Geophys. J. Int. (2008) **173,** 25–48 doi: 10.1111/j.1365-246X.2007.03722.x

Geodetic inversion for space–time distribution of fault slip with time-varying smoothing regularization

Jun'ichi Fukuda,[1,*] Shin'ichi Miyazaki,[1] Tomoyuki Higuchi[2] and Teruyuki Kato[1]

[1] *Earthquake Research Institute, University of Tokyo, Yayoi* 1-1-1, *Bunkyo-ku, Tokyo* 113-0032, *Japan. E-mail: jfukuda@indiana.edu*
[2] *The Institute of Statistical Mathematics,* 4-6-7 *Minami-Azabu, Minato-ku, Tokyo* 106-8569, *Japan*

Accepted 2007 December 29. Received 2007 December 28; in original form 2007 March 10

SUMMARY

We have developed a new geodetic inversion method for space–time distribution of fault slip velocity with time-varying smoothing regularization in order to reconstruct accurate time histories of aseismic fault slip transients. We introduce a temporal smoothing regularization on slip and slip velocity through a Bayesian state space approach in which the strength of regularization (temporal smoothness of slip velocity) is controlled by a hyperparameter. The time-varying smoothing regularization is realized by treating the hyperparameter as a time-dependent stochastic variable and adopting a hierarchical Bayesian state space model, in which a prior distribution on the hyperparameter is introduced in addition to a conventional Bayesian state space model. We have tested this inversion method on two synthetic data sets generated by simulated aseismic slip transients. Results show that our method reproduces well both rapid changes of slip velocity and steady-state velocity without significant oversmoothing and undersmoothing, which has been hard to overcome by the conventional Bayesian approach with time-independent smoothing regularization. Application of this method to transient deformation in 2002 caused by a silent earthquake off the Boso peninsula, Japan, also shows similar advantages of this method over the conventional approach.

Key words: Time series analysis; Inverse theory; Probability distributions; Transient deformation.

1 INTRODUCTION

In recent years, arrays of continuously operated Global Positioning System (GPS) receivers have recorded transient crustal deformations resulting from aseismic fault slip events. For example, transient deformation caused by earthquake afterslip has widely been observed by GPS arrays (e.g. Heki *et al.* 1997; Miyazaki *et al.* 2004). In addition, silent earthquakes (also called slow slip events), which are spontaneous aseismic slip events, have been inferred from GPS data in central Japan (e.g. Ozawa *et al.* 2002), southwest Japan (e.g. Hirose *et al.* 1999), Cascadia (e.g. Dragert *et al.* 2001) and southern Mexico (e.g. Kostoglodov *et al.* 2003), and space–time fault slip inversions have been applied to some of these events.

Imaging time evolution of aseismic slip events is important for inferring spatiotemporal relationships between aseismic slip and seismicity. It has been demonstrated that slow slip events in some subduction zones are correlated with non-volcanic tremors (Rogers & Dragert 2003; Obara *et al.* 2004). Slow slip events on the San Andreas fault and on Kilauea volcano are associated with increased levels of microseismicity (Linde *et al.* 1996; Segall *et al.* 2006). Furthermore, it has been suggested that aseismic slip events may trigger relatively large earthquakes (Murakami *et al.* 2006; Pritchard & Simons 2006). Stressing due to afterslip can also contribute to the temporal evolution of aftershocks (e.g. Schaff *et al.* 1998; Perfettini & Avouac 2004; Perfettini *et al.* 2005). In order to investigate causal relationships between aseismic slip and seismicity, we require detailed images of the time evolution of aseismic slip. Finally, the time evolution of aseismic slip velocity can be used to infer fault frictional properties (Miyazaki *et al.* 2004, 2006a; Hsu *et al.* 2006).

Segall & Matthews (1997) presented a space–time fault slip inversion method, the Network Inversion Filter (NIF), based on the state space (Kalman filtering) method (e.g. Kalman 1960; Kitagawa & Gersch 1996), one of the Bayesian sequential estimation methods (e.g. Kitagawa 1984, 1987; Kitagawa & Gersch 1996). In their method, a temporal smoothness prior is employed to regularize the inversion. Specifically,

*Now at: Department of Geological Sciences, Indiana University, 1001 East 10th Street, Bloomington, IN 47405-1405, USA.

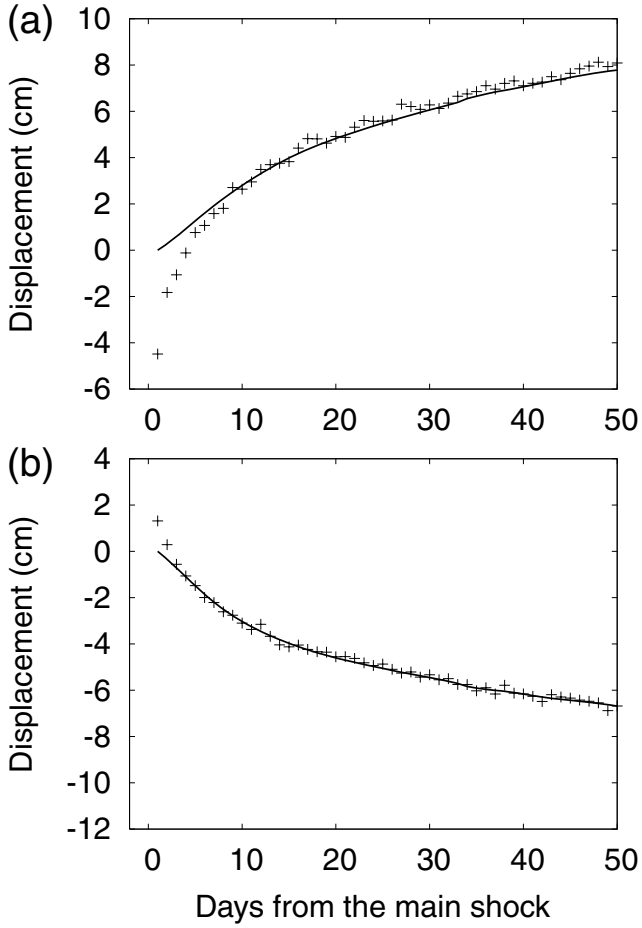

Figure 1. GPS time-series (pluses) at Erimo-1 (0019) (Miyazaki *et al.* 2006b). Solid lines show predicted displacements from the afterslip distribution obtained by the NIF (Miyazaki *et al.* 2006b). (a) East component, (b) North component.

slip acceleration is assumed to be a Gaussian white noise process whose variance governs the temporal smoothness of slip. Thus the variance is often called temporal smoothing parameter. It is also called a hyperparameter in the Bayesian modelling framework. This hyperparameter is estimated by maximizing the likelihood (e.g. Kitagawa & Gersch 1996; Segall & Matthews 1997). The NIF and its modified version, the Extended Network Inversion Filter (ENIF) (McGuire & Segall 2003), have been applied to earthquake afterslip (e.g. Segall *et al.* 2000; Bürgmann *et al.* 2001; Miyazaki *et al.* 2004; Hsu *et al.* 2006) and to slow slip events (e.g. Ozawa *et al.* 2001, 2002; McGuire & Segall 2003; Miyazaki *et al.* 2003; Murray & Segall 2005).

In the NIF and the ENIF frameworks, the hyperparameter (temporal smoothing parameter) is assumed to be constant over time. Under this assumption, rapid changes of slip velocity may be oversmoothed because the hyperparameter is too small (i.e. temporal smoothing regularization is too strong) to trace such rapid changes. On the other hand, the estimated slip velocity during periods of steady slip may be too oscillatory because the hyperparameter is too large (i.e. temporal smoothing regularization is too weak). Therefore, it may be difficult to image, for example, the nucleation of slow slip events or rapid decay of slip velocity in the very early stage of post-seismic slip. For example, Hsu *et al.* (2006) applied the ENIF to GPS time-series after the 2005 Nias-Simeulue earthquake to estimate time evolution of afterslip. They showed that fit of the estimated afterslip to the GPS time-series is poor for a short period immediately after the main shock (see fig. S4 of Hsu *et al.* 2006). Miyazaki *et al.* (2006b) employed the NIF to image afterslip for 7 months following the 2003 Tokachi-oki earthquake in Japan. Fig. 1 shows comparison of observed GPS time-series and predicted displacements from their NIF inversion for 50 d following the earthquake at a GPS station close to their estimated afterslip area. From Fig. 1, it is clear that their afterslip model cannot fit the rapid displacements immediately after the earthquake. We infer that these poor fits are due to oversmoothing of rapid afterslip decay.

In order to overcome this difficulty, Fukuda *et al.* (2004) developed an inversion method for temporal variation of fault slip with time-varying smoothing regularization by introducing a temporal change of the hyperparameter. They applied this method to synthetic data sets generated by transient aseismic slips and compared the inverted slip histories with those obtained by the conventional inversion method. They showed that the slip histories estimated by their inversion method did not suffer from significant oversmoothing and undersmoothing of slip in the temporal domain, whereas those obtained by the conventional inversion method with a constant hyperparameter were oversmoothed and undersmoothed. Although their method has such an advantage, fault slip was assumed to be spatially uniform. It is necessary to relax this assumption in order to obtain detailed images of spatiotemporal evolution of aseismic slip events.

In this study, we extend the work of Fukuda *et al.* (2004) to include spatial variability of fault slip and slip velocity. Following Fukuda *et al.* (2004), we introduce a time-varying hyperparameter for temporal smoothness of slip velocity to avoid the oversmoothing/undersmoothing problem. We combine the hierarchical Bayesian approach (e.g. MacKay 1992) with state space Kalman filtering (e.g. Kitagawa & Gersch 1996; Segall & Matthews 1997) to estimate the temporal variation of the hyperparameter. We call this approach hierarchical Bayesian state space approach. This approach treats the hyperparameter as a stochastic variable and estimates the posterior distribution of the hyperparameter in the Bayesian framework. In the following section, we present the new geodetic inversion method for space–time distribution of fault slip. In Section 3, the method is applied to simulated data sets and results are compared with the conventional method that assumes a constant hyperparameter. Finally, in Section 4, we apply the inversion method to a slow slip event off the Boso peninsula, central Japan, and demonstrate the advantages over the conventional method.

2 INVERSION METHOD

2.1 Statistical model

We use a hierarchical Bayesian state space model to estimate space–time distribution of fault slip and slip velocity with a time-varying hyperparameter. This method consists of two models: a state space model and a stochastic model of the hyperparameter evolution. The state space model is further comprised of an observation model and a system model. The observation model relates fault slip at depth to observed surface displacements and the system model describes the temporal smoothness prior on the fault slip. The observation model and the system model are formulated in Sections 2.1.1 and 2.1.2, respectively. The stochastic model of the hyperparameter evolution specifies the prior distribution of the hyperparameter and is presented in Section 2.1.3. Then we briefly describe the sequential estimation method for evolution of the hyperparameter, slip and slip velocity in Section 2.2.

2.1.1 Observation model

We consider displacement discontinuity, or slip $\mathbf{s}(\boldsymbol{\xi}, t)$, at a point $\boldsymbol{\xi}$ and at a time t, on a fault surface $A(\boldsymbol{\xi})$ embedded in a homogeneous, isotropic, elastic half-space. Let $\mathbf{u}(\mathbf{r}, t)$ be the cumulative surface displacement as a function of station position $\mathbf{r}$ and time of measurement t. We model the surface displacement at $\mathbf{r} = \mathbf{r}_j$ and $t = t_n$, $\mathbf{u}(\mathbf{r}_j, t_n)$, by

$$u_i(\mathbf{r}_j, t_n) = \sum_{k=1}^{2} \int_A G_{ik}(\mathbf{r}_j, \boldsymbol{\xi}) s_k(\boldsymbol{\xi}, t_n) d\,A(\boldsymbol{\xi}) + \mathcal{L}_i(\mathbf{r}_j, t_n) + f_i(t_n) + e_{ijn}, \tag{1}$$

where the left-hand side is the ith component of $\mathbf{u}(\mathbf{r}_j, t_n)$. The first term on the right-hand side of (1) represents the surface displacement due to cumulative slip $\mathbf{s}(\boldsymbol{\xi}, t)$ on the fault surface $A(\boldsymbol{\xi})$, where $G_{ik}(\mathbf{r}, \boldsymbol{\xi})$ is the elastostatic Green's function for dislocation (e.g. Okada 1985). The summation over the index k shows contributions from strike and dip slip components. The second term $\mathcal{L}_i(\mathbf{r}_j, t_n)$ represents coloured noise due to non-tectonic random benchmark motions (e.g. Langbein & Johnson 1997). Segall & Matthews (1997) modelled coloured noise as random benchmark motions which were assumed to follow Brownian random walk processes and estimated random benchmark motions together with slip distribution in the NIF. They showed that this approach can distinguish fault slip-induced displacements from coloured noise. Further, Fukuda *et al.* (2004) demonstrated that this approach also works well for the fault slip inversion with time-varying smoothing regularization. Therefore, we incorporate random benchmark motions into our model. The third term on the right-hand side of (1), $f_i(t_n)$, represents errors that are common to all stations, the common-mode-errors, which are well known to be contained in GPS time-series (e.g. Wdowinski *et al.* 1997; Miyazaki *et al.* 2003). The final term of (1), e_{ijn}, represents the observation error that is assumed to follow a Gaussian distribution with zero mean and covariance $\sigma^2\Sigma_n$, where Σ_n is a data covariance matrix and σ^2 is an unknown scale factor. In the case of GPS data, Σ_n is a covariance matrix of GPS site coordinates derived from GPS data processing. Because the GPS processing software does not fully model all error sources such as multipath or azimuthally varying path delays, the scale factor σ^2 is introduced to account for these error sources.

We divide the fault surface $A(\boldsymbol{\xi})$ into M_f small subfaults and assume that slip is spatially uniform in each subfault. Because estimating the spatial distribution of slip is an ill-posed inverse problem, solutions may be extremely oscillatory due to measurement errors without any constraints on the spatial slip distribution. In order to obtain a slip distribution that is spatially smooth to some degree, spatial smoothing constraints are necessary. Here we adopt the method developed by Segall *et al.* (2000) in which spatial smoothing is implemented by representing the slip distribution with spatial basis functions. Following Segall *et al.* (2000), the slip distribution at time t_n, $\mathbf{a}_n$, is expanded in some spatial basis set $\mathbf{B}$,

$$\mathbf{a}_n = \mathbf{B}\mathbf{c}_n, \tag{2}$$

where $\mathbf{B}$ is a $2M_f \times M$ matrix, $\mathbf{c}_n$ is a M dimensional vector of time-varying coefficients, and M is the number of basis functions. The spatial basis functions $\mathbf{B}$ are given in (A15). These spatial basis functions are based on the singular value decomposition (SVD) solution to the fault slip inversion with spatial smoothing constraints (Harris & Segall 1987) and enforces spatial smoothing on the slip distribution. The degree of smoothness is controlled by the number of basis functions, M (see A16 and Segall *et al.* 2000). Derivation of the basis functions is summarized in Appendix A.

Using the basis function representation of the slip distribution, the observation model (1) can be written in state space form as follows:

$$\mathbf{y}_n = \mathbf{H}_n \mathbf{x}_n + \mathbf{w}_n, \qquad \mathbf{w}_n \sim N(\mathbf{0}, \mathbf{R}_n). \tag{3}$$

Here $\mathbf{x}_n$ is an unknown state vector that contains $\mathbf{c}_n$, $\dot{\mathbf{c}}_n$, $\mathcal{L}_i(\mathbf{r}_j, t_n)$ and $f_i(t_n)$, $\mathbf{y}_n$ is a data vector, $\mathbf{H}_n$ is a known matrix, and $\mathbf{R}_n$ is defined by $\mathbf{R}_n = \sigma^2 \Sigma_n$. The subscript n corresponds to time t_n. An explicit expression of the observation model (3) is given in Appendix A.

