tornadoes, floods, tsunamis, landslides, and earthquakes, to the modeling and data assimilation of the challenges associated with long-term climate change. Applications include disaster risk planning, urban planning, disaster response, and mitigation.

In this paper, we present a new big data set for earthquake risk and mitigation in the form of ground motion predictions at high spatial resolution for large events in a large urban setting. In Section 2 we present a high-performance computing technique for rapid generation of very large data sets in near real-time. Section 3 details the implementation of that methodology for producing high-resolution ground motions for scenario earthquakes. Results are provided in Section 4 and Section 5 summarized the conclusions from this work.

2 HIGH PERFORMANCE STREAMING PROCESSING

Big data analytics problems generally are classified as either stream-processing or batch-processing. In batch-processing applications, big data are stored and then analysis is performed separately, generally at a later time (e.g. process billing by credit card companies). This is most applicable in those cases in which data generation is constrained, such as when a model is too computationally-intensive model to be run on demand. This type of batch processing benefits from parallelism and is structured such that those steps which result in the greatest efficiencies occur in the early stages. In addition, data providers are moving to cloud computing, a model where consumers perform computation on the same infrastructure that stores the data [6].

Streaming processing is most commonly applied when the data needs to be analyzed quickly because the results are needed in real-time or near real-time. It is most effective in those cases where data is available as a continuous stream from an online source and rapid processing is required (e.g. disaster evaluation and emergency response). As a result, it is most effective if the analysis is understood and pre-programmed prior to access. The continuous nature of the data stream and its inherent high velocity coupled with large volumes means that only a small amount of the data is stored in local memory. The analysis also is most effective if it is computationally inexpensive. Applications that require rapid results benefit the most from streaming processing, which again makes it particularly useful for near real-time hazard assessment [6].

The large volumes and a high delivery velocity of big data present a significant challenge when trying to respond quickly and effectively to natural catastrophes. Real-time and near real-time products derived from a broad range of remote sensing images can provide important and timely information to disaster response officials. In particular, seismic hazard and ground shaking maps have a wide variety of practical applications. They are widely used by policymakers to specify earthquake-resistant building codes, insurance companies to evaluate risk and set insurance rates, by civil engineers to ensure structural stability, and by emergency response agencies to evaluate hazard potential and coordinate response. In an increasingly competitive insurance market, companies are looking for better assessment of short-term variability around hazard and risk. Better estimates of seismic hazard and ground-shaking will improve our ability to quantify the potential damage and outcomes from a large earthquake. In this work, we provide an example of how advanced streaming methods can automate the generation of those maps while improving their resolution, providing better estimates of risk and loss. In addition, estimation of ground-shaking in near real-time can inform effective and rapid deployment of hazard responders and other emergency response resources.

An advanced parallel computing platform, IBM InfoSphere Streams (Streams) is designed for deployment across a scalable set of computing clusters, enabling continuous, rapid analysis of big data volumes with low latency [10]. InfoSphere Streams provides a runtime platform, toolkits and a programming model. Together these process the data as streams; processing procedures are implemented as operators and connected to form processing pipelines. Stream Processing Language (SPL) is the programming language for building applications in stream processing – a data-centric programming model. For each application, the applicable streams and operators and streams are assembled into a data-flow graph which defines the connections between the inputs (streaming data sources), operators (procedure), and outputs (data sinks) [11, 12].

Probabilistic seismic hazard analysis (PSHA) is the standard for estimating seismic hazard around the world [13–15]. Seismic hazard maps quantify the probability of exceeding a particular ground motion value at any given site for some specified time period. This hazard is estimated by combining the probability of exceedance for all possible events, at any distance and magnitude value as prescribed by the local and regional seismic history. The spatial resolution of these hazard or ground-shaking maps is a very important parameter. Details will be lost if the grid is too coarse, and potentially result in an underestimation of the maximum ground motion level value [16]. However, a finer grid requires significant computational resources, potentially resulting in significant delays in map generation. For example, for eastern Canada, more than 6.5 trillion locations are required to generate a hazard map at a one-meter resolution [11]. Streams allows a complex problem such as this to be divided into smaller tasks (programs) that are pipelined and deployed on large computer clusters.

