Geophysical Journal International

Geophys. J. Int. (2019) **221,** 334–350
Advance Access publication 2019 December 12
GJI General Geophysical Methods

doi: 10.1093/gji/ggz559

Outlier-insensitive Bayesian inference for linear inverse problems (OutIBI) with applications to space geodetic data

Yu Hang,[1] Sylvain Barbot,[2] Justin Dauwels,[1] Teng Wang,[3] Priyamvada Nanjundiah[4] and Qiang Qiu[2]

[1] *School of Electrical & Electronic Engineering, Nanyang Technological University,* 50 *Nanyang Ave, Block N3.1,* 639789, *Singapore*
[2] *University of Southern California,* 3651 *Trousdale Pkwy, Los Angeles, CA* 90089, *Los Angeles, CA, USA. E-mail: sbarbot@usc.edu*
[3] *Peking University, Beijing, Beijing, China*
[4] *Asian School of the Environment, Nanyang Technological University,* 62 *Nanyang Dr,* 637452, *Singapore*

Accepted 2019 December 10. Received 2019 September 25; in original form 2019 April 17

SUMMARY

Inverse problems play a central role in data analysis across the fields of science. Many techniques and algorithms provide parameter estimation including the best-fitting model and the parameters statistics. Here, we concern ourselves with the robustness of parameter estimation under constraints, with the focus on assimilation of noisy data with potential outliers, a situation all too familiar in Earth science, particularly in analysis of remote-sensing data. We assume a linear, or linearized, forward model relating the model parameters to multiple data sets with *a priori* unknown uncertainties that are left to be characterized. This is relevant for global navigation satellite system and synthetic aperture radar data that involve intricate processing for which uncertainty estimation is not available. The model is constrained by additional equalities and inequalities resulting from the physics of the problem, but the weights of equalities are unknown. We formulate the problem from a Bayesian perspective with non-informative priors. The posterior distribution of the model parameters, weights and outliers conditioned on the observations are then inferred via Gibbs sampling. We demonstrate the practical utility of the method based on a set of challenging inverse problems with both synthetic and real space-geodetic data associated with earthquakes and nuclear explosions. We provide the associated computer codes and expect the approach to be of practical interest for a wide range of applications.

Key words: Inverse problem; Inverse theory; Geodesy; Fault model; Crustal deformation.

1 INTRODUCTION

Data assimilation and model optimization are central to many fields of scientific inquiry (Sarvas 1987; Wunsch 1996; Bertero & Boccacci 1998; Bai *et al.* 2012; Halmos 2012; Santilli 2013; Ito & Jin 2015). In particular, inverse theory has proved an invaluable framework to unravel Earth's physical properties (Tarantola & Valette 1982; Parker 1994; Tarantola 2005; Aster *et al.* 2012). For example, the modelling of Earth's internal deformation from remote sensing has been cast as an inverse problem (Harris & Segall 1987; Segall & Matthews 1988, 1997; Fialko *et al.* 2001; Johanson *et al.* 2006; Rolandone *et al.* 2008; Barbot *et al.* 2008, 2013; Atzori *et al.* 2009; Lorito *et al.* 2011; Moore *et al.* 2017; Marchandon *et al.* 2018; Wang *et al.* 2018; Silverii *et al.* 2019; Tang *et al.* 2019). Despite the superficial simplicity of the linear, or linearized, forward model, inversion of geophysical data is often complicated by the sparsity of observations, the weak sensitivity of data to model parameters, the contamination of white and coloured noise and the presence of outliers, which make the inversion underdetermined and unstable without additional regularization (Yabuki & Matsu'Ura 1992; Fukahata & Wright 2008; Nocquet 2018). Markov chain Monte Carlo (MCMC) methods and Bayes theory have allowed us to avoid physically unjustified regularization (Simons *et al.* 2011; Minson *et al.* 2013, 2014; Bletery *et al.* 2016; Daout *et al.* 2016b,a; Gombert *et al.* 2018; Amey *et al.* 2018; Bagnardi & Hooper 2018). Other approaches have used sparsity (Evans & Meade 2012) or model discretization (Barnhart & Lohman 2010; Wang *et al.* 2016a) as a form of regularization.

Although MCMC sampling provides the means to investigate various point estimates (e.g. mean, median, mode and variance) and incorporate non-Gaussian statistics, optimization of model parameters is still widely used, perhaps owing to the computational convenience. Recent developments include the estimation of hyperparameters (Yabuki & Matsu'Ura 1992; Fukahata & Wright 2008), assimilation of heterogeneous data sets (Funning *et al.* 2014; Bletery

© The Author(s) 2019. Published by Oxford University Press on behalf of The Royal Astronomical Society. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Downloaded from https://academic.oup.com/gji/article-abstract/221/1/334/5674118 by sbarbot@usc.edu on 03 March 2020

et al. 2014, 2016; Jiang & Simons 2016; Gombert *et al.* 2018), errors in model predictions (Amiri-Simkooei & Jazaeri 2013; Duputel *et al.* 2014, 2015; Bletery *et al.* 2016; Gombert *et al.* 2017), errors in regularization (Wang *et al.* 2016a), optimization under positivity or bounds constraints (Nocquet 2018) and detection of outliers (Baarda 1968; Prószyński 1997; Snow & Schaffrin 2003; Rangelova *et al.* 2009; Lehmann 2013; Amiri-Simkooei & Jazaeri 2013; Lehmann & Lösler 2016; Wadehn *et al.* 2016; Lehmann & Voß-Böhme 2017; Milliner *et al.* 2018). The motivation behind this work is to combine some of the above considerations in a single, self-consistent framework.

With the advance of space-borne geodesy (Tronin 2010; Di Traglia *et al.* 2014; Blewitt 2015; Crosetto *et al.* 2016), multiple independent data sets may constrain a single phenomenon. These observations may be gathered with different instruments varying in sensitivity and resolution. In many practical cases of interest, the relative uncertainties among observations obtained through a consistent set of procedures and physical assumptions are well described. Their absolute uncertainties may be poorly documented, however, due to complex data processing steps and epistemic uncertainties. Meanwhile, the assimilation of large data sets may be complicated by the presence of outliers (Davies & Blewitt 2000; Nocquet & Calais 2004), which can arise because of gross processing errors, but more often because of the contribution of natural processes that are not included in the physical model, or poorly approximated.

To improve the robustness of inverse models, it is sometimes useful to incorporate additional constraints from the physics of the problem (Stark & Parker 1995; Schafer *et al.* 2003; Nocquet 2018). Indeed, Earth's response is bounded, and so are its spatial variations. For example, fault slip accommodates plate tectonics. Therefore, slip must be broadly aligned with the long-term motion of a fault. The small variations of stress drops over a wide range of earthquake magnitudes imply reasonable bounds on stress change and by extension, slip gradients. These constraints may be implemented by penalization of slip in directions orthogonal to plate motion, by imposing hard bounds, or both. But as the amount of oblique slip and stress drop are unknown *a priori*, the inversion method must estimate the optimal weight of these constraints.

With these considerations in mind, we describe a novel approach to linear inverse problems that aims to provide a robust estimate of model parameters under linear equality and inequality constraints based on a robust characterization of data uncertainty for multiple data sets, inference of the posterior distribution of the hyperparameters (i.e. free of tuning after model design) and outliers detection. In the next sections, we describe an algorithm based on the Bayesian perspective and Gibbs sampling that provides this information. The approach is based on MCMC methods, which is potentially computationally intensive. However, the assumption of a linear forward model allows a particularly efficient implementation of the random walk that does not require Metropolis sampling. Therefore, the proposed approach incorporates the flexibility of MCMC methods while converging faster to the true posterior than with Metropolis sampling.

In Section 2, we present a Bayesian model that defines the relationships between the relevant random variables. In Section 3, we discuss a post-processing algorithm that provides the modes of the posterior probability density of the model parameters. In Section 4, we document the performance and limitations of the proposed approach with synthetic tests that reproduce typical inversion settings based on space geodetic data. In Section 5, we provide examples of inversion of global positioning system (GPS) and synthetic aperture radar (SAR) data for fault slip and distributed internal strain that illustrate the benefits and shortcomings of the method.

2 BAYESIAN FORMULATION OF THE FORWARD MODEL

We start by considering D sets of observations $\boldsymbol{d}_i$, $i = 1, ..., D$, that were obtained conditionally independently of each other. For example, $\boldsymbol{d}_1$ may represent the displacements within a given GPS network and $\boldsymbol{d}_2$ the line-of-sight displacements from SAR interferometry (InSAR). Suppose that for each of these sets of observations, the forward model can be expressed in the linear form

$$\boldsymbol{d}_i = \mathbf{G}_i \boldsymbol{m} + \boldsymbol{\epsilon}_i + \boldsymbol{\delta}_i \,, \tag{1}$$

where $\boldsymbol{d}_i$ is a $N_{d_i} \times 1$ vector of observations, $\boldsymbol{m}$ is a $M \times 1$ vector of model parameters, both related through the forward-model matrix $\mathbf{G}_i$ of dimension $N_{d_i} \times M$, $\boldsymbol{\epsilon}_i$ denotes measurement noise, and $\boldsymbol{\delta}_i$ denotes the outliers. The level of measurement noise for these data sets may be different and is assumed to follow a Gaussian distribution

$$\boldsymbol{\epsilon}_i \sim \mathcal{N}\left(0, (\lambda_{d_i}\mathbf{W}_i)^{-1}\right) \,, \tag{2}$$

where $\lambda_{d_i}\mathbf{W}_i$ is the inverse covariance matrix (i.e., the precision matrix) representing the overall noise level of the measurements, $\mathbf{W}_i$ is a given weight matrix describing the relative uncertainties within the data set $\boldsymbol{d}_i$ at N_i different locations. In the simplest case, we have $\mathbf{W}_i = \mathbf{I}$. First, we consider the case whereby no *a priori* information on the noise level λ_{d_i} of a data set is available, except that it is positive. Under this scenario, we consider a non-informative Jeffrey's prior for λ_{d_i} (Jeffreys 1939)

$$p(\lambda_{d_i}) \propto \frac{1}{\lambda_{d_i}} \,, \tag{3}$$

which is scale invariant (Figueiredo & Nowak 2001; Berger 2013) and free of hyperparameters. Since the prior is non-informative, the resulting posterior distribution will reflect the information about the variance brought by the data (e.g. Carlin & Louis 2008; Liang *et al.* 2011). In a second case, the inverse covariance matrix $\lambda_{d_i}\mathbf{W}_i$ may be known, for example, by estimating the amplitude and spatial structure of atmospheric noise in InSAR data (Agram & Simons 2015; Jolivet *et al.* 2015; Elliott *et al.* 2016) or the temporal correlation of GPS time-series (Langbein 2008). In this case, we fix $\lambda_{d_i} = 1$ and set $\mathbf{W}_i$ to the provided value. When considering multiple data sets, some may have a properly estimated noise structure, others not, and both cases may be considered simultaneously.

In addition to measurement noise, we consider possible outliers, whose values are larger than the noise level. To remain meaningful, the outliers must be relatively rare, that is $\boldsymbol{\delta}_i$ should be a sparse vector with most entries zero. If the components of $\boldsymbol{\delta}_i$ are independent identically distributed random variables, each entry δ_{ij} of the outlier vector may follow a zero-mean normal distribution with precision γ_{ij}, following:

$$p(\delta_{ij}|\gamma_{ij}) \propto \sqrt{\gamma_{ij}} \exp\left(-\frac{1}{2}\gamma_{ij}{\delta_{ij}}^2\right) . \tag{4}$$

The outlier precision γ_{ij} is also a random variable (Fig. 1), and as we have no *a priori* knowledge on its value, we assume a non-informative prior

$$p(\gamma_{ij}) \propto \frac{1}{\gamma_{ij}} \,. \tag{5}$$

Downloaded from https://academic.oup.com/gji/article-abstract/221/1/334/5674118 by sbarbot@usc.edu on 03 March 2020

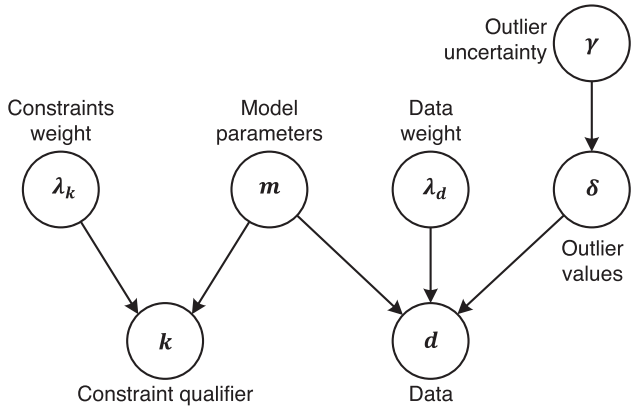

Figure 1. Directed graph (i.e. Bayesian network) corresponding to the proposed Bayesian model. The nodes denote the random variables and the edges characterize the conditional dependence among them. A random variable only has conditional dependence with its parents, children and the coparents of its children. The Markov chain Monte Carlo (MCMC) method described in Section 3 provides distributions of all these random variables.

Considering both γ_{ij} and δ_{ij} as random variables is appropriate to promote sparsity in δ_{ij} because of the following considerations. Imagine that the outlier precision follows a conjugate Gamma distribution

$$p(\gamma_{ij}) = \text{Gamma}(\gamma_{ij}; a_0, b_0) \tag{6}$$

$$= \frac{b_0^{a_0}}{\Gamma(a_0)} {\gamma_{ij}}^{a_0-1} \exp(-b_0\gamma_{ij}) \,, \tag{7}$$

where $\Gamma(a_0)$ denotes a complete gamma function. A marginal distribution for each outlier could then be obtained by integrating out γ_{ij}, as follows

$$\vdots$$

$$p(\delta_{ij}) = \int_0^\infty p(\delta_{ij}|\gamma_{ij}) p(\gamma_{ij}) \, \mathrm{d}\gamma_{ij} \tag{8}$$

$$= \frac{\Gamma(a_0 + \frac{1}{2})}{\Gamma(a_0)\sqrt{2\pi b_0}} \left(\frac{1}{1 + \frac{{\delta_{ij}}^2}{2b_0}} \right)^{a_0 + \frac{1}{2}} \,, \tag{9}$$

which corresponds to a Student's t-distribution. On the one hand, the Student's t-distribution will shrink most of δ_{ij} to zero as it has a very high peak at zero. On the other hand, the long tail of the distribution allows some elements δ_{ij} to be distant from zero. Hence, the resulting vector $\boldsymbol{\delta}_i$ will be sparse. Actually, in the particular case $a_0 = b_0 = 0$, the distribution of the precision γ_{ij} becomes eq. (5). The corresponding marginal distribution for each outlier δ_{ij} after integrating out γ_{ij} becomes

$$p(\delta_{ij}) \propto \frac{1}{|\delta_{ij}|} \,, \tag{10}$$

that is, the limiting case of the Student's t-distribution (9) for $a_0 = b_0 = 0$. This distribution also promotes sparsity, and is commonly used in the framework of sparse Bayesian learning (Figueiredo & Nowak 2001; Tipping 2001; Babacan *et al.* 2012; Yu & Dauwels 2015; Wang *et al.* 2016b).

