A Strategy for Earthquake Catalog Relocations Using a Maximum Likelihood Method

Ka Lok Li

Supervisors: Ólafur Gudmundsson and Ari Tryggvason
Abstract

A strategy for relocating earthquakes in a catalog is presented. The strategy is based on the argument that the distribution of the earthquake events in a catalog is reasonable a priori information for earthquake relocation in that region. This argument can be implemented using the method of maximum likelihood for arrival time data inversion, where the a priori probability distribution of the event locations is defined as the sum of the probability densities of all events in the catalog. This a priori distribution is then added to the standard misfit criterion in earthquake location to form the likelihood function. The probability density of an event in the catalog is described by a Gaussian probability density. The a priori probability distribution is, therefore, defined as the normalized sum of the Gaussian probability densities of all events in the catalog, excluding the event being relocated. For a linear problem, the likelihood function can be approximated by the joint probability density of the a priori distribution and the distribution of an unconstrained location due to the misfit alone. After relocating the events according to the maximum of the likelihood function, a modified distribution of events is generated. This distribution should be more densely clustered than before in general since the events are moved towards the maximum of the posterior distribution. The a priori distribution is updated and the process is iterated. The strategy is applied to the aftershock sequence in southwest Iceland after a pair of earthquakes on 29th May 2008. The relocated events reveal the fault systems in that area. Three synthetic data sets are used to test the general behaviour of the strategy. It is observed that the synthetic data give significantly different behaviour from the real data.
## Contents

1  Introduction .................................................. 1

2  Methodology .................................................. 3
   2.1  MLECL .................................................. 3
   2.2  Termination of iteration .................................. 6
   2.3  Approximation of a priori distribution .................. 7

3  Example: synthetic data ........................................ 8
   3.1  Origin time of earthquakes ................................ 9
   3.2  Computation of traveltime ................................ 9
   3.3  Computation of time residual .............................. 12
   3.4  Random measurement error ................................. 17
   3.5  Locating earthquakes ..................................... 17
   3.6  Results ................................................ 21

4  Example: aftershock sequence in SW Iceland .................. 23
   4.1  Results ................................................ 23

5  Discussion .................................................. 24

6  Conclusions .................................................. 27

7  Acknowledgements ........................................... 28
1 Introduction

Earthquake location is a common task in seismology. During an earthquake location process, a velocity model, which may be three-dimensional, is assumed and the arrival times predicted from this velocity model are matched as well as possible to the measured arrival times. A point location is obtained based on an optimization criterion. Since the problem is non linear, the location process can be done by applying a linearized approach or a grid search approach. The optimization criterion often applied is the quadratic measure of misfit, i.e. the least-squares misfit or the weighted-least-squares misfit. The problem is generally over determined since the number of measurements from the earthquake stations is generally larger than the parameters that have to be computed in the location process. Therefore, no regularization is needed in general. The location of a specific event is therefore independent of the other events in the earthquake catalog.

Earthquakes usually occur on faults. The earthquake locations are, however, diffused around the faults. The diffuseness of the earthquake locations can be a natural property of the earthquake process. But, the diffuseness can also be artificial. It may be due to the imprecise measurements of arrival times or an oversimplified velocity model is used and this model does not account for the three-dimensional variation of velocity in the region. Besides, an oversimplification in computing the predicted traveltime by ray theory can also lead to diffuseness of earthquake locations. The above errors can, of course, be minimized during the location process, e.g. by using a velocity model that better describes the region, but it is never possible for the uncertainties to be eliminated, since arrival times, being observational data, are themselves subject to uncertainties. In summary, a large portion of the diffuseness of earthquake locations is due to artificial causes while the natural property of earthquake process may also contribute to the diffuseness.

Due to the diffuseness of earthquake locations in a catalog, one may ask if the information in the earthquake catalog can be utilized in a better way to obtain a "better" estimate of locations such that the locations can better reflect possible structures of the region. This is the essential of earthquake relocation. A simple way to relocate earthquakes is to use the so-called nearest neighbour approach. In this scheme, the locations which seem
to detract the structures are being ignored. Besides, if some specific structures are expected based on geological or geophysical information, they can be included as a priori information. For example, if the earthquakes occur on a fault, one would expect that the structures to be sought are planes. A strategy, called the collapsing method, to reduce the diffuseness of earthquake locations was proposed by Jones and Stewart (1997). They argued that the earthquake locations are subjected to uncertainties and they assumed these uncertainties are known. The earthquake locations are then relocated in a way that each event is attracted towards the centre of mass of all events within a specific confidence interval distance from it. The idea behind this strategy is to obtain a simpler alternative distribution of seismicity without compromising the data fit significantly. Nicholson et al. (2000) quantified this simplification of earthquake distribution using the concept of entropy. They also modified the method to make use of the additional temporal information in the earthquake distribution. Asanuma et al. (2001) improved the collapse method by generalizing it to have one or two dimensional attraction in additional to a point attraction.

In this report, we propose another strategy for relocating earthquake locations in a catalog. Although the strategy presented in this report has the same objective as the collapsing method, i.e. to relocate the events in an earthquake catalog such that possible structures can be probed from the relocated catalog of events, the strategy we propose is conceptually different from that proposed by Jones and Stewart (1997). Our strategy is based on a well-known probabilistic concept, namely, the method of maximum likelihood subject to a set of a priori constraints on the locations. The a priori constraints are in the form of an a priori probability distribution of the model parameters. We will later refer to this strategy as the Maximum Likelihood Earthquake Catalog Locations (MLECL).

An example from synthetic data sets is used to explore the general behaviour of MLECL. It also helps us to better understand the convergence properties of the strategy and provides guidelines to determine when to terminate the iteration. An example from an aftershock sequence in southwest Iceland is used to demonstrate the application of this strategy to real data.
2 Methodology

The method of maximum likelihood is a common method for estimating the model parameters of a statistical model given a data set. To use the method, the joint probability density function for all observations in the data set is first specified. With the observed values fixed, the estimators for the model parameters are then varied until the joint probability density function is maximized. The values are then taken to be the estimates of the model parameters.

2.1 MLECL

Consider an existing catalog of earthquake locations with information about the data and the uncertainties of both the data and the locations. The catalog may be obtained from the registered events of a local seismic network. Now consider a scenario where all the seismometers in the network malfunction due to power failure. A felt earthquake occurs and you are asked where that earthquake has occurred. A reasonable guess would be the location where most earthquakes have occurred in the past. We can therefore conclude that the distribution of the past events in an earthquake catalog can be treated as the a priori distribution of the event location in the region.

We then consider the location of an event based on data fit under some optimization criterion. The location of this event is uncertain. It is because the arrival times measured from the seismometers are uncertain and the velocity model used is also an incomplete description of the region. These lead to an uncertainty in computing the predicted arrival times. Therefore, the location should be defined as a point with uncertainty, or more precisely, a peaked probability distribution with the width of the distribution described by the uncertainty. In other words, the location is most likely somewhere within a finite volume specified by the confidence level. One can say it is equally possible for the true location to be at the left side or right side of the volume. However, if many of the previous events occurred at the right side, then it is reasonable to say that the right side should be a more favourable candidate for the location. This is consistent with the statement, concluded from the last paragraph, that the distribution of the past events is reasonable.
a priori information for earthquake relocation in the region.

