13th World Conference on Earthquake Engineering
Vancouver, B.C., Canada
August 1-6, 2004
Paper No. 2492

OBJECTIVE VALIDATION OF SEISMIC HAZARD SOURCE MODELS

R.M.W. Musson[1]

SUMMARY

In many cases, the process of conducting probabilistic seismic hazard studies is something of a black box. A seismic source model is designed, it is fed into a seismic hazard program, and a hazard curve is the output. It does not seem to be the case that there are established formal procedures for checking that the model is a reasonable reflection of reality. Aside from obvious errors such as typing mistakes in the model file, it is easy to introduce flaws into a hazard model through design decisions that seemed proper at the time, but serve to make the model unrealistic. A simple example is the forcing of an inappropriate magnitude-recurrence model onto data that can't be fit by such a model. The hazard results at the end of the process may look "reasonable", but if they come from a model that is actually incompatible with the observed data, they cannot be considered very robust. Forward simulation provides an excellent means of testing source models in an objective way. A source model can be used to construct large numbers of synthetic earthquake catalogues, which represent different possible outcomes of the seismicity over a future period. These catalogues can then be compared to the historical catalogue in terms of spatial and magnitude distribution, and various statistical tests can be run to determine if the future predictions are compatible with the historical observations. If this is not the case, the model needs to be reviewed closely to determine the source of the discrepancy. This paper provides some worked examples in the context of UK seismic hazard studies.

INTRODUCTION

Considering the potential impact of seismic hazard studies on design costs, it is not unreasonable to raise the issue of validation of the results of such studies. Often seismic source models are highly opaque. A complex decision-making process results ultimately in a hazard curve, but what comes in between can

[1] British Geological Survey, West Mains Road, Edinburgh EH9 3LA, UK

often be obscure. The standard "checking" procedure involves a peer review team, who consider the subjective opinions (or expert judgments) of the original analysts and interpret them in the light of their own subjective opinions. The problem is, that it is possible to take decisions in seismic hazard analysis which, considered purely in their own terms, seem reasonable, but which have hidden implications that are really undesirable. As long as these decisions are evaluated purely in terms of the processes by which they were arrived at, and not in terms of their impact on the model, their undesirable nature will not show. Conventional sensitivity analyses will not suffice here. A sensitivity analysis will show the impact of a decision on the *results* but not on the model, and the results, in the sense of a hazard curve, are a highly derived product of a model.

What is needed is a set of procedures to make objective tests of the realism of the seismic model. The model is a complete description of regional seismicity in aggregate terms, allowing for uncertainty. The seismic hazard results show the probability of different levels of ground motion at site, given that the seismicity during the lifetime of the structure will accord with the general description given by the model. For the hazard results to be considered valid, this description must be physically realistic. As a test of this realism, one can compare the predictions of the model against historical experience.

The objection that might be raised against this, in principle, is that historical experience is usually short relative to the seismic cycle, and if one were to become fixated on the replication of recent seismicity, the model might be inadequate in accounting for future seismicity, the pattern of which will be controlled by long-term rates. Nevertheless, it is the case that the historical seismicity was also a result of the same long-term rates. The seismic source model seeks to describe those rates, including uncertainty. It ought to be possible to use the model to reconstruct a reasonable approximation of the historical seismicity. If this is not possible, if the historical earthquake record and the model are actually incompatible, then one cannot trust that what the model has to say about future hazard levels will be reliable.

SIMULATION AS A TOOL FOR HAZARD EVALUATION

The tool that one can use here is Monte Carlo simulation. This has been identified as a means to seismic hazard evaluation as long ago as ago as 1983, Rosenhauer [1], Shapira [2]. Early use was limited by available computer power, and it was in the 1990s that routine use became more practical, Sólnes et al [3], Musson [4], Musson [5]. The use of Monte Carlo simulation (or stochastic modelling, to give at an alternative name) to compute hazard involves taking a standard seismic source model and using it to generate a large number of synthetic earthquake catalogues representing possible future outcomes of regional seismicity in a period representing the lifetime of the structure being designed. Each synthetic event is attenuated to site (with a scatter applied to mimic the natural scatter of strong ground motion attenuation), and one then obtains a set of n annual ground motion maxima for the site, all possible outcomes of the model, where n is a very large number, typically 10,000,000. To find the ground motion with annual probability of 10^{-4} of being exceeded, one need only sort the outcomes and pick the value that is exceeded 1,000 times, Musson [5]. This procedure is completely compatible with conventional PSHA in the manner derived from Cornell [6] and McGuire [7]; the same seismic input model will give identical results from both methods, even though they are totally different in approach.

