SHARE

Grant Agreement no. 226967

Seismic Hazard Harmonization in Europe

Project Acronym: SHARE

SP 1-Cooperation

Collaborative project: Small or medium-scale focused research project

THEME 6: Enviroment

Call: ENV.2008.1.3.1.1 Development of a common methodology and tools to evaluate earthquake hazard in Europe

D3.7a – Assessment of activity rates for seismic source zones

Due date of deliverable: 01 September 2011

Actual submission date: 26 July 2011

Start date of project: 2009-06-01 Duration: 36

RMW Musson

British Geological Survey (BGS)

Revision: 1

Dissemination Level		
PU	Public	x
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

1. Introduction

This report describes the basis on which magnitude-frequency statistics have been calculated for seismic source zones (SSZs) within the SHARE project conducted for the European Union under the 7th Framework Project Grant Agreement no. 226967. It also serves as a user guide and manual for operating the program Attic Ivy (current version 1.2 as of July 2011). This software was developed within the SHARE project, and is distributable amongst project partners. It is written in FORTRAN and designed to run under MS-DOS (or at a Windows command prompt) but has also been tested under UNIX. The program name is an anagram of "activity". Part of the program is based on code written by R.R. Youngs circa 1992, and made available to SHARE by kind permission for use within the project. The original code has been simplified, and in some places modified, but some legacy features have been retained in the code that are not actually utilised.

A companion report will be issued detailing the decisions made in actually computing activity rates for Europe in the SHARE project, and listing the results.

2. Methodological background

The process of assessing seismic hazard can be divided into three principle tasks, as shown in the often-reproduced figure from TERA Corporation (1980), shown here as Figure 1. These tasks are the identification of seismic sources, the description of seismic sources in terms of the characteristics of the earthquake activity, and the characterisation of the propagation of strong ground motion from earthquakes occurring in those sources. It is often the case that a seismic hazard project is divided into two areas of work: seismic source characterisation (SSC) and ground motion characterisation (GMC). The SSC part of a project is a combination of the first two tasks, while the GMC part is the third task.

The purpose of this report is to discuss the second task, the characterisation of earthquake recurrence within a seismic source; in particular, within a seismic source zone (SSZ), taken to be an area containing one or more populations of faults such that one can reasonably conclude that there is an equal chance of an earthquake occurring anywhere within the SSZ. For SSZs, the key to assessing earthquake activity is largely through analysis of the relevant section of the regional earthquake catalogue. For fault sources, other approaches can be used based on slip rate.

Two assumptions are generally made about earthquake occurrence within a SSZ: firstly, that seismicity follows a Poisson process, and secondly, that seismicity follows a Gutenberg-Richter power law model according to equation (1).

$$\text{Log } N = a - b\,M \quad (1)$$

In equation (1), N is the cumulative number of earthquakes per year equal to or greater than magnitude M, and *a* and *b* are constants. Equation (1) can also be written

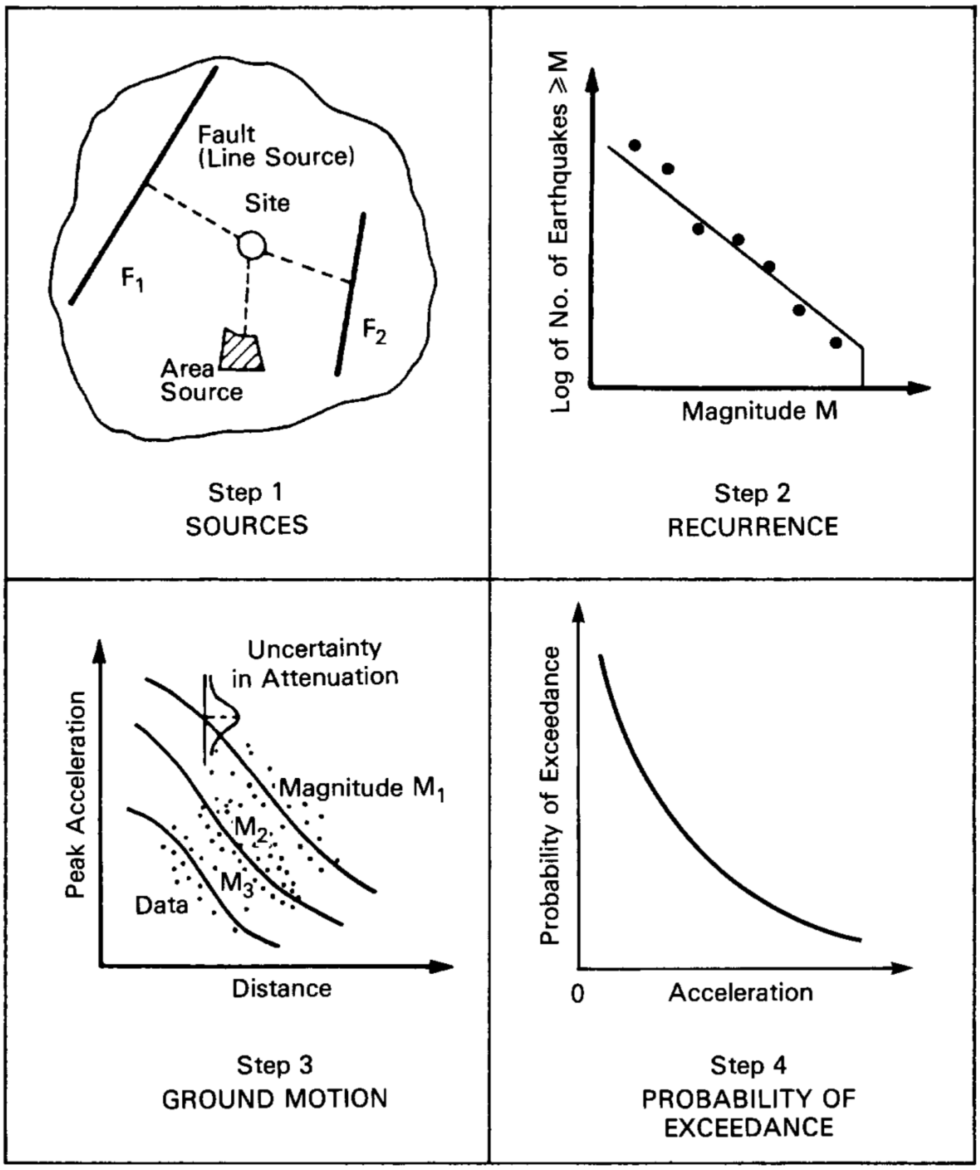

Figure 1 - Elements of seismic hazard assessment (TERA 1980)

$$\text{Log } N = a - b\,(M - m0) \quad (2)$$

where m0 represents a threshold magnitude of interest. Note that neither assumption is absolutely necessary in probabilistic seismic hazard assessment (PSHA), but they are usually found to be an adequate representation of reality.