2.1.2 System model

Next, we introduce a system model that describes the temporal smoothness prior of the fault slip. We assume slip velocity follows a Brownian random walk process (i.e. slip velocity is locally nearly constant in time) following Segall & Matthews (1997). For the kth spatial basis functions (the kth column of $\mathbf{B}$), temporal evolution of slip is determined by the kth component of the time-varying coefficient $\mathbf{c}_n$, $c_k(t_n)$, and its first derivative with respect to time, $\dot{c}_k(t_n)$. Temporal variation of $c_k(t_n)$ is expressed as follows:

$$\begin{bmatrix} c_k(t_n) \\ \dot{c}_k(t_n) \end{bmatrix} = \begin{bmatrix} 1 & \Delta t_n \\ 0 & 1 \end{bmatrix} \begin{bmatrix} c_k(t_{n-1}) \\ \dot{c}_k(t_{n-1}) \end{bmatrix} + \mathbf{v}_n^{(k)}, \quad \mathbf{v}_n^{(k)} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \alpha_n^2 \begin{bmatrix} \Delta t_n^3/3 & \Delta t_n^2/2 \\ \Delta t_n^2/2 & \Delta t_n \end{bmatrix} \right). \tag{4}$$

Here Δt_n is defined as $t_n - t_{n-1}$, $\mathbf{v}_n^{(k)}$ is a 2-D vector of system noise that represents deviation from the constant slip velocity, and α_n^2 is a time-varying hyperparameter that controls the magnitude of $\mathbf{v}_n^{(k)}$.

The random benchmark motions are assumed to follow Brownian random walk processes with scale parameter τ (unit length/time$^{1/2}$) (e.g. Langbein & Johnson 1997) following Segall & Matthews (1997):

$$\mathcal{L}_i(\mathbf{r}_j, t_n) = \mathcal{L}_i(\mathbf{r}_j, t_{n-1}) + \epsilon_{ijn}, \quad \epsilon_{ijn} \sim N(0, \tau^2 \Delta t_n), \tag{5}$$

We assume that τ has the same value for all components of all sites.

The common mode errors [third term in the right-hand side of (1)] are assumed to follow a Gaussian white noise process with system noise variance τ_f^2:

$$f_i(t_n) = \eta_{in}, \quad \eta_{in} \sim N(0, \tau_f^2). \tag{6}$$

The variance τ_f^2 is treated as a known constant in this study.

Combining (4), (5) and (6) yields the state space form of the system model:

$$\mathbf{x}_n = \mathbf{F}_n \mathbf{x}_{n-1} + \mathbf{v}_n, \quad \mathbf{v}_n \sim N(\mathbf{0}, \mathbf{Q}_n), \tag{7}$$

where

$$\mathbf{v}_n = \left[\mathbf{v}_n^{(1)T}, \mathbf{v}_n^{(2)T}, \ldots, \mathbf{v}_n^{(M)T}, \epsilon_{11n}, \epsilon_{21n}, \epsilon_{31n}, \ldots, \epsilon_{1N_s n}, \epsilon_{2N_s n}, \epsilon_{3N_s n}, \eta_{1n}, \eta_{2n}, \eta_{3n} \right]^T \tag{8}$$

$$\mathbf{F}_n = \begin{bmatrix} \mathbf{F}_n^W & \ldots & \mathbf{O} & \mathbf{O} & \mathbf{O} \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ \mathbf{O} & \ldots & \mathbf{F}_n^W & \mathbf{O} & \mathbf{O} \\ \mathbf{O} & \ldots & \mathbf{O} & \mathbf{I}_{3N_s \times 3N_s} & \mathbf{O} \\ \mathbf{O} & \ldots & \mathbf{O} & \mathbf{O} & \mathbf{O}_{3\times 3} \end{bmatrix} \tag{9}$$

$$\mathbf{Q}_n = \begin{bmatrix} \alpha_n^2 \mathbf{Q}_n^W & \ldots & \mathbf{O} & \mathbf{O} & \mathbf{O} \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ \mathbf{O} & \ldots & \alpha_n^2 \mathbf{Q}_n^W & \mathbf{O} & \mathbf{O} \\ \mathbf{O} & \ldots & \mathbf{O} & \tau^2 \Delta t_n \mathbf{I}_{3N_s \times 3N_s} & \mathbf{O} \\ \mathbf{O} & \ldots & \mathbf{O} & \mathbf{O} & \tau_f^2 \mathbf{I}_{3\times 3} \end{bmatrix}. \tag{10}$$

Here $\mathbf{F}_n^W$ and $\mathbf{Q}_n^W$ are given by

$$\mathbf{F}_n^W = \begin{bmatrix} 1 & \Delta t_n \\ 0 & 1 \end{bmatrix} \tag{11}$$

and

$$\mathbf{Q}_n^W = \begin{bmatrix} \Delta t_n^3/3 & \Delta t_n^2/2 \\ \Delta t_n^2/2 & \Delta t_n \end{bmatrix}, \tag{12}$$

$\mathbf{I}_{3N_s \times 3N_s}$ and $\mathbf{I}_{3\times 3}$ are $3N_s \times 3N_s$ and 3×3 identity matrices, respectively, and N_s is the number of observation sites.

2.1.3 Stochastic model of hyperparameter evolution

To allow temporal variations of the hyperparameter α_n, we treat the hyperparameter as a time-varying stochastic variable following Fukuda *et al.* (2004). Specifically, we introduce a time-varying stochastic variable $\mathcal{A}_n$ whose realization corresponds to $\tilde{\alpha}_n^2 = \alpha_n^2/\sigma^2$. $\mathcal{A}_n$ is assumed

to be a discrete stochastic variable which takes a value among $\{\tilde{\alpha}^2_{(1)}, \tilde{\alpha}^2_{(2)}, \ldots, \tilde{\alpha}^2_{(m)}\}$ at time t_n where m denotes the number of candidate values.

In order to estimate the temporal evolution of $\mathcal{A}_n$, we introduce a stochastic model which describes a prior information for the temporal evolution of the hyperparameter. In this study, $\mathcal{A}_n$ is assumed to follow a Markov process, that is,

$$p(\mathcal{A}_n|\mathcal{A}_{1:n-1}) = p(\mathcal{A}_n|\mathcal{A}_{n-1}), \tag{13}$$

where $\mathcal{A}_{i:j} = \{\mathcal{A}_i, \mathcal{A}_{i+1}, \ldots, \mathcal{A}_j\}$ is a set of $\mathcal{A}_n$ from time t_i to time t_j. The stochastic model is realized by a Markov switching model with transition probability given by

$$\Pr\left(\mathcal{A}_n = \tilde{\alpha}^2_{(l)}|\mathcal{A}_{n-1} = \tilde{\alpha}^2_{(k)}\right) = \begin{cases} \beta & l = k \\ (1-\beta)/(m-1) & l \neq k, \end{cases} \tag{14}$$

where $0 \leq \beta \leq 1$. This means that the probability of $\mathcal{A}_n$ taking the same value as $\mathcal{A}_{n-1}$ is β, and the probability of $\mathcal{A}_n$ taking a different value from $\mathcal{A}_{n-1}$ is $(1-\beta)/(m-1)$. Hence the value of β controls the temporal smoothness of the hyperparameter evolution. We present a method to determine the value of β in Section 2.1.5.

2.1.4 Hierarchical Bayesian state space model

Let $\tilde{\alpha}^2_n$ be a realization of $\mathcal{A}_n$ at time t_n. Given a realization $\tilde{\alpha}^2_n$, the observation model (3), the system model (7), and the stochastic model of hyperparameter evolution (14), form a hierarchical Bayesian state space model as follows:

$$\mathbf{x}_n = \mathbf{F}_n\mathbf{x}_{n-1} + \mathbf{v}_n, \quad \mathbf{v}_n \sim N\left[\mathbf{0}, \sigma^2\tilde{\mathbf{Q}}_n\left(\mathcal{A}_n = \tilde{\alpha}^2_n\right)\right] \tag{15}$$

$$\mathbf{y}_n = \mathbf{H}_n\mathbf{x}_n + \mathbf{w}_n, \quad \mathbf{w}_n \sim N(\mathbf{0}, \sigma^2\tilde{\mathbf{R}}_n) \tag{16}$$

$$\Pr\left(\mathcal{A}_n = \tilde{\alpha}^2_{(j)}\middle|\mathcal{A}_{n-1} = \tilde{\alpha}^2_{(i)}\right) = \begin{cases} \beta & i = j \\ (1-\beta)/(m-1) & i \neq j, \end{cases} \tag{17}$$

where

$$\tilde{\mathbf{Q}}_n\left(\mathcal{A}_n = \tilde{\alpha}^2_n\right) = \begin{bmatrix} \tilde{\alpha}^2_n\mathbf{Q}^W_n & \ldots & \mathbf{O} & \mathbf{O} & \mathbf{O} \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ \mathbf{O} & \ldots & \tilde{\alpha}^2_n\mathbf{Q}^W_n & \mathbf{O} & \mathbf{O} \\ \mathbf{O} & \ldots & \mathbf{O} & \tau^2/\sigma^2\Delta t_n\mathbf{I}_{3N_s\times 3N_s} & \mathbf{O} \\ \mathbf{O} & \ldots & \mathbf{O} & \mathbf{O} & \tau^2_f/\sigma^2\mathbf{I}_{3\times 3} \end{bmatrix} \tag{18}$$

$$\tilde{\mathbf{R}}_n = \Sigma_n \tag{19}$$

with $\mathbf{Q}^W_n$ defined in (12).

2.1.5 Model selection

We can determine the number of basis functions, M, and the value of β which controls the temporal smoothness of the hyperparameter evolution using a Bayesian model comparison framework (e.g. MacKay 1992, 2003). In this framework, relative plausibilities of models with different (M, β) values are evaluated by posterior probabilities of (M, β). If we have two different models with $(M = M_1, \beta = \beta_1)$ and $(M = M_2, \beta = \beta_2)$, the relative plausibility is given by the ratio of posterior probabilities of (M_1, β_1) and (M_2, β_2). Using Bayes' theorem, we obtain

$$\frac{p(M_1, \beta_1|\mathbf{y})}{p(M_2, \beta_2|\mathbf{y})} = \frac{p(\mathbf{y}|M_1, \beta_1)}{p(\mathbf{y}|M_2, \beta_2)}\frac{p(M_1, \beta_1)}{p(M_2, \beta_2)}, \tag{20}$$

where $\mathbf{y}$ is data, $p(M, \beta|\mathbf{y})$ is the posterior probability, $p(\mathbf{y}|M, \beta)$ is the marginal likelihood, and $p(M, \beta)$ is prior probability of (M, β). The second ratio, $p(M_1, \beta_1)/p(M_2, \beta_2)$, on the right-hand side measures how much we favour the model with (M_1, β_1) over that with (M_2, β_2) before we collect the data. The first ratio, $p(\mathbf{y}|M_1, \beta_1)/p(\mathbf{y}|M_2, \beta_2)$, expresses how well the model with (M_1, β_1) predicts the data compared to that with (M_2, β_2). When the posterior ratio on the left-hand side is greater than 1, we prefer (M_1, β_1) over (M_2, β_2). Similarly, if we have many models with different (M, β) values, the optimal value of (M, β) is the one that maximizes the posterior probability $p(M, \beta|\mathbf{y})$.

In this study, we assume a uniform prior probability, that is, $p(M, \beta)$ is constant for all the possible values of (M, β). In this case, the prior ratio, $p(M_1, \beta_1)/p(M_2, \beta_2)$, in (20) is equal to unity and, as a result, the ratio of the posterior probabilities is proportional to the ratio of the marginal likelihoods. Therefore, the optimal value of (M, β) is the one that maximizes the marginal likelihood, $p(\mathbf{y}|M, \beta)$. A method for computing the marginal likelihood is given in appendix C of Fukuda *et al.* (2004).

2.2 Bayesian recursive estimation

For estimation of the hyperparameter $\mathcal{A}_n$ and the state vector $\mathbf{x}_n$ based on the hierarchical Bayesian state space model given in Section 2.1.4, we use a Bayesian recursive filtering algorithm, Monte Carlo mixture Kalman filter (MCMKF) introduced by Fukuda *et al.* (2004). Here we briefly summarize the method. Readers are referred to section 3 of Fukuda *et al.* (2004) for more details. Let $\mathbf{y}_{i:j}$ be a set of data vectors from time t_i to time t_j, that is, $\mathbf{y}_{i:j} = \{\mathbf{y}_i, \mathbf{y}_{i+1}, \ldots, \mathbf{y}_j\}$. The purpose of the MCMKF algorithm is to estimate the joint posterior distribution of the hyperparameter $\mathcal{A}_{1:N_e}$ and the state vector $\mathbf{x}_n$ conditioned on data $\mathbf{y}_{1:N_e}$, $p(\mathbf{x}_n, \mathcal{A}_{1:N_e}|\mathbf{y}_{1:N_e})$, for $n = 1, \ldots, N_e$ where N_e is the number of observation epochs. As shown in (15) and (16), the state vector $\mathbf{x}_n$ is linearly related to data while the hyperparameter $\mathcal{A}_n$ is non-linearly related. The linear and the non-linear parameters in the joint posterior distribution, $p(\mathbf{x}_n, \mathcal{A}_{1:N_e}|\mathbf{y}_{1:N_e})$, can be separated using an identity for joint probability,

$$p(\mathbf{x}_n, \mathcal{A}_{1:N_e}|\mathbf{y}_{1:N_e}) = p(\mathbf{x}_n|\mathbf{y}_{1:N_e}, \mathcal{A}_{1:N_e})p(\mathcal{A}_{1:N_e}|\mathbf{y}_{1:N_e}). \tag{21}$$

The MCMKF estimates the linear part, $p(\mathbf{x}_n|\mathbf{y}_{1:N_e}, \mathcal{A}_{1:N_e})$, using the Kalman filtering algorithm and the non-linear part, $p(\mathcal{A}_{1:N_e}|\mathbf{y}_{1:N_e})$, using a Bayesian sequential Monte Carlo method which is conceptually similar to the particle filter (Kitagawa 1996).