We now turn our attention to different constraints on the model parameters. We consider both equality and inequality constraints. For convenience, we consider a set of L statistically independent equality constraints, taking the form

$$\mathbf{K}_i \boldsymbol{m} - \boldsymbol{k}_i = \mathbf{0} \,, \tag{11}$$

where $i = 1, ..., L$. From a Bayesian perspective, we can regard $\boldsymbol{k}_i$ as the pseudo observations from the forward model

$$\boldsymbol{k}_i = \mathbf{K}_i \boldsymbol{m} + \boldsymbol{\xi}_i \,, \tag{12}$$

where the noise follows a normal distribution

$$\boldsymbol{\xi}_i \sim \mathcal{N}\left(0, (\lambda_{k_i}\mathbf{I})^{-1}\right) \,, \tag{13}$$

and λ_{k_i} is a parameter that controls the strength of the constraint. Hence, the distribution of $\boldsymbol{k}_i$ given $\boldsymbol{m}$ and λ_{k_i} can be written as

$$\begin{aligned} p(\boldsymbol{k}_i|\boldsymbol{m}_i, \lambda_{k_i}) &= \mathcal{N}\left(\boldsymbol{k}_i; \mathbf{K}_i\boldsymbol{m}, (\lambda_{k_i}\mathbf{I})^{-1}\right) \\ &\propto {\lambda_{k_i}}^{\frac{N_{k_i}}{2}} \exp\left[-\frac{1}{2}\lambda_{k_i}(\mathbf{K}_i\boldsymbol{m} - \boldsymbol{k}_i)^T(\mathbf{K}_i\boldsymbol{m} - \boldsymbol{k}_i) \right] \,, \end{aligned} \tag{14}$$

where N_{k_i} denotes the length of vector $\boldsymbol{k}_i$. To infer the distribution of λ_{k_i} from the data, we also impose a non-informative Jeffrey's prior

$$p(\lambda_{k_i}) = \frac{1}{\lambda_{k_i}} \,, \tag{15}$$

similar to (3). The inequality constraints can be expressed as

$$\mathbf{A}\boldsymbol{m} \geq \boldsymbol{a} \,, \tag{16}$$

where $\mathbf{A}$ is a $N_a \times M$ matrix. Depending on the definition of $\mathbf{A}$, the inequality can represent positivity constraints, bounds, or other domain boundaries on the model parameters.

The directed graphical model corresponding to the proposed Bayesian model is shown in Fig. 1. The overall Bayesian model can be factorized as

$$\begin{aligned} &p(\boldsymbol{d}, \boldsymbol{k}, \boldsymbol{m}, \boldsymbol{\delta}, \boldsymbol{\lambda}_d, \boldsymbol{\lambda}_k, \boldsymbol{\gamma}) \\ &\quad \propto \; p(\boldsymbol{d}|\boldsymbol{m}, \boldsymbol{\delta}, \boldsymbol{\lambda}_d) p(\boldsymbol{\delta}|\boldsymbol{\gamma}) p(\boldsymbol{\lambda}_d) p(\boldsymbol{\gamma}) p(\boldsymbol{k}|\boldsymbol{m}, \boldsymbol{\lambda}_k) \, p(\boldsymbol{\lambda}_k) \,, \end{aligned} \tag{17}$$

where we have assumed $p(\boldsymbol{m}) \propto 1$ over the domain defined by $\mathbf{A}\boldsymbol{m} \geq \boldsymbol{a}$. In other words, the prior $p(\boldsymbol{m})$ follows a uniform distribution defined on the domain of $\boldsymbol{m}$, so it is non-informative. Alternatively, we can regard one of the equality constraints $p(\boldsymbol{k}_i|\boldsymbol{m}, \lambda_{k_i})$ as the prior of $\boldsymbol{m}$, that is, $p(\boldsymbol{m}|\lambda_{k_i}) \propto p(\boldsymbol{k}_i|\boldsymbol{m}, \lambda_{k_i})$, since $\boldsymbol{k}_i$ is given and fixed, and the remaining ones $p(\boldsymbol{k}_j|\boldsymbol{m}, \lambda_{k_j})$ as the pseudo observations (e.g. Rue & Held 2005; Yu *et al.* 2014; Yu & Dauwels 2016). Accordingly, we can factorize the Bayesian model as

$$\begin{aligned} &p(\boldsymbol{d}, \boldsymbol{k}, \boldsymbol{m}, \boldsymbol{\delta}, \boldsymbol{\lambda}_d, \boldsymbol{\lambda}_k, \boldsymbol{\gamma}) \\ &\quad = \; p(\boldsymbol{d}|\boldsymbol{m}, \boldsymbol{\delta}, \boldsymbol{\lambda}_d) p(\boldsymbol{\delta}|\boldsymbol{\gamma}) p(\boldsymbol{\lambda}_d) p(\boldsymbol{\gamma}) \\ &\qquad \times p(\boldsymbol{m}|\lambda_{k_i}) \prod_{j \neq i} p(\boldsymbol{k}_j|\boldsymbol{m}, \lambda_{k_j}) \, p(\boldsymbol{\lambda}_k) \,. \end{aligned} \tag{18}$$

Since N_{K_i} is typically smaller than M, the resulting prior $p(\boldsymbol{m}|\lambda_{k_i})$ may be rank-deficient. In practice, this does not cause any problems (e.g. Rue & Held 2005; Yu *et al.* 2014; Yu & Dauwels 2016). By application of Bayes' theorem, the posterior distributions of unknown parameters resulting from the formulas of eqs (17) and (18) are strictly equivalent.

Our objective is to infer the posterior distribution

$$p(\boldsymbol{m}, \boldsymbol{\lambda}_d, \boldsymbol{\lambda}_k, \boldsymbol{\delta}, \boldsymbol{\gamma}|\boldsymbol{d}, \boldsymbol{k}) \tag{19}$$

and further truncate it according to the inequality constraint $\mathbf{A}\boldsymbol{m} \geq \boldsymbol{a}$. As there is no closed-form expression of the posterior distribution with or without truncation, we draw samples from the distribution using the Gibbs sampling method, which is described in the next section.

Downloaded from https://academic.oup.com/gji/article-abstract/221/1/334/5674118 by sbarbot@usc.edu on 03 March 2020

Algorithm 1 Gibbs Sampling

Input: Initial value $\boldsymbol{x}^{(0)} = \{x_1^{(0)}, x_2^{(0)}, ..., x_M^{(0)}\}$, the conditional probability of x_i, the number of iterations I.
Output: A collection of $\boldsymbol{x}$ providing the marginal distribution of x_i.
for $\kappa = 1, 2, ..., I$ **do**
  for $j = 1, 2, ..., M$ **do**
    - Sample $x_j^{(\kappa)} \sim p(x_j|x_1^{(\kappa)}, ..., x_{j-1}^{(\kappa)}, x_{j+1}^{(\kappa-1)}, ..., x_M^{(\kappa-1)})$.
  end for
end for

3 BAYESIAN INFERENCE OF THE INVERSE PROBLEM WITH GIBBS SAMPLING

In this section, we describe an MCMC method to draw samples from the multivariate probability distribution of eq. (17) with the graphical model of Fig. 1. For numerical efficiency, we apply the Gibbs sampling method (Raftery & Lewis 1991; Casella & George 1992), which is typically more efficient than Metropolis–Hasting sampling, as its mixing rate is higher and fewer samples are required to cover the same parameter space. Gibbs sampling draws samples from a conditional probability distribution. The linearity of the forward model allows us to derive this truncated distribution in closed form, resulting in a minimal computational burden.

In the following subsection, we describe our overall method to draw samples from the truncated posterior distribution (19). Next, we describe how to identify point estimates and how to describe the marginal and the conditional distributions of $\boldsymbol{m}$ from the Gibbs samples.

3.1 Sampling the parameter space

Let us start with a general description of Gibbs sampling. Consider a multivariate distribution

$$p(\boldsymbol{x}) = p(x_1, x_2, ..., x_M)\,, \tag{20}$$

and suppose that we have found a proper initial value of $\boldsymbol{x}^{(0)} = \{x_1^{(0)}, x_2^{(0)}, ..., x_M^{(0)}\}$. Gibbs sampling proceeds by recursively drawing a new sample for each variable x_j in every iteration from the conditional distribution of this variable x_j conditioned on the remaining variables $\boldsymbol{x}_{-j}$, where the subscript $-j$ denotes all indices in $\{1, ..., M\}$ except j. Specifically in iteration κ, the conditional distribution can be expressed as:

$$p(x_j|\boldsymbol{x}_{-j}) = p(x_j|x_1^{(\kappa)}, ..., x_{j-1}^{(\kappa)}, x_{j+1}^{(\kappa-1)}, ..., x_M^{(\kappa-1)})\,. \tag{21}$$

This procedure is then repeated until there are enough samples to cover the entire parameter space, as summarized in Algorithm 1 in pseudo-code. The distribution of the Gibbs samples converges to the true joint posterior when it becomes stationary. In practice, since consecutive Gibbs samples are correlated, a thinning process is required to obtain independent samples by retaining every nth sample instead of all samples (Casella & George 1992). Moreover, samples from the beginning of the Markov chain (the burn-in period) are biased towards the initial value and are usually discarded (Raftery & Lewis 1991).

We now apply the Gibbs sampling approach to draw samples from the posterior (19) with or without the inequality constraints $\mathbf{A}\boldsymbol{m} \geq \boldsymbol{a}$. To apply the concepts of eqs (20) and (21), we consider the ensemble of random variables $\boldsymbol{x} = \{\boldsymbol{m}, \boldsymbol{\lambda}_d, \boldsymbol{\lambda}_k, \boldsymbol{\delta}, \boldsymbol{\gamma}\}$. Accordingly, we sequentially draw samples of $\boldsymbol{m}$, $\boldsymbol{\lambda}_d$, $\boldsymbol{\lambda}_k$, $\boldsymbol{\delta}$, and $\boldsymbol{\gamma}$ from the corresponding conditional distributions, which are described in closed-form in the following paragraphs.

First, we consider the model parameters $\boldsymbol{m}$. If there is no inequality constraint, the conditional distribution of $\boldsymbol{m}$ is a Gaussian distribution, such that

$$\begin{aligned} p(\boldsymbol{m}|\boldsymbol{\lambda}_d, \boldsymbol{\lambda}_k, \boldsymbol{\delta}, \boldsymbol{\gamma}, \boldsymbol{d}, \boldsymbol{k}) &= p(\boldsymbol{m}|\boldsymbol{\lambda}_d, \boldsymbol{\lambda}_k, \boldsymbol{\delta}, \boldsymbol{d}, \boldsymbol{k}) \\ &\propto \exp\left[-\frac{1}{2}\boldsymbol{m}^T\left(\sum_{i=1}^{D}\lambda_{d_i}\mathbf{G}_i^T\mathbf{W}_i\mathbf{G}_i + \sum_{i=1}^{L}\lambda_{k_i}\mathbf{K}_i^T\mathbf{K}_i\right)\boldsymbol{m}\right. \\ &\left. +\boldsymbol{m}^T\left(\sum_{i=1}^{D}\lambda_{d_i}\mathbf{G}_i^T\mathbf{W}_i(\boldsymbol{d}_i - \boldsymbol{\delta}_i) + \sum_{i=1}^{L}\lambda_{k_i}\mathbf{K}_i^T\boldsymbol{k}_i\right)\right], \end{aligned} \tag{22}$$

where we have used that $\boldsymbol{m}$ is conditionally independent of $\boldsymbol{\gamma}$ given $\boldsymbol{\delta}$ (Fig. 1). The above Gaussian distribution can be parametrized in the information form (Yu *et al.* 2014)

$$p(\boldsymbol{m}|\boldsymbol{\lambda}_d, \boldsymbol{\lambda}_k, \boldsymbol{\delta}, \boldsymbol{d}, \boldsymbol{k}) = \mathcal{N}(\boldsymbol{m}; \mathbf{J}^{-1}\boldsymbol{h}, \mathbf{J}^{-1})\,, \tag{23}$$

where the precision matrix $\mathbf{J}$ and the potential vector $\boldsymbol{h}$ can be written as

$$\mathbf{J} = \textstyle\sum_{i=1}^{D}\lambda_{d_i}\mathbf{G}_i^T\mathbf{W}_i\mathbf{G}_i + \sum_{i=1}^{L}\lambda_{k_i}\mathbf{K}_i^T\mathbf{K}_i\,, \tag{24}$$

and

$$\boldsymbol{h} = \textstyle\sum_{i=1}^{D}\lambda_{d_i}\mathbf{G}_i^T\mathbf{W}_i(\boldsymbol{d}_i - \boldsymbol{\delta}_i) + \sum_{i=1}^{L}\lambda_{k_i}\mathbf{K}_i^T\boldsymbol{k}_i\,. \tag{25}$$

The corresponding mean and covariance of the above Gaussian distribution can be computed as $\boldsymbol{\mu} = \mathbf{J}^{-1}\boldsymbol{h}$ and $\boldsymbol{\Sigma} = \mathbf{J}^{-1}$ as in (23). The resulting density function can be expressed as:

$$\begin{aligned} &p(\boldsymbol{m}|\boldsymbol{\lambda}_d, \boldsymbol{\lambda}_k, \boldsymbol{\delta}, \boldsymbol{d}, \boldsymbol{k}) \\ &\propto \exp\left[-\frac{1}{2}\left(\boldsymbol{m} - \boldsymbol{\mu}\right)^T\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{m} - \boldsymbol{\mu}\right)\right] \\ &\propto \exp\left[-\frac{1}{2}\left(\boldsymbol{m} - \mathbf{J}^{-1}\boldsymbol{h}\right)^T\mathbf{J}\left(\boldsymbol{m} - \mathbf{J}^{-1}\boldsymbol{h}\right)\right] \\ &\propto \exp\left[-\frac{1}{2}\boldsymbol{m}^T\mathbf{J}\boldsymbol{m} + \boldsymbol{m}^T\boldsymbol{h} - \frac{1}{2}\boldsymbol{h}^T\mathbf{J}^{-1}\boldsymbol{h}\right] \\ &\propto \exp\left[-\frac{1}{2}\boldsymbol{m}^T\mathbf{J}\boldsymbol{m} + \boldsymbol{m}^T\boldsymbol{h}\right], \end{aligned} \tag{26}$$

where the constant term in the exponential has been absorbed by use of the proportionality sign. In this case, we can sample the entire vector $\boldsymbol{m}$ as *one block* directly from the above Gaussian distribution by using the following transform:

$$\boldsymbol{m} = \mathbf{L}^{-T}(\boldsymbol{e} + \mathbf{L}^{-1}\boldsymbol{h})\,, \tag{27}$$

where $\boldsymbol{e} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ is a random variable that follows a standard multivariate normal distribution, $\mathbf{L}$ is the Cholesky decomposition of the precision matrix $\mathbf{J}$, that is, $\mathbf{J} = \mathbf{L}\mathbf{L}^T$, and $\mathbf{L}^{-T}$ denotes the transpose of the inverse of $\mathbf{L}$.

