The Hazard Component of the GEM Modeller's Toolkit: A Framework for the Preparation and Analysis of Probabilistic Seismic Hazard (PSHA) Input Tools

G. A. Weatherill
Global Earthquake Model (GEM) Facility, Pavia, Italy

M. Pagani
Global Earthquake Model (GEM) Foundation, Pavia, Italy

D. Monelli
Global Earthquake Model (GEM) Facility, ETH Zurich, Switzerland

SUMMARY:
The development of the seismic source model is a fundamental component in the modelling of seismic hazard on both a site-specific and regional scale. The characterisation of different source typologies (e.g. fault sources, area sources or grid sources) requires careful consideration of information from different fields of research, including geology, statistical seismology and geodesy. This presents many challenges in reconciling this information and a subsequent understanding and treatment of the respective uncertainties. To support both the regional programs and wider hazard modelling community, the GEM Model Facility is developing the Hazard Modeller's Toolkit: a suite of tools for the preparation and development of seismogenic source models for earthquake hazard analysis. Functionalities include: i) algorithms for preparation of earthquake catalogues, ii) pre-processing of the earthquake recurrence model (e.g. declustering, statistical estimation of completeness), iii) earthquake recurrence for area sources and iv) earthquake recurrence from geological and/or geodetic slip data.

Keywords: seismic hazard analysis, model development, software

1. THE GLOBAL EARTHQUAKE MODEL SEISMIC HAZARD INITIATIVE

1.1 The need for a Modeller's Toolkit

The development of a new global model of earthquake hazard is one of several important outcomes of the Global Earthquake Model (GEM) initiative. Building on foundations laid by earlier projects, such as the Global Seismic Hazard Assessment Program (GSHAP) (Giardini *et al.*, 1999), the new model will utilise new global data sets of earthquake input data provided by the GEM Global Components (GCs, see Pinho, 2012), including a new global instrumental earthquake catalogue, a global historical earthquake catalogue, a global database of active faults, a global geodetic strain model and a global selection of GMPEs for different tectonic environments. Whilst the GEM Global Components will produce new data sets for model preparation and development, the seismic hazard model preparation and application is undertaken by a set of regional programmes (RPs) for different areas of the globe. The RPs cover the main continents of the globe, and include existing projects such as the Seismic Hazard Harmonisation of Europe (SHARE), the Earthquake Model of the Middle East (EMME) and the Earthquake Model of Central Asia (EMCA). At the time of writing, new regional programs in progress including Central America, the Caribbean, South America, North Africa, Sub-Saharan Africa and South-East Asia. The motivation of the regional programmes is to ensure a greater level of regional knowledge, participation and ownership of the seismic hazard and risk models and outputs than may have been found in earlier projects.

Within the development of a global seismic hazard model, it is necessary to understand the various methodologies that are used in the practice of model development, and where appropriate to make them available to the earthquake hazard and risk community. It is within this context that the GEM Model Facility, in close collaboration with the widest scientific community, is developing the Hazard Modeller's Toolkit. This is a suite of software tools that should provide an active and working repository for algorithms that are commonly used in the seismogenic hazard model development process. In the process of reviewing existing seismic hazard models for different countries and regions, a substantial divergence in the model development process is evident. This is driven in large part by the differences in the quality and volume of data available (e.g., earthquake catalogues, geological data etc.). Whilst such considerations play an important role, there is a substantial degree of heterogeneity and a lack of transparency in the model development process also arises due to the availability of different methodologies and the software in which algorithms are implemented. These issues create limitations upon the extent to which it is possible to discern the potential impact that each modelling decision made in the development process will have on the resulting hazard model (i.e. in terms of source characterisation and recurrence) and ultimately on the hazard and risk analysis.

The GEM Hazard Modeller's Toolkit has been developed with the motivation to address issues of consistency, transparency and availability in seismic hazard model development. It is intended to provide, within an open and transparent test-driven development environment, a repository for algorithms that are widely used when building a seismic hazard model. These tools can be chained together to form a "workflow" that can be used for sensitivity testing modelling decisions, or they can be used as standalone tools for implementing specific algorithms for usage in a research-based context. The initial functionalities are outlined within this paper, in addition to some of the seismic hazard model development issues that are consequently raised.