In the MCMKF, we first estimate the posterior distribution of $\mathcal{A}_{1:N_e}$, $p(\mathcal{A}_{1:N_e}|\mathbf{y}_{1:N_e})$. For this purpose, two conditional distributions of $\mathcal{A}_{1:n}$, (i) predictive distribution $p(\mathcal{A}_{1:n}|\mathbf{y}_{1:n-1})$ and (ii) filter distribution $p(\mathcal{A}_{1:n}|\mathbf{y}_{1:n})$, are approximated by many 'particles' that can be considered to be independent realizations from each conditional distribution in a similar way as the particle filter (Kitagawa 1996). Let $\mathcal{A}^{(j)}_{n|k}$ be the jth realization of the distribution of $\mathcal{A}_n$ given data $\mathbf{y}_{1:k}$, $p(\mathcal{A}_n|\mathbf{y}_{1:k})$, and let $\mathcal{A}^{(j)}_{1:i|k} = \{\mathcal{A}^{(j)}_{1|k}, \mathcal{A}^{(j)}_{2|k}, \ldots, \mathcal{A}^{(j)}_{i|k}\}$ denote the jth realization of the conditional distribution $p(\mathcal{A}_{1:i}|\mathbf{y}_{1:k})$. Each of the predictive and filter distributions is approximated by $N_p (N_p \gg 1)$ realizations (particles) as follows:

$$\left\{\mathcal{A}^{(1)}_{1:n|n-1}, \mathcal{A}^{(2)}_{1:n|n-1}, \ldots, \mathcal{A}^{(N_p)}_{1:n|n-1}\right\} \sim p(\mathcal{A}_{1:n}|\mathbf{y}_{1:n-1}) \tag{22}$$

$$\left\{\mathcal{A}^{(1)}_{1:n|n}, \mathcal{A}^{(2)}_{1:n|n}, \ldots, \mathcal{A}^{(N_p)}_{1:n|n}\right\} \sim p(\mathcal{A}_{1:n}|\mathbf{y}_{1:n}). \tag{23}$$

(22) and (23) mean that the predictive and the filter distributions are approximated by the following probability density functions:

$$p(\mathcal{A}_{1:n}|\mathbf{y}_{1:n-1}) = \frac{1}{N_p}\sum_{j=1}^{N_p}\delta\left(\mathcal{A}_{1:n} - \mathcal{A}^{(j)}_{1:n|n-1}\right) \tag{24}$$

$$p(\mathcal{A}_{1:n}|\mathbf{y}_{1:n}) = \frac{1}{N_p}\sum_{j=1}^{N_p}\delta\left(\mathcal{A}_{1:n} - \mathcal{A}^{(j)}_{1:n|n}\right), \tag{25}$$

where $\delta(x)$ is a discrete version of the delta function defined by $\delta(x) = 1$ if $x = 0$ and $\delta(x) = 0$ otherwise. Fukuda *et al.* (2004) showed that starting from $\{\mathcal{A}^{(1)}_{0|0}, \ldots, \mathcal{A}^{(N_p)}_{0|0}\}$ which are realizations of an initial distribution for $\mathcal{A}_n$, $p(\mathcal{A}_0)$, a set of particles approximating the predictive distribution and the filter distribution are obtained recursively by implementing the following prediction and filtering algorithms for $n = 1, \ldots, N_e$:

$$\text{Prediction}: \quad \left\{\mathcal{A}^{(1)}_{1:n-1|n-1}, \ldots, \mathcal{A}^{(N_p)}_{1:n-1|n-1}\right\} \longrightarrow \left\{\mathcal{A}^{(1)}_{1:n|n-1}, \ldots, \mathcal{A}^{(N_p)}_{1:n|n-1}\right\} \tag{26}$$

$$\text{Filtering}: \quad \left\{\mathcal{A}^{(1)}_{1:n|n-1}, \ldots, \mathcal{A}^{(N_p)}_{1:n|n-1}\right\} \longrightarrow \left\{\mathcal{A}^{(1)}_{1:n|n}, \ldots, \mathcal{A}^{(N_p)}_{1:n|n}\right\}. \tag{27}$$

The prediction step propagates the filter distribution at the previous epoch, $\{\mathcal{A}^{(1)}_{1:n-1|n-1}, \ldots, \mathcal{A}^{(N_p)}_{1:n-1|n-1}\} \sim p(\mathcal{A}_{1:n-1}|\mathbf{y}_{1:n-1})$, through the stochastic model of the hyperparameter evolution (17) to form the predictive distribution, $\{\mathcal{A}^{(1)}_{1:n|n-1}, \ldots, \mathcal{A}^{(N_p)}_{1:n|n-1}\} \sim p(\mathcal{A}_{1:n}|\mathbf{y}_{1:n-1})$. Then the filtering step incorporates the new data $\mathbf{y}_n$ and updates the predictive distribution through Bayes' theorem to form the new filter distribution, $\{\mathcal{A}^{(1)}_{1:n|n}, \ldots, \mathcal{A}^{(N_p)}_{1:n|n}\} \sim p(\mathcal{A}_{1:n}|\mathbf{y}_{1:n})$. More specifically, particles approximating the predictive and the filter distributions can be obtained by repeating (i) and (ii) for $n = 1, \ldots, N_e$:

(i) Repeat (a), (b) and (c) for $j = 1, \ldots, N_p$.
(a) Generate a realization $\mathcal{A}^{(j)}_{n|n-1}$ from $p(\mathcal{A}_n|\mathcal{A}_{n-1} = \mathcal{A}^{(j)}_{n-1|n-1})$ which is given by (17).
(b) Obtain a particle $\mathcal{A}^{(j)}_{1:n|n-1}$ by $\mathcal{A}^{(j)}_{1:n|n-1} = \left[\mathcal{A}^{(j)}_{1:n-1|n-1}, \mathcal{A}^{(j)}_{n|n-1}\right]$.
(c) Compute the importance weight for jth particle, $w^{(j)}_n = p(\mathbf{y}_n|\mathcal{A}_{1:n} = \mathcal{A}^{(j)}_{1:n|n-1}, \mathbf{y}_{1:n-1})$.
(ii) Obtain particles $\{\mathcal{A}^{(1)}_{1:n|n}, \ldots, \mathcal{A}^{(N_p)}_{1:n|n}\}$ by the sampling with replacement of $\{\mathcal{A}^{(1)}_{1:n|n-1}, \ldots, \mathcal{A}^{(N_p)}_{1:n|n-1}\}$ with sampling probabilities $w^{(j)}_n / \sum_{i=1}^{N_p} w^{(i)}_n$.

Intuitively, the prediction step [step (i)(a)–(b)] imposes temporal smoothing of the hyperparameter using (17), and then, in the filtering step [step (ii)], particles $\{\mathcal{A}^{(1)}_{1:n|n-1}, \ldots, \mathcal{A}^{(N_p)}_{1:n|n-1}\}$ with high/low importance weights are multiplied/discarded to generate particles $\{\mathcal{A}^{(1)}_{1:n|n}, \ldots, \mathcal{A}^{(N_p)}_{1:n|n}\}$. The details of these steps are described in Fukuda *et al.* (2004). Using the prediction and the filtering algorithms recursively, we finally obtain N_p particles $\{\mathcal{A}^{(1)}_{1:N_e|N_e}, \ldots, \mathcal{A}^{(N_p)}_{1:N_e|N_e}\}$ that approximate $p(\mathcal{A}_{1:N_e}|\mathbf{y}_{1:N_e})$, the posterior distribution of $\mathcal{A}_{1:N_e}$ conditioned on all of the available data.

Next, the MCMKF estimates the linear part in (21), $p(\mathbf{x}_n|\mathbf{y}_{1:N_e}, \mathcal{A}_{1:N_e})$. The posterior distribution of $\mathbf{x}_n$ conditioned on each particle, $p(\mathbf{x}_n|\mathbf{y}_{1:N_e}, \mathcal{A}_{1:N_e} = \mathcal{A}^{(j)}_{1:N_e|N_e})$, is a Gaussian distribution because $\mathbf{x}_n$ is linearly related to data and the observation and the system noises are assumed to be Gaussian [see (15) and (16)]. Therefore, the Kalman filtering algorithm can be used to estimate $p(\mathbf{x}_n|\mathbf{y}_{1:N_e}, \mathcal{A}_{1:N_e} = \mathcal{A}^{(j)}_{1:N_e|N_e})$. Specifically, mean vector and covariance matrix of $p(\mathbf{x}_n|\mathbf{y}_{1:N_e}, \mathcal{A}_{1:N_e} = \mathcal{A}^{(j)}_{1:N_e|N_e})$ can be obtained by implementing the Kalman filter for (15) and (16) with a time history of the hyperparameter fixed to the each particle, $\mathcal{A}^{(j)}_{1:N_e|N_e}$. From (21), the joint posterior distribution of $\mathbf{x}_n$ and $\mathcal{A}_{1:N_e}$, $p(\mathbf{x}_n, \mathcal{A}_{1:N_e}|\mathbf{y}_{1:N_e})$, is given by

$$\Pr\left(\mathbf{x}_n, \mathcal{A}_{1:N_e} = \mathcal{A}^{(j)}_{1:N_e|N_e}\middle|\mathbf{y}_{1:N_e}\right) = p\left(\mathbf{x}_n\middle|\mathbf{y}_{1:N_e}, \mathcal{A}_{1:N_e} = \mathcal{A}^{(j)}_{1:N_e|N_e}\right)\Pr\left(\mathcal{A}_{1:N_e} = \mathcal{A}^{(j)}_{1:N_e|N_e}\middle|\mathbf{y}_{1:N_e}\right). \tag{28}$$

Estimate of the state vector $\mathbf{x}_n$ is obtained from the posterior distribution of $\mathbf{x}_n$, $p(\mathbf{x}_n|\mathbf{y}_{1:N_e})$. This distribution is derived by marginalizing the joint posterior distribution $p(\mathbf{x}_n, \mathcal{A}_{1:N_e}|\mathbf{y}_{1:N_e})$ over $\mathcal{A}_{1:N_e}$ as follows:

$$\begin{aligned} p(\mathbf{x}_n|\mathbf{y}_{1:N_e}) &= \sum_{j=1}^{N_p} p\left(\mathbf{x}_n, \mathcal{A}_{1:N_e} = \mathcal{A}^{(j)}_{1:N_e|N_e}|\mathbf{y}_{1:N_e}\right) \\ &= \sum_{j=1}^{N_p} p\left(\mathbf{x}_n|\mathbf{y}_{1:N_e}, \mathcal{A}_{1:N_e} = \mathcal{A}^{(j)}_{1:N_e|N_e}\right)\Pr\left(\mathcal{A}_{1:N_e} = \mathcal{A}^{(j)}_{1:N_e|N_e}|\mathbf{y}_{1:N_e}\right) \\ &= \frac{1}{N_p}\sum_{j=1}^{N_p} p\left(\mathbf{x}_n|\mathbf{y}_{1:N_e}, \mathcal{A}_{1:N_e} = \mathcal{A}^{(j)}_{1:N_e|N_e}\right) \end{aligned} \tag{29}$$

(Fukuda *et al.* 2004). Eq. (29) means that $p(\mathbf{x}_n|\mathbf{y}_{1:N_e})$ is given by the sum of the Gaussian distributions each of which is obtained by the Kalman filter in which the hyperparameter history $\mathcal{A}_{1:N_e}$ is fixed to each particle $\mathcal{A}^{(j)}_{1:N_e|N_e}$. It should be noted that we do not choose one particular particle as an optimal hyperparameter history. Instead, the information from the full posterior distribution of the hyperparameter, $p(\mathcal{A}_{1:N_e}|\mathbf{y}_{1:N_e})$, is incorporated into the posterior distribution of $\mathbf{x}_n$, $p(\mathbf{x}_n|\mathbf{y}_{1:N_e})$.

A summary of the MCMKF algorithm is as follows:

(i) Generate $\{\mathcal{A}^{(1)}_{0|0}, \mathcal{A}^{(2)}_{0|0}, \ldots, \mathcal{A}^{(N_p)}_{0|0}\}$ from an initial distribution $p(\mathcal{A}_0)$.
(ii) Obtain $\{\mathcal{A}^{(1)}_{1:N_e|N_e}, \ldots, \mathcal{A}^{(N_p)}_{1:N_e|N_e}\} \sim p(\mathcal{A}_{1:N_e}|\mathbf{y}_{1:N_e})$ by repeating (a) and (b) for $n = 1, \ldots, N_e$
(a) Prediction: $\{\mathcal{A}^{(1)}_{1:n-1|n-1}, \ldots, \mathcal{A}^{(N_p)}_{1:n-1|n-1}\} \longrightarrow \{\mathcal{A}^{(1)}_{1:n|n-1}, \ldots, \mathcal{A}^{(N_p)}_{1:n|n-1}\}$
(b) Filtering: $\{\mathcal{A}^{(1)}_{1:n|n-1}, \ldots, \mathcal{A}^{(N_p)}_{1:n|n-1}\} \longrightarrow \{\mathcal{A}^{(1)}_{1:n|n}, \ldots, \mathcal{A}^{(N_p)}_{1:n|n}\}$.
(iii) Repeat (c) for $j = 1, \ldots, N_p$
(c) Compute $p(\mathbf{x}_n|\mathbf{y}_{1:N_e}, \mathcal{A}_{1:N_e} = \mathcal{A}^{(j)}_{1:N_e|N_e})$ for $n = 1, \ldots, N_e$ using the Kalman filter.
(vi) Compute $p(\mathbf{x}_n|\mathbf{y}_{1:N_e})$ for $n = 1, \ldots, N_e$ using (29).

2.3 Kalman filter-based inversion method

In order to demonstrate the effectiveness of the new inversion method, it is necessary to compare results obtained by the new method with those obtained by the conventional inversion method. For this purpose, we introduce a Kalman filter-based inversion method which is similar to the conventional NIF method (Segall & Matthews 1997). In this section, we briefly describe the Kalman filter-based inversion method used in this study.

We employ a state space model (e.g. Kitagawa & Gersch 1996; Segall & Matthews 1997) to formulate a statistical model. The state space model used in this study is the same as (15) and (16) except that the hyperparameter α_n^2/σ^2 is treated as a temporally invariable constant, not a time-varying stochastic variable.

In the framework of the state space modelling, mean vectors and covariance matrices of the posterior distributions of the state $p(\mathbf{x}_n|\mathbf{y}_{1:N_e}, \alpha^2, \sigma^2)$ given the hyperparameters $\alpha^2 \equiv \alpha_n^2$ and σ^2,

$$\mathbf{x}_{n|N_e} = \mathrm{E}(\mathbf{x}_n|\mathbf{y}_{1:N_e}, \alpha^2, \sigma^2) \tag{30}$$

$$\mathbf{V}_{n|N_e} = \mathrm{Cov}(\mathbf{x}_n|\mathbf{y}_{1:N_e}, \alpha^2, \sigma^2), \tag{31}$$

are obtained using the Kalman filter algorithm (e.g. Kitagawa & Gersch 1996; Segall & Matthews 1997). The hyperparameters α^2 and σ^2 are estimated by maximizing the likelihood, $p(\mathbf{y}_{1:N_e}|\alpha^2, \sigma^2)$. A recursive method for computing the likelihood is given by Kitagawa & Gersch (1996) and Segall & Matthews (1997).

3 NUMERICAL EXAMPLES

In this section, we examine the validity and advantages of the new inversion method through two numerical examples in which the method is applied to simulated data. The procedure for inverting simulated data is as follows. First, the fault geometry and the 'true' space–time slip distribution on the fault surface are specified. Second, surface displacement time-series at observation sites are computed based on the true slip distribution. Third, simulated data are generated by adding Gaussian white noise, random benchmark motions, and common mode errors

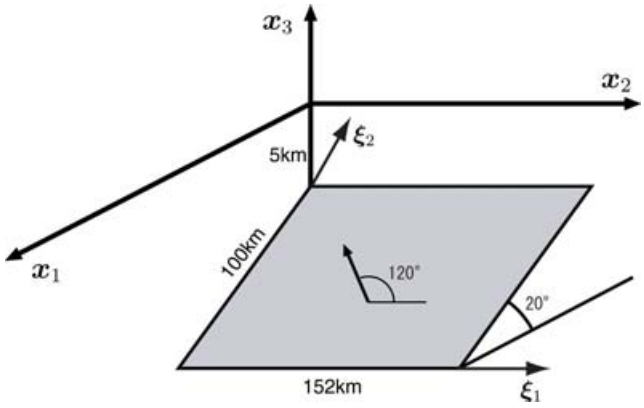

Figure 2. Fault geometry used in the numerical simulation. The bold arrow on the fault plane represents true slip direction (the true direction of the movement of the hanging wall relative to the foot wall).

to the surface displacement time-series. Finally, the simulated data are inverted to reconstruct the space–time distribution of slip and slip velocity. The validity of our inversion method is examined by comparing the estimated slip velocity with the true value and with that obtained by the conventional Kalman filter-based inversion method.

