# MULTIPLE-EVENT LOCATION USING THE MARKOV-CHAIN MONTE CARLO TECHNIQUE

Stephen C. Myers, Gardar Johannesson, and William Hanley

Lawrence Livermore National Laboratory

Sponsored by National Nuclear Security Administration
Office of Nonproliferation Research and Engineering
Office of Defense Nuclear Nonproliferation

Contract No. W-7405-ENG-48

## ABSTRACT

The goal of next-generation seismic location is to ascertain a consistent set of event locations and travel-time corrections through simultaneous analysis of all relevant data. Towards that end, we are developing a new multiple-event location algorithm that utilizes the Markov-Chain Monte Carlo (MCMC) method for solving large, non-linear inverse problems. Unlike most inverse methods, the MCMC approach produces a suite of solutions, each of which is consistent with seismic and other observations, as well as prior estimates of data and model uncertainties. In the MCMC multiple-event locator (MCMCloc), the model uncertainties consist of prior estimates on the accuracy of each input event location, travel-time prediction uncertainties, phase measurement uncertainties, and assessments of phase identification. The prior uncertainty estimates include correlations between travel-time predictions, correlations between measurement errors, and the probability of misidentifying one phase for another (or bogus picks). The implementation of prior constraints on location accuracy allows the direct utilization of ground-truth events in the location algorithm. This is a significant improvement over most other grid multiple-event locators (GMEL is an exception), for which location accuracy is achieved through post-processing comparisons with ground-truth information. Like the double-difference algorithm, the implementation of a correlation structure for travel-time predictions allows MCMCloc to operate over arbitrarily large geographic areas. MCMCloc can accommodate non-Gaussian and multi-modal pick distributions, which can enhance application to poorly recorded events. Further, MCMCloc allows for ambiguous determination of phase assignments, and the solution includes the probability that phases are properly assigned. The probabilities that phase assignments are correct are propagated to the estimates of all other model parameters. Posteriori estimates of event locations, path corrections, pick errors, and phase identifications are made through analysis of the posteriori suite of acceptable solutions.

We test the MCMC locator on a regional data set of Nevada Test Site nuclear explosions. Event locations and origin times are known for these events, allowing us to tests the features of MCMCloc against a true ground truth (GT0) data set. Preliminary tests suggest that MCMCloc provides excellent relative locations (similar to other algorithms), and excellent absolute locations when constraints from one or more ground truth event are included. Tests also include realistic phase misidentification, where phase assignments are switched for phases that arrive within a few seconds of one another. We find that MCMCloc is a promising method for simultaneously locating large, geographically distributed data sets. Because we allow for input of prior knowledge on many aspects of the data set, MCMCloc is ideal for combining trusted and lesser-quality data.

## OBJECTIVES

Traditionally, multiple event locators simultaneously determine the optimal, relative location for a geographically clustered set of events (e.g., Douglas, 1967; Dewey, 1971; Jordan and Sverdrup, 1981; Pavlis and Booker, 1981). These methods work best when there is substantial overlap in the network of stations that recorded each event and inter-events distances are short relative to event/station distances. These conditions allow for robust estimation of a travel-time correction for each station/phase pair and the determination of relative event locations. In a recent effort Waldhauser and Ellsworth (2000) expand the applicability of multiple-event methods to larger geographic regions by allowing for spatially varying travel-time corrections. Myers and Schultz (2000) also use spatially varying travel-time corrections but with a modeled correlation structure, thus approximating a multiple-event result with a single-event algorithm. Each method has it advantages and challenges. The HypoDD method of Waldhauser and Ellsworth (2000) employs an ad hoc correlation structure, so the transition between relative and absolute location is based on a static preset parameter. Also, the location estimates are determined by iteratively inverting a system of linear equations, requiring approximately exponential growth in computational power as the number of observations grows. Therefore, application to very large problems can become computationally restrictive. In the case of travel-time corrections based on Bayesian Kriging, Myers and Schultz, 2000 demonstrate that single-event locations can closely match multiple event estimates. Because the application is through a single-event location algorithm, application to large data sets is “embarrassingly” parallel. However, development of a self-consistent set of correction surfaces for each station and phase is time intensive for the analyst. Also, the correction surface method is not applicable to the local-distance case, where slowness changes rapidly with event depth.

