DEVELOPING A GLOBAL RISK ENGINE

VITOR SILVA
Doctoral Student
University of Aveiro
Aveiro - Portugal

HELEN CROWLEY
Risk Coordinator
GEM Foundation
Pavia – Italy

RUI PINHO
Secretary General
GEM Foundation
Pavia – Italy

HUMBERTO VARUM
Associate Professor
University of Aveiro
Aveiro - Portugal

SUMMARY

Risk analysis is a critical link in the reduction of casualties and damages due to earthquakes. Recognition of this relation has led to a rapid rise in demand for accurate, reliable and flexible risk assessment software. However, there is a significant disparity between the high quality scientific data developed by researchers and the availability of versatile, open and user-friendly risk analysis tools to meet the demands of end-users. In the past few years several open-source software have been developed that play an important role in the seismic research, such as *OpenSHA* and *OpenSEES*. There is however still a gap when it comes to open-source risk assessment tools and software. In order to fill this gap, the Global Earthquake Model (GEM) has been created. GEM is an internationally sanctioned program initiated by the OECD that aims to build independent, open standards to calculate and communicate earthquake risk around the world. This initiative started with a one-year pilot project named GEM1, during which an evaluation of a number of existing risk software was carried out. After a critical review of the results it was concluded that none of the software were adequate for GEM requirements and therefore, a new object-oriented tool was to be developed. This paper presents a summary of some of the most well known applications used in risk analysis, highlighting the main aspects that were considered for the development of this risk platform. The research that was carried out in order to gather all of the necessary information to build this tool was distributed in four different areas: information technology approach, seismic hazard resources, vulnerability assessment methodologies and sources of exposure data. The main aspects and findings for each of these areas will be presented as well as how these features were incorporated in the up-to-date risk engine. Currently, the risk engine is capable of predicting human or economical losses worldwide considering both deterministic and probabilistic-based events, using vulnerability curves.
A first version of GEM will become available at the end of 2013. Until then the risk engine will continue to be developed by a growing community of developers, using a dedicated open-source platform.

1. INTRODUCTION

Great improvements have been made in the fields of structural analysis, hazard prediction or vulnerability assessment in the past decades. However, the world has experienced a significant increase of the economic and human losses due to earthquakes. The exponential growth of the population in developing countries located in high seismic hazard zones have greatly contributed to this situation and it is likely that for the next few decades the risk will continue to rise in these regions unless measures for risk mitigation are taken [Bommer, 2002]. The reduction of the seismic risk can be done by the identification of the zones with higher risk and consequent improvement of the earthquake resistance of the exposed elements and establishment of regulations that will force new structures to be built according to seismic design codes. These studies can also help national governments to provide financial incentives for the good practices of construction, to implement emergency plans and relief funds, to create insurance systems and to forbid the construction in particularly dangerous regions. Spence [2004] proved that these types of regulation could provide benefits on the reduction of economic and human losses. Unfortunately, due to the lack of resources and available data, the identification of regions with high seismic risk has only been done in some parts of the world, and in only of a few cases these results

were used to propose risk mitigation actions (e.g.: Balassanian *et al.*, 1999, Kreimer *et al*., 1999, Bakıra and Boduroglu , 2004, CSSC, 2007).
There is a clear bottleneck between the developments that have been done by researchers and the demands of end-users. This is reflected by the lack of a transparent, reliable and flexible seismic risk platform capable of meeting the wide variety of end-users that can go from a simple house owner to an international institution [Porter and Scawthorn, 2007]. Each type of user will naturally have different expertise on the subject and needs regarding the types of results. In order to understand all of the possible situations that a global risk engine needs to be capable of support, the following parameters were considered:

- Number of Events: This parameter distinguishes the analysis in terms of deterministic scenario-based (only one earthquake considered) or probabilistic scenario-based (set of earthquakes within a certain time span).
- Number of locations: This variable is faced as a geographic site with a specific seismic hazard that might contain one or more assets;
- Number of Assets: An asset in this study is interpreted as an instance of an element at risk that is exposed to the seismic hazard. A collection of assets might exist at a single location.
- Spatial correlation: The methods necessary to employ in order to estimate the losses due to seismic events might vary considerably if spatial correlation of the ground motion is required or not.

Using different combinations between the above variables, it was possible to achieve eight different cases. This set of combinations can be seen in the proposed diagram:

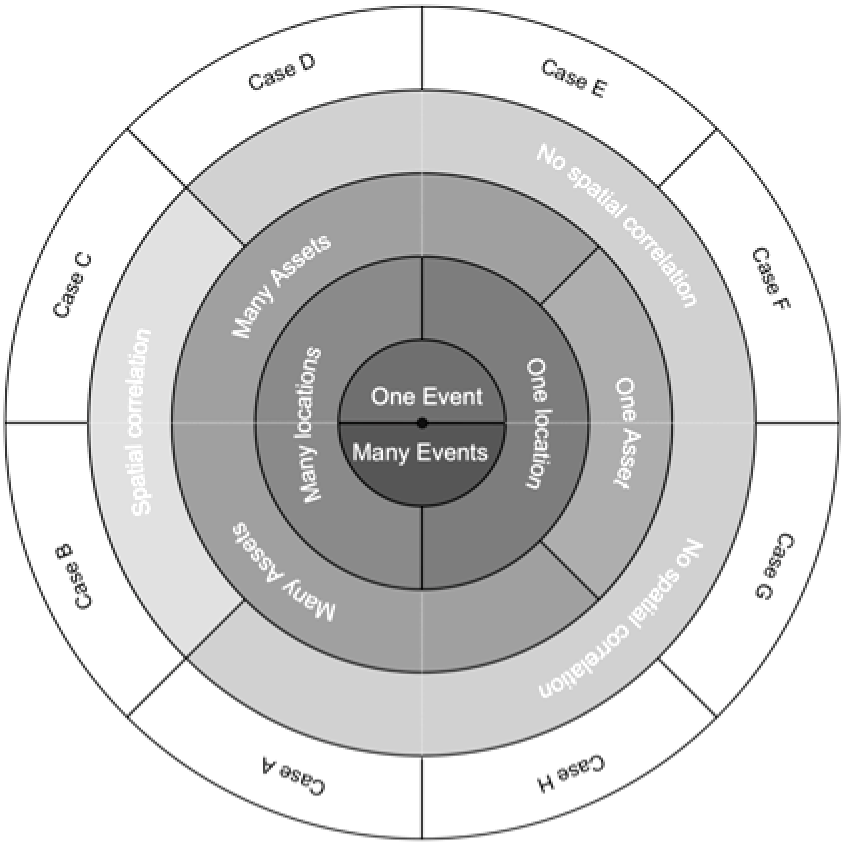


Figure 1: Possible cases in risk assessment.

