Earthquake location

J. Havskov, Institute of Solid Earth Physics, University of Bergen, Norway

Introduction

Earthquake location is one of most important tasks in practical seismology and most seismologists have been involved in this tasks from time to time. The intention with this document is describe the most common methods without going into the mathematical details, which has been described in numerous text books, and to give some practical advice on earthquake location.

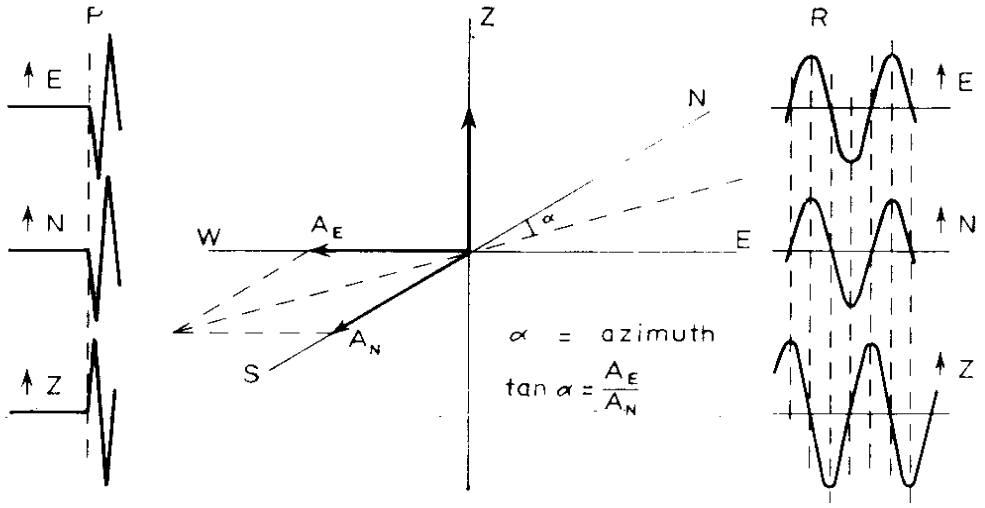
The earthquake location is defined by the earthquake hypocenter (x_0,y_0,z_0) and the origin time t_0. The hypocenter is the physical location, usually longitude(x_0), latitude(y_0) and depth below the surface (z_0,km). The epicenter is (x_0,y_0). For simplicity, the hypocenter will be labeled x_0, y_0, z_0 with the understanding that it can be either geographical or Cartesian coordinates. The origin time is the time of occurrence of the earthquake. When the earthquake is large, the physical dimension can be several hundred kilometers and the hypocenter can in principle be located anywhere on the rupture surface. Since the hypocenter and origin time are determined by arrival times of seismic phases initiated by the first rupture, the computed location will correspond to the point where the rupture initiated and the origin time to the time of the initial rupture. This is also true using any P or S-phases since the rupture velocity is smaller than the S-wave velocity so that P or S-wave energy emitted from the end of a long rupture will always arrive later than energy radiated from the beginning of the rupture. Standard earthquake catalogs (such as from the International Seismological Center, ISC) report location based primarily on arrival times of high frequency P-waves. This location can be quite different from the centroid time and location obtained by moment tensor inversion of long period waves. The centroid location represent the average time and location for the entire event.

Single station location

In general, epicenters are determined using many arrival times from different seismic stations and phases. However, it is also possible to locate an earthquake using a single 3-component station. Since the P-waves are vertically and radially polarized, the vector of P-wave motion can be used to calculate azimuth to the epicenter, see Figure 1. The radial component of P will be recorded on the 2 horizontal seismometers and the ratio of the amplitudes on the horizontal components can be used to calculate the azimuth of arrival. There is then an ambiguity of 180° since the first polarity can be up or down so the polarity must also be used in order to get the correct azimuth. If the first motion (on vertical component) of the P is upward, the radial component of P is directed away from the hypocenter, and opposite if P-polarity is negative. The amplitude of the Z-component can, together with the amplitude on the horizontal components, also be used to calculate the angle on incidence.

Figure 1

Left: Method of determining the azimuth to the source using the amplitudes of the P-wave as recorded on the 3 components. Right: The method can also be used to get the arrival azimuth of surface waves. The Raleigh waves corresponding to the P-waves are shown. Note the 90° phase shift between Z and the horizontal components.. (From Båth, 1979).



With modern high frequency data it might be difficult to manually read the amplitudes of the first break or sometimes the first P-swings are emergent. Since the amplitude ratio between the components should remain constant not only for the first swing of the P-phase but also for the following oscillations of the same phase, we can, with digital data, use the predicted coherence method (Roberts et al, 1989) to automatically calculate azimuth as well as the angle of incidence. Since this is much more reliable and faster than using the manually readings of the first amplitudes, calculation of azimuth has again become a routine practice. As we shall see later, azimuth observations are useful in restricting epicenter locations.

With a single station we have now the direction to the seismic source. The distance can be obtained from the difference in arrival time of two phases, usually P and S. If we assume a constant velocity, and origin time t_0, the P and S-arrival times can then be written as

$$t_p = t_0 + \Delta/v_p \qquad t_s = t_0 + \Delta/v_p \qquad (1)$$

where t_p and t_s are the P and S-arrival times respectively, v_p and v_s are the P and S-velocities respectively and Δ is the distance. By eliminating t_0 from (1), the distance can be calculated as

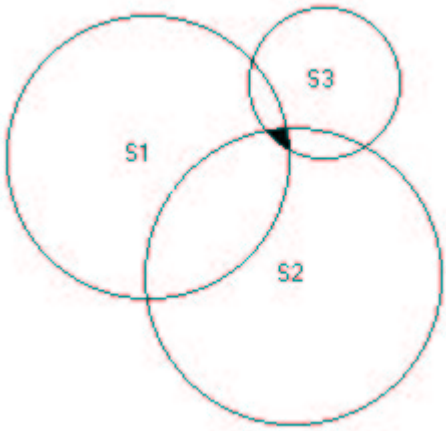
$$\Delta = (t_s - t_p)\frac{v_p * v_s}{v_p - v_s} \qquad (2)$$

For a Poisson solid, we can assume that $v_s = v_p/\sqrt{3}$. As a rule of thumb $\Delta = (t_s - t_p) * 9.0$ for many local models. This corresponds to an average P-velocity of 6.6 km/sec. For larger distances, travel time tables can be used to calculate the distance. With both azimuth and distance, the epicenter can be obtained by measuring off the distance along the azimuth of approach. Finally, knowing the distance, we can calculate the P-travel time and thereby get the origin time using the P-arrival time.

