Geophysical Journal International

Geophys. J. Int. (2014) **199,** 735–751
GJI Seismology

doi: 10.1093/gji/ggu280

Trans-dimensional finite-fault inversion

Jan Dettmer,[1,2] Roberto Benavente,[1] Phil R. Cummins[1] and Malcolm Sambridge[1]

[1]*Research School of Earth Sciences, Australian National University, Canberra ACT0200, Australia. E-mail: dettmer.jan@gmail.com*
[2]*School of Earth and Ocean Sciences, University of Victoria, Victoria BC V8W3P6, Canada*

Accepted 2014 July 18. Received 2014 July 17; in original form 2014 March 26

SUMMARY

This paper develops a probabilistic Bayesian approach to the problem of inferring the spatiotemporal evolution of earthquake rupture on a fault surface from seismic data with rigorous uncertainty estimation. To date, uncertainties of rupture parameters are poorly understood, and the effect of choices such as fault discretization on uncertainties has not been studied. We show that model choice is fundamentally linked to uncertainty estimation and can have profound effects on results. The approach developed here is based on a trans-dimensional self-parametrization of the fault, avoids regularization constraints and provides rigorous uncertainty estimation that accounts for model-selection ambiguity associated with the fault discretization. In particular, the fault is parametrized using self-adapting irregular grids which intrinsically match the local resolving power of the data and provide parsimonious solutions requiring few parameters to capture complex rupture characteristics. Rupture causality is ensured by parametrizing rupture-onset time by a rupture-velocity field and obtaining first rupture times from the eikonal equation. The Bayesian sampling of the parameter space is implemented on a computer cluster with a highly efficient parallel tempering algorithm.

The inversion is applied to simulated and observed W-phase waveforms from the 2010 Maule (Chile) earthquake. Simulation results show that our approach avoids both over- and underparametrization to ensure unbiased inversion results with uncertainty estimates that are consistent with data information. The simulation results also show the ability of W-phase data to resolve the spatial variability of slip magnitude and rake angles. In addition, sensitivity to spatially dependent rupture velocities exists for kinematic slip models.

Application to the observed data indicates that residual errors are highly correlated and likely dominated by theory error, necessitating the iterative estimation of a non-stationary data covariance matrix. The moment magnitude for the Maule earthquake is estimated to be ∼8.9, with slip concentrated in two zones updip of and north and south of the hypocentre, respectively. While this aspect of the slip distribution is similar to previous studies, we show that the slip maximum in the southern zone is poorly resolved compared to the northern zone. Both slip maxima are higher than reported in previous studies, which we speculate may be due to the lack of bias caused by the regularization used in other studies.

Key words: Inverse theory; Probability distributions; Earthquake source observations; Computational seismology; Statistical seismology; Early warning.

1 INTRODUCTION

Finite-fault slip inversion provides estimates of the spatiotemporal evolution of earthquake rupture over one or more discretized fault surfaces and represents a way to study the physics of rupture (Olsen & Apsel 1982; Hartzell & Heaton 1983). The fundamental problem of estimating slip on an unknown fault surface from incomplete and noisy data is highly non-linear, non-unique and challenging to address without making substantial assumptions about some of the unknowns. Over the last three decades, significant work has been directed at estimating slip for linearized and non-linear parametrizations to obtain optimal parameter estimates (Hartzell 1989; Beresnev 2003; Ide 2007).

While the inverse problem can be linearized by constraining the parametrization to special cases, for example, the multiple time-window method (Hartzell & Heaton 1983), the consequences of such simplifications are difficult to quantify. Linearized inversion approaches are typically overparametrized, that is, the spatial and temporal discretization are below the resolving power of the data, and require some form of regularization (often through spatial

© The Authors 2014. Published by Oxford University Press on behalf of The Royal Astronomical Society.

Downloaded from http://gji.oxfordjournals.org/ by guest on September 2, 2014

smoothing) since the inverse problems are ill-posed and underdetermined. The impacts of regularization constraints on inversion results are difficult to quantify which can result in substantial difficulty in interpreting results. In particular, regularization has been shown to produce biased and underestimated uncertainties (Aster *et al.* 2005). In addition, regularization requires specification of parameters that scale the solution between the regularized solution and that due to the data information. While these parameters can be estimated based on misfit (e.g. L-curve test) or Bayesian model selection [e.g. Akaike's Bayesian Information Criterion (BIC, Yabuki & Matsu'ura 1992; Matsu'ura *et al.* 2007; Guo *et al.* 2011)], rigorous estimation is often overlooked and is associated with additional computational cost. Finally, regularization is specified *a priori* for the whole model, which may be inadequate in cases where data constraint varies between model parameters. For example, if a global smoothing matrix is assumed all regions of the model are exposed to the same smoothing which can over- and/or underregularize various aspects of the model. In addition, these undesirable aspects can not only affect parameters of identical meaning (various regions of slip may be resolved to varying degrees) but also parameters of different physical meaning (different parameter types may have different spatial/temporal resolution).

Non-linear optimization techniques (e.g. simulated annealing, Hartzell *et al.* 1996) have also been extensively applied to account for some non-linear effects, such as treating rise time as unknown. While regularization is sometimes applied in non-linear inversion, it is not required. For unregularized inversions, optimal parameter vector estimation requires quantitative model selection. However, both linear and non-linear models are typically specified *ad hoc* based on empirical estimates of data resolution. Consequently, slip inversion results for identical earthquakes and data often vary substantially and show strong dependence on the choice of parametrization (Beresnev 2003). While non-linear approaches have relaxed some of the assumptions at higher computational cost, non-linear parameter uncertainties of slip models have been only recently investigated (Minson *et al.* 2013).

In Bayesian inversion, the solution to the inverse problem is given probabilistically by the posterior probability density (PPD) of model parameters (MacKay 2003; Tarantola 2005) which expresses the state of information about the parameters. Bayesian inversion considers prior information (independent information about the parameters expressed as a probability) and updates the prior with data information (expressed in terms of the likelihood function) to obtain the PPD. Here, the term model is considered to be general and includes a particular choice of physical theory (e.g. wave propagation that predicts the response of the system to a signal), an appropriate set of model parameters (e.g. the spatiotemporal fault parameters), and a statistical representation of the uncertainty distribution for the data (which includes measurement errors and errors due to model limitations). The uncertainty distribution for the data (the data covariance matrix) may also depend on parameters. A realization of the model is given by a parameter vector. The data are generally noisy and incomplete (e.g. limited frequency band, azimuthal coverage) which causes uncertainty in the parameters. In a Bayesian approach, it is crucial to define and differentiate between variability and uncertainty. Variability is a measure of the inherent spatial and/or temporal heterogeneity in an environmental property. Uncertainty is a measure of our knowledge of an environmental parameter value, and is quantified by a probability density. To study parameter variability (the ultimate goal of inversion work), rigorous uncertainty estimation is a pre-requisite.

Examples of application of Bayesian inference to linear slip inversion include Yabuki & Matsu'ura (1992) and Matsu'ura *et al.* (2007) for geodetic data, regularized linear inversion of geodetic data with positivity constraints (Fukuda & Johnson 2008), and regularized mixed linear–non-linear joint inversion of multiple data sets (Fukuda & Johnson 2010). The leading work of Minson *et al.* (2013) developed a fully non-linear Bayesian inversion including a kinematic source model on a fixed, regular grid. The source-time function is parametrized with a rise time and a rupture velocity and causality is satisfied by solving the eikonal equation. A non-linear Bayesian approach was successfully applied to joint inversion of geodetic, seismic and tsunami waveform data from the 2011 Tohoku-Oki earthquake (Simons *et al.* 2011; Minson *et al.* 2014). In these applications, the model is assumed to be fixed and model selection is not applied.

The uncertainties from Bayesian inversion quantify the accuracy of parameter estimates for the particular model choice (including choice of physical theory, appropriate set of model parameters and statistical representation of the data-uncertainty distribution), and can depend strongly on that choice. Therefore, it is important to estimate the model quantitatively and objectively from the data (a process referred to as model selection). Model selection must avoid both over- and underparametrization (MacKay 2003; Dettmer *et al.* 2009; Brooks *et al.* 2011; Menke 2012). Overparametrization is associated with models that include too many parameters which cannot be properly constrained by the data. Such models overfit data (fitting noise features) and parameter inference leads to results with unconstrained structure and unrealistically high uncertainty estimates. Underparametrization is associated with models that include too few parameters to properly fit the data, leaving structure unresolved and biased. This requirement for parsimony is intrinsically satisfied by Bayes' theorem (MacKay 2003) and quantitative model selection can be carried out by evaluating the denominator in Bayes' theorem (referred to as Bayesian evidence). The extension of Bayesian inference to model selection provides the ability to discriminate between models and make quantitative statements about which model is preferred. It follows that the inferences carried out with the posterior density of the preferred model are superior to those obtained from a posterior for an inferior model. Such model selection can be addressed efficiently by trans-dimensional (trans-D) models (Brooks *et al.* 2011). Trans-D inversion is based on a hierarchical Bayesian formulation (Green 1995, 2003) where model specification is relaxed so that a group of models is simultaneously considered. Group members contribute to the solution according to their support by the data and prior information. The trans-D model may contain multiple members that fit the data and should be considered jointly in uncertainty estimation. The result is a probability distribution which extends over all group members and provides ensemble-parameter estimates with rigorous uncertainties that inherently include uncertainty due to this model-selection ambiguity.

However, it should be noted that all Bayesian inferences, including model selection, are fundamentally tied to the assumption about the statistical distribution form (e.g. Gaussian) of the residual errors (a combination of theory and measurement errors) and the parameters estimated for that distribution (e.g. variances and covariances). Poorly met assumptions and/or poor distribution parameter estimates can cause biases in inversion results (MacKay 2003). Inverse problems in seismology can be particularly challenging in this context, since theory errors are often dominant in the residuals. While these issues cannot be fully alleviated in all cases, statistical analysis of posterior residual errors should be applied to examine the

Downloaded from http://gji.oxfordjournals.org/ by guest on September 2, 2014

validity of the assumptions about residual-error statistics (Dettmer *et al.* 2007). Posterior residuals that are reasonably consistent with the assumptions raise confidence in the results, while poor consistency may require application of an alternative likelihood function.

This work develops a novel non-linear Bayesian approach to slip inversion that rigorously addresses model selection with a trans-D model which efficiently and locally self-adapts to data resolution. In addition, strongly correlated data errors are quantified in the likelihood function by a combination of hierarchical and empirical Bayesian estimation. The inversion is validated with simulated data and then applied to the inversion of W-phase data from the 2010 Maule (Chile) earthquake.

Trans-D models have been applied widely in geophysics (Malinverno 2002; Sambridge *et al.* 2006; Dettmer *et al.* 2010a; Minsley 2011; Ray & Key 2012; Bodin *et al.* 2012) and are particularly effective for 2-D inverse problems with irregular grids (Bodin & Sambridge 2009; Dettmer & Dosso 2013), where the curse of dimensionality can preclude the use of regular-grid parametrization. Sambridge *et al.* (1995) proposed irregular grids based on Voronoi cells (Voronoi 1908) for spatial interpolation, grouping regions in terms of nearest neighbours. Voronoi cells can be used as efficient parametrizations since nodes can move freely and nodal density vary to be sparse where little parameter variability and/or low data information content exist and dense where variability and/or information is high (Bodin *et al.* 2009). A substantial generalization is to allow grids where the number of nodes sampled in the inversion is variable (Bodin & Sambridge 2009). In such trans-D models, the parameters include the number of nodes, node positions and environmental properties at each node, which can result in substantially fewer parameters compared to regular grids. Trans-D models have been shown to self-adapt locally to the data information based on Bayesian parsimony (Bodin & Sambridge 2009; Minsley 2011). The parsimony, which is intrinsic to a Bayesian trans-D model, controls structure so as to be consistent with prior and data information, and no subjective choices/regularizations are required.

We account for residual errors (including those due to measurement process and model limitations) in the inversion by a combination of empirical and hierarchical estimation. Residual-error estimates from an initial inversion that assumes uncorrelated errors are used to estimate the data covariance matrix for an optimal model. The data covariance matrix is a block matrix that treats the multiple seismic stations as independent and results in no requirement for weighting of data for the various stations. Rather, weights are estimated from the data (i.e. from the noise level which becomes an unknown in the inversion). The residual-error estimation process also accounts for non-stationary effects which are common for waveform data because the error statistics change as a function of time at each station. Since the optimal model is a point estimate limited to one fixed-dimensional member of the trans-D model, a hierarchical scaling is applied to each station. Note that many of these residual-error characteristics (non-stationarity, strong correlations) are common in waveform inversion and are largely due to errors stemming from limitations in the model. Accounting for these model limitations is particularly important in Bayesian inference when uncertainty estimation is of interest.

