

Constraining Earthquake Source Inversions with GPS Data 1: Resolution Based Removal of Artifacts

Morgan T. Page

U.S. Geological Survey, Pasadena, California, USA
Department of Physics, University of California at Santa Barbara, Santa Barbara, California, USA

Susana Custódio

Institute for Crustal Studies and Department of Earth Science, University of California at Santa Barbara, Santa Barbara, California, USA

Ralph J. Archuleta

Institute for Crustal Studies and Department of Earth Science, University of California at Santa Barbara, Santa Barbara, California, USA

J. M. Carlson

Department of Physics, University of California at Santa Barbara, Santa Barbara, California, USA

Abstract.

We present a resolution analysis of an inversion of GPS data from the 2004 M_w 6.0 Parkfield Earthquake. This earthquake was recorded at 13 1-Hz GPS receivers, which provides for a truly co-seismic dataset that can be used to infer the static-slip field. We find that the resolution of our inverted slip model is poor at depth and near the edges of the modeled fault plane that are far from GPS receivers. The spatial heterogeneity of the model resolution in the static field inversion leads to artifacts in poorly resolved areas of the fault plane. These artifacts look qualitatively similar to asperities commonly seen in the final slip models of earthquake source inversions, but in this inversion they are caused by a surplus of free parameters. The location of the artifacts depends on the station geometry and the assumed velocity structure. We demonstrate that a nonuniform gridding of model parameters on the fault can remove these artifacts from the inversion. We generate a nonuniform grid with a grid spacing that matches the local resolution length on the fault, and show that it outperforms uniform grids, which either generate spurious structure in poorly resolved regions or lose recoverable information in well-resolved areas of the fault. In a synthetic test, the nonuniform grid correctly averages slip in poorly resolved areas of the fault while recovering small-scale structure near the surface. Finally, we present an inversion of the Parkfield GPS dataset on the nonuniform grid and analyze the errors in the final model.

1. Introduction

Kinematic inversions of seismic data are routinely used to create models of the temporal evolution of slip on finite faults. While kinematic inversions remain our best tool with which to image the earthquake source, they are problematic because quantifying model error is difficult. Traditional estimates of goodness of fit to seismic data are not necessarily reliable measures of model error. Exact solutions may depend on poorly determined features of the data, and thus limiting free parameters can improve the model [Jackson, 1972]. Furthermore, constraints and stability criteria are necessary to produce physically meaningful solutions [Olson and Apsel, 1982]. Still, a stable solution that matches physical constraints is not necessarily closely related to the true slip distribution. This is evident in comparing source models for a given earthquake produced by competing research groups. Alternative models that fit the data may appear to be virtually uncorrelated [Mai, 2007].

Due to the limited number of seismic stations close to the earthquake source, in practice the inverse problem is often

underdetermined. This means that the available data are not sufficient to uniquely determine every source parameter [Menke, 1989]. Thus resolution is of vital importance for the reliability of the final slip model. Specifically, in this paper, we analyze model resolution, which is the ability of the inversion to uniquely determine model parameters. Note that resolution error is a different type of error than perturbation error, which is the error in the final model due to errors in the data. Both types of errors affect the inversion result, and both must be quantified in order to assess the validity of the final model.

Because the data cannot resolve tradeoffs between model parameters, not all structure seen in the final model is required by the data. The inversion result depends on choices made during the inversion process, including grid size and any smoothing or damping used (e.g., [Hartzell, 1989; Hartzell and Langer, 1993; Das and Suhadolc, 1996b; Liu and Archuleta, 2004]). In addition, the near-field station geometry can strongly affect the solution [Saraó *et al.*, 1988; Olson and Anderson, 1988]. We will discuss optimal smoothing and gridding choices that can remove artifacts in the final model that are due to the station distribution and inversion parameterization.

Kinematic inversions can employ a variety of different data types. In this paper we focus of the resolving power

of Global Position System (GPS) data. GPS data measure the displacement at a particular point on the surface. The sampling rate of GPS is typically too poor to resolve temporal details of the earthquake rupture. Thus, unlike dynamic measurements such as accelerograms, which contain information about the dynamics of faulting, GPS receivers measure the static field (the final displacement). GPS data are ideally suited to invert for the final slip distribution on the fault, rather than the temporal evolution of slip.

GPS data provide a particularly interesting case to study, as the static field decays very quickly with distance from the source. (The static field decays as the inverse square of the source-to-receiver distance, whereas dynamic waves decay as the inverse of this distance [Aki and Richards, 2002].) Thus, the resolving power of GPS data varies significantly for different areas of the fault plane. Synthetic tests of different data types confirm this (e.g., [Delouis et al., 2002]). Compared to static measurements, seismic waves can resolve slip that occurs deeper and further from stations. However, GPS data can still provide important information about the rupture that dynamic data cannot. Because of low-frequency noise, accelerometer records typically must be filtered, which eliminates the possibility of retrieving information at 0 Hz. In contrast, GPS data and other static field measurements are sensitive to the final slip. Thus static and dynamic measurements provide complementary information about the rupture process. In the companion to this paper, Custódio et al. [companion paper] describe a two-step inversion scheme that uses both GPS and strong-motion data to image the temporal evolution of slip.

We use data from the 2004 M_w 6.0 Parkfield Earthquake to investigate the resolution of GPS inversions. The details of this earthquake rupture are of particular interest as the Parkfield earthquake sequence is extremely important for testing ideas of earthquake recurrence and predictability [Bakun and Lindh, 1985; Bakun et al., 2005]. Historically, the Parkfield earthquake series provided the impetus for formulating the characteristic earthquake hypothesis that still greatly impacts ideas used in seismic hazard analysis [Jackson and Kagan, 2006]. By comparing kinematic inversions of past earthquakes at Parkfield we can determine to what extent these earthquakes are similar, and thus, to what extent ideas developed in this region can be extrapolated to future seismicity on other faults [Murray and Langbein, 2006; Custódio and Archuleta, 2007].

In addition, the quantity of data available at this site allows us to probe the earthquake source at a finer resolution than has been possible in less well-recorded earthquakes. Thus the 2004 M_w 6.0 Parkfield earthquake provides the ideal setting to investigate a central issue for earthquake source physics: the resolving power of kinematic inversions. To compare the source processes of different earthquakes or interpret individual source models, a quantitative measure of the uncertainty associated with different models is needed. The goal of this paper is to separate robust features of inversions from artifacts so that kinematic inversions can provide more reliable, easily interpretable images of the earthquake process at depth.

We study the impact that nonuniqueness has on the solution in the case of the Parkfield GPS data inversion. In Section 2, we compute the resolution matrix for an inversion of the Parkfield GPS data and demonstrate how the spatial variability in the resolution affects the model in a series of synthetic tests. Section 3 contains an alternative method using a nonuniform grid that removes artifacts caused by poor resolution. We show via synthetic tests that the nonuniform grid outperforms uniform grids of various sizes, even when smoothing constraints are imposed to destroy artifacts in the model. We then apply the nonuniform grid to an inversion of the Parkfield GPS dataset in Section 6, and quantitatively assess the effect of both resolution errors and data errors

on the solution. Finally, in Appendix A, we demonstrate that bootstrapping, a technique that has been used in past attempts of model error quantification, fails to distinguish true structure from artifacts, and in fact gives incorrect error bounds that are lower in areas where the model error is large due to poor resolution. This underscores the importance of the alternative techniques that we present here.

1.1. Formulation of the Inverse Problem

Kinematic inversions image the earthquake rupture by exploiting the linearity between slip on a fault plane and ground motions recorded in the surrounding medium. In the full spatial and temporal problem the displacement u recorded by a seismograph on the surface is represented by

$$u_i(\vec{x}, t) = \int_0^t \int_{\Sigma} \Delta u_j(\vec{\xi}, \tau) G_{ij}(\vec{x} - \vec{\xi}, t - \tau) d\vec{\xi} d\tau, \quad (1)$$

where Δu is the slip on the fault surface Σ , and G is the Green's function for a point dislocation and a given crustal structure.

