FINAL TECHNICAL REPORT

"Grid-Search Location Methods for Ground-Truth Collection From Local and Regional Seismic Networks"

William Rodi, Principal Investigator
Earth Resources Laboratory
Department of Earth, Atmospheric, and Planetary Sciences
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

Craig A. Schultz (Principal Investigator), Gardar Johannesson. and Stephen C. Myers
Lawrence Livermore National Laboratory
Livermore, CA 94550

Submitted to
The U.S. Department of Energy
National Nuclear Security Administration
Office of Nonproliferation Research and Engineering
Office of Defense Nuclear Nonproliferation

Grant/Contract Nos. DE-FC03-01SF22397 (MIT) and W-7405-ENG-48 (LLNL)
Project Duration: September 1, 2001-August 31, 2004

May 13, 2005

Contents

1 Executive Summary **5**
- 1.1 Objectives . . . 5
- 1.2 Research Accomplished . . . 6
  - 1.2.1 Multiple-Station Kriging . . . 6
  - 1.2.2 Event Location Experiments at NTS . . . 7
  - 1.2.3 Mixture-of-Gaussians Error Model . . . 7
- 1.3 Conclusions and Recommendations . . . 8
- 1.4 Personnel Supported . . . 9

2 Multiple-Event Location With Multiple-Station Kriging: Theory **10**
- 2.1 Multiple-Event Location Problem: Calibration and Location . . . 10
- 2.2 Parameterization of Path Corrections . . . 10
- 2.3 Multiple-Station Kriging . . . 12
  - 2.3.1 Kriging for Time-Term Parameter Functions . . . 12
  - 2.3.2 Differential Operators for a Class of Correlation Functions . . . 13
  - 2.3.3 Examples of Numerical Correlation Functions . . . 14
  - 2.3.4 Kriging for Mislocation-Vector Parameter Functions . . . 15
- 2.4 Combining Multiple-Station Kriging With Event Location . . . 17
  - 2.4.1 GMEL . . . 17
  - 2.4.2 KMEL . . . 18

3 Calibration and Event Location Tests Using Pn Arrival Times From Nevada Test Site Explosions **20**
- 3.1 Introduction . . . 20
- 3.2 Travel-Time Corrections From Multiple-Station Kriging . . . 20
- 3.3 Experiments With GT0-Calibrated Single-Event Location . . . 21
- 3.4 Experiments With Multiple-Event Location (Joint Calibration/Location) . . . 25

4 Using a Mixture of Gaussians to Model Seismic Arrival Time Error Distributions: Estimation and the Impact on Seismic Event Inference **31**
- 4.1 Mixture of Gaussian Distributions . . . 31
  - 4.1.1 Maximum Likelihood Estimation: The EM Algorithm . . . 32
  - 4.1.2 The Bayesian Approach: Markov Chain Monte Carlo (MCMC) . . . 32
- 4.2 Example . . . 33
  - 4.2.1 The Setup . . . 33
  - 4.2.2 Mixture Estimation via EM and MCMC . . . 35

4.2.3 Bayesian Estimation of Seismic Event Location and Time of Origin 39

List of Figures

2.1 A 2-D correlation function generated with our ML kriging algorithm. The case shown is for order $\ell = 3$ and correlation length $\lambda = 500$ km. The result has the correct properties of a correlation function, with maximum value at the synthetic observation point (30° N, 72.5° E). 16

2.2 Profile (vs. latitude) through a 2-D correlation function (left) and a 3-D one (right), computed with our ML kriging approach. Both cases use order $\ell = 2$ and correlation length $\lambda = 500$ km. The longitude in each case, and depth in the 3-D case, are evaluated at the synthetic observation point: 72.5°E. $z = 0$. In 3-D, the analytic solution is an exponential function, while in 2-D it is a function which is less sharply peaked at the center. 16

2.3 Cross-sections of 2-D correlation functions computed with orders $\ell = 2$, 3 and 4 (left, center, right) and correlation length $\lambda = 500$ km. 17

2.4 Cross-sections of 2-D correlation functions computed with order $\ell = 2$ and correlation lengths $\lambda = 250$, 500 and 750 km (left, center, right). 17

3.1 Mislocation-vector parameter function (in km) for the region surrounding NTS: north, east and depth components (top left, top right, and bottom frames, respectively). This parameter function was derived by multiple-station kriging of Pn travel-time residuals, using 200 km for the correlation length and 5 km for the prior error on the mislocation components (case A in Table 3.1). Station locations (some labeled) are marked with circles. 22

3.2 Same as Figure 3.1, except the mislocation-vector function was obtained using 100 km for the correlation length and 10 km for the prior error on the mislocation components (case D). 23

3.3 Travel-time correction functions (in seconds) for stations ELK (left) and BMN (right), generated from the mislocation-vector parameter function shown in Figure 3.2. Station locations (some labeled) are marked with circles, and the NTS events show as a small cluster of small white dots near the center of the plot. . . 24

3.4 Travel-time correction functions for stations KNB (left) and LAC (right), generated from the mislocation-vector parameter function shown in Figure 3.2. 25

3.5 NTS event locations resulting from a single-event location algorithm applied with no station corrections. The blue circles are the seismic solutions and the red circles are the ground-truth (GT0) locations. A line connects the solution and GT0 location of each event. Only Pn arrival times were included in the data set. 26

3.6 NTS event locations resulting from a single-event location algorithm applied with station time-term corrections (top) and with travel-time correction functions derived from multiple-station kriging, case D (bottom). In each case, the corrections were derived from travel-time residuals from these same events relative to their GT0 locations. Plotting conventions are the same as in Figure 3.5. 27
3.7 NTS event locations resulting from multiple-event location performed with GMEL (top) and KMEL (bottom). GMEL solves for travel-time corrections that are simple station time terms. KMEL solves for travel-time correction functions that are parameterized in terms of a mislocation-vector function. The KMEL solution shown uses the case D kriging parameters (see Table 3.1). Plot conventions are the same as in Figure 3.6. 28

4.1 Location of stations (red circles) and events (green triangles). 34
4.2 The measurement error mixture density (left) and a scatter plot of observed arrival time versus the true arrival time. 35
4.3 Histogram of the generated measurement errors along with the true mixture density (black, solid line), a kernel density estimate (red, dashed line), and a fitted (ML) Gaussian density (green, dotted line). 36
4.4 Histogram of the generated measurement errors along with the true mixture density (black, solid line), a kernel density estimate (red, dashed line), and the MLE via EM of a Gaussian mixture density with 3, 4, and 5 components. 37
4.5 Histogram of the generated measurement errors along with the true mixture density (black, solid line), a kernel density estimate (red, dashed line), and a Bayesian estimate (via MCMC) of a Gaussian mixture density with 3, 4, and 5 components. 38
4.6 The true mixture distribution of the arrival time measurement errors along with a simple (one-component) Gaussian approximation. 40
4.7 The generated MCMC samples for the three parameters $\mathbf{x} = (x, y)$ and t in the case of a *simple* (one-component) Gaussian measurement error distribution. The red horizontal lines show the true value of each parameter (equal to 1/2 in all cases). 41
4.8 The generated MCMC samples for the three parameters $\mathbf{x} = (x, y)$ and t in the case of a *4-component* mixture-of-Gaussians measurement error distribution. The red horizontal lines show the true value of each parameter (equal to 1/2 in all cases). 42
4.9 The MCMC sampled event locations, $\{\mathbf{x}^{(i)} : i = 2,001, \ldots, 10,000\}$, when assuming a Gaussian measurement error distribution (left) and when assuming a 4-component mixture measurement error distribution (right). The true location of the event is shown at $(1/2, 1/2)$. 43
4.10 The MCMC sampled event origin times, $\{t^{(i)} : i = 2,001, \ldots, 10,000\}$, when assuming a Gaussian measurement error distribution (left) and when assuming a 4-component mixture measurement error distribution (right). The true origin of the event is shown at $t = 1/2$. 44

Chapter 1

Executive Summary

1.1 Objectives

The overall objective of this project was to develop improved seismic event location techniques that can be used to generate more and better quality reference events, using arrival time data from local and regional seismic networks. Toward this end, we addressed some of the limitations of the multiple-event location methods that are currently used to analyze such data. These limitations pertain to the simplifying assumptions made about the observational errors in seismic arrival times ("pick" errors) and about the errors in the travel-time forward model used to analyze the data ("model" errors).

Multiple-event location methods, including the original master-event technique (Evernden, 1969) and its more general descendants such as hypocentroidal decomposition (HDC: Jordan and Sverdrup, 1981; Engdahl and Bergman, 2001), solve a version of what we will call here the *joint calibration/location* inverse problem. The general process of joint calibration and event location combines information from multiple events and stations in such a way that—from the point of view of ground-truth collection—location information from well-recorded events, or events whose locations are well-constrained from other information, "feeds" through the network of seismic stations, via estimated path travel-time corrections, to constrain the locations of other events. The travel-time corrections estimated in this process are intended to account for the effects of velocity anomalies in the Earth that are not represented in the travel-time forward model (travel-time tables).

However, the parameterization of travel-time corrections used by multiple-event location methods has typically been very restrictive. The most commonly applied methods in nuclear monitoring applications solve what this report will call the "basic" multiple-event location problem, in which corrections to travel-times are represented as a simple time term for each station/phase combination in the data set, independent of the event location. This assumption, in practice, restricts the application of these methods to clusters of events that are small in spatial dimension; the cluster aperture must be much less than the epicentral distances to the stations. The methods so restricted include HDC, JHD (joint hypocentral decomposition: Dewey, 1971) and PMEL (progressive multiple-event location: Pavlis and Booker, 1983)). The double-differencing method of multiple-event location (Waldhauser and Ellsworth, 2000) relaxes the small-cluster requirement but does so with the expense of eliminating an explicit representation of the travel-time corrections. One of the objectives of this project was to relax the requirement within the framework of joint calibration/location inversion.

Another shortcoming of the basic multiple-event location methods is of the opposite nature. The time-term corrections for different stations are allowed to vary independently, even when the stations are spatially close. This assumption does not restrict the allowed event-station geometry but it precludes the use of important physical information concerning the correlation between travel-times over similar travel paths.

Almost all location methods, single-event and multiple-event, assume that observational errors follow a simple Gaussian distribution. However, measurements of the onset time of seismic arrivals are subject to a number of different error processes, related to such difficulties as identifying the onset of low signal-to-noise signals and misidentifying secondary phases as the first arrival (Jeffreys,1932; Buland, 1986; Engdahl et al., 1998). As a result, residual distributions are often asymmetric and heavy-tailed, and may even be multi-modal. Only limited efforts have been made to incorporate distributions other than Gaussian into event location procedures (e.g. Billings et al., 1994; Dreger et al., 1998; Rodi and Toksöz, 2001). Another objective of this project was to explore an alternative class of error distributions, first proposed by Jeffreys (1932), in which the error distribution is represented as a "mixture", or weighted sum, of Gaussian distributions having different means and variances. Such distributions can describe the complexities affecting arrival picks from low magnitude events and, thus, their use in event location can potentially expand the event base which can be incorporated into multiple-event analyses for ground-truth collection.

1.2 Research Accomplished

1.2.1 Multiple-Station Kriging

A significant accomplishment of this project was the mathematical formulation of the multiple-event location problem in a more general setting than the basic problem, allowing for more general parameterizations of travel-time corrections. The formulation identifies the problem of multiple-event location with a joint inverse problem that combines travel-time calibration and event location, each being a special case of the joint problem. Our formulation of the problem is described in Chapter 2 of this report.

To specifically address our objective of developing a multiple-event location method with fewer restrictions on event-station geometry, Chapter 2 proceeds with a specific new parameterization of travel-tine corrections. The key concept involved is that of a correction *function* which has an explicit dependence on the location of events *and* stations. This concept had already been established for the event location by Schultz et al. (1998) in the context of travel-time calibration, which they performed by applying the method of kriging to fit station-specific functions, or *surfaces,* to the travel-time residuals observed at each station of a seismic network. We generalized this approach to allow for an explicit dependence on the station location by invoking the concept of a generic parameter function which generates a travel-time correction for an arbitrary event-station path. Accordingly, we recast the kriging method as a method of maximum-likelihood inversion (which others have done in different contexts), resulting in a method we call *multiple-station kriging.* This method accomplishes travel-time calibration with allowance for both event-side and station-side correlation between corrections, based on event or station separation. Chapter 2 describes the theory and numerical implementation of our multiple-station kriging method, while Chapter 3 shows examples of the method applied to Pn travel-time residuals observed from explosions at the Nevada Test Site (NTS).

1.2.2 Event Location Experiments at NTS

In a different project overlapping this one, M.I.T. developed a computer code for multiple-event location named GMEL (for "grid-search multiple-event location"). GMEL solves the basic multiple-event location problem, finding maximum-likelihood estimates for the hypocenters and origin times of multiple events, and for time-term corrections for the station/phase combinations observed from the events. The event location task of GMEL is performed with a grid-search technique.

In this project, we developed a prototype extension of GMEL, called KMEL, that replaces time-term corrections with a generic parameter function from which path-dependent travel-time corrections can be generated, as described above. KMEL thus solves a more general form of the joint calibration/location problem than the basic multiple-event location problem solved by GMEL. We note that KMEL was not developed as a stand-alone computer program, but was emulated by iterating back and forth between grid-search single-event location and our new multiple-station kriging algorithm.

Chapter 3 reports the results of event location experiments that compare the two approaches to multiple-event location (GMEL and KMEL). The experiments were performed with a data set of high quality Pn picks from 71 NTS explosions, assembled by LLNL (Walter et al., 2003). Precise GT0 locations and origin times are available for these events, so the outcome of the location experiments can be evaluated unambiguously in terms of mislocation distances.

We performed two types of experiments. The first performed calibration and event location as separate tasks. The calibration task developed travel-time corrections from the observed travel-time residuals (relative to IASP91) from the 71 explosions, using the GT0 locations of the explosions to compute the residuals. Two types of corrections were derived: simple time-term corrections (obtained by averaging residuals at stations), and correction functions obtained by our multiple-station kriging method. The event location task then consisted of performing single-event location on each event, applying the pre-derived travel-time corrections in the location process.

The second type of location experiments involved full multiple-event location with the two types of corrections, as performed by GMEL and KMEL. Thus, in each case, the corrections and event locations were estimated in a joint inversion process. No ground-truth prior information was used to constrain the solutions.

The results of both sets of experiments (single-event and multiple-event) indicated that correction functions, as generated from a generic parameter function, lead to moderately smaller event mislocations than simple time-terms. In particular, significantly more events were mislocated by less than one kilometer (e.g., 13 more in the single-event experiment, 7 more in the multiple-event experiment). Given the good data coverage, the mislocations themselves were small in all cases (mostly less than 3 km), so the small improvement was meaningful. Our conclusion is that the improvement is due to the fact that the kriged correction functions include information about the correlation between the travel-time corrections at stations that are near one another. The station time-term corrections do not include this information.

1.2.3 Mixture-of-Gaussians Error Model

Chapter 4 reports our efforts in developing a new, more general, stochastic model for arrival time pick errors. The model replaces the customary Gaussian probability distribution with a mixture-of–Gaussians distribution. A mixture of Gaussians can more realistically describe the

complexities of the signal onset picking process (e.g. missing the first cycle of a low signal-to-noise arrival) by allowing for such effects as asymmetry and multi-modality in the probability density function.

A mixture-of-Gaussians distribution can contain many more than the two free parameters (mean and variance) of a simple Gaussian distribution. For example, a four-competent mixture has eleven free parameters: four means, four variances, and three independent weights. Chapter 4 formulates the estimation problem for these parameters, which must be fit to data sets of observed travel-time residuals. Two approaches, and a numerical algorithm for each, are presented: a maximum-likelihood approach, solved with the Expectation-Maximization (EM) algorithm, and a Bayesian approach, solved with the Markov Chain Monte Carlo (MCMC) algorithm. The methods are applied to a synthetic data set to demonstrate and compare their effectiveness.

Chapter 4 also reports the results of a numerical experiment with event location to test the sensitivity of location results to the error model that is assumed in the location procedure. The experiment used synthetic arrival time data based on an assumed "true" event location and contaminated with pseudo-random noise drawn from a four-component Gaussian mixture. A Metropolis sampling technique was applied to the synthetic data to determine the Bayesian posterior distribution of the event location. When the sampling technique was applied with the correct four-component mixture distribution, the inferred posterior included the true event location near its peak. When a simple Gaussian distribution (preserving mean and variance) was used in place of the true distribution, however, the true location ended up in the tails of the infrared posterior distribution. This experiment demonstrates the importance of using the correct error model in an event location procedure.

