JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 106, NO. B6, PAGES 11,217–11,237, JUNE 10, 2001

Estimating source parameters from deformation data, with an application to the March 1997 earthquake swarm off the Izu Peninsula, Japan

P. Cervelli, M. H. Murray,[1] and P. Segall

Department of Geophysics, Stanford University, Stanford, California

Y. Aoki and T. Kato

Earthquake Research Institute, University of Tokyo, Tokyo, Japan

Abstract. We have applied two Monte Carlo optimization techniques, simulated annealing and random cost, to the inversion of deformation data for fault and magma chamber geometry. These techniques involve an element of randomness that permits them to escape local minima and ultimately converge to the global minimum of misfit space. We have tested the Monte Carlo algorithms on two synthetic data sets. We have also compared them to one another in terms of their efficiency and reliability. We have applied the bootstrap method to estimate confidence intervals for the source parameters, including the correlations inherent in the data. Additionally, we present methods that use the information from the bootstrapping procedure to visualize the correlations between the different model parameters. We have applied these techniques to GPS, tilt, and leveling data from the March 1997 earthquake swarm off of the Izu Peninsula, Japan. Using the two Monte Carlo algorithms, we have inferred two sources, a dike and a fault, that fit the deformation data and the patterns of seismicity and that are consistent with the regional stress field.

1. Introduction

The increasingly widespread use of space geodesy has resulted in numerous, high-quality surface deformation data sets. For example, a large continuous GPS network of nearly 1000 stations covers most of Japan [*Kato et al.,* 1998], and a network of more than 250 stations is under construction in the greater Los Angeles area [*Bock et al.,* 1997]. Many geologically active areas such as Kilauea volcano and Long Valley caldera also have regional continuous GPS networks [*Owen et al.,* 2000; *Dixon et al.,* 1997]. Survey mode GPS data spanning more than a decade exist in many areas around the globe. Interferometric synthetic aperture radar (INSAR) promises even more deformation data with potentially worldwide spatial extent [*Massonet and Feigl,* 1998]. These geodetic data can provide important constraints on fault geometry and slip distribution. For this reason, we have developed robust and nearly automatic methods for rigorously inverting surface deformation fields for source type and geometry.

Many past attempts to infer source geometry from deformation fields have used elasticity theory and a trial-and-error approach to find geologically plausible deformation models that fit the major features of the observed deformation field [e.g., *Okada and Yamamoto,* 1991; *Owen et al.,* 1995]. Other workers have systematically searched through a large set of feasible models, comparing the predictions to the data and choosing the model that minimizes the misfit [e.g., *Ward and Barrientos* 1986; *Marshall et al.,* 1991]. Derivative-based searching algorithms have also been used; for example, *Árnadóttir et al.* [1992] used a quasi-Newton method to invert for the fault plane of the 1989 Loma Prieta earthquake. Finally, we have applied the inversion methods discussed below to the source geometry estimation problem. See, for example, *Murray et al.* [1996], *Jònsson et al.* [1999], *Freymueller et al.* [1999], or *Aoki et al.* [1999]. The aims of this paper are (1) to evaluate, using synthetic data sets, a variety of inversion methods for robustness and efficiency, (2) to develop techniques for assigning meaningful uncertainties to the estimated source parameters, and (3) to apply the inversion methods to the deformation from the March 1997 Izu Peninsula earthquake swarm. The study of the Izu swarm is included here both for pedagogical reasons and as an expansion on our previous work [*Aoki et al.,* 1999].

2. Theory

2.1. Earth and Source Models

Estimating source geometry from geodetic data requires a forward model of how the crust responds to various kinds

[1]Now at Seismological Laboratory, University of California, Berkeley, California.

Paper number 2000JB900399.
0148-0227/01/2000JB900399$09.00

of deformation sources. The most commonly used crustal model is the homogeneous, isotropic, linear, elastic half-space, which we adopt here. In spite of its limitations, the elastic half-space model is widely used primarily because of the simplicity of the expressions for the surface deformation caused by uniform, rectangular dislocations [e.g., *Okada,* 1985] and point sources [*Mogi,* 1958]. Moreover, until recently, most geodetic data were not of sufficiently high quality to justify more complex crustal models.

2.2. Inversion as Optimization

The relationship between the deformation field and the source geometry can be expressed by the following observation equation:

$$\mathbf{d} = \mathbf{G}(\mathbf{m}) + \epsilon, \tag{1}$$

where $\mathbf{d}$ is the deformation data vector, $\mathbf{m}$ is the source geometry vector (e.g., for a dislocation, length, width, depth, dip, strike, location, slip), and $\mathbf{G}$ is the function that relates the two. The ϵ term is a vector of observation errors. For the source geometry estimation problem the data are related nonlinearly to the source parameters. For this reason, source estimation reduces to nonlinear optimization. The optimal source model, $\hat{\mathbf{m}}$, will minimize the misfit between observation and prediction. Therefore we systematically search the finite dimensional parameter space for $\hat{\mathbf{m}}$, using $\mathbf{G}$ to predict the deformation field for a given $\mathbf{m}$.

How to quantify the misfit is somewhat arbitrary. Here, we follow common convention and use the weighted residual sum of squares, which can be normalized by the number of individual data, n, minus the number of estimated model parameters, p. This forms the "mean square error"(MSE), expressed as:

$$\mathrm{MSE} = \frac{\mathbf{r}^T \boldsymbol{\Sigma}^{-1} \mathbf{r}}{n - p}, \tag{2}$$

where $\mathbf{r} = \mathbf{d} - \mathbf{G}(\mathbf{m})$ and $\boldsymbol{\Sigma}$ is the presumably known data covariance. Several statistical tests exist for evaluating the significance of the MSE (e.g., the χ^2 and F tests), but it should be emphasized that the validity of these tests, as well as the meaningfulness of the $n - p$ normalization, depends on Gaussian noise and a linear relationship between model and data. Statistical tests on the MSE should therefore be regarded skeptically. The geodetic signal contains unmodeled deformation components such as those arising from elastic heterogeneity or anisotropy. This unmodeled signal contributes to the misfit, so the best fitting model for a particular data set may yield an MSE that is quite high. The "optimality" of our estimated source model is thus always conditional on the assumptions intrinsic to the forward model.

2.3. Solving the Optimization Problem

Derivative-based algorithms, Levenberg-Marquardt or the method of conjugate gradients, offer the most straightforward and efficient approach to solving this optimization problem [*Gill et al.,* 1981]. However, because these algorithms depend on the gradient and higher-order derivatives to guide them through misfit space, they can get trapped in the first local minimum that they encounter and never find or even approach the global minimum. Consequently, these algorithms work well only when the initial guess is near the global minimum. A priori information can often provide a good initial guess. For example, when *Árnadóttir et al.* [1992] used a quasi-Newton method to invert for the fault plane of the 1989 Loma Prieta earthquake they had a very good idea, from geologic and seismologic data, of where to begin the inversion. Moreover, they found the same model estimate when starting from several different initial guesses.

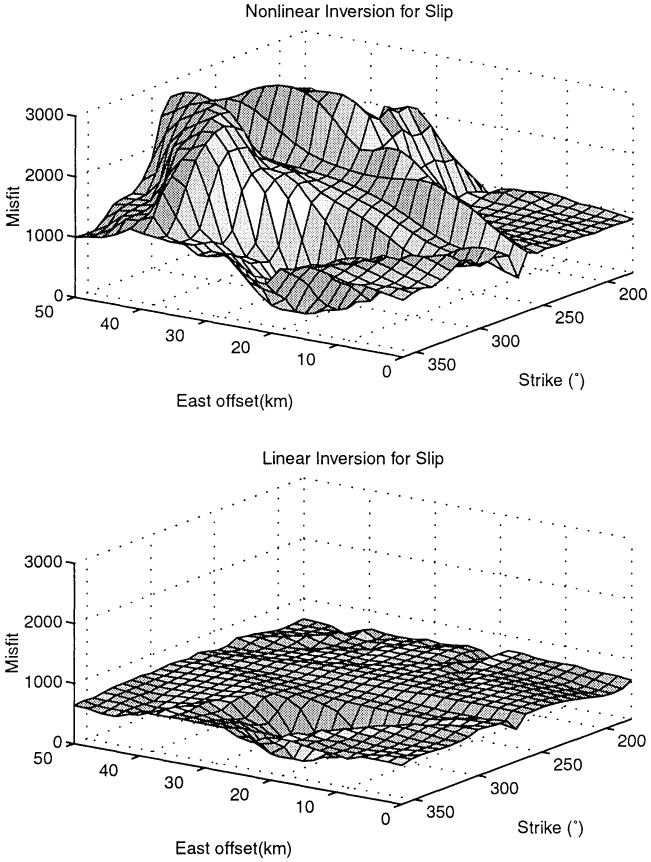

Figure 1. (top) Misfit space depicted in three dimensions, i.e., misfit as a function of two model parameters. Note that the space contains multiple minima, which makes the result of a derivative-based algorithm dependent on the initial guess. (bottom) The misfit surface for the same two model parameters when all the linear parameters are set at their optimal values, in a least squares sense. The effect is to flatten the misfit space, which degrades the performance of Monte Carlo algorithms.

Figure 1 illustrates the topologic complexity of misfit space in three dimensions (misfit as a function of two parameters). Clearly, whether a derivative-based method (or any method that always moves "downhill") reaches the global minimum depends on where it starts. Moreover, we have found that particularly in cases with low signal-to-noise ratios (SNR), the misfit space often contains numerous local minima and lacks a deep, well-defined global minimum. We can therefore only endorse derivative-based methods as a solution to the geodetic inversion problem in cases characterized by both a high SNR and good geologic insight into the probable type and location of the

deformation source. Even in these cases, it is essential to show that the optimization converges to the same minimum from many different starting values.

Another way to solve the optimization problem involves discretizing the misfit space and then searching exhaustively through the resulting grid. Exhaustive searches prove viable, however, only when the number of estimated parameters stays small. The number of possible combinations equals Q^p, where Q is the number of divisions in the grid and p is the number of parameters. Adopting a coarse grid does little to help the problem both because of the exponential growth and because an overly coarse grid might miss the part of the misfit space containing the global minimum. Random searches prove more efficient than exhaustive searches, but the increase in efficiency is not sufficient to overcome the exponential nature of the problem. For these reasons, neither exhaustive nor random searching seems useful for anything but the most modest source geometries and the simplest Earth models.

In spite of their inefficiency, exhaustive and random searches do not suffer from the local minimum problem. Beginning in the 1940s, mathematicians sought algorithms that combined the efficiency of a derivative-based method with the robustness of a random search. The result was the Monte Carlo class of algorithms. The common feature that all algorithms of this class share is an element of randomness that permits an occasional uphill move, that is, the algorithms will not always move from a candidate model with higher misfit to a model with lower misfit. This feature permits Monte Carlo algorithms to escape local minima. Monte Carlo algorithms differ from "conventional" random search methods in that the former retain a directivity that engenders a high level of efficiency. The two Monte Carlo algorithms explored in this paper are simulated annealing [*Metropolis et al.,* 1953] and the random cost algorithm of *Berg* [1993]. Another class of Monte Carlo algorithm includes the genetic algorithms [e.g., *Yu and Rundle,* 1995]. These are considerably more difficult to implement than either of the algorithms explored here.

2.3.1. Simulated annealing. In simulated annealing, the likelihood of choosing a higher misfit model over a lower one depends not only on the misfit difference between the two but also on the state of the annealing process at the time of the choice. The algorithm quantifies this state dependence in terms of a temperature. At high temperatures all source models have roughly equal chances of getting picked, while at low temperatures the algorithm favors low misfit models.

