Geophys. J. Int. (2000) **142,** 37–51

On entropy and clustering in earthquake hypocentre distributions

T. Nicholson, M. Sambridge and Ó. Gudmundsson*

Research School of Earth Sciences, Australian National University, Canberra, ACT 0200, *Australia. E-mails: todd@rses.anu.edu.au; malcolm@rses.anu.edu.au*

Accepted 2000 January 12. Received 2000 January 6; in original form 1999 May 28

SUMMARY

The degree of clustering or disorder within earthquake distributions may be measured using the concept of entropy. A method for calculating the entropy of any 3-D point set (e.g. earthquake foci) is presented. This makes use of Voronoi cells (convex polyhedra representing nearest neighbour regions) to measure point density in three dimensions. An estimate of event density can be determined directly from the size of Voronoi cells. Normalizations are introduced to the definition of entropy that allow data sets containing different numbers of events and occupying different volumes to be compared quantitatively, for example, earthquake catalogues from different tectonic regimes. Our results show a clear correlation between earthquake entropy and tectonic regime. The most ordered are the mid-ocean ridges, followed by the subduction zones and finally intraplate seismicity.

We show how entropy may be used to quantify the simplification of earthquake distributions, for example, due to relocation procedures. A recently published algorithm called the collapsing method is used as an example of a technique that reduces entropy while respecting data fit. Modifications to this method are made that reduce artefacts and use additional temporal information in the earthquake distribution. These methods are applied to a global catalogue of 85 000 events, and a local catalogue from the SIL network in Iceland containing 43 300 events. The entropy of both catalogues is reduced. Results from the Hengill region within the SIL network show lineations whose orientations agree with independent studies using relative location techniques and surface faulting.

Key words: earthquake location, entropy, seismicity.

1 INTRODUCTION

Earthquake distributions have been used to infer a great deal about Earth processes and structure, for example, the delineation of plate boundaries and the distribution of faults in the upper crust. Earthquake distributions often fall into clusters, varying in size from small scale such as earthquake swarms in volcanic areas to large scale along plate boundaries. The size and shape of these clusters provide valuable constraints on the Earth processes involved; however, the amount of clustering varies significantly, both geographically and over time.

It is clear that any underlying pattern in the distribution of earthquakes in a particular region will always be blurred by hypocentral errors. While a great deal of effort has been put into reducing the effect of both random and systematic errors (see Buland 1976 and Billings *et al.* 1994 for reviews), uncertainty in the individual locations makes the examination of clusters difficult and has lead to most interpretations being of a qualitative nature.

Many of the patterns that appear within earthquake clusters are very ordered features. For example, the aftershocks of a large event may fall on the same fault plane. However, the random component of location is likely to cause an earthquake distribution to be more disperse and more disordered than if the locations were perfect. It would, therefore, be useful to have a quantitative measure of the degree of clustering within a distribution. Such a measure might allow us to compare and contrast the distribution of seismicity from one region to another. To this end, we propose to use the statistical quantity known as entropy as a measure of scatter in seismicity.

Entropy was first used as a scientific term in thermodynamics by Clausius (1850), who coined it from the Greek word 'trope' meaning to transform. The probabilistic interpretation in statistical mechanics is, however, attributed to Boltzmann (1871). Planck (1906) later recorded the explicit relationship between probability and entropy. The concept was

* Now at: Danish Lithospheric Centre, Øster Voldgade 20, 1350 Copenhagen K, Denmark. E-mail: og@dlc.ku.dk

used by Shannon (1948) to give an economical description of the properties of long sequences of symbols. He applied the concept to a number of basic problems in coding theory and data transmission.

In studies of earthquake distributions in a regional or local setting, it is common to relocate each hypocentre, for example, using regional refinements to global velocity models Smith & Ekstrom (1996) and Gudmundsson & Sambridge (1998). Other approaches include joint-hypocentre determination (Douglas 1967; Dewey 1972) and techniques involving master events (Evernden 1969; Peppin *et al.* 1989). One possible use of entropy in this context is to compare the effects of two location methods, or location with different earth models. By comparing the entropy of a distribution before and after relocation, we can obtain a quantitative measure of the difference and hence the effect of the method (or earth model) on the clustering of events.

Another quantity that has been used to measure earthquake clustering is fractal dimension. Hirata (1989) investigated the fractal dimension of fault systems in Japan, while Okuba & Aki (1987) and Aviles *et al.* (1987) conducted similar studies on the San Andreas Fault region. It has been noted by Aviles *et al.* (1987) that fractal dimension is sometimes difficult to measure because of the underlying assumption of self-similarity. Another difficulty arises from the limited resolvable range of spatial scales over which many earthquake clusters occur. It is important to keep in mind that neither entropy nor fractal dimension completely characterizes a distribution. They do, however, provide potentially useful indications about the nature of a distribution and allow quantitative comparisons to be made.

In this paper, we show how entropy may be used to measure the disorder of a distribution of earthquakes. We make use of the Voronoi diagram (Voronoi 1908; Sambridge *et al.* 1995) to measure the earthquake density within a region. This is well suited to the task because it provides a unique estimate of density at any point in a medium without the introduction of any arbitrary smoothing parameters. As well as being a measure of structural detail, entropy may also be used to delineate structure in the distribution. This is a new concept (introduced in this paper) in which earthquakes may be jointly relocated by simultaneously fitting the observed data and minimizing the overall entropy. The objective is to seek the simplest distribution of events consistent with the data, where complexity is measured by entropy. We discuss the details and difficulties in using entropy in this way. We show that a reduction in entropy of both global and local seismicity corresponds to a simplification of event distributions. This is achieved using the 'collapsing method' of Jones & Stewart (1997), to which we make some improvements.

2 THE ENTROPY OF A POINT SET IN THREE DIMENSIONS

2.1 Definitions and concepts

Here we propose a new definition and procedure for calculating the entropy of a point set, e.g. a cloud of earthquakes. Papoulis (1984) defines the entropy of a point set as

$$S = -\int_V P(\mathbf{x}) \log (P(\mathbf{x}))\, d\mathbf{x}, \tag{1}$$

where P is the probability density, $\mathbf{x}$ is a position vector in the domain and log is the natural logarithm (which is our convention throughout this paper). Skilling (1989) generalized this definition by incorporating a reference distribution, m,

$$S = \int_V P(\mathbf{x})\, d\mathbf{x} - \int_V m(\mathbf{x})\, d\mathbf{x} - \int_V P(\mathbf{x}) \log\left(\frac{P(\mathbf{x})}{m(\mathbf{x})}\right) d\mathbf{x}. \tag{2}$$

If P and m are normalized probability densities (which we shall assume) the first two integrals in eq. (2) are unity and cancel, so we are left with

$$S = -\int_V P(\mathbf{x}) \log\left(\frac{P(\mathbf{x})}{m(\mathbf{x})}\right) d\mathbf{x}. \tag{3}$$

However, for a point set (or hypocentral cloud) we can write

$$P(\mathbf{x}) = \frac{\rho(\mathbf{x})}{N}, \tag{4}$$

where N is the number of points and $\rho(\mathbf{x})$ is the density of the point set, which must be measured in some way. Substituting eq. (4) into eq. (3) and rearranging the logarithm term, we obtain

$$S = \log (N) - \frac{1}{N}\int_V \rho(\mathbf{x}) \log (\rho(\mathbf{x}))\, d\mathbf{x} + \frac{1}{N}\int_V \rho(\mathbf{x}) \log (m(\mathbf{x}))\, d\mathbf{x}. \tag{5}$$

The estimation of density from a finite set of arbitrarily distributed points is an underdetermined problem, which has been much studied in statistics (e.g. Silverman 1986). In order to estimate 'density' between the events, a smoothing criterion must be applied. One method to obtain a measure of density is simply dividing up the domain into boxes and counting the number of nodes in each box. Another takes the smallest sphere around each node that contains at least a certain number of points. These two methods, and most others, have the disadvantage that an arbitrary choice of smoothing parameter is required. In the case of the box-counting method the smoothing parameter is the size of the box, and for the sphere method it is the number of nodes in each sphere. The smoothing parameter may have a strong influence on the density estimate. For example, if we use the box-counting method and make the boxes too big we will experience a loss of resolution in measuring variability of density. If the boxes are too small there will be a large increase in density error. This is analogous to the well-known trade-off between resolution and variance in linear inverse problems (Backus & Gilbert 1968).

2.2 The use of Voronoi cells to estimate density and entropy

To avoid the need for an arbitrary smoothing parameter in determining the density, we invoke the concept of Voronoi diagrams, which were originally developed by Dirichlet (1850) and Voronoi (1908). A Voronoi diagram is a partition of the domain. Fig. 1 shows an example of a Voronoi diagram, generated from the earthquake distribution in Fig. 8(a). Each Voronoi cell contains one event in the distribution (we often refer to these as nodes). The Voronoi diagram is unique because any point within a cell must be closer to that cell's node than any other node. Hence, Voronoi cells are 'nearest neighbour' regions (see Fig. 1). Nodes that have cells that are in contact are called 'natural neighbours'. The nodes around the exterior