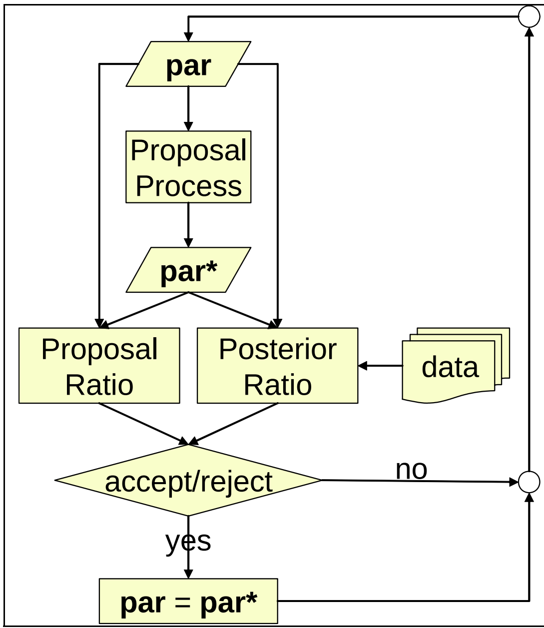
In this study we introduce a new multiple-event method that uses the Markov-Chain Monte Carlo (MCMC) method for solving large inverse problems. The new method (MCMCloc) allows us to simultaneously estimate event location, travel-time correction, pick precision, and phase identification. We present a brief outline of the forward problem, the statistical framework for assessing data fit, and a “first cut” example application to the Nevada Test Site (NTS) data set of nuclear explosions with known source parameters.

## RESEARCH ACCOMPLISHED

### The MCMC method and its applicability to seismic location

In single-event location, where the full space of potential locations can be interrogated, grid-search and related methods have a proven record of performance (Billings et al., 1994; Rodi et al., 2003, Lomax, 2005). However, for multiple-event location the full-combination of potential solutions quickly out strips even modern computational capabilities. Markov Chain Monte Carlo (MCMC) is particularly well suited to characterize (albeit not fully explore) the multiple-event solution space, because it is a “smart” search through a potentially enormous space, and the efficiency of the search can be further improved by taking advantage of prior information about each event location (and other parameters).

The MCMC approach, while not new to geophysics (e.g., Shapiro and Ritzwoller, 2002; Pasyanos et al., submitted), has not to our knowledge been used in multiple event locations. General overviews of MCMC can be found in, for example, Gilks et al. (1996) and Gelman et al. (2004). At its core, the MCMC technique is a Markovian proposal process, meaning that a new parameter configuration (i.e., hypocenters, arrival-time measurement errors, travel-time corrections, and phase assignments) is made by modifying the current configuration. The series of configurations is referred to as a chain. The new configuration is, by design, consistent with prior parameter constraints. A configuration is kept as a potential solution if predictions based on that configuration (e.g., travel-time predictions) are within prescribed uncertainties of observed data (Figure 1). The MCMC approach is distinct from traditional (linearized) inversion techniques that return a point for each location with an associated confidence ellipsoid. The MCMC solution for each event location is a cloud of points that defines a probability density function. The “best” or “most probable” location and a confidence region can be derived from the non-linear probability distribution.




**MCMC** algorithm:
The current parameter configuration (**par**) is modified via a (probabilistic) proposal process to yield a new parameter configuration (**par***). Depending on the fidelity of the new parameter configuration relative to data and the prior constraints on the parameters, the new configuration is either accepted or rejected.

**Figure 1. Flow chart depicting the proposal and assessment of parameter configurations in the MCMC inversion procedure.**

## Formulation of the MCMCloc forward problem

We have developed a new location algorithm (MCMCloc) with the MCMC solver at its core. In MCMCloc the seismic location forward problem is by dividing into 3 components.

1. Hypocenter-Model: Conditional distribution of event location parameters given prior estimates and resulting fit to data.

2. Data-Model: Conditional distribution on the observed data given the true configuration of arrival-times and phase assignments (akin to pick error but now includes phase assignments).

3. Prediction-Model: Conditional distribution for the true arrival-times and phase assignments given a particular configuration of event locations (path corrections).