3.1 Simulated data

We consider a dipping thrust fault in a homogeneous, isotropic, elastic half-space as shown in Fig. 2. The fault surface is discretized into 4 km × 4 km rectangular subfaults and slip on each subfault is assumed to be spatially uniform. Let (ξ_1, ξ_2) be a coordinate of the centre of a subfault in ξ_1–ξ_2 coordinate system defined in Fig. 2. Let $s(\xi_1, \xi_2, t_n)$ be the true slip magnitude on a subfault whose centre is located at (ξ_1, ξ_2). $s(\xi_1, \xi_2, t_n)$ as a function of (ξ_1, ξ_2) and the time of measurement t_n is assumed as follows:

$$s(\xi_1, \xi_2, t_n) = A_s \varphi(t_n) \exp\left\{-\frac{1}{2}\left[\left(\frac{\xi_1 - \xi_1^{(s)}}{\sigma_1^{(s)}}\right)^2 + \left(\frac{\xi_2 - \xi_2^{(s)}}{\sigma_2^{(s)}}\right)^2\right]\right\}, \tag{32}$$

where A_s is a constant, $\varphi(t_n)$ is a function which describes time dependence of the slip magnitude, and $\xi_1^{(s)}, \xi_2^{(s)}, \sigma_1^{(s)}$ and $\sigma_2^{(s)}$ determine spatial distribution of slip magnitude. These constants are given by $A_s = 0.1$ m, $\xi_1^{(s)} = 74$ km, $\xi_2^{(s)} = 54$ km, $\sigma_1^{(s)} = 30$ km and $\sigma_2^{(s)} = 20$ km. The time dependence $\varphi(t_n)$ is assumed as

$$\varphi(t_n) = \begin{cases} 0 & n = 1, \ldots, 45 \\ \frac{1}{10}(t_n - 45) & n = 46, \ldots, 55 \\ 1 & n = 56, \ldots, 97, \end{cases} \tag{33}$$

where $n = 1, 2, \ldots, N_e$, $N_e = 97$ and $t_1 = 1, t_2 = 2, \ldots, t_{97} = 97$ d. The true slip direction is assumed to be uniform in space and time with the rake angle of 120° as shown in Fig. 2.

The distribution of 81 observation sites on the ground surface is shown in Fig. 3. We sample the horizontal and vertical displacements with observation noise, random benchmark motions, and common mode errors at the 81 stations with sampling interval of 1 d for 97 d. Surface displacement time-series at the observation sites are computed as the sum of contributions from slip on each subfault using the analytical expressions that relate surface displacements to a rectangular dislocation in a homogeneous elastic half-space (Okada 1985). Finally, the observation noise (e_{ijn} in eq. 1), the random benchmark motions [$\mathcal{L}_i(\mathbf{r}_j, t_n)$ in eq. 1], and the common mode errors [$f_i(t_n)$ in eq. 1] are added to the time-series to make a set of simulated data. The observation noise and the common mode errors are generated from Gaussian distribution with zero mean and standard deviation of 2 mm. The random benchmark motions are computed from Brownian random walk processes with scale parameter 1 mm yr$^{-1/2}$.

3.2 Inversion

We first discretize the slip distribution on the fault surface with 4 km × 4 km subfaults. Second, the spatial basis functions for slip distribution is generated following (A15). Third, the observation model (16) and the system model (15) is constructed. The random walk scale parameter, τ, divided by σ and the system noise standard deviation of the common mode errors, τ_f, divided by σ are fixed to $\tau/\sigma = 1$ mm yr$^{-1/2}$ and $\tau_f/\sigma = 2 \times 10^{-2}$ m, respectively. Finally, in the model of hyperparameter evolution, (17), we employ $m = 18$ candidate values for $\mathcal{A}_n$, $\{\tilde{\alpha}^2_{(1)}, \tilde{\alpha}^2_{(2)}, \ldots, \tilde{\alpha}^2_{(18)}\} = \{10^{-4}, 10^{-3}, \ldots, 10^{13}\}$. For this inversion, $N_p = 2000$ particles are used. The initial distribution of the hyperparameter $\mathcal{A}_n$, $p(\mathcal{A}_0)$, is assumed to follow a uniform distribution: $\Pr(\mathcal{A}_0 = \tilde{\alpha}^2_{(i)}) = 1/m (i = 1, 2, \ldots, m)$.

The number of spatial basis functions, M, and the parameter β in the hyperparameter evolution model are unknown in this inversion. These parameters can be determined by maximizing the marginal likelihood, $p(\mathbf{y}_{1:N_e}|M, \beta)$, as shown in Section 2.1.5. It is expected that

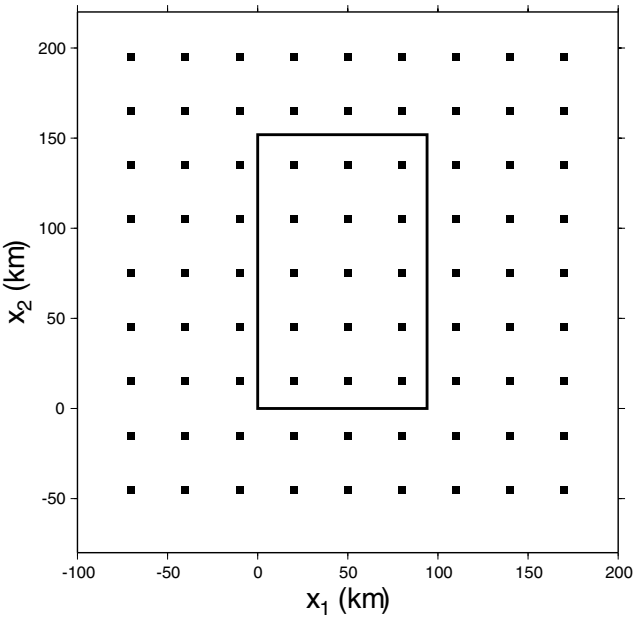

Figure 3. Map of the station distribution and fault plane for the numerical simulation. The solid squares are the observation sites and the rectangle is the surface projection of the fault plane whose geometry is shown in Fig. 2.

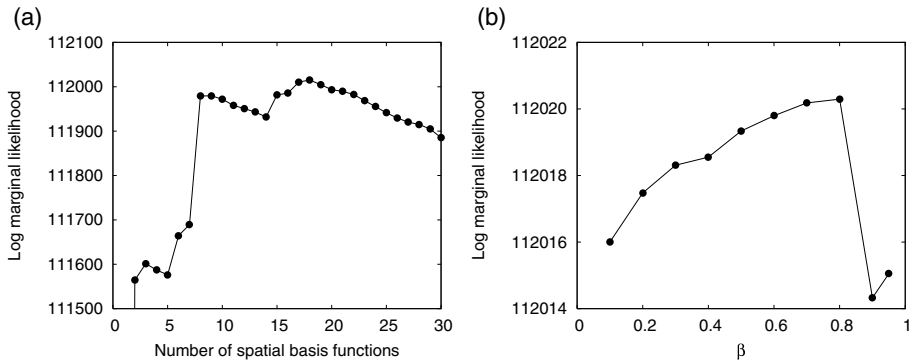

Figure 4. (a) Log marginal likelihood as a function of the number of spatial basis functions, M, for $\beta = 0.95$. (b) Log marginal likelihood as a function of β. The number of spatial basis functions is fixed to $M = 18$.

the correlation between M and β is small because M determines spatial smoothness of slip distribution, whereas β controls the temporal smoothness of the hyperparameter evolution. Therefore, we first determine M with a fixed value of β and then estimate β with M fixed to the selected value. More specifically, we first fix $\beta = 0.95$ and apply the inversion method for $M = 1, 2, \ldots, 30$. Fig. 4(a) shows the log marginal likelihood, $\log[p(\mathbf{y}_{1:N_e}|M, \beta)]$, as a function of M for $\beta = 0.95$, showing that the marginal likelihood is maximized at $M = 18$. Based on this result, we employ $M = 18$ basis functions below. Next, we compute the log marginal likelihood for $\beta = 0.1, 0.2, \ldots, 0.9$ with the number of basis functions fixed to $M = 18$. Fig. 4(b) shows the log marginal likelihood as a function of β. The β value that maximizes the marginal likelihood is $\beta = 0.8$. Below, we show the inversion result for $M = 18$ and $\beta = 0.8$.

3.3 Results and discussion

Fig. 5(a) shows the estimated temporal variation of the posterior probability distribution of the hyperparameter $\mathcal{A}_n$, $\Pr(\mathcal{A}_n = \tilde{\alpha}_n^2|\mathbf{y}_{1:N_e})$, where $\tilde{\alpha}_n^2$ is a realization of $\mathcal{A}_n$ and takes a value among $\{\tilde{\alpha}_{(1)}^2, \tilde{\alpha}_{(2)}^2, \ldots, \tilde{\alpha}_{(m)}^2\}$. Because we employ the stochastic model of slip evolution shown in (4), the magnitude of $\tilde{\alpha}_n^2$ characterizes the magnitude of slip velocity change. As shown in Fig. 5, the estimated posterior probability of the hyperparameter $\mathcal{A}_n$ is distributed around smaller values before the transient slip event starts, then concentrates on larger values around 10^3 when slip velocity suddenly changes, and finally returns to smaller values after the event ends. This temporal variation of $\mathcal{A}_n$ is consistent with the true slip history shown in (33) and Fig. 5(b).

The true and estimated slip velocity distributions for Day 43, 46, 50, 54 and 57 are shown in Figs 6(a) and (b), respectively. For comparison, the slip velocity distributions estimated by the conventional Kalman filter-based method (see Section 2.3 for details) are also

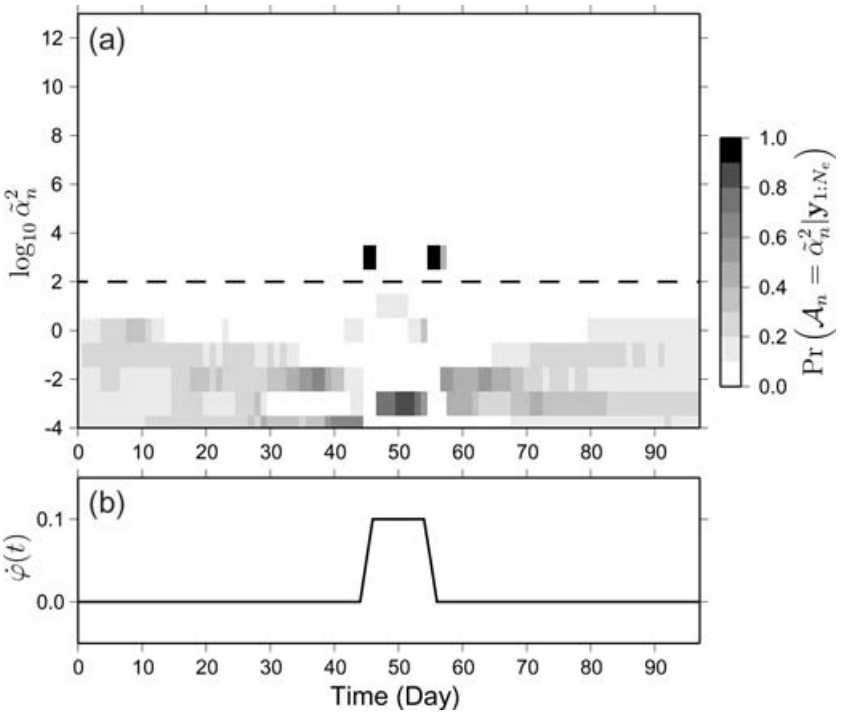

Figure 5. (a) Estimated temporal variation of posterior probability distribution of the hyperparameter $\mathcal{A}_n$, $\Pr\left(\mathcal{A}_n = \tilde{\alpha}_n^2 | \mathbf{y}_{1:N_e}\right)$. The dashed line shows the value of α^2/σ^2 that is used in the Kalman filter-based method. (b) The first derivative of $\varphi(t)$ that govern the time dependence of true slip velocity [See (33) for the definition of $\varphi(t)$].

presented in Fig. 6(c). In the Kalman filter analysis, the hyperparameter α^2/σ^2 is estimated by maximizing the likelihood and the estimated value is $\alpha^2/\sigma^2 = 10^2$. The slip velocity distributions estimated by the conventional method have some considerable differences from true slip velocity distributions and those obtained by the new method. First, the estimated slip velocity is not zero at Day 43 when the true transient slip event has not started yet. Second, the estimated slip velocity at Day 46 and 54 is considerably smaller than true slip velocity. Finally, the estimated slip velocity has not fallen to zero after the end of the transient slip (Day 57). In other words, the slip velocity estimated by the conventional method is temporally oversmoothed during the transient slip event. This problem has been conquered by the new method (Fig. 6b). Fig. 7 compares time histories of slip velocity magnitude at two points on the fault surface, $(\xi_1, \xi_2) = (74, 54)$ km and $(\xi_1, \xi_2) = (106, 54)$ km, obtained by the two methods with the true slip velocity history. The slip velocity history obtained by the conventional method is rather oscillatory during the steady-state periods, and is excessively smoothed during the true slip velocity changes. On the other hand, the slip velocity history estimated by the new method much better reproduces both slip velocity changes and the steady-state velocity. We measure the reproducibility of the two methods in terms of the rms residual between the true slip velocity distributions and the recovered ones:

$$r = \sqrt{\frac{1}{N_e M_f} \sum_{n=1}^{N_e} (\hat{\dot{\mathbf{s}}}_n - \dot{\mathbf{s}}_n)^T (\hat{\dot{\mathbf{s}}}_n - \dot{\mathbf{s}}_n)}, \tag{34}$$

where $\hat{\dot{\mathbf{s}}}_n$ and $\dot{\mathbf{s}}_n$ are M_f-dimensional vectors of the estimated and the true slip velocities at time t_n, respectively. The values of r for the new and the conventional methods are 1.4×10^{-2} and 4.2×10^{-2} cm d^{-1}, respectively, indicating that the new method provides a better recovery of the true slip velocity than the conventional method. These results show that the differences between slip velocity estimated by the conventional method and true slip velocity are due to inappropriate temporal smoothing caused by the constant hyperparameter assumption. These results also demonstrate that introducing the time-varying hyperparameter $\mathcal{A}_n$ considerably improves the ability of reproducing time history of slip velocity in slow slip events.

To examine the effect of the choice of β on the solution, we show in Fig. 8 the estimated temporal variations of the hyperparameter and slip velocity for $\beta = 0.1$ and 0.9. Comparison of the time histories of the hyperparameter for $\beta = 0.1$ (Fig. 8a), $\beta = 0.8$ (Fig. 5a) and $\beta = 0.9$ (Fig. 8b) shows that temporal roughness of the hyperparameter is controlled by β. Fig. 8(c) shows slip velocity histories at $(\xi_1, \xi_2) = (74, 54)$ km for $\beta = 0.1$, 0.8 and 0.9. Although the recovered slip velocity histories are generally similar to each other, the slip velocity history for $\beta = 0.9$ is smoother than that for $\beta = 0.8$ at which the log marginal likelihood is maximized, whereas that for $\beta = 0.1$ is slightly rougher than the case of $\beta = 0.5$. Therefore, the choice of β affects the estimated slip velocity histories, although the effect is not large in this inversion.