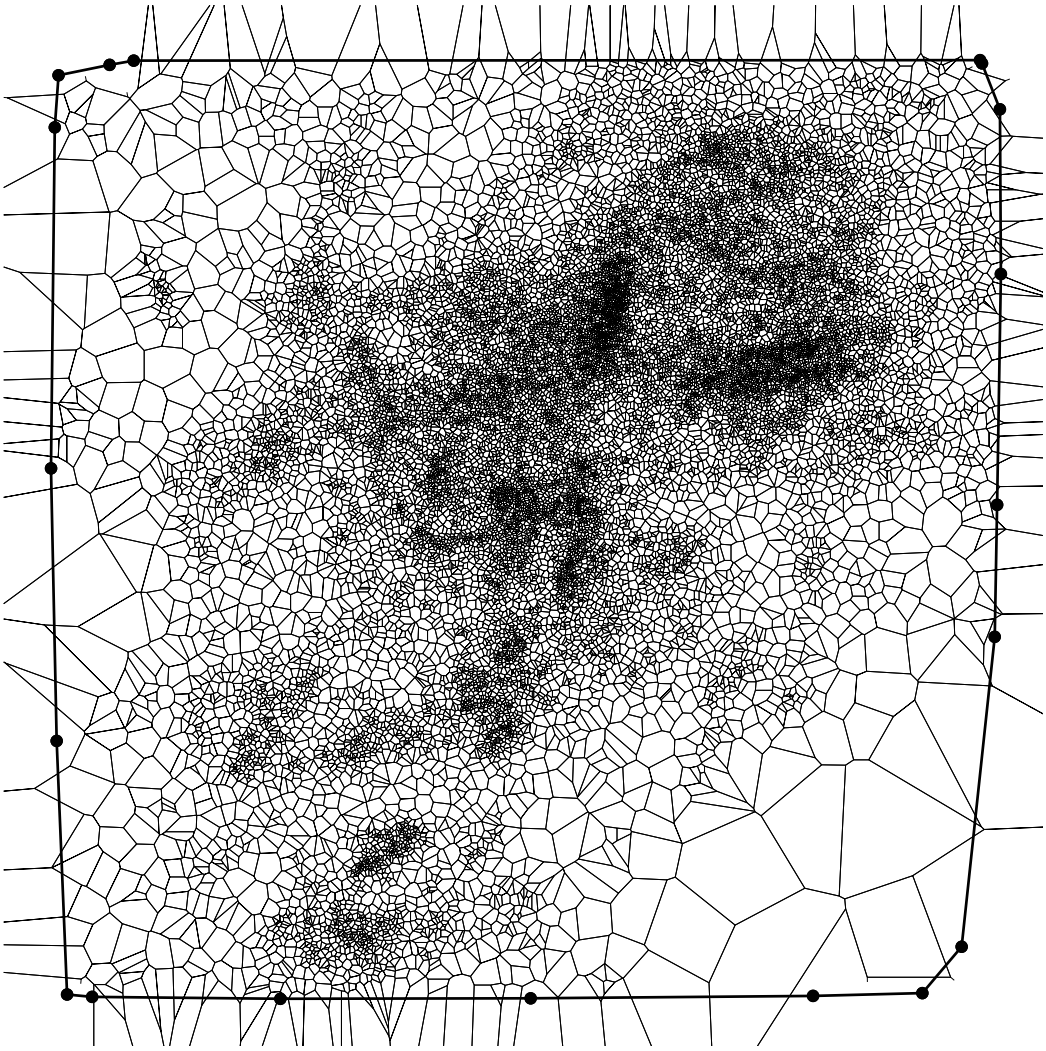

Figure 1. The Voronoi diagram for the earthquake distribution shown in Fig. 8. Note that the larger Voronoi cells in the southeast corner and along the western side reflect the low event density in these regions. Conversely, the Voronoi cells in the centre are much smaller, indicating a higher event density. The thicker line is the convex hull and the dots are points on the convex hull.

of the distribution have unbounded Voronoi cells and form the 'convex hull' of the nodes. The only choice required in generating a Voronoi diagram is the form of the distance norm (we use Euclidean distance throughout). The density of the Voronoi cells reflects the density of the nodal distribution (Okabe *et al.* 1992). In the centre of Fig. 1 the Voronoi cells are very densely distributed; however, to the west lies an area of low seismicity and the cells are much larger. A complete description of the properties of Voronoi cells is provided in Okabe *et al.* (1992).

A Delaunay triangulation (Delaunay 1934) is formed by connecting the nodes whose Voronoi cells have common boundaries (i.e. the natural neighbours). Over the past 20 years a number of algorithms have been developed to calculate the Delaunay triangulation, and from this the details of the Voronoi diagram can be determined. Popular Delaunay generation methods include an 'edge flipping' algorithm (see Fortune 1992) for the 2-D case, and the quickhull algorithm of Barber & Huhdanpaa (1994) for the 3-D case.

By definition the size of the Voronoi cell about each earthquake is a reflection of the 'closeness' to neighbouring events. An estimate of the density of the earthquake distribution can be found at an arbitrary point $\mathbf{x}$ using

$$\rho(\mathbf{x}) = \frac{1}{v(\mathbf{x}_i)}\,, \tag{6}$$

where $v(\mathbf{x}_i)$ is the volume of the Voronoi cell about the ith event, $\mathbf{x}_i$ is the position vector of the ith event and $\mathbf{x}$ is assumed to lie inside Voronoi cell i, that is, $\mathbf{x}_i$ is the event closest to point $\mathbf{x}$. Eq. (6) results in a simple, unique measure of density with spatial variability controlled directly by the distribution of earthquakes. Note that it requires no arbitrary smoothing parameter to be introduced, unlike the box-counting method. The Voronoi cell of any node on the convex hull has infinite volume so the density estimate becomes zero. However, we choose to limit the volume of these cells to the portion that lies inside the convex hull, hence all the Voronoi cells have finite volume and their sum is equal to the volume inside the convex hull.

Calculating the volumes of the Voronoi cells is a simple procedure once the natural neighbours have been determined for each Voronoi cell. This is because they are just convex

polyhedra. Watson (1992) gives a method for determining the natural neighbour list from the Delaunay tetrahedra. A suitable method for calculating the volumes of the Voronoi cells can be found in Braun & Sambridge (1995), which uses the approach of Lasserre (1983) to calculate the volumes of convex polygons.

Since the Voronoi cells cover the domain and the density within each Voronoi cell is assumed to be constant, the integral in (5) may be replaced with a sum over the nodes (in doing so $d\mathbf{x}_i$ becomes v_i). Therefore, eq. (5) becomes

$$S = \log(N) - \frac{1}{N}\sum_{i=1}^{N} \rho_i(\mathbf{x}) \log\left(\frac{\rho_i(\mathbf{x})}{m(\mathbf{x})}\right) v_i \,. \tag{7}$$

Using eq. (6), we have

$$S = \log(N) + \frac{1}{N}\sum_{i=1}^{N} \log(v_i) + \frac{1}{N}\sum_{i=1}^{N} \log(m_i(\mathbf{x})) \,. \tag{8}$$

This amounts to taking the event density to be constant in each of the irregularly distributed convex polyhedra in Fig. 1. Note that if the probability density $P(\mathbf{x})$ were sampled uniformly over the domain, for example, the case of a 2-D image (Skilling & Bryan 1984), then a similar equation would be obtained that will be more familiar to users of maximum entropy techniques (see e.g. Skilling & Bryan 1984; Gull & Skilling 1983):

$$S = \log(N) - \frac{\Delta\mathbf{x}}{N}\sum_{i=1}^{N} \rho_i(\mathbf{x}) \log(\rho_i(\mathbf{x})) + \frac{\Delta\mathbf{x}}{N}\sum_{i=1}^{N} \rho_i(\mathbf{x}) \log(m_i(\mathbf{x})) \,, \tag{9}$$

where $\Delta\mathbf{x}$ is the pixel area and is effectively a scaling constant.

To make eq. (6) useful we still need to choose a reference distribution (i.e. the m_is). A natural choice is to make the reference distribution uniform and structureless. With this choice a uniform distribution has zero entropy and is the maximum value that can be obtained (all other distributions have negative entropy). With this choice we have

$$m_i = \frac{1}{V_0} \,, \tag{10}$$

where V_0 is the size of the domain. Using (10) in (6) we obtain

$$S = \log(N) - \log(V_0) + \frac{1}{N}\sum_{i=1}^{N} \log(v_i) \,. \tag{11}$$

It is clear from eq. (11) and the properties of the natural logarithm function that S is always non-positive. The value used for the size of the domain, V_0, is the volume inside the convex hull. Notice that if all the volume terms in eq. (11) are multiplied by a scale factor ϵ then all the terms in ϵ cancel and the expression is unchanged. Therefore, the entropy expression in eq. (11) is scale-independent. Notice also that eq. (11) is independent of the number of events used to sample the underlying structure. Appendix A contains some examples of entropy determined from simple distributions.

Eq. (11) now gives a definition of entropy that is normalized by both the number of events, N, and the total volume, V_0, and may be used to compare directly distributions containing differing numbers of events and spanning volumes of vastly different sizes. A caveat here is that in order to make use of (11) a distribution must contain a sufficient number of events to ensure that the volume of the Voronoi cells can be robustly determined, for example, not all are on the convex hull. Once this is assured then we can make direct comparisons between two clusters, whether they are related or not. Possible uses of entropy are to compare the degree of clustering for only the well-constrained locations in a data set with that of all the locations within a data set. Another potential application would be to compare entropy within a region over two different time periods to measure an evolution over time, independent of the number of events. The time periods could be of completely different lengths without affecting the result. We may even compare two distributions that differ by orders of magnitude in spatial scale. In this way the difference in character between local or regional seismicity and global seismicity could be analysed. In the next section, we consider an application of earthquake entropy in detail. The entropy values of samples taken from a global earthquake data set are calculated and the results interpreted to illustrate the usefulness of entropy in the analysis of earthquake clusters.

3 EXAMPLE OF APPLICATION OF THE ENTROPY CALCULATION

Our calculation of the entropy of a set of earthquake locations involves four steps.

(1) A coordinate transformation is performed to convert the earthquake locations from their standard spherical polar coordinates (latitude, longitude, depth) into a Cartesian coordinate system.

(2) The quickhull algorithm of Barber *et al.* (1996) is used to calculate the Delaunay triangulation of the set of 3-D nodes.

(3) The volume of each Voronoi cell is calculated from the list of natural neighbours determined in step 2. Each volume is calculated from formulae for the volume of convex polyhedra.

(4) The volumes are substituted into eq. (11)

The time taken to calculate each step is proportional to the number of events and step 3 is also dependent on the number of natural neighbours of each event (that is, the number of sides of each Voronoi cell). This calculation is practical on a desktop workstation [for example, we calculated the Voronoi cell volumes of 10 000 globally distributed events on an Ultrasparc 5 (270 MHz, 128 Mbyte RAM) using a total of 3 min CPU time.

The quickhull algorithm can have numerical difficulties calculating the list of natural neighbours in certain pathological cases, for instance, when four neighbouring nodes lie on a plane. In this case, the Voronoi diagram is unique, but the Delaunay triangulation is not. These problems can be easily overcome, for example, by adding small-amplitude noise to the distribution (see Barber & Huhdanpaa 1994; Barber *et al.* 1996).

To illustrate the application of this process we applied it to 10 subsets of the EHB (Engdahl *et al.* 1998) global earthquake catalogue. The results are shown in Table 1 and Fig. 2. Each of the subsets of events was chosen to demonstrate the range of entropy values that can occur for different tectonic regimes. The first three are sections of mid-ocean ridges from the Atlantic and Eastern Pacific. The next four are subduction zones from the Pacific and South America, and the final three are zones of continental intraplate seismicity in Tibet, Africa and the Western United States.

Table 1. Entropy values of earthquake distributions in 10 seismically active regions.