Component (1) is the traditional location inversion problem, allowing for constrains on hypocenter parameters. Any hypocenter component of any event can be constrained using any specified probability density function. The default is a flat prior (no constraint). Component (2) is a model of observational error. We extend traditional assessments of observation error – random measurement or "pick" error – to include the chance that phase assignments may be incorrect and the chance that any datum may be altogether bogus. Component (3) accounts for errors in travel-time prediction. The model for travel-time predictions can be quite general for MCMCloc, but in this initial implementation it is simply a static adjustment to the travel-time curve for each phase.

MCMCloc iteratively modifies each *component* of the forward model (as defined above). That is, MCMCloc alternates between proposing a new *component* of the solution. Therefore, MCMCloc will propose one of the following: (a) a new hypocenter model, (b) a new data model, (c) or a new travel time prediction model. Choosing a succession of components that are modified can greatly speed convergence.

In order to ensure coverage of the solution space, multiple chains are generated independently. Each chain is run for an initial burn-in period after which the chain continues for additional adaptive training that fine-tunes the proposal process. The multiple chains are merged to form a global distribution.

**Example Application**

We relocated 9 nuclear tests at the Nevada Test Site (NTS) (Walter et al., 2003) for demonstration purposes (Figure 2). The hypocenter is known for each event, providing unambiguous assessment of location accuracy. For this demonstration we locate using Pn, Pg, and Lg arrival times. The number of observed phases for each event ranges from 7 (3 stations) to 23 (8 stations), with a total of 128 arrivals for the 9 events.

The prior distribution for the hypocenter parameters (lat, lon, depth, time) was taken to be multivariate normal, with depth log-transformed. The travel-time model is IASP91 (Kennett and Engdahl, 1991), which is known to result in location bias of ~5km for the NTS explosion data set (Anderson and Myers, 2005). The travel-time error-model is, in this instance, very simple: a travel-time shift for each phase (with a normal prior on the travel-time shift parameters). The travel-time residual model (pick error) was taken to be normal, with independent variance for each station and phase. More specifically, log[Var(i,j,k)] = A(k) + B(j), where Var(i,j,k) is the variance associated with the i-th event, the j-th station, and the k-th phase, and A(k) and B(j) are phase- and station-specific parameters, respectively.

We consider three test cases:

1. A vague prior on all parameters and phase-assignments assumed correct (standard relative location).
2. A strong (narrow) prior for the hypocenter parameters of two events and phase-assignments assumed correct (relative location with two fixed locations).
3. A strong (narrow) prior for the location parameters of two events, phase-assignments *not* assumed correct. In this instance, three Pn/Pg phase-assignments were switched and 10 sec was added to another three Pg arrival times (i.e., corrupted data).

*Case 1*

Figure 3 shows a map of epicenter posterior distributions for the 9 events. For well-observed events, the posterior distribution is relatively tight around the true location. As expected, for poorly observed events the distribution is wider and the mean of the distribution is not centered on the known location. However, it is important to note that the distribution still covers the true event location for the most poorly observed events. The “mass” of the origin time posterior distributions was generally within 1 second of the true origin-time (not shown). Instances with broad origin posterior distributions still covered the known value. Posterior distributions for event depth were broad (not shown), owing to inherent uncertainties. The depth distributions, although broad, included the true depth in all cases. The phase-specific variance parameter was largest for Lg and smallest for Pn. As a result Pn arrivals get considerable more weight that Lg arrivals. In other words, the uncertainty in measuring the Lg arrival overwhelms the increased sensitivity due to the slow Lg velocity. Lastly, we find a considerable variation in the station-specific variance parameters (pick error), which is likely the result of better quality signals at some stations that results in high-precision arrival-time measurements.

For comparison purposes, we show conventional relative locations in Figure 4. Conventional relative methods assume a static travel-time adjustment for each station/phase pair. Unfortunately, addition of the station/phase parameters comes at the cost of lost resolution in absolute location accuracy, and in this case epicenters are displaced to the east of the known locations. Our choice of a simple travel-time adjustment is, in this instance, beneficial because location accuracy is maintained. In the near future, we plan to extend the simple travel-time correction model to include station/phase terms; however, our plan is for a hierarchic approach in which adjustments to the travel-time curve are made first, followed by station/phase specific adjustments. The station/phase terms will include spatial correlation, which will allow application over an arbitrarily large area. Using this general approach we aim to maintain as much accuracy as possible, while introducing the terms that improve relative location precision.