1.3 Conclusions and Recommendations

This project demonstrated the feasibility and promise of some new techniques for improving seismic event locations derived from regional and local networks. We developed a new approach to empirical travel-time calibration that simultaneously fits data from multiple stations and events, using a generalization of the kriging method, and predicts travel-time corrections for arbitrary event-station paths. We combined this calibration approach with grid-search event location to produce a prototype new multiple-event location method that allows the use of spatially well-distributed events and takes into account correlations between the travel-time corrections from proximate event-station paths. Preliminary tests with a high quality data set from NTS explosions indicated that our new calibration/location method offers improvement over the conventional multiple-event location methods now in common use, and is applicable to more general event-station geometries than the conventional methods. However, our tests were limited and further research is needed to fully evaluate, and improve, our approach. Moreover, the full integration of the multiple-station kriging technique we developed into a multiple-event location algorithm, to create a stand-alone module and optimized algorithm, remains to be done.

Our project also demonstrated the importance of using a realistic error model in an event location procedure, and we took the initial steps in developing the inference algorithms needed to estimate such models. Our main development was a Markov Chain Monte Carlo algorithm for fitting a mixture-of-Gaussian distribution to observed travel-time residuals. Mixture-of-Gaussian distributions possess the properties necessary to characterize the arrival time error

processes that can occur when picking low signal-to-noise arrivals, as well as the complex signals that result from propagation along complex paths. To greatly facilitate further research into the use of these error models, we recommend that the next step in their development be to incorporate them into the grid-search location algorithms we have developed.

1.4 Personnel Supported

The following MIT personnel were supported by funds derived from the project:

1. William L. Rodi (Principal Research Scientist and PI).
2. Frank Dale Morgan (Professor of Geophysics).
3. Dr. Xu Li (Postdoctoral Associate).
4. Mr. Youshun Sun (Graduate Research Assistant)

Chapter 2

Multiple-Event Location With Multiple-Station Kriging: Theory

2.1 Multiple-Event Location Problem: Calibration and Location

We consider the multiple-event location problem involving m events and n station/phase combinations for which arrival times have been observed from one or more of the events. This problem can be stated as

$$d_{ij} = T_i(\mathbf{x}_j) + t_j + c_{ij} + e_{ij} \tag{2.1}$$

where d_{ij} is the observed arrival time for the (i,j)th path; $\mathbf{x}_j$ and t_j are origin parameters (hypocenter and time, respectively) of the jth event; T_i is a model-based travel-time function for the ith station/phase pair; c_{ij} is a correction to this function; and e_{ij} is an observational (picking) error in d_{ij}. It is understood that equation (2.1) holds only for the (i,j) pairs that have been observed, among the mn possibilities.

The unknowns in this inverse problem are the event parameters, $(\mathbf{x}_j, t_j)$, and the path corrections, c_{ij}. In addition to the arrival time data, other information may be available in the form of prior constraints on some or all of the unknown parameters. A special case of this inverse problem is the travel-time *calibration* problem, which occurs when the event locations are known (i.e., are GT0 events) and the corrections are to be estimated. Another special case is the *single-event location* problem, in which the c_{ij} are assumed known and the event locations are to be estimated. In this case, the problem decouples between events (hence the descriptor *singe-event*). In general, the problem can be called the *joint calibration/location* inversion problem. The general problem may also be called the *multiple-event location* problem since the event location problem couples across events when the c_{ij} are not known.

2.2 Parameterization of Path Corrections

Multiple-event location methods (e.g., Dewey, 1971; Jordan and Sverdrup, 1981; Pavlis and Booker, 1983) have generally assumed that the path corrections are independent of the event location and, hence, the event index, i. In this case we can write

$$c_{ij} = a_i \tag{2.2}$$

where a_i is a time term associated with the ith station/phase. We will refer to the inverse problem resulting from this assumption as the *basic* multiple-event location problem. The goal is then to solve simultaneously for the m event locations and the n path corrections. The assumption of event-independent path corrections is appropriate when the event cluster is sufficiently small in spatial dimension that the raypaths from individual events to a particular station, for a particular phase, are affected by the same velocity anomalies in the Earth. This can impose a severe restriction on the event-station geometries that can be used in the multiple-event analysis, e.g. disallowing stations too close to the event cluster.

The more general parameterization of path corrections we investigated in this project extends the approach of Schultz et al. (1998). They allowed an explicit dependence of c_{ij} on the location of an event as

$$c_{ij} = C_i(\mathbf{x}_j) \tag{2.3}$$

where $C_i(\mathbf{x})$ is a correction *function* specific to the ith station/phase. Schultz et al. (1998) treated the correction functions for different i as being independent of one another. Our extension links them by allowing an explicit dependence on the station location as well as event location, as we now describe.

Let ϕ indicate a distinct seismic phase, e.g. ϕ = P, Pn, S, etc. We parameterize travel-time corrections for each phase in terms of a *generic* function, $C_\phi(\mathbf{x}, \mathbf{y})$, where $\mathbf{x}$ denotes the event location and $\mathbf{y}$ denotes the station location. We thus have

$$c_{ij} = C_{\phi_i}(\mathbf{x}_j, \mathbf{y}_i) \tag{2.4}$$

where ϕ_i is the phase, and $\mathbf{y}_i$ the station location, associated with the ith station/phase pair. The calibration problem then becomes to solve for the functions $C_\phi(\mathbf{x}, \mathbf{y})$ corresponding to the various seismic phases in the data set. The multiple-event location problem entails doing this in conjunction with estimating the m event locations.

Inferring one or more generic functions, C_ϕ, from a finite number of data poses a highly ill-posed inverse problem. It becomes necessary to either restrict the functional form of the C_ϕ, impose smoothness constraints on their spatial dependence, link them between phases, or invoke all three of these mitigation tactics. To restrict the functional form of C_ϕ, we derive it from one or more generic *parameter* functions, which reduce the separate dependence on event and station location to a single spatial dependence. In this project, we have considered two such parameter functions. The first is a a *time-term* function for each phase, a_ϕ, which generates the travel-time correction for a path as

$$C_\phi(\mathbf{x}, \mathbf{y}) = a_\phi(\mathbf{x}) + a_\phi(\mathbf{y}). \tag{2.5}$$

It is clear that the time-term function generates reciprocal travel-time corrections, since the event location $\mathbf{x}$ and station location $\mathbf{y}$ can be interchanged in equation (2.5). The second parameter function we have considered, which is employed in the examples shown later, is a *mislocation-vector* function, $\mathbf{a}(\mathbf{x})$, such that

$$C_\phi(\mathbf{x}, \mathbf{y}) = \mathbf{p}_\phi(\mathbf{x}, \mathbf{y}) \cdot \mathbf{a}(\mathbf{x}) + \mathbf{q}_\phi(\mathbf{x}, \mathbf{y}) \cdot \mathbf{a}(\mathbf{y}). \tag{2.6}$$

Here, $\mathbf{p}_\phi$ is the gradient of the model-based travel-time (which we can denote as $T_\phi(\mathbf{x}, \mathbf{y})$) with respect to the event location, while $\mathbf{q}_\phi$ is the gradient with respect to the station location:

$$\mathbf{p}_\phi = \nabla_1 T_\phi(\mathbf{x}, \mathbf{y}) \tag{2.7}$$

$$\mathbf{q}_\phi = \nabla_2 T_\phi(\mathbf{x}, \mathbf{y}). \tag{2.8}$$

(∇_k means gradient with respect the kth argument of a function.) Each is a slowness vector pointed along the raypath at one of its endpoints, and each depends on both $\mathbf{x}$ and $\mathbf{y}$ since the raypath depends on the location of both its endpoints. The interpretation of $\mathbf{a}$ as a mislocation vector follows from the fact that equation (2.6) is the first order change in travel-time that results from perturbing the event location by $\mathbf{a}(\mathbf{x})$ and the station location by $\mathbf{a}(\mathbf{y})$.

Since T_ϕ is a reciprocal function of its arguments, we have $\mathbf{p}_\phi(\mathbf{x}, \mathbf{y}) = \mathbf{q}_\phi(\mathbf{y}, \mathbf{x})$, which shows that the mislocation-vector parametrization also generates travel-time corrections that are source-receiver reciprocal.

We point out that station-specific correction functions of the type considered by Schultz et al. (1998) are easily defined in terms of generic parameter functions. In the case of the mislocation-vector parameterization:

$$C_i(\mathbf{x}) = \mathbf{p}_i(\mathbf{x}) \cdot \mathbf{a}(\mathbf{x}) + \mathbf{q}_i(\mathbf{x}) \cdot \mathbf{a}(\mathbf{y}_i) \tag{2.9}$$

where we have used the contractions $\mathbf{p}_i(\mathbf{x}) \equiv \mathbf{p}_{\phi_i}(\mathbf{x}, \mathbf{y}_i)$, $\mathbf{q}_i(\mathbf{x}) \equiv \mathbf{q}_{\phi_i}(\mathbf{x}, \mathbf{y}_i)$, This points up an important aspect of generic parameterizations such as equations (2.5) and (2.6). They define a station-specific correction function for any station location, whether or not a station actually exists at that location or, if one does, it has observed travel-time residuals to constrain the function. This is accomplished by allowing correlations between the corrections among stations and between stations and events.

2.3 Multiple-Station Kriging

In their formulation of travel-time corrections in terms of station/phase-specific correction functions, $C_i(\mathbf{x})$, Schultz et al. (1998) considered only the calibration problem, not the joint calibration/location problem. They solved the problem with an inversion method know as *kriging*. Their approach determines the correction function for each station/phase pair independently of the other stations and phases, i.e. the problem decouples over i. In this section, we describe an extension of their method to the problem of estimating a generic parameter function, such as $a(\mathbf{x})$ and $\mathbf{a}(\mathbf{x})$ above, from which station/phase-specific corrections can be generated. This problem does not decouple over i and we refer to it as *multiple-station* kriging ("multiple-station/phase" kriging would be more accurate but too unwieldy).

2.3.1 Kriging for Time-Term Parameter Functions

We describe the method first for the case of time-term parameter functions, defined in equation (2.5). To simplify the discussion we consider only a single seismic phase and drop the subscript ϕ, implying there is a single unknown parameter function, $a(\mathbf{x})$, to infer.

Referring to equation (2.1), let us define the travel-time residuals over observed paths with respect to a given set of event locations and times as

$$r_{ij} = d_{ij} - T_i(\mathbf{x}_j) - t_j. \tag{2.10}$$

The inverse problem of estimating $a(\mathbf{x})$, with the event parameters fixed, can then be stated by rewriting (2.1) as

$$r_{ij} = a(\mathbf{x}_j) + a(\mathbf{y}_i) + e_{ij}. \tag{2.11}$$

This inverse problem is an interpolation problem that is somewhat more complicated than the interpolation problem solved by the standard kriging method (see, for example, Deutsch and Journel, 1998). Each observation, r_{ij}, samples the unknown function, $a(\mathbf{x})$, at two points instead of one. This notwithstanding, the problem can be solved with geo-statistical constraints imposed on $a(\mathbf{x})$, as are used in conventional kriging. We accomplish this here through a maximum-likelihood (ML) approach to the inverse problem, instead of the minimum-variance approach used in conventional kriging.

For Gaussian data errors, the ML approach takes the solution of (2.11) to be the function a that minimizes the objective function given by

$$\Psi = \sum_{ij} |r_{ij} - a(\mathbf{x}_j) - a(\mathbf{y}_i)|^2 / \sigma_{ij}^2 + \int a(\mathbf{x}) \, [Da](\mathbf{x}) \, d\mathbf{x} \tag{2.12}$$

where σ_{ij} is the assigned standard deviation of e_{ij} and D is a specified differential operator. Numerically, we have implemented this approach with the function a approximated by a sampled version on a latitude/longitude/depth grid, with the operator D approximated by a difference operator, and with the integration in the second term of Ψ performed as a weighted summation over the grid. Given this discrete approximation to the problem, the minimization of Ψ is performed with a conjugate gradients technique.

The maximum-likelihood inversion approach is equivalent to conventional kriging with a given correlation function when D is the *inverse* of the *covariance* operator associated with the correlation function (see Tarantola, 1987). That is, the covariance operator, $C(\mathbf{x}, \mathbf{x}')$, is the Green's function for the differential operator:

$$DC(\mathbf{x}, \mathbf{x}') = \delta(\mathbf{x} - \mathbf{x}') \tag{2.13}$$

where δ is the Dirac-delta function. When $C(\mathbf{x}, \mathbf{x}')$ depends only on the separation between $\mathbf{x}$ and $\mathbf{x}'$, it can be identified with a correlation function, R:

$$C(\mathbf{x}, \mathbf{x}') = R(\mathbf{x} - \mathbf{x}') \tag{2.14}$$

2.3.2 Differential Operators for a Class of Correlation Functions

In this project we developed ML kriging using the following operators. For three-dimensional parameter functions, varying with latitude (θ), longitude (ϕ) and depth (z), we let

$$D_3 = \frac{(2\ell - 3)^{3/2}}{(\ell - 1) 2^{2\ell - 1} \pi \lambda_1^2 \lambda_2 \sigma_0^2} C_{\ell-2}^{2\ell-4} \left[\delta(\mathbf{x}) - \frac{1}{(2\ell - 3)} \left(\frac{\lambda_1^2}{r^2} \nabla_1^2 + \lambda_2^2 \frac{\partial^2}{\partial z^2} \right) \right]^\ell, \quad \ell \geq 2 \tag{2.15}$$

where $r \equiv r_0 - z$ (r_0 being the Earth's mean radius), and where ∇_1^2 is the 2-D Laplacian operator in geographical coordinates,

$$\nabla_1^2 = \frac{\partial^2}{\partial \theta^2} + \frac{1}{\cos^2 \theta} \frac{\partial^2}{\partial \phi^2}. \tag{2.16}$$

C_n^m is the binomial coefficient for numbers m and n. For 2-D functions having no depth dependence, we let

$$D_2 = \frac{1}{2\pi \lambda_1^2 \sigma_0^2} \left[\delta(\mathbf{x}) - \frac{1}{(2\ell - 2)} \frac{\lambda_1^2}{r_0^2} \nabla_1^2 \right]^\ell, \quad \ell \geq 2. \tag{2.17}$$

For completeness, the analogous operator for 1-D functions of depth is

$$D_1 = \frac{(2\ell-1)^{1/2}}{2^{2\ell-1}\lambda_2\sigma_0^2} C_{\ell-1}^{2\ell-2} \left[\delta(\mathbf{x}) - \frac{1}{(2\ell-1)}\lambda_2^2 \frac{\partial^2}{\partial z^2}\right]^{\ell}, \quad \ell \geq 1. \tag{2.18}$$

In these expressions, σ_0^2 is a prior variance of the parameter function; λ_1 is a correlation length in the horizontal direction; λ_2 is a correlation length in the vertical direction; and ℓ is an integer order of the operator.

Various choices of ℓ correspond to correlation functions of various types. For example, in 3-D with $\ell = 2$ we have

$$D_3 = \frac{1}{8\pi\lambda_1^2\lambda_2\sigma_0^2} \left[\delta(\mathbf{x}) - \frac{\lambda_1^2}{(r_0 - z)^2}\nabla_1^2 - \lambda_2^2 \frac{\partial^2}{\partial z^2}\right]^2. \tag{2.19}$$

Setting $\lambda_1 = \lambda_2 = \lambda$ (to avoid defining more notation), the Green's function for D_3 in a *wholespace* is the covariance operator given by

$$C(\mathbf{x}, \mathbf{x}') = \sigma_0^2 \exp(-|\mathbf{x} - \mathbf{x}'|/\lambda). \tag{2.20}$$

Comparing to (2.14), we recognize the correlation function (R) of the *exponential* type. In a bounded domain, the wholespace solution for $C(\mathbf{x}, \mathbf{x}')$ will apply for points far from a boundary, as measured against the correlation length λ. Near a boundary (e.g., the Earth's surface) the wholespace solution does not apply and $C(\mathbf{x}, \mathbf{x}')$ does not have the stationary structure of (2.14). In 1-D, an exponential correlation function results with $\ell = 1$:

$$D_1 = \frac{1}{2\lambda\sigma_0^2} \left[\delta(\mathbf{x}) - \lambda^2 \frac{\partial^2}{\partial z^2}\right]. \tag{2.21}$$

whose wholespace Green's function also is an exponential-type correlation function:

$$C(z, z') = \sigma_0^2 \exp(-|z - z'|/\lambda). \tag{2.22}$$

In 2-D, an exponential correlation function cannot be achieved with any integer ℓ. However, $\ell = 2$ yields a correlation function that is similar but less peaked at zero lag.

In 1, 2 or 3 dimensions, a correlation operator of the *Gaussian* type results as $\ell \to \infty$, e.g. in 3-D,

$$C(\mathbf{x}, \mathbf{x}') \to \sigma_0^2 \exp(-\frac{1}{2}|\mathbf{x} - \mathbf{x}'|^2/\lambda^2). \tag{2.23}$$

(Note that this formula defines the correlation length to be the lag for which the covariance drops to $e^{-1/2} = 0.61$ of its peak. It is usually defined to be at $e^{-1} = 0.37$ of the peak.)