One should note that an assumption has to be made for the above statement to be qualified for argument sake. We must assume that the catalog of earthquake locations is nearly complete, i.e. earthquakes have occurred in all regions where they are likely to occur.

If the above assumption is satisfied, we can quantify the statement into a scheme for earthquake relocation. In a single-event earthquake location we measure the arrival times from different seismic stations and put them into a data vector $d$. We seek a model in a vector $m$ which describes the location of the event. If the data errors are Gaussian, we can define a probability distribution for the model,

$$P_d(m) = k_1 \exp \left[ -\frac{1}{2} \left[d - G(m)\right]^T C_d^{-1} \left[d - G(m)\right] \right], \quad (1)$$

where a subscript $d$ is used to indicate that the probability distribution is based on data fit. $G$ is a generalized function of the model $m$ that predicts the observations based on, e.g. ray theory. $C_d$ is the data covariance matrix and $k_1$ is a normalization constant. To find $m$ we maximize eq. (1) by minimizing the misfit,

$$Q = \left[d - G(m)\right]^T C_d^{-1} \left[d - G(m)\right]. \quad (2)$$

Let us denote the model estimate by $\hat{m}$. Near the optimal model $\hat{m}$ we can apply a linear approximation to $G$ such that the probability distribution of model can be expressed as

$$P_m(m) = k_2 \exp \left[ -\frac{1}{2} (m - \hat{m})^T C_m^{-1} (m - \hat{m}) \right] \quad (3)$$

in the model space. $C_m$ is the model covariance matrix, which can be derived from the data covariance matrix $C_d$ (Menke, 1989).

In our strategy, we add an a priori probability distribution to the above optimization
criterion. The model probability distribution is then described as

\[ P_j(m) = P_{mj}(m)P_{aj}(m), \]

where \( P_{aj}(m) \) is the a priori distribution and \( P_j(m) \) is the posterior distribution. Here we add a subscript \( j \) to distinguish between different events. Eq. (4) assumes that the a priori distribution \( P_{aj}(m) \) is independent of the distribution based on misfit \( P_{mj}(m) \).

The distribution of events in the catalog can be represented in various ways. Our strategy first defines a probability density for every event location. We then sum the probability distributions for all events and use the normalized sum as the a priori probability, i.e.

\[ P_{aj}(m) = k_3 \sum_{k \neq j} P_{mk}(m) \]

\[ = k_3 \sum_{k \neq j} \exp \left[ -\frac{1}{2}(m_k - \hat{m}_k)^T C_{mk}^{-1}(m_k - \hat{m}_k) \right]. \]

Note that the distribution for the \( j \)th event is excluded from the a priori probability density for that particular event. It is vital to exclude this distribution. Otherwise, the information would be used twice.

We now have a re-formulated description of the distribution for the \( j \)th event. This distribution includes both the contribution from the data misfit and the a priori distribution, and is used as the optimization criterion for relocating events. The relocation is done event by event. After relocating all events in the catalog, a new distribution of events is obtained. This new distribution is used to construct a new a priori distribution, and the process is repeated. This iterative relocating scheme is what we refer to as MLECL. In the above discussion we used eq. (3) to formulate the iterative scheme. One can, of course, use eq. (1), rather than the approximated equation, to formulate the scheme.

MLECL gives a simpler distribution of events in a catalog. This simplified distribution should be a more likely distribution than the initial one provided that the catalog is nearly complete, i.e. the earthquakes have occurred in all seismogenic regions.
The whole process is iterated in a self consistent manner without any arbitrary smoothing parameters being introduced. The process converges to a tighter distribution of events with reduced error bounds. The peripheral events around a concentration of events are moved towards the centre of that concentration. Nevertheless, two clusters of events are not drawn to each other as in the collapsing method suggested by Jones and Stewart (1997).

2.2 Termination of iteration

When we apply MLECL to an earthquake catalog, we can repeat the strategy for any number of iteration. The strategy itself does not give a criterion for terminating the process. To make the strategy become practical, we must define a criterion to terminate the process. Another reason for the need of such a criterion is that the strategy will give an unphysical structure if too many iterations are performed. This phenomenon is explained as followed. For a confined cluster in the original distribution, it gives rise to a peak in the initial a priori distribution. This peak will attract all the neighbouring events towards it, thus creating a more confined peak in the a priori distribution for the next iteration. Eventually, after many iterations, the peak will evolve into an arbitrarily narrow cluster. This is, of course, not a physical and meaningful structure that we seek.

We follow the same argument as Jones and Stewart (1997) when we develop a criterion to terminate the iteration. Assume the data errors are Gaussian, if the problem is nearly linear, then the location uncertainties will also be Gaussian. The displacement moved by each event from the peak of a Gaussian distribution, normalized by the width of the distribution in the direction of displacement, is then $\chi^2_3$ distributed. Therefore, it is justified to terminate the iteration when the distribution of the normalized displacements approaches a $\chi^2_3$ distribution.

In some cases, however, the distribution of the normalized displacements may not approach a $\chi^2_3$ distribution, e.g. when the data errors are not Gaussian. In this situation, we can artificially set a criterion in which the iteration is terminated if the convergence rate is below some limit.
2.3 Approximation of a priori distribution

In the above section, we define the a priori probability distribution as

\[
P_{aj}(\mathbf{m}) = k_3 \sum_{k \neq j} P_{mk}(\mathbf{m})
= k_3 \sum_{k \neq j} \exp \left[ -\frac{1}{2} (\mathbf{m}_k - \hat{\mathbf{m}}_k)^T \mathbf{C}_{mk}^{-1} (\mathbf{m}_k - \hat{\mathbf{m}}_k) \right].
\] (6)

However, when we apply MLECL to our examples from the synthetic data sets and the aftershock sequence in Iceland, instead of the above equation, we use an approximated density distribution for the a priori distribution. In this approximated distribution, the contribution from the current event is not excluded. In other words, we use the following approximation:

\[
P_{aj}(\mathbf{m}) = k_3 \sum_{k \neq j} P_{mk}(\mathbf{m}) \approx k_3 \sum_{k} P_{mk}(\mathbf{m})
= k_3 \sum_{k} \exp \left[ -\frac{1}{2} (\mathbf{m}_k - \hat{\mathbf{m}}_k)^T \mathbf{C}_{mk}^{-1} (\mathbf{m}_k - \hat{\mathbf{m}}_k) \right].
\] (7)

The effect of this approximation is to introduce a local contribution to the a priori distribution. If the catalog is big, the effect of this local contribution on the overall a priori distribution is small. Including it does not change the overall structure of the a priori distribution significantly. The effect, however, can be significant locally. For an event which is close to the density cluster, the contribution of the event is small on the properties of a local peak. For a stand-alone event, which is far away from the density cluster, the effect will be significant locally. A local peak, which should not be there, will be introduced into the a priori distribution. This local peak will have two consequences for the relocation of events. First, the stand-alone events are moved less than they should since the local peak tends to reduce the effect of the density cluster. Second, the error estimates of stand-alone events become smaller than they should be. These two effects are more significant for the stand-alone events than the events inside the density cluster.