The specific annealing algorithm adopted here follows from the work of *Metropolis et al.* [1953] and *Creutz* [1984]. Called the "heat bath" algorithm, it proceeds as follows. The initialization procedure consists of two steps: (1) set bounds on the values for all the model parameters (these bounds can come from geologic constraints or physical limitations) and (2) randomly pick an initial starting model. Cycle through the individual model parameters, m_1 through m_p. At each m_i, compile a list of candidate models by varying m_i while holding parameters $m_{j\neq i}$ fixed at their current values. Next, calculate the misfit associated with each model, and form a probability distribution according to

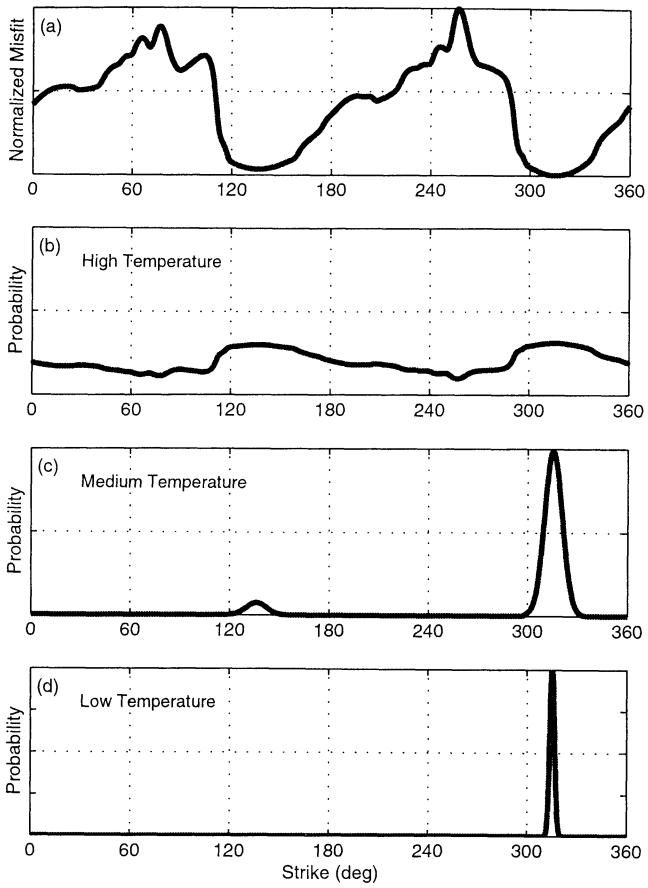

Figure 2. Probability density functions for simulated annealing at three different temperatures. (a) Misfit as a function of a single model parameter (all the other parameters are held fixed). (b) Corresponding probability density function at high temperature. The density function is quite flat; lower misfit models are only slightly more favored than higher misfit models. (c and d) probability density function evolution as the annealing progresses from medium to low temperature. At low temperatures, the parameter value that minimizes the misfit becomes overwhelmingly probable.

$$p_k \sim e^{-\frac{\mathrm{MSE}_k}{T}}, \tag{3}$$

where T refers to the current temperature and MSE_k refers to the MSE of the model corresponding to the kth permissible value of the ith parameter. To update the value for the ith model parameter, randomly sample from the distribution. Continue updating until all the parameters have been cycled through, then lower the temperature according to a "cooling schedule" and repeat until some termination criterion obtains. Usually, that criterion is some preset number of model updates or some time period. Alternatively, a "freezing" criterion can be used where the algorithm stops if the current model has remained unchanged for some specified number of iterations.

Figure 2 illustrates the effect of temperature on the probability distribution. At high temperatures the distribution is very flat. At this stage of the annealing, the algorithm essentially functions as a random search. As the temperature cools, low misfits are preferred, but the random sampling ensures that they are not required. Oc-

casional uphill moves occur; local minima are visited and exited. In the final stages of annealing, when the temperature is very low, the probabilities associated with low misfits become very high, making uphill moves extremely unlikely.

The most significant complication to the simulated annealing algorithm is the cooling schedule, i.e., how the temperature changes as the annealing progresses. This plays a crucial role in the success or failure of the optimization. *Rothmann* [1985] defined a critical temperature at which the bulk of the annealing should, for maximum efficiency, occur. In brief, at the critical temperature the system remains cool enough to favor low misfits but still high enough to escape local minima. *Basu and Frazer* [1990] developed a method to quickly find the critical temperature that greatly improved on Rothmann's trial and error approach. This method runs the annealing algorithm at a number of different fixed temperatures and defines the critical temperature as the temperature at which the algorithm retains a high degree of directivity yet can still easily escape local minima. At temperatures much higher than the critical temperature, the algorithm behaves very much like a random search, while at temperatures much lower than the critical temperature it behaves like (an inefficient implementation of) a derivative-based algorithm.

We have found that for the geodetic inversion problem it is generally not necessary to calculate the critical temperature using the algorithm described by *Basu and Frazer* [1990]. Instead, using extensive trials, we have derived a fast method of forming a cooling schedule that works well for the geodetic inversion problem: calculate the average MSE for 100 randomly selected models. The approximate critical temperature T_c can be found from

$$\log_{10} T_c = \log_{10}(\overline{\text{MSE}}) - 2.5. \tag{4}$$

The first term on the right side of (4) provides a natural scale for the misfit function, while the second term is derived empirically. For several test cases, Figure 3 compares our approximations of the critical temperature to those obtained from the rigorous method of *Basu and Frazer* [1990]. Once a good approximation of the critical temperature has been found, a cooling schedule can be easily generated. Like *Basu and Frazer* [1990], we recommend that the bulk of the annealing time be spent at or near the T_c. Specifically, the cooling schedule should begin 1 order of magnitude higher than T_c and end 1 order of magnitude lower. The number of iterations spent at a particular temperature should decay rapidly with distance from T_c. We have found that a Gaussian distribution centered at T_c works well.

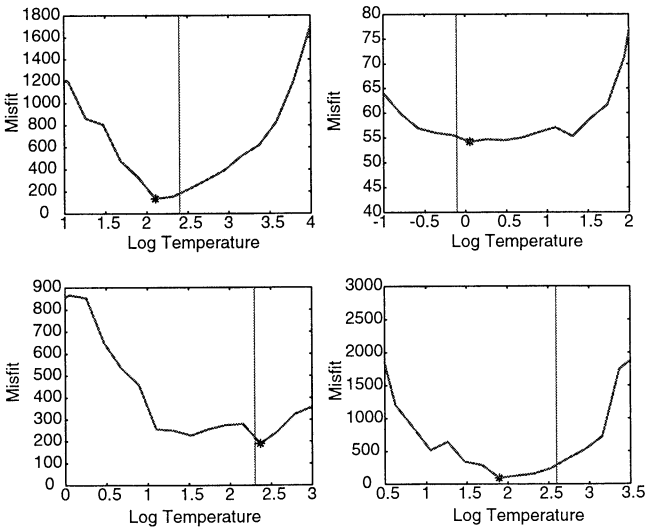

Figure 3. A comparison of robustly determined critical temperature (star) to the rule of thumb approach (vertical line). Note the good agreement between the two methods for finding the critical temperature.

2.3.2. Random cost. Random cost is an alternative Monte Carlo approach for nonlinear optimization problems with many local minima in a broad misfit space [*Berg,* 1993]. It uses a simple stochastic process to enforce a random walk in misfit space, which enables it to overcome local increases in misfit to find the global minimum. It preferentially samples minima (or maxima) while thoroughly sampling the misfit space. Our tests indicate that it is significantly less efficient than simulated annealing, but it is much easier to implement because it does not require a specific cooling schedule.

The random cost approach begins by generating a set of trial models that span a region about an arbitrary a priori model. The method for generating the trial model set is not unique. As described in section 3 we follow *Berg* [1993] by defining a geometric grid in parameter space centered on the a priori model. A geometric grid provides a convenient means to simultaneously sample the parameter space both broadly away from the a priori model (i.e., thorough sampling) and densely near the a priori model (i.e., high accuracy).

In the discussion that follows, the "cost" function is the MSE misfit. The misfit difference, Δ, is determined for each trial model, such that Δ is the difference between the misfit of the trial model and the misfit of the a priori model. Δ is negative for trial models with an MSE smaller than the a priori model and positive for trial models with higher MSE. For simplicity, we ignore complications that arise when the trial model MSE equals the misfit of the a priori model (see *Berg* [1993] for discussion). Let f^- equal the average of all the negative Δ (i.e., better trial models) and f^+ equal the average of all the positive Δ (i.e., worse trial models). If the a priori model is not at a misfit extremum, the random cost method chooses a new a priori model from the existing trial models based on a probability function determined from f^- and f^+. Let two probabilities p^- and p^+ ($p^- + p^+ = 1$) be defined by

$$p^- f^- = p^+ f^+, \tag{5}$$

which yields

$$p^- = \frac{f^+}{f^+ + f^-}. \tag{6}$$

Let r be a random number chosen from a uniform distribution between 0 and 1. If $r \leq p^-$, a new a priori model is randomly selected from the set of trial models with negative Δ; otherwise, it is selected from the set of trial models

with positive Δ. A new set of trial models is generated about the new a priori model, and the process is repeated until a local minimum (or maximum) is encountered, that is, until no better (or worse) trial models are found. This model is recorded and then the process can be restarted at another randomly chosen a priori model.

When the a priori model is near a misfit minimum, negative misfit differences will, on average, have smaller magnitude than the positive cost differences ($f^- \leq f^+$). From (6) this implies that p^- will be greater than p^+, so there is a greater probability that the next a priori model will be chosen from the set with negative cost differences and therefore be closer to the cost minimum. As the a priori models get closer to the misfit minimum, the probability increases that an even better a priori model will be chosen for the next iteration. Thus the random cost method preferentially samples the local minima. However, there always remains a finite probability that a poorer fitting a priori model will be chosen, enabling the random cost method to overcome local increases in the misfit function in order to locate the global minimum.

From the symmetry of (5) and (6), it should be noted that maximum misfits will also be preferentially sampled, that is, the algorithm preferentially visits extrema, not just minima. Since we are seldom interested in the model that fits the data worst, we simply discard those models corresponding to maxima in misfit space.

Ideally, when not at a minimum or maximum, f^+ and f^- (and thus p^- and p^+ in (5) and (6)) should be roughly equal, giving equal likelihood that either better or worse new a priori models would be chosen. This enforces a random walk in misfit space, enabling it to be thoroughly sampled. Thus the random cost method is most effective for studying broad (i.e., nearly flat) cost functions with many local minima or maxima. Because the geometric grid of trial models cannot possibly sample the entire cost space, there is no guarantee that the global minimum will be the first minimum detected. Therefore the above process needs to be repeated, beginning with other randomly chosen a priori models, to ensure that all local minima (including the global minimum) are properly sampled.

3. Using Monte Carlo Optimization

Given infinite time, Monte Carlo optimization methods will always find the global minimum of the misfit space. Our experience with these algorithms is that over finite time, however, they sometimes fail. The reason for the occasional failures stems from the intrinsic randomness of these algorithms. The main way to address the algorithm failure problem is to let the algorithms run for a long time. Of course, what constitutes "a long time" remains an open question. Our experience has been that for either algorithm 50,000 to 100,000 evaluations of the misfit function are almost always sufficient.

The update step in both simulated annealing and random cost requires picking from a list of models, each one corresponding to a particular permissible state of the parameter in question. Deciding how to define the permissible states for model parameters is arbitrary, though plausible upper and lower limits often exist. The simplest list of permissible parameter values consists of a regular, linear spacing between the lower and upper limits. We prefer a geometric grid similar to that proposed by *Berg* [1993]: For a parameter at value m_0, the list of permissible values contains $m_0 \pm 2^{-q}\Delta$, for integer values of q up to the grid precision, Q. Δ equals half the interval between the lower and upper parameter limits. There may be fewer permissible values if $m_0 \pm 2^{-q}\Delta$ yields results lying outside the bounds on the parameter in question. Note that in contrast to a linear spacing scheme, the range of values from which an update can be chosen depends on the parameter value prior to the update. This feature permits a more thorough search of parameter space by combining broad coverage and high precision. Moreover, we have found that letting $Q = 4$, which entails at maximum eight choices at each update, works very well for both simulated annealing and random cost. To achieve a similar robustness with a linear spacing requires a fine enough grid that the maximum number of choices at each update becomes an order of magnitude greater.

Most deformation models have at least one and perhaps several parameters that are linearly related to the deformation field, e.g., the "slip," "opening," or "potency" parameters. Thus, for a given set of the remaining nonlinear parameters it is possible to solve, using conventional least squares, for the optimal values of the linear parameters. This leads to an optimization strategy that involves projecting the linear parameters out of the misfit space. We can rewrite equation (1) as

$$\mathbf{d} = s\mathbf{G}(\mathbf{m}) + \epsilon, \tag{7}$$

where s is the slip or opening, and simply solve for the optimal s given some geometric model m. This strategy permits the optimization algorithm to operate only on the nonlinear parameters, implicitly solving for the linear parameters at each iteration. The obvious advantage to this approach is that it reduces the dimension of the misfit space, which should increase optimization efficiency. Nonetheless, we recommend against using this approach for the following reasons: (1) Absurd parameter values from unconstrained least squares estimation can compensate for an otherwise deficient set of nonlinear model parameters. For example, dislocation model $\mathbf{m_a}$ might fit that data as well as model $\mathbf{m_b}$ but only if $\mathbf{m_a}$ has 100 m of slip. (2) Projecting the linear parameters out of the optimization problem flattens the misfit space (see Figure 1), which will degrade the performance of any directed search algorithm. (3) Least squares is not without computational cost. Our experience has been that when the Monte Carlo algorithms operate on all the parameters, linear parameters included, it is no less efficient than separate linear inversion. Further, in the test cases, the Monte Carlo algorithms more consistently converge to the global minimum when they are permitted to operate over all the parameters. It should be noted that in some cases, separating the linear and nonlinear parameters is called for, e.g., simultaneous inversion for source geometry and spatially variable slip.

