Objective assessment of source models for seismic hazard studies

R.M.W. Musson,[a] **M.EERI, P.W. Winter,**[b]

Up to now, the search for increased reliability in probabilistic seismic hazard analysis (PSHA) has concentrated on ways of assessing expert opinion and subjective judgement. Although in some areas of PSHA subjective opinion is unavoidable, there is an all too present danger that assessment procedures and review methods simply pile up further subjective judgements on top of those already elicited. Such reviews are in danger of assessing only the form and neglecting the content. The time has come to find ways to demonstrate objectively, where possible, if interpretations are valid or not. This can be done by studying what these interpretations physically mean in terms of seismicity. Experience shows that well-meaning but flawed design decisions can lead to source models that are actually incompatible with seismic history. One such method is as follows: from a seismic source model one can generate large numbers of synthetic earthquake catalogues that match the completeness thresholds of the historical catalogue. The question is then posed, is the historical earthquake catalogue a credible member of the set of all possible catalogues derived from the model? If the answer to this is no, and this can be determined statistically, then one can reject, with a specified confidence level, the hypothesis that the model is a valid depiction of the long-term seismicity rates that will govern the future hazard.

[a] BGS, West Mains Road, Edinburgh EH9 3LA, UK;
[b] ESR Technology, Whittle House, Risley WA3 6FW, UK.

INTRODUCTION

The practice of probabilistic seismic hazard analysis (PSHA) in the past has often tended to the arcane and is seldom fully transparent. Since the majority of PSHA studies are conducted commercially for engineering design purposes, rather than in the open spaces of the scientific literature, many significant methodological issues are seldom aired in public debate (Musson 2004). As a simple example, both in commercial hazard reports and published studies the decision made as to the minimum magnitude used in the hazard calculations is usually announced baldly without discussion or justification – or is not mentioned at all. Yet by arbitrarily changing this parameter one can substantially alter the hazard results. Hazard software mostly resembles a black box; one has input (a seismic source model) and output (a hazard curve), but what goes on in between is usually obscure and taken on trust. The danger is that decisions are made in the modelling process that are well-meaning, and can easily be justified in writing, but which have unforeseen consequences when it comes to the hazard. It can be hard to spot when this occurs in the normal run of things. If the hazard curve looks about right, it is easy to assume that all is well, because the processes within the hazard software that convert input to output are not deconstructed.

As an example, a hazard map for Italy (Slejko et al 1998) suffered from a decision made about completeness rates in some source zones, which, on the face of it, seemed a reasonable and rational decision. It was only later when subsequent analysis was made of some of the implications of this decision for the Italian earthquake catalogue (Stucchi and Albini 2000) that it was found that the method used for assessing completeness was flawed in such a way that it increased the hazard substantially in some parts of the map. The method predicated an inflated number of large earthquakes in certain zones, which raised the hazard, though this was not apparent just from reviewing the original study and its results in the normal way. In fact, this problem was also detected by a study by Mucciarelli et al (2000) who found significant discrepancies between the hazard map values and those derived by analysis of historical experience of certain towns. However, Mucciarelli et al (2000) declined to draw the correct conclusion, that there was actually something wrong with the source model used for the probabilistic calculations.

The difficulty is that problems of this kind can be hard to spot in the course of the normal peer review process, hence the need for approaches such as those of Mucciarelli et al (2000) and Stucchi and Albini (2000). In the course of a PSHA study one has to deal with many

highly uncertain issues, and various procedures have been devised for dealing with the elicitation of expert opinion (Budnitz et al 1997). The more experts are polled, the more one glimpses the size and range of the uncertainty, and the use of multiple experts can ensure that the spectrum of opinion in the broader community is fully sampled. However, when it comes down to drawing zones on a map, or, in the case of non-zoned methods, defining kernel shapes and sizes, an expert or team has finally to formulate their subjective judgement into a set of numbers that will be fed to the hazard software.

The process can be summed up by the diagram in Figure 1. The team conducting the study has access to some relevant data; they have some beliefs of their own and they have listened (hopefully) to the beliefs of others. They make some interpretations, and out of these three, beliefs data and interpretations, the source model is created. It is then fed to the hazard software. How the hazard software works may or may not be fully understood by the team, depending often on whether they wrote it or not, but either way, the internal processes of the software are usually not transparent. The results can be interrogated to some extent by sensitivity analyses and disaggregation. This can tell you which elements made the greater contribution to the hazard results, but not necessarily why this was the case. It will not tell you, for instance, if one zone has anomalously high activity because of an assumption that was made in computing the activity rates for that zone.

When it comes to the review process, at least judging from personal experience, peer review tends to focus on what happens above the dashed line in Figure 1. The review team will measure the beliefs and interpretations of the hazard team against their own beliefs and interpretations. They may vet the range of procedures undertaken by the hazard team and suggest extra work. In the process, a new layer of subjectivity is added to the study. But who reviews the reviewers? Cases where gross errors in hazard studies have slipped past a peer review team are not unknown (but obviously, cannot be cited). What tends not to be examined in detail is the part below the line in Figure 1, which is where the model is turned into results. The critical question is, what is the physical meaning of the model, and how does this compare to reality?

To return to Stucchi and Albini (2000), the specific question asked there was, if the activity rates are in reality as depicted in the model, how many large earthquakes should have occurred in the historical period compared to the number that are recorded? If the number of

predicted events is very much higher than the number observed, is this credible taking into account what is known about levels of historical documentation in Italy?

This is one instance of a general class of question, the aim of which is to show that the output of a seismic hazard study is compatible with the input. Establishing this goal is actually quite modest. There are many facets of a hazard study that can't be tested in this way, such as decisions made about ground motion models, or the arbitrary issue of minimum magnitude already referred to. Still, making these checks between input and output increases the reliability and robustness of any study, which is something to be desired, not least by the client who has to implement the results.

EVALUATING ZONE MODELS

There are two things that such tests can trap: misguided decision-making and gross errors. The latter category includes such things as critical typing errors in the input file. Such things should not happen, but occasionally they do.