In this Streams application, parallel processing of individual PSHA programs results in faster production of real-time seismic hazard maps at higher resolutions. The PSHA programs employed here are based on the open-source EqHaz software suite of [17]. EqHaz consists of three programs. Synthetic earthquake catalogs based on user-specified seismicity parameters are generated using EqHaz1. The resulting ground motions, hazard curves and motions at each location and for specified return periods are generated using EqHaz2. The relative contributions of the different earthquake sources as a function of distance and magnitude are de-aggregated using EqHaz3. Here, EqHaz1 and EqHaz2 source codes were linked into Streams as shared system object libraries [11]. This innovative computational technique for PSHA mapping produces high-resolution maps of the probability of exceeding certain ground motion levels across eastern Canada (Fig. 1) and can be used to produce near real-time maps of ground shaking for damage response and estimation [11].

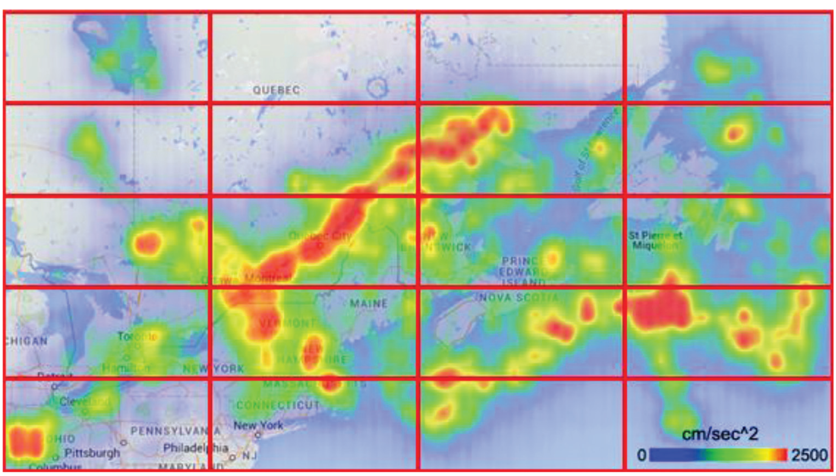

Figure 1: Mean hazard map for eastern Canada with a return period of 2475 years for pseudo -spectral acceleration at T = 0.1 sec period (modified from [11]). Red lines demonstrate a spatial decomposition of map grid. Each red rectangle contains 100,000 points.

3 HIGH-RESOLUTION GROUND SHAKING MAPS

While hazard maps, such as those in Fig. 1, are useful for regional hazard estimation, here we demonstrate the application of high-resolution mapping of ground shaking in dense urban areas for a series of scenario earthquakes. As a result, we have adapted this procedure to generate synthetic earthquake catalogs from the historical earthquake record. In this example, the procedure is applied to the city of Montréal, in eastern Canada. This synthetic catalog then is used to estimate the distribution of ground motions at a large number of sites using selected ground motion prediction equations (GMPEs) [18]. The resulting maps take into account uncertainties in earthquake location, size, and soil type.

We again use the open-source EqHaz software suite of programs to generate the suite of Monte Carlo simulations [17]. The seismicity zoning and occurrence model used in this work as input for synthetic data generation are a composite model of thirty-nine Eastern Canada zones provided by GSC [19–21]. This Streams GMPE application generates synthetic catalogs using the operator developed from the EqHaz1 program and produces GMPEs based on the EqHaz2 program operator.