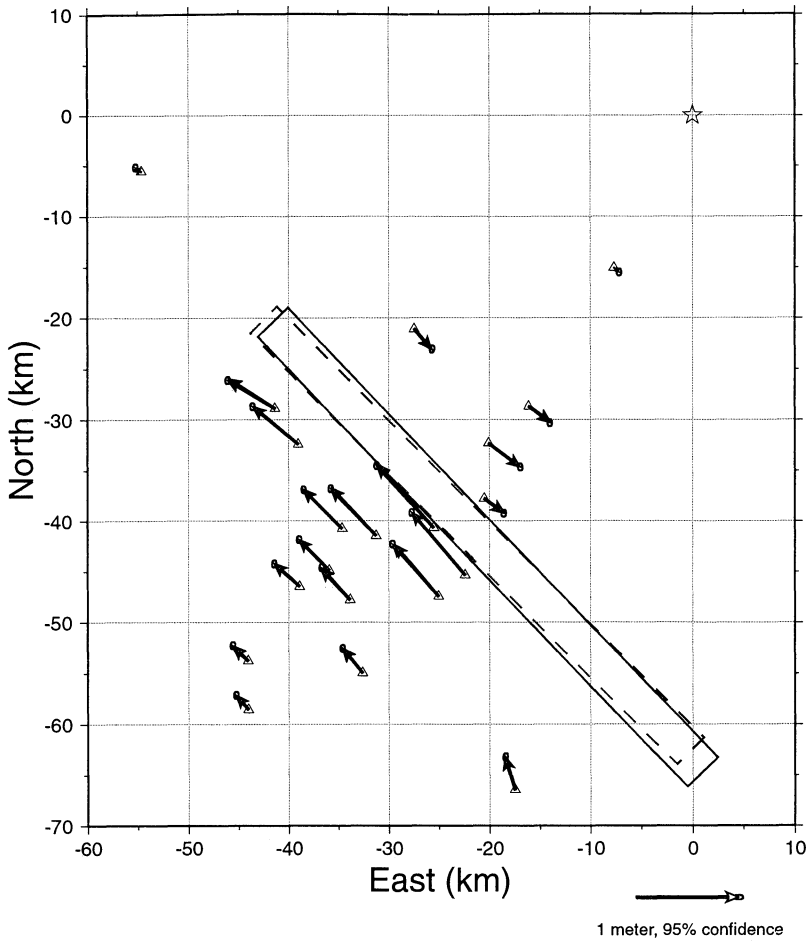

Figure 4. The observed displacement field for the simulated earthquake represented by test case 1 are depicted in white; the displacements from the model are depicted in black. (At this scale, observation and prediction are difficult to distinguish.) Displacements are relative to the far-field station marked with a star. The actual dislocation surface is depicted by the solid line, while the model dislocation is depicted by the dashed lined. Correlated noise, indicated by the error ellipses, has been added to the displacement vectors.

4. Experiments With Synthetic Data

4.1. Test Cases

To test the robustness and efficiency of the optimization methods described in sections 2 and 3 we present two scenarios based on synthetic data. These two represent the end-members from a variety of scenarios that span a wide range of SNRs. We simulated the surface displacements from two earthquakes using elastic dislocations with uniform slip. We added correlated noise to the displacements based on an actual scaled GPS data covariance matrix, which contained typical, if not high, amounts of noise. The station distribution was taken from a real geodetic network in northern California. Figure 4 and Figure 5 depict the surface projections of the model faults and the corresponding displacements with error ellipses. All displacements are given relative to the coordinate system origin, which lies near the upper right corner and is indicated by a star. Table 1 and Table 2 give the specific model parameters for the two test cases.

The first test model (Figure 4) represents a relatively easy inversion problem: the synthetic displacements are very large, corresponding to those from a $M7$ earthquake. The high SNR of this case leads to a very deep minimum in misfit space, which should be quickly found by any optimization algorithm. Moreover, the displacement field lends itself to straightforward geological interpretation, which permits reasonable a priori estimate of the underlying deformation source. Several "downhill" algorithms were tested: constrained and unconstrained Levenberg-Marquardt least squares and a Nelder-Mead simplex method. The constrained Levenberg-Marquardt method was able to recover the original model; the other two algorithms converged to local minima. Both simulated annealing and random cost found the global minimum very rapidly, often converging in fewer than 5000 calls to the misfit function (somewhat less than a minute of computer time on a 300 MHz Pentium II processor). Because of the added noise, the minimum in the misfit function does not precisely correspond to the actual model. Nonetheless, as Figure 4 illustrates, the recovered model mimics the actual model almost perfectly. Is it necessary to use Monte Carlo algorithms on "easy" inversion problems characterized by high SNR and plentiful prior geological information? Per-

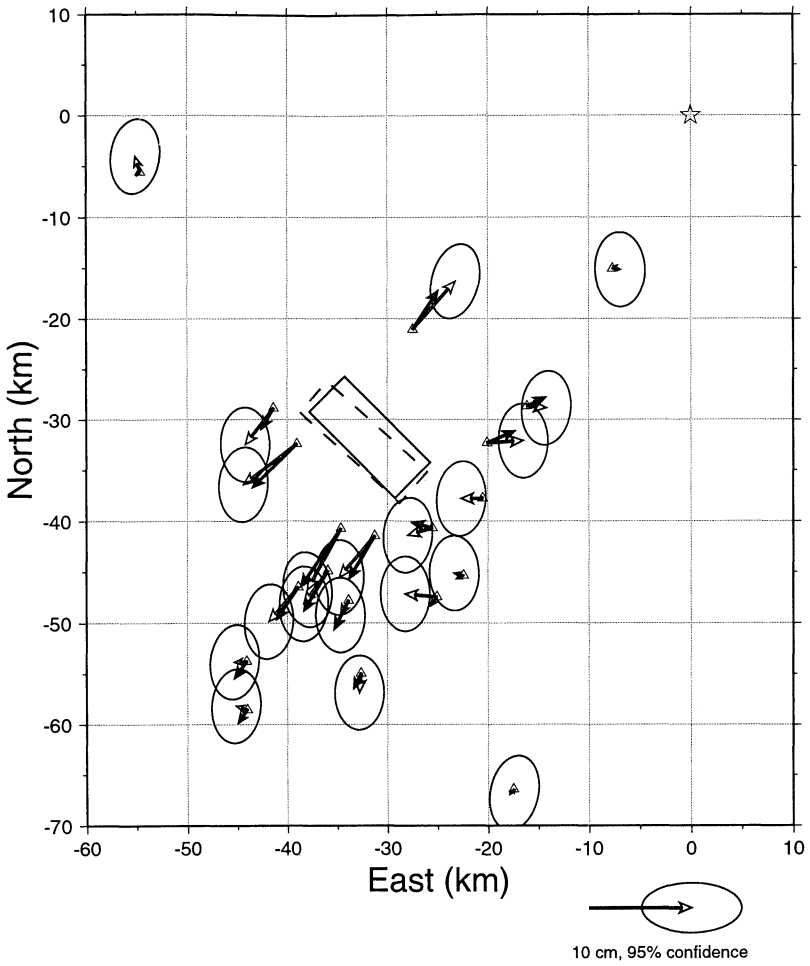

Figure 5. The observed displacement field for the simulated earthquake represented by test case 2 are depicted in white; the displacements from the model are depicted in black. Displacements are relative to the far-field station marked with a star. The actual dislocation surface is depicted by the solid line, while the model dislocation is depicted by the dashed line. Correlated noise, indicated by the error ellipses, has been added to the displacement vectors.

haps not, but the important point is that with a Monte Carlo algorithm, no prior model is required. If a source estimate from a Monte Carlo inversion agrees with geologic observation, this is an independent confirmation of the result.

The second test case (Figure 5) represents a much smaller earthquake, $\sim M6$, with corresponding surface displacements on the threshold of what can be easily detected and modeled. The small SNR makes for a topographically subdued misfit space with many minima of approximately

Table 1. Test Case 1[a]

		Inversion Bounds		Inversion Method			Confidence Intervals	
Parameter	True Value	Lower	Upper	SA	RC	HY	Lower	Upper
Length, km	60	20	100	58.0	59.8	59.7	59.5	60.9
Width, km	12	5	15	11.4	10.7	11.8	11.2	12.6
Depth to top, km	1	0	5	1.2	1.3	1.1	0.8	1.2
Dip, deg	70	50	120	71.2	72.2	70.3	68.9	71.2
Strike, deg	315	270	360	315.1	315.7	314.8	314.6	315.7
East offset, km	-20	-50	0	-20.8	-20.4	-20.0	-20.1	-19.5
North offset, km	-40	-50	0	-39.5	-40.0	-40.0	-40.6	-39.8
Strike slip, m	2	-5	5	2.1	2.2	2.0	1.92	2.11
Dip slip, m	0.2	-5	5	0.21	0.22	0.2	0.19	0.21
Moment,[b] 10^{19} N m	4.34	N/A	N/A	4.18	4.23	4.24	N/A	N/A

[a]N/A, not applicable; SA, simulated annealing; RC, random cost; HY, hybrid.
[b]Moment is not a model parameter; it is included for convenience.

Table 2. Test Case 2[a]

Parameter	True Value	Inversion Bounds		Inversion Method			Confidence Intervals	
		Lower	Upper	SA	RC	HY	Lower	Upper
Length, km	12	1	15	13.1	13.8	13.4	8.9	14.6
Width, km	7	1	15	5.2	5.0	5.7	4.4	9.8
Depth to top, km	0.5	0	5	0.85	0.09	0.75	0.0	1.4
Dip, deg	45	5	175	44.8	50.0	43.8	38.9	153.3
Strike, deg	315	180	360	309	311.2	311.9	302	335.4
East offset, km	-30	-50	0	-31.5	-30.3	-31.0	-33.8	-28.2
North offset, km	-30	-50	0	-30.8	-30.9	-30.7	-33.8	-29.1
Strike slip, m	0	-2	2	0.07	0.12	0.01	-0.61	0.18
Dip slip, m	0.75	-2	2	0.87	0.84	0.78	-1.0	1.0
Moment,[b] 10^{19} N m	0.19	N/A	N/A	0.178	0.178	0.180	N/A	N/A

[a]N/A, not applicable; SA, simulated annealing; RC, random cost; HY, hybrid.
[b]Moment is not a model parameter; it is included for convenience.

equal depth. All of the "downhill" algorithms we tested (the same three as above) failed to converge on the global minimum even when seeded with good initial guesses. The Monte Carlo algorithms usually converged on the global minimum, though occasionally they were fooled by the conjugate plane, which predicts surface displacements that fit the data nearly as well as the true fault plane. The lesson here is to allow the Monte Carlo algorithms ample time to explore the misfit space, particularly when the SNR is small. However, if the algorithms consistently fail to converge on the same answer, that is, if this misfit space is very flat over a large range of parameter values, then the problem may lie not in the algorithms, but rather in the data themselves. The uncertainty analysis methods discussed in section 5 are particularly useful for distinguishing such cases.

4.2. Discussion

We find that the random cost and simulated annealing algorithms work well for finding the vicinity of global minima in misfit space. For efficiently reaching the exact minimum we recommend a hybrid approach to this optimization problem: First, use a Monte Carlo algorithm and let it converge on a model, which is then passed to a derivative-based algorithm as an a priori starting guess.

Comparisons between the two Monte Carlo algorithms show several differences in operational efficiency. Random cost is the easier of the two algorithms to use, with no metaparameters to set. Simulated annealing requires an appropriate cooling schedule to operate properly, so this algorithm involves either a fair amount of user experience or a subalgorithm to calculate a cooling schedule. Once an appropriate cooling schedule has been found, simulated annealing generally runs much more quickly than random cost, though finding the appropriate cooling schedule can more than offset this increased efficiency. For this reason, we favor simulated annealing when essentially the same optimization problem needs to be run many times, as, for example, in the bootstrap discussed in section 5.3. Random cost seems to perform substantially better in cases with low SNR than it does in cases of high SNR. The reason for this probably stems from the nature of the random cost algorithm, which preferentially samples extrema. Misfit spaces characterized by high SNR tend to have few extrema; hence the strength of the random cost algorithm is not realized.