1.2. The Software Infrastructure of the Modeller's Toolkit

The GEM Modeller's Toolkit is eventually intended to sit as part of a wider set of earthquake hazard and risk functionalities within the GEM OpenQuake platform. It is recognised, however, that within such a framework the software itself is limited in the extent to which is can be developed by the earthquake community. It is for this purpose that a command line version "Command-Line Modeller's Toolkit (CLMT)" is also in active development (www.github.com/gem/oq-hazard-modeller). This is a lighter version of the software that can be compiled and run inside a Unix/Linux (including Mac OS X) environment, in which the scientific code is accessible to the community. This is intended to form a more agile development platform that should allow for rapid prototyping and development of new functionalities, and to which community contributions to the code base can be accepted and tested. Those functionalities that are well-established, tested and stable will then form part of the GEM OpenQuake platform. It is the command-line version of the toolkit that is the focus of this paper, although it should be recognised that many of the considerations, and indeed the functionalities themselves, are common to both.

The complete GEM OpenQuake software suite, including the CLMT, is developed in the Python programming language. For mathematical calculations, considerable use is made of Python's libraries for numerical and scientific computing (Scipy/Numpy), which, when used effectively, adopt a scripting syntax that is recognisable to users of numerical computing languages such as Matlab or R. As the name indicates, the CLMT is executed from the command line, with the workflow and parameter settings established in a configuration file (Figure 1). The scientific functionalities themselves are maintained in a scientific module that is embedded within the code structure, and which is accessible for model integration and development. As with GEM's OpenQuake software, the code for the most stable release of the toolkit can be found inside GEM's open Github repository (www.github.com/gem/oq-hazard-modeller). Additional python dependencies are needed for execution of the code, including the Python YAML library (pyyaml) (http://pyyaml.org), and Shapely (http://pypi.python.org/pypi/shapely) a library of tools for spatial calculation.

A crucial element of the Modeller's Toolkit development process is in understanding how the earthquake hazard community will need to apply the tools. Furthermore, if a key motivation of the toolkit development is to form a potential repository for the integration of new algorithms and functionalities, then it is important that integration of new features can be handled in an accessible manner. Being developed as part of a wider GEM software suite, the Modeller's Toolkit must also conform to test-driven software development process, which undergoes review from the GEM Model

Facility development team.

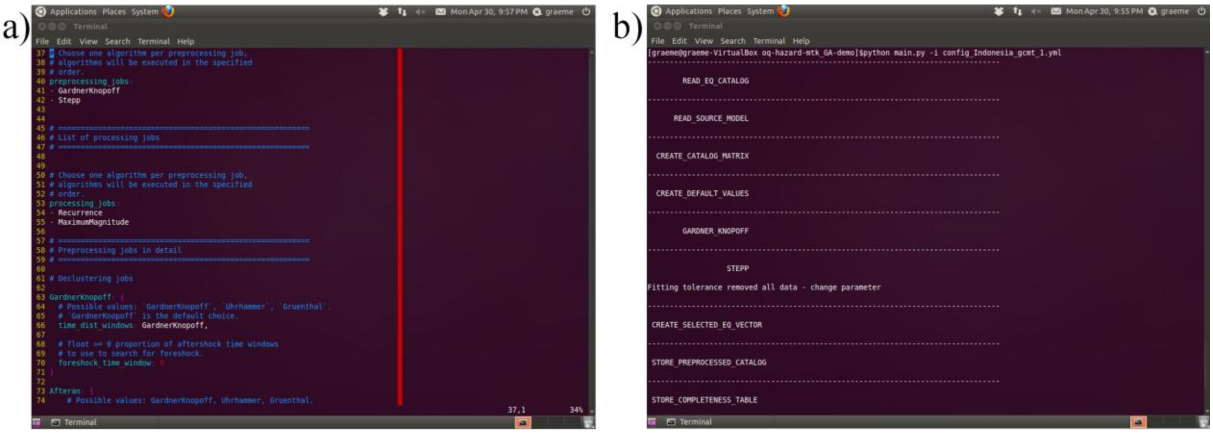

Figure 1 Screenshots of a) the Modeller's Toolkit Configuration File, and b) the execution on the command line

1.3. Supported Data Formats

For the current functionalities of the Modeller's Toolkit, Table 1 lists the current data formats that are supported or are in the process of being supported, where further details can be found in the Modeller's Toolkit documentation (http://docs.openquake.org/mtoolkit) or in the OpenQuake User Manual (Crowley *et al.*, 2011).

