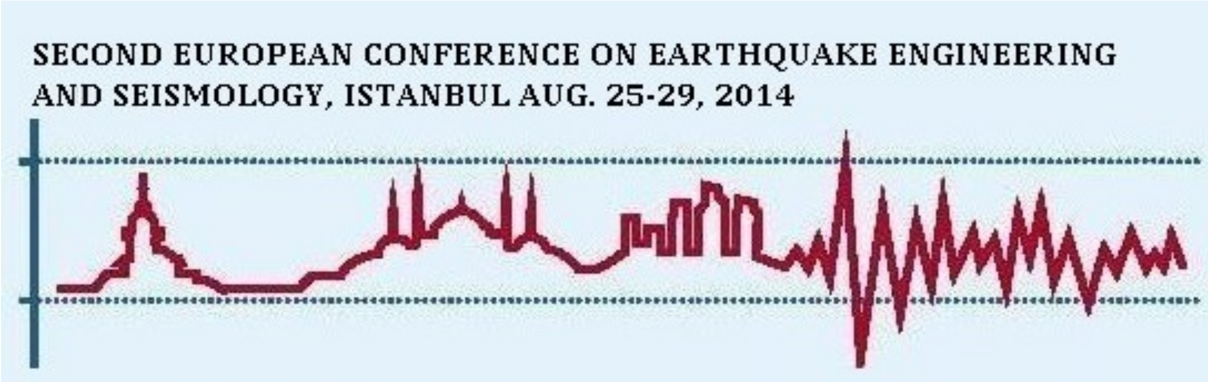

THE GEM DATABASE OF SEISMIC HAZARD MODELS

Damiano MONELLI[1], Marco PAGANI[2], Graeme WEATHERILL[1], Julio GARCIA[1]

ABSTRACT

We present the current status of the database of seismic hazard models promoted by the Global Earthquake Model (GEM) initiative. Through this database GEM aims at providing access - within a single framework – to hazard models originally published by different organisations or scientific consortia and characterized by different geographic coverage, tectonic regimes and modelling approaches. The end goal is to facilitate access to the various hazard studies done around the world and to introduce a first level of standardization in the data model definition and in the calculation procedure.

A first challenge in the construction of such a database is the high heterogeneity present in the primary data. Such heterogeneity can be managed only by defining a conceptual model (in a database terminology usually referred to as 'schema') that can accommodate all the different ways in which seismicity is described for the purpose of hazard assessment. A second challenge is in the definition of quality assurance tests able to guarantee that the information present in the database can be used to reproduce the original model results. This means that the database must be associated to a computational infrastructure able to process the stored information to derive seismic hazard estimates. The computational infrastructure must be flexible enough to deal with the different modelling strategies identified by the broad scientific community and validated against the different software packages currently used to perform seismic hazard analysis.

The main benefit of such a database is not only having a source of information that can be used to estimate seismic risk in any region of the world. Being able to represent heterogeneous models inside a single, common framework is a first step in the direction of standardizing seismic hazard assessment. As a consequence, models become more easily comparable because represented using the same data model and differences in hazard estimates due to differences in the calculation procedure are removed because the same computational framework is used.

INTRODUCTION

Nowadays the seismological and earthquake engineering communities have access to seismic hazard models for various regions of the world as developed by different agencies, geological surveys or scientific consortia. Models are typically constructed following different approaches, which take into account different tectonic regions and data availability. Although the end products are the same for all models (e.g. hazard curves and maps, uniform hazard spectra), the original information utilized to derive hazard estimates is often represented in terms of different data models. In some cases, the software adopted for the model implementation is not available, posing a serious barrier to