If some inequality constraints $\mathbf{A}\boldsymbol{m} \geq \boldsymbol{a}$ are present, we instead sample $\boldsymbol{m}$ from a multivariate truncated Gaussian distribution. To do so, we consider the random variable

$$\boldsymbol{e} = \mathbf{L}^T\boldsymbol{m} - \mathbf{L}^{-1}\boldsymbol{h}\,. \tag{28}$$

that corresponds to the inverse transform used in eq. (27). With this definition, the elements of $\boldsymbol{e}$ are decoupled, that is, $\boldsymbol{e}$ follows the truncated normal distribution $\mathcal{N}(\mathbf{0}, \mathbf{I})$ following the transformed inequality constraints

$$\tilde{\mathbf{A}}\boldsymbol{e} \geq \tilde{\boldsymbol{a}}\,, \tag{29}$$

where

Downloaded from https://academic.oup.com/gji/article-abstract/221/1/334/5674118 by sbarbot@usc.edu on 03 March 2020

Algorithm 2 Draw one Gibbs sample of $\boldsymbol{m}$ from the Truncated Gaussian Distribution

Input: The multivariate Gaussian distribution $\mathcal{N}(J^{-1}h, J^{-1})$, the inequality constraints $\mathbf{A}\boldsymbol{m} \geq \boldsymbol{a}$, and $\boldsymbol{m}^{(\kappa-1)}$ satisfying the constraints.
Output: $\boldsymbol{m}^{(\kappa)}$ satisfying the constraints.
Decouple the correlation between $\{m_1, \cdots, m_P\}$ by projecting it as $\boldsymbol{e} = \mathbf{L}^T\boldsymbol{m} - \mathbf{L}^{-1}\boldsymbol{h}$, where $\mathbf{L}$ is the Cholesky decomposition of $\mathbf{J}$ and so $\boldsymbol{e} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
Transform $\mathbf{A}$ and $\mathbf{a}$ accordingly as $\tilde{\mathbf{A}} = \mathbf{A}\mathbf{L}^{-T}$ and $\tilde{\boldsymbol{a}} = \boldsymbol{a} - \tilde{\mathbf{A}}\mathbf{L}^{-1}\boldsymbol{h}$
for $j = 1, ..., M$ **do**
- Obtain the upper bound u_j and lower bound l_j from the set of constraints
$\mathcal{S} = \{e_j : \tilde{\mathbf{A}}_{:,j}m_j \geq \tilde{a} - \tilde{\mathbf{A}}_{:,-j}\boldsymbol{e}_{-j}\}$,
where $\tilde{\mathbf{A}}_{:,j}$ denotes the j-th column in $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{A}}_{:,-j}$ denotes the remaining columns.
- Compute the value of cumulative distribution function (CDF) of both the upper and lower bound:
$u_j^C = \Phi(u_j; 0, 1)$,
$l_j^C = \Phi(l_j; 0, 1)$,
where $\Phi(*; 0, 1)$ denotes the CDF of a univariate standard normal distribution.
- Draw samples from $p(e_j|\boldsymbol{e}_{-j})$ with constraints $l_j \leq z_j \leq u_j$ using inverse transform sampling:
- Draw a sample from the standard uniform distribution, i.e., $v \sim U(0, 1)$.
- $e_j = \Phi^{-1}(l_j^C + v(u_j^C - l_j^C); 0, 1)$.
end for
Compute $\boldsymbol{m}^{(\kappa)} = \mathbf{L}^{-T}(\boldsymbol{e} + \mathbf{L}^{-1}\boldsymbol{h})$

$$\tilde{\mathbf{A}} = \mathbf{A}\mathbf{L}^{-T}\,, \tag{30}$$

$$\tilde{\boldsymbol{a}} = \boldsymbol{a} - \tilde{\mathbf{A}}\mathbf{L}^{-1}\boldsymbol{h}\,. \tag{31}$$

Next, we draw a sample of $\boldsymbol{e}$ following the Gibbs sampling scheme. Specifically, the conditional distribution of one variable e_j conditioned on the remaining variables $\boldsymbol{e}_{-j}$ reduces to a univariate standard normal distribution $\mathcal{N}(0, 1)$ truncated by bounds $[l_j, u_j]$. The bounds on e_j that satisfy $l_j \leq e_j \leq u_j$ are derived from the inequality $\tilde{\mathbf{A}}\boldsymbol{e} \geq \tilde{\boldsymbol{a}}$ given the values of $\boldsymbol{e}_{-j}$. To draw a sample from this univariate truncated normal distribution, we follow the inverse transform sampling approach (Geweke 1991). Specifically, we first evaluate the value of the cumulative distribution function (CDF) l_j^C and u_j^C corresponding to the bounds l_j and u_j on e_j. We then draw a uniformly distributed sample in the interval $[l_j^C, u_j^C]$, and convert this sample back to the Gaussian domain by applying the inverse CDF. Finally, we can obtain a sample of $\boldsymbol{m}$ from the sample $\boldsymbol{e}$ via eq. (27). For clarity, the resulting modified Gibbs sampling algorithm is listed in Algorithm 2.

Given the samples of $\boldsymbol{m}$, the conditional distribution of the remaining unknown parameters can find closed-form expressions following the relationships illustrated in Fig. 1. The conditional probability density of the outliers is described by

$$\begin{aligned} p(\boldsymbol{\delta}_i|\boldsymbol{m}, \boldsymbol{\lambda}_d, \boldsymbol{\lambda}_k, \boldsymbol{\gamma}, \boldsymbol{d}, \boldsymbol{k}) &= p(\boldsymbol{\delta}_i|\boldsymbol{m}, \lambda_{d_i}, \boldsymbol{\gamma}_i, \boldsymbol{d}_i) \\ &= \mathcal{N}\Big(\big[\lambda_{d_i}\mathbf{W}_i + \mathbf{diag}(\boldsymbol{\gamma}_i)\big]^{-1}\lambda_{d_i}\mathbf{W}_i(\boldsymbol{d}_i - \mathbf{G}_i\boldsymbol{m}), \\ &\quad \big[\lambda_{d_i}\mathbf{W}_i + \mathbf{diag}(\boldsymbol{\gamma}_i)\big]^{-1}\Big)\,, \end{aligned} \tag{32}$$

Algorithm 3 Bayesian Inference via Gibbs Sampling to Solve the Inverse Problem with Outliers, Equality and Inequality Constraints

Input: the data and the forward model characterized by $\boldsymbol{d}_i$, $\mathbf{G}_i$, and $\mathbf{W}_i$ for $i = 1, ..., D$, the equality constraints characterized by $\mathbf{K}_i$ and $\boldsymbol{k}_i$ for $i = 1 : L$, and the inequality constraints characterized by $\mathbf{A}$ and $\boldsymbol{a}$.
Output: Gibbs samples of $\boldsymbol{m}$, $\boldsymbol{\lambda}_d$, $\boldsymbol{\lambda}_k$, $\boldsymbol{\delta}$, and $\boldsymbol{\gamma}$.
Initialize $\boldsymbol{m}^{(0)}$, $\boldsymbol{\lambda}_d^{(0)}$, $\boldsymbol{\lambda}_k^{(0)}$, $\boldsymbol{\delta}^{(0)}$, and $\boldsymbol{\gamma}^{(0)}$.
for $\kappa = 1, 2, \cdots$ **do**
if There are no equality constraints **then**
Sample the entire vector $\boldsymbol{m}^{(\kappa)} \sim \mathcal{N}(\mathbf{J}^{-1}\boldsymbol{h}, \mathbf{J}^{-1})$,where $\mathbf{J}$ and $\boldsymbol{h}$ are defined in eqs (24) and (25), respectively.
else
Sample $\boldsymbol{m}^{(\kappa)}$ from the truncated Gaussian distribution using Algorithm 2.
end if
for $i = 1, ..., D$ **do**
Sample $\boldsymbol{\delta}_i^{(\kappa)} \sim \mathcal{N}(\mathbf{J}_{\delta_i}^{-1}\boldsymbol{h}_{\delta_i}, \mathbf{J}_{\delta_i}^{-1})$, where
$\mathbf{J}_{\delta_i} = \lambda_{d_i}^{(\kappa-1)}\mathbf{W}_i + \mathbf{diag}(\boldsymbol{\gamma}_i^{(\kappa-1)})$,
$\boldsymbol{h}_{\delta_i} = \lambda_{d_i}^{(\kappa-1)}\mathbf{W}_i(\boldsymbol{d}_i - \mathbf{G}_i\boldsymbol{m}^{(\kappa)})$.
for $j = 1, ..., N_{d_i}$ **do**
Sample $\gamma_{ij}^{(\kappa)} \sim \text{Gamma}(0.5, {\delta_{ij}^{(\kappa)}}^2/2)$.
end for
Sample $\lambda_{d_i}^{(\kappa)} \sim \text{Gamma}(N_{d_i}/2, (\boldsymbol{d}_i - \mathbf{G}_i\boldsymbol{m}^{(\kappa)} - \boldsymbol{\delta}_i^{(\kappa)})^T \mathbf{W}_i(\boldsymbol{d}_i - \mathbf{G}_i\boldsymbol{m}^{(\kappa)} - \boldsymbol{\delta}_i^{(\kappa)})/2)$.
end for
for $i = 1, ..., L$ **do**
Sample $\lambda_{k_i}^{(\kappa)} \sim \text{Gamma}(N_{k_i}/2, (\boldsymbol{k}_i - \mathbf{K}_i\boldsymbol{m}^{(\kappa)})^T (\boldsymbol{k}_i - \mathbf{K}_i\boldsymbol{m}^{(\kappa)})/2)$.
end for
end for

where the operator $\mathbf{diag}(\boldsymbol{x})$ places the elements of the vector $\boldsymbol{x}$ along a diagonal matrix. Following eqs (4) and (5), the conditional probability density for the outlier precision follows:

$$\begin{aligned} p(\gamma_{ij}|\boldsymbol{m}, \boldsymbol{\lambda}_d, \boldsymbol{\lambda}_k, \boldsymbol{\delta}, \boldsymbol{d}, \boldsymbol{k}) &= p(\gamma_{ij}|\delta_{ij}) \\ &= \text{Gamma}\left(\frac{1}{2}, \frac{\delta_{ij}^2}{2}\right). \end{aligned} \tag{33}$$

The conditional probability density of the data weights is given by

$$\begin{aligned} p(\lambda_{d_i}|\boldsymbol{m}, \boldsymbol{\lambda}_k, \boldsymbol{\delta}, \boldsymbol{\gamma}, \boldsymbol{d}, \boldsymbol{k}) &= p(\lambda_{d_i}|\boldsymbol{m}, \boldsymbol{\delta}_i, \boldsymbol{d}_i) \\ &= \text{Gamma}\left(\frac{N_{d_i}}{2}, \frac{(\boldsymbol{d}_i - \mathbf{G}_i\boldsymbol{m} - \boldsymbol{\delta}_i)^T\mathbf{W}_i(\boldsymbol{d}_i - \mathbf{G}_i\boldsymbol{m} - \boldsymbol{\delta}_i)}{2}\right). \end{aligned} \tag{34}$$

Finally, the weights on the equality constraints have the probability density

$$\begin{aligned} p(\lambda_{k_i}|\boldsymbol{m}, \boldsymbol{\lambda}_d, \boldsymbol{\delta}, \boldsymbol{\gamma}, \boldsymbol{d}, \boldsymbol{k}) &= p(\lambda_{k_i}|\boldsymbol{m}, \boldsymbol{k}_i) \\ &= \text{Gamma}\left(\frac{N_{k_i}}{2}, \frac{(\boldsymbol{k}_i - \mathbf{K}_i\boldsymbol{m})^T(\boldsymbol{k}_i - \mathbf{K}_i\boldsymbol{m})}{2}\right). \end{aligned} \tag{35}$$

We then sample from the above conditional distributions recursively until the distribution of the Gibbs samples becomes stationary and until there is a sufficient number of samples to cover the entire parameter space. The overall Gibbs sampling algorithm is summarized in Algorithm 3. We use standard numerical functions to generate random numbers following Gaussian or Gamma distributions.

Downloaded from https://academic.oup.com/gji/article-abstract/221/1/334/5674118 by sbarbot@usc.edu on 03 March 2020

The resulting computational complexity of the proposed algorithm per iteration scales as

$$C = \mathcal{O}\left(M^3 + M^2\left(N_a + N_d + N_k\right) + MN_{d^2}\right) , \tag{36}$$

where M is the number of model parameters, N_a is the number of inequality constraints, $N_d = \sum_{i=1}^{D} N_{d_i}$ is the total number of measurements, $N_{d^2} = \sum_{i=1}^{D} N_{d_i}^2$, and $N_k = \sum_{i=1}^{L} N_{k_i}$ is the total number of equality constraints. The first term $\mathcal{O}(M^3)$ derives from the Cholesky decomposition. The terms $\mathcal{O}(M^2 N_d)$ and $\mathcal{O}(MN_{d^2})$ arise from the matrix product $\mathbf{G}_i^T \mathbf{W}_i \mathbf{G}_i$ in (24). Note that $\mathcal{O}(MN_{d^2})$ reduces to $\mathcal{O}(MN_d)$ when $\mathbf{W}_i$ is a diagonal matrix. Similarly, the term $\mathcal{O}(M^2 N_k)$ results from the matrix product $\mathbf{K}_i^T \mathbf{K}_i$ in (24). Finally, the last term in (36) stems from the operations on the inequality constraints in Algorithm 2. In the case without the inequality constraints, $N_a = 0$ and this term can be removed.

3.2 Central tendency description: mean, median and mode

Evaluating the moments of the Gibbs samples, for example, the mean and median, is straightforward. The mode of the posterior distribution is of particular interest in regard to explaining the data. Specifically, we seek to estimate the mode of the marginal of the model parameters, taking into account the range of possible values of the hyperparameters and their associated uncertainties. Still referring to Fig. 1, the corresponding marginal distribution can be written

$$\begin{aligned} p(\boldsymbol{m}|\boldsymbol{d}, \boldsymbol{k}) &= \int p(\boldsymbol{m}, \boldsymbol{\lambda}_d, \boldsymbol{\lambda}_k, \boldsymbol{\delta}, \boldsymbol{\gamma}|\, \mathrm{d}, \boldsymbol{k})\, \mathrm{d}\boldsymbol{\lambda}_d\, \mathrm{d}\boldsymbol{\lambda}_k\, \mathrm{d}\boldsymbol{\delta}\, \mathrm{d}\boldsymbol{\gamma} \\ &= \int p(\boldsymbol{m}|\boldsymbol{\lambda}_d, \boldsymbol{\lambda}_k, \boldsymbol{\delta}, \boldsymbol{\gamma}, \boldsymbol{d}, \boldsymbol{k}) \\ &\quad p(\boldsymbol{\lambda}_d, \boldsymbol{\lambda}_k, \boldsymbol{\delta}, \boldsymbol{\gamma}|\boldsymbol{d}, \boldsymbol{k})\, \mathrm{d}\boldsymbol{\lambda}_d\, \mathrm{d}\boldsymbol{\lambda}_k\, \mathrm{d}\boldsymbol{\delta}\, \mathrm{d}\boldsymbol{\gamma} , \end{aligned} \tag{37}$$

which is a mixture of Gaussian distributions. To estimate the mode of (37), we utilize the annealed mean shift method (Shen *et al.* 2007), a particularly suited algorithm to find the mode of the empirical density from Gibbs samples.

As a preliminary, we introduce the kernel density and the mean shift algorithm. Take the model parameters $\boldsymbol{m}$ as an example. Given N samples of the model parameters $\boldsymbol{m}^{(1:N)}$, the multivariate kernel density can be expressed as (Silverman 1981)

$$\hat{p}(\boldsymbol{m}) = \frac{1}{N}\sum_{i=1}^{N} \mathcal{K}_H\left(\boldsymbol{m} - \boldsymbol{m}^{(i)}\right), \tag{38}$$

$$= \frac{\det(\mathbf{H})^{-\frac{1}{2}}}{N}\sum_{i=1}^{N} \mathcal{K}\left(\mathbf{H}^{-\frac{1}{2}}\left(\boldsymbol{m} - \boldsymbol{m}^{(i)}\right)\right) , \tag{39}$$

where $\mathcal{K}$ is a kernel function that is non-negative, and $\mathbf{H}$ is the bandwidth matrix that is positive definite. $\mathcal{K}_H$ is a scaled kernel defined as $\mathcal{K}_H(\boldsymbol{m}) = \det(\mathbf{H})^{-\frac{1}{2}}\mathcal{K}(\mathbf{H}^{-\frac{1}{2}}\boldsymbol{m})$. One commonly used kernel is the Gaussian kernel, that is, $\mathcal{K}(\boldsymbol{m}) = \phi(\boldsymbol{m})$, where $\phi(\boldsymbol{m})$ is the probability density function of a M-variate standard normal distribution with mean $\mathbf{0}$ and covariance $\mathbf{I}$. In practice, one typically assumes that the bandwidth matrix is isotropic, that is, $\mathbf{H} = h^2\mathbf{I}$, where h is called the bandwidth (Silverman 1981). Under these assumptions, the kernel density can be equivalently written as (Silverman 1981)

$$\hat{p}(\boldsymbol{m}) = \frac{1}{Nh}\sum_{i=1}^{N} \mathcal{K}\left(\left\|\frac{\boldsymbol{m} - \boldsymbol{m}^{(i)}}{h}\right\|^2\right) . \tag{40}$$

The mean shift algorithm seeks the mode of $\hat{p}(\boldsymbol{m})$ by setting the gradient $\nabla_{\boldsymbol{m}}\hat{p}(\boldsymbol{m}) = 0$ in every iteration, resulting in the following update rule for $\boldsymbol{m}$ (Shen *et al.* 2007)

$$\boldsymbol{m}^{\{t+1\}} = \frac{\sum_{i=1}^{N} \mathcal{G}\left(\left\|\frac{\boldsymbol{m}^{\{t\}} - \boldsymbol{m}^{(i)}}{h}\right\|^2\right)\boldsymbol{m}^{(i)}}{\sum_{i=1}^{N} \mathcal{G}\left(\left\|\frac{\boldsymbol{m}^{\{t\}} - \boldsymbol{m}^{(i)}}{h}\right\|^2\right)} , \tag{41}$$

where t denotes the iteration number, and $\mathcal{G}(\cdot) = -\mathcal{K}'(\cdot)$

Since the mean shift is a gradient-based algorithm, it is sensitive to local maxima. To mitigate the problem, we resort to annealed mean shift (Shen *et al.* 2007). Specifically, we initiate the distribution by a kernel density of sufficiently large bandwidth, and apply mean shift to find the mode of the kernel density function. We then slowly decrease the bandwidth, and for each bandwidth, we further maximize the kernel density via mean shift starting from the previous mode, until the mode does not change within a threshold. Note that setting the bandwidth of the kernel to be large is equivalent to smoothing the true density of $\boldsymbol{m}$, thus removing the local maxima. It has been observed that the oversmoothed density function with a sufficiently large bandwidth is unimodal (Shen *et al.* 2007). Using a continuation principle, the influence of the global peak in the true density function is introduced gradually when reducing the bandwidth. In this way, the algorithm is more likely to attain the global maximum.