Table 1 Data formats supported by the Modeller's Toolkit (April 2012)

Data	Supported Data Formats	References
Earthquake Catalogue (prior to catalogue homogenisation)	i. IMS 1.0 (text format) ii. Comma separated value (.csv) file	i. International Seismological Centre Online Bulletin (www.isc.ac.uk) ii. Modeller's Toolkit Documentation
Homogenised Earthquake Catalogue	i. .csv	i. Modeller's Toolkit Documentation
Area Source Model	i. Natural Risk Markup Language (nrml) (.xml) ii. Shapefile	i. Crowley *et al.* (2011)
Fault Source Model	i. Nrml ii. Prototype YAML (.yml) format iii. Shapefile	i. Crowley *et al.* (2011) ii. Modeller's Toolkit Documentation

2. MODELLER'S TOOLKIT FUNCTIONALITIES

2.1. Earthquake Catalogue Preparation

The GEM Global Component for the Global Instrumental Catalogue, developed by the International Seismological Centre (ISC), may represent the most accurate global seismicity catalogue available to the community. As a global catalogue, however, the expected worldwide threshold magnitude levels are relatively high: $M_S \geq 5.5$, $M_S \geq 6.25$ and $M_S \geq 7.5$ for the 1960 – 2009, 1918 – 1959 and 1900 – 1917 periods respectively. In the practice of seismic hazard model development, catalogue thresholds on the order of M_W 4.5 to M_W 4.0 or even lower are routinely used for estimation of recurrence parameters. Whilst it should be emphasised that the Global Instrumental Catalogue provides the most accurate homogenous representation of seismicity across the globe, a need has emerged in many regions of the world to supplement this catalogue with other, possibly more local, catalogues at lower magnitudes. It is expected, therefore, that local catalogues, including those from the contributing agencies to the ISC bulletin, will play a role in the calculation of recurrence. Furthermore, it is also important for a seismic hazard modeller to understand and fully characterise the magnitude uncertainty that emerges from the catalogue homogenisation process. It is this that has motivated the development of these catalogue homogenisation tools, yet it should be recognised that this process needs to be treated carefully, and that for the optimum accuracy the merging and homogenisation should be undertaken at the waveform level – a process that requires a considerably greater investment of resources. It may be suggested therefore, that the catalogue homogenisation tools developed here are not, in the long term, a substitute for more direct homogenisation and integration of local catalogues into the ISC bulletin.

The process of catalogue preparation and homogenisation of multiple catalogues from a non-waveform approach can essentially be broken down into four steps:

i) Identification of the event solutions (time and location) common to the catalogues (duplicate finding)
ii) For a set of solutions (time and location) representing an event, selection of the preferred solution for the event
iii) Given a catalogue containing the full set of solutions and magnitudes for an event, development of empirical or physical relations between the magnitudes in the originally recorded scales ("native" magnitudes) by the original agencies ("native" agencies), and those in the magnitude scale to which the catalogue should be homogenised ("target" magnitude and agency)
iv) Given a set of empirical models relating the native magnitudes to the corresponding target magnitude, the hierarchy in which conversion models should be applied for the purposes of homogenisation

Steps i) and ii) should be optimally solved by referring to the original waveforms for the event and relocating, where possible. In practice, it may be difficult to retrieve the original waveform data; hence, it may be common to apply more general user-judgement or statistical procedures. These tools are currently in development and are shall not be the primary focus of the discussion here. Instead the initial catalogue homogenisation tools address steps iii) and iv): the magnitude conversion and homogenisation problem.

For conversion and homogenisation, the initial starting point is the full ISC bulletin, containing the solutions from the contributing agencies, typically represented in the IMS 1.0 format. Within this format each event may contain one or more locations (time and hypocentre) and each location may be associated with zero, one or more magnitude estimates in different magnitude scales. The catalogue homogenisation tool allows the user to query the catalogue of events, returning a set of native and target magnitudes subject to filtering criteria (i.e. agency, geographical location, time etc.). For those magnitudes matching the query criteria, the user can specify the form of the function preferred for regression analysis. At present, the functional forms supported are linear, n^{th}-order polynomial and n-segment piecewise linear. Orthogonal distance regression is used to take into consideration uncertainty in the native and target magnitude values.

2.1.1 Homogenisation problems

When implementing tools to undertake orthogonal distance regression regression, two potential problems can emerge in the process. The first problem is that, depending on the filtering and catalogue querying criteria set by the user, the query may return multiple solutions for each event. For example, a simple query of the catalogue considering M_S as the native magnitude and M_W as the target may, for some events, return multiple solutions that match the criteria if no agency filtering is applied. The regressions tools need to understand how to treat this data. Consider an example case in which a single M_W target magnitude is given to the event, and three native M_S values are found. A total of three M_S-M_W pairs are returned. One option is to treat all three estimates independent for the purposes of regression. Alternatively, a solution could be selected at random from the set, or ideally subject to a set of conditions (the most likely being the selection of the M_S-M_W pair that contains uncertainty estimates on both native and target magnitudes, or subsequently selecting the solution with the smallest total uncertainty ($\sqrt{\sigma_{native}^2 + \sigma_{target}^2}$). Good practice would suggest that the user should simply re-define their selection criteria in order to reduce the possibility of returning multiple solutions for an event, but in doing so this may risk returning too few data points to which a stable regression can be applied.