One of the largest cities in eastern Canada, Montréal with a population of 1.65 million inhabitants [22] is the second most seismically vulnerable cities in Canada after Vancouver [23]. The Montréal area is in a zone of moderate seismic hazard, where a M5-6 earthquake occurs approximately every 25 years and an event of M > 6 has a recurrence period of 100 years. A recent study performed by AIR Worldwide assessed the probability of a major earthquake in the Montréal area and determined that there is a 5–15% chance of a moderate earthquake in the next 50 years [24]. In addition, the aftermath of the 1988 Saguenay earthquake demonstrated that the primary aggravating factor in Montréal are site effects, i.e. the effect of local soil conditions on earth ground-motion / shaking (response) and damage patterns.

In eastern Canada, earthquakes are believed to be caused primarily by a northeast-to-east oriented compressive stress field reactivating zones of crustal weakness. The Charlevoix region of eastern Canada is the site of five large earthquakes (M > 6) since 1663, the most recent in 1925. The offshore Atlantic margin, southeast of Newfoundland, experienced a magnitude 7.2 earthquake in 1929 [25].

Because we are interested in all earthquake sources capable of producing damaging ground motions in Montréal, we implement a probabilistic approach. In this region, individual faults are not identifiable, and as a result earthquake sources are described by areal regions in which earthquakes may occur anywhere. All possible sources are identified with the distribution of magnitudes and source-to-site distances associated with earthquakes from each source [26]. This model takes into consideration events in historically-active clusters and wandering large events along the rift model. Small-to-moderate events are characterized by the Gutenberg-Richter relations for smaller zones [26]. Figure 2 shows the distribution of events in these seismic zones for eastern Canada.

The simulation generates a synthetic seismic catalog of all events, M > 4.5, for a one-million-year time period in eastern Canada with ground motion estimates for the city of Montréal on a grid with 2093 locations. The resolution of the analysis (2093 points) is limited by the resolution of the soil map used in this study [27]. Again, the simulation is performed using a Monte Carlo approach based on the open-source program suit EqHaz and IBM InfoSphere Streams platform, as detailed above [11, 18, 28]. The simulations are performed in three steps: (1) One million years of events are generated using the seismic

sources and rates of [26]; (2) ground motion estimates are generated from that catalog for each event in the synthetic catalog [31]; (3) ground motion values are corrected based on site

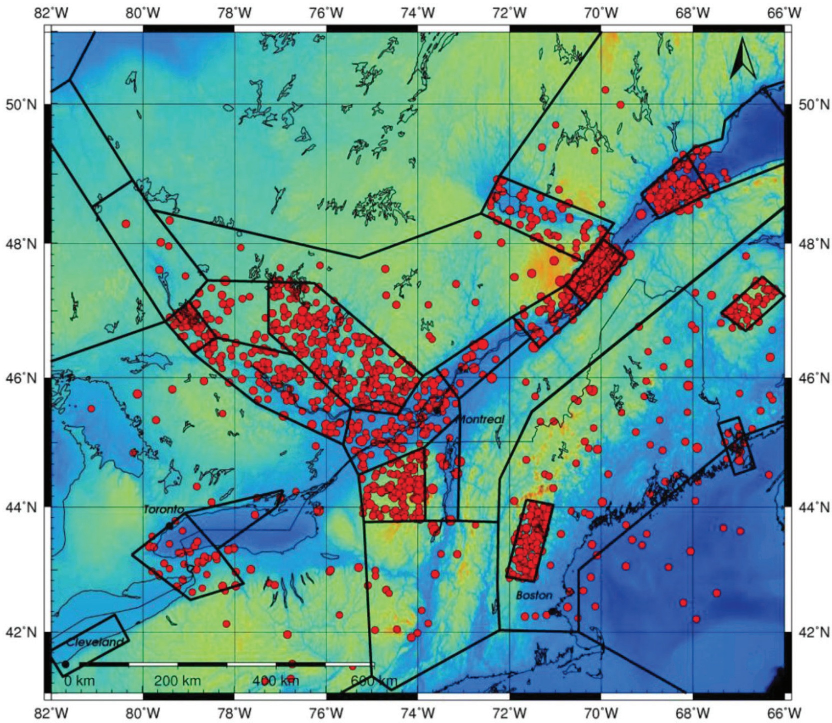

Figure 2: 10,000 year synthetic seismic catalog generated for eastern Canada; one realization out of 100 sets used for the generation of a one-million-year catalog.

amplification factors [27, 29]. The Gutenberg-Richter relation for the complete catalog is shown in Fig. 3. Note that the largest anticipated event is approximately M7.5, as expected from the historic and paleoseismic record [30].