3.3 Conditional distribution of the model parameters

After obtaining the mode of $\boldsymbol{\lambda}_d$, $\boldsymbol{\lambda}_k$ and $\boldsymbol{\delta}$, denoted $\hat{\boldsymbol{\lambda}}_d$, $\hat{\boldsymbol{\lambda}}_k$ and $\hat{\boldsymbol{\delta}}$, respectively, we can use the Gaussian posterior conditional distribution

$$\theta(\boldsymbol{m}) = p(\boldsymbol{m}|\hat{\boldsymbol{\lambda}}_d, \hat{\boldsymbol{\lambda}}_k, \hat{\boldsymbol{\delta}}, \boldsymbol{d}, \boldsymbol{k}) , \tag{42}$$

given by eq. (23) to describe the behaviour of the model parameters $\boldsymbol{m}$. Equivalently, we follow the empirical Bayes framework (Casella 1985), in which the hyper parameters $\boldsymbol{\lambda}_d$, $\boldsymbol{\lambda}_k$ and the outliers $\boldsymbol{\delta}$ are set to their most probable values instead of being integrated out. The resulting conditional posterior distribution (42) provides a practical way to characterize $\boldsymbol{m}$, as it is simply defined by the mean and the covariance.

In the case where the inequality constraint $\mathbf{A}\boldsymbol{m} \geq \boldsymbol{a}$ are present, the distribution (42) will be a truncated Gaussian. The truncated Gaussian distribution is parametrized by three parameters: the mean and covariance of the original Gaussian distribution that is not truncated, and the truncation. In our case, the truncation is given by $\mathbf{A}\boldsymbol{m} \geq \boldsymbol{a}$ and the mean and covariance of the untruncated Gaussian distribution is the same with those in eq. (23).

3.4 Posterior data uncertainty

The posterior data uncertainty can be estimated given the Bayesian framework of Fig. 1 following two interpretations of the nature of outliers. From a first standpoint, we assume that the outliers have been modelled out, that is, $\boldsymbol{\delta}_i$ represents a forward model. In this case, we can write the conditional probability distribution of the observed data $\boldsymbol{d}_i$ given the model parameters, noise, and outliers, that is, the elements that collectively form the forward model, as

$$\begin{aligned} p(\boldsymbol{d}_i|\boldsymbol{m}, \lambda_{d_i}, \boldsymbol{\delta}_i) &= \mathcal{N}\left(\boldsymbol{d}_i; \mathbf{G}_i\boldsymbol{m} + \boldsymbol{\delta}_i, (\lambda_{d_i}\mathbf{W}_i)^{-1}\right) \\ &\propto \lambda_{d_i}^{\frac{N_{d_i}}{2}} \exp\left[-\frac{1}{2}\lambda_{d_i}(\boldsymbol{d}_i - \mathbf{G}_i\boldsymbol{m} - \boldsymbol{\delta}_i)^T\mathbf{W}_i(\boldsymbol{d}_i - \mathbf{G}_i\boldsymbol{m} - \boldsymbol{\delta}_i)\right]. \end{aligned} \tag{43}$$

Downloaded from https://academic.oup.com/gji/article-abstract/221/1/334/5674118 by sbarbot@usc.edu on 03 March 2020

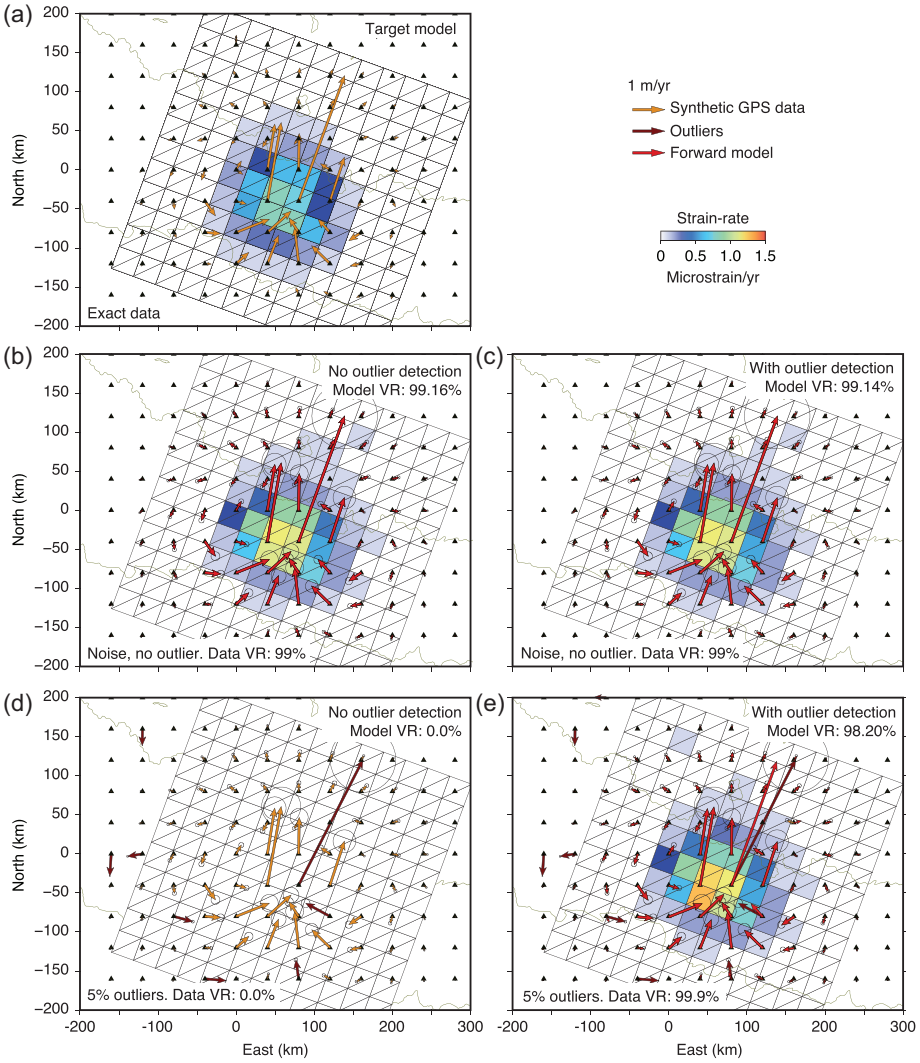

Figure 2. Verification of the inversion technique with synthetic data: inferring the distribution of subsurface strain from space-borne geodesy. (a) A surface velocity field (orange vectors) is calculated along a rectilinear array (black triangles) assuming internal deformation (strain-rate) within the Earth's interior corresponding to viscoelastic flow in the Indian lower crust below Nepal. (b) Inverted distribution of strain-rate (background colours) and corresponding forward model for surface displacements (red vectors) for a data set with white noise without outlier detection. (c) Inverted strain-rate and forward model for a noisy data set with outlier detection. (d) Inverted strain-rate and forward model for a data set with outliers (brown vectors), but without outlier detection. In both d and e, large outliers are added to 18 different stations, about 5 per cent of the data set, representing 4.2 per cent of the data variance. The colour of the volume elements indicates the magnitude of the deviatoric strain tensor.

The above function $p(\boldsymbol{d}_i|\boldsymbol{m}, \lambda_{d_i}, \boldsymbol{\delta}_i)$ is the distribution of $\boldsymbol{d}_i$, but also the likelihood function of $\boldsymbol{m}$.

From another perspective, the outliers represent a gross data error. Following eq. (1), as the noise $\boldsymbol{\epsilon}_i$ and the outliers $\boldsymbol{\delta}_i$ are independent, the overall covariance of $\boldsymbol{d}_i$ is the sum of the covariances of $\boldsymbol{\epsilon}_i$ and $\boldsymbol{\delta}_i$. The posterior distribution of the data can then be written

$$\begin{aligned} p(\boldsymbol{d}_i|\boldsymbol{m}, \lambda_{d_i}, \boldsymbol{\gamma}_i) &= \int p(\boldsymbol{d}_i|\boldsymbol{m}, \lambda_{d_i}, \delta_i) p(\delta_i|\gamma_i)\, \mathrm{d}\delta_i, \\ &\propto \exp\left\{ -\frac{1}{2}(\boldsymbol{d}_i - \mathbf{G}_i\boldsymbol{m})^T \left[(\lambda_{d_i}\mathbf{W}_i)^{-1} \right.\right. \\ &\left.\left. + \mathbf{diag}(\gamma_i)^{-1}\right]^{-1} (\boldsymbol{d}_i - \mathbf{G}_i\boldsymbol{m}) \right\}, \end{aligned} \tag{44}$$

where $\mathbf{diag}(\gamma_i)$ denotes a diagonal matrix with γ_i on the diagonal, $\mathbf{diag}(\gamma_i)^{-1}$ is the covariance of δ_i, and $(\lambda_{d_i}\mathbf{W}_i)^{-1}$ is the covariance of the noise $\boldsymbol{\epsilon}_i$.

The formulation (44) is reminiscent of other approaches whereby the overall covariance of $\boldsymbol{d}_i$ is decomposed into the sum of the covariance of the measurement noise and the covariance of the prediction error (Tarantola 2005; Duputel *et al.* 2014; Jolivet *et al.* 2015). The latter measures the inadequacies of using the proposed linear model $\mathbf{G}_i\boldsymbol{m}$ to describe the data $\boldsymbol{d}_i$, and can be approximated based on a perturbation method (Duputel *et al.* 2014). Our approach is more general, as we simply regard the outliers as the prediction error. This allows us to detect outliers wherever the forward model is locally inadequate. This approach provides the spatial distribution, amplitude and uncertainty of the detected outliers, but no information about the underlying cause for their presence.

4 VERIFICATION

We illustrate the potential and limitation of the inversion approach by considering a synthetic case where we control the data uncertainty, number of outliers, and the target model. We consider the

Downloaded from https://academic.oup.com/gji/article-abstract/221/1/334/5674118 by sbarbot@usc.edu on 03 March 2020

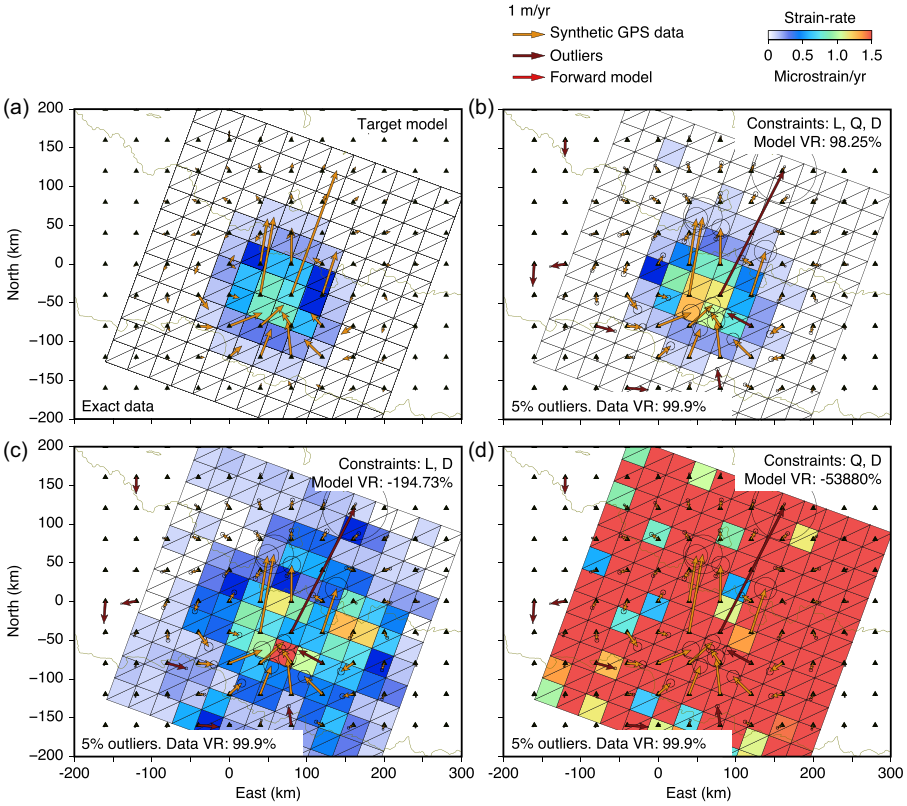

Figure 3. Impact of modelling assumptions (regularizing constraints) on data misfit and model recovery. (a) Target model (background strain-rate) and exact data (orange vectors). (b) Inverted distribution of strain-rate using data with 5 per cent outliers (brown vectors), outlier detection, and three regularizing constraints: a smoothing matrix **L**, a constraint on the direction of strain **Q**, and a penalization of isotropic components **D**. (c) Inverted strain-rate with only two regularizing constraints: **L** and **D**. (d) Inverted strain-rate with two constraints: **Q** and **D**. In all cases the data can be fit with more than 99.9 per cent variance reduction. However, the model is recovered satisfactorily only when appropriate physical assumptions are included in the inversion.

scenario where the distribution of subsurface strain must be imaged based solely on the measurement of displacement vectors available at the Earth's surface. These measurements may originate from a land-based geodetic network, or from space geodesy (Leprince *et al.* 2007), assuming that the observation network was properly designed (Sathiakumar *et al.* 2017). This scenario captures the general problem of imaging the accelerated viscoelastic flow in the lower crust and upper mantle that follows large earthquakes (Tsang *et al.* 2016; Moore *et al.* 2017; Qiu *et al.* 2018; Tang *et al.* 2019).

4.1 Problem description

We consider the tectonic context of the Nepalese Himalaya with the collision of the Indian plate beneath the Eurasian Plate. We simulate a synthetic data set of surface displacements at a 10×12 network of three-component GPS stations organized in a rectilinear array assuming viscoelastic flow confined in the Indian lower crust (Landry & Barbot 2019). The surface displacements are calculated by meshing the subsurface into $12 \times 12 \times 2 = 288$ cuboidal and dodecahedral volume elements and using the associated Green's functions (Barbot *et al.* 2017; Barbot 2018). Each volume element has six degrees of freedom, resulting in $6 \times 288 = 1728$ model parameters. We then add Gaussian noise and outliers to the data. The combination of 3×120 measurements and 1728 model parameters gives rise to an underdetermined inverse problem.