2.3.3 Examples of Numerical Correlation Functions

To check the validity of our alternative approach to kriging, we applied it to synthetic inverse problems whose theoretical solution is the covariance operator itself. Specifically, let

$$\Psi_1 = (1 - a(\mathbf{x}_0))^2 + \int a(\mathbf{x})\,[Da](\mathbf{x})\,d\mathbf{x}. \tag{2.24}$$

Comparing to (2.12), we see that this is the objective function when $a(\mathbf{x})$ has been observed at the single point $\mathbf{x}_0$. It is not too difficult to show that the minimum of Ψ_1 is obtained with

$$a(\mathbf{x}) = \frac{C(\mathbf{x}, \mathbf{x}_0)}{1 + C(\mathbf{x}_0, \mathbf{x}_0)}. \tag{2.25}$$

Clearly then,

$$a(\mathbf{x}_0) = \frac{C(\mathbf{x}_0, \mathbf{x}_0)}{1 + C(\mathbf{x}_0, \mathbf{x}_0)}. \tag{2.26}$$

We can combine (2.25) and (2.26) and get

$$C(\mathbf{x}, \mathbf{x}_0) = \frac{a(\mathbf{x})}{1 - a(\mathbf{x}_0)}. \tag{2.27}$$

Thus, the solution of the single-datum inverse problem in (2.24) yields a slice of the covariance function that is the Green's function of the differential operator D.

We applied this technique on the discretized problem using our conjugate gradients minimization algorithm. The parameter grid covered the portion of Asia from 5°–55°N and from 40°–105°E, with a 0.5° spacing in both latitude and longitude. The point $\mathbf{x}_0$ was chosen near the geographic center (30°N, 72.5°E) and at zero depth. This leads to a stationary covariance function for suitably small choices of the correlation lengths. Both 2-D and 3-D correlation functions were generated for $\sigma_0 = 1$ and various values of the parameters ℓ and λ. The results of the calculations are shown in Figures 2.1–2.4. Fair conclusions from these results are, first, that maximum-likelihood inversion provides a valid approach to kriging and, second, the family of differential operators D defined in equations (2.15) and (2.17) provide a reasonable range of correlation functions for practical applications.

2.3.4 Kriging for Mislocation-Vector Parameter Functions

For the mislocation-vector parametrization, the multiple-station kriging inverse problem is

$$r_{ij} = \mathbf{p}_{ij} \cdot \mathbf{a}(\mathbf{x}_j) + \mathbf{q}_{ij} \cdot \mathbf{a}(\mathbf{y}_i) + e_{ij} \tag{2.28}$$

where we have defined the further contractions $\mathbf{p}_{ij} \equiv \mathbf{p}(\mathbf{x}_j, \mathbf{y}_i)$ and $\mathbf{q}_{ij} \equiv \mathbf{q}(\mathbf{x}_j, \mathbf{y}_i)$. This is an even more complicated interpolation problem than in the time-term case, in which each observation samples a projection of the unknown vector function $\mathbf{a}$ at two points. Nonetheless, the maximum-likelihood approach defines an optimal solution in the same way, with the objective function becoming

$$\Psi = \sum_{ij} \left| r_{ij} - \mathbf{p}_{ij} \cdot \mathbf{a}(\mathbf{x}_j) - \mathbf{q}_{ij} \cdot \mathbf{a}(\mathbf{y}_i) \right|^2 / \sigma_{ij}^2 + \int \mathbf{a}(\mathbf{x}) \cdot [D\mathbf{a}](\mathbf{x})\, d\mathbf{x}. \tag{2.29}$$

In the second term, D operates on $\mathbf{a}$ component-wise, and the vector dot product of $\mathbf{a}$ with $D\mathbf{a}$ leads to a sum over components. Geo-statistically, this means that the components of $\mathbf{a}$ are assumed to be independent random functions, each having the covariance operator implied by D.

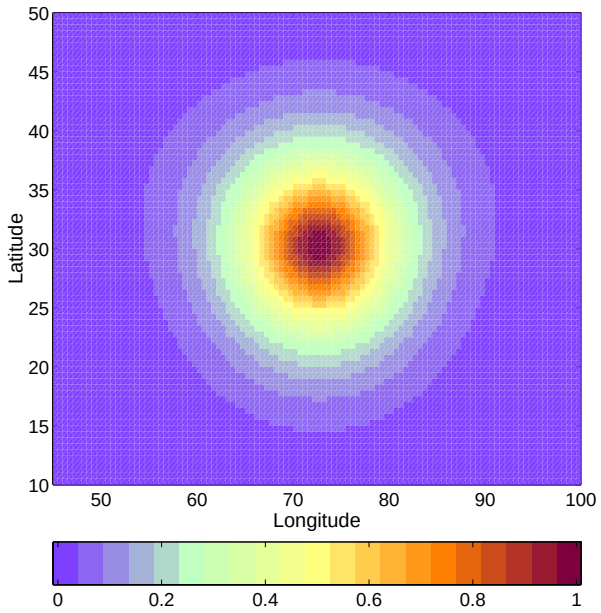

Figure 2.1: A 2-D correlation function generated with our ML kriging algorithm. The case shown is for order $\ell = 3$ and correlation length $\lambda = 500$ km. The result has the correct properties of a correlation function, with maximum value at the synthetic observation point (30° N, 72.5° E).

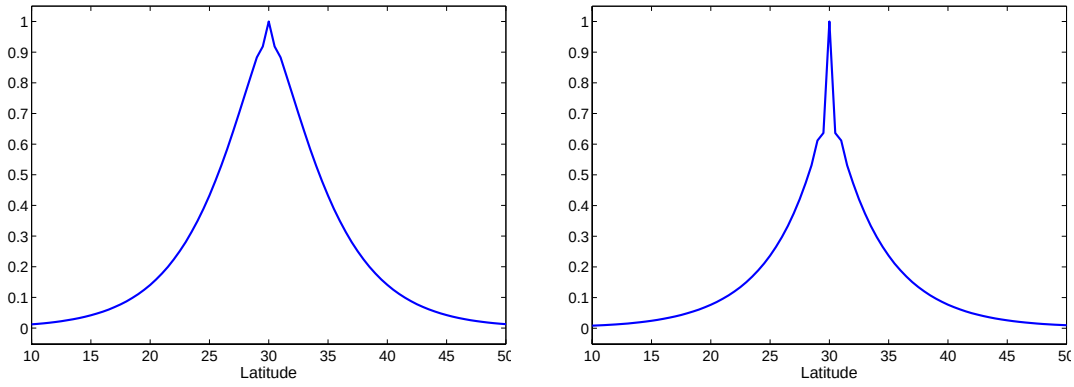

Figure 2.2: Profile (vs. latitude) through a 2-D correlation function (left) and a 3-D one (right), computed with our ML kriging approach. Both cases use order $\ell = 2$ and correlation length $\lambda = 500$ km. The longitude in each case, and depth in the 3-D case, are evaluated at the synthetic observation point: 72.5°E. $z = 0$. In 3-D, the analytic solution is an exponential function, while in 2-D it is a function which is less sharply peaked at the center.

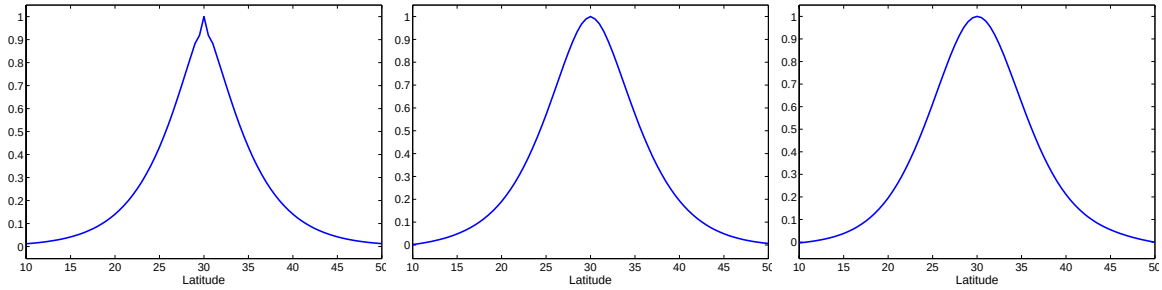

Figure 2.3: Cross-sections of 2-D correlation functions computed with orders $\ell = 2$, 3 and 4 (left, center, right) and correlation length $\lambda = 500$ km.

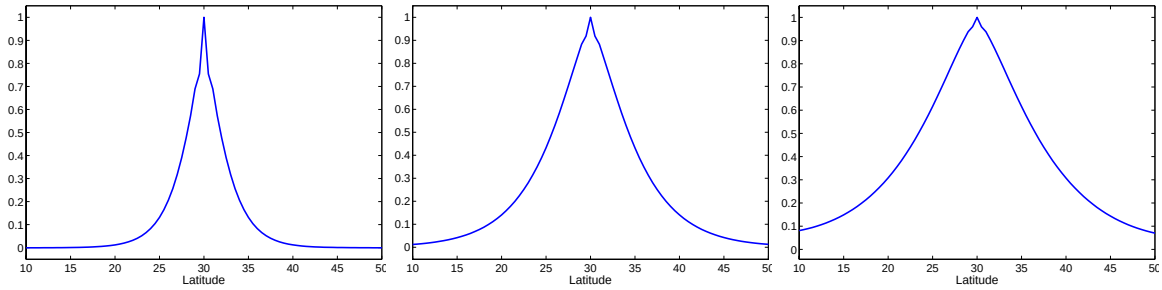

Figure 2.4: Cross-sections of 2-D correlation functions computed with order $\ell = 2$ and correlation lengths $\lambda = 250$, 500 and 750 km (left, center, right).

2.4 Combining Multiple-Station Kriging With Event Location

Many investigators have developed techniques for solving the "basic" multiple-event location problem, i.e. jointly estimating event locations and a simple travel-time correction (time term) for each station/phase combination. Examples include Dewey (1971), Jordan and Sverdrup (1981) and Pavlis and Booker (1983). When more complicated and less restrictive parameterizations of path corrections are involved, workers have generally fixed the event locations, assuming they are GT0 or attempting to factor their location uncertainty into the uncertainty analysis of the resulting estimated path corrections. A notable exception occurs in earthquake tomography, in which path corrections are parametrized in terms of a velocity model. Some tomography studies using earthquake arrival times have solved for updated earthquake locations in conjunction with the velocity model (e.g. Reiter et al., 2005).

This project took the first steps toward combining event location and travel-time calibration with correction functions into a joint inversion procedure. The next subsections describe the basic multiple-event location algorithm, GMEL, which was the starting point for this effort and the the new prototype algorithm, KMEL, that combines event location and multiple-station kriging.

2.4.1 GMEL

In basic multiple-event location, one finds event locations, $(\mathbf{x}_j, t_j)$, $j = 1, \ldots, m$, and time terms, c_i, $i = 1, \ldots, n$, which minimize the objective function given by (in the case of Gaussian

errors)

$$\Psi = \sum_{ij} |d_{ij} - T_i(\mathbf{x}_j) - t_j - c_i|^2 / \sigma_{ij}^2. \tag{2.30}$$

Rodi et al. (2002) described and compared the algorithm used in various basic multiple-event location methods (e.g., JHD, Dewey, 1971; PMEL, Pavlis and Booker, 1983) for performing this minimization. In this project we used an algorithm called GMEL (grid-search multiple-event location) developed at M.I.T. under support of DOE/NNSA and the Air Force Research Laboratory. The algorithm alternates between pure event location and pure travel-time calibration by iterating the following two-step process:

1. With the travel-time corrections fixed to their current values, minimize Ψ with respect to the event location parameters ($\mathbf{x}_j$, t_j).

2. With the event locations fixed to their current values, minimize Ψ with respect to the station time terms (c_i).

Step 1 above is performed by applying single-event location to each event in term. GMEL does this with a grid-search algorithm (Rodi and Toksöz, 2000). Step 2 is performed for each station/phase combination in turn. For Gaussian errors, each c_i is simply a weighted average of the residuals across events. The two steps are repeated until a convergence is reached.

The GMEL algorithm is not particularly efficient compared to previous multiple-event location algorithms, but it is simple, fully nonlinear, and can also handle the case of non-Gaussian data errors. Further, the use of grid search simplifies the use of parameter constraints in the minimization of Ψ. Thus, for example, one can include ground-truth constraints on event locations when they are available.

2.4.2 KMEL

One can replace c_i in (2.30) with other parameterizations of travel-time corrections. In the case of the mislocation-vector parameterization, we replace c_i with $C_i(\mathbf{x}_j)$, as given by equation (2.9). Further, we can add the regularization term we used in multiple-station kriging to Ψ, which then becomes

$$\begin{aligned}\Psi = \sum_{ij} |d_{ij} - T_i(\mathbf{x}_j) - t_j - \mathbf{p}_i(\mathbf{x}_j) \cdot \mathbf{a}(\mathbf{x}_j) - \mathbf{q}_i(\mathbf{x}_j) \cdot \mathbf{a}(\mathbf{y}_i)|^2 / \sigma_{ij}^2 \\ + \int \mathbf{a}(\mathbf{x}) \cdot [D\mathbf{a}](\mathbf{x})\, d\mathbf{x}.\end{aligned} \tag{2.31}$$

The joint calibration/location problem is solved now by minimizing Ψ with respect to the $(\mathbf{x}_j, t_j)$ and the mislocation-vector parameter function $\mathbf{a}(\mathbf{x})$.

Our prototype algorithm for solving this minimization problem follows a similar scheme to GMEL, iterating the following *three* steps:

1. Using the current $\mathbf{a}(\mathbf{x})$, generate a 3-D travel-time correction table, $C_i(\mathbf{x})$, for each station/phase, based on equation (2.9).

2. Perform single-event location on each event in turn, applying travel-time corrections interpolated from the current 3-D tables.

3. Using the current event locations, perform multiple-station kriging on the travel-time residuals (r_{ij} of equation (2.10)) to obtain an updated parameter function $\mathbf{a}(\mathbf{x})$.

Step 2 minimizes Ψ in (2.31) with respect to the $(\mathbf{x}_j, t_j)$ with $\mathbf{a}(\mathbf{x})$ fixed. Step 3 minimizes Ψ with respect to $\mathbf{a}(\mathbf{x})$ with the event locations fixed. Iterating the three steps until convergences minimizes Ψ with respect to all the parameters. We note that our grid-search single-event location algorithm has the capability to use 3-D travel-time correction tables (correction vs. latitude, longitude and depth), as required in Step 2.

In the next chapter, we applied GMEL and KMEL to real data from the Nevada Test Site in order to compare the relative merits of simple time-term corrections and correction functions parameterized by a mislocation-vector function. We point out that we did not develop a stand-alone computer program that repeats the three steps above. Rather, KMEL was emulated with scripts that invoke our event location program and multiple-station kriging program in alternating batch runs.

Chapter 3

Calibration and Event Location Tests Using Pn Arrival Times From Nevada Test Site Explosions

3.1 Introduction

In previous work (Walter et al., 2003), Lawrence Livermore National Laboratory (LLNL) developed a data set of seismic arrival times picked from 74 Nevada Test Site (NTS) nuclear explosions having precisely known locations and origin times (i.e. GT0 events). The explosions were recorded at a network of regional and local stations, and a LLNL analyst re-picked the phases and supplemented them with selected picks from the National Earthquake Information Center (NEIC) to produce a high quality data set. The data set has been used in studies of event location accuracy at LLNL (e.g. Myers et al., 2003) and we used it in this project to test our new approach to multiple-event location and kriging. The examples here use only the Pn picks from the data set and we removed events with fewer than four Pn arrivals. The resulting subset comprises 674 arrivals from 71 events and 59 stations.

3.2 Travel-Time Corrections From Multiple-Station Kriging

The first part of our analysis applied our multiple-station kriging approach to the NTS data to estimate a generic parameter function of the mislocation-vector type (see equation (2.6)). Depth dependence of the parameter function was not allowed since the events are all shallow. However, we point out that the mislocation-vector itself, $\mathbf{a}(\mathbf{x})$ is 3-D, with a north, east and depth component at each latitude/longitude point $\mathbf{x}$. We also note that that travel-time corrections generated from $\mathbf{a}(\mathbf{x})$ will be mildly depth-dependent owing to the depth dependence of the slowness vectors in equation (2.6).

We applied multiple-station kriging to travel-time residuals relative to the IASP91 model-based times. The residuals were calculated by setting the event locations and origin times to their GT0 values. We used four sets of the geo-statistical input parameters defined in equation (2.17). The correlation length (λ_1) was set to either 100 or 200 km. The prior error (σ_0) on each mislocation-vector component (north, east or depth) was set to 5 or 10 km. In each case, the order of the correlation operator (ℓ) was set to 2. This corresponds to a 2-D correlation function

Table 3.1: Posterior RMS Residual vs. Kriging Parameters

(Prior RMS = 0.93 s)

Case	*Correlation Length (km)*	*Prior Error (km)*	*RMS Residual (s)*
A	200	5	0.46
B	200	10	0.44
C	100	5	0.44
D	100	10	0.43

that is more sharply peaked at zero distance than a correlation function of the Gaussian type, but less sharply peaked than a correlation function of the exponential type. Table 3.1 lists the four correlation length/prior error combinations, which are labeled A through D.