We check the effect of the initial distribution of the hyperparameter, $p(\mathcal{A}_0)$, on the solution. Figs 9(a)–(c) show temporal variations of the hyperparameter for three initial distributions, $p(\mathcal{A}_0) = U(10^4, 10^{13})$, $\delta(10^8)$ and $\delta(10^{-2})$, where $U(10^4, 10^{13})$ is a discrete uniform distribution over the interval $[10^4, 10^{13}]$, and $\delta(10^8)$ and $\delta(10^{-2})$ are the delta functions concentrated on 10^8 and 10^{-2}, respectively. These temporal variations are generally similar to Fig. 5(a) in which $p(\mathcal{A}_0) = U(10^{-4}, 10^{13})$ is employed. The initial distributions employed in

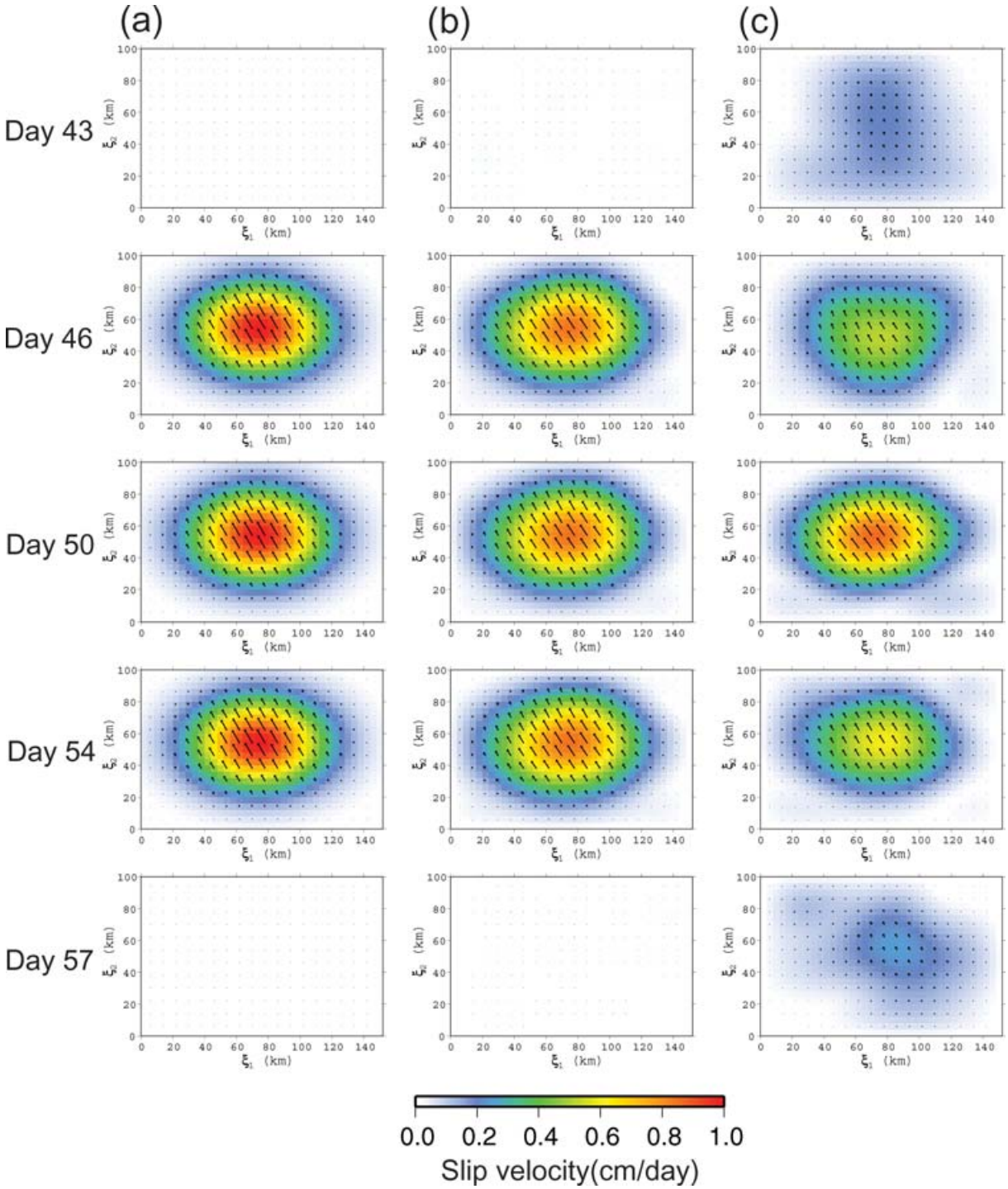

Figure 6. True and estimated slip velocity distributions for Day 43, 46, 50, 54 and 57. (a) True slip velocity distributions. (b) Estimated slip velocity distributions using the new method with the time-varying hyperparameter $\mathcal{A}_n$. (c) Estimated slip velocity distributions using the conventional Kalman filter-based method with the constant hyperparameter.

Figs 9(a) and (b) take significantly larger values than the estimated hyperparameter distribution shown in Fig. 5(a). However, influence of the initial distribution can only be seen at the first epoch. On the other hand, influence of the initial distribution continues for a longer period in Fig. 9(c) in which the initial distribution is close to the estimated hyperparameter distribution. Fig. 9(d) shows slip velocity histories for the four initial distributions, $p(\mathcal{A}_0) = U(10^{-4}, 10^{13})$, $U(10^4, 10^{13})$, $\delta(10^8)$ and $\delta(10^{-2})$. The slip velocity histories are very similar to each other and the small differences in the hyperparameter histories do not have significant effects on the slip velocity histories. This indicates that slip velocity is insensitive to the selection of the initial distribution.

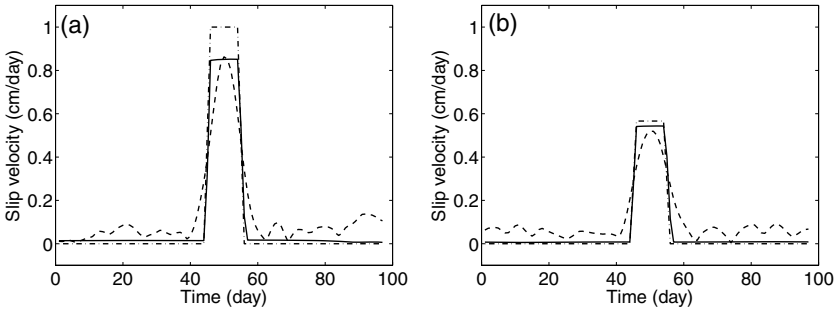

Figure 7. Temporal variation of slip velocity magnitude at two points on the fault plane. (a) Slip velocity magnitude at the location of the largest slip, (ξ_1, ξ_2) = (74, 54) km. (b) Slip velocity magnitude at (ξ_1, ξ_2) = (106, 54) km. The dash-dotted curve is true slip velocity, the solid curve represents slip velocity obtained by the new method with time-varying hyperparameter $\mathcal{A}_n$, and the dashed curve denotes slip velocity obtained by the conventional method with constant hyperparameter.

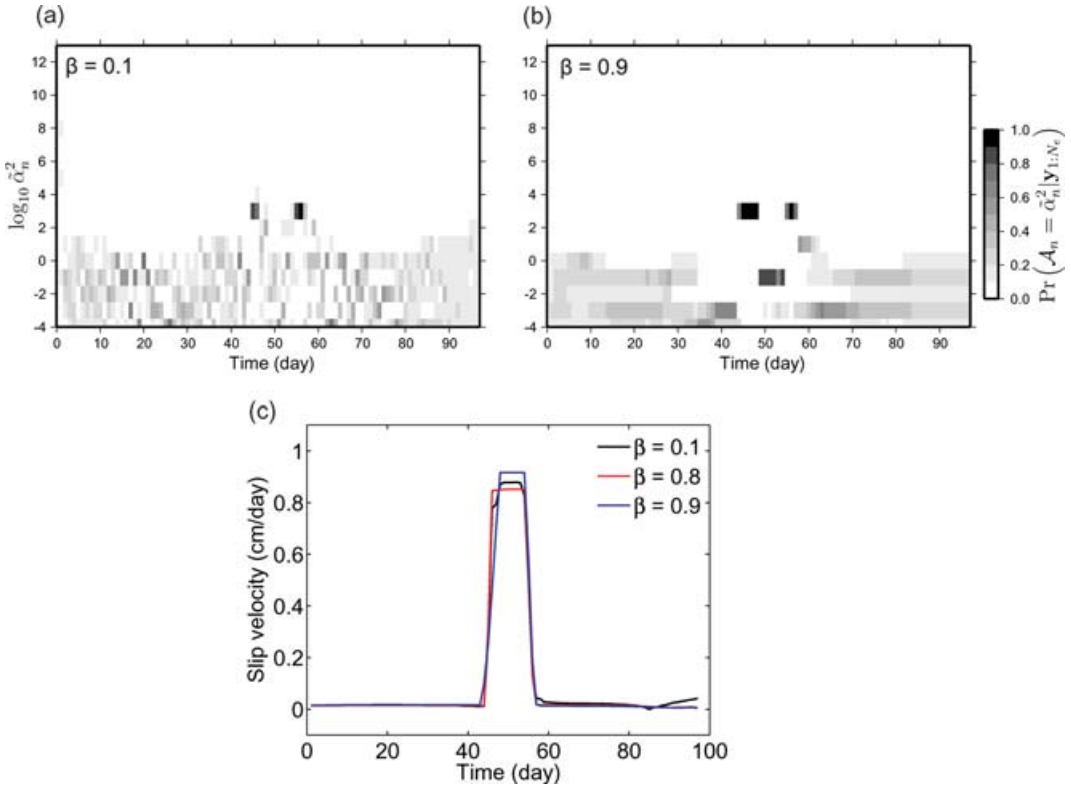

Figure 8. Temporal variations of the hyperparameter and slip velocity for different β values. (a) Estimated temporal variation of posterior probability distribution of the hyperparameter $\mathcal{A}_n$, $\Pr\left(\mathcal{A}_n = \tilde{\alpha}_n^2 | \mathbf{y}_{1:N_e}\right)$ for $\beta = 0.1$. (b) Same as (a) but for $\beta = 0.9$. (c) Temporal variations of slip velocity magnitude at (ξ_1, ξ_2) = (74, 54) km for $\beta = 0.1$, 0.8, 0.9. Black, red, and blue curves are slip velocity histories for $\beta = 0.1$, 0.8 and 0.9, respectively. The slip velocity history for $\beta = 0.8$ is the same as the solid curve in Fig. 7(a).

3.4 A more complex example

In the previous example, spatial distribution and time history of the true slip were very simple in that the spatial distribution has only one large slip area and the true slip accelerates and decelerates simultaneously everywhere on the fault. Here we test the new method with a more complex example in which the true slip distribution has two patches, each of which has a different time history. We use the same fault geometry and the same station distribution as used in the previous example (Figs 2 and 3). We assume the true slip distribution for $t_n = 1, 2, \ldots, 97$ (d). Temporal variation of the true slip velocity distribution for the period between $t_n = 29$ and 77 (d) is shown in Fig. 10. Simulated daily time-series that contain fault slip-induced displacements, observation noise, random benchmark motions, and common mode errors are generated in the same way as described in Section 3.1. We apply the new inversion method to the simulated data in the same way as described in Section 3.2. The values of M and β are determined by maximizing the marginal likelihood. The resulting values are $M = 30$ and $\beta = 0.55$. We also apply the conventional Kalman filter-based inversion method for comparison. In the Kalman filter analysis, we use the same value of M as in the new method, and the hyperparameter α^2/σ^2 is estimated by maximizing the likelihood and the estimated value is $\alpha^2/\sigma^2 = 10^2$.

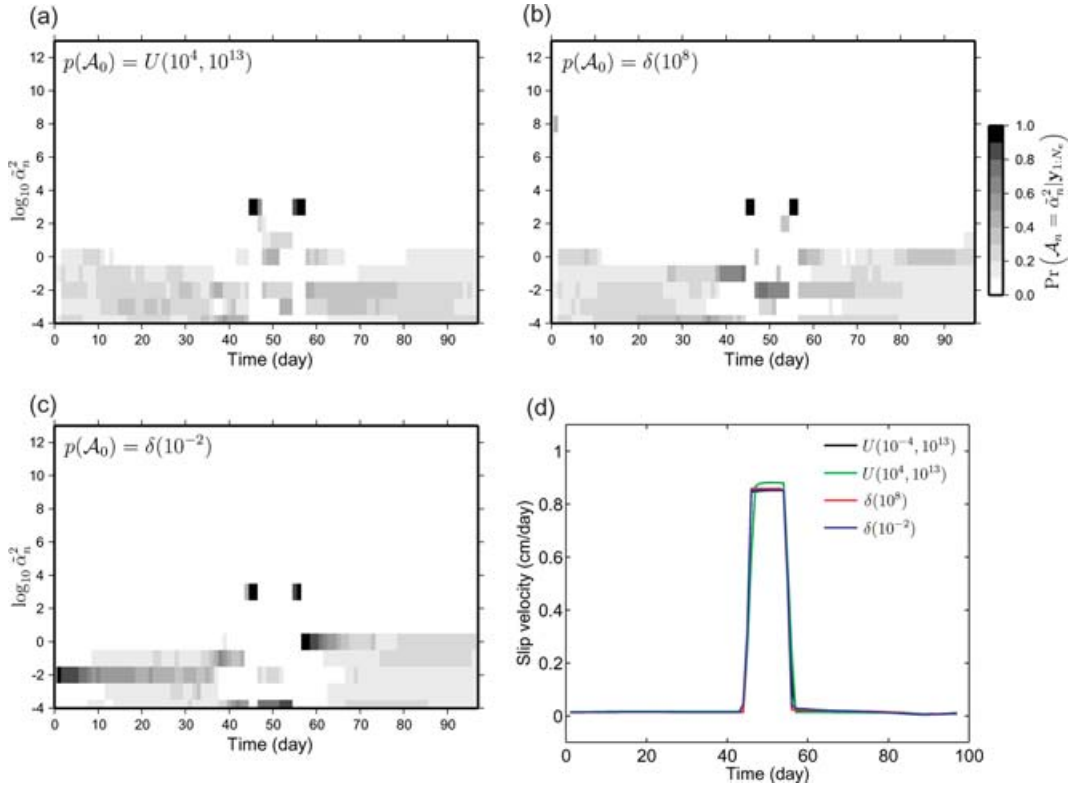

Figure 9. Temporal variations of the hyperparameter and slip velocity for different initial distributions of the hyperparameter, $p(\mathcal{A}_0)$. (a) Estimated temporal variation of posterior probability distribution of the hyperparameter $\mathcal{A}_n$, $\Pr(\mathcal{A}_n = \tilde{\alpha}_n^2 | \mathbf{y}_{1:N_e})$ for $p(\mathcal{A}_0) = U(10^4, 10^{13})$. (b) Same as (a) but for $p(\mathcal{A}_0) = \delta(10^8)$. (c) Same as (a) but for $p(\mathcal{A}_0) = \delta(10^{-2})$. (d) Temporal variations of slip velocity magnitude at $(\xi_1, \xi_2) = (74, 54)$ km for four different initial distributions. Black, green, red, and blue curves are slip velocity histories for $p(\mathcal{A}_0) = U(10^{-4}, 10^{13})$, $U(10^4, 10^{13})$, $\delta(10^8)$ and $\delta(10^{-2})$, respectively. The slip velocity history for $p(\mathcal{A}_0) = U(10^{-4}, 10^{13})$ is the same as the solid curve in Fig. 7(a).

The posterior probability distribution of the hyperparameter $\mathcal{A}_n$ estimated by the new method concentrates on values around 10^3 when slip velocity suddenly changes ($t_n = 40, 50, 60, 70$). During other periods, the posterior probability distribution of the hyperparameter is distributed around smaller values (10^{-4}–10^0). The Kalman filter-based estimate of the hyperparameter value, $\alpha^2/\sigma^2 = 10^2$, is between the two ranges. This behaviour is similar to the previous example. Figs 11 and 12 show temporal variations of slip velocity distribution estimated by the new method and the Kalman filter-based method, respectively. Both methods underestimate the magnitude of slip velocity. This is because we impose spatial smoothing of slip by expressing the slip distribution in terms of a finite number of spatial basis functions. Comparison of the slip velocity distributions estimated by the Kalman filter-based method (Fig. 12) and the true slip velocity distributions (Fig. 10) shows that the Kalman filter-based method temporally oversmoothes the slip accelerations and decelerations and undersmoothes the slip velocity during the steady-state periods. On the other hand, Fig. 11 shows that the new method does not suffer from significant oversmoothing and undersmoothing and better reproduces the true temporal evolution. These results are qualitatively similar to the previous example. The rms residuals between the true and the estimated slip velocity distributions, r, defined in (34) for the new and the Kalman filter-based methods are 3.0×10^{-2} and 4.7×10^{-2} cm d^{-1}, respectively, indicating that the new method better reproduces the true slip velocity. This example suggests that the new method is useful not only for transient slip events with very simple slip evolution, but also for events with complex slip evolution.