In order to understand the above diagram, readers should start from the centre and proceed from tier to tier until the most peripheral layer (which contains the case type) is achieved. For instance, Case A represents a probabilistic scenario-based analysis where many locations that contain various assets are being considered and there is no need in taking into account the spatial correlation of the ground motion. The influence that each of these parameters has in designing a risk engine will be highlighted throughout this paper.

The development of such a tool also requires a good planning of the architecture of the code and development philosophy. A group of experts on the areas of risk and seismic assessment (most of them with great experience on the development of similar tools) were consulted in order to better understand what should be the requirements of a global risk engine. Based on their suggestions, the developing risk team compiled a list of requirements [Crowley *et al*,, 2010] that can be summarized as follows:

- Open-source software development: the source code of the global risk calculator should be available to any user and the development of the code should be a product of the efforts of a community, and not just limited to a working group;
- Platform independent: this tool should be able to be used in any operative system;
- Flexible: this code needs to be developed with the purpose of creating a platform for risk assessment, instead of another static risk application.
- Multi-hazard: Although only losses due to earthquakes are being considered at the moment, this platform should in the future allow the estimation of losses due to other hazards such as floods or hurricanes.
- Dynamic: this calculator should allow users to update their results based on newer models, datasets or hazard inputs;
- Modular and expandable: this risk calculator should be developed in a way that any user can easily implement and combine different methodologies. To make this attribute possible, an object-oriented philosophy should be adopted.
- Scalable: this tool should allow one to perform risk assessment at different levels of resolution from an urban level to a global scale.

It was also suggested that the risk team should not start the development of the software from scratch. A critical evaluation of existing risk software should be carried out in order to understand the strengths and weaknesses of each application and use the knowledge gained from this review to avoid past mistakes. The following section describes some of the results from this study.

2. RISK ASSESSMENT TOOLS

As recommended by a group of scientific experts, a critical evaluation of the existing risk software was conducted within the GEM1 project. A list of ten applications were selected and distributed among three institutions (EUCENTRE, GFZ and NORSAR) to be evaluated. With the objective of learning from these risk software, several tests were carried and parameters such as the IT characteristics, hazard typologies, vulnerability methodologies, exposure elements supported and type of possible outputs were approached. Different test-bed applications were also performed using these tools, with the purpose of analysing how the outputs were being produced and to evaluate the computational performance. It is important to understand that the objective of this part of the study was not to validate the tools or to conclude which ones were providing more accurate results, but to simply understand their capacities and functionalities. A detailed description of all of the results and conclusions can be found in Crowley *et al*. [2010]. The following table summarizes some of the main results of this evaluation:

Table 1 - Summary of the evaluation of the risk software

		SELENA (NORSAR)	EQRM (G.A.)	ELER (KOERI)	QLARM (WAPMERR)	CEDIM (CEDIM)	CAPRA (World Bank)	RISKSCAPE (GNS/NIWA)	LNECLOSS (LNEC)	MAEVIZ (MAE Center)	OPENRISK (SPA Risk)
IT details	Programming language	Matlab / C	Python	Matlab	Java	Visual Basic	Visual Basic	Java	Fortran	Java	Java
	Availability	OS	OS	SA	SC	SC	SC	SA	SC	OS	SA
	GUI	Yes	No	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes
Hazard	Multi-hazards	No	No	No	No	No	Yes	Yes	No	Yes	No
	Type of scenarios	D / P	D / P	D	D	D	D / P	D	D	D	P
Vulnerability	Approach	A	A	A / E1 / E2	E2	E2	A	E2	A / E1 / E2	A	A
	Non structural damage	No	Yes	No	No	No	No	No	No	Yes	No
Exposure	Building inventory	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
	Economic inventory	Yes	Yes	No	No	Yes	Yes	Yes	Yes	Yes	Yes
	Population inventory	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
	Region applicability	User defined	User defined	User defined	World	User defined	User defined	New Zealand	User defined	User defined	California

OS - Open Source (code can be acquired freely)

SA - Standalone application (available under request)

SC - Source code only (available under request)

D - Deterministic scenario-based

P - Probabilistic scenario-based

A - Analytical approach (e.g. capacity spectrum, displacement coefficient)

E1 - Empirical approach (using intensity measures, e.g.: MMI, MSK, EMS)