All four cases yielded similar solutions and fits to the data (last column of Table 3.1). Figures 3.1 and 3.2 show the mislocation-vector functions for cases A and D, respectively. The functions are similar, but the smaller correlation distance combined with the larger prior error for case D (Figure 3.2) results in a rougher function with somewhat larger values, especially in the east component. Case D also fits the travel-time residuals better than case A, resulting in 0.428 s posterior RMS compared to 0.455 s. The prior RMS of the residuals is 0.926 s.

Next we show some station-specific travel-time correction functions generated from the second parameter function. Figure 3.3 shows the correction functions for stations ELK and BMN. These two stations are north of NTS and close to one another, and we see that their travel-time corrections are very similar (highly correlated). Figure 3.4 displays the correction functions for KNB, a station east of NTS, and LAC, south of NTS. Their pattern of corrections show significant differences from each other and from the northern stations.

3.3 Experiments With GT0-Calibrated Single-Event Location

We performed some experiments with event location to compare the location accuracy achieved with various types of path travel-time corrections. In each test, events were located individually with the path corrections fixed, i.e. *single*-event location. Three types of path corrections were tested:

1. Zero path corrections.
2. A time-term correction for each station.
3. A travel-time correction *function* for each station, generated from a mislocation-vector parameter function.

In both 2 and 3, the corrections were developed from travel-time residuals with respect to the given GT0 locations of the NTS explosions. Simple averaging of residuals over events produce the time terms for the type 2 corrections. The correction functions (type 3) are those we obtained by multiple-station kriging, as described in Section 3.2 (see Figures 3.3 and 3.4 for examples). Four tests of type 3 were performed, one for each combination of kriging parameters

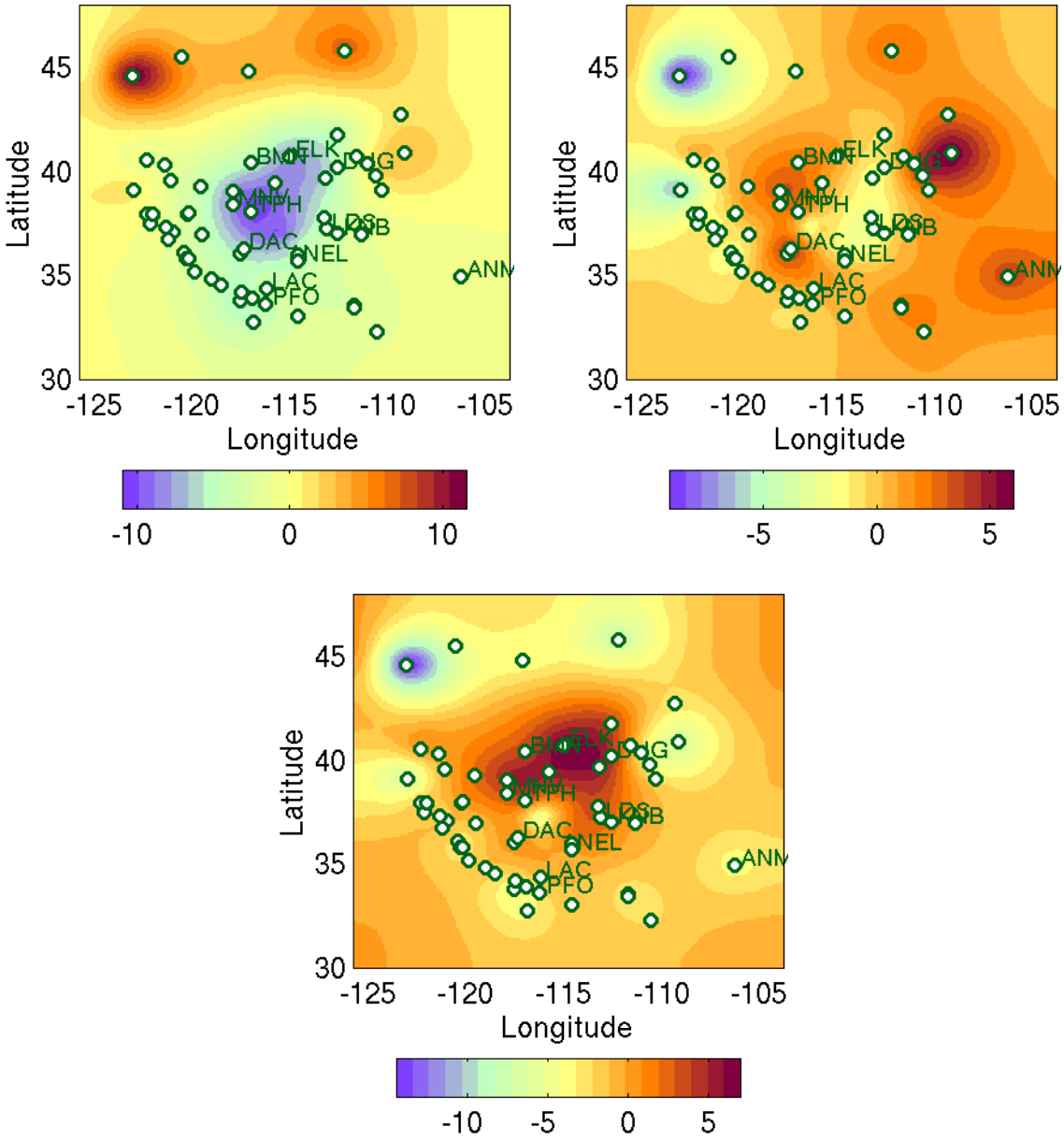

Figure 3.1: Mislocation-vector parameter function (in km) for the region surrounding NTS: north, east and depth components (top left, top right, and bottom frames, respectively). This parameter function was derived by multiple-station kriging of Pn travel-time residuals, using 200 km for the correlation length and 5 km for the prior error on the mislocation components (case A in Table 3.1). Station locations (some labeled) are marked with circles.

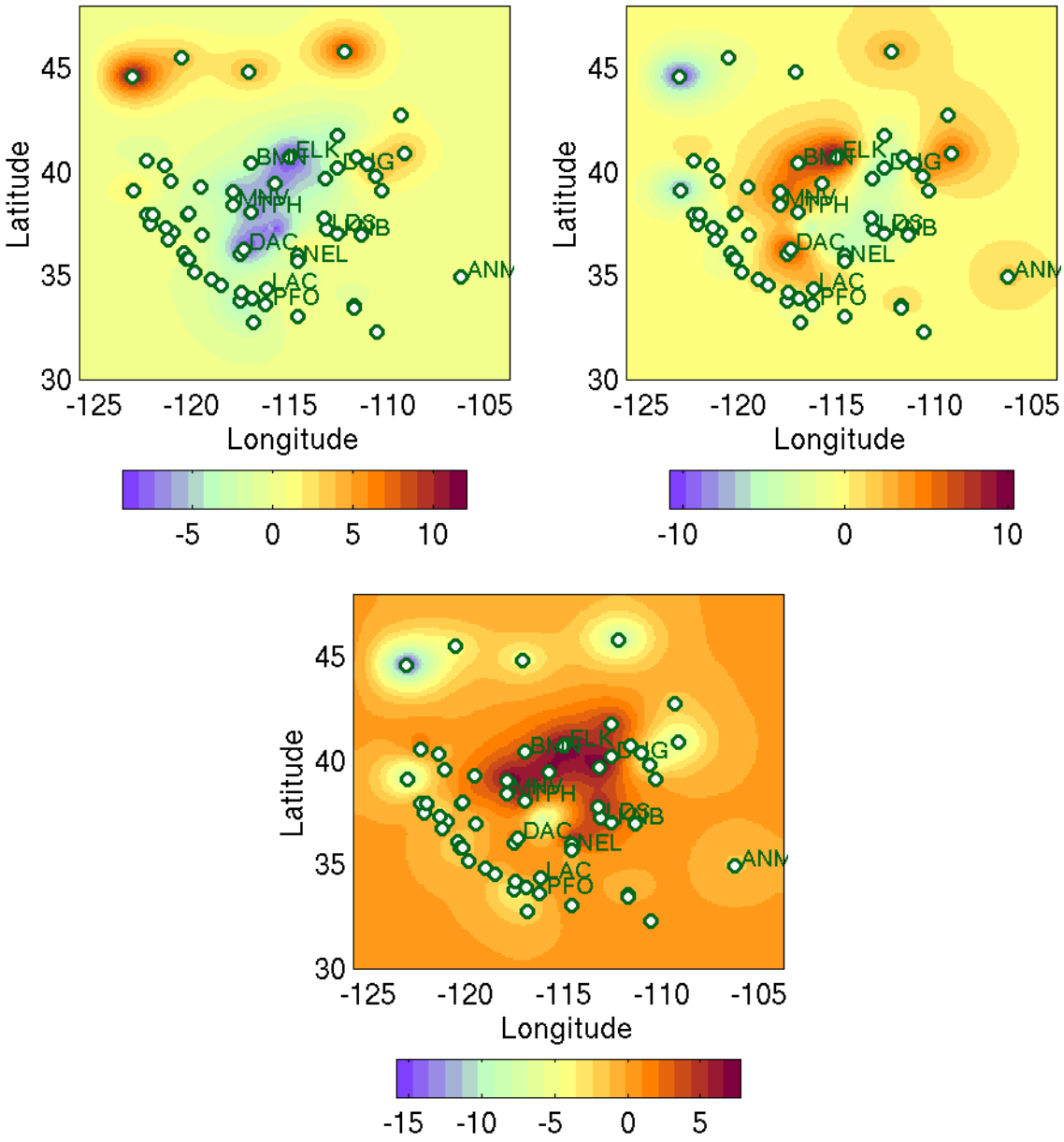

Figure 3.2: Same as Figure 3.1, except the mislocation-vector function was obtained using 100 km for the correlation length and 10 km for the prior error on the mislocation components (case D).

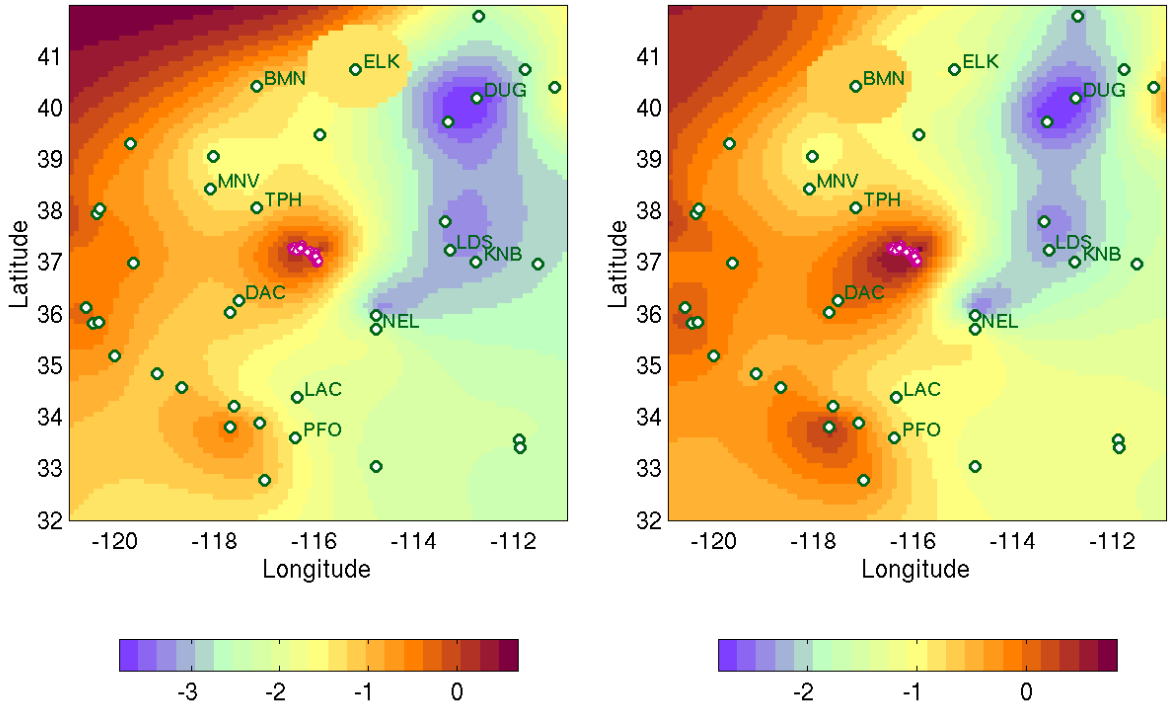

Figure 3.3: Travel-time correction functions (in seconds) for stations ELK (left) and BMN (right), generated from the mislocation-vector parameter function shown in Figure 3.2. Station locations (some labeled) are marked with circles, and the NTS events show as a small cluster of small white dots near the center of the plot.

listed in Table 3.1. Therefore, six different solutions for the NTS event locations were generated among the three types of corrections.

We point out that, in this application, the event aperture is small (approx. 50 km). Therefore, the main difference between correction terms and correction functions is that the latter allow for correlation between stations, based on their separation. Station terms, in contrast, are independent between stations regardless of their separation.

In all the location tests, event depths were fixed to zero. IASP91 travel-time tables were used for model-based times to be consistent with the use of residuals relative to IASP91 in developing the travel-time corrections. The location runs were done with our grid-search single-event location algorithm, described in Rodi and Toksöz (2000). This algorithm has the capability to add time-term corrections to the model-based travel-times, or to interpolate travel-time corrections from 3-D tables, which is required for correction functions derived from multiple-station kriging (type 3).

Figure 3.5 shows the location solution obtained when zero travel-time corrections are applied (uncalibrated), comparing them to the known (GT0) event locations. Figure 3.6 shows two of the five solutions obtained with corrections applied: the one in which the corrections are station time terns (top), and one of the four solutions (case D) obtained with kriged corrections. Table 3.2 summarizes the results of all the tests in terms of various epicentral mislocation statistics. Mislocations are with respect to the known GT0 epicenters. The three columns to the left list

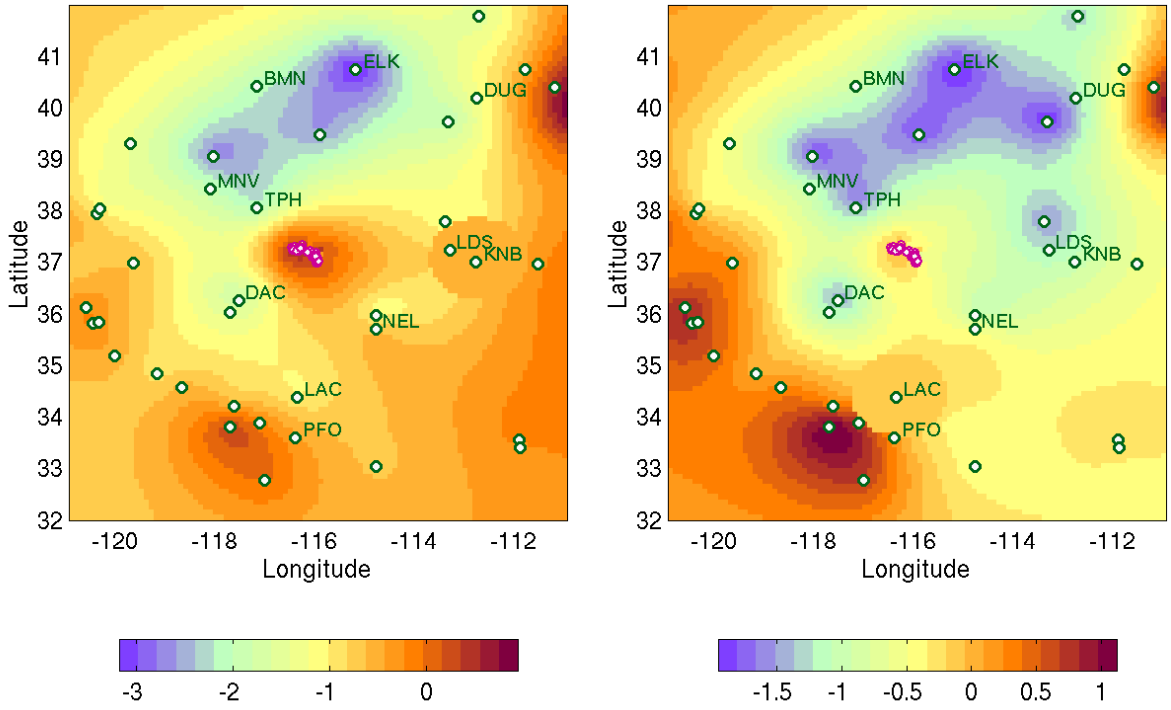

Figure 3.4: Travel-time correction functions for stations KNB (left) and LAC (right), generated from the mislocation-vector parameter function shown in Figure 3.2.

the mean, median and 90%-tile of the historgram of epicentral mislocation distances. The three columns to the right state the number of events (out of 71) that are within certain epicentral distances of the GT0 locations. We repeat that the same ground-truth locations were used to develop the corrections of various types, i.e. these are *not* cross-validated tests.