[1] Hazard Scientist, Global Earthquake Model (GEM), Pavia, Italy, damiano.monelli@globalquakemodel.org
[2] Hazard Coordinator, Global Earthquake Model (GEM), Pavia, Italy, marco.pagani@globalquakemodel.org
[1] Hazard Scientist, Global Earthquake Model (GEM), Pavia, Italy, graeme.weatherill@globalquakemodel.org
[1] Hazard Scientist, Global Earthquake Model (GEM), Pavia, Italy, julio.garcia@globalquakemodel.org

reproducibility. In those cases in which the software can be obtained, other issues such as lack of documentation, support, and software maintenance may prevent, in practice, any attempt to reproduce the original results. These issues become more and more important as the time from the initial release of the model increases. Any scientist or analyst outside of the group of the original modellers has limited possibilities to fully understand how a model has been implemented and how models results would change with different assumptions. Lack of a common data model, proprietary or undocumented/unmaintained software, make some of the currently available seismic hazard models difficult to be fully understood, and thus tested, constructively criticized and potentially improved by the wide scientific community.

In this paper we present the effort of the Global Earthquake Model initiative (Crowley *et al.* 2013) in building a database of seismic hazard models, which, described using a common data model and associated to the same computational infrastructure, can respond to the need of open access and reproducibility of seismic hazard assessment. Firstly we give a general description of the different models currently available in the database, the different required modelling approaches and the computational infrastructure utilized for the model implementation. Then, we describe examples of quality assurance tests implemented to verify the correctness of the model definition and finally present sensitivity analysis performed considering different hazard models in the database.

DATABASE CONSTRUCTION

The GEM database hosts national and regional scale hazard models that are either public or made available to GEM. Currently 11 models covering different regions in the Americas, Europe, Middle East and Oceania (see **Table 1**) are present. Most of the models represent the most updated models available in each region, with the oldest model dating back to 2002. Most of the models describe seismicity occurring under different tectonic settings (roughly categorized as active shallow crust, stable continental regions and subduction) and follow different approaches for modelling earthquake activity.

Models produced by the United States Geological Survey (USGS) represent about 30% of the database, that is 4 models out of 11: Conterminous US (Petersen *et al.* 2008), Alaska (Wesson *et al*. 2007), South America (Petersen *et al*. 2010) and South East Asia (Petersen *et al*. 2007). All these models are described with a similar data structure: background seismicity and deep intraslab seismicity are taken into account by using smoothed seismicity models while large subduction interface events and known active faults are taken into account by modelling explicitly 3D fault surfaces. A similar approach is also taken in the 2012 national seismic hazard model for Japan (http://www.j-shis.bosai.go.jp/en/). Both the SHARE (Danciu *et al*. 2013) and EMME (http://www.emme-gem.org) models use multiple modelling approaches in a logic tree structure; three source models are defined: one consisting of area sources only (that is zones of uniform seismicity), one based on a pure smoothed seismicity approach, and one consisting of background zones and fault sources. The remaining models (Australia, Canada, Central America, Cuba and Lesser Antilles) are based on area sources only. All models define time-independent (Poissonian) occurrence probabilities. In few cases, for particular sources (in Japan) or regions (New Madrid Zone in US) time-dependent and clustering models are adopted, respectively.

Apart from the SHARE and EMME models, the remaining hazard models have been translated from the original data format to the 'Natural hazard' Risk Markup Language (NRML) format defined by the OpenQuake-engine (Pagani et al. 2014). This translation process is a critical step, which allows porting the heterogeneous definitions of seismic hazard models into a single data model. Each model is originally defined in a set of one or more files (ASCII and or binary) following a model specific format. In the database, models are represented by one or more NRML files containing seismic sources and epistemic uncertainties definition. We want to emphasize that the usage of a single data model is not only facilitating comparisons between models but allows using a single computational infrastructure (the OpenQuake-engine) for all.

QUALITY ASSURANCE TESTS

Quality assurance (QA) tests are needed to verify that models stored in the database provide results in agreement with the results provided by the original modellers (i.e the OpenQuake-engine implementation reproduces the original behaviour of a model). This is an on-going process, which requires careful investigation of the original model, of the algorithms adopted by the software utilized for the original calculation (if this is available) and close collaboration with the original modelling team. We want to emphasize that this is the most important and time-consuming part in the database construction process and a required step to guarantee that errors or changes are not introduced when translating the original data to the NRML format.

