Geophys. J. R. astr. Soc. (1986) **87**, 679–697

A novel method of hypocentre location

M. S. Sambridge and B. L. N. Kennett *Research School of Earth Sciences, Australian National University, GPO Box 4, Canberra, ACT 2601, Australia*

Accepted 1986 May 16. Received 1986 May 16; in original form 1986 March 6

Summary. The determination of earthquake locations requires a good velocity model for the region of interest, appropriate statistics for the residuals encountered and an efficient, stable inversion algorithm.

A direct nonlinear inversion scheme has been constructed which can use any velocity model for which travel times can be calculated from an arbitrary source position to the receivers in the seismic network. The procedure is based on the minimization of a misfit function depending on the residuals between observed and calculated arrival times. Different statistics, e.g. Gaussian and Jeffreys distributions, can be accommodated by the choice of misfit function. The algorithm is based on a directed grid search which narrows down the range of possible origin times whilst carrying out a spatial search in the neighbourhood of the current minimum of the misfit function. No numerical differentiation of travel times is required, and convergence is rapid, stable and tolerant of occasional large errors in reading observed travel times. A useful product of the method is that the misfit function values are available in the neighbourhood of the minimum, so that a fully nonlinear treatment of the statistical confidence regions for a particular location can be made.

A prerequisite for the use of the algorithm is the delineation of bounds on the four hypocentral parameters. Epicentral bounds are constructed using a variant of the 'arrival order' technique, and rapid scanning in depth and origin time over this region yields useful bounds on these parameters.

The new nonlinear algorithm is illustrated by application to the SE Australian seismic network, for an event in the most active seismic zone. Two different velocity models are used with both Gaussian and Jeffreys statistics and good convergence for the algorithm is achieved despite significant non-linearity in the behaviour. The Jeffreys statistics are more tolerant of large residuals and are to be preferred when the requisite velocity model is not too well known.

Key words: hypocentre location, nonlinear inversion

1 Introduction

Most earthquake location procedures commonly used today may be viewed in terms of an optimization problem. Although an underdetermined problem, due to our lack of knowledge of the real Earth, the nonlinear inversion of seismic data for hypocentral parameters is generally treated as one of overdetermined type. This is achieved by assuming knowledge of the compressional and shear wave velocity structure of the Earth in the form of a fixed seismic velocity model, which, therefore, reduces the number of unknowns in the problem to the four hypocentral parameters of an earthquake. The data consist of observed first arrival times of the body wave phases of an event (say $t_{0i}, i = 1 \ldots N$) at a network of seismic stations in the area. By using the assumed velocity structure with an initial set of hypocentral parameters (h_k, $k = 1, 4$), theoretical arrival times (t_{ci}) may be calculated. As an indication of the success of the trial hypocentre we may produce an error or 'misfit' statistic C using observed and calculated arrival times. If we define the 'best' or solution set of hypocentral parameters as that combination which produces a minimum misfit statistic, then the location problem reduces to the optimization of C in the hypocentral parameter space.

The non-linearity in the problem arises because of the non-linear dependence of the travel times of seismic waves with respect to variations in both the hypocentral and velocity model parameters. The commonly used misfit statistic is an L_2-norm or squared residual statistic

$$C = \sum_{i=1}^{N} [r_i(\mathbf{h})]^2, \quad r_i(\mathbf{h}) = (t_{oi} - t_{ci})/\sigma_i, \tag{1.1}$$

where σ_i is the standard deviation of the i-th datum. It can easily be shown (Buland 1976) that the least-squares optimization is equivalent to assuming a Gaussian distribution for the picking errors on the observed arrival times with no cross-correlation of errors (see Fig. 1)

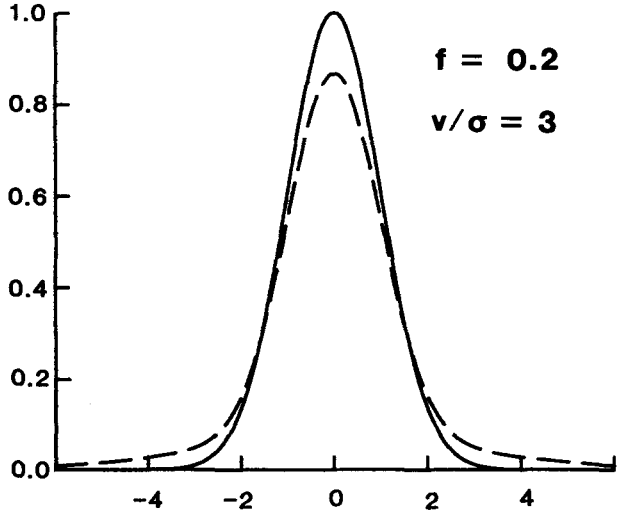

Figure 1. Comparison of the Gaussian and Jeffreys distributions for picking errors. The Gaussian is shown in a solid line and the Jeffreys distribution which imposes a wider Gaussian beneath the normal distribution is shown in a dashed line.

i.e. a probability distribution

$$\exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(t_{oi}-t_{ci})^2/\sigma_i^2\right\}. \tag{1.2}$$

The squared residual statistic (1.1) is usually minimized by a Gauss–Newton type of algorithm, or a related variant. The predominant techniques used in the hypocentre problem at the present time are: (a) Gauss–Newton, least-squares

$$\delta\mathbf{x} = -[\mathbf{A}^T\mathbf{A}]^{-1}\mathbf{A}^T\mathbf{r} \tag{1.3}$$

(b) Damped least-squares

$$\delta\mathbf{x} = -[\mathbf{A}^T\mathbf{A} + \theta^2\mathbf{A}]^{-1}\mathbf{A}^T\mathbf{r} \tag{1.4}$$

where $\delta\mathbf{x}$ is the estimated adjustment vector for the hypocentral parameters, $\mathbf{r}$ is the residual vector and $\mathbf{A}$ is the $(n \times 4)$ matrix of partial derivatives of the residual vector with respect to the hypocentral parameters. Both these algorithms define an improvement to an existing vector, $\mathbf{x}$.

The matrix $\mathbf{A}$ describes the way in which the travel times (and, hence, the residual vector) are related to the hypocentral parameters. Due to the nonlinearity of this dependence $\mathbf{A}$ is itself a function of the hypocentral parameters and varies with position. The problem is therefore solved iteratively. The numerical problems associated with the matrix inversions in (1.3), (1.4) are well known (Lee & Stewart 1981). A local linearization is made at each iterative step and derivatives of travel times constitute the entries of $\mathbf{A}$. In cases where the data poorly constrain the parameters, such as areas outside the network of stations, the similar dependence of derivatives at different stations on the hypocentre parameters may cause the matrix $\mathbf{A}^T\mathbf{A}$ to become rank defective. In this case an attempt to calculate its inverse as in (1.3) leads to wild oscillations in the estimated adjustment vector $\delta\mathbf{x}$, and, hence, numerical instabilities occur. The instability may be avoided by using a damped least-squares approach as in (1.4) but now θ must be chosen carefully to avoid a reduction in parameter resolution.

Another factor common to earthquake location procedures is that events are located relative to a fixed Earth model and all errors associated with that model will be treated as errors in the position of the event, with a consequent shift in the hypocentre. This has the result that the minimum of the misfit statistic, which we consider as the goal of the optimization process, will only give us a solution which is as good as the velocity model will allow. The statistic used to estimate the picking errors (which is usually assumed to be Gaussian) is in fact inherently being used to model the errors in our velocity model, as compared with the real Earth, which may be far from Gaussian.