Poor decision-making is a more common problem. A simple example is any case where the hazard analyst decides that all seismicity follows a linear Gutenberg-Richter relationship even where sometimes it doesn't.

The aspects of PSHA that are best suited to testing are the spatial interpretation of seismicity and the activity rates. The former is more commonly a zone model, but may be a spatially smoothed model, or even a combination of the two. Spatially smoothed models do not free the analyst from the burden of decision-making; just as there are an infinite number of possible zone models that could be drawn for an area, so too there are an infinite number of combinations of kernel shape/size, isotropy/anisotropy, uniform/non-uniform smoothing, etc. If one were to be cynical, one could say that the real advantage of the spatially smoothed methods is that the subjectivity of the decisions is better concealed. For the rest of this paper the discussion will concentrate mostly on zone models.

Zone models for PSHA have in the past provided an easy target. On the one hand, they have been attacked by proponents of deterministic hazard (e.g. Krinitszky 1993) and on the other hand by proponents of spatial smoothing (e.g. Woo 1996). The method employed is usually the same: some specific model is held up for criticism with the implication that, hence or otherwise, all models are equally bad. The usual straw man is the study of Bernreuter et al (1989), in which a large number of different zonations of Central/Eastern

USA were obtained from different experts, which disagree wildly from one another. The argument then runs, "if these models are all equally valid, when they are totally incompatible with one another, then surely the whole process, and the end results, is both meaningless and valueless?" (our paraphrase).

In fact, it is logically invalid to argue that because some study applies a methodology badly, therefore the methodology is bad. It is like arguing that because some people drive badly therefore cars are badly designed. To counter this argument, it is necessary to be brutally honest and admit that not all source models are equally valid. Just because it is someone's interpretation doesn't mean that it must be admissible. Some models are in fact worthless, and it's time that this fact was admitted, and procedures devised for sorting out the bad ones from the good ones.

One can discern a basic principle behind all PSHA studies that there exist some controlling parameters behind the long-term seismicity of any area. Seismicity is not totally chaotic; some places are persistently more active than others, and average rates over a long enough period do not fluctuate wildly. The processes of PSHA are therefore in a large part ways of estimating those controlling parameters on the basis of what has gone before, and then using the estimates to find the likelihood of strong earthquakes in the immediate future. The hazard values are valid only so much as these estimates are credible.

Since the parameters that control the future earthquakes around a site are the same as those that controlled the earthquakes of the last (say) 200 years, then it should be the case that the historical seismicity is compatible with those parameters. This is something that can be tested. If the model is valid, then the historical earthquake catalogue must be a possible outcome of the model. If one can show that the historical catalogue could never have resulted from the model, then probably the future seismicity won't follow the model either. In that case, the hazard predicted by the model is not in any way reliable.

Hence, one needs ways of testing whether the historical catalogue could have resulted from the model. This is perfectly feasible.

WORKED EXAMPLE

To demonstrate this, a worked example is presented, taking the UK as a case in point. Two possible zone models are shown in Figure 2. The first model (Figure 2a) was created especially for this study. It is intended to be an example of misguided decision-making; it is a

model behind which there is a consistent interpretation; it just so happens that this is deliberately not a very good one. The assumption made is that the key factor in partitioning UK seismicity is the gross crustal structure; therefore, each major geological terrane (following Pharaoh et al 1996, with some trimming and simplification) is a seismic source zone. It is not intended for a moment that anyone should take this as a serious hypothesis; nevertheless, it is just credible that if someone should arrive at this idea or something similar, they could write up a sufficiently eloquent defence of it that a reviewer could be persuaded to acknowledge that it might be justified. This will be referred henceforth as the terrane model.

The second model (Figure 2b) is the zonation that was used in the mapping projects GSHAP (Grünthal et al 1996) and SESAME (Jiménez et al 2001). This will be referred to as the GSHAP model. The zonation was slightly simplified for the purposes of this study. The GSHAP zone model covers a slightly different area than the terrane model; there is a background zone that extends over Ireland, and zones that overlapped France and Belgium were not included. This is not significant for the purposes of the present study.

The object of the exercise, therefore, is to demonstrate objectively that the GSHAP model is better than the terrane model. It will also be a good result if it can be shown objectively that the terrane model can be completely rejected as a valid model (and correspondingly, that this is not the case for the GSHAP model).

Activity rates for both models were assessed in the same way from the UK earthquake catalogue (Musson 1994 with later additions). An extract from the catalogue was prepared that is considered to be a complete data set – this is the 200 years of data following 1800, counting magnitude 4 ML and above. This should be complete or near complete for the land area of the UK (Musson and Winter 1996). This historical data file is the "ground truth" for this study. We need to determine if the historical file could have resulted from either of the models.

Since both models describe fully supposed long-term seismicity parameters for the UK, they can be used to generate synthetic catalogues by applying a Monte Carlo process (Musson 2000). Figure 3 shows some preliminary results. At the left is the historical set of events (200 years ML $\geq$ 4). The top row shows four maps of synthetic catalogues from the Terrane model (also 200 years ML $\geq$ 4) and the bottom row shows the first four simulations prepared using the GSHAP model. Could the map on the left be considered a member of the set of maps in the top row? In the bottom row? Visual inspection suggests the maps in the

bottom row are quite realistic, while those in the top are not. However, one would like more than visual inspection to go on. Note that in the bottom row, the Channel Islands area is blank because this area was not included in the model.

In both cases, 1000 catalogues were now prepared. For statistical analysis, it is necessary to compare like number of events. As the historical file contains 71 earthquakes, the synthetic catalogues were constructed to be of whatever length necessary to generate 71 events. (This has implications that will be returned to later.) This is different from the situation in Figure 3, where the synthetic catalogues were all set at 200 years.

The area was now divided into a 5x5 grid. The area covered was 49 to 59 N, 8 W to 2 E for the Terrane model; and 50 to 59 N, 8 W to 2 E for the GSHAP model. For every cell, the number of events in each catalogue was counted; cells with fewer than five historical events were not used in the analysis.