The second problem that emerges is that magnitude uncertainty is not always reported by a substantial number of agencies. As missing uncertainty values cannot be treated by standard orthogonal distance regression techniques, it is again necessary for the user to inform the tool how to rectify this problem. Available options may include assigning a user-defined σ_M term to missing magnitude uncertainties, or assigning values such as the largest observed uncertainty or a quantile of the observed uncertainties. Alternatively, the user may opt to remove from consideration those events for which the uncertainties are missing, but again this may result in too few events from which a stable regression estimate can be applied.

2.2. Earthquake Catalogue Processing

2.2.1 Declustering

Upon compilation of a magnitude-homogenised earthquake catalogue for a region, it is necessary for the purposes of time-independent seismic hazard calculation, to identify non-Poissonian events from the catalogue (i.e. foreshocks, aftershocks, transient clusters etc.) and to remove these from the calculation of the “steady-state” recurrence rate. The declustering process has been widely studied in statistical seismology, and many different algorithms are available to undertake this task, each operating in different ways or utilising different assumptions about the nature of the earthquake clustering process. A summary of many of the different algorithms available for this purpose can be found in van Stiphout *et al.* (2010). Although, many declustering algorithms are available, in practice the majority of hazard studies tend to utilise a few of the most common and established algorithms, such as Gardner & Knopoff (1974) or Reasenberg (1985). Often the motivation for using a particular algorithm may not always be clear, or at least may not always be made explicit in the documentation of the hazard model. As established previously, a key scientific motivation for the development of the Modeller’s Toolkit is to understand how many of the most fundamental modelling decisions impact both the key elements of the hazard model, such as the parameters of the recurrence model, and ultimately the resulting hazard analysis. For the first implementation of the toolkit, three declustering algorithms are available (Gardner & Knopoff, 1974; Reasenberg, 1985; Musson, 1999 [“AFTERAN”]), which are based on the family of “window” declustering algorithms. More declustering algorithms will be implemented in future, as they become available to us and as time and user requirements allow. In addition to the implementation of the declustering algorithms themselves, a basic χ^2 testing procedure to determine the confidence level in the Poisson assumption for the declustered catalogue (e.g. Luen & Stark, 2012) is also intended for future development.

2.2.2. Statistical Completeness

For calculation of the recurrence parameters for a region, the Modeller’s Toolkit implements several algorithms to allow for an estimation of the temporal variation in catalogue completeness for an input earthquake catalogue. In the initial implementation of the Modeller’s Toolkit, two algorithms are given: i) Stepp (1971) and ii) Assumed Magnitude-Frequency Distribution (MFD), the latter of which contains a family of methods that estimate the completeness magnitude based on the point of deviation from classic Gutenberg and Richter (1944) power law (Woessner & Wiemer, 2005), the choice of method being a user-configurable parameter. Other catalogue-based algorithms such as that of Alborello *et al.* (2001) are intended for future versions.

As shall be seen shortly, if the intention of the user is to create a “workflow” approach, effectively chaining together the processes, then it is necessary to take steps to ensure consistency in the outputs of each step so that different algorithm choices can be used interchangeably. Completeness estimation is one such example. For calculation of the a- and b-value for a catalogue, it is necessary to take into account time-variation of completeness, assuming spatial variation could be accounted for by selection of events from the catalogue according to some geospatial criteria. Unlike the Stepp (1971) or Albarello *et al.* (2001) methodologies, which give time-dependent estimates of completeness, the assumed MFD methodologies provide only a single estimate of completeness magnitude for a given event set. Therefore, in order to ensure consistency of output, the time-variation in completeness using the assumed MFD methodologies must be accounted for by subdividing the catalogue into overlapping time periods according to user-defined criteria (e.g. window size and increment, or minimum number of events).

In practice, the potential shortcomings of catalogue-based statistical estimators of completeness are recognised (Mignan & Woessner, 2012), and it may often be the case that completeness estimation can be undertaken using alternative methods. The probabilistic network-based approach of Schorlemmer & Woessner (2008), and a corresponding historical seismicity implementation (Felzer, 2008), are not supported directly at present. This is simply because they require a far more detailed description of the recording conditions than can be accounted for in the currently supported data formats, although we hope that this will be revised in future. Instead, if independent estimators of completeness are required, the user can input the time-varying completeness properties in a separate input file, which will be used instead of the completeness estimation calculators.

