Bull Earthquake Eng (2012) 10:367–378
DOI 10.1007/s10518-011-9299-6

ORIGINAL RESEARCH PAPER

Objective assessment of source models for seismic hazard studies: with a worked example from UK data

R. M. W. Musson · P. W. Winter

Received: 25 February 2010 / Accepted: 23 June 2011 / Published online: 7 July 2011

Abstract 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 a danger that assessment procedures and review methods contribute further subjective judgements on top of those already elicited. It is helpful to find techniques for objectively assessing seismic source models that show what the interpretations physically mean in terms of seismicity. Experience shows that well-meaning but flawed design decisions can lead to source models that are incompatible with the seismic history that was used as input. In this paper a method is demonstrated in which large numbers of synthetic earthquake catalogues, that match the completeness thresholds of the historical catalogue, are generated. The study area can be divided into a grid of uniform cells, and the number of earthquakes in each cell in both the historical catalogue and each simulated catalogue are then counted. Comparison of the historical pattern and a set of 1,000 simulated patterns, using a X^2 test, shows if the historical pattern is credibly a member of the set of outcomes obtainable from the seismic source model. A second method is to chart the distribution of a large sample of simulated catalogues in terms of magnitude frequency, and observe whether the historical catalogue is comfortably within this distribution, or an outlier. If it proves impossible to replicate the historical catalogue using the model, it casts doubt on whether the model is a valid depiction of the seismicity rates that will govern the future hazard. At the very least, the disparity needs careful investigation to ensure the model is error-free. A worked example is presented here for the UK, using a source model that was used in Global Seismic Hazard Map (GSHAP), compared to one that was artificially constructed to be credible but flawed. Two tests find the GSHAP model to be an acceptable representation of the pattern of seismicity in the UK, while the artificial model is conclusively rejected.

R. M. W. Musson (✉)
British Geological Survey, West Mains Road, Edinburgh EH 93LA, UK
e-mail: rmwm@bgs.ac.uk

P. W. Winter
ESR Technology, Whittle House, 410 Birchwood Park, Warrington, Cheshire WA3 6FW, UK
e-mail: paul.winter@esrtechnology.com

Keywords Seismic hazard · PSHA · Validation · Seismic source models · Stochastic modelling

1 Introduction

The practice of probabilistic seismic hazard analysis (PSHA) in the past has seldom been 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 often announced without discussion—or is even not mentioned at all. Yet by arbitrarily changing this parameter one can substantially alter the hazard results. Hazard software can resemble 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; not the most recent one) 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 conclusion that there was actually something wrong with the source model used for the probabilistic calculations, which subsequently turned out to be the case.

Many more examples could be adduced; however, the majority will be found in the grey literature, since most PSHA studies are undertaken for specific engineering projects, not for publication. Where engineering decisions rely on results of PSHA studies, published or not, quality is essential.

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.

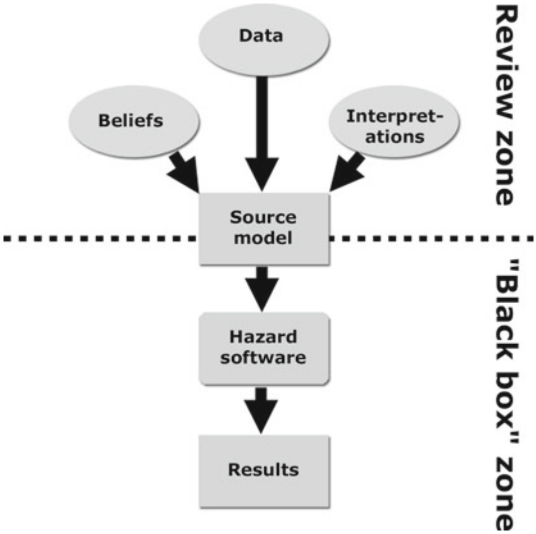

Fig. 1 Schematic depiction of the processes involved in a PSHA study. Typically, a review of a PSHA study evaluates the inputs to the seismic source model in terms of the data and judgements that go into its construction. What happens between the formulation of the model and the actual results is largely hidden

The process can be summed up by the diagram in Fig. 1. The team conducting the study has access to some relevant data; they have some beliefs of their own and they have listened 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 or low 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 Fig. 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. What tends not to be examined in detail is the part below the line in Fig. 1, which is where the model is turned into results. Cases where errors in hazard studies (for instance, a typing error in an input file) have slipped past a peer review team are not unknown (but obviously, cannot be cited).