For non-linear problems, no analytic solution exists for the posterior density and numerical sampling methods must be applied. Here, the sampling is implemented with the reversible-jump Metropolis–Hastings–Green (MHG) algorithm (Geyer & Moller 1994; Green 1995) which samples by probabilistically accepting/rejecting new model vectors generated by a proposal density. The proposal choice can strongly affect sampling efficiency and finding efficient densities has historically been one of the greatest challenges in applying Bayesian inference to large-scale problems (Brooks *et al.* 2011). Here, trans-D interacting Markov chains (also referred to as parallel tempering, PT, Geyer 1991; Jasra *et al.* 2007; Dettmer & Dosso 2012; Sambridge 2014) are applied to improve sampling efficiency by simulating many Markov chains in parallel on a computer cluster. The likelihood functions of the various chains are tempered (raised to a power between zero and one) such that chains with an increasing level of tempering (low powers) increasingly de-emphasize data information and sample more widely. Chains with no tempering sample from the posterior of interest and are used for posterior inference. The parallel chains exchange information by swapping tempering parameters, and the exchange is governed by the Metropolis–Hastings (MH) criterion for multiple chains.

The fault is parametrized by a variable number of nodes which define homogeneous nearest neighbourhood regions. Each node position has two unknown parameters (along-strike and along-dip positions). The slip vector at each node has two components (symmetric around an initial guess for the rake) of unknown magnitude which also estimates the rake (Hartzell *et al.* 1996). The rupture time at each node is parametrized by a rupture velocity and the rise time of the triangular source-time function is fixed at 24 s. Green's functions (GFs) are defined on an underlying grid which is finer than the resolving power of the data. The node values of the irregular grid always guarantee a natural partitioning of the fault which is used to copy the slip parameters from Voronoi nodes onto the underlying regular grid to carry out data predictions. The underlying fine grid is also used to compute first arrival times for the rupture wave front by solving the eikonal equation.

The inversion is applied to simulated W-phase (Kanamori 1993) data and W-phase observations from the 2010 Maule (Chile) earthquake. W-phase data are of particular interest in tsunami early warning systems (Kanamori & Rivera 2008). Recently, Benavente & Cummins (2013) have shown that W-phase data can be used for efficient estimation of the slip distribution. However, uncertainties may have significant impact on the utility of W-phase data for early warning but are poorly understood. This study is in part motivated by studying the uncertainties associated with slip inversion of W-phase data. The results for simulated data show the principal ability of W-phase data to resolve slip in finite-fault models well (including peak slip) and also its ability to resolve spatially dependent rupture velocities.

Application to the Maule earthquake shows that the slip pattern is well resolved by the data and peak slip is well constrained in some areas but not in others. These data display strongly correlated residual errors that are likely due to a combination of the long-period bandpass filter typically applied in W-phase data processing and model limitations. Significant attention is directed to estimating data covariance matrices for all stations. Posterior residual-error analysis is carried out to judge the quality of the covariance estimates. The covariance estimates significantly improve the residual statistics but do not remove all effects of correlation.

2 INVERSION METHOD

2.1 Bayesian inference with trans-D models

A trans-D model (Green 1995, 2003) is used here to treat the number of grid nodes as unknown. The underlying theory is briefly reviewed below and more detailed discussion on trans-D inversion

Downloaded from http://gji.oxfordjournals.org/ by guest on September 2, 2014

in geophysics can be found elsewhere (Malinverno 2002; Sambridge *et al.* 2006; Bodin & Sambridge 2009; Dettmer *et al.* 2010a; Minsley 2011; Dettmer *et al.* 2012; Ray & Key 2012). In particular, we follow Dettmer & Dosso (2012, 2013) to implement trans-D sampling by combining the reversible-jump algorithm with PT on a computer cluster to achieve efficient sampling of a highly non-linear, high-dimensional 2-D problem. Trans-D models allow a group of parametrizations to be considered simultaneously for analysis. The state space of the inversion is composed of the union of subspaces for the various models. These subspaces can vary in dimension and also in the physical meaning of parameters. Here, the inversion is carried out such that the subspaces correspond to models that differ in the number of Voronoi nodes and hence dimension. These subspaces are indexed with parameter k with a uniform prior and the posterior density is defined over all subspaces (of various dimensions). To carry out inferences on the posterior, k is typically treated as a nuisance parameter and marginalized over. The ensemble inferences then include all models considered in the analysis, and account for how limited knowledge about the parametrization affects the posterior-parameter and uncertainty estimates. Inferences obtained in this manner avoid the inherent biases involved in selecting a single fixed parametrization and are hence more realistic and reflective of the state of knowledge about model parameters.

Let $\mathbf{d}$ be a random variable of N observed data and $\mathcal{M}_k$ denote a group of models specifying particular choices of physical theory, model parametrization, and error statistics, where $k \in \mathcal{K}$ and $\mathcal{K}$ is a countable set. In this particular application, k will index the number of Voronoi nodes in the parametrization. Let $\mathbf{m}_k$ be a random variable of M_k parameters representing a realization of model $\mathcal{M}_k$. Green (1995) shows that Bayes' rule can be written for a Bayesian hierarchical model to include parameter k

$$p(k, \mathbf{m}_k|\mathbf{d}) = \frac{p(k)p(\mathbf{d}|k, \mathbf{m}_k)p(\mathbf{m}_k|k)}{\sum_{k' \in \mathcal{K}} \int p(k')p(\mathbf{d}|k', \mathbf{m}'_{k'})p(\mathbf{m}'_{k'}|k')\mathrm{d}\mathbf{m}'_{k'}}, \tag{1}$$

where $p(k)$ is the prior over the $\mathcal{K}$ models considered. The state variables $(k, \mathbf{m}_k)$ are of dimension M_k and the state space is trans-D and given by the union of all fixed-dimensional spaces in $\mathcal{K}$, that is, $\bigcup_{k \in \mathcal{K}}(\{k\} \times \mathbb{R}^{M_k})$. A Markov chain that samples this state space can be defined and converges to the trans-D posterior $p(k, \mathbf{m}_k|\mathbf{d})$. Note that the PPD $p(k, \mathbf{m}_k|\mathbf{d})$ intrinsically addresses model selection and typical inferences about expectations do not require the computation of normalizing constants (the denominator in eq. 1). This is a substantial advantage over model selection by way of normalizing constants/evidence for geophysical problems (Dettmer *et al.* 2010b). The conditional probability $p(\mathbf{d}|k, \mathbf{m}_k)$ in eq. (1) describes the residual-error statistics, where residual errors are defined as the difference of observed and predicted data. For observed (fixed) data, $p(\mathbf{d}|k, \mathbf{m}_k)$ is interpreted as the likelihood function $L(\mathbf{x})$, where $\mathbf{x} = (k, \mathbf{m}_k)$ is the parameter (state) vector. The likelihood is a function of only $\mathbf{x}$ and quantifies the probability that a given set of parameters gave rise to the observed data: for low L, $\mathbf{x}$ is judged unlikely to have given rise to the observed data; if L is high, $\mathbf{x}$ is likely to be the source of the data.

The slip and rupture-velocity distributions on the fault are described fully by the node positions and the parameters associated with the nodes, resulting in five unknowns for every node: along-strike node position u, along-dip node position w, first slip component s_1, second slip component s_2 and rupture velocity v_{rup}. This hierarchical partition model allows data and prior to determine the spatial complexity of rupture, and adapts locally to the detail resolved by the data.

The model is sampled with the MHG algorithm (Green 1995), where the creation and deletion of nodes are implemented as reversible-jump steps. The MHG algorithm is a generalization of the MH algorithm and is implemented by proposing new states $\mathbf{x}' = (k', \mathbf{m}'_{k'})$ of the Markov chain using a proposal distribution q which is generally centred on the current state $\mathbf{x}$. The proposed state is then accepted/rejected based on the acceptance probability

$$\alpha = \min\left[1, \frac{p(\mathbf{x}')}{p(\mathbf{x})}\left(\frac{L(\mathbf{x}')}{L(\mathbf{x})}\right)^{\beta} \frac{q(\mathbf{x}|\mathbf{x}')}{q(\mathbf{x}'|\mathbf{x})}|\mathbf{J}|\right], \tag{2}$$

where β is an annealing parameter (see below for details), and $|\mathbf{J}|$ is the determinant of the Jacobian for the diffeomorphism from $\mathbf{x}$ to $\mathbf{x}'$ (assures proper dimension mapping of proposals, see Green 2003,for a detailed discussion). A new node is proposed by uniformly sampling a position on the fault (drawing along-strike and along-dip positions from a uniform density) and perturbing the slip parameters that were previously associated with that position. A Gaussian proposal distribution centred on the previous parameters is used for the perturbation (although other choices are possible). The proposed node is accepted/rejected according to eq. (2) if α is greater than a uniform random number. To delete, a random node is selected and removed and the resulting model is again evaluated according to eq. (2). A detailed derivation of the acceptance probability for adding/deleting nodes is given by eq. (20) in Dettmer *et al.* (2010a). Upon creation and deletion of nodes, the Voronoi cells provide a natural repartitioning of the space that allows straightforward evaluation of the proposal ratio (which requires the step size of the perturbation). In addition to jumps that create/delete nodes, perturbations of node positions and the associated parameters are performed as additional steps in the Markov chain.

2.2 Non-parametric data covariance matrix with hierarchical scaling

To evaluate eq. (2), the likelihood function must be defined which requires specifying the data uncertainty distribution for the measured data. However, since data uncertainties include measurement and theory errors, their distribution is rarely known independently, but must be estimated from residual errors which are the difference of observed and predicted data. For observed data $\mathbf{d}$ and predicted data $\mathbf{d}(\mathbf{x})$, the residuals are given by $\mathbf{r} = \mathbf{d} - \mathbf{d}(\mathbf{x})$. Hence, these residual errors are model-based estimates. Lack of specific knowledge suggests assuming a simple distribution form for the residuals, for example, a multivariate Gaussian which is supported by the Central Limit Theorem. For Gaussian-distributed residuals, the likelihood function is given by

$$L(\mathbf{x}) = \prod_{i=1}^{S} \frac{1}{(2\pi)^{N_i/2} \mid \mathbf{C}_i \mid^{1/2}} \exp\left(-\frac{1}{2}\mathbf{r}_i^{\top}\mathbf{C}_i^{-1}\mathbf{r}_i\right), \tag{3}$$

where i indexes S data sets $\mathbf{d}_i$ at multiple stations and/or data types that are assumed to have independent noise, N_i is the number of data points for that data set, and $\mathbf{C}_i$ are covariance matrices of the residual errors $\mathbf{r}_i$. Note that eq. (3) is general and can be applied to any number of stations and various types of data so that joint inversion is straightforward. The relative weighting of data sets/types is intrinsically given by the $\mathbf{C}_i$ with no requirement for *ad hoc* scaling/empirical weights.

However, the $\mathbf{C}_i$ are generally not known independently because the sources of theory errors are poorly understood. Therefore, a crucial aspect of Bayesian inversion is the estimation of $\mathbf{C}_i$. Note that covariance estimation based directly on the data (e.g. a quiet

Downloaded from http://gji.oxfordjournals.org/ by guest on September 2, 2014

section of a seismogram, ensemble averages of multiple measurements) fails to consider the theory error which is intrinsic to most geophysical problems and can in fact be important. Therefore, it is desirable to estimate $\mathbf{C}_i$ from residual errors that include the effects of theory error. This estimation can be carried out by parametric approaches, where the covariance matrix is parametrized with few parameters, such as autoregressive error models (Dettmer *et al.* 2012; Steininger *et al.* 2013). While hierarchical estimation of covariances is possible and desirable for trans-D models, to date only simple models have been applied (Bodin *et al.* 2012; Dettmer *et al.* 2012). For the W-phase data considered in this work, such simple models (exponential decay of correlation with lag) are insufficient since the correlations appear more complex.

Alternatively, non-parametric approaches (Holland *et al.* 2005; Dettmer *et al.* 2008) can be applied to iteratively estimate $\mathbf{C}_i$ under the initial assumption of uncorrelated Gaussian errors of unknown standard deviation. To derive such an implicit likelihood function, we temporarily assume uncorrelated errors of standard deviation σ_i ($\mathbf{C}_i = \sigma_i^2\mathbf{I}$, where $\mathbf{I}$ is the identity matrix). In this case, the likelihood function simplifies and can be used to derive maximum likelihood estimates $\hat{\sigma}_i$ by setting $\partial L_i/\partial\sigma_i = 0$ and solving for σ_i. After substituting $\hat{\sigma}_i$ into eq. (3) it can be shown (Dosso & Wilmut 2006) that the likelihood function becomes

$$L(\mathbf{x}) \propto \prod_{i=1}^{S} \exp\left(-\frac{N_i}{2}\log_{\mathrm{e}}\left(\mathbf{r}_i^{\top}\mathbf{r}_i\right)\right), \tag{4}$$

where the normalizing constant is omitted for simplicity, resulting in the proportionality sign. Note that eq. (4) represents a hierarchical model, that is, the standard deviation of the residual error is estimated by the data. However, instead of explicitly sampling over S standard deviation parameters, the analytic expression $\hat{\sigma}_i$ is used for more efficient estimation (for large data sets with many stations, this approach can substantially reduce the number of parameters). Hierarchical error models (Malinverno & Briggs 2004; Dosso & Wilmut 2006) are a substantial advantage of a Bayesian approach, since error statistics are estimated as part of the inverse problem which can greatly simplify the inversion.