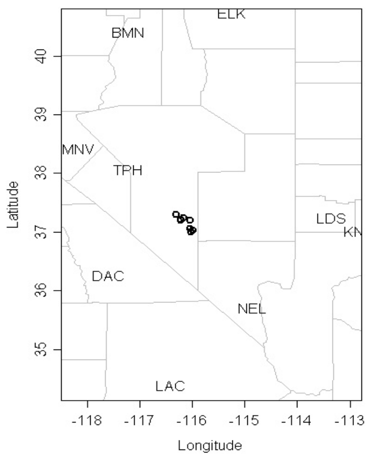



**Figure 2. Test event (circles) and stations (shown as station abbreviations) for the regional test case.**

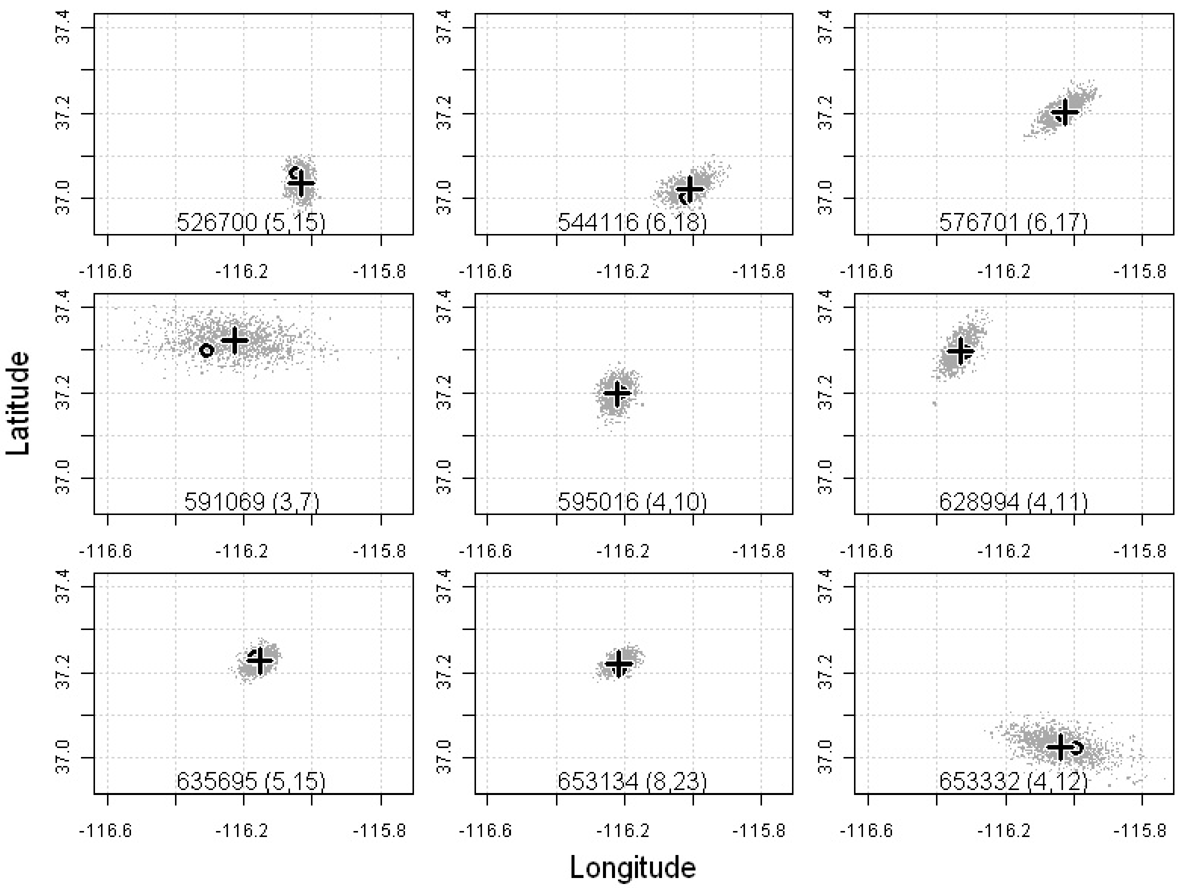



**Figure 3. Epicenter posterior distribution of the 9 events in case 1 (relative locations). Each panel shows the distribution of a single event. Each gray dot is an epicenter realization that was "accepted" by MCMCloc. Black circles show the true location of the event, and black crosses show the posterior average (mean) location. The bottom of each panel shows an event identification number along with the number of stations observing the event and the total number of observed phases in parentheses.**

