JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 102, NO. B4, PAGES 8245–8254, APRIL 10, 1997

A method for determining significant structures in a cloud of earthquakes

R. H. Jones

CSM Associates Ltd., Cornwall, England

R. C. Stewart[1]

Department of Geoscience and Technology, Tohoku University, Sendai , Japan

Abstract. The interpretation of a cloud of earthquake hypocenters in terms of causative structures is not a simple task. Locations are subject to uncertainties, which will not be the same for every earthquake. The data should therefore not be interpreted simply by inspection, which is difficult in the case of three-dimensional data anyway. Instead, we propose using the location uncertainties as a guide in processing the data. Earthquake locations are moved inside their uncertainty or confidence ellipsoids until a simplified picture of the earthquake cloud is obtained, which can then be interpreted in terms of some simplified structure such as faults. The aim of the approach is to give the simplest possible structure that is consistent with all the location and confidence ellipsoid data. The method is applied to three synthetic sets of data. These illustrate the potential and limitations of the method. Application to a real earthquake data set from Rabaul Caldera in Papua New Guinea gives an image of the caldera ring fault that suggests departures from the simple ring-fault structure previously assumed. Sensitivity analysis on the Rabaul data shows that the method is not unduly sensitive to the assumptions that have to be made in applying it.

Introduction

The interpretation of a group of earthquake locations, a "cloud," to determine any causative structure or structures is a common problem in seismology. The more numerous and complex the data, the more complex we would expect the interpretation to be. However, earthquake locations are subject to uncertainties and errors, arising, for example, from bad timings, an incorrect velocity structure, or a poor network geometry. Errors can, of course, be minimized by exercising due care during the location process, but uncertainties can never be eliminated as the observational data, i.e., the arrival times, are themselves subject to uncertainties. Any interpretation of the data should therefore take account of the estimates of the location uncertainties.

More numerous and complex-looking location data need not imply a more complex structural interpretation. Faced with such a set of locations and their uncertainties, we need to know how complicated an interpretation can be supported by that data set.

Note that we are addressing the interpretation of a cloud of hypocenters and not the location problem itself. Location methods exist which can be applied to improve a set of locations; e.g., the joint-epicenter and joint-hypocenter determination methods of *Douglas* [1967] and *Frohlich* [1979], the relative location procedure of *Spence* [1980], the Bayesian procedure of *Matsu'ura* [1984], or the genetic algorithm of *Sambridge and Gallagher* [1993]. All of these will still result in a set of locations and their associated uncertainties. Ideally, the technique presented in this paper should only be applied to a set of locations once all reasonable efforts have been made to remove mispicks and systematic errors and the only remaining uncertainties are those due to random observational uncertainties in picking arrival times.

A crude method of accounting for location uncertainties in the interpretation process is to discard locations whose uncertainties are so large that it is felt that they will probably be displaced too far from the underlying structures [e.g., *Mori and McKee*, 1987]. This is, however, highly subjective, and it is difficult to avoid needless discarding of data. All the earthquake locations will contain some information on the underlying structures, whatever their uncertainties, and should not be discarded so easily.

An alternative method is to display the uncertainties along with the locations. This commonly involves plotting error bars [e.g., *Savage and Meyer*, 1985; *Mori*, 1989; *Willis-Richards*, 1990], although confidence ellipses are to be preferred [e.g., *Aster and Meyer*, 1988]. It is then perhaps possible to interpret the locations while taking account of the uncertainty. Two problems with this approach are that the diagrams can become overly complicated and difficult to interpret, especially with a large number of earthquakes, and that the method is not easy to implement in three dimensions.

Earthquake location data are well suited to computational methods of analysis, and a number of methods have been proposed to help interpretation by searching for patterns or for volumes of greater density within a seismic cloud. The simplest method is to use a "nearest neighbor" approach, stripping out locations which seem to detract from the structure, i.e., those that are not in volumes of dense seismicity. Alternatively, specific structures can be sought, perhaps based on a priori geophysical or geological information. Since earthquakes occur on faults, it is not surprising that the most sought after structures are planes. *Michelini and Bolt* [1986] have applied a method for determining hypocenter "trends" as a function of time. *Fehler et al.* [1987] have developed a search method for finding statistically significant planes of any orientation in a cloud of earthquake locations. Statistical techniques can also be used to check the compatibility of a location cloud with a plane of known

[1]Now at Montserrat Volcano Observatory, Montserrat, West Indies.

Paper number 96JB03739.
0148-0227/97/96JB-03739$09.00

orientation, say, a fault plane or a Wadati-Benioff zone [e.g., *Frohlich and Willeman*, 1987; *Michael*, 1989].

Here we propose a method that uses the location uncertainties to simplify the earthquake cloud; earthquake hypocenters are moved within bounds imposed by their location confidence ellipsoids. The aim is structural simplicity, and so a small volume of dense seismicity is simpler than a large volume of diffuse seismicity. In terms of the number of active faults present, one fault slipping twice is a simpler causative model than two separate faults each slipping once. This can be extended to any number of earthquakes by proposing that one fault slipping a number of times is simpler than a number of faults slipping once each. We can therefore generate a simplified cloud by moving hypocenters together as much as is possible within the limits of their uncertainties. There is no imposition of any a priori structures, and there is very little discarding of data.

Simplifying the Earthquake Cloud

Consider two hypocenters. If these have been determined using arrival times, each will have a spatial uncertainty ellipsoid, which is a three-dimensional probability density function. If the two ellipsoids only overlap for high confidence levels, then the obvious interpretation would be that the earthquakes were spatially distinct. If, however, the uncertainty ellipsoids overlap to a large degree for low confidence levels, then the simplest structural interpretation is that the two earthquakes occurred on the same fault at the same location. The technique developed below extends this concept to a large number of earthquakes, moving hypocenters to positions that simplify the earthquake cloud.

We assume that all errors are normally distributed. The probability density function is then nonzero over all space, although it tends to zero rapidly with distance. In order to make computations tractable, we truncate the uncertainty ellipsoids at some high confidence level. Thus for each hypocenter in the data set, only hypocenters which occur within its truncated confidence ellipsoid are considered.

The aim of the method is to try to "collapse" the locations of the earthquakes toward each other. The extent to which this collapsing can take place is governed by the location uncertainties for each earthquake and the distribution of surrounding locations. The location uncertainties are, in turn, a function of the data measurement uncertainties, the velocity model uncertainties, and the spatial distribution of seismometers. If the location uncertainties are sufficiently large in relation to the separation of the locations, then the earthquake cloud would be completely collapsible to a point and the arrival time data alone cannot disprove the null hypothesis that random normal uncertainties are all that is required to explain the location cloud.