2.3. Earthquake Recurrence based on Instrumental/Historical Seismicity

2.3.1 a- and b-value

For the observational seismicity data, the current Modeller’s Toolkit calculates a- and b-value of the Gutenberg & Richter (1944) model using one of two methods:

i) Weichert (1980)
ii) Maximum Likelihood (e.g. Bender, 1983)

The Weichert (1980) algorithm accounts explicitly for the time-variation in completeness, whilst the Maximum Likelihood methods apply to a single catalogue. Again, for consistency of implementation, the Maximum Likelihood methods are fitted to different subsets of the catalogue corresponding to different time periods and corresponding completeness levels. The “final” recurrence values are derived based on a weighted mean of the different subsets. As expected, in tests on synthetic catalogues with known completeness and recurrence properties, the Weichert (1980) method provides a more accurate and stable estimate of completeness than the adapted Maximum Likelihood approach. Consequently, it is strongly recommended to use that specific methodology in practice. A Bayesian adaptation of the Weichert (1980) method has been applied in several major hazard initiatives (e.g. Johnston et al., 1994) and is being considered for implementation subject to further testing and evaluation. Similarly, maximum likelihood estimators for alternative recurrence models, such as the tapered Gutenberg & Richter (1944) model (Kagan, 2002) are also in the development stage.

2.3.2 Instrumental Estimators of Maximum Magnitude

In addition to the recurrence calculators, and in order to provide a comprehensive set of parameters describing the truncated Gutenberg & Richter (1944) distribution, the user can opt to calculate maximum magnitude for the zone using one of two statistical techniques: i) the non-parametric Gaussian method (Kijko, 2004), or ii) a bootstrapped cumulative-moment methodology adapted from Makropoulos & Burton (1983).

The bootstrapped cumulative moment approach is an adaptation of the cumulative strain energy method first proposed by Markopoulos & Burton (1983). This method considers the total moment release over the catalogue to identify upper and lower moment release limits. The difference between these limits provides an estimate of the maximum moment release for a catalogue. The uncertainty in M_{max} is estimated by bootstrapping the catalogue, sampling the magnitude uncertainty on each bootstrap. For both methodologies, the estimators give similar results when a large ($n > 100$) number of events are given. The estimate of maximum magnitude variance using the Kijko (2004) non-parametric Gaussian approach is heavily conditional on the variance of the observed maximum magnitude, whereas the variance estimated using the bootstrapped cumulative moment approach take into account the uncertainties on all the magnitudes in the catalogue. In both cases, however, the estimators are very sensitive to the catalogue duration and only become stable if the catalogue captures the full strain cycle in a region (ideally over several cycles). Consequently, these methods perform poorly in stable continental or low seismicity zones. In these cases the EPRI approach (Johnston *et al*. 1994) is often preferred, and this methodology is a high priority for the future.

2.4. Earthquake Recurrence on Active Faults from Geological Slip

The main body of the scientific tools in the Modeller's Toolkit have so far addressed the "instrumental seismicity" aspects of the model development. In many of the state-of-the-art models such as the Uniform California Earthquake Rupture Forecast Version 2 (Field *et al.,* 2009), the 2002 New Zealand national hazard map (Stirling *et al*., 2002), the 2009 Japan National Seismic Hazard Map (NIED, 2009) and the current SHARE project, there is clear precedent for the integration of active fault data into the hazard model. GEM obviously continues to develop this initiative, as evident in the GEM Global Component on the Faulted Earth. Following this process, it is important that the Modeller's Toolkit provides tools calculating recurrence on active fault sources utilising geological slip information.

2.4.1 Defining the fault source

The crucial stage in defining a geological recurrence model calculator is to provide the means of defining the input fault source. At present, two fault source typologies are supported:

i) Simple Fault Source: where the geometry is defined by the fault trace (a multi-segment on the surface), a dip value and an upper and lower seismogenic depth
ii) Complex Fault Source: where geometry is defined by multiple edges (a multi-segment laying on the fault surface), to support less regular fault types such as subduction interfaces or listric faults.

A web-based GIS tool has been developed as part of the GEM Faulted Earth project, from which the user can define the fault source following a process of interpretation and iteration of the fault geometry and its corresponding attributes; beginning with point observations and fault traces, eventually resulting in complete fault geometries and interpreted seismogenic fault sources. Within this framework, each fault is associated with geological attributes including slip, slip-type and proportion of aseismic slip. The definition of a standard data format for representing the corresponding fault information is in development.