One of the advantages of the simulation approach to hazard is that it is inherently more transparent. One can collect statistics on the synthetic catalogues derived from the model and use these to investigate how the hazard is being derived. One obvious application of this is in the study of design earthquakes, Musson [8], Musson [9], where one can analyse the population of earthquakes that cause the design ground motion, as an alternative to the analytical disaggregation procedure of McGuire [10]. But another application is in testing the model. A typical seismic source model manifests itself as many pages of abstruse numbers. Checking that the numbers have been entered correctly is merely a tedious mechanical chore. Checking that they really do make seismological sense is another matter, notwithstanding the best intentions of the model's authors. Modelling decisions made in good faith sometimes turn out to have unintended consequences, which may not be visible from the presentation of the model, nor from the results – as long as the final hazard curve looks approximately right, no suspicions are raised. However, inspection of the synthetic catalogues that result from the model provides a good way to examine the supposed physical basis on which the hazard calculations actually rest.

Two levels of checking can be done. The simplest consists of simply taking a small number of synthetic earthquake catalogues that have a match to the historical catalogue in terms of completeness, and looking at them. Some problems with the model may be immediately evident, even with a small sample (say, about ten catalogues). For more robust statistical analysis, a sample of 100 catalogues is sufficient according to Diggle [11]; a sample of 1,000 was used in a study by Musson and Winter [12].

In the following examples, the seismicity of the UK is used, based largely on Musson [13]. Rather than castigate any actual studies, made-up simple source models are used for illustrative purposes only. However, the problems discussed have all been encountered "in the wild". Note also that there are some potential problems that can't be detected in this way, so the methods proposed here are not a complete solution.

SOME ISSUES IN SEISMIC SOURCE MODELLING

Zone model design

Anyone involved in PSHA studies will be well familiar with the common lack of agreement on the appropriate size for seismic source zones. To what extent should one continue to divide up a region into smaller and smaller zones? Or should they be aggregated into as large zones as possible? In intraplate areas, the hypothesis is sometimes advanced that a large area (a whole country) may plausibly be regarded as having random seismicity, with any apparent local clustering being by chance. In Figure 1a, the actual seismicity of the UK is shown; all events > 4 ML since 1800 are plotted, this criterion giving reasonable completeness for the whole land area. Figure 1b shows a single source zone for the land mass of Great Britain, and a synthetic catalogue that might result for 200 years of data > 4 ML. One only has to look at a few instances of Figure 1b, which shows true random seismicity, in comparison to Figure 1a, to conclude that there is a significant difference.

There are also more formal ways of demonstrating this. A simple one is to take the zone boundary from Figure 1b, use it to select the seismicity contained within it in Figure 1a (the historical data), and test whether the epicentres are evenly distributed within it by applying nearest-neighbour statistics. As is shown by Musson [14], this demonstrates conclusively that the seismicity within this large zone is clustered; use of such a zone in PSHA would therefore violate the principle that seismicity has an equal probability of appearing anywhere within each zone.

One can also perform a statistical comparison of the spatial distributions of simulated and real seismicity, Musson and Winter [12]. The basis of this method is the division of the study area into a grid of n x n cells. The number of events in each cell in the historical data set represents a target that the synthetic catalogues try to match. Over a large series of simulations, a X^2 test is used to evaluate whether or not the population of simulations and the real data set are compatible.