Multiple station location

When at least 3 stations are available, a simple manual location can be made from drawing circles (the circle method) with the center at the station locations and the radius equal to the epicentral distance calculated from the S-P-times (Figure 2).

Figure 2
Location by the circle method. The stations are located in S1, S2 and S3. The epicenter is found within the black area where the circles cross.



These circles will rarely cross in one point which indicates errors in the observations and/or that we have wrongly assumed a surface focus. Methods exist (e.g. Båth, 1979) to deal with this depth problem, however since it is rarely used, it will not be discussed here.

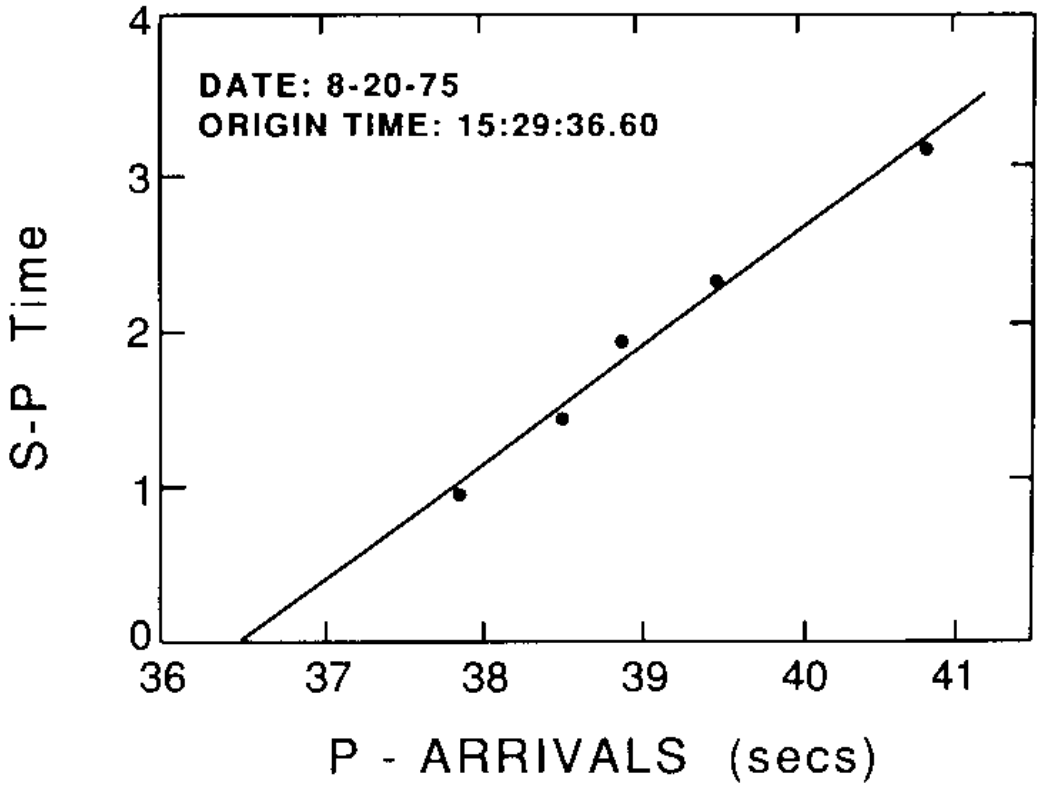
With several stations available from a local earthquake, the origin time can be determined by a very simple graphical technique called a Wadati diagram (Figure 3). Using equation (1) and eliminating Δ, the S - P travel time can be calculated as

$$t_s - t_p = (v_p/v_s - 1) * (t_p - t_0) \qquad (3)$$

The S-P times are plotted against the absolute P-time. Since $t_s - t_p$ goes to zero at the hypocenter, a straight line fit on the Wadati diagram gives the origin time at the intercept with the P-arrival axis and from the slope of the curve, we get v_p/v_s. Note that is thus possible to get a determination of both the origin time and v_p/v_s without any prior knowledge of the crustal structure, the only assumption is the v_p/v_s is constant and that the P and S-phases are of the same type like Pg and Sg or Pn and Sn. Independent determination of these parameters can be very useful when using other methods of earthquake location.

Figure 3

An example of a Wadati diagram. The intercept on the x-axis gives the origin time. The slope of the line is 0.72 so the v_p/v_s ratio is 1.72. Note that the points do not exactly fit the line indicating model or observation errors. (From Lay and Wallace, 1995).



The Wadati diagram can be very useful in making independent checks of the observed arrival times. Any points not fitting the linear relationship might be badly identified, either by not being of the same type or misread.

Manual location methods provide insight into the location problems, however in practice we use computer methods. In the following, the most common ways of calculating hypocenter and origin time by computer will be discussed.

The calculated arrival time t_i^c at station i can be written as

$$t_i^c = T(x_i y_i, z_i, x_0, y_0, z_0) + t_0 \qquad (4)$$

where T is the travel time as a function of the location of the station location (x_i,y_i,z_i) and the hypocenter location. This equation has 4 unknowns, so in principle 4 arrival time observations from at least 3 stations are needed in order to determine the hypocenter and origin time. If we have n observations, there will be n equations of the above type and the system is

overdetermined and has to be solved in such a way that the misfit or residual r_i at each station is minimized. r_i is defined as the difference between the observed and calculated travel times

$$r_i = t_i^o - tc_i \qquad (5)$$

Thus the problem seems quite simple in principle. However, since the travel time function T is a nonlinear function of the model parameters, it is not possible to solve (4) with any analytical methods. So even though T can be quite simple to calculate, particularly when using a 1D earth model or travel time tables, the non linearity of T greatly complicates the task of inverting for the best hypocentral parameters. The non linearity is evident even in a simple 2D location where the travel time t_i from the point (x,y) to a station (x_i,y_i) can be calculated as

$$t_i = \frac{\sqrt{(x-x_i)^2 + (y-y_i)^2}}{v} \qquad (6)$$

where v is the velocity. It is obvious that t_i does not scale linearly with either x and y so it is not possible to use any set of linear equations to solve the problem and standard linear methods cannot be used. This means that given a set of arrival times, there is no simple way of finding the best solution. In the following, some of the methods of solving this problem will be discussed.