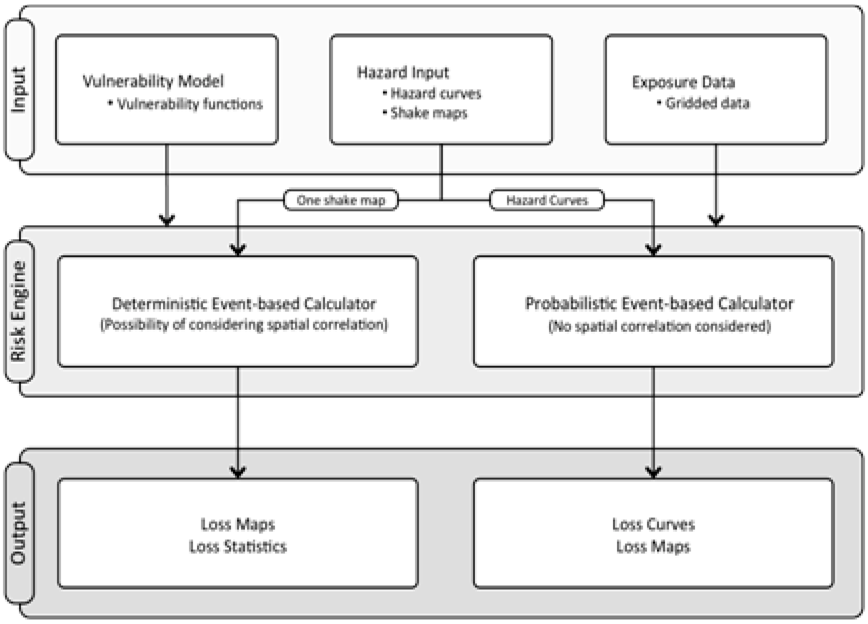
E2 - Empirical approach (using instrumental measures, e.g.: PGA, PGV, SA)

From the evaluation of the existing software it was clear that a significant effort should be made in order to simplify the way the data is provided to the applications and how each parameter is defined. Testing these applications was not a simple task even for the experts in risk assessment and for this reason, it is important to focus a good portion of the development enhancing the ways this tool will interact with users with different degrees of expertise. Several options can be choosen to minimize these issues such as: make the tool available through a web interface that can be customized based on user's capabilities, produce standalone applications with good tutorials and help systems, make the source code available in public code repositories along with high quality documentation and even more important, produce some pre-computed results that can be easily interpreted by users less experienced with risk assessment. All of these alternatives are being considered and have been documented in Crowley *et al*. [2010].
Regarding the need to cover all of the cases that were represented in Figure 1, it is possible to conclude that some of the applications were quite close to fulfilling this requirement. A third of the evaluated tools were capable of performing both deterministic and probabilistic scenario-based analyses (using different hazard inputs such as hazard curves or sets of ground-motion fields, also known as shake maps) and although all of their development was done with the purpose of assessing several assets per analysis, it is also possible to run the calculations for a single element. Thus, a great amount of information can be used from the evaluation of these tools and furthermore, from interacting with the developers responsible for the creation of those tools.
HAZUS [FEMA, 2003], which is one of the most popular tools for risk assessment, was also among the applications that was considered for testing. However, due to the fact that this application is extremely computational demanding, it was decided not to proceed with its evaluation. Nevertheless, it is important to understand that this tool has been the basis of some of the codes that were tested herein and therefore, its evaluation was implicit.
The insurance and reinsurance industry have developed similar applications to estimate future losses due to earthquakes in order to fix their annual premiums. Some examples of this type of tools include RISKLINK®, EQECAT, ALLRISK and Advanced Component Method™. One of the major drawbacks of the models developed by this industry is the lack of information regarding the methodologies that are being used, that usually are only revealed to the client, if at all [Crowley *et al*., 2006].
Overall, the experience gained from the evaluation of the existing risk software was precious in order to identify the aspects that should be avoided, the features that must be incorporated and to define an efficient development strategy for the GEM global risk engine.

3. GLOBAL RISK ENGINE

3.1. Current status

The current risk engine is capable of performing deterministic scenario-based loss assessment using ground-motion fields and probabilistic scenario-based loss assessment using hazard curves. The way the different parameters are defined, the calculations that need to be employed and how the data is provided to the risk engine varies significantly depending of which scenario was chosen. However, due to the object-oriented design of the engine, a number of common pieces of code are used for both analyses. Concerning the vulnerability, the collection of curves that comprises the model needs to be defined at a discrete number of intensity measure levels, as will be described in the following section. The uncertainty of each value can be expressed as a standard deviation or as a coefficient of variation for both the seismic hazard (on the ground-motion fields) and the vulnerability model. If the user does not want to consider the uncertainty, these parameters can simply be set to zero and all of the calculations are carried out considering the respective values as deterministic. The implementation of this feature was quite important since from the evaluation of the existing software it was noticed that most of the applications compute losses without the option of considering uncertainties in the vulnerability functions or ground-motion fields. Regarding the coverage of the list of cases proposed in Figure 1, the current risk engine only covers case A, C and D and not fully since the exposure data needs to be defined in a specific format that might not be compatible with every exposure inventories. The assessment of single locations with one or many assets and a probabilistic loss assessment capable of considering the spatial correlation of the ground motion are not yet implemented. The limitations and future developments of the risk engine will be further discussed in the following sections. The up-to-date architecture followed by the risk engine is illustrated in Figure 2.



Risk engine architecture.

The classes and objects that contain all of the calculations are located in a core that is shielded from the way the data is provided to the risk engine or how the results are going to be exported. This way, should users have their data in a specific format not currently supported by the risk engine, a new interface would be created and the calculations would still be valid. This is an important feature since some of the data that needs to be provided to the risk engine can vary significantly as it will be shown later (e.g. hazard input, exposure data).

3.2. Seismic hazard input

The risk engine has been developed in a way that it can read the seismic hazard input from a file, a database or directly from a hazard engine. For the time being, all of the calculations have been performed by reading this data from customized file formats. However, due to its modular architecture, other ways of inserting the seismic hazard can be easily incorporated into the calculator. This feature might be particularly important for users who have their own hazard data already and do not want to re-compute these values but simply to introduce them into the risk calculator.

As previously mentioned, the risk engine uses ground-motion fields in order to perform deterministic scenario-based analysis and hazard curves for the probabilistic scenario-based analysis. Both hazard inputs can be acquired through many resources such as seismic hazard applications, web-based tools or online archives. While testing the calculator, it was decided to use data from different sources in order to evaluate how flexible the risk engine should be regarding the introduction of the hazard input. It was concluded that each resource tend to provide the values in its own format and that it was necessary to develop a standard format capable of storing this information. A file format as been proposed for each type of seismic hazard and various technologies have been studied (e.g. ASCII, XML, YAML) [Crowley *et al.* 2010].