In the last step two steps, ground motion parameters are generated at each site of interest for every earthquake scenario in the catalog. GMPEs are the key input that drives the resulting ground motion maps, or ground-shaking scenarios. For this study, we employed the eastern Canada GMPE solutions developed by [31]. Finally, the Montréal soil map [28] was used to determine the correct site amplification coefficients. The final result is a synthetic catalog of ground motions at 2093 locations, each slightly less than 500 meters square, in the city of Montréal.

4 RESULTS

The construction of the scenario shakemaps is performed over a regular grid covering the Island of Montréal. Figure 4 shows the location of three individual events in the synthetic catalog, all of M~7.7.

At each grid point, the expected horizontal component ground motions for a given earthquake scenario are calculated for a reference site condition. Figure 5 presents the ground motions for each of the events in Fig. 4, before application of the amplification factors necessary to modify the ground motions for the likely effects of local soil conditions at that point; i.e. after the second step described in Section 3, above. Note that the pattern of ground shaking is dominated by both distance and direction from each event.

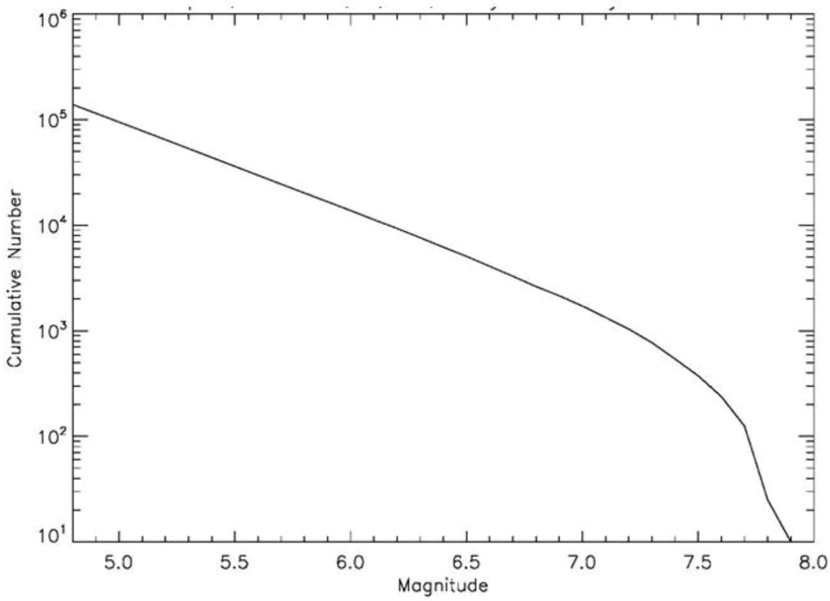

Figure 3: Gutenberg-Richter relation, cumulatie number of events, for a one-million-year synthetic catalog of Montréal.

Figure 6 again shows the same three events as seen in Fig. 4, and the estimated ground motions *after* soil amplification factors are applied to the shakemaps of Fig. 5. Figure 5, right, also shows the estimated ground motions after application of soil amplification factors. Here the ground motion pattern changes significantly, reflecting the importance of soft soils in the seismic response of structures in Montréal.

The maps shown in Fig. 6 are designed to provide emergency management personnel information on the distribution of strong ground shaking to facilitate an informed and effective emergency response in the event of a catastrophic earthquake. The data that are generated via these shakemaps can be used as inputs for the casualty and damage assessment routines for rapid earthquake loss estimation.