The match of cell counts in the historical file was then compared to the cell counts in the 1000 simulated catalogues, and the X2 statistic computed to evaluate the hypothesis that there is no significant difference between them. The value obtained for the terrane model was 29.74, indicating that the null hypothesis can be rejected with 99.5% confidence. Therefore one can also reject the terrane model as an adequate representation of seismic processes in Britain. If it were true, one could not have obtained the historical result; except perhaps as a freak occurrence.

The equivalent value for the GSHAP model was 1.73. For this model, one cannot reject the null hypothesis, and this model is therefore a viable depiction of seismicity patterns in the UK.

A further test was made by taking the terrane model and deriving the 1001st simulated catalogue. This catalogue was then compared to the previous 1000 simulations, in the certain knowledge that it was produced by the same model. Because the correct answer is known in advance (that the 1001st simulation is compatible with the other 1000) this provides a check on the method. The X^2 statistic in this case was 4.94 – one cannot reject the null hypothesis.

It could be objected that it is possible to fool this test: if one deliberately created a model that mimicked the historical result very closely – say by taking the historical catalogue as a series of point sources with a very small amount of scatter – then one could be sure of a good result from the test. However, this does not mean that the model would be a good one if it

just recapitulates the historical result over and over again; the chances are that seismicity over the next few hundred years will not be an exact recapitulation of the historical catalogue.

The answer is that this test, and indeed, any test, has to be applied sensibly and not as a blindly applied procedure. If a model is far too precise and lacks the degree of generalisation necessary for an adequate coverage of possible future seismicity, this should be self-evident and can be addressed in other ways. If one thinks of three possible problems, over-generalisation (zones are too large), mis-generalisation (zones are in the wrong place) and under-generalisation (zones are too small), then this test will trap the first two problems but not the third. However, in our experience, the first two problems are far more common in actual practice.

It was pointed out by Abrahamson (2004 pers. comm.) that this has useful implications in the case of smoothed seismicity models; one can detect the largest smoothing radius that is compatible with the historical data (though not the smallest).

A poor result from this test, leading to rejection of a model, can result in two ways. Firstly, if the zones are drawn inappropriately; secondly, if the activity rates are poorly derived. Both problems may be present.

Activity rates can be examined using another procedure. For any catalogue, one can compute the number of events > Mmin within a known completeness margin, and for the same set of earthquakes, the mean magnitude. For a series of catalogues derived from the same model, one will obtain a range of values, which will form a distribution. One can consider this as a 2D probability distribution. In addition, there is the historical result, which gives a point to be compared to this distribution. If the historical result appears as an extreme outlier, this suggests that the activity rates in the model are not truly realistic. If the historical result falls comfortably (a rigid definition of this will not be proposed here) within the distribution formed by the simulations, then one cannot reject the model.

This is shown in Figure 4. Here the two test models have been used to generate 1000 simulated catalogues exactly 200 years long (Mmin = 4). A series of bins was used to grid the data in a parameter space formed by the average number of events and the mean magnitude, and the distribution was contoured according to the number of simulations that generated any given pairing of number of events and mean magnitude. The historical result for the period 1800-2000 is plotted on each contour diagram as a star. In the case of the terrane model, the

historical result actually lies outside the distribution, showing that, whatever problems may exist with the zone boundaries, the overall activity rates are not well determined. In the case of the GSHAP model, although the historical result is not right at the centre of the distribution, it is sufficiently well within the distribution that one would not reject it. One can also notice from the distribution that the GSHAP model is slightly conservative as regards activity rates, and inclines towards steeper b-values (the centre of the distribution has a lower mean magnitude value than the historical result).

DISCUSSION

In the case of a real study in which the terrane model was somehow proposed, the previous tests could be used in an iterative process to evaluate and improve the model. Having discovered from the first test that the model as first designed is incompatible with past seismicity, one would then find from the second test that the activity rates are too high. This would lead to a revision of the activity rates. If the amended model still gives a high X2 statistic (as it would in this case) one would then focus on the geometry. Finally one would arrive at an improved model, and hence, more robust hazard estimates. The same analytical process would not be possible if all one had to depend on was the opinions of a peer review team.

The applicability of the X2 statistic test depends to some extent on the density of seismicity within a given region. It is possible to apply it, as shown here, to a hazard map model covering the whole of Britain. For a site hazard study where the model covered only a small area (say, 200 km radius around the site) there would be too few events > Mmin to fill the grid cells, at least in a low-to-moderate seismicity region. It might be possible to increase the number of earthquakes for analysis by reducing the magnitude threshold, but this is likely to introduce problems with regard to completeness thresholds.

The distribution test, on the other hand, can be applied to individual zones. Given the bad result for the terrane zone shown in Figure 4, a logical next step would be to repeat the analysis for each zone in turn to determine which ones were the chief contributors to the poor result.

Figure 5 shows a real example taken from a commercial seismic hazard site study for a location which, for reasons of confidentiality, need not be identified here; nor is it necessary to identify the authors. The hazard at site is controlled (as is usually the case) by the source

zone containing the site. In this case, this source zone also contains a single active fault. The distribution in Figure 5 is computed for the site source zone and the fault combined. Again, the historical result is shown as a star, and is seen to be an outlier of the whole distribution. Most of the simulations gave lower event counts than the historical result, suggesting that the activity rates for this zone have been underestimated. This should prompt a reinvestigation of the way in which activity rates were derived in the first place, in order to seek an explanation for this discrepancy. If the study were being peer reviewed in the conventional way, without the benefit of this sort of analysis, such a problem could easily be overlooked. This example shows that the sort of problems trapped by these analyses do occur in actual practise.

Figure 6 is another real example. In this case the number of events > Mmin in the zone containing the site was zero, so the average magnitude values cannot be compared and only the number of events per 200 years generated by the model are shown. In the case where the real number of events is zero, the distribution of model-predicted events cannot flank the real value, as one cannot have negative numbers of events. However, faced with Figure 6, one might reasonably question whether such strong conservatism is reasonable, particularly in a case where the number of events < Mmin in the modern period is also very low. Figure 6 makes very clear what is not clear when one is only reading the study report or scanning the hazard input file.