The above integral, when discretized in space and time, becomes a linear matrix equation of the form $A\vec{x} = \vec{f}$, where \vec{f} is the data vector composed of displacements measured on the surface, \vec{x} is the fault slip to be recovered, and A is the matrix describing the system response (Green's function) at the locations where measurements are available. The goal of the inversion is to find the slip \vec{x} , given the known data \vec{f} and the system response matrix A . To write the inverse problem in the linear form $A\vec{x} = \vec{f}$, one must specify the discretization of \vec{x} a priori; that is, one must specify the rupture time windows in advance. For the full inversion in space and time, this can be difficult, because the temporal problem is nonlinear in rupture velocity (i.e., the correct rupture time windows at each fault location are in fact not known). One must either linearize the problem by choosing a discretization of \vec{x} (for example one could choose a constant rupture velocity, although this is not required) or by moving to a nonlinear algorithm that may not necessarily find the global minimum. There are disadvantages to either approach. The linearized solution may not be close to the true solution if extensive nonlinearities are present, or the nonlinear algorithm may be too computationally intensive to adequately search the parameter space. The nonlinearity of the dynamic problem complicates the resolution issue as well; conventional methods (e.g., [Backus and Gilbert, 1968; Wiggins, 1972]) for assessing model resolution fail under strong nonlinearity.

As the resolution of nonlinear inversions is difficult to assess, we focus on the static inverse problem in this paper. GPS data give final displacements, which eliminates the time component in Equation 1 and renders the problem linear. With static GPS data, we can only image the final slip on the fault plane. The linearity of the static inverse problem allows us to describe the resolving power of the Parkfield GPS dataset quantitatively.

We find the minimum-length least-squares solution of our inverse problem, $A\vec{x} \approx \vec{f}$, using the Moore-Penrose Generalized Inverse of A , \hat{A} . Also termed the "natural" inverse, \hat{A} gives one of the least-squares solutions (it minimizes the data residual $\|A\vec{x} - \vec{f}\|_2$). In addition, the solution is termed minimum-length because it is the least-squares solution with the smallest solution length ($\|\vec{x}\|_2$ is minimized).

We can find \hat{A} using Singular Value Decomposition (SVD) [Nash, 1990]. First we find the SVD of A :

$$A = U\Lambda V^T, \quad (2)$$

where U and V are orthogonal and Λ is diagonal with elements λ_{ii} . The singular values λ_{ii} are unique for a given

A. Once A is decomposed in this manner, the generalized inverse of A is given by:

$$\tilde{A} = V\Lambda^{-1}U^T, \text{ where } (\Lambda^{-1})_{ii} = \begin{cases} \lambda_{ii}^{-1}, & \text{if } \lambda_{ii} > 0. \\ 0, & \text{if } \lambda_{ii} = 0. \end{cases} \quad (3)$$

While there is some freedom in the choice of U and V , the generalized inverse \tilde{A} is unique. Our preferred solution satisfies $\tilde{x} = \tilde{A}f$.

Notice that the preferred solution given by SVD eliminates the parts of the solution that correspond to zero singular values of A (by setting $(\Lambda^{-1})_{ii} = 0$ when the singular value $\lambda_{ii} = 0$). This eliminates unstable parts of the solution, as zero singular values correspond to vectors in the model space (the solution space) that have no effect on the data. As such, the data can place no constraint on the weight given to zero singular vectors in the solution. The Moore-Penrose solution gives these unconstrained vectors zero weight. This is why the solution is minimum-length, as adding unstable vectors to the solution does not improve the data residual but does increase $\|\tilde{x}\|_2$. Using the Moore-Penrose inverse achieves the same result as a small amount of damping, which also removes small singular vectors from the solution.

1.2. Inversion Parameters

Our modeled fault plane and the surrounding GPS stations are shown in map view in Figure 1. We use 1-Hz

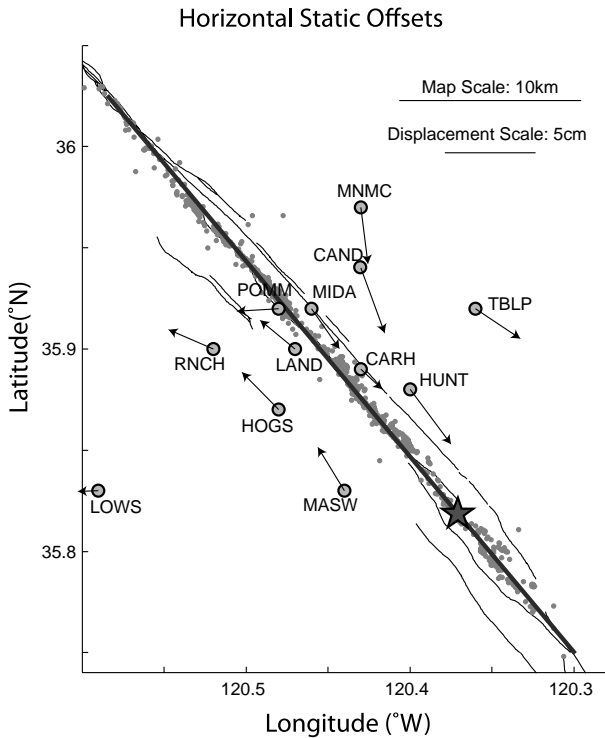


Figure 1. Horizontal static offsets recorded at GPS stations during the 2004 M_w 6.0 Parkfield earthquake. The dark grey solid line shows the horizontal extent of our modeled fault plane. The surface trace of the San Andreas Fault is shown in black [Rymer *et al.*, 2006], aftershocks are shown in light grey [Thurber *et al.*, 2006], and the epicenter of the earthquake is denoted by the star. Note that the stations are primarily located near the center of the fault trace. This results in poor spatial resolution near the edges of the fault plane and at depth, as shown in Figure 2. Figure 8b shows the vertical GPS offsets.

GPS data from 13 stations located close to the fault. The GPS offsets used were computed from 1-Hz data using the method of Larson *et al.* [2003] with added modified side-real filtering [Choi *et al.*, 2004]. Co-seismic displacements were calculated as differences in average displacements between 40-100 seconds after the earthquake and 5-100 seconds before the earthquake. The sampling rate of the GPS data allow us to use a truly co-seismic ground displacement (rather than a daily average, which would be contaminated by post-seismic deformation) with minimal afterslip. We assume the rupture is constrained to a fault plane 40 km in the strike direction \times 13.65 km down-dip, in accordance with aftershock locations (grey dots, Figure 1) [Thurber *et al.*, 2006]. Source parameters are specified on a grid with sub-faults (discretized areas of the fault plane) that are 2 km along-strike by 1.95 km down-dip, and then interpolated to a $500 \text{ m} \times 487.5 \text{ m}$ grid before convolution with Green's functions. We assume that the fault strikes 140° SE and dips 87° SW, and the ruptured area is deeper than 0.5 km (that is, there is no surface rupture). The velocity structure is approximated by a 1D bilateral model with slower material to the northeast side of the fault [Thurber *et al.*, 2003, 2006]. This velocity structure provides a good approximation to the true three-dimensional structure, which has been shown to be less important for static problems [Wald

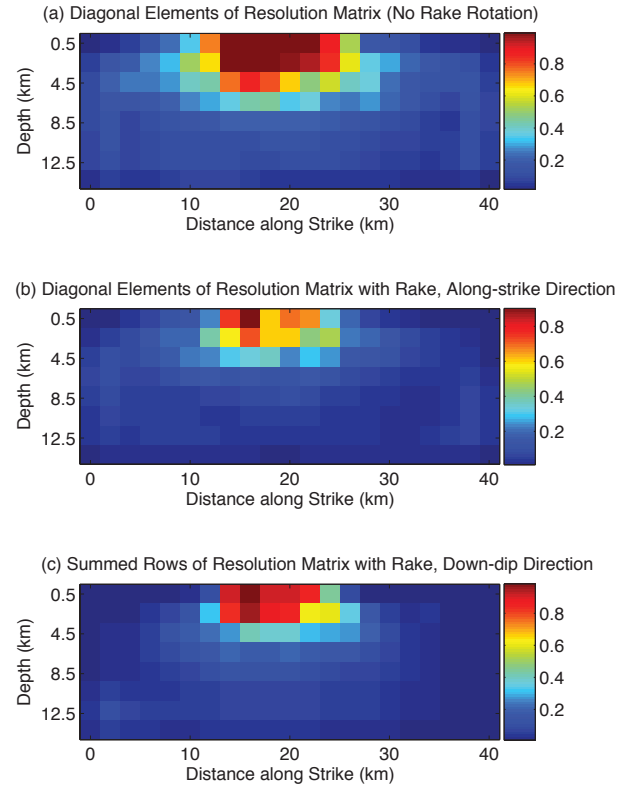


Figure 2. The diagonal elements of the Parkfield GPS resolution matrix plotted on the fault plane. A value of unity indicates perfect resolution for a particular sub-fault. a) In the case where slip is assumed to occur in the along-strike direction only, there are 168 model parameters (compared to 39 data points). The problem is severely underdetermined, and resolution is particularly poor near the edges of the fault and at depth, far from the GPS stations. Subfigures b) and c) show how the resolution suffers when rake rotation is allowed. Allowing rake rotation doubles the number of free parameters in the inversion.

and Graves, 2001]. Following *Liu et al.* [2006], Green’s functions for this layered-velocity structure are computed using the frequency-wavenumber method [*Zhu and Rivera*, 2002].