Grid search

Since it is so simple to calculate the travel times to any point in the model, given a lot of computer power, a very simple method is to perform a grid search over all possible locations and origin times and compute the arrival time at each station. The hypocentral location and origin time would then be the point with the best agreement between the observed and calculated times. This means that some measure of best agreement is needed, particularly if many observations are used. The most common approach is least squares which is to find the minimum of the sum of the squared residuals e from the n observations:

$$e = \sum_{i=1}^{n} (r_i)^2 \qquad (7)$$

The root mean squared residual *RMS*, is defined as $\sqrt{e/n}$. *RMS* is given in almost all location programs and commonly used a guide to location accuracy. If the residuals are of similar size, the *RMS* gives the approximate average residual. As will be seen later, *RMS* only gives an indication of the fit of the data, and a low *RMS* does not automatically mean an accurate hypocenter determination.

The average squared residual e/n is called the variance of the data. Formally, n should here be the number of degrees of freedom, *ndf*, which is the number of observations – number of parameters in fit (here 4). Since n usually is large, it can be considered equal to number of degrees of freedom. This also means that RMS^2 is approximately the same as the variance.

The least squares approach is the most common measure of misfit since it leads to simple forms of the equations in the minimization problems (see later). It also works quite well if the residuals are caused by uncorrelated Gaussian noise. However in real problems this is often not the case. A particularly nasty problem is outliers, which is individual large

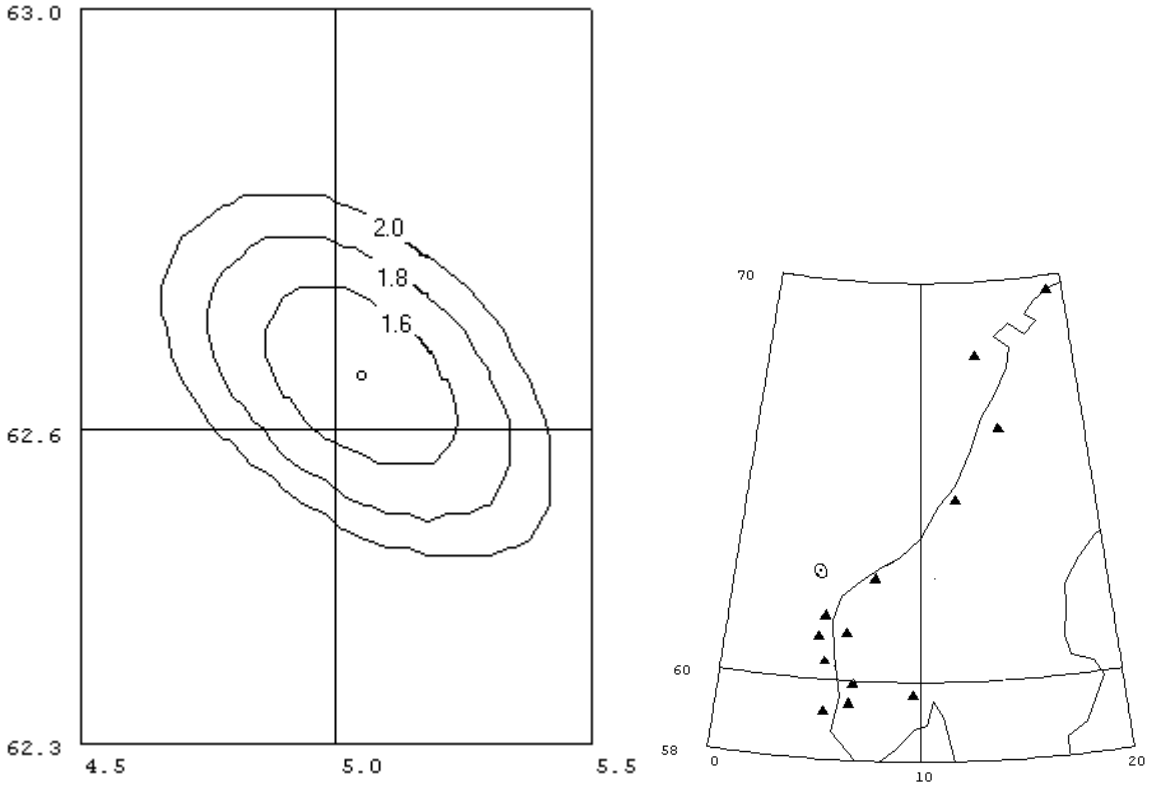
residuals. A residual of 4 will contribute 16 times more to the misfit, e, than a residual of 1. This problem could partly be solved by using the sum of the absolute residuals as a norm for the misfit

$$e1 = \sum_{i=1}^{n} |r_i| \qquad (8)$$

This is called the L1 norm and is considered more robust when there are large outliers in the data. It is not much used in standard location programs since the absolute sign creates complications in the equations. This is of course not the case for grid search. Most location programs will have some scheme for weighting out large residuals (see later), which partly solves the problem.

Once the misfits (e.g. *RMS*) has been calculated at all grid points, one could assign the point with the lowest *RMS* as the 'solution'. For well behaved data, this would obviously be the case, but with real data, there might be several points, even far apart, with similar *RMS* and the next step is therefore to estimate the probable uncertainties of the solution. The simplest way to get an indication of the uncertainty, is to contour the *RMS* as a function of *x* and *y* (2D case) in the vicinity of the point with the lowest *RMS* (Figure 4).

Figure 4
Right: RMS contours (sec) from a grid search location of an earthquake off western Norway (left).The grid size is 2 km. The circle in the middle indicates the point with the lowest RMS (1.4 sec). Left: The location of the earthquake and the stations used. Note the elongated geometry of the stations. The RMS ellipse from the figure on the left is shown as a small ellipse. Latitudes are degrees North and longitudes degrees West.



Clearly, if *RMS* is growing rapidly when moving away from the minimum, a better solution has been obtained than if *RMS* grows slowly. If *RMS* is contoured in the whole search area, other minima of similar size might be found indicating not only large errors but a serious ambiguity in the solution.

Location by iterative methods

Despite increasing computer power, earthquake locations are mainly done by other methods than grid search. These methods are based on linearizing the problem. The first step is to make a guess of the hypocenter and origin time (x_0,y_0,z_0,t_0). In its simplest form, this can be done by using a location near the station with the first arrival time and using that arrival time as t_0. Other methods also exists, see later. In order to linearize the problem, it is now assumed that the true hypocenter is close enough to the guessed value so that residuals at the trial hypocenter is a linear function of the correction we have to make in hypocentral distance.

