Geophys. J. Int. (2008) **173,** 220–232 doi: 10.1111/j.1365-246X.2008.03733.x

Bayesian inference of kinematic earthquake rupture parameters through fitting of strong motion data

D. Monelli and P. M. Mai

Institute of Geophysics, ETH Zurich, Schafmattstr. 30, CH-8093 *Zurich, Switzerland. E-mail: monelli@tomo.ig.erdw.ethz.ch*

Accepted 2008 January 9. Received 2007 October 25

SUMMARY

Due to uncertainties in data and in forward modelling, the inherent limitations in data coverage and the non-linearity of the governing equation, earthquake source imaging is a problem with multiple solutions. The multiplicity of solutions can be conveniently expressed using a Bayesian approach, which allow to state inferences on model parameters in terms of probability density functions. The estimation of the posterior state of information, expressing the combination of the *a priori* knowledge on model parameters with the information contained in the data, is achieved in two steps. First, we explore the model space using an evolutionary algorithm to identify good data fitting regions. Secondly, using a neighbourhood algorithm and considering the entire ensemble of models found during the search stage, we compute a geometric approximation of the true posterior that is used to generate a second ensemble of models from which Bayesian inference can be performed. We apply this methodology to infer kinematic parameters of a synthetic fault rupture through fitting of strong motion data. We show how multiple rupture models are able to reproduce the observed waveforms within the same level of fit, suggesting therefore that the solution of the inversion cannot be expressed in terms of a single model but rather as a set of models which show certain statistical properties. For all model parameters we compute the posterior marginal distribution. We show how for some parameters the posterior do not follow a Gaussian distribution rendering the usual characterization in terms of mean value and standard deviation not correct. We compare the posterior marginal distributions with the 'raw' marginal distributions computed from the ensemble of models generated by the evolutionary algorithm. We show how they are systematically different proving therefore that the search algorithm we adopt cannot be directly used to estimate uncertainties. We also analyse the stability of our inferences comparing the posterior marginals computed by different independent ensembles. The solutions provided by independent explorations are similar but not identical because each ensemble searches the model space differently resulting in different reconstructed posteriors. Our study illustrates how uncertainty estimates derive from the topology of the objective function, and how accurate and reliable resolution analysis is limited by the intrinsic difficulty of mapping the 'true' structure of the objective function.

Key words: Inverse theory; Earthquake source observations.

1 INTRODUCTION

Current earthquake source imaging studies use different data sets (strong motion, teleseimic, GPS and InSAR) and inference methods (linear or linearized data inversions, direct search techniques) in order to retrieve kinematic rupture parameters. A fault rupture can be described, kinematically, as a shear dislocation propagating along a surface within an elastic medium. Using seismic data the dislocation process at each point on the fault is usually parametrized in terms of slip (or slip-velocity), rake angle (direction of slip), rupture time (time at which the slip process starts) and rise time (duration of slip). These parameters enter in the slip function which in turn determines the ground motion through the representation theorem (Aki & Richards 2002).

The mathematical parametrization of the slip function is not unique in inverse modelling studies, although the chosen functional form has important implications from the dynamic point of view. It determines in fact the traction evolution over the fault surface (Piatanesi *et al.* 2004). Two main methods are used for representing the slip function: the multitime window and the single time-window approach. In the former, the slip function is not prescribed *a priori* but is expanded into a number of basis functions (Olson & Apsel

1982; Wald & Heaton 1994; Ide *et al.* 1996; Sekiguchi *et al.* 2000; Delouis *et al.* 2002; Salichon *et al.* 2003). In the latter the slip function is forced to assume a predefined functional form, like a triangle (Hartzell & Heaton 1983), a boxcar (Emolo & Zollo 2005) or a more complex form involving, for instance, trigonometric (Hartzell *et al.* 1996) or power (Liu & Archuleta 2004) functions.

Fixing, for each location on the fault, rise time and rupture time (for a multi time-window approach, rise time and rupture time for each basis function), the relation between slip and ground motion becomes linear. A solution can then be obtained using the linear least-square method (Olson & Apsel 1982; Hartzell & Heaton 1983; Wald *et al.* 1991; Ide *et al.* 1996; Sekiguchi *et al.* 2000; Sekiguchi & Iwata 2002). This methodology requires the inversion of the forward modelling operator. Because of uncertainties in both data and theory and limited data coverage, this is often an ill-posed and ill-conditioned problem (multiple solutions may exist due to the presence of a null space in the model space and small change in the data may lead to large variations in the parameter estimates). Damping parameters are therefore additionally required in order to get a unique solution. Possible constrains are: moment minimization, smoothing of slip and filtering of singular values (Hartzell & Heaton 1983).

Relaxing the assumptions on rupture time and rise time render the inversion non-linear. Under these conditions a linearized inversion can be used to infer, together with slip, rupture time (Beroza & Spudich 1988) and also rise time values (Cotton & Campillo 1995). The main drawback of this approach is that the inversion results depend on the starting model and, requiring the computation of the generalized inverse, damping parameters are again needed.

As computational resources improved, optimization methods like simulated annealing (Hartzell *et al.* 1996; Bouchon *et al.* 2000; Delouis *et al.* 2002; Salichon *et al.* 2003; Liu & Archuleta 2004), neighbourhood (Vallee & Bouchon 2004) and genetic (Emolo & Zollo 2005) algorithms started to be adopted in earthquake source imaging studies. With such methods no assumptions on the objective function are made and good data-fitting models are found by directly searching the model space. Only the forward modelling operator is computed and no matrix inversion is needed (hence no damping parameters are required). Despite these benefits, all these randomized search techniques require a certain number of tuning parameters in order to guide the search, but no general theories are available that help to chose optimal values (Mosegaard & Sambridge 2002). Each problem often requires its own tuning parameters values. Moreover, even if some algorithms are guaranteed to converge to the global minimum (like some simulated annealing algorithms with certain cooling schedules, Sen & Stoffa 1995), this convergence is only asymptotic, that is, the true global minimum is found only after an infinite number of iterations. Practically, finite computational resources limit our ability in searching the model space so that the solution found can never be proved to be optimal.

A key issue in any parameter-estimation technique is the assessment of uncertainties which affect the inferred model parameters. In linear or linearized least-square inversions the objective function is a quadratic function with a single minimum. Uncertainties on model parameters can be obtained by computing the curvature of this function around the minimum (Menke 1989).

In non-linear inversions the structure of the objective function is actually unknown and it may presents multiple (and even degenerate valley-like) minima. Using optimization algorithms we can efficiently identify good data-fitting models but we cannot directly estimate uncertainties. For this purpose different strategies have been proposed. Emolo & Zollo (2005) used a genetic algorithm to search the model space and estimated resolution making a Gaussian approximation of the objective function around the best-fitting model. In this approach uncertainties are estimated only locally, in the neighbourhood of the best-fitting model, forcing the objective function to be Gaussian around it. Other approaches estimate uncertainties by statistically analysing the set of models visited during the search of the model space. From the set of models produced by a neighbourhood algorithm, Peyrat & Olsen (2004) selected 19 models that fit the data almost equally well, and then computed the standard deviation for each model parameter from this ensemble. Piatanesi *et al.* (2007) computed weighted mean and standard deviation for each model parameter considering the whole ensemble of models produced by a simulated annealing algorithm. The main limitation of these approaches is that they derive resolution estimates by statistically analysing the ensemble (or subensemble) of models produced by an optimization algorithm without taking into account that this ensemble does not reflect in general the actual uncertainties, that is the topology of the misfit function, but rather the operators adopted by the search algorithm. Moreover, all these techniques assume uncertainties to be Gaussian, which is generally not true for non-linear problems.

The major goal of this paper is to estimate resolution on kinematic earthquake rupture parameters taking into account the full non-linearity of the problem, without invoking any approximation on the objective function and hence allowing for possible non-Gaussian uncertainties. We consider a synthetic test so that we can control uncertainties in data and in forward modelling. In order to express the multiplicity of the solutions we adopt a Bayesian approach (Tarantola 2005). Inferences on inverted parameters are derived from the posterior probability density function. It is obtained as the conjunction of 'states of information' (expressed in terms of probability densities) reflecting our prior information on model parameters, data and their correlation (the physical law governing the forward modelling). We compute the posterior using the strategy proposed by Sambridge (1999). First, using a direct search algorithm, we explore the model space in order to discover the structure of the posterior probability density function and to identify good data fitting regions. In this study we use an evolutionary algorithm (Beyer 2001) to perform this task. Secondly, using a neighbourhood algorithm and considering the whole ensemble of models produced during the search stage, we compute a geometric approximation of the true posterior from which samples are generated and Bayesian inference performed. Hence, the solution we provide for each model parameter is stated in terms of a marginal probability density function from which uncertainty estimates can be derived.