We see from Table 3.2 that in all cases the event mislocations are not very large, owing to the good station coverage around NTS and high quality of the Pn arrival picks. However, there are noticeable differences in the relative performance of different types of corrections. As expected, locations performed with any version of travel-time corrections improve location accuracy significantly, compared to zero corrections. We also see that the mislocations obtained using kriged correction functions are somewhat better than those using station time terms, most notably in the number of events mislocated by less than 1 km. We believe this is because the kriged correction functions allow for correlation between stations based on their separation, while station time terms do not.

3.4 Experiments With Multiple-Event Location (Joint Calibration/Location)

The travel-time corrections developed in the previous section were derived from the 71 NTS events, using their GT0 locations, and then were tested by applying the corrections in a relocation of these same events. This section performs similar experiments with the same data

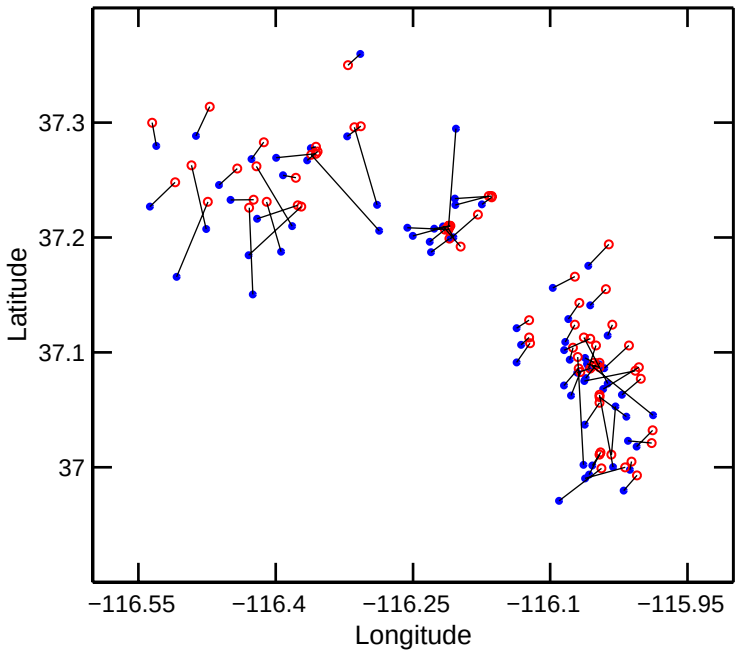

Figure 3.5: NTS event locations resulting from a single-event location algorithm applied with no station corrections. The blue circles are the seismic solutions and the red circles are the ground-truth (GT0) locations. A line connects the solution and GT0 location of each event. Only Pn arrival times were included in the data set.

set more realistically. Namely, we employ *multiple*-event location techniques to simultaneously derive travel-time corrections and relocate the events. We applied no location constraints in the process, which thus embodies a limiting case of joint calibration/location; in general, some events would have GT constraints while others are the target of the relocation process. In this limiting case, information about the event locations, beyond the arrival time data themselves, arises only indirectly from prior information one might impose on the travel-time corrections and from the requirement that the event locations and corrections be self-consistent.

The experiments reported in this section compare two multiple-event location techniques. The first solves jointly for the event locations and a time-term correction at each station (i.e. "basic" multiple-event location). We performed this task with the GMEL algorithm, described in Section 2.4.1. The second technique solves jointly for the event locations and a generic parameter function of the mislocation-vector type, from which station-specific travel-time functions can be generated. This entails solving a joint location/kriging problem, which we did with the prototype algorithm KMEL described in Section 2.4.2. KMEL was applied four ways, corresponding to four cases of the geo-statistical parameters used in multiple-station kriging. They are the same four cases considered in the previous section (cases A–D in Table 3.1).

Two of the multiple-event location solutions are shown in Figure 3.7. Plotted on the top is the solution obtained with GMEL (station time terms). The bottom panel is one of the four KMEL solutions that use a mislocation-vector parameter function (case D kriging parameters). It is obvious that the epicentral mislocations for the station time-term solution (GMEL, top)

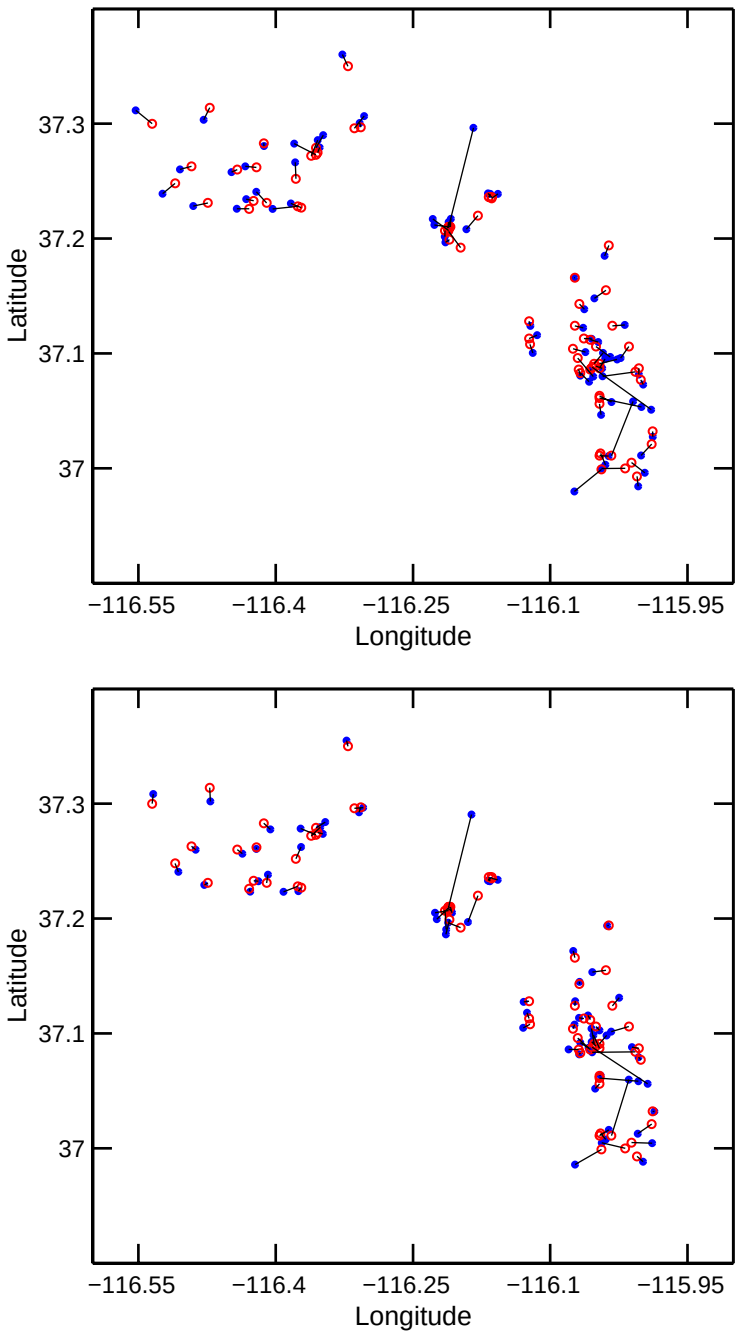

Figure 3.6: NTS event locations resulting from a single-event location algorithm applied with station time-term corrections (top) and with travel-time correction functions derived from multiple-station kriging, case D (bottom). In each case, the corrections were derived from travel-time residuals from these same events relative to their GT0 locations. Plotting conventions are the same as in Figure 3.5.

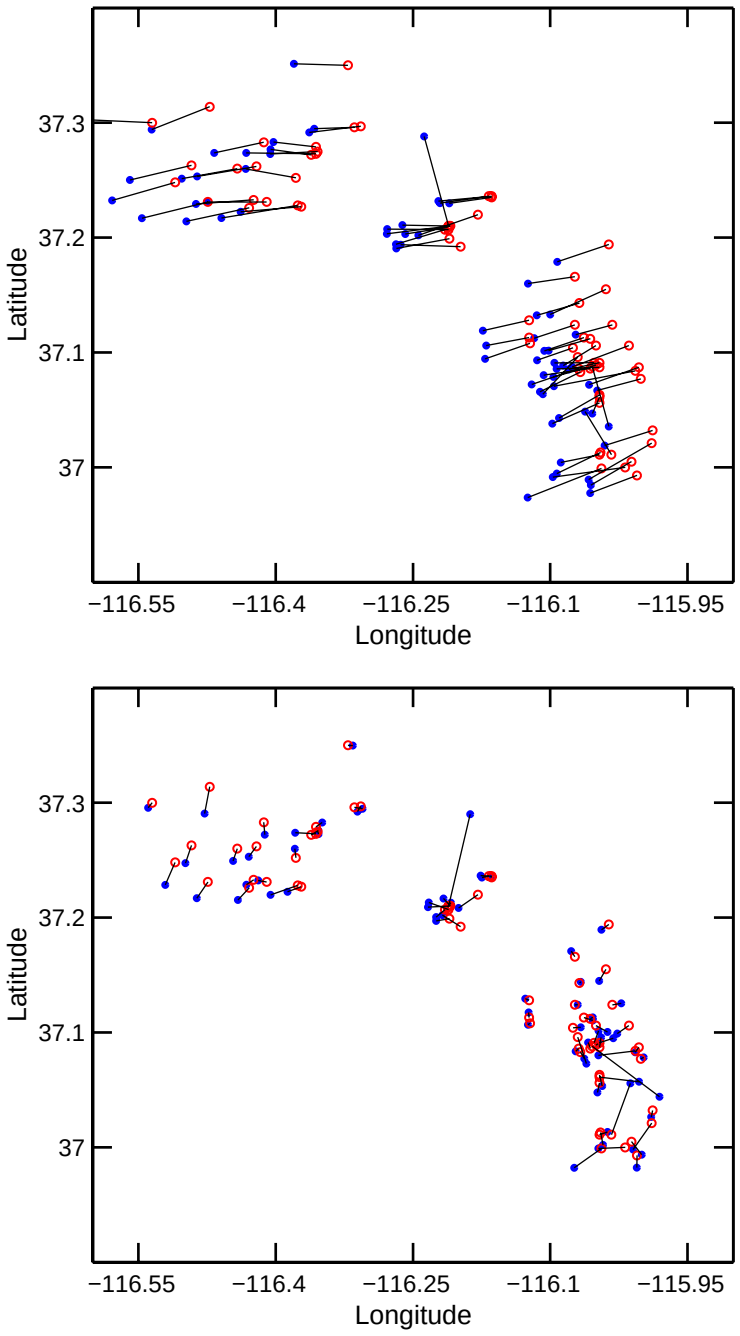

Figure 3.7: NTS event locations resulting from multiple-event location performed with GMEL (top) and KMEL (bottom). GMEL solves for travel-time corrections that are simple station time terms. KMEL solves for travel-time correction functions that are parameterized in terms of a mislocation-vector function. The KMEL solution shown uses the case D kriging parameters (see Table 3.1). Plot conventions are the same as in Figure 3.6.

Table 3.2: Mislocations of NTS Events: Comparison of Path-Corrections in Single-Event Location

Correction	*Mislocation Metric (km)*			*Number of Events*		
Type	median	mean	90%-tile	< 1	< 2	< 3 km
Zero corrections	2.4	3.3	7.1	3	23	42
Station terms	1.2	1.5	2.4	28	60	65
Kriged A	0.9	1.3	2.2	37	62	65
Kriged B	0.8	1.3	2.4	42	63	66
Kriged C	0.8	1.3	2.4	40	63	65
Kriged D	0.8	1.3	2.3	41	62	66

Table 3.3: Mislocations of NTS Events: Comparison of Path-Corrections in Multiple-Event Location

Correction	*Centroid*	*Rel. Misloc. Metric (km)*			*Number of Events*		
Type	*Misloc. (km)*	median	mean	90%-tile	< 1	< 2	< 3 km
Station terms	4.8	1.2	1.6	3.0	29	59	63
Kriged A	2.3	1.4	1.7	3.0	25	54	63
Kriged B	2.1	1.0	1.4	2.7	36	58	65
Kriged C	2.2	1.0	1.5	2.8	37	59	65
Kriged D	0.3	1.0	1.5	2.6	35	57	65

are dominated by a west-southwestward bias consistent across events. This is a consequence of the ambiguity between the station corrections and the centroid location of the event cluster that occurs when neither the event locations nor travel-time corrections are constrained with prior information (Jordan and Sverdrup, 1981). We do not see this bias in the locations obtained with kriged corrections (KMEL, bottom) because the geo-statistical constraints on the mislocation-vector function serve to constrain the travel-time corrections. Since, in practice, the station time-term approach would also use some constraints on either some of the event locations or on travel-time corrections, we have evaluated the event mislocation results of our multiple-event location experiments mainly in terms of *relative* mislocations.

These results are shown in Table 3.3, including the case with station time-term corrections (line 1) and the four cases with kriged corrections (lines 2–5). The first column of numbers is the mislocation in the centroid of the 71 events, where in line 1 we see numerically the centroid/correction ambiguity mentioned above (i.e. 4.8 km centroid mislocation). The remaining six columns show the same statistics we showed for the single-event location experiments (Table 3.2), except they are determined from the relative mislocations, i.e. the mislocations remaining after correcting for the centroid mislocation (via a uniform translation of the solution locations). We observe that the multiple-event mislocations (Table 3.3) are generally larger than the single-event mislocations (Table 3.2), even though the former were forgiven for a centroid shift. This is because, in the multiple-event solutions, the travel-time corrections fit the residuals with respect to the estimated event locations, whereas the single-event locations we derived used

corrections that were based on the GT0 event locations.

The relative mislocation results in Table 3.3 indicate that multiple-event location done jointly with kriging (KMEL) is somewhat better, in three of four cases, than with correction terms (GMEL). The worst of the four kriging cases was A, whose combination of geo-statistical parameters was most restrictive on the travel-time corrections (see Table 3.1). While kriging cases B–D only show a slight improvement over the station-term case on average (about 10% smaller mislocations), we note that the number of events mislocated by less than 1 km is significantly improved, dropping by about 6 events. The performance difference we see in these experiments may have been diminished by the crudeness of our prototype KMEL procedure, and may increase when KMEL is developed into an integrated, optimized algorithm. Further, it remains to compare the performance of GMEL and KMEL when some GT constraints are applied in the joint location/calibration process.

Chapter 4

Using a Mixture of Gaussians to Model Seismic Arrival Time Error Distributions: Estimation and the Impact on Seismic Event Inference

4.1 Mixture of Gaussian Distributions

In what follows, we give a brief description of modeling the distribution of a random variable via a mixture of univariate Gaussian distributions. Both Bayesian (MCMC) and frequentest (EM) inference approaches will be outlined.

Let y be a random variable with probability density given by

$$p(y \mid k, \mathbf{w}, \boldsymbol{\mu}, \boldsymbol{\sigma}^2) = \sum_{j=1}^{k} w_j \varphi(y \mid \mu_j, \sigma_j^2), \tag{4.1}$$

where $w_j > 0$, $\sum_{j=1}^{k} w_j = 1$, and $\varphi(\cdot \mid \mu_j, \sigma_j^2)$ denotes the univariate Gaussian probability density with mean μ_j and variance σ_j^2. The components of the mixture in (4.1) can either have a specific interpretation, with each component representing a "group", or the mixture can just be viewed as a semi-nonparametric approach to approximate the distribution of y. In either case, it is convenient to think of the k components of (4.1) as representing different groups or populations. As such, drawing a random variable from the density in (4.1) consists of two steps: first draw a group label, say z, with $P(z = j) = w_j$, and then draw y randomly from a Gaussian distribution with mean μ_z and variance σ_z^2. The group label z is then discarded and y is only available to the observer. Note that conditional on z, in addition to k, $\boldsymbol{\mu}$, and $\boldsymbol{\sigma}^2$,

$$p(y \mid z, k, \boldsymbol{\mu}, \boldsymbol{\sigma}^2) = \varphi(y \mid \mu_z, \sigma_z^2).$$

Given observed data $\mathbf{y} = (y_1, \ldots, y_n)$ our goal is to make inferences on $(k, \boldsymbol{\mu}, \boldsymbol{\sigma}^2)$ and in particular, if y^* is a new observation, being able to conclude something about the probability of the newly observed y^*.

4.1.1 Maximum Likelihood Estimation: The EM Algorithm

Assume k is known (or fixed). The likelihood of the observed data is given by

$$L(\mathbf{w}, \boldsymbol{\mu}, \boldsymbol{\sigma}^2) = \prod_{i=1}^{n} p(y_i \mid \mathbf{w}, \boldsymbol{\mu}, \boldsymbol{\sigma}^2) = \prod_{i=1}^{n} \left(\sum_{j=1}^{k} w_j \varphi(y_i \mid \mu_j, \sigma_j^2) \right). \tag{4.2}$$

The maximum likelihood estimate (MLE) of $(\mathbf{w}, \boldsymbol{\mu}, \boldsymbol{\sigma}^2)$, which we shall denote by $(\hat{\mathbf{w}}, \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\sigma}}^2)$, is the set of parameter values that maximize (4.2). The maximization of (4.2) is typically achieved using the Expectation-Maximization (EM) algorithm of Dempster et al. (1977) The EM algorithm takes advantages of the fact that, if the (missing) group label vector $\mathbf{z} = (z_1, \ldots, z_n)$ were available, then the MLEs are easily obtained. Hence, the algorithm treats $\mathbf{z}$ as missing data and iterates between computing the expected (log) likelihood of the complete data ($\mathbf{y}$ and $\mathbf{z}$), with respect to the $\mathbf{z}$ (the E-step), and maximizing the expected likelihood (the M-step); see Dempster et al. (1977) for details.