2.4.2 Geological Recurrence Calculators

Utilising the geological slip information, the Modeller's Toolkit defines the earthquake recurrence model according to the evenly discretised magnitude frequency distribution (i.e. a discrete representation of the fault recurrence to allow for the support of different time-independent models). Calculators corresponding to two families of time-independent recurrence models are implemented: those based on a Gutenberg & Richter (1944) recurrence model (e.g. Anderson & Luco, 1983; Bungum, 2007) and those based on a characteristic earthquake model (e.g. Field *et al.*, 2009). The choice of recurrence model must be defined by the user, but the toolkit can support multiple recurrence model types provided they are associated with corresponding weights.

2.4.3 Tectonic Regionalisation

The implementation of active fault recurrence calculators may, depending on the choice of calculator, require the specification of parameters that are not necessarily known for every source, but which could be inferred on the basis of the tectonic region in which the fault is found. The problem of tectonic regionalisation from both a seismogenic source and a ground motion prediction equation (GMPE) perspective is currently being addressed by a GEM Working Group in Tectonic Regionalisation. The particulars of this project need not be addressed here, but it is recognised that the regionalisation can assist in providing a more complete definition of the fault for the purposes of recurrence calculation. Attributes of the fault that may be associated with a tectonic region may include the choice of the magnitude scaling model, the aspect ratio of rupture on the fault, the shear modulus and the strain drop (displacement to length ratio of the fault). In many cases, it is more likely that the tectonic regionalisation will define multiple parameters or models, which will be associated with weights that can later be treated within the logic tree analysis of the hazard calculation.

2.5. Model Integration – Directions for Combining Calculators

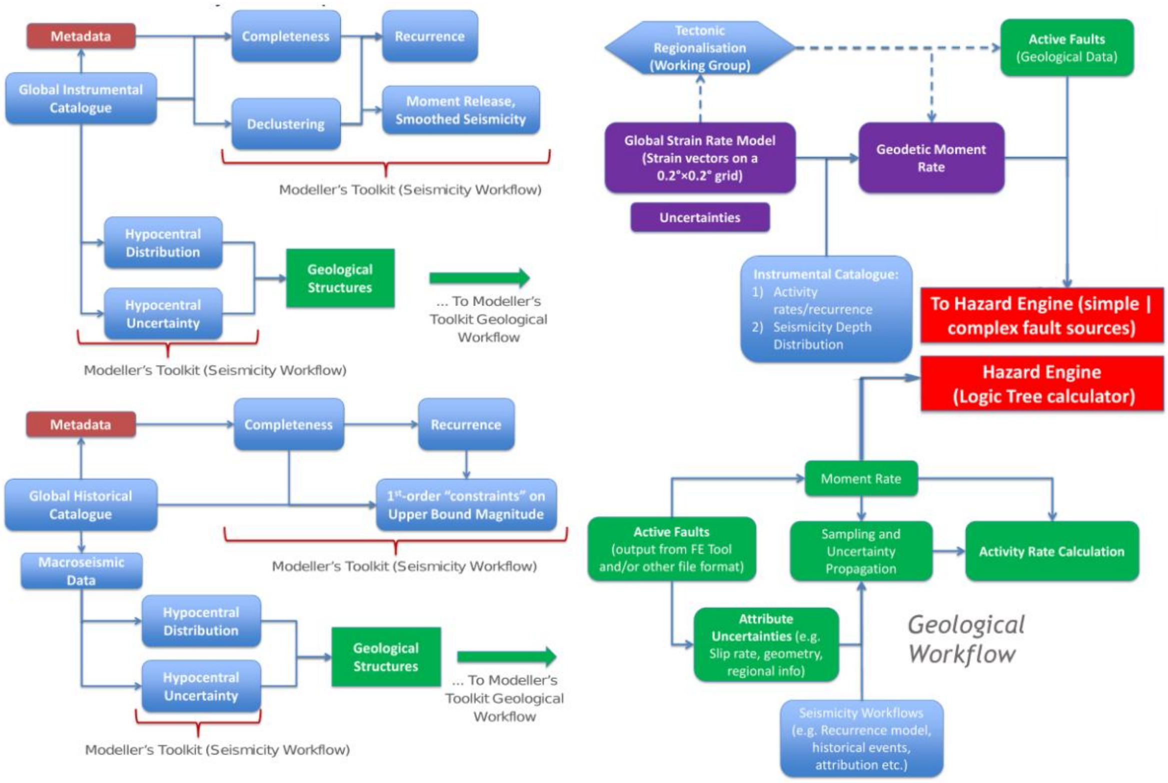

Figure 1.2. Integration of the different Modeller's Toolkit workflows

At the heart of the design the Modeller's Toolkit, and a key motivation behind its conception, is recognition that in building a global earthquake hazard model it is important to attempt to reconcile information coming from the various sources. In particular, it is necessary to ensure consistency between conceptual definitions that are common to multiple lines of data input, the tectonic regionalisation being one such case. Ultimately, however, the aim of the seismic hazard modeller should focus on reconciliation of the various data sets. The toolkit can be steered toward this objective. One such way is to calculate and compare, for a given zone or region, the seismic activity rates using different input data, such as geological slip, geodetic strain and observed seismicity. A tentative step in this direction has already been undertaken as part of the SHARE project, albeit using a customised set of tools for that purpose. Alternatively, the toolkit functionalities can be extended in order to test, within a probabilistic framework, the impact of the underlying uncertainties in the contributing data sets. This can provide the basis for integration into the OpenQuake hazard calculation tools, by establishing the logic tree for epistemic uncertainty analysis.