If these assumptions are taken as applicable, it follows that the earthquake behaviour of a SSZ can be adequately described by three parameters: the *a* and *b* constants in equation (1), and the maximum possible earthquake, Mmax. This last is required; from equation (1) it would be possible to infer that there is an infinitely small possibility of an infinitely large earthquake, which is not the case. Equation (1) has to be considered as truncated at some upper bound magnitude that expresses the physical limit on earthquake size that could occur given the tectonic properties of the SSZ.

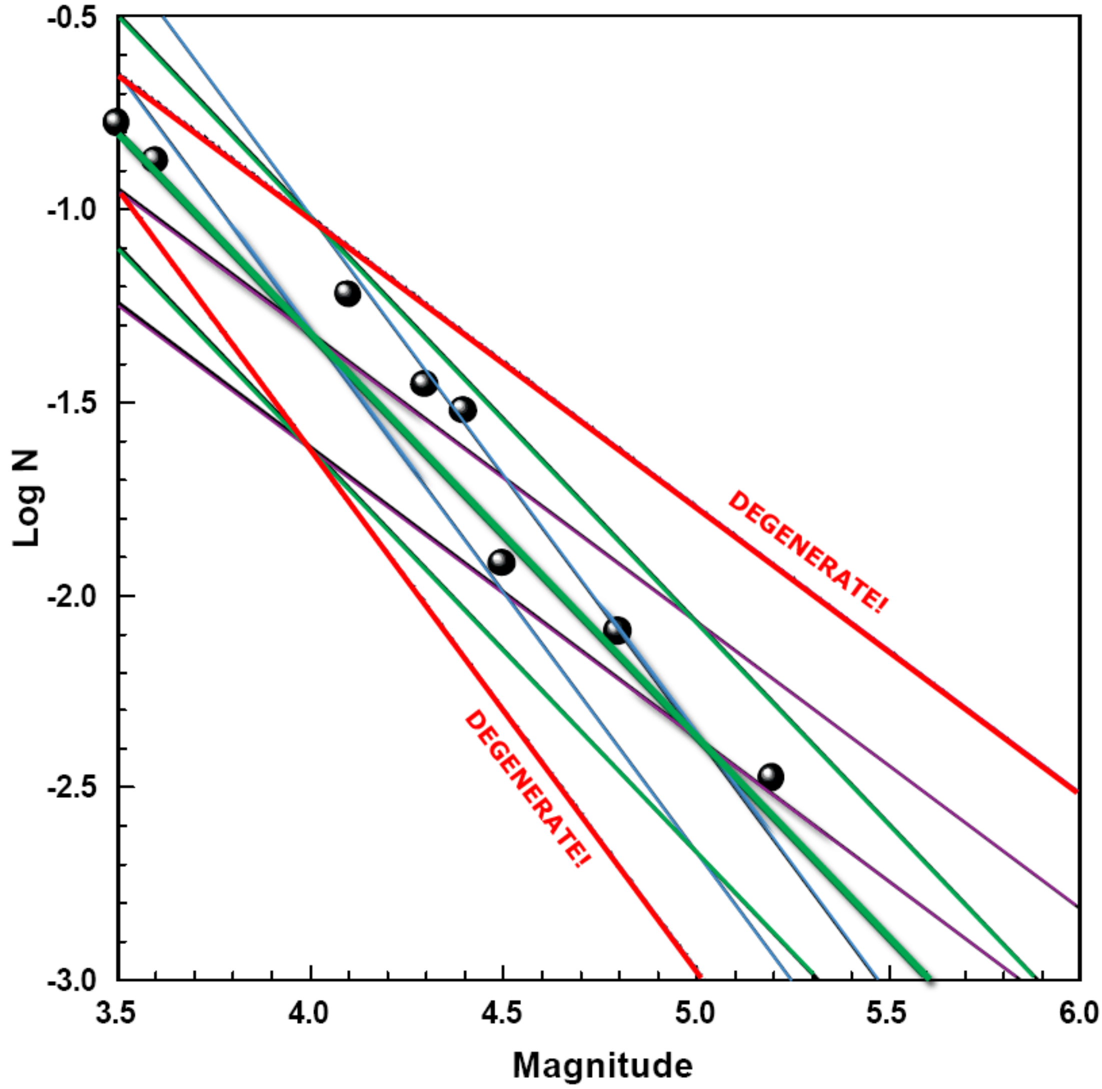

Figure 2 - Effects of varying a and b parameters

These three parameters, *a* (the activity rate), *b* (slope) and Mmax, are linked. In fitting equation (1) to actual seismicity data, it will be found that a correlation exists between *a* and *b*, as shown in Figure 2. In this figure, actual catalogue data are shown by black bullets. A best-fit of equation (1) is shown by the bold green line. If one allows for some uncertainty in both *a* and *b*, other fits are possible. Changing the activity rate alone while conserving the *b* value results in the other two green lines. Increasing the *b* value gives the blue lines, and decreasing it gives the purple lines. The combination of high *b* and low *a*, and low *b* and high *a* result in the red lines, which fail to intersect with the data and can therefore be considered degenerate.

The involvement of Mmax is less immediately obvious, but is demonstrated in Figure 3.