Figure 1 attempts to illustrate how this approach might work with a real data set. It shows a plan view of earthquakes located by a local seismograph network at Rabaul Caldera, Papua New Guinea (details of the seismicity and the recording network are given below and by *Mori et al.* [1986]). The locations of 1363 earthquakes are shown in Figure 1, about one third of the full data set discussed below. The earthquakes excluded are those with the larger location uncertainties. Each earthquake is represented by a solid uncertainty ellipsoid, drawn at the 1 standard deviation level; we have used this level, as opposed to 4 standard deviations used in the method developed below, and only the better located earthquakes in Figure 1 for clarity. This figure is very instructive, once one takes into account that small

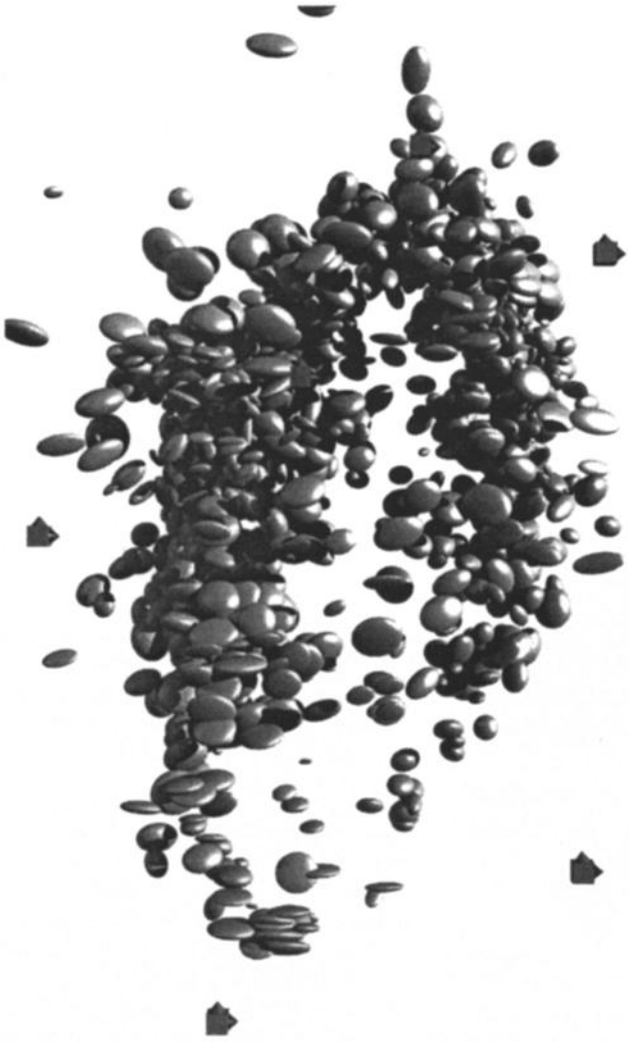

Figure 1. Schematic diagram of location uncertainties for the Rabaul earthquake data. The uncertainty ellipsoid for each earthquake is represented by a solid object. These have then been drawn using an optical raytracing algorithm with the light source shining over the right shoulder of the viewer. Station locations are represented by cubes with cones on the north, east, and lower faces. A total of 1,363 earthquake hypocenters were used in generating this diagram. Confidence ellipsoids are drawn at the 1 standard deviation level for clarity.

ellipsoids are more important than large ones. Clearly, the elliptical shape of the ring fault is not an artifact of the location process, as the uncertainty ellipsoids do not span it. However, any structures within the fault zones may or may not be real since the sizes of the uncertainty ellipsoids are similar to the apparent thickness of the fault zone. Graphically, Figure 1 is the equivalent of plotting error bars. Unfortunately, it is not possible to easily view this in three dimensions and so interpretation is limited.

Two other important features can be seen in Figure 1. First, the ellipsoids are seldom near-spherical. This is a network geometry effect; the observation that most of the ellipsoids are elongated east-west may well be linked to the network layout being elongated north-south [see *Mori et al.*, 1986]. Thus a set of closely spaced earthquake locations with no structure subject to random timing uncertainties could produce an ellipsoidal-

shaped cloud of hypocenters which could easily be misinterpreted. The second feature to note is that the size, shape, and orientation of the ellipsoids show significant variation. This is probably both due to changes made to the recording network with time and due to a lack of arrival time readings at all stations for some earthquakes. Such variations in the uncertainties will make any attempt at visual interpretation while allowing for the uncertainties very difficult. These two issues are almost universally apparent in microearthquake studies.

Collapsing Method

In order to develop the collapsing method, we need to consider the details of error ellipsoid estimation. It is the joint uncertainty distribution of the three spatial variables that is of greatest interest, as this determines the geometry of the spatial uncertainty ellipsoids and therefore the geometry of the joint spatial confidence ellipsoid. *Flinn* [1965] uses the *F* distribution to link the data uncertainties to the location-parameter joint confidence ellipsoids when the data variances are unknown. *Evernden* [1969], however, criticizes this approach for producing unrealistically large uncertainty ellipsoids and suggests the χ^2 distribution for the case where the variances are known. We follow *Evernden* [1969]; the joint uncertainty distribution of the three spatial errors, when each is normally distributed, and the data uncertainty variance is known, is a χ^2 distribution with 3 degrees of freedom (χ^2_3). The joint spatial confidence uncertainty ellipsoid for a certain confidence level signifies the confidence with which we can state that the true location lies within the ellipsoid. For a large number of locations the expectation would be that the number of true locations which lie within the confidence ellipsoids at the various levels of significance would be consistent with the probability density function of the uncertainty ellipsoids.

Normally distributed uncertainty in the data space implies normally distributed uncertainties in the solution space, in the region where the problem is linear [*Buland*, 1976]. Thus the location uncertainties in the data space have a χ^2_3 distribution. The data collapsing should therefore proceed until the distribution of hypocenter movements matches a χ^2_3 distribution as closely as possible. To allow for the fact that each earthquake is located with a different degree of accuracy, the movement is normalized in terms of the uncertainties of its own solution-space parameters. For the explanation that follows, it is assumed that the data uncertainties are known exactly. This search for a χ^2_3 distribution in the hypocenter movements removes any need for a subjective measure of "simplicity" in the earthquake cloud.

To apply the collapsing technique, we first require an estimate of the uncertainty for each hypocenter. The approach followed here is that used by *Aki and Richards* [1981, p. 694] for calculating model variance and confidence ellipsoids. The variance for the model parameters is related to the data uncertainty variance by the equation

$$(\mathbf{M}^T\mathbf{M}) = \sigma^2\mathbf{C}^{-1} \qquad (1)$$

where σ^2 is the data variance, **M** is the solution vector, and **C** is the covariance matrix given by $\mathbf{A}^T\mathbf{A}$; **A** is the matrix of partial differentials at the solution. Since we are only interested in the three-dimensional spatial uncertainty ellipsoid, the joint confidence ellipsoid for the three spatial solution parameters can be constructed from

$$(\mathbf{M}_s^T\mathbf{M}_s) = \sigma^2\mathbf{C}_s^{-1}\chi^2_3(\alpha) \qquad (2)$$

where the subscript **s** denotes the spatial partition and $\chi^2_3(\alpha)$ is the value of the χ^2 distribution with 3 degrees of freedom for some confidence level α.

The first step in the collapsing method is to calculate the three-dimensional spatial uncertainty ellipsoid which bounds the required confidence level, i.e., a certain number of standard deviations. We use 4 standard deviations. The data collapsing is then implemented iteratively, producing successive generations of hypocenters. It consists of two loops, an inner loop around all earthquakes and an outer loop over generations of collapsed hypocenters.