The evaluation of existing seismic hazard software was also one of the main goals of the GEM1 project. In order to do so, a list of seismic hazard applications were collected and tested with the purposed of evaluating factors such as: availability, quality of the documentation, computational performance, flexibility, programming language, existence of a graphical user interface and software requirements. All of the results are documented in Danciu *el al.* [2010] and the following table describes some of the characteristics of the existing seismic hazard software that were evaluated in the previously mentioned study:

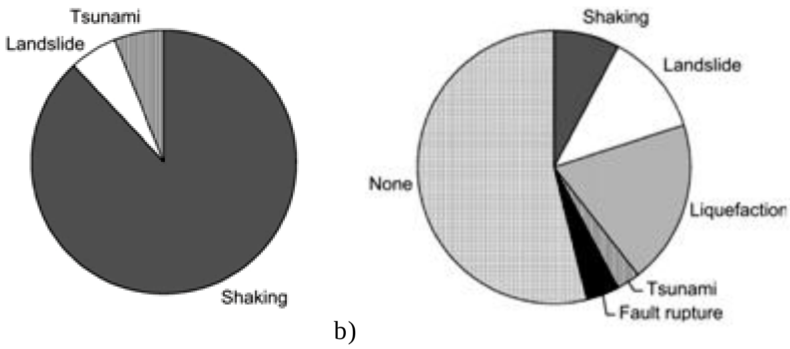
Table 2 - Characteristics of existing seismic hazard software.

<table>
<tr><th rowspan="2">Software</th><th rowspan="2">Developers / Affiliation</th><th rowspan="2">Availability</th><th rowspan="2">GUI</th><th rowspan="2">PSHA approach</th><th colspan="3">Output</th></tr>
<tr><th>Hazard curves</th><th>Uniform Haz. Spectra</th><th>Hazard maps</th></tr>
<tr><td>CRISIS</td><td>M. Ordaz / UNAM</td><td>Free by request</td><td>Yes</td><td>Classic</td><td rowspan="10">Yes</td><td>Yes</td><td rowspan="10">Yes</td></tr>
<tr><td>EQRM</td><td>D. Robinson / GA</td><td>Open Source</td><td>No</td><td>Monte Carlo</td><td>No</td></tr>
<tr><td>FRISK 88M</td><td>R. K. McGuire / Risk Engineering, Inc.</td><td>Proprietary</td><td>No</td><td>Classic</td><td>Yes</td></tr>
<tr><td>MoCaHAZ</td><td>S. Wiemer / ETH Zurich</td><td>Free by request</td><td>No</td><td>Monte Carlo</td><td>Yes</td></tr>
<tr><td>MRS</td><td>R. LaForge / USBR</td><td>Free by request</td><td>No</td><td>Classic</td><td>Yes</td></tr>
<tr><td>NSHMP</td><td>Frankel et al. / USGS</td><td>Free download</td><td>No</td><td>Classic</td><td>Yes</td></tr>
<tr><td>OHAZ</td><td>B. Zabukovec / GSS</td><td>Free download</td><td>Yes</td><td>Classic</td><td>No</td></tr>
<tr><td>OpenSHA</td><td>N. Field / USGS</td><td>Open Source</td><td>Yes</td><td>Classic</td><td>Yes</td></tr>
<tr><td>SEISRISK</td><td>Bender, Perkins & LaForge</td><td>Free download</td><td>No</td><td>Classic</td><td>No</td></tr>
<tr><td>SEISHAZ</td><td>M. Stirling / GNS</td><td>Proprietary</td><td>No</td><td>Classic</td><td>Yes</td></tr>
</table>

3.2.1. Future developments

The employment of ground-motion fields and hazard curves to represent the seismic hazard was decided mainly due to its simplicity, which was fundamental considering the time constrains that the risk team had. A single ground-motion field with the spatial distribution of the ground motion and a second map with the respective uncertainty are being provided to perform deterministic scenario-based assessments. However, this type of calculations can also be done using for the same seismic event hundreds of ground-motion fields whose ground motion distribution reflect the uncertainty from the ground motion prediction equations (GMPE). This way of estimating losses for single-event scenarios is currently being implemented on the risk engine. Another important future development that will be incorporated into the calculator is the capacity of performing probabilistic scenario-based analysis using thousands of shakes maps generated through a stochastic event set. This approach brings advantages such as the possibility of computing statistical parameters (e.g. aggregated losses per ground-motion field, total aggregated losses, total standard deviation) or the consideration of the spatial correlation of the ground motion.

Another aspect that will be discussed in the future developments of the risk engine is the consideration of damage due to ground failure. Existing risk software clearly lacks of the capacity of considering such effects since only MAEVIZ and RISKSCAPE consider losses due to liquefaction or tsunamis. Although ground shaking is considered to be the primary cause of economical and human losses, there are numerous examples of earthquakes where the losses due to landslides, liquefaction and ground rupture have been significant. These effects are frequently the cause of major disruptions, particularly to lifelines, which can lead to prolonged loss of function and income. Bird and Bommer [2004] studied the losses due to ground failure in 50 earthquakes that happened since 1989 and the primary and secondary causes of the damages that occurred in those events were identified. It was concluded that ground shaking was the primary cause of building damage in 88% of the earthquakes reviewed, while tsunamis and landslide had a toll of 12%. The scenario changes when evaluating the second most significant cause of damage in which landslides and liquefaction are responsible for 32% of the losses. The following figure illustrates the distribution of the causes of building damage observed throughout the reviewed 50 events:

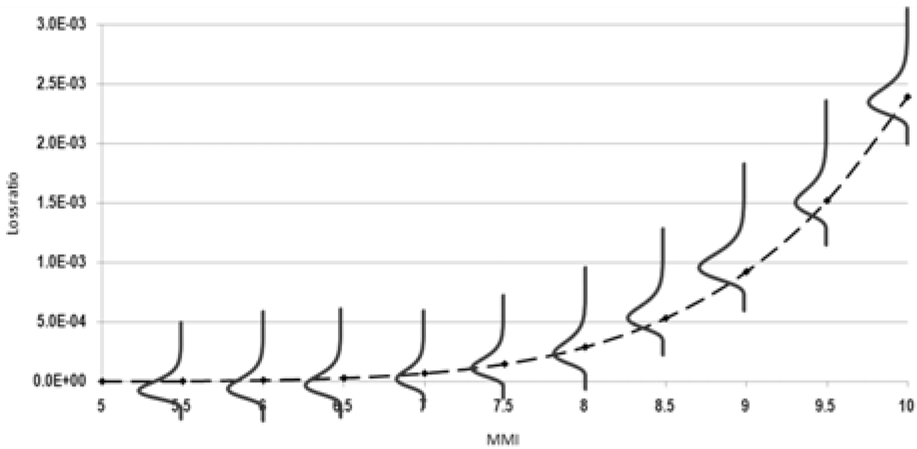


a) b)

Causes of earthquake damage to buildings: a) primary cause, b) secondary cause.

3.3. Vulnerability module

The influence of the vulnerability of the exposed elements to seismic events is fundamental to identify the magnitude of losses. A simple comparison between earthquakes occurred in Third World regions and developed countries reveals the critical importance of the vulnerability. For instance, the Spitak (Armenia) earthquake of December 1988 had a magnitude of M_s 6.7 and left a death toll of about 25.000 casualties. Less than a year after, an earthquake with a greater magnitude (M_s 7.0) occurred in Loma Prieta (California, USA) causing a number of human losses smaller than 70 [Bommer, 2002]. Several methodologies to represent the vulnerability were considered within the GEM1 project and for simplicity reasons, it was decided that in this initial phase vulnerability curves would be used to estimate losses. These vulnerability functions relate a list of ground motion levels with the respective loss ratios and associated coefficients of variation. The uncertainty on the loss ratio is assumed to have a lognormal distribution. In the following figure, an example of how the vulnerability curves are being defined is presented:



Vulnerability function as defined on the GEM risk engine.

3.3.1. Future developments

A great amount of improvements can be done in order to expand the risk engine to be compatible with other ways of relating seismic hazard with losses or damage states. The following methodologies are in the future plans of the risk engine:

- **Continuous vulnerability functions**: Currently the risk engine requires the vulnerability functions to be defined in a discrete number of points as shown in Figure 3. However, users might have vulnerability functions whose behaviour can be described by an analytical expression and therefore, enabling the capacity of supporting continuous functions is fundamental.

- **Fragility functions**: This type of functions can relate ground motion with the probability of exceedance a certain damage or limit state. Through this relation is possible to build two types of maps: one that can present the expected damage state for a given probability of exceedance and time span and another that can provide the probability of exceedance for a given damage state and time span. The later has been already developed by the United States Geological Survey (USGS) [1] through a web interface, where a user starts by choosing a set of parameters that characterizes the construction type and the desired damage state (from slightly to completely damaged) and then, a map can be extracted with the spatial distribution of the estimated probability of exceedance. Unfortunately, this tool currently only covers the United States territory.
- **Damage probability matrices**: These matrixes can establish a direct relation between macroseismic intensities and expected loss ratio for a given building typology. Each value of the matrix indicates the probability of a certain building class be affected by a specific damage state. Damage probability matrices (DPM) are often created based on post-seismic data (Whitman *et al.*, 1974, Braga *et al.*, 1986, Chávez, 1998) and since the damage suffered by each building is evaluated based on observations, the expert's judgement is fundamental. With the development of accurate analytical model and due to the lack of past earthquake data, non-linear dynamic analysis can also be used to produce DPMs [Kappos et al., 1995, 1998].

The methods that are employed in order to produce the previously mentioned vulnerability methodologies can also vary significantly. In the early 70's, vulnerability/fragility curves and damage probability matrixes were created through the employment of **empirical methods** that used macroseismic intensities since at the time, hazard maps were mainly defined in terms of these discrete damage scales. The emergence of more ground motion prediction equations in terms of instrumental measures such as spectral acceleration (SA) or velocity (SV), as opposed to macroseismic intensity or peak ground acceleration (PGA), has given rise to the development of **analytical methods**. These approaches tend to feature slightly more detailed and transparent vulnerability assessment algorithms with direct physical meaning, that not only allow detailed sensitivity studies to be undertaken, but also cater to straightforward calibration to various characteristics of building stock and seismic hazard [Calvi *et al.*, 2006]. Within the various analytical methods, the Capacity Spectrum Method and Displacement-Based methods are planned for future developments of the risk engine.

The developments done on the analytical models can be used not to fully replace the empirical methods but rather to improve them. This leads to **hybrid methods** that combine post-earthquake damage statistics with data produced through mathematical models to create vulnerability curves or DPM. This approach is particularly advantageous when there is lack of damage data at certain intensity levels or when calibration of analytical models is necessary. Furthermore, the usage of post-earthquake data might also reduce the computational effort that would be required if only analytical methods would be carried out.

3.4. Sources of Exposure data

3.4.1. Global buildings distribution

A spatial distribution database of buildings at a global scale does not currently exist, but methods to estimate the built area/number of buildings from census data and remote sensing will be developed and applied within GEM together with crowd data collection methods. A solution to this problem is to use national building databases that might be enough for deterministic event-based assessment or even for probabilistic scenario-based analysis if a user only wants to estimate losses for a restricted region. Such databases have been created already for some large cities exposed to regular seismic activity such as Istanbul, Lisbon, Wellington, Managua or Los Angeles. The risk team has used some of these building inventories in the software evaluation study.