2 THE BAYESIAN APPROACH

The general idea of a Bayesian approach to inverse theory is that a certain amount of information or knowledge about the physical system under investigation (represented by the model parameter vector **m**) and the data (**d**) is available before the inversion, and can be expressed in terms of a probability density function. Together with this '*a priori*' knowledge, another source of information is given by the correlation between model parameters and data expressed by a physical law [$\mathbf{d} = \mathbf{g}(\mathbf{m})$]. The solution of the inverse problem is then obtained by combining these two states of information. The main difficulty in computing the solution is in extracting information contained in the correlation between **d** and **m**, in particular when **m** is defined in a large dimensional space and the forward modelling operator **g** is computationally expensive. Under these conditions computing the equation $\mathbf{d} = \mathbf{g}(\mathbf{m})$ on a regular grid of points in the model space is unfeasible and one is forced to use randomized

techniques in order to evaluate the above equation in a limited number of points which should be representative of the most important regions of the model space (where the correlation between **d** and **m** is high). However, finite computation time and finite computing resources will always limit our ability in extracting this information. The consequence is that the solution of an inverse problem will be, for any realistic large scale problem, incomplete and always subject to a certain amount of variability that decreases as the exploration of the model space becomes more and more extensive.

2.1 The posterior state of information

In presenting the Bayesian approach, we follow the theoretical formulation of Tarantola (2005). We assume the M-dimensional model space and D-dimensional data space, $\mathbb{M}$ and $\mathbb{D}$, respectively, to be linear spaces. Indicating with $\rho_M(\mathbf{m})$ and $\rho_D(\mathbf{d})$ the prior probability density functions on model parameters and data respectively, while with $\theta(\mathbf{d}\,|\mathbf{m})$ the conditional probability density representing the correlation between **d** and **m**, the posterior state of information on the model space is given by:

$$\sigma_M(\mathbf{m}) = k\rho_M(\mathbf{m})L(\mathbf{m}), \tag{1}$$

where k is a normalization constant and $L(\mathbf{m})$ is the likelihood function:

$$L(\mathbf{m}) = \int_{\mathbb{D}} d\mathbf{d}\, \rho_D(\mathbf{d})\theta(\mathbf{d}|\mathbf{m}). \tag{2}$$

Assuming that our *a priori* knowledge on model parameters consists of the only information that each model parameter is strictly bounded by two values $m^\alpha_{\min}$ and $m^\alpha_{\max}$, where $\alpha \in I_M$, $I_M = \{1, \ldots, M\}$, we write:

$$\rho_M(\mathbf{m}) = \prod_{\alpha_M \in I_M} \rho_\alpha(m^\alpha), \tag{3}$$

where

$$\rho_\alpha(m^\alpha) = \begin{cases} \frac{1}{m^\alpha_{\max} - m^\alpha_{\min}} & \text{for} \quad m^\alpha_{\min} \le m^\alpha \le m^\alpha_{\max} \\ 0 & \text{otherwise} \end{cases}$$

is the prior marginal for each model parameter (that is a uniform probability density function).

In our synthetic test we add Gaussian noise to the seismograms produced by the true model. Our prior information on the data can therefore be expressed through a Gaussian probability density function. Defining $\mathbf{r} = \mathbf{d} - \mathbf{d}^{\text{obs}}$ (where **d** are the actual data and $\mathbf{d}^{\text{obs}}$ are the observed data, that is, actual data contaminated with noise), we write:

$$\rho_D(\mathbf{d}) = ((2\pi)^D \det\mathbf{C}_D)^{-1/2} \exp\left[-\frac{1}{2}\mathbf{r}^T\mathbf{C}_D^{-1}\mathbf{r}\right] \tag{4}$$

where $\det\mathbf{C}_D$ is the determinant of the data covariance matrix.

In our synthetic test we do not introduce any modelling uncertainties; the correlation between data and model parameters is therefore represented by a Dirac delta function:

$$\theta(\mathbf{d}|\mathbf{m}) = \delta[\mathbf{d} - \mathbf{g}(\mathbf{m})]. \tag{5}$$

Substituting eqs (5) and (4) into eq. (2) and the result of the integration together with eq. (3) into eq. (1), we obtain:

$$\sigma_M(\mathbf{m}) = \begin{cases} k \exp\left[-\frac{1}{2}\mathbf{r}^T\mathbf{C}_D^{-1}\mathbf{r}\right] & m^\alpha_{\min} \le m^\alpha \le m^\alpha_{\max} \\ 0 & \text{otherwise,} \end{cases} \tag{6}$$

where now $\mathbf{r} = \mathbf{g}(\mathbf{m}) - \mathbf{d}^{\text{obs}}$. Eq. (6) represents, for our synthetic test, the solution of the inverse problem. Being a multidimensional probability density function it can be characterized in terms of its properties in the model space. We can identify the maximum likelihood model (in our case corresponding to the best-fitting model). We can also compute the mean model:

$$\langle\mathbf{m}\rangle = \int_{\mathbb{M}} d\mathbf{m}\, \mathbf{m}\sigma_M(\mathbf{m}) \tag{7}$$

and the covariance matrix:

$$\mathbf{C}_M = \int_{\mathbb{M}} d\mathbf{m}\, (\mathbf{m} - \langle\mathbf{m}\rangle)(\mathbf{m} - \langle\mathbf{m}\rangle)^T \sigma_M(\mathbf{m}) \tag{8}$$

Eqs (7) and (8) give useful results only if σ_M is Gaussian. In a Bayesian approach this is possible only if $\rho(\mathbf{m})$, $\rho(\mathbf{d})$ and $\theta(\mathbf{d}\,|\mathbf{m})$ are Gaussian and the equation $\mathbf{d} = \mathbf{g}(\mathbf{m})$ is linear. In the case these conditions are not satisfied, we can still look at the information provided on a single parameter computing its corresponding marginal probability density function:

$$M(m^\alpha) = \int \ldots \int \sigma_M(\mathbf{m}) \prod_{\substack{k=1 \\ k\neq\alpha}}^{M} \mathrm{d}m^k \tag{9}$$

Eq. (9) involves computing the integral of the posterior probability density function in all the dimensions of the model space except the one corresponding to the parameter of interest.

If additional knowledge on model parameters is available, this methodology allows to introduce more complex *a priori* distributions and if the Gaussian assumption for data uncertainties is not valid also different norms can be used. We emphasize that eq. (6) has been derived assuming no uncertainties in the forward modelling. This may be valid for a synthetic test. For a real case where uncertainties and approximations are present in the modelling, and if these effects can be quantified, the correlation between model parameters and data can be represented in terms of a more complex probabilistic correlation rather then a simple Dirac delta function.

2.2 Computing the posterior: searching and appraising the ensemble

In practise, solving an inverse problem from a Bayesian viewpoint implies computing integrals in a multidimensional space (eqs 7–9). This can be done using Monte Carlo techniques which basically require generating samples according to the posterior probability density function. A variety of sampling methods can be used for this purpose (for a review, see for instance, Tarantola 2005). The applicability of each of these algorithms depends on the problem (if a small or large model space is considered, if an analytical, explicit expression of the posterior is available or not). Here, rather than directly using a sampling algorithm, we address the problem adopting a two stage procedure (Sambridge 1999): first, using an optimization algorithm, we explore the model space, possibly identifying its good data fitting regions. Secondly, using the whole ensemble of models found during the search stage, we compute a geometric approximation of the true posterior that is used for generating a new ensemble of models from which Bayesian inference can be performed. Sambridge (1999) validate this methodology using both a neighbourhood and a genetic algorithm to perform the search of the model space. Here we use an evolutionary algorithm (Beyer 2001). In principle, any other direct search method can be used. Within this approach we can exploit the efficiency of optimization algorithms in identifying good data-fitting regions of the model space and compute the forward modelling operator only during the search stage and not during the sampling process which usually requires larger number of evaluations (in this study 160 100 models have been visited during the search stage, whereas the sampling process required generating 475 000 models).

2.2.1 Searching the model space

The optimization algorithm we use to explore the model space is an evolutionary algorithm (EA) (Beyer 2001). EA is the current denomination used to identify all those population-based stochastic optimization methods inspired by the Darwinian paradigm of evolution. Among EAs there are genetic algorithms, evolutionary strategies and evolutionary programming techniques. According to these methods an optimization problem is considered similar to the process of evolution of a population of individuals that, through an evolutionary loop defined by a series of mechanisms like recombination, mutation and selection, improve their characteristics (fitness) in order to better survive in the environment where they are located. In our problem an individual is a model belonging to the model space and its 'fitness' is given by the misfit value $[\mathbf{g}\,(\mathbf{m}) - \mathbf{d}^{\rm obs}]^T \mathbf{C}_D^{-1}[\mathbf{g}(\mathbf{m}) - \mathbf{d}^{\rm obs}]$ expressing the discrepancy between predictions and observations.

Among the many EAs available, we use, following the notation of Beyer (2001), a $(\mu/\mu_D, \lambda)$—Evolutionary Strategy.[1] According to this algorithm, the exploration of the model space starts with generating an initial population, corresponding to the generation $g = 0$, of μ *parent* models $\mathcal{P}_\mu^{(0)}$:

$$\mathcal{P}_\mu^{(0)} := \left\{\mathbf{m}_1^{(0)}, \mathbf{m}_2^{(0)}, \ldots, \mathbf{m}_\mu^{(0)}\right\}. \tag{10}$$

This set of models, obtained through uniform random sampling of the model space, then evolves through the subsequent repeated application of three operators: *Dominant* μ-recombination, *Gaussian* mutation and *Truncation* selection.

