

# OPEN-SOURCE SOFTWARE AND TREATMENT OF EPISTEMIC UNCERTAINTIES IN EARTHQUAKE LOSS MODELLING

**R.** Pinho<sup>1</sup>, H. Crowley<sup>2</sup>, J.J. Bommer<sup>3</sup>

<sup>1</sup>Assistant Professor, Department of Structural Mechanics, University of Pavia, Via Ferrata 1, Pavia, Italy <sup>2</sup>European Centre for Training and Research in Earthquake Engineering (EUCENTRE), Pavia, Italy <sup>3</sup>Department of Civil and Environmental Engineering, Imperial College London, SW7 2AZ, UK Email: rui.pinho@unipv.it, helen.crowley@eucentre.it, j.bommer@imperial.ac.uk

# **ABSTRACT :**

Earthquake loss modelling is a multidisciplinary activity which requires the collection of databases of earthquake activity, ground conditions, ground-motion prediction models, building stock and infrastructure exposure, vulnerability characteristics of the exposed elements, repair costs and human casualties. Each component of the model carries large associated uncertainties which need to be propagated through the analysis, such that the total uncertainty in the resulting loss estimates can be obtained. However, a number of epistemic (i.e. knowledge-based) uncertainties present within a given method are often not accounted for. This paper describes various studies which have looked at the influence of these uncertainties in the loss model results and makes proposals for how these uncertainties could be included in future earthquake loss models.

**KEYWORDS:** Uncertainties, loss modeling, cat modeling, open-source software

# **1. INTRODUCTION**

Earthquake loss estimation has evolved into a major area of research activity. This is, however, a relatively recent phenomenon: a search through the literature to identify publications on topics such as 'earthquake loss estimation' and 'earthquake loss modelling' reveals that there were very few studies on these topics published before 1990. During the past 15 years the total research activity in earthquake loss modelling has grown exponentially and the topic now features prominently in most conferences on earthquake engineering and seismology, as well as being the subject of papers in most issues of journals in this field. There can be little doubt that the impetus for the rapid expansion of this field of research has come from the reinsurance industry and the growth of commercial earthquake insurance, there have also been several governmental projects for earthquake insurance, which have required loss models to guide decision-making with respect to the purchase of reinsurance.

Although it is the case that earthquake loss modelling is now an established area of research, with many groups in research institutes and universities engaged in the field, often with government funding, the needs of the insurance and reinsurance industries have been met mainly by a few well-known companies specializing in catastrophe (CAT) modelling. The achievements of these companies, both commercially and technically, are not to be underestimated, but the loss models that they produce and distribute are widely viewed as being 'black boxes': the user is provided with very little information about the inner details of the model. A number of arguments could be put forward to defend the policy of concealing the data, assumptions and algorithms that constitute the core of a commercially-produced earthquake loss model, the most obvious perhaps being that the investment in data collection and model development needs to be protected from competitors. However, the opaque nature of these commercial loss models has produced growing concern and disquiet in many circles: recently, for example, it is understood that the Financial Services Authority in the UK "expressed concerns about London market insurers relaying on 'black box' catastrophe modelling and underwriting solutions that mean little to anyone but the trained mathematicians, actuaries and geoscientists that produced them" (BGHRC, 2006).

In addition to such expressions of concern regarding the current situation, there are two other factors that



constitute significant challenges to the practice of vending 'black box' loss models, the first being the HAZUS loss estimation methodology developed by FEMA (FEMA, 2003); regardless of whatever technical criticisms may be made of HAZUS, it has three major merits: it is comprehensive, (mostly) transparent and freely available. The second factor is the growth of academic research in this field, in which transparency in terms of data, assumptions, methods and results is a fundamental requirement. In particular, academic research in the field of earthquake loss modelling is starting to produce detailed sensitivity studies, providing valuable insight into how much loss results depend on the models, data, uncertainties and assumptions employed.

The key point is that risk management is decision-making in the face of uncertainty, and hence the financial analyst using an earthquake model as the basis for risk management decisions not only needs to understand the components and workings of the loss model, but also the sensitivity of the results it yields to the uncertainties associated with the inputs (Bommer *et al.*, 2006).