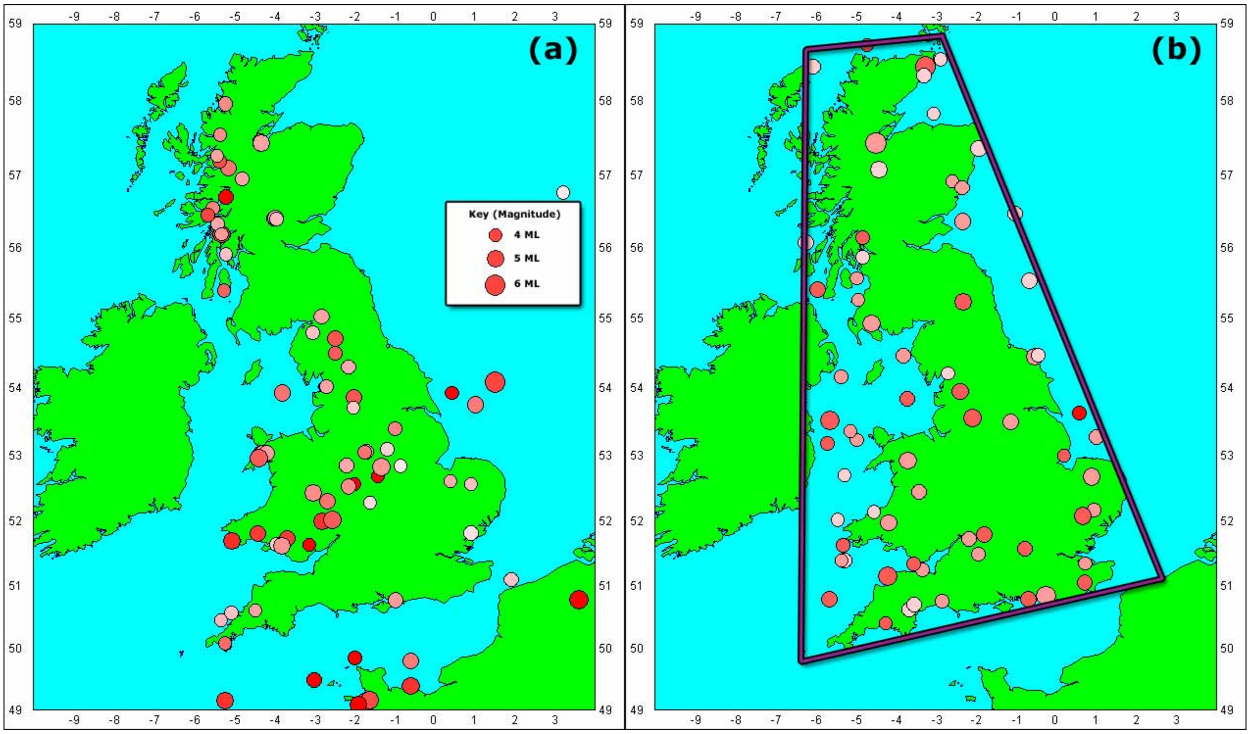

Figure 1: (a) Historical seismicity since 1800, magnitude 4 ML and above; (b) Sample 200 years of synthetic data > 4 ML on the assumption that seismicity occurs randomly within the single source zone shown. In both cases, symbol size is proportional to magnitude, colour to depth, lighter events being shallower.

This method can also be used, not just to evaluate whether a source zone model is insufficiently fine, but also on wider issues. For example, one might hypothesise that geological terrane boundaries are fundamental seismotectonic divides in the British Isles, and construct a source zone model in which each zone was mapped to a specific terrane. The synthetic catalogues so generated could be compared to the historical data using this form of analysis as a check on whether this hypothesis is viable. If the X^2 test shows that statistically, the simulations and historical data set are incompatible, then the hypothesis has probably to be rejected, whatever the geological arguments in its favour. This method forms a quantitative and statistical counterpart to the simple procedure of plotting up a sample of synthetic catalogues and evaluating them visually. In the case of post hoc evaluation of an existing study, the statistical approach is really necessary; but as a tool in the development of a working model, the visual check is useful and usually sufficient.

A related issue concerns the shape of seismic source zones. It has often been the case, especially in earlier studies using relatively unsophisticated software, that analysts would tend to draw source zones that are more or less rectangular, often aligned N-S. In this way one can produce source zone maps that look pleasantly neat. What is not so pleasant is to see the maps of the synthetic catalogues produced by such models. When one sees seismicity maps in which epicentres evenly fill rectangular boxes, it becomes clear that such models produce unrealistic catalogues, as shown by Musson [14]. And since the hazard computed from such models is determined by the future seismicity being one of those unrealistic catalogues, it follows that the hazard is also unrealistic.