However, attempts to use other error statistics, or combinations of error distributions are very uncommon. Jeffreys (1932) introduced a modification to the Gaussian distribution which he noted to be representative of teleseismic traveltime residuals. The Jeffreys distribution, which consists of a Gaussian combined with a smaller background Gaussian of greater width (see Fig. 1), is known to be robust (unlike the simple Gaussian distribution) i.e. it tolerates occasional outlying values in the residuals. It is less influenced by a single relatively large station residual which may arise from a misrecording of a time pick or, perhaps, an anomaly in the velocity model error due to local lateral heterogeneity. An attempt to model both observational and velocity model errors simultaneously has been made by Tarantola & Valette (1982). They combined probability density functions (p.d.f.'s) of both observational and velocity model errors and then integrated over origin-time to

produce a statistic which is independent of the origin-time of the event. Few attempts seem to have been made to incorporate these variations on the Gaussian statistic into routine location algorithms, most probably due to the difficulty of formulating the problem in terms of the usual matrix inversion procedure.

In this paper we describe a fully non-linear hypocentral location algorithm which is completely independent of the type of statistics employed. It may, therefore, be used to examine the effect the chosen statistic has on the final location of events. No local linearization of the problem is made and derivatives of travel times are not required. All matrix inversions are avoided and so the method has beneficial stability in cases where the hypocentral parameters may be ill-constrained.

The algorithm falls into the class of grid search methods, which are usually deemed too inefficient in cases where more than one dimension of parameter space must be searched. However, by separating the spatial and temporal searches and using some simple optimization techniques, we have been able to produce an efficient and robust algorithm. As with most search type algorithms, a function evaluation is the only type of repetitive information used and so the resulting method is rather general in nature. Just as one may choose any type of objective function to be minimized, one is at liberty to replace the velocity model with any other desired, including 3-D laterally heterogeneous models. This full non-linear approach to inversion allows us to make a direct estimate of the size of the confidence region about the solution. We are able to examine the distortions in the confidence region imposed by the nonlinearity in the problem, and also look at the relative constraints placed on the solution in the epicentral and depth planes.

The method requires an initial region of hypocentral parameters on which the grid search procedure acts. Although this region need not necessarily enclose the actual hypocentre of the event, the size and accuracy of the initial region affects the rate of convergence of the algorithm. A technique has therefore been devised which quickly generates a fairly accurate first estimate of the region of parameter space containing the solution, which will be described before the treatment of the main algorithm.

The new algorithm is illustrated by application to hypocentral estimates for local and regional events using the network of seismic stations in southeastern Australia. However, once a suitable travel time routine has been supplied for the phases picked from the seismograms, the algorithm may be used for teleseismic problems or in regions of strong lateral heterogeneity.

2 Hypocentral bounds: the arrival order method

All earthquake location algorithms require an initial guessed set of hypocentral parameters, which will hopefully be improved during the course of the inversion. In this section we describe a procedure which leads to upper and lower bounds on all four hypocentral parameters.

The method employs the time picks of the first arriving *P*-wave phases at each station and, where available, the corresponding initial *S*-wave times. A quick, robust and fairly accurate estimate of a region of parameter space surrounding the hypocentre is generated and used as the starting point for the main nonlinear location algorithm described below.

2.1 EPICENTRAL BOUNDS

The estimation of bounds on the epicentral position is based on a variation of the technique proposed by Anderson (1981) using the order of arrival of the *P*-waves at the different

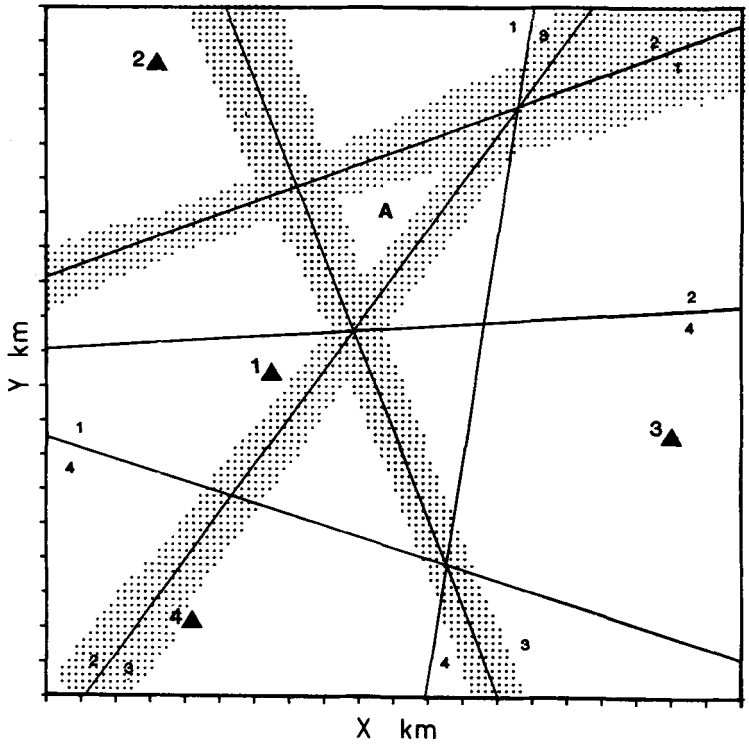

Figure 2. Construction of arrival order zone for a theoretical network of four stations, including the effect of picking errors. The constraints are labelled by the pairs of station used.

stations of the network. The basic assumption is that for any particular phase those stations with the earliest time picks are closest to the epicentre.

The system is best illustrated by a simple example (Fig. 2) using a four station network. The stations are numbered according to the arrival order of the *P*-waves. Each pair of stations then provides a geometrical constraint on the position of the epicentre. If we take the two earliest stations 1 and 2 then the epicentre must lie closer to 1 than 2, i.e. the epicentre must lie below the perpendicular bisector of the line joining 1 and 2. Similarly taking all the other possible combinations of pairs of stations in turn, we define a set of geometrical constraints on the epicentre location. We select the region which satisfies the maximum number of constraints, region A in Fig. 2 which is determined by the relative arrival times of the pairs 1 and 2, 2 and 3, 3 and 4.

With an n station network there are $n(n-1)/2$ such geometrical constraints and we would choose the zone satisfying the maximum number of constraints as that most likely to contain the actual epicentre. Inaccurate time picks may mean that some stations get interchanged from the true order of arrival and so prevent the constraints from being entirely consistent.

Anderson (1981) seeks only an initial epicentral solution and takes the vertex of the zone satisfying the largest number of constraints which lies nearest to the station which detects the event first. The problem then becomes one of linear programming and is solved using a Simplex algorithm (Whittle 1971). However, we wish to examine the entire zone and determine the smallest rectangle that encloses the arrival order zone, thereby giving our epicentral bounds. A quite different method of solution is therefore required. We first

represent the region containing the network of seismic stations by a system of equally spaced grid points. To each point a rank is then assigned according to the number of geometrical constraints satisfied by its position. We then define the arrival order zone as the region containing the points of highest rank. In this way we obtain a 'discretized' version of the required zone. This approach has the advantage of being tolerant of a set of inconsistent constraints brought about by errors in the observed arrival times. In practice, however, due to the size of the average station separation in most local to regional seismic networks, this rarely becomes a problem. We usually obtain an epicentral zone satisfying all or nearly all possible constraints.

A further advantage of this type of approach is that one may perturb the boundaries that define the arrival order zone, to make allowance for the estimated picking errors in arrival times. This may be achieved by sampling grid points on either side of a boundary (perpendicular bisector) and finding all those which satisfied a 'perturbed constraint'. The perturbed or 'fuzzy' constraint is generated by looking at a set of points along the boundary, calculating the *P*-wave travel times from the point to both stations and then selecting those epicentral points whose travel times would lie in a specified range about the travel times for the boundary point. Fig. 2 shows this perturbation of boundaries about the arrival order zone for a four station network.