The Prompt Assessment of Global Earthquakes for Response (PAGER) [2] group has already taken the first steps in the development of a global building inventory database [Jaiswal and Wald, 2009]. This inventory consists of estimates of the fractions of building types observed in each country, their functional use and their average day and night occupancy. Fours tables, each reflecting a combination of rural or urban and residential or non-residential categories essentially comprises this database. The fraction of building types or dwellings and their occupancy characteristics have been collated for each country to represent a country-based distribution using the PAGER structure taxonomy. This construction type classification was a product of an evaluation of several sources that classify buildings according to attributes such as structural system, load transfer mechanisms, predominant construction material, performance during past earthquakes, etc. The PAGER group recognized that none of the existing sources could provide an adequate classification of the buildings since most of them required building-specific information that was not available for most of the inventory data. Hence, it was necessary to adopt a classification based on the material used for the construction of the walls and roofs. This classification is similar to the one used by most of the housing census and surveys carried out for a large number of countries in the past. In order to produce this building inventory, a great number of data sources were studied and rated according to the process used to gather the data. The following table describes all the sources that were used, the coverage of each one and the assigned quality rating:

Table 3 - Inventory data sources.

Source of data	Quality Rating	Global Coverage
World Housing Encyclopedia	Medium	110 residential construction types in 37 countries. Exact fraction of each housing type in a given country is not known The day and night time occupancy by construction type is available.
UN Database	Low	44 countries with construction type description based on external walls and 96 countries with type of housing units. About 110 countries with the average occupancy estimated based on total building stock.
Census of Housing	Medium	197 countries conducted housing census in 1990. Several countries do not publish housing statistics even though housing census was conducted.
Published Literature	High	About 10 countries have been identified that contains high quality information based on the conducting survey and the verification of other published information such as census/tax assessor's data. The day and night occupancy by construction type is not available.
WHE-PAGER Survey	High	Inventory information for about 22 countries has been gathered in the first phase of WHE- PAGER expert opinion survey. In order to facilitate their judgment, country-specific inventory information gathered from general internet research and housing censuses was provided to these experts.

The described database will be constantly updated as more data become available and therefore, it is expected that some parts of the inventory will be replaced by better quality data.

3.4.2. Global population distribution

Databases that can provide population count or density for each location in the world already exist. Within the GEM1 project, two databases were reviewed: LandScan^{tm} [3] and GRUMP [4]. A description of both sources is provided herein highlighting the main advantages of each one.

LandScan^{tm}

The Oak Ridge National Laboratory produces this database and it can be obtained in an ESRI grid format or ESRI binary raster format. It has a 30 arc second spatial resolution (about one square kilometre at the equator) and a great number of data sources were used to create this database such as: Digital Chart of the World [5], VMap1 (a map of major roads and rail networks, drainage networks, utility systems, elevation contours, coastlines, international boundaries and populated places), night-time lights, Global Land Cover Characterization (GLCC) [6] and high-resolution aerial photography and satellite imagery. The dataset provides population estimates based on aggregate data for second order administrative units from the US Census Bureau's International Program Center [7]. An algorithm is used that assesses the likelihood of population occurrence in grid cells based on parameters such as road proximity, slope, land cover or night-time lights. There is no specific distinction made between urban and rural areas, though urban areas can be inferred by analyzing population density [Dobson *et al*., 2000]. The most recent version of LandScan^{tm} is from 2008 and it has a world coverage that goes from North 84 degrees to South 90 degrees and West 180 degrees to East 180 degrees. The values of each cell are integers, which represent average population count. LandScan^{tm} is free of charge for educational institutions and non-profit organizations.

GRUMP
The Global Rural Urban Mapping Project (GRUMP) dataset is produced by the Center for International Earth Science Information Network (CIESIN) and it is currently an alpha version with a 30 arc second resolution. A combination of datasets have been used to produce this database such as Census data with georeferenced human settlements data (55,000) with a population of more than 1000, Gazetteer [8], City Population [9], Digital Chart of the World [5], administrative boundaries datasets and United Nations (UN) national estimates [10]. The population is assigned to the grid cells considering the classification of the areas in urban or rural. After this, checks are made on the total population in each administrative unit according to UN estimates. The urban extents are defined by a combination of datasets and not just night-time lights (recorded in a period between 1994 and 1995), that can miss small settlements in less developed countries and greatly overestimate urban extents for large settlements. [Nachtergaele and Petri, 2007]. The value assigned to each cell is an integer that represents the estimated population count. GRUMP can be acquired free of charge from the CIESIN website but its licence does not allow redistribution of the data without CIESIN clearance.

Critical evaluation
Although a few spot checks of population databases can be made, there will be some areas of the world where one dataset performs better than another and the opposite might be true in another area. A study by Gunasekera et al. [2009] has shown that in Istanbul the LandScan^{tm} data was closer to the ground truth, but more comparisons are needed to understand which population database is more reliable. For this reason, considering that the population density is thus an epistemic uncertainty in loss calculations, both the GRUMP and LandScan^{tm} databases were implemented in the risk engine.
One problem with LandScan^{tm} concerns the road database. The model processes the input layers by country without taking into consideration the spatial continuity of the roads networks between them, resulting in uneven changes of population density at country boundaries. Another problem is that, owing to the way in which LandScan^{tm} processing methods evolved, population comparisons between available revisions of the database is not possible. Also, the underlying models have not been published or few information can be found about them and hence, assumptions employed by LandScan^{tm} to distribute population counts per cell are not known. Another problem that has been identified within this dataset is the fact that night-time lights can be more linearly correlated with GDP (gross domestic product) and electrification than population density and there is a clear "blooming" effect which means that the extents of urban areas are often overestimated and some small settlements are not clearly identified [Elvidge *et al*., 2004].
Regarding GRUMP, the main advantage is that it uses population data from the census, rather than predicting it based only on probability coefficient or lighted areas. Also, it makes use of other geographic information systems data to identify urban areas, compensating for the small settlements in poor countries that are not detected by the night-time lights. The resulting grid is a dataset of population distribution that takes into account the urban and rural areas. Some limitations of GRUMP are the fact that although it recognizes that applying a threshold would reduce the number of small settlements that are not frequently identified by night-time lights, due to the complexity of finding a single threshold that would work globally, no light threshold was applied yet. [Salvatore *et al*., 2005]. This database also lacks of a continuous updating and therefore, the population distribution for three time periods can only be found: 1990, 1995 and 2000. Besides this, it is also important to understand that the lights factor refers to a time between 1994 and 1995 for both databases and hence, care should be taken when using this database in countries that experienced a fast growth during the last decade (e.g. Kuala Lumpur, Malaysia).