To regularize the problem, we impose additional constraints. Since viscoelastic flow is deviatoric, that is, incompressible, we constrain the trace of the strain tensor to vanish. This is enforced by the algebraic constraint $\mathbf{D}\boldsymbol{m} = \mathbf{0}$. As the viscoelastic flow relaxes a known stress field, we penalize the direction of strain in five directions orthogonal to the local stress tensor at the centre of each volume element. The direction of a strain or stress tensor is meaningful considering that any symmetric second-order tensor can find a six-component vector representation. These constraints on the direction of flow can be written $\mathbf{Q}\boldsymbol{m} = \mathbf{0}$. We also impose a smooth distribution of strain based on the assumption that the wavelength of the forces driving the flow is large. This leads to the regularizing constraints $\mathbf{L}\boldsymbol{m} = \mathbf{0}$. These modelling assumptions yield a linear inverse problem (the forward model is algebraic) with equality constraints and a regularization constraint of unknown weights. Specifically, we have $L = 3$, $\mathbf{K}_1 = \mathbf{D}$, $\mathbf{K}_2 = \mathbf{Q}$ and $\mathbf{K}_3 = \mathbf{L}$, with $\mathbf{k}_i = \mathbf{0}, i = 1 \ldots L$.

The benchmark considered here is challenging due to the presence of a large number of data and parameters, the disproportion of model parameters compared to the available data, measurement noise, the presence of outliers and the weak sensitivity of the data to the model parameters resulting from the *Saint Venant* principle. We use Algorithm 3 to estimate the data uncertainty for each measurement, and the individual weights of the constraints. This configuration allows us to document the quality of fit to the data, but also the quality of model recovery, since the latter is also known. We estimate the posterior distribution by drawing $\kappa = 200\,000$ samples in total. We regard the first 140 000 samples as the burn-in samples. For each case considered in this section, the simulation takes about 5 hr on a 64-bit windows OS laptop with two Intel i7-4940MX CPU, 3.10GHz processors with 32.0 GB RAM.

Downloaded from https://academic.oup.com/gji/article-abstract/221/1/334/5674118 by sbarbot@usc.edu on 03 March 2020

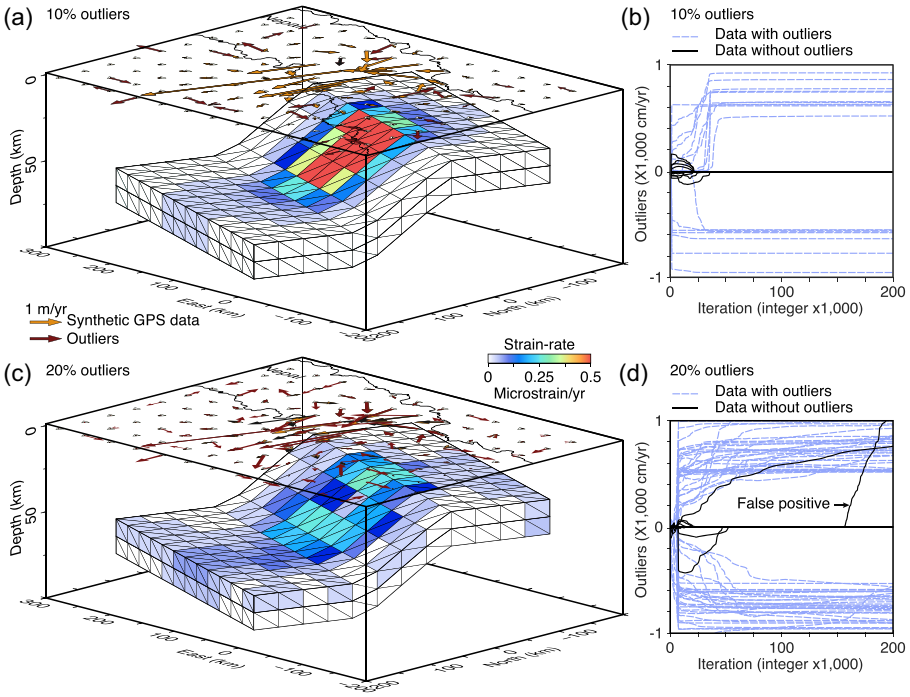

Figure 4. Effect of outliers. (a) A surface velocity field (orange vectors) is calculated along a rectilinear array (white triangles) assuming internal deformation (strain-rate) within the Earth's interior (mesh of dodecahedra). Large outliers are added to 36 different stations (dark red vectors), representing about 10 per cent of the data set. The colour of the volume elements indicates the magnitude of the deviatoric strain tensor. With the outlier detection the distribution of strain is recovered within less than 1 per cent error with a corresponding variance reduction of the exact data of more than 99 per cent. The true data and forward model are indistinguishable at this scale. The true and recovered models also appear extremely similar. (b) The evolution of the outlier vector (running median of δ) as a function of iterations during Gibbs sampling. The data altered with outliers (dashed blue profiles) are all detected. The data with white noise (black profiles) are not biased by false positive detection. (c) Inferred distribution of strain with 20 per cent of outliers in the data set. The model is only 20 per cent similar to the target model. (d) Corresponding evolution of the running median of δ. Although many outliers are properly detected, some outliers are not, and some non-outliers are inadequately flagged as such. Without outlier detection, the model cannot be recovered satisfactorily (not shown). Vertical exaggeration (x5).

4.2 Outlier detection

We first demonstrate the potential importance of outlier detection (Fig. 2). A target model is shown in Fig. 2(a), with the associated surface displacement vectors at the location of the synthetic geodetic network. In a first case, we only add white noise to the synthetic data and follow the algorithm without outlier detection. The model is recovered with a variance reduction of 99.16 per cent, that is, the target and recovered model differ by less than 1 per cent. The data are also well fit by the forward model. We also follow the algorithm with outlier detection, which only has a modest impact on the model recovery (99.14 per cent in this case). The small difference in performance is expected from the stochastic nature of the algorithm.

We then consider a data set where large outliers are added to 18 different stations, about 5 per cent of the data set, representing 4.2 per cent of the data variance. Without outlier detection, the model recovery and data variance reduction are 0 per cent. In contrast, with outlier detection, the model recovery reaches 98.20 per cent with a data variance reduction of 99.9 per cent. These results indicate the importance of outlier detection, the capacity of the proposed algorithm to detect a small set of outliers, and the minimal loss of performance in doing so.

The model is well recovered if no outliers are present, or if they are properly detected. In contrast, models are poorly recovered when outliers are not taken into account in the analysis. For example, simpler approaches using weighted least squares (Parker 1994), regularized least squares (Tarantola & Valette 1982; Tarantola 2005), pseudo-inverse (Aster *et al.* 2012), a QR decomposition (Golub & Van Loan 1996) or the lasso and basis pursuit techniques (van den Berg & Friedlander 2008), all fail to recover the target model in the presence of even a few outliers. We do not illustrate these negative results because the differences between the target and recovered models are so large that they are meaningless.

4.3 Limitations

We now illustrate some shortcomings of the approach. First, as the inverse problem considered is underdetermined, the solution depends on the modelling assumptions. Fig. 3 shows the model recovery as a function of the type and number of constraints assumed in the inversion. We consider data with white noise and 5 per cent outliers. Considering constraints on the direction of strain, incompressibility of flow, and smoothness of the deformation field, the model is recovered with a 98.25 per cent variance reduction. Accordingly, the data is reduced at the 99.9 per cent level. If the constraint on the direction of strain is relaxed, the model recovery deteriorates significantly, despite a 99.9 per cent variance reduction of the data. If the constraints of spatial smoothness are relaxed, the model recovery is catastrophically wrong. In all these cases the data can be fit with more than 99.9 per cent variance reduction. However, the model is recovered satisfactorily only when appropriate physical assumptions are included in the inversion.

Downloaded from https://academic.oup.com/gji/article-abstract/221/1/334/5674118 by sbarbot@usc.edu on 03 March 2020

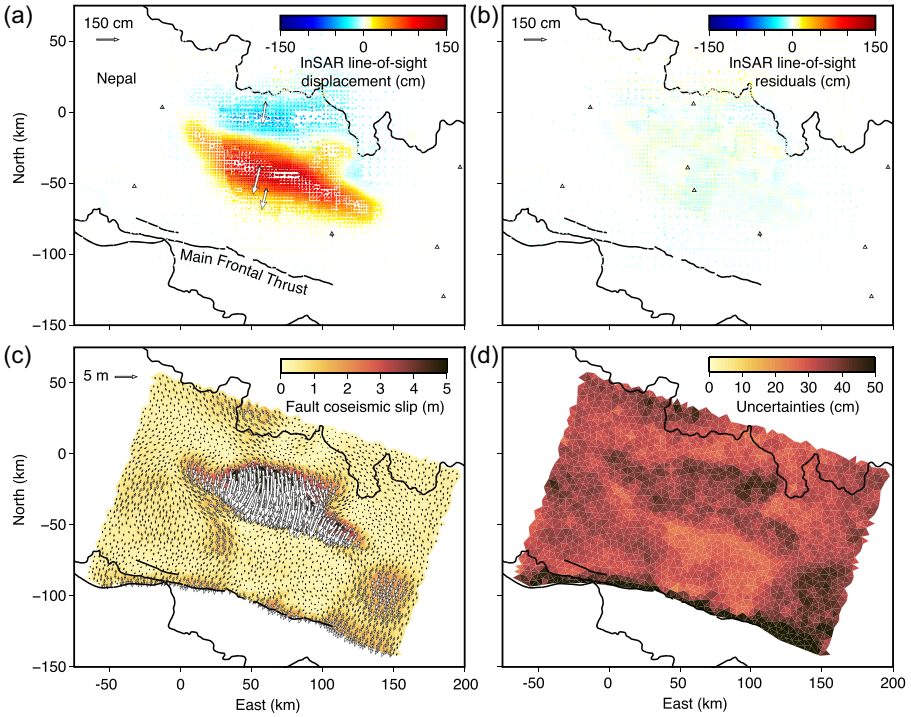

Figure 5. Inversion of space-borne geodetic data: case of the 2015 $M_w = 7.8$ Gorkha, Nepal earthquake. (a) Interferometric synthetic aperture radar (InSAR) line-of-sight displacements and GPS displacements (white vectors) for the 2015 $M_w = 7.8$ Gorkha earthquake. (b) Residuals after inversion for the distribution of coseismic slip. (c) Fault slip corresponding to the mean of the Gibbs sampling. The vectors indicate the amplitude and direction of fault slip. (d) The standard deviation of the fault slip based on Gibbs sampling.

Another limitation of our approach is in the presence of a large number of outliers (Fig. 4). For example, Fig. 4(b) shows the evolution of outlier identification during Gibbs sampling when 10 per cent outliers are present. All outliers are correctly identified without any false detection. The outlier detection converges after about 50 000 iterations. The estimation of the remaining parameters converges after 140 000 iterations. The distribution of internal strain can be recovered and the surface data can be reduced within 1 per cent variability when up to 10 per cent of the data contain outliers. Above this threshold, our assumptions on the statistics of outliers break down and the recovered model shows substantial deviations from the original. Below 10 per cent of outliers, the model is recovered with less than 1 per cent variations. For example, for 5 per cent outliers, the data variance reduction is 99.999 per cent and the model variance reduction is 98.206 per cent. Fig. 2(c) shows the model recovery for the case of 20 per cent outliers, showing a substantial smearing of the strain distribution. Fig. 2(d) illustrates the corresponding outlier detection. Even though many outliers are correctly detected, some are not and some outlier-free data are incorrectly labeled as outliers (false positives).

Despite these limitations, these results give us confidence that the method is correctly implemented and that the approach can be useful to tackle inversion of real data when only a small fraction of the data is contaminated by outliers.

5 EXAMPLES

We now consider various examples of natural and man-made deformation that illustrate the potential of the method. All examples include a heterogeneous data set consisting of measurements collected by different instruments that may be corrupted by outliers. In Section 5.1, we quantify various point estimates of the model parameters. In Section 5.2, we investigate the truncated posterior distribution of the model parameters. Examples 5.1 and 5.2 both include inequality constraints. In Section 5.3, we consider a case of man-made deformation with a relative large number of outliers.

5.1 Coseismic slip distribution of the 2015 $M_w = 7.8$ Gorkha, Nepal earthquake

We now illustrate the algorithm considering surface displacements caused by large earthquakes. We first consider the case of the 2015 $M_w = 7.8$ Gorkha, Nepal earthquake (Avouac *et al.* 2015; Galetzka *et al.* 2015; Lindsey *et al.* 2015; Wang & Fialko 2015; Hubbard *et al.* 2016; Mencin *et al.* 2016; Qiu *et al.* 2016; Castaldo *et al.* 2017; Wang & Fialko 2018), a relatively small earthquake that shook the Kathmandu region due to the intense seismic activity on the Main Himalayan Thrust that accommodates the shortening between the Indian and the Eurasian plates (Cattin & Avouac 2000; Bollinger *et al.* 2004; Grandin *et al.* 2012; Bollinger *et al.* 2013; Sapkota *et al.* 2013; Stevens & Avouac 2015). The surface deformation was captured by InSAR (Lindsey *et al.* 2015; Qiu *et al.* 2016) and GPS data (Avouac *et al.* 2015; Galetzka *et al.* 2015; Feng *et al.* 2017). These data are not associated with well identified uncertainties and may contain outliers due to contamination of the interferometric phase by tropospheric and ionospheric delays and loss of coherence due to changing scattering properties of the ground caused by seasonal changes and strong shaking from the earthquake near the epicentre (Massonnet & Rabaute 1993; Rosen *et al.* 2000; Jolivet *et al.* 2011; Hu *et al.* 2014; Agram & Simons 2015).

Downloaded from https://academic.oup.com/gji/article-abstract/221/1/334/5674118 by sbarbot@usc.edu on 03 March 2020

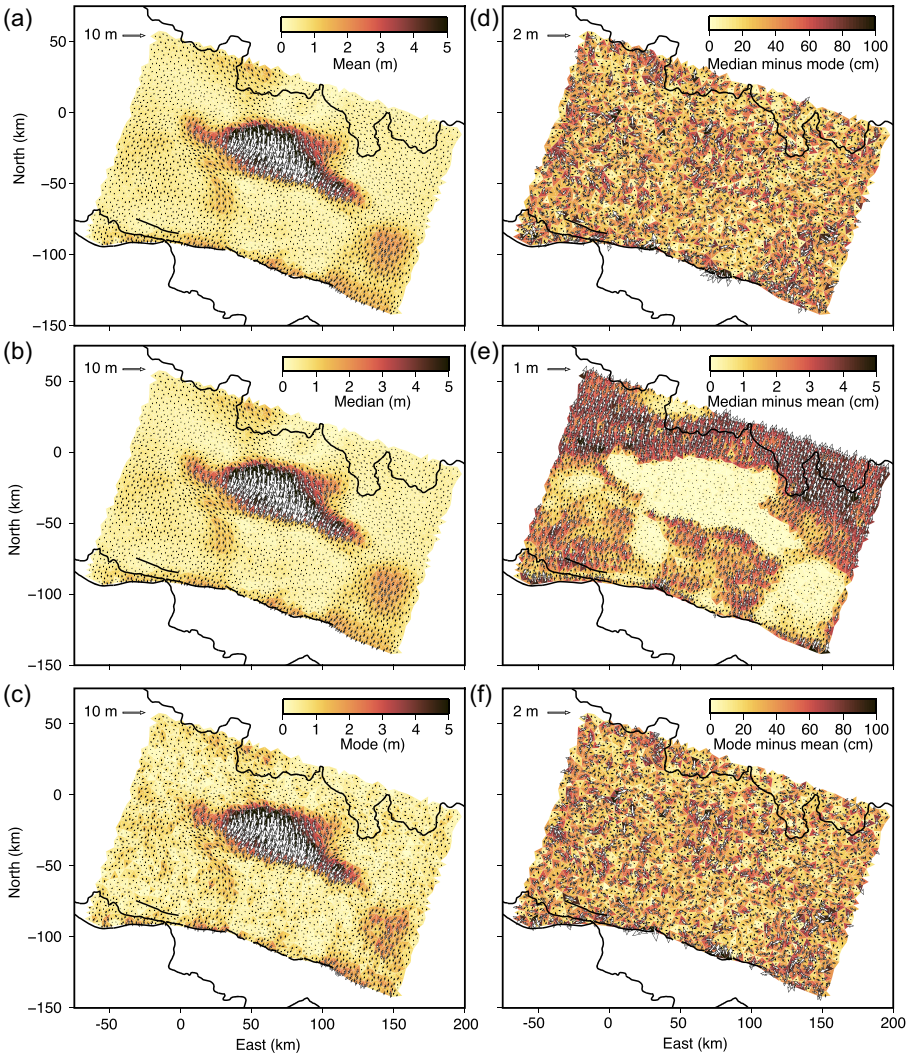

Figure 6. Modes of the slip distribution. (a) Mean of the statistical distribution of fault slip. The arrows represent the vector and the background colour, its norm. (b) Median of the statistical distribution of fault slip. (c) Mode of the statistical distribution of fault slip. (d) Difference between the median and the mode. (e) Difference between the median and the mean. (f) Difference between the mode and the mean.