The method described here will generate a pair of upper and lower latitude and longitude

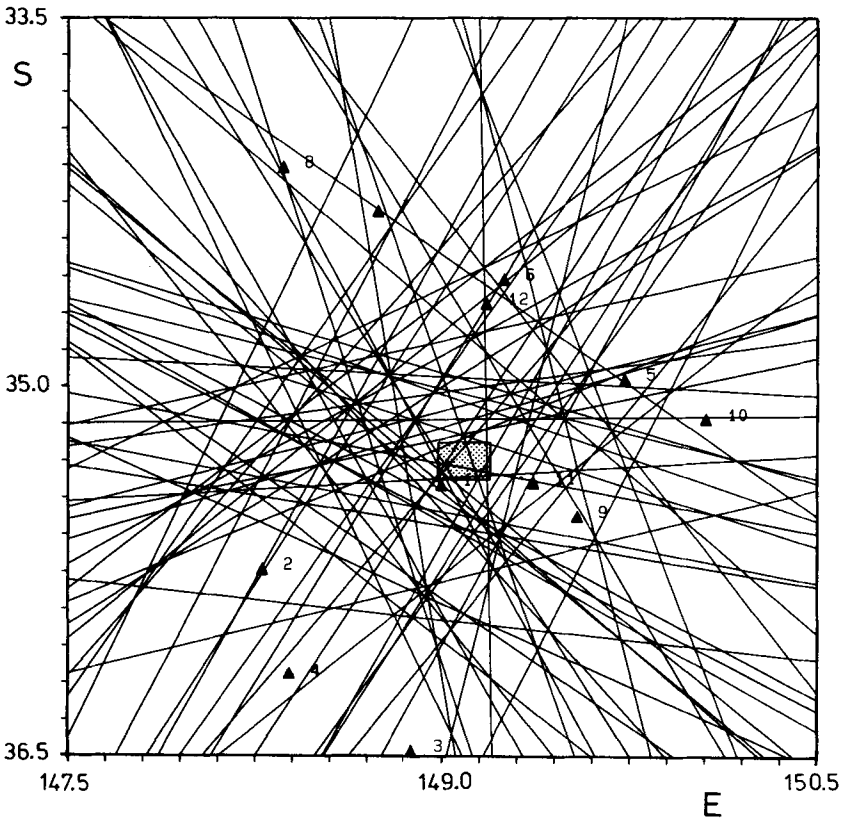

Figure 3. Practical example of the arrival order method for an event near station CAN in Fig. 5, illustrating the 66 constraints imposed by time picks at 12 stations. The rectangular zone needed for further inversion is stippled.

bounds on the epicentre efficiently and quite accurately. The underlying assumption in the method is that the arrival order implies distance order, which is true for most velocity models. The arrival order is therefore insensitive to changes in velocity model and also to picking errors. As a result we get a very robust result. An example with real data is shown in Fig. 3 where a 12 station network is used. The small, shaded, rectangle represents the epicentral bounds determined for the event, while the triangular region within the rectangle is the arrival order zone produced by the combination of the 66 geometrical constraints.

2.2 DEPTH AND ORIGIN TIME BOUNDS

An extension of the above approach can be used to obtain an estimate of depth and subsequently origin time bounds, with the use of both *P*- and *S*-arrival times. Given a velocity model we may calculate the travel times of *P*- and *S*-phases from each of the points previously used to define the arrival order zone to each station in the network. If *S* wave data are available at several stations then the time separation $S-P$ may be calculated and compared with the theoretical predictions. If this is done for a fixed range of depths for each grid point then we may determine the depth which produces a minimum $S-P$ discrepancy between the observed and calculated values. In this way a depth may be associated with each point across the zone and an upper and lower depth bound may be found. In practice, however, *S*-wave data may be of poor quality or unavailable at most stations and in this case *a priori* depth bounds are used.

For the crustal events located using the southeast Australian network *a priori* depth bounds of 0 to 5 km have proven to be useful in generating robust origin time bounds. These may be calculated by a similar procedure to that above. Once again we sample the grid points that span the arrival order zone at both upper and lower depth bounds and calculate the *P*-wave travel times to each station in the network. By subtracting the observations from the theoretical times and averaging over stations we may associate an origin time with each point sample. The variation across the arrival order zone provides the upper and lower origin time bounds. We have found that a very simple velocity model is usually adequate to give effective bounds so the required calculations can be carried out very quickly. A more refined model may be appropriate for the actual hypocentre locations.

2.3 RESULTS OF TESTS

The Dalton/Gunning seismic zone in SE Australia about 60 km north of Canberra is a region of considerable repetitive activity (Cleary 1967). 45 events were selected from this region and hypocentral bounds were generated for each event using the arrival order method we have just described. The velocity model was that proposed by Finlayson & McCracken (1981) derived from a refraction experiment with a line of stations passing through the Dalton/Gunning zone.

These estimates of the hypocentral parameters were compared with the conventional initial estimates obtained from the direct use of $S-P$ times and the standard regional travel time model (Doyle, Everingham & Hogan 1959). In addition a full location was made for each event using the nonlinear algorithm described in the next section.

For the events tested 70 per cent of the final origin times from the nonlinear algorithm described below, were closer to the mean of the upper and lower bounds estimated from the arrival order method than the standard estimate. The mean error in the arrival order estimates was 0.25 s. For the epicentral locations, 10 per cent of the conventional estimates lay outside the corresponding arrival order zones, and all the final locations were within the

epicentral bounds produced by the arrival order method. All depth bounds were set at 0 to 5 km, as S-wave data was not abundant. The conventional initial estimates are based on all available $S-P$ time separations. However, in the arrival order method, to warrant the use of S-wave data we require sufficient information to constrain the depth bounds, typically readings at about three-quarters of the stations in the network.

3 Nonlinear grid search procedure

We start from a set of bounds on hypocentral parameters $(x_j, y_j, z_j, t_j, j = 1, 2)$ which define the initial region of parameter space. These bounds may be determined by the method we have just described (or via some other technique e.g. by assigning generous error bounds to an estimated solution).

We have also to establish some criterion, C; a function of the observed times and estimated hypocentral parameters which we shall minimize to find our optimal solution. The choice of misfit function C depends on the class of statistics we wish to employ to describe the observational errors in the data and the imprecision of our assumed velocity model. We denote the observed travel times by t_{oi}, $i = 1, \ldots, N$ and their theoretical equivalents, calculated for a fixed velocity model and some particular choice of hypocentral parameters by t_{ci}, $i = 1, \ldots, N$. The residual for the i-th datum is then

$$r_i = t_{oi} - t_{ci}. \tag{3.1}$$

The commonly used assumption of a Gaussian distribution for the observational errors leads to a sum of weighted squared residuals for C:

$$C = \sum_{i=1}^{N} (t_{oi} - t_{ci})^2/\sigma_i^2 = \mathbf{r}^T \mathbb{C} \mathbf{r}, \tag{3.2}$$

where $\mathbb{C}$ is the inverse covariance matrix, which with the assumption of no correlation between data picks reduces to a diagonal matrix of inverse variances.

An alternative C function results from the use of the Jeffreys error statistics with a superposition of Gaussians:

$$C = -\sum_{i=1}^{N} \log_e \left\{ \frac{(1-f)}{\sigma_i(2\pi)^{1/2}} \exp\left(\frac{-r_i^2}{2\sigma_i^2}\right) + \frac{f}{\upsilon(2\pi)^{1/2}} \exp\left(\frac{-r_i^2}{2\upsilon^2}\right) \right\}, \tag{3.3}$$

where the small proportion f of a broader based probability distribution with variance υ^2 can account for outlying misfit values. In a regional study this extra term can give a general account of the deficiencies of a single fixed velocity model in a three-dimensionally varying real world.