The aim of the first two operators is to generate, from the current parent population, a new set of λ models, the *offsprings* population. In the *Dominant* μ-recombination, every ith component of the offspring $\tilde{\mathbf{m}}$ is obtained by uniform random selection from the μ i-components of the current parents. At each generation g we have:

$$\tilde{\mathbf{m}}_j^{(g)} := \sum_{i=1}^{M} \left[\mathbf{e}_i^T \mathbf{m}_{k_i}^{(g)}\right]\mathbf{e}_i, \quad j = 1, \ldots, \lambda, \tag{11}$$

where k_i is an integer uniform random number between $\{1, ..., \mu\}$ and the symbol $\mathbf{e}_i$ stands for the unit vector in the ith direction of the model space. The scalar product gives the ith component of the uniformly random selected parent $\mathbf{m}_{k_i}$.

In the *Gaussian* mutation an additional perturbation is added using a normal distribution $\mathcal{N}$ with zero expectation value:

$$\hat{\mathbf{m}}_j^{(g)} := \tilde{\mathbf{m}}_j^{(g)} + [\sigma_1 \mathcal{N}(0, 1), \ldots, \sigma_M \mathcal{N}(0, 1)], \tag{12}$$

where $j = 1, \ldots, \lambda$ and $\mathcal{N}(0, 1)$ represents a normal random number with zero expectation value and unit standard deviation. The final offspring $\hat{\mathbf{m}}$ is therefore obtained around the parental recombination result $\tilde{\mathbf{m}}$ through the addition of a Gaussian random vector. The mutation can be isotropic, that is for all the parameters the standard deviation is the same, or anisotropic (in case model parameters have different physical meanings therefore requiring different standard deviations).

The aim of the selection operator is to choose among the final set of offsprings a new ensemble of models to be used as a parent population for the next generation. In the *Truncation* selection this is done in a deterministic way. The new parent population is formed by selecting the μ best-fitting models among the only λ offsprings. This requires $\lambda \geq \mu$. This series of steps is repeated until a stop criterion is reached (e.g. a stationary level of fit). Evidently, the last step of the algorithm is the most expensive in terms of computation time because it requires the calculation of the misfit function for each offspring. Great improvement can be achieved parallelizing the computation, that is, distributing the calculation of the misfit over several processors and, once collected the results, performing the selection.

The EA requires a certain number of parameters to be tuned. The number of parents and offsprings, μ and λ, respectively, and the standard deviations for the mutation operator. Unfortunately, no general theory is available that helps to choose optimal values for these parameters, essentially because the performance of the algorithm is strictly dependent on the unknown 'fitness landscape'. However, some guidelines are available. The ratio μ/λ determines the trade-off between exploration/exploitation. Clearly the condition $\mu = \lambda$ basically means pure exploration (no selection among offsprings) and as the ratio μ/λ decreases the exploitation tendency increases. For the mutation operator, the algorithm allows to choose a different standard deviation for each model parameter. To limit the number of tuning parameters, we choose to use different standard deviations only for those parameters that represents different physical quantities. The 'strength' of the mutations (the magnitude of the standard deviations) is another important factor. They should not be too small, to ensure population diversity, and not too large, to allow convergence towards good data fitting regions of the model space. However, following these guidelines is not sufficient to properly set the algorithm's parameters, and additional trial and error work is usually required.

2.2.2 Appraising the ensemble

The models produced by the evolutionary algorithm cannot be used directly for Bayesian inference, because they are not generated according to the posterior probability density function. However all these models, together with their corresponding values of $\sigma_M(\mathbf{m})$ (easily computed knowing the value of the misfit, eq. 6) constitute an important source of information about the structure of the actual posterior; this can be used to compute a geometric approximation of it, from which samples can be drawn. This is the basic idea behind the appraising methodology developed by Sambridge (1999). The ensemble of models found during the search stage constitute an irregular distribution of points in the model space. Around each of these points a nearest-neighbour region can be calculated using a geometrical construct known as Voronoi cell. For any distribution of irregular points in any number of dimensions, Voronoi cells are unique, space-filling, convex polyedra, whose size and shape are automatically adapted to the distribution of the point set. This implies that the size (volume) of each cell is inversely proportional to the density of the points. A geometric approximation of the true posterior is then calculated setting the known value of the posterior of each model to be constant inside its Voronoi cell.

A new ensemble of models generated according to the approximated posterior is produced using a Gibbs sampler. A Gibbs sampler generate samples performing a random walk in the model space. From a given starting point, the algorithm sequentially performs a step along each parameter axis generating a random deviate from the conditional probability density function of the approximated posterior along the considered direction. An iteration is completed when all dimensions has been cycled through once, and a new model

[1] In this notation μ denotes the number of parents and λ the number of offsprings. The comma symbol ',' indicates that the μ parents for the next generation are selected among the only λ offsprings of the current generation. Note that this implies $\lambda \geq \mu$. The notation μ/μ_D denotes that all the μ parents are used for *Dominant* (D) recombination.

has been generated. After many iterations, the random walk will generate models with a distribution that tends towards the target distribution, that is the approximated posterior.

The practical applicability of this methodology is limited by the memory and computation time needed to perform this appraising step. The storage S required by the algorithm is controlled by the number of models constituting the ensemble N_e and the number of dimensions of the model space M:

$$S \propto N_e M. \tag{13}$$

Computation time T is additionally dependent on the resampled ensemble N_r, that is by the set of models sampled from the approximated posterior:

$$T \propto N_r N_e M. \tag{14}$$

As in the in the search stage, computational time can be greatly decreased distributing the resampling process on several processors.

For the synthetic test we present, the dimension of the model space is $M = 38$, the number of models visited during the search is $N_e = 160\,100$. The number of models constituting the resampled ensemble is $N_r = 475\,000$. The resulting computation time (on a 20 CPUs Linux cluster) is $T \sim 1$ day.

3 THE SYNTHETIC TEST

To control uncertainties in data and in forward modelling we consider a synthetic test. The kinematic rupture model we use as 'true' model is shown in Fig. 1. We represent the fault as a 32 km long and 12 km deep, vertically dipping, plane surface. The fault's upper edge is at 2.75 km depth. The rupture process is characterized by a heterogeneous distribution of peak slip-velocity, whereas rake angle and rise time are constant (0°, 0.8 s, respectively). Peak slip-velocity values are defined on a 4×4 km grid (nodes represented by black dots). The time evolution of the rupture process is prescribed in terms of a circular front that propagates from the hypocentre (12.5 km deep) with constant rupture velocity ($V_r = 2.7$ km s^{-1}).

The observational network we use for the inversion is depicted in Fig. 2. The fault strikes at 150°, station locations and velocity model are adapted from the 2000 Western Tottori earthquake (Semmane *et al.* 2005). All stations are located within 60 km from the epicentre.

We compute ground velocities using the frequency-domain representation theorem (Spudich & Archuleta 1987):

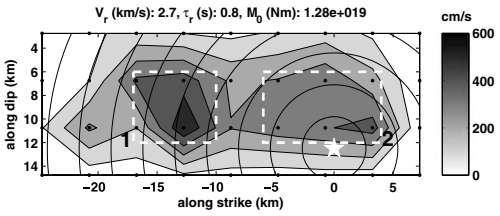


Figure 1. The 'true' kinematic rupture model. Only the maximum slip-rate amplitude is heterogeneous. Rake angle is everywhere zero (pure left-lateral strike slip event) and rise time is constant, $\tau_r = 0.8$ s. Rupture times are given by the arrival times of a circular rupture front expanding from the hypocentre (white star) with constant rupture velocity $V_r = 2.7$ km s^{-1}. The corresponding seismic moment is $M_0 = 1.28e19$ Nm. Black dots represent locations where peak slip-velocity values are defined. Dashed white rectangles delimit the two main large-slip regions characterizing the slip distribution. In the article we will refer to them as asperity 1 (the one on the left) and asperity 2 (the one on the right).

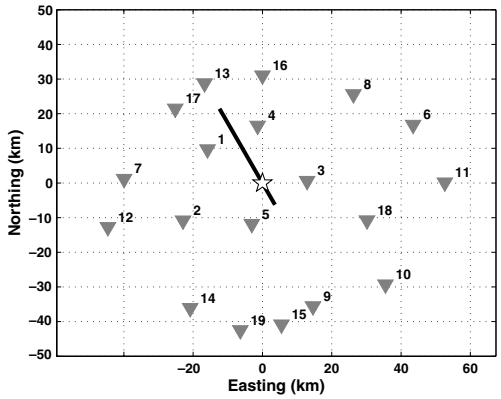


Figure 2. The observational network. 19 stations (grey triangles) are located near the fault strike (black solid line) within 60 km from the epicentre (white star). Station locations and velocity model are adapted from the 2000 Western Tottori earthquake (Semmane *et al.* 2005).

$$\dot{u}_m(\mathbf{y}, \omega) = \iint_{\Sigma} \dot{\mathbf{s}}(\mathbf{x}, \omega) \cdot \mathbf{T^m}(\mathbf{x}, \omega; \mathbf{y}, \mathbf{0})\, \mathbf{d\Sigma}, \tag{15}$$

where $\dot{u}_m$ is the m component of ground velocity at the receiver location $\mathbf{y}$, $\dot{\mathbf{s}}$ is the slip-velocity function, $\mathbf{T^m}$ is the traction exerted across the fault surface Σ at point $\mathbf{x}$ generated by an impulsive force applied in the mth direction at the receiver and $\omega = 2\pi f$ is the angular frequency.