Figure 6 presents the results of several trials we ran to compare the efficiency of the two algorithms. In these synthetic cases, simulated annealing consistently, and some-

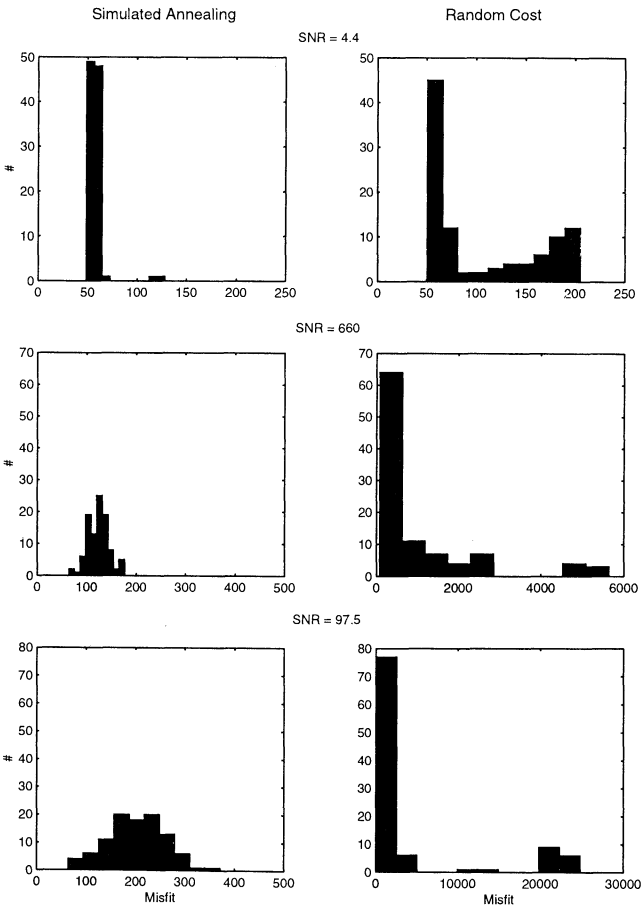

Figure 6. Histograms depicting the results of 100 trials on three different test cases. The two algorithms were permitted to run until a set number of calls to the misfit function was reached. The histograms show the distribution of the misfits corresponding to the optimal model for each of the 100 trials. Note the horizontal scale change in the bottom two panels from the left to the right column.

times dramatically, outperforms random cost. We suspect that this behavior may be an artifact of the synthetic nature of the misfit space. In the absence of noise, the forward model predicts the synthetic data perfectly. We suspect that this leads to an artifically smooth misfit space that biases the comparison between the two algorithms against random cost. In spite of its relative inefficiency, the random cost algorithm did ultimately converge to the global minimum for each of the trial cases. Therefore we emphasize that for maximum robustness, both algorithms ought to be used for all inversion problems. Because the two algorithms function so differently, they provide good independent cross-checks of one another.

5. Confidence Intervals

5.1. Uncertainty in Nonlinear Models

Assessing the uncertainties, or confidence intervals, associated with the estimated source parameters remains an important problem, since without information about uncertainties, evaluating the practical significance of an inversion result becomes difficult. We would like to know not only the uncertainty of the individual parameters but also which parameters correlate with one another, and to what degree.

In the analysis of confidence intervals and standard errors that follows, one important point bears remembering: We are estimating parameters and their uncertainties given a particular model of the the Earth, e.g., a homogeneous elastic half-space. In the test cases, the synthetic data come from the very model subsequently used for inversion. For these cases, therefore, parameter misestimation stems only from the contaminating effect of noise in the data. In reality, noise accounts for only part of the difference between parameter estimates and their actual values. Unmodeled complexities, such as material heterogeneities, comprise the rest of the difference. The distinction between uncertainties about source model parameters and uncertainties about the model itself must be kept clear.

Ellipsoidal confidence regions result only from linear cases where noise follows a Gaussian probability distribution. In general, confidence regions will be much more complex and much harder to assess. This is even true for individual confidence intervals, or standard errors. Misfit will not be a parabolic function of parameter value, so confidence intervals will not be symmetric. The $\pm\sigma$ notation becomes obsolete in favor of $^{+\sigma_1}_{-\sigma_2}$. Indeed, even this notation does not capture the most general case where the confidence intervals are not continuous (see, for example, Figure 10).

5.2. The Bootstrap on Correlated Data

To estimate the individual confidence intervals associated with the inverted model parameters, we employ the bootstrap method [*Árnadóttir et al.,* 1992; *Efron and Tibshirani,* 1993], which works as follows. Start with a set of noisy data $\mathbf{d}$ and some function of the data, f, that returns one or more values. Clearly, $f(\mathbf{d})$ will depend not just on the signal within $\mathbf{d}$ but also on the noise, so a confidence interval must bracket the result of $f(\mathbf{d})$. The bootstrap builds confidence intervals by randomly resampling from the data, with replacement (i.e., a particular datum can be resampled more than once, others not at all) and then reevaluating $f(\mathbf{d}^*)$, where $\mathbf{d}^*$ is the resampled set. The values of f for a large number of resamples ($B > 1000$) are then sorted from lowest to highest. To find, for example, the 95% confidence interval, simply discard the top and bottom 2.5% of the sorted, bootstrap resampled values of $f(\mathbf{d}^*)$.

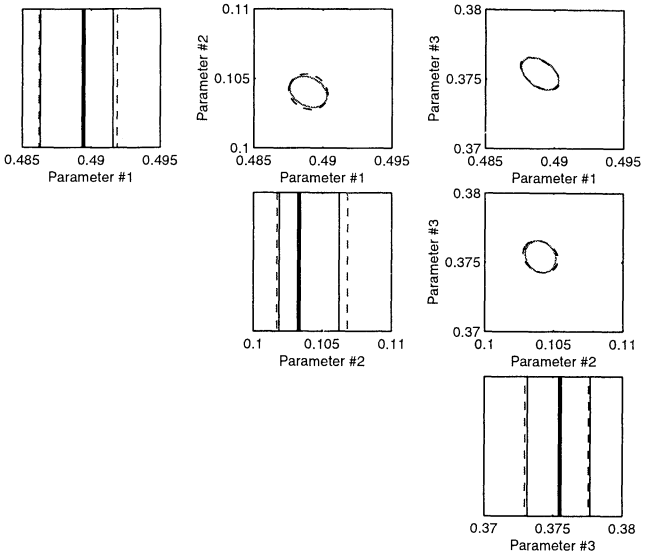

Figure 7. Empirical comparison of bootstrap values for correlations and uncertainties to analytic estimates. A three parameter linear model was used to generate synthetic data, to which correlated noise was added. The confidence intervals (regions) from linear inversion are depicted as solid lines; the intervals (regions) inferred from the bootstrap are as dashed lines. The thick solid lines indicate the actual parameter values.

Implemented in this way, the bootstrap assumes that all the data are equally weighted and that they are uncorrelated with one another. Geodetic data, however, are almost always unequally weighted and significantly correlated. These facts complicate the application of the bootstrap algorithm to the source estimation problem.

Beginning with a data vector $\mathbf{d}$ and its matrix covariance $\mathbf{\Sigma_d}$, how do we find a resampled data vector $\mathbf{d}^*$ that has appropriately resampled covariance matrix $\mathbf{\Sigma_d^*}$? One approach that does not work is to form a resampling operator $\mathbf{J}$, which is an identity matrix of the same dimension as $\mathbf{\Sigma_d^*}$ with its rows resampled with replacement, and to apply it to the data and its covariance forming $\mathbf{d}^* = \mathbf{Jd}$ and $\mathbf{\Sigma_d^*} = \mathbf{J\Sigma_d J^T}$. Using this approach will produce, in general, a rank deficient $\mathbf{\Sigma_d^*}$. It is easy to see why this is so: Twice resampled data points become completely correlated, which creates repeated rows in $\mathbf{\Sigma_d^*}$. To circumvent this problem, we resample the weighted residuals rather than the data themselves. The optimization methods described above find the model that (effectively) minimizes the squared weighted residual norm, $\|\mathbf{Wr}\|^2$, where $\mathbf{r} = \mathbf{d} - \mathbf{G}(\mathbf{m})$ and $\mathbf{W} = \mathbf{\Sigma_d^{-\frac{1}{2}}}$. For bootstrap-

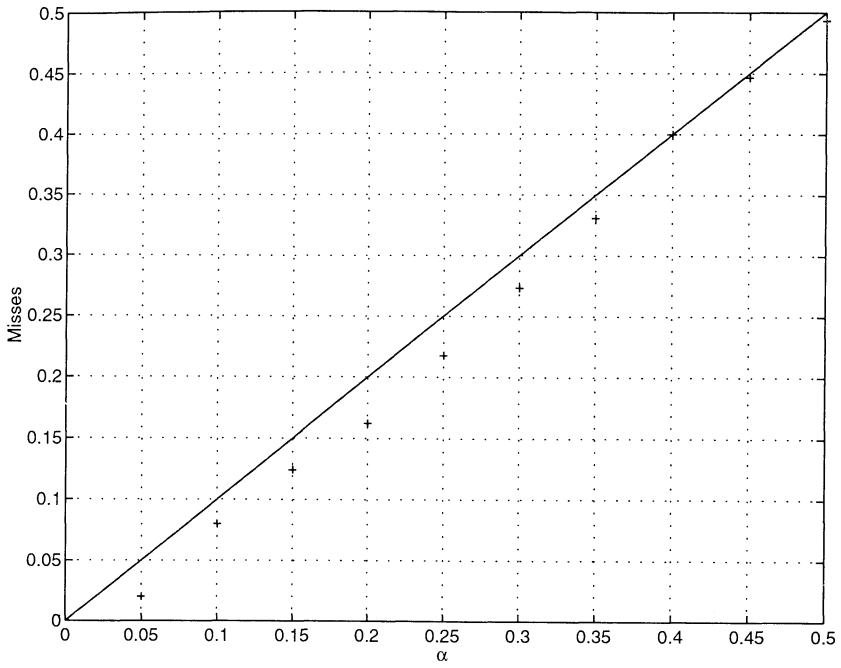

Figure 8. Actual versus expected behavior of bootstrap confidence intervals. If we choose 95% confidence ($\alpha = 0.05$), we expect the true parameter value to fall outside the confidence interval ~ 1 time in 20. The solid line depicts this expectation over a range of α values, while the crosses indicate the behavior observed in our numerical tests. For this problem the bootstrap seems to slightly overpredict confidence intervals, probably because of unresolved algorithm failures.

ping purposes we employ the optimization methods in exactly the same way, except that we now strive to minimize the squared norm of the resampled weighted residual, $\|\mathbf{JWr}\|^2$. We have tested this method extensively on linear problems with correlated Gaussian noise and find excellent agreement between the bootstrapped uncertainties and those predicted by linear propagation of errors (see Figure 7).

We expect the actual parameter values to fall outside the 95% confidence intervals ~ 1 time in 20. To ensure that the bootstrap was performing consistently with this expectation, we conducted a simple test: We added 50 different noise vectors, each derived from the same covariance matrix, to the displacement data vector for test case 1. We then obtained bootstrap confidence intervals (using a quasi-Newton bootstrap, with $B = 2000$) based on each of the 50 new noisy data vectors. In effect, we estimated the confidence intervals for 450 (9 x 50) parameters to see how often they fell outside their confidence intervals. Figure 8 plots the number of times (as a percent of the total) that a parameter fell outside its bootstrap confidence interval against the expectation based on α, where (1-α)% defines the confidence interval. The "correct" answer is plotted as the diagonal line; that is, if the bootstrap is performing as designed, we would expect the crosses to plot on this line. Note that bootstrap seems to overestimate confidence intervals slightly (e.g., when $\alpha = 0.05$, we expect the true model parameter to fall outside its interval 5% of the time; in fact, the percentage is closer to 2.5%). Raising the number of resamples, B, did not eliminate the disagreement.