	Region	Longitude	Latitude	No. of events	No. on hull	Entropy
1	North Atlantic	50W–20W	20S–60N	1229	93	−4.26
2	South Atlantic	20W–0	60S–10N	608	99	−3.88
3	East Pacific Rise	130W–85W	60S–0N	586	97	−4.10
4	Aleutians	170E–145W	40N–65N	6701	61	−3.34
5	Marianas–Japan–Kuriles	135E–165E	10N–60N	16 100	65	−3.84
6	Tonga–Kermadec	175E–170W	40S–10S	10 367	58	−2.73
7	Andes	78W–60W	40S–0	4525	52	−3.16
8	Tibet	80E–110E	30N–60N	882	64	−1.91
9	African Rift	20E–45E	25S–10N	221	54	−1.75
10	Western USA	125W–100W	30N–60N	265	42	−2.78

The mid-ocean ridges have the lowest entropy (−3.88 to −4.26), followed by the subduction zones (−2.73 to −3.84) and the continental regions (−1.75 to −2.78). The mean of each set seems to be well separated. These values represent both the disorder within the actual earthquake distribution (that is, the distribution we would get if we could locate earthquakes perfectly) and the disorder introduced by location errors. We would expect that the actual earthquake distributions for the subduction zones and mid-ocean ridges would be relatively ordered and, therefore, that these two groups should have lower entropy. On the other hand, the African rift and Tibet regions have much higher entropy due to their more complicated and disordered nature.

The order of the entropy ranges for the three tectonic regimes can be explained by considering their underlying structure. Mid-ocean ridges are the simplest and most ordered since they consist of a series of zigzagging vertical planes; it is reasonable, therefore, that they should have the lowest entropy. Subduction zones are also relatively well-ordered features, but deformation (for example, departures away from a simple dipping plane) from one end of a region to the other seems to cause them to have higher entropy. The difficulty in determining the depth of events could also have a greater effect on subduction zones since they tend to be sloped. Any depth error in an event from a subduction zone will cause it to move off the 'plane' of subduction, whereas for mid-ocean ridge events, moving the event up or down will not move it off the plane of faulting (indeed, it is not unusual to constrain these events to lie at depths of not less than 10 km). Intraplate regions have the highest entropy since their seismicity is quite disordered compared to that within the other two tectonic regimes. Reassuringly the entropy values show no correlation with the number of events, number of points on the hull or the size of the regions.

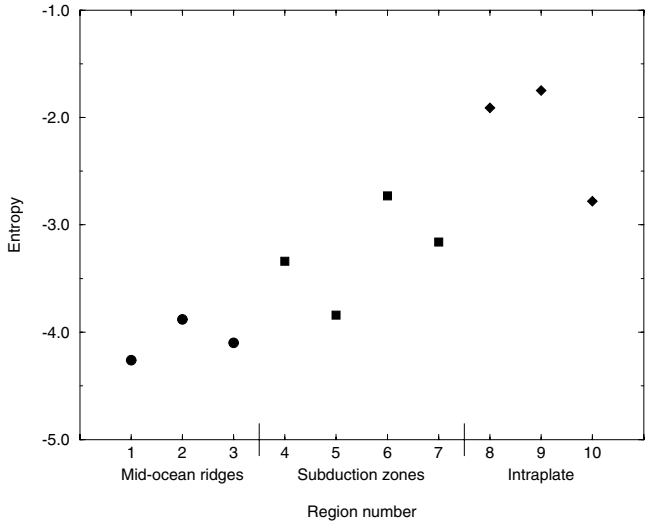

Figure 2. The entropy values of the 10 subsets of the EHB catalogue shown in Table 1.

The entropy of the Western USA sample is somewhat of an anomaly given that it is significantly higher than that of the other intraplate regions and lies within the range of entropies characteristic of subduction zones. This low entropy value is possibly due to the inclusion of interplate seismicity at the border between the Pacific and North American plates, which is clearly of a more organized structure.

We have seen that entropy may be used to characterize earthquake distributions from various tectonic regimes and also as a measure of organization or complexity within earthquake distributions. In the next section we show how entropy may be used to seek a simpler distribution of hypocentres.

4 USING ENTROPY TO REDUCE SCATTER IN EARTHQUAKE DISTRIBUTIONS

It is self-evident that hypocentral errors tend to make earthquake distributions dispersed and hence obscure the underlying 'structure' or pattern in the seismicity. One way to reduce this 'noise scatter' is to try to minimize individual location errors. However, there are limits on how accurately we can locate each earthquake by standard techniques (see Billings *et al.* 1994 for a discussion of the issues). Instead of directly trying to reduce the error in each location we propose an indirect approach, which is to minimize the entropy, and hence scatter, in a distribution of events while requiring that they fit the data to a satisfactory level. It is commonplace among seismologists to interpret earthquake distributions in terms of simplified structures. In doing so we are qualitatively removing scatter from earthquake distributions. However, by combining data misfit with entropy we can add a quantitative side to this analysis and in doing so reduce the need for educated guessing (or, perhaps more accurately, make the guesses more educated).

Many earthquake location procedures locate earthquakes individually (e.g. Geiger 1912). Such methods as they stand are of no use here because scatter can only be defined across a distribution of events, and so we must use a joint earthquake location procedure. Here we present two possible approaches to linking data misfit and entropy and study one of them in

detail. The first is to proceed in a similar vein to single earthquake location, except that we use multiple events and add an entropy term (as a measure of scatter or disorder) to the minimization criterion,

$$\psi = \phi + \lambda S, \tag{12}$$

where ϕ is a measure of the data misfit of all events, S is the entropy of the distribution and λ is a 'tuning parameter' that reflects the relative weight given to data misfit and entropy.

By minimizing ψ we are reducing the disorder in a given earthquake cluster while requiring that the distribution of events fits its observational data to within a satisfactory level. The data misfit terms are independent for any two earthquakes. However, the additional entropy term acts to couple the locations together so that each location is affected by those around it (see Appendix B for a full formulation of this approach). In a typical application there may be between 1000 and 200 000 earthquakes, which results in up to 8×10^5 unknowns (i.e. four coordinates for each event). This is a very large number of unknowns and is comparable to the size of systems used in recent tomographic studies (Widiyantoro & van der Hilst 1997).

While in theory this is an interesting and viable option there are several difficult computational problems that must be overcome before it can be applied. These problems make the application of this theory cumbersome with the tools currently available. However, Voronoi cell theory is an area of very active research at present and so we have included a full formulation of the theory and a brief discussion of the computational problems in Appendix B.

An alternative to entropy in (12) is to use the fractal dimension of the hypocentres. Fractal dimension can be calculated using the approach described by Mandelbrot (1982). The formulation for this case is the same as that in Appendix B. This approach, however, is likely to be difficult to implement because replacing entropy with fractal dimension only adds to the difficulty since derivatives of fractal dimension have no natural definition and therefore are not suitable for use with gradient or matrix inversion algorithms.

An alternative to the minimization of ψ in (12) is to make use of a more indirect approach. The 'collapsing method' of Jones & Stewart (1997), which is summarized in Section 4.1, attempts to give, in their own words, 'the simplest possible structure [of an earthquake distribution] that is consistent with all the location and confidence ellipsoid data', therefore producing simpler structures while not significantly affecting data misfit. This method was originally developed for use with local seismicity, but is, in principle, applicable to any tectonic regime. They apply the method and interpret their results without defining a quantitative measure of 'simplicity'. Entropy may be used as such a measure. In our experiments we found that the collapsing method was able to perturb earthquake foci so as to reduce the entropy of the distribution, despite the fact that it was not specifically designed to do so. At the same time it does not significantly affect the misfit of arrival times. The collapsing method has also empirically been found to reduce the fractal dimension of the distribution (Jones, personal communication, 1999).

In the following sections we study the collapsing method and its effects on entropy in detail. We propose two modifications to the method that overcome some of its limitations and increase the range of problems to which it may be applied. We also apply each of our modified collapsing methods to a data set of interest and discuss the results. We find that entropy is reduced in all cases and some interesting tectonic features are brought into focus.

4.1 The collapsing method

The basic idea of the collapsing method is to perturb each hypocentre while trying to ensure that the overall perturbed distribution is as statistically likely as the unperturbed one in terms of its fit to the data. In this method approximate data consistency is imposed by requiring that the hypocentral movements fit a χ^2 distribution to within some tolerance.

The collapsing method consists of three simple steps.

(1) For each event one finds all the hypocentres falling within its four standard deviation confidence ellipsoid and calculates their centroid.

(2) Each hypocentre is then perturbed 0.61803 of the distance from the location to the centroid in the direction of the centroid (mimicking the golden section search of Press *et al.* 1986).

(3) Return to step 1 unless the distribution of movements from the original locations best matches a χ^2 distribution within some tolerance.

Jones & Stewart (1997) provide a complete description of the method.

They applied the collapsing method to a series of synthetic data sets and discussed possible artefacts, before applying it to events beneath the Rabaul Caldera, in Papua New Guinea. Here we supplement their discussion of artefacts before applying a modified collapsing method to both regional and global data sets. In our examples we show that at each iteration it reduces the overall entropy of the event distribution while forcing the statistics of the hypocentral movements to be consistent with the noise in the original data. With this approach we hope to obtain an impression of what might be achieved by linking data misfit and entropy as well as a better understanding of the collapsing method.

4.2 Artefacts in the collapsing method

The main artefact of the collapsing method noted by Jones & Stewart (1997) is that clouds of seismicity tend to shrink slightly from their initial distribution. This effect is particularly evident in depth. Distributions with very shallow or deep events tend to have the shallowest events pulled down and the deepest events pulled up. This occurs because just below the shallow events there is a region of very high event density and above them is an area of zero event density. As a result, shallow events experience no force of attraction upwards to counteract the attraction downwards. A similar effect occurs for the deepest events but is usually less prominent because of the fewer 'nearby events' above. This problem is accentuated by the fact that most earthquakes have a much larger uncertainty in depth than in epicentral position and hence can move further up or down and still be consistent with observed data.

We found that this 'shrinkage' could also occur in other situations, for example, for aftershock sequences associated with major fault planes. In this case, events at the edge of the cloud of seismicity tend to move towards the centre of the distribution and hence the imaged fault plane tends to shrink.

In our experiments discussed in section 6 we found this effect to depend on the ratio of location uncertainty to the size of the fault plane. If the uncertainty in location is of the same order of magnitude as the feature in question, then the collapsing method would tend to shrink it quite significantly and in extreme cases the whole pattern collapses into a single point. While this appears to be a clear artefact of the method, it is consistent with the goal of the collapsing method, that is, to find the simplest structure, which is as likely as the original distribution. In the case where all the uncertainty ellipsoids overlap, a single point is the simplest structure.

Jones & Stewart (1997) touch on the issue of shortened features in their synthetic cross-example but do not suggest any general remedy. In their example using data from the Rabaul Caldera, however, the size of the cloud is reduced. They provide a 'correction' procedure, whereby events are then moved away from the centre of the Caldera so as to preserve the mean distance from the centre. This correction to the shrinking problem would appear to be inconsistent with the statistical assumptions upon which the method is based because there is a potential for hypocentres to move outside their uncertainty ellipsoid.