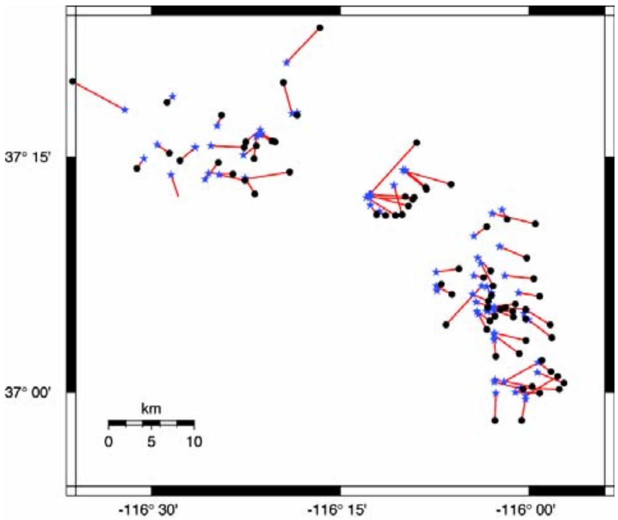



**Figure 4. Relative locations of the full NTS set of explosions using a conventional relative location method. Blue stars are the known location and black dots are estimated epicenters. Note the eastward bias of ~5 km. (From Anderson and Myers, 2005).**

*Case 2*

Figure 5 shows epicenter posterior distributions when two events are constrained to the known location (tight priors). Because locations in case 1 are excellent, there is not a large visual difference between Figures 3 and 5, but closer inspection shows slightly tighter posteriors in case 2, as expected. Posterior epicenter distributions in case 2 tighten because the travel-times are "calibrated" with more certainty by the two constrained events.