The reason for this approximation is solely to reduce the demand on computational
power. If eq. (6) is used, we have to store a three-dimensional density array for every event in the catalog during a whole iteration, and these densities should be densely sampled to avoid discretization artefacts. These require a large amount of random-access memory (RAM) even for a catalog of 8000 events. This big demand on computational power is, of course, because the MLECL program is not yet optimized for an effective usage of memory resources. One may consider, e.g. to parallelize the program, to speed up the process in the long term. However, it is necessary to use the approximation as stated in eq. (7) for the time being.

3 Example: synthetic data

In the synthetic data a catalog of 10000 earthquakes is generated. The events in the catalog are contained in a $40 \times 40$ km$^2$ region and extended to a depth of 12 km. Two faults, with length 20 km (fault 1) and 10 km (fault 2) along the $y$ direction, are located at 18.5 km and 21.5 km along the $x$ direction, respectively. Both faults extend to a depth of 7.5 km. Since earthquakes usually occur around a fault plane, the events are generated randomly around the two faults, with 60 % of the events around fault 1 and 40 % around fault 2. Figure 1 shows the distribution of earthquakes around the two faults. The events are distributed around the centres of the faults in the $x$ direction, described by two Gaussian distributions with specific standard deviations. In the $y$ direction, the events are randomly distributed along the whole fault planes and in the $z$ direction, the events are randomly distributed between a depth of 0.5 km and 7.5 km, in both cases as uniform densities.

In the region a network of 10 seismic stations is used. The stations are randomly located in the study region. The map view of the locations of the stations is shown in Figure 2. This geometry is chosen to simulate the example of real data discussed later.

To generate synthetic data for earthquake locations, we first have to generate the arrival times at different seismic stations for each event. The arrival times are then used to compute the locations and the uncertainties of event locations. In our synthetic data, a measurement of arrival time consists of the following components: (1) the origin time of
the event; (2) the traveltime for the seismic wave from the source to the station; (3) the time residual that accounts for three-dimensional heterogeneity of the velocity model and (4) the random error due to imprecise measurement. In the following subsections, we will discuss how we generated different components of the synthetic arrival time measurements.

3.1 Origin time of earthquakes

Since we are only using the locations and the uncertainties of locations in the relocation process, the origin times of the events are not our concern. We can, therefore, simply assume all the events occur at the same instant. A simple choice would be the origin times of all events equal zero. In other words, we remove the component of arrival time due to the origin time.

3.2 Computation of traveltime

The traveltime of the seismic wave is computed using a one-dimensional velocity model where velocity depends only on depth, i.e. \( v = v(z) \). This velocity model consists of different layers in depth. For each layer, the velocity is assumed to be a linear function of depth,

\[
v_i(z) = a_iz + b_i
\]

for the \( i \)th layer. \( a_i \) and \( b_i \) are constants and take different values for different layers. The constants \( a_i \) and \( b_i \) are so chosen that the velocity function is continuous at the boundaries between the layers. Figure 3 shows the velocity model used in the traveltime computation.

Since the velocity model is continuous and is linearly depended on depth only, the traveltime and the offset distance can be computed analytically given the source coordinates and the take-off angle (Slotnick, 1959). For any given source in space, the offset distance
$X(p)$ away from the source at which a ray dives to the maximum depth of penetration is

$$X = \int_0^{z_t} \frac{pv(z)}{\sqrt{1 - p^2 v(z)^2}} dz,$$  \hspace{1cm} (9)

where $z_t$ is the maximum depth of penetration and $p$ is the ray parameter. Substitute eq. (8) into eq. (9), we have

$$X = \int_0^{z_t} \frac{p(az + b)}{\sqrt{1 - p^2 (az + b)^2}}dz = \frac{1}{a} \int_b^{v_a} \frac{pv}{\sqrt{1 - p^2 v^2}}dv,$$  \hspace{1cm} (10)

where $v_a$ is the velocity at $z = z_t$. Here, for ease of notation, we drop the subscript $i$ with the implication that we are evaluating the offset distance $X(p)$ in the $i$th layer. To evaluate the above integral, we can set

$$S = pv.$$  \hspace{1cm} (11)

Then,

$$X = \frac{1}{ap} \int_{pb}^{pv_a} \frac{S}{\sqrt{1 - S^2}} dS = \frac{1}{ap} \left[ \sqrt{1 - S^2} \right]_b^{pv_a} = \frac{1}{ap} \left[ \sqrt{1 - p^2 b^2} - \sqrt{1 - p^2 v_a^2} \right].$$

Therefore, the offset distance $X(p)$ away from the source at which a ray dives to the maximum depth of penetration is

$$X = \frac{1}{ap} \left[ \sqrt{1 - p^2 b^2} - \sqrt{1 - p^2 (az_t + b)^2} \right].$$  \hspace{1cm} (12)

The traveltime from the given source to the point of maximum depth of penetration is
given by

\[ T = \int_0^{z_t} \frac{dz}{v(z) \sqrt{1 - p^2 v(z)^2}}. \] (13)

With similar derivation as for the offset distance, one can obtain the equation for evaluating traveltime, namely,

\[ T = \frac{1}{a} \ln \left[ \frac{(az_t + b)(1 + \sqrt{1 - p^2 b^2})}{b(1 + \sqrt{1 - p^2 (az_t + b)^2})} \right]. \] (14)

With application of eq. (12) and eq. (14) to every layer in the velocity model, one can evaluate the total traveltime and total offset distance for an up-going ray, i.e. a ray without diving into a deeper layer. However, for a down-going ray, the result is not so obvious since one has to take account of the turning point of the ray. To find out if the ray turns in a particular layer, one should first evaluate the maximum depth of penetration for that layer. By definition, a ray turns when the angle of refraction is 90°. Therefore, from Snell’s law, we have

\[ p = \frac{\sin \theta}{v_s} = \frac{\sin 90°}{v_{zt}} = \frac{1}{az_t + b} \] (15)

at the turning point. \( \theta \) is the take-off angle. \( v_s \) and \( v_{zt} \) are the velocities at source depth and maximum depth of penetration respectively. From eq. (15), we have

\[ z_t = \frac{1 - pb}{ap}. \] (16)

Then the condition to judge if the \( i \)th layer is the layer in which the ray turns is given by

\[ z_t < h_i, \] (17)

where \( h_i \) is the thickness of the \( i \)th layer. If the above condition is satisfied, the ray turns at this layer and the maximum depth of penetration for this layer is \( z_t \). Otherwise, the ray continues to dive into the deeper layer. After knowing the layer in which the ray turns,
one can compute the total traveltime and the total offset distance by summing up the individual traveltimes and offset distances for all layers which the ray has passed through.