Two possible objections to this type of analysis could be raised. Firstly, one can argue that, at least in areas of low to moderate seismicity, the length of the historical earthquake catalogue is insufficient to show the full seismic cycle, and therefore it would be wrong to pin too much emphasis on the historical record. The second argument goes that usually a seismic hazard model is constructed to compute the hazard at low probabilities (such as 10^{-4} per annum), and therefore the short-term seismic record should not be used to evaluate the model. Both these arguments are specious.

In the first case, whatever the full seismic cycle might be like, the historical catalogue is a part of it. The historical earthquake catalogue is also usually a major input to the development of the model. If one cannot recreate from the model something approximating to its own input data, this suggests that something is wrong. In the case where regional seismicity is clustered, it is sometimes argued that this clustering is short-term and by chance, and if one could see the “full seismic cycle” one would realise that the seismicity is, in fact, even distributed. Such a hypothesis admits of simple statistical testing, and if testing shows

that, at a given confidence level, this hypothesis can be rejected, continuing to adhere to it becomes perverse.

There is an exception to this, which is where one can show good reasons external to the earthquake catalogue why spatial stationarity is not to be expected. In some tectonic situations one can identify zones of activity that may alternate, switching on and off over periods of a few hundred years. Since (as already stated) these tests are tools to be applied intelligently, not rigidly, in cases where a bad result can actually be explained, it may be that the model can still be retained. The test result in this case indicates the need for careful justification of the model rather than its rejection.

The second objection rests on a fallacious view of probability. One does not (or not usually) take into account, when designing a hazard model, the design probability that needs to be computed. One does not make one model for high-probability hazard and another one for low-probability hazard. If one did, one could never have a hazard curve. In a hazard study one is interested in events that have a very low probability, but one does not design a hazard model that only produces low probability outcomes. The basis of PSHA is calculating, given a model that describes the common seismic processes, what is a rare event. If the model can't compute common things accurately, it certainly can't compute rare things accurately.

A third, and more germane, reservation can be made concerning the X^2 analysis in the context of site studies. It is common, for a site study, to model an area extending up to 300 km in radius from the site. However, the influence on the hazard exerted by seismicity at the edges of the model is slight, and it is perfectly reasonable to adopt a simplistic approach to zoning the seismicity round the edges of the model, on the grounds that pursuing a more detailed approach would be unnecessary. In such a case, the X^2 test will give the model a poor score even though it is adequate for the purposes for which it was intended. This argument does not apply to models intended for hazard maps.

CONCLUSIONS

We believe that greater accountability and transparency in seismic hazard analysis is to be encouraged. When expensive engineering decisions are made on the basis of PSHA studies, the client should be able to see some evidence that the results obtained are robust and reliable. Peer review, as it is frequently practised, may not be sufficient to give a guarantee of this, when the review team look mostly at the input and not much at the output and how it is

derived from the input. Some aspects of PSHA, especially those connected with the attenuation of strong ground motion, are not easily amenable to objective testing, but methods for evaluating source zone models are practical.

This paper has presented two such tools, which can detect inconsistencies between the PSHA model and the input earthquake data on which it is based. If a model is incompatible with its own input data, it is not very likely that it will give accurate estimates of hazard from future earthquakes. When a model scores badly in any of these tests, further analysis is indicated to explain the result. In a few cases it may be that special circumstances justify the model. More likely, it is a result of modelling decisions that were well-intentioned but which had unforeseen consequences; in a few cases it may even be due to typing or compilation errors in the hazard input file.

The routine use of such procedures, especially in high-consequence projects, should improve the overall reliability of seismic hazard results.

ACKNOWLEDGEMENTS

The development of some of the methods presented here was supported by British Energy and British Nuclear Fuels. This paper is published with the permission of the Executive Director of the British Geological Survey (NERC).

REFERENCES

Bernreuter, D. L., Savy, J. B., Mensing, R. W. & Chen, J. C., 1989. Seismic hazard characterisation of 69 nuclear power plant sites east of the Rocky Mountains. In: *Report*, US Nuclear Regulatory Commission.

Budnitz, R. J., Apostolakis, G., Boore, D. M., Cluft, L. S., Coppersmith, K. J., Cornell, C. A. & Morris, P. A., 1997. Recommendations for probabilistic seismic hazard analysis: guidance on uncertainty and use of experts. In: *Report*, U.S. Nuclear Regulatory Commission.

Grünthal, G., Bosse, C., Musson, R. M. W., Gariel, J.-C., Crook, T. d., Verbeiren, R., Camelbeeck, T., Mayer-Rosa, D. & Lenhardt, W., 1996. Joint seismic hazard assessment for the central and western part of GSHAP Region 3 (Central and Northwest Europe). In: *Seismology in Europe* (ed Thorkelsson, B.), pp. 339-342, Icelandic Met. Office, Reykjavik.

Jiménez, M. J., Giardini, D., Grünthal, G. & the SESAME Working Group, 2001. Unified seismic hazard modeling throughout the Mediterranean region. *Bolletino di Geofisica Teorica ed Applicata*, **42**, 3-18.

Krinitzsky, E. L., 1993. Earthquake probability in engineering - Part 1: The use and misuse of expert opinion. *Engineering Geology*, **33**, 257-288.

Mucciarelli, M., Peruzza, L. & Caroli, P., 2000. Tuning of seismic hazard estimates by means of observed site intensities. *Journal of Earthquake Engineering*, **4**, 141-159.

Musson, R. M. W., 1994. A catalogue of British earthquakes, British Geological Survey.

Musson, R. M. W., 2000. The use of Monte Carlo simulations for seismic hazard assessment in the UK. *Annali di Geofisica*, **43**, 1-9.

Musson, R. M. W., 2004. Objective validation of seismic hazard source models. In: *13th World Conference on Earthquake Engineering Proceedings*, pp. 2492, Vancouver.