There is, of course, an exception to this, and the exception is: when it doesn't matter. In the case of Figure 1a, if one imagines a hypothetical case of a hazard study for a site in the west of Ireland, such a site is so far away from the centres of earthquake activity that it makes little difference what source geometry is used. One could design a detailed tessellation of zones, or three rectangles, or just the one zone of Figure

1b, and probably the hazard result would be just the same; really, the only requirement in the case of distant seismicity is to have roughly the right number of earthquakes at roughly the right distance.

In other cases, inspection of synthetic catalogues focuses attention on what the model means in terms of physical seismicity, and if this is unrealistic, it should prompt checks on the influence of the zone design on the hazard.

While such analyses can identify problems with zones that are too big, or are badly shaped, it cannot identify zones that are too small. It is a methodological question whether zones can actually *be* too small, as the various gridding methods used sometimes to replace zonation, starting with the work of Jacob et al [15], are equivalent to having a large number of very small zones evenly distributed over the study area.

Earthquake magnitude-frequency

A source model contains four types of information about seismic sources. The first is the geometry, already discussed. Second is the depth distribution for each source. This is usually very poorly constrained, which makes evaluation of depth distributions difficult. In the case of high-seismicity areas it may be irrelevant anyway. The third is the occurrence model, i.e. is the seismicity Poissonian in behaviour or does it follow some other pattern? Since this question is almost invariably ignored in PSHA, it will be ignored here too. This leaves the recurrence rates, which will now be considered. At a minimum, four pieces of information are needed: the form of the recurrence relation (linear, double-truncated exponential, characteristic, etc), the slope or *b* value (the rate at which large earthquakes occur relative to small earthquakes), the activity rate (how many earthquakes occur per year above magnitude M_0), and the maximum magnitude (M_{max}, the upper bound to the seismicity distribution). These four parameters control how the zones in the model are populated with earthquakes. It is particularly here that it is easy to take modelling decisions that lead to unrealistic results. Consequently, formal evaluation methods provide a very useful safeguard. (Note: no account here is taken of minimum magnitude, despite the fact that this parameter has a large impact on hazard, because it is an arbitrary, conventional value.)

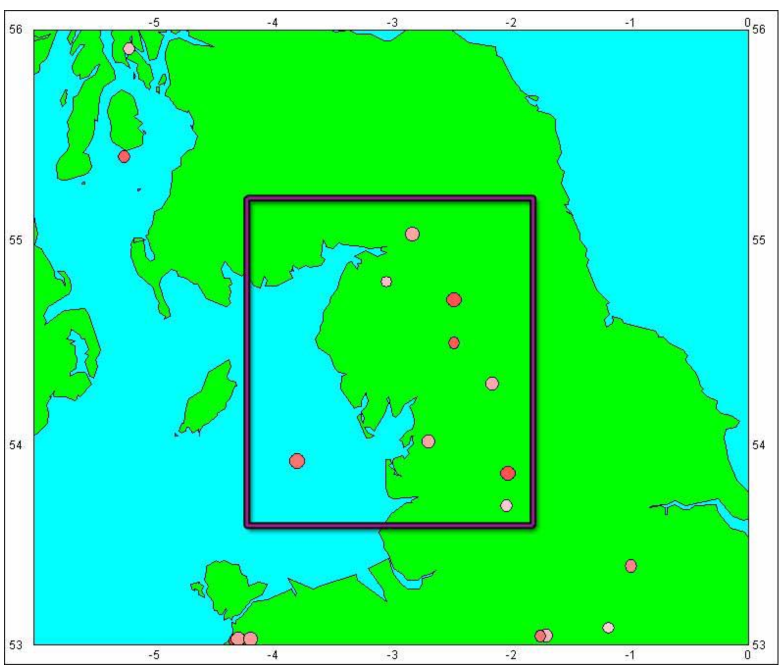

Figure 2: Hypothetical source zone for N England, with historical seismicity since 1800.