Tractions $\mathbf{T^m}$ are computed, up to a frequency of 2 Hz, using a Discrete Wavenumber / Finite Element method (Compsyn package, (Spudich & Xu 2002)), for a 1-D flat layered Earth model without attenuation. A trapezoidal-rule quadrature of the product $\dot{\mathbf{s}} \cdot \mathbf{T^m}$ is performed separately for each frequency, with the quadrature points being the sample points where $\mathbf{T^m}$ have been computed. Peak slip-velocity values at integration points are derived through bilinear interpolation of values of surrounding grid nodes. The slip-velocity function is assumed to be an isosceles triangle. With this parametrization, the maximum slip-rate corresponds to the hight of the isosceles triangle and the rise time to the base length. Each computed synthetic seismogram contains 4096 data points, from 0 to 40.95 s, with a time sampling of 0.01 s.

We do not introduce any uncertainties in the forward modelling but we perturb synthetic seismograms produced by the true model with Gaussian noise so that a data covariance matrix C_D can be computed. We assume noise statistics to be the same for each waveform and without correlation between different stations and between different components of the same station. Thus the covariance matrix for the whole set of data reduces to a block diagonal matrix where each block matrix represent the covariance matrix for each single waveform. To compute the covariance matrix we follow the approach of Gouveia & Scales (1998). We treat each synthetic seismogram produced by our true model as a 'mean' seismogram $\mathbf{s}^{\text{mean}}$. We then compute several realizations of noisy seismograms $\mathbf{s}^{\text{noise}}$ simply adding to the mean seismogram a Gaussian time-series $\mathbf{s}^{\text{gauss}}$ with zero mean and fixed standard deviation ($\mathbf{s}^{\text{noise}} = \mathbf{s}^{\text{mean}} + \mathbf{s}^{\text{gauss}}$). If N is the number of realization done, an estimate of the covariance matrix for each waveform is given by:

$$\hat{C}_D = \frac{1}{N} \sum_{i=1}^{N} \left(\mathbf{s}_i^{\text{noise}} - \mathbf{s}^{\text{mean}}\right)\left(\mathbf{s}_i^{\text{noise}} - \mathbf{s}^{\text{mean}}\right)^T = \mathbf{s}_i^{\text{gauss}} (\mathbf{s}_i^{\text{gauss}})^T \tag{16}$$

from which we see that $\hat{C}_D$ is the same for all inverted seismograms depending on the Gaussian time-series only. For our synthetic test

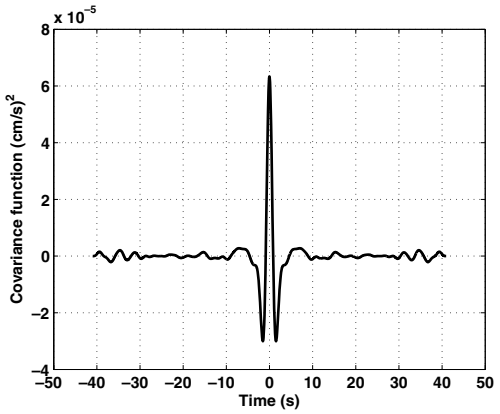


Figure 3. The noise covariance function. The correlation is almost zero after 10 s. This is consistent with the fact that the covariance matrix has been estimated considering Gaussian time-series filtered in the frequency range [0.1 0.5] Hz, containing therefore periods between 2 and 10 s.

we generate Gaussian time-series with zero mean and standard deviation equal to 1 cm s^{-1} which are then filtered in the frequency range 0.1–0.5 Hz. The resulting standard deviation of the noise is very small, about 0.01 cm s^{-1}. The corresponding signal-to-noise ratio (SNR) (calculated as the ratio between the maximum value of the signal and the maximum value of noise) varies depending on the waveforms. The minimum SNR observed is about 7. We performed $N = 500$ noise realizations and the resulting $\hat{C_D}$ was smoothed by replacing each element with the average of its diagonal. In Fig. 3, we show the resulting noise covariance function (i.e. the cross-diagonal terms). Note how the filtering has introduced a certain level of correlation in the noise that almost disappears after 10 s, consistent with the fact that noise below 0.1 Hz has been filtered out.

We invert all components for all stations in order to retrieve peak slip-velocity values at gridpoints, rupture velocity and rise time. Rake angle and hypocentre location are fixed to their true values. We define peak slip-velocity values on the same grid used for calculating the true seismograms. As we mentioned in Section 2.1, for each model parameter the prior marginal is uniform, inside a predefined range of values. Model parameter ranges are [0 600] cm s^{-1} for peak slip-velocity, [2 3] km s^{-1} for rupture velocity and [0.5 1.5] s for rise time. The total number of model parameter we invert for is therefore 38.

The fitness function used during the search is calculated as the reduced χ^2_ν value of the data fit, where ν is the number of degrees of freedom (number of data minus number of parameters):

$$\chi^2_\nu = \frac{1}{\nu}(\mathbf{g}(\mathbf{m}) - \mathbf{d}^{\rm obs})^T \mathbf{C}_D^{-1}(\mathbf{g}(\mathbf{m}) - \mathbf{d}^{\rm obs}). \quad (17)$$

Eq. (17) contains the inverse of the covariance matrix $\mathbf{C}_D^{-1}$. In our case each waveform contains 4096 data points so that the covariance matrix for each waveforms is a 4096 × 4096 matrix. As a first order approximation we consider, in the calculation of the misfit, only the main diagonal (i.e. the variance of the noise).

From equation (17) we also see that the misfit value depends, through $\mathbf{d}^{\rm obs}$, on the particular noise realization added to the 'mean' seismograms. In this study we present results obtained using a single data realization. Clearly a different data realization would produce, for the same model, a different value of fit. However is out of the scope of this paper investigating the effect of different noise realizations in the computed posterior.

4 INVERSION RESULTS

4.1 The maximum likelihood model

As explained in Section 2.2 the first step in our inversion consists of searching the parameter space. After several trial inversions the evolutionary algorithm parameters have been fixed to the following values: $\mu = 100, \lambda = 4000$. The standard deviations for the mutation operator, for peak slip-velocity, rupture velocity and rise time are, respectively: $\sigma_{A_{\max}} = 10$ cm s^{-1}, $\sigma_{V_r} = 0.3$ km s^{-1} and $\sigma_{\tau_r} = 0.3$ s. We do not expect these values to be optimal (in rendering the search the most efficient) and as already stated in Section 2.2.1, even if some guidelines are available trial and error work is usually required to set these parameters.

In Fig. 4, we show the best-fitness function value for each generation versus the generation number. After about the 20th generation the misfit reaches an approximately stationary level that lasts until the search is stopped. The total number of models visited is 160 100. On a 20-CPU Linux cluster the search required about 1 day of computation time.

The first result of the search we may look at is the maximum likelihood model (corresponding to the best-fitting model in our problem, the one with the lowest χ^2_ν value). We show it in Fig. 5. Comparing with the true model (Fig. 1) we can see that the general characteristics of the rupture process are retrieved. The locations of the two slip patches are correctly imaged and also rupture velocity and rise time values are close to the true ones. These similarities produce also a corresponding seismic moment near the true value. However, we can also see that even if the large scale features are correctly imaged, the details are not, for example, at the bottom of the fault the peak slip-velocity is significantly overestimated. Despite

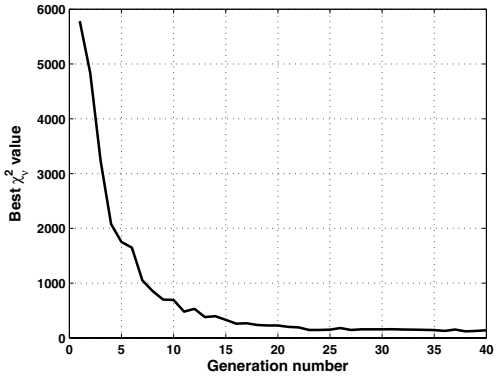


Figure 4. χ^2_ν reduction during the search. The best-fitness function value for each generation versus generation number is shown. After about the 20th generation the misfit reaches an approximately stationary level.

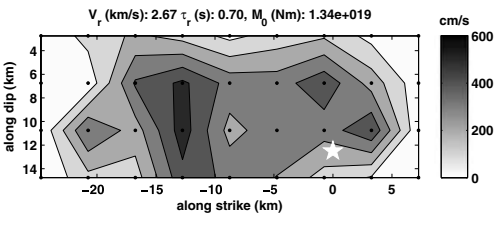


Figure 5. The maximum likelihood model (corresponding to the lowest χ^2_ν value). The general shape of the slip distribution is correctly retrieved and rupture velocity, rise time and seismic moment values are close to the true ones. However the maximum slip-rate is overestimated at the bottom of the fault.