Table 1. Seismic Hazard Models currently included in the GEM database

Author	Region/Nation	Year	Modeling Approach	Tectonic Regions
USGS (United States Geological Survey)	Conterminous US	2008	Gridded seismicity Fault based seismicity	Active Shallow Crust Stable Continental Subduction
USGS	Alaska	2007	Gridded seismicity Fault based seismicity	Active Shallow Crust Stable Continental Subduction
USGS	South America	2010	Gridded seismicity Fault based seismicity	Active Shallow Crust Stable Continental Subduction
USGS	Indonesia	2007	Gridded seismicity Fault based seismicity	Active Shallow Crust Stable Continental Subduction
GA (Geoscience Australia)	Australia	2012	Distributed seismicity (Area Sources)	Stable Continental
GSC (Geological Survey of Canada)	Canada	2010	Distributed seismicity (Area sources) Fault based seismicity	Active Shallow Crust Stable Continental Subduction
NIED (National Research Institute for Earth Science and Disaster Prevention, Japan)	Japan	2012	Gridded seismicity Fault based seismicity	Active Shallow Crust Stable Continental Subduction
SHARE (Seismic Hazard Harmonization in Europe – EU project)	Europe	2013	Distributed seismicity (Area Sources) Gridded seismicity Fault based seismicity	Active Shallow Crust Stable Continental Subduction
EMME (Earthquake Model for Middle East)	Middle East	2012	Distributed seismicity (Area Sources) Gridded seismicity Fault based seismicity	Active Shallow Crust Subduction
CENAIS (National Centre for Seismological Research, Cuba)	Cuba and surrounding	2007	Distributed seismicity (Area Sources)	Active Shallow Crust
GEOTER (Geology Tectonic Environment et Risks, France)	Lesser Antilles	2002	Distributed seismicity (Area Sources)	Active Shallow Crust Subduction

Depending on the level of information available, different QA tests can be performed. In the best case, models are provided in terms of input data and software used for the calculation. In this situation it is possible to check not only the final results but also the consistency between intermediate results, such as source occurrence rates and GMPEs values (mean and standard deviation). By considering simple test cases, it is possible to investigate differences in the calculation procedure; for instance the effect of different approximations in the modelling of earthquake ruptures for different source typologies. In the current database, the software utilized for the original calculation is available for the USGS and the Geoscience Australia (GA) models. Also the models for Cuba and surrounding, and Lesser Antilles have been implemented using the freely available CRISIS software (Ordaz *et al.* 2013). Apart from the EMME and SHARE models, which have been implemented natively using the OQ-engine, the remaining models – Canada and Japan - lack of the original software. In these cases, QA tests can be built as comparisons against final results as published in reports or available in electronic format.

As an example of QA test, we present the comparison between hazard curves for a number of cities in the United States as provided by the United States Geological Survey for the 2008 national seismic hazard model (Petersen *et al.* 2008) and the hazard curves computed using the NRML implementation and the OQ-engine (Figure 1). Considering the level of complexity in the model and the complete independence of the software used for the calculations, we consider the level of agreement between the curves satisfactory. Moreover, the current NRML implementation of the model does not contain all the features defined in the original model; in particular the OQ-engine calculation does not take into account the epistemic uncertainties on the ground motion median value for the Western US GMPEs, and the for the New Madrid region only the non-cluster model is considered. Despite the lack of these components the level of agreement is encouraging. Future work on this model will concentrate on the implementation the missing features. As an additional example of QA test, we present the comparison of hazard map values for the capital cities in Australia as provided by the GA 2010 (Burbidge *et al*. 2012) national seismic hazard model and the corresponding NRML implementation (Figure 2). For a return period of 500 years, the hazard levels are in good agreement. Larger discrepancies can be seen for the longer return period. This may be explained by the fact that hazard levels at large return periods are dominated by rare, large magnitude events and the way ruptures are modelled plays a larger role. In particular, it is worth noticing that the GA calculation procedure adopts a Monte Carlo approach for simulating earthquake ruptures, while the OQ-engine uses a Cornell-like integration procedure.