The calculated arrival times at station i, t_i^c from the trial location are, as given in (4), $t_i^c = T(x_0,y_0,z_0,x_i,y_i,z_i) + t_0$ and the travel time residuals r_i are $r_i = t_i^o - t_i^o$ .We now assume that these residuals are due to the error in the trial solution and the corrections needed to make them zero are Δx, Δy, Δz and Δt. If the corrections are small, we can calculate the corresponding corrections in travel times by approximating the travel time function by a Taylor series and only using the first term. The residual can now be written:

$$r_i = (\partial T/\partial x_i) * \Delta x + (\partial T/\partial y_i) * \Delta y + (\partial T/\partial z_i) * \Delta x + \Delta t \qquad (9)$$

In matrix form we can write this as

$$\boldsymbol{r} = \boldsymbol{G} * \boldsymbol{X} \qquad (10)$$

where $\boldsymbol{r}$ is the residual vector, $\boldsymbol{G}$ the matrix of derivatives (with 1 in the last column corresponding to the time correction term) and $\boldsymbol{X}$ is the unknown correction vector in location and origin time.

This is a set of linear equation with 4 unknowns (corrections to hypocenter and origin time), and there is one equation for each observed phase time. Normally there would be many more equations than unknowns (e.g. 4 stations with 3 phases each would give 12 equations). The best fit to equation (9) is usually obtained with standard least squares techniques and the corrections to the hypocenter and origin time is obtained. The original trial solution is then corrected and this new solution is used as trial solution for the next iteration etc. This method was first proposed by Geiger (1912) and is called the Geiger method of earthquake location. The iterative process usually converges rapidly unless the data is badly configured or the initial guess is very far from the true solutions (see later). However, it also happens that the solution converges to a local minimum and this would be hard to detect in the output unless the residuals are very bad. A test with a grid search program could tell if the minimum is local or tests could be made with several start locations.

So far we have only dealt with observations in terms of arrival times. Many 3-component stations and arrays now routinely report azimuth of arrival ϕ. It is then possible to locate events with only one station and P and S-times (Figure 1), however, the depth must be fixed. If one or several azimuth observations are available, they can be used together with the

arrival time observations in the inversion and the additional equations for the azimuth residual are

$$r_i^{\phi} = (\partial\phi/\partial x_i) * \Delta x + (\partial\phi/\partial y_i) * \Delta y \qquad (11)$$

Equations of this type are then added to (9) or (10). The Δx and Δy in (11) are the same as for equations (9), however the residuals are in degrees so in order to make an overall *RMS*, the degrees must be 'converted to seconds' in terms of scaling. In one program, Hypocenter (Lienert and Havskov, 1995), a 10 deg azimuth residual was optionally made equivalent to 1 sec travel time residual. Using e.g. 20 degrees as equivalent to 1 sec would lower the weight of the azimuth observations.

Equations (9) are written without discussing whether working with a flat earth or a spherical earth. However, the principle is exactly the same. The travel times and derivatives are often calculated by interpolating in tables and in principle it is possible to use any earth model including 2D and 3D models. However, in practice 1D models are used, since 2D and 3D models are not known well enough and the computations are very time consuming. For local seismology, it is a common practice to specify a 1D crustal model and calculate arrival times for each ray while for global models, interpolation in travel time tables is the most common.

Example of location in a homogeneous model

The simplest case for earthquake location is a homogeneous medium. The travel times can be calculated as

$$T_i = \frac{\sqrt{(x-x_i)^2 + (y-y_i)^2 + (z-z_i)^2}}{v} + t_0 \qquad (12)$$

where v is the velocity. The partial derivatives can be made from (12) and for x, the derivative is

$$\frac{\partial T_i}{\partial x} = \frac{(x-x_i)}{v} * \frac{1}{\sqrt{(x-x_i)^2 + (y-y_i)^2 + z^2}} \qquad (13)$$

Similar expressions can be made for y and z. Table 1 gives an example of locating an earthquake with 10 stations in a model with constant velocity (from Stein, 1991). The stations are from 11 to 50 km from the hypocenter. The earthquake has an origin time of 0 seconds at the point (0,0,10) km. The starting location is at (3,4,20) km at 2 seconds. The exact travel times were calculated using a velocity of 5 km/sec and the iterations were done as indicated above. At the initial guess, the sum of the squared residuals were 92.4 sec^2, after the first iteration it was reduced to 0.6 sec^2 and already at the second iteration, the 'correct' solution was obtained. This is hardly surprising since the data had no errors. We shall later see how this works in the presence of errors.

Table 1
Inversion of error free data. Hypocenter is the correct location, Start is the start location, and the location is shown for the 2 following iterations. All units are km and seconds.

	Hypocenter	Start	1. iteration	2. iteration
x	0.0	3.0	-0.5	0.0
y	0.0	4.0	-0.6	0.0
z	10.0	20.0	10.1	10.0
t_0	0.0	2.0	0.2	0.0
e		94.2	0.6	0.0
RMS		3.1	0.25	0.0

Location errors

Since earthquakes are located with arrival times that contain observational errors and the travel times are calculated assuming we know the model, all hypocenters will have errors. Contouring the grid search *RMS* (Figure 4) gives an indication of the uncertainty of the epicenter. Likewise it would be possible to make 3D contours to get and indication of the 3D uncertainty. The question is now how to quantify this measure. The *RMS* of the final solution is very often used as a criteria for 'goodness of fit'. Although it can be an indication, *RMS* depends on the number of stations and does not in itself give any indication of errors and *RMS* is not reported by e.g. PDE and ISC.

From Figure 4 it is seen that the contours of equal *RMS* are not circles. We can calculate contours within which there is a 67 % probability (or any other desired probability) of finding the epicenter (see below). We call this the error ellipse. This is the way hypocenter errors normally are represented. It is therefore not sufficient to give one number for the hypocenter error since it varies spatially. Standard catalogs from PDE and ISC give the errors in latitude, longitude and depth, however, that can also be very misleading unless the error ellipse has the minor and major axis NS or EW. In the example in Figure 4, this is not the case.. Thus the only proper way to report error is to give the full specification of the error ellipsoid.