5 CONCLUSIONS

While big data presents important challenges today in the field of hazards modelling and response, innovative technological advancements can provide cost-effective solutions. Here we have presented a method for creating a synthetic catalog of seismic events for a dense city of two million people, Montréal, much of which has not been seismically retrofitted. This catalog, generated using state-of-the-art, high-performance computing methodologies, contains 150,635 events over a one-million-year time period. Each of these events then is converted into a ground shaking map, using GMPE equations and well-quantified soil amplification factors for the city, producing 308,860,665 high-resolution shake maps.

The resulting suite of ground shaking maps can be analyzed to provide both the recurrence rates for the most likely locations of strong ground motions on a high-resolution grid, producing maps of the probability of ground shaking, and even recurrence plots, for each location. These can be combined with insurance company exposure values to provide the probability

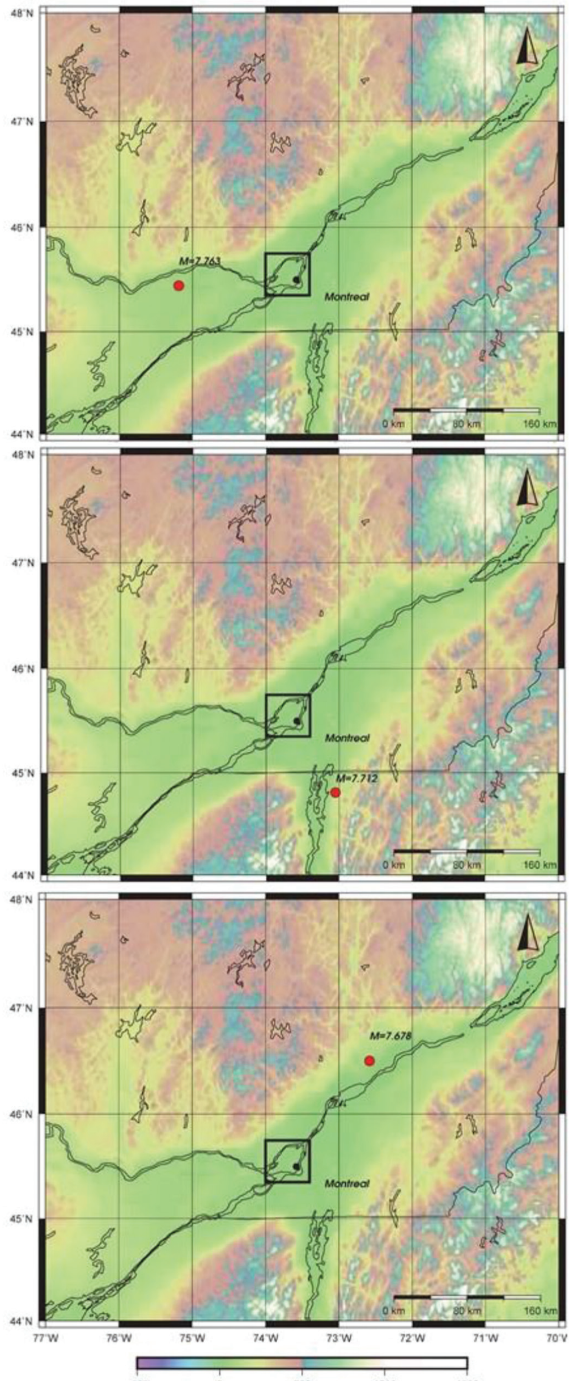

Figure 4: Three selected events (M~7.7) from the one-million-year synthetic catalog. Events locations are shown in red.