3. GROUND MOTION PREDICTION EQUATION (GMPE) TOOLS

The primary focus of the Modeller's Toolkit to this point has been on the development of the seismogenic source model. The selection of GMPE's has also been the subject of substantial discussion in the scientific literature (e.g. Bommer *et al.,* 2010). To assist in the selection process, there are several tools that are in the planning stages for development. It is important to recognise, however, that GMPE testing tools are contingent on the implementation of the GMPEs themselves inside the OpenQuake hazard calculation engine. On this basis, possible tools include visualisation functionalities, including higher-dimensional visualisation tools such as those outlined in Scherbaum *et al*. (2010). The current GEM Global Component on GMPEs is currently implementing tools for GMPE testing. As this component reaches completion, it is anticipated that some of the testing code can be moved into the Modeller's Toolkit for future usage.

4. CONCLUSIONS AND FUTURE DIRECTIONS

Whilst it is evident that the GEM Modeller's Toolkit remains in an early stage of development, there are many open issues in the hazard model development process that are being addressed. Also, the need for more rigorous definition and integration of the uncertainty on many attributes of the input data is becoming self-evident. As one possible conduit for knowledge exchange between GEM's Global Components and its Regional Programmes, the Modeller's Toolkit can serve an important function in understanding the needs of the hazard modelling community.

Several of the highest priority future development plans have already been mentioned, but there are more features that are already in the early development stage. These include the integration of tools for calculating recurrence and activity rates from geodetic strain data (Bird & Liu, 2007), Bayesian calculators for probabilistic attribution of observed seismicity to active faults (Wesson *et al.,* 2003) and a greater suite of functionalities for integration of parameter and model epistemic uncertainties with a view towards logic tree development.

Whilst the Modeller's Toolkit development may be driven primarily by the needs of the GEM initiatives and the Regional Programmes, the software itself is intended to represent a community driven tool. The software is distributed under a GNU Affero General Public Licence 3.0, which makes it available for use by the scientific community. User-testing, requirements and scientific feedback on the tools is an essential part of the development process, and it is hoped that this will continue to grow in the future.

ACKNOWLEDGEMENTS

Several members GEM Model Facility development team have been heavily instrumental in constructing the software architecture and we are particularly grateful to Giuseppe Vallarelli and Andrea Cerisera for their involvement. We have also benefitted from ongoing discussions with scientists and participants of many regional programmes and those involved in a wide variety of GEM initiatives and collaborations.

REFERENCES

Anderson, J. G. and Luco, J. E. (1983). Consequences of Slip Rate Constraints on Earthquake Occurrence Relations. *Bulletin of the Seismological Society of America.* **73.** 471 - 496

Albarello, D., Camassi, R. and Rebez, A. (2001). Detection of Space and Time Heterogeneity in the Completeness of a Seismic Catalogue by a Statistical Approach: An Application to the Italian Area. *Bulletin of the Seismological Society of America.* **91**.1694 - 1703

Bender, B. (1983). Maximum Likelihood Estimation of b Values for Magnitude Grouped Data. *Bulletin of the Seismological Society of America.* **73**. 831 - 851

Bird, P., and Liu, Z. (2007). Seismic Hazard Inferred from Tectonics: California *Seismological Research Letters.* **78**. 37 - 48

Bommer, J. J., Douglas, J., Scherbaum, F., Cotton, F., Bungum, H., Fäh, D. (2010). On the selection of ground-motion prediction equations for seismic Hazard analysis. *Seismological. Research Letters.* **81.** 783-793.