# 2. IMPACT OF UNCERTAINTIES IN LOSS MODELS

Earthquake loss modelling carries many uncertainties in the models related to several elements: the collection of databases of earthquake activity; the ground conditions; the ground-motion prediction; the characterisation of the building stock and infrastructure exposure; the definition of the vulnerability characteristics of the exposed elements; and the estimation of repair costs and human casualties.

An earthquake loss model attempts to estimate the economic impact of future earthquakes on a defined area of exposure, which may be a city, region or country, or in some cases a portfolio of buildings and facilities within such a geographical area. This is an ambitious aim since the problem is very complex and there is major uncertainty associated with each element of the loss model. If one considers the basic elements of the simplest possible loss model, one that considers only direct losses (i.e., cost of repair or replacement) to a building stock due to the effects of strong ground shaking induced by earthquakes, clearly the first level of uncertainty concerns the choice of damage estimation methodology and one could expect significantly different results to be obtained using approaches based on macroseismic intensity, single ground-motion parameters (such as PGA or PGV), or displacement spectral ordinates. Durukal et al. (2006) compare loss results for a single earthquake scenario affecting Istanbul using intensity-based and spectral displacement-based vulnerabilities; the probable maximum loss (PML) ratio is estimated at 14% with the former method and as 28% using the latter, demonstrating just how sensitive the results can be to the adopted loss estimation approach. Spence et al. (2003) also looked at the difference in predicted damage distributions at three different locations considering the Kocaeli 1999 earthquake using an intensity-based method and a spectral displacement-based method. For the RC framed buildings the intensity model predicted fewer buildings with moderate and extensive damage, but more with complete damage compared with the spectral displacement model, but the two models predicted overall damage (Mean Damage Ratio) at about the same level: 40–60% for 'good' buildings, and 60–70% for 'poor' buildings. For the masonry buildings, the intensity model consistently predicted higher damage than the displacement model, by 50-60% at all three sites.

In both of these examples, different hazard definitions, vulnerability functions and even building class definitions have been required in the two methodologies that were compared. Hence, it is not clear whether the larger damage estimation in one method as compared to the other is due to an overestimation in the demand, an underestimation of the vulnerability or the chosen division of the exposure into different building classes. The separate impact of each of these components is discussed in the following sections through the presentation of a number of studies which have looked at the influence of uncertainties on single loss models.

# 2.1 Sensitivity to Choices in Hazard Model

Uncertainty in seismic hazard analysis is widely recognized as being significant and a key challenge of probabilistic seismic hazard assessment is to identify, quantify, and incorporate the sensitivities into the analysis. The two main sources of uncertainty are the models for seismicity and for ground motion, in other

# The 14<sup>th</sup> World Conference on Earthquake Engineering October 12-17, 2008, Beijing, China



words where and how often will earthquakes of different magnitudes occur, and what levels of ground motion will each of these events produce at each site of interest.

For a single scenario earthquake, the damage distribution is highly sensitive to the epistemic uncertainty in the ground-motion prediction equation. As part of a systematic sensitivity study for losses in the Sea of Marmara region due to single earthquake scenario, Crowley *et al.* (2005) explored this sensitivity by varying the gradient of the ascending linear branch of the displacement spectrum – as a surrogate for using alternative prediction equations – and found, for this particular case, an almost direct reflection of the changes in the Mean Damage Ratio (MDR, which is the ratio of cost of repair to cost of replacement), as shown in Figure 1.





Grossi (2000) looked at quantifying the uncertainty in the average annual loss (AAL) to residential structures in Oakland, California, using the HAZUS software (FEMA, 2003). For the seismic hazard estimation, the parameters/models considered in the sensitivity study included updated recurrence of earthquake events, an updated attenuation relationship, and an updated soil classification scheme. The modifications to the recurrence of earthquakes as compared to the default model led to an 18% reduction in the AAL, whilst the change in attenuation (ground-motion prediction) equation led to a 20% reduction in the loss, and the change in site classification resulted in a 6% reduction in the loss.