Consider the case shown in Figure 2. Here a hypothetical analyst interested in the North of England has decided on tectonic grounds that the seismicity of the Pennines and Irish Sea are related, and has delineated a single zone for all the North of England seismicity. This has been drawn as a nice neat rectangle.

The zone contains nine earthquakes. The analyst reasons that this is insufficient to calculate a *b* value from, so he elects to use a regional *b* value, which according to Musson [13] is -1.03. He will obtain the activity rate by fitting this value to the data within the zone. In the normal course of events, the appropriate hazard values would be obtained for the site, and that would be that. Here, we start by looking at a sample catalogue, as shown in Figure 3.

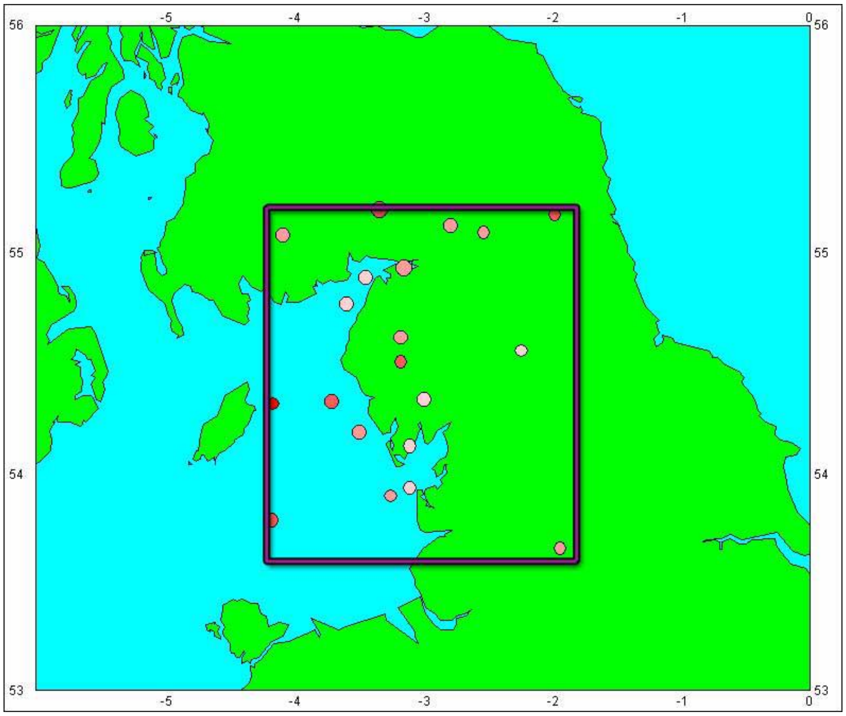

Figure 3: Hypothetical source zone for N England, with sample synthetic seismicity.

This doesn't look very much like Figure 2, but the differences are not conclusive. The X^2 test can't be used because there are insufficient earthquakes; if one dropped the magnitude threshold it would yield more events to work with, but completeness issues would intrude. (If Figure 3 were plotted down to magnitude 3, it would also become much more apparent how boxy the seismicity is due to the rectangular zone; one can begin to see this in Figure 3 as it is.) A further test is to compare the distribution of parameters of the synthetic catalogues with the same parameters of the historical data. Here, two such parameters are used: the mean magnitude and the number of events (in 200 years > 4 ML). Mean magnitude is simply the mean of the magnitudes of all the events in a catalogue; it is related to the *b* value, but is uniquely defined for a given data set, whereas *b* is not. Figure 4 shows the distribution of the means of 1,000 synthetic catalogues. The true (historical) mean, within the same catalogue constraints, is 4.5. If the true mean were in the tail of the distribution of the sample means, this would indicate a problem. In fact, as Figure 4 shows, it is acceptably central, albeit with a tendency for the samples (synthetic catalogues) to produce lower mean values, indicating a higher absolute value of *b* and a slight unconservatism.