The inner loop consists of the following four steps and is repeated for each earthquake: (1) Consider an earthquake location, the object earthquake. (2) Find all other earthquakes whose locations lie within the surface of this earthquakes uncertainty ellipsoid. (3) Find the centroid for all these locations, including the object earthquake. All locations are given equal weighting. (4) Move the object earthquake toward the calculated centroid. Moving it only a fraction of the distance between the earthquake and the centroid seems to dampen oscillations in the data and reduce the number of iterations. We move the earthquake 0.61803 of the calculated distance, mimicking the golden section search in one dimension [*Press et al.*, 1986], although the actual value used does not seem to be critical. Let this new position be the object earthquake's location for the next generation; it is not considered in this iteration. The position and size of the uncertainty ellipsoid remain unchanged, centered on the original location of the object earthquake. Note that each location is treated in exactly the same manner, and so the order in which the data are considered is unimportant.

The outer loop consists of the following three steps and is repeated for each generation of collapsed hypocenters: (1) Create the next generation of earthquake hypocenters. (2) For each earthquake, calculate the distance from the original location to the current location in terms of standard deviations of the error ellipsoid for that earthquake. This is the earthquake's normalized measure of movement. (3) Compare the distribution of movements with the χ^2_3 distribution using the Kolmogorov test [*Press et al.*, 1986], which is based on the greatest difference in the cumulative probability distributions. Since the standard deviation of each earthquake is used to normalize the hypocenter movement, the theoretical distribution used is actually the square root of the χ^2_3 distribution. The outer loop is repeated until a minimum in the misfit between the distributions is reached. If locations are moved too far; that is, we go beyond the minimum and the misfit starts to increase, the process is taken back two generations, the fraction of the earthquake centroid distance moved is multiplied by 0.61803, and the process is restarted. Iteration proceeds until either the fraction moved has been reduced 4 times or the decrease in the misfit is less than 1%.

Some simple effects of this method can be imagined easily. Very isolated locations will be unaffected by the process. A pair of nearby locations will move to their common center. A location on the edge of a large group will be drawn toward that group.

The only variables in the collapsing method are the level of confidence used to truncate the confidence ellipsoid and the estimated variance in the data. For the former we use 4 standard deviations, the 99.86% confidence level, which seems appropriate for data sets of a few thousand earthquakes. We have

tried using 5 standard deviations, the 99.998% confidence level, but no significant change was produced, apart from an increase in computer run time. The sensitivity of the technique to the data variance is explored in detail below.

Synthetic Data Sets

Three simple synthetic data sets can be used to illustrate the collapsing method. In the first synthetic a cloud of 1000 hypocenters was generated by random perturbations of a single point in x, y and z; the perturbation distances were taken from a normal distribution. This simulates the smearing out of a single location by random location uncertainties. The resulting hypocenter cloud is shown in Figure 2a. We now apply the collapsing method described above to determine whether this cloud represents a seismically active volume or just one feature which is repeatedly activated; the latter explanation is our null hypothesis.

The results of applying the collapsing technique to this data set are shown in Figures 2b, 2c, and 2d for 1, 3, and 5 iterations respectively. A confidence level of 99.86% (4 standard deviations) was used to construct the uncertainty ellipsoids, and the data variance was, of course, known exactly. The distribution of actual data movements, in terms of standard deviations, is shown compared to the theoretical χ^2_3 distribution sought in Figures 2e to 2h.

After five iterations the two distributions match and the earthquakes have collapsed back to a point, as we expect. Thus the seismic cloud is incapable of disproving the null hypothesis

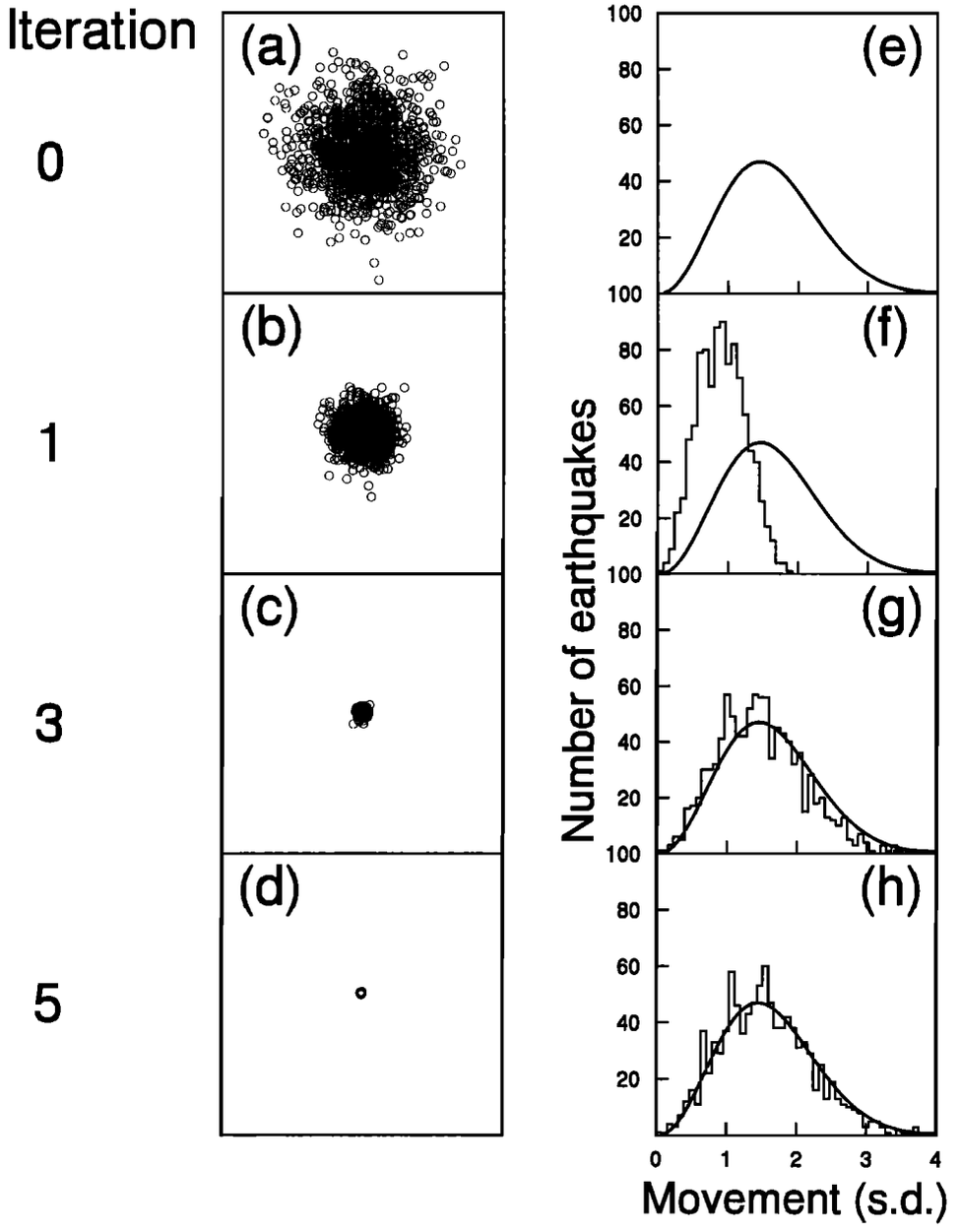

Figure 2. Application of the collapsing method to a synthetic data set consisting of 1000 earthquake locations perturbed from a common point. Views of (a) the original locations and the collapsed locations after (b) one, (c) three, and (d) five iterations and histograms of the (e) original hypocenter movements and those after (f) one, (g) three, and (h) five iterations, comparing them with the desired χ^2_3 distribution (solid curves).

that it could have been caused by a single feature failing repeatedly.