We use synthetic tests to investigate the resolution errors present in a typical formulation of the Parkfield GPS inversion. Resolution errors are due to the underdetermined nature of the inversion, and are independent of data errors (also termed perturbation errors), which we do not focus on in this work. To isolate the effects of resolution errors on the inversion result, we do not perturb the data in our synthetic tests. This allows us to use the generalized-inverse solution with no additional damping or smoothing. In our final inversion of the Parkfield GPS dataset, we do sample the actual GPS data errors via a Monte Carlo procedure, as the nonuniform grid that we introduce allows both resolution and data errors to be taken into account. This is explained further in Section 6.

2. The Parkfield GPS Resolution Matrix

Recall that our solution satisfies $\tilde{x} = \tilde{A}f = \tilde{A}(Ax) = (\tilde{A}A)x$. The matrix $R = \tilde{A}A$ is the resolution matrix, which maps the true solution to the inversion result. If the resolution matrix equals the identity matrix, then the estimated model is perfectly resolved. For inverse problems that are underdetermined, there are insufficient data to resolve all the model parameters. Hence, the estimated model parameters are actually averages of the true model parameters [*Menke*, 1989]. The i th row of the resolution matrix shows how the i th model parameter is mapped to other estimated model parameters by the inversion process. In most geophysical problems, as in ours, there is a natural spatial ordering of model parameters. (In our problem, the entries of \tilde{x} represent slip on various subfaults, which are arranged spatially on a two-dimensional grid.) In these cases, a length scale can be determined from each row of R , and this resolution length gives the distance over which slip from a subfault (gridded area of the fault) is “smeared out” by the inversion. For example, if there is a one-dimensional spatial ordering of model parameters (if we were only inverting for offsets on the fault trace), then if we were to plot the entries for a particular row of the resolution matrix, we would expect something similar to a Gaussian centered at the diagonal element. In our case, the model parameters have a two-dimensional spatial ordering, so in fact, we might expect to see the elements in a given row to resemble a two-dimensional Gaussian when plotted onto their corresponding subfaults. The resolution length for a given subfault (which may be different in the along-strike direction and down-dip direction) can be found by explicitly fitting a Gaussian distribution to the elements of the row of R corresponding to that subfault, or alternatively, estimating the spread of the Gaussian from the magnitude of the diagonal element (assuming that the Gaussian normalizes to unity).

The Parkfield GPS inverse problem, when gridded into 2-km \times 1.95-km subfaults, is severely underdetermined. There are 13 static GPS stations, with 3 components each, thus giving 39 data points to invert. However, with source parameters located every ~ 2 km, there are 168 subfaults. Solving for rake rotation in addition (allowing slip to occur in either the strike or the dip direction) doubles the number of free parameters, giving 336 model parameters. In either case, the null space of A is large, which means there are many vectors in the model space that are unconstrained by the data. Thus, we can only solve for averages of the model parameters, even if the data and Green’s function contain no errors.

As shown in Figure 1, the GPS stations in the Parkfield region are primarily located in the middle of the fault section. Thus we expect resolution to be poor near the northwest and southeast edges of the modeled fault plane. This is confirmed in Figure 2, which shows the diagonal elements of

the Parkfield GPS resolution matrix mapped onto the fault plane. The diagonal elements show how much slip in a particular subfault is correctly mapped to that subfault by the inversion; thus, a value of unity indicates perfect resolution. Given the Parkfield GPS data, slip in a small region of the center part of the fault and near the surface is well-resolved. Slip placed by the inversion elsewhere on the fault has large spatial uncertainties. When rake rotation is also solved for, as in Figures 2b and 2c, the problem becomes even more underdetermined and resolution worsens. This can be seen graphically in the figure, as the well-resolved areas (the red areas) are substantially smaller in in Figures 2b and 2c than in Figure 2a.

Even though the resolution is near zero in many of the deeper subfaults, this does not mean that the data contain no information about slip at depth. Rather, slip is unlikely to be recovered in the correct spatial location. This is because the data places no constraint on unstable slip patterns in the model space that correspond to small-scale structure at depth.

The diagonal elements of the resolution matrix are a simple measure of the resolution of each subfault, but neglect subfault correlations. The off-diagonal structure in R affects the inversion result as well. To aid the reader in visualizing this structure, we present in Figure 3 three rows of the resolution matrix, corresponding to subfaults with varying resolutions. While what is being shown are *rows* of the resolution matrix, we have plotted the elements in these rows two-dimensionally, because each element corresponds to a particular subfault on the fault plane. A single row of the resolution matrix shows the result of the following synthetic

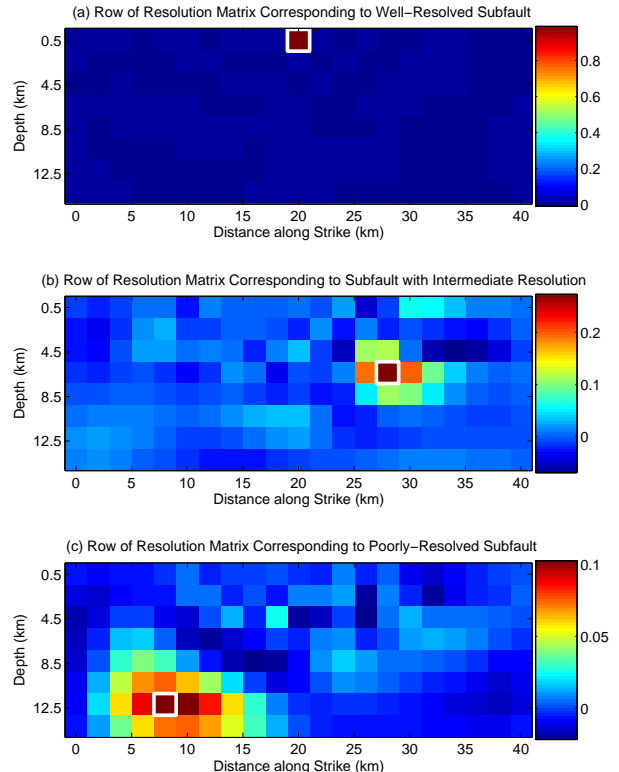


Figure 3. The rows of the resolution matrix corresponding to three sample subfaults. The row shown corresponds to the subfault that is outlined in white. For a well-resolved subfault (a), all of the slip is recovered in the correct location. However, for less-resolved subfaults (b) and (c), some of the slip is mapped to other subfaults. Note that the color scale is different in each subfigure.

test: data is generated for slip (with a magnitude of one) for a single subfault (shown in the white box in the figure); this data is then inverted using the Moore-Penrose inverse. The extent to which the final image “smears out” the input slip to neighboring subfaults is indicative of the resolving power of the data for that subfault. Figure 3a shows that in well-resolved regions of the fault plane, the inversion maps the slip to the correct subfault. Slip from deeper subfaults (see Figures 3b and 3c) is “smeared out” by the inversion over progressively larger areas of the fault as depth increases.

Due to structure in R , it is possible to obtain structure even by inverting a random field. In Figure 4, we show a synthetic test to probe the resolving power of the Parkfield GPS data set. Synthetic data are generated from an input distribution of random, uncorrelated slip (with uniform probability between zero and one for both the along-strike and the up-dip direction) at each subfault. The slip model inverted from the synthetic data shows apparent structure not present in the original slip model. This happens due to structure in the resolution matrix. The rows of R do not, in general, sum to unity. This means that some slip is lost in the mapping of the true solution x to the final model

\tilde{x} ; thus the inversion does not preserve moment without an additional constraint. Even with this constraint, however, additional structure in R can allow slip to be preferentially mapped to certain subfaults rather than others. The underdetermined nature of the inverse problem leads to spatial uncertainty in slip. The inversion “smears out” slip from its true location to surrounding subfaults. Often there is a spatial bias in this process due to structure in R . The elliptical (rather than circular) shape of the resolution kernels shown in Figure 3 shows this bias. As a result, some subfaults have more slip incorrectly mapped to them in the inversion process than other subfaults. This leads to spurious structure in the final inverted image. It is common for resolution kernels to have this spatial bias [Harris and Segall, 1987; Du et al., 1992].