Eq. (4) is used in an initial phase of the inversion to obtain optimal point estimates of $\mathbf{x}$ to provide data residuals $\mathbf{r}_i$. Under the assumption of ergodicity, the residuals are considered to represent a realization of the errors. The (biased) autocovariance functions of the $\mathbf{r}_i$ are then used to approximate the data covariance matrices. This approach results in diagonal-constant matrix estimates that can be applied subsequently via eq. (3). Note that the approach is typically applied iteratively until the $\mathbf{C}_i$ are stable (Dettmer *et al.* 2007), which can be performed during the burn-in phase of the sampling algorithm. This approach requires some damping of the autocovariance function to ensure the resulting matrix is positive definite. The damping is applied here by employing the biased estimate for the autocovariance function (i.e. always normalize by the total number of data for all lags) which dampens values at large lags, where only few data pairs are available for estimation. The covariance matrix is updated every 1000 iterations during the burn-in phase and then held fixed during sampling. Each update of the matrix is based on the highest likelihood model for the most probable k value. Since the covariance matrices are estimated from optimal models, estimates change little once the algorithm samples appropriate high-likelihood regions (Dettmer *et al.* 2008). Here, covariance matrix estimates were empirically found to change little after a few thousand Markov chain steps, while burn-in periods for the sampler were much longer (100 000 burn-in steps were used in the Maule case).

Diagonally constant matrices assume errors to be stationary at each seismic station. However, the W-phase data considered here are strongly non-stationary, likely due to theory errors that scale with the magnitude of the data. Non-stationarity is accounted for by estimating the standard deviation in a running window of length of 1/5 of the number of data at each station (to ensure a sufficient number of data to reasonably estimate the standard deviation). The initial $\mathbf{C}_i$ estimates are then normalized to give correlation matrices and finally each element is scaled according to the non-stationary standard deviation estimates (Dettmer *et al.* 2007). The resulting covariance matrices are for a single model that may not be representative for all members of the trans-D model. Therefore, we include a hierarchical scaling parameter for each station in the likelihood function, so that the magnitude of the covariance matrix is scaled by a single parameter ξ_i:

$$L(\mathbf{x}) \propto \prod_{i=1}^{S} \exp\left(-\frac{N_i}{2}\log_{\mathrm{e}}\xi_i - \frac{1}{2\xi_i}\mathbf{r}_i^{\top}\mathbf{C}_i^{-1}\mathbf{r}_i\right). \tag{5}$$

The assumption about the distribution form of residual errors should be examined *a posteriori*. This can be carried out by examining standardized residuals (by multiplying $\mathbf{r}_i$ with the Cholesky decomposition of $\mathbf{C}_i$, e.g. Dettmer *et al.* 2007), which should be uncorrelated, Gaussian distributed with unit standard deviation. Absence of significant evidence against the assumptions raises confidence in the inversion results. If significant evidence against the original assumption exists, covariance estimates are likely poor or other distribution forms (e.g. Laplacian) should be considered.

2.3 Fault parametrization

The fault model extends over a distance of interest along strike and dip and is approximately centred around an initial hypocentre location. The hypocentre location is itself treated as unknown (parametrized in terms of along-strike and along-dip position) within a small region around an initial location. The prior for the hypocentre location is uniform over that region. The fault extent is based on independent information, such as the distribution of aftershocks. The plane orientation in space is given by fixed strike and dip angles obtained here from independent point-source inversion results.

The fault is parametrized here as a partition model in terms of Voronoi cells which are non-overlapping nearest neighbour regions that are fully described by node position (along-strike and along-dip distance) and a norm that defines the distance between the node position and any point in space (Fig. 1). The norm is given here by the Euclidian distance $d^{(i)}$ for the ith node in normalized coordinates u_{n} (along strike) and w_{n} (along dip)

$$d^{(i)} = \sqrt{(u_{\mathrm{n}}^{(i)} - u_{\mathrm{n}})^2 + (w_{\mathrm{n}}^{(i)} - w_{\mathrm{n}})^2}\,. \tag{6}$$

The subfault parameters associated with a node are constant within the cell area.

A vector of Voronoi nodes can represent a 2-D environment efficiently with a small number of parameters. This form of partitioning is particularly suited to cases where some reasonably homogeneous regions are expected in the environment. In such situations, Voronoi nodes can be arranged (driven by data) so that few nodes represent large areas. The trans-D hierarchical formulation of the model ensures parsimony, constraining the number of nodes to be consistent

Downloaded from http://gji.oxfordjournals.org/ by guest on September 2, 2014

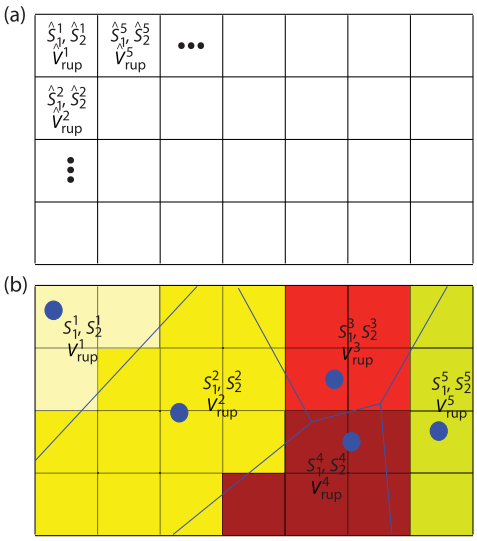

Figure 1. Two grids are used to carry out the inversion: (a) The regular subfault grid on which Green's functions are computed and (b) the variable dimension, irregular grid that is used to define the inversion parametrization. For each data prediction, the Voronoi-node values are copied onto the regular grid nodes within the cell to match the Green's functions representation. Quantities on the regular grid are denoted with the ˆ symbol in (a) and the coloured regions in (b) indicate which parts of the grid are constrained by the various Voronoi nodes.

with data and prior information. We choose cells to represent regions of constant values since previous work (Bodin & Sambridge 2009) has shown that such cells have the ability to capture both abrupt changes and smooth transitions in posterior ensemble estimates. Alternatively, smooth interpolation may be advantageous in some cases while it may cause smooth representation where the actual process is discreet. Smooth models are not investigated in this work.

The prediction of waveforms is based on a GF library. The GF are computed for W phases by normal-mode summation (Kanamori & Rivera 2008) for a reference earth model (Dziewonski & Anderson 1981). To obtain a smooth representation of the irregular inversion grid, GFs are given on an underlying fine grid with regular 12-km spacing (Fig. 1). The grid spacing of the fine grid is chosen to be below the resolution power of the data. Due to the finite regular spacing, some changes in Voronoi-node positions can result in no change to the parameters of the fine grid. However, this does not significantly affect the algorithm. For each data prediction, the rupture parameters (s_1, s_2 and v_{rup}) for each node are applied to all regular grid nodes in its nearest neighbour region, resulting in a homogeneous area (Fig. 1). However, since initiation time of the rupture depends on v_{rup} and the distance from the hypocentre, it is not constant within the cell. Rather, the eikonal equation is solved for the v_{rup} distribution on the regular GF grid with a fast finite-difference solver (Podvin & Lecomte 1991). The eikonal equation ensures causality for arbitrary spatial distributions of v_{rup}. Due to the long periods, W-phase data are not sensitive to the rise time of the source-time function (Kanamori & Rivera 2008) and a triangular function with a fixed rise time of 24 s is applied here. In addition, rise times between 15 and 30 s were found to not change linear inversion results based on the algorithm of Benavente & Cummins (2013). For computational efficiency, this source-time function is convolved with the GFs in the library and variations in initiation time (given by the spatial v_{rup} distribution) are accounted for by shifting the GFs in time.

2.4 Sampling efficiency

A fundamental challenge in non-linear Bayesian inference is the efficient sampling of the posterior density. This issue is particularly difficult for high-dimensional problems, such as 2-D parametrizations with several unknown parameters as a function of space. In addition, sampling is generally more difficult for data with high information content, where the likelihood function is highly peaked with high-likelihood regions being confined to extremely small parts of the parameter space. Since high data-information content is associated with low noise levels, sampling is typically more difficult for data with low noise level.

To date, little work exists in non-linear Bayesian sampling for slip inversions. Minson *et al.* (2013) develop a highly efficient fixed-dimensional sampling algorithm for joint inversion that employs sequential Monte Carlo techniques (resampling, annealed posterior sequence) similar to methods developed in other fields (Jarzynski 1997; Neal 2001; Godsill & Clapp 2001). In addition, Minson *et al.* (2013) estimate parameter covariances from the sample to allow efficient proposal of new steps which is similar to proposal densities in principal component space (Dosso 2002). However, to date efficient proposals in principal component space have not been developed for trans-D problems.

Here, population Monte Carlo/PT (Geyer 1991; Jasra *et al.* 2007) is applied to improve sampling from the trans-D space. PT samples from a sequence of tempered densities which has been shown to improve sampling efficiency for trans-D inverse problems with peaked likelihoods (Dettmer & Dosso 2012, 2013). Tempering relaxes the likelihood function by raising it to the power of an annealing parameter ($0 \leq \beta \leq 1$, see eq. 2) which effectively flattens peaks, making transitions of the Markov chain easier. By simulating many parallel chains for various values of β and allowing information exchange between chains, sampling efficiency can be dramatically improved for difficult problems. While additional computational cost is associated with the additional chains, the accelerated rate of convergence often dwarfs that cost. PT is particularly well suited for a parallel implementation and large problems. It is formally based on augmenting the posterior with additional densities to obtain π^*

$$\pi^* = \prod_{i=1}^{N} \pi_i \,. \tag{7}$$

Here, N is the number of parallel Markov chains, π_i is the posterior sampled by Markov chain i for a given value of the tempering parameter β_i, and at least one $\pi_i \equiv \pi = p(\mathbf{x}|\mathbf{d})$ (i.e. the posterior sampled with $\beta = 1$). The idea is that the additional densities aid in sampling from π to a degree that the additional computational effort is insignificant. Note that the π_i-sequence can be chosen in any meaningful way and is not limited to a sequence of tempered densities (hence, some literature prefers the more general term of interacting Markov chains). Here, the sequence is chosen to be tempered. To allow proper application to trans-D models (Jasra *et al.* 2007), only the likelihood function is raised to the power of the tempering parameter $\left(\pi_i \propto p(\mathbf{x})L(\mathbf{x})^{\beta_i}, \beta_i \in [0, 1]\right)$. Each density is targeted by one Markov chain and the multiple chains are simulated in parallel on a computer cluster. While high-β chains sample locally, low-β chains sample much more widely due to the relaxed likelihood structure. Importantly, chains can interact by exchange updates where randomly selected pairs are allowed to swap temperatures (Geyer 1991). For two chains n and q with parameter vectors $\mathbf{x}_n$ and $\mathbf{x}_q$, respectively, updates are

Downloaded from http://gji.oxfordjournals.org/ by guest on September 2, 2014

accepted/rejected based on the MH criterion for exchange updates given by

$$\alpha = \min\left[1, \frac{\pi_n(\mathbf{x}_q)\pi_q(\mathbf{x}_n)}{\pi_n(\mathbf{x}_n)\pi_q(\mathbf{x}_q)}\right] = \min\left[1, \left(\frac{L(\mathbf{x}_n)}{L(\mathbf{x}_q)}\right)^{\beta_q-\beta_n}\right]. \quad (8)$$

Note that chains n and q can have states of arbitrary dimension during exchange moves since eq. (8) simplifies to a tempered likelihood ratio when only likelihoods are tempered (i.e. prior probabilities cancel). The interacting population of chains is implemented here on a parallel computer cluster with fast interconnect between processors. The algorithm has a simple master-worker structure and each chain is run on an individual processor. Exchange moves are carried out by the master when workers communicate with the master after each step. Since the order of workers communicating with the master is random, proposing exchange moves for consecutively reporting worker pairs results in proper sampling. Note that the reporting order can depend on the particular computer cluster and should be verified by saving the chain number as part of the posterior sample. Note that PT works best when exchange moves are proposed frequently. Since exchange moves carry virtually no computational overhead (do not require likelihood evaluation) these moves should be proposed as often as possible. The only limitation on exchange frequency on a computer cluster can be high-latency/low-bandwidth networks.