Another choice which can be made when calculating the damage to geographically distributed exposure with ground-motion prediction equations is whether to include the spatial correlation of the ground motions, and if it is included, the way in which it is modelled. Crowley et al. (2008a) have looked at the influence of including the spatial correlation of the aleatory variability in the ground motion for a single earthquake scenario in the Sea of Marmara region, Turkey. The scenario earthquake was repeated many times, each time considering a different ground-motion field based on different samples of intra-event (within-earthquake) aleatory variability at each site in the model. The variability in the mean damage ratio (MDR) considering the 1000 repetitions of the scenario earthquake, without the consideration of spatial correlation, was found to be 11%. Including the spatial correlation of the intra-event variability, based on the Wang and Takada (2005) model with a correlation length of 5km (where the correlation length is the distance between sites at which the correlation coefficient is equal to e<sup>-1</sup>), led to an increase in the variability of the losses from 11% to 19%, whilst a correlation length of 10km produced a variability of 27% in the MDR (see Figure 2). The model was thus seen to be highly sensitive to the chosen correlation length. Molas et al. (2006) have also looked at the influence of the correlation length on the variability of losses to a portfolio of buildings with the coefficient of variation of the loss varying from 9% to 30% due to an increase in the correlation length. Correlation lengths of about 5km have been calculated using Californian strong-motion data (Boore, 1997) whilst Wang and Takada (2005) have found lengths of up



to 40km in the data from Japan and Taiwan. Hence, the influence of the correlation length is certainly worth considering in loss models due to the large uncertainty in the definition of this parameter.



Figure 2 Variability in the mean damage ratio for a loss model with a single earthquake scenario considering spatial correlation of the ground motions (a) correlation length = 5km and (b) correlation length = 10km (Crowley *et al.*, 2008a)

In earthquake loss modelling there is an additional uncertainty, which relates to the choice of hazard calculation method. If the output is required in the form of loss curves (i.e., relationships between different levels of loss and their associated return periods), single earthquake scenarios are not appropriate. The most attractive option is to model the hazard using probabilistic seismic hazard assessment (PSHA) to produce maps of the selected ground-motion parameter for a number of selected return periods, and then to convolve these maps with the exposure. The alternative method, which is much more computationally intensive, is to model future earthquakes using Monte Carlo simulations to generate synthetic earthquake catalogues of very long duration. For calculating the seismic hazard at a particular location, the two methods will produce identical results, provided the stochastic earthquake catalogue is sufficiently long (Crowley and Bommer, 2006). For calculating losses in a geographically distributed portfolio, however, the methods yield different loss curves as the use of PSHA maps to define ground-motion fields at a given return period assumes that the ground motions are fully correlated at all sites. The use of multiple scenario earthquakes allows the inter-event (earthquake-to-earthquake) and intra-event variability of the ground motions at different sites to be explicitly modelled. In this way the correlation of ground-motions due to the same inter-event variability being considered for a given scenario earthquake can be accounted for whilst the intra-event variability at different sites can be modelled as independent or spatially correlated. The three sites which were considered were sufficiently distant from one another that spatial correlation was not included in the model. The results showed that the consideration of perfectly correlated ground motions at all sites (through the use of PSHA maps) leads to an underestimation of the losses at high annual frequencies of exceedance and an overestimation of the losses at low annual frequencies of exceedance.



Figure 3 Mean rate of exceedance curves for losses to a given portfolio computed using the six proposed models for simulating random fields of ground-motion intensity for each earthquake (Park *et al.*, 2007)



The influence of spatial correlation on loss exceedance curves has been considered by Park *et al.* (2007). Six different procedures for generating ground-motion fields were used: Case 1 used independent intra-event variability at each site in the model, Cases 2 to 4 considered different spatial correlation models, and Case 6 was based on perfectly positively correlated intra-event variability. The loss exceedance curves are presented in Figure 3; these curves show that neglecting to model the spatial correlation of ground-motion intensity (i.e., Case 1) systematically distorts the trend of the curve. The rare losses are consistently underestimated and the frequent losses are overestimated compared to those produced by Cases 3 and 4, which were considered as benchmarks for the exercise.

# 2.2 Sensitivity to Choices in Vulnerability Model

The main difficulty in defining the sensitivity of loss results to changes in the vulnerability predictions arises because vulnerability methods often require different definitions of the hazard, and thus it is not possible to consider exclusively the impact of changing the vulnerability assessment method. As discussed previously, Spence *et al.* (2003) and Durukal *et al.* (2006) both compared damage and loss predictions for Turkey based on different vulnerability functions, but one of the methods required an intensity-based definition of the hazard and the other a spectral-displacement based definition.