There are several possible explanations for the overestimation of the bootstrap confidence intervals. First, the implicit assumption in using a quasi-Newton algorithm to find the minimum in the misfit space of the resampled data is that the prior model inverted from the actual data is close to (i.e., in the same minimum as) the optimal model under the resample. To the extent that this assumption is incorrect, the confidence intervals will be biased toward being too large since the quasi-Newton algorithm will converge to a suboptimal model in the resampled misfit space. Another possible explanation for this overestimation is that this relatively small data set is not sufficient to completely illuminate the tails of the posterior probability distribution. That is, the expectation that we can specify precise 95% confidence intervals is not reasonable given the data strength of this case.

5.3. Using the Bootstrap

Geodetic data often consist of vector quantities, e.g., displacement vectors. The theory behind the bootstrap requires that resampling mimic actual data collection, so vector quantities should be resampled as whole vectors rather than by their individual components. To take the GPS example, the components of the residual vector will come in triples, e.g., east, north and up, corresponding to one particular receiver. These triples should not be separated but rather would be resampled as units.

Each bootstrap resample of the data creates a new p-dimensional misfit space. In general, the new space has a different set of minima (and a different global minimum)

from the cost space of the original data. Hence each bootstrap resample poses a new nonlinear optimization problem. Unfortunately, the Monte Carlo optimization methods discussed here cannot be considered as functions. That is, they do not necessarily return a unique answer for a given set of arguments. Because these algorithms contain an element of randomness and because they operate over finite timescales, there is no guarantee of convergence to the same minimum at every execution. In principle, they will always find the global minimum of misfit space, but in practice, this is often not the case. For these reasons, great care must be taken when applying the bootstrap to model parameters estimated with nonfunctional algorithms. We would like bootstrap confidence intervals for the estimated parameters that reflect the limitations of the data, not the limitations of the estimation algorithm.

We propose two alternate methods of applying the bootstrap to the results of Monte Carlo optimization. The first method applies when the signal-to-noise ratio appears very high, $\gg$100, which probably indicates that the misfit space has a large, well-defined global minimum. Under these circumstances a bootstrap resample will probably not change the gross shape of cost space, changing instead only the precise location of the global minimum. For this reason, a derivative-based method, with the best model from the original optimization as an initial guess, should find the new global minimum easily. Derivative-based algorithms are functional, so the bootstrap confidence intervals found with them should not be biased by algorithm failures. We have applied this bootstrapping method to test case 1 and have found that for each resample, $B = 2000$, a Levenberg-Marquardt algorithm converged to the new minimum in a very small number of iterations. The danger in using this method lies in its dependency on a starting guess. Should the resampled cost space contain a global minimum quite different from the original minimum, a derivative-based method may never find it.

The alternative method to applying the bootstrap is to perform a full Monte Carlo optimization at each resample. In general, this method provides much greater robustness but at the expense of drastically increased computational time. Circumventing the nonfunctional nature of Monte Carlo algorithms and preserving the data dependence of the bootstrap confidence interval estimates requires a lengthy and detailed bootstrapping process. Begin as normal by forming a resampled data vector. Next, perform a Monte Carlo optimization (preferably a hybrid) and save the best model and its misfit (evaluated using the bootstrap resample). Also save the resampled data vector. After B resamplings, review the resulting B misfit values looking for any that look abnormally high. These probably represent algorithm failures. Rerun the Monte Carlo optimization on all the resamples that failed to converge (i.e., those that had abnormally high misfits), using the same resampled data vector. This amounts to giving the Monte Carlo algorithm more time to find the global minimum for a particular resample. Repeat this procedure until satisfied that a global minimum has been reached at each resampling.

The bootstrap can also estimate, at least in a qualitative way, the correlations between parameters. A bootstrapping run produces B estimates (based on B data resamplings) for each model parameter. For each parameter pair, form a scatter plot of all the bootstrap estimates. The resulting shapes provide an excellent way of visualizing the correlation between the parameter pairs. Unlike the linear case, there is no unique way of defining a 95% confidence region. Nonetheless, plots of this kind can set the individual confidence intervals into the proper context and help prevent the acceptance of unreasonable models.

5.4. Bootstrap Applied to Test Cases

We have applied both bootstrapping methods to the previously described test cases. Test case 1 has a very high signal-to-noise ratio ($\sim$ 660), so we used the derivative-based bootstrapping method described above. Tables 1 and 2 show the confidence intervals obtained using the bootstrap. Notice that in each case, the bootstrap confidence interval brackets the actual value (from the hybrid algorithm). Figure 9 depicts a matrix of scatter plots that show the correlations between the various model parameters. The bottom row of Figure 9 shows histograms for each of the individual model parameters, with vertical lines marking the bootstrap confidence intervals. Recall that only Gaussian noise was added to the synthetic displacement data for both test cases. Nonetheless, the nonlinearity of the problem means that the model parameters are not normally distributed. The histograms reflect this fact, particularly for length and east offset, both of which show marked asymmetry. Moreover, the scatter plots depict complicated shapes that only vaguely resemble ellipses. However, compared to the results from test case 2 presented below, both the histograms and the scatter plots reflect a relatively simple probability distribution for the model parameters. In our experience, this is typical for cases with a high SNR.

Figure 10 depicts a scatter plot matrix for the bootstrap performed on test case 2, which has a low SNR ($\sim$ 5). For this reason, we used Monte Carlo estimation for each resample. Two features appear most striking: The two-dimensional scatter plots do not form even vaguely elliptical shapes, and the shapes they do form are discontinuous. This suggests the presence of at least two almost equally deep minima in misfit space, corresponding to the actual fault plane and its conjugate. The one-dimensional confidence interval plots also reflect this bimodal distribution, most clearly in the dip and the east and north offsets. None of the model parameters in this test case are well constrained, particularly when compared to the results from the first case. However, careful examination of the histograms shows that the actual parameter value lies beneath the distribution peak for each unimodal distribution. Moreover, for the bimodal distributions the actual value lies below the peak of the larger mode. In this sense, the bootstrap not only reveals how well the model parameters are resolved but also offers a qualitative distribution of the parameter values.

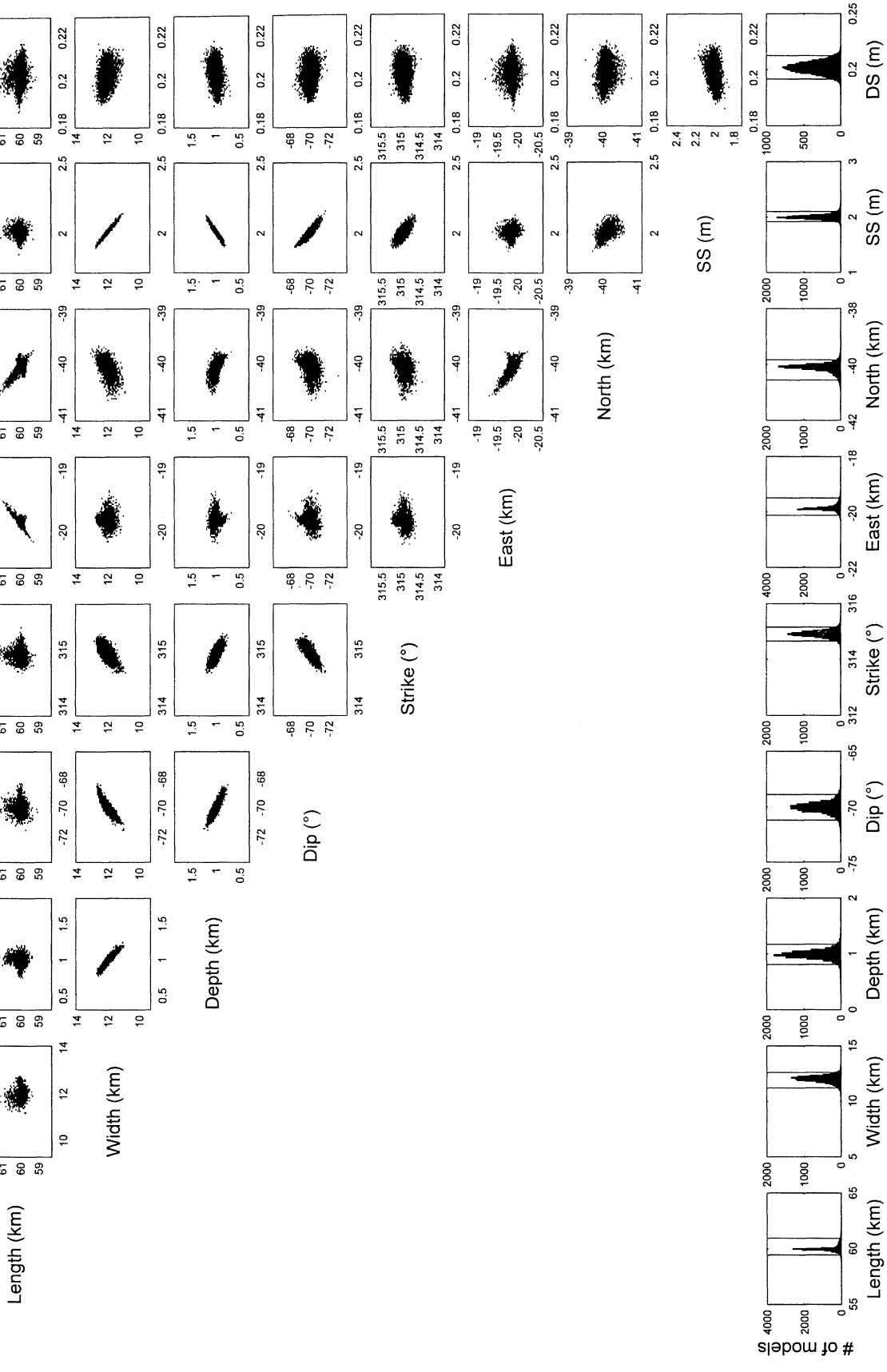

Figure 9. A covariance scatter plot for test case 1. The bottom row depicts the a posteriori distribution of the model parameters. The other rows show correlations between parameter pairs. The vertical lines shown in the last row bracket the 95% individual confidence intervals.

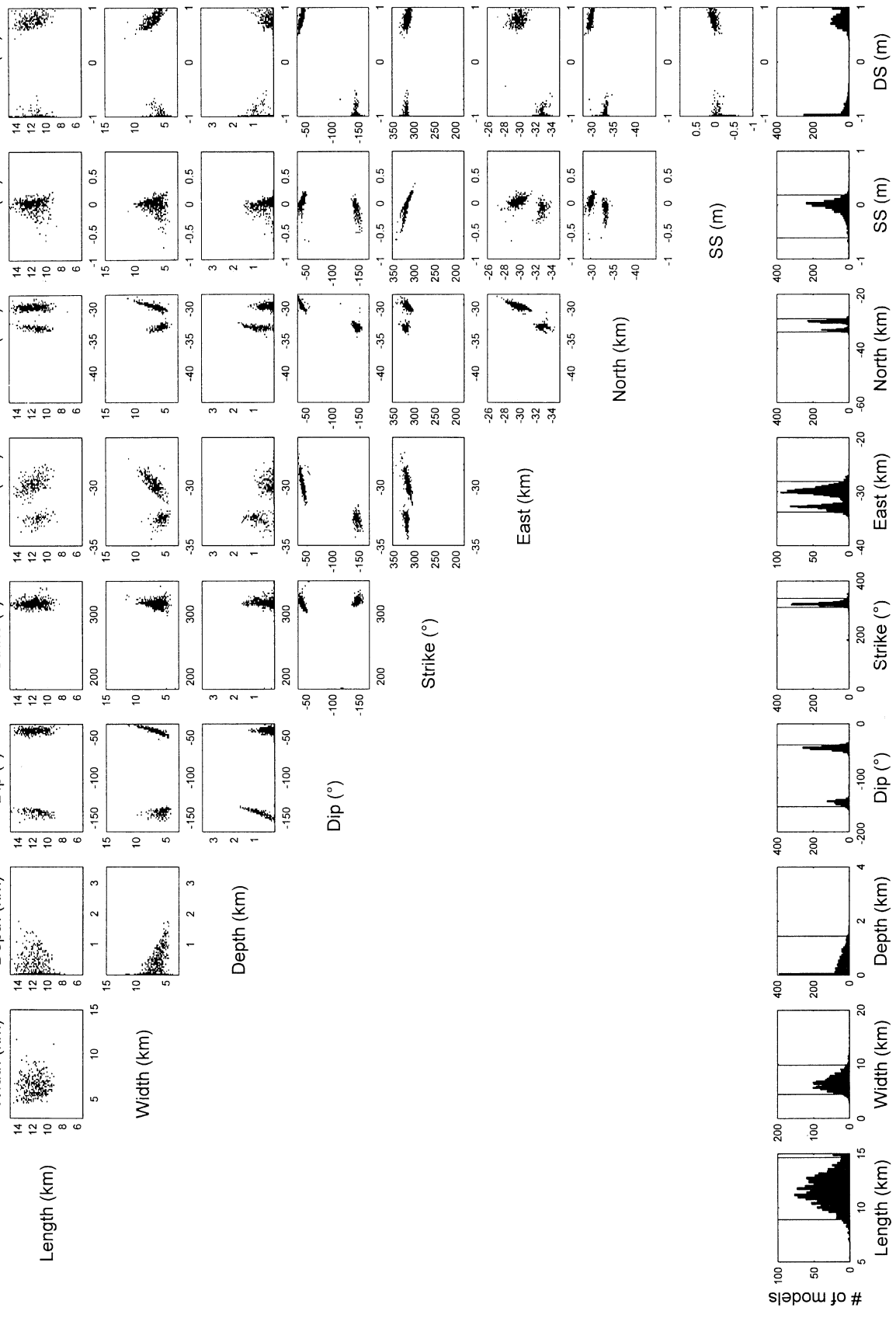

Figure 10. A covariance scatter plot for test case 2. The bottom row depicts the a posteriori distribution of the model parameters. The other rows show correlations between parameter pairs. The vertical lines shown in the last row bracket the 95% individual confidence intervals.