The choice of proposal distribution for perturbation steps can significantly impact algorithm efficiency. While arbitrary adaptation of the proposal distribution is not allowed and can cause biased sampling, diminishing adaptation (Brooks *et al.* 2011) is applied to allow automated tuning of the standard deviations of the proposal densities. The adaptation of step sizes is such that MH acceptance rates are targeted to be in the range of 20–30 per cent. Each chain is adapted individually, so that proposal distributions are scaled appropriately for the various tempering levels in the algorithm. Therefore, highly tempered chains propose much larger steps than those with little or no tempering. This is consistent with the relaxation of the tempered likelihood function. The adaptation is automated such that it diminishes over time and does not violate detailed balance. In practice, most adaptation also occurs in the burn-in phase where chains do not contribute to the posterior estimates. Finally, Cauchy distributions are used as proposal distributions for perturbation moves which often perform better for difficult problems than Gaussian proposals due to a higher probability of proposing large steps (heavy tails) while having similar probability of proposing small steps (Dosso & Wilmut 2008).

3 INVERSION RESULTS

This section considers three data sets, two of which are simulations and one for the 2010 February 27 $M_w = 8.8$ Maule (Chile) earthquake. All data sets are based on the configuration of 30 stations obtained for the Maule earthquake as shown in Fig. 2. Vertical-component W-phase data were obtained for distances of 5°–90° with good azimuthal coverage (maximum azimuthal gap of 50°, Fig. 2). Since a large number of data are available, we limit the inversion to vertical components which are typically less noisy than horizontal components. This approach is favourable for the computational expense which increases with the number of data. The data are bandpass filtered to 200–1000 s periods and W-phase time windows are chosen according to Kanamori & Rivera (2008). The sampling rate for all data sets is 4 s (∼6500 data points). The fault-plane geometry is considered fixed with a strike of 18°, a dip of 18°, along-strike extent of 600 km and along-dip extent of 300 km (Benavente & Cummins 2013). The orientation of the two components of the slip vector is identical to that in the simulation study. Hence, the initial rake is 104° and depending on the scale of the two components, rake can vary between 59° and 149° (i.e. slip component one is at 149°, component two at 59°). Data predictions are based on a GF library, computed for W phases by normal-mode summation (Kanamori & Rivera 2008) for a reference earth model (Dziewonski & Anderson 1981).

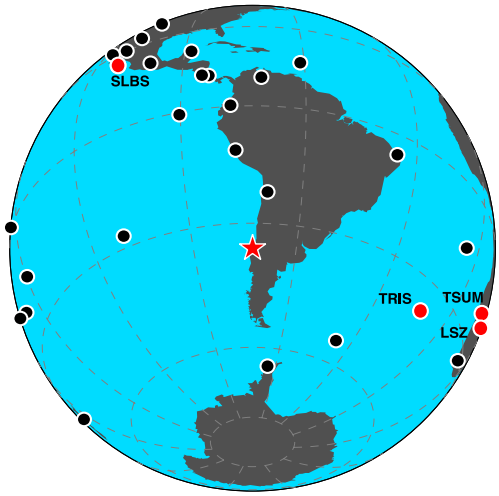

Figure 2. Station configuration (dots) used for all inversions in Section 3. Stations highlighted in red are considered in more detail in Section 3.3 and the hypocentre location is indicated by a star.

Convergence of all inversions is judged by examining the chain history. First, a burn-in period is discarded. The burn-in period includes the adaptive phase of the algorithm, where step sizes are tuned (automated) based on the acceptance rate. Once acceptance rates for all parameters are between 20 and 30 per cent, the algorithm is run as long again and these steps are discarded. For the sampling phase, sections of chain history (the first third and the last third) are compared in terms of marginal densities and chain plots (see Dettmer *et al.* 2013,for a detailed description). Once the practitioner is satisfied by the results, the inversion is terminated.

This section obtains inferences about parameters by marginalizing the high-dimensional PPD (Sivia & Skilling 2006). This results in various marginal densities that integrate over all parameters but those that are inferred. Therefore, the marginal estimates include the integrated effects (covariances) due to the other uncertain parameters and give quantities that can be interpreted in a more straightforward manner than the high-dimensional PPD.

3.1 Importance of model selection and irregular grids

The first data set considers finite-fault inversion of simulated W-phase data for a simple slip distribution to illustrate the importance of quantitative model selection for rupture parameter and uncertainty estimation. W-phase data are simulated for a single, rectangular slip patch located between 330- and 420-km along-strike distance and 30- and 90-km along-dip distance. The slip magnitude in this patch is 28.6 m and is parametrized as two components with 25- and 14-m slip. All other areas have zero slip. Gaussian-distributed noise with a standard deviation of 15 per cent of the mean absolute amplitude at each station was added to the data. Three inversions with different parametrizations are carried out to investigate the

Downloaded from http://gji.oxfordjournals.org/ by guest on September 2, 2014

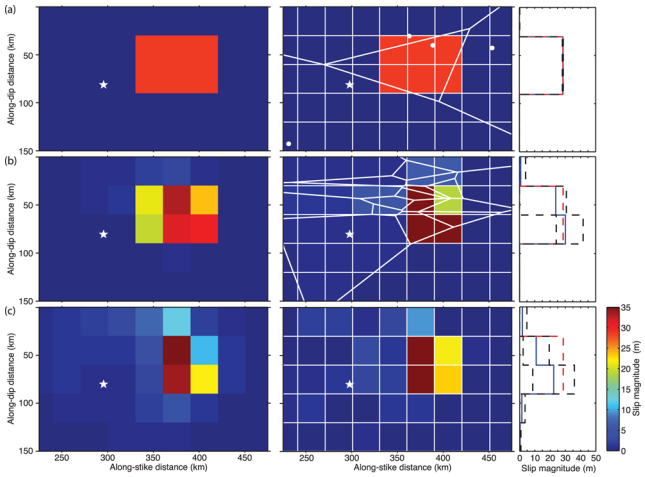

Figure 3. Inversion results for simulated data comparing (a) trans-D results, (b) irregular grid with 20 nodes and (c) regular grid with 120 nodes. Ensemble mean model (left-hand panels), MAP model (middle panels) and uncertainty estimates (right panels) are shown (solid blue: posterior mean, dashed black: 95 per cent CIs, dashed red: true values). Fine white lines indicate the Green's function grid (subfault boundaries) and heavy white lines the position of cell boundaries for irregular grids. For clear visual presentation, Voronoi-node positions are only shown in (a) as white points and omitted in (b). The hypocentre location is given by the white star.

impact of parametrization choice on the results for this problem: (i) A trans-D, irregular grid as outlined in Section 2 is applied. (ii) The number of irregular grid nodes is fixed at 20 (much higher than required for such a simple slip distribution), otherwise the inversion is identical to case (i). (iii) The inversion grid is regular, fixed and identical to the underlying GF grid. Uniform priors were applied for slip with bounds of 0 and 40 m. The rupture velocity and hypocentre location were assumed as known. The three choices have substantially different numbers of parameters. In case (i), the number of parameters is variable but the algorithm predominantly samples models with four grid nodes, hence the dimension of the parameter vector is close to 16 (two slip parameters and two position parameters per node). In case (ii), the dimension is 80 and in case (iii) it is 240 (a subfault size of 30 by 30 km). These differences have a significant impact on the ability to infer information from the data. Convergence for all three inversions was judged based on chain history after burn-in (Section 3). Case (iii) converged after ∼5000 steps. Note that one step in the chain includes proposing perturbations to all parameters and, hence, computational time for one Markov-chain step increases linearly with the number of parameters. Case (ii) converged after ∼15 000 steps, while case (iii) required ∼20 000 steps. Hence, the total number of forward-model evaluations for case (i) is ∼80 000, for (ii) $\sim 1\times 10^6$ and for (iii) $\sim 4\times 10^6$.

Fig. 3 compares the three inversion results in terms of posterior mean and maximum *a posteriori* (MAP) models, and 95 per cent credibility intervals (CI) for the slip as a function of along-dip distance at 405-km along-strike distance. The trans-D inversion has the ability to capture the rupture complexity with only four nodes (16 parameters) which results in the lowest parameter uncertainties (∼1-m CIs). Such low uncertainties appear reasonable considering the simple slip pattern and the large number of data. Note that the detailed shape of the Voronoi cells does not reflect the data prediction, since the cells are only used to copy the appropriate rupture parameters onto the regular GF grid (see Fig. 1). When the number of irregular grid nodes is arbitrarily increased to (but fixed at) 20, posterior estimates are smeared/smoothed and uncertainties substantially increase. Interestingly, node positions cluster around the slip centre and the large areas without slip are still parametrized by few nodes, illustrating how irregular grids can locally adapt to the structure supported by the data. In the case of a fixed grid with spacing of 30 km (240 parameters), the model is so overparametrized for the given data that the inversion loses ability to resolve parameter values and results are even more smeared out/smoothed. In addition, uncertainties are large and in some cases reach over the full prior width.

This example illustrates the importance of quantitative model selection to obtain rigorous posterior estimates. While specific details of the solution may vary with noise level and data type, the important result is that uncertainty estimates depend strongly on parametrization choice. For models of 2-D/3-D structures such as faults, empirically choosing the parametrization/grid spacing can be extremely difficult. In addition, parameter resolution typically varies as a function of space, making spatially regular parametrizations undesirable. The parsimonious self-parametrization developed here is ideally suited to address these issues which is desirable for non-linear Bayesian inversion.

3.2 Simulated data inversion

This section considers W-phase data simulated for a realistic test case based on the Maule earthquake (station configuration and fault geometry are identical) to verify the algorithm. The goal of this study is to validate the algorithm and to study fault-parameter uncertainties in an ideal test case (without theory error). Fig. 4 shows the model parameters used to simulate vertical components of W-phase data. The rake is parametrized with the two component slip vectors which are rotated such that the mean rake is centred between them (i.e. slip component one is at 149°, component two at 59°). The simulated slip distribution has two centres of high slip with midpoints at ∼150 and 475 km along-strike distance, and ∼70 and 40 km along-dip distance, respectively. Changes in rake of up to ±45° around the centre value are included by varying the two slip components independently. Note that the two slip centres are located at different locations from the top of the fault to examine the ability of W-phase data to differentiate between such small differences. Prior densities were chosen to be uniform between 0 and 30-m slip for each component. In addition, the rupture velocity is a function of space across the fault with a uniform prior with bounds of 1.5 and 3.5 km s^{-1}. Large rupture velocities exist near the hypocentre and velocity decreases with increasing distance from the hypocentre. The hypocentre is located on the fault at 300-km along-strike and 80-km along-dip distance. Uncorrelated, Gaussian-distributed noise with standard deviations of 5 per cent of the mean absolute amplitude at each station was added to the data, resulting in a different level of noise for each waveform. We chose to carry out the inversion for data with uncorrelated errors as an ideal test case, where no theory error exists. However, the noise level of 5 per cent of mean amplitude is high compared to seismic noise on typical observed data. For example, estimating the seismic noise standard deviation from 24 hr of data before the Maule event (for 200–1000-s periods), gives much lower noise levels (∼1 per cent). However, such estimates are unrealistically low since theory errors are ignored.