Before going into a slightly more formal discussion of errors, let us try to get a feeling what elements affects the shape and size of the epicentral error ellipse. If we have no arrival time errors, there is no epicenter errors so the magnitude of the error (size of error ellipse) must be related to the arrival time uncertainties. If we assume that all arrival time errors are equal, only the size and not the shape of the error ellipse can be affected. So what would we expect to give the shape of the error ellipse ?. Figure 4 is an example of an elongated network with the epicenter off to one side. It is clear that in the NE direction, there is a good control of the epicenter since S-P times control the distances in this direction due to the elongation of the network. In the NW direction, the control is poor for the same reason. We would therefore expect an error ellipse with the major axis NW as observed. Another way of understanding why the error is larger in NW than in NE direction, is to look at equation (9). The derivatives $\partial T/\partial x$ will be much smaller then $\partial T/\partial y$ so the Δy-terms will have larger weight then the Δx-terms in the equations (strictly speaking derivatives with respect to NW and NE). Consequently, errors in arrival times will affect Δx more than Δy. Note, that if azimuth observations were available for any of the stations far North or South of the event, this would

drastically reduce the error estimate in the EW directions since $\partial\phi/\partial x$ is large while $\partial\phi/\partial y$ is nearly zero.

Another geometry of the stations would give another shape of the error ellipse. It is thus possible for any network to predict the shape and orientation of the error ellipses, and given a arrival error, also the size of the ellipse for any desired epicenter location. This could e.g. be used to predict how a change in network configuration would affect earthquake locations at a given site.

In all these discussions, it has been assumed that the errors have Gaussian distribution and that there is no systematic errors like clock error. It is also assumed that there is no errors from the structure. This is of course not true in real life, however error calculations become too difficult if we do not assume a simple error distribution and that all stations have the same arrival time error.

The previous discussion gave a qualitative description of the errors. We will now show how to calculate the actual hypocentral errors from the errors in the arrival times and the network configuration. The most common approach to earthquake location is based the least squares inversion and a Gaussian distribution of the arrival time errors in which case the statistics is well understood and we can use the Chi-Square probability density distribution to calculate errors. For a particular earthquake location, χ^2 can be calculated as:

$$\chi^2 = \frac{1}{\sigma^2}\sum_{i=1}^{n} r_i^2 \qquad (14)$$

where σ is the assumed same standard deviation of any one of the residuals and n is the number of observations. We can now look in standard statistical tables (extract in Table 2) to find the expected value of χ^2 within a given probability. As it can be seen from the table, within 50% probability, χ^2 is approximately the number of degrees of freedom (*ndf*), which in our case is *n-4*.

Table 2.
The percentage points of the χ^2 distribution for different number of degrees of freedom (*ndf*)
.

ndf	χ^2 (95%	χ^2 (50%)	χ^2 (5%)
5	1.1	4.4	11.1
10	3.9	9.3	18.3
20	10.9	19.3	31.4
50	34.8	49.3	67.5
100	77.9	99.3	124.3

If e.g. an event is located with 24 stations (*ndf*=20), there is only a 5% chance that χ^2 will exceed 31.4. The value of χ^2 will grow as we move away from the best fitting epicenter and in the example above, the contour within which χ^2 is less than 31.4 will show the error ellipse within which there is 95 % chance of finding the epicenter. In practice, errors are mostly reported within 67 % probability.

The errors in the hypocenter and origin time can also formally be defined with the variance – covariance matrix σ_X^2 of the hypocentral parameters. This matrix is defined as

$$\sigma_X^2 = \begin{Bmatrix} \sigma_{xx}^2 & \sigma_{xy}^2 & \sigma_{xz}^2 & \sigma_{xt}^2 \\ \sigma_{yx}^2 & \sigma_{yy}^2 & \sigma_{yz}^2 & \sigma_{yt}^2 \\ \sigma_{zx}^2 & \sigma_{zy}^2 & \sigma_{zz}^2 & \sigma_{zt}^2 \\ \sigma_{tx}^2 & \sigma_{ty}^2 & \sigma_{tz}^2 & \sigma_{tt}^2 \end{Bmatrix} \qquad (15)$$

The diagonal elements are variances of the location parameters *x,y,z* and t_0 while the off diagonal elements give the coupling between the errors in the different hypocentral parameters. For more details, see .e.g. (Stein, 1991). The nice property about σ_X^2 is that it is simple to calculate:

$$\sigma_X^2 = \sigma^2 * (\boldsymbol{G}^T\boldsymbol{G})^{-1} \qquad (16)$$

where σ^2 is the variance of the arrival times multiplied with the identity matrix and $\boldsymbol{G}^T$ is $\boldsymbol{G}$ transposed. The standard deviation of the hypocentral parameters are thus given by the square root of the diagonal elements and these are the usual errors reported. So how can we use the off diagonal elements ?. Since σ_X^2 is a symmetric matrix, it can be represented by a diagonal matrix in a coordinate system which is rotated relative the reference system. We now only have the errors in the hypocentral parameters, and the error ellipse simply have semi axes σ_{xx}, σ_{yy}, and σ_{zz}. The main interpretation of the off diagonal elements is thus that they define the orientation and shape of the error ellipse. A complete definition therefore requires 6 elements. Equation (15) also shows, as earlier stated intuitively, that the shape and orientation of the error ellipse only depends on the geometry of the network and the crustal structure while the standard deviation of the observations is a scale factor.

The critical variable in the error analysis is thus the arrival time variances σ^2. This value is usually larger than would be expected from timing and picking errors alone, however it might vary from case to case. Setting a fixed value for a given data set could result in unrealistic error calculations. Most location programs will therefore estimate σ from the residuals of the best fitting hypocenter:

$$\sigma^2 = \frac{1}{ndf}\sum_{i=1}^{n} r_i^2 \qquad (17)$$

Division by *ndf* rather then by *n* compensates for the improvement in fit resulting from the use of the arrival times from the data. However, this only partly works and some programs allow setting an priori value which is only used if the number of observations is small. For small networks this can be a critical parameter.

Example of error calculation

We can use the previous error free example (Table 1) and add some errors (from Stein 1991). We add Gaussian errors with a mean of zero and a standard deviation 0.1 sec to the arrival times. Now the data are inconsistent and cannot fit exactly. As it can be seen from the results in Table 3, the inversion now requires 3 iterations (2 before) before the location do not change anymore. The final location is not exactly the location used to generate the arrival times and the deviation from the correct solution is 0.2, 0.4 and 2.2 km for x, y and z respectively and 0.2 sec for origin time. This gives an indication of the location errors.

Table 3
Inversion of arrival times with a 0.1 sec standard error. Hypocenter is the correct location, Start is the start location, and the location is shown for the 3 following iterations. All units are km and seconds.