One should notice that the total traveltime and the total offset distance cannot be evaluated unless a given ray parameter $p$, which in turn depends on the take-off angle $\theta$ for a particular ray, is specific in advance. Since the problem here is to find the traveltime for a given source and receiver configuration, the take-off angle should be known. A simple method to find the take-off angle is the shooting method. In this method, an arbitrary take-off angle is chosen. The angle is used to compute the ray parameter $p$ and eventually the total offset distance travelled by this specific ray. If the ray is overshot, a smaller take-off angle is then chosen. However, if the ray is undershot, a larger take-off angle is then chosen. The process is repeated until the error in offset distance is within certain limit,

$$|X_{est} - X_{true}| < \epsilon,$$  \(18\)

where $X_{est}$ is the offset distance computed from eq. (12), $X_{true}$ is the actual offset distance between the source and the receiver and $\epsilon$ is the error limit. With the take-off angle known, we can now compute the traveltime given the coordinates of the source and receiver.

### 3.3 Computation of time residual

When we compute the traveltime, only a one-dimensional velocity model is used. To make the situation more realistic and account for three-dimensional heterogeneity in the study region, a small random perturbation in velocity is added to the one-dimensional velocity model, which we refer to as the background velocity model.

A perturbation can be described by an autocorrelation function (ACF). An ACF is a statistical measure of the spatial correlation and the magnitude of the fluctuations in the medium. The generation of the perturbation is done by the realization of a random process. According to the Wiener-Khinchin theorem (Goodman, 1984), the ACF and the
power spectral density function (PSDF) form a Fourier transform pair, i.e.

\[
P(k) = \iiint_{-\infty}^{\infty} c(x)e^{ikx}dx
\]

(19)

\[
c(x) = \iiint_{-\infty}^{\infty} P(k)e^{-ikx}dk,
\]

(20)

where \(P(k)\) is the PSDF in the wavenumber domain and \(c(x)\) is the ACF in the space domain. Therefore, one can generate the perturbation according to a specific ACF by directly filtering the original spectrum of a white noise process with the square root of the PSDF in the wavenumber domain.

In our case, a pseudorandom realization \(W(x)\) of the white noise of a standard normal distribution is first generated in a rectangular grid. A three-dimensional Fourier transform \(W(k)\) of the white noise is then computed. The Fourier transform is multiplied by the square root of the PSDF of a Gaussian ACF, namely,

\[
U(k) = F(k)W(k),
\]

(21)

where

\[
F(k)F^*(k) = P(k) = \epsilon^2\lambda^3\sqrt{\pi^3}e^{-k^2\lambda^2/4}
\]

for a Gaussian ACF. \(F^*(k)\) represents the complex conjugate of \(F(k)\). \(\epsilon\) is the standard derivation of the ACF and \(\lambda\) is the correlation length. The three-dimensional inverse Fourier transform \(U(x)\) of the product \(U(k)\) is computed. \(U(x)\) is the perturbation we seek for the velocity model, and is added to the background velocity,

\[
v(x) = v_0(z) + U(x),
\]

(22)

where \(v_0(z)\) is the background velocity model and \(v(x)\) is the overall velocity model including the three-dimensional heterogeneity.

Now we have a velocity model with perturbation to the background velocity. To
account for this effect on our synthetic arrival time measurements, one has to compute
the time residual for every source-receiver configuration.

In general, the change in slowness in a velocity model has two effects on the traveltime
calculation. The first one is the effect of the change in slowness without changing the ray
and the second one is the effect of the changing ray path due to the change in slowness.
Mathematically,

$$\delta t = \int_{\text{ray}} \delta u(x) \, dl + \int_{\delta \text{ray}} u(x) \, dl ,$$

(23)

where $u(x)$ is the slowness. It turns out that the second term on the right hand side of
the equation is a second order term. Since the velocity (or slowness) perturbation in our
model is very small, the second order term will be much smaller than the first order term,
we are then justified to ignore the second term and the equation becomes

$$\delta t = \int_{\text{ray}} \delta u(x) \, dl .$$

(24)

In discrete case, the integral becomes a summation over all the ray segments along the
ray path, namely,

$$\delta t = \sum_{\text{ray}} \delta u(x) \Delta l .$$

(25)

In order to compute $\delta u$, one has to trace the ray from the source to the receiver and
compute $\delta u$ at every ray segment on the ray. The problem then becomes a ray tracing
problem. Since the velocity is a linear function of depth in each layer, the ray path is the
arc of a circle. For a two dimensional case, this circle has a centre located at

$$C = \left( \frac{\sqrt{1 - p^2 b^2}}{ap}, \frac{-b}{a} \right)$$

(26)
and radius

\[ R = \frac{1}{ap} \]  \hspace{1cm} (27)

Consider a ray passing through the \(i\)th layer of the velocity model, as shown in Figure 4, the coordinates of points A and B, where the ray enters and exits the layer, are known from eq. (12). The angle \(\alpha\) subtended by the arc \(l\) at the centre is given by

\[ \alpha = \sin^{-1} [pv(z_2)] - \sin^{-1} [pv(z_1)] \]  \hspace{1cm} (28)

Since \(l = R\alpha\) and \(R = 1/ap\) from eq. (27), we have

\[ l = \frac{1}{ap} \left\{ \sin^{-1} [pv(z_2)] - \sin^{-1} [pv(z_1)] \right\} \]  \hspace{1cm} (29)

Now, if we divide the ray into a number of small ray segments \(\delta l\), the region through which the ray segment passes can be considered to have a constant velocity perturbation. Then the effect of the velocity perturbation on the traveltime simply becomes the difference between the overall velocity and the background velocity at that region. However, this statement holds only when the length of the ray segment is much smaller than the correlation length in the velocity perturbation. We therefore choose \(\delta l\) to be 10% of the correlation length.

For a chosen value of \(\delta l\), the number of ray segments \(n\) in the layer is \(l/\delta l\) by definition. From eq. (29),

\[ n = \frac{l}{\delta l} = \frac{\sin^{-1} [pv(z_2)] - \sin^{-1} [pv(z_1)]}{ap\delta l} \]  \hspace{1cm} (30)

The angle \(\delta \alpha\) subtended by \(\delta l\) at the centre is

\[ \delta \alpha = \frac{\alpha}{n} = ap\delta l \]  \hspace{1cm} (31)

Here we estimate \(\delta u\) using the slowness difference at the midpoint of a ray segment, i.e.
the position where the arc length of the ray segment is \( \delta l/2 \). Accordingly, the angle subtended by that segment is \( \delta \alpha/2 \). Consider an up-going ray segment \( \delta l \), as shown in figure 5, the small increment \( \delta x_u \) in the \( x \) direction is

\[
\delta x_u = R \cos(\alpha_1 - \frac{\delta \alpha}{2}) - R \cos(\alpha_1)
= \frac{1}{ap} \left\{ \cos[\sin^{-1}(pv(z_1)) - \frac{\delta \alpha}{2}] - \cos[\sin^{-1}(pv(z_1))] \right\}
\]

and the small increment \( \delta z_u \) in the \( z \) direction is

\[
\delta z_u = R \sin \alpha_1 - R \sin(\alpha_1 - \frac{\delta \alpha}{2})
= \frac{1}{ap} \left\{ \sin[\sin^{-1}(pv(z_1))] - \sin[\sin^{-1}(pv(z_1)) - \frac{\delta \alpha}{2}] \right\}
= \frac{1}{ap} \left\{ pv(z_1) - \sin[\sin^{-1}(pv(z_1)) - \frac{\delta \alpha}{2}] \right\}.
\]