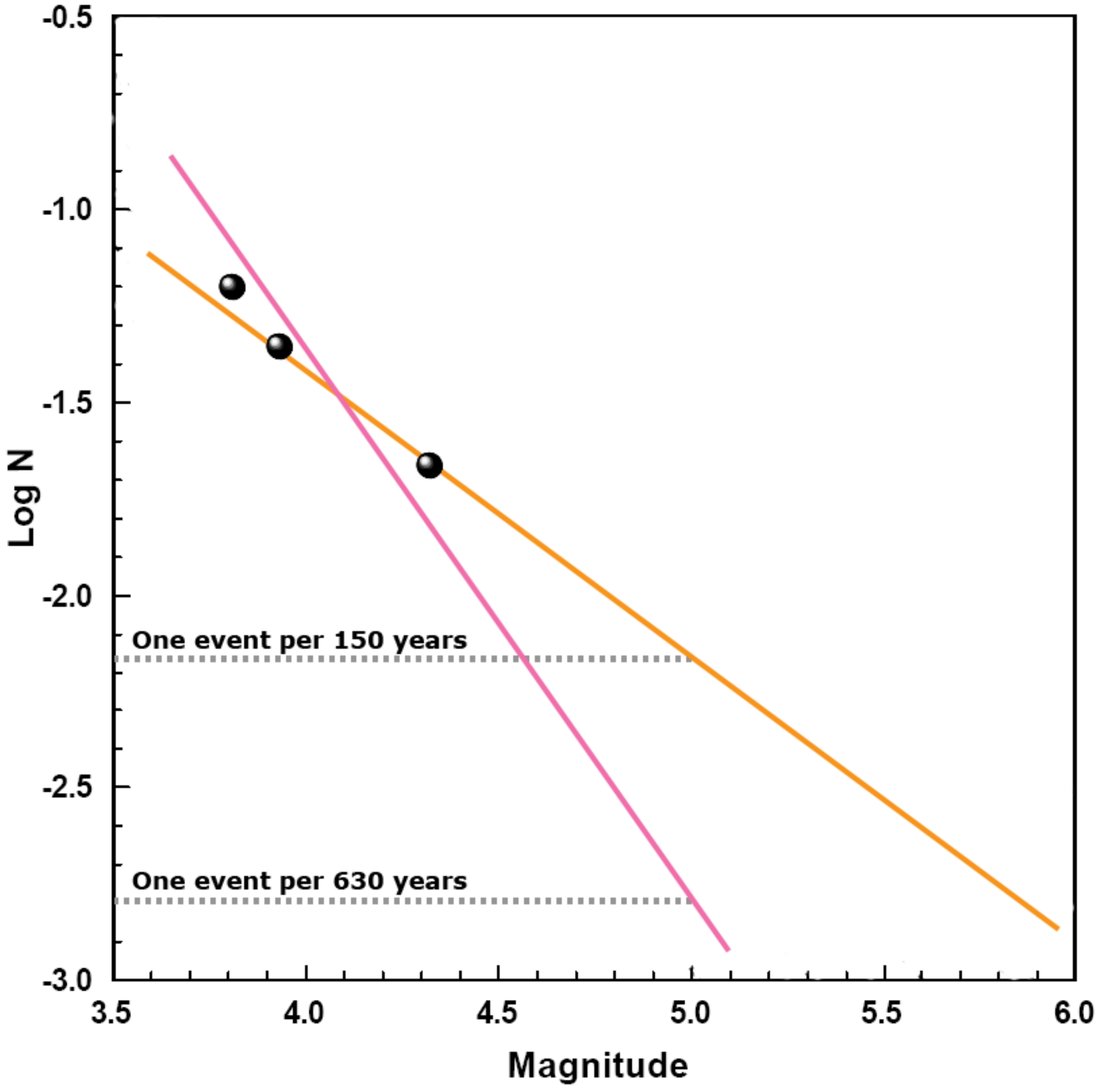

Figure 3- b values and Mmax

This figure shows a fairly typical low-seismicity case. There are no observed earthquakes above 4.4 Mw, despite a reasonably lengthy period of observation. The orange and pink lines represent two attempts to fit equation (1). If the orange line is correct, there should be one event > 5.0 Mw every 150 years, whereas none have been observed. If the pink line is correct, an event > 5.0 Mw is expected only once every 630 years, which is easier to reconcile with the absence of observations. The area between the two lines to the right of their intersection effectively represents the amount of sesimicity predicted by the orange line above that predicted by the pink line, which, above 4.5 Mw, represents an overprediction with respect to observation. The higher the maximum magnitude, the larger this area becomes, and the less likely it is that the orange line is credible. In other words, the higher the Mmax, the more probable it is that the *b* value is high.

However, in practice, the correlation between Mmax and *b* is much weaker than that between activity rate and *b*, and common practice is to determine Mmax independently.

A way of jointly determining all three parameters is proposed by Musson (2004), and it is helpful to consider this because it provides a simple conceptual framework that is helpful in understanding the full issues.

Suppose, for a given SSZ, the earthquake catalogue contains five earthquakes in the range 4.0-4.4 Mw, three 4.5-4.9 Mw events, a 5.3 Mw and a 5.6 Mw. This can be expressed as a vector Vh, where

$$Vh = [\ 5, 3, 1, 1, 0\ , 0] \qquad (3)$$

each element giving the number of events observed historically within magnitude bins of 0.5 units width, extending from 4.0 to 6.9 Mw.

Now arbitrarily take some credible values for each of *a*, *b* and Mmax. Assuming seismicity to be Poissonian, generate stochastically a simulated earthquake catalogue equal in length to the historical catalogue (it is assumed here that the historical catalogue can be considered complete above 4.0 Mw for some suitable catalogue length). This simulated catalogue can be expressed as a vector Vn and compared to Vh. In the simplest case, one can propose an error parameter ε such that

$$\varepsilon n = Vh - Vn \qquad (4)$$

This can then be repeated over n = 1, 2, 3 … 1,000. The number of times equation (4) returns a value of zero is a direct expression of the probability of obtaining the historical catalogue by chance, given the chosen values of *a*, *b* and Mmax. The exercise can then be systematically repeated over all credible combinations of values for *a*, *b* and Mmax. The best-fitting values will return lower values of ε, and one can use these scores for weighting preferences for any *a*, *b*, Mmax triplet. The basic question being asked is, if these parameters are correct, how likely is it that the historical result would be observed?

Estimating magnitude recurrence parameters using this stochastic method, and applying the results in PSHA, is rather cumbersome, but possible. It is simplified if one decouples the *a* and *b* results from the Mmax distribution, and because the correlation with Mmax is weaker, one loses relatively little information by doing this. Such an approach was one of several implemented in the PEGASOS project (Musson et al 2009).

The overall effect of taking this quasi-observational approach is to arrive at a maximum likelihood estimator by another route.

The use of maximum likelihood as an estimator of *b* value was first proposed by Aki (1965), and ever since has been considered to be superior to the use of least squares regression, which is what has been used as an alternative. The problem with least squares regression is that it simply minimises deviation of all points from the regression line. In doing so, it treats all points as equal. But because seismicity follows a cumulative power law distribution, one point on a graph like Figure 3 represents the number of events > 4.0 Mw, including events > 4.5 Mw, whereas the point representing events > 4.5 Mw represents a smaller number of events. The highest magnitude point usually only represents one event, and it is unlikely to be the case that, if for instance it is the largest earthquake in a 500 year catalogue, it is exactly the event with a recurrence of 1 in 500 years. It may be that the 1 in 1,000 year event has occurred within the 500 years of observation, or it may be that the 1 in 500 year event has not occurred in the last 500 years (37% chance of this). Either way, as a data point it is quite unreliable as an indicator of recurrence, and allowing it to influence a least-squares regression would be very undesirable. A maximum likelihood estimator gives the correct importance to all points, because it represents the likelihood of observing the actual catalogue.

The method was refined by Weichert (1980) to handle varying levels of catalogue completeness. In the exposition of the stochastic approach, it was assumed that one would posit a period *n* for which the catalogue was complete above 4.0 Mw, and that only this window would be used. In many cases one might have data referring back to earlier earthquakes even though for the earlier period the catalogue is not complete for 4.0 Mw. The danger is that this information will simply be lost. It may be that while the catalogue is complete back to 1800 for earthquakes of 4.0 Mw, for earthquakes of 5.0 Mw is could be considered complete to 1700. In which case it is possible to divide the catalogue into two parts, > 5.0 Mw, 300 years long, and 4.0-49 Mw, 200 years long.