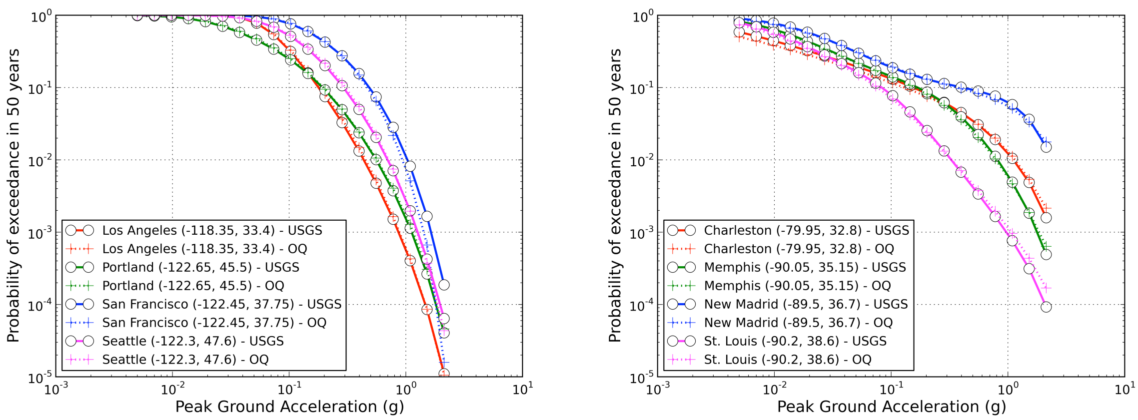

Figure 1 Comparison between hazard curves for a selected number of cities in the United States as provided by the United States Geological Survey (USGS) for the 2008 national seismic hazard model and the hazard curves computed with the NRML implementation of the model and the OQ-engine.

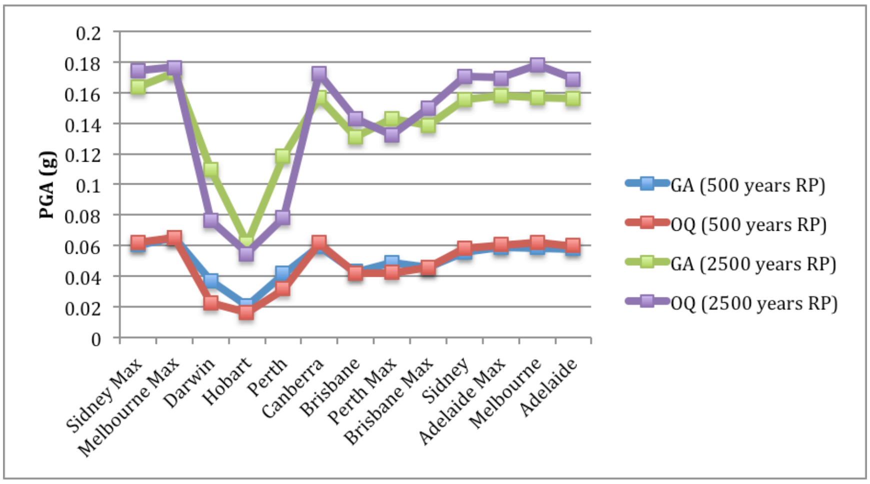

Figure 2 Comparison of seismic hazard map values for capital cities in Australia at two return periods (RP 500 and 2500 years) as provided by Geoscience Australia (GA) and the OQ-engine model implementation (OQ).