Fig. 5 shows four representative waveforms and the MAP predictions from the inversion. In addition, 95 per cent CIs for the data predictions are shown. The CIs are obtained by predicting data for all posterior samples (i.e. marginalizing the posterior in data space)

Downloaded from http://gji.oxfordjournals.org/ by guest on September 2, 2014

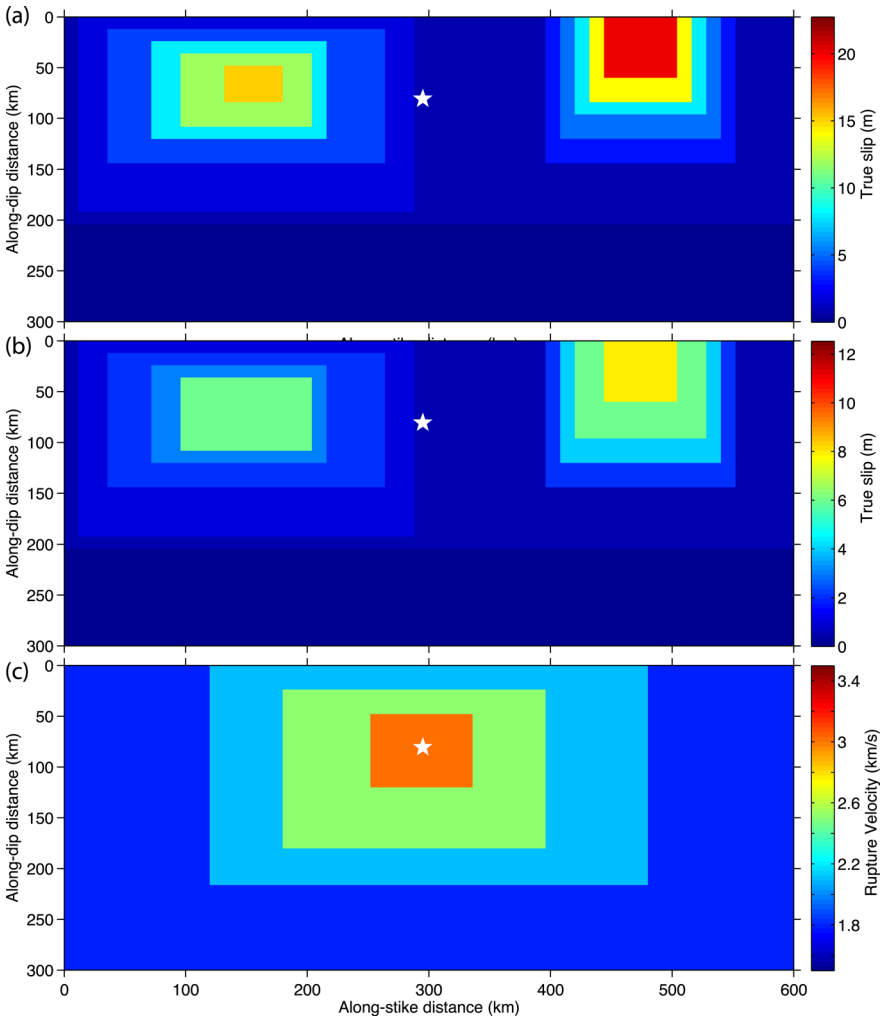

Figure 4. True parameter values and hypocentre location (★) for the finite-fault inversion simulation: Slip component (a) one and (b) two and (c) rupture velocity. Note that spatial changes of rupture velocity and slip do not coincide.

and indicate what range of data predictions the inversion produces. The data are well fit by the inversion and the range of data predictions is consistent with the level of noise on the simulated data. In particular, the implicit sampling of the standard deviation at each station accurately captures the various noise levels without any requirement to specify weights or scale amplitudes. The likelihood values sampled by the inversion are of the order of the likelihood value when using the true model. The trans-D model has the highest probability for group members with 10 nodes but has some probability from 8 to 12 nodes (Fig. 6a). In addition, the true moment magnitude of ∼8.99 is also well estimated by the inversion with low uncertainty (Fig. 6b).

Fig. 7 considers the inversion results in terms of ensemble-mean models for both slip components and for the rupture velocity. The spatial distributions of both slip components are well recovered and the mean model clearly distinguishes between areas of shallow and deeper slip. However, the slip is not precisely recovered due to noise, and since the self-parametrization cannot capture the complexity of the true model precisely in this more complex case. It is important to note that this is a limitation for all inversions where the model is simpler than the true environment. Here, a parsimonious approximation of the true environment is successfully found and can be verified by the fact that the true likelihood value is in the range of sampled likelihood values. The results capture the main features of the true model well, agreeing in magnitude, direction and position of slip. The self-parametrizing inversion is able to adapt locally to the structure supported by the data and capture different spatial scales in different parts of the model without any requirement for regularization. Fig. 7 also shows the posterior mean model for the rupture velocity and the rupture wave fronts for that model. The posterior rupture velocity agrees reasonably well with the true model but some differences exist. This particular example is challenging, since the true spatial distribution of v_{rup} is on a different scale than that for slip. Therefore, some nodes may be required for v_{rup} but not for slip and vice versa. However, the inversion always requires all parameters to be present for each node, a limitation that is non-trivial to overcome.

The Bayesian approach applied here also allows examination of uncertainties for all parameters in the inversion. Fig. 8 shows the slip magnitude and v_{rup} along-dip profiles at selected locations along the strike of the fault (locations are indicated in Fig. 7). Posterior mean profiles, 95 per cent CIs and true values are shown. The uncertainty of slip magnitude is generally of the order of metres but is as high as ∼10 m in some areas of high slip. In the transition zones between low- and high-slip areas, the uncertainty of slip magnitude can be very high due to uncertainty in resolving the location of the transition (e.g. the slice at 480-km along-strike distance, Fig. 8). The uncertainty estimates for v_{rup} indicate that W-phase data contain

Downloaded from http://gji.oxfordjournals.org/ by guest on September 2, 2014

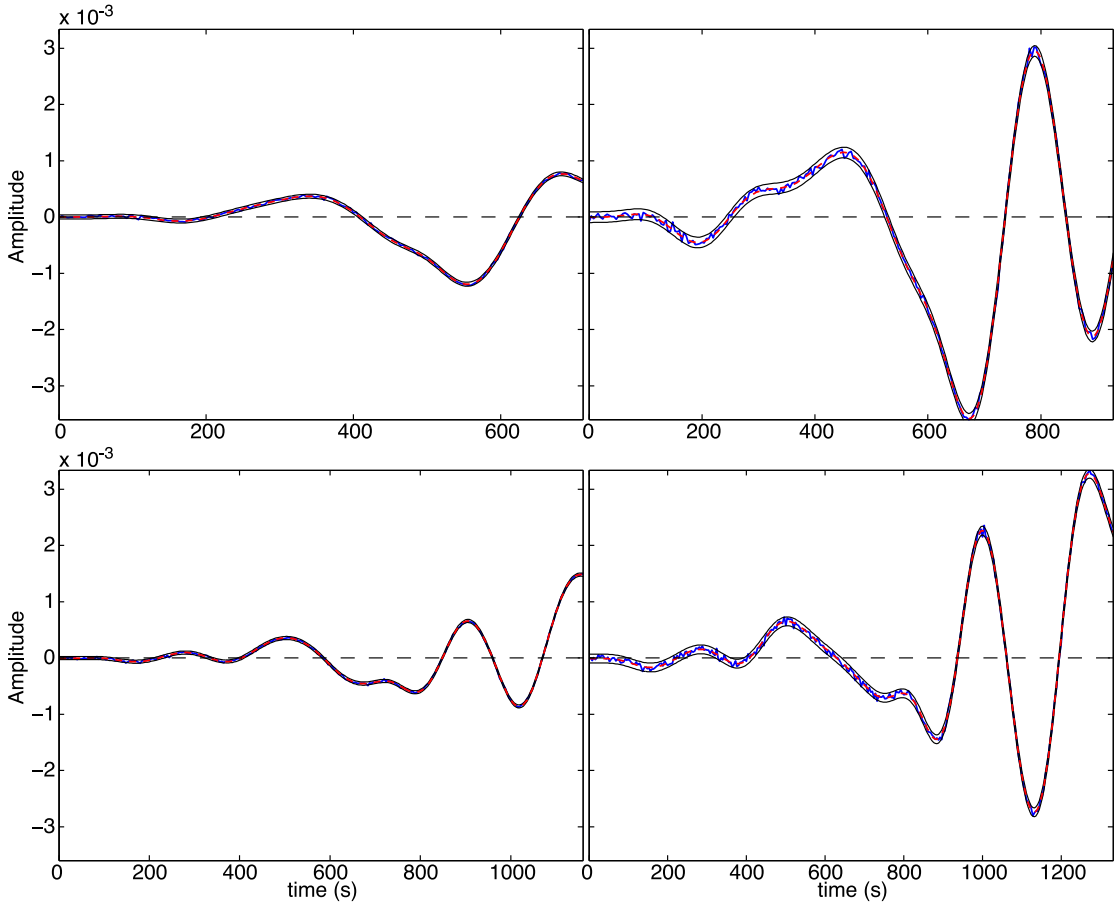

Figure 5. Simulated vertical component W-phase data for four selected stations (blue). The MAP predictions (red) and 95 per cent CIs (black) of the predictions are also shown.

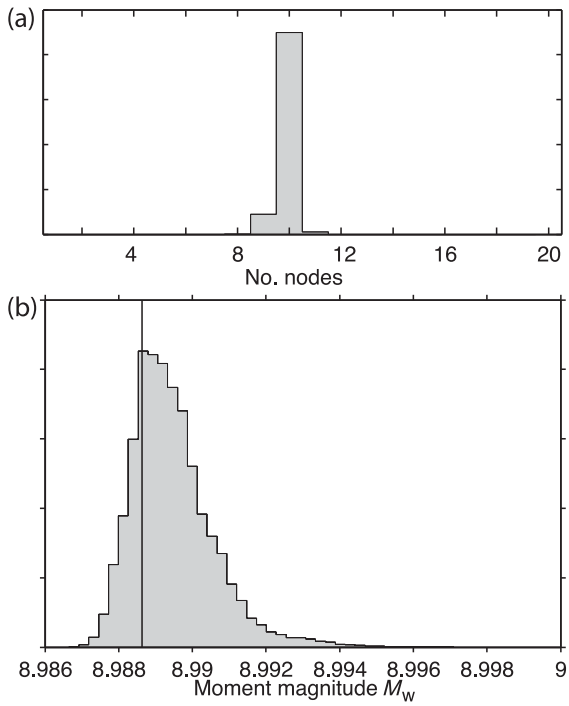

Figure 6. Simulation results for (a) model complexity of inferred trans-D model and (b) moment magnitude. The solid vertical line in (b) indicates M_w for the true model.

little information about the spatial distribution of v_{rup}. However, while uncertainty is large, some sensitivity appears to exist to rupture velocity. To examine v_{rup} resolution further, the inversion was also carried out by parametrizing v_{rup} with a single unknown parameter that quantifies v_{rup} in space (referred to as inversion two). Hence, while the value of v_{rup} can be uncertain and is inferred as part of inversion two, v_{rup} does not change spatially. The inversion results were compared using the BIC (Schwarz 1978), which identified the more complex model (inversion one) as preferable. In addition, convergence was much more difficult in inversion two and results showed strong divergence from the true model. Therefore, we conclude that W-phase data can be sensitive to significant variability in v_{rup} and accounting for such variability in the model should be examined by model selection for observed data. However, inferences on v_{rup} variability are associated with large uncertainties.

3.3 Application to the Maule ($M_w = 8.8$, Chile) earthquake

This section applies the inversion to the 2010 February 27 $M_w = 8.8$ Maule (Chile) earthquake. Prior bounds for both slip components are uniform between 0- and 70-m slip and for rupture velocity uniform between 2.0 and 3.0 km s^{-1}. Two complete inversions were carried out for these data to investigate the optimal parametrization for the rupture velocity. The first inversion assumes a simpler model where rupture velocity is not a function of space but given by a single unknown parameter. The second inversion assumed a more complex

Downloaded from http://gji.oxfordjournals.org/ by guest on September 2, 2014

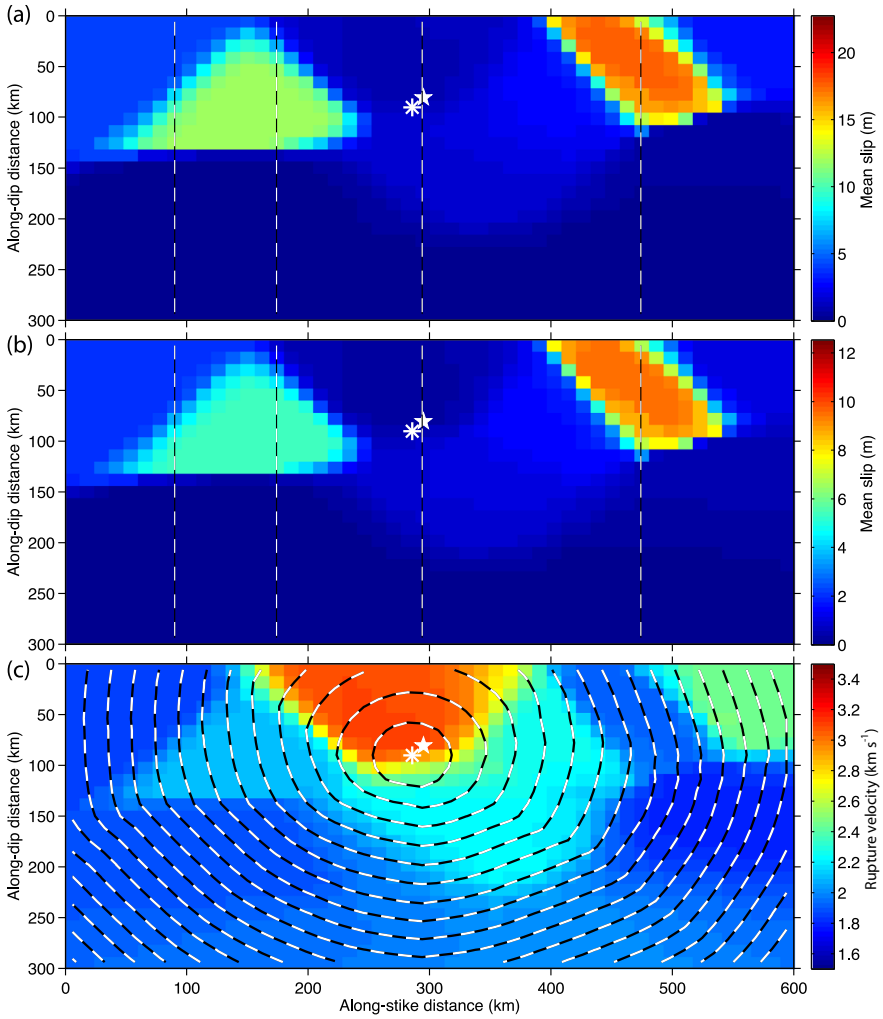

Figure 7. Inversion results in terms of the posterior mean model: Slip component (a) one and (b) two, and (c) rupture velocity. Dashed vertical lines indicate the location of slices for examining uncertainty estimates (see Fig. 8) and dashed lines in (c) are inferred rupture wave fronts for the posterior mean model. True hypocentre location (⋆) and MAP hypocentre location (∗) are also shown.

model where v_{rup} is a function of space and given by an additional parameter for each grid node. The results for both inversions are compared with the BIC and BIC values for the simpler model are 15 units lower than those for the more complex model. Therefore, the simpler model is chosen as optimal and the following detailed discussion of results only considers results for a single unknown for v_{rup}.