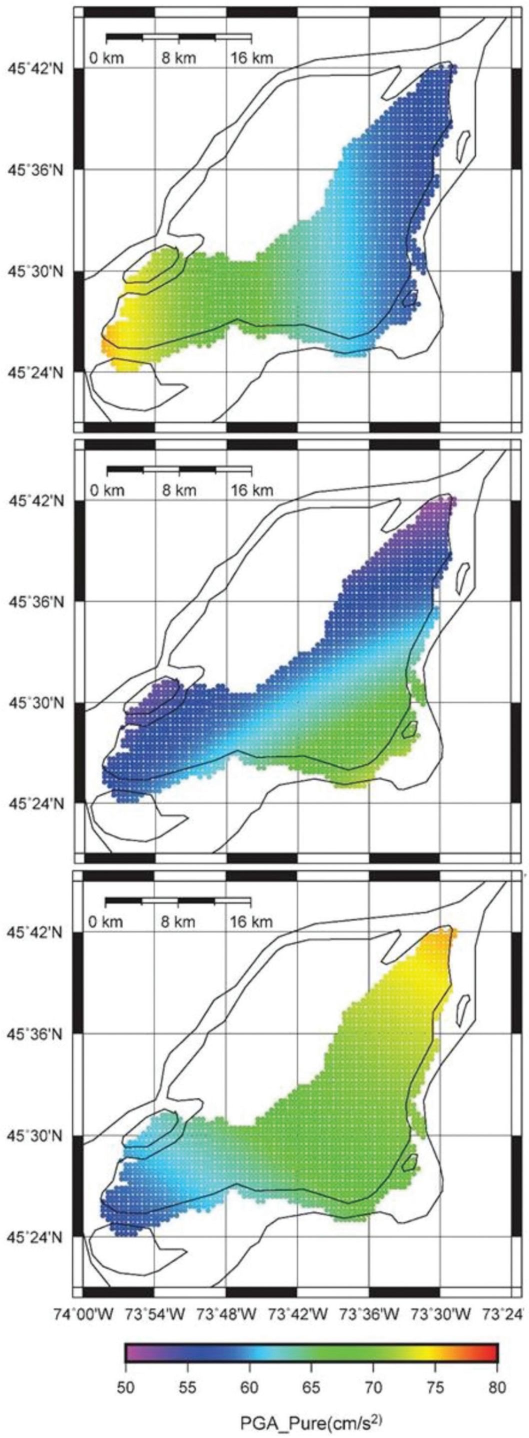

Figure 5: Ground shaking before soil amplification, in PGA (cm/sec^2), for each of the three events from Fig. 4, in the same order as shown in Fig. 4.