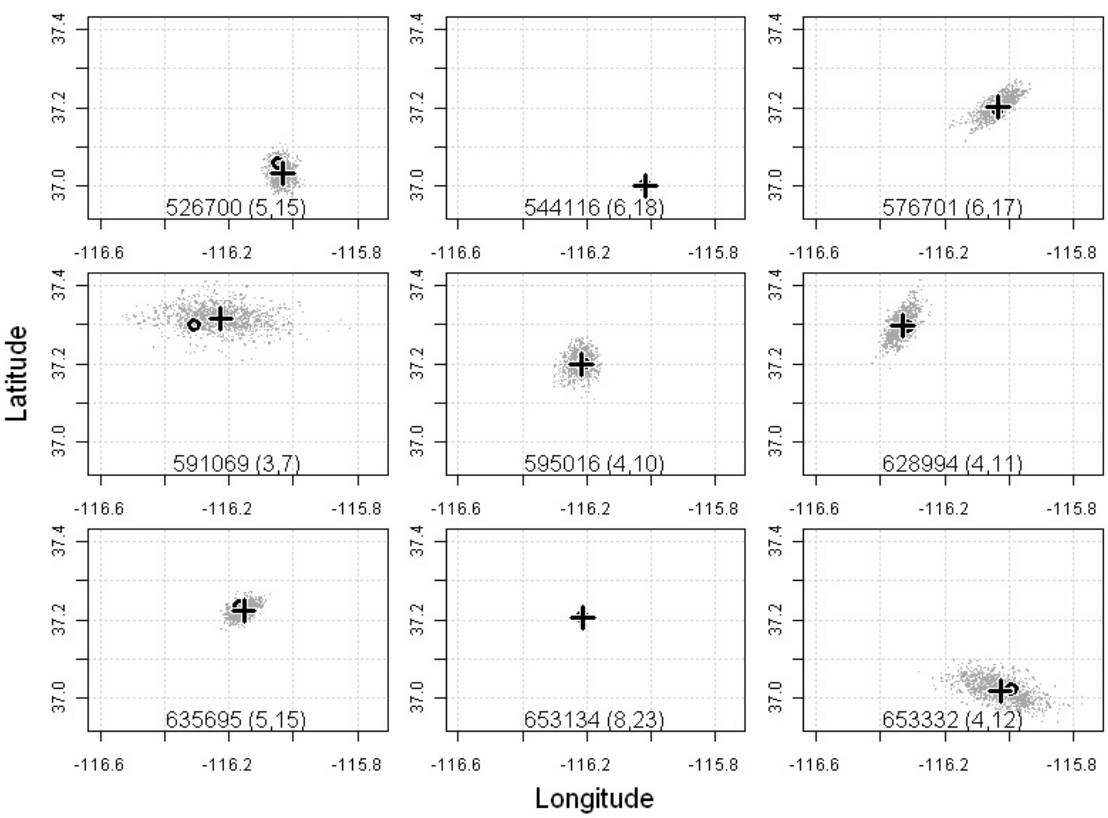



**Figure 5. Latitude-longitude location posterior distribution of the nine events in case 2. See Figure 3 for caption.**

*Case 3*

In this case, Pg/Pn phase-assignments are switched for events numbers 628994, 576701, and 635695. The confusion of Pg and Pn is presented only to demonstrate the capabilities of MCMCloc, as this mistake may not be realistic for most data sets. Pn and Pg phases arrive within 6, 10, and 1 second of one another for these event numbers, respective to the listed order above. For the same events, we also delayed 3 Lg arrivals by 10 seconds (not at the same stations for which phase-assignments were switched). Hence, for example, event 628994 has 11 phases at 4 stations. Of those 11 observations, Pn and Pg phase names were switched at one station and one Lg observation was delayed by 10 seconds (i.e., 3 out of 11 observation were corrupted).

In case 3 the phase-assignments were viewed as a model parameter with a prior based on the analyst's phase assignment. Phase assignments were sampled in MCMCloc, and posteriori assessments of the phase assignments were made. In addition to allowing for phase mislabeling in our model, we also allow for bad data. Bad data are reassigned to the "X" phase label.

Figure 6 shows the epicenter posterior plot for case 3. The most striking difference between Figures 6 and 3 is event 591069, which had no data corrupted in case 3. The change in event 591069 is the result of corrupted data from one station whose weight was decreased by the data corruption.

For the mislabeled phases (Pn and Pg switched), the marginal posterior phase-assignments are reported in Table 1. Unfortunately, in our test run we inadvertently set the probability of the correct phase assignment to 99%. We find that our prior was too "tight"; MCMCloc diligently followed the tight prior by maintaining the input phase assignment; however, it is interesting that MCMCloc reported a significant probability that the tampered arrivals are bogus.

**Table 1. Assigned phase (rows) with posterior probability of phase assignment (columns). In each of these cases, Pn and Pg phase assignments were reversed. Each column shows the posterior probability for assigning a given arrival observation (labeled by O[X], where X is the observed phase-label) to the three possible phase-labels (P[Lg], P[Pg], P[Pn]), plus the 'bogus' label (P[NA]) – each column should sum up to 1. Recall that the observed Pg (O[Pg]) are actually Pn and the observed Pn are Pg. As can be seen, where the phase mislabeling occurred, the observation tends to be removed.**

| 628994 to KNB | 576701 to BMN | 635695 to TPH |
|---|---|---|
| O[Lg] O[Pg] O[Pn] | O[Lg] O[Pg] O[Pn] | O[Lg] O[Pg] O[Pn] |
| P[Lg] 0.888 0.000 0.000 | P[Lg] 0.958 0.000 0.000 | P[Lg] 0.882 0.003 0.000 |
| P[Pg] 0.000 0.706 0.038 | P[Pg] 0.000 0.975 0.002 | P[Pg] 0.000 0.000 0.013 |
| P[Pn] 0.000 0.000 0.000 | P[Pn] 0.000 0.000 0.000 | P[Pn] 0.000 0.000 0.961 |
| P[NA] 0.112 0.294 0.962 | P[NA] 0.042 0.025 0.998 | P[NA] 0.118 0.997 0.026 |

The marginal posteriori phase-assignments for the arrivals that were corrupted by 10 seconds are reported in Table 2. The posterior phase assignment suggests that corruption of the Lg arrival time does not call for its removal. Although it may be unexpected, 3 instances of 10 second error in this data set is consistent with the overall posterior distribution of Lg arrival-time uncertainty. In this instance, MCMCloc pointed out a flaw in our perception that such a large arrival time error would be inconsistent with other observations. We note, however, that MCMCloc down weighted the Lg phase considerably, and the location estimates were minimally affected by the Lg data.