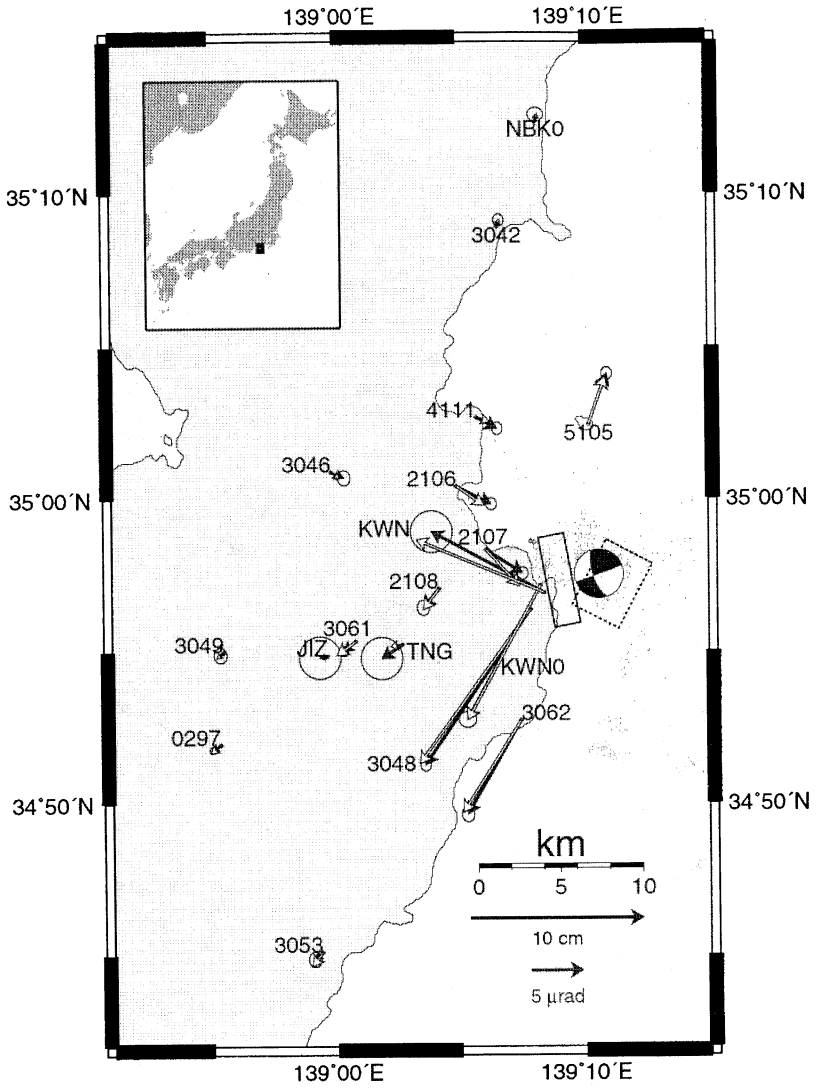

Figure 11. Observed and predicted deformation from the March 1997 Izu Peninsula (see inset for location of Izu Peninsula) earthquake swarm. Black vectors depict GPS and tilt measurements (tilt stations have three-character IDs); white vectors are model predictions. Error ellipses are 2σ. The two planes are the surface projections of the shear source (solid) and the dike (dashed). The focal mechanism is from the largest event in the swarm, $M5.3$.

6. Application to the March 1997 Earthquake Swarm off Izu Peninsula, Japan

6.1. Introduction

Aoki et al. [1999] used the simulated annealing algorithm to estimate the source geometry for the 1997 Izu Peninsula earthquake swarm, which they interpreted as arising from a dike intrusion and a strike-slip fault. After estimating the source geometry they solved for the spatial and temporal evolution of slip and opening on these structures. Here, we extend the geometry estimation part of their work, applying both random cost and simulated annealing to the problem. We compare the two algorithms to one another for efficiency and ease of use and address the efficacy of hybrid methods. Further, we correct an error in the forward calculation of the misfit function that led to a biased estimate of the source geometry of *Aoki et al.* [1999]. Our preferred model, discussed in detail below, differs from the model of *Aoki et al.* [1999] as follows: Our model fault dips more shallowly and lies farther to the east, and our model dike has a taller aspect ratio, less area, and more opening.

6.2. Geological Background

The Izu Peninsula, located in central Japan (Figure 11), has been the source of numerous seismic swarms, with one occurring almost every year since 1978. A swarm in 1989 led to a submarine eruption [*Okada and Yamamoto*, 1991]. This study targets the swarm that began on March 2, 1997, and continued for $\sim$ 10 days. The onset of activity occurred at 1500 (UTC) near the coastline and the center of activity moved seaward at around 2300 (UTC). The earthquake focal depths are concentrated in the upper 10 km

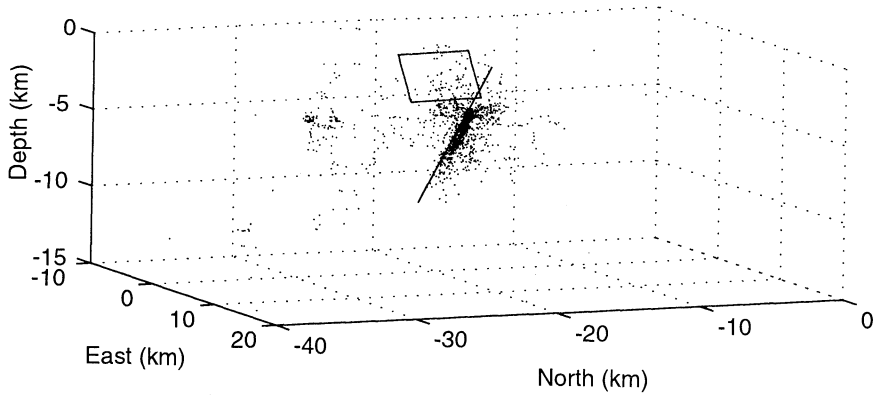

Figure 12. A depiction of the preferred model in 3-D perspective. The plane shown in cross section represents the dike, the plane shown nearly head on is the shear source. Catalog seismicity is also plotted. The view is parallel to the strike of the dike (roughly toward the northwest). Note good agreement between the dip of the dike and the seismicity.

of the crust (Figure 12). Seismic activity subsided after March 9, 1997.

Almost all of the earthquakes have P axes oriented NW-SE, which is consistent with the regional tectonic stress pattern [*Ukawa,* 1991]. This swarm was considerably more energetic than previous swarms, with 25 earthquakes larger than $M4$ and 4 larger than $M5$ [*Earthquake Research Institute,* 1997]. The largest earthquake $M5.3$ occurred at 0351 (UTC) on March 4. The hypocentral distribution of the events is shown in Figure 12.

Significant crustal deformation was also associated with the swarm (Figure 11). Permanent Global Positioning System (GPS) observations show $\sim$ 120 mm of NE-SW extension between Hatsushima island and the Izu Peninsula (Figure 11). Repeated leveling reveals a maximum uplift of 30 mm along the east coast of the Izu Peninsula (Figure 13). Continuously recording borehole tiltmeters showed significant tilts, reaching a maximum of 13 μrad NW tilt at the station nearest the swarm, KWN (Figure 11). Many researchers have suggested a volcanic origin for the swarm activity off the Izu Peninsula. For example, *Okada and Yamamoto* [1991] have shown that the deformation associated with the 1989 event resulted from a combination of an intruding dike and the coseismic offset from the swarm's largest earthquake, $M5.3$. Generally, swarm activity off the Izu Peninsula lasts from 2 to 10 days. This activity is characterized by a sudden burst of seismicity followed by a gradual decrease. Although *Okada and Yamamoto* [1991] were able to satisfactorily fit the crustal deformation from the 1989 event using forward modeling, the detailed processes operative during the swarms are still not well understood. Recently, however, the num-

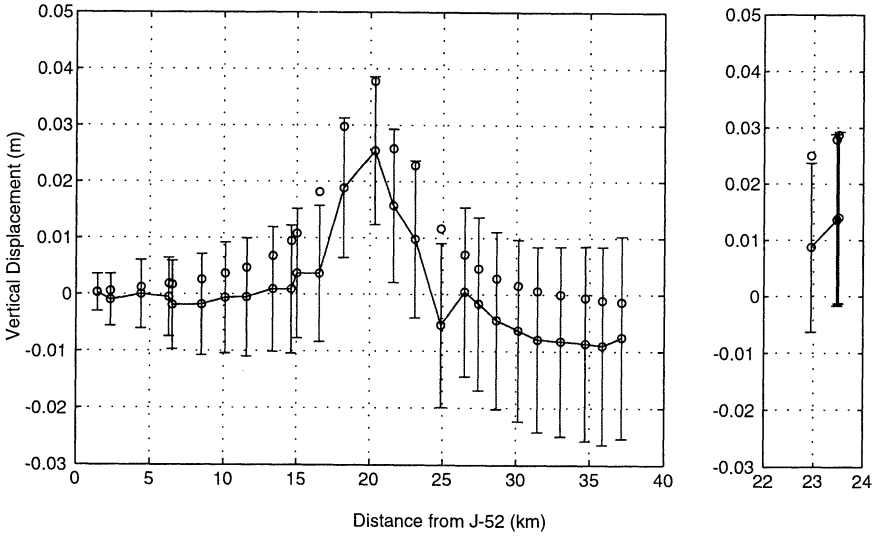

Figure 13. Observed and predicted vertical deformation from the March 1997 Izu Peninsula earthquake swarm. Circles connected by a line depict observed height changes from leveling measurements. The remaining circles depict model prediction. Error bars are 2σ. (left) Main leveling line; (right) spur (see Figure 11). Note that because of correlations in the leveling data, it is effectively the derivative of the profile that is fit by the inversion.

ber and density of permanent geodetic stations, including GPS and borehole strain and tilt, have increased tremendously in Japan [e.g., *Kato et al.,* 1998], so that we now can model the crustal deformation with higher resolution [e.g., *Aoki et al.,* 1999]. Here we estimate the geometry of the sources active during the March 1997 seismic swarm using the methods discussed previously.

6.3. Data

The available geodetic data consist of GPS, leveling, and tilt measurements (see Figure 11 for station locations). In the eastern part of the Izu Peninsula, 12 permanent GPS stations are operated by the Geographical Survey Institute of Japan (GSI) and one station is operated by the National Institute of Earth Science and Disaster Prevention of Japan (NIED). Site coordinates were estimated once a day. Displacements were estimated by taking 5-day weighted averages of the station positions centered on February 23, 1997, and March 22, 1997, dates chosen to bracket the active deformation period. The formal uncertainties were scaled by the repeatability about the mean over the two 5-day periods, when little or no deformation is presumed to have occurred. The repeatability is $\sim$ 3-4 mm for the horizontal components and around 10 mm for the vertical components. Although GPS time series can contain temporally correlated errors, which have been modeled as either random walk or flicker noise *Zhang et al.,* 1997; *Mao et al.,* 1999], this effect is safely neglected since the time interval is so short. For this data set the contribution from colored noise is at most submillimeter.