The subscript \( u \) in the above expressions refer to the increments for the up-going rays. The small increments for a down-going ray can be derived in a similar manner. Here we just list the results. The small increment \( \delta x_d \) in the \( x \) direction is

\[
\delta x_d = \frac{1}{ap} \left\{ \cos[\sin^{-1}(pv(z_1))] - \cos[\sin^{-1}(pv(z_1)) + \frac{\delta \alpha}{2}] \right\}
\]

and the small increment \( \delta z_d \) in the \( z \) direction is

\[
\delta z_d = \frac{1}{ap} \left\{ \sin[\sin^{-1}(pv(z_1)) + \frac{\delta \alpha}{2}] - pv(z_1) \right\}.
\]

The perturbation in slowness for a ray segment then becomes

\[
\delta u = u(x_1 + \delta x, z_1 + \delta z) - u_0(x_1 + \delta x, z_1 + \delta z),
\]

where \( u \) and \( u_0 \) are the overall and background slowness respectively. The results are multiplied by the length of the ray segment and summed over all ray segments and all layers according to eq. (25). The time residuals are then added to the traveltimes computed in
the last section for every source-receiver configuration.

3.4 Random measurement error

Every observational measurement contains a measurement error. No matter how good a measurement is, the measurement error can never be eliminated. One can only increase the precision of a measurement to a certain extent using, e.g. a better data-collection technique or better equipment. Since our synthetic arrival times simulate observational data, one has to account for this uncertain nature of the data. We, therefore, introduce random measurement errors to our synthetic arrival times. We choose the distribution of the random errors to be a Gaussian distribution with zero mean and a specific standard deviation. This random error is added to every arrival time measurement.

3.5 Locating earthquakes

With all the arrival time measurements at hand, we are now able to locate the earthquakes. As we mentioned in the introduction section, there are many approaches for earthquake location. In our location process, we use a nested grid search approach.

In the nested grid search approach, a grid of points is first constructed in the study area. Every node on the grid is treated as a potential location for the event. The predicted traveltimes, and hence the misfit measures, of all the potential locations are computed. The location with the smallest misfit is picked as the best location for this iteration. A finer grid of points is then constructed around the best location in the previous iteration. The predicted traveltimes and the misfit measures are computed again to obtain a refined estimate. The whole process is repeated to obtain a further refined estimate of event location.

To pick the best location in the grid search, we have to define a misfit measure. The misfit measure $Q$ is given by the sum of squared differences between the predicted arrival times and the observed arrival times, averaged over all seismic stations or seismic phases,
namely,

$$Q = \frac{1}{N} \sum_{i=1}^{N} (t_{\text{pre}i} - t_{\text{obs}i})^2,$$

(37)

where $N$ is the number of seismic stations or seismic phases, $t_{\text{pre}i}$ and $t_{\text{obs}i}$ are the predicted and observed arrival times for the $i$th station or $i$th phase respectively.

The predicted arrival time at the $i$th seismic station or for the $i$th seismic phase is

$$t_{\text{pre}i} = G(x_i, x_0) + t_0,$$

(38)

where $G(x_i, x_0)$ is the traveltime, which is a function of the location of the $i$th station $x_i$ and the potential location of the event $x_0$, and $t_0$ is the origin time. Since the origin time is linear with the arrival times, we can estimate the origin time as

$$\hat{t}_0 = \frac{1}{N} \sum_{i=1}^{N} [t_{\text{obs}i} - G(x_i, x_0)].$$

(39)

This estimated origin time is used to compute the predicted arrival times according to eq. (38). We then use the predicted arrival times, together with the observed arrival times to compute the misfit for each potential location.

It is noted that only the P wave arrivals are used in the location process. It is known that including the S wave arrivals has an advantage of resolving the ambiguity between the source depth and the origin time estimation. For an event occurs nearly at the same distance from all seismic stations in a network, the estimation of the origin time and source depth for such an event is usually unsatisfactory. It is because a predicted arrival time with an earlier origin time and shallower depth will match the measured arrival time as well as a predicted arrival time with later origin time and deeper depth. This is what we usually refer to as the trade-off between the origin time and the source depth. A solution to this problem is to include the S wave arrivals. The difference between the P wave and S wave arrivals can be used to provide an independent estimation of the origin time which is used as a constraint in the location process. Although including S wave arrivals
can reduce the trade-off between the origin time and the source depth, it is not obvious
to decide suitable weights for the P wave and S wave arrivals. This usually requires a
trial-and-error process. Since our main objective of this project is to generate synthetic
data sets to test the relocation strategy, to avoid digging into the question endlessly, we
simply use the P wave arrivals to locate the events.

After the event location is estimated, we have to estimate the uncertainty of that
estimated location, which is manifested by the standard deviations of location in the $x$,
y and $z$ directions. In inverse theory, to estimate the variances of the model parameters,
one can first estimate the variances of data and map them into the model parameters
using the generalized inverse. Mathematically, it is

$$C_m = A C_d A^T,$$

(40)

where $C_m$ is the model covariance matrix, $C_d$ is the data covariance matrix and $A$ is the
generalized inverse for a particular inversion scheme. Given a data kernel $G$ that can be
written in an explicit linear form, i.e.

$$d = Gm,$$

(41)

where $d$ and $m$ are the data and the model parameter vectors respectively, the generalized
inverse is

$$A = [G^T G]^{-1} G^T$$

(42)

for the least square inversion scheme. However, because the location process is non linear,
the data kernel is not in an explicit linear form of the model parameter vector. We,
therefore, have to linearize it in order to use the above approach to compute the model
covariance matrix. From eq. (38), let $t_{pre}$ be the data vector $d$ and $x_0 = (x_0, y_0, z_0)^T$ be
the model parameter vector \( m \), we have

\[
d = G(m) + t_0.
\] (43)

Here we ignore the dependence of the station locations \( x_i \) on \( G \) since the station locations are fixed and not related to our discussion here.

At a location close to the optimized estimate of the event location \( m_0 \), the predicted arrival time is

\[
d \approx G(m_0 + \delta m) + t_0 + \delta t
\]

\[
\approx G(m_0) + t_0 + \frac{\partial G}{\partial x} \delta x + \frac{\partial G}{\partial y} \delta y + \frac{\partial G}{\partial z} \delta z + \delta t + \text{high order terms}
\]

\[
\approx G(m_0) + t_0 + \nabla G \cdot \delta m + \delta t
\]

\[
\approx G(m_0) + t_0 + G \delta m_y
\] (44)

where

\[
G = \begin{pmatrix}
\frac{\partial G_1}{\partial x} & \frac{\partial G_1}{\partial y} & \frac{\partial G_1}{\partial z} & 1 \\
\frac{\partial G_2}{\partial x} & \frac{\partial G_2}{\partial y} & \frac{\partial G_2}{\partial z} & 1 \\
\vdots & \vdots & \vdots \\
\frac{\partial G_N}{\partial x} & \frac{\partial G_N}{\partial y} & \frac{\partial G_N}{\partial z} & 1
\end{pmatrix}
\] (45)

and

\[
\delta m_y = \begin{pmatrix}
\delta x, \ \delta y, \ \delta z, \ \delta t
\end{pmatrix}^T.
\] (46)

From eq.(44), we have

\[
\delta d = G \delta m_y,
\] (47)

where \( G \) is now in an explicit linear form such that eq. (40) and eq. (42) can be applied to find the model covariance matrix given the data covariance matrix.
To estimate the data covariance matrix, we assume all data uncertainties are independent of each other so that the off diagonal elements in the data covariance matrix are all zeros. We also assume all data have the same variance. Hence, the data covariance matrix becomes

$$C_d = \sigma_d^2 (I),$$  \hspace{1cm} (48)

where

$$\sigma_d^2 = \frac{1}{N - 1} \sum_{i=1}^{N} (t_{obs_i} - t_{pre_i})^2 = \frac{N}{N - 1} Q$$  \hspace{1cm} (49)

and $I$ is a $N$ by $N$ identity matrix.