3.4.3. Future developments

The value of the exposed elements can be presented in two main ways: in a gridded format where this information is indicated for each grid cell, or in a discrete format where these values are defined at each location by a pair of geographical coordinates. Currently, the risk engine is capable of reading the data regarding the exposure inventory in the gridded format that usually appears in a raster binary file (the previously mentioned databases for population distribution use this type of format). However, it is more common to have information regarding buildings in a discrete format (portfolio of buildings), since it allows users to store more detailed information for each asset such as occupancy type (e.g. residential, commercial, industrial), number of occupants, contents value, replacement cost and even the type of soil which allows the treatment of site effects with more accuracy. For these reasons, it is fundamental to extend the risk engine to allow the introduction of the exposure data also in this format. It is also important to understand that the parameters that define each asset may vary considerably depending of how the vulnerability is handled. If an analytical approach is used in the assessment of the vulnerability, parameters regarding the geometry, structural behaviour and construction materials might be necessary to incorporate in the exposure data. Furthermore, a third format will be necessary to develop in order to allow users to store information regarding elements that have a large spatial distribution such as lifelines (e.g. pipelines, electrical networks, roads, water supply). Recognizing that there is not a unique way of storing exposure data, the development of the exposure module of the risk engine will be done considering several formats, knowing that a flexible interface will reduce the effort that any user will have to employ in order to introduce his data into the calculator.

Another issue that will be approached in the future is the lack of dynamics that current databases have and how fast they can get outdated. The majority of the databases reflect the value of a certain type of element at a certain instant and in environments that experience heavy changes over time, the results produced using that datasets might not be valid or become obsolete too quickly. As an example, according to the Population Reference Bureau of the United States [PRB, 2010], in 40 years India will be the largest country in the world regarding population achieving 1.748 millions people, followed by China with 1.437 and then the United States with 423 millions. The opposite will also happen in countries such as Japan, Russia and Norway where a decrease of the population is expected. Overall, it is estimated that in 2025 the world population will increase to 8.108 millions and in 2050 will reach a value of more than 9.400 millions of people. These changes might also cause a significant variation on the value of building stock since different social needs will be felt. In such a dynamic environment, the estimated losses for a time span of 50 years or an emergency plan that is meant to be used for the next decades should not be produced based on a static database.

4. CONCLUSIONS

The initial development of the risk calculator that is being developed for GEM has been described herein. A review of the existing risk software was presented, highlighting the main features regarding the IT characteristics, hazard input, vulnerability methodologies and exposure data. It was concluded that none of the tools were completely fulfilling GEM requirements and therefore, a new risk calculator was to be developed. Through the study of these tools, it was possible to understand the fundamental features that should be implemented and the different ways that the results should be presented to the wide variety of the users.

The up to date architecture of the risk engine was presented and currently the calculator is capable of performing deterministic scenario-based loss assessment using shake maps and probabilistic scenario-based loss assessment using hazard curves. It was acknowledged that other ways of doing these calculations exist and will be incorporated in the risk engine in the future, such as the use of thousands of ground motion fields generated using stochastic event sets. Regarding the hazard module, a list of different tools that can provide seismic hazard input was presented and the importance of considering other types of hazard such as landslides and liquefaction was described.

Concerning the vulnerability, it was recognized that there are several ways of defining this property and that the most widely used methodologies such as vulnerability curves, fragility functions and damage probability matrixes should be implemented. The way these vulnerability measures can be generated can also vary significantly and hence, it was concluded that the development of the risk engine should allow the incorporation of modules that can produce those parameters through the introduction of empirical data, the employment of analytical methods or through the hybrid combination of the last two approaches.

Finally, a study was done with the purpose of understanding the availability of exposure data regarding population and building distribution. It was concluded that datasets capable of providing population count or density for any region in the world already exist and should be used in this first stage of the development. However, the same does not happen when it comes to building inventories where a database that could provide a spatial distribution of buildings globally does not currently exist. A first attempt to create such database has been

done by the PAGER group and some important sources of information regarding buildings distribution were presented.
This extensive study of the different components required for risk assessment allowed the risk team not just to gather valuable resources, information and data, but even more importantly, to understand that the development of a global risk engine needs to be a collective effort of a community of developers and not just a product of a localized team.

5. ACKNOWLEDGMENTS

The authors are very grateful for the information obtained from conversations with Miriam Colombi and Marco Pagani, especially on the evaluation of the existing risk and hazard software. It is also appreciated the review from Ben Wyss that led to a significant improvement of the paper.

6. REFERENCES

[1] Balassanian, S., Melkoumian, M., Arakelyan, A. and Azarian,A. (1999) Seismic risk assessment for the territory of Armenia and strategy of its mitigation. *Natural Hazards*, **20**(1), 43–55.

[2] Bakıra, P., Boduroglu , H. (2004) Earthquake Risk and Hazard Mitigation in Turkey. *Earthquake Spectra*, **18**(3), 427–447.

[3] Bird, J. and Bommer, J. (2004) Earthquake Losses due to Ground Failure. *Engineering Geology*, **75**(2), 147-179.