It helps to ask, 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. 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.

In the course of this paper, some general issues will be discussed, and it will be argued that consistency of a seismic source model with the seismicity data on which it is based is a fundamental check. A worked example will be given using the seismic hazard modelling of the UK as a test case, showing how a X^2 statistic can be used to compare actual input (the earthquake catalogue) and what the model forecasts as possible future patterns of earthquake activity.

2 Overview of formal testing

Techniques for formally testing seismic source models can trap two types of error: misguided decision-making and gross errors. The former category can include making assumptions not justified by the data; the latter includes such things as critical typing errors in the input file. The aspects of PSHA that are best suited to testing of this sort are the spatial interpretation of seismicity and the activity rates.

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. Krinitzsky 1993) and on the other hand by proponents of spatial smoothing (e.g. Woo 1996). The method employed tends to be that some specific model is held up for criticism with the implication that, hence or otherwise, all models are equally bad. The usual convenient target 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 fallacious to argue that because some study applies a methodology badly, therefore the methodology is bad. Some models are in fact not acceptable representations of reality, and this should be recognised, and procedures devised for sorting out the bad ones from the good ones.

A note on terms here is important. It should not be thought that describing a model as "valid" is claiming it to be true; there are no "true" models. The objective here is to identify models that cannot even reproduce their own input data. Such models are effectively falsifiable. A "valid" model is simply one that is "not capable of being invalidated by available tests".

One can discern a basic principle behind all PSHA studies that there exist some controlling parameters behind the long-term seismicity of any area. 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. If the model is realistic, 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 questionable.

The essence of formal testing is therefore one of considering the set of all possible outcomes of the model in terms of earthquake catalogues equal in length to the historical catalogue. Is the historical catalogue credibly a member of this set? This is a fairly common type of question in statistics, and tools are available to assess significant differences between a sample (in this case, the historical outcome) and a population (all the possible outcomes of the model).

3 A worked example: inputs

To demonstrate this, a worked example is presented, taking the UK as a case in point. Two possible zone models are shown in Fig. 2. The first model (Fig. 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 1996), with some trimming and simplification) is a seismic source zone. It is not intended 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 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 (Fig. 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.

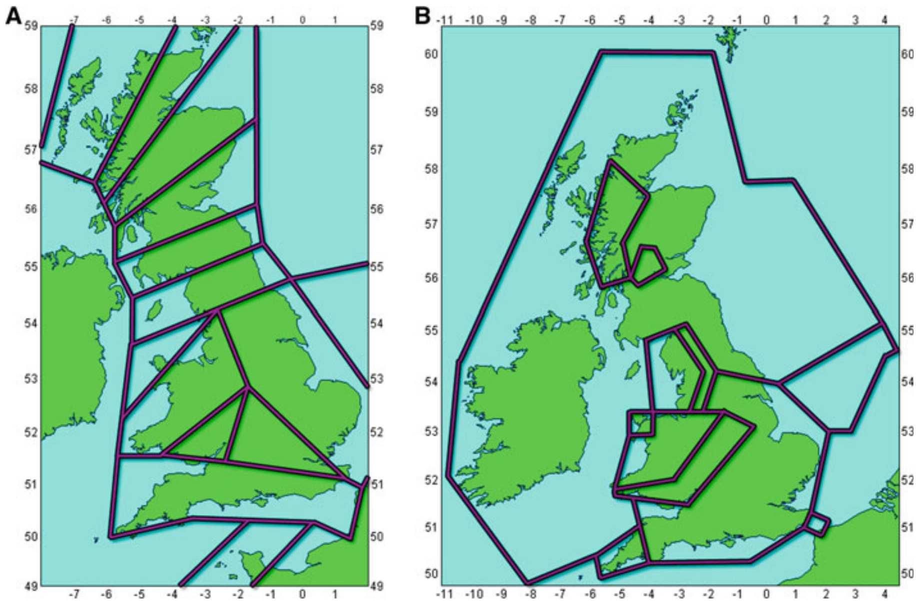

Fig. 2 **a** Source model for the UK based on geological terranes. This is intended to be an example of a model that could perhaps be considered credible by someone—it is not suggested for actual use. **b** Source model used for GSHAP, as shown in Grünthal et al. (1996), but slightly simplified for the 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 using 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). Since this is a demonstration exercise, epistemic uncertainty in magnitude-frequency parameters is not relevant, and each source zone was given a single activity rate—*b* value pair determined by maximum likelihood analysis. A single maximum magnitude value of 6.5 ML was assumed for all zones in both models.