An important point to remember when studying distributions produced using the collapsing method is that the method is designed to resolve patterns in seismicity. It is not designed to pull out new structures or detail that is not present in the original distribution. Therefore, any entirely new structure is probably an artefact of the method. In our experiments such new structures are rare and reasonably obvious. In most cases the final collapsed distribution appears more like a 'focused' version of the original.

4.3 Modifications to the collapsing method

To enhance the statistical interpretation of the collapsing method we introduced a weighting scheme into the calculation of the centroid at step 1 of the algorithm. The purpose of this is to allow the events near the centre of an uncertainty ellipsoid to produce a greater attraction on the original event than those at the edge on the ellipsoid. The original collapsing method does not distinguish between distant and nearby events, and so all events within the prescribed confidence ellipsoid exert the same attraction on the central event (i.e. in step 2). This causes the centroid to depend on outliers in a non-robust manner. We would like to improve the robustness of the collapsing method by reducing the influence of outliers. Here we use a Gaussian distribution as our weighting scheme. This allows the 'force of attraction' between two events to be proportional to the probability that the event we are considering is at a point in the middle. This gives a weighting function

$$w(d) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-d^2}{2}\right), \tag{13}$$

where d is the distance (in units of σ) from the event to the centre of the uncertainty ellipsoid.

To illustrate the differences between the unweighted collapsing (UC) method and the weighted collapsing (WC) method, we conducted a synthetic test using a random distribution of 8000 points in three dimensions as the initial data set (see Fig. 3). All the points had 4σ error spheres of radius 100 units. The depth varied from 0 to 200 units and both collapsing methods moved the distribution so that it was almost collapsed onto the horizontal plane at 100 units depth.

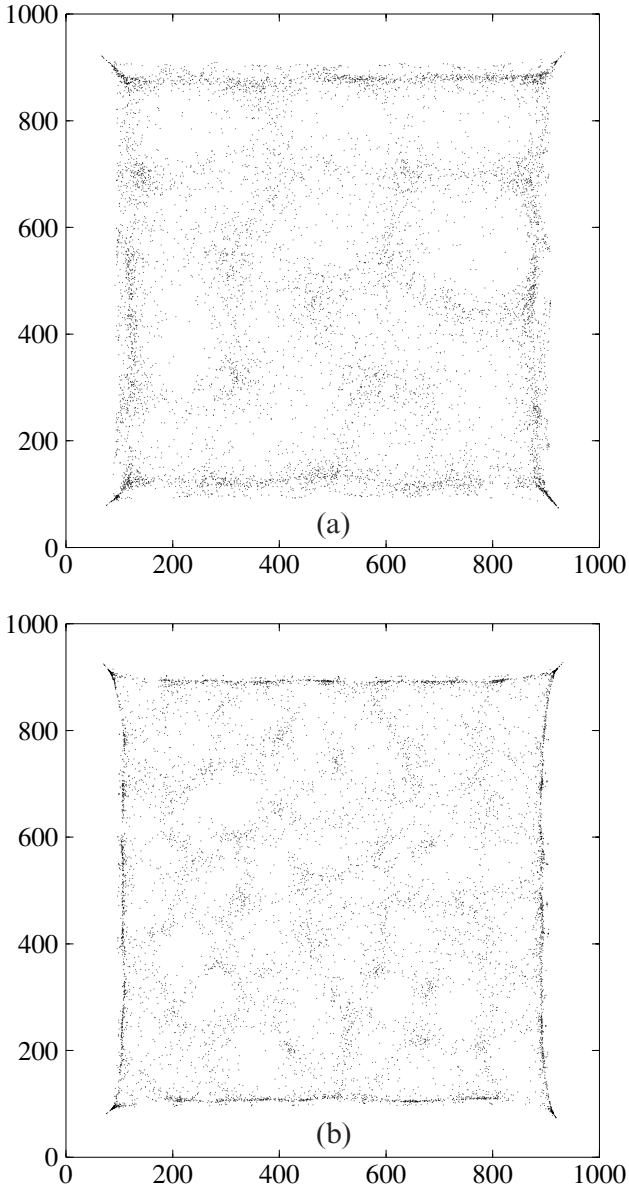

Figure 3. The collapsing methods when applied to a random distribution. The initial distribution of events ranged from 0 to 1000 units in the x and y directions, and 0 to 200 units in the z direction. After both methods were applied the distributions collapsed so that they were contained between 95 and 105 units in the z direction. In both cases the z direction is not shown. (a) The results of the UC method; (b) the results of the WC method.

This test distribution has no underlying structure, so we are only looking at artefacts. We see, in the UC method results, a number of 'holes' of low event density 200 units across. These are approximately 8σ in diameter (the diameter of a confidence ellipsoid) so that an event that began near the centre of one of these low-density areas may have been moved almost 4σ. By changing the size of the confidence ellipsoid to 3σ, we found a similar characteristic pattern with holes of diameter 6σ. Therefore, the choice of cut-off point used in the UC method can result in a 'characteristic distance' in the size of features

in the resulting distribution. The results for the WC method also show a number of low-density regions. These regions are smaller and more numerous and it may seem that we have merely halved the characteristic distance. However, for a χ^2 distribution, a large number of events moving just under two standard deviations is many times more likely than a large number of events moving just under four standard deviations. Therefore, the weighted method results in a distribution that is far more likely (and hence statistically acceptable) than that produced by the unweighted method.

A consequence of the more complicated weighted scheme is that the computation time for the overall procedure is increased by up to a factor of 5. This is because at each iteration the centroids calculated by the WC method are generally closer to the central events than in the UC method. Therefore, at each iteration the events move a shorter distance and hence the modified algorithm takes more iterations to converge. In practice, however, this was not a severe problem. This synthetic test shows how artificial structure can be generated with the method and also indicates the typical form this may take. The 'patchwork pattern' in Fig. 3 is also observed in real data studies (see next section).

In summary, the introduction of the weighting function does not remove the artefacts produced by the collapsing method; however, it does make the method more robust and intuitive. The collapsing method and its variants do not explicitly minimize the entropy. However, we found that they achieved the lesser goals of reducing entropy and fitting the data as well as the original distribution (given the assumptions made about the Gaussian nature of error statistics).

5 APPLICATION OF THE WEIGHTED COLLAPSING METHOD TO A GLOBAL DATA SET

Engdahl *et al.* (1998) (hereafter referred to as EHB) relocated nearly 100 000 earthquakes from the ISC catalogue that occurred between 1964 and 1995 using the ak135 velocity model. The EHB data set improves on the ISC and NEIC data sets by including arrival times of more phases, particularly *pP* and *pwP*. EHB also filtered out those events they thought were poorly constrained. Only those events with at least 10 usable first arrivals from teleseismic stations and with azimuthal coverage of these stations greater than 180° were included. We use those events from the EHB data set that were not fixed in depth during their relocation procedure (84 616 events).

In applying the WC method to a global data set, we are examining the effect it has on large-scale features. A large proportion of the events in the EHB catalogue are on, or near, plate boundaries. We therefore concentrate our study on features associated with subduction zones and mid-ocean ridges.

The size and shape of features, as well as the density of hypocentres, varies enormously over the globe. Jones & Stewart (1997) applied the collapsing method to a local earthquake problem containing a single feature, or rather a pair of features as they postulated. However, the collapsing method is equally applicable to a larger number of events distributed across several tectonic regimes.

The EHB data set contains information about the errors in each location. It gives a value for the standard error in longitude and latitude combined, as well as the standard error in depth. We assume that the error in longitude and latitude is equally distributed between the two. The EHB catalogue does not give any information about the orientation of the error ellipses and so all our ellipsoids have vertical depth axes. Error estimates given in earthquake catalogues may be optimistic as absolute error estimates. As relative error estimates it is less clear whether they are optimistic or not because some sources of systematic error are similar for nearby events, for example, errors due to an inaccurate velocity model. However, following Jones & Stewart (1997) we increased the size of the error ellipsoids by factors between 1.25 and 7.0. These factors act as tuning parameters, and they are the most influential choice made in applying the collapsing methods. We will see below the effect this arbitrary scaling factor has on the resulting distribution.

When using a data set that is spread over such a large volume it is likely that parts of the distribution will be too sparse for the method to collapse the locations. Often an event will not have any other events in its error ellipsoid and such locations cannot move. Fig. 4 shows a graph of entropy versus iteration number. Note that the entropy decays as a smooth function of iteration number, indicating that the method does indeed reduce entropy. We can see that most of the reduction in entropy is achieved in the first few iterations, which is where most of the hypocentral movements occur.

Figs 5, 6 and 7 show a series of diagrams containing subsets of the global results. Figs 5 and 6 show results from cases where we increased the reported estimates of epicentral error by a factor of 1.5 and depth error by a factor of 3.0. A full interpretation of the patterns appearing in the global results is beyond the scope of this paper. Here we concentrate on subsets and point out some of the general features of interest.

Mid-ocean ridges are one of the classes of features we are most interested in when studying the EHB data set. Earthquakes along ridge segments are sometimes difficult to locate accurately because of their great distance from recording stations. Fig. 5 shows an area of the Atlantic off the coast of Brazil. Included in this picture are the St. Paul (just west of centre), Romanche (just east of centre) and Ascension (south-west corner) fracture zones. The ridge segments were inferred from magnetic lineations in the seafloor (Cande *et al.* 1989). Most of the seismicity in this region is due to strike-slip faulting

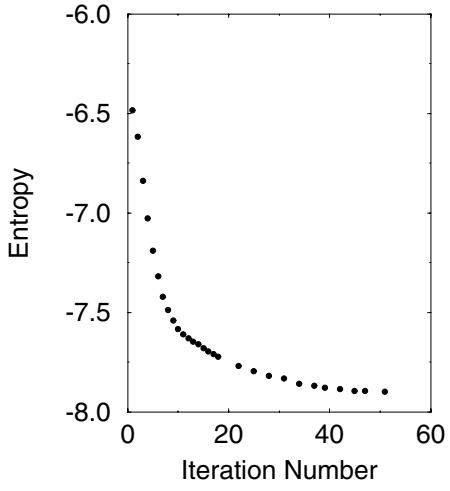

Figure 4. Entropy as a function of iteration number for the application of the WC method to the EHB catalogue. Note the smooth decay of entropy with iteration number.