Crowley *et al.* (2005) looked at the impact of the epistemic uncertainty in the definition of the geometric and material properties of reinforced concrete structures on the losses from a single earthquake scenario in the Sea of Marmara region. The vulnerability prediction was based on the DBELA methodology (Crowley *et al.*, 2004) which calculates the variability in the limit state period of vibration and displacement capacity of classes of buildings based on simple geometrical and material properties (section dimension, steel yield strength, etc.). The period of vibration is used to calculate the displacement demand which is compared with the displacement capacity; the variability in each of these parameters is used to calculate the probability that the demand is higher than the capacity, and that the limit state is exceeded. In the aforementioned sensitivity study, modifications of 5-20% to the mean values of the geometric and material properties used in the base model were made, and the damage distribution was recalculated. A variation of up to 50% in the MDR from the base model was noted when all parameters were changed simultaneously to produce a lower or upper bound capacity.

Crowley *et al.* (2008b) applied two different mechanics-based methods in the calculation of seismic risk maps for Italy: DBELA (Crowley *et al.*, 2004) and SP-BELA (Borzi *et al.*, 2008). These two methods have many similarities as they both use the same hazard definition and the capacity definition of both methods is based on the calculation of the secant period of vibration and displacement capacity at different limit states. Figure 4 shows that despite their similarities, there is a notable difference in the mean number of buildings predicted to exceed the significant damage limit state per year in each municipality when calculated with the two different methodologies. The maximum difference in the number of buildings predicted to exceed the limit state with the two methods is around 34, which is a 98% difference in the number of buildings for that municipality.

As part of the NERIES project (Network of Research Infrastructures for European Seismology) a comparison of European vulnerability models has been carried out for a single earthquake scenario close to the city of Istanbul, Turkey (Strasser *et al.*, 2008). The same ground motions and exposure data were provided to all participants such that the vulnerability models could be explicitly compared. For the empirical vulnerability methods, MSK intensity was provided, whilst peak ground acceleration (PGA) and 5%-damped spectral accelerations at response periods of 0.2 and 1.0 s were provided to those using analytical vulnerability assessment methods. The results of the study were similar for the four intensity-based methods with predictions of between about 3% and 8% of buildings damaged beyond repair (KOERI-MSK, ESCENARIS Level 0 and 1 and SIGE-DPC), whilst the predictions for the three analytical methods varied to a much higher degree: 6.4% and 4.7% for the proportion of collapsed buildings for DBELA and KOERI-SD, but around 30% for the SELENA method (see Figure 5).





Figure 4 Mean annual seismic risk maps in terms of the number of buildings which exceed the significant damage limit state per municipality using (left) DBELA and (right) SP-BELA vulnerability predictions (Crowley *et al.*, 2008b)



Figure 5 Comparison of the distribution of collapsed buildings for a scenario earthquake near Istanbul predicted with three analytical methods: DBELA, KOERI-SD and SELENA (<u>http://www.neries-eu.org/</u>)

# 3. RECOMMENDATIONS FOR FUTURE OF LOSS MODELLING

In view of the very high sensitivity of loss estimates to the methodology employed, the seismicity and ground-motion models, the vulnerability functions, and assumed replacement costs, it is clear that future loss models should explicitly account for these epistemic uncertainties. Indeed, a cause of frequent concern in the insurance and reinsurance industries is precisely the fact that for certain regions and perils, the three main commercial CAT models (AIR, EQECAT and RMS) often yield significantly different loss estimates. Equally unsettling to many users is the fact that updates of the models sometimes lead to very significant changes in the losses compared to the previous version of the software. This situation arises because it is simply impossible to produce a precise deterministic estimate of the future losses to a given portfolio due to a specific peril. Molina and Lindholm (2005) have suggested that a logic-tree approach similar to that used in seismic hazard



assessment should be used for loss models. Such schemes might be feasible to treat the epistemic uncertainties within a given model (for example, the choice between different ground-motion prediction equations, or description of the capacity curves for the building classes within the model), but if the aim is to understand the variability that arises when different methodologies are employed then it may not be so straightforward. Different methodologies may use different descriptions of the hazard, which will give rise to the need to employ certain vulnerability functions which in turn may require particular building classes to be identified. However, if each methodology is clearly and transparently applied, with the shortcomings and benefits of each method clearly exposed by the developers, then the end-users can begin to compare the results and the uncertainty in these results from different models.