A specific attempt to introduce the error in the modelling stage into the formulation of the problem was made by Tarantola & Valette (1982) who suggest a mismatch function

$$C = \sum_{i=1}^{N} \sum_{j=1}^{N} [\tilde{t}_{oi} - \tilde{t}_{ci}]^T P_{ij} [\tilde{t}_{oj} - \tilde{t}_{cj}], \tag{3.4}$$

in terms of the matrix

$$\mathbb{P} = [\mathbb{C}_o + \mathbb{C}_c]^{-1},$$

where $\mathbb{C}_o$ and $\mathbb{C}_c$ are the covariance matrices for observed and calculated travel times. The

travel times $\tilde{t}_{oi}$, $\tilde{t}_{ci}$ have a weighted mean removed

$$\tilde{t}_{oi} = t_{oi} - \sum_j p_j t_{oj} \Big/ \sum_j p_j,$$

$$\tilde{t}_{ci} = t_{ci} - \sum_j p_j t_{cj} \Big/ \sum_j p_j,$$

and the vector $\mathbf{p}$ has components

$$p_j = \sum_i P_{ij}.$$

The specific choice of C does not affect the structure of the algorithm and so we are able to compare the results of different choices for the same data set.

3.1 SEARCH ALGORITHM

We partially separate the spatial and temporal parts of the search in our attempt to construct the global minimum of $C(\mathbf{h})$ where $\mathbf{h}$ denotes a 4-D vector in hypocentral parameter space. We will denote the 4-vector at the minimum by $\hat{\mathbf{h}}$. At fixed origin time, T, we will denote the minimum of C over the three spatial parameters by $C(\mathbf{h}_T, T)$, where $\mathbf{h}_T$ is the 3-D spatial location of the minimum.

We may now set up the problem of minimising C as one of finding $C(\mathbf{h}_t, t)$ for any given time t and then minimizing $C(\mathbf{h}_t\ t)$ along the temporal axis (see Fig. 4).

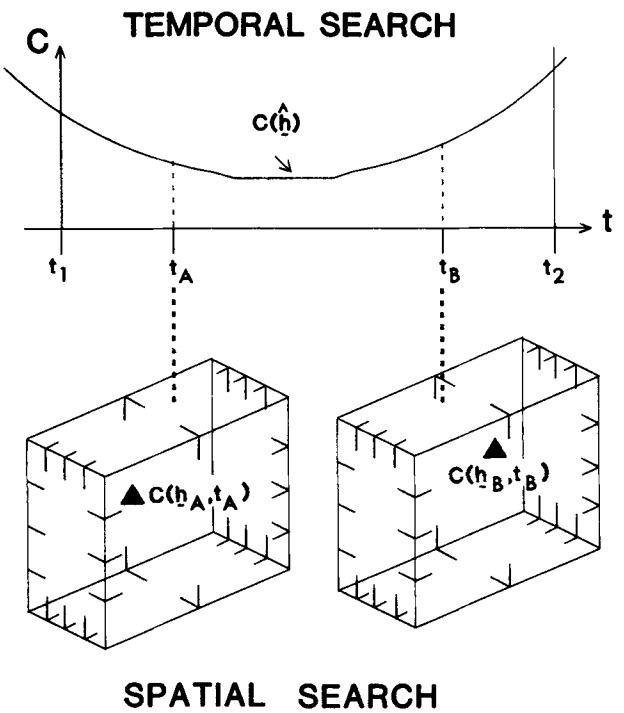

Figure 4. Schematic illustration of the nonlinear procedure, showing the interaction between temporal and spatial searches for a minimum of the misfit function C.

3.1.1 Temporal minimization

We assume, for the moment, that we can find $C(\mathbf{h}_t, t)$, for any t, to any required accuracy, and then we have to perform a 1-D minimization over t. Although a variety of methods are available we would like to avoid the use of derivative information, in keeping with our aim of keeping the entire algorithm free of numerical differentiation. A suitable technique which requires only function evaluations is the Golden Section Search (Whittle 1971). We assume that the position of the global minimum of $C(\mathbf{h})$ lies between the time bounds t_1 and t_2 where C takes the values $C(\mathbf{h}_1, t_1)$, $C(\mathbf{h}_2, t_2)$ respectively.

Consider now two points t_A, t_B symmetrically placed about the midpoint of (t_1, t_2) i.e. $1/2(t_1 + t_2)$ (see Fig. 4), and evaluate the spatial minima $C(\mathbf{h}_A, t_A)$, $C(\mathbf{h}_B, t_B)$. If

$$C(\mathbf{h}_A, t_A) \leqslant C(\mathbf{h}_B, t_B), \tag{3.5}$$

then the minimum value must be confined to the temporal region (t_1, t_B), and we can restrict attention to this zone in which we already have one function evaluation at t_A. Choosing a further point t_c such that

$$t_c - t_1 = t_B - t_A. \tag{3.6a}$$

we have a similar problem to the original one but over a smaller interval in time. When (3.5) does not hold, we restrict attention to (t_A, t_2) and define t_D such that

$$t_2 - t_B = t_D - t_A. \tag{3.6b}$$

The Golden section search gains its name because the subintervals are always divided in the same ratio

$$\frac{t_A - t_1}{t_B - t_1} = \frac{t_B - t_1}{t_2 - t_1} = \rho,$$

but

$$t_B - t_1 = (t_2 - t_1)\,\rho, \tag{3.8}$$

$$t_A - t_1 = (t_2 - t_1)\,(1 - \rho), \tag{3.9}$$

as t_A, t_B are symmetrically disposed about the midpoint of (t_1, t_2). Thus we require

$$(1 - \rho)/\rho = \rho, \tag{3.10}$$

and solving for the Golden section number ρ we obtain

$$\rho = 0.618\ldots. \tag{3.11}$$

We define t_B by (3.8) and t_A by (3.9) which we may also write as

$$t_2 - t_A = \rho(t_2 - t_1). \tag{3.12}$$

When (3.5) holds we apply the procedure to the new interval (t_1, t_B) to define a new value for t_A by

$$t_B - t_A = \rho(t_B - t_1). \tag{3.13}$$

If, however, (3.5) is untrue we examine the interval (t_A, t_2) and construct a new t_B by

$$t_B - t_A = \rho(t_2 - t_A). \tag{3.14}$$

In either case, we have arrived back at the original situation of having a minimum confined to a time interval with an interior function value at the interval ratio $1:\rho$. We can therefore

repeat the process by considering the new inequality of the type (3.5). At each stage the size of the interval which needs to be considered is reduced by a factor of ρ. After k evaluations of $C(\mathbf{h}_t, t)$ we will be able to locate the minimum in an interval reduced by ρ^{k-1} of the original.

The assumption that the minimum of C lies between $C(\mathbf{h}_1, t_1)$ and $C(\mathbf{h}_2, t_2)$ along the origin time axis is not essential to the algorithm. If the global minimum lies outside this time interval the Golden Section search will converge on the closest of the two boundaries to the minimum. One may actually monitor the boundary points t_1 and t_2 during the progress of the search to detect this situation. Alternatively, a simple procedure adopted here is to determine if

$$C(\mathbf{h}_1, t_1) < C(\mathbf{h}_A, t_A) < C(\mathbf{h}_B, t_B) < C(\mathbf{h}_2, t_2), \tag{3.15}$$

or

$$C(\mathbf{h}_2, t_2) < C(\mathbf{h}_B, t_B) < C(\mathbf{h}_A, t_A) < C(\mathbf{h}_1, t_1), \tag{3.16}$$

holds for the initial time interval. If (3.15) holds then this is an indicator that the global minimum $C(\hat{\mathbf{h}})$ may lie below $C(\mathbf{h}_1, t_1)$ on the time axis. Although this condition is not sufficient to determine whether the initial interval of interest requires expansion, it seems prudent to decrease t_1 (in this case) until (3.15) no longer holds i.e. $C(\mathbf{h}_1, t_1) \geqslant C(\mathbf{h}_A, t_A)$. The upper limit may be treated in a similar manner using (3.16).