[4] Bommer, J., (2002) *Basics of Seismology and Seismic Hazard Assessment*. Course Notes at European School for Advanced Studies in Reduction of Seismic Risk (ROSE School), Pavia, Italy.

[5] Braga, F., Dolce, M. and Liberatore, D. (1982) A Statistical Study on Damaged Buildings and an Ensuing Review of the MSK-76 Scale. Proceedings of the Seventh European Conference on Earthquake Engineering, Athens, Greece, 431-450.

[6] Calvi, M., Pinho, R., Magenes, G., Bommer, J., Restrepo-Vélez, L., Crowley, H. (2006) Development of Seismic Vulnerability Assessment Methodologies over the past 30 years. *ISET Journal of Earthquake Technology*, **43**(3), 75-104.

[7] Chávez, J. (1998) Evaluación de la vulnerabilidad y el riesgo sísmico a escala regional: Aplicación a Cataluña, *PhD thesis,* Universidad Politécnica de Cataluña, Barcelona, Spain.

[8] Crowley, H., Pinho, R., Bommer, J.J. and Bird, J.F. (2006) Development of a Displacement-Based Method for Earthquake Loss Assessment. Report 2006/01, European School for Advanced Studies in Reduction of Seismic Risk (ROSE School), Pavia, Italy.

[9] Crowley H., Cerisara A., Colombi M., Crempien J., Erduran E., Field N., Jaiswal K., Keller N., Lopez M., Liu H., Luco N., Pagani M., Mayfield M., Milanesi M, Porter K., Silva V., Wald D., Wyss B. (2010) GEM1 Risk Report. GEM Technical Report No 9, GEM Foundation, Pavia, Italy.

[10] CSSM (2007) *California Earthquake Loss Reduction Plan*. Alfred E. Alquist Seismic Safety Commission, Sacramento, USA.

[11] Danciu, L., Pagani, M., Monelli, D., Wiemer, S. (2010) Evaluation of Seismic Hazard Codes: GEM1 perspective. GEM Technical Report No 1, GEM Foundation, Pavia, Italy.

[12] Dobson, J., Bright, E., Coleman, P., Durfee, R., Worley, B. (2000) LandScan: A Global Population Database for Estimating Populations at Risk. *Photogrammetric Engineering and Remote Sensing*, **66**(7), 849-857.

[13] Elvidge, C., Safran, J., Nelson, I., Tuttle, B., Hobson, V., Baugh, K., Dietz, J., Erwin, E. (2004) Area and position accuracy of DMSP nighttime lights data. *Remote Sensing and GIS Accuracy Assessment*, 281-292.

[14] FEMA (2003) *HAZUS-MH technical manual*. Federal Emergency Management Agency, Washington DC, USA.

[15] Gunasekera, R., Toyos, G., Saito, K. (2009) Sensitivity analysis and application of remotely derived global population datasets in risk assessment of natural phenomena. Remote Sensing and Photogrammetry Society Annual Conference 2009, Leicester, United Kingdom.

[16] Jaiswal, K.S., and Wald, D.J. (2009) A Global Building Inventory for Earthquake Loss Estimation and Risk Management. Technical Report 2008-1160, U.S. Geological Survey, Colorado, United States of America.

[17] Kappos, A., Pitilakis, K., Stylianidis, K. (1995) Cost-Benefit Analysis for the Seismic Rehabilitation of Buildings in Thessaloniki, Based on a Hybrid Method of Vulnerability Assessment. Proceedings of the Fifth International Conference on Seismic Zonation, Nice, France, 1, 406-413.

[18] Kappos, A.J., Stylianidis, K.C. and Pitilakis, K. (1998) Development of Seismic Risk Scenarios Based on a Hybrid Method of Vulnerability Assessment. Natural Hazards, **17**(2), 177- 192.

[19] Kreimer, A., Arnold, M., Barham, C., Freeman, P., Gilbert, R., Krimgold, F., Lester, R., Pollner, J., Vogt, T. (1999) *Managing disaster risk in Mexico: Market Incentives for mitigation investment*. The World Bank, Washington D.C., USA.

[20] Nachtergaele, F., Petri, M. (2007) Mapping Land Use Systems at global and regional scales for Land Degradation Assessment Analysis. Internal report, LADA project, Rome, Italy.

[21] Porter, K. and Scawthorn, C. (2007) Open-Source Risk Estimation Software. Technical Report, SPA Risk, Pasadena, United States of America.

[22] PRB (2010) *2010 World Population Data Sheet*. Population Reference Bureau, Washington, USA.

[23] Spence, R. (2004) Risk and regulation: can improved government action reduce the impacts of natural disasters?. *Building Research & Information*, **32**(5), 391-402, London, United Kingdom.

[24] Salvatore, M., Pozzi, F., Ataman, E., Huddleston, B. and Bloise, M. (2005) *Mapping global urban and rural population distributions*. Food and Agriculture Organization of the United Nations, Rome, Italy.

[25] Whitman, R., Reed, J., Hong, S. (1973) Earthquake Damage Probability Matrices. Proceedings of the Fifth World Conference on Earthquake Engineering, Rome, Italy, **2**, 2531-2540.

WEB REFERENCES

1. Probabilistic Seismic Risk Maps: http://earthquake.usgs.gov/hazards/risk
2. PAGER: http://earthquake.usgs.gov/earthquakes/pager
3. LandScan^{tm}: http://www.ornl.gov/sci/landscan
4. GRUMP (Global Rural-Urban Mapping Project): http://sedac.ciesin.columbia.edu/gpw
5. Digital Chart of the World: http://www.maproom.psu.edu/dcw
6. GLCC (Global Land Cover Characterization): http://edc2.usgs.gov/glcc/glcc.php
7. US Census Bureau’s International Program Center: http://www.census.gov/ipc/www/idb
8. Gazetteer: http://www.world-gazetteer.com
9. City Population: http://www.citypopulation.de
10. UN Population Division: http://www.un.org/popin