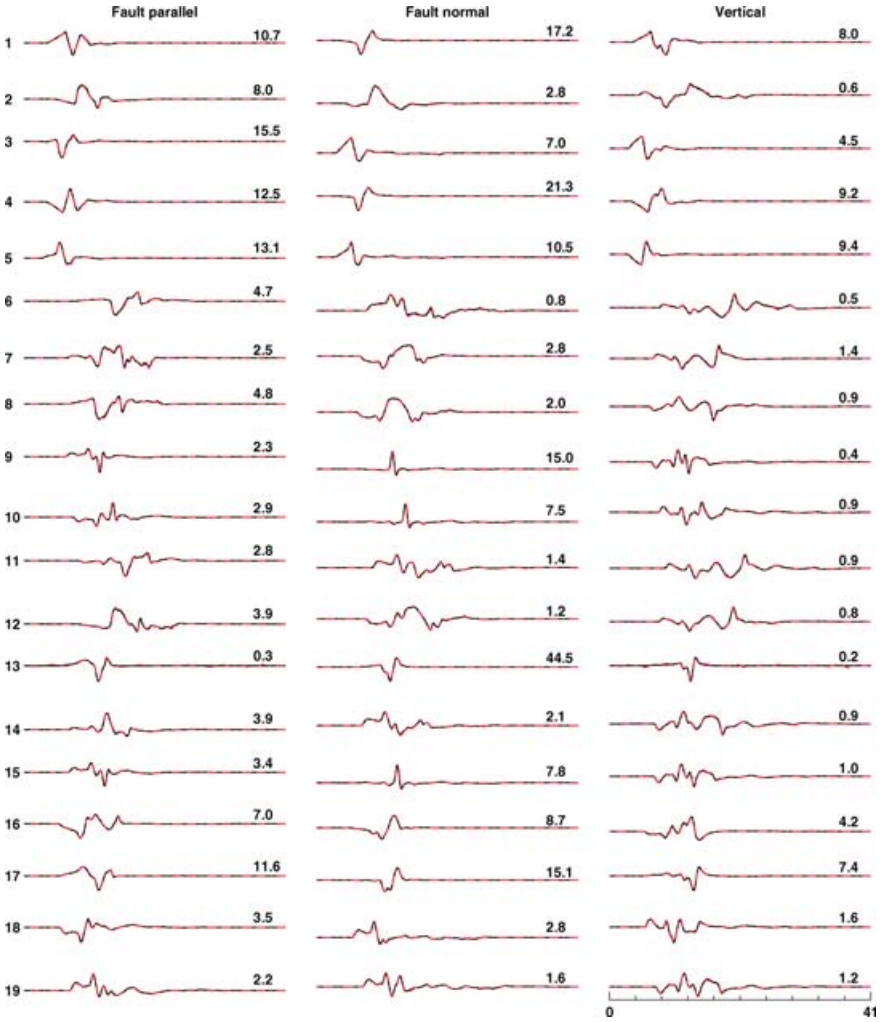


Figure 6. Data fit between data-prediction (solid black) and observed data (dash–dotted red) for the maximum likelihood model. For each waveform the maximum amplitude (cm s^{-1}) of the observed ground velocity is shown.

these differences the corresponding level of fit is visually very good (Fig. 6). Numerically it corresponds to $\chi^2_\nu \simeq 118$. This high value (for uncorrelated noise $\chi^2_\nu > 1$ means that predicted data are not able to reproduce, in average, the observed data whitin the assumed standard deviation) is basically due to the very small uncertainties we consider in measuring the data-fit (we recall that the standard deviation of noise is $\sim$0.01 cm s^{-1}).

4.2 Uncertainties estimates

The need for estimating uncertainties comes from the fact that the maximum-likelihood model is not the only model that produces a good level of fit to the data. In Fig. 7, we show peak-slip velocity distribtuions for 40 models, found during the search, with a $\chi^2_\nu \leq 1000$. The visual analysis of the peak slip-velocity distributions shows that all these models share some large scale features also present in the best-fitting solution: low slip-rate at the top, right and left borders of the fault and near the hypocentre; a major slip patch located between -20 and -10 km along strike; and a second slip patch above the hypocentre. Despite this common characteristics, the details of each peak slip-velocity distribution varies from model to model. In Fig. 8, we show the level of fit produced by all these models. They all generate waveforms very similar to the observed ones. From this example it can be seen that, within a certain level of fit, the inverted data cannot constrain a single model but rather a set of models which are different one from another but share some common properties. Quantifying and expressing these common properties is the ultimate goal of the inversion.

Following the methodology described in Section 2.2 we compute for each model parameter its corresponding 1-D posterior marginal probability density function. In Fig. 9(a), we show the posterior and the prior marginals for the peak slip-velocity, together with the true value, for each grid node on the fault surface. We also plot the raw marginals computed from the ensemble of models generated by the evolutionary algorithm.

Comparing raw and posterior marginals we can see that they are in general different, that is, they do not follow the same distribution. The raw marginals often present a much better defined peak then the posterior suggesting therefore better resolution then the actual one [see, for instance, posteriors at $(-20.75, 2.5)$, $(-16.75, 2.5)$]. This shows that the statistical properties of the ensemble of models produced by the evolutionary algorithm do not represents the actual uncertainties affecting model parameters.

We also notice that in general posteriors do not show a Gaussian shape (especially for those parameters for which the true value is close to 0 or to the maximum boundary value, like the posteriors

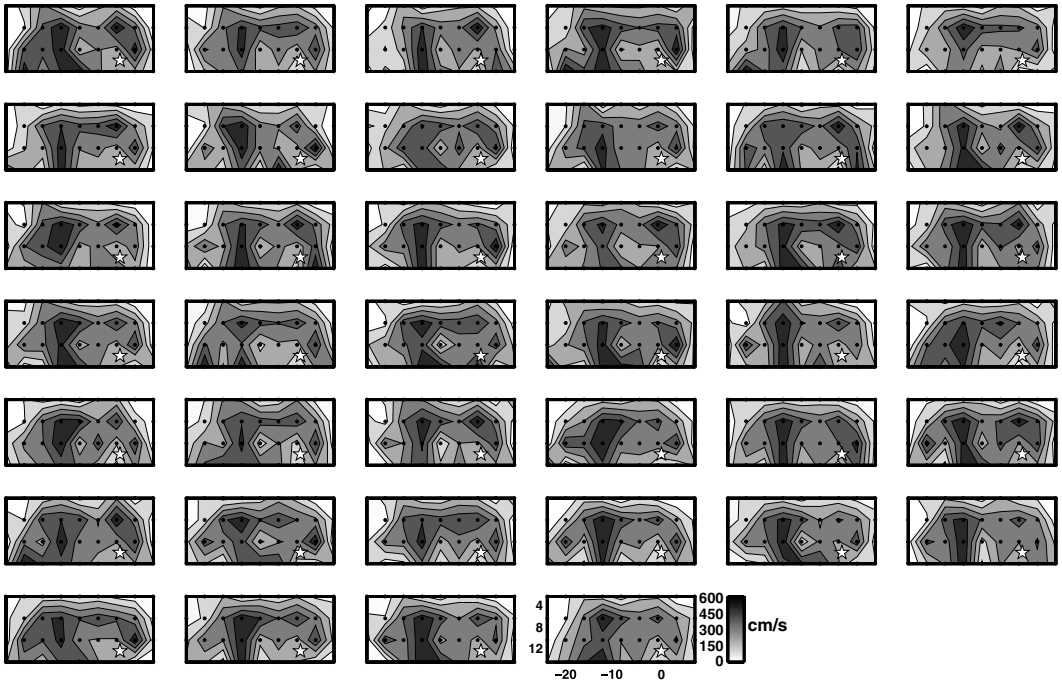


Figure 7. Peak slip-velocity distributions (cm s^{-1}) for a set of models, found during the search, with $\chi^2_\nu \leq 1e3$. Simple visual analysis shows that all models share some large scale features. Low slip-rates at top, left- and right-hand borders of the fault and near the hypocentre. Two main slip patches, the one on the left characterized by higher values (between 400 and 600 cm s^{-1}).

at (−24.75, 2.5) and (−12.75, 10.5)). For these cases, the standard characterization in terms of mean value and standard deviation is not really meaningful: the mean value would not correspond to the maximum likelihood value and the standard deviation cannot be interpreted as a symmetric error bar on the mean value. For these parameters we therefore cannot use the Gaussian uncertainty hypothesis.

Without the support of the Gaussian assumption resolution on model parameters can be better understood by looking at the difference between priors and posteriors. At some fault locations a single well-defined peak in the posterior can be identified (at the right and left sides of the fault surface, for instance), at some others locations there is little difference with respect to the uniform prior [see posteriors at (−20.75, 2.5), (−16.75, 2.5), (−20.75, 10.5), (16.75, 10.5) for instance], suggesting therefore poor resolution. We can also see that at the lower edge of the fault (nodes at (−12.75, 14.75), (−8.75, 14.75)) and at node (−0.75, 10.75) the true value is located on the tail of the computed marginal posterior. For these parameters the posterior seems to miss the true value. A tentative explanation for these results can be that for these parameters the search algorithm did not reach the true values but got locked into a solution prematurely. Assuming these parameters to be very poorly resolved (something that we can expect for nodes located in the bottom part of the fault) the 'fitness' landscape for those parameters will be something similar to a valley. If then the search is stopped before exploring all the valley and therefore without reaching the true values, the reconstructed posterior will be incomplete and will contain that valley only partially. Therefore, even if the true posterior is constant for these parameters, the approximated posterior will be peaked only around the best-fitting models found during the search. This is important to bear in mind. The reconstructed posterior reflects only what the search algorithm illuminated. This implies that the reconstructed posterior may not completely reflect the true, data-determined posterior. A similar behaviour can also be found in the results provided by Sambridge (1999). In the synthetic receiver function problem he considers, the marginal posterior for the thickness of the bottom layer completely misses the true value (fig. 7, p. 738).