These deformation data can be reduced assuming that the surface displacements were caused by slip on the fault surface, which can be discretized into 2841 individual triangular fault patches. Slip on the fault has a component in the strike and dip directions (no opening is allowed), giving rise to 5682 unknowns. The surface displacements caused by the earthquake are measured by InSAR, in particular, the Japanese Advanced Land Observing Satellite (ALOS2) L-band SAR satellite (Lindsey *et al.* 2015). We keep 8306 line-of-sight measurements from the interferogram (Fig. 5a) based on noise and distance from the epicentre considerations (Qiu *et al.* 2016; Feng *et al.* 2017). An additional data set includes a sparse GPS network of 18 continuously recording stations (Galetzka *et al.* 2015), providing three-component displacement vectors based on the difference of position before and after the earthquake (Fig. 5a). The absolute uncertainty of these measurements is unknown, but GPS processing provides an estimate of relative errors between the displacement components among all stations.

The distribution of slip on the fault can be smooth or rough depending on the rupture physics, so we only allow as much roughness as required by the data (Shirzaei & Bürgmann 2013; Amey *et al.* 2018; Marchandon *et al.* 2018). We also require a relative uniform direction of fault slip, so we penalize slip in directions orthogonal to the overall dip direction and we strictly forbid slip that deviates more than 45° from the dip direction (Barbot *et al.* 2013). This results in a linear inverse problem with regularization and inequality constraints that require outlier detection with estimation of uncertainties and weights.

Here, we draw 100 000 samples and discard the first 10 000 samples. Our algorithm produces a probability distribution of model parameters and there is an infinite set of solutions, each associated with a probability. From this set, the mode and the first and second moments of the distribution are of particular interest, corresponding to the model producing the best fit to the data (the most likely model), the average model and the model variance. Fig. 5(c) shows the mean model parameters, showing a concentration of fault slip, of the order of 5 m, at the centre of the modelled fault, with slip aligned with the direction of convergence. The residuals between the data and the forward model are shown in Fig. 5(b). The model variance (Fig. 5d) correlates with the amplitude of slip.

Fig. 6 showcases the mean, median, and mode of the posterior distribution and their differences. Overall, these point estimates feature a similar spatial distribution. However, the mode is a less

Downloaded from https://academic.oup.com/gji/article-abstract/221/1/334/5674118 by sbarbot@usc.edu on 03 March 2020

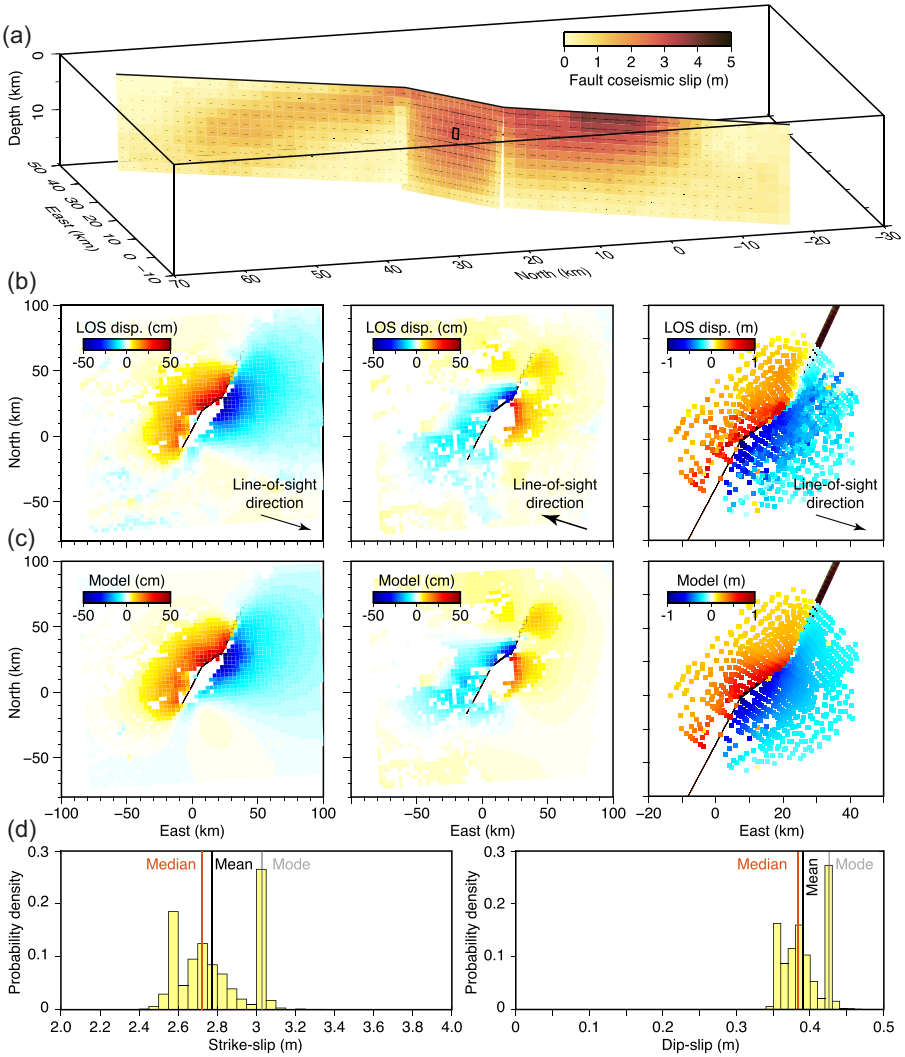

Figure 7. Inversion of heterogeneous data sets (SAR interferograms and SAR amplitude cross-correlation) for coseismic slip. (a) The fault geometry with 700 elementary fault patches, each with two degrees of freedom (strike-slip and dip-slip components of the slip vector) and the recovered slip distribution (mean value of the model distribution). (b) InSAR line-of-sight (LOS) displacements for interferograms 1, 2 and 3. (c) Forward model based on the first moment of fault slip for data sets 1, 2 and 3. (d) Histograms of the strike-slip and dip-slip components with joint mode, mean and median, of the patch situated at the centre of the fault, highlighted by a rectangle in A.

robust estimate, leading to high-wavenumber noise in the mode spatial distribution. The median and mean are similar in regions of high slip, but exhibit systematic differences of the order of 5 cm in regions of low slip. The mode differs from the mean and median by a spatially random noise with a standard deviation of the order of 1 m. Due to the large variance of the mode, applications requiring a smooth distribution of slip may prefer the mean or median estimates.

These results indicate that the method is adequate to tackle the inversion of heterogeneous data sets, providing a detailed description of the data and model statistics in cases with a relatively large parameter space.

5.2 Coseismic slip distribution of the 2015, $M_w = 7.2$ Sarez, Tajikistan earthquake

The December 2015, $M_w = 7.2$, Tajikistan earthquake (Metzger *et al.* 2017; Sangha *et al.* 2017) was the largest one to rupture in the Pamir Plateau since the namesake 1911 Lake Sarez earthquake in 1911 (Ambraseys & Bilham 2012; Kulikova *et al.* 2016). The Pamir Plateau is part of a complex tectonic setting on the western end of India Eurasia collision zone that experienced close to 300 km of shortening since the collision between India and Eurasian in the past 50 Ma (Mohadjer *et al.* 2010; Schurr *et al.* 2014; Kufner *et al.* 2018). The 2015 $M_w = 7.2$ earthquake was a predominantly left-lateral strike-slip rupture that started on the shores of Lake Sarez and propagated unilaterally for approximately 60 km.

Due to the remoteness, the near-field surface displacements associated with fault slip were only captured by remote sensing (Sangha *et al.* 2017). The surface displacements are measurements from InSAR and pixel offsets derived from amplitude cross-correlation (Leprince *et al.* 2007; Wang *et al.* 2014; Wang & Jónsson 2015, Fig. 7b), including two coseismic interferograms with complementary look angles. The first interferogram is derived from a pair of Sentinel1 images acquired in ascending orbit (track 100) in 6 December 2015 and 30 December 2015. The second interferogram is from descending orbit (track 5) from 18 November 2015 and 12 December 2015 acquisitions. The pixel tracking data in the near

Downloaded from https://academic.oup.com/gji/article-abstract/221/1/334/5674118 by sbarbot@usc.edu on 03 March 2020

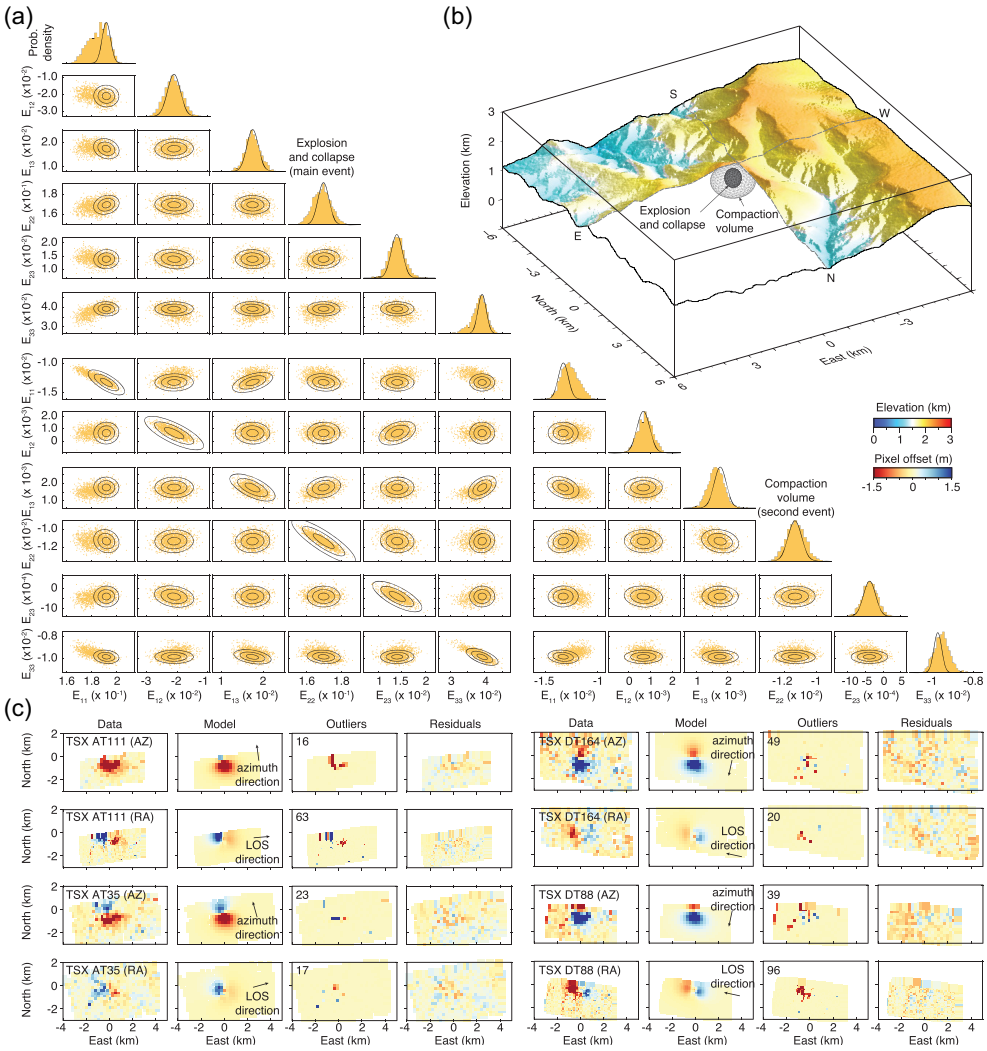

Figure 8. Kinematic inversion of explosion, collapse and compaction source for the 2017 North Korean nuclear explosion at Mount Mantap. (a) Pair-wise distribution (orange dots in off-diagonal figures) and histograms (orange bars in the diagonal figures) of the random samples drawn from the marginal distribution (37). The dashed ellipses (in the off-diagonal figures) represent the 66, 95 and 99 per cent confidence intervals of the pair-wise distribution of the conditional probability density in (42) and the dashed curves (in the diagonal figures) represent its marginal. The marginal and conditional distributions vary slightly. E_{11}, E_{22} and E_{33} represent uniaxial strain in the horizontal and depth directions. E_{12}, E_{13} and E_{23} represent the shear components. (b) Geometry of the explosive and compaction sources. The model parameters for the explosion and collapse do not correlate with each other. The parameters for the compaction do not correlate with each other either. However, the parameters for the explosion and the subsequent collapse correlate due to the similar displacements that they produce. (c) Data, mean model, outliers (the number of outliers is indicated in the upper left-hand corner), and residuals of the optimization based on the mean model. The variance reduction is greater than 80 per cent with outlier detection and only 50 per cent without outlier detection.

field (data set 3) is based on the amplitude images of the first interferogram. The slip model is discretized with 700 rectangular fault patches (Okada 1985, 1992), each with two degrees of freedom (the strike-slip and dip-slip components). The dip-slip component is penalized and the slip vector is bounded by a rake angle of $\pm 45^{\circ}$ from the strike direction. The slip distribution is assumed smooth. The inversion involves heterogeneous data sets, regularization with unknown weights and inequality constraints. All results presented below are based on 80 000 iterations of the Gibbs sampler after 8000 burn-in iterations.

The mean model is shown in Fig. 7(a), showing a coseismic rupture spread out on three different fault segments, running across a major fault bend. The largest slip occurs near the surface. The forward models based on the mean slip model for the data sets 1 and 2, and pixel offsets along the line-of-sight in the near-field from cross-correlation of the amplitude image pair forming interferogram 1 are shown in Fig. 7(c). Because of the inequality constraint, the posterior probability is truncated. We illustrate the posterior marginal distribution for a single patch (Fig. 7d) located at the centre of the fault bend (bold square in Fig. 7a). The distribution indicates the dominance of strike-slip faulting, with the distribution of strike slip and dip slip covering 4 and 0.5 m, respectively. Various point estimates can be extracted from the Gibbs samples, including the joint mean, median and mode of the distribution. Similarly as for the Nepal case (Fig. 6), the mean and median of the joint distribution are similar, but differ significantly from the joint mode.

5.3 Application to the 2017, North Korea nuclear test

We now consider man-made internal deformation to demonstrate the importance of outlier detection. On 3 September 2017, two

Downloaded from https://academic.oup.com/gji/article-abstract/221/1/334/5674118 by sbarbot@usc.edu on 03 March 2020

seismic events separated by about 8.5 min were detected at North Korea's Punggye-ri nuclear test site (Wang *et al.* 2018). Analysis of seismic waveforms revealed a predominantly isotropic explosive source located beneath Mount Mantap for the first event (Liu *et al.* 2018). Because of the large surface displacements in the near field (Fig. 8) the coherence is not sufficient to form radar interferometry. Instead, correlation of the amplitude of space-borne radar images determined up to 3.5 m of divergent horizontal motion with 0.5 m of subsidence (Wang *et al.* 2018). These data, however, are contaminated by the many small-scale landslides triggered by the violent nuclear explosion. This represent a scenario where the data are sufficient to reconstruct the 3-D displacement of the surface pending the successful identification of outliers.