In order to allow comparisons of different loss models to be carried out in this way, there is a compelling case to open (i.e., make transparent to users and the scientific community) the 'black boxes' currently used very widely in the insurance and reinsurance industries. Open-source software is now widely available and used in many applications (e.g., AGORA, OpenSHA, MAEVIZ, EQRM). The idea is that the source code is made freely available to all users, who are then free to modify, improve and adapt it, and incorporate it into their own systems for redistribution. The argument is that the ability of the user community to 'own' the code enables much more rapid development, the spotting and removal of bugs, and ultimately produces better software. In other words, the benefits associated with open-source loss models may be split into two main categories; (i) advancing the state-of-the-art of catastrophe risk modelling and (ii) improved information sharing. There are some pitfalls with open-source modeling when it is not centrally controlled, as there is a danger that multiple copies of a single software may be in circulation and the modifications which have been carried out are not documented in any way. However, these shortcomings can be addressed by ensuring that there is a protocol to the modification of open-source software.

# 4. CLOSURE

Earthquake loss models are affected by many sources of uncertainty and these uncertainties can exert a very strong influence on the results obtained from the models. The key issue therefore is enabling the users of loss models to view the uncertainty in the loss estimates produced, in order to usefully inform decision making. The question that then arises is whether loss models should switch to displaying ranges of estimated loss curves, with associated confidence levels, rather than single best estimates? In part, the commercial loss models are employed single best estimates because the insurance and reinsurance industries are currently based on using single values as the basis for decision making. Even when two or more of the commercial CAT models are employed, the output will often be averaged (sometimes with uneven weighting) to obtain a single best estimated losses, there will need to be a paradigm shift in the industry to using such distributions, rather than single values, as the basis for operational decision-making. This represents a major challenge, but also could bring considerable benefits to the industry, since the decisions would be more informed, not only in terms of the different levels of risk being accumulated but also in the confidence in the quantifications of this risk. The confidence levels on the more pessimistic estimates could counsel against certain investments and help safeguard against major 'unforeseen' losses and bankruptcy.

There is also a commercial challenge to the CAT modelling companies. If one of the modellers were to break ranks now and opt to display complete distributions of loss estimates, reflecting the full range of uncertainties, then this could – in the current climate – easily have the undesirable result of simply undermining confidence in the model. Until the industry itself has come to terms with the nature and extent of the uncertainties in loss modelling, any loss modelling company opting to display the true extent of uncertainty in their loss estimates, whilst their competitors do not, is very likely to lose customers since their model will simply appear less reliable. However, choosing to not accurately model uncertainty in order to portray a greater-than-realistic confidence in a model estimate is not a sustainable approach. If the magnitude of the uncertainty does not sit comfortably with end users then greater pressure should be applied to reducing this uncertain through research into methodological development and a stronger emphasis on data quality. The principle of consistent



crudeness (Elms, 1992) should dictate the development of CAT modelling. If it is found that the crudest element in the modelling process is the exposure information, then direct incentive will be provided to insurance companies characterising their portfolios. Rather than viewing uncertainties as an impediment, stakeholders in catastrophe modelling should look upon uncertainty as an avenue via which greater understanding of the modelling process may be achieved. Hence, ideally, the user should be informed of the sources of these uncertainties, which could be readily achieved through provision of the full details of the model or, even better, the provision of the source-code. It is thus strongly believed that the development of open-source catastrophe modelling tools, carried out through a collaborative effort of all interested players (e.g., researchers, reinsurance brokers, commercial modeling companies, stakeholders), will facilitate significantly the advancement in the state-of-the-art of catastrophe risk modelling and the provision of more reliable loss estimates to individual, corporate or governmental end-users.

#### REFERENCES

BGHRC 2006. Black Box Blues. Alert 18, Benfield Greg Hazard Research Centre, University College London, p.2.