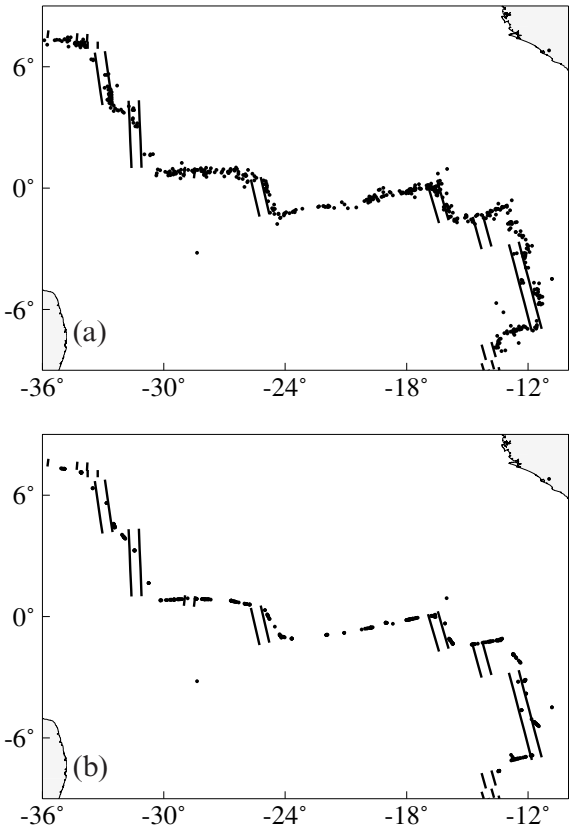

Figure 5. Earthquakes near mid-ocean ridges in the Atlantic. A number of fracture zones are show in this picture including the St. Paul, Romanche and Ascension zones. The ridge segments are shown as double lines. (a) The EHB locations; (b) the EHB locations after the WC method has been applied with uncertainty scaled up by 2.0 in longitude and latitude and 4.0 in depth.

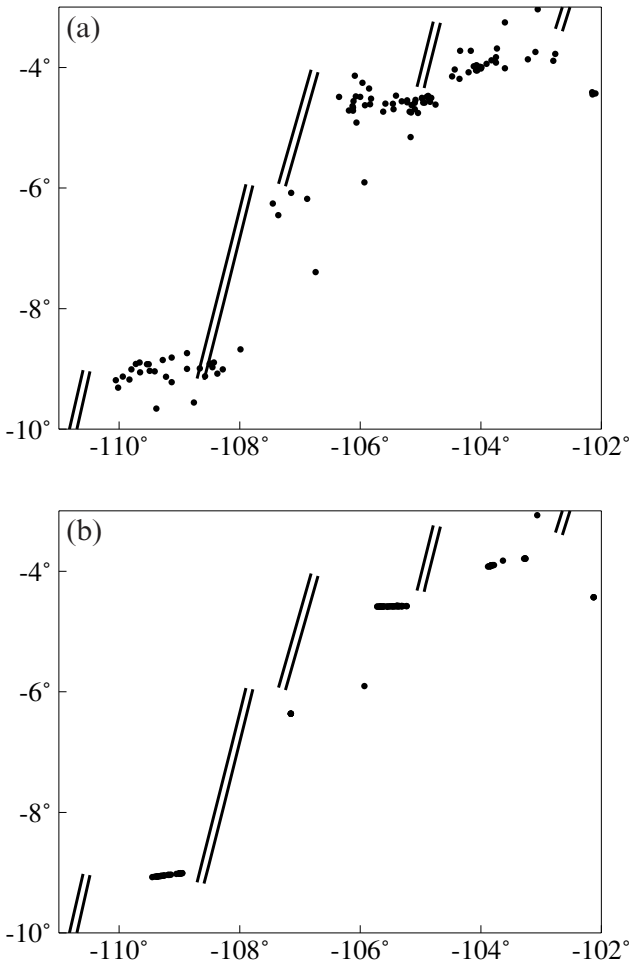

Figure 6. Earthquakes near mid-ocean ridges on the East Pacific rise. This portion of the East Pacific rise is part of the rise where the Nazca and Pacific plates meet. (a) The EHB locations; (b) the EHB locations after the WC method has been applied with uncertainty scaled up by 2.0 in longitude and latitude and 4.0 in depth.

along the fracture zones. However, significant normal faulting also occurs along the ridge segments. The WC method causes the distribution to tighten, removing some of the scatter about the ridge segments and transform faults. Almost all of the seismicity in the processed diagram falls on straight lines along either a transform fault or a ridge segment, which is what is expected for a mid-ocean ridge. There is a slight smoothing in the connection between a ridge segment and the western end of the Romanche fracture zone (at approximately 25°W). We believe this is an artefact of the WC method. In some small areas of low density all of the events have collapsed to a point. Note that, for this subset of the EHB data set, the entropy of the processed subset (-3.96) is lower than that of the initial subset (-2.41). This was calculated using the method in Section 2.

A smaller region from the East Pacific rise is shown in Fig. 6. This area includes the Wilkes (southwest corner) and Discovery (northeast corner) fracture zones. The processed data has a lower degree of disordering because it is essentially two lines with a few other points nearby as opposed to a series of clusters. In this case the entropy values were -1.50 for the initial distribution and -6.97 for the processed data. Fig. 6 shows some interesting features. The initial data has events falling outside the area between the ridge segments. However, the processed events show that it is equally likely that they all fall between the lineations and, furthermore, they can all fall on a line. Note that the two regions above are small subregions of those presented in Table 1 with substantially fewer events, which leads to a less robust measure of entropy in an absolute sense.

Fig. 7 shows the Molucca Sea region of Indonesia with two different error ellipse scaling factors. These are 1.25 in epicentral coordinates and 2.0 in depth (Fig. 7b), compared to 2.0 in epicentral coordinates and 4.0 in depth (Fig. 7c). Fig. 7(c) hints at a link between a number of the features that Fig. 7(b) shows as separate features. Some of these linkages are likely to be artificial. For instance, Fig. 7(c) shows a hint of a continuous hook that starts in the centre and moves towards the southwest, south of the Minahassa Peninsula on Sulawesi Island. This is unlikely to be accurate because they are different tectonic boundaries and at different depths (Wilson *et al.* 1998). In Fig. 7(b) this hook appears to be separated, whereas other areas that may well be linked are not linked in Fig. 7(b). Figs 5, 6 and 7 show how widely the WC method may be applied, while highlighting the limitations of the method. All of the figures show a clarification of structure and in all cases a corresponding decrease in entropy.

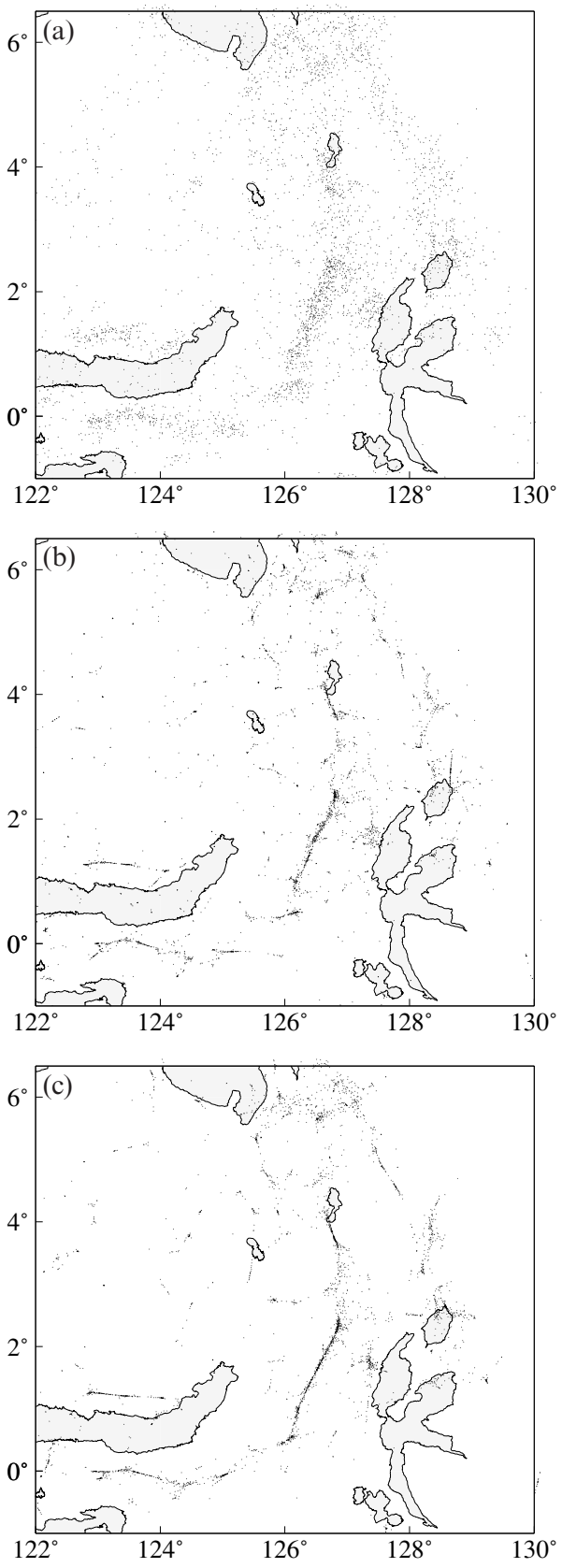

Figure 7. Earthquakes in the Molucca Sea region of Indonesia. (a) EHB locations; (b) EHB locations after the WC method has been applied with uncertainty scaled up by 1.25 in the longitude and latitude directions and by 2.0 in depth; (c) uncertainty scaled up by 2.0 in longitude and latitude and 4.0 in depth.

6 APPLICATION OF THE WEIGHTED COLLAPSING METHOD TO DATA FROM THE SIL NETWORK

We use data from the Southern Iceland Lowlands (SIL) network (Bodvarsson *et al.* 1996) recorded between 1 January 1996 and 31 December 1997 to demonstrate a further modification of the collapsing method. During this period 43 300 events were recorded, not including those that were assigned a fixed depth. The majority of the events occurred near the Hengill triple junction in southwest Iceland. This is a tectonically complex area and the junction between the Reykjanes rift, the Western Iceland rift zone and the South Iceland transform. Surface faulting in the area is dominated by NE-striking normal faults congruent with the dominant strike of volcanic fissures associated with the Hengill volcano. Focal mechanisms of microearthquakes in the area are most commonly strike-slip on near-vertical N–S or E–W faults, although normal faulting is also common (Rognvaldsson *et al.* 1998). Upper crustal fabric as inferred from shear wave splitting is complex but mostly subparallel to the strike of the Hengill fissure swarm (i.e. NE) (Evans *et al.* 1996).

6.1 Using temporal information

This data set contains a large number of earthquake swarms of duration less than about four days (Rognvaldsson *et al.* 1999). Detailed relocation of events in these swarms have shown that they cluster tightly around near-vertical planes (Rognvaldsson *et al.* 1998). Many of these planes cross, but the swarms are separated in time and so should be treated separately. The WC method has no mechanism to do this, and so we make a further modification and add a new method of attraction that makes use of temporal information.