This works perfectly well with the stochastic method; the simulated catalogues are generated in such a way as to match the historical completeness situation, and the vectors compared as before.

In Weichert (1980) this is translated into maximum likelihood terms by sorting the catalogue into a series of magnitude intervals, assessing the period of complete reporting for each interval, and counting the number of events. The method can be extended to non-uniform intervals by specifying interval boundaries individually (Johnston et al 1994). For each magnitude interval, assuming equation (1) and a Poisson process, one can calculate the probability of observing the historical number of earthquakes, given any possible value of *a* and *b*. From this, one can compute the likelihood of observing the entire catalogue, for any values of *a* and *b*, rather as was done in the example of the stochastic example, but algorithmically rather than quasi-observationally. The equations are given in Weichert (1980) and Johnston et al (1994) and will not be repeated here. By maximising the log likelihood over *a* and *b* one can arrive at a best fitting solution.

This approach was subsequently developed further by Veneziano and van Dyke (1985) and Youngs (1992, *unpublished*). Two modifications were introduced. The first was intended to handle the special case of estimating recurrence parameters for very large regions, where the catalogue completeness actually varies from one part of the region to another. It involves dividing the region up into sub-zones with their own completeness windows, and then proceeding much as before, with the different sub-zones being handled in a similar way to the different magnitude-specific completeness periods. This had a specific application in Johnston et al (1994), but is probably not generally useful in PSHA, since SSZs are usually defined with a view to being homogeneous with respect to completeness anyway.

The second modification is much more significant, and handles cases where there are so few earthquakes within a SSZ that the *b* value is very poorly constrained. This is handled by introducing a prior estimate of *b* in the form of a penalty term for which a weight can be specified. The weight and the deviation of estimated *b* from prior *b* are then factored in to the likelihood function to produce the penalised likelihood function. This is maximised with respect to an expression that introduces the total number of events, with the effect that the fewer number of events on which the estimate is made, the greater the penalty for deviating from the prior. Again, for the precise equations, Johnston et al (1994) is a useful reference, and it is not necessary to repeat them here.

The net effect is that the higher the weight and the fewer the number of events, the more the resulting *b* value will be conditioned by the prior. The solution of the penalised maximum likelihood function is obtained iteratively for *b*, after which it is relatively straightforward to obtain the rate density.

A useful feature of this approach is that it can also compute the uncertainty on *a* and *b*, which can then be used for weighting purposes in a logic tree within the PSHA calculations. The uncertainty is obtained from the asymptotic covariance matrix for the coefficients, which is estimated by the inverse of the second derivative of the log likelihood function evaluated at the maximum likelihood estimates of *a* and *b*. This is described in more detail in Veneziano and van Dyke (1985).

3. Low seismicity cases

The penalised maximum likelihood method can still be used in cases where only very few earthquakes fall within the periods for which the catalogue is complete. Even if there is only one event, the absence of larger events within a known completeness period can be used. Obviously, in such cases one has inadequate control over the *b* value. However, so long as a prior is specified for *b*, it is possible to constrain the *b* value to an appropriate value.

However, there are cases of empty SSZs. One of the problems with attempting a maximum likelihood solution of such cases is that of infinite divisibility. Suppose an aseismic area is modelled as a single SSZ. Following some principle, one can propose a hypothetical activity rate. Now suppose that (perhaps on geological grounds) the zone is divided into two. The same calculation now applies to each new zone, with the net effect that double the original seismicity is modelled; the same values occur in each half of the original zone. One can continue this ad infinitum, increasing the seismicity every time one splits a zone into two parts. While one could argue that no-one would sensibly do this, it is a defect in the model that it is even possible to abuse the procedures in this way.

The second general problem is that of the hazard of the unknown. One can contrast, for instance, the situation of Ireland with that of an equal-sized area somewhere in the Atlantic.

Neither area has any known earthquakes. Is the offshore SSZ inherently more hazardous than the onshore SSZ? On the one hand, it could be argued that both are aseismic SSZs and should be treated the same way. On the other hand, it could be argued that the long-term seismicity of Ireland is known to be low, but the long-term behaviour of the offshore SSZ is unknown. Therefore, although there is no reason whatever to suppose the seismic activity of the offshore SSZ *is* higher than the onshore zone, there is no way to disprove that it *could* be higher. Therefore, taking uncertainty fully into account, the probability of earthquakes in the offshore zone is higher than that of the onshore zone, and therefore the hazard is higher as well. There are certainly seismologists who would take the latter view, and hazard maps exist in which contours indicate systematically increasing hazard with distance from land, for precisely this reason alone (e.g. EQE 2002).

But while from a strictly mathematical viewpoint, this is a realistic interpretation of probabilistic hazard, from a geophysical perspective it is less attractive. If seismic hazard is intended to represent in some way seismogenic processes, worldwide experience suggests that, in general low seismicity cases, hazard does not increase as one ventures further away from land. (Leaving aside passive margin events, which are a different issue.)

Consideration of these two problems suggests that a solution should have two characteristics: firstly that activity rates for asesimic SSZs should be dependent on area, and secondly, that they should not be dependent on completeness periods. (Though it is conceded that some might disagree with the second item.)

To facilitate investigation of these issues, Europe was divided into two as shown in Figure 4. Also shown on the map is seismicity ≥ 4.0 Mw since 1970. (Note that the southern zone is missing seismicity for Algeria and Tunisia.)

We can consider, as a first approximation, that the northern zone is broadly indicative of activity in low seismicity areas of Europe generally, including low seismicity areas in the Mediterranean. The northern zone in Figure 4 contains 137 events in 36 years, over an area of slightly more than 7.5 million sq km. (In contrast, the southern zone has 3,555 events in the same period, in about half the area). This gives an overall annual rate, for low seismicity areas, of 0.497 events per 10^6 sq km.

If one takes just the onshore area of Ireland, it has an area of 120,000 sq km. Therefore, if it experienced earthquakes at the average rate, it should have an annual rate of 0.06 events ≥ 4.0 Mw, corresponding to one event every 17 years. In practice, there have been no such events for at least 250 years. Therefore, an adequately conservative solution would be to rule that aseismic zones experience earthquakes ≥ 4.0 Mw at one-tenth the average rate for low-seismicity Europe, i.e. at a rate of 0.05 events per year per 10^6 sq km. One could argue that the area of the northern zone in Figure 4 should be reduced on account of the offshore area in the north-west, but given that one is working to an acceptable approximation, a value of 0.05 is adequate for current purposes, and is applied to all empty SSZs in Attic Ivy.

Obviously, no *b* value can be calculated in such cases, and it is necessary simply to assign a value equal to the chosen prior.