	Hypocenter	Start	1. iteration	2. iteration	3. iteration
x	0.0	3.0	-0.2	0.2	0.2
y	0.0	4.0	-0.9	-0.4	-0.4
z	10.0	20.0	12.2	12.2	12.2
t_0	0.0	2.0	0.0	-0.2	-0.2
e		93.7	0.33	0.04	0.04
RMS		3.1	0.25	0.06	0.06

It is now interesting to compare to what is obtained with the formal error calculation. Table 4 gives the variance – covariance matrix. Taking the square root of the diagonal elements we get a standard deviation of x,y,z and t_0 of 0.3, 0.3 and 1.1 km and 0.1 secs respectively. This is close to the 'true' error so the solution is quite acceptable. Also note the *RMS* is close to the standard error.

Table 4
Variance – covariance matrix for the example in Table 3

	x	y	z	t
x	0.06	0.01	0.01	0.00
y	0.01	0.08	-0.13	0.01
z	0.01	-0.13	1.16	-0.08
t	0.00	0.01	-0.08	0.0

The variance – covariance matrix shows some interesting features. As seen from above, the error is much larger in the depth estimate than in x and y. This clearly reflects that the depth is less well constrained then the epicenter which is quite common unless there are stations very close to the epicenter. We have for simplicity calculated the standard deviations from the diagonal terms, however since the off diagonal terms are not zero, the true errors might be larger. In this example it can be shown that the semimajor and semiminor axis have lengths of 0.29 and 0.24 km respectively and the semimajor axis trends N22°E so the difference from the original diagonal terms is small.

The zt term, the covariance of the depth and origin time, is negative, indicating a trade-off between the focal depth and the origin time. This is commonly observed in practice

and is more prone to happen if only P-phases are used such that there is no strong limitation in distances.

Error calculation is a fine art, there are endless variations on how it is done and different location programs will usually give different results.

Master event technique

The relative location between events within a certain region can often be made with a much greater accuracy than the absolute location of any of the events. This is the case when velocity variations outside the local region is the major cause of the travel time residuals such that residuals measured at distance stations will be very similar for all of the local events. Usually, the events in the local area are relocated relative one specially well located event called the master event. It should be clear that the Master Event Technique only can be used when the distance to the stations is much larger than the distance between the events.

Most programs can be used for master event location. The individual station residuals for the master event are all assumed to be caused by velocity variations outside the region. By using these station residuals as station corrections, location of the remaining events will be made relative to the master event since all relative changes in arrival times are now entirely due to changes in location within the local region. It is obvious that only stations and phases for which observations are available for the master event can be used for the remaining events. Ideally, the same stations and phases should be used for all events.

Joint hypocenter location

In the Master Event Technique, it was assumed that true structure dependent residuals could be obtained absolute correct from the master event, however other errors could be present in the readings for the master event. A better way is to determine the most accurate station residuals using the whole local data set. This is what Joint Hypocenter Determination (JHD) is about. Instead of determining one hypocenter and origin time, we will jointly determine *m* hypocenters and origin times and *n* station corrections. This is done by adding the stations residuals Δt_i^s to equation 9 and writing the equations for all *m* earthquakes (index *j*):

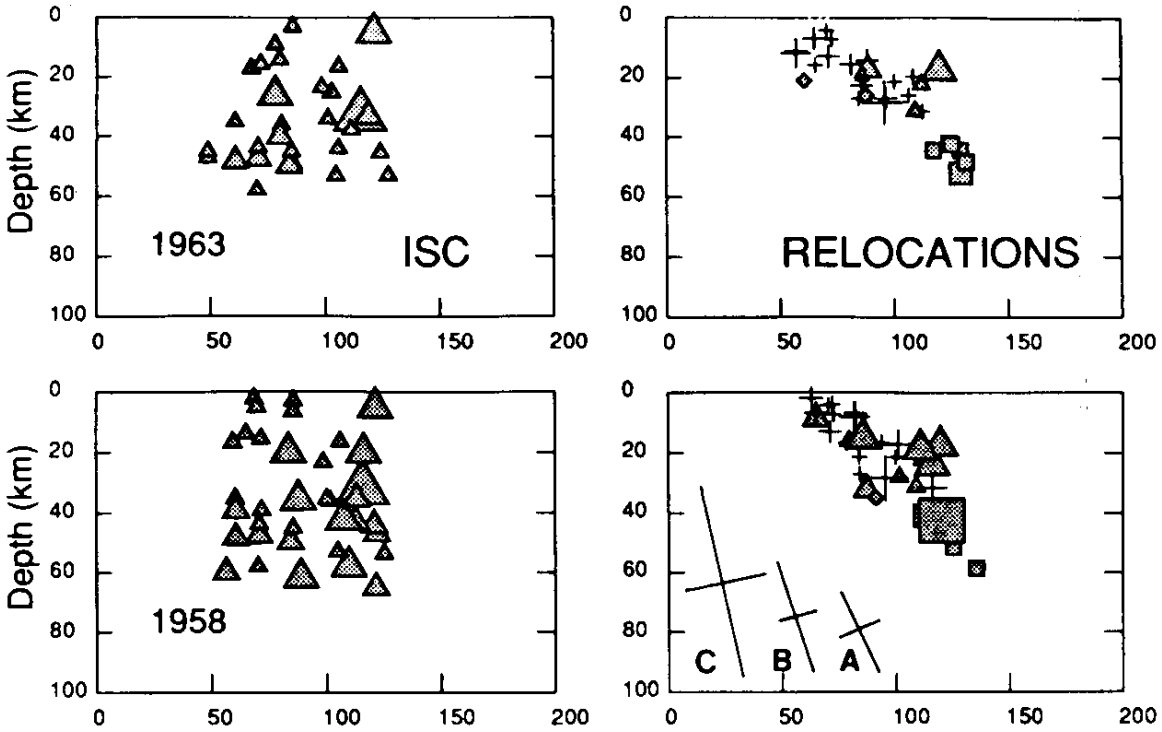
$$r_{ij} = (\partial T/\partial x_{ij}) * \Delta x + (\partial T/\partial y_{ij}) * \Delta y + (\partial T/\partial z_{ij}) * \Delta x + \Delta t_i^s + \Delta t_j \qquad (18)$$

The method of JHD was first proposed by Douglas (1967). Since the matrix $\boldsymbol{G}$ is now much larger than the 4 x 4 matrix for the a single event location, efficient inversion schemes must be used. If we e.g. use 20 stations with 2 phases each for 10 events, there will be 20 *10 *2 = 400 equations and 60 unknowns (10 hypocenters and origin times, and 20 station residuals).