We can use the same figure to consider the effects of the technique on a finite-sized cloud, i.e., when the earthquakes are genuinely distributed in a volume. To do so, we use the same earthquake cloud but a lower data variance. If the variance were half of that used before, then the histogram of hypocenter movements would most closely match the χ^2_3 distribution when the cloud still had a volume and resembled that shown for one iteration. In this case the null hypothesis of a single feature repeatedly failing is disproved.

The second and third synthetic data sets are more complex and approximate realistic structures that may be encountered. They are a circle and a pair of intersecting lines and are shown at the top of Figure 3. The circle contains 1000 hypocenters, and the cross contains 2000. These have both been perturbed in three dimensions using random directions and random distances taken from a normal distribution. The perturbed locations, shown in the middle of Figure 3, retain the structures seen in the original data. The collapsed hypocenters are shown at the bottom of Figure 3. The data from the circle have collapsed back to a near-perfect replica of the original data. The only real discrepancy is that the collapsed circle is slightly smaller. This is because the density of earthquakes on the inside of the perturbed circle is greater than on the outside and so is more attractive during the collapsing process.

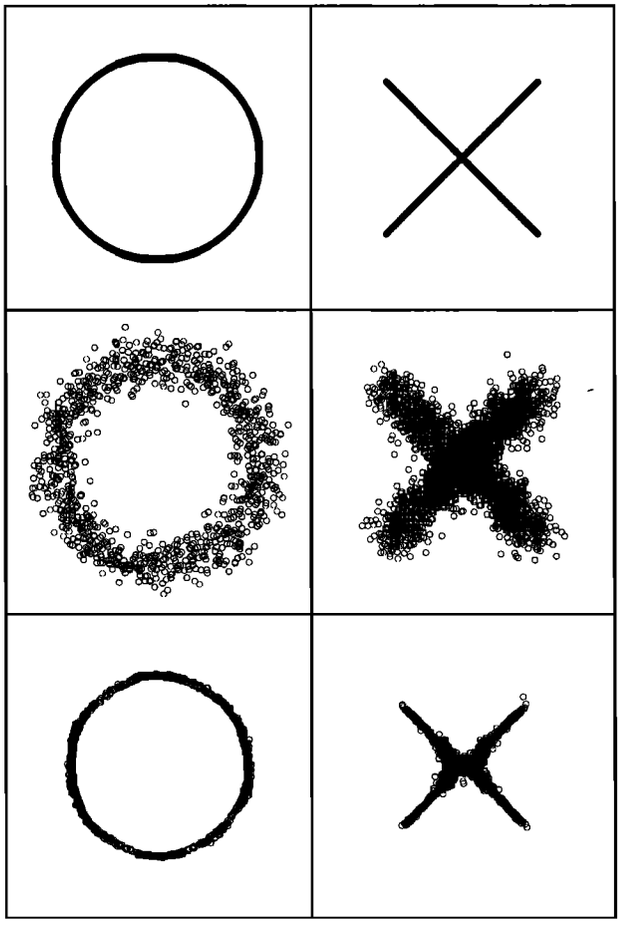

Figure 3. Application of the collapsing method to two synthetic data sets, (left) a circle of 1000 earthquakes and (right) two intersecting lines containing 2000 hypocenters, showing (top) original locations (middle) the perturbed locations, and (bottom) the collapsed locations.

The collapsed cross also replicates the features of the original data, with the arms of the cross having near-zero thickness for much of their length. The lengths of the two arms are noticeably smaller than in the original data. The reason for this is the same as for the shrinking of the circle: the density of hypocenters is significantly lower at the ends of the arms. The collapsing process is less successful near the intersection of the two lines. This is because some of the perturbed hypocenters have moved into the cloud generated by the adjacent arm. They then collapse back into the wrong line. This illustrates that there is a limit to the size of features that can be extracted from the data. It is interesting to note that the collapsing has generated a small artifact in that the center of the cross seems to be a short, horizontal line. Repeated trials with different data sets showed that this was not systematic, the line was as likely to be vertical as horizontal.

The above discussion illustrates that the method is useful when the data variance is known exactly. For real data, of course, the data variance for each earthquake is not known exactly, but we do have an estimate for it, the standard error, the sum of the squares of the residuals divided by (n-4), where n is the number of timings. This is thought to be reasonable for locations inside a network [*Peters and Crosson,* 1972]. The reliability of the standard error as an estimate for the data variance is examined below.

Rabaul Data Set

Seismicity in Rabaul Caldera, Papua New Guinea, has been monitored using a local network since 1967. Earthquakes have magnitudes M_L ranging from negative values up to about 5 [*Mori et al.*, 1989]. Earthquake occurrence rates are highly variable and swarms are common, as in the 1983-1985 "seismo-deformational crisis" [*McKee et al.*, 1984; *Mori et al.*, 1986, 1989]. The distribution of the seismicity defines an elliptical caldera ring fault, as mentioned above. The earthquake locations have been used to study the structure of the caldera ring fault. In order to reduce the scatter on the seismicity plots, *Mori and McKee* [1987] only showed earthquakes recorded at seven or more stations, with horizontal location errors smaller than 1 km and with vertical errors less than 1.5 km. They then showed that the ring fault defined by the earthquake locations has an outward dip. Here we apply the collapsing technique described above to investigate the finer structure of the caldera ring fault.

The Rabaul data used here consist of 4442 earthquakes recorded between 1971 and early 1992. Earthquakes were recorded on a Develocorder (strip film recorder), and all picks were made manually by analysts at Rabaul Volcano Observatory (RVO). This has resulted in a very consistent data set in terms of the picking errors. Clear S arrivals are very rare, as all stations use just a vertical sensor, and so only P picks are used in our analysis. The seismometer network used to record the Rabaul data [*Mori et al.*, 1986] has grown through time, from seven stations in 1971 to eleven in 1992. Locations of nine stations are shown in Figure 1 (the other three are outside the diagram). The location uncertainties will have changed along with the network; the collapsing method accommodates such changes in the network.