We model the deformation using a pair of cuboidal volume elements (Barbot *et al.* 2017; Barbot 2018), each with six degrees of freedom corresponding to the associated strain components. The first volume element represent the main event. The second one, the smaller event that followed a few minutes later, a few hundred meters to the southeast. In general, strain can be positive or negative, corresponding to extension or compression, or various directions of shear, so we do not impose inequality constraints. The surface deformation data include the pixel-tracking of four pairs of radar amplitude images from the TerraSAR-X satellite of the German Aerospace Agency (DLR). The correlation of each pair provides the displacements in the azimuthal direction of satellite flight and the line-of-sight displacements in the range direction (Michel *et al.* 1999; Hu *et al.* 2014; Wang *et al.* 2014; Wang & Fialko 2015). These data were acquired in ascending and descending modes, providing enough constraints to recover the full vector-valued displacement field.

We applied the proposed algorithm and use 90 000 Gibbs samples after 10 000 burn-ins. The Gibbs samples for the strain components are shown in Fig. 8(a). The joint posterior marginal distribution (eq. 37) indicates that the strain components of each source are weakly correlated to each other. However, there is a negative correlation between the components of the two sources. This is due to the fact that the sources are close to each other and produce similar surface displacements, but with different wavelengths and slightly offset. The analysis confirms the strong horizontal extension of the main source associated with little vertical deformation (Wang *et al.* 2018). The secondary source corresponds to a secondary phase of collapse associated with the breakdown of underground infrastructure. The distribution of surface deformation, model, outliers and residuals are shown in Fig. 8(c). The algorithm detects a large number of outliers, some scattered around the images, some closely distributed (the number of outlier points is indicated in each panel). With outlier detection, the data variance reduction exceeds 80 per cent. Without it, it reaches only 50 per cent.

The conditional posterior probability distribution of eq. (42) is indicated by the ellipses and the dashed curves in Fig. 8(a). Although the marginal and conditional probability density should not match exactly, their close similarity indicates that the closed-form expressions available for the conditional probability density can be a useful tool to describe the statistics of the model parameters in practice, include for propagating uncertainties for further processing and analysis. In particular, the conditional probability captures qualitatively well the mean and correlation between the model parameters. These results provide additional validation of the self-consistency of the method and its potential to help reduce complex, heterogeneous data sets with a large number of outliers.

6 CONCLUSIONS

We have derived a general method to solve the linear inverse problem for multiple data sets possibly contaminated with outliers within the framework of Bayes' theory. The approach allows to determine the weights of *a priori* constraints without intervention from the practitioner, that is, free of tuning. If relevant from the physics of the problem, inequality constraints can be incorporated. The method relies on the Gibbs sampling method, which is a particular flavour of MCMC methods. The prior distribution for the model parameters is uniform within the range allowed by the inequality constraints.

The assumptions regarding the priors and the linearity of the forward model allows us to derive the conditional probability of the random variables in closed form, each associated with well-established random number generators. This allows us to draw samples from the posterior distribution without relying on Metropolis sampling. The approach is therefore only modestly computationally expensive. This is demonstrated by examples of geophysical data inversion involving large data sets and a large model parameter space. By casting the problem as a general linear inverse problem under equality and inequality constraints with outlier detection, we expect the solution method to be of practical interest for a wide range of applications in various fields of science.

ACKNOWLEDGEMENTS

We are grateful for the constructive comments from Romain Jolivet and an anonymous reviewer. The study benefited from funding from the National Science Foundation, under award number EAR-1848192. The study was funded in part by the Singapore Ministry of Education under projects number MOE2018-T2-1-030 and MOE2017-T2-2-126. The computer codes used in this study are available at http://bitbucket.org/sbarbot/outibi. Part of the code is implemented using the Armadillo C++ template library (Sanderson & Curtin 2016, 2019).

REFERENCES

Agram, P. & Simons, M., 2015. A noise model for InSAR time series, *J. geophys. Res.,* **120**(4), 2752–2771.

Ambraseys, N. & Bilham, R., 2012. The Sarez-Pamir earthquake and landslide of 18 February 1911, *Seismol. Res. Lett.,* **83**(2), 294–314.

Amey, R., Hooper, A. & Walters, R., 2018. A Bayesian method for incorporating self-similarity into earthquake slip inversions, *J. geophys. Res.,* **123**(7), 6052–6071.

Amiri-Simkooei, A.R. & Jazaeri, S., 2013. Data-snooping procedure applied to errors-in-variables models, *Stud. Geophys. Geod.,* **57**(3), 426–441.

Aster, R.C., Borchers, B. & Thurber, C.H., 2012. *Parameter Estimation and Inverse Problems,* 2nd edn, Academic Press.

Atzori, S. *et al.*, 2009. Finite fault inversion of DInSAR coseismic displacement of the 2009 L'Aquila earthquake (central Italy), *Geophys. J. Int.,* **36**(15), doi:10.1029/2009GL039293.

Avouac, J.-P., Meng, L., Wei, S., Wang, T. & Ampuero, J.-P., 2015. Lower edge of locked Main Himalayan Thrust unzipped by the 2015 Gorkha earthquake, *Nat. Geosci,* **8,** 708–711.

Baarda, W., 1968. A testing procedure for use in geodetic networks, *Netherlands Geod. Commission,* **2**(5).

Babacan, S.D., Luessi, M., Molina, R. & Katsaggelos, A.K., 2012. Sparse Bayesian methods for low-rank matrix estimation, *IEEE Trans. Signal Process.,* **60**(8), 3964–3977.

Bagnardi, M. & Hooper, A., 2018. Inversion of surface deformation data for rapid estimates of source parameters and uncertainties: a Bayesian approach, *Geochem. Geophys. Geosyst.,* **19**(7), 2194–2211.

Downloaded from https://academic.oup.com/gji/article-abstract/221/1/334/5674118 by sbarbot@usc.edu on 03 March 2020

Bai, Y. *et al.*, 2012. *Computational Methods for Applied Inverse Problems,* Vol. **56,** Walter de Gruyter.

Barbot, S., 2018. Deformation of a half-space from anelastic strain confined in a tetrahedral volume, *Bull. seism. Soc. Am.,* **108**(5A), 2687.

Barbot, S., Hamiel, Y. & Fialko, Y., 2008. Space geodetic investigation of the coseismic and postseismic deformation due to the 2003 Mw 7.2 Altai earthquake: implications for the local lithospheric rheology, *J. geophys. Res.,* **113**(B3), doi:10.1029/2007JB005063.

Barbot, S., Agram, P. & De Michele, M., 2013. Change of apparent segmentation of the San Andreas fault around Parkfield from space geodetic observations across multiple periods, *J. geophys. Res.,* **118**(12), 6311–6327.

Barbot, S., Moore, J.D. & Lambert, V., 2017. Displacement and stress associated with distributed anelastic deformation in a half-space, *Bull. seism. Soc. Am.,* **107**(2), 821–855.

Barnhart, W. & Lohman, R., 2010. Automated fault model discretization for inversions for coseismic slip distributions, *J. geophys. Res.* **115**(B10), doi:10.1029/2010JB007545.

Berger, J.O., 2013. *Statistical Decision Theory and Bayesian Analysis,* Springer Science & Business Media.

Bertero, M. & Boccacci, P., 1998. *Introduction to Inverse Problems in Imaging,* CRC Press.

Bletery, Q., Sladen, A., Delouis, B., Vallée, M., Nocquet, J.-M., Rolland, L. & Jiang, J., 2014. A detailed source model for the Mw 9. 0 Tohoku-Oki earthquake reconciling geodesy, seismology, and tsunami records, *J. geophys. Res.,* **119**(10), 7636–7653.

Bletery, Q., Sladen, A., Jiang, J. & Simons, M., 2016. A bayesian source model for the 2004 great Sumatra-Andaman earthquake, *J. geophys. Res.,* **121**(7), 5116–5135.

Blewitt, G., 2015. GPS and space-based geodetic methods, in *Treatise on Geophysics,* 2nd edn, **Vol. 3,** pp. 307–338, Elsevier.

Bollinger, L., Avouac, J.P., Cattin, R. & Pandey, M.R., 2004. Stress buildup in the Himalaya, *J. geophys. Res.,* **109**(B11405), 8.

Bollinger, L., Klinger, Y., Tapponnier, P., Gaudemer, Y. & Tiwari, D., 2013. Estimating the return times of great Himalayan earthquakes in eastern Nepal: evidence from the Patu and Bardibas strands of the Main Frontal Thrust, *Nat. Geosci.,* **6,** 71–76.

Carlin, B.P. & Louis, T.A., 2008. *Bayesian Methods for Data Analysis,* Chapman and Hall/CRC.

Casella, G., 1985. An introduction to empirical Bayes data analysis, *Am. Stat.,* **39**(2), 83–87.

Casella, G. & George, E.I., 1992. Explaining the Gibbs sampler, *Am. Stat.,* **46**(3), 167–174.

Castaldo, R. *et al.*, 2017. Finite element modelling of the 2015 Gorkha earthquake through the joint exploitation of Dinsar measurements and geologic-structural information, *Tectonophysics,* **714,** 125–132.

Cattin, R. & Avouac, J.P., 2000. Modeling mountain building and the seismic cycle in the Himalaya of Nepal, *J. geophys. Res.,* **105**(B6), 13 389–13 407.

Crosetto, M., Monserrat, O., Cuevas-González, M., Devanthéry, N. & Crippa, B., 2016. Persistent scatterer interferometry: a review, *ISPRS J. Photogram. Remote Sens.,* **115,** 78–89.

Daout, S., Barbot, S., Peltzer, G., Doin, M.-P., Liu, Z. & Jolivet, R., 2016a. Constraining the kinematics of metropolitan Los Angeles faults with a slip-partitioning model, *Geophys. Res. Lett.,* **43**(21), 11–192.

Daout, S. *et al.*, 2016b. Along-strike variations of the partitioning of convergence across the Haiyuan fault system detected by InSAR, *Geophys. Suppl. Mon. Not. R. Astron. Soc.,* **205**(1), 536–547.

Davies, P. & Blewitt, G., 2000. Methodology for global geodetic time series estimation: a new tool for geodynamics, *J. geophys. Res.,* **105**(B5), 11 083–11 100.

Di Traglia, F., Nolesini, T., Intrieri, E., Mugnai, F., Leva, D., Rosi, M. & Casagli, N., 2014. Review of ten years of volcano deformations recorded by the ground-based InSAR monitoring system at Stromboli volcano: a tool to mitigate volcano flank dynamics and intense volcanic activity, *Earth-Sci. Rev.,* **139,** 317–335.

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**(1), 464–482.

Duputel, Z. *et al.*, 2015. The Iquique earthquake sequence of April 2014: Bayesian modeling accounting for prediction uncertainty, *Geophys. Res. Lett.,* **42**(19), 7949–7957.

Elliott, J., Jolivet, R., González, P., Avouac, J.-P., Hollingsworth, J., Searle, M. & Stevens, V., 2016. Himalayan megathrust geometry and relation to topography revealed by the Gorkha earthquake, *Nat. Geosci.,* **9**(2), 174.

Evans, E.L. & Meade, B.J., 2012. Geodetic imaging of coseismic slip and postseismic afterslip: sparsity promoting methods applied to the great Tohoku earthquake, *Geophys. Res. Lett.,* **39**(11), doi:10.1029/2012GL051990.

Feng, W. *et al.*, 2017. Source characteristics of the 2015 Mw 7.8 Gorkha (Nepal) earthquake and its Mw 7.2 aftershock from space geodesy, *Tectonophysics,* **712,** 747–758.

Fialko, Y., Simons, M. & Agnew, D., 2001. The complete (3-D) surface displacement field in the epicentral area of the 1999 M_w 7.1 Hector Mine earthquake, southern California, from space geodetic observations, *Geophys. Res. Lett.,* **28,** 3063–3066.

Figueiredo, M.A. & Nowak, R.D., 2001. Wavelet-based image estimation: an empirical Bayes approach using Jeffrey's noninformative prior, *IEEE Trans. Image Process.,* **10**(9), 1322–1331.

Fukahata, Y. & Wright, T.J., 2008. A non-linear geodetic data inversion using ABIC for slip distribution on a fault with an unknown dip angle, *Geophys. J. Int.,* **173**(2), 353–364.

Funning, G.J., Fukahata, Y., Yagi, Y. & Parsons, B., 2014. A method for the joint inversion of geodetic and seismic waveform data using ABIC: application to the 1997 Manyi, Tibet, earthquake, *J. geophys. Int.,* **196**(3), 1564–1579.

Galetzka, J. *et al.*, 2015. Slip pulse and resonance of the Kathmandu basin during the 2015 Gorkha earthquake, Nepal, *Science,* **349**(6252), 1091–1095.

Geweke, J., 1991. Efficient simulation from the multivariate normal and Student-t distributions subject to linear constraints and the evaluation of constraint probabilities, *Comput. Sci. Statist.,* **23**.

Golub, G.H. & Van Loan, C.F., 1996. *Iterative Methods for Linear Systems,* 3rd edn: Matrix Computations, Johns Hopkins U. Press.

Gombert, B., Duputel, Z., Jolivet, R., Doubre, C., Rivera, L. & Simons, M., 2017. Revisiting the 1992 Landers earthquake: a Bayesian exploration of co-seismic slip and off-fault damage, *Geophys. J. Int.,* **212**(2), 839–852.

Gombert, B., Duputel, Z., Jolivet, R., Simons, M., Jiang, J., Liang, C., Fielding, E. & Rivera, L., 2018. Strain budget of the Ecuador–Colombia subduction zone: a stochastic view, *Earth planet. Sci. Lett.,* **498,** 288–299.

Grandin, R., Doin, M.-P., Bollinger, L., Pinel-Puysségur, B., Ducret, G., Jolivet, R. & Sapkota, S.N., 2012. Long-term growth of the Himalaya inferred from interseismic InSAR measurement, *Geology,* **40** (12), 1059–1062.

Halmos, P.R., 2012. *A Hilbert Space Problem Book,* Vol. **19,** Springer Science & Business Media.

Harris, R.A. & Segall, P., 1987. Detection of a locked zone at depth on the Parkfield, California, segment of the San Andreas Fault, *J. geophys. Res.,* **92**(B8), 7945–7962.

Hu, J., Li, Z., Ding, X., Zhu, J., Zhang, L. & Sun, Q., 2014. Resolving three-dimensional surface displacements from InSAR measurements: a review, *Earth-Sci. Rev.,* **133,** 1–17.

Hubbard, J., Almeida, R., Foster, A., Sapkota, S.N., Bürgi, P. & Tapponnier, P., 2016. Structural segmentation controlled the 2015 Mw 7.8 Gorkha earthquake rupture in Nepal, *Geology,* **44**(8), 639–642.

Ito, K. & Jin, B., 2015. *Inverse Problems: Tikhonov Theory and Algorithms,* World Scientific.

Jeffreys, H., 1939, *Theory of Probability,* Clarendon Press.

Jiang, J. & Simons, M., 2016. Probabilistic imaging of tsunamigenic seafloor deformation during the 2011 Tohoku-Oki earthquake, *J. geophys. Res.,* **121**(12), 9050–9076.

Johanson, I.A., Fielding, E.J., Rolandone, F. & Bürgmann, R., 2006. Coseismic and postseismic slip of the 2004 Parkfield Earthquake from Space-Geodetic data, *Bull. seism. Soc. Am.,* **96**(4B), S269–S282.

Jolivet, R., Grandin, R., Lasserre, C., Doin, M.-P. & Peltzer, G., 2011. Systematic InSAR tropospheric phase delay corrections from

Downloaded from https://academic.oup.com/gji/article-abstract/221/1/334/5674118 by sbarbot@usc.edu on 03 March 2020

global meteorological reanalysis data, *Geophys. Res. Lett.,* **38**(17), doi:10.1029/2011GL048757.