While the input slip model in this particular synthetic test is physically unrealistic, this test demonstrates an important point: Because inversions are stabilized and smoothed, they may return apparently “sensible” results even if the data or Green’s functions are not at all sensible. This test contains no perturbation error, and in fact, the inversion result fits the data perfectly. In underdetermined problems, data fit is not a good measure of model resolution. With enough free parameters, one can always fit the data; but this does not mean that the model error is small.

The off-diagonal structure of R can be visualized by plotting the summed rows of R onto the fault plane, as shown in Figure 4c. It is no accident that Figure 4b and 4c are highly correlated. The resolution matrix as a whole contains all the information about how the true slip model is mapped to the inversion result. Each row provides an image of how slip in one subfault is “smeared out” by the inversion. The inversion will not recover the correct solution due to lack of data; furthermore, damping of unstable modes results in slip accumulation in certain areas of the fault plane. This can be visualized by looking at the summed rows of the resolution matrix. These summed rows show which subfaults tend to have more slip incorrectly mapped to them. The structure seen in the output slip model is a function of the Green’s function, and thus the station locations, rather than structure associated with the seismic source. The resolution

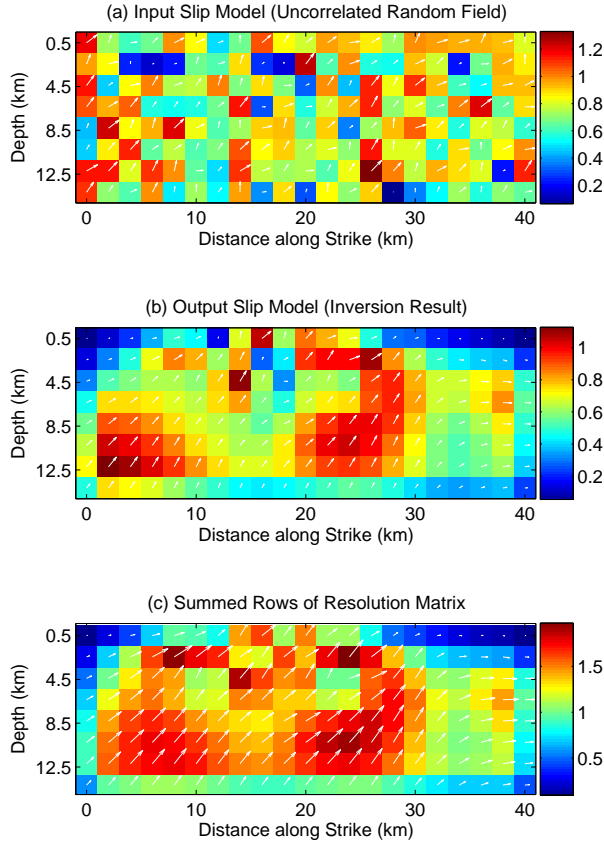


Figure 4. Due to structure in the resolution matrix, it is possible to obtain structure by inverting a random field. The slip field in this synthetic test (a) is uncorrelated noise, but the inversion result (b) shows apparent “asperities” that are due to off-diagonal structure in R rather than structure in the data. Off-diagonal structure in the resolution matrix R can play a large role in apparent heterogeneity that appears in the inversion result. The summed columns of R for an inversion with rake rotation is shown in (c) for comparison. Note that (b) correlates strongly to (c), showing that inverted structure can be due to structure in R , rather than structure in the data.

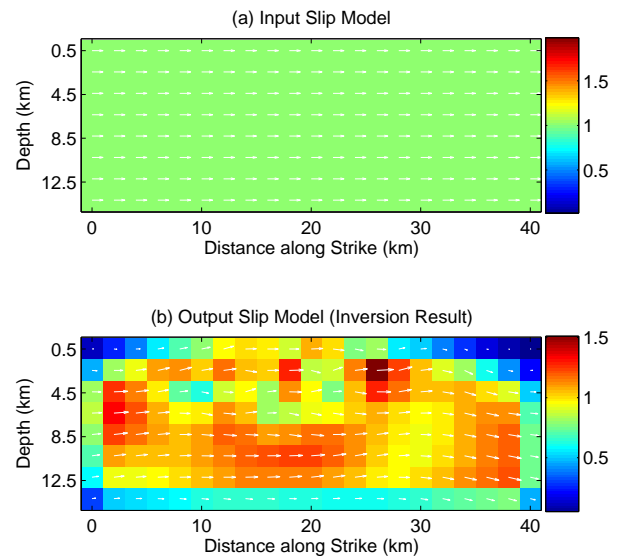


Figure 5. A synthetic test showing the resolving power of the Parkfield GPS stations for a rectangular slip patch in the input model with slip only along strike (a). The inversion result (b) shows strong spurious structure because the problem is very underdetermined.

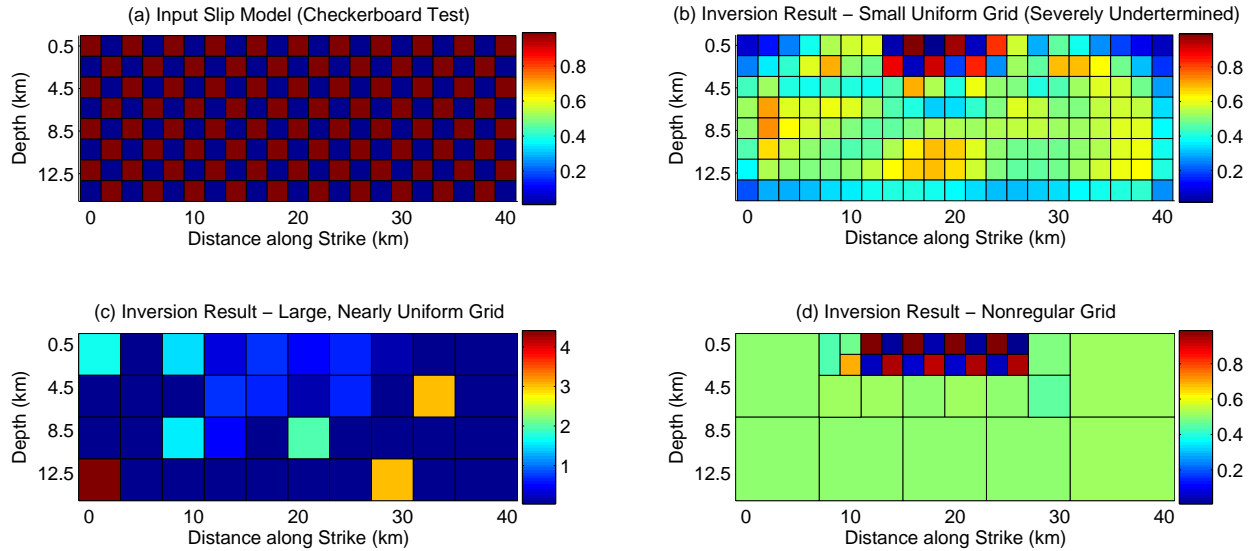


Figure 6. We generate data from a synthetic slip model (a checkerboard test) shown in (a) and invert the data onto three different grids. b) With a small, uniform grid, the inverse problem is severely underdetermined. While the small scale structure is correctly recovered in the well-resolved portion of the fault near the center of the fault trace, spurious structure is generated at depth. c) With a larger uniform gridding, the problem is not underdetermined. However, structure near the surface is lost and spurious structure is again generated at depth in part because the large subfaults near the surface are removing structure that is within the resolution length of the problem. d) With a nonuniform grid with spacing that approximates the local resolution length on the fault plane, structure is adequately recovered in well-resolved portions of the fault and spurious slip is avoided in poorly resolved areas. This inversion parameterization correctly averages small-scale structure in poorly resolved areas while recovering small-scale structure near the surface.

matrix explains this finding, which has been seen in synthetic tests for other inverse problems as well [Saraó *et al.*, 1988; Olson and Anderson, 1988].