One of the drawbacks of the EM algorithm, as with many maximization algorithm, is that it only guaranties a local maximum. Hence, EM can be very sensitive to the starting values provided, particularly to the starting value of $\boldsymbol{\mu}$. In the current implementation of EM, the starting value for μ_j is taken to be equal to the $(j/(k+1))$-th quantile of the observed data, the starting value for σ_j is half the empirical standard deviation of the data, and the starting value for w_j is $1/k$.

4.1.2 The Bayesian Approach: Markov Chain Monte Carlo (MCMC)

The Bayesian inference approach differs considerably from the frequentest MLE approach above. Instead of treading $(\mathbf{w}, \boldsymbol{\mu}, \boldsymbol{\sigma}^2)$ as deterministic and unknown, they are assumed stochastic (random) and given a *prior* distribution, $p(\mathbf{w}, \boldsymbol{\mu}, \boldsymbol{\sigma}^2)$, which summarizes in a probabilistic way our knowledge about the parameters before observing the current data $\mathbf{y}$. Parameter inference is in terms of the posterior distribution, the distribution of the parameters given the data,

$$p(\mathbf{w}, \boldsymbol{\mu}, \boldsymbol{\sigma}^2 \mid \mathbf{y}) = \frac{p(\mathbf{y} \mid \mathbf{w}, \boldsymbol{\mu}, \boldsymbol{\sigma}^2) p(\mathbf{w}, \boldsymbol{\mu}, \boldsymbol{\sigma}^2)}{p(\mathbf{y})}.$$

Only in the most trivial cases is one able to derive the posterior distribution in closed form, which is equivalent to being able to compute $p(\mathbf{y})$ above. And even if $p(\mathbf{y})$ is available in closed form and, hence the joint posterior distribution above, it can still be quite difficult to make inferences using the posterior distribution (e.g., computing the marginal distribution of μ_j, or just computing its posterior mean and variance). This has led to sampling-based techniques for Bayesian inference.

In the case of a mixture of Gaussians, a popular set of prior distributions (and natural in terms of conjugacy) are (see e.g. Robert, 1996):

$$\begin{aligned} w_j, &\sim \mathrm{Dir}(\gamma_j) \\ \mu_j &\sim \mathrm{Gau}(\xi_j, \tau_j^2 \sigma_j^2) \\ \frac{1}{\sigma_j^2} &\sim \mathrm{Gam}(\alpha_j, \beta_j), \quad j = 1, \ldots, k, \end{aligned} \tag{4.3}$$

where $\mathrm{Dir}(\cdots)$ denotes the Dirichlet distribution (with density proportional to $w_j^{\gamma_j - 1}$), Gau denotes the Gaussian distribution (with mean ξ_j and variance $(\tau_j \sigma_j)^2$), and Gam denotes the Gamma distribution (with mean and variance given by α/β and α/β^2). The parameters $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_k)$, $\boldsymbol{\tau}^2 = (\tau_1^2, \ldots, \tau_k^2)$, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_k)$, and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)$ are assumed known. Note that the prior variance of μ_j is given relative to σ_j^2, the variance of jth component of the Gaussian mixture. Another approach (see Richardson and Green, 1997) is to drop this dependence and let

$$\mu_j \sim \mathrm{Gau}(\xi_j, \tau_j^2), \quad j = 1, \ldots, k. \tag{4.4}$$

Note that the likelihood in (4.2) is invariant to relabeling of the groups. As such, the prior distribution of $(\mathbf{w}, \boldsymbol{\mu}, \boldsymbol{\sigma}^2)$ is often also made invariant to relabeling by not letting the prior parameters vary between groups:

$$w_j \sim \mathrm{Dir}(\gamma), \quad \mu_j \sim \mathrm{Gau}(\xi, \tau^2), \text{ and } \frac{1}{\sigma_j^2} \sim \mathrm{Gam}(\alpha, \beta).$$

Our implementation is that of Richardson and Green (1997) in the case of a fixed k (which was also used by Stephens, 2000). As in Richardson and Green (1997), the rate parameter β, associated with the prior distribution of $1/\sigma_j^2$ above, is not assumed fixed, but assumed random with its own prior distribution,

$$\beta \sim \mathrm{Gam}(g, h).$$

Inference for the Bayes model is done via the Markov Chain Monte Carlo method (MCMC). The full conditional distribution of all parameters is available and, hence all MCMC proposal distributions are Gibbs kernels (see Richardson and Green, 1997). Hyperparameters were fixed as in Richardson and Green (1997), with the assistance of the range and the median of the observed data (making the method partly empirical Bayes).

The output from MCMC is a sample from the posterior distribution. One drawback of the MCMC approach is that it does not yield (naturally) estimates of the weight, mean, and standard deviation of each "group", as both the prior and the likelihood are invariant to permutation of the group labels. That is, if one looks at the MCMC samples for say μ_1, the mean of the first "group", it typically "jumps" between the various modes seen in the data. Only when the different modes of the mixture are well separated and a large number of data is available does one get a "natural" separation in the MCMC output. There are sup-optimal methods to post-process the MCMC samples so the MCMC of each parameter is associated with one group label. These are based on ordering the MCMC samples according to, for example, the μ_j's of each iteration. Another approach is to constrain the MCMC sampler to yield samples of the μ_j which always satisfy $\mu_1 < \cdots < \mu_k$; our current implementation can perform both constrained and unconstrained MCMC sampling with respect to the μ_j.

4.2 Example

We shall now compare the EM and the MCMC estimation approach for artificial data generated from a known mixture of univariate Gaussians.

4.2.1 The Setup

With the final application in mind, seismic event location estimation, we shall generate data that could have potentially come from a seismic network.

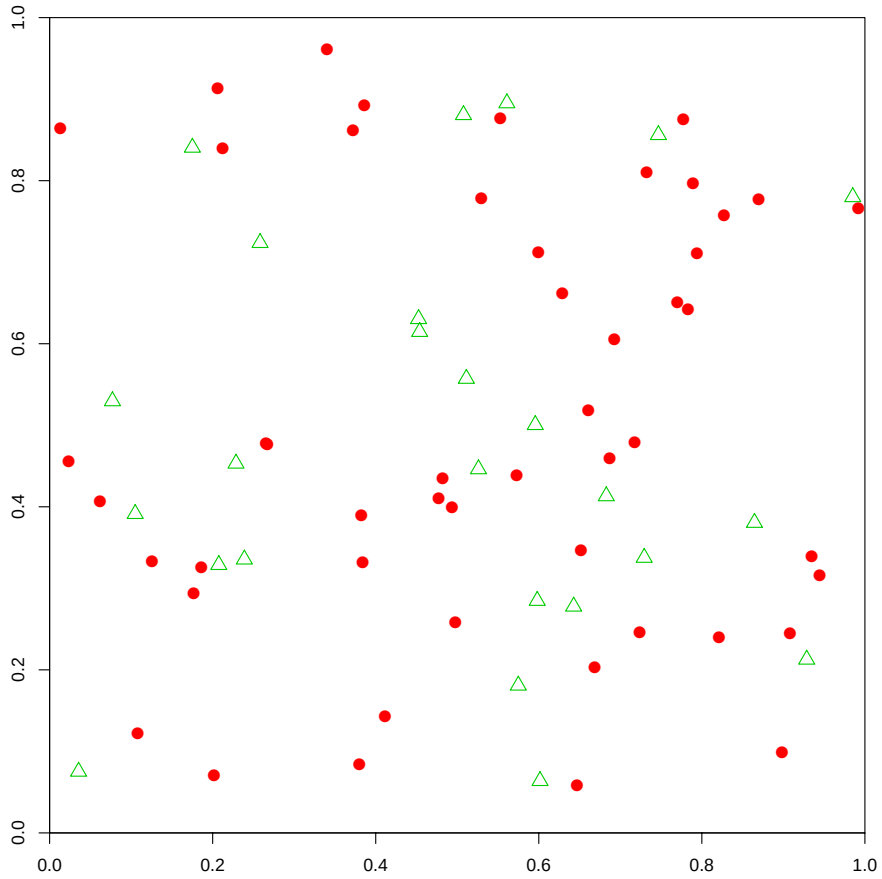

Figure 4.1: Location of stations (red circles) and events (green triangles).

Assume we have observations from m seismic events, $j = 1, \ldots, m$, from a network of stations, $i = 1, \ldots, n$, for a single seismic phase. We then have (see equation (2.1))

$$d_{ij} = t_j + T_i(\mathbf{x}_j) + c_{ij} + e_{ij} = t_j + T_i^*(\mathbf{x}_j) + e_{ij},$$

where t_j and $\mathbf{x}_j$ are the origin time and hypocenter of the jth event; d_{ij} is the measured arrival time of the jth event at the ith station; $T_i(\cdot)$ is the travel-time model for the ith station; c_{ij} is a travel-time corrections for the ijth path; and e_{ij} is the measurements error in d_{ij}. In this mini-example, we assume that the travel-time model and corrections are known, and given by

$$T_i^*(\mathbf{x}) = \|\mathbf{x} - \mathbf{s}_i\|^{1/2},$$

where $\mathbf{s}_i$ is the location of the ith station. The norm above is just the Euclid-norm in 2-D, so travel-time is simply set equal to the event-station distance. In this exercise, $n = 50$ and $m = 25$, and both the locations of the 50 stations and the 25 events were generated uniformly within $[0, 1] \times [0, 1]$ (see Figure 4.1). Similarly, the origin time of the 25 events were generated randomly within $[0, 1]$.

The 1,250 ($= 25 \cdot 50$) measurement errors, $\{e_{ij} : i = 1, \ldots, 50,\ j = 1, \ldots, 25\}$, were generated from a single 4-component Gaussian mixture distribution with the parameters listed in Table

Table 4.1: Assumed Mixture-of-Gaussian Parameters

	component (j)			
	1	2	3	4
mean (μ_j)	-0.01875	0.00000	0.01875	0.03749
st. dev. (σ_j)	0.00469	0.00469	0.00469	0.00469
prob. (w_j)	0.10000	0.50000	0.30000	0.10000

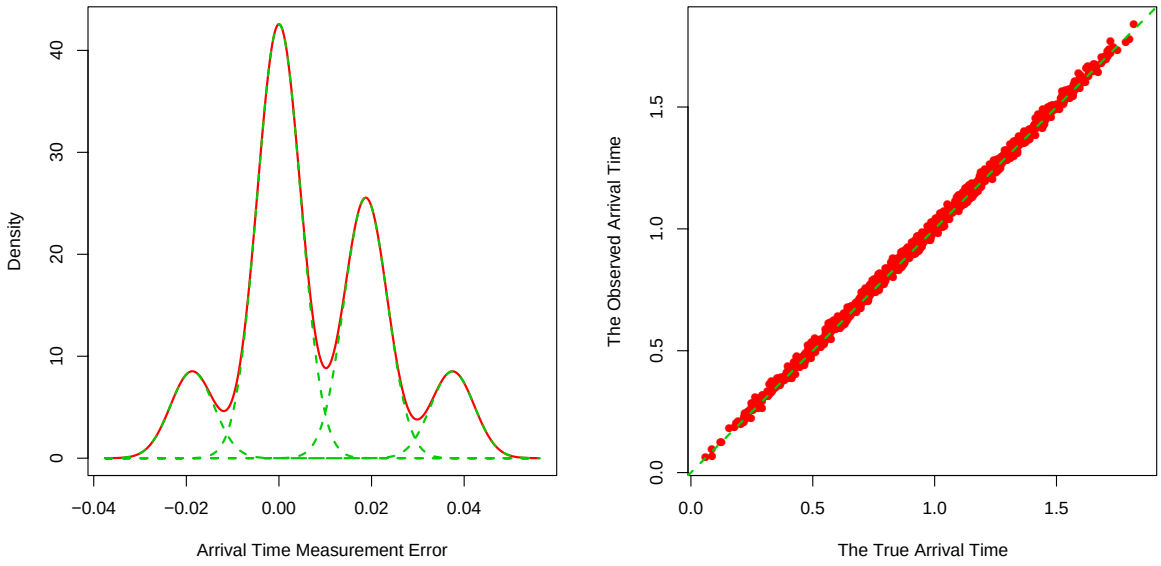

Figure 4.2: The measurement error mixture density (left) and a scatter plot of observed arrival time versus the true arrival time.

4.1. The standard deviation of each component is equal to 1% of the mean of the 1,250 travel times with a separation between adjacent components taken to be equal to 4 times the standard deviation. Component number two has mean zero and the largest weight and can be thought of as the error distribution associated with picking the right phase. The resulting mixture density and a plot of observed (generated) arrival times ($d_{ij} = t_j + T_i^*(\mathbf{x}_j) + e_{ij}$) versus the "true" arrival times ($d_{ij} = t_j + T_i^*(\mathbf{x}_j)$) are shown in Figure 4.2. Figure 4.3 shows then a histogram of the generated measurement errors $\{e_{ij}\}$ along with a superimposed line showing the true mixture density, a kernel density estimate, and a fitted (MLE) single-component Gaussian density.

4.2.2 Mixture Estimation via EM and MCMC

The EM algorithm was used to obtain MLEs of 3, 4, and 5 component Gaussian mixtures fitted to the measurement errors $\{e_{ij}\}$. Figure 4.4, similar to Figure 4.3, shows the MLEs of the three Gaussian mixture distributions. Figure 4.5, on the other hand, shows the Bayesian estimate of the 3, 4, and 5 component Gaussian mixture in a similar way to Figure 4.4. Both the EM

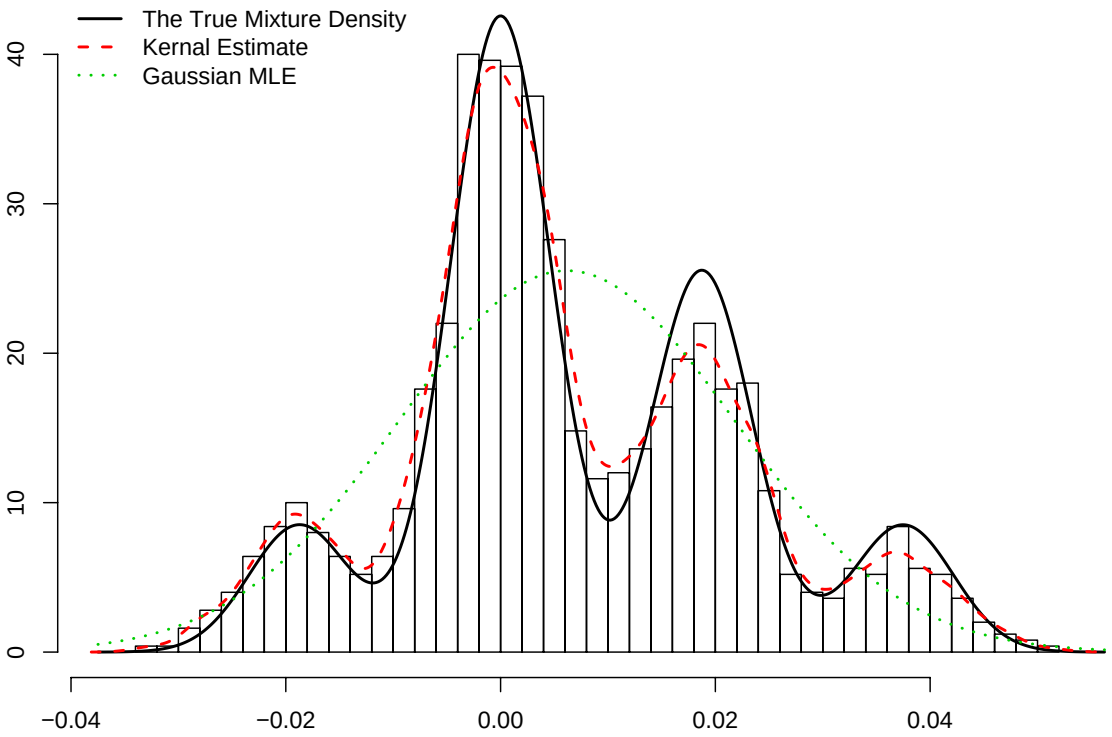

Figure 4.3: Histogram of the generated measurement errors along with the true mixture density (black, solid line), a kernel density estimate (red, dashed line), and a fitted (ML) Gaussian density (green, dotted line).