As the WC method progresses we identify those events that are members of swarms. An earthquake is a member of a swarm if within four days of the earthquake more than 20 other earthquakes occur inside its uncertainty ellipsoid. For each swarm, we determine the best-fitting vertical plane by a least-squares procedure and then discard events that are more than 300 m from the plane as outliers. If fewer than 20 events remain we no longer have a swarm. Those events that are still members of swarms are then perturbed towards the closest point on the plane through their swarm. The proportion of the distance to the plane that these events are moved is chosen to be the same as the proportion of the distance to the centroid for events not considered to be members of swarms (i.e. 0.61803 initially). We chose 20 events and 300 m as our cut-off values because they are realistic values for both the spread of events and number of events within a cluster for this region. A consequence of the new 'temporal' attraction mechanism is that events that are attracted by a plane may be moved outside their original uncertainty ellipsoids. When this occurs, we move them in the direction of their original location until they lie on the edge of their uncertainty ellipsoid. Those events that are not members of swarms are perturbed by the collapsing method in the usual way, that is, towards the centroid of nearby events. We apply the same stopping criteria to the procedure as before.

This approach of using temporal information by attracting swarm activity to planes is only one way of incorporating time

information. An alternative would be to weight events according to origin-time separation when computing the centroid location for a given event, and thus limit the memory span of the method.

6.2 Application to SIL data

Fig. 8 shows the results of the collapsing method in the region around Hengill. The area plotted is about 15 km by 15 km. Fig. 8(a) shows the original distribution as located by the SIL network. Time is colour-coded from red in early 1996, through green at the end of 1996 to blue at the end of 1997. Uncertainties are reported to be about 0.5 km in the horizontal for most events, but may be as large as 3–5 km for the worst-constrained events. Significant structure is evident in the figure. Early on during the two-year period lineations striking approximately N70E–N90E are visible in red. Clear lineations striking about N–N20E are visible later (in green and blue). A hint of a NE lineation is seen around the edges of the distribution where it is less dense.

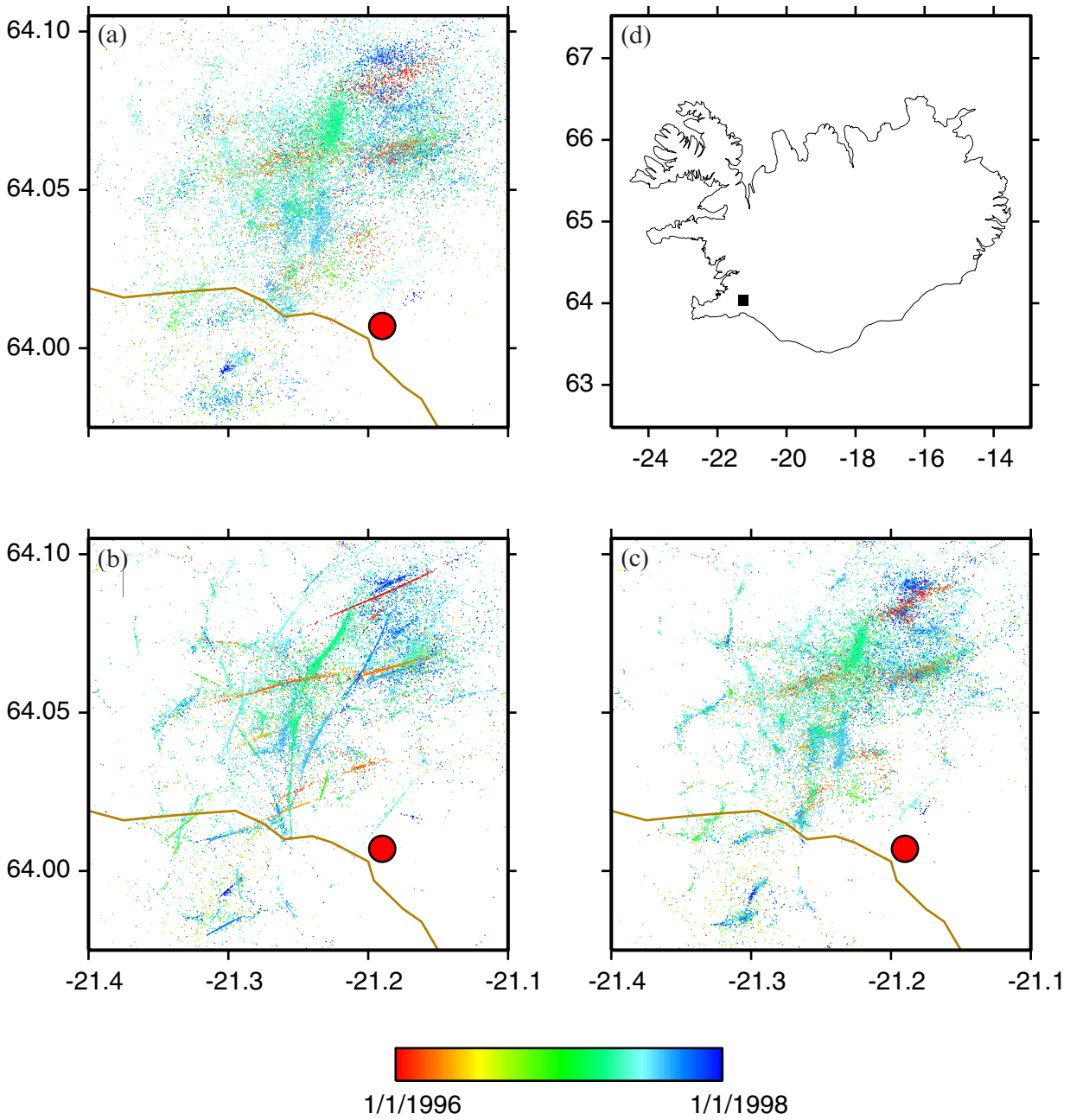

Figure 8. Earthquakes in the Hengill region of southwestern Iceland. (a) The SIL locations, which have entropy −2.41; (b) SIL locations after the WC method including temporal information has been applied, which have entropy −3.96; (c) SIL locations after the WC method has been applied; (d) the area in (a), (b) and (c) is shown as a black square. The town of Hveragerdi is shown as a large red dot and highway 1 as an orange line. In both (b) and (c) no rescaling of uncertainty ellipsoids was performed. The epicentres are drawn in random order so as not to obscure earlier clusters.

Fig. 8(b) shows the result of the collapsing method with the modification to make use of temporal information. A time span of four days is chosen to focus on earthquake swarms. Fig. 8(c) shows the result when temporal information is ignored. In both cases the pattern of the seismicity is significantly focused. Lineations are particularly clear in Fig. 8(b). N70E trends are especially clear and individual swarms line up along similar directions to those estimated by Rognvaldsson *et al.* (1998). The centre of the domain has some clear nearly north-striking features but also clear lineations striking N30E. This direction is not far off the dominant surface faulting in the area. An interesting lineation in the western half of the domain with a direction close to N20W stretches from the centre south to the northeast corner of the domain.

It is important to remember what this method attempts to do before venturing to interpret its outcome. We are seeking a simpler (lower-entropy) pattern of seismicity than in the original distribution without significantly sacrificing fit to arrival-time data (as events are not moved outside their error ellipses). Therefore, the outcome is an alternative distribution, which is not significantly less likely than the original. Structure is introduced into the distribution, faults may be highlighted and faults may be artificially connected.

It is beyond the scope of this paper to study the tectonics of the Hengill region in detail. The results in Fig. 8 suggest some interesting interpretations relating to the tectonics. Rognvaldsson *et al.* (1998) interpreted the near-N–S- and E–W-striking lineations as strike slip on near-vertical faults. This interpretation is in part based on constraints on focal mechanisms. The N20W trend has a geometrical distribution that connects the southern tip of the Hengill fissure swarm with the northern tip of the adjacent Brennisteinsfoll fissure swarm and may accommodate strike-slip motion in the deformation zone between the two rift segments as the complimentary strike-slip direction to the N70E direction in the bookshelf rotational tectonics proposed for this region by Einarsson (1991).

7 DISCUSSION

In this paper we have introduced the concept of entropy as a measure of disorder in earthquake distributions. An analytical expression for entropy has been presented that allows earthquake distributions to be compared with differing numbers of events and over differing spatial scales. A practical algorithm has also been presented that allows calculation of this entropy measure without any need for arbitrary scaling parameters. This makes use of the concept of Voronoi cell (i.e. nearest-neighbour region) volumes to approximate the density of the seismicity at any point in space.

We have suggested that the entropy measure may have several uses in quantifying structure or patterns occurring in seismicity and we have explored one of these possibilities in detail. Our main focus has been on the potential use of entropy as a tool for quantifying scatter within earthquake distributions as well as its possible use in conjunction with other methods in reducing this scatter. This is a new, and in our view, interesting development in earthquake location. Since entropy provides a quantitative measure of disorder in a seismicity pattern, it allows an inverse problem to be posed where one seeks the least disordered pattern of earthquakes consistent with the observed data. In this way we hope to suppress unwarranted scatter in the distribution of seismicity. This approach is motivated somewhat by the successful use of entropy in image enhancement algorithms (e.g. Gull & Skilling 1983).

Here we make use of a simple algorithm that simultaneously minimizes entropy and data misfit. The collapsing method of Jones & Stewart (1997) serves this purpose, but also exhibits a number of artefacts that we have discussed in some detail. Our first modification of their method was to introduce a weighting scheme for the attraction between events, which lessened some of the artefacts. From our experiments with both global and local data sets, we conclude that the method has a tendency to be overzealous in collapsing the position of events onto a line or even a single point. This has the effect of elongating the apparent position of faults and potentially artificially condensing some clouds of seismicity. The second modification we introduced was to take advantage of the temporal information in a local earthquake catalogue to aid the alignment of events on fault planes. In the SIL data set this appears to have been rather successful. The resulting distribution of faults imaged is consistent with previous seismic studies and other geophysical information from the region.

In summary, the concept of entropy may be used to measure, or help reduce the disorder of, an earthquake distribution. An accurate knowledge of earthquake locations is important in seismotectonic interpretations, earthquake hazard assessment and seismic tomography. In each area the information from the collective distribution is more important than the accuracy of any single event. Entropy provides a means of characterizing this collective information and may have a useful role in a number of seismic studies.

ACKNOWLEDGMENTS

We acknowledge the Icelandic Meteorological Survey for making the SIL data available to us, and Bob Engdahl for providing us with the EHB catalogue. Keith Priestley made some useful suggestions regarding applications to real data. We also acknowledge useful discussions with Brian Kennett and Rob Jones, as well as constructive reviews from A. Jackson and R. Willemann.