Results from another synthetic test are shown in Figure 5. The input model in this test is more physically realistic than the test shown in Figure 4 but yields a similar result. Here the input model has uniform slip in the strike direction only, but as the inversion allows for rake rotation the problem is extremely poorly constrained. The resulting slip model obtained by the inversion shows structure – *not present in the input model* – that correlates strongly with structure in the off-diagonal elements of the resolution matrix in the strike direction.

In practice the inverted slip model will contain true structure from the data and artificial structure due to resolution problems. While consideration of the structure of the resolution matrix may help one interpret the final model, it does not provide a definitive means to determine which features are robust. Structure in well-resolved areas is certainly more trustworthy than structure elsewhere, but in GPS inversions even large-scale features at depth may not be an accurate reflection of the true slip field. In Appendix A, we demonstrate that bootstrapping fails to determine which slip features are reliable. Improving the resolution of source models requires a spatially heterogeneous approach, which we discuss in Section 3.

3. Optimal Gridding of the Fault Plane

As we have seen, formulating the inverse problem in a way that is severely underdetermined can lead to spurious structure in the final model. In the inversion of Parkfield GPS data, the resolution is highly spatially variable, with a much smaller resolution length near the top and center of the fault plane. We can improve the model resolution by making the subfaults larger in poorly resolved areas. A nonuniform grid

with cells that match the local resolution length on the fault plane simultaneously maximizes the recoverable information in well-resolved areas of the fault while avoiding spurious structure in poorly resolved areas.

We use a checkerboard synthetic slip model (Figure 6a) to generate data at the Parkfield GPS stations, and invert it with different subfault parameterizations to test how the gridding of the fault plane can affect the inversion result. As Figure 6b demonstrates, when we use a small uniform grid of ~ 2 km, information from the well-resolved portion of the fault plane is recovered, but spurious structure is generated elsewhere. Figure 6c shows results using a large approximately uniform grid such that the inverse problem is not underdetermined. Here information from the well-resolved portion of the fault plane is lost; also spurious structure is generated at depth because the inversion is sensitive to structure at a smaller scale than the grid size near the surface allows. Finally, in Figure 6d we show the inversion result using a nonuniform grid in which the scale of the subfaults was chosen to approximate the local resolution length on the fault plane. In this optimal gridding of the fault, we see that information from the well-resolved portion of fault plane is recovered, and in other areas of the fault the inverted slip represents a spatial average (in the larger subfaults, the recovered slip is near 0.5, which is the average slip of the input checkerboard model in a subfault of that size). No spurious, larger-scale structure is generated. This is an optimal result, for it resolves small-scale features in areas that they can be recovered without adding artifacts elsewhere.

This irregular grid is similar to nonuniform grids presented in other static inversions, which also seek to more closely match grid size and resolution length [Sagiya and Thatcher, 1999; Pritchard *et al.*, 2002; Simons *et al.*, 2002]. A similar multi-scale inversion approach is discussed by Uchide and Ide [2007]; it jointly performs the inversion on

three uniform grids of different sizes with constraints to ensure that they are mutually compatible. An iterative algorithm for creating a grid to match the local resolution length is described in Lohman and Simons [2005], who use an irregular grid for InSAR data down-sampling based on data resolution (rather than model resolution, as we discuss in this work).

Smoothing, of course, can help to eliminate spurious structure in poorly resolved inversions. None of the inversions shown in Figure 6 contain smoothing (except for the larger subfault size in the Figures 6c and 6d, which in a sense is smoothing with a boxcar function). While spatially uniform smoothing in the regular grid inversion shown in Figure 6b eliminates spurious structure, it also eliminates true structure in the well-resolved portion of the fault. The nonuniform grid eliminates this tradeoff by accounting for the spatially varying resolution length in the structure of the grid itself. Spatially variable smoothing can achieve a similar result. Smoothing explicitly via the grid has the advantage that the resolution length scale is readily apparent in the final image.

4. Inversion of Parkfield GPS Data

We perform a traditional inversion of the Parkfield GPS data on a uniform $2\text{km} \times 1.95\text{km}$ grid and compare this result to an inversion performed on the nonuniform grid shown in Figure 6d. We weight the data in proportion to the inverse square of the data errors given by the GPS processing, which are 1.49 mm for the east component, 2.97 mm for the north component, and 3.77 mm for the up component of displacement.

Unconstrained inversions of this data show backslip, which is unphysical. This may be a sign of Green's function errors (i.e. deficiencies in the velocity structure), as we did not see backslip in synthetic tests. To prevent backslip, we use the nonnegative least squares (NNLS) algorithm of Lawson and Hanson [1974]. The NNLS algorithm, unlike the generalized-inverse solution, is not minimum-norm and is not inherently smooth. Therefore we append Laplacian smoothing constraints to the system of equations [Hartzell and Heaton, 1983] and choose the amount of smoothing via finding the corner in the data residual vs. model roughness tradeoff curve [Harris and Segall, 1987]. In addition, we apply a moment constraint to match the observed teleseismic moment of 1×10^{18} Nm. This has the effect of preventing the slip from reaching the bottom edge of the modeled fault plane. With a moment constraint applied, edge constraints (constraining the slip to go to zero at the edges of the fault plane) were not needed. Slip is assumed to be purely right-lateral in agreement with past inversions [Custódio *et al.*, 2005; Liu *et al.*, 2006; Murray and Langbein, 2006; Johanson *et al.*, 2006], which reduces the number of free parameters and improves resolution considerably. Thus the resolution in this inversion is similar to that shown in as Figure 2a, rather than Figures 2b and 2c.

Figures 7a and 7b show the slip model from the uniform-grid inversion and the associated perturbation error from a Monte Carlo sampling of the errors in the GPS data. For the Monte Carlo error sampling, we generate 1000 perturbed sets of GPS data. Each set of perturbed data is a random realization of the data errors, which we assume are uncorrelated and Gaussian, with standard deviations as given by the GPS processing. We then invert the perturbed datasets and generate a suite of slip models. Next, we compute the standard deviation of slip for each subfault using the range of slips produced for that subfault from the suite of slip models. It should be noted that this data error is only part of the error in this model because the resolution error is not adequately captured in this uniform-grid inversion.

The inversion on the nonuniform grid, as shown in Figure 7c, captures the resolution error in the gridding of the fault

plane. The perturbation error in this model is shown in Figure 7d, which we computed in the same way as for the regular grid. Thus both types of uncertainty have been captured in this final model. The maximum slip in the two models is comparable, as is the variance reduction, which is 90% for the uniform grid inversion and 89% for the nonuniform grid inversion. This small difference in variance reduction is despite the fact that the uniform-grid inversion employs significantly more free parameters. The perturbation error in the nonuniform grid inversion is slightly higher than in the regular-grid inversion, which demonstrates the tradeoff between resolution and variance [Menke, 1989]. The fit to the data is similar in both inversions, and is shown for the nonuniform grid inversion in Figure 8.

In our view, the nonuniform grid inversion of the Parkfield GPS data is superior because it assesses both resolution and perturbation errors. In addition, it is less likely to contain artifacts because the larger subfaults at depth limit the number of free parameters. In the companion to this paper, the uniform-grid and nonuniform grid GPS inversion results are each used to constrain the final slip in two inversions of the strong-motion data from the 2004 M_w 6.0 Parkfield earthquake [Custódio *et al.*, companion paper]. The two-step inversion using the GPS slip model on the uniform grid is unable to match strong-motion data at stations to the southeast, in contrast to the successful two-step inversion using the GPS slip model on the nonuniform grid. This confirms our view that the nonuniform grid produces more reliable results with fewer artifacts.

5. Conclusion

Underdetermined inversions do not have enough data to constrain all of the model parameters. When it is not possible to obtain more data, the way to remedy this problem is to limit the number of free parameters. As we have shown, a surplus of free parameters leads to artifacts in the final slip model. In poorly resolved regions, these artifacts can look very similar to slip asperities. These artifacts can even appear to be robust features, because they depend upon the station distribution, which does not change even as other model assumptions are varied or the data values are perturbed. Thus it is important to limit the number of free parameters in inversions. While extra free parameters can improve the data fit, we have shown that the data fit is not a good measure of model error, which is what modelers want to minimize.