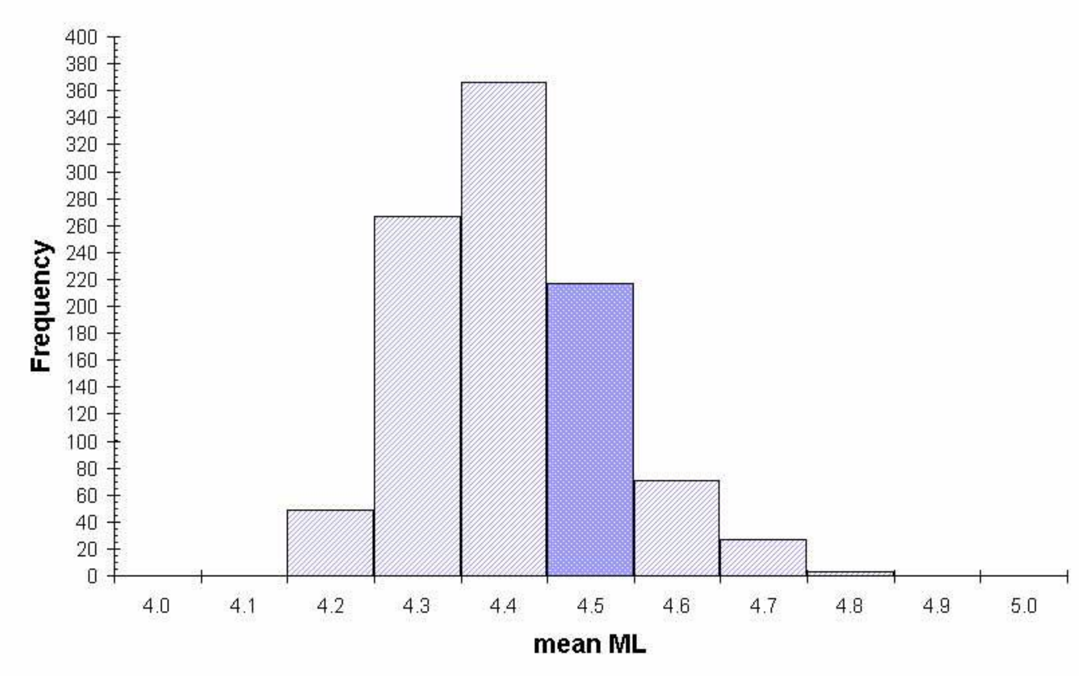

Figure 4: Distribution of mean magnitudes from 1,000 synthetic catalogues. The column containing the corresponding historical value is shaded more darkly.

Figure 5 shows the same thing for number of events, and here the story is rather different. The mean number of events per catalogue in the synthetics is 15 ± 4.0, compared to the historical value of 9. This is not quite in the tail of the distribution, but close to it: 925 out of 1,000 samples had more events than the historical catalogue.