This then completes the temporal part of the minimization process. We discuss the convergence criteria we have employed in section 3.1.3. At each stage of the temporal minimization we have assumed that for a given t the minimum of C over the three spatial coordinates, i.e. $C(\mathbf{h}_t, t)$, is available to us. It is here that the spatial search part of the algorithm is put to use.

3.1.2 Spatial search

The optimization of $C(\mathbf{h}_t, t)$ over the spatial coordinates is performed by searching on a spatial lattice with a fixed interval in all three dimensions (e.g. for regional studies we have taken the spacing to be 1 km). The major problem faced when attempting this type of procedure is in limiting the amount of redundant work. We need to be able to guide the algorithm in searching over regions of parameter space. If possible, we would prefer to always search in regions close to our desired minimum, and we therefore need to place as many constraints as possible on the choice of search regions.

The procedure we have adopted is aimed at reducing the overall time taken in searching over the spatial grid. Initially we define a fixed grid in all three dimensions for the region of parameter space given by our bounds on the spatial coordinates ($x_j, y_j, z_j, j = 1, 2$). This gives us a 3-D rectangular grid system over the arrival order zone. The centre of each unit cube in the lattice is taken to represent the region of parameter space occupied by the cube, and the C statistic is evaluated at each central point. This search over the entire arrival order zone is performed for an origin time t_M, which is the mean of the initial origin time bounds, and requires a considerable proportion of the overall computation time. However, it provides an initial point on the $C(\mathbf{h}_t, t)$ curve (Fig. 4), which acts as a 'handrail' for subsequent spatial searches. Once this initial large scale search is completed then all further searches are carried out over smaller portions of the 3-D grid.

A box is set up with its centre at the location of the minimum at the time t_M i.e. $\mathbf{h}_M$, with sides of a suitable size for the class of location problem at hand. We have found 7 x 7 km in latitude and longitude by 5 km in depth to be adequate for the southeast Australian

network. The next spatial search occurs for $t = t_1$ (see Fig. 4) and is carried out over this smaller region to obtain $C(\mathbf{h}_1, t_1)$. The location of this minimum provides the central point for the next box search at $t = t_2$ and so on. Currently we have used a fixed size of box; ideally we would like the box dimensions to be always large enough to just contain the next minimum $C(\mathbf{h}_t, t)$, where t is given by the temporal search procedure. However, as the algorithm progresses the spatial distance between these minima decrease as the change in origin time parameter decreases with subsequent searches. The size of the box is therefore aimed at being efficient in the initial stages of the algorithm when the movement of the minimum is largest.

If at any stage we find that $C(\mathbf{h}_t, t)$ is on the boundary of the box suggesting that the true minimum lies outside, we use a simple tracking procedure. The box is moved in the direction of the minimum and its dimensions are reduced in the other directions (e.g. to 3 km). For example, if we sample over an epicentral region at a shallow depth and the spatial minimum is located in the middle of the epicentral zone but at the deepest level sampled, then we may expect that the movement of $C(\mathbf{h}_t, t)$ will be greater in depth than in the horizontal coordinates. We would therefore restrict the sampling in the horizontal directions during the next search. The interference of possible local minima of C can to some extent be avoided by increasing the box dimensions from the beginning but at the cost of increased computation.

Overall, we have a progressive narrowing of the search region as the algorithm progresses. We have to strike a balance between the effort needed to perform the search over a box, which is least for a small box, and the number of times tracking is invoked, which increases the total computational effort and may occur frequently for too small a box.

3.1.3 Convergence criteria and the final location

One may define a convergence criterion for the algorithm by one or more means. We could choose a threshold on the misfit statistic C below which discrimination is not possible because of the likely errors in the data. However, such a condition would be dependent on the choice of misfit function C, a property which we would prefer to avoid.

We have chosen therefore to define the convergence of the algorithm by imposing bounds on the spatial region within which the solution must lie. This is done by testing the difference in the 4-vectors at the minima of $C(\mathbf{h})$ at the current origin time bounds

$$\mathbf{b} = \mathbf{h}_2 - \mathbf{h}_1.$$

The algorithm is terminated when the $\mathbf{b}$ vector has components less than specified length and time measures (e.g. we have taken 1 km and 0.1 s for our 1 km grid). $\mathbf{b}$ is an indication of the size of our best region of parameter space as measured by our 3-D grid system, and is therefore constrained by the lattice spacing.

To obtain a single final location we resort to a quadratic fit of the C function over the region spanned by $\mathbf{b}$ and take the minimum found. Since this fit is performed over a small region of parameter space the effects of nonlinearity on the final location estimate are negligible.

3.2 CONFIDENCE LIMITS

The estimation of a point solution to the location problem is not very informative unless some reliable estimate of its accuracy can be given. The relative errors in the solution can be determined by examining the variation of the C function in the region of parameter space

about the proposed location. The approach we have adopted is particularly convenient since the values of the misfit function C are available on a regular grid about the calculated location. We may therefore readily examine the shape of the contours of C in this region. Figs 7 and 8 show such contours expressed as confidence levels on the probability of the location, when one assumes Gaussian and Jeffreys statistics for the picking errors.

Buland (1976) describes a method by which we may obtain an estimate of the 95 per cent confidence limits for such contours. We introduce

$$\Gamma(\mathbf{h}) = C(\mathbf{h}) - C(\hat{\mathbf{h}})$$

where $C(\hat{\mathbf{h}})$ is the value of C at our solution point. Then we say that the surfaces of constant $\Gamma(\mathbf{h})$ represent joint confidence levels for the estimation of hypocentral parameters as suggested by the data statistic. Buland (1976) notes that one may assign confidence levels to the raw C values when a least-squares misfit function (3.2) is assumed, by considering a chi-squared distribution with four degrees of freedom. If we assume that the dependence of C on the hypocentral parameters is approximately linear in a region close to the solution, then the chi-squared distribution may be used to determine values of $\Gamma(\mathbf{h})$ corresponding to a 95 per cent confidence limit. However, this scheme is only valid if the picking errors are truely Gaussian, and we have modelled them correctly by assuming accurate σ_i in formulating the C function.

The effect of the errors in the velocity model on the solution is unclear in most circumstances. The direct use of the chi-squared estimation of confidence regions does not take into account any errors from this cause. In an attempt to do so, Tarantola & Valette (1982) treat the whole problem from a probabilitistic point of view and effectively increase the relative weights σ_i on the observational errors to account for errors in the velocity model. The result is to broaden the C function contours and then these may be modelled as a chi-squared distribution with $(n-4)$ degrees of freedom. In this way the confidence limits may be more representative of the combined velocity model and picking time errors.