4 APPLICATION TO THE 2002 SLOW SLIP EVENT OFF THE BOSO PENINSULA, JAPAN

4.1 Brief description of the event

The Boso peninsula is located in the southeast Kanto region in central Japan beneath which the Philippine Sea plate is subducting at the Sagami trough to the northwest (Fig. 13a). In addition, the Pacific plate is subducting to the west beneath the Philippine Sea plate at the Japan trench (e.g. Ishida 1992). Recent continuous GPS measurements have revealed that the interseismic surface velocity in this region is about 2 cm yr^{-1} to the northwest relative to the landward plate in the southern part and smaller value in the northern part, reflecting interplate coupling between the Philippine Sea plate and the landward plate (Sagiya 2004). Sagiya (2004) estimated interplate coupling rate between

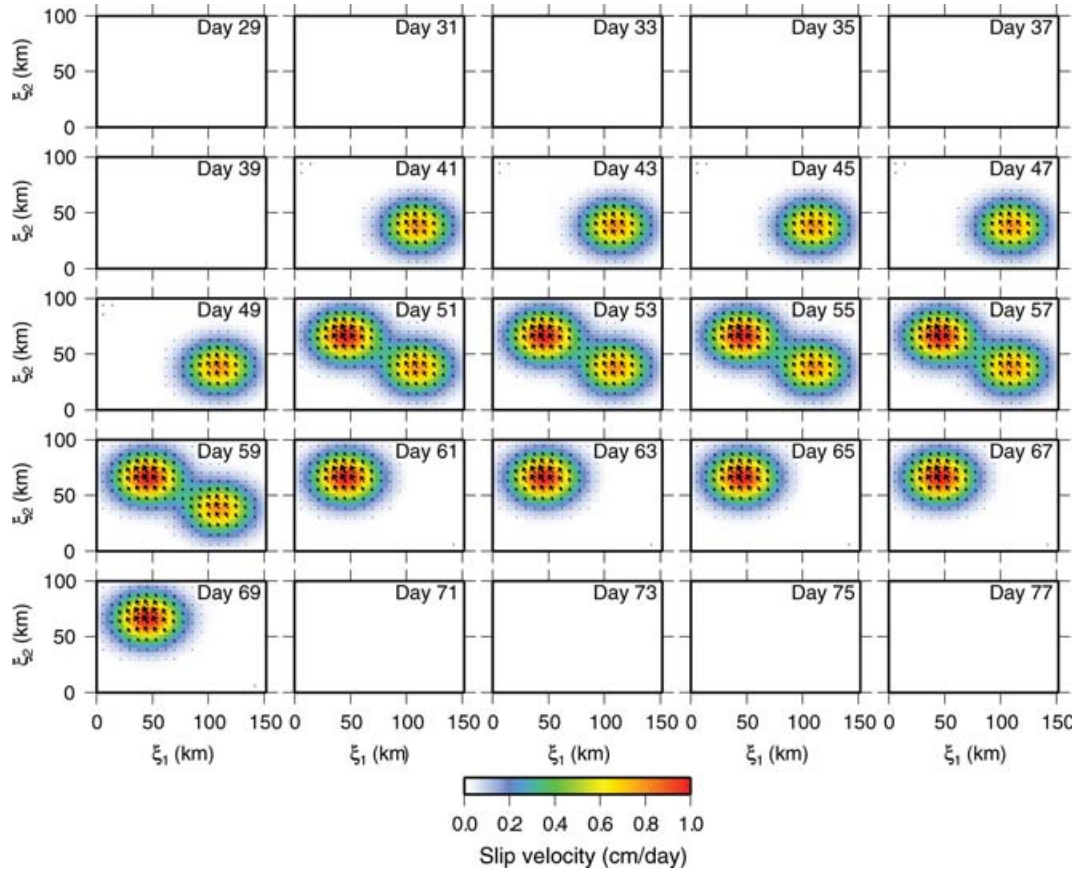

Figure 10. True slip velocity distributions for the period between Day 29 and 77.

the Philippine Sea plate and the landward plate. According to his results, the estimated coupling rate off the east coast of the Boso peninsula is over 30 mm yr^{-1} in southern part and gradually decreases northward. The two plates are decoupled in the northern part.

In 1996 May, a continuous GPS array, GPS Earth Observation Network (GEONET) (Miyazaki *et al.* 1997), detected aseismic transient displacements in the Boso peninsula (Sagiya 2004). The observed transient displacements were southeastward, opposite to the displacements caused by interplate coupling. Sagiya (2004) analysed the transient deformation and found that the observed displacements are reasonably well fit by aseismic slip on the Philippine Sea plate interface off the east coast of the Boso peninsula. The aseismic slip region found by Sagiya (2004) is located between the strongly coupled and decoupled regions. In October 2002, six years after the 1996 transient, similar transient displacements caused by aseismic slip were recorded by GEONET at GPS sites in and around the Boso peninsula (Ozawa *et al.* 2003). Ozawa *et al.* (2003) applied the NIF to estimate spatiotemporal evolution of aseismic slip and found that the observed displacements are well fit by aseismic slip on the Philippine Sea plate similar to the 1996 event.

In this section, we analyse the surface deformation data during the 2002 transient slip event off the east coast of the Boso peninsula using the new inversion method presented in Section 2 in order to demonstrate the effectiveness of the method.

4.2 Data

We use the daily GPS data from 77 stations of GEONET in Kanto district for the period between 16 August 2002 and 23 November 2002. The GPS phase and pseudo-range data are processed with the GIPSY-OASIS II software using the precise point positioning technique (Zumberge *et al.* 1997), with fiducial-free satellite orbits, satellite clock offsets, and Earth orientation parameters provided by Jet Propulsion Laboratory (JPL). The obtained fiducial-free solutions are then transformed into ITRF2000 (Altamimi *et al.* 2002) reference frame using network transformation parameters given by JPL to yield daily coordinate time-series at the 77 stations.

Fig. 13(b) shows observed horizontal displacements for the period between 20 August 2002 and 23 November 2002. Horizontal displacements up to 3.7 cm were observed along the east coast of the Boso peninsula. On the other hand, observed vertical displacements are at most 0.8 cm, much smaller than horizontal displacements and within data uncertainties. Figs 14(a)–(c) show daily time-series of observed displacement at selected GPS sites shown in Fig. 14(d). Rapid transient displacements were observed between day of year (DOY) 276–285 (October 3–12), and the transient continued after DOY 285 with smaller displacement rates.

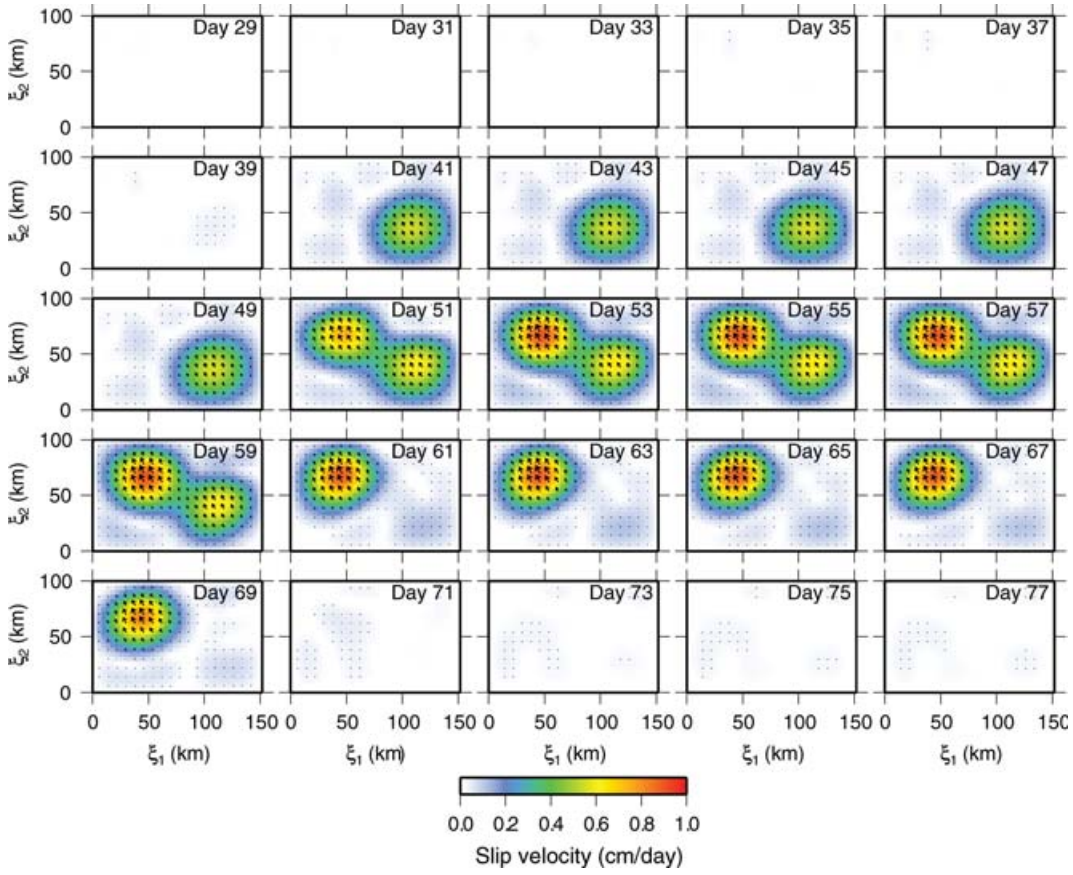

Figure 11. Slip velocity distributions for the period between Day 29 and 77 estimated by the new method with the time-varying hyperparameter $\mathcal{A}_n$.

4.3 Inversion

In this study, we assume that the observed transient displacements were caused by transient aseismic slip on the Philippine Sea plate interface following Ozawa *et al.* (2003). We approximate the geometry of the plate interface by a planar fault summarized in Table 1 following the plate interface configuration given by Ishida (1992).

We invert the observed displacement time-series at the 77 stations for the period between 2002 August 16 and 2002 November 23 to infer space–time distribution of aseismic slip and slip velocity using the new inversion method. We approximate a continuous slip distribution on the fault plane (Table 1) with a discrete slip distribution by dividing the fault plane into 31 $\times$ 25 of 4 km$\times$4 km rectangular segments. Then the spatial basis functions for slip distribution (A15) and the observation and the system models are constructed. The random walk scale parameter, τ, divided by σ and the system noise standard deviation of the common mode errors, τ_f, divided by σ are fixed to $\tau/\sigma = 1$ mm yr$^{-1/2}$ and $\tau_f/\sigma = 2 \times 10^{-2}$ m, respectively. As for time-evolution model for the hyperparameter $\mathcal{A}_n$, (17), we employ $m = 13$ candidate values for $\mathcal{A}_n$, $\{\tilde{\alpha}^2_{(1)}, \tilde{\alpha}^2_{(2)}, \ldots, \tilde{\alpha}^2_{(13)}\} = \{10^{-4}, 10^{-3}, \ldots, 10^{8}\}$. For this inversion, $N_p = 2000$ particles are used. The initial distribution of the hyperparameter $\mathcal{A}_n$, $p(\mathcal{A}_0)$, is assumed to follow a uniform distribution: $\Pr(\mathcal{A}_0 = \tilde{\alpha}^2_{(i)}) = 1/m (i = 1, 2, \ldots, m)$.

The number of spatial basis functions, M, and the parameter β in the hyperparameter evolution model are determined by the same procedure as described in Section 3.2. Fig. 15(a) shows the log marginal likelihood, $\log[p(\mathbf{y}_{1:N_e}|M, \beta)]$, as a function of M for $\beta = 0.95$, indicating that the marginal likelihood is maximized at $M = 15$. Fig. 15(b) shows the log marginal likelihood as a function of β. The β value that maximizes the marginal likelihood is $\beta = 0.5$. Below, we show the inversion result for $M = 15$ and $\beta = 0.5$.

4.4 Results and discussion

Fig. 16 shows the estimated temporal variation of posterior probability distribution of the hyperparameter $\mathcal{A}_n$, $\Pr(\mathcal{A}_n = \tilde{\alpha}^2_n|\mathbf{y}_{1:N_e})$, where $\tilde{\alpha}^2_n$ is a realization of $\mathcal{A}_n$ and takes a value among $\{\tilde{\alpha}^2_{(1)}, \tilde{\alpha}^2_{(2)}, \ldots, \tilde{\alpha}^2_{(13)}\}$. As shown in Fig. 16, the estimated probability density of the hyperparameter $\mathcal{A}_n$ is distributed around smaller values until DOY 273 (September 30), then concentrates on larger values around 10^1 and 10^2 between DOY 276 and DOY 285 (October 3–12) and gradually returns to smaller values. The period with larger hyperparameter value has good correspondence to that of the transient displacements (see Fig. 14). As seen in (4), the magnitude of the hyperparameter characterizes the magnitude of slip velocity change. Therefore, the time-variation of the hyperparameter shown in Fig. 16 is the expected result because of the relatively rapid acceleration and deceleration of slip during the slow slip event.

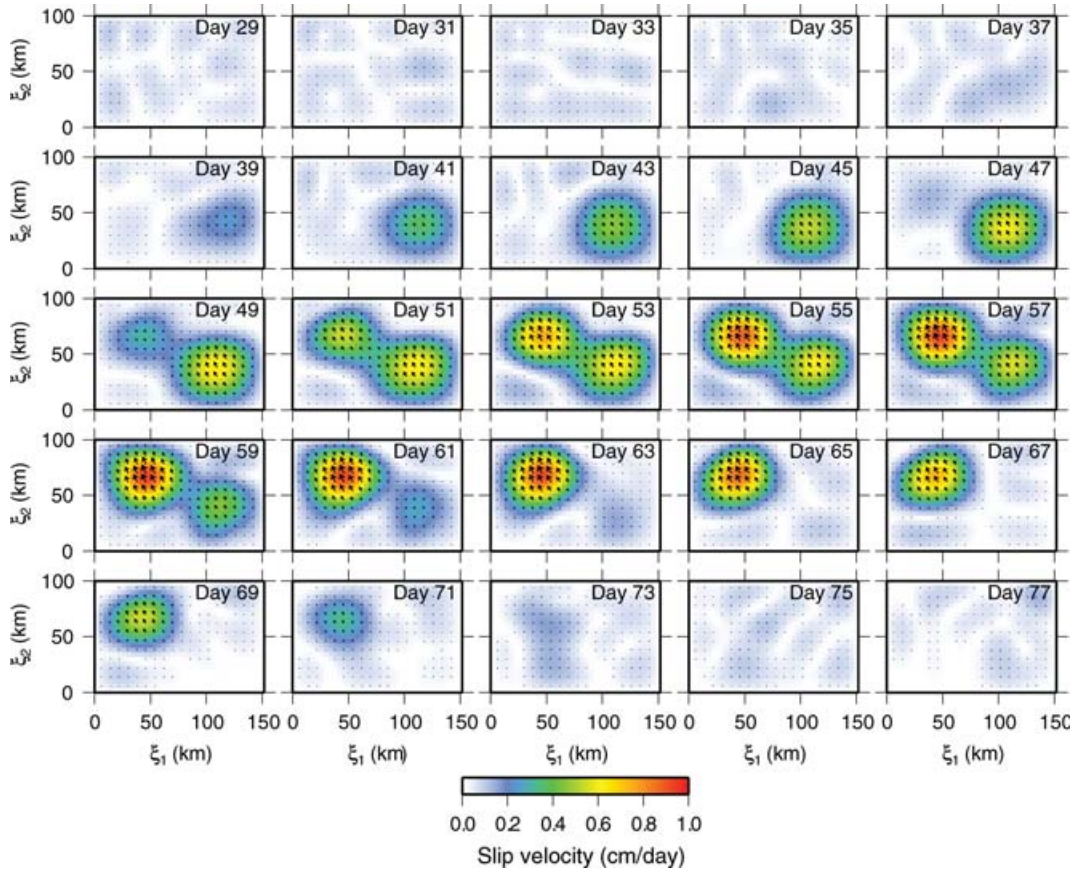

Figure 12. Slip velocity distributions for the period between Day 29 and 77 estimated by the conventional Kalman filter-based method with the constant hyperparameter.