The relative locations obtained by Master Event Technique and JHD are usually much better than the individual relative locations. However, only if we have the absolute location of one of the local events (like a known explosion) will we be able to convert the relative locations to absolute locations while for JHD, absolute locations are obtained. Accurate relative locations are useful to study e.g. the structure of a subduction zone or the geometry of aftershocks which might indicate the orientation of the fault. Figure 5 shows an example of using JHD.

Figure 5

Comparison of earthquake locations using the normal procedure at ISC (left) and JHD relocations (right). The events are located in the Kurile subduction zone along the rupture zones of large thrust events in 1963 and 1958. The vertical cross sections shown traverse the thrust zone from left to right. Note that the JHD solutions reduce scatter and makes it possible to define a dipping plane. (From Swartz et al,1989).



Practical consideration in earthquake locations

This section is intended to give some practical hints on earthquake location. The section does not refer to any particular location program, but most of the parameters discussed can be used with the Hypocenter program (Lienert and Havskov, 19945.

Phases

The most unambiguous phase to pick is usually P and P is the main phase used in most teleseismic locations. For local earthquakes, usually S-phases are also used. Using phases with different velocities has the effect of constraining the distances and there is then little

trade off between depth and origin time or epicenter location and origin time if the epicenter is outside the network. In general, it is thus an advantage to use as many different phases as possible under the assumption that they are correctly identified. One very wrong phase can throw off an otherwise well constrained solution.

The majority of location programs for local earthquakes only use first arrivals (e.g. HYPO71). This is ok is many cases. In some distance ranges, Pn is the first arrival, and it usually has a small amplitude. This means that the corresponding Sn phase automatically used by the program, might have a very small amplitude and the phase read is Sg or Lg. Since the program automatically assumes a first arrival, a wrong travel time curve is used for the observed phase, resulting in a systematic location error. This error is amplified by the fact that the S-phase, due to its low velocity, has a larger influence on the location than the P-phase. It is therefore important to use location programs where all crustal phases can be specified.

Starting location

Iterative location programs commonly start at a point near the station recording the first arrival. This is good enough for most cases, particularly when the station coverage is good and the epicenter is near or within the network. However, this can also lead to problems when using least squares techniques, which converge slowly or sometimes not at all for events outside the limits of a regional network (Buland, 1976). Another possibility is that the solution converges to a local minima which might be far from the correct solution. For small elongated networks, two potential solutions may exist at equal distance from the long axis. A starting location close to the first arrival station can then bias the final solution to the corresponding side of such a network. Although this bias usually is on the correct side, any systematic error in the first-arrival station's time can have a disproportionately large effect on the final location. Thus in many cases, it is desirable to use a better start location than the nearest station. There are several possibilities:

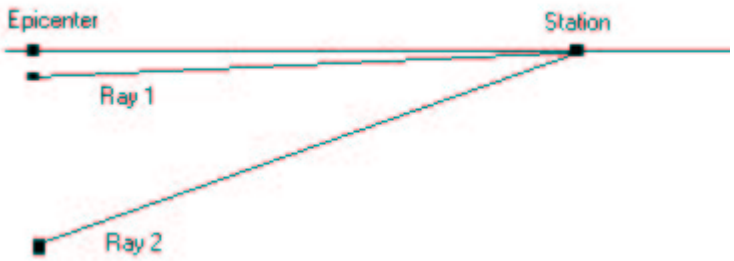
a) In many cases the analyst knows by experience the approximate location and can then manually give a start location. Most programs have this option.
b) Similar phases at different stations can be used to determine the apparent velocity and azimuth of a plane wave using linear regression on the arrival times with respect to the horizontal station coordinates. With the apparent velocity and/or S-P times, an estimate of start location can be made. This method is particularly useful when locating events fart away from the network (regionally or globally).
c) Azimuth information is frequently available from 3 component stations or seismic arrays and can be used as under b,
d) S-P and the circle method can be used with pairs of stations to get an initial location.

The starting depth is usually a fixed parameter and set to the most likely depth for the region. For local earthquakes that is usually in the range 10-20 km while for distant events it is often set to 33 km. If depth phases like e.g. pP are available for distant events, these phases can be used to set or fix the depth, see next section.

Hypocentral depth

| The hypocentral depth is the most difficult parameter to determine due to the fact that the travel time derivative with respect to depth changes very slowly as a function of depth (see Figure 6) unless the station is close to the epicenter. In other words, the depth can be moved up and down without changing the travel time. Figure 6 shows a shallow (ray 1) and a deeper event (ray 2). It is clear that the travel time derivatives with respect to depth is nearly zero for ray 1 and but not for ray 2. In this example, it would thus be possible to get an accurate depth estimate for the deep event but not for the shallow earthquake. Unfortunately, most rays are like ray 1 and locations are therefore often made with a fixed 'normal' start depth and only after the a reliable epicenter is obtained will the program try to iterate for the depth. Another possibility is to locate the event with several starting depths and use the depth that gives the best fit to the data. Although one depth will give a best fit to the data, the depth estimate might still be very uncertain and the error estimate must be checked.

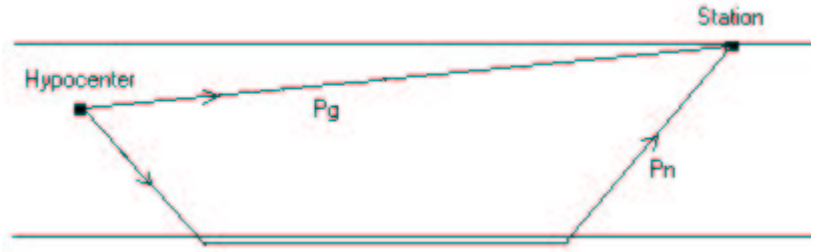
Figure 6
The depth – distance trade off in determination of focal depth.



For teleseismic events, the best way to improve depth determination is to include readings from the so called depth phases pP, sP. The differential time pP-P is quite constant at a range of epicentral distance for a given depth so the depth can be determined nearly independently of the epicenter. Another way of getting a reliable depth estimate for teleseismic locations is to have both near and far stations available, however this is unfortunately only the case for a few events.