We now have a catalog of synthetic data with all estimated locations and uncertainties of the events. Before applying MLECL to the synthetic data, we test the validity of the location process by generating a few data sets with different random measurement errors. Figure 6b, 6c, 6d and 6e show the distributions of the located events for data sets with random measurement errors of 0 s, 0.1 s, 0.2 s and 0.3 s respectively. They are compared to the distribution of events with true locations (Figure 6a). The distributions of the located events recover the two faults in general, except for the case with random error of 0.3 s, where the diffuseness of events is too large to make the faults identifiable. The diffuseness of the events increases with the random measurement error as we expect.

### 3.6 Results

We have generated three sets of synthetic data with different random measurement errors (Table 1) and relocated the events in the catalogs using MLECL. When we applied MLECL to these synthetic data sets, we used the approximated density distribution for the a priori density distribution, as stated in eq. (7).

In order to quantify the amount of attraction due to MLECL, a histogram showing the distribution of the $x$ coordinates of the event locations is plotted for every iteration.
We then fit these histograms with two Gaussian functions,

\[ f(x) = a_1 \exp \left[ -\frac{(x - b_1)^2}{2c_1^2} \right] + a_2 \exp \left[ -\frac{(x - b_2)^2}{2c_2^2} \right], \quad (50) \]

where \( a_1, a_2, b_1, b_2, c_1, \) and \( c_2 \) are the constants found by curve fitting. We are justified to approximate the data using two Gaussian functions because the initial distributions of events in the \( x \) direction are Gaussian distributions around the faults. The amount of the attraction is then reflected on the values of the standard deviations, \( c_1 \) and \( c_2 \), of the Gaussian functions.

Figure 7, 8 and 9 show the evolutions of the event locations in the catalog as the iteration proceeds for synthetic data set A, B and C respectively. In these synthetic data sets, it is not so clear that the patterns of seismicity evolve towards simpler distributions. The standard deviations of the event locations in the \( x \) direction do not decrease in general as the event locations evolve (Table 2). The standard deviations increase for some cases. Therefore, it is not likely that the events are attracted to the central lines of the faults. The error bound of each event reduces gradually as the iteration proceeds. The events are attracted towards the locations where the densities are the highest among the neighbourhood. The distribution of relocation displacements also gradually approaches a \( \chi^2 \) distribution. As the iteration proceeds, many small local maxima of the probability density develop, which make the density become quite rough.

Comparing results from different data sets, one can notice that the initial location uncertainties for events in set B and C are in general larger than those in set A. The initial locations are also more diffuse. Some events from two different faults are overlapping each other at the middle of the faults. However, the initial probability densities for set B and C are smoother than that for set A. As the event locations evolve, the events in set B and C are more diffuse than those in set A. The events in set B and C are not attracted to the central lines of the two faults as well as those in set A. Nevertheless, we can still observe that the initially overlapping events are separated and the structure can be identified clearly after the fifth iteration. However, as the iteration proceeds further, the events are relocated in a way that the structure becomes unclear again.
4 Example: aftershock sequence in SW Iceland

We also apply the MLECL relocation method to a sequence of aftershocks after a pair of $M \sim 6$ earthquakes in southwest Iceland in May 2008. Within 12 hours of the events a dense network of seismographs was deployed in the area. This network included 11 seismographs in addition to the four nearest seismographs in the permanent network of the Icelandic Meteorological Office (IMO). The network was operated for approximately one month. More than 19000 events were registered and located with an automated procedure (Brandsdóttir et al., 2010). 7850 of them have estimated location uncertainties of less than 1 km in lateral directions and 2 km in vertical direction.

The locations of these aftershock events are mainly concentrated along three lineations. The north-south striking lineation furthest to the east coincides the initial rupture on 29th May 2008. The second lineation slightly to the west represents the rupture that occurred a few seconds after the first event. The third lineation, which strikes approximately N80E, corresponds to a fault zone that is a part of the complex plate boundary deformation in the area. The events on this lineation appear to be triggered by the two events on 29th May 2008.

4.1 Results

In this catalog of aftershocks, we also applied the approximation in eq. (7). Figure 10 shows the evolution of the catalog as iterations proceed. Compare to the synthetic result in Figure 7, it is more obvious that the pattern of seismicity in the aftershock sequence evolves into a simpler distribution. The events are clearly attracted towards the three lineations, except for the third lineation, which strikes approximately N80E, some N30E and north striking lineations appear at the ends. The initial distribution of the aftershocks around the two main faults is 2 to 5 km wide. After the relocation process, the distribution is reduced to only 0.5 to 1 km wide. The error bounds of the events decrease. The distribution of relocation displacements also gradually approaches a $\chi^2_3$ distribution.
5 Discussion

For both examples, one can observe that the events are attracted towards the regions where the probability densities are high as expected. The error bounds of the events are reduced in both examples. The outlying events, which are located far away from the central fault lines, are not attracted to the centre of the catalog. These events form small clusters at the outlying region. As we proposed in the methodology section, two clusters with the similar values of probability density should not be attracted to each other. This observation confirms our proposition, and shows the difference between MLECL and the collapsing method proposed by Jones and Stewart (1997).

The distributions of the relocation displacements approach a $\chi^2$ distribution for both examples. This observation implies that MLECL is actually converging. But, if we compare the distributions from the two examples for the same iteration, we notice that the distribution from the aftershock sequence in Iceland is not as close to the $\chi^2$ distribution as the one from the synthetic data. This may be because the initial distribution of the events in the aftershock catalog from Iceland does not follow a Gaussian distribution, but the distribution of synthetic data does since we define it to be Gaussian.

For the synthetic data set A with a random measurement error of 0.1 s, we can see there is a small improvement in terms of the simplicity of the seismicity in the catalog. The initial distribution of the events is quite broad in the $x$ direction. After a few iterations, the distribution becomes narrower because of the decrease in the error bounds and the relocation of events towards the centres of the faults. At the same time, many local maxima of the probability density appear in the region. These local maxima cause neighbouring events to move towards them, instead of the central lines of the faults. As a result, the events are not attracted further to the fault lines after further iterations.

For the synthetic data set B with a random measurement error of 0.2 s, a similar phenomenon is observed. Initially, the events are very diffuse such that some events from different faults are overlapping each other. After a few iterations, these events are separated so that one can identify the faults clearly. However, further iteration focuses the relocations at local maxima and the separation of the faults is not clarified again.
because of the appearance of the local maxima of the probability density. The faults are barely distinguishable after the 15th iteration.