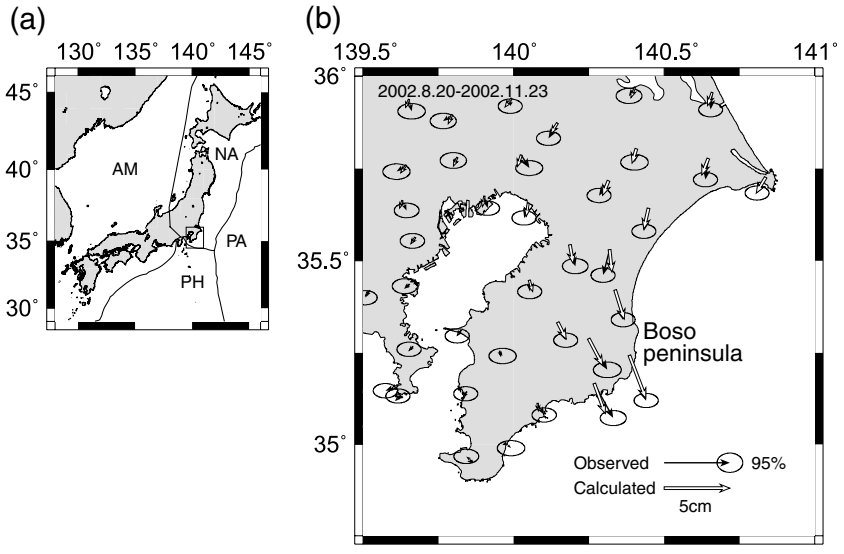

Figure 13. (a) Tectonic setting of the Japanese islands. Solid lines indicate plate boundaries. AM, PH, PA, and NA denote Amurian, Philippine Sea, Pacific and North American plates, respectively. (b) Magnified map of a rectangular area in (a). Black and white arrows show observed and calculated horizontal displacements, respectively, in ITRF 2000 for the period between 20 August 2002 and 23 November 2002. Error ellipses show 95 per cent confidence limits.

Fig. 17(a) shows the slip velocity distributions for the period between DOY 273 and DOY 287 (September 30–October 14), indicating that aseismic slip started on DOY 277 (October 4), accelerated and shifted southward until DOY 281 (October 8), and decelerated until DOY 287 (October 14). The aseismic slip continued after DOY 287 with much smaller slip velocity. The location of aseismic slip is consistent with the analysis of Ozawa *et al.* (2003). For comparison, we also apply the conventional Kalman filter-based method (see Section 2.3 for

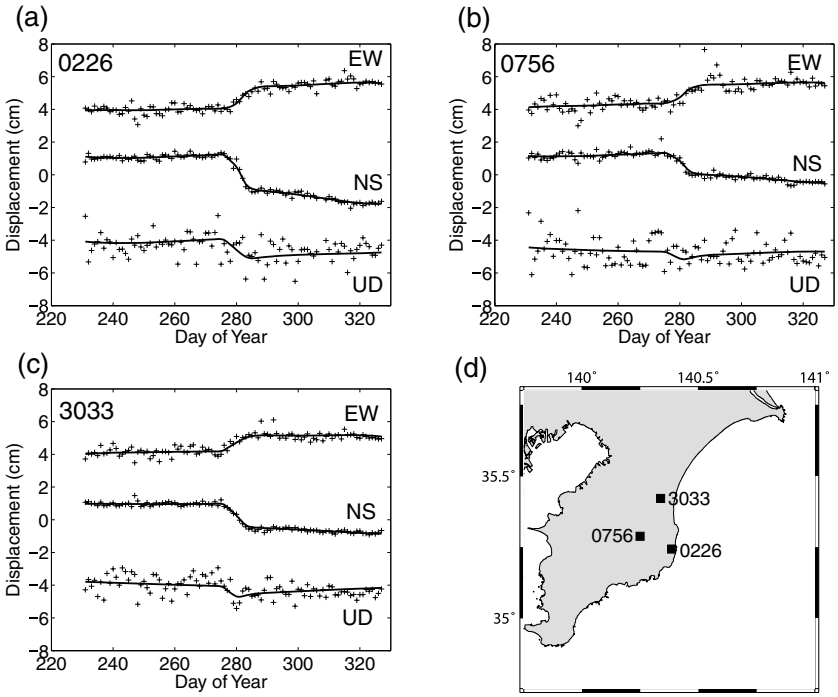

Figure 14. Observed and computed displacements at selected stations, (a) 0226, (b) 0756 and (c) 3033. Data are shown with pluses. Solid lines show computed displacements estimated by the new inversion method. Common mode errors are subtracted from both observed and computed displacements. (d) Locations of the selected stations.

Table 1. Fault parameters for the slow slip event off the east coast of the Boso peninsula. Latitude and longitude are position of the upper left-hand edge, D is depth of the upper edge, W is width, L is length, ϕ is strike measured clockwise from the north, and δ is dip angle of the fault plane, respectively.

Latitude (°N)	Longitude (°E)	D (km)	W (km)	L (km)	ϕ (°)	δ (°)
45.700	141.300	10.0	100.0	124.0	200.0	11.5

details) to estimate slip and slip velocity. In the Kalman filter-based analysis, we use the same number of spatial basis functions as in the new method ($M = 15$), and the hyperparameter α^2/σ^2 is optimized as $\alpha^2/\sigma^2 = 10^0$. Resulting slip velocity distributions are shown in Fig. 17(b). Comparison of Figs 17(a) and (b) shows the slip velocity distribution on DOY 281 (October 8) estimated by the new method is significantly larger than that estimated by the conventional method. The maximum slip velocity of 2.1 cm d^{-1} is obtained on DOY 281 by the new method, whereas that obtained by the conventional method is 1.2 cm d^{-1}. Figs 18(a)–(c) show the temporal variation of slip velocity magnitude at 3 points on the fault plane shown in Fig. 18(d). The slip velocity history obtained by the conventional method is rather oscillatory before and after the slow slip event, whereas the slip velocity transient is very smooth. Consequently, it is difficult to identify the onset time of the slow slip event. In contrast, the slip velocity history estimated by the new method is constant before and after the event and has larger maximum value compared with that estimated by the conventional method. Furthermore, the slip velocity estimated by the conventional method begins to accelerate around DOY 269 (September 26), whereas that estimated by the new method does on DOY 276 (October 3) (see Fig. 18a).

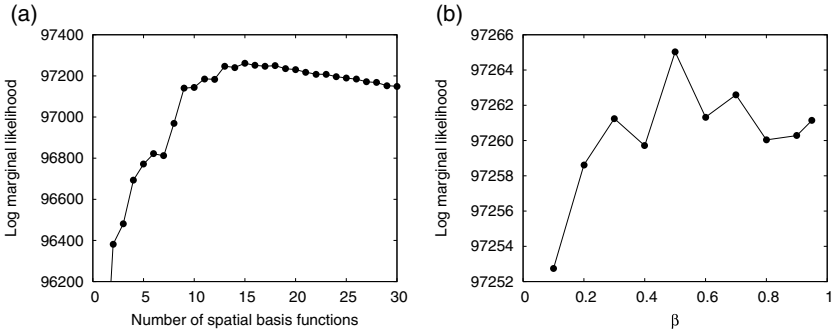

Figure 15. (a) Log marginal likelihood as a function of the number of spatial basis functions, M, for $\beta = 0.95$. (b) Log marginal likelihood as a function of β. The number of spatial basis functions is fixed to $M = 15$.

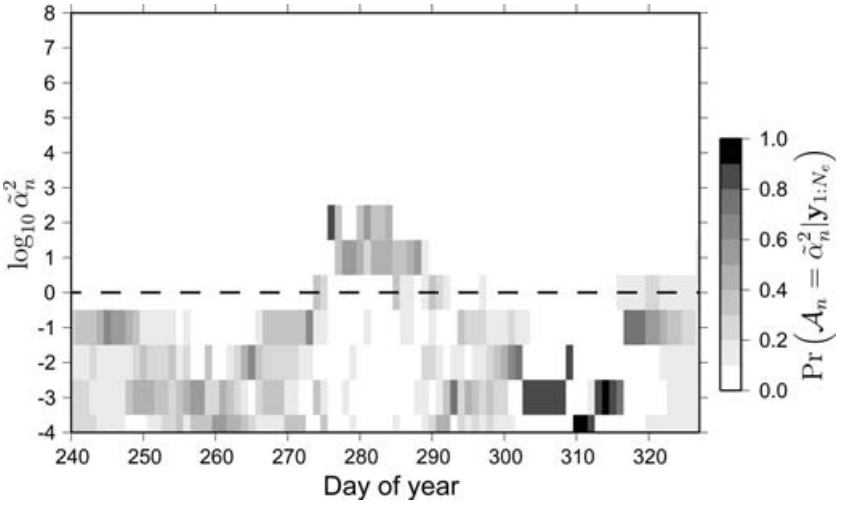

Figure 16. Estimated temporal variation of posterior probability distribution of the hyperparameter $\mathcal{A}_n$, $\Pr(\mathcal{A}_n = \tilde{\alpha}_n^2 | \mathbf{y}_{1:N_e})$. The dashed line shows the value of α^2/σ^2 that is used in the Kalman filter-based method.

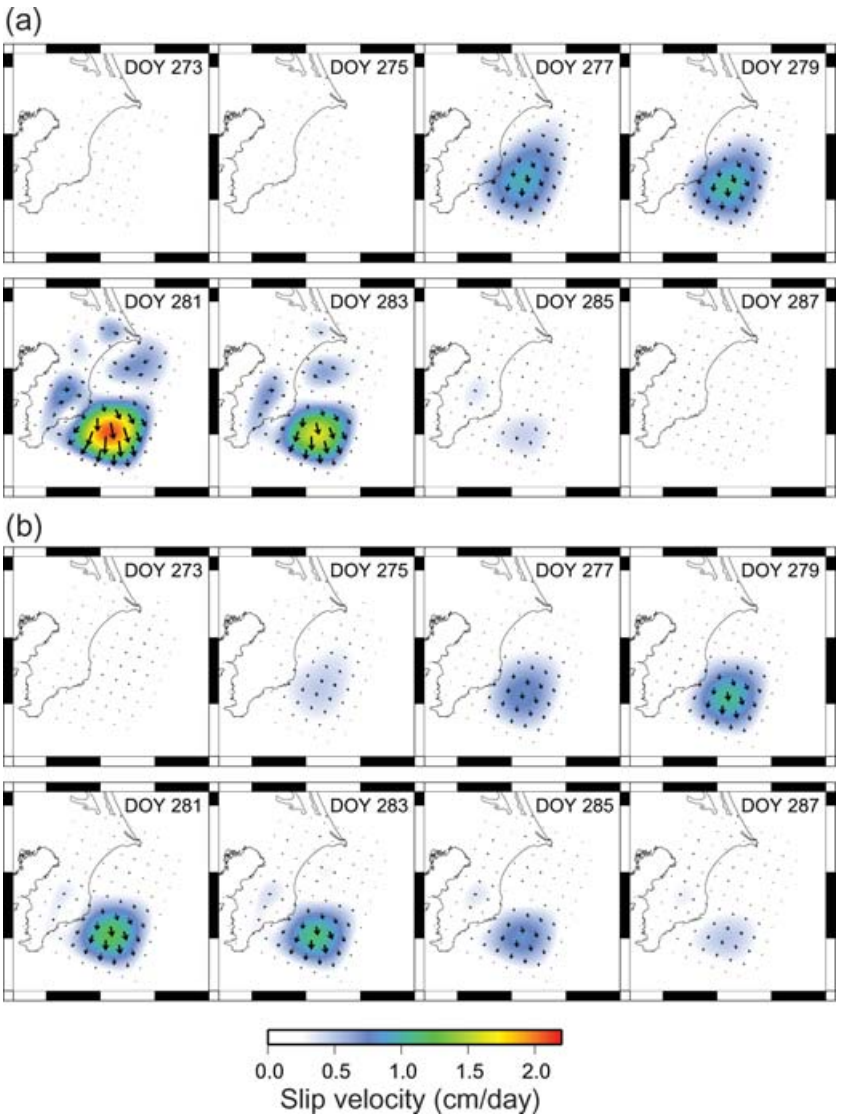

Figure 17. (a) Space–time distribution of slip velocity estimated by the new method with the time-varying hyperparameter $\mathcal{A}_n$. (b) Space–time distribution of slip velocity estimated by the conventional Kalman filter-based method with the constant hyperparameter.

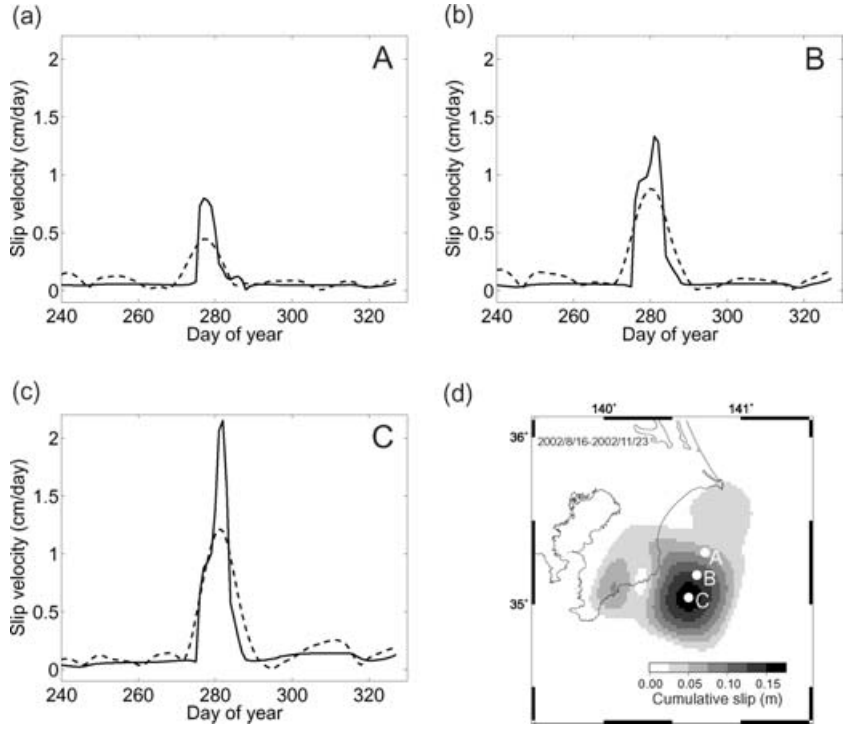

Figure 18. (a), (b) and (c) Temporal variation of slip velocity magnitude at points A, B and C on the fault plane shown in (d). The solid curve represents slip velocity obtained by the new method with time-varying hyperparameter $\mathcal{A}_n$, and the dashed curve denotes slip velocity obtained by the Kalman filter-based method with constant hyperparameter. (d) Cumulative slip inferred from the new method for the period between 2002 August 16 and 2002 November 23. White circles show locations of the points A, B and C.