**Table 2. Assigned Phase (rows) with posterior probability of phase assignment. In each of these cases 10 seconds was added to the Pg arrival time. See table 1 caption for more detail.**

| 628994 to NEL | 576701 to ELK | 635695 to NEL |
|---|---|---|
| O[Lg] O[Pg] O[Pn] | O[Lg] O[Pg] O[Pn] | O[Lg] O[Pg] O[Pn] |
| P[Lg] 0.949 0.000 0.000 | P[Lg] 0.783 0.000 0.000 | P[Lg] 0.939 0.000 0.000 |
| P[Pg] 0.000 0.947 0.000 | P[Pg] 0.000 0.922 0.000 | P[Pg] 0.000 0.979 0.000 |
| P[Pn] 0.000 0.000 0.979 | P[Pn] 0.000 0.000 0.979 | P[Pn] 0.000 0.000 0.972 |
| P[NA] 0.051 0.053 0.021 | P[NA] 0.217 0.078 0.021 | P[NA] 0.061 0.021 0.029 |

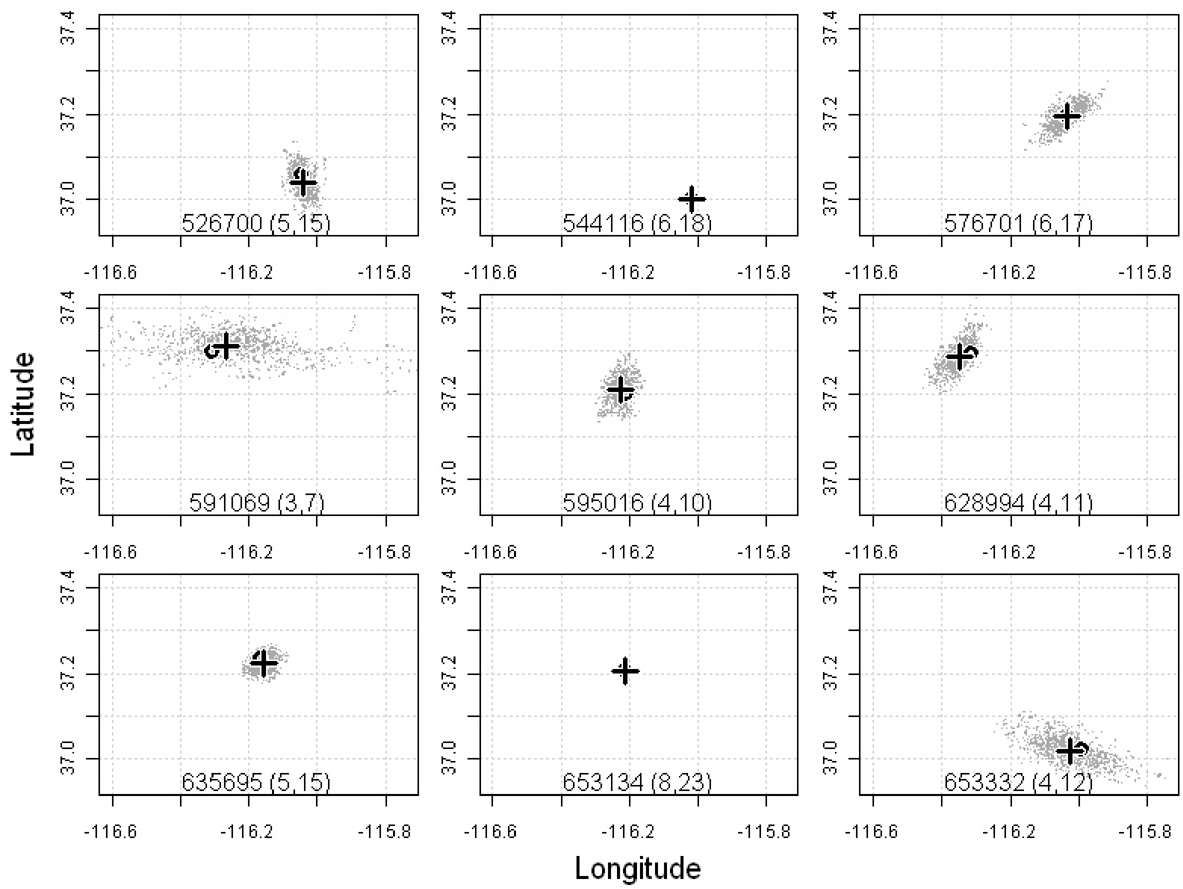



**Figure 6. Epicenter posterior distribution of the nine events in case 3. See Figure 3 for caption.**

## CONCLUSIONS AND RECOMMENDATIONS

We have developed a multiple-event location algorithm (MCMCloc) that employs the Markov Chain Monte Carlo method for solving large, non-linear inverse problems. The MCMC method is an efficient way to probe large model spaces, and it enables us to execute a smart search over the numerous combinations of model parameters. The introduction of prior constraints on model parameters further improves the efficiency of the parameters search and (like conventional techniques) helps to improve the accuracy of parameter estimates.