Static inversions such as GPS inversions have a highly nonuniform resolution in space. Thus, to prevent information in well-resolved regions from being lost, free parameters must be limited via a spatially nonuniform approach. The ideal solution would have no resolution artifacts, and would still resolve source structure at the finest scale possible. We have presented a method that achieves these goals. By employing a nonuniform grid, free parameters are limited in regions where the model parameters are most underdetermined. This solution allows for artifacts to be removed from regions with poor resolution, while preserving information recoverable from areas with superior resolution.

Two types of uncertainty must be quantified in kinematic inversions: resolution error and perturbation error. These types of error differ in character, and therefore require different types of quantification. Perturbation errors are errors in the solution resulting from errors in the data. They can be captured through Monte Carlo sampling, and can be described in terms of error bounds on the slip at each subfault (as well as subfault-to-subfault correlations, which are more difficult to describe concisely). Resolution errors, which are due to insufficient constraints and data, cannot be captured

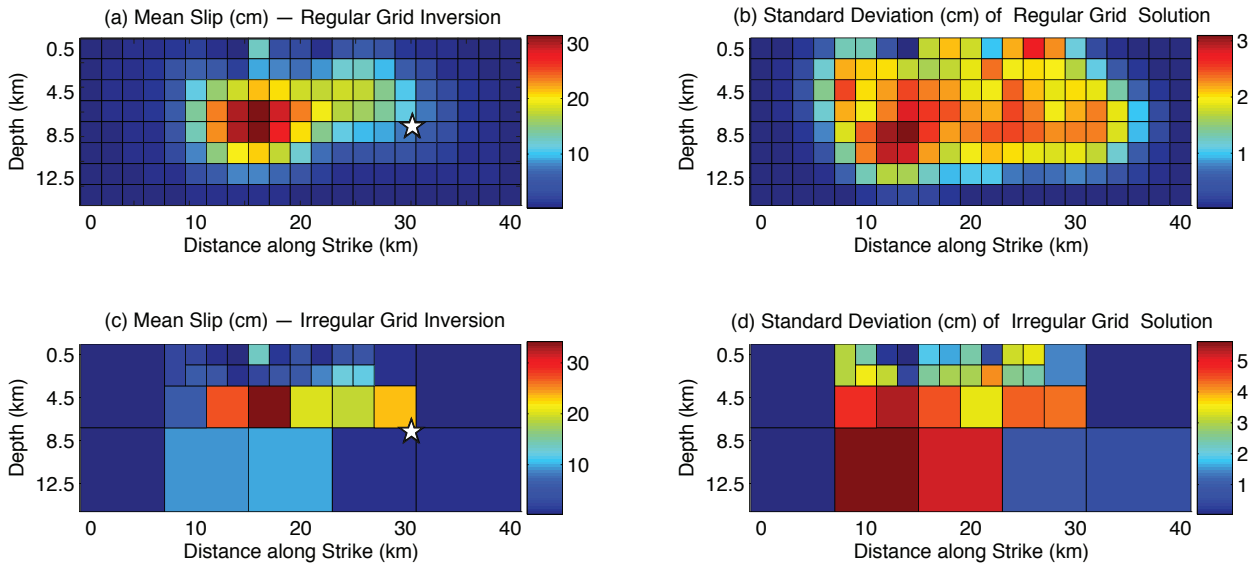


Figure 7. Inversion of Parkfield GPS data on a regular grid (a), and on an irregular grid (c), with associated perturbation errors found via Monte Carlo sampling of GPS errors (b and d). Both inversions give similar fits to the data with a variance reduction of 89-90%. The fit to GPS data for the irregular-grid inversion is shown in Figure 8.

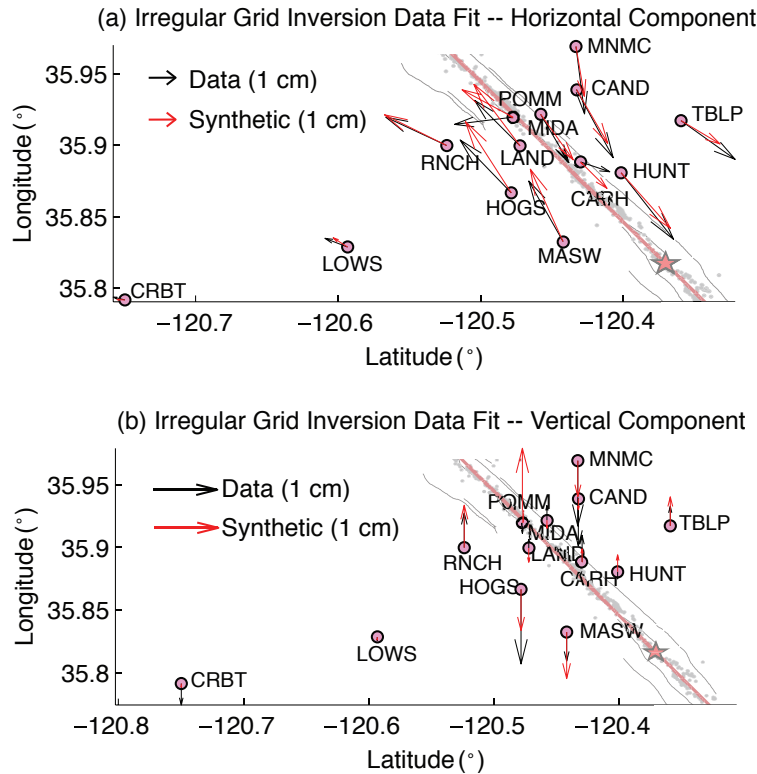


Figure 8. The fit to GPS data from the nonuniform-grid inversion. Data is shown in black, and synthetics are shown in red. The variance reduction is 89%, compared to 90% for the inversion on the uniform grid.

via error bounds because in a strict mathematical sense they are unbounded (that is, the null vectors can be of any magnitude without affecting the data). Because resolution error

is best described in terms of a length scale, it is appropriate to capture this type of error in the grid itself. We have shown that a nonuniform grid, with grid spacing equal to the

local resolution length, can capture this type of uncertainty, and thus prevent spurious structure from contaminating the solution. Subsequent Monte Carlo sampling of data errors on the nonuniform grid can allow both types of errors to be quantified, leading to a more easily interpreted slip model.

The impact of other types of perturbation error on this inversion have not been included, but merit further study. Errors in the Green’s function due to incorrect velocity structure or fault location are highly nonlinear, and can change the final slip model significantly [Das and Suhadolc, 1996a; Sekiguchi et al., 2000]. A thorough quantification of errors in kinematic inversions will allow for the determination of robust features in the models, which will allow researchers to draw firmer conclusions from this information.

Acknowledgments. We would like to thank Peng-Cheng Liu for use of his Green’s function and kinematic inversion code. M. T. P. acknowledges the support of a LEAPS fellowship as part of an NSF GK-12 grant to UCSB. In addition, this work was supported by the James S. McDonnell Foundation (grant 21002070), NSF Grant No. DMR-0606092, the David and Lucile Packard Foundation, and USGS NEHRP Grant No. 06HQGR0046.

S. Custódio acknowledges a PhD fellowship from *Fundação para a Ciência e Tecnologia* (Portuguese Foundation for Science and Technology). This work was supported by a grant from the National Science Foundation EAR-0512000.

This research was supported by the Southern California Earthquake Center. SCEC is funded by NSF Cooperative Agreement EAR-0106924 and USGS Cooperative Agreement 02HQAG0008. This is SCEC contribution number 1127 and ICS contribution number 0812.