The approach we have adopted here is to reduce the C statistic by a factor such that the value of $C(\hat{\mathbf{h}})$ is rescaled to the expectation value of a chi-squared distribution with $(n-4)$ degrees of freedom which is just $(n-4)$, where n is the number of data values. After the rescaling we take $\Gamma(\mathbf{h})$ as a chi-squared function with four degrees of freedom, i.e. we calculate the region of parameter space which satisfies

$$\Gamma(\mathbf{h})/K \leq \chi_4^2(0.95),$$

where

$$K = C(\hat{\mathbf{h}})/(n-4),$$

or

$$\frac{[C(\mathbf{h}) - C(\hat{\mathbf{h}})]}{C(\hat{\mathbf{h}})} \leq \frac{\chi_4^2(0.95)}{(n-4)}. \tag{3.17}$$

K is the factor by which we reduce the function so that its value at our solution $\hat{\mathbf{h}}$ is $(n-4)$. By dividing by K we are in fact trying to eliminate the effect of velocity model errors on the spread of the C function about $\hat{\mathbf{h}}$. A drawback of this approach is that all the confidence regions are in fact based on the assumption of Gaussian statistics for picking errors; however the resulting regions seem quite reasonable.

Koch (1985) has produced a similar method for estimating confidence regions with an allowance for inadequacies in the velocity model. He has defined his 95 per cent confidence

regions by the region such that the 4-vector **h** satisfies

$$\frac{[C(\mathbf{h}) - C(\hat{\mathbf{h}})]}{C(\hat{\mathbf{h}})} \leqslant \frac{4\, F^{n}_{n-4}(0.95)}{(n-4)}, \tag{3.18}$$

where F^{n}_{n-4} is the Fisher distribution with n and $(n-4)$ degrees of freedom. The Fisher distribution is usually used to distinguish between hypotheses and has smaller values than appearing in (3.17), our estimates of confidence regions are therefore more conservative (i.e. larger) than those used by Koch.

All confidence regions are dependent on the statistics used to model picking errors and also therefore on the optimization function C. We may follow an analogous route to that described above to generate confidence regions with the Jeffreys statistic if the fraction of the broader distribution, f, is small. In this case the analogue of chi-squared is

$$X_J = -2 \sum_{i=1}^{n} \log_e \left\{ \frac{(1-f)}{\sigma_i (2\pi)^{1/2}} \exp\left(\frac{-r_i^2}{2\sigma_i^2}\right) + \frac{f}{v(2\pi)^{1/2}} \exp\left(\frac{-r_i^2}{2v^2}\right) \right\}$$

$$+ 2 \sum_{i=1}^{n} \log_e \{\sigma_i (2\pi)^{1/2}\}.$$

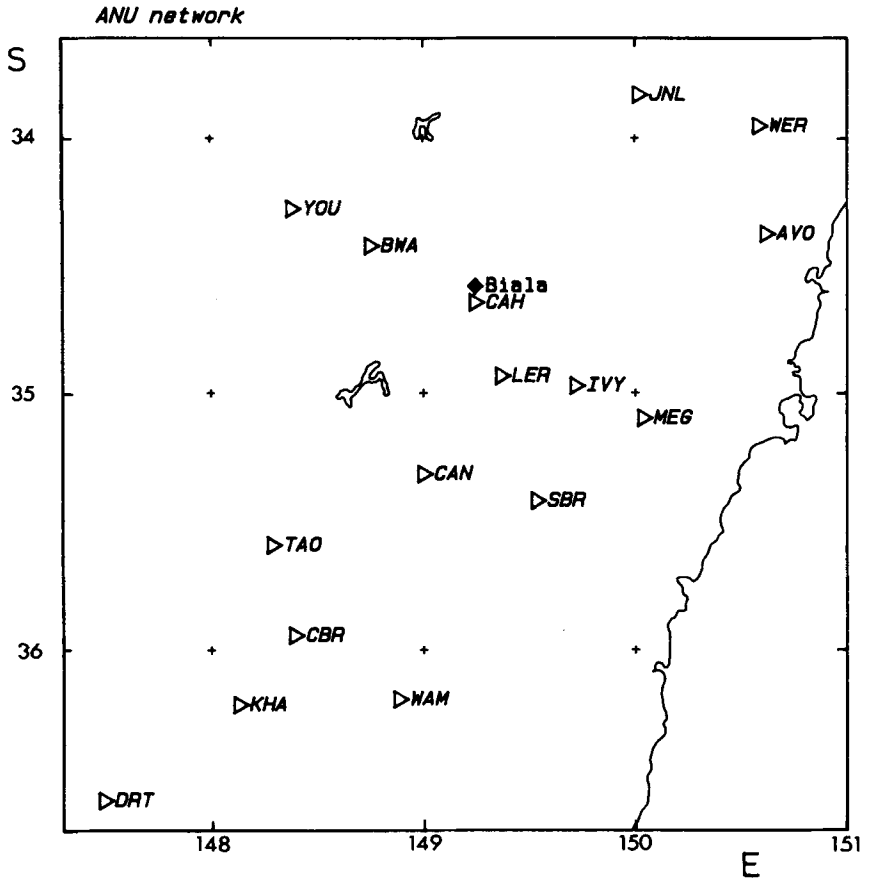

Figure 5. The SE Australian seismic network administered by the Australian National University, with the location of the earthquake at Biala illustrated in Figs 7 and 8.

The expectation value of X_J on $(n-4)$ degrees of freedom is approximately

$$(n-4)\ \{(1-f)+v^2 f/\sigma^2\}$$

for small f, and equal σ_i^2, and so is only slightly shifted from the former value. The 95 per cent confidence level for 4 degrees of freedom can be estimated for small f as

$$\chi_4^2\ (0.95)\ \{(1-f)+fg(v/\sigma)\}$$

where $g(v/\sigma)$ is typically of the order of 2–4. Thus only a slight modification of the right side of (3.17) is required.

It should be noted that any estimate of such confidence regions is heavily dependent on the velocity model. A poor choice of velocity model may give a false impression of precision.

4 Results of nonlinear inversion

We will illustrate our nonlinear algorithm for the determination of hypocentres by examining the location of events in southeastern Australia using the seismic network shown in Fig. 5. The stations follow the southeastern highlands and mostly lie in the palaeozoic Lachlan fold belt. The three northernmost stations operated by the Sydney Water Board lie in the younger sediments of the Sydney Basin.

The network was started in 1958 by the Australian National University in association with the major engineering works of the Snowy Mountain project and has subsequently been expanded to its present status of 16 stations. The seismicity of the area is in general quite widespread and diffuse with no clear correlations with the major faults of the fold belt (Lambeck *et al.* 1985). There are concentrations of events near the stations YOU and WAM, and an exceptionally active zone in the Dalton/Gunning region about 60 km north of

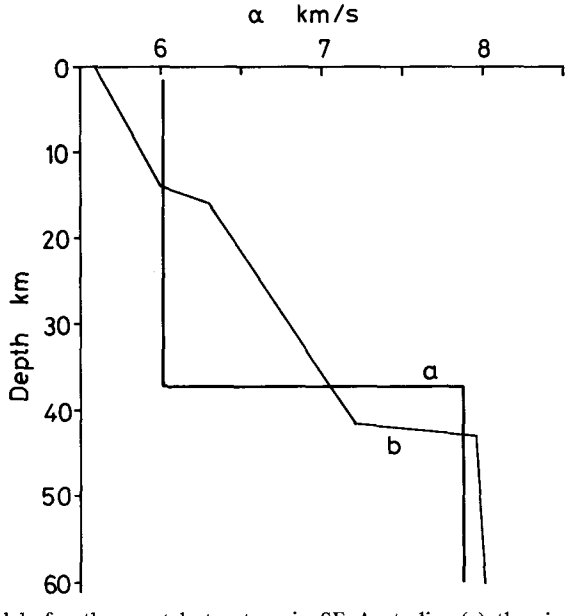

Figure 6. Velocity models for the crustal structure in SE Australia: (a) the simple regional model of Doyle, Everingham & Hogan (1959), (b) averaged velocity model for the Dalton/Gunning zone derived from the work of Finlayson & McCracken (1981).