Musson, R. M. W. & Winter, P. W., 1996. Seismic hazard of the UK. In: *AEA Technology Report AEA/CS/16422000/ZJ745/005*, Risley.

Pharaoh, T. C., 1996. Tectonic map of Britain, Ireland and adjacent areas, 1:1 500 000, British Geological Survey, Keyworth.

Slejko, D., Peruzza, L. & Rebez, A., 1998. Seismic hazard maps of Italy. *Annali di Geofisica*, **41**(2), 183-214.

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

Woo, G., 1996. Kernel estimation methods for seismic hazard area source modelling. *Bulletin of the Seismological Society of America*, **86**(2), 353-362.

FIGURES

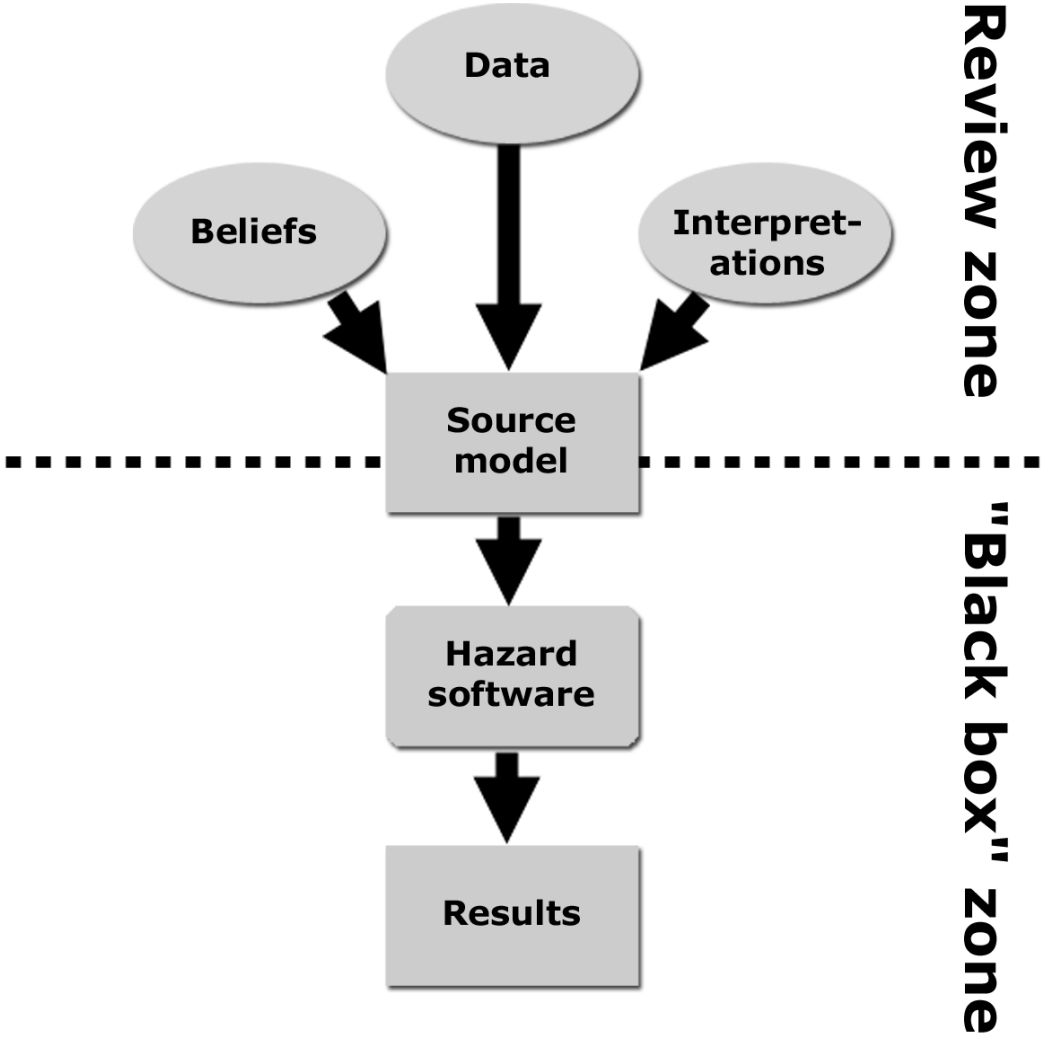

Figure 1. Schematic depiction of the processes involved in a PSHA study.