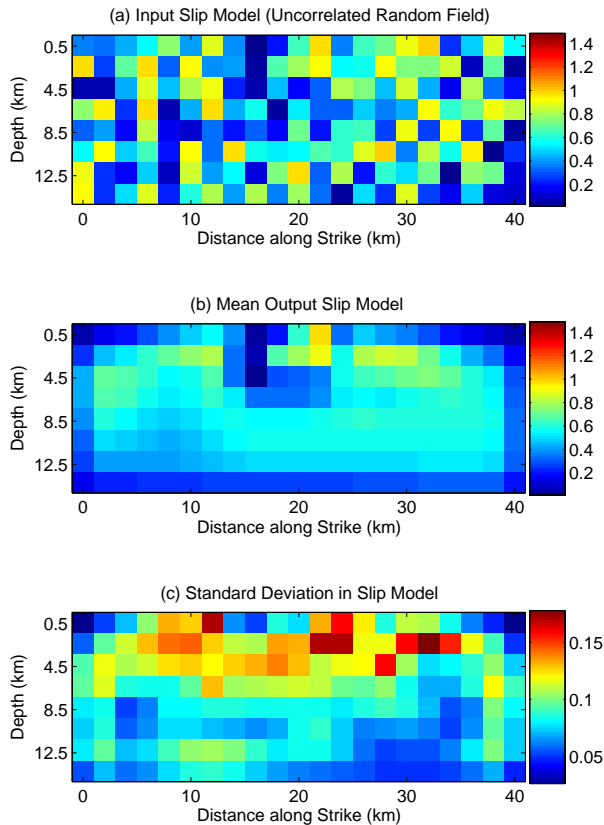


Figure 9. Bootstrapping fails to give the correct error bounds for underdetermined problems. Here we use a random synthetic slip model (a) without any rake rotation. The average inversion result from the bootstrapping analysis (b) contains spurious structure. Also, 60% of the subfaults fall outside of the 95% confidence bounds (the standard deviation is shown in (c)) given by the bootstrapped inversions.

Appendix A: The Failure of Bootstrapping

Previously we showed that in an underdetermined inversion the slip model could contain artifacts due to structure in R (and thus A) rather than structure in the data. In this appendix we investigate the ability of a commonly used error-estimation method, bootstrapping, to distinguish artifacts from true structure.

In Figure 9, we demonstrate that a random field input again yields structure, this time for an inversion without rake rotation, which is better resolved than the inversion with rake rotation shown in Figure 4. A bootstrapping analysis is a resampling plan that consists of inverting datasets that are randomly resampled from the original, complete dataset [Efron, 1982]. The length of each bootstrapped data vector equals the length of the original data vector, but as each of the 39 data points are chosen randomly, some data are included multiple times while others are omitted. Then each of the bootstrapped data vectors are inverted separately to create a family of slip models. If every slip model shows a particular feature, it is assumed that that feature is robust. We invert 1000 bootstrapped datasets and calculate statistics on the resulting slip models. Figure 4b shows the average slip values obtained, and Figure 4c shows the standard deviation of slip for each individual subfault.

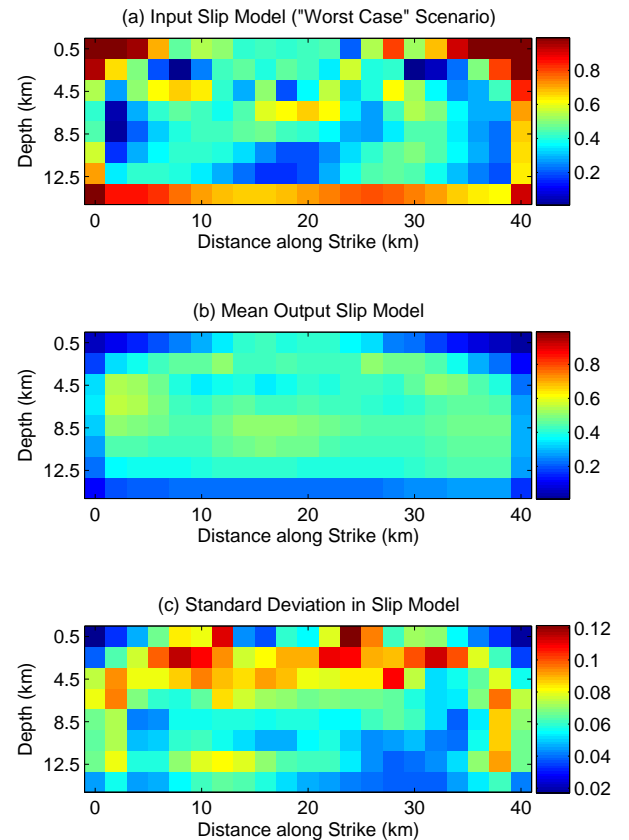


Figure 10. Here we show a “worst case” scenario for inverting an underdetermined problem. We formulated a synthetic test that exploits the structure in R and places slip in regions with little off-diagonal weight in the resolution matrix. The synthetic slip model is shown in (a). The inversion tends to put high slip in regions where the input model had little slip, as shown in (b), which is nearly the converse of our input model. Even though this example has a smoother input model than the example shown in Figure 9, bootstrapping again severely underestimates the standard deviation (c) of model parameters.

Bootstrapping substantially underestimates the uncertainty for this synthetic model, in part because slip is correlated on a smaller scale (the subfault size) than the resolution length for much of the fault. In fact, approximately 60% of the subfaults fall outside of the 95% confidence window given by the bootstrapping analysis. Bootstrapping fails because the null spaces in the individual bootstrapped inversions are highly correlated. As such, certain unstable slip distributions (corresponding to zero singular values of A) are given zero weight in every bootstrapped inversion. The uncertainty stemming from this poor resolution is never sampled.

In addition to substantially underestimating the uncertainty in the model parameters, the bootstrapping analysis also fails to capture the spatial distribution of model error. In fact, there is a negative correlation ($r = -0.1819$) between the absolute model error and the standard deviation given by the bootstrapping analysis. This means that subfaults given a *lower* variance by the bootstrapping analysis are more likely to have *larger* absolute errors.

In Figure 10, we show a “worst case” inversion scenario exploiting the structure of R . The input slip model in this synthetic test has low slip in subfaults with high off-diagonal elements in R , i.e., in subfaults where slip tends to be mapped due to the structure of R . As a result, the inverted model appears to be closer to the complement of the true slip model. As before, the bootstrapping analysis fails to fully capture the model uncertainty. Despite the input slip model being spatially smoother in this example, 57% of the subfaults fall outside of the 95% confidence window given by bootstrapping.

Bootstrapping is designed for datasets that contain redundant information. Poorly constrained inversions are not good candidates for this procedure, and bootstrapping can easily give a false picture of the model error. Bootstrapping fails to capture resolution error – the model error that is due to the infinite class of solutions present in an underdetermined inversion. This failure occurs because the null space of A is never sampled by the bootstrapping, and the null space is in fact the source of the resolution error.

The jackknife method is a similar resampling technique that omits one data point with each resampling. This also fails to quantify resolution error because the null space is never sampled.