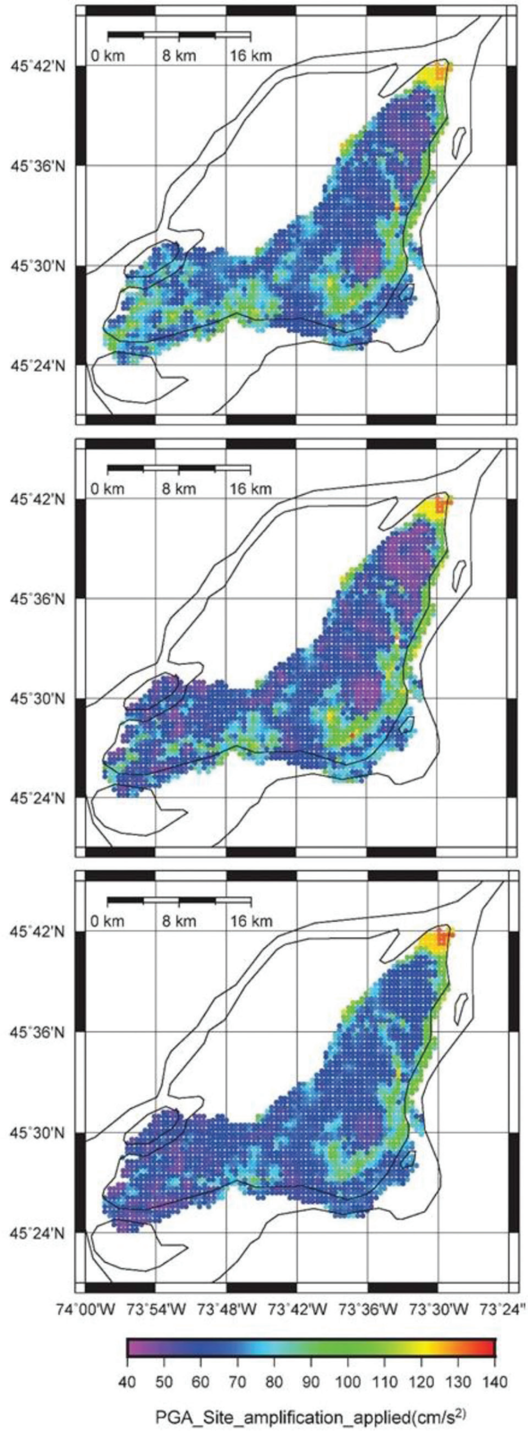

Figure 6: Maps of ground shaking, in PGA (cm/sec^2), for the three selected events from Figs. 4 and 5, that incorporate site amplification effects.

of specific loss values at individual locations. These same risk maps can be used to provide estimates of damage and human losses, to site and allocate resources, and emergency response planning. Most importantly, here we provide a high-performance computing methodology to estimate risk from hazard that is flexible and can be implemented in regions where detailed fault information and deterministic simulations are not feasible.

REFERENCES

[1] Kouzes, R.T., Anderson, G.A., Elbert, S.T., Gorton, I. & Gracio, D.K., The changing paradigm of data-intensive computing. *Computer*, **42**(1), pp. 26–34, 2009. https://doi.org/10.1109/mc.2009.26

[2] Ma, Y., Wang, L., Liu, P. & Ranjan, R., Towards building a data-intensive index for big data computing – A case study of Remote Sensing data processing. *Information Sciences,* **319**, pp. 171–188, 2015. https://doi.org/10.1016/j.ins.2014.10.006

[3] Moore, R., Prince, T.A. & Ellisman, M., Data-intensive computing and digital libraries, *Communications of the ACM*, **41**(11), pp. 56–62, 1998. https://doi.org/10.1145/287831.287840

[4] Miller, S.D., Kleese van Dam, K. & Li, D., *Challenges in Data Intensive Analysis at Scientific Experimental User Facilities*. Springer: Berlin, pp. 249–284, 2011.

[5] McAfee, A. & Brynjolfsson, E., Big data: The management revolution. *Harvard Business Review,* 2012.

[6] Hu, H., Wen, Y., Chua, T.-S. & Li, X., Toward scalable systems for big data analytics: a technology tutorial. *IEEE Access,* **2**, pp. 652–687, 2014. https://doi.org/10.1109/access.2014.2332453

[7] U.S. Department of Energy, The Office of Science data-management challenge, *Technical Report 1–2,* 2004.

[8] U.S. Department of Energy, Synergistic challenges in data-intensive science and exascale computing, *Summary Report of the Advanced Scientific Computing Advisory Committee (ASCAC) Subcommittee*, 2013.

[9] Dilling, L. & Lemos, M.C., Creating usable science: opportunities and constraints for climate knowledge use and their implications for science policy. *Global Environmental Change,* **21**, pp. 680–689, 2011. https://doi.org/10.1016/j.gloenvcha.2010.11.006

[10] IBM, *IBM Knowledge Center* available at http://www-01.ibm.com/support/knowledgecenter, 2014.

[11] Kropivnitskaya, Y., Qin, J., Tiampo, K.F. & Bauer, M.A., A pipelining implementation for high resolution seismic hazard maps production. *Procedia Computer Science*; **51**, pp. 1473–1482, 2015. https://doi.org/10.1016/j.procs.2015.05.337

[12] Bauer, M.A., Biem, A., McIntyre, S. & Xie, Y., A pipelining implementation for parsing X-ray diffraction source data and removing the background noise. *Journal of Physics: Conference Series*, **256**, p. 012017, 2010. https://doi.org/10.1088/1742-6596/256/1/012017

[13] Cornell, C.A., Engineering seismic risk analysis. *Bulletin of the Seismological Society of America*, **58**(5), pp. 1583–1606, 1968.