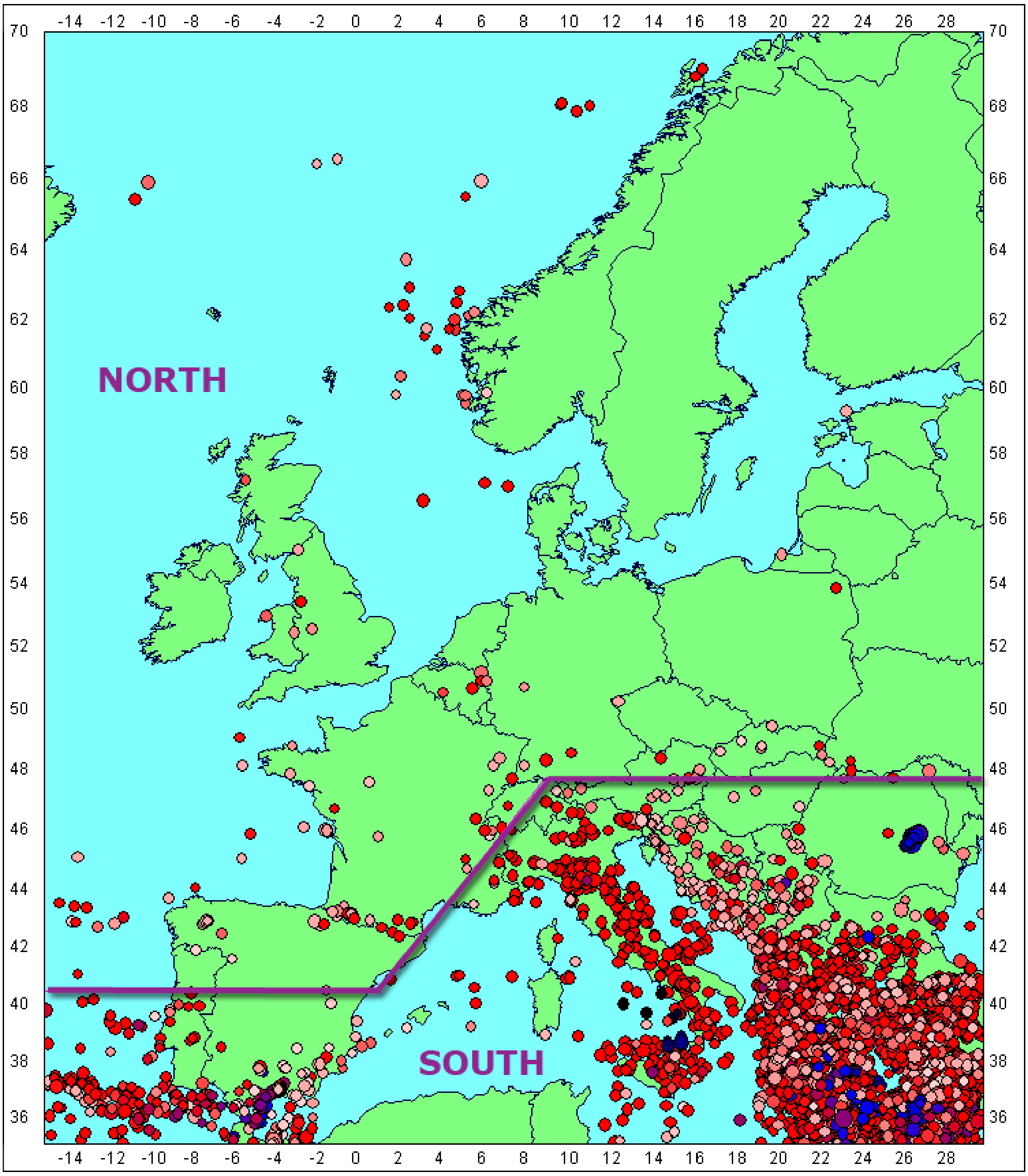

Figure 4 - Zones for analysis of post 1970 seismicity

4. Handling uncertainty in input parameters

It is often the practice in PSHA to treat the parameters of earthquakes in any catalogue as given, without taking into consideration that they may be poorly determined, and that this could influence the hazard calculations. One of the objectives in developing Attic Ivy for SHARE was to provide a way of fully representing parameter uncertainty in the recurrence calculations.

Two uncertainties are important: uncertainty in location, and uncertainty in magnitude. In addition, parameter uncertainty need not be Gaussian. It can take the form of discrete alternatives. This is particularly the case for historical earthquakes. If one takes some particular method of locating a historical earthquake, that method may return estimated uncertainty in the form of the epicentre being ± 20 km and the magnitude ± 0.4 (for example). But these could be further conditioned by some basic assumption, for instance, whether an earthquake occurred offshore or onshore. Assuming an offshore epicentre can give a completely different epicentral estimate, again with an error radius, and a different magnitude value (again with an uncertainty). If there is no good way to discriminate as to which option is the correct one, there is a real uncertainty that is best captured by including both options with suitable weights.

4.1 Uncertainty in epicentre

Uncertainty in epicentre has no impact in the case that whatever the true location of an event, it falls within the same seismic source zone. Where it becomes important is in cases where it is uncertain which source zone it falls in. Attributing the earthquake firmly to one zone is to ignore this issue. It is more accurate to allow the event to contribute to all zones that it could feasibly belong to, with a weighting corresponding to the probability that the event actually occurs in that zone.

This can be handled by adopting a bootstrapping approach. This involves repeating the maximum likelihood calculations a large number of times (e.g. 1,000 times). At each run, each earthquake is evaluated stochastically as to its "true" location with respect to that particular run. If the event has two weighted alternatives, one is selected randomly with a probability equal to the weight assigned to that alternative. A specific location is then assigned by perturbing the selected epicentre according to the Gaussian error radius. This location is then tested to see if it falls within the SSZ being evaluated.

The net effect is that each run will calculate *a* and *b* values on the basis of slightly different sets of earthquakes, each of which is a possible representation of the actual earthquakes that occurred within the SSZ over the observed catalogue.

The final values for *a* and *b* are then taken as the mean values of those calculated over the whole series of bootstrap runs.

This approach is an innovation within the current project. It is not known that any previous PSHA study has attempted to account for epicentre uncertainty.

4.2 Uncertainty in magnitude

The effect that uncertainty in magnitude has on activity rates, unlike epicentral uncertainty, has been discussed several times in the literature, notably Tinti and Mulargia (1985), Veneziano and van Dyke (1985), Rhoades (1996), Rhoades and Dowrick (2000), McGuire (2004) and Castellaro et al (2006). Unfortunately, some of these authors give contradictory views, and an attempt has been made to resolve the differences in Musson (2011).