References

- Aki, K., and P. G. Richards (2002), *Quantitative Seismology*, second ed., University Science, Sausalito, California.
- Backus, G., and F. Gilbert (1968), The resolving power of gross earth data, *Geophys. J. R. astr. Soc.*, *16*, 169–205.
- Bakun, W. H., and A. G. Lindh (1985), The Parkfield, California, earthquake prediction experiment, *Science*, *229*(4714), 619 – 624, doi:10.1126/science.229.4714.619.
- Bakun, W. H., et al. (2005), Implications for prediction and hazard assessment from the 2004 Parkfield earthquake, *Nature*, *437*(13), 969–974, doi:10.1038/nature04067.
- Choi, K., A. Bilich, K. M. Larson, and P. Axelrad (2004), Modified sidereal filtering: Implications for high-rate GPS positioning, *Geophys. Res. Lett.*, *31*(L22608), doi:10.1029/2004GL021621.
- Custódio, S., and R. J. Archuleta (2007), Parkfield earthquakes: Characteristic or complementary?, *J. Geophys. Res.*, *112*(B05310), 969–974, doi:10.1029/2006JB004617.
- Custódio, S., P. Liu, and R. J. Archuleta (2005), 2004 Mw 6.0 Parkfield, California, earthquake: Inversion of near-source ground motion using multiple datasets, *Geophys. Res. Lett.*, *32*(L23312), L23,312, doi:10.1029/2005GL024417.
- Custódio, S., M. T. Page, and R. J. Archuleta (companion paper), Constraining earthquake source inversions with GPS data 2: A two-step approach to combine seismic and geodetic datasets, *J. Geophys. Res.*
- Das, S., and P. Suhadolc (1996a), On the inverse problem for earthquake rupture: The Haskell-type source mode, *J. Geophys. Res.*, *101*(B3), 5725–5738.
- Das, S., and P. Suhadolc (1996b), On the inverse problem for earthquake rupture: The Haskell-type source model, *J. Geophys. Res.*, *101*(B3), 5725–5738.
- Delouis, B., D. Giardini, P. Lundgren, and J. Salichon (2002), Joint inversion of InSAR, GPS, teleseismic, and strong-motion data for the spatial and temporal distribution of earthquake slip: Application to the 1999 İzmit mainshock, *Bull. Seismol. Soc. Am.*, *92*(1), 278–299.
- Du, Y., A. Aydin, and P. Segall (1992), Comparison of various inversion techniques as applied to the determination of a geophysical model for the 1983 Borah Peak earthquake, *Bull. Seismol. Soc. Am.*, *82*(4), 1840–1866.
- Efron, B. (1982), *The Jackknife, the Bootstrap and Other Resampling Plans*, Society for Industrial and Applied Mathematics, Philadelphia.
- Harris, R. A., and P. Segall (1987), Detection of a locked zone at depth on the Parkfield, California, segment of the San Andreas Fault, *J. Geophys. Res.*, *92*(B8), 7945–7962.
- Hartzell, S. (1989), Comparison of seismic waveform inversion results for the rupture history of a finite fault: Application to the 1986 North Palm Springs, California, earthquake, *J. Geophys. Res.*, *94*(B6), 7515–7534.
- Hartzell, S., and C. Langer (1993), Importance of model parameterization in finite fault inversions: Application to the Mw 8.0 Peru earthquake, *J. Geophys. Res.*, *98*(B12), 22,123–22,134.
- Hartzell, S. H., and T. H. Heaton (1983), Inversion of strong ground motion and teleseismic waveform data for the fault rupture history of the 1979 Imperial Valley, California, earthquake, *Bull. Seismol. Soc. Am.*, *73*(6), 1553–1583.
- Jackson, D. D. (1972), Interpretation of inaccurate, insufficient and inconsistent data, *Geophys. J. R. Soc.*, *28*, 97–109.
- Jackson, D. D., and Y. Y. Kagan (2006), The 2004 Parkfield earthquake, the 1985 prediction, and characteristic earthquakes; lessons for the future, *Bull. Seismol. Soc. Am.*, *96*(4B), S397–S409, doi:10.1785/0120050821.
- Johanson, I. A., E. J. Fielding, F. Rolandone, and R. Burgmann (2006), Coseismic and postseismic slip of the 2004 Parkfield earthquake from space-geodetic data, *Bull. Seismol. Soc. Am.*, *96*(4B), S269–S282, doi:10.1785/0120050818.
- Larson, K. M., P. Bodin, and J. Gomberg (2003), Using 1-Hz GPS data to measure deformations caused by the Denali fault earthquake, *Science*, *300*(5624), 1421 – 1424, doi:10.1126/science.1084531.
- Lawson, C. L., and R. J. Hanson (1974), *Solving Least Squares Problems*, Prentice-Hall, Englewood Cliffs, New Jersey.
- Liu, P., and R. J. Archuleta (2004), A new nonlinear finite fault inversion with three-dimensional Green’s functions: Application to the 1989 Loma Prieta, California earthquake, *J. Geophys. Res.*, *109*(B02318), B02,318–B02,318, doi:10.1029/2003JB002625.
- Liu, P., S. Custódio, and R. J. Archuleta (2006), Kinematic inversion of the 2004 M 6.0 Parkfield earthquake including an approximation to site effects, *Bull. Seismol. Soc. Am.*, *96*(4B), S143–S158, doi:10.1785/0120050826.
- Lohman, R. B., and M. Simons (2005), Some thoughts on the use of InSAR data to constrain models of surface deformation: Noise structure and data downsampling, *Geochem. Geophys. Geosyst.*, *6*(1), doi:10.1029/2004GC000841.
- Mai, P. M. (2007), SRCMOD: A database of finite-source rupture models, www.seismio.ethz.ch/srcmod.
- Menke, W. (1989), *Geophysical Data Analysis: Discrete Inverse Theory*, Academic Press, San Diego.
- Murray, J., and J. Langbein (2006), Slip on the San Andreas Fault at Parkfield, California, over two earthquake cycles, and the implications for seismic hazard, *Bull. Seismol. Soc. Am.*, *96*(4B), S283–S303, doi:10.1785/0120050820.
- Nash, J. C. (1990), *Compact Numerical Methods for Computers: Linear Algebra and Function Minimization*, second ed., 30-48 pp., Adam Hilger, Bristol, England.
- Olson, A. H., and J. G. Anderson (1988), Implications of frequency-domain inversion of earthquake ground motions for resolving the space-time dependence of slip on an extended fault, *Geophys. J.*, *94*(3), 443–455.

- Olson, A. H., and R. J. Apsel (1982), Finite faults and inverse theory with applications to the 1979 Imperial Valley earthquake, *Bull. Seismol. Soc. Am.*, *72*(6), 1969–2001.
- Pritchard, M. E., M. Simons, P. A. Rosen, S. Hensley, and F. H. Webb (2002), Co-seismic slip from the 1995 July 30 Mw = 8.1 Antofagasta, Chile, earthquake as constrained by InSAR and GPS observations, *Geophys. J. Int.*, *150*(2), 362–376.
- Rymer, M. J., et al. (2006), Surface fault slip associated with the 2004 Parkfield, California, earthquake, *Bull. Seismol. Soc. Am.*, *96*(4B), S11–S27, doi:10.1785/0120050830.
- Sagiya, T., and W. Thatcher (1999), Coseismic slip resolution along a plate boundary megathrust: The Nankai Trough, southwest Japan, *J. Geophys. Res.*, *104*(B1), 1111–1129.
- Saraó, A., S. Das, and P. Suhadolc (1988), Effect of non-uniform station coverage on the inversion for earthquake rupture history for a Haskell-type source model, *J. Seism.*, *2*(1), 1–25.
- Sekiguchi, H., K. Irikura, and T. Iwata (2000), Fault geometry at the rupture termination of the 1995 Hyogo-ken Nanbu earthquake, *Bull. Seismol. Soc. Am.*, *90*(117-133), 117–133.
- Simons, M., Y. Fialko, and L. Rivera (2002), Coseismic deformation from the 1999 Mw 7.1 Hector Mine, California, earthquake as inferred from InSAR and GPS observations, *Bull. Seismol. Soc. Am.*, *92*(4), 1390–1402.
- Thurber, C., S. Roecker, K. Roberts, M. Gold, L. Powell, and K. Rittger (2003), Earthquake locations and three-dimensional fault zone structure along the creeping section of the San Andreas fault near Parkfield, CA: Preparing for SAFOD, *Geophys. Res. Lett.*, *30*(3), 1112, doi:10.1029/2002GL016004.
- Thurber, C., H. Zhang, F. Walfhauser, J. Hardebeck, A. Michael, and D. Eberghart-Phillips (2006), Three-dimensional compressional wavespeed model, earthquake relocations, and focal mechanisms for the Parkfield, California, region, *Bull. Seismol. Soc. Am.*, *96*(4B), 8767–8788, doi:10.1029/2000JB900435.
- Uchide, T., and S. Ide (2007), Development of multiscale slip inversion method and its application to the 2004 mid-Niigata Prefecture earthquake, *J. Geophys. Res.*, *112*(B06313), doi:10.1029/2006JB004528.
- Wald, D. J., and R. W. Graves (2001), Resolution analysis of finite fault source inversion using one- and three-dimensional Green's functions: 2. combining seismic and geodetic data, *J. Geophys. Res.*, *106*(B5), 8767–8788.
- Wiggins, R. A. (1972), The general linear inverse problem: Implication of surface waves and free oscillations for earth structure, *Geophys. J. R. astr. Soc.*, *10*(1), 251–285.
- Zhu, L., and L. A. Rivera (2002), A note on the dynamic and static displacements from a point source in multilayered media, *Geophys. J. Int.*, *148*(3), 619–627, doi:10.1046/j.1365-246X.2002.01610.x.

M. T. Page, U.S. Geological Survey, 525 South Wilson Ave., Pasadena, CA, 91106-3212 (pagem@caltech.edu)

S. Custódio, Institute for Crustal Studies, Girvetz Hall 1140, University of California at Santa Barbara Santa Barbara, CA 93106-1100 (susana@crustal.ucsb.edu)

R. J. Archuleta, Institute for Crustal Studies, Girvetz Hall 1140, University of California at Santa Barbara Santa Barbara, CA 93106-1100 (ralph@crustal.ucsb.edu)

J. M. Carlson, Department of Physics, University of California at Santa Barbara, Santa Barbara, CA 93106-9530 (carlson@physics.ucsb.edu)