We present also the 1-D marginals for rise time and rupture velocity (9b and c). Again, a well-defined single peak of the raw marginals contrasts with a smoother and broader *a posteriori* distribution. For these two parameters the posteriors shows approximately a Gaussian shape so that they can be characterized in terms of mean value and standard deviation. The mean rise time underestimates the true value of about 0.1 s. The true rupture velocity is inside one standard deviation (about 0.1 km s^{-1}) from the estimated mean rupture value.

Besides single model parameters, we can also analyse resolution on combination of model parameters. As we have noticed before often much more resolution is achieved on the large scale features of the slip distribution rather then on the local details. In Figs 9(d) and (e), we present 1-D marginals for the average peak slip-velocity on the two main asperity regions characterizing the true model (asperities extensions are: 7 by 6 km for asperity 1 and 10 by 6 km for asperity 2). Here we see that our *a priori* marginal is not uniform anymore because it represents information on a combination of the original parameters. In both cases the true values are correctly retrieved with a good resolution (standard deviations of the order of 50 cm s^{-1}, corresponding to relative error of 14 per cent). Good resolution is achieved also for the seismic moment (standard deviation equal to 2.44e18 Nm, relative error 18 per cent) (Fig. 9f).

5 RECONSTRUCTING THE POSTERIOR

Our resolution analysis derives from the reconstructed posterior computed from the ensemble of models visited during the search stage. This implies that our uncertainty estimates depend on the way the search developed in the model space. To further elucidate this point we perform three independent searches, with the same settings for the evolutionary algorithm parameters, but with different seeds

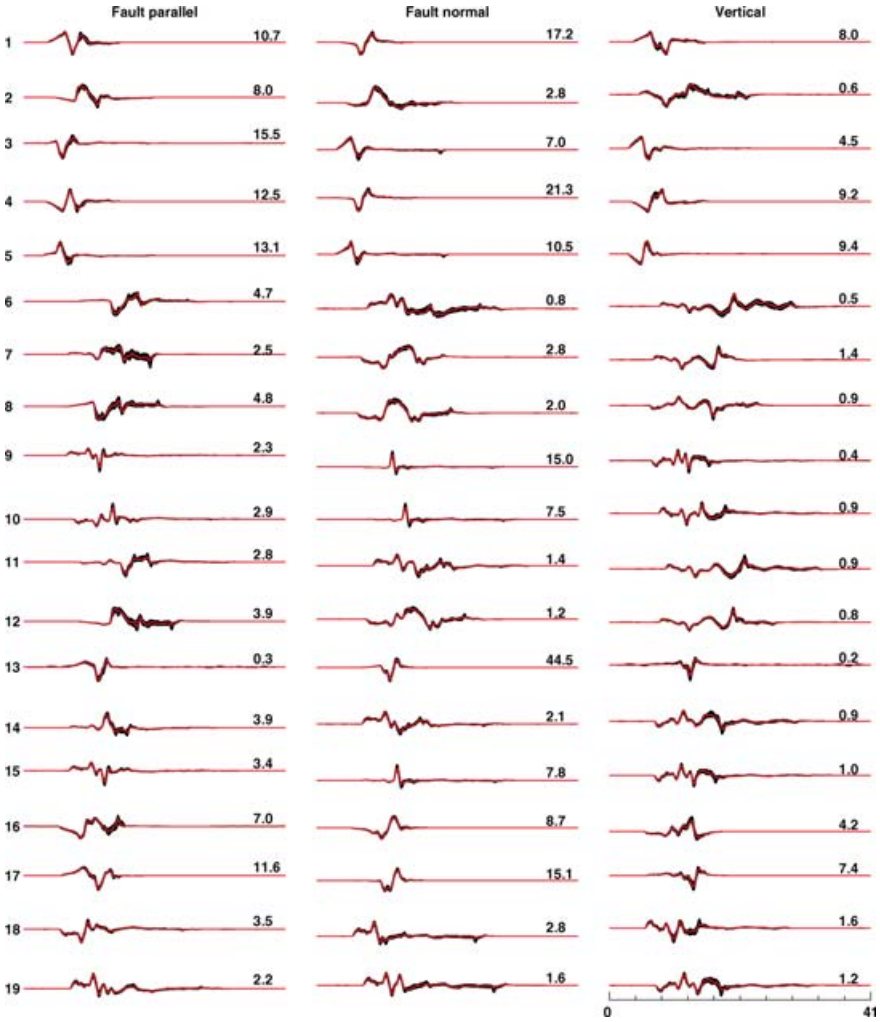


Figure 8. The level of fit produced by the rupture models shown in Fig. 7 (black data-predictions, red observed). For each waveform the maximum observed ground velocity (cm s^{-1}) is shown.

for the random number generator. We carry out the searches for the same number of generation. In Fig. 10, we show posterior marginals for all the original parameters investigated in this study considering the three independent ensembles produced. We can see some variability affecting especially the marginal probability densities for local peak slip-velocity parameters, but the general features of the inverse solution are maintained. The variability we observe comes from the fact that these three ensembles search the model space in different ways so that each of them provides different approximation of the actual posterior. This is an inherent difficulty because an exhaustive search is unfeasible and we are forced to explore the parameter space only in a limited number of points. This is especially true for large dimensional model spaces. Merging the set of models produced by independent searches into one single ensemble can be a good strategy to increase the results' stability. However one has to bear in mind that, for this kind of analysis, memory requirement and computation time scale with the size of the ensemble (see eqs 13 and 14).

6 DISCUSSION

Accurate estimates of uncertainties are needed in order to asses the reliability of the inverted solutions. As it has been pointed out by different authors (Cohee & Beroza 1994; Beresnev 2003; Ide *et al.* 2005) and is also represented in the online database of earthquake rupture models (http://www.seismo.ethz.ch/srcmod), for the same earthquake, acceptable fit to the data can be provided by different rupture models. The discrepancies between models may be due to the different choices adopted during the inversion concerning the forward modelling, the model parametrization, the inversion methodology, the type of data set and processing used. However, independently of the particular approach, intrinsic reasons render imaging the earthquake source a problem with multiple solutions: uncertainties in data and in forward modelling (which allow multiple models to be considered acceptable) and lack of resolution (due to the always limited data coverage). For a linear or linearized inversion, these factors render the problem ill-conditioned and ill-posed. For instance, Graves & Wald (2001), considering a linear slip inversion, explicitly showed that uncertainties in Green's funtions increase ill-conditioness of the problem, requiring increasing value of damping parameter (smoothing of slip in their case) to stabilize the matrix inversion.

In the context of earthquake source inversions real data are contaminated with ambient noise and also by uncertainties in the alignment of the recording sensors. More important, in our opinion, are the uncertainties due to approximations in the forward modelling.