Leveling surveys were also conducted by GSI in October 1996 and March 1997 (Figure 13). The only activity that occurred during this interval was the March 1997 swarm, to which we attribute all measured height changes. The covariance of the leveling data is calculated following *Árnadóttir et al.* [1992], and scaled according to the circuit misclosure in the surveys by $\sigma = 1.20$ mm/$\sqrt{\text{km}}$. The error in the tilt data is reasonably approximated by Gaussian random walk [*Wyatt et al.,* 1988].

We analyzed the power spectrum from a long run of tilt data not affected by swarm events and concluded that the tilt errors could be adequately modeled by a sum of white noise with $\sigma_W = 0.01$ μrad and a random walk process with scale $\sigma_{RW} = 3$ μrad/$\sqrt{\text{yr}}$, so that the variance in the tilt error is

$$\sigma^2_{\text{total}} = \sigma^2_W + \sigma^2_{RW}\sqrt{t}. \quad (8)$$

Using a value for time of $t = 0.077$ years (28 days) and the values for the other terms given above, (8) yields a value of σ_{total} of 0.83 μrad. The two components of the tilt vector were assumed to be independent.

6.4. Model

We first attempted to model the deformation with a single dike source, represented by a uniform opening dislocation. We placed very broad bounds on the parameter values (Table 3), since we had little a priori information about the deformation source(s). The results of two inversion runs, one for simulated annealing (model a), the other for random cost (model b), are given in Table 3. Note that the source parameter estimates differ significantly between the two algorithms. However, both source parameter estimates, if used as a priori values for a derivative based algorithm, lead to the same final model (model c) in Table 3. This reinforces our conclusion from the tests on synthetic data that Monte Carlo algorithms operate most effectively when combined with a derivative based algorithm. We repeatedly ran both algorithms to explore the misfit space thoroughly. A strong global minimum, corresponding to model c in Table 3, dominates the misfit space, though several weaker local minima also appear. Before finding the vicinity of the global minimum, the random cost algorithm required a factor of $\sim$ 3 more calls to the misfit function than simulated annealing. This is in keeping with our observation from the synthetic data that the random cost algorithm performs comparatively poorly when the misfit space is dominated by a large global minimum.

The NW-SE trending dike represented by model c in Table 3 is consistent with the regional tectonic stress field [*Ukawa,* 1991]. The single dike model fits northeast extension in the GPS data but does not explain the southeast directed motion at GPS sites along the east coast of the Izu Peninsula, north of the seismic swarm (Figure 11). These dike parallel motions are not predicted by any dike-like source. The MSE associated with model c was $\sim$ 11. Permitting both shear and opening on a single dislocation lowered the misfit significantly (MSE 5.6), but we rejected this model outright for its physical implausiblility. Dikes

Table 3. Source Parameters for the March 1997 Izu Peninsula Earthquake Swarm for Dike Only

	Inversion Bounds		Models		
Parameter	Lower	Upper	a	b	c
Length, km	0.01	25	7.63	9.21	7.06
Width, km	0.01	15	6.91	4.22	6.87
Depth to top, km	0	5	0.16	0.74	0.21
Dip, deg	45	135	89.5	90.4	90.0
Strike, deg	240	350	300	298	302
East offset, km	2	15	5.44	6.79	5.44
North offset, km	-30	-10	-23.0	-22.3	-22.9
Opening, m	0	5	0.34	0.52	0.38
MSE			14.1	16.3	11.2

Table 4. Source Parameters for the March 1997 Izu Peninsula Earthquake Swarm for Dike and Fault

Parameter	Inversion Bounds		Models	
	Lower	Upper	a	b
Dike				
Length, km	0.01	25	8.65	3.17
Width, km	0.01	15	0.744	9.67
Depth to top, km	0	5	0.992	0.80
Dip, deg	45	135	73.2	62.6
Strike, deg	240	350	298	298
East offset, km	2	15	6.26	7.94
North offset, km	-30	-10	-22.0	-19.7
Opening, m	0	5	1.90	0.60
Fault				
Length, km	1	15	6.57	5.53
Width, km	0.01	15	1.40	3.30
Depth to top, km	0	6	4.99	0.29
Dip, deg	0	180	50	62
Strike, deg	310	25	353	349
East offset, km	0	15	5.5	2.92
North offset, km	-30	-15	-21.6	-21.6
Strike Slip, m	-1.5	1.5	-1.50	-0.49
MSE			2.8	4.1

are expected to intrude normal to the least compressive stress and are unlikely to exhibit substantial shear displacement.

To capture the presumably independent shearing motion associated with the swarm earthquakes, we increased the complexity of the model and included a second dislocation in the inversion. As above, we set very broad bounds on the parameter values (Table 4). Introducing eight new dimensions to the misfit space eliminated the single strong global minimum in favor of several different weak minima each having approximately the same depth (MSE < 5). In terms of algorithm efficiency, random cost was the clear winner in this case. Simulated annealing took about twice as many calls to the misfit function for consistent convergence to one of the low-valued minima. We found that by perturbing the critical temperature slightly from the nominal value provided by our shortcut method, we could improve the effectiveness of the algorithm. Indeed, in our implementation of the simulated annealing algorithm the default behavior is to make the shortcut guess but then to run the algorithm three times at three different critical temperatures that bracket the nominal value. Random cost reliably found low-valued minima, often with surprisingly few calls to the misfit function ($< 10,000$). Again, this is consistent with our experience from the test cases, which indicated that random cost performs well in misfit spaces characterized by multiple minima.

Using hybrid versions of both algorithms, we assembled a list of $\sim$ 20 source models all having low misfit (MSE < 5). We then examined each of the models, judging them for consistency with the earthquake locations and mechanisms, physical and geological plausibility, and the degree and nature of the misfit. We do not use the MSE alone as the sole criterion for choosing a source model for two main reasons: First, the feasible region of the misfit space is defined very simplistically: we impose upper and lower bounds on the value of each parameter but impose no constraint equations that interrelate parameters with one another (e.g., excluding models that have stress changes that exceed a certain value). This is analogous to the army imposing on its recruits height limits of 150-200 cm and weight limits of 60-140 kg. Under such a plan, soldiers who were 155 cm tall, but weighed 135 kg would be deemed battleworthy. Likewise, models may be permissible, in that none of their parameters fall outside the bounds but at the same time be rejectable for other reasons. Second, not all the available data are accounted for in the misfit function. There are qualitative geologic concerns that would be very hard to include, but there are also quantitative measurements like the size and location of earthquakes that are not modeled.

Our preferred source model for the earthquake swarm is listed as model b in Table 4. There is very good agreement between the location and orientation of the dike source and the seismicity (Figures 11 and 12). The shear source agrees fairly well with the focal mechanism of the largest earthquake, though there is a discrepancy in dip direction. However, the shear source does lie in a region of seismicity separate from the main cloud that presumably arose from the dike intrusion. The deformation data are fit well by this model, with no evidence of any systematic misfit (Figures 11 and 13). Though the plot of leveling data (Figure 13) appears to indicate systematic overfit, this is a plotting artifact. Because of the correlations associated with leveling, it is essentially the first derivative of the data that is fit.

We also list another model, model a in Table 4, which we believe represents the minimum in misfit space for this

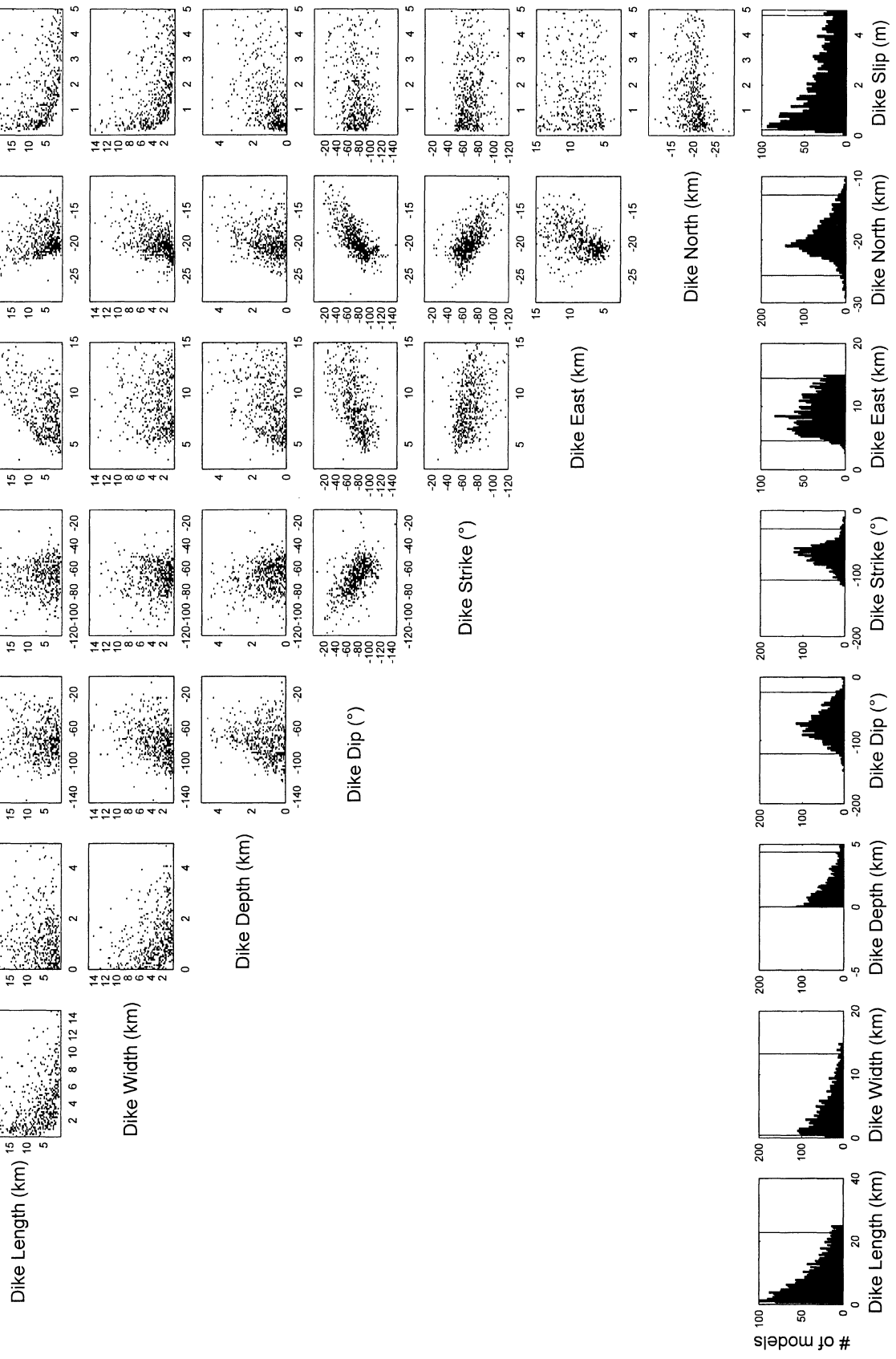

Figure 14. A covariance scatter plot for the dike from the preferred Izu model. The bottom row depicts the a posteriori distribution of the model parameters. The other rows show correlations between parameter pairs. The vertical lines in the bottom row bracket the 95% individual confidence intervals.