For our analysis all the earthquakes were first relocated using the joint-hypocenter determination (JHD) method of *Frohlich* [1979], which allows the analysis of very large data sets. This

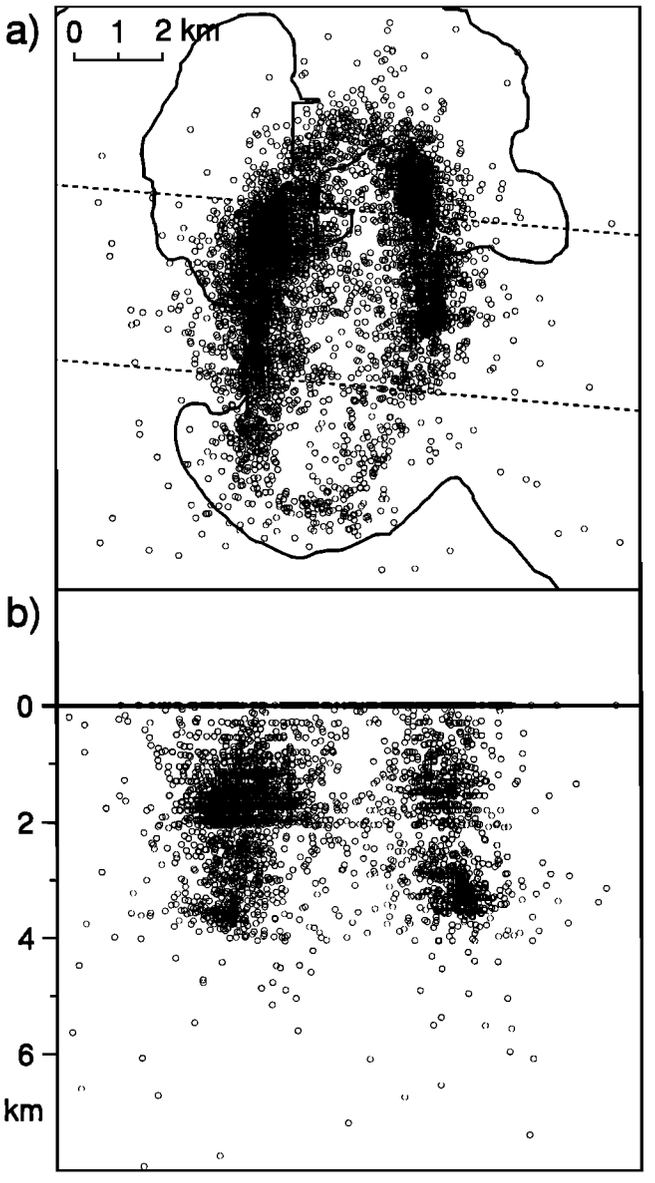

Figure 4. Earthquakes recorded at Rabaul Caldera, Papua New Guinea, between 1971 and early 1992. Data have been relocated using the joint-hypocenter method of *Frohlich* [1979]. (a) Plan view of all 4442 epicenters. The coastline is shown for reference. (b) Sectional view from an azimuth of 185°, i.e., looking along the axis of the caldera from the south. The section used is a 4 km thick slice through the middle of the seismicity and is delimited by the dashed lines in Figure 4a. Note that the scales of Figures 4a and 4b are different.

removes systematic errors associated with individual stations as the station residuals are allowed to change, with the restriction that the aggregate value of the station corrections must remain constant throughout the iterative process. The station corrections determined by JHD are very similar to those used in routine locations [*Mori*, 1988]. Figure 4 shows the hypocenters determined by the JHD relocation in plan and section view. The JHD locations are little different from the original locations the data in Figure 4 have more scatter than those shown by *Mori and McKee* [1987] since we show all the hypocenters and they show only the better ones, i.e., those with horizontal errors less than 1 km and vertical errors less than 1.5 km.

Table 1. Rabaul Velocity Structure

Layer	Depth to Top of Layer, km	*P* Wave velocity, km/s
1	0.00	1.70
2	0.25	2.06
3	0.50	2.43
4	0.75	2.79
5	1.00	3.15
6	1.25	3.51
7	1.50	3.88
8	1.75	4.24
9	2.00	4.60
10	4.00	6.11

Data are after Almond and McKee [1982].

The sectional view (Figure 4b) shows some artifacts of the location process; poorly determined depths which are probably due to inadequacies in the velocity model (Table 1). First, there are some earthquakes with a depth of 0 km. These are probably very shallow earthquakes but cannot be properly located owing to incorrect velocities in the upper layer or layers of the model. There is also a distinct banding of the earthquakes at different depths, coincident with boundaries in the velocity model. Note that neither of these artifacts is seen in the data presented by *Mori and McKee* [1987], presumably owing to their more selective data set. The final point to note about the data in Figure 4b is that the earthquakes appear to show a fairly abrupt cutoff at a depth of 4 km. This has been remarked upon [*Mori et al.*, 1989] as being due to the termination of the caldera-collapse ring fault at the top of the underlying magma chamber. It is important to note, however, that there is a major discontinuity in the velocity model at 4 km depth (Table 1).

Figure 5 explores the effect of the depth of this discontinuity on the hypocenters. It shows a view, from the south, of the joint-hypocenter determination for three different velocity models. These are the standard Rabaul velocity model, where the interface is at a depth of 4 km, and two similar models, where the interface is at depths of 5 and 3 km. It is clear from Figure 5 that the depth of the base of the seismicity depends on the depth of the velocity interface. The reason for this becomes clear after close examination of the location results; the majority of the deep earthquakes have at least one station where the first arrival is treated by the location software as a head wave from this interface.

The provenance of the model listed in Table 1 is not entirely clear. *Mori et al.* [1986, 1989] state that it is "after *Almond and McKee* [1982]." *Almond and McKee* [1982] carried out a refraction survey in Rabaul harbor and observed a clear velocity gradient with depth. *Almond and McKee* [1982, p.12] proposed a velocity model which had a continuous velocity increase with depth in the top 2 km and then a constant velocity of 4.11 km/s which was "assumed to continue down to a depth of 6 km"; they assumed that no earthquakes were deeper than 6 km. The layered velocity model was introduced with the use of standard location software at RVO. The top eight layers of the model clearly mimic the *Almond and McKee* [1982] velocity gradient, and the deepest interface is probably an attempt to incorporate the results from larger-scale refraction surveys carried out by *Cifali et al.* [1969] and *Finlayson and Cull* [1973]. Both of these showed a major refractor whose depth and velocities varied with location.

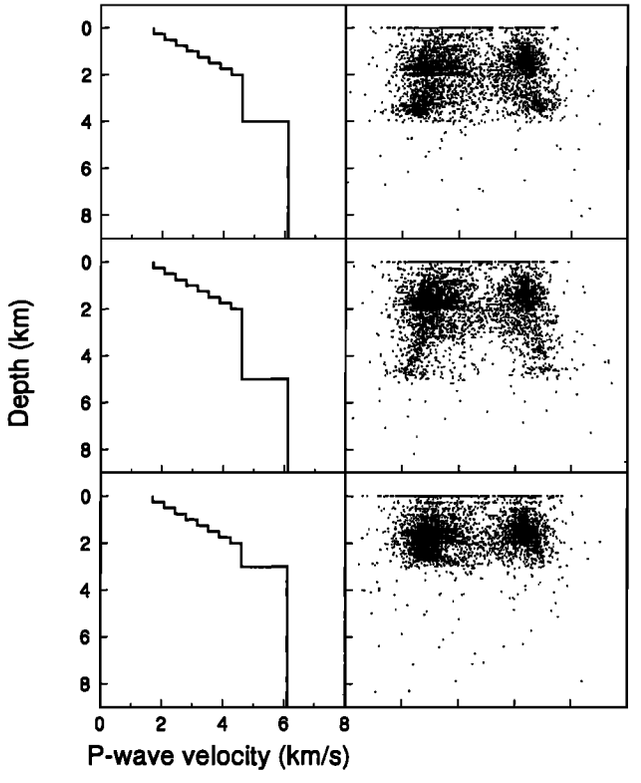

Figure 5. (left) Three different velocity models for Rabaul and (right) the resulting joint-hypocenter locations viewed from the south. The models are the (top) standard model, where the major interface is at a depth of 4 km, and two similar models where the interface is at (middle) 5 km and (bottom) 3 km.