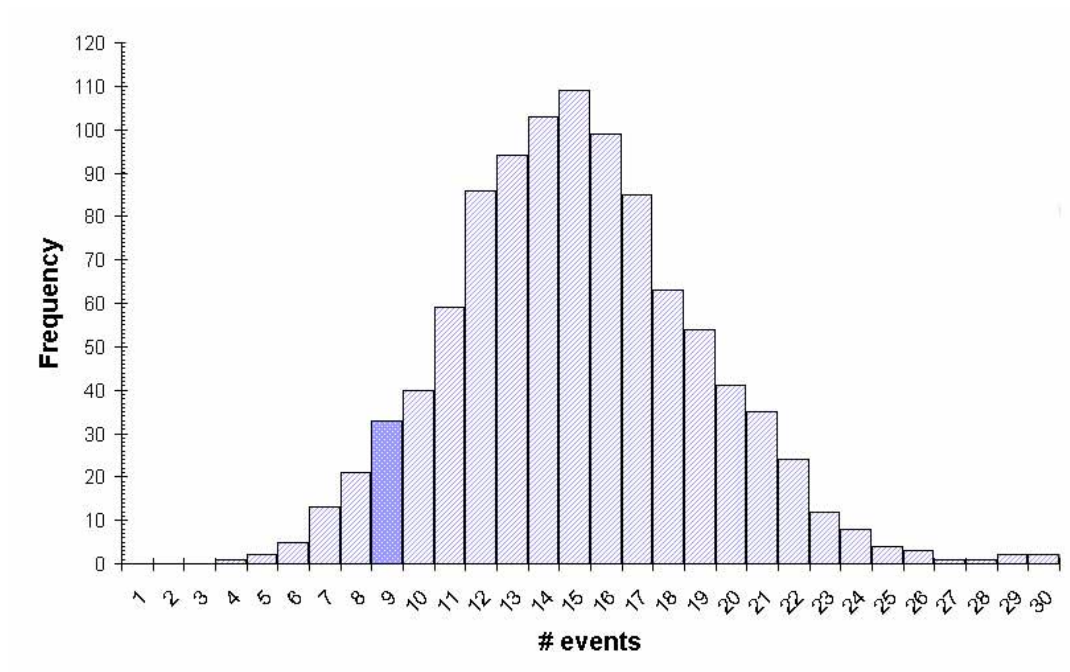

Figure 5: Distribution of number of event totals from 1,000 synthetic catalogues. The column containing the corresponding historical value is shaded more darkly.

Figure 5, therefore, shows substantial conservatism, while Figure 4 shows slight unconservatism. This opposition, combined with the eccentricity of the historical value in the distribution shown in Figure 5, should prompt a careful re-examination of how the magnitude-frequency curve fitting was done for the zone. What happened was, a method for curve fitting was defined in advance, and applied mechanically to this zone as it would have been to other zones, assuming Figure 2 to be just part of a larger model. This is common procedure. The method in question (fitting a predetermined *b* value to the median magnitude) would work well enough in other zones, but it doesn't work well in this one. (The reader may object that the method used in this example for deriving magnitude-frequency parameters is not a good one. This may well be true, but the object of this paper is to discuss ways of evaluating PSHA studies that may not always use the most up-to-date procedures.)

If one wished to defend the model in the face of Figure 5, one argument that could be used is that the estimation of completeness for the historical data set may not be accurate, and that if one had knowledge of the missing events that have escaped the historians, the number of events in the truly complete data set would be closer to the middle of the distribution in Figure 5. This immediately raises the interesting but difficult question of proving a historical negative, Musson [16]. From a consideration of the historical "footprint" of an event > 4 ML in 1800, and a familiarity with the source materials available for this period, one would probably conclude that, at most, one could believe one missing event might exist, but even this is doubtful. An interesting study by Stucchi and Albini [17] considers from a historian's perspective the number of large earthquakes in parts of Italy that would have to be missing from the historical record to reconcile the Italian earthquake catalogue with the magnitude-frequency implications of the seismic source model used for hazard calculations by Slejko et al [18]. Their conclusion, that the number of implied missing earthquakes was incompatible with the quality of the historical source materials, prompted a re-examination of the procedures used to populate the hazard model.

CONCLUSIONS

In papers designed to criticise the PSHA method in general, e.g. Krinitzsky [19], it is common practice to hold up as a warning the study of Bernreuter et al [20], and argue that because different experts in that study came up with wildly different seismic models for the Eastern and Central USA, the whole methodology must be flawed. In truth, just because a procedure is sometimes badly executed, it does not follow that the procedure itself must be at fault. Faced with conflicting interpretations of seismic sources, the response should not be to abandon the method, but to look for ways of discriminating better source models from poorer source models. The present paper, following on from Musson [14] and Musson and Winter [12], demonstrates some approaches that can be used for this important task.

In the past, the evaluation of models by peer review has tended to concentrate on a critique from scientific principles of the data, assumptions and arguments involved in the decision-making process. It might be thought that if the modelling decisions are defensible (i.e. one can advance a reasonable argument for their justification) then the results must be realistic. This is optimistic. In practice, "reasonable" arguments can sometimes lead to models that are questionable or even incompatible with reality. This can be hard to spot so long as hazard results are derived directly from a numerical model by processes concealed within the black box of the hazard software. The actual hazard to a structure comes not from an abstracted model, but from real, physical earthquakes. The model is a description of the seismicity, but what earthquakes does it actually describe? The use of synthetic catalogues provides a way to access the "missing link" between the numerical model and the final hazard curve.