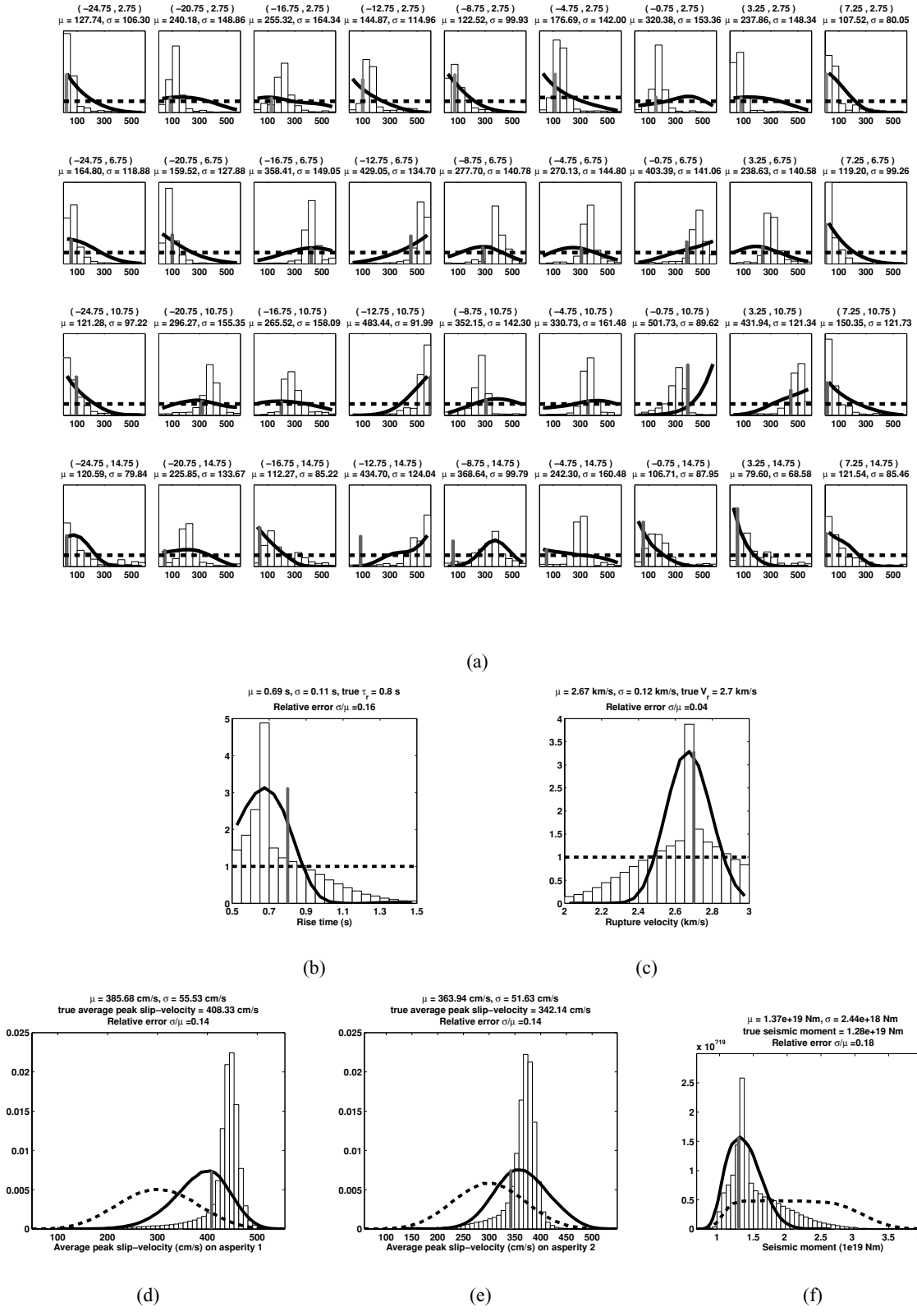
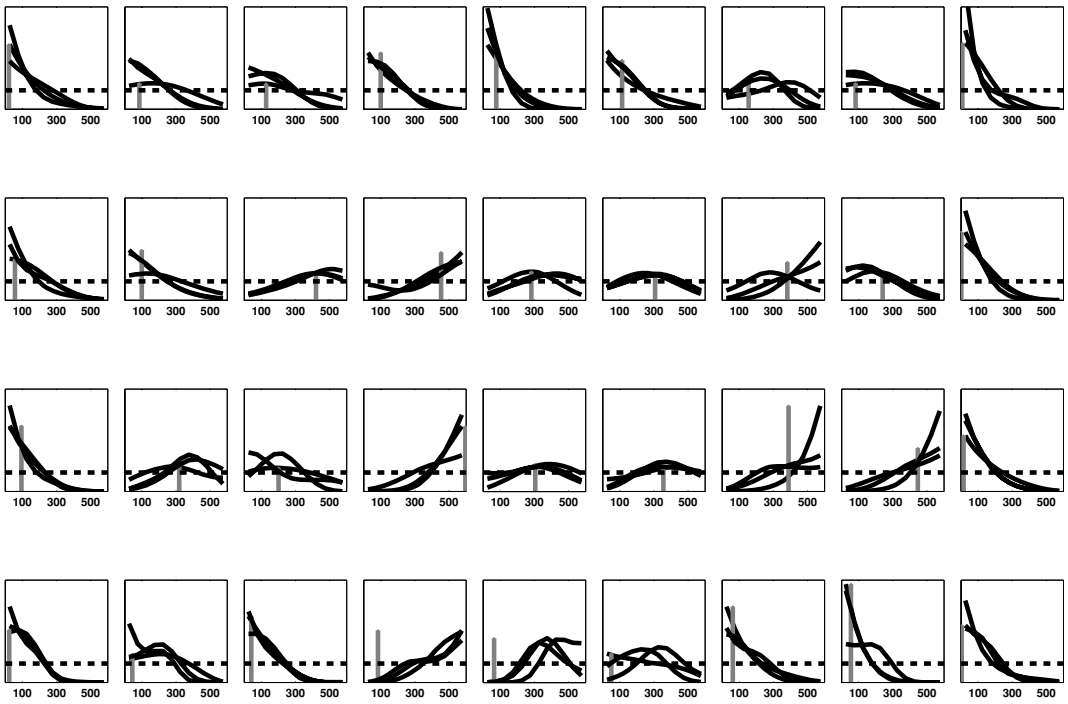
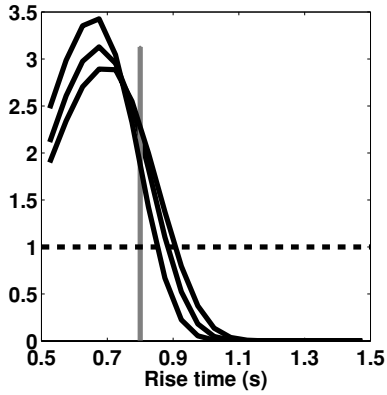


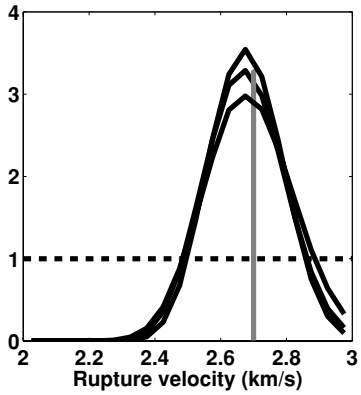
Figure 9. (a) 1-D posterior (black solid line) and prior (black dashed line) marginal probability density functions for peak slip-velocity at grid nodes on the fault surface. Each subplot corresponds to a node position. For each subplot we indicate node's coordinates (along strike, along dip) in km, with respect to a reference system centred at the epicentre, located at (0,0), and pointing towards southeast. In this reference system the hypocentre is at (0,12.5). True values are represented by vertical grey bars. For each posterior we compute the mean value μ and standard deviation σ although most of them do not show a Gaussian distribution but rather a skewed one. For comparison we plot also 'raw' marginals produced by the evolutionary algorithm (white histograms). All marginals are normalized to unit area so that relative information can be compared. (b), (c), (d), (e) and (f) present the same quantities for rise time, rupture velocity, average peak-slip velocity on asperities 1 and 2 and seismic moment, respectively.



(a)



(b)



(c)

Figure 10. (a), (b) and (c) represent 1-D posterior and prior marginal probability density functions for peak slip-velocity at grid's nodes, rise time and rupture velocity, computed considering three independent ensembles.

Real waveforms often show complexities (due to source, path and site effects) which the adopted modelling is not able to explain. The best-fitting model (the model which provides the best-numerical fit to the data) is therefore not so meaningful because we do not know precisely to what extent the best-fitting model is reproducing the modelled part of the data rather than the unmodelled one. Providing the best-fitting model as an image of the earthquake source can be therefore misleading. We suggest therefore that a better way to show results of an earthquake source estimation is to provide multiple models which are able to reproduce the data within a certain level of fit (determined by the accuracy of our data and modelling). In such a way we can visually identify what are the main features of the inverted solutions without trying to draw conclusions from the unstable details.

Lack of resolution is another important factor to bear in mind. The fact that linear inversions practically always require damping parameters implies the presence of a null space in the model space (or in other words of very close-to-zero singular values). In physical terms what happens is that the data we consider may contain very little information about certain parameters we want to invert for. In our methodology, which does not require any matrix inversion, we try to measure this lack of resolution rather than reducing it through the addition of damping parameters.

Considering a simple synthetic test, we point out that imaging the earthquake source implies a process of extraction of information from a set of data (in our case waveforms) which cannot be reduced to simply providing a best-fitting model. Efforts should be put in estimating resolution on inverted parameters. Multiple rupture models may in fact produce very similar waveforms. We want to stress that uncertainty analysis should be carried out using an appropriate theoretical framework in order to get meaningful results. We have shown how the use of an optimization algorithm to estimate

uncertainties is not suitable. We suggest that a Bayesian approach instead provides a possible way to face this problem.

The main consequence in using this approach is that our knowledge of the earthquake rupture process, as derived by the fitting of some kind of data, can be only probabilistic. In other words, available data and theoretical knowledge do not allow us to identify a single model but rather a set of models which share certain statistical properties. Identifying and quantifying these statistical properties should be the real aim of any inversion.

We used this approach considering only strong motion data. Clearly, this methodology can be applied also to investigate resolution on model parameters considering different data sets (teleseismic data, geodetic data) which all together can improve the quality of our inferences. Wald & Graves (2001) showed, for a linear slip inversion, that adding geodetic data to seismic data has a significant contribution. They found that features imaged by inversion of individual data sets alone may not be recognized when using combined data.