EXPLORING HAZARD MODELS' ASSUMPTION

The main benefit of having seismic hazard models associated with an open and transparent computational infrastructure is the ability to explore model assumptions. As an example, we present here a sensitivity analysis exploring how hazard curves change when taking into account rupture finiteness when modelling background seismicity. The importance of properly modelling rupture finiteness in seismic hazard calculations has been recently raised by Bommer and Akkar (2012), who showed that if a GMPE based on Joyner-Boore distance (or closest distance to the rupture) is used as if it were based on distance metrics related to point ruptures this would lead to a considerable underestimation of the seismic hazard. Because of the greater complexity in the distance calculations, taking into account rupture finiteness increases the computation time associated with a seismic hazard analysis. It this therefore important to investigate in real seismic hazard models how much the inclusion of rupture finiteness can impact the final results. To answer this question, we considered the NRML definition of the 2008 national seismic hazard model for US (Petersen et al. 2008) as currently present in the GEM database of seismic hazard models. We calculated hazard curves for two sites in the western United States (Los Angeles and San Francisco) and two sites in the central and eastern United States (New Madrid and Charleston) using, for background seismicity, once the scaling relationship originally defined in the model (Wells and Coppersmith 1994) and once point ruptures (Figure 3).

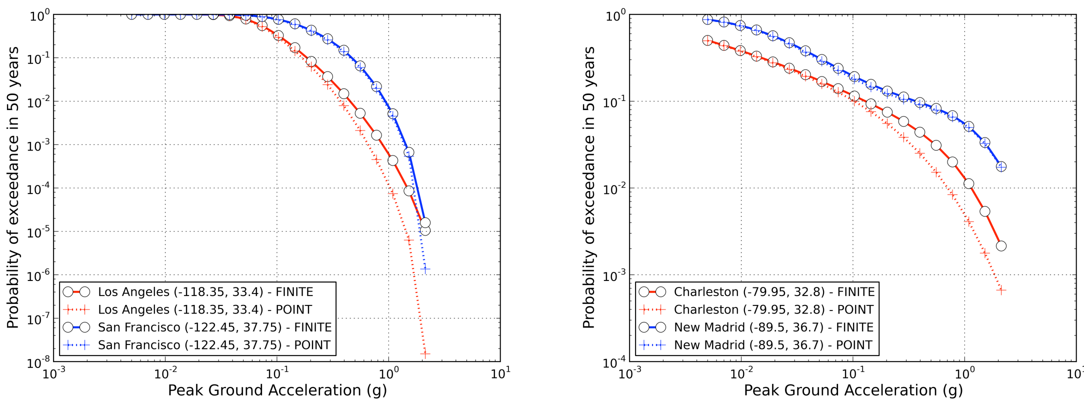

Figure 3 Sensitivity of hazard curves on rupture finiteness considering the 2008 US national seismic hazard model.

We can see that the effect of rupture finiteness is site dependent. In the western United States, the San Francisco site is practically not affected by considering point ruptures, while for the Los Angeles site a significant decrease in the probabilities of exceedance (below 10^{-2}) is visible. A disaggregation analysis shows that the hazard at long return periods at the San Francisco site is mostly dominated by fault sources and that background seismicity has little influence, while for the Los Angeles site background seismicity has a higher impact. This can explain the different influence that rupture finiteness has on the two sites. We observe a similar situation in the central and eastern United States, where the New Madrid site is not influenced at all by considering point ruptures (and again this can be explained in terms of the vicinity of the New Madrid fault zone). On the contrary, for the Charleston site, differences are clearly visible when approximating ruptures as points; this can be explained by the fact that seismic hazard around the Charleston site is modelled purely as background seismicity and no active fault structures are defined in the vicinity.