The 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.

4 X^2 testing

Since both models describe fully the supposed long-term seismicity parameters for the UK, they can be used to generate synthetic catalogues by applying a Monte Carlo process (described in 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 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, 1,000 catalogues were now prepared. For statistical analysis, it is necessary to compare like numbers 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 is different from the situation in Fig. 3, where the synthetic catalogues were all set at 200 years.

The area was now divided into a 5 $\times$ 5 grid. 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 to ensure the integrity of the approximation of the chi-squared distribution. This is not a rigid limit; one can use a lower limit provided a balance is struck between the total number of cells and the fraction of cells with counts of less than five (Yarnold 1970).

The procedure is depicted in Fig. 4. If the cells are numbered counting from top-left as number 1, then the significant cells in the historical outcome are numbers 2, 7, 13, 18, 19, 22 and 24, with counts of 6–9–8–9–5–5–6. The corresponding counts in the particular simulation shown are 1–4–6–13–2–0–8.

The match of cell counts in the historical file was then compared to the cell counts in each of the 1,000 simulated catalogues, and the X^2 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.

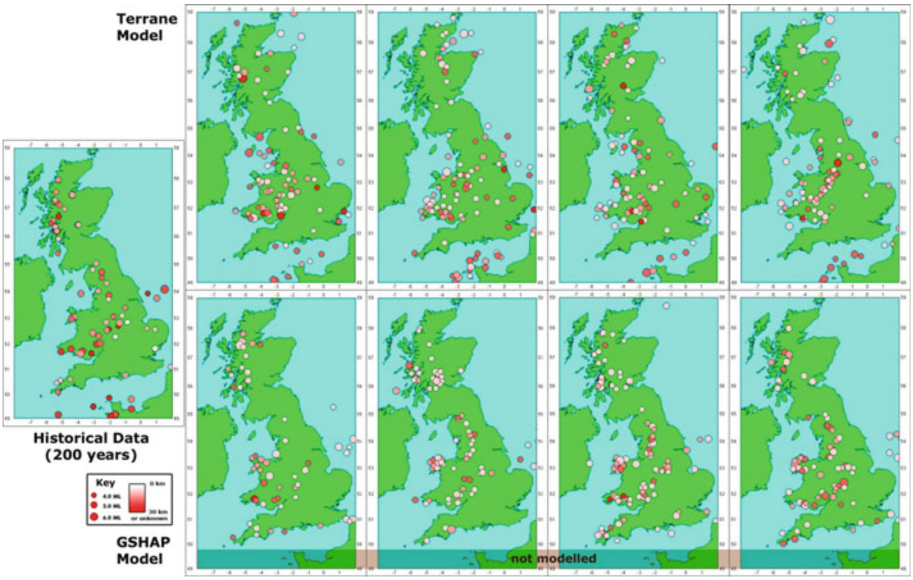

Fig. 3 Sample simulations from the two models, selected at random. The *upper row* is taken from the "terrane model", the *lower row* from the GSHAP model, and the comparable historical map is shown at the *left*. Even cursory inspection suggests that it is unlikely that the historical result could be a member of the set represented by the six maps on the *upper row*. It looks credible that it could belong with the *lower row*

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 1,001st simulated catalogue. This catalogue was then compared to the previous 1,000 simulations, in the certain knowledge that it was produced by the same model. Because the correct answer is known in advance (that the 1,001st simulation is compatible with the other 1,000), 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