Fig. 9 shows W-phase waveforms for four arbitrarily chosen stations (IU.TRIS, IU.SLBS, IU.TSUM and IU.LSZ). In addition, MAP data predictions and prediction CIs are shown. The predictions appear to fit the observed data well. However, while the observed data lie mostly within the range of data predictions, close inspection shows that areas of systematic misfit appear at all stations, suggesting strongly correlated residual errors. It is crucial to address such correlations in the covariance matrix of the likelihood function to avoid biases in parameter estimates and uncertainties. Following the approach in Section 9, the matrices for each station are estimated here from residuals obtained from an initial inversion that assumes uncorrelated errors. The matrices are then periodically updated during the burn-in phase of the following inversion. The final estimates are shown for four representative stations in Fig. 10. Note that the residuals are strongly non-stationary with smaller standard deviations at early times (low waveform amplitudes) and high standard deviations at later times (high waveform amplitudes). This suggests that misfit increases with W-phase amplitude (i.e. errors scale with the magnitude of the data). This effect is likely due to theory errors stemming from limitations in the GF library (Duputel *et al.* 2014).

The trans-D inversion captures the complexity of this rupture with 12 nodes and uncertainties range from 11 to 14 nodes (Fig. 11). Inversion results in terms of the posterior-mean model for total slip (Fig. 12a) suggest that three spatially distinct areas of significant slip exist. Two of these areas show peak-slip magnitudes of >30 m and are located at the shallowest parts of the fault plane. While the southern slip concentration (at ∼190 km along strike) shows the highest mean slip of 58 m, the uncertainty associated with this region is also high (∼30 m CI, Figs 12b and 13). In areas of such high uncertainty, the posterior mean model does not provide meaningful estimates of slip and the credibility intervals in Fig. 13 should be considered instead. Fig. 13 shows that uncertainties in the southern slip concentration are so high that meaningful peak-slip estimation is not possible in that area. The northern slip centre shows much lower uncertainty with slip of ∼35 ± 7.5 m. The slip pattern is consistent with other published results, in particular the geodetic inversion by Vigny *et al.* (2011) who observed significant slip south of the hypocentre. However, peak slip is much higher than previously

Downloaded from http://gji.oxfordjournals.org/ by guest on September 2, 2014

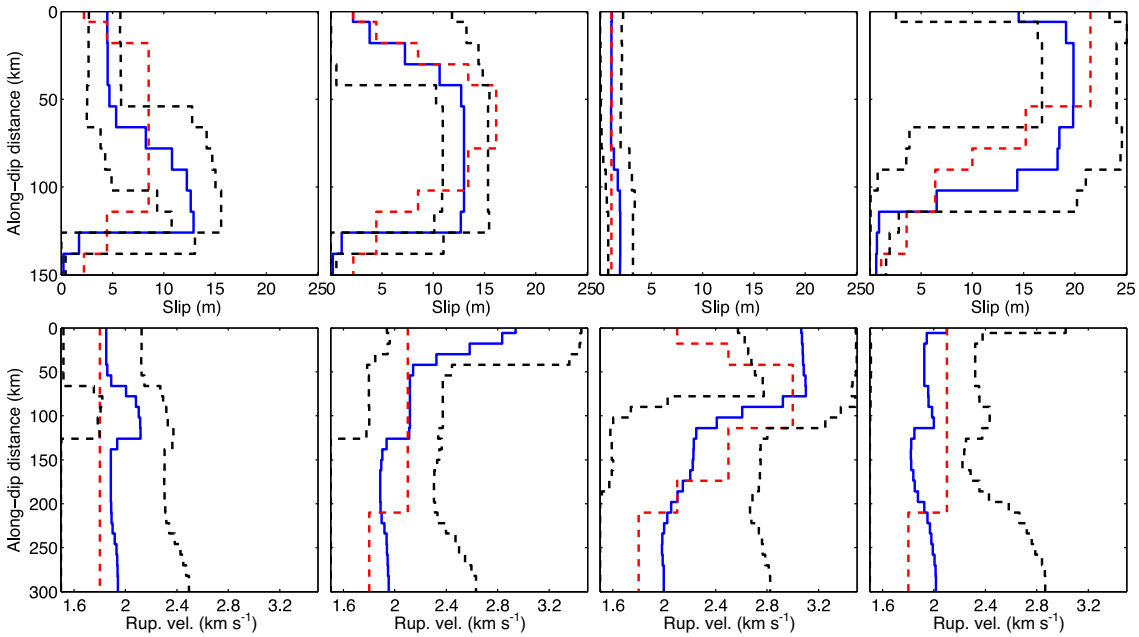

Figure 8. Posterior mean profiles (blue), 95 per cent CIs (dashed) and true model (red), at arbitrarily selected along-strike distances for slip magnitude (top) and rupture velocity (bottom). The top row is cropped to 150-km along dip distance to focus on the area where most slip occurs.

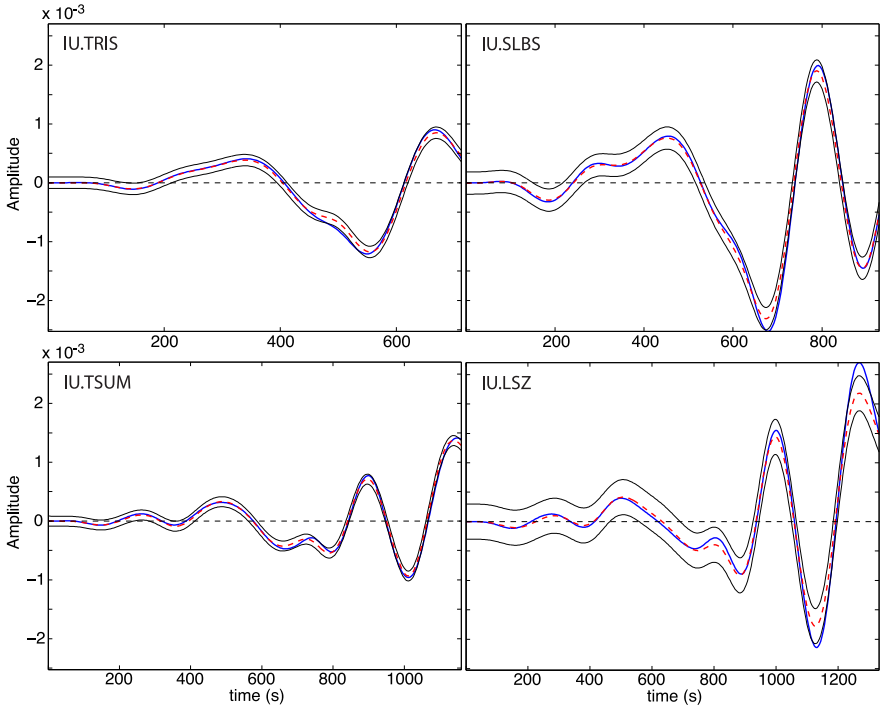

Figure 9. W-phase waveforms for select stations as indicated in Fig. 2. Observed (blue), MAP data predictions (red) and 95 per cent CIs (black) for three stations for the Maule earthquake inversion.

reported (e.g. Koper *et al.* 2012; Benavente & Cummins 2013). This work is significantly different from all previous work on slip estimation for this event since no smoothing or regularization is applied. In addition, this is the most rigorous model selection applied to slip inversion to date and the simulation in Section 3.2 has clearly shown that W-phase data can resolve peak slip within reasonable uncertainties. Therefore, it is possible that previous work underestimated the peak slip of this event due to restrictive assumptions made in the inversion. In particular, it is fundamentally not possible to quantify the effects of global regularization on the inferred parameters. In addition, previous work has ignored correlated residual errors for various data types. It is important to note that correlated errors can substantially impact linear and non-linear inversion results (Dettmer *et al.* 2007). However, the W-phase data used here show such strongly correlated residuals that some concern remains about the reliability of the estimation process for the data covariance matrices. Recent work on physical modelling of some aspects of theory error (Yagi & Fukahata 2011; Duputel *et al.* 2014) shows

Downloaded from http://gji.oxfordjournals.org/ by guest on September 2, 2014

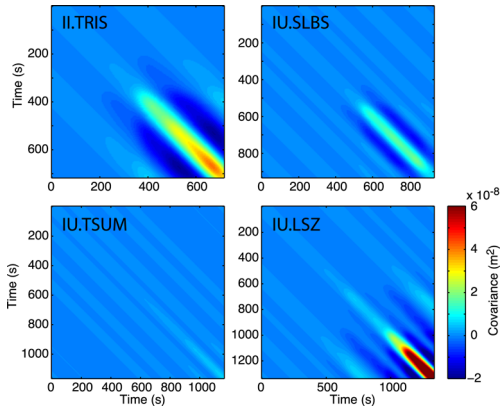

Figure 10. Examples of non-stationary data covariance matrix estimates for the same stations as shown in Fig. 9. Note that the range of time lags varies for the various matrices.

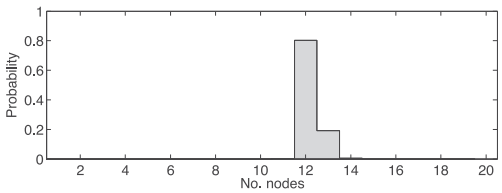

Figure 11. Posterior distribution of number of nodes in rupture model (indicating model complexity).

promise but intrinsically cannot address all theory error. A combination of the approach presented here and physical modelling of theory error may address the problem better but poses a significant research task. Joint inversion may also provide further insights since additional data types may help constrain slip magnitudes better.

In the northern updip slip centre, the rake (Fig. 12c) is estimated at 140° with a 95 per cent CI from 130° to 149°. These values are consistent with previously reported results (Benavente & Cummins 2013) and are also consistent with the tsunamigenic character of the earthquake. The uncertainty in rake is much higher in the southern slip centre where the 95 per cent CI is from 110° to 149°. A third region of smaller slip (∼25 ± 7.5 m) exists at 480-km along-strike distance and 100-km along-dip distance. The posterior rake estimate in this region is ∼85° but there is no conclusive explanation for the difference in rake compared to the patch at shallower depth. Since the slip is significantly smaller than elsewhere in the model, the data features constraining this patch are likely small and are more likely to be affected by limitations in the noise estimation process. Fig. 12(d) shows the node density on the fault which is an indication for where the data support structure in the slip distribution. Importantly, no significant structure is supported at along-dip distances larger than 150 km.

Fig. 14(a) shows marginal densities for the hypocentre location. The density is highly multimodal which is likely due to the influence of theory errors in the inversion. While the multimodal density increased sampling difficulty, it is important to account for the uncertainty in hypocentre location in the parameter estimates of primary interest (slip). Fig. 14(b) shows the marginal estimate for rupture velocity with a MAP value of 2.15 km s^{-1}. The MAP moment magnitude is $M_w \approx 8.9$ with a 95 per cent CI of 0.02.