- Bommer, J.J., Spence R., and Pinho R. 2006. Earthquake Loss Estimation Models: time to open the black boxes? *Proceedings of 1st European Conference on Earthquake Engineering and Seismology*, Geneva, Switzerland, Paper no. 834.
- Boore, D.M. 1997. Estimates of average spectral amplitudes at FOAKE sites, Appendix C in An evaluation of methodology for seismic qualification of equipment, cable trays, and ducts in ALWR plants by use of experience data.
  K. K. Bandyopadhyay, D. D. Kana, R. P. Kennedy, and A. J. Schiff (Editors), U.S. Nuclear Regulatory Commission NUREG/CR-6464 and Brookhaven National Lab BNL-NUREG- 52500, C-1-C-69.
- Borzi, B., Pinho, R. and Crowley, H. 2008. Simplified pushover-based vulnerability analysis for large scale assessment of RC buildings. *Engineering Structures*, **30**:**3**, 804-820.
- Crowley, H., and Bommer, J.J. 2006. Modelling seismic hazard in earthquake loss models with spatially distributed exposure, *Bulletin of Earthquake Engineering*, **4**:**3**, 249-273.
- Crowley, H., Pinho, R. and Bommer, J.J. 2004. A probabilistic displacement-based vulnerability assessment procedure for earthquake loss estimation. *Bulletin of Earthquake Engineering*, **2**:**2**, 173-219
- Crowley, H., Bommer, J.J., Pinho, R., and Bird, J. 2005. The impact of epistemic uncertainty on an earthquake loss model. *Earthquake Engineering & Structural Dynamics*, **34**:**14**, 1653-1685.
- Crowley, H., Stafford, P.J. and Bommer, J.J. 2008a. Can earthquake loss models be validated using field observations?. *Journal of Earthquake Engineering*, in press.
- Crowley, H., Colombi, M., Borzi, B., Faravelli, M., Onida, M., Lopez, M., Polli, D., Meroni, F., Pinho, R. 2008b. Comparison of seismic risk maps for Italy. *Submitted to Bulletin of Earthquake Engineering*.
- Durukal, E., Erdik, M., Sesetyan, K., and Fahjan, Y. 2006. Building Loss Estimation for Earthquake Insurance Pricing. Proceedings of the 8th US National Conference on Earthquake Engineering, San Francisco.
- Elms, D.G. 1992. Consistent crudeness in system construction. In: Topping, B.H.V. (ed.) Optimisation and Artificial Intelligence in Civil Engineering, Dordrecht: Kluwer 1: 61-70.
- FEMA. 2003. HAZUS-MH Technical Manual, Federal Emergency Management Agency, Washington D.C.
- Grossi, P. 2000. Quantifying the uncertainty in seismic risk and loss estimation. *Proceedings of the Second EuroConference on Global Change and Catastrophe Risk Management: Earthquake Risks in Europe*, Austria.
- Molas, G.L., Anderson, R., Seneviratna, P. and Winkler, T. 2006. Uncertainty of portfolio loss estimates for large earthquakes. *Proceedings of the First European Conference of Earthquake Engineering and Seismology*, paper no. 1117.
- Molina, S. and Lindholm, C. 2005. A logic tree extension to the capacity spectrum method developed to estimate seismic risk in Oslo, Norway. *Journal of Earthquake Engineering*, **9:6**, 877–897.
- Park, J., Bazzurro, P. and Baker, J.W. 2007. Modelling spatial correlation of ground motion intensity measures for regional seismic hazard and portfolio loss estimation. *Proceedings of the 10th International Conference of Application* of Statistics and Probability in Civil Engineering (ICASP10), Tokyo, Japan, 8pp.
- Spence, R., Bommer, J.J., del Re, D., Bird, J., Aydinoğlu, N. and Tabuchi, S. 2003. Comparing loss estimation with observed damage: a study of the 1999 Kocaeli earthquake in Turkey. *Bulletin of Earthquake Engineering*, **1**:**1**, 83-113.
- Strasser, F.O., Bommer, J.J., Scedilescediletyan, K., Erdik, M., Ccedilagbrevenan Z., Irizarry, J., Goula, X., Lucantoni, A., Sabetta, F., Bal, I.E., Crowley, H., Lindholm, C. 2008. A Comparative Study of European Earthquake Loss Estimation Tools for a Scenario in Istanbul. *Journal of Earthquake Engineering*, **12:S2**, pp. 246 – 256.
- Wang, M. and Takada, T. 2005. Macrospatial correlation model of seismic ground motions. *Earthquake Spectra*, **21**:**4**, 1137-1156.