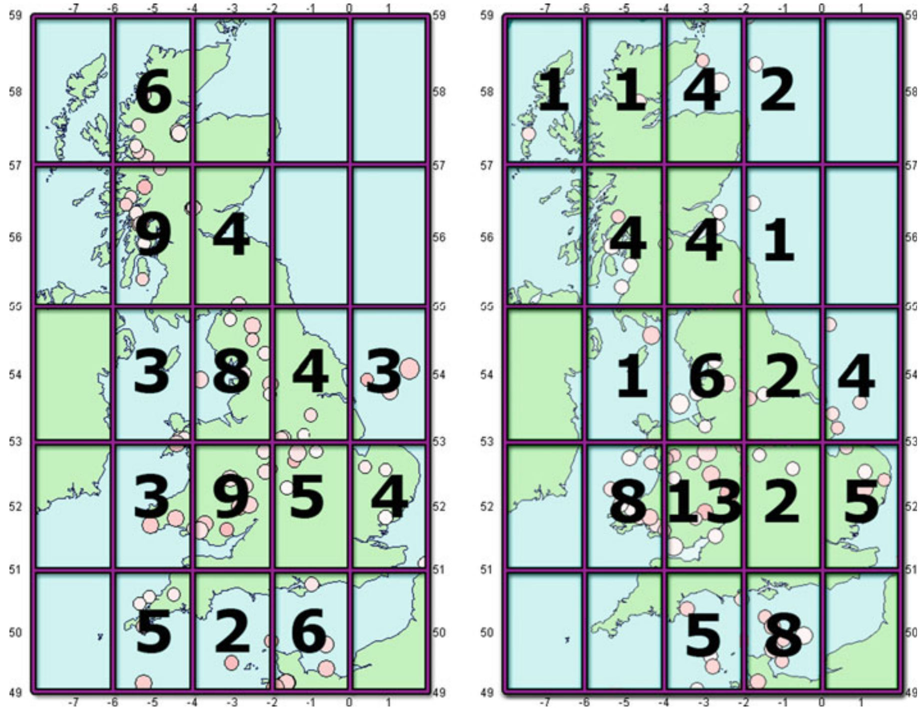

Fig. 4 Illustration of the event counting procedure. The *left* side shows the historical number of events in each cell (total 71). The *right-hand side* shows the counts for the first 71 events of one of the simulated catalogues from the terrane model. The historical cell counts can be compared to the patterns of cell counts from the simulated catalogues using a standard X^2 test

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.

5 Aggregate testing of activity rates

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. Mean magnitude, for sets of events with the same completeness and bounds, is a simple analogue of *b*-value, which has the advantage of being uniquely determined. 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

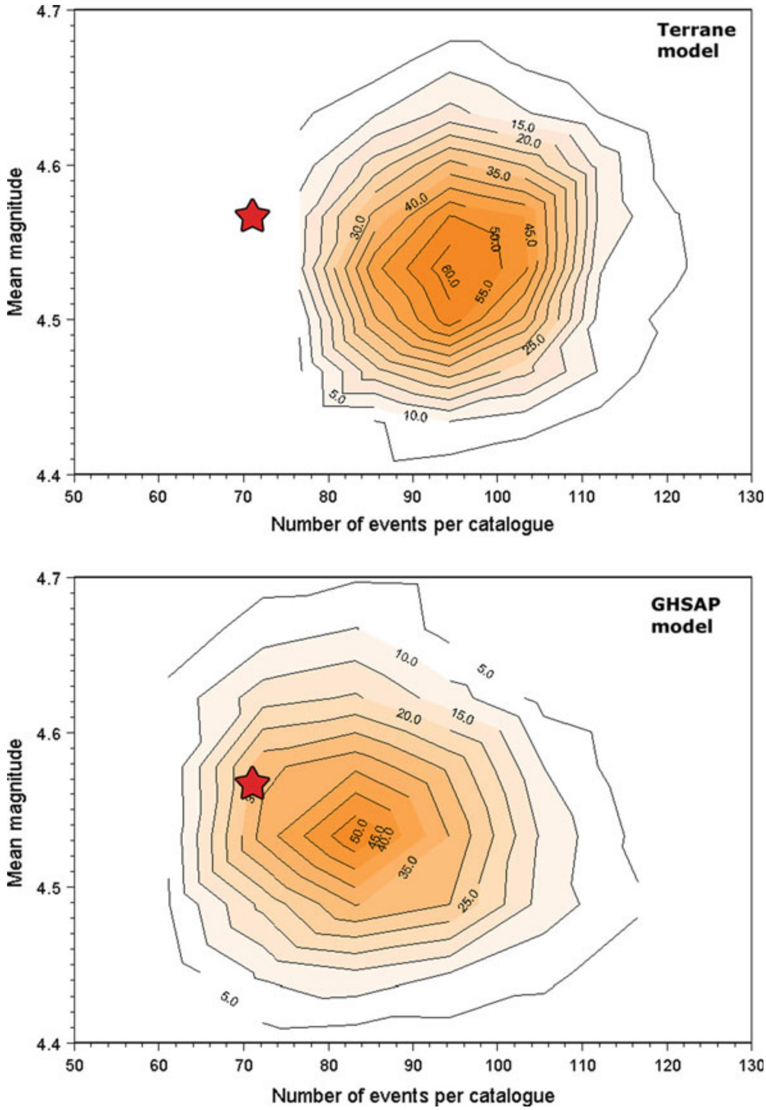

Fig. 5 "Rate space" plots for UK seismicity in 200 years generated by the two models. The axes are the number of events and the mean magnitude (a simple analogue of the *b* value). The contours show the number of results (out of 1,000) giving the different possible combinations.The *star* indicates the historical values, which in the *upper plot* are outside the whole distribution