[14] McGuire, R., Seismic hazard and risk analysis. *Earthquake Engineering Research Institute*, Oakland, California, 2004.

[15] McGuire, R., Probabilistic seismic hazard analysis: early history. *Earthquake Engineering & Structural Dynamics,* **37**, pp. 329–338, 2008. https://doi.org/10.1002/eqe.765

[16] Musson, R.M.W. & Henni, P.H.O., Methodological considerations of probabilistic seismic hazard mapping. *Soil Dynamics and Earthquake Engineering*, **21**, pp. 385–403, 2001. https://doi.org/10.1016/s0267-7261(01)00020-3

[17] Assatourians, K. & Atkinson, G.M., EqHaz: An open-source probabilistic seismic-hazard code based on the Monte Carlo simulation approach. *Seismological Research Letters,* **84**, pp. 516–524, 2013. https://doi.org/10.1785/0220120102

[18] Musson, R.M.W., Determination of design earthquakes in seismic hazard analysis through Monte Carlo simulation. *Journal of Earthquake Engineering*, **3**, pp. 463–474, 1999. https://doi.org/10.1080/13632469909350355

[19] Adams, J. & Halchuk, S., 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*, 2003.

[20] Adams, J. & Halchuk S., Fourth-generation seismic hazard maps for the 2005 national building code of Canada. *Proceedings of the 13th World Conference on Earthquake Engineering, Vancouver, Canada.* Paper 2502 on CD-ROM, 2004.

[21] Halchuk, S. & Adams, J., Fourth generation seismic hazard maps of Canada: maps and grid values to be used with the 2005 National Building Code of Canada, *Geological Survey of Canada Open File 5813,* 2008.

[22] UNSD Demographic Statistics, UnData, available at http://data.un.org/, 2011.

[23] National Building Code of Canada, Associate Committee on the National Building Code, National Research Council of Canada, Ottawa, ON, 2016.

[24] AIR Worldwide, Study of impact and the insurance and economic cost of a major earthquake in British Columbia and Ontario/Québec, IBC of Canada, 2013.

[25] Adams, J. & Basham, P.W., The seismicity and seismotectonics of eastern Canada. *Neotectonics of North America*, ed. D.B. Slemmons, E.R. Engdahl, M.D. Zoback, D.D. Blackwell, Geological Society of America, Boulder, CO, 1991.

[26] Atkinson, G. & Goda K., Effects of seismicity models and new ground motion prediction equations on seismic hazard assessment for four Canadian cities. *Bulletin of the Seismological Society of America*, **101**, pp. 176–189, 2011. https://doi.org/10.1785/0120100093

[27] Rosset, P., Bour-Belvaux, M. & Chouinard, L., Microzonation models for Montréal with respect to VS30. *Bulletin of Earthquake Engineering,* 13(8), pp. 2225–2239, 2015. https://doi.org/10.1007/s10518-014-9716-8

[28] Hong, H.P. & Goda, K., A comparison of seismic-hazard and risk deaggregation. *Bulletin of the Seismological Society of America*, **96**(6), pp. 2021–2039, 2006. https://doi.org/10.1785/0120050238

[29] Boore, D.M. & Atkinson, G.M., 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. *Earthquake Spectra*, 24(1), pp. 99–138, 2008. https://doi.org/10.1193/1.2830434

[30] Halchuk, S., Allen, T.I., Adams, J. & Rogers, G.C., 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*, 2014.

[31] Pezeshk, S., Zandieh, A. & Tavakoli, B., Hybrid empirical ground-motion prediction equations for eastern North America using NGA models and updated seismological parameters. *Bulletin of the Seismological Society of America*, 101(4), pp. 1859–1870, 2011.
https://doi.org/10.1785/0120100144