In Rabaul, *Cifali et al.* [1969] put this at a depth of 6 km and *Finlayson and Cull* [1973] put it at 4 km. The model is probably based on the latter study (their refractor had a velocity of 6.11 km/s), but it cannot be regarded as accurate.

It is not possible to construct a more appropriate velocity model for Rabaul without a detailed inversion of the data or further experimental work. This is outside the scope of this paper, and we therefore persist with the velocity model given in Table 1 and presume that the locations are not grossly in error due to its inadequacies. Because of this, any interpretation of the results, especially the earthquake depths, will need qualification.

In order to apply the collapsing method to the data shown in Figure 4, the spatial uncertainty ellipsoid for each earthquake must be calculated from the data variance using (2). In the development of the method it was assumed that the data variances were known exactly. Data variances are not known for the Rabaul data, and so we need to estimate them.

Pavlis [1986] considers the following three sources of uncertainty in the regional location problem: measurement uncertainties, Earth model uncertainties, and uncertainties due to the nonlinearity of the problem. As pointed out above, there do seem to be significant Earth model uncertainties in the Rabaul location data. Since we have no formal way to treat these, we are forced to ignore them, along with any uncertainties due to nonlinearity. This is consistent with the generally conservative approach adopted here, and any contribution to the uncertainties from these sources is treated as a random uncertainty rather than a systematic one and so will tend to increase the location uncertainties.

This leaves us to consider the measurement uncertainty. There are a number of different approaches that could be adopted here. We have chosen the simplest way to estimate the variance for each earthquake; this will undoubtedly vary with earthquake size, the network used, etc.. We use the standard error for the location residuals for individual earthquakes, i.e., the sum of the squares of the residuals divided by n-4, where n is the number of P timings used to locate the earthquake.

The analysis of the Rabaul data proceeded for nine iterations until a satisfactory fit between the hypocenter movements and the $\sqrt{\chi^2_3}$ distributions was obtained, shown in Figure 6. The resulting locations are shown in Figure 7, in the same plan and section views as Figure 4. We can immediately see that the method has, as hoped, sharpened up the view of the caldera ring fault.

A detailed interpretation of the data in Figure 7 is only valid within the context of other data from Rabaul Caldera and is beyond the scope of this paper; it will be reported elsewhere. We only discuss some of the more important features.

Previous interpretations of the Rabaul location data [e.g., *Mori and McKee*, 1987] refer to the ring fault as a single structure. This requires the assumption that the width of the seismic zone can be fully explained by a planar structure and location uncertainties. The collapsing method empirically tests this assumption. The collapsed seismicity in Figure 7 shows that the single-fault model is appropriate for most of the seismogenic zone.

The collapsed seismicity also shows that the ring fault is not as simple as had been previously interpreted. Instead of one elliptical fault, there appears to be an outer elliptical fault which embraces a smaller inner elliptical fault at its northern end. The two ring faults are at different depths, with the inner one shallower. In the section shown in Figure 7b, we can see that the ring fault generally dips outward as proposed by *Mori and McKee* [1987]. However, this dip is not constant with depth and the eastern wall appears to have a near-vertical upper section.

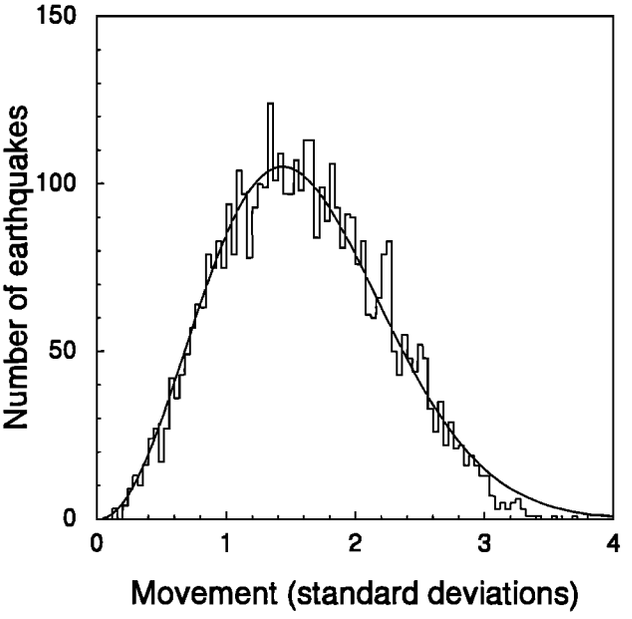

Figure 6. Histogram of the distribution of movements obtained when applying the collapsing method to the Rabaul earthquake data (result after nine iterations). The earthquake movements are expressed in terms of standard deviations. The solid curve shows the χ^2_3 distribution sought in the method.

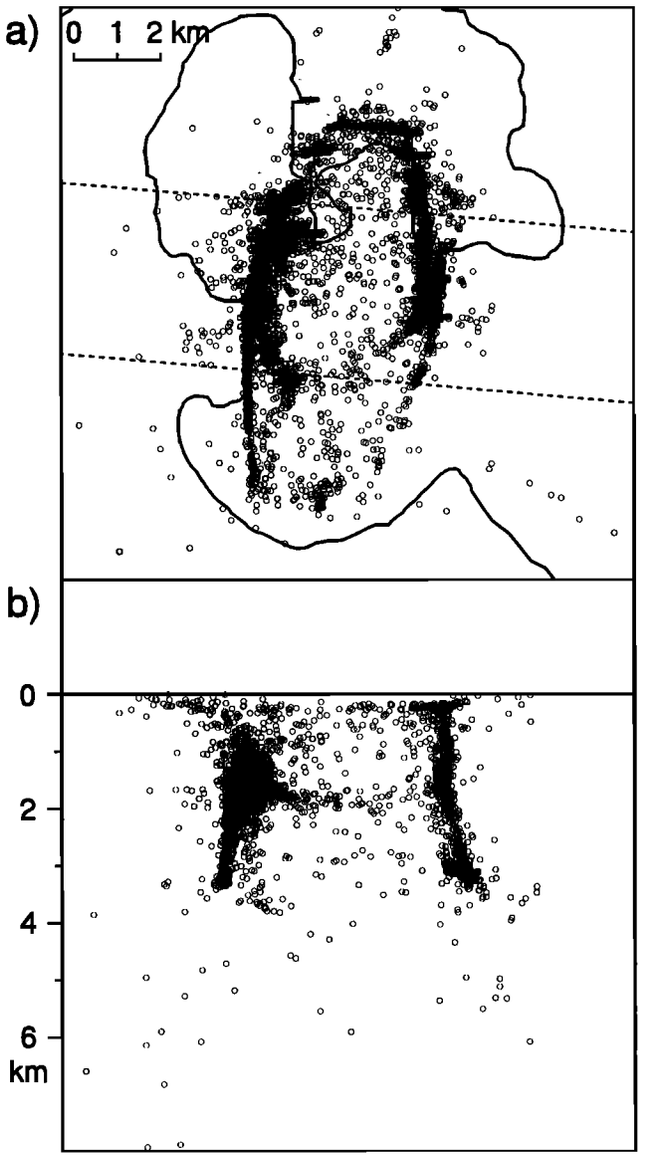

Figure 7. Earthquakes recorded at Rabaul Caldera after application of the collapsing method showing (a) plan view and (b) sectional view, as in Figure 4.