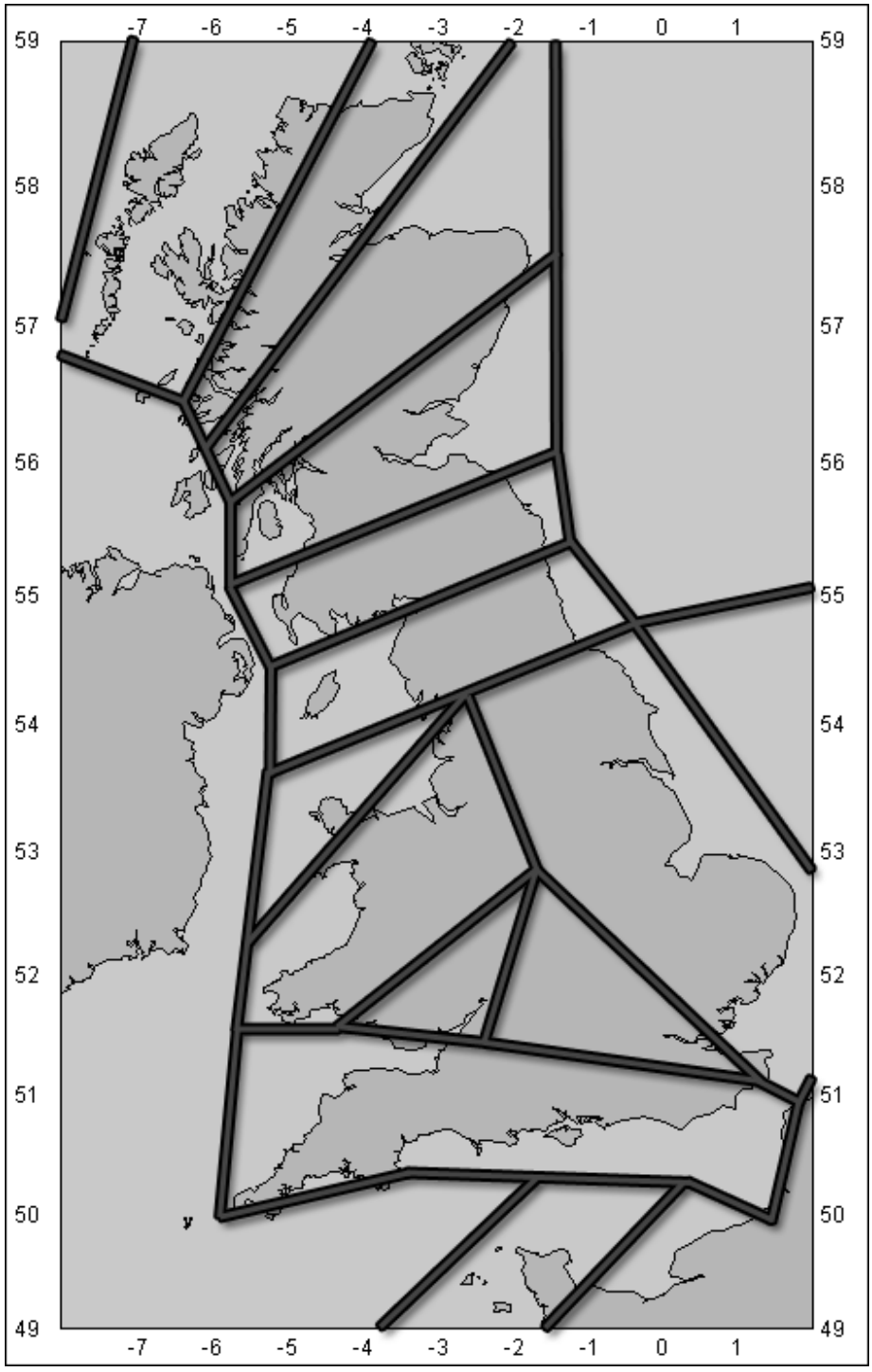

Figure 2a. Source model for the UK based on geological terranes.

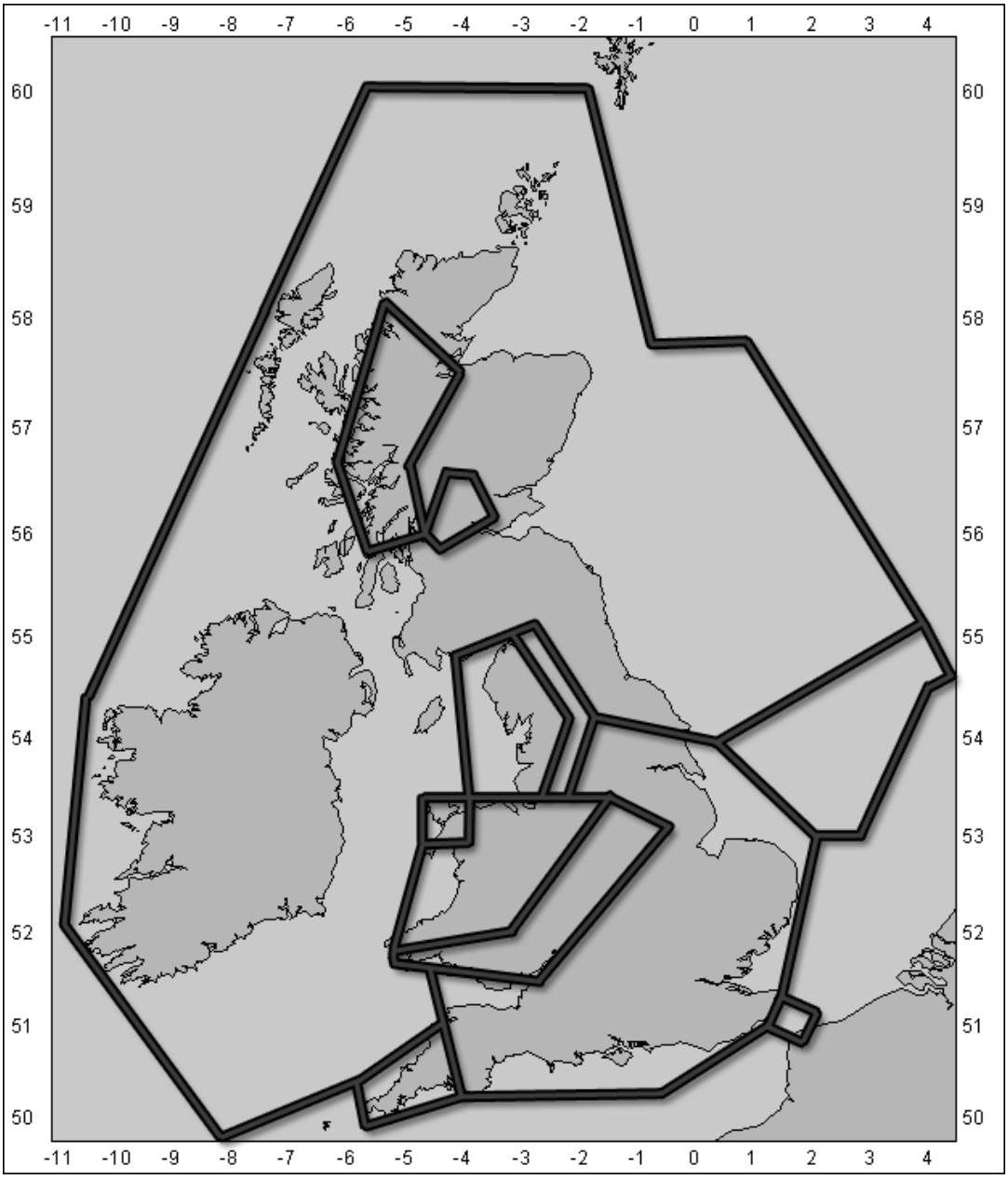

Figure 2b. Source model used for GSHAP (slightly simplified).