For the synthetic data set C with a random measurement error of 0.3 s, the distribution is more diffuse than that in set B. The two faults are not distinguishable. One may note that there is a significant reduction in the error bounds of the events. However, this should not be considered as an improvement since the reduction in error bounds is a consequence of the concentration of events. One should consider it to be artificial.

Consider the standard deviations of event locations in the $x$ direction, the values do not decrease significantly for all synthetic data sets, implying there is no significant attraction along the fault lines. For some cases, the value increases, probably due to the irregular shapes of the distributions.

Compared the results from the synthetic data sets to that from the aftershock sequence in Iceland, one can observe a different behaviour. The events are attracted to the three main lineations, except for small-scale lineations appearing along the third earthquake zone. Small local maxima do not appear and the probability densities are very smooth even after many iterations.

The reason for the difference in the synthetic data sets and the real data set is probably that the numerous local maxima of the probability density appear in the synthetic data as the catalog evolves. Since our strategy does not have any pre defined geometry for the attractor, and the way that the events are relocated is totally governed by the a priori distribution and the distribution from the data misfit, each local maximum in the distribution acts as an individual attractor for the neighbouring events. For the case of synthetic data, these local maxima are quite randomly distributed, and are not aligned in the direction of the fault lines. When the events are attracted towards these maxima, they are not aligned in that direction too. As a result, the events are scattered over the region. However, for the case of the aftershock sequence in Iceland, the maxima are mainly aligned along the three lineations. This results in a simpler distribution as the iteration proceeds.

The reasons for the appearance of the local maxima are still not clearly known. Here,
we offer a few possible explanations to the phenomenon. First, the distribution of events in the real data may not follow a Gaussian distribution. Instead, it may follow a more focused distribution, e.g. an exponential distribution. Using a Gaussian distribution to simulate real data in the synthetic data may produce a different behaviour. Second, the distribution of events in the real data is anisotropically aligned with the fault lines on all scales. However, in our synthetic data sets, we simply use three distributions, a Gaussian distribution for the $x$ direction and two uniform distributions for the $y$ and $z$ directions, for the three coordinates of the events. This configuration only affects the distribution of events on large scale. On a small scale, the events are basically isotropic. This oversimplification of our synthetic data may contribute to the phenomenon. Third, the size of synthetic catalog may be too small to cover the whole fault planes uniformly. In other words, the events on the fault planes may not be uniformly distributed due to the small population of events. Besides, since the location uncertainties of events scale with the size of the population, small population size may result in small uncertainties of events, which make the a priori distribution less smooth in general.

The above explanations of the appearance of the local maxima may provide some directions for further research. First, one may consider using an exponential distribution or a power law distribution to simulate the distribution of events in the real data. Second, a probability distribution, which contains small scale anisotropy, may be used to describe the events. Third, one may consider increasing the catalog size to obtain more uniformly distributed events.

One assumption is made when we derive the methodology for MLECL. Recall eq. (4), in the derivation of the posterior distribution, we simply state that the posterior distribution is the product of the distribution from data misfit $P_{mj}(\mathbf{m})$ and the a priori distribution $P_{aj}(\mathbf{m})$. This statement assumes that the two probabilities $P_{mj}(\mathbf{m})$ and $P_{aj}(\mathbf{m})$ are independent. However, a careful inspection reveals that this assumption can only be fulfilled for the first iteration. For the first iteration, because of the exclusion of the $j$th event in the a priori distribution, the a priori distribution is obviously independent of the distribution from the data misfit. But, when the iteration continues, the a priori
distribution for the subsequent iteration is no longer independent of the distribution from the data misfit since we define the a priori distribution to be the normalized sum of the probability distributions of all events in the previous iteration, and this sum contains a contribution from the $j$th event. The dependence for the first few iterations would be insignificant, but when the iteration proceeds, the dependence will continue to build up and finally to an extent that it can no longer be considered to be insignificant. Therefore, one has to develop a strategy for monitoring this dependence. If the dependence becomes too large, one may have to terminate the process.

6 Conclusions

We present a strategy, called MLECL, for relocating events in an earthquake catalog. In MLECL, we first define a probability density function for the misfit of the model parameters, i.e. the coordinates of locations, of every event in the catalog. This density is then multiplied by an a priori distribution to give a likelihood function for relocation. The a priori distribution is the normalized sum of the probability densities of all events in the catalog except the event that is being relocated. The events are relocated using the method of maximum likelihood. After relocating all events, the a priori distribution is updated and the process is repeated. The iteration continues until a $\chi^2$ distribution is approached for the distribution of the relocation displacements or until the convergence rate is below some limit.

We used two examples to demonstrate the application of MLECL. The first example is from three synthetic data sets and the second example is from an aftershock sequence in Iceland after a pair of earthquakes on 29th May 2008. Both examples show some features that are expected from the strategy, e.g. the attraction of events towards the region where the probability density is the highest, and the reduction of error bounds for each event. The examples also confirm that two clusters of events are not drawn to each other.

The events from the synthetic data sets, however, are not relocated as well as those from the aftershock sequence. The events in the synthetic data sets are not attracted to the fault lines as the events in the aftershock sequence. Many small local maxima appear
in the probability density of the synthetic data sets. These maxima attract the events to them and result in a scatter of events in the catalog. Three possible explanations are proposed to tackle this question. First, the events in the real data may not be Gaussian distributed as we assume in the synthetic data. Second, the distributions of events in the synthetic data do not account for the anisotropy on small scale. Third, the size of the synthetic data may be too small. These explanations give directions for further research.

7 Acknowledgements

I would like to express my wholehearted thanks to my supervisors, Ólafur Gudmundsson and Ari Tryggvason, for providing me with such a precious opportunity to learn so much and for their guidance, patience, and support throughout my master project. Ólafur taught me the fundamental tools that are commonly used in seismology. His patience allowed me to strengthen my knowledge in seismology. Ari gave me the insight into the topic of parallel computing, which provided a direction for optimizing the MLECL program.

Special thanks go to Arnaud Pharasyn who provided computer support for me and Roland Roberts who handled a lot of administrative work for me at the beginning of my project.

I would also like to offer my heartfelt thanks to my parents and my fiancée, Kwan Yin Lau, who have always been there for me, for their unlimited love and care and for always giving me the power to overcome every difficulty I have encountered or will encounter in the future.
References


Figure 1: Distribution of earthquake events around two faults, fault 1 (left) and fault 2 (right). The events are distributed around the centres of the faults in the $x$ direction, described by two Gaussian distributions with specific standard deviations.

Figure 2: The locations of the seismic stations in map view.
Figure 3: The one-dimensional velocity model used for computing traveltime. The model consists of different layers in depth. For each layer, the velocity is a linear function of depth.