Sensitivity of the Method

Two assumptions are made when applying the collapsing method, the confidence level used in the analysis and the variance of the observational data, i.e., the timing uncertainties. Trials with higher confidence levels for truncation appear to have little effect on the results. However, we can see from (2) that the variance affects the size of the confidence ellipsoid used. We can therefore explore the likely effects of changes in the variance by systematically scaling the error ellipsoids in the analysis.

Figure 8 shows, in plan view, the collapsed hypocenters for different analyses of the Rabaul data, varying the uncertainty ellipsoid semimajor axes by scaling factors ranging from 0.25 to 4.0, relative to that used above. These are equivalent to scaling the variance by the squares of these factors. The original cloud of earthquakes is also shown in Figure 8. For large scaling factors there is a tendency for the method to shrink the overall size of the cloud moving hypocenters toward the middle of the plot, as seen in the synthetic circular data set. This, however, does not change the shape or internal structure of the cloud. In order to counteract this unwanted effect, the final locations have been moved away from the center so that the mean distance to the center is unchanged by the collapsing process.

Figure 8 shows that the amount of movement is related to the confidence ellipsoid size; if the scaling factor is small, then hypocenters cannot move and so the image is not modified. If the factor is large, then the data do not just collapse to a more structured image, as might be expected. Instead, as the scaling factor increases, a complex image again results. The increase in complexity appears to develop when the truncated confidence ellipsoids of many of the locations span the whole caldera and so the less accurate locations in both walls start to move together. The data look most structured and simple for an inflation factor between 0.8 and 1.25.

This result suggests two things about the effects of the estimated variance levels on the collapsing method. First, the method is not overly sensitive to the values assumed. Second, the values used are probably appropriate; that is, the standard error is a sufficiently reliable estimate of the data variance.

Conclusions

We have developed a technique in this paper which uses a simple attraction rule to "collapse" a set of earthquake locations. The rule is applied iteratively until the movements are consistent with the estimated uncertainties. The aim of the technique is to find the simplest structural interpretation which is consistent with all of the location data, in accordance with Occam's Razor: "it is vain to do with more what can be done with less" [*Mackay*, 1991, p. 185]. The interpreter should be attempting to produce the simplest structural interpretation i.e., the fewest number of structures consistent, with all of the data.

It is felt that this method is to be preferred over more traditional methods of examining the data by eye or searching for specific structures within the cloud. We discard very little of the available data and make no a priori assumptions about the interpretation. The method works well with synthetic and real data. The synthetic data sets show the strengths of the technique and some of its limitations, mainly, that it is not possible to resolve the very fine detail that has been smoothed out by the location errors. The real data come from Rabaul Caldera in Papua New Guinea, and we have used the method to generate an image of the caldera ring fault which lends new insight to the processes at Rabaul. It is not within the scope of this paper to discuss the detailed interpretation of that picture, but the existence of two ring faults rather than one is revealed.

There are two main assumptions made when applying the method; the confidence limit used to truncate the uncertainty ellipsoid and the standard error is a reliable estimate of the variance of the observational data. We have had success using a reasonable level for the confidence limit. The results of a sensitivity analysis support the second assumption, at least for the Rabaul data set.

The method does not assume that everything must collapse back to simple features. The synthetic data can be used to show that if the true locations were diffuse, then they would be collapsed back to a diffuse zone, not a point, line, or plane. The technique therefore may be used to resolve whether seismicity does fall on sharp features or not.

We have concentrated on the use of the method with earthquake data recorded on a small local seismograph network.