7 CONCLUSIONS

In this paper, we address the problem of inferring kinematic earthquake rupture parameters following a Bayesian approach. Imaging the earthquake source is seen as a problem of combination of information: *a priori* information (available before the inversion) and information contained in the data. This combination gives the posterior state of information, represented by a probability density function over the model space. We compute the posterior using a two step procedure. First we explore the model space through an evolutionary algorithm. The search of the parameter space reveals that within the same level of fit the observed waveforms can be reproduced by multiple models. All of them, though being different one from another, share some similarities. Quantifying and expressing these similarities is the aim of the second step. We use the ensemble of models found during the search to compute a geometric approximation of the true posterior and we use it to compute marginal probability density functions for each model parameter. Each marginal represents the combination of the prior information with the information that we have been able to extract from the data. From each marginal we can derive uncertainty estimates. We point out how this second step of the procedure is particularly important in order to correctly compute resolution on inverted parameters. The search algorithm alone, though being effective in finding good data fitting models, does not provide direct information about uncertainties. Misleading results can be obtained if simple statistical analysis of the ensemble of models is used to estimate resolution. We also point out how the information structure on the inverted parameters cannot be always represented in terms of Gaussian probability density functions. We show explicitly how for some parameters the posterior marginal does not follow a Gaussian shape: for these parameters the standard characterization in terms of mean value and standard deviation is not meaningful. The fact that Gaussian uncertainty hypothesis is not valid for non-linear problems is widely known but still current non-linear source estimations adopt this approximation. We also point out how estimating resolution can be limited by our ability in reconstructing the true structure of the posterior. This is an intrinsic difficulty due to the fact that exhaustive search is unfeasible and that we are always forced to explore the model space on a limited number of points. The consequence is that uncertainties estimates will be always subject to a certain amount of variability which decreases as the exploration of the model space becomes more and more extensive.

ACKNOWLEDGMENTS

We thank Malcolm Sambridge for providing the code for Bayesian inference. We thank Sigurjon Jonsson for reviewing the manuscript. Comments from three anonymous reviewers helped to improve the manuscript. This study was funded through ETH-grant TH-16/05-1.

REFERENCES

Aki, K. & Richards, P.G., 2002. *Quantitative Seismology*, Universitary Science Books, Sausalito, CA.

Beresnev, I.A., 2003. Uncertainties in finite-fault slip inversions: to what extent to believe? (A critical review), *Bull. seism. Soc. Am.,* **93,** 2445–2458.

Beroza, G.C. & Spudich, P., 1988. Linearized inversion for fault rupture behavior: application to the 1984 Morgan Hill, California, earthquake, *J. geophys. Res.,* **93,** 6275–6296.

Beyer, H.G., 2001. *The Theory of Evolution Strategies*, Springer, Berlin.

Bouchon, M., Tosöz, N., Karabulut, H., Bouin, M.-P., Dietrich, M., Aktar, M. & Edie, M., 2000. Seismic imaging of the 1999 Izmit (Turkey) rupture inferred from the near-fault recordings., *Geophys. Res. Lett.,* **27,** 3013–3016.

Cohee, B.P. & Beroza, G., C., 1994. A comparison of two methods for earthquake source inversion using strong motion seismogram, *Ann. Geof.,* **XXXVII,** 1515–1538.

Cotton, F. & Campillo, M., 1995. Frequency domain inversion of strong motions: application to the 1992 Landers earthquake, *J. geophys. Res.,* **100,** 3961–3975.

Delouis, B., Giardini, D., Lundgren, P. & Salichon, J., 2002. Joint inversion of InSAR, GPS, teleseismic, and strong-motion data for the Spatial and temporal distribution of earthquake slip: application to the 1999 Izmit mainshock, *Bull. seism. Soc. Am.,* **92,** 278–299.

Emolo, A. & Zollo, A., 2005. Kinematic source parameters for the 1989 Loma Prieta earthquake from the nonlinear inversion of accelerograms, *Bull. seism. Soc. Am.,* **3,** 981–994.

Gouveia, W.P. & Scales, J.A., 1998. Bayesian seismic waveform inversion: parameter estimation and uncertainty analysis, *J. geophys. Res.,* **103**(B2), 2759–2779.

Graves, R.W. & Wald, D.J., 2001. Resolution analysis of finite fault source inversion using one- and three-dimensional Green's functions 1. Strong motions, *J. geophys. Res.,* **106**(B5), 8745–8766.

Hartzell, H.S. & Heaton, T.H., 1983. Inversion of strong ground motion and teleseismic waveform data for the fault rupture history of the 1979 Imperial Valley, California, earthquake, *Bull. seism. Soc. Am.,* **73,** 1553–1583.

Hartzell, S., Liu, P. & Mendoza, C., 1996. The 1994 Northridge, California, earthquake: investigation of rupture velocity, rise time, and high-frequency radiation, *J. geophys. Res.,* **101,** 20 091–20 108.

Ide, S., Beroza, G.C. & McGuire, J.J., 2005. Imaging earthquake source complexity, in *Seismic Earth: Array Analysis of Broadband Seismograms*, **Vol. 157,** pp. 117–135, eds Levander, A. & Nolet, G., Geophysical Monograph Series.

Ide, S., Takeo, M. & Yoshida, Y., 1996. Source process of the 1995 Kobe earthquake: determination of spatio-temporal slip distribution by bayesian modeling, *Bull. seism. Soc. Am.,* **86,** 547–566.

Liu, P. & Archuleta, R.J., 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, doi: 10.1029/2003JB002625.

Menke, W., 1989. *Geophysical Data Analysis: Discrete Inverse Theory*, Academic Press, San Diego, CA.

Mosegaard, K. & Sambridge, M., 2002. Monte carlo analysis of inverse problems, *Inverse Problems,* **18**, R29–R54.

Olson, H.A. & Apsel, R.J., 1982. Finite faults and inverse theory with applications to the 1979 Imperial Valley earthquake, *Bull. seism. Soc. Am.,* **72,** 1969–2001.

Peyrat, S. & Olsen, K.B., 2004. Nonlinear dynamic rupture inversion of the 2000 Western Tottori, Japan, earthquake, *Geophys. Res. Lett.,* **31**, L05604, doi: 10.1029/2003GL019058.

Piatanesi, A., Tinti, E. & Cocco, M., 2004. The dependence of traction evolution on the earthquake source time function adopted in kinematic rupture models, *Geophys. Res. Lett.,* **31**, L04609, doi: 10.1029/2003GL019225.

Piatanesi, A., Cirella, A., Spudich, P. & Cocco, M., 2007. A global search inversion for earthquake kinematic rupture history: application to the 2000 western Tottori, Japan earthquake, *J. Geophys. Res.,* **112**, B07314, doi: 10.1029/2006JB004821.

Salichon, J., Delouis, B., Lundgren, P., Giardini, D., Costantini, M. & Rosen, P., 2003. Joint inversion of broadband teleseismic and interferometric synthetic apertur radar (InSAR) data for the slip history of the mw = 7.7, Nazca ridge (Peru) earthquake of 12 november 1996, *J. geophys. Res.,* **108**(B2), 2085, doi: 10.1029/2001JB000913.

Sambridge, M., 1999. Geophysical inversion with a neighbourhood algorithm–II. Appraising the ensemble, *Geophys. J. Int.,* **138,** 727–746.

Sekiguchi, H. & Iwata, T., 2002. Rupture process of the 1999 Kocaeli, Turkey, earthquake estimated from strong-motion waveforms, *Bull. seism. Soc. Am.,* **92,** 300–311.

Sekiguchi, H., Irikura, K. & Iwata, T., 2000. Fault geometry at the rupture termination of the 1995 Hyogo-ken Nambu earthquake, *Bull. seism. Soc. Am.,* **90,** 117–133.

Semmane, F., Cotton, F. & Campillo, M., 2005. The 2000 Tottori earthquake: a shallow earthquake with no surface rupture and slip properties controlled by depth, *J. geophys. Res.,* **110,** B03306.

Sen, M. & Stoffa, P.L., 1995. *Global Optimization Methods in Geophysical Inversion*, Elsevier, Amsterdam.

Spudich, P. & Archuleta, R., 1987. Techniques for earthquake ground motion calculation with applications to source parametrization of finite faults, in *Seismic Strong Motion Synthetics*, Vol. 37, pp. 205–265, ed. Bolt, B.A., Academic Press, Orlando, Florida.

Spudich, P. & Xu, L., 2002. *Documentation of Software Package Compsyn sxv3.11: Programs for Earthquake Ground Motion Calculation using Complete 1-D Green's Functions*, Academic Press, San Diego, CA.

Tarantola, A., 2005. *Inverse Problem Theory and Methods for Model Parameter Estimation*, Society for Industrial and Applied Mathematics.

Vallee, M. & Bouchon, M., 2004. Imaging coseismic rupture in far field slip patches, *Geophys. J. Int.,* **156,** 615–630.

Wald, D.J. & Graves, R.W., 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.

Wald, D.J. & Heaton, T.H., 1994. Spatial and Temporal Distribution of Slip for the 1992 Landers, California, Earthquake, *Bull. seism. Soc. Am.,* **84,** 668–691.

Wald, D.J., Helmberger, D.V. & Heaton, T.H., 1991. Rupture model of the 1989 Loma Prieta earthquake from the inversion of strong motion and broadband data, *Bull. seism. Soc. Am.,* **81,** 1540–1572.