Canberra near station CAH. 400 events were recorded from this latter zone during 1984. We have chosen an event at Biala, with magnitude 2, from this region for detailed study. The geometry of most of the network has been dictated by the needs of line-of-sight radio telemetry and hence is not ideal for events in this neighbourhood.

As the network covers a considerable linear extent from the Sydney Basin up into the Snowy Mountains and beyond, it is difficult to construct a single, laterally homogeneous, model for the velocities in the entire region. However, a simple model which gives reasonable results is that proposed by Doyle, Everingham & Hogan (1959) from refraction experiments in the area (see Fig. 6a). More detailed refraction work has been carried out by Finlayson & McCracken (1981) using quarry blasts and profiles crossing much of the region of interest. We have used an averaged model for profiles passing near the Dalton/Gunning zone to try to improve the location of events in that region, the resulting gradient model is shown in Fig. 6b.

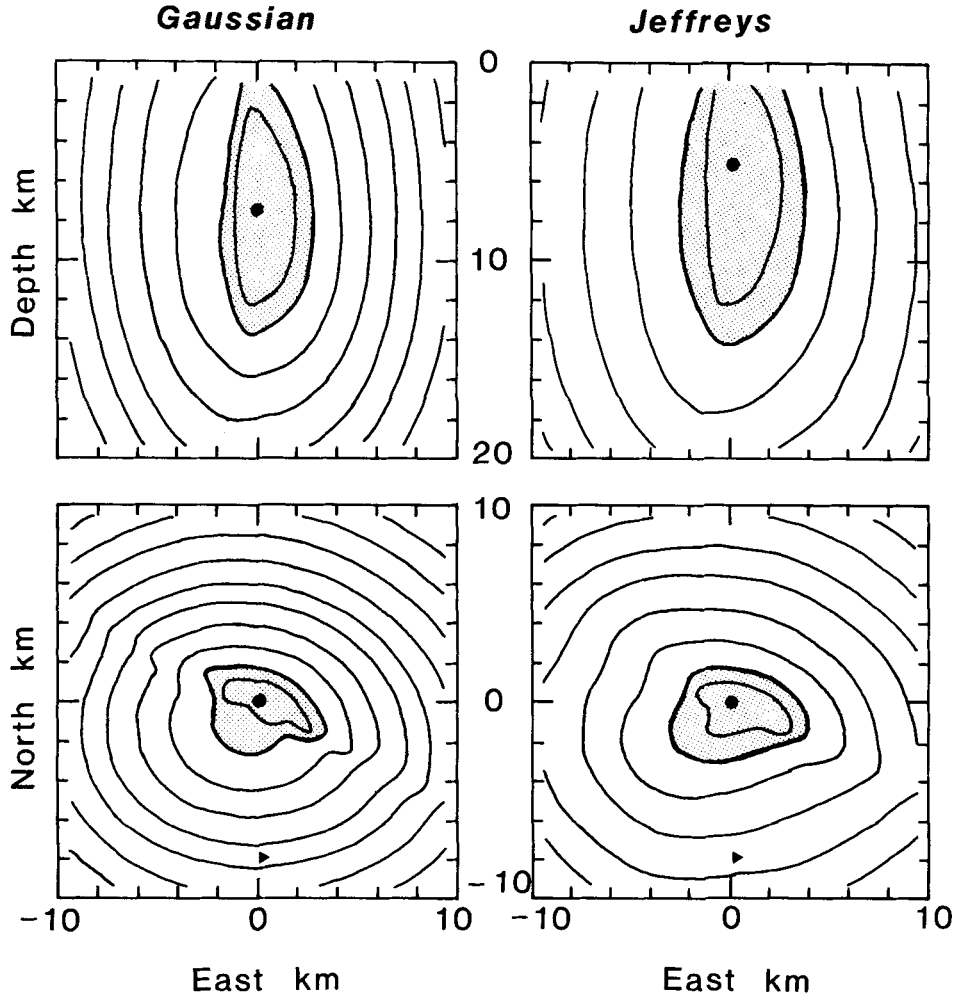

Figure 7. Comparison of nonlinear hypocentre locations using Gaussian and Jeffreys statistics for the Biala event. Contours of the misfit function are shown in vertical and horizontal sections for each case using the regional model (Fig. 6a). The 95 per cent confidence regions are emphasized by shading.

The arrival times for the *P*- and *S*-phases for the events were taken directly from the network data file. The typical picking errors are relatively uniform across all the stations and so we took standard deviations of 0.1 s for *P*-wave readings, 0.2 s for *S*-wave readings. These can be used directly in the construction of the mismatch function (3.2) based on Gaussian statistics. When we employ the Jeffreys form (3.3), we need to introduce the size and 'width' of the underlying broader distribution, which we take to be representative of our modelling errors. A wide variety of trials with Gaussian forms showed that it was very difficult to reduce the minimum misfit below a level corresponding to errors of the order of two to three times our estimate of the reliability of our picks. We have therefore taken the background distribution to have a standard deviation of 0.3 s, with a weighting factor of 0.005 chosen so that the dominant role in the inversion is played by the data constraints.

We now consider a comparison of the use of the squared-residual (3.2) and Jeffreys (3.3) mismatch functions for the Biala event shown on Fig. 5. We could equally well include the Tarantola & Valette mismatch function (3.4) but results using this function are similar to the squared residual case. This arises because in the formulation of (3.4) both the 'modelling' and observational errors are assumed to have Gaussian probability densities.

In Fig. 7, we use the regional model of Doyle *et al.* (1959), shown in Fig. 6a, with the non-linear inversion algorithm and draw contours of the mismatch functions in horizontal and vertical sections through the proposed hypocentres. The vertical sections are taken along east–west lines through the hypocentres, extending 20 km both horizontally and vertically. The horizontal sections are taken at the depth of the proposed locations on a fixed grid 20 km in both northerly and easterly extent. The use of the fixed grid enables any shift in the epicentres to be more readily judged. The contours are drawn at confidence levels of 75, 95 and 99.9 per cent and then at equal intervals in the mismatch function. The 95 per cent confidence region is emphasized by a bolder outline and shading.

For this Biala event, the contours of the mismatch function in the squared-residual (Gaussian) case show significant departures from ellipses, even at the 75 per cent confidence level. This indicates substantial nonlinearity in the location problem arising from the network geometry. We have one relatively close station (CAH) whose spatial location is marked by a triangle in Fig. 7, and a number of other stations on a line through the epicentral region. Algorithms of the least-squares type (such as (1.3), (1.4)) assume that in the neighbourhood of the hypocentre the contours of the mismatch function are ellipsoids (see e.g. Kennett 1976). The nonlinear character shown in Fig. 7, shows that for the Biala event such error estimates using local linearization would be, at best, misleading.

From micro-earthquake surveys in the Biala region we know that the activity mostly lies shallower than 4 km. However, the hypocentre suggested by the use of the squared residual fit (3.2) is rather deep at 8.5 km. This discrepancy is probably mostly due to the use of the over simplified velocity model (c.f. Fig. 8). With the use of the Jeffreys function (3.3) the proposed hypocentre is shallower (4.9 km) because less weight is attached to the poorest data fits arising from the inadequacies of the velocity model. We note that the pattern of the misfit contours is rather different in the two cases, for example the 95 per cent confidence region for this simple model intersects the surface for the Jeffreys function but is almost closed in the Gaussian case.