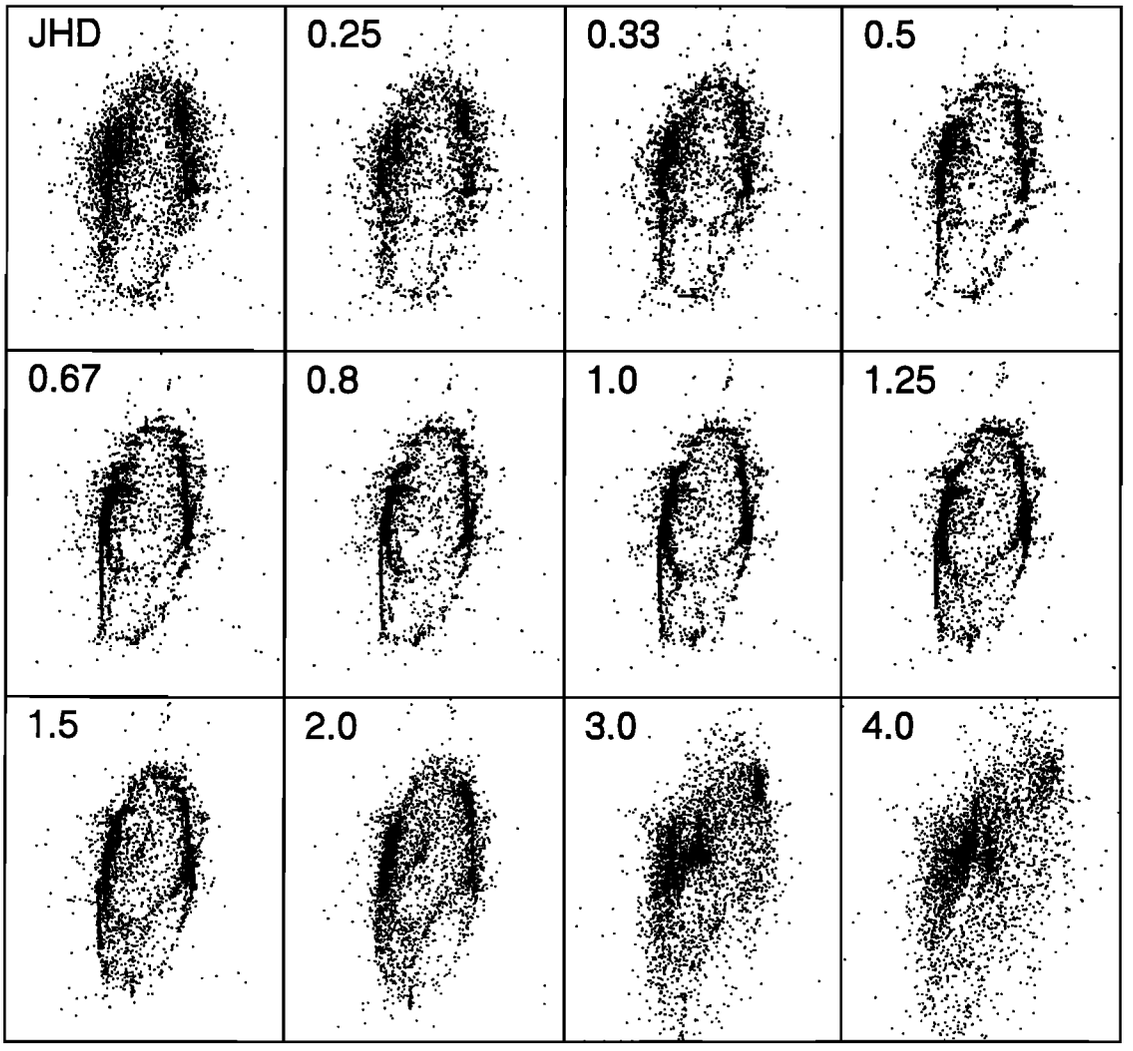

Figure 8. Plan views of collapsed epicenters for different size uncertainty ellipsoids. The scaling factor (see text for details) ranges from 0.25 to 4.0. The original joint-hypocenter locations are shown at top left for comparison.

The method is applicable to any set of seismic location data for which uncertainty estimates are available.

Acknowledgments. We would like to thank Hiroaki Niitsuma of Tohoku University, who has encouraged us in this work. We acknowledge the record readers employed at Rabaul Volcano Observatory (RVO) who have, over the years, produced a very consistent set of arrival time data. Most of the work described here was carried out while R.S. was working at RVO.

References

Aki, K., and P.G. Richards, *Quantitative Seismology*, vol. 2, W. H. Freeman New York, 1980.

Almond, R.A., and C.O. McKee, Location of volcano-tectonic earthquakes within the Rabaul caldera, Rep. 82/19, Geol. Surv. of Papua New Guinea, Port Moresby, NCD, 1982.

Aster, R.C., and R.P. Meyer, Three-dimensional velocity structure and hypocenter distribution in the Campi Flegrei caldera, Italy, *Tectonophysics, 149*, 195-218, 1988.

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

Cifali, G., G.W. d'Addario, E.J. Polak, and W.A. Wiebenga, Rabaul preliminary crustal seismic test, New Britain 1966, *Aust. Rec. 1969/125*, Bur. of Miner. Resour., Canberra, ACT, 1969.

Douglas, A., Joint epicentre determination, *Nature, 215*, 47-48, 1967.

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

Fehler, M., L. House, and H. Kaieda, Determining planes along which earthquakes occur: Method and application to earthquakes accompanying hydraulic fracturing, *J. Geophys. Res., 92*, 9407-9414, 1987.

Finlayson, D.M. and J.P. Cull, Time-term analysis of New Britain / New Ireland island arc structures, *Aust. Rec. 1973/21*, Bur. of Miner. Resour., Canberra, ACT, 1973.

Flinn, E.A., Confidence regions and error determination for seismic event location, *Rev. of Geophys., 3*, 157-185, 1965.

Frohlich, C., An efficient method for joint hypocenter determination for large groups of earthquakes, *Comput. Geosci., 5*, 387-389, 1979.

Frohlich, C., and R.J. Willeman, Statistical methods for comparing directions to the orientations of focal mechanisms and Wadati-Benioff zones, *Bull. Seismol. Soc. Am., 77*, 2135-2142, 1987.

Mackay, A.L., *A Dictionary of Scientific Quotations*, Adam Hilger, Bristol, England, 1991.

Matsu'ura, M., Bayesian estimation of hypocenter with origin time eliminated, *J. Phys. Earth, 32*, 469-483, 1984.

McKee, C.O., P.L. Lowenstein, P. de Saint Ours, B. Talai, I. Itikarai, and J.J. Mori, Seismic and ground deformation crises at Rabaul caldera: Prelude to an eruption?, *Bull. Volcanol., 47*, 397-411, 1984.

Michael, A.J., Spatial patterns of aftershocks of shallow focus earthquakes in California and implications for deep focus earthquakes, *J. Geophys. Res., 94*, 5615-5626, 1989.

Michelini, A., and B.A. Bolt, Application of the principal parameters method to the 1983 Coalinga, California aftershock sequence, *Bull. Seismol. Soc. Am., 76*, 409-420, 1986.

Mori, J., Re-location of Rabaul caldera earthquakes using S-wave data, *Rep. 88/20* Geol. Surv. of Papua New Guinea, Port Moresby, NCD, 1988.

Mori, J., The New Ireland earthquake of July 3, 1985 and associated seismicity near the Pacific-Solomon Sea-Bismarck Sea triple junction, *Phys. Earth Planet. Inter., 55*, 144-153, 1989.

Mori, J., and C.O., McKee, Outward-dipping ring-fault structure at Rabaul caldera as shown by earthquake locations, *Science, 235*, 193-195, 1987.

Mori, J., C.O. McKee, I. Itikarai, P. Lowenstein, P. de Saint Ours, and B. Talai, Account and interpretation of the seismicity during the Rabaul seismo-deformational crisis September 1983 to July 1985, *Rep. 86/26*, Geol. Surv. of Papua New Guinea, Port Moresby, NCD, 1986.

Mori, J., C. O. McKee, I. Itikarai, P. Lowenstein, P. de Saint Ours, and B. Talai, Earthquakes of the Rabaul seismo-deformational crisis September 1983 to July 1985: Seismicity on a caldera ring fault, in *IAVCEI Proceedings in Volcanology*, vol. 1, *Volcanic Hazards*, edited by J.H. Latter, pp. 429-462, Springer-Verlag, New York, 1989.

Pavlis, G.L., Appraising earthquake hypocenter location errors: A complete practical approach for single-event locations, *Bull. Seismol. Soc. Am., 76*, 1699-1717, 1986.

Peters, D.C., and R.S. Crosson, Application of prediction analysis to hypocenter determination using a local array, *Bull. Seismol. Soc. Am., 62*, 775-788, 1972.

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

Sambridge, M., and K. Gallagher, Earthquake hypocenter location using genetic algorithms, *Bull. Seismol. Soc. Am., 83*, 1467-1491, 1993.

Savage, M.K., and R.P. Meyer, Aftershocks of an M=4.2 earthquake in Hawaii and comparison with long-term studies of the same volume, *Bull. Seismol. Soc. Am., 75*, 759-777, 1985.

Spence, W., Relative epicenter determination using P-wave arrival-time differences, *Bull. Seismol. Soc. Am., 70*, 171-183, 1980.

Willis-Richards, J., The effect of a high heat production granite batholith on lithospheric stress and strength in south west England, in *Rock at Great Depth, vol. 3*, edited by V. Maury and D. Fourmaintroux, pp. 1369-1378, A.A. Balkema, Rotterdam, Netherlands, 1990.

R.H. Jones, CSM Associates Ltd., Rosemanowes, Herniss, Penryn, Cornwall, TR10 9DU, England. (e-mail: rjones@csm.ex.ac.uk)

R.C. Stewart, Montserrat Volcano Observatory, Montserrat, West Indies.

(Received April 10, 1996; revised October 25, 1996; accepted November 26, 1996.)