Finally, residual errors are examined to judge how appropriate the assumptions in the likelihood function are. Raw and standardized residuals are considered. The raw residuals are obtained for the MAP model from the initial inversion that ignored correlated errors (only scaled by a single standard deviation that is implicitly sampled in the inversion according to eq. 4). Standardized residuals are obtained for the MAP model of the final inversion that included the hierarchically scaled covariance estimate. These residuals are standardized by multiplication with the Cholesky decomposition of the covariance matrix (Dettmer *et al.* 2007). If assumptions are well met, the standardized residuals should be Gaussian distributed with zero mean and unit standard deviation. We examine the randomness of the residuals by considering the autocorrelation function which should have a sharp central peak for truly random residuals. Fig. 15(a) shows the autocorrelation for four stations for raw and standardized residuals. The standardized residuals are substantially narrower in the main peak but still show correlation. Hence, the covariance estimation carried out here improved the inversion. To examine Gaussianity, residual histograms are considered. Fig. 15(b) shows histograms for raw and standardized residuals, and a Gaussian density as reference. The standardized residuals appear to be much closer to the Gaussian reference than the raw residuals, which raised confidence in the results. Note that the results in Fig. 15

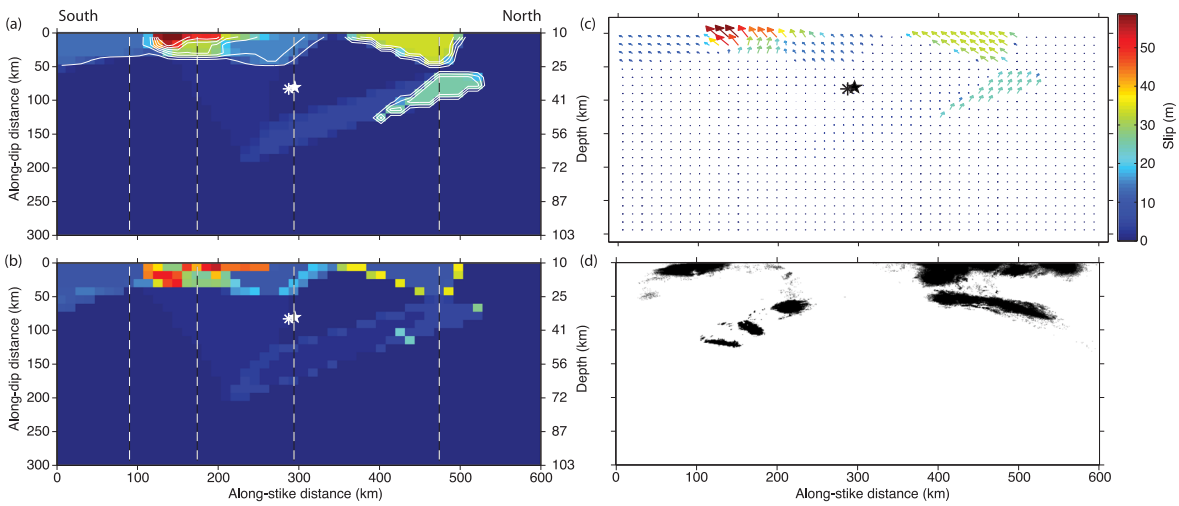

Figure 12. Inversion results for the Maule event: (a) Posterior mean slip magnitude and 7.5-m slip contours (white), (b) 95 per cent CI width for slip magnitude on the fault, (c) rake for posterior mean model and (d) Voronoi-node position probability (darker values correspond to higher density). Point-source hypocentre location (★) and MAP hypocentre location (∗) are also shown. Vertical dashed lines indicate locations of marginal profiles in Fig. 13.

Downloaded from http://gji.oxfordjournals.org/ by guest on September 2, 2014

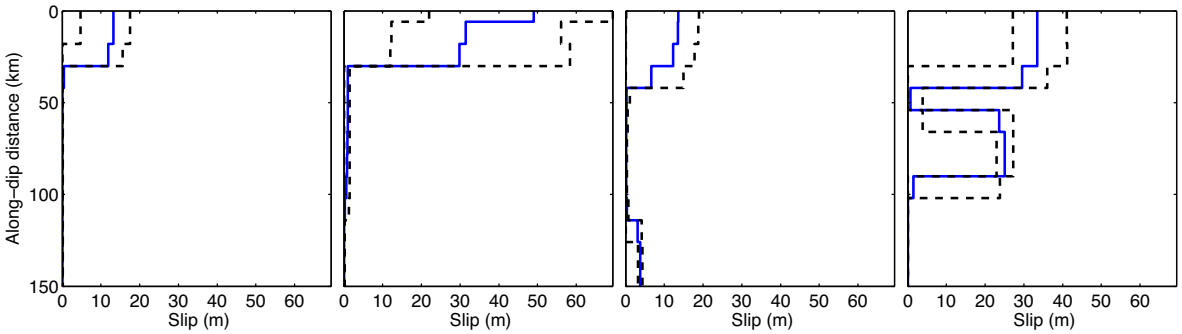

Figure 13. Marginal profiles of slip magnitude for select along-strike distances. Posterior mean (solid blue) and 95 per cent CIs (dashed black) are shown. Note the large uncertainty of ∼50 m associated with the southern slip centre. Along-dip distance is cropped to 150 km to focus on the area where most slip occurs.

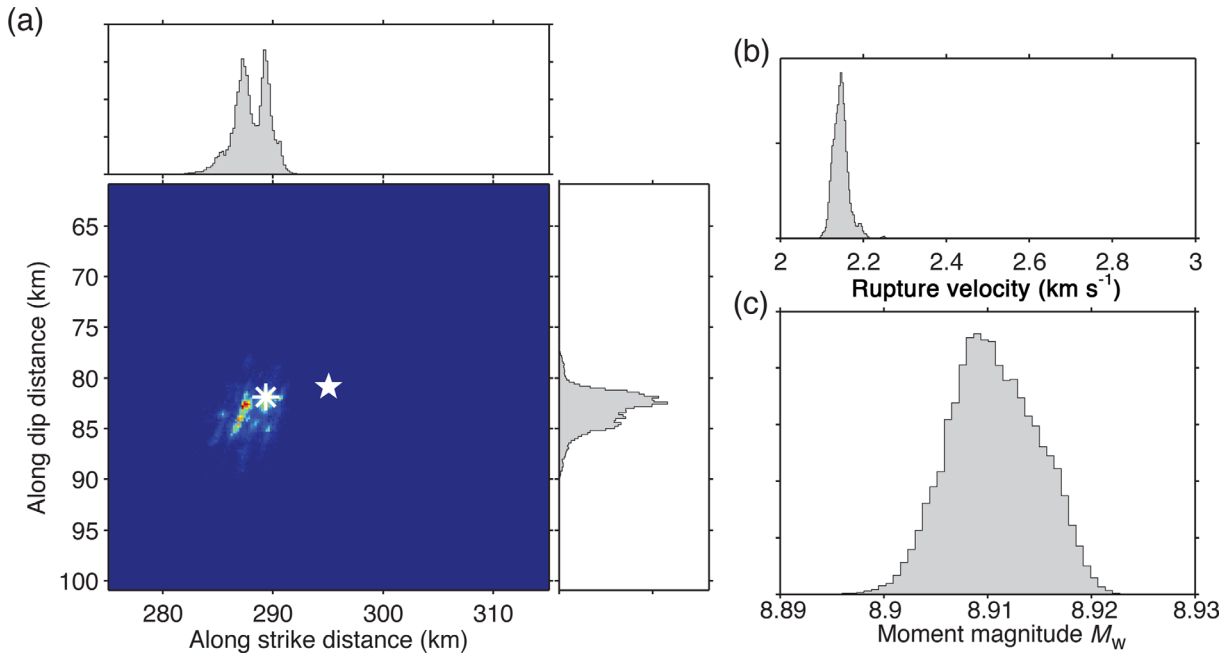

Figure 14. Marginal densities for (a) hypocentre location (point-source hypocentre: ⋆, MAP hypocentre: ∗), (b) rupture velocity and (c) moment magnitude.

are representative for all 30 stations. In summary, the advanced covariance estimation applied here significantly improves residuals but does not address all issues. More research is required to better understand the source of theory errors.

4 SUMMARY AND DISCUSSION

This paper developed a probabilistic finite-fault inversion based on a trans-D self-parametrization of the fault. The fault is parametrized using self-adapting irregular grids which intrinsically match the local resolving power of the data and provide parsimonious solutions requiring few parameters to infer complex rupture characteristics. The self-adapting parametrization is implemented by a trans-D irregular grid based on an unknown number of Voronoi nodes. The model structure is controlled by Bayesian parsimony, and hence does not require subjective choices of regularization.

The time dependence of the rupture was parametrized by rupture-velocity parameters and a single triangular source-time function with fixed rise time of 24 s. Fixing the rise time is justified by the long periods of W-phase data which are not sensitive to changes in rise time (Benavente & Cummins 2013). Rupture causality is ensured by solving the eikonal equation on the regular GF grid. For the parametrization of v_{rup}, we selected two possible models: (1) a rupture-velocity parameter being present for all Voronoi nodes, and (2) a spatially independent v_{rup}. However, the optimal parametrization is likely different but non-trivial to estimate. The data information on the spatial variability of various types of parameters is likely different for each type. Model one cannot treat spatial variability in v_{rup} as independent of that in slip distribution. Rather, the spatial dependence is parametrized with a single set of nodes with three slip-related and two position parameters. If data information about spatial variability in v_{rup} is substantially different from that about the slip components, the resulting trans-D model may be overparametrized. An alternative approach would be to decouple the types of parameters and allow a separate trans-D model for v_{rup}. However, this leads to similar problems where the overparametrization is due to additional node-position parameters.

The inversion was applied to simulated W-phase data and W-phase observations from the 2010 Maule (Chile) earthquake. The simulated-data results showed the ability of W-phase data to resolve slip distributions and spatially dependent rupture velocities of slip models. The complexity of the true model was chosen such that the spatial variability of rupture velocity does not coincide with the

Downloaded from http://gji.oxfordjournals.org/ by guest on September 2, 2014

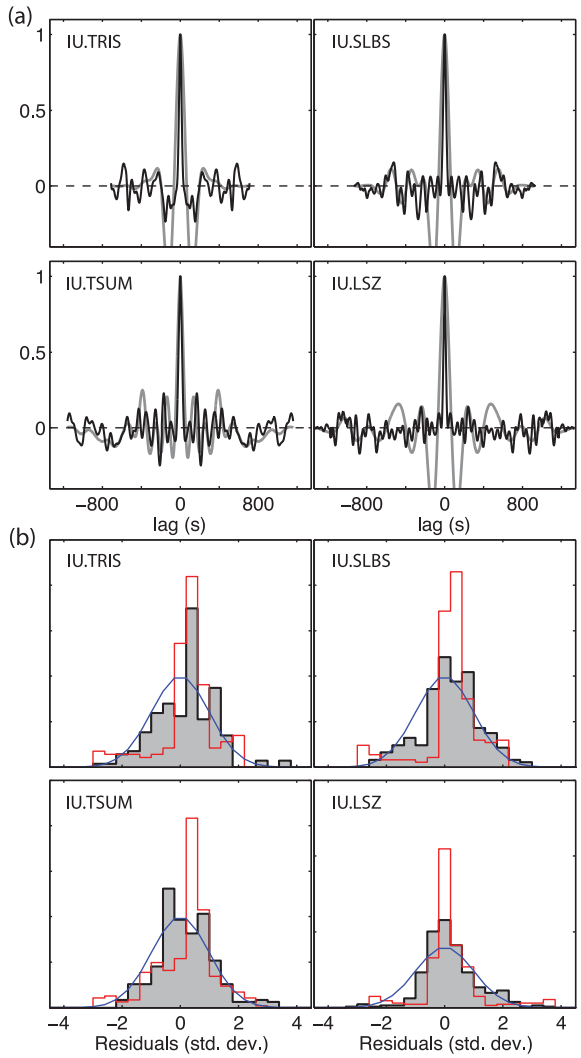

Figure 15. Residual analysis in terms of (a) residual autocorrelation plots (raw residuals: grey, standardized residuals: black), and (b) residual histograms (raw residuals: red, standardized residuals: black and standard normal distribution: blue) for four stations.

spatial variability of the slip. Despite this challenge, the complexity of the true model was captured with just a few Voronoi nodes (∼10). The estimated moment magnitude agreed closely with the true moment magnitude of the simulation and slip uncertainties are mostly of the order of metres. However, some areas of large uncertainties on slip exists where slip transitions from low to high values.

Application to the Maule earthquake estimates a moment magnitude of ∼8.9 with 95 per cent CIs of 0.02. For the Maule event, the data do not support rupture velocity varying as a function of space. A slip pattern of three major slip centres was clearly resolved by the data, showing two shallow areas of large slip to the south and to the north of the hypocentre. The southern centre has large uncertainty associated with the slip (almost as large as the magnitude of the slip), while the northern slip centre is well determined with uncertainties of only ±7.5 m. While the results of the spatial variability of slip are consistent with other published results (Vigny *et al.* 2011), peak slip is significantly higher (∼35 m in the well-determined northern slip centre) than previously reported. Benavente & Cummins (2013) recently reported finite-fault results from a linear regularized inversion of W-phase data. While uncertainty estimation was not considered, general similarity exists with the slip pattern presented here. A significant difference is our ability to identify the southern slip centre as poorly constrained. In addition, the northern area resolves two distinct regions of slip due to the ability to adapt the parametrization to the local resolving power of the data. The global regularization in Benavente & Cummins (2013) precludes such resolution.