COMPARING SEISMIC HAZARD GLOBALLY

Compiling seismic hazard models developed for different regions of the world into a single database and describing them using a single data model is a first concrete step towards being able to compare, in a homogeneous way, seismic hazard assessment in different places. As an example of comparative study that the database allows, we present disaggregation analysis for three locations in a subduction environment: Seattle, Santiago de Chile, and Jakarta. We performed the analysis using the USGS models for United States, South America and South East Asia. For simplicity, we considered only one GMPE per tectonic region type: Chiou and Youngs 2008 for active shallow crust, Zhao et al. 2006 for subduction interface, and Atkinson and Boore 2003 for subduction intralslab. By using the disaggregation calculator of the OQ-engine we computed contributions to the probability of exceeding the ground motion level corresponding to 10% in 50 years (corresponding to 475 years return period), in terms of longitude and latitude of rupture closest point (geographic disaggregation) in conjunction with magnitude and tectonic region type (Figure 4). Such analysis allows comparing the geographic and magnitude/ tectonic region type distribution of the contributions for the three different locations. The analysis shows that the proximity to the subduction zone drives the seismic hazard. In Santiago del Chile, the largest contribution is associated with Mw > 9 subduction interface events and shallow seismicity provides only limited contributions. This can be explained by the fact that the subduction interface events considered in the model can propagate until below Santiago del Chile. In the case of Seattle, the subduction interface events can propagate only until the coastline. Therefore, although the contribution from the subduction interface events is still significant, the highest contributions are coming from shallow crustal sources in the vicinity of the site. For the Jakarta site, the subduction interface events play a very minor role in contributing to the hazard at 475 years return period. The largest contributions are coming from active shallow crust sources and, secondly, from deep intraslab sources in the vicinity of the site.