In order to show the very considerable influence of the choice of velocity model on the character of the solution, we show in Fig. 8 a similar set of misfit diagrams but now using the gradient model in Fig. 6(b). Once again we compare the use of the two choices (3.2) and (3.3) for the misfit criterion. With the improved velocity model we observe that the solutions have moved in space, especially in depth. The proposed locations in Fig. 8 are concordant with the micro-earthquake results. The contours for the squared-residual

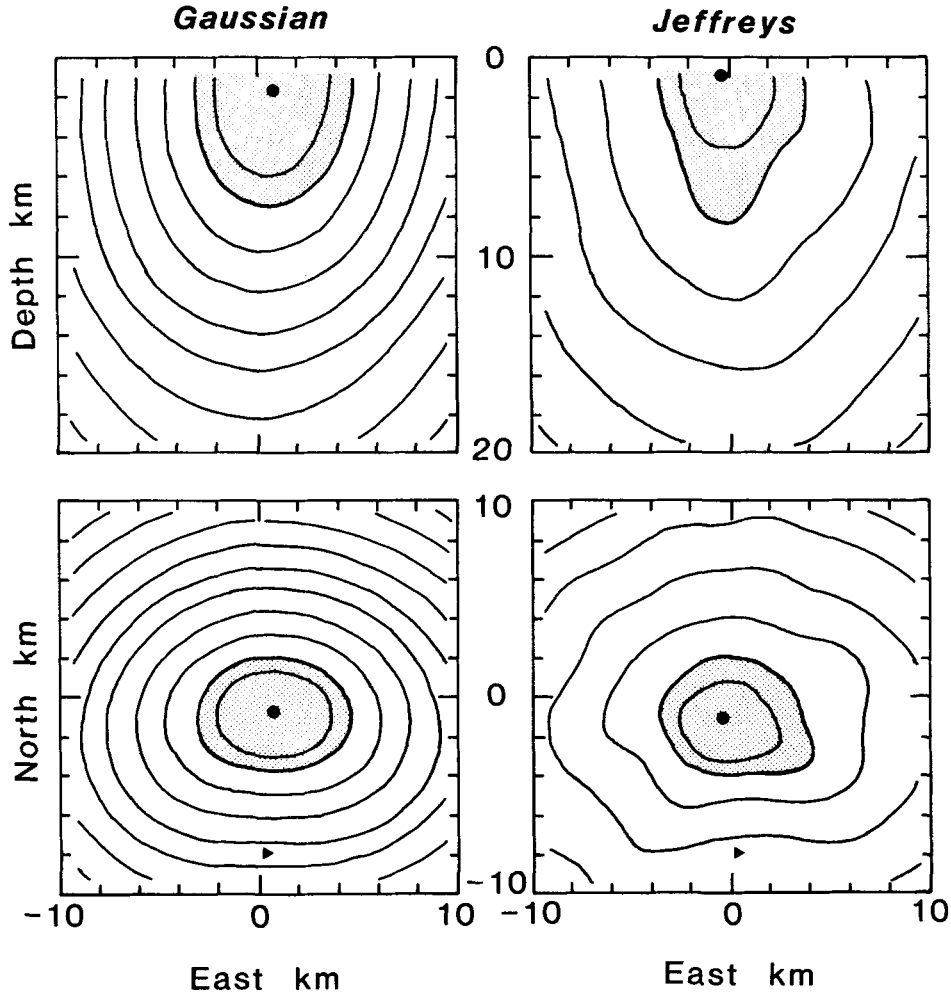

Figure 8. Nonlinear hypocentre locations and misfit contours for the Biala event using the velocity gradient model shown in Fig. 6(b). Misfit functions based on both Gaussian and Jeffreys statistics are shown. The 95 per cent confidence regions are emphasized by shading.

function (3.2) are somewhat smoother than in Fig. 7, and show significant flattening in the neighbourhood of the station CAH whose projection is indicated by the triangle. The character of the misfit contours for the Jeffreys function (3.3) is more complex but also shows flattening near CAH. The 95 per cent confidence regions estimated for this better velocity model are in much closer agreement for the two classes of statistics, even though the estimates of the optimum epicentres differ by about 1 km. Similar behaviour is displayed for many other events in the Dalton/Gunning seismic zone.

The inversion algorithm has been successfully employed for events over the whole region covered by the network. The estimates of the 95 per cent confidence regions show a strong dependence on velocity model (e.g. a much shallower depth with the gradient model) and a weaker dependence on the statistics employed. Even though the behaviour of the misfit contours implies very nonlinear dependence of the C function on the hypocentral parameters, the non-linear algorithm we have developed performs well and will converge even if

the initial hypocentral bounds are misplaced. The algorithm is, however, at its most efficient when C increases monotonically away from a unique global minimum.

The tolerance of the Jeffreys function (3.3) to outlying residuals is particularly beneficial where the velocity distribution is less well determined. Since little more computation is needed with our non-linear algorithm, the Jeffreys function is to be preferred for general use.

5 Discussion

The non-linear approach adopted here avoids some of the major problems associated with iterative linear schemes for hypocentre locations. The generation of corrections to a best guess hypocentral parameter is avoided, and our algorithm deals with regions of parameter space which are more informative than point estimates. We also avoid numerical instabilities associated with matrix inversion. Due to the rather simple requirements of the algorithm i.e. evaluations of the C function alone, our location scheme is very versatile and rather general in character. We may substitute one misfit function for another without changing the form of the algorithm. The only requirement is that we have the ability to generate values of C for any given set of hypocentral parameters. In exactly the same way we may substitute any type of velocity model for which the travel times can be calculated. Layered velocity models, gradient zones or three-dimensionally laterally heterogeneous models may be used equally well. This independence of the form of the velocity model is a result of bypassing the use of analytical expressions for the derivatives of travel times with respect to the hypocentral coordinates. However, we do require an efficient forward modelling program for the travel times to avoid excessive computation times in complex models.

The results of experiments with real data from a regional network indicate that we have been very successful in our main objective of finding a hypocentral location by the minimization of a misfit function, to some specified accuracy. An indication of the reliability of that solution is obtained directly by the examination of the contours of the misfit function about the proposed solution. The ability to relocate events with different C functions, and to examine the constraints placed on the location, means that we have a very powerful tool for investigating the distribution of earthquakes.

References

Anderson, K. R., 1981. Epicentral location using arrival time order, *Bull. seism. Soc. Am.*, **30**, 119–130.

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

Cleary, J. R., 1967. The seismicity of the Gunning and surrounding areas 1958–1961, *J. geol. Soc. Austr.*, **14**, 23–29.

Doyle, H. A., Everingham, I. B. & Hogan, T. K., 1959. Seismic recordings of large explosions in south-eastern Australia, *Aust. J. Phys.*, **12**, 222–230.

Finlayson, D. M. & McCracken, H. M., 1981. Crustal structure under the Sydney Basin and Lachlan Fold Belt, determined from explosion seismic studies, *J. geol. Soc. Austr.*, **28**, 177–190.

Jeffreys, H., 1932. An alternative to the rejection of observations, *Proc. R. Soc.*, **137A**, 78–87.

Kennett, B. L. N., 1976. Some aspects of nonlinearity in inversion, *Geophys. J. R. astr. Soc.*, **55**, 373–391.

Koch, M. 1985. A numerical study on the determination of the 3-D structure of the lithosphere by linear and nonlinear inversion of teleseismic travel times, *Geophys. J. R. astr. Soc.*, **80**, 73–93.

Lambeck, K., Stephenson, R., McQueen, H. & Denham, D., 1985. The state of stress in the Australian continent, *Ann. Geophysicae*, **2**, 723–742.

Lee, W. H. K. & Stewart, S. W., 1981. Principles and Applications of Microearthquake Networks, *Advances in Geophysics, supplement 2*, Academic Press, London.

Tarantola, A. & Valette, B., 1982. Inverse problems: quest for information, *J. Geophys.*, **50**, 159–170.

Whittle, P., 1971. *Optimisation under constraints*, Wiley, London.