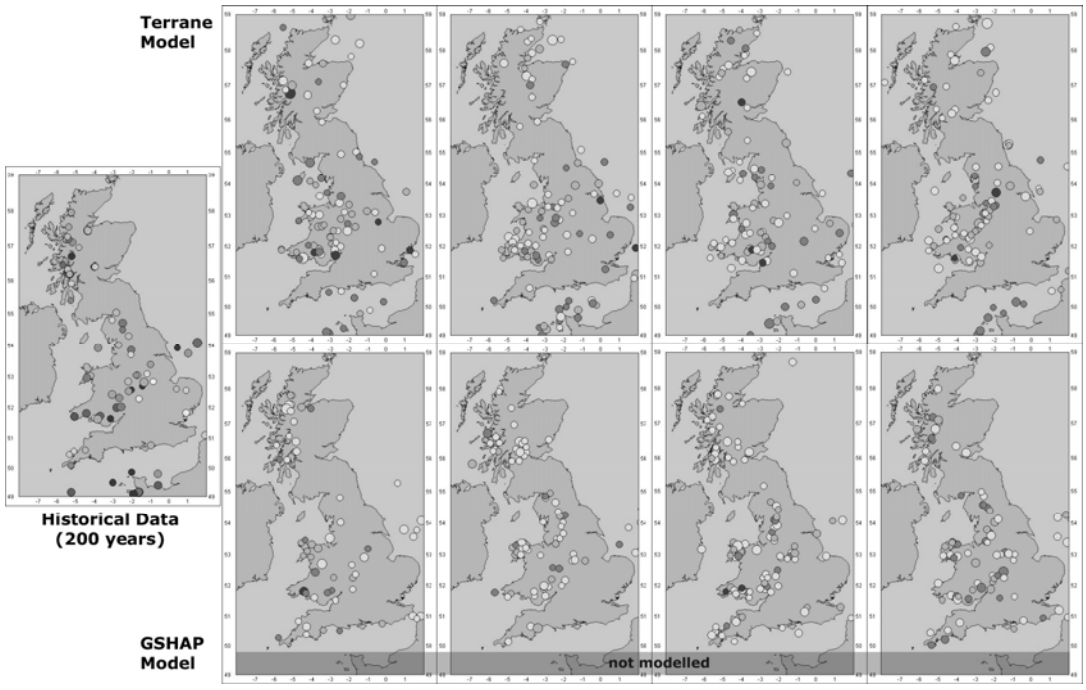

Figure 3. Sample simulations from the two models, compared to the historical data (left).

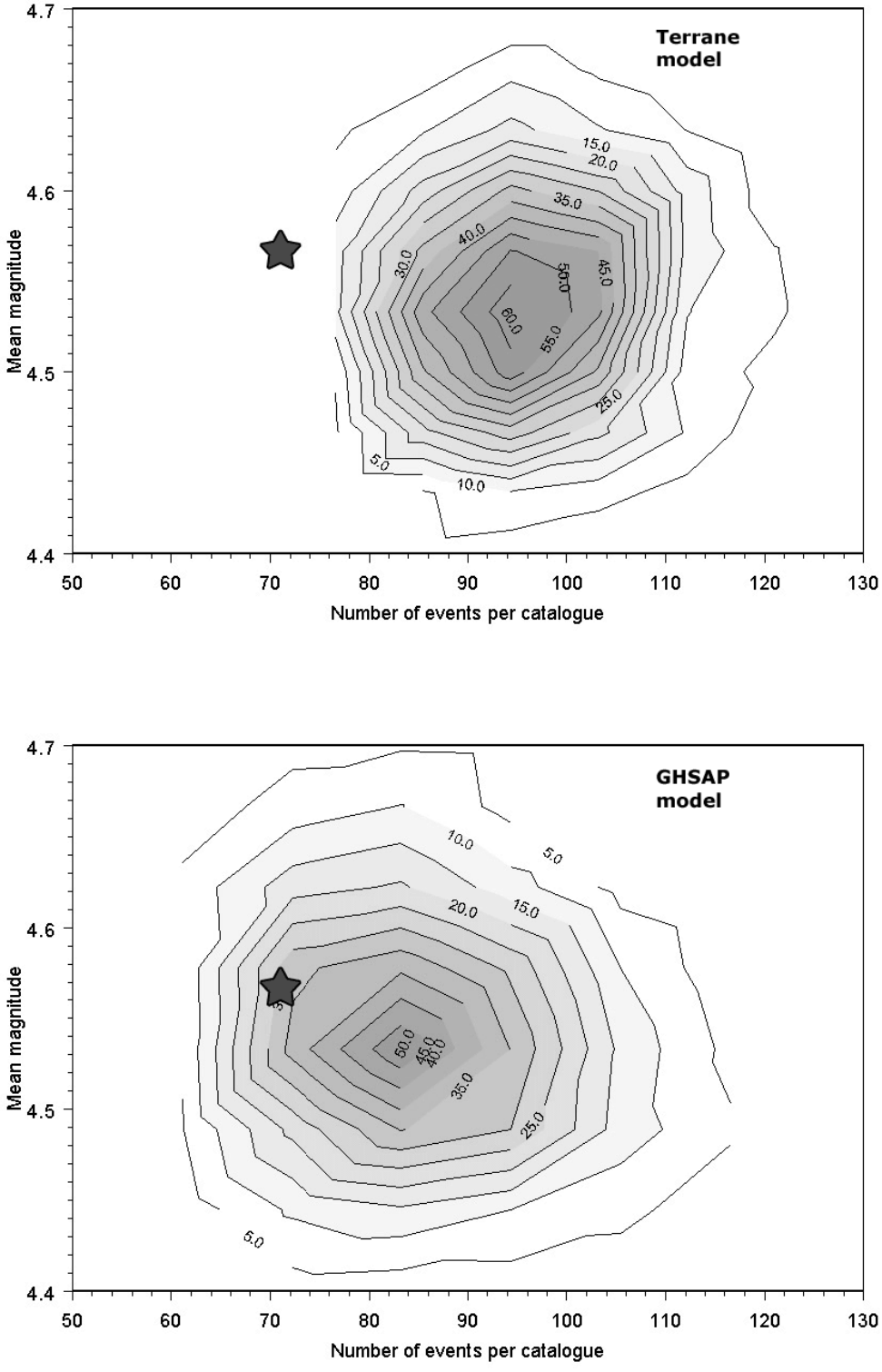

Figure 4. Sample simulations from the two models, compared to the historical data (left).