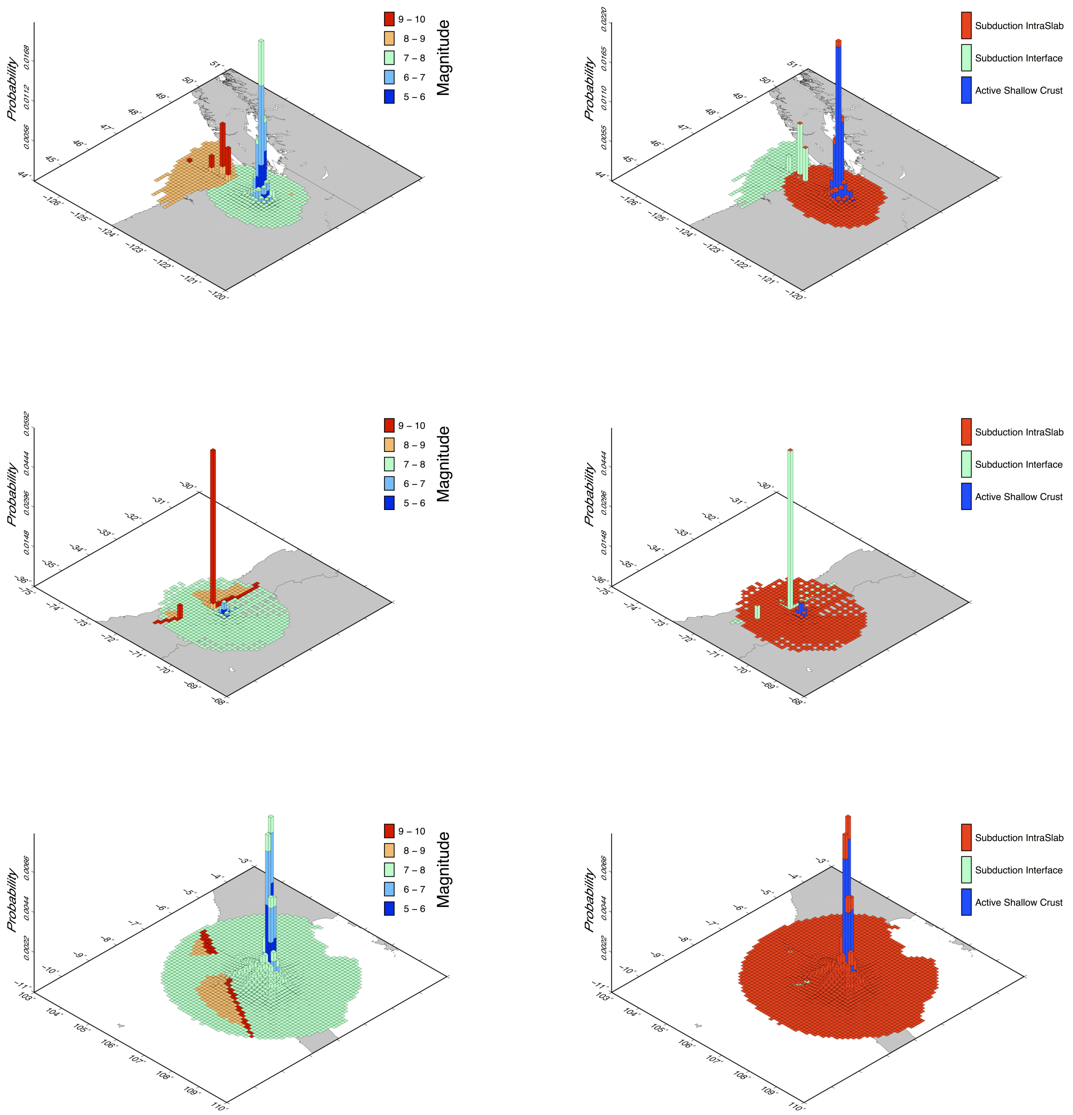

Figure 4 Disaggregation analysis (left panels - geographic and magnitude; right panels – geographic and tectonic region type) for Seattle (top panels), Santiago del Chile (middle panels) and Jakarta (bottom panels)

CONCLUSIONS

We presented the effort of the GEM initiative in building a database of seismic hazard models aiming at facilitating access to the various hazard studies done around the world and to introduce a first level of standardization in the data model definition and in the calculation procedure. The database consists of national and regional scale seismic hazard models developed for various regions in the Americas, Europe, Middle East and Oceania. Models as produced by geological surveys or scientific consortia are translated from the original data format to the 'Natural hazard' Risk Markup Language (NRML) format defined by the OQ-engine. Such data translation process allows bringing all models into the same data format, which facilitates comparisons. Moreover, models can be implemented using the same software, the OQ-engine. This means that models results can be also better compared because differences in the calculation procedure are removed by using the same computational infrastructure. To make sure that models in the database can be used to reproduce the original results, QA tests are implemented that verify that seismic hazard estimates obtained using the OQ-engine are in agreement with the ones produced by the original modellers. To showcase the major benefits that the database of seismic hazard models may provide to the seismic risk community we presented a sensitivity study performed using the 2008 national seismic hazard model for United States investigating the effect of modelling finite ruptures in background seismicity. Such sensitivity analysis has been possible thanks to the open implementation of the model present in the database and with the possibility, provided by the OQ-engine, to test different rupture modelling options. The sensitivity shows that the effect of rupture finiteness is site dependent and is influenced by the proximity of active fault structures giving higher contributions than background seismicity. The benefit of using a common computational infrastructure for different models is also exemplified by a comparative study of hazard disaggregation results in different subduction zones. By utilizing the disaggregation capabilities of the OQ-engine we could show how for the three locations considered (Seattle, Santiago de Chile, and Jakarta) and for the same return period of 475 years, the proximity of the site to the subduction interface events is a key factor in determining which type of events is contributing the most to the hazard. The GEM database of seismic hazard models will be publically released at the end of 2014 and will be maintained by incorporating seismic hazard studies as developed by the global seismic hazard and risk community.

REFERENCES

Atkinson, G.M., and Boore, D.M. (2003) "Empirical ground motion relations for subduction-zone earthquakes and their application to Cascadia and other regions", *Bulletin of the Seismological Society of America*, v. 93, p. 1703–1729.