The majority view, and according to Musson (2011) the correct view, is that uncertainty in magnitude has the effect of raising the apparent activity rate while leaving the *b* value unchanged. The reason is simply stated. The apparent magnitude of any earthquake may be larger or smaller than the true magnitude with equal probability. But, since magnitudes are distributed according to a power law, there are more small earthquakes than large ones. So the number of earthquakes with apparent magnitude of M will include more events with true magnitude $< M$ than events with true magnitude $> M$. So the number of earthquakes with apparent magnitude M is always an overestimate of the number of earthquakes with true magnitude M. Furthermore, the overestimation, δa, can be calculated (following Tinti and Mulargia 1985, Castellaro et al 2006) from

$$\delta a = (b^2 \sigma_m^2) / (2 \log_{10} e) \qquad (5)$$

where σ_m is the standard deviation of the magnitude uncertainty (assuming this is constant over the whole catalogue).

From this, it follows that one can calculate a correction factor m' from

$$m' = \delta a / b \qquad (6)$$

One then proceeds to subtract m' from each magnitude prior to performing the maximum likelihood calculation (Rhoades and Dowrick 2000).

There is an apparent impasse in that the magnitude correction factor is needed before the maximum likelihood calculations can be performed, but *b* needs to be known in order to compute the correction factor. Veneziano and van Dyke (1985) propose that the results are sufficiently insensitive to errors in *b* that it is sufficient to use a first approximation to *b*. Rhoades and Dowrick (2000) suggest using an iterative approach to converge to a final value, stating that three or four iterations are usually sufficient. The approach adopted in Attic Ivy is to perform two passes: the first calculates *b* on the unmodified data, and this first value is then used in the correction factor to compute the final results.

4.2.1 Testing the correction factor

It can be demonstrated that this method works. For the purposes of this exercise, four synthetic catalogues were constructed in such a way that the earthquakes were complete above magnitude 3.0 Mw and fitted perfectly a Gutenberg-Richter power law, so far as is possible with integer numbers of events. Problems of granularity of integer numbers were mitigated by using large catalogues containing over 12,000 events. The equation used as the basis for catalogue construction was the following:

$$\log N = 4.4 - 1.1\ Mw \qquad (7)$$

From this it follows that the expected rate of earthquakes of 4.0 Mw and greater is one per year. For ease of compatibility with existing software, each earthquake was given a random date between 1 January 1000 and 31 December 1999, a time of 1h 01m, a depth of 10 km, and a random epicentre within a 10 degree square. The maximum magnitude was set to be 6.5 Mw.

Catalogue #1 was generated exactly as above. Use of this catalogue presumes perfect knowledge of a perfect dataset.

Catalogue #2 was generating supposing a proxy measure of Mw, which can be called Mx. This can be taken to represent some other parameter, either a different magnitude scale or some macroseismic parameter. For convenience, the relationship between Mw and Mx is taken to be:

$$Mx = Mw \qquad (8)$$

The standard deviation of equation (8) was set to be 0.4. Thus Catalogue #2 can be considered a dataset comprising Mw magnitudes converted from Mx observations according to equation (8). However, since there is variability in the value of Mx for any actual Mw, there is in practice a uniform uncertainty of 0.4 in all the magnitude values in this catalogue. This was achieved by simply perturbing all the magnitude values randomly following a Gaussian distribution with standard deviation of 0.4.

Catalogue #3 was created in a similar way to Catalogue #2, except that it was assumed that the variance in equation (8) is time dependent. The values used are given in Table 1. Obviously, these are not intended to be realistic, and because of the way the catalogue is created, the effect is to have random but known errors throughout the catalogue.

Date range	Variance
1900-1999	0.2
1700-1899	0.3
1500-1699	0.4
1000-1499	0.5

Table 1 - Values used in constructing Catalogue #3

Catalogue #4 was created in a similar way to Catalogue #3, except that it was assumed that the variance in equation (8) is magnitude dependent. The following values were used:

Mw (true value)	Variance
≥ 5.0	0.2
4.5-4.9	0.3
4.0-4.4	0.4
<4.0	0.5

Table 2 - Values used in constructing Catalogue #4

The four catalogues were processed using Attic Ivy. In each case a single zone was used encompassing the whole area. A single Mmax value of 6.5 (the true value) was specified. In practice, Catalogue #1 is complete for all years above magnitude 3.0 Mw, but the maximum likelihood routine expects more than one magnitude band in order to calculate *b*. The input file therefore specified six magnitude ranges, from 4.0 to 6.5 at intervals of 0.5 of a magnitude unit, all with a completeness date of 1000.

Catalogues #2, #3 and #4 are effectively complete in terms of real Mw, but appear incomplete at low magnitudes in Mx, since, whereas in Catalogue #1 there is a sharp cutoff at 3.0 Mw, in the other catalogues some events with magnitude ≥ 3.0 Mw will scatter downwards to values < 3.0 Mx, but as events < 3.0 Mw are not treated, none will be perturbed upwards into values above 3.0 Mx. Since the magnitude range of interest is 4.0 and above, these issues at the lower end of the magnitude scale are unimportant, and in fact, the reason for generating events as small as 3.0 Mw was precisely to provide a buffer between the lower magnitude cutoff of the catalogue and the magnitude range of interest.

No prior for *b* was specified.

The results from the initial run of AtticIvy are as given in Table 3. Only the highest weighted values are shown. The activity rate is given as the annual number of earthquakes of 4.0 Mw (real or estimated).

Catalogue	a	b
#1	0.999	1.099
#2	1.616	1.068
#3	1.725	1.070
#4	2.017	1.096

Table 3 - Initial analyses of three synthetic catalogues

In reading Table 3, one should recall that the values for the activity rate and *b* value that underlie the catalogue construction are 1.0 and 1.10. The first point to take from Table 3 is that when the magnitudes are known precisely, the results from Attic Ivy are accurate to 0.001. When Mw is estimated from the surrogate parameter Mx, which has an uncertainty with respect to the true Mw, the *b* value is reasonably accurate but the activity rate is greatly overestimated, as expected from equation (5). In the case of Catalogue #4, the apparent frequency of events ≥ 4.0 Mw is double the true rate. It is clear that this is a serious issue.

When using the correction factor as described above, the results shown in Table 3 change to those shown in Table 4. Obviously the results for Catalogue #1 are unchanged, since this catalogue has no errors in it.

Catalogue	a	b
#1	0.999	1.099
#2	0.965	1.090
#3	1.008	1.106
#4	1.063	1.039

Table 4 - Corrected analyses of three synthetic catalogues

The most difficult case is Catalogue #4, since the magnitude-dependency of the error imparts a slight curvature to the Gutenberg-Richter plot. The *b* value estimate is the worst of the four, but the error is only 0.06, and the activity rate is quite accurate. Certainly, Table 4 is much to be preferred to Table 3. It is also worth noting that when the uncertainty is date-dependent, which is likely to be more common than magnitude-dependence, the correct values are retrieved very well. Also, this is only the results of a single perturbation. If one repeated the exercise many times, one could expect to see other variations, with the mean values approximating to the true values (Rhoades 1996). The single-run exercise here is useful because it is analogous to the situation in real life where one is dealing with one catalogue only.