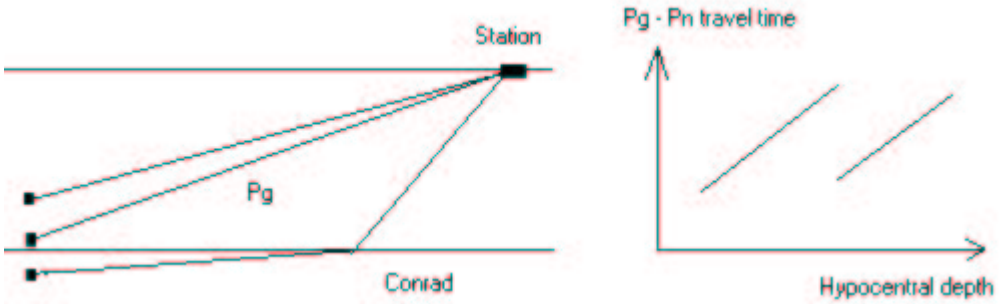
For local events, a rule of thumb is that stations should be no further away than 2 times the depth in order to get a reliable estimate (Figure 6). This is very often not possible, particularly for regional events. At distance larger than 2 * depth, the depth derivative change very little with depth if the first arriving phase is Pg, but at distances where Pb or Pn arrives, there is again a sensitivity to depth due to the steeply down going rays of Pb or Pn (Figure 7). So if stations are available at distances with both direct and refracted rays are first arrivals, reasonably reliable solutions might be obtained. An even better solution is when both Pg and Pn are available at the same station and the location capability could be similar to using P and pP for teleseismic events. The problem is that it might be difficult to identify secondary P-phases and a wrong identification might make matters worse.

Figure 7
Example of both Pg and Pn. Note the steep incidence angle for Pn as compared to Pg.



The depth estimate using a layered model remains problematic even with a mix of phases. Checking catalogs with local earthquakes, it will often be noted that there is a clustering of hypocenters at layer boundaries. This is caused by the discontinuities in the travel time curves of the direct phase Pg as a function of depth at layer boundaries, see Figure 8 for an example. The Pg travel time suddenly decreases when the hypocenter crosses a boundary (here Moho) since a larger part of the ray suddenly is in a higher velocity layer, while the Pn travel time continuously decrease as the depth increases. This gives rise to the discontinuities in the Pg –Pn travel time curve. So one Pn-Pg travel time is not enough to ensure a reliable depth estimate, several such phase arrivals must be available.

Figure 8
Ray paths of Pg phases as they cross the Conrad discontinuety (right). On the left is sketched the travel time curve of Pg-Pn as a function of depth.



Even when several Pg and Pn phases are available, depth estimates still remains a problems of regional distances due to the uncertainty in the crustal models. Since the depth estimates are critically dependent on the accurate calculation of Pg and Pn travel times, small uncertainties in the model can quickly throw off the depth estimate.

Many location programs give the *RMS* of the travel time residuals in a grid around the calculated hypocenter. This, in addition to the error estimates, gives an idea about the accuracy and a local minimum might be found in this way. A more direct way of estimating the quality of the depth estimate is to calculate the *RMS* as a function of depth in order to check if a local minimum has been reached.

Outliers

The largest residuals have disproportionally large influence on the fit of the arrival times due to the commonly use least squares fit. Most location programs will have some kind

residual weighting scheme in which observations where large residuals are given lower or even no weight. Bisquare weighing is often used for teleseismic events (Anderson, 1982). The residual weighting works very well if the residuals are not extreme since the residual weighting can only be used after a few iterations so that the residuals are close the final ones. Individual large residuals can often lead to completely wrong solutions, even when 90 % of the data is good and residual weighting will not help in these cases. Some programs will try to scan the data for gross errors (like minute errors) before starting the iterative procedure. If an event has large residuals, try to look for obvious outliers. A Wadati diagram can often help in spotting bad readings for local earthquakes, see Figure 3.

Weighting

The arrival time observations will by default always have different weight in the inversion. A simple case is that S-waves will have larger weight than P-waves due to the lower velocity. An extreme case is T-waves (guided waves in the ocean), which with their low velocity (1.5 km/sec) completely can dominate the solution. Considering, that the accuracy of the picks is probably is best for the P-waves, it is natural, that the P-arrivals have more importance than S-arrivals in the location. However, the default in most location programs is to leave the original weights unless the user actively changes the weight. It is normally possible to apriori give all S-phases a low weight and in addition, all phases can be given individual weights, including being weighted out.

When working with local earthquakes, the nearest stations will usually provide the most accurate information due to the clarity of the phases. In addition, uncertainty in the local model has less influence on the results at short distances than at larger distances, this is particularly true for depth. It is therefore desirable to put more weight on data from near stations than distance stations and this is usually done by using a distance weighting function of

$$w_d = \frac{x_{far} - \Delta}{x_{far} - x_{near}} \qquad (19)$$

where Δ is the epicentral distance, x_{near} is the distance to which full weight is used and x_{far} is the distance where weight is zero. The constants x_{near} and x_{far} are adjusted to fit the size of the network and x_{near} should be about the diameter of the network and x_{far} about twice x_{near} . For a dense network, x_{near} and x_{far} might be made even smaller to more accurate solutions. Distance weighting is not used for teleseismic locations since the global travel time tables have no particular distance bias and clarity of phases is not distance dependent.

Acknowledgements

This text has to a large extent followed a similar description in Shearer (1999). Figures and ideas has also been taken from Stein (1991) and Lay and Wallace (1995). Peter Bormann has also helped with text and ideas.

References

Anderson, K. R. (1982). Robust earthquake location using M-estimates, Phys. Earth. Plan. Inter, 30,119-130.

Buland,R. (1976). The mechanics of locating earthquakes, Bull. Seism. Soc. Am., 66, 173-187.

Båth, M.(1979). Introduction to seismology, 428 pp., Birkhauser Verlag. Basel.

Douglas, A. (1967). Joint hypocenter determination. Nature 215, 47-48.

Geiger, L. (1912). Probability method for determination of earthquake epicenters from the arrival times only. Bull. St. Louis Univ. 8,60-71.

Lay, T. and T. C. Wallace (1995). Modern global seismology, 521 pp., Academic Press.

Lienert, B. R. and J. Havskov (1995). A computer program for locating earthquakes both locally and globally. Seismological Research letters, 66, 26-36.

Roberts, R. G., A. Christofferson and F. Cassedy (1989). Real time event detection, phase identification and source location using single station 3 component seismic data and a small PC. Geophysical Journal, 97, 471-480.

Shearer, P. M. (1999). Introduction to seismology, 260 pp., Cambridge University Press.

Stein, S. (1991). Introduction to seismology, earthquakes and earth structure. Lecture notes, Department of Geological sciences, Northwestern University, USA.

Schwartz, S. Y., J. W Dewey and T. Lay (1989). Influence of fault plane heterogeneity on the seismic behavior in southern Kurile Islands arc. J. Geophys. Res. 94,5637-5649.