Bommer, J. J., and Akkar, S. (2012). "Consistent Source-to-Site Distance Metrics in Ground-Motion Prediction Equations and Seismic Source Models for PSHA", *Earthquake Spectra* **28** 1-15

Burbidge, D. R., Leonard, M., Robinson, D. and Gray, D., (2012) "The 2012 Australian Earthquake Hazard Maps", in *The 2012 Australian Earthquake Hazard Map* D. R. Burbidge (Editor), Geoscience Australia Record 2012/71, 108-116.

Chiou, B. S-J and Youngs, R. R. (2008) "An NGA Model for the Average Horizontal Component of Peak Ground Motion and Response Spectra", *Earthquake Spectra*, **24** 173-215.

Crowley, H., R. Pinho, M. Pagani, N. Keller (2013) "Assessing global earthquake risks: the Global Earthquake Model (GEM) initiative", in *Handbook of seismic risk analysis and management of civil infrastructure systems*, Tesfamariam, S., and Goda, K. (Editor), Woodhead Publishing Limited.

Danciu, L., J. Woessner, D. Giardini and the SHARE consortium (2013) "A community-based probabilistic seismic hazard model for the European region", Vienna Congress on Recent Advances in Earthquake Engineering and Structural Dynamics, Tagung, Vienna.

Ordaz, M., F. Martinelli, V. D'Amico, and C. Meletti (2013) "CRISIS2008: A Flexible Tool to Perform Probabilistic Seismic Hazard Assessment", *Seismological Research Letters*, 84(3): 495-504, doi: 10.1785/0220120067.

Wells, D.L. and K.J. Coppersmith (1994) "New Empirical Relationships among Magnitude, Rupture Length, Rupture Width, Rupture Area, and Surface Displacement", *Bull. Seism. Soc. Am.* **84** 974-1002

Pagani, M., D. Monelli, G. Weatherill, L. Danciu, H. Crowley, V. Silva, P. Henshaw, L. Butler, M. Nastasi, L. Panzeri, M. Simionato, D. Vigano, (2014). OpenQuake-engine: An open hazard (and risk) software for the Global Earthquake Model. In press in *Seismological Research Letters*.

Petersen, M., Harmsen, S., Mueller, C., Haller, K., Dewey, J., Luco, N., Crone, A., Lidke, A., Rukstales, Kenneth. (2007) "Documentation for the Southeast Asia Seismic Hazard Maps",U.S. Geological Survey Administrative Report.

Petersen, M., Harmsen, S., Haller, K., Mueller. C., Luco, N., Hayes, G., Dewey, J., Rukstales, K. (2010) "Preliminary Seismic Hazard Model for South America", in *La sismología en sudamérica y los mecanismos de prevención y mitigación del peligro y riesgo sísmico, Homenaje a Alberto Giesecke*, Daniel Huaco Editor, Lima, Peru.

Petersen, M, A. Frankel, S. Harmsen, C. Mueller, K. Haller, R. Wheeler, R. Wesson, Y. Zeng, O. Boyd, D. Perkins, N. Luco, E. Field, C. Wills, K. Rukstales, (2008) "Documentation for the 2008 Update of the United States National Seismic Hazard Maps", U.S. Geological Survey, *Open-File Report 2008–1128***.**

Wesson, R., Boyd, O., Mueller, C., Bufe, C., Frankel, A., Petersen, M. (2007) "Revision of time-Independent probabilistic seismic hazard maps for Alaska", U.S. Geological Survey *Open-File Report 2007-1043*.

Zhao J.X., Zhang, J., Asano, A., Ohno, Y., Oouchi, T., Takahashi, T., Ogawa, H., Irikura, K., Thio, H., Somerville, P., Fukushima, Y., and Fukushima, Y., (2006) "Attenuation relations of strong ground motion in Japan using site classification based on predominant period", *Bulletin of the Seismological Society of America*, v. 96, p. 898–913.