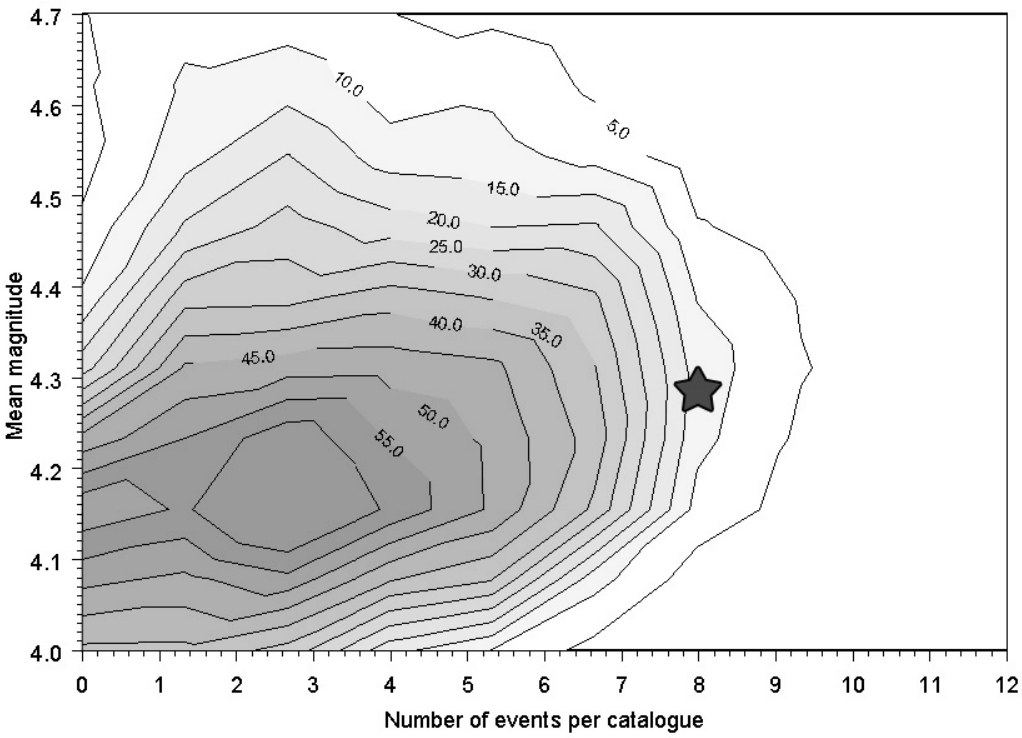

Figure 5. Rate space plot for a single source zone from an actual study. The star indicates the historical values.

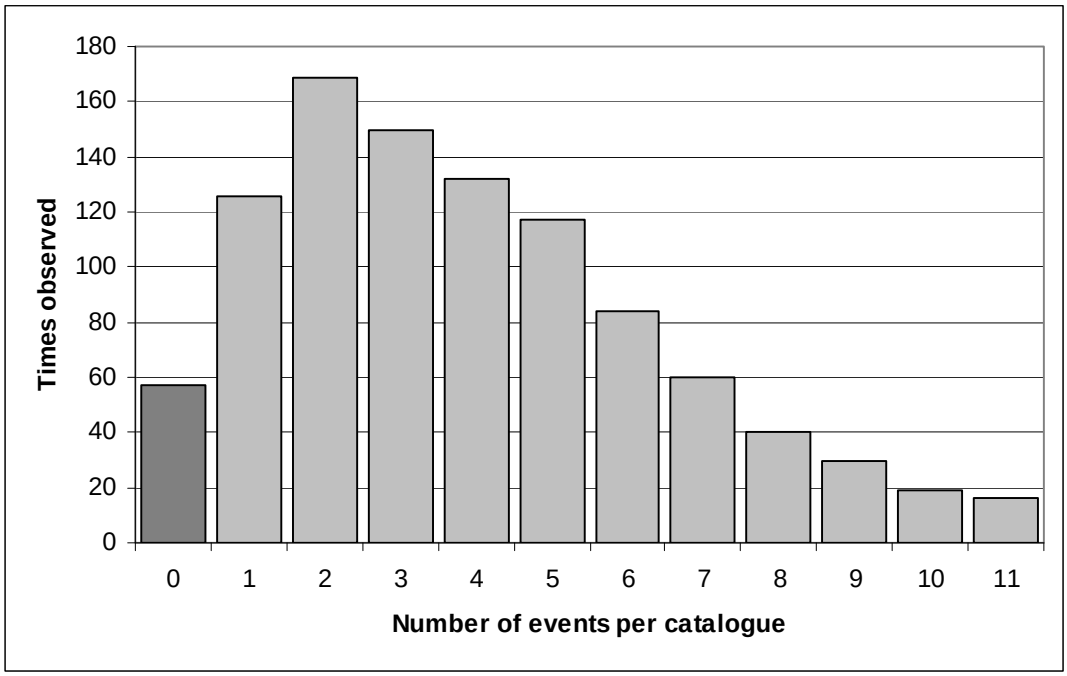

Figure 6. Rate plot (number of events per 200 years only) for a single source zone from an actual study. The actual number of observed events is zero.

.