Even simple inspection of such catalogues provides a quick check against gross errors or highly unrealistic assumptions. Statistical analysis comparing synthetic against historical data provides a way to expose poor modelling decisions, and ultimately, may assist to distinguish between competing models.

REFERENCES

1. Rosenhauer, W. "Methodological aspects encountered in the Lower Rhine area seismic hazard analysis". Ritsema, A.R. and Gürpinar, A. Editors. Seismicity and seismic risk in the offshore North Sea area. Dordrecht: Reidel 1983, 385-396.
2. Shapira, A. "Potential earthquake risk estimations by application of a simulation process". Tectonophysics 1983; 95: 75-89.
3. Sólnes, J., Sigbjörnsson, R. and Halldórsson, B. "Assessment of earthquake hazard and seismic risk based on synthetic and upgraded earthquake catalogues of Iceland". Proc. 9th Int. Sem. on Eq. Prognostics, San José, Costa Rica, Sept 1994.
4. Musson, R.M.W. "On the use of Monte Carlo simulations for seismic hazard assessment". Proc. 6th US National Conf. on Eq. Eng., 1998.
5. Musson, R.M.W. "The use of Monte Carlo simulations for seismic hazard assessment in the UK". Annali di Geofisica 2000; 43: 1-9.
6. Cornell, C.A. "Engineering Seismic Risk Analysis". Bull. Seis. Soc. Am. 1968; 58 (5): 1583-1606.
7. McGuire, R.K. "FORTRAN computer program for seismic risk analysis". USGS Open File Report 76-67, 1976
8. Musson, R.M.W. "Determination of design earthquakes in seismic hazard analysis through Monte Carlo simulation". Jnl. Eq. Eng. 1999; 3: 463-474.
9. Musson, R.M.W. "Design earthquakes in the UK". In preparation, 2004.
10. McGuire, R.K. "Probabilistic seismic hazard analysis and design earthquakes: closing the loop". Bull. Seism. Soc. Am 1995; 85 (5): 1275-1284.
11. Diggle, P.J. "Statistical analysis of spatial point patterns". New York: Academic Press, 1983.
12. Musson, R.M.W. and Winter, P.W. "Application of formal testing methods to models for site-specific seismic hazard studies in the UK: The case of Sellafield". BGS Technical Report No. CR/02/207, 2003
13. Musson, R.M.W. "A catalogue of British earthquakes". BGS Global Seismology Report No WL/94/04, 1994.
14. Musson, R.M.W. "Evaluation of seismic hazard source models". Lapajne, J. and Vidrih, R. Editors. Seismicity modelling in seismic hazard mapping, Ljubljana: Slovenian Geophysical Survey, 2000: 53-66.
15. Jacob, K., Armbruster, J., Barstow, N. and Horton, S. "Probabilistic ground motion estimates for New York: Comparison with design ground motions in national and local codes". Proc. 5th US Nat. Conf. on Eq. Eng., 1994; 3: 119-128.
16. Musson, R.M.W. "The significance of absent earthquakes: What is the weight of a historian's opinion?". Eos Trans. AGU, 2002, 83 (47), Fall Meet. Suppl.: F1062.

17. Stucchi, M. and Albini, P. “Quanti terremoti distruttivi abbiamo perso nell’ultimo millenio? Spunti per la definizione di un approccio storico alla valutazione della completezza”. Galadini, F., Meletti, C. and Rebez, A. Editors. Richerche del GNDT nel campo della pericolosita sismica (1996-1999), Rome: CNR-GNDT, 2000: 333-343.

18. Slejko, D., Peruzza, L. and Rebez, A. “Seismic hazard maps of Italy”. Annali di Geofisica 1998; 41 (2): 183-214.

19. Krinitzsky, E.L. “Earthquake probability in engineering – Part 1: The use and misuse of expert opinion”. Eng. Geol. 1993; 33, 257-288.

20. Bernreuter, D.L., Savy, J.B., Mensing, R.W. and Chen, J.C. “Seismic hazard characterisation of 69 nuclear power plant sites east of the Rocky Mountains”. US Nuc. Reg. Comm. Report NUREG/CR-5250, 1989.