Figs 17(a) and (b) show that the two methods give fairly different spatial distributions of slip velocity on DOY 281 and 283. Specifically, the new method images slip areas on the deeper and northern parts of the main slip area, whereas the conventional method does not. During the slow slip event, the hyperparameter obtained from the new method becomes larger than that obtained from the conventional method (Fig. 16). Therefore, temporal smoothing of slip velocity is weaker in the new method during the event. This weak temporal smoothing tends to make slip velocity oscillatory in regions of low slip. On the other hand, the conventional method gives stronger temporal smoothing during the event, thus oscillation of slip velocity in low slip regions is small. We think that this effect generates the deeper and northern slip areas on DOY 281 and 283 in Fig. 17(a), resulting in different slip velocity distribution from that shown in Fig. 17(b). This effect is clearly a limitation of our method and the deeper and northern slip areas in Fig. 17(a) are probably artefacts.

Fig. 13(b) shows a comparison of the observed and computed cumulative displacements for the period between 2002 August 20 and 2002 November 23, indicating that the fit is reasonable. Figs 14(a)–(c) show comparison of the observed and computed displacement time-series at selected stations shown in Fig. 14(d). The observed and computed time-series demonstrate good agreements. Fig. 19 shows time-series of displacements computed from the slip distributions and the random benchmark motions estimated by both methods for the period between DOY 260 and DOY 300, together with the data corrected for the common mode errors, at the selected stations (Fig. 14d). Although the computed displacements obtained from the conventional method generally fit the observed data, they oversmooth the transient displacements for several time-series (e.g. NS component time-series shown in Fig. 19). On the other hand, the computed displacements obtained from the new method better fit the transient displacements.

Results presented in Figs 17–19 are qualitatively similar to the numerical simulation shown in Section 3. The numerical simulation shows that the conventional method oversmoothes slip velocity changes associated with transient slip events on the one hand and undersmoothes slip velocity during steady-state periods on the other. In contrast, the new method can better reproduce the true slip velocity history because the time-varying hyperparameter $\mathcal{A}_n$ enables the estimates not to be oversmoothed or undersmoothed. In view of the simulation results, the slip velocity estimated by the new method is probably a better approximation to the true slip velocity than that estimated by the conventional method. The differences of the maximum slip velocity and the start time of the transient slip event between the two methods result from the different smoothing strategies employed in the two inversion methods.

Figs 20(a) and (b) show the estimated temporal variations of the hyperparameter for $\beta = 0.1$ and 0.9, respectively. Comparison of the time histories of the hyperparameter for $\beta = 0.1$ (Fig. 20a), $\beta = 0.5$ (Fig. 16) and $\beta = 0.9$ (Fig. 20b) shows that temporal roughness of the hyperparameter strongly depends on β. Fig. 20(c) shows slip velocity histories at point C in Fig. 18(d) for $\beta = 0.1$, 0.5 and 0.9. The slip velocity history for $\beta = 0.1$ is rougher than that for $\beta = 0.5$ at which the log marginal likelihood is maximized, whereas that for $\beta = 0.9$ is smoother. This result shows the importance of properly determining β to obtain the optimal solution.

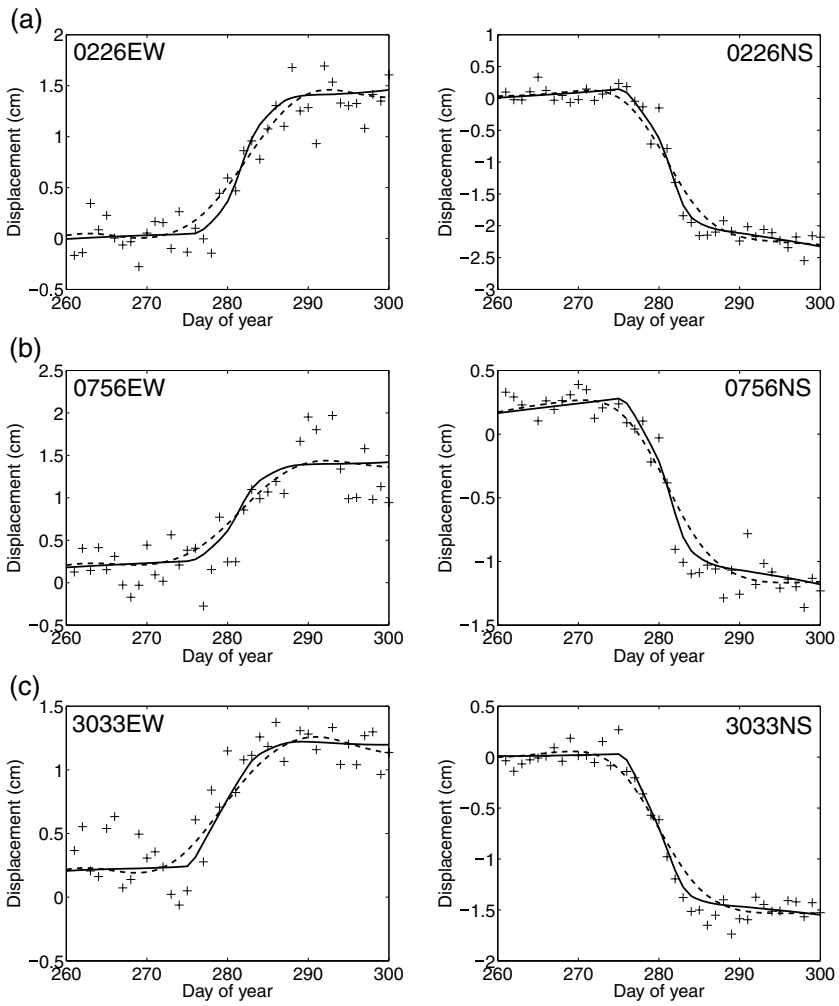

Figure 19. Comparison of observed displacements (pluses) and computed displacements estimated by the new method (solid line) and the conventional method (dashed line) for selected stations for the period between DOY 260 and 300. Common mode errors are subtracted from both observed and predicted data. Left- and right-hand panels show EW and NS components, respectively. Note that the vertical scale is different for each plot. (a) 0226, (b) 0756 and (c) 3033. Locations of the stations are shown in Fig. 14(d).

It has been reported that a swarm-like seismic activity occurred associated with the slow slip event around the estimated slip area. (Japan Meteorological Agency 2003; National Research Institute for Earth Science and Disaster Prevention 2003; Ozawa *et al.* 2003). This seismic activity started on DOY 274 (October 1) and high seismicity rate continued for 8 d, with the largest earthquake of magnitude 3.8 (Japan Meteorological Agency 2003). These earthquakes are considered to occur on the upper surface of the Philippine Sea plate based on hypocentre distribution and focal mechanisms (National Research Institute for Earth Science and Disaster Prevention 2003). We compare the onset times of the slow slip event obtained by the two inversion methods with the start time of the seismic activity. The onset time of the slow slip event obtained by the new inversion method, DOY 276 (October 3), indicates that the slow slip event started immediately after the onset of the seismic activity. On the other hand, the onset time of the slow slip event obtained by the conventional method is approximately DOY 269 (September 26), well before the onset of the seismic activity. This difference indicates that it is difficult to precisely determine the onset time of the slow slip event relative to the seismic activity using the conventional method. It is important to precisely determine the relative timing of the aseismic and seismic events because the relative timing may significantly affect the interpretation of mechanical relationships between them (Segall *et al.* 2006). Therefore, the new method may provide insights into the association of the slow slip event and the seismic activity.

5 CONCLUSION

We have developed an inversion method for estimating the distribution of fault slip velocity in space and time from geodetic data. We introduced a state space model that relates observed deformation to an aseismic fault slip event. In this model, temporal smoothness of slip velocity is controlled by a hyperparameter, which is assumed to be a constant in conventional geodetic inversion methods. The assumption of constant hyperparameter often causes oversmoothing of rapid slip velocity changes and/or undersmoothing of steady-state slip velocity,

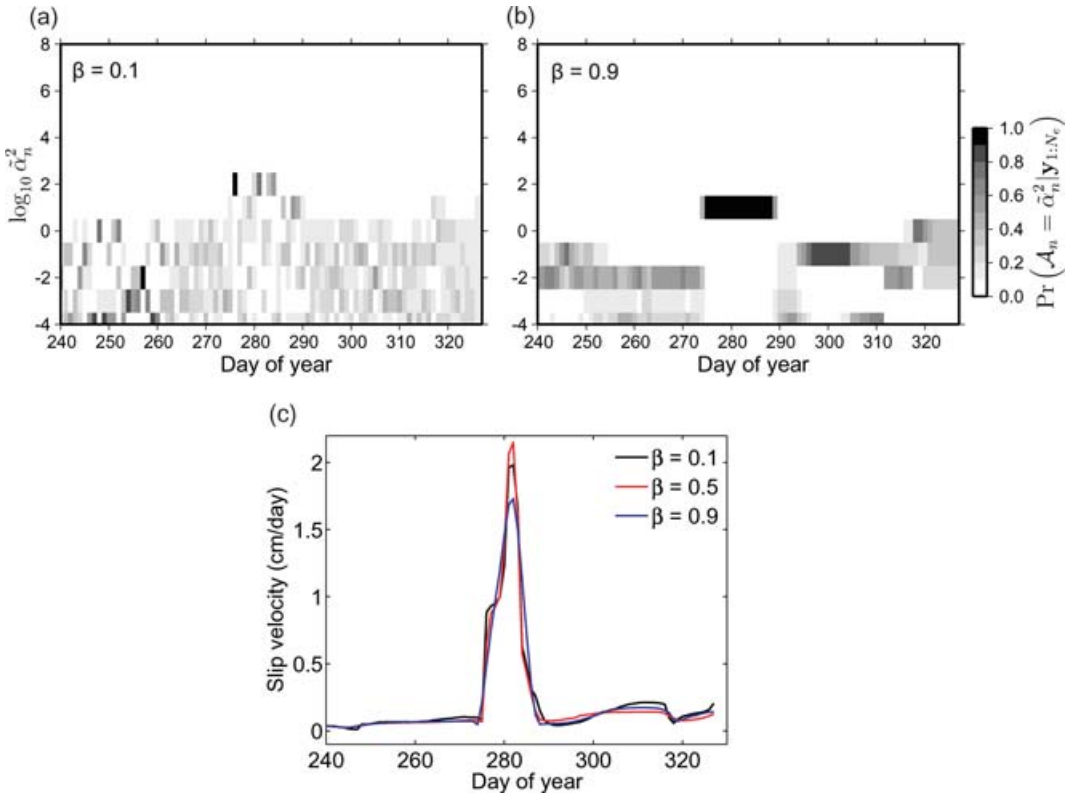

Figure 20. Temporal variations of the hyperparameter and slip velocity for different β values. (a) Estimated temporal variation of posterior probability distribution of the hyperparameter $\mathcal{A}_n$, $\Pr(\mathcal{A}_n = \tilde{\alpha}_n^2|\mathbf{y}_{1:N_e})$ for $\beta = 0.1$. (b) Same as (a) but for $\beta = 0.9$. (c) Temporal variations of slip velocity magnitude at point C in Fig. 18(d) for $\beta = 0.1$, 0.5, 0.9. Black, red and blue curves are slip velocity histories for $\beta = 0.1$, 0.5 and 0.9, respectively. The slip velocity history for $\beta = 0.5$ is the same as the solid curve in Fig. 18(c).

which obscures the estimated slip velocity history. In order to avoid temporally oversmoothed/undersmoothed slip velocity estimates caused by this assumption, we treated a hyperparameter as a time-varying stochastic variable. To estimate temporal variation of the hyperparameter, we employed the hierarchical Bayesian state space approach. Specifically, we introduced a stochastic model of hyperparameter evolution in addition to the state space model. Application of a Bayesian recursive filtering method, Monte Carlo mixture Kalman filter, yields the posterior distribution of the hyperparameter and slip velocity.

The validity of the new inversion method was examined through numerical simulations using synthetic data sets generated by simulated slow slip events. The results were compared with those obtained by the conventional method with constant hyperparameter. The space–time distribution of slip velocity estimated by the conventional method is temporally oversmoothed when slip velocity rapidly changes, whereas it is excessively oscillatory during steady-state periods due to undersmoothing. On the other hand, the new method does not undergo significant oversmoothing nor undersmoothing, and consequently better reproduces the true slip velocity.

Finally, we applied the new inversion method to transient deformation observed in the Boso peninsula, Japan, in October 2002, caused by a slow slip event. Estimated slip velocity was compared with that obtained by the conventional method with constant hyperparameter. The space–time distribution of slip velocity obtained by the conventional method is temporally smoother than that obtained by the new method during the slow slip event, whereas it is more oscillatory before and after the slow slip event. In view of the simulation results, we infer that slip velocity estimated by the conventional method suffers from the oversmoothing/undersmoothing problems, whereas that estimated by the new method does not. Therefore, the slip velocity estimated by the new method is probably better approximation to the true slip velocity. These results demonstrate that introducing the time-dependent smoothing regularization through the time-varying hyperparameter is effective to solve the oversmoothing/undersmoothing problems and consequently enables us to reconstruct detailed image of slow slip events.

ACKNOWLEDGMENTS

We thank the Geographical Survey Institute for providing the GPS data. We also thank Kaj Johnson and two anonymous reviewers for valuable comments. This study was carried out under the ISM Cooperative Research Program (2005-ISM·CPR-1026). For this study, we used the computer systems of the Earthquake Information Center of the Earthquake Research Institute, University of Tokyo, and the Institute of Statistical Mathematics. A software package, Generic Mapping Tools (Wessel & Smith 1998), was used to plot the figures.

REFERENCES

Altamimi, Z., Sillard, P. & Boucher, C., 2002. ITRF2000: a new release of the International Terrestrial Reference Frame for earth science applications, *J. geophys. Res.,* **107**(B10), 2214, doi:10.1029/2001JB000561.

Bürgmann, R., Kogan, M.G., Levin, V.E., Scholtz, C.H., King, R.W. & Steblov, G.M., 2001. Rapid aseismic moment release following the 5 December, 1997 Kronotsky, Kamchatka earthquake, *Geophys. Res. Lett.,* **28,** 1331–1334.

Dragert, H., Wang, K. & James, T.S., 2001. A silent slip event on the deeper Cascadia subduction interface, *Science,* **292,** 1525–1528.

Fukuda, J., Higuchi, T., Miyazaki, S. & Kato, T., 2004. A new approach to time-dependent inversion of geodetic data using a Monte Carlo mixture Kalman filter, *Geophys. J. Int.,* **159,** 17–39.

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,** 7945–7962.

Heki, K., Miyazaki, S. & Tsuji, H., 1997. Silent fault slip following an interplate thrust earthquake at the Japan Trench, *Nature,* **386,** 595–598.

Hirose, H., Hirahara, K., Kimata, F., Fujii, N. & Miyazaki, S., 1999. A slow thrust slip event following the two 1996 Hyuganada earthquake beneath the Bungo Channel, southwest Japan, *Geophys. Res. Lett.,* **26,** 3237–3240.

Hsu, Y.-J. *et al.*, 2006. Frictional afterslip following the 2005 Nias-Simeulue earthquake, Sumatra, *Science,* **312,** 1921–1926.

Ishida, M., 1992. Geometry and relative plate motion of the Philippine Sea plate and Pacific plate beneath the Kanto-Tokai district, Japan, *J. geophys. Res.,* **97,** 489–513.

Japan Meteorological Agency, 2003. Seismic activity in east off Chiba prefecture (October 2002) (in Japanese), *Rep. Coord. Comm. Earthquake Predict