REFERENCES

Aviles, C.A., Scholz, C.H. & Boatwright, J., 1987. Fractal analysis applied to characteristic segments of the San Andreas fault, *J. geophys. Res.,* **92,** 331–344.

Backus, G. & Gilbert, F., 1968. The resolving power of gross Earth data, *Geophys. J. R. astr. Soc.,* **16,** 169–205.

Barber, B. & Huhdanpaa, H., 1994. *Qhull* (computer program), The Geometry Center, Minneapolis, MN.

Barber, B., Dobkin, D.P. & Huhdanpaa, H., 1996. The quickhull algorithm for convex hulls, *ACM Trans. Math. Software,* **22,** 4, 469–483.

Billings, S.D., Sambridge, M.S. & Kennett, B.L.N, 1994. Errors in hypocentre location: picking, model and magnitude dependence, *Bull. seism. Soc. Am.,* **84,** 1978–1990.

Bodvarsson, R., Rognvaldsson, S.T., Jakobsdottir, S.S., Slunga, R. & Stefansson, R., 1996. The SIL data acquisition and monitoring system, *Seism. Res. Lett.,* **67,** 35–46.

Boltzmann, L., 1871. *Sitzungsber. K. Akad. Wiss. Wien.,* **63,** 679–732.

Braun, J. & Sambridge, M., 1995. A numerical method for solving partial differential equations on highly irregular evolving grids, *Nature,* **376,** 655–660.

Buland, R., 1976. The mechanics of locating earthquakes, *Bull. seism. Soc. Am.,* **66,** 173–187.

Cande, S.C., LaBrecque, J.L., Larson, R.L., Pitman, W.C., Golovchenko, X. & Haxby, W.F., 1989. *Magnetic Lineations of the World's Ocean Basins,* AAPG Map Series, cat. no. 506.

Clausius, R., 1850. On the motive power of heat, and on the laws which can be deduced from it for the theory of heat, *Ann. Physik,* **LXXIX,** 368, 500.

Delaunay, B.N., 1934. Sur la sphere vide, *Bull. Acad. Sci. USSR,* **VII,** 793–800.

Dewey, J.W., 1972. Seismicity and tectonics of Western Venezuela, *Bull. seism. Soc. Am.,* **62:** 1711–1751.

Dirichlet, G.L., 1850. Uber die Reduction der positiven quadratischen Formen mit drei unbestimmten ganzen Zahlen, *J. Rein Angew. Math.,* **40,** 209–227.

Douglas, A., 1967. Joint hypocentre determination, *Nature,* **215,** 47–48.

Einarsson, P., 1991. Earthquakes and present-day tectonism in Iceland, *Tectophysics,* **189,** 261–279.

Engdahl, R.E., van der Hilst, R. & Buland, R., 1998. Global teleseismic earthquake relocations with improved travel times and procedures for depth determination, *Bull. seism. Soc. Am.,* **88,** 722–743.

Evans, J.R., Foulger, G.R., Julian, B.R. & Miller, A.D., 1996. Crustal shear-wave splitting from local earthquakes in the Hengill triple junction, southwest Iceland, *Geophys. Res. Lett.,* **23,** 455–458.

Evernden, J.F., 1969. Precision of epicentres obtained by small numbers of world-wide stations, *Bull. seism. Soc. Am.,* **59,** 1365–1398.

Fortune, S., 1995. Voronoi diagrams and Delaunay triangulations, in *Computing in Euclidean Geometry,* 2nd edn, pp. 225–265, eds Hwang, F. & Du, D.Z., World Scientific, Singapore.

Geiger, L., 1912. Probability method for the determination of earthquake epicentres from arrival time only, *Bull. St Louis Univ.,* **8,** 60–71.

Gudmundsson, O. & Sambridge, M., 1998. The regionalized upper mantle (RUM) seismic model, *J. geophys. Res.,* **103,** 7121–7136.

Gull, S.F. & Skilling, J., 1983. *Maximum Entropy method, Indirect Imaging,* Cambridge University Press, Cambridge.

Hirata, T., 1989. A correlation between the b value and the fractal dimension of earthquakes, *J. geophys. Res.,* **94,** 7507–7514.

Jones, R.H. & Stewart, R.C., 1997. A method for determining significant structures in a cloud of earthquakes, *J. geophys. Res.,* **102,** 8245–8254.

Lasserre, J.B., 1983. An analytical expression and an algorithm for the volume of a convex polyhedra in R^n, *J. opt. Theory Appl.,* **39,** 363–377.

Mandelbrot, B.B., 1982. *The Fractal Geometry of Nature,* W. H. Freeman, San Francisco.

Menke, W., 1984. *Geophysical Data Analysis: Discrete Inverse Theory,* Academic Press, Orlando.

Okabe, A., Boots, B. & Sugihara, K., 1992. *Concept and Applications of Voronoi Diagrams,* John Wiley & Sons, Chicester.

Okubo, P.G. & Aki, K., 1987 Fractal geometry in the San Andreas fault system, *J. geophys. Res.,* **92,** 345–355.

Papoulis, A., 1984. *Probability, Random Variables and Stocastic Processes,* 2nd edn, McGraw-Hill, New York.

Parker, R., 1994. *Geophysical Inverse Theory,* Princeton University Press, Princeton.

Peppin, W.A., Honjas, W., Somerville, M. & Vetter, U.R., 1989. Precise master-event locations of aftershocks of the 4 October 1978 Wheeler Crest earthquake sequence near Long Valley, California, *Bull. seism. Soc. Am.,* **79,** 67–76.

Planck, M., 1906. Entropy, *Elet., Lond.,* **50,** 694–695.

Press, W.H., Flannery, B.P., Teukolsky, S.A. & Vetterling, W.T., 1986. *Numerical Recipes: The Art of Scientific Computing,* Cambridge University Press, New York.

Rognvaldsson, S.T., Gunnarsson, G.B., Agustsson, K., Jakobsdottir, S.S., Slunga, R. & Stefansson, R, 1998. Overview of the 1993-1996 seismicity near Hengill, *Icelandic Meteorological Service Rept,* **R98006-JA05**.

Rognvaldsson, S.T., Gudmundsson, A. & Slunga, R., 1998. Seismotectonic analysis of the Tjornes Fracture Zone, an active transform fault in north Iceland, *J. geophys. Res.,* **103,** 30 117–30 129.

Sambridge, M., Braun, J. & McQueen, H., 1995. Geophysical parametrization and interpolation of irregular data using natural neighbours, *Geophys. J. Int.,* **122,** 837–857.

Shannon, C.E., 1948. A mathematical theory of communication, *Bell Syst. Tech. J.,* **27,** 379–423, 623–656.

Silverman, B.W., 1986. Density estimation for statistics and data analysis, *Monographs on Statistics and Applied Probability,* Chapman & Hall, London.

Skilling, J., 1989. *Maximum Entropy and Bayesian Methods,* ed. Skilling, J., pp. 45–52, Klewer Academic, London.

Skilling, J. & Bryan, R.K., 1984. Maximum entropy image reconstruction: general algorithm, *Mon. Not. R. astr. Soc.,* **211,** 111–124.

Smith, G.P. & Ekstrom, G., 1996. Improving teleseismic event locations using a three-dimensional earth model, *Bull. seism. Soc. Am.,* **86,** 788–796.

Tarantola, A., 1987. *Inverse Problem Theory,* Elsevier, Amsterdam.

Voronoi, M.G., 1908. Nouvelles applications des parametres continus a la theorie des formes quadratiques, *J. reine Angew Math.,* **134,** 198–287.

Watson, D.F., 1992. *Contouring: A Guide to the Analysis and Display of Spatial Data,* Pergamon, Oxford.

Widiyantoro, S. & van der Hilst, R.D., 1997. Mantle structure beneath Indonesia inferred from high resolution topographic imaging, *Geophys. J. Int.,* **130,** 167–182.

Williamson, P.R., 1990. Tomographic inversion in reflection seismology, *Geophys. J. Int.,* **100,** 255–274.

Wilson, P. *et al.*, 1998. Study provides data on active plate tectonics in southeast Asia region, *EOS, Trans. Am. geophys. Un.,* **79,** 545–549.

APPENDIX A: ENTROPY OF SOME SIMPLE DISTRIBUTIONS

To aid in understanding the implications of our entropy definition, we have applied it to two synthetic distributions. While these are far simpler than the real distributions we consider in Section 3, they illustrate some important properties of entropy.

(1) The random distribution. This is simply a set of points randomly chosen within a cube (the length of the side of this cube does not influence the results). This example is used to show the effect of changing the number of points in the distribution on the entropy value. Theoretically, changing the number of points in eq. (11) will not affect the entropy value obtained. However, when the number of points is small the limited sampling causes some regions to have a higher point density than others. As the number of points increases the distribution becomes more uniform and hence less 'structured'. As a result the entropy increases slightly as the number of points increases (see Table A1). While there is a trend observed it is relatively small when compared to the differences observed in Table 1 and there is no correlation observed in Table 1

Table A1. Entropy of synthetic data sets.

Example number	Number of points	Std dev of added error	Entropy
1	100	N/A	−0.232
1	200	N/A	−0.274
1	500	N/A	−0.146
1	1000	N/A	−0.132
1	2000	N/A	−0.119
2	1000	30 units	−0.448
2	1000	20 units	−0.424
2	1000	10 units	−0.586
2	1000	5 units	−0.898
2	1000	3 units	−1.116

between the number of points and the entropy. Indeed, the mid-ocean ridges and intraplate seismicity have similar numbers of points but are at opposite ends of the entropy spectrum.

(2) The mid-ocean ridge. Two examples of this distribution are shown in plane (xy) view in Figs A1(a) and (b). This is essentially six noisy vertical planes connected end to end in a zigzag pattern. The extent of this distribution in the z-direction is small (10 units), while each of the noisy planes is 100 units long. Normally distributed noise is added perpendicular to the vertical planes. We see from Fig. A1(c) that the entropy increases with the width of the noise distribution. This is because the distribution of points becomes less structured and more uniform as the width of the noise distribution increases (e.g. Fig. A1b). In this case we clearly see that the entropy measures the level of structure in a distribution; that is, the larger entropy corresponds to less structure and a more diffuse cloud.

The earthquake entropy is unaffected by a rescaling of all the axes by a constant factor, hence the entropy measure is scale-independent. Also, if the pattern is repeated and placed alongside the original distribution, the overall entropy will be only slightly changed. In this case the Voronoi cells of all of the points inside the convex hull (see Fig. 1) are unchanged; however, the Voronoi cells of the points on the hull may change, and hence the overall entropy will also change. This is essentially an 'edge effect' and it decreases rapidly as the number of points in the distribution increases.