Jolivet, R., Simons, M., Agram, P., Duputel, Z. & Shen, Z.-K., 2015. Aseismic slip and seismogenic coupling along the central San Andreas fault, *Geophys. Res. Lett.,* **42**(2), 297–306.

Kufner, S.-K., Schurr, B., Ratschbacher, L., Murodkulov, S., Abdulhameed, S., Ischuk, A., Metzger, S. & Kakar, N., 2018. Seismotectonics of the Tajik basin and surrounding mountain ranges, *Tectonics,* **37**(8), 2404–2424.

Kulikova, G., Schurr, B., Krüger, F., Brzoska, E. & Heimann, S., 2016. Source parameters of the Sarez-Pamir earthquake of 1911 February 18, *Geophys. J. Int.,* **205**(2), 1086–1098.

Landry, W. & Barbot, S., 2019. Fast, accurate solutions for 3D strain volumes in a heterogeneous half space, *Comput. Geosci,* **125,** 109–114.

Langbein, J., 2008. Noise in GPS displacement measurements from Southern California and Southern Nevada, *J. geophys. Res.,* **113**(B05405), doi:10.1029/2007JB005247.

Lehmann, R., 2013. 3σ-rule for outlier detection from the viewpoint of geodetic adjustment, *J. Surv. Eng.,* **139**(4), 157–165.

Lehmann, R. & Lösler, M., 2016. Multiple outlier detection: hypothesis tests versus model selection by information criteria, *J. Surv. Eng.,* **142**(4), 04016017.

Lehmann, R. & Voß-Böhme, A., 2017. On the statistical power of Baarda's outlier test and some alternative, *J. Geod. Sci.,* **7**(1), 68–78.

Leprince, S., Barbot, S., Ayoub, F. & Avouac, J.P., 2007. Automatic, precise, ortho-rectification and coregistration for satellite image correlation, application to ground deformation measurement, *IEEE Trans. Geosc. Rem. Sens.,* **45**(6), 1529–1558.

Liang, F., Liu, C. & Carroll, R., 2011. *Advanced Markov Chain Monte Carlo Methods: Learning from Past Samples,* Vol. **714,** John Wiley & Sons.

Lindsey, E.O., Natsuaki, R., Xu, X., Shimada, M., Hashimoto, M., Melgar, D. & Sandwell, D.T., 2015. Line-of-sight displacement from ALOS-2 interferometry: Mw 7.8 Gorkha Earthquake and Mw 7.3 aftershock, *Geophys. Res. Lett,* **42** (16), 6655–6661.

Liu, J., Li, L., Zahradník, J., Sokos, E., Liu, C. & Tian, X., 2018. North Korea's 2017 test and its nontectonic aftershock, *Geophys. Res. Lett.,* **45**(7), 3017–3025 .

Lorito, S. *et al.*, 2011. Limited overlap between the seismic gap and coseismic slip of the great 2010 Chile earthquake, *Nat. Geosci.,* **4,** 173–177.

Marchandon, M., Vergnolle, M., Sudhaus, H. & Cavalié, O., 2018. Fault geometry and slip distribution at depth of the 1997 Mw 7.2 Zirkuh earthquake: contribution of near-field displacement data, *J. geophys. Res.,* **123**(2), 1904–1924.

Massonnet, D. & Rabaute, T., 1993. Radar interferometry: limits and potential, *IEEE Trans. Geosci. Rem. Sens.,* **31**(2), 455–464.

Mencin, D. *et al.*, 2016. Himalayan strain reservoir inferred from limited afterslip following the gorkha earthquake, *Nat. Geosci.,* **9,** 533–537

Metzger, S. *et al.*, 2017. The 2015 Mw 7.2 Sarez strike-slip earthquake in the Pamir interior: response to the underthrusting of India's western promontory, *Tectonics,* **36**(11), 2407–2421.

Michel, R., Avouac, J.-P. & Taboury, J., 1999. Measuring ground displacements from SAR amplitude images: application to the Landers earthquake, *Geophys. Res. Lett.,* **26**(7), 875–878.

Milliner, C., Materna, K., Bürgmann, R., Fu, Y., Moore, A.W., Bekaert, D., Adhikari, S. & Argus, D.F., 2018. Tracking the weight of hurricane Harvey's stormwater using GPS data, *Sci. Adv.,* **4**(9), eaau2477, doi:10.1126/sciadv.aau2477.

Minson, S., Simons, M. & Beck, J., 2013. Bayesian inversion for finite fault earthquake source models I—theory and algorithm, *J. geophys. Int.,* **194**(3), 1701–1726.

Minson, S. *et al.*, 2014. Bayesian inversion for finite fault earthquake source models–II: the 2011 Great Tohoku-Oki, Japan earthquake, *J. geophys. Int.,* **198**(2), 922–940.

Mohadjer, S. *et al.*, 2010. Partitioning of India-Eurasia convergence in the Pamir-Hindu Kush from GPS measurements, *Geophys. Res. Lett.,* **37**(4), doi:10.1029/2009GL041737.

Moore, J.D. *et al.*, 2017. Imaging the distribution of transient viscosity after the 2016 Mw 7.1 Kumamoto earthquake, *Science,* **356**(6334), 163–167.

Nocquet, J.-M., 2018. Stochastic static fault slip inversion from geodetic data with non-negativity and bound constraints, *J. geophys. Int.,* **214**(1), 366–385.

Nocquet, J.-M. & Calais, E., 2004. Geodetic measurements of crustal deformation in the western Mediterranean and Europe, *Pure Appl. Geophys.,* **161**(3), 661–681.

Okada, Y., 1985. Surface deformation due to shear and tensile faults in a half-space, *Bull. seism. Soc. Am.,* **75**(4), 1135–1154.

Okada, Y., 1992. Internal deformation due to shear and tensile faults in a half-space, *Bull. seism. Soc. Am.,* **82,** 1018–1040.

Parker, R.L., 1994. *Geophysical Inverse Theory,* Princeton Univ. Press.

Prószyński, W., 1997. Measuring the robustness potential of the least-squares estimation: geodetic illustration, *J. Geod.,* **71**(10), 652–659.

Qiu, Q. *et al.*, 2016. The mechanism of partial rupture of a locked megathrust: the role of fault morphology, *Geology,* **44**(10), 875–878.

Qiu, Q., Moore, J.D.P., Barbot, S., Feng, L. & Hill, E., 2018. Transient viscosity in the Sumatran mantle wedge from a decade of geodetic observations, *Nat. Commun.,* **9**(995), doi:10.1038/s41467-018-03298-6.

Raftery, A.E. & Lewis, S., 1991. How many iterations in the Gibbs sampler?, Tech. rep., Dept of Statistics, Washington Univ. Seattle.

Rangelova, E., Fotopoulos, G. & Sideris, M., 2009. On the use of iterative re-weighting least-squares and outlier detection for empirically modelling rates of vertical displacement, *J. Geod.,* **83**(6), 523–535.

Rolandone, F. *et al.*, 2008. Aseismic slip and fault-normal strain along the central creeping section of the San Andreas fault, *Geophys. Res. Lett.,* **35**(L14305), 5.

Rosen, P.A., Hensley, S., Joughin, I.R., Li, F.K., Madsen, S.N., Rodriguez, E. & Goldstein, R.M., 2000. Synthetic aperture radar interferometry, *Proc. IEEE,* **80**(3), 333–382.

Rue, H. & Held, L., 2005. *Gaussian Markov Random Fields: Theory and Applications,* Chapman and Hall/CRC.

Sanderson, C. & Curtin, R., 2016. Armadillo: a template-based C++ library for linear algebra, *J. Open Source Software,* **1**(2), 26.

Sanderson, C. & Curtin, R., 2019. Practical sparse matrices in C++ with hybrid storage and template-based expression optimisation, *Math. Comput. Appl.,* **24**(3), 70.

Sangha, S., Peltzer, G., Zhang, A., Meng, L., Liang, C., Lundgren, P. & Fielding, E., 2017. Fault geometry of 2015, Mw 7.2 Murghab, Tajikistan earthquake controls rupture propagation: insights from InSAR and seismological data, *Earth planet. Sci. Lett.,* **462,** 132–141.

Santilli, R.M., 2013. *Foundations of Theoretical Mechanics I: The Inverse Problem in Newtonian Mechanics,* Springer Science & Business Media.

Sapkota, S., Bollinger, L., Klinger, Y., Tapponnier, P., Gaudemer, Y. & Tiwari, D., 2013. Primary surface ruptures of the Great Himalayan earthquakes in 1934 and 1255, *Nat. Geosci.,* **6**(1), 71–76.

Sarvas, J., 1987. Basic mathematical and electromagnetic concepts of the biomagnetic inverse problem, *Phys. Med. Biol.,* **32**(1), 11.

Sathiakumar, S., Barbot, S.D. & Agram, P., 2017. Extending resolution of fault slip with geodetic networks through optimal network design, *J. geophys. Res.,* **122**(12), 10 538–10 558.

Schafer, C.M., Stark, P.B., Evans, S. & Hansen, B., 2003. Using what we know: inference with physical constraints, in *Proceedings of the PHYSTAT2003, Statistical Problems in Particle Physics, Astrophysics, and Cosmology (SLAC),* 8–11 September 2003, SLAC, Stanford, CA, pp. 25–34.

Schurr, B., Ratschbacher, L., Sippl, C., Gloaguen, R., Yuan, X. & Mechie, J., 2014. Seismotectonics of the pamir, *Tectonics,* **33**(8), 1501–1518.

Segall, P. & Matthews, M., 1988. Displacement calculations from geodetic data and the testing of geophysical deformation models, *J. geophys. Res.,* **93**(B12), 14 954–14 966.

Segall, P. & Matthews, M., 1997. Time dependent inversion of geodetic data, *J. geophys. Res.,* **102**(B10), 22 391–22 409.

Shen, C., Brooks, M.J. & Van Den Hengel, A., 2007. Fast global kernel density mode seeking: applications to localization and tracking, *IEEE Trans. Image Process.,* **16**(5), 1457–1469.

Shirzaei, M. & Bürgmann, R., 2013. Time-dependent model of creep on the hayward fault from joint inversion of 18 years of InSAR and surface creep data, *J. geophys. Res.,* **118**(4), 1733–1746.

Downloaded from https://academic.oup.com/gji/article-abstract/221/1/334/5674118 by sbarbot@usc.edu on 03 March 2020

Silverii, F., D'Agostino, N., Borsa, A.A., Calcaterra, S., Gambino, P., Giuliani, R. & Mattone, M., 2019. Transient crustal deformation from Karst aquifers hydrology in the Apennines (Italy), *Earth planet. Sci. Lett.,* **506,** 23–37.

Silverman, B.W., 1981. Using kernel density estimates to investigate multimodality, *J. R. Stat. Soc., B,* **43**(1), 97–99.

Simons, M. *et al.*, 2011. The 2011 magnitude 9.0 Tohoku-Oki earthquake: mosaicking the megathrust from seconds to centuries, *Science,* **332**(6036), 1421–1425.

Snow, K.B. & Schaffrin, B., 2003. Three-dimensional outlier detection for GPS networks and their densification via the Blimpbe approach, *GPS Solut.,* **7**(2), 130–139.

Stark, P.B. & Parker, R.L., 1995. Bounded-variable least-squares: an algorithm and applications, *Comput. Stat.,* **10,** 129–129.

Stevens, V. & Avouac, J., 2015. Interseismic coupling on the main Himalayan thrust, *Geophys. Res. Lett.,* **42**(14), 5828–5837 .

Tang, C.-H., Hsu, Y.-J., Barbot, S., Moore, J. D.P. & Chang, W.-L., 2019. Lower-crustal rheology and thermal gradient in the Taiwan orogenic belt illuminated by the 1999 Chi-Chi earthquake, *Sci. Adv.,* **5** (2), doi:10.1126/sciadv.aav3287.

Tarantola, A., 2005. *Inverse Problem Theory and Methods for Model Parameter Estimation,* Soc. Ind. App. Math.

Tarantola, A. & Valette, B., 1982. Generalized nonlinear inverse problems solved using the least squares criterion, *Rev. Geophys.,* **20**(2), 219–232.

Tipping, M.E., 2001. Sparse bayesian learning and the relevance vector machine, *J. Mach. Learn. Res.,* **1**(Jun), 211–244.

Tronin, A., 2010. Satellite remote sensing in seismology. A review, *Remote Sens.,* **2**(1), 124–150.

Tsang, L.L., Hill, E.M., Barbot, S., Qiu, Q., Feng, L., Hermawan, I., Banerjee, P. & Natawidjaja, D.H., 2016. Afterslip following the 2007 Mw 8.4 Bengkulu earthquake in Sumatra loaded the 2010 Mw 7.8 Mentawai tsunami earthquake rupture zone, *J. geophys. Res.,* **121**(12), 9034–9049.

van den Berg, E. & Friedlander, M.P., 2008. Probing the pareto frontier for basis pursuit solutions, *SIAM J. Scient. Comput.,* **31**(2), 890–912.

Wadehn, F., Bruderer, L., Dauwels, J., Sahdeva, V., Yu, H. & Loeliger, H.-A., 2016. Outlier-insensitive kalman smoothing and marginal message passing, in *2016 24th European Signal Processing Conference (EUSIPCO),* pp. 1242–1246, IEEE.

Wang, C., Ding, X., Li, Q., Shan, X., Zhu, J., Guo, B. & Liu, P., 2016a. Adaptive regularization of earthquake slip distribution inversion, *Tectonophysics,* **675,** 181–195.

Wang, H., Yu, H., Hoy, M., Dauwels, J. & Wang, H., 2016b. Variational Bayesian dynamic compressive sensing, in *Proceedings of the 2016 IEEE International Symposium on Information Theory (ISIT),* pp. 1421–1425, IEEE.

Wang, K. & Fialko, Y., 2015. Slip model of the 2015 Mw 7.8 Gorkha (Nepal) earthquake from inversions of ALOS-2 and GPS data, *Geophys. Res. Lett.,* **42** (18), 7452–7458.

Wang, K. & Fialko, Y., 2018. Observations and modeling of coseismic and postseismic deformation due to the 2015 Mw 7.8 Gorkha (Nepal) earthquake, *J. geophys. Res.,* **123**(1), 761–779.

Wang, T. & Jónsson, S., 2015. Improved sar amplitude image offset measurements for deriving three-dimensional coseismic displacements, *IEEE J. Select. Topics Appl. Earth Observ. Remote Sens.,* **8**(7), 3271–3278.

Wang, T., Jónsson, S. & Hanssen, R.F., 2014. Improved sar image coregistration using pixel-offset series, *IEEE Trans. Geosc. Rem. Sens.,* **11**(9), 1465–1469.

Wang, T. *et al.*, 2018. The rise, collapse, and compaction of Mt. Mantap from the 3 September 2017 North Korean nuclear test, *Science,* **361** (6398), 166–170.

Wunsch, C., 1996. *The Ocean Circulation Inverse Problem,* Cambridge Univ. Press.

Yabuki, T. & Matsu'Ura, M., 1992. Geodetic data inversion using a bayesian information criterion for spatial distribution of fault slip, *J. geophys. Int.,* **109**(2), 363–375.

Yu, H. & Dauwels, J., 2015. Variational bayes learning of graphical models with hidden variables, in *Proceedings of the 2015 IEEE 25th International Workshop on Machine Learning for Signal Processing (MLSP),* pp. 1–6, IEEE.

Yu, H. & Dauwels, J., 2016. Modeling spatio-temporal extreme events using graphical models, *IEEE Trans. Signal Process.,* **64**(5), 1101–1116.

Yu, H., Dauwels, J. & Jonathan, P., 2014. Extreme-value graphical models with multiple covariates, *IEEE Trans. Signal Process.,* **62**(21), 5734–5747.

Downloaded from https://academic.oup.com/gji/article-abstract/221/1/334/5674118 by sbarbot@usc.edu on 03 March 2020