The residual errors for W-phase data appear to be dominated by theory error (errors due to model limitations) which results in highly correlated errors that we account for by a combination of empirical and hierarchical estimation: The non-stationary covariance matrix at each station is iteratively estimated from the residuals and applied in the inversion with a hierarchical scaling factor. While this approach is an advanced treatment of residual errors, not all correlation is removed and further work is required to better understand W-phase residual errors. While some recent contributions in the literature show promise (Yagi & Fukahata 2011; Duputel *et al.* 2014), a comprehensive solution to this problem does not currently exist.

While computationally intensive, the methods developed here can in principle be applied to any combination of earthquake observations, including seismic, geodetic and tsunami data. The hierarchical-empirical approach to determining noise statistics provides natural scaling of the contribution each data set makes to the solution, avoiding the need to apply subjective weighting factors. Better understanding of earthquake sources is key to improving tsunami-hazard assessments which typically base predictions on overly simplistic characterizations of earthquake rupture and poor understanding of uncertainty. Quantitative study of parameter uncertainty in rupture models is a crucial component for better understanding of rupture and better hazard assessments. The efficient parametrization and rigorous error treatment developed here are among the first such steps towards better inferences on rupture.

ACKNOWLEDGEMENTS

We thank Luis Rivera, Hiroo Kanamori and Zacharie Duputel for providing the GFs library we used for predicting W-phase data and Stan E. Dosso for valuable comments and discussion on the material in this manuscript. Inversions were performed on the Terrawulf III computational facility and on the Raijin system of the Australian National Computational Infrastructure. Terrawulf III is supported through the AuScope Australian Geophysical Observing System. AusScope is funded under the National Collaborative Research Infrastructure Strategy, and the Education Investment Fund (EIF3), both Australian Commonwealth Government Programmes. This research was funded by Australian Research Grant DP110101983. RB is supported by a scholarship from CONICYT, Chile. Finally, we would like to thank two anonymous reviewers and editor Eiichi Fukuyama for thorough reviews that substantially improved the quality of the manuscript.

REFERENCES

Aster, R.C., Borchers, B. & Thurber, C.H., 2005. *Parameter Estimation and Inverse Problems,* pp. 195–402, International Geophysics, Elsevier.

Benavente, R. & Cummins, P.R., 2013. Simple and reliable finite fault solutions for large earthquakes using the W-phase: the Maule (Mw = 8.8) and Tohoku (Mw = 9.0) earthquakes, *Geophys. Res. Lett.,* **40,** 3591–3595.

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.

Downloaded from http://gji.oxfordjournals.org/ by guest on September 2, 2014

Bodin, T. & Sambridge, M., 2009. Seismic tomography with the reversible jump algorithm, *Geophys. J. Int.,* **178,** 1411–1436.

Bodin, T., Sambridge, M. & Gallagher, K., 2009. A self-parametrizing partition model approach to tomographic inverse problems, *Inverse Problems*, **25,** 055009, doi:10.1088/0266-5611/25/5/055009.

Bodin, T., Sambridge, M., Tkalčić, H., Arroucau, P., Gallagher, K. & Rawlinson, N., 2012. Transdimensional inversion of receiver functions and surface wave dispersion, *J. geophys. Res.,* **117,** B02301, doi:10.1029/2011JB008560.

Brooks, S., Gelman, A., Jones, G. & Meng, X., Eds., 2011. *Handbook of Markov Chain Monte Carlo,* pp. 1–592, Springer.

Dettmer, J. & Dosso, S.E., 2012. Trans-dimensional matched-field geoacoustic inversion with hierarchical error models and interacting Markov chains, *J. acoust. Soc. Am.,* **132,** 2239–2250.

Dettmer, J. & Dosso, S.E., 2013. Probabilistic two-dimensional water-column and seabed inversion with self-adapting parameterizations, *J. acoust. Soc. Am.,* **133,** 2612–2623.

Dettmer, J., Dosso, S.E. & Holland, C.W., 2007. Uncertainty estimation in seismo-acoustic reflection travel-time inversion, *J. acoust. Soc. Am.,* **122,** 161–176.

Dettmer, J., Dosso, S.E. & Holland, C.W., 2008. Joint time/frequency-domain inversion of reflection data for seabed geoacoustic profiles, *J. acoust. Soc. Am.,* **123,** 1306–1317.

Dettmer, J., Dosso, S.E. & Holland, C.W., 2009. Model selection and Bayesian inference for high resolution seabed reflection inversion, *J. acoust. Soc. Am.,* **125,** 706–716.

Dettmer, J., Dosso, S.E. & Holland, C.W., 2010a. Trans-dimensional geoacoustic inversion, *J. acoust. Soc. Am.,* **128,** 3393–3405.

Dettmer, J., Dosso, S.E. & Osler, J.C., 2010b. Bayesian evidence computation for model selection in non-linear geoacoustic inference problems, *J. acoust. Soc. Am.,* **128,** 3406–3415.

Dettmer, J., Molnar, S., Steininger, G.A.M.W., Dosso, S.E. & Cassidy, J.F., 2012. Trans-dimensional inversion of microtremor array dispersion data with hierarchical autoregressive error models, *Geophys. J. Int.,* **188,** 719–734.

Dettmer, J., Holland, C.W. & Dosso, S.E., 2013. Trans-dimensional uncertainty estimation for dispersive seabed sediments, *Geophysics,* **78,** WB63–WB76.

Dosso, S.E., 2002. Quantifying uncertainty in geoacoustic inversion. I. A fast Gibbs sampler approach, *J. acoust. Soc. Am.,* **111,** 129–142.

Dosso, S.E. & Wilmut, M.J., 2006. Data uncertainty estimation in matched-field geoacoustic inversion, *IEEE J. Ocean. Eng.,* **31,** 470–479.

Dosso, S.E. & Wilmut, M.J., 2008. Uncertainty estimation in simulataneous Bayesian tracking and environmental inversion, *J. acoust. Soc. Am.,* **124,** 82–97.

Duputel, Z., Agram, P.S., Simons, M., Minson, S.E. & Beck, J.L., 2014. Accounting for prediction uncertainty when inferring subsurface fault slip, *Geophys. J. Int.,* **197,** 464–482.

Dziewonski, A.M. & Anderson, D.L., 1981. Preliminary reference Earth model, *Phys. Earth planet. Inter.,* **25,** 297–356.

Fukuda, J. & Johnson, K.M., 2008. A fully Bayesian inversion for spatial distribution of fault slip with objective smoothing, *Bull. seism. Soc. Am.,* **98,** 1128–1146.

Fukuda, J. & Johnson, K.M., 2010. Mixed linear-non-linear inversion of crustal deformation data: Bayesian inference of model, weighting and regularization parameters, *Geophys. J. Int.,* **181,** 1441–1458.

Geyer, C. & Moller, J., 1994. Simulation procedures and likelihood inference for spatial point processes, *Scand. J. Stat.,* **21,** 359–373.

Geyer, C.J., 1991. Markov chain Monte Carlo maximum likelihood, in *Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface,* pp. 156–163.

Godsill, S. & Clapp, T., 2001. Improvement strategies for Monte Carlo particle filters, in *Sequential Monte Carlo in Practice,* pp. 139–158, Springer.

Green, P.J., 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination, *Biometrika,* **82,** 711–732.

Green, P.J., 2003. Trans-dimensional Markov chain Monte Carlo, in *Highly Structured Stochastic Systems,* pp. 179–198, Oxford Statistical Science Series, Oxford University Press.

Guo, R., Dosso, S., Liu, J., Dettmer, J. & Tong, X., 2011. Nonlinearity in Bayesian 1-D magnetotelluric inversion, *Geophys. J. Int.,* **185,** 663–675.

Hartzell, S.H., 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,** 7515–7534.

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

Holland, C.W., Dettmer, J. & Dosso, S.E., 2005. Remote sensing of sediment density and velocity gradients in the transition layer, *J. acoust. Soc. Am.,* **118,** 163–177.

Ide, K., 2007. Slip inversion, in *Treatise on Geophysics,* Vol. 4, pp. 193–223, Elsevier.

Jarzynski, C., 1997. Nonequilibrium equality for free energy differences, *Phys. Rev. Lett.,* **78,** 2690–2693.

Jasra, A., Stephens, D.A. & Holmes, C.C., 2007. Population-based reversible jump Markov chain Monte Carlo, *Biometrika,* **94,** 787–807.

Kanamori, H., 1993. W phase, *Geophys. Res. Lett.,* **20,** 1691–1694.

Kanamori, H. & Rivera, L., 2008. Source inversion of W phase: speeding up seismic tsunami warning, *Geophys. J. Int.,* **175,** 222–238.

Koper, K.D., Hutko, A.R., Lay, T. & Sufri, O., 2012. Imaging shortperiod seismic radiation from the 27 February 2010 Chile (Mw 8.8) earthquake by back-projection of P, PP, and PKIKP waves, *J. geophys. Res.,* **117,** B02308, doi:10.1029/2011JB008576.

MacKay, D.J.C., 2003. *Information Theory, Inference, and Learning Algorithms,* pp. 343–386, Cambridge Univ. Press.

Malinverno, A., 2002. Parsimonious Bayesian Markov chain Monte Carlo inversion in a nonlinear geophysical problem, *Geophys. J. Int.,* **151,** 675–688.

Malinverno, A. & Briggs, V.A., 2004. Expanded uncertainty quantification in inverse problems: hierarchichal Bayes and empirical Bayes, *Geophysics,* **69,** 1005–1016.

Matsu'ura, M., Noda, A. & Fukahata, Y., 2007. Geodetic data inversion based on Bayesian formulation with direct and indirect prior information, *Geophys. J. Int.,* **171,** 1342–1351.

Menke, W., 2012. *Geophysical Data Analysis: Discrete Inverse Theory,* Academia Press.

Minsley, B.J., 2011. A trans-dimensional Bayesian Markov chain Monte Carlo algorithm for model assessment using frequency-domain electromagnetic data, *Geophys. J. Int.,* **187,** 252–272.

Minson, S.E., Simons, M. & Beck, J.L., 2013. Bayesian inversion for finite fault earthquake source models I—theory and algorithm, *Geophys. J. Int.,* **194,** 1701–1726.

Minson, S.E. *et al.*, 2014. Bayesian inversion for finite fault earthquake source models II: the 2011 great Tohoku-Oki, Japan earthquake, *Geophys. J. Int.,* **198,** 922–940.

Neal, R.M., 2001. Annealed importance sampling, *Stat. Comp.,* **11,** 125–139.

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

Podvin, P. & Lecomte, I., 1991. Finite difference computation of traveltimes in very contrasted velocity models: a massively parallel approach and its associated tools, *Geophys. J. Int.,* **105,** 271–284.

Ray, A. & Key, K., 2012. Bayesian inversion of marine CSEM data with a trans-dimensional self parametrizing algorithm, *Geophys. J. Int.,* **191,** 1135–1151.

Sambridge, M., 2014. A parallel tempering algorithm for probabilistic sampling and multimodal optimization, *Geophys. J. Int.,* **196,** 357–374.

Sambridge, M., Brain, J. & McQueen, H., 1995. Geophysical parametrization and interpolation of irregular data using natural neighbours, *Geophys. J. Int.,* **122,** 837–857.

Downloaded from http://gji.oxfordjournals.org/ by guest on September 2, 2014

Sambridge, M., Gallagher, K., Jackson, A. & Rickwood, P., 2006. Trans-dimensional inverse problems, model comparison and the evidence, *Geophys. J. Int.,* **167,** 528–542.

Schwarz, G., 1978. Estimating the dimension of a model, *Ann. Stat.,* **6,** 461–464.

Simons, M. *et al.*, 2011. The 2011 magnitude 9.0 Tohoku-Oki earthquake: mosaicking the megathrust from seconds to centuries, *Science,* **332,** 1421–1425.

Sivia, D.S. & Skilling, J., 2006. *Data Analysis: A Bayesian Tutorial,* Oxford Science Publications.

Steininger, G.A.M.W., Holland, C.W., Dosso, S.E. & Dettmer, J., 2013. Seabed roughness parameters from joint backscatter and reflection inversion at the Malta Plateau, *J. acoust. Soc. Am.,* **134,** 1833–1842.

Tarantola, A., 2005. *Inverse Problem Theory and Methods for Model Parameter Estimation,* pp. 1–57, SIAM.

Vigny, C. *et al.*, 2011. The 2010 Mw 8.8 Maule megathrust earthquake of Central Chile, monitored by GPS, *Science,* **332,** 1417–1421.

Voronoi, M.G., 1908. Nouvelles applications des parametres continus a la theorie des formes quadratiques, *J. rein. Angew. Math.,* **134,** 198–287.

Yabuki, T. & Matsu'ura, M., 1992. Geodetic data inversion using a Bayesian information criterion for spatial distribution of fault slip, *Geophys. J. Int.,* **109,** 363–375.

Yagi, Y. & Fukahata, Y., 2011. Introduction of uncertainty of Green's function into waveform inversion for seismic source processes, *Geophys. J. Int.,* **186,** 711–720.

Downloaded from http://gji.oxfordjournals.org/ by guest on September 2, 2014