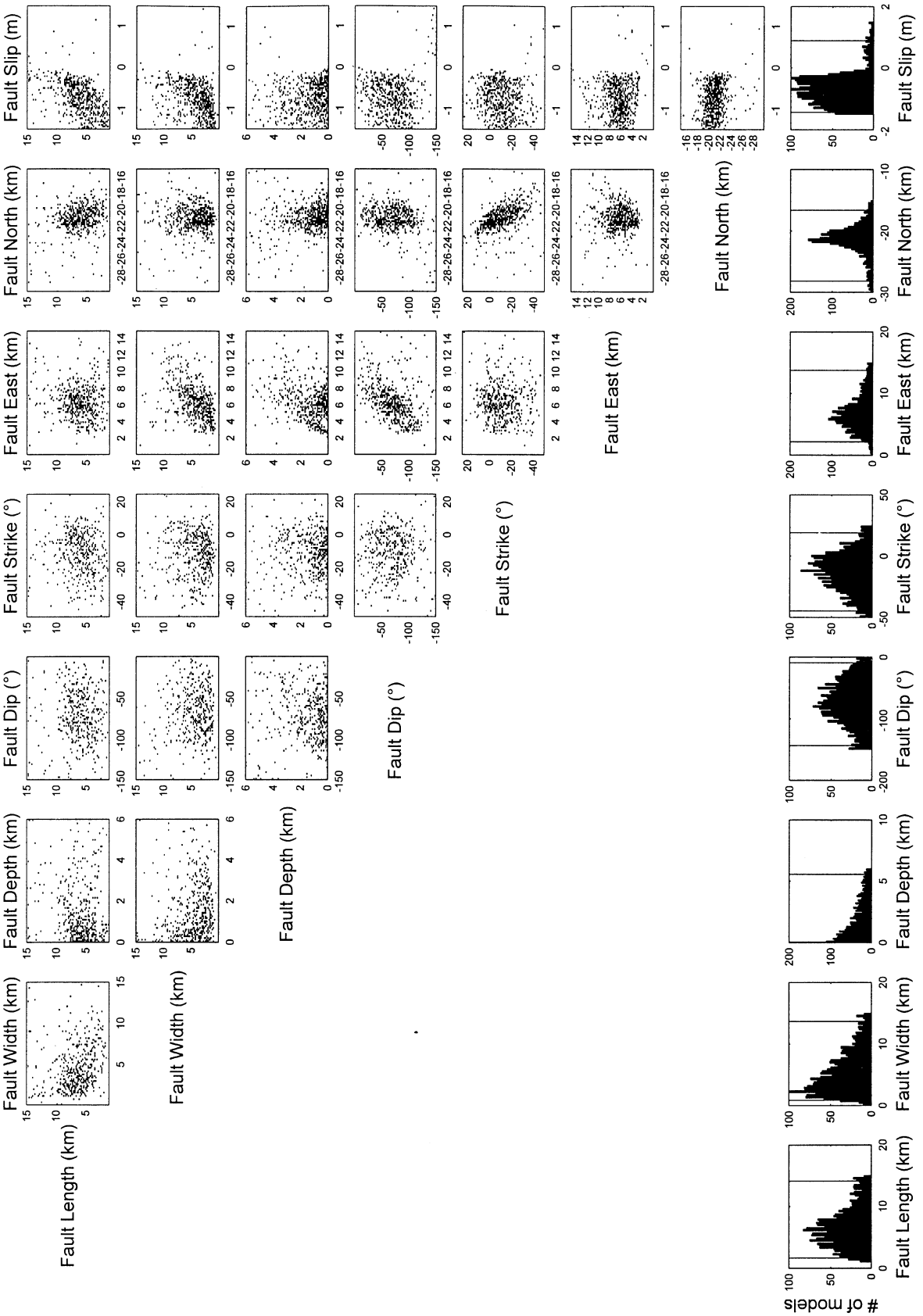

Figure 15. A covariance scatter plot for the fault from the preferred Izu model. The bottom row depicts the a posteriori distribution of the model parameters. The other rows show correlations between parameter pairs. The vertical lines in the bottom row bracket the 95% individual confidence intervals.

problem. Despite its low misfit (MSE = 2.9), we rejected this model for a number of reasons. First, the model shows very poor agreement with both the location and orientation of the swarm seismicity. Second, the model dike reaches the surface, but there was no evidence of a submarine eruption. Third, the dimensions of the model dike are very small for such a large amount of opening. (This suggests that a constraint on the dike overpressure might screen for unrealistic models). Fourth, the moment of the shear source ($M_w = 5.8$) seems quite high.

To estimate uncertainties on the model parameters, we performed a bootstrap analysis as described above. We performed 2000 resamples and used a hybrided simulated annealing algorithm at each iteration to find the minimum of the resampled misfit space. To enhance algorithm performance, we first used the rigorous method of *Basu and Frazer* [1990] to estimate the critical temperature. The results are depicted in Figure 14 and Figure 15. The model parameters are quite poorly constrained by the data set, which is not surprising given that nearly all of the data were measured on only one side of the earthquake swarm. In addition to the obvious trade-offs between the area of the deformation sources and their opening/slip, there are several others worth noting: Both the fault and the dike show significant trade-off between their spatial coordinates and their orientation. This suggests additional source complexity; for example, perhaps the dike propagated as distinct segments or shearing occurred on several different planes. Alternatively, this could reflect limitations in the data, given their uneven distribution.

7. Conclusions

We have applied Monte Carlo optimization methods to the problem of estimating source parameters from deformation data. The misfit space that characterizes this optimization problem often contains multiple minima, making the use of derivative-based methods impractical. We have tested two very different Monte Carlo algorithms: simulated annealing and random cost. Our testing on synthetic data sets indicates that both of these algorithms can solve the global optimization problem of source parameter estimation. The simulated annealing algorithm is more complicated to implement than the random cost algorithm, though our test cases indicate that the former algorithm is more robust than the latter. We have presented a method for rapidly approximating the critical temperature, which is an important metaparameter for simulated annealing.

We have used the bootstrap method of approximating the a posteriori covariance of the estimated source parameters. Earlier implementations of the bootstrap did not account for correlated data. We have developed a resampling technique that properly handles correlation, which we have empirically verified with tests on linear problems where the a posteriori covariance is known analytically. Understanding the implications of the a posteriori covariance of the estimated source parameter can be very difficult; because of the nonlinearity of the problem the a posteriori covariance is not a matrix, nor does it describe a hyperellipsoid. We present a visualization method that provides both a qualitative representation of the correlations among the estimated source parameters and a depiction of approximate two-dimensional confidence regions.

Finally, using the optimization methods described above, we have analyzed deformation data resulting from an earthquake swarm off of the Izu Peninsula, Japan. This analysis led to an interpretation of the earthquake swarm as a dike intrusion and associated shear on a nearby fault. The deformation data, consisting of GPS observations, leveling measurements, and readings from borehole tiltmeters, are fit well by these source models. Moreover, the source locations agree with both the earthquake hypocenters and the purported regional stress field. To find estimates of the confidence intervals for the source model parameters, we used the bootstrap, the results of which are depicted by Figures 14 and 15.

Acknowledgments. NASA grant NAG5-6167-0001 provided the funding for the research. We acknowledge Mark Matthews for helpful advice about the bootstrap method, Susan Owen for extensive discussion about inversion in general, Jessica Murray for substantive commentary and proof reading, and the reviewers, Jeffrey Freymueller, Kristine Larson, and Richard Bennett, for recommendations that greatly improved the paper.

References

Aoki, Y., P. Segall, T. Kato, P. Cervelli, and S. Shimada, Upward dike migration during the 1997 seismic swarm off the Izu Peninsula, Japan, from inversion of deformation data, *Science, 286*, 927–930, 1999.

Árnadóttir, T., P. Segall, and M. Matthews, Resolving the discrepancy between geodetic and seismic fault models for the 1989 Loma Prieta, California, earthquake, *Bull. Seismol. Soc. Am., 82*, 2248–2255, 1992.

Basu, A., and L. N. Frazer, Rapid determination of the critical temperature in simulated annealing inversion, *Science, 249*, 1409–1412, 1990.

Berg, B., Locating global minima in optimization problems by a random-cost approach, *Nature, 361*, 708–710, 1993.

Bock Y., et al., Southern California Permanent GPS Geodetic Array: Continuous measurements of crustal deformation between the 1992 Landers and 1994 Northridge earthquakes, *J. Geophys. Res., 102*, 18,013–18,033, 1997.

Cruetz, M., Monte-Carlo study of quantized SU(2) gauge theory, *Phys. Rev. D, 21*, 2308–2315, 1984.

Dixon, T. H., et al., Continuous monitoring of surface deformation at Long Valley Caldera, California, with GPS, *J. Geophys. Res., 102*, 12,017–12,034, 1997.

Earthquake Research Institute, Seismic activities in the Izu Peninsula and its vicinity (in Japanese), *Rep. Coord. Comm. Earthquake Prediction, 58*, 239–249, 1997.

Efron, B., and R. J. Tibshirani, *An Introduction to the Bootstrap*, 436 pp., Chapman and Hall, New York, 1993.

Freymueller, J. T., M. H. Murray, P. Segall, and D. Castillo, Kinematics of the Pacific-North America plate boundary zone, northern California, *J. Geophys. Res., 104*, 7419–7441, 1999.

Gill, P. E., W. Murray, and M. H. Wright, *Practical Optimization*, 401 pp., Academic, San Diego, Calif., 1981.

Jónsson, S., H. Zebker, P. Cervelli, P. Segall, H.Garbeil, P. Mouginis-Mark, and S. Rowland, A shallow-dipping dike fed the 1995 flank eruption at Fernandina Volcano, Galapagos, observed by satellite radar interferometry, *Geophys. Res. Lett., 26*, 1077–1080, 1999.

Kato, T., G. S. El-Fiky, E. N. Oware, and S. Miyazaki, Crustal

strains in the Japanese Islans as deduced from dense GPS array, *Geophys. Res. Lett., 25*, 3445–3448, 1998.

Mao, A., C. G. A. Harrison, and T. H. Dixon, Noise in GPS coordinate time series, *J. Geophys. Res., 104*, 2797–2816, 1999.

Marshall, G.A., R.S. Stein, and W. Thatcher, Faulting geometry and slip from co-seismic elevation changes: the 18 October 1989, Loma Prieta, California, Earthquake, *Bull. Seismol. Soc. Am., 81*, 5, 1660–1693, 1991.

Massonnet, D., and K. L. Feigl, Radar interferometry and its application to changes in the Earth's surface, *Rev. Geophys., 36*, 441–500, 1998.

Metropolis, N., A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller, Equation of state calculations by fast computing machines, *J. Chem. Phys., 21*, 1087–1092, 1953.

Mogi, K., Relations between the eruptions of various volcanoes and the deformations of the ground surfaces around them, *Bull. Earthquake Res. Inst., Univ. Tokyo, 36*, 99–134, 1958.

Murray, M. H., G. Marshall, M. Lisowski, R. Stein, The 1992 M=7 Cape Mendocino, California, earthquake: Coseismic deformation at the south end of the Cascadia megathrust, *J. Geophys. Res., 101*, 17707–17725, 1996.

Okada, Y., Surface deformation due to shear and tensile faults in a half-space, *Bull. Seismol. Soc. Am., 75*, 1135–1154, 1985.

Okada, Y., and E. Yamamoto, A Model for the 1989 Seismo-Volcanic Activity off Ito, Central Japan, derived from crustal movement data, *J. Phys. Earth, 39*, 177–195, 1991.

Owen, S., P. Segall, J. Freymueller, A. Miklius, R. Denlinger, T. Arnadottir, M. Sako, and R. Bürgmann, Rapid deformation of the south flank of Kilauea volcano, Hawaii, *Science, 267*, 1328–32, 1995

Owen, S., P. Segall, M. Lisowski, M. Murray, M. Bevis, and J. Foster, January 30, 1997 eruptive event on Kilauea Volcano, Hawaii, as monitored by continuous GPS, *Geophys. Res. Lett., 27*, 2757–2760, 2000.

Rothmann, D.H., Nonlinear inversion, statistical mechanics, and residual statics estimation, *Geophysics, 50*, 2784–2796, 1985.

Ukawa, M., Collision and fan-shaped compressional stress pattern in the Izu block at the northern edge of the Philippine Sea Plate, *J. Geophys. Res., 96*, 713–728, 1991.

Ward, S. N., and S. E. Barrientos, An inversion for slip distribution and fault shape from geodetic observations of the 1983, Borah Peak, Idaho, earthquake, *J. Geophys. Res., 91*, 4909–4919, 1986.

Wyatt, F. K., S. Morrissey, and D. C. Agnew, Shallow borehole tilt: A reprise, *J. Geophys. Res., 93*, 9197–9201, 1988.

Yu, T. T., and J. B. Rundle, Inverting for fault zone geometry using genetic algorithms (abstract), International Union of Geodesy and Geophysics; XXI General Assembly, Week B56, 1995.

Zhang, J., Y. Bock, H. Johnson, P. Fang, S. Williams, J. Genrich, S. Wdowinski, and J. Behr, Southern California permanent GPS geodetic array: Error analysis of daily position estimates and site velocities, *J. Geophys. Res., 102*, 18,035–18,055, 1997.

Y. Aoki and T. Kato, Earthquake Research Institute, University of Tokyo, 1-1 Yayoi 1 Bunkyo-ku Tokyo 112, Japan. (aoki@eri.u-tokyo.ac.jp; teru@eri.u-tokyo.ac.jp)

P. Cervelli and P. Segall, Department of Geophysics, Stanford University, Stanford, CA 94035-2215. (cervelli@stanford.edu; segall@stanford.edu)

M. H. Murray, Seismological Laboratory, University of Califonia, 215 McCone Hall, Berkeley, CA 94720-4760. (mhmurray@seismo.berkeley.edu)

(Received January 27, 2000; revised October 25, 2000; accepted November 1, 2000.)