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 Fig. 5. Here the two test models have been used to generate 1,000 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 these two parameters. 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; most of the possible outcomes forecast by the model involve more earthquakes ≥ 4.0 ML in 200 years than actually occurred in the last 200 years. The modal value is 83 events, compared to the historical outcome of 71. The model also inclines very slightly towards steeper *b*-values. The mean magnitude for the maximum point of the whole distribution is 4.53, compared to the historical value of 4.57. However, if one disregards the number of events per simulation and looks only at the distribution of mean magnitude per simulation, the median is 4.56, very close to the historical value.

6 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 X^2 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 X^2 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 Fig. 5, 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.

Experience with unpublished seismic hazard studies suggests that such problems are not confined to artificial examples of the type shown here. Cases have been seen in which activity rates in critical source zones have been either significantly underestimated, or sometimes, overestimated, for different reasons (which can be as simple as a misplaced decimal point in a long and hard-to-read input file). Often such problems can easily be overlooked.

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.

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 at least careful investigation needs to be made to explain and justify the divergence.

As to the second objection, 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 not compute a hazard curve. In a hazard study one is interested in events that have a very low probability, but one does not normally design a hazard model that is only valid for low probability outcomes. The PSHA processes calculates rare events, from a model that describes the common seismic processes. If the model can't compute common things accurately, it may not be reliable in computing 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, and the testing method is still appropriate for the close-in zones that control the hazard in a site-specific model.

It should be stressed that testing methods are not prescriptive; they are principally tools for identifying potential problems that require further checking.

7 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. 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 raises questions as to if 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 some cases it may be that particular circumstances justify the model. But it can be 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. At the very least, it demonstrates that there are issues that need to be looked into. A good result does not validate the model in the sense of proving it, but does give one increased confidence that it is not obviously flawed.

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

Acknowledgments The development of some of the methods presented here was supported by British Energy and British Nuclear Fuels. The paper was also significantly improved by the comments of an anonymous reviewer. This paper is published with the permission of the Executive Director of the British Geological Survey (NERC).

References

Bernreuter DL, Savy JB, Mensing RW, Chen JC (1989) Seismic hazard characterisation of 69 nuclear power plant sites east of the rocky mountains. US Nuclear Regulatory Commission, NUREG/CR-5250

Budnitz RJ, Apostolakis G, Boore DM, Cluft LS, Coppersmith KJ, Cornell CA, Morris PA (1997) Recommendations for probabilistic seismic hazard analysis: guidance on uncertainty and use of experts. US Nuclear Regulatory Commission, NUREG/CR-6372

Grünthal G, Bosse C, Musson RMW, Gariel J-C, Crook TD, 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: Thorkelsson B (ed) Seismology in Europe. Icelandic Met. Office, Reykjavik, pp 339–342

Jiménez MJ, Giardini D, Grünthal GThe Sesame Working Group (2001) Unified seismic hazard modeling throughout the Mediterranean region. Bolletino di Geofisica Teorica ed Applicata 42:3–18

Krinitzsky EL (1993) Earthquake probability in engineering - Part 1: the use and misuse of expert opinion. Eng Geol 33:257–288

Mucciarelli M, Peruzza L, Caroli P (2000) Tuning of seismic hazard estimates by means of observed site intensities. J Earthq Eng 4:141–159

Musson RMW (1994) A catalogue of British earthquakes. British geological survey global seismology report, WL/94/04

Musson RMW (2000) The use of Monte Carlo simulations for seismic hazard assessment in the UK. Annali di Geofisica 43:1–9

Musson RMW (2004) Objective validation of seismic hazard source models. Proceedings of the 13th world conference on earthquake engineering, Vancouver 2492

Musson RMW, Winter PW (1996) Seismic hazard of the UK Report AEA/CS/16422000/ZJ745/005

Pharaoh TC (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 41:183–214

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: Galadini F, Meletti C, Rebez A (eds) Richerche del GNDT nel campo della pericolosita sismica (1996–1999). CNR-GNDT, Rome, pp 333–343

Woo G (1996) Kernel estimation methods for seismic hazard area source modelling. Bull Seismol Soc Am 86:353–362

Yarnold JK (1970) The minimum expectation in X^2 goodness of fit tests and the accuracy of approximations for the null distribution. J Am Stat Assoc 65:864–886