Bungum, H. (2007). Numerical Modelling of Fault Acitivities. *Computers & Geosciences*. **33**. 808 - 820

Crowley, H., D. Monelli, M. Pagani, V. Silva, G. Weatherill. (2011). OpenQuake User's Manual. The GEM Foundation, Pavia, Italy (http://openquake.org/users)

Felzer, K. R. (2008). The Uniform California Earthquake Rupture Forecast, version 2 (UCERF 2) Appendix I: Calculating California Seismicity Rates. *U.S. Geological Survey Open File Report 2007-01347I*

Field, E. H., Dawson, T. E., Felzer, K. R., Frankel, A. D., Gupta, V., Jordan, T. H., Parsons, T., Petersen, M. D., Stein, R. S., Weldon II, R. J. and Willis, C. J. (2009). Uniform California Earthquake Rupture Forecast, Version 2 (UCERF 2). *Bulletin of the Seismological Society of America.* **99**. 2053-2107

Gardner, J. K. and Knopoff, L. (1974). Is the sequence of earthquakes in Southern California, with aftershocks removed, Poissonian? *Bulletin of the Seismological Society of America*. **64**. 1363 - 1367

Giardini, D., Grünthal, G., Shedlock, K., M. and Zhang, P. (1999). The GSHAP Global Earthquake Hazard Map. *Annali di Geofisica*, **42:6**, 1225 - 1230

Gutenberg, B. and Richter, C. F. (1944). Frequency of Earthquakes in California. *Bulletin of the Seismological Society of America.* **34**.185 - 188

Johnston, A. C., Coppersmith, K. J., Kanter, L. R. and Cornell, C. A. (1994). The Earthquakes of Stable Continental Regions. *Electrical Power Research Institute (EPRI) Technical Report.* TR-102261-V1.

Kagan, Y. Y. (2002). Seismic moment distribution revisited: I. Statistical Results. *Geophysical Journal International.* **148**. 520 - 541

Kijko, A. (2004). Estimation of the Maximum Earthquake Magnitude, m_{max}. *Pure and Applied Geophysics*. **161**. 1655 – 1681

Luen, B., and Stark, P. B. (2012). Poisson tests of declustered catalogues. *Geophysical Journal International*. **189**. 691 – 700.

Makropoulos , K. C., and Burton, P. W. (1983). Seismic Risk of Circum-Pacific Earthquakes I. Strain Energy Release. *Pure and Applied Geophysics*. **121:2**. 247 - 267

Mignan, A., and Woessner, J. (2012). Estimating the magnitude of completeness in earthquake catalogs. *Community Online Resource for Statistical Seismicity Analysis.* doi:10.5078/corssa-00180805.

Musson, R. M. W. (1999). Probabilistic Seismic Hazard Maps for the North Balkan Region. *Annali di Geofisica,* **42**.1109 - 1124

National Research Institute for Earth Science and Disaster Prevention, Japan (NIED). (2009). Technical Reports on National Seismic Hazard Maps for Japan. *Technical Note of the National Research Institute for Earth Science and Disaster Prevention*. **336**

Pinho, R. (2012). GEM: a participatory framework for open, state-of-the-art models and tools for earthquake risk assessment worldwide", *Proceedings of the 15th World Conference on Earthquake Engineering*, Lisbon, Portugal.

Reasenberg, P. (1985). Second-order moment of central California seismicity, 1969-82. *Journal of Geophysical Research.* **90.** 5479 - 5495

Scherbaum, F., Keuhn, N. M., Ohrnberger, M. and Koehler, A. (2010). Exploring the Proximity of Ground Motion Models Using High-Dimensional Visualisation Techniques. *Earthquake Spectra***. 26:4**. 1117 - 1138

Schorlemmer, D. and Woessner, J. (2008). Probability of Detecting an Earthquake. *Bulletin of the Seismological Society of America.* **98**. 2103 - 2117

Stepp, J. C. (1971). An investigation of earthquake risk in the Puget Sound area by the use of the type I distribution of largest extremes. *PhD Thesis.* Pennsylvania State University

Stirling, M. W., McVerry, G. H., Berryman, K. (2002). A New Seismic Hazard Model for New Zealand. *Bulletin of the Seismological Society of America*. **92:5**. 1878 - 1903

van Stiphout, T., Zhuang, J. and Marsan, D. (2012). Seismicity declustering. *Community Online Resource for Statistical Seismicity Analysis*. doi:10.5078/corssa-52382934.

Weichert, D. H. (1980). Estimation of the Earthquake Recurrance Parameters for Unequal Observation Periods for Different Magnitudes. *Bulletin of the Seismological Society of America.* **70**. 1337 – 1346

Wesson, R. L., Bakun, W. H. And Perkins, D. M. (2003). Association of Earthquakes and Faults in the San Francisco Bay Area Using Beyesian Inference. *Bulletin of the Seismological Society of America*. **93:3**. 1306 - 1332

Woessner, J. and Wiemer, S. (2005). Assessing the quality of earthquake catalogues: estimating the magnitude of completeness and its uncertainty. *Bulletin of the Seismological Society of America.* **95**. 684 - 698