Figure 4: Schematic graph for the passage of the ray at a particular layer. The ray enters the layer at point A, with coordinates $x_1$ and $z_1$, and exits at point B, with coordinates $x_2$ and $z_2$. The locus of the ray is the arc of a circle with centre at point C and radius R. $\alpha$ is the angle subtended by the arc $l$. $\delta \alpha$ is the angle subtended by a small ray segment $\delta l$ and $\delta z$ is the vertical length of the ray segment $\delta l$. 
Figure 5: Schematic graph for a small ray segment $\delta l$ at a particular layer. Point D, E and F represent the start, the midpoint and the end of the ray segment. $\alpha_1$ is measured from the start of the ray segment to the top of the layer. $\delta x$ and $\delta z$ are the horizontal and the vertical length of the half of the ray segment.
Figure 6: The comparison of the distributions of located events and true events. The upper graph shows the histogram of the distribution of event locations in the $x$ direction. The lower graph shows the distribution of the events in map view. (a) the events with true locations in the catalog. The located events from data sets with (b) no random measurement error and random measurement errors of (c) 0.1 s, (d) 0.2 s and (e) 0.3 s.
Figure 7: The evolution of the catalog locations for the synthetic data set A, with random measurement error of 0.1 s. The left frame shows the epicentres (red dots) of more than 8800 events with original error estimates less than 2 km in the horizontal and 4 km in the vertical. The epicentres are drawn on top of one standard deviation contours of the posterior distribution for every event. The small graph on the top of the left frame represents the histogram of the distribution of event locations in the $x$ direction. The right frame shows a contour graph of the a priori distribution. The contour levels are at 0.999, 0.99, 0.95, 0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2 and 0.1 of the maximum value. The distribution of the relocation displacements is shown at the top of the right frame. The red solid line represents a $\chi^2_3$ distribution. The number of iteration is given in each figure.
Figure 8: The evolution of the catalog locations for the synthetic data set B, with random measurement error of 0.2 s. The left frame shows the epicentres (red dots) of more than 8800 events with original error estimates less than 2 km in the horizontal and 4 km in the vertical. The epicentres are drawn on top of one standard deviation contours of the posterior distribution for every event. The small graph on the top of the left frame represents the histogram of the distribution of event locations in the $x$ direction. The right frame shows a contour graph of the a priori distribution. The contour levels are at 0.999, 0.99, 0.95, 0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2 and 0.1 of the maximum value. The distribution of the relocation displacements is shown at the top of the right frame. The red solid line represents a $\chi^2_3$ distribution. The number of iteration is given in each figure.
Figure 9: The evolution of the catalog locations for the synthetic data set C, with random measurement error of 0.3 s. The left frame shows the epicentres (red dots) of more than 8800 events with original error estimates less than 2 km in the horizontal and 4 km in the vertical. The epicentres are drawn on top of one standard deviation contours of the posterior distribution for every event. The small graph on the top of the left frame represents the histogram of the distribution of event locations in the $x$ direction. The right frame shows a contour graph of the a priori distribution. The contour levels are at 0.999, 0.99, 0.95, 0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2 and 0.1 of the maximum value. The distribution of the relocation displacements is shown at the top of the right frame. The red solid line represents a $\chi^2_3$ distribution. The number of iteration is given in each figure.
Figure 10: The evolution of the catalog locations for the aftershock sequence after a pair of events on 29th May 2008 in southwest Iceland. The upper frame shows the epicentres (red dots) of more than 18000 events with original error estimates less than 2 km in the horizontal and 4 km in the vertical. The epicentres are drawn on top of one standard deviation contours of the posterior distribution for every event. The lower frame shows a contour graph of the a priori distribution. The contour levels are at 0.999, 0.99, 0.95, 0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2 and 0.1 of the maximum value. The distribution of the relocation displacements is shown at the right of the lower frame. The red solid line represents a $\chi^2_3$ distribution. The number of iteration is given in each figure.
<table>
<thead>
<tr>
<th>Set</th>
<th>The standard deviation of the Gaussian distribution describing the true event locations (km)</th>
<th>Correlation length (km)</th>
<th>Random measurement error (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.3</td>
<td>5</td>
<td>0.1</td>
</tr>
<tr>
<td>B</td>
<td>0.3</td>
<td>5</td>
<td>0.2</td>
</tr>
<tr>
<td>C</td>
<td>0.3</td>
<td>5</td>
<td>0.3</td>
</tr>
</tbody>
</table>

Table 1: Parameters used in different synthetic data sets
Table 2: The standard deviations of the event locations in the $x$ direction as the iteration proceeds. (a), (b) and (c) are for data set A, B and C with random measurement errors of 0.1 s, 0.2 s and 0.3 s respectively.
<table>
<thead>
<tr>
<th>Nr</th>
<th>Titel</th>
<th>Författare</th>
<th>Datum</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Geomorphological mapping and hazard assessment of alpine areas in Vorarlberg, Austria</td>
<td>Marcus Gustavsson</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Verification of the Turbulence Index used at SMHI</td>
<td>Stefan Bergman</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Forecasting the next day’s maximum and minimum temperature in Vancouver, Canada by using artificial neural network models</td>
<td>Magnus Nilsson</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>The tectonic history of the Skyttorp-Vattholma fault zone, south-central Sweden</td>
<td>Anna Victoria Engström</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Investigation on Surface energy fluxes and their relationship to synoptic weather patterns on Storglaciären, northern Sweden</td>
<td>Yvonne Kramer</td>
<td></td>
</tr>
<tr>
<td>240</td>
<td>Mobility of multi-walled carbon nanotubes in unsaturated porous media</td>
<td>Abenezer Mekonen</td>
<td>August 2012</td>
</tr>
<tr>
<td>241</td>
<td>Re-processing of Shallow and Deep Crustal Reflection Seismic Data along BABEL Line7, Central Sweden</td>
<td>Hanieh Shahrokhi</td>
<td>August 2012</td>
</tr>
<tr>
<td>242</td>
<td>Usability of Standard Monitored Rainfall-Runoff Data in Panama, Juan Diaz River Basin</td>
<td>José Eduardo Reynolds Puga</td>
<td>August 2012</td>
</tr>
<tr>
<td>244</td>
<td>Regional Precipitation Study in Central America, Using the WRF Model</td>
<td>Tito Maldonado</td>
<td>September 2012</td>
</tr>
<tr>
<td>245</td>
<td>Reduktion av järn, mangan och CODMn i dricksvatten – Ett pilotförsök vid Högåsens vattenverk</td>
<td>Tommy Olausson</td>
<td>September 2012</td>
</tr>
<tr>
<td>246</td>
<td>Short-term Variations in Ice Dynamics During the Spring and Summer Period on Storglaciären, Kebnekaise, Sweden</td>
<td>Helena Psaros</td>
<td>September 2012</td>
</tr>
<tr>
<td>247</td>
<td>Reassessment of the ‘last’ Goniopholidid Denazinosuchus kirtlandicus Wiman, 1932 from the Late Cretaceous of New Mexico</td>
<td>Oskar Bremer</td>
<td>October 2012</td>
</tr>
<tr>
<td>248</td>
<td>Variability of Intermittent Headwater Streams in a Boreal Landscape Influence of different discharge conditions</td>
<td>Tum Nihm</td>
<td>October 2012</td>
</tr>
<tr>
<td>249</td>
<td>2D Reflection Seismic Imaging at a Quick-clay Landslide Site, in Southwest Sweden</td>
<td>Muhammad Umar Saleem</td>
<td>November 2012</td>
</tr>
</tbody>
</table>