One practical consideration was discovered, related to the fact that the magnitude data in the catalogues is arranged in discrete steps of 0.1 units, whereas the correction factors computed from equation (6) have no particular resolution. This led to edge effects, for instance, when an earthquake of magnitude 4.2 was corrected to 3.9845, less than 4.0. This was obviated by rounding corrected magnitude values to one decimal place, so 3.9845 was treated as 4.0. Not making this correction results in significant underestimates of the activity rate, and also, less accurate *b* values.

4.2.2 Non-Gaussian magnitude uncertainty

In the foregoing discussion, as in all the literature cited, it is assumed that uncertainty in magnitude is Gaussian. As remarked at the outset of this section, there are cases in historical seismicity where one may have two competing interpretations of the same earthquake, with significantly different epicentres and magnitudes.

In such cases, the discrete uncertainty is handled by Attic Ivy using the bootstrapping approach described in section 4.1. The magnitude correction factor is then applied to the selected magnitude value at each bootstrap run.

5. Running Attic Ivy

To analyse a SSZ model, two input files are needed. One contains the actual SSZ information, and by default has the extension .inp. The other contains the earthquake catalogue, and has the default extension .dat. The program is written so that if these extensions are followed, it is not necessary to type the extension when prompted for a file name. So if the catalogue file is called Sheec.dat, it is sufficient to type Sheec. The program also asks for the number of bootstraps runs. If this feature is not needed, type 1. If no number is entered, the default of 1000 is used. If there are no multiple determinations of events, or no epicentral uncertainties, there is no point in using bootstrapping, so 1 is sufficient.

The input files will now be described in turn, followed by an explanation of the output files that are generated.

5.1 The catalogue file

The catalogue input uses an extension of WIZMAP format (Musson 1998). This allows flexibility in file format, since the various parameters do not need to appear in a given order or follow a predetermined column width. The placing of parameters is conveyed to the program using a single file header line.

The start of a file might look like this:

```
YYYY  MM  DD  HH  IISSSSS PPPPPP LLLLLLL   KKK   RRR  AA  BB  WWWW  FFF     EE
1000   3  29   0   0   0   50.18    4.24         3.7   1   2  0.65  0.40    20
1000   3  29   0   0   0   49.50    4.00         4.2   2   2  0.35  0.40    20
1005   1   1   0   0   0   43.46   11.88         5.2   1   1  1.00  0.40    20
1005   1   1   0   0   0   41.49   13.83         5.2   1   1  1.00  0.40    20
1010   3   9   0   0   0   40.80   28.80         5.5   1   1  1.00  0.40    15
1013  11  18   0   0   0   50.65    5.58         3.5   1   1  1.00  0.40    20
1014   1   1   0   0   0   45.65    0.15         5.4   1   1  1.00  0.40    20
1019   4   1   0   0   0   41.13   14.78         4.7   1   1  1.00  0.40    20
```

The letter codes in the header line are effectively labels of the columns below. The first earthquake in the list has two possible sets of parameters. The column marked BB contains the total number of alternative interpretations (two for the first event, one for all the others), and the AA column gives the number of this particular alternative, 1 for the first, 2 for the second, etc. The maximum number of interpretations allowed for one earthquake is three. The column marked WWWW is the weight assigned to this interpretation.

The other columns are labelled as follows:

YYYY	Year
MM	Month
DD	Day
HH	Hour
II	Minute
SSSSS	Second
PPPPPP	Latitude (phi) in decimal degrees, south negative
LLLLLLL	Longitude (lambda) in decimal degrees, west negative
KKK	Depth (kilometres)
RRR	Magnitude ("Richter")
FFF	Magnitude uncertainty
EE	Epicentre uncertainty, in km

The last two values may be somewhat subjective, though there are ways of evaluating them (e.g. Bakun and Wentworth 1997, Musson and Jiménez 2008). It would be reasonable to suggest, in the case of magnitude uncertainty, that it would be inconceivable for this to be less than 0.2 degrees in any circumstances.

5.2 The SSZ file

The input file for the SSZ model needs three types of information for each SSZ. Firstly, the geographical co-ordinates. Secondly, the maximum magnitude distribution. Thirdly, the magnitude completeness ranges. In addition, it is necessary to specify, for the whole model, what is the base magnitude with respect to which values of the activity rate will be calculated, i.e. the value for m0 in equation (2).

The start of the input file could look as follows:

```
Mmin.......:4.0
# zones....:21
Cornwall  ,   4
  49.850 ,   -6.000
  50.280 ,   -3.360
  51.370 ,   -4.880
  50.200 ,   -6.000
# Mmax.....: 1
 6.5   1.0
# Periods..: 5
 3.5 1970
 4.0 1810
 4.5 1765
 5.0 1700
 7.0 1500
A prior and weight
 0.0   0.0
B prior and weight
 1.0  25.0
```

The first line, Mmin, gives the base magnitude for the activity rates, which will be expressed as the number of events ≥ Mmin per year. This need not be the same as the lower bound magnitude in the hazard calculations, nor the lowest magnitudes in the catalogue. It is the same as m0 in equation (2).

Then follows the number of SSZs in the file. In this example, only the data for the first SSZ is shown; the same data blocks will be repeated for the other SSZs in the model.

The first line for each zone starts with an identifying code so that the output can be clearly labelled (and the input can be checked more easily) - it helps if the identifier is meaningful and not simply "Zone 13". After this comes the number of vertices defining the SSZ. The next lines contain the vertices, in latitude/longitude pairs, decimal degrees, south/west negative.

The next section contains the maximum magnitude distribution. Several values can be specified, one to a line, with the first line of the block specifying how many values there are, and each line giving a magnitude and a weight. However, in practice, because the correlation between Mmax and the other parameter is rather weak, and because specifying activity rate/*b* value distributions for each Mmax would greatly increase the number of branches in a PSHA logic tree, only the largest Mmax value is actually used, so this is all that is needed to be specified. The functionality for calculating activity rates for a suite of Mmax values is retained in the program code; it is only necessary to amend the output routine if this extra data becomes wanted at some time in the future.

The next section is for magnitude-completeness windows. In this example, five periods are identified. Data for magnitude 3.5 Mw are deemed to be complete since 1970, for 4.0 Mw since 1810, and so on. It is not necessary to specify only the dates where the completeness changes; indeed, it is often disadvantageous to do so. For instance, if one was using a catalogue that only contained modern instrumental data, starting in 1965, and was complete for all magnitudes above 4.0 Mw, then it would be appropriate to write something like this:

```
# Periods..: 5
 4.0 1965
 4.5 1965
 5.0 1965
 6.0 1965
 7.0 1965
```

This is because the *b* value is calculated on the basis of the magnitude values given in this section. So the more values that are given, the more points are used in the slope-fitting computation.