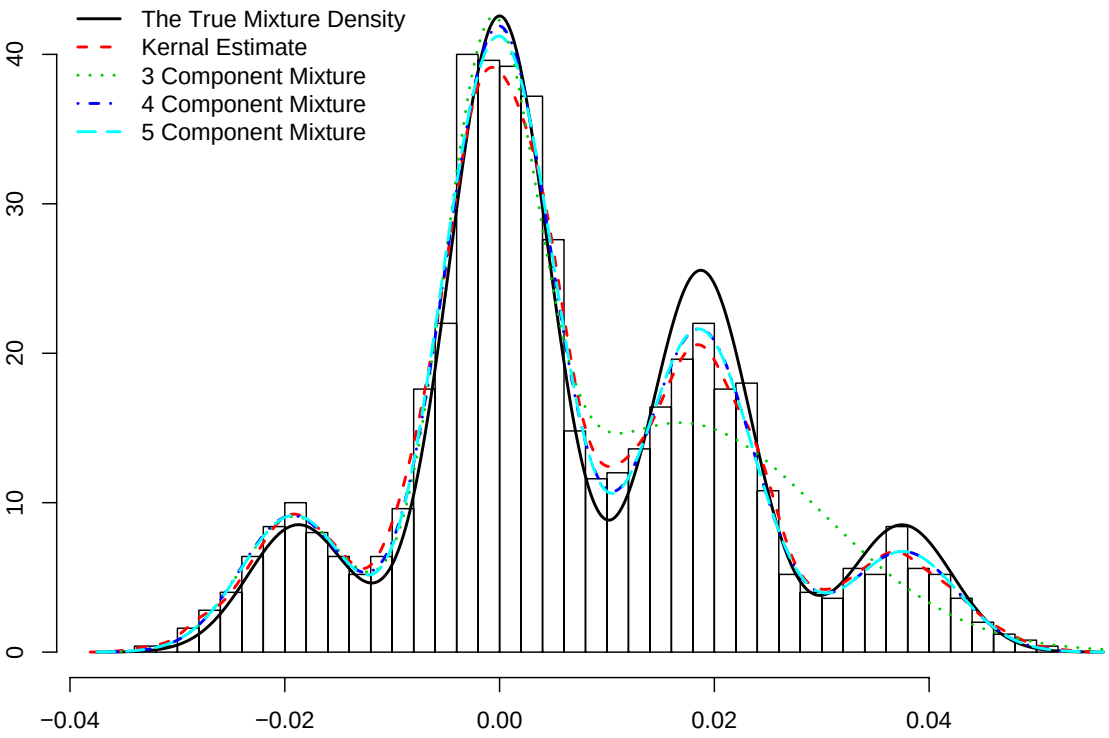

Figure 4.4: Histogram of the generated measurement errors along with the true mixture density (black, solid line), a kernel density estimate (red, dashed line), and the MLE via EM of a Gaussian mixture density with 3, 4, and 5 components.

algorithm and the MCMC were run for 10,000 iterations (or rather full sweeps through the parameters).

In computing the MCMC estimates shown in Figure 4.5, only sweeps 2,000 to 10,000 were used. The MCMC estimates are computed using

$$p(e \mid \text{data}) = \frac{1}{8,001} \sum_{i=2,000}^{10,000} \sum_{j=1}^{k} w_j^{(i)} \varphi(y \mid \mu_j^{(i)}, \sigma_j^{2(i)}), \tag{4.5}$$

where data $= \{e_{ij}\}$ and e is a new measurement error. The above expression is an approximation of the predictive measurement error density

$$p(e \mid \text{data}) = \int p(e \mid \mathbf{w}, \boldsymbol{\mu}, \boldsymbol{\sigma}^2) p(\mathbf{w}, \boldsymbol{\mu}, \boldsymbol{\sigma}^2 \mid \text{data}) d\mathbf{w} d\boldsymbol{\mu} d\boldsymbol{\sigma}^2,$$

where $p(\mathbf{w}, \boldsymbol{\mu}, \boldsymbol{\sigma}^2 \mid \text{data})$ is the posterior distribution of the mixture parameters. The predictive measurement error density above summarizes what we know about the density of the measurement error given the observed data (and would be used for a new event).

The MLEs via EM in Figure 4.4 are simply given by $\sum_{j=1}^{k} \hat{w}_j \varphi(y \mid \hat{\mu}_j, \hat{\sigma}_j^2)$, where $\hat{w}_j$, $\hat{\mu}_j$, and $\hat{\sigma}_j^2$ are the final values of the EM algorithm.

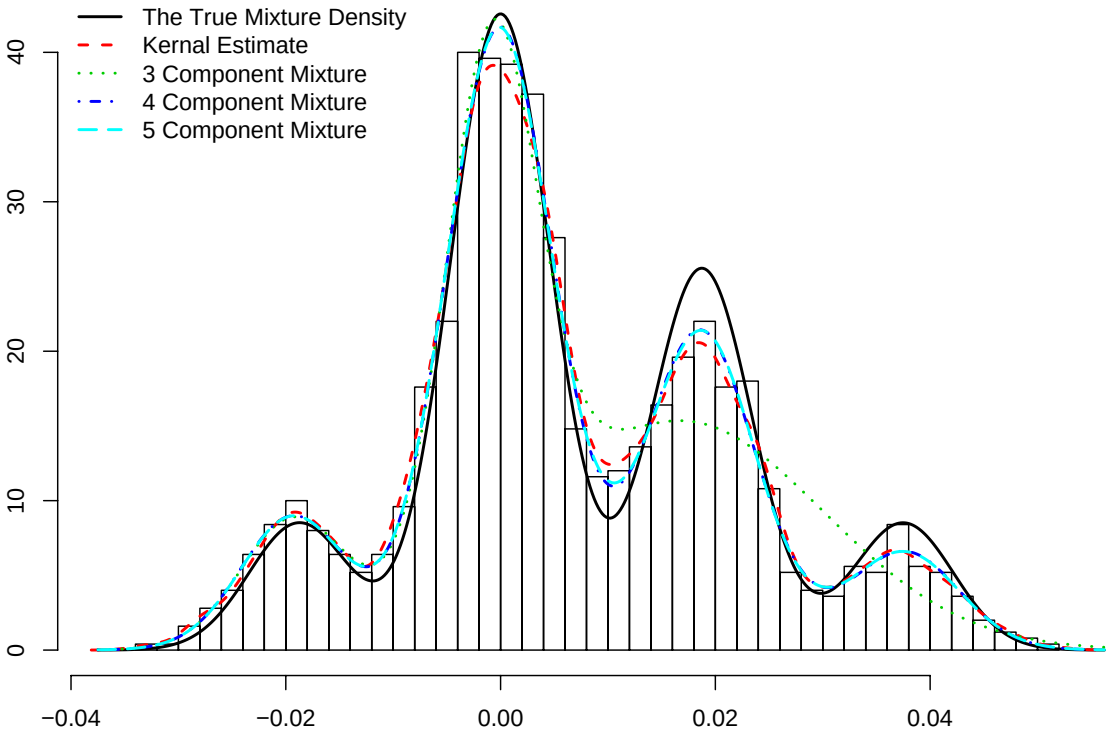

Figure 4.5: Histogram of the generated measurement errors along with the true mixture density (black, solid line), a kernel density estimate (red, dashed line), and a Bayesian estimate (via MCMC) of a Gaussian mixture density with 3, 4, and 5 components.

Due to the abundance of data (1,250 observations), the MLEs and the Bayes estimate look nearly identical. This is due to the fact that the prior distributions used in the Bayes approach are rather non-informative (flat) and provide little information in addition to what is seen in the data.

4.2.3 Bayesian Estimation of Seismic Event Location and Time of Origin

Here we report the results of a simple study to analyze the impact of mis-specifying the measurement error distribution in a Bayesian setting.

Assume that the measurement error distribution of a seismic event arrival time is known at all stations. We shall set up a Bayesian model for estimating the location of a (new) seismic event and its time of origin using a MCMC sampler.

Given the same network of 50 stations as shown in Figure 4.1, assume that

$$d_i = t + T_i^*(\mathbf{x}) + e_i, \quad i = 1, \ldots, 50,$$

where $(\mathbf{x}, t)$ is the new seismic event location and time of origin, $T_i^*(\mathbf{x})$ is the travel time to the ith station, and e_i is the measurement error at the ith station (coming from a known measurement error distribution).

The unknown parameters of the model are only $(\mathbf{x}, t)$, which we give the following prior distributions:

$$\mathbf{x} \sim \text{Unif}([0,1] \times [0,1]) \text{ and } t \sim \text{Unif}([0,1]).$$

To sample from the posterior distribution of $(\mathbf{x}, t)$ we simply use a random-walk Metropolis sampler. That is, given that the chain is at $(\mathbf{x}^{(i)}, t^{(i)})$ at step i, a new state, $(\mathbf{x}', t')$, is proposed via,

$$(\mathbf{x}', t') = (\mathbf{x}^{(i)} + \boldsymbol{\epsilon}_x, t^{(i)} + \epsilon_t), \quad \boldsymbol{\epsilon}_x \sim \text{Gau}(\mathbf{0}, \sigma_x^2 \mathbf{I}) \text{ and } \epsilon_t \sim \text{Gau}(0, \sigma_t^2).$$

Then u is drawn form Unif$([0,1])$ and the next state of the chain, at step $i+1$, is given by

$$(\mathbf{x}^{(i+1)}, t^{(i+1)}) = \begin{cases} (\mathbf{x}', t') & \text{if } u \leq \frac{L(\mathbf{x}', \mathbf{t}')}{L(\mathbf{x}^{(i)}, t^{(i)})} \\ (\mathbf{x}^{(i)}, t^{(i)}) & \text{otherwise,} \end{cases} \tag{4.6}$$

where $L(\cdot)$ is the likelihood. When the measurement error distribution is assumed to be given by a 4 component Gaussian mixture, L is given by

$$L(\mathbf{x}, t) = p(\mathbf{d} \mid \mathbf{x}, t) = \prod_{i=1}^{50} \sum_{j=1}^{4} w_j \varphi(d_i - (t + T_i^*(\mathbf{x})) \mid \mu_j, \sigma_j^2). \tag{4.7}$$

The variance parameters σ_x^2 and σ_t^2, associated with the random-walk Metropolis sampler, were tuned such that the new proposed state was accepted between 30% and 40% of the time in (4.6).

In this study, an event at $\mathbf{x} = (1/2, 1/2)$ was assumed to have occurred at $t = 1/2$. Exact travel-times were then computed as $T_1^*(\mathbf{x}), \ldots, T_{50}^*(\mathbf{x})$ and arrival time data $d_1, \ldots, d_{50}$ generated by adding randomly generated measurement noise from the 4-component mixture distribution used previously (see Figure 4.6)

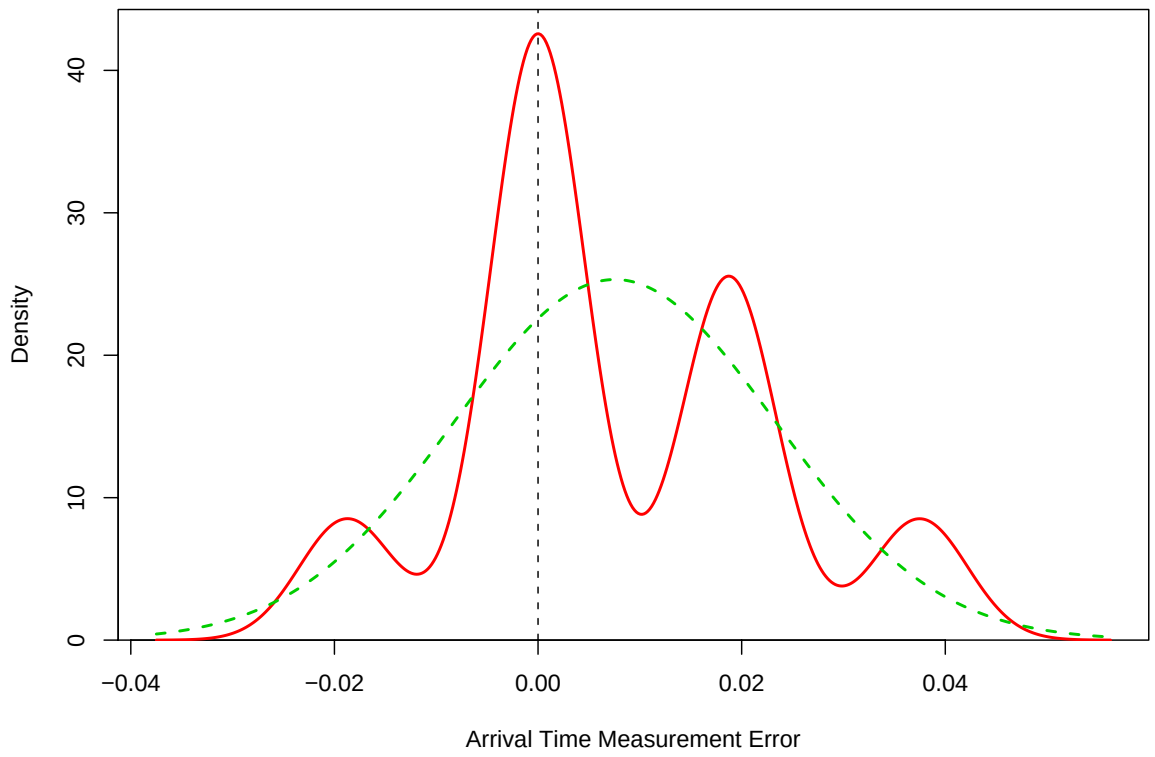

Figure 4.6: The true mixture distribution of the arrival time measurement errors along with a simple (one-component) Gaussian approximation.

We compared two Bayesian solutions for the event parameters, $\mathbf{x}$ and t. In the first, the correct measurement error distribution was used, i.e. the 4-component mixture (i.e., the likelihood is as in (4.7)). In the second, an incorrect measurement error distribution was used, i.e. a single-component Gaussian distribution. In the latter case, the likelihood is given by

$$L(\mathbf{x}, t) = p(\mathbf{d} \mid \mathbf{x}, t) = \prod_{i=1}^{50} \varphi(d_i - (t + T_i^*(\mathbf{x})) \mid \mu_u, \sigma_u^2), \tag{4.8}$$

where μ_u and σ_u^2 were selected to match the mean and variance of the composite, 4-component mixture distribution as well as possible; see Figure 4.6. Both solutions (using correct vs. incorrect error distribution) were obtained with an MCMC of length 10,000.

Figures 4.7 and 4.8 show the trace (the sample path) of $\mathbf{x} = (x, y)$ and t in the case of the Gaussian approximation and the 4-component mixture, respectively. A burn-in of 2,000 seems to be sufficient for both cases.

Figure 4.9 shows a scatter plot of the MCMC sampled event locations, $\{\mathbf{x}^{(i)} : i = 2,001, \ldots, 10,000\}$, for the two cases along with the exact location of the event ($\mathbf{x}_0 = (1/2, 1/2)$). As is clear from Figure 4.9, even though the Gaussian approximation does not seem to be too far off from the 4-component mixture (see Figure 4.6), the final impact on the posterior distribution of $\mathbf{x}$ is severe. The MCMC event location samples, when assuming a simple Gaussian error distribution, have

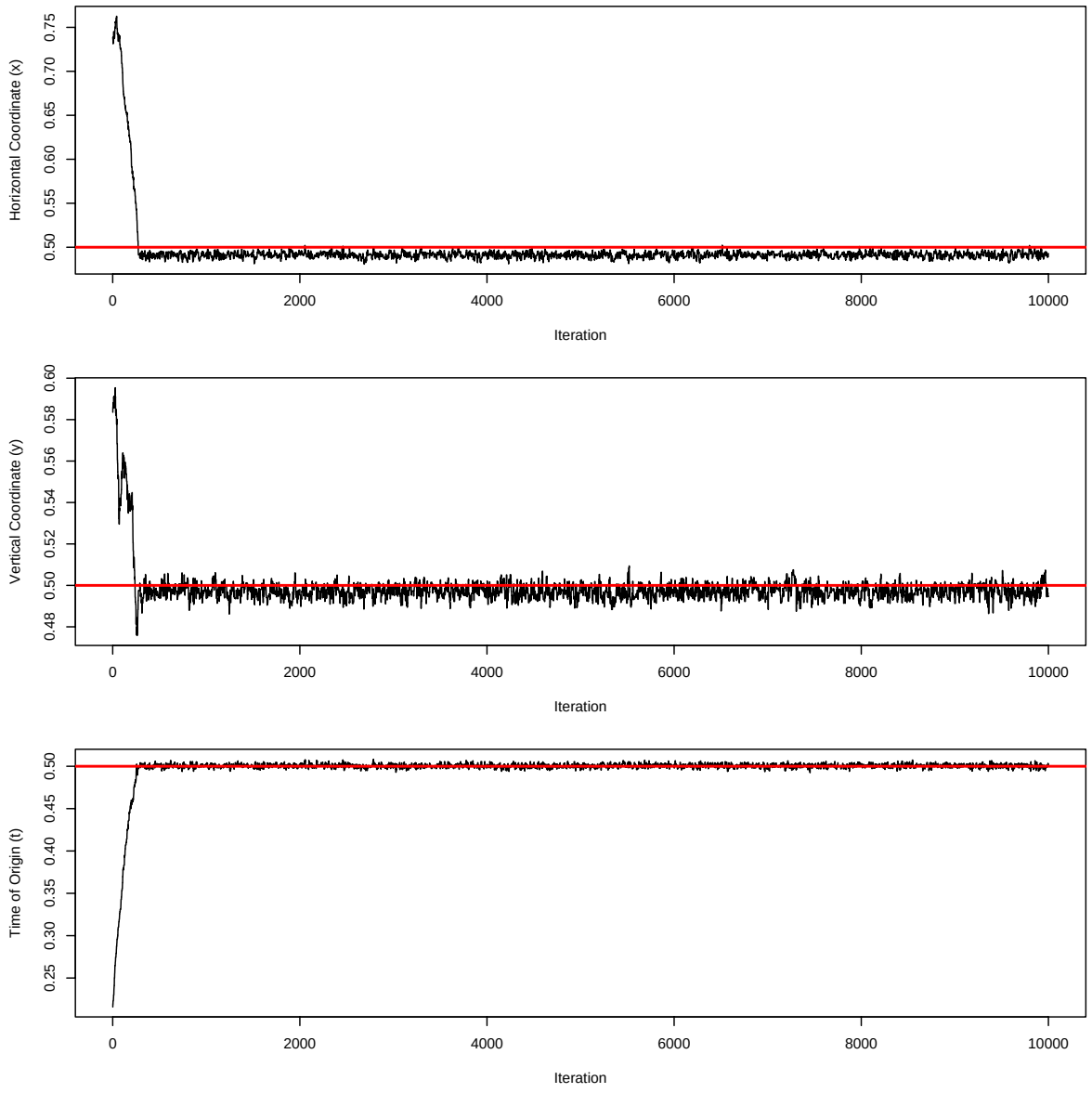

Figure 4.7: The generated MCMC samples for the three parameters $\mathbf{x} = (x, y)$ and t in the case of a *simple* (one-component) Gaussian measurement error distribution. The red horizontal lines show the true value of each parameter (equal to 1/2 in all cases).

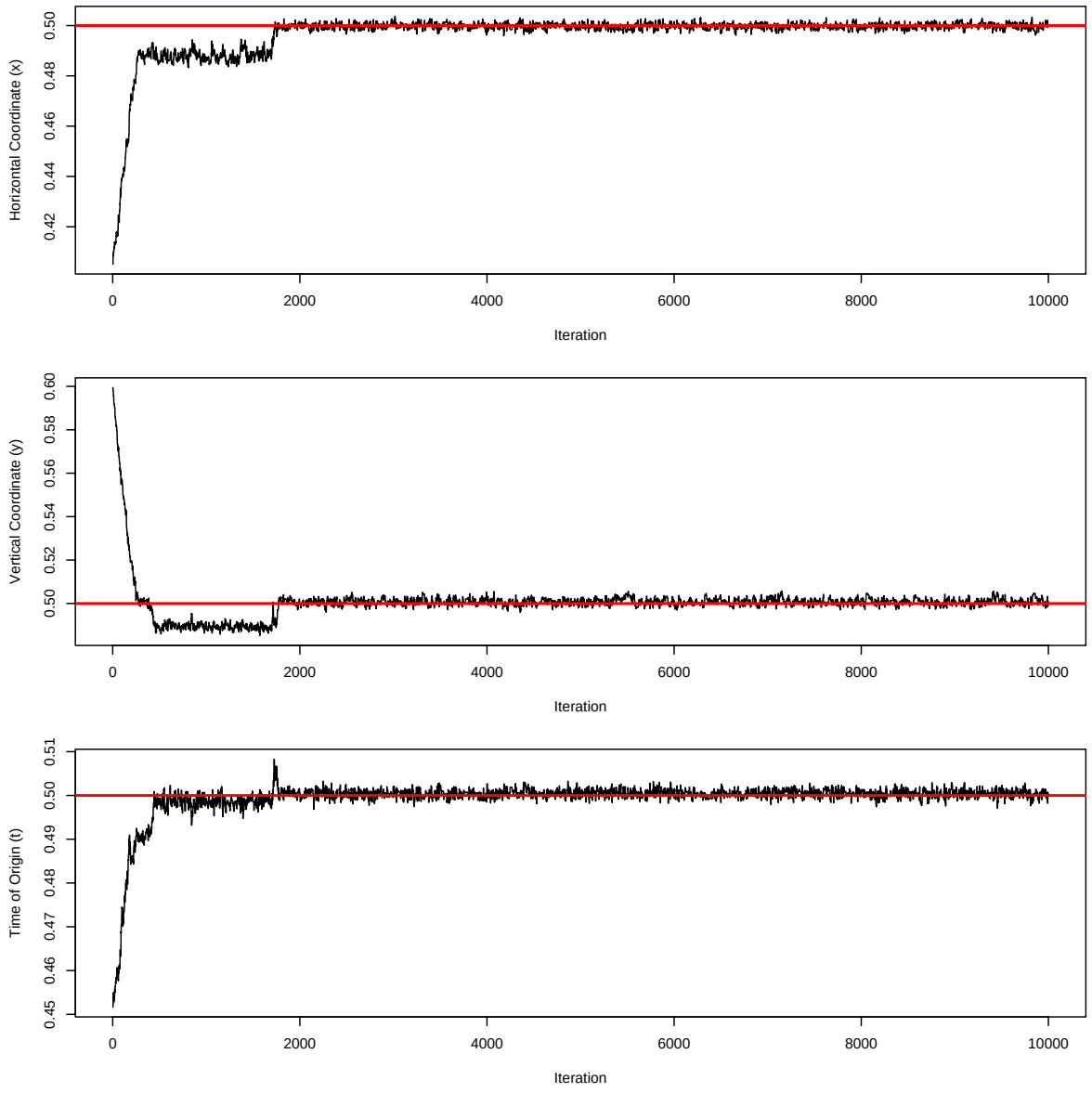

Figure 4.8: The generated MCMC samples for the three parameters $\mathbf{x} = (x, y)$ and t in the case of a *4-component* mixture-of-Gaussians measurement error distribution. The red horizontal lines show the true value of each parameter (equal to 1/2 in all cases).

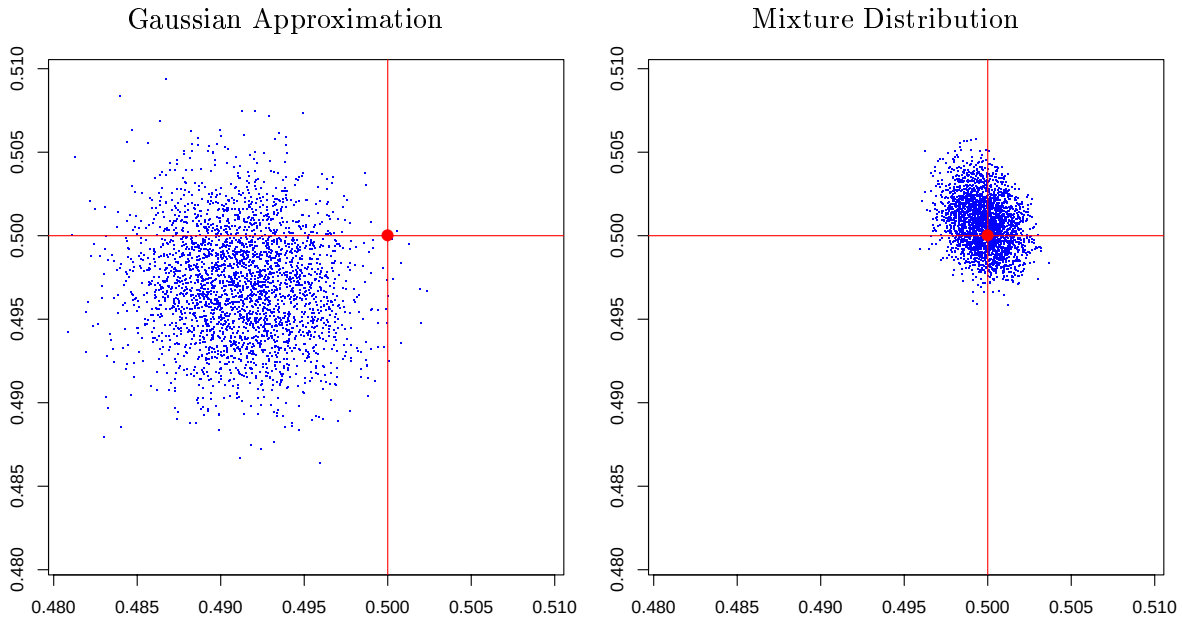

Figure 4.9: The MCMC sampled event locations, $\{\mathbf{x}^{(i)} : i = 2,001, \ldots, 10,000\}$, when assuming a Gaussian measurement error distribution (left) and when assuming a 4-component mixture measurement error distribution (right). The true location of the event is shown at $(1/2, 1/2)$.

a considerably larger spread (i.e., variance). Additionally, they show the actual location to be rather unlikely (i.e., bias). The expected distance to the true location, $\mathbf{x}_0 = (1/2, 1/2)$, given by $\int \|\mathbf{x} - \mathbf{x}_0\|^{1/2} p(\mathbf{x} \mid \text{data}) d\mathbf{x}$, is 0.0018 when using the mixture-of-Gaussians distribution, versus 0.0099 when using the simple Gaussian distribution.

Figure 4.10 shows histograms of the MCMC sampled event time, $\{t^{(i)} : i = 2,001, \ldots, 10,000\}$, for the two cases along with the exact origin of the event ($t = 1/2$). In both cases the true event time is captured by the posterior, but the spread of the posterior in the case of a mixture distribution is considerably smaller than in the case of a Gaussian distribution.

We can conclude from this simple numerical experiment that mis-specifying the arrival time measurement error distribution can have severe consequences for the estimation of the location and origin time of a seismic event.

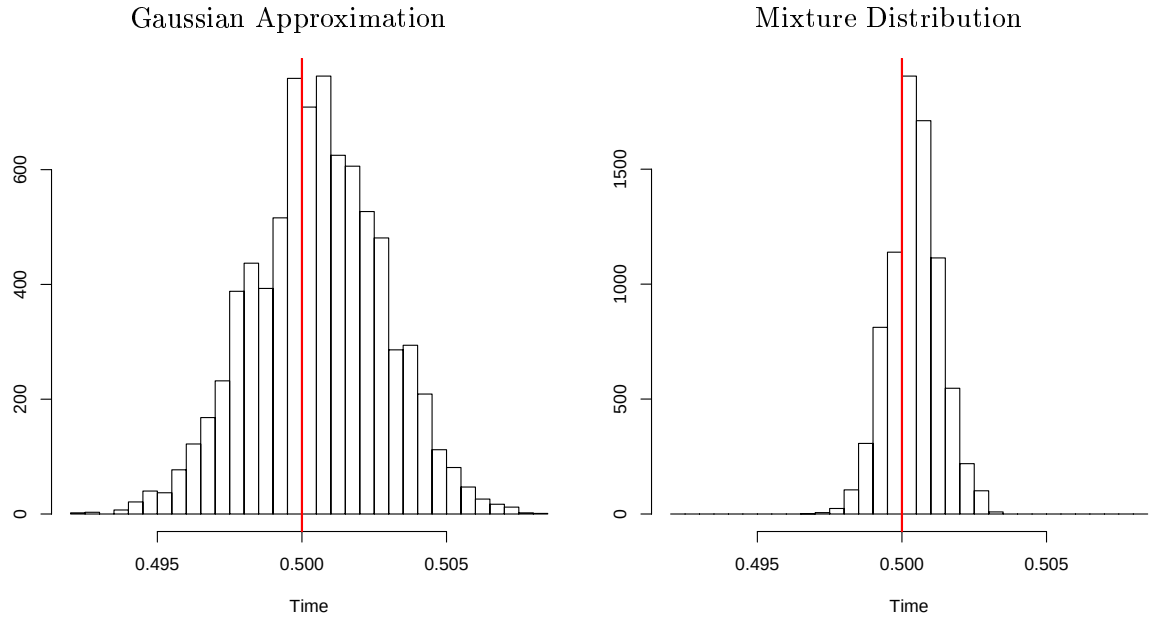

Figure 4.10: The MCMC sampled event origin times, $\{t^{(i)} : i = 2,001, \ldots, 10,000\}$, when assuming a Gaussian measurement error distribution (left) and when assuming a 4-component mixture measurement error distribution (right). The true origin of the event is shown at $t = 1/2$.

References

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

Buland, R. (1986), Uniform reduction error analysis, *Bull. Seism. Soc. Am.,* 76, 217–230.

Dempster, A.P., N.M. Laird, and D.B. Rubin (1977). Maximum likelihood from incomplete data via the em algorithm (with discussion), *J. Roy. Statistical Soc. B, 39*, 1–38.

Deutsch, C.V. and A.G. Journel (1998). *GSLIB: Geostatistical Software Library and User's Guide,* 2nd ed., Oxford University Press, Inc., New York, 369 pp.

Dewey, J.W. (1971). Seismicity studies with the method of joint hypocenter determination, *Ph.D. Thesis,* University of California, Berkeley.

Dreger, D., R. Uhrhammer, M. Pasyanos, J. Franck and V. Romanowicz (1998), Regional and far-regional earthquake locations and source parameters using sparse broadband networks: a test on the Ridgecrest sequence, *Bull. Seism. Soc. Am.,* 88, 1353–1362.

Engdahl E.R., R. van der Hilst and R. Buland (1998). Global teleseismic earthquake relocation with improved travel times and procedures for depth determination, *Bull. Seism. Soc. Amer., 88,* 722–743.

Engdahl, E.R. and E.A. Bergman (2001), Validation and generation of reference events by cluster analysis, *Proceedings,* 23rd Annual Seismic Research Review, Jackson Hole, Wyoming.

Evernden, J.F. (1969). Identification of earthquakes and explosions by use of teleseismic data, *J. Geophys. Res., 74*, 3828–3856.

Jeffreys, H.. (1932). An alternative to the rejection of observations, *Mon. Not. R. Astr. Soc, Geophys. Suppl. 2,* 78–87.

Jordan, T.H. and K.A. Sverdrup (1981). Teleseismic location techniques and their application to earthquake clusters in the south-central Pacific, *Bull. Seism. Soc. Am., 71*, 1105–1130.

Myers, S.C., D.B. Harris, M.L. Anderson, W.R. Walter, M.P. Flanagan and F. Ryall (2003). LLNL location and detection research, *Proceedings*, 25th Seismic Research Review, Tucson, Arizona, this volume.

Pavlis, G.L, and J.R. Booker (1983). Progressive multiple event location (PMEL), *Bull. Seism. Soc. Am., 73,* 1753–1777.

Reiter, D., W. Rodi and M. Johnson (2005). Development of a tomographic upper mantle velocity model beneath Pakistan and northern India for improved regional seismic event location, *Bull. Seism. Soc. Am.,*, in press.

Richardson, S. and P.J. Green (1997). On Bayesian analysis of mixtures with an unknown number of components, *J. Roy. Statistical Soc. B, 59*, 731–792.

Robert, C.P. (1996). Mixture of distributions: inference and estimation, In W.R. Gilks, S. Richardson, and D.J. Spiegelhalter (Eds.), *Markov Chain Monte Carlo in Practise* (pp. 441–464). London: Chapman and Hall.

Rodi, W. and M.N. Toksöz (2000). Grid-search techniques for seismic event location, *Proceedings,* 22nd Annual Seismic Research Review, New Orleans.

Rodi, W. and M.N. Toksöz (2001). Uncertainty analysis in seismic event location, *Proceedings,* 23rd Annual Seismic Research Review, Jackson Hole, Wyoming, Defense Threat Reduction Agency.

Rodi, W., E.R. Engdahl, E.A. Bergman, F. Waldhauser, G.L. Pavlis, H. Israelsson, J.W. Dewey and M.N. Toksöz (2002). A new grid-search multiple-event location algorithm and a comparison of methods, *Proceedings,* 24th Annual Seismic Research Review, Ponte Vedra Beach, Florida, 403–411.

Schultz, C.A., S.C. Myers, J. Hipp and C.J. Young (1998). Nonstationary Bayesian kriging: a predictive technique to generate spatial corrections for seismic detection, location, and identification, *Bull. Seism. Soc. Am., 88,* 1275–1288.

Stephens, M. (2000). Bayesian analysis of mixture models with an unknown number of components—an alternative to reversible jump methods, *The Analysis of Statistics, 28*, 40–74.

Tarantola, A. (1987). *Inverse Problem Theory.* Elsevier.

Waldhauser, F. and W.L. Ellsworth (2000). A double-difference earthquake location algorithm: method and application to the northern Hayward fault, California, *Bull. Seism. Soc. Am., 90,* 1353–1368.

Walter, W.R., K. D. Smith, J. L. O'Boyle, T. F. Hauk, F. Ryall, S.D. Ruppert, S.C. Myers, M. Anderson, and D.A. Dodge (2003). Improving the fundamental understanding of regional seismic signal processing with a unique western United States dataset, *Proceedings*, 25th Seismic Research Review, Tucson, Arizona, this volume.