The MCMCloc forward model is divided between event location, travel-time prediction, and arrival time data. The most novel aspect of the MCMCloc forward model is the introduction of phase assignments into the arrival time model component. Therefore, we assess the likelihood that phases are properly named in the location algorithm. Although MCMCloc is in its formative stage, we find that inclusion of the phase assignment into the list of unknowns is desirable, and can lead the user to discover flaws in the data set. Further, MCMCloc tests for “bad” data that do not match any known arrival. Bad data may be removed from some instances of the solution or, if the data are persistently bad, they will be removed from every solution.

We test MCMCloc on a 9-event subset of the NTS nuclear explosions (Walter et al., 2003). We find that our preliminary version of MCMCloc outperforms conventional relative location methods. By implementing travel-time corrections in a hierarchic fashion in which travel-time curves are adjusted first, we find a distinct improvement in absolute location accuracy. In the case where 2 events are used to calibrate surrounding locations, we do not find a notable difference in absolution location error, but we do see smaller uncertainty bounds. In our last example, we demonstrate the ability of MCMC loc to reassign phase names and identify questionable data. This feature, while still in need of tuning, shows promise for automating the tedious task of grooming travel-time data sets.

## ACKNOWLEDGEMENTS

We have benefited from many conversations with Bill Rodi on the subject of seismic location. Thanks also to Bill Walter for leading the effort to compile and quality control the NTS data set.

## REFERENCES

Anderson, M., and S.C. Myers (2005), Assessment of regional-distance location calibration using a multiple event location algorithm, *In preparation for Bull. Seism. Soc. Am.*

Billings, S.D., B.L.N Kennett, M.S. Sambridge (1994), Hypocenter location: genetic algorithms incorporating problem-specific information, *Geophys. Jour. Int.* 118: 693-706.

Dewey, J.W. (1971), Seismicity and tectonics of western Venezuela, *Bull. Seism. Soc. Am*. 62: 1711-1751.

Douglas, A. (1967), Joint epicenter determination, *Nature* 215: 47-48.

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.

Kennett, B.L.N., and E.R. Engdahl (1991), Traveltimes for global earthquake location and phase identification, *Geophys. Jour. Int*., 105, 429-465.

Lomax, A. (2005), NLLoc - Non-linear, earthquake location program, http://alomax.free.fr/nlloc/.

Myers, S.C., S.C. Schultz (2000), Improving sparse network seismic location with Bayesian kriging and teleseismically constrained calibration events *Bull. Seism. Soc. Am*., 90: 199-211.

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

Pasyanos, M.E., G.A. Franz, and A.L. Ramirez (2005), Reconciling a geophysical model to data using a Markov Chain Monte Carlo algorithm: An application to the Yellow Sea - Korean Peninsula region, submitted to J. Geophys. Res.

Rodi, W., S. Myers, and C. Schultz (2003), Grid-Search Location Methods for Ground-Truth Collection from Local and Regional Seismic Networks, in *Proceedings of the 25th Seismic Research Review—Nuclear Explosion Monitoring: Building the Knowledge Base,* LA-UR-03-6029, Vol. 1, pp. 311-319.

Shapiro, N. and M. Ritzwoller (2003), Monte-Carlo inversion for a global shear-velocity model of the crust and upper mantle, *Geophys. J. Int.* (2002) 151, 88–105.

Waldhauser F, Ellsworth WL (2000), A double-difference earthquake location algorithm: Method and application to the northern Hayward fault, California *Bull. Seism. Soc. Am*., 73: 1753-1777.

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 of the 25th Seismic Research Review - Nuclear Explosion Monitoring: Building the Knowledge Base*, LA-UR-03-6029, Vol. 1, pp. 486-494.