It is also important to note that even the lowest entropy value in Table A1 is higher than the highest value in Table 1. This is again an indication of the complexity displayed by the real distributions, and while it is easy to compare structure within these synthetic examples without the need for a quantitative measure, it is far more difficult for real distributions.

APPENDIX B: MINIMIZATION OF EARTHQUAKE ENTROPY

In this Appendix we describe a formal approach to the minimization of a combination of 'hypocentral entropy' and arrival-times residual. This leads to a familiar least-squares-type iterative procedure.

We seek the hypocentral distribution that minimizes a linear combination of data fit and entropy:

$$\psi = \phi + \lambda S . \tag{B1}$$

Here λ is an arbitrary damping parameter that determines the relative weight placed on the two minimization quantities, ϕ and S. Parker (1994) called λ 'a trade-off parameter because it balances, or trades off, the two undesirable properties of the model' and we will continue with this convention. While the trade-off parameter results in a non-unique distribution, it allows us to change the weights to reflect our perception of the quality of the data. In geographical areas where we know the data is of good quality we can reduce λ accordingly. Of course, it is important not to get into a circular argument. If we expect simple structures, we might increase λ until we get the structures we expect and use this as proof of their existence. A responsible and conservative choice of λ should avoid any such circularity.

Suppose we have arrival-time data, t_{kj} (for the jth earthquake at the kth station), that are non-linear functionals of the earthquake location parameters, $\mathbf{x}_j$. Each event $\mathbf{x}_j$ has four elements. We can string a whole data set of n events into vectors $\mathbf{t}$ and $\mathbf{x}$, so that $\mathbf{x}$ has a dimension $4n$.

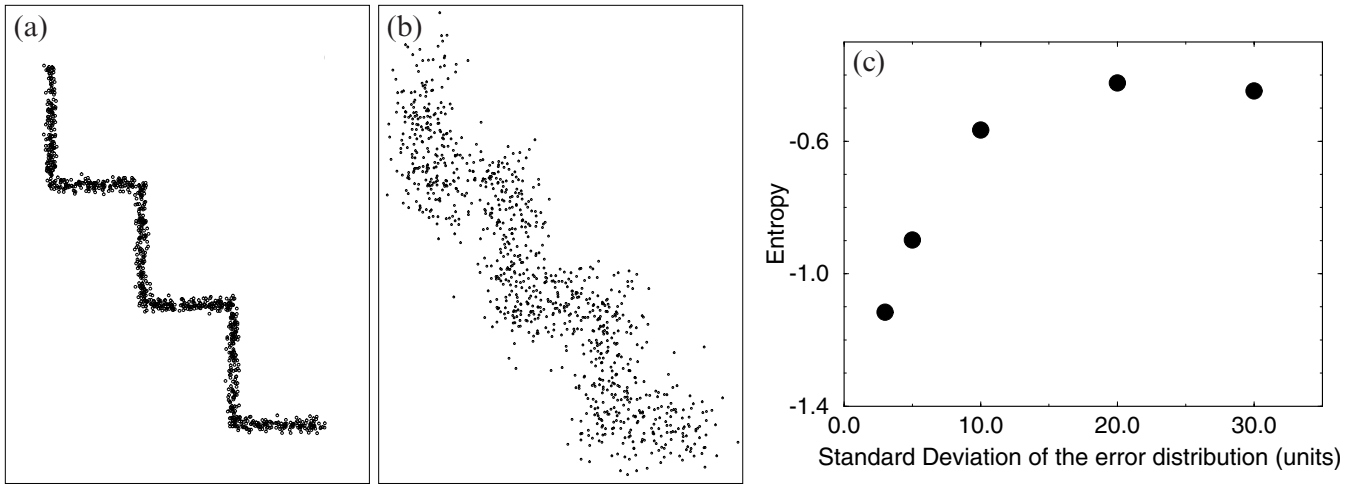

Figure A1. Synthetic data sets used to test entropy. The synthetic mid-ocean ridge with (a) 3 units of noise added and (b) 20 units of noise added, and (c) entropy as a function of noise added for the synthetic mid-ocean ridge.

We relate the arrival times to the hypocentre and write

$$t_i = g_i(\mathbf{x}), \tag{B2}$$

where t_i is the ith element of the data vector $\mathbf{t}$ and g_i is the solution to the forward problem for the ith arrival time calculated by ray tracing.

Given $\mathbf{x}$ we can predict t_i and thus write the prediction error

$$e_i = t_i - g_i(\mathbf{x}). \tag{B3}$$

We want to find a solution $\mathbf{x}$ that minimizes $\mathbf{e}$ in some sense. A common measure is

$$\phi = \mathbf{e}^{\mathrm{T}}\mathbf{C}_d^{-1}\mathbf{e}, \tag{B4}$$

which is a weighted sum of predicted errors with the weights chosen as the inverse standard error estimate. In this equation $\mathbf{C}_d$ is the data covariance matrix describing noise in the observations (Tarantola 1987; Menke 1984). Minimizing ϕ, we arrive at a set of decoupled least-squares solutions. However, we choose to regularize the procedure by minimizing the entropy, $S(\mathbf{x})$. Note that entropy is also a non-linear functional of the hypocentres and is independent of origin time.

A tractable way of proceeding is to linearize the problem about a starting model $\mathbf{x}$,

$$\mathbf{e} = \mathbf{t} - \mathbf{g}(\mathbf{x}) \approx \mathbf{t} - \mathbf{g}(\mathbf{x}_0) - \mathbf{G}\delta\mathbf{x} = \delta\mathbf{t} - \mathbf{G}\delta\mathbf{x}, \tag{B5}$$

where $\delta\mathbf{t}$ is a vector of time residuals or prediction errors for the model $\mathbf{x}_0$ and

$$G_{ij} = \left.\frac{\partial g_i}{\partial x_j}\right|_{\mathbf{x}_0}. \tag{B6}$$

We seek the model $\mathbf{x} = \mathbf{x}_0 + \delta\mathbf{x}$ that perturbs the model $\mathbf{x}_0$ towards a more accurate solution. Define $\mathbf{e} = \delta\mathbf{x} - \mathbf{G}\delta\mathbf{x}$. This first-order expansion enters a second-order norm and, therefore, we do not need any higher-order terms. Similarly, for the entropy term we have

$$S(\mathbf{x}) = \boldsymbol{\nabla}^{\mathrm{T}} S\delta\mathbf{x} + \delta\mathbf{x}^{\mathrm{T}}\boldsymbol{\Omega}\delta\mathbf{x} + S(\mathbf{x}_0) + \ldots, \tag{B7}$$

where $\boldsymbol{\nabla}S$ is the gradient of S and $\boldsymbol{\Omega}$ is its Hessian matrix,

$$\Omega_{ij} = \frac{\partial}{\partial x_i}\frac{\partial}{\partial x_j} S\big|_{\mathbf{x}_0}.$$

ψ depends on S to first order so we must expand S to second order,

$$\psi = (\delta\mathbf{t}^{\mathrm{T}} - \delta\mathbf{x}^{\mathrm{T}}\mathbf{G}^{\mathrm{T}})\mathbf{C}_d^{-1}(\delta\mathbf{t} - \mathbf{G}\delta\mathbf{x}) + \lambda(S_0 + \boldsymbol{\sigma}^T\delta\mathbf{x} + \delta\mathbf{x}^{\mathrm{T}}\boldsymbol{\Omega}\delta\mathbf{x}), \tag{B8}$$

where for simplicity we write the gradient of S as $\boldsymbol{\sigma} = \boldsymbol{\nabla}S|_{\mathbf{x}_0}$.

Differentiating with respect to elements of the vector $\delta\mathbf{x}$,

$$0 = -2\mathbf{G}^{\mathrm{T}}\mathbf{C}_d^{-1}\delta\mathbf{t} + 2\mathbf{G}^{\mathrm{T}}\mathbf{C}_d^{-1}\mathbf{G}\delta\mathbf{x} + \lambda\boldsymbol{\sigma} + \lambda 2\boldsymbol{\Omega}\delta\mathbf{x} \tag{B9}$$

$$\Rightarrow (\mathbf{G}^{\mathrm{T}}\mathbf{C}_d^{-1}\mathbf{G} + \lambda\boldsymbol{\Omega})\delta\mathbf{x} = \mathbf{G}^{\mathrm{T}}\mathbf{C}_d^{-1}\delta\mathbf{t} - \frac{\lambda\boldsymbol{\sigma}}{2}. \tag{B10}$$

Therefore,

$$\delta\mathbf{x} = (\mathbf{G}^{\mathrm{T}}\mathbf{C}_d^{-1}\mathbf{G} + \lambda\delta\boldsymbol{\Omega})^{-1}\left(\mathbf{G}^{\mathrm{T}}\mathbf{C}_d^{-1}\delta t - \frac{\lambda}{2}\boldsymbol{\sigma}\right). \tag{B11}$$

When $\lambda = 0$ these equations decouple into sets of four that are the non-regularized least-squares location equations (see e.g. Menke 1984). Eq. (B11) constitutes an algorithm for regularized earthquake location that can be applied to any earthquake catalogue. In a typical application there may be between 1000 and 200 000 earthquakes, which results in up to 8×10^5 unknowns (that is, four coordinates for each event). This is comparable to the size of systems used in recent tomographic studies (e.g. Widiyantoro & van der Hilst 1997).

In line with λ's role as a trade-off parameter, the linearized system should be calculated and solved for several different λ values so that the sensitivity of the solution can be judged. This means that for several different λ the gradient $\boldsymbol{\sigma}$ and the Hessian of entropy $\boldsymbol{\Omega}$ must be calculated after each iteration, neither of which are trivial. From eq. (11) we see that in order to calculate $\boldsymbol{\sigma}$ and $\boldsymbol{\Omega}$ one must be able to determine the first and second derivatives of the volume of a Voronoi cell with respect to the position vectors of its defining node and those of its neighbours. Braun & Sambridge (1995) described a method for calculating the first derivative of the volume of each Voronoi cell with respect to the nodal coordinates using the formulae of Lasserre (1983). No algorithm is currently available for the second derivatives, although we note that in seismic tomography it is common to ignore the analogous second-derivative terms in the Hessian (see Williamson 1990 for a discussion). Implementation of (B11) as an iterative algorithm would appear to be rather numerically expensive and may not be practical for routine use.

It may be possible to avoid the need for second derivatives by using gradient descent methods of optimization of ψ (for example, steepest descent or conjugate gradients). Skilling & Bryan (1984) followed this approach in an image enhancement problem. That type of approach may well be useful for the hypocentral 'image enhancement' problem proposed here. This would need to be tested numerically before any general conclusions could be drawn.