Note also that the largest magnitude for which a completeness is specified *must* be equal to or larger than the largest Mmax value specified in the previous block.

The final section is for the priors to be used in the calculation for this zone. It is possible to specify a prior for the activity rate *a* as well as the *b* value. This is included largely as a legacy feature, and it is not recommended that it be used.

Where a prior is not used, the input is 0.0 0.0. In the example shown, a *b* value of 1.0 is specified (this should not be written -1.0) with a weight of 25.0. Weights can vary from 0.0

(no prior) to 100.0 (forcing). It is suggested by Johnston et al (1994) that the weight be inverse to the variance of the prior value; in practice it is a subjective decision.

5.3 Output files

When Attic Ivy runs, it writes to screen some values pertaining to each SSZ as it processes it. In particular, it gives the number of earthquakes found within each specified completeness window. If multiple bootstrap runs have been used, it is possible that in some iterations of a zone it will be empty, and in others not. The number of runs for each zone that did not result in the zone being empty is printed.

If a very large model is being run, the zone information will scroll off the screen, but it shows that the program is still running, or, in the case of an error, the last SSZ to be processed.

The program creates two output files, the names of which are derived automatically from the name of the input file. If the input file is called Share.inp, the output files will be Share_out.txt and Share_short.txt.

The first of these is the main output file, supplying the input for PSHA. It will look like this:

```
Cornwall

   25
     0.007    0.0011    0.606
     0.031    0.0009    0.803
     0.057    0.0007    1.000
     0.041    0.0006    1.197
     0.012    0.0004    1.394
     0.012    0.0032    0.606
     0.060    0.0025    0.803
     0.117    0.0020    1.000
     0.088    0.0016    1.197
     0.026    0.0013    1.394
     0.010    0.0053    0.606
     0.059    0.0042    0.803
     0.126    0.0034    1.000
     0.101    0.0027    1.197
     0.032    0.0021    1.394
     0.003    0.0152    0.606
     0.023    0.0121    0.803
     0.061    0.0096    1.000
     0.060    0.0077    1.197
     0.022    0.0061    1.394
     0.000    0.0251    0.606
     0.005    0.0200    0.803
     0.017    0.0159    1.000
     0.020    0.0127    1.197
```

```
0.009    0.0101    1.394
```

This block is repeated once for each SSZ, identified by the SSZ name (in this case, Cornwall). The number 25 is the count of the following lines, which consist of triplets of logic-tree weight, activity rate and *b* value. The reason for 25 is that the optimal values are calculated, together with plus and minus one and two standard deviations in each of the two parameters. The best-fit values are the highest weighted pair, which will always be the central value of the block; in this case 0.0034 and 1.000.

This is different for empty SSZs, where only one triplet will be printed instead of 25.

The other output file contains summary data for each SSZ, one line per SSZ, suitable for mapping. The information for the SSZ above would appear as:

```
Cornwall     50.67    -4.96     1   0.0034  1.000    9.877236
```

The columns here are: the SSZ identifier, the latitude and longitude of the central point of the zone (approximated by the mean of the highest and lowest latitude and longitude values of the vertices), the total number of earthquakes in the SSZ falling within the completeness periods (in this case only one), the activity rate (number of events per year) and the *b* value. The last number is the area-adjusted seismic moment release, in Newton metres, expressed as the log of the annual rate per square kilometre.

6. References

AKI, K. 1965. Maximum likelihood estimation of b in the formula log N = a - bM and its confidence limits. *Bulletin of the Earthquake Research Institute, Tokyo University*, Vol. 43, 237-239.

BAKUN, W H, and WENTWORTH, C M. 1997. Estimating earthquake location and magnitude from seismic intensity data. *Bulletin of the Seismological Society of America*, Vol. 87, 1502-1521.

CASTELLARO, S, MULARGIA, F, and KAGAN, Y Y. 2006. Regression problems for magnitudes. *Geophysical Journal International*, Vol. 165, 913-930.

EQE INTERNATIONAL LTD. 2002. Seismic hazard: UK continental shelf. *Health and Safety Executive Offshore Technology Report 2002/005*.

JOHNSTON, A C, COPPERSMITH, K J, KANTER, L R, and CORNELL, C A. 1994. The earthquakes of stable continental regions. *Electric Power Research Institute*, TR-102261-V4.

MCGUIRE, R K. 2004. *Seismic hazard and risk analysis*. (Oakland CA: Earthquake Engineering Research Institute.)

MUSSON, R M W. 1998. A self-parsing file format for earthquake catalogue and data files. *Seismological Research Letters*, Vol. 69, 248-250.

MUSSON, R M W. 2004. Joint solution of seismicity parameters for seismic source zones through simulation. *Bolletino di Geofisica Teorica ed Applicata*, Vol. 45, 1-13.

MUSSON, R M W. 2011. The effect of magnitude uncertainty on activity rates. *Bulletin of the Seismological Society of America*, Submitted.

MUSSON, R M W, and JIMÉNEZ, M-J. 2008. Macroseismic estimation of earthquake parameters. *NERIES project report, Module NA4, Deliverable D3*.

MUSSON, R M W, SELLAMI, S, and BRÜSTLE, W. 2009. Preparing a seismic hazard model for Switzerland: The view from PEGASOS Expert Group 3 (EG1c). *Swiss Journal of Geology*, Vol. 102, 107-120.

RHOADES, D A. 1996. Estimation of the Gutenberg-Richter relation allowing for individual earthquake magnitude uncertainties. *Tectonophysics*, Vol. 258, 71-83.

RHOADES, D A, and DOWRICK, D J. 2000. Effects of magnitude uncertainties on seismic hazard estimates. *Proceedings of the 12th World Conference on Earthquake Engineering*, Wellington, NZ, Paper 1179.

TERA CORPORATION. 1980. Seismic hazard analysis: A methodology for the Easrtern United States. NUREG/CR-1582.

TINTI, S, and MULARGIA, F. 1985. Effects of magnitude uncertainties on estimating the parameters in the Gutenberg-Richter frequency-magnitude law. *Bulletin of the Seismological Society of America*, Vol. 75, 1681-1697.

WEICHERT, D H. 1980. Estimation of the earthquake recurrence parameters for unequal observation periods for different magnitudes. *Bulletin of the Seismological Society of America*, Vol. 70, 1337-1346.

VENEZIANO, D, and VAN DYKE, J. 1985. Seismic parameter estimation methods. Appendix A in *Seismic Hazard Methodology for Nuclear Facilities in the Eastern United States (Draft 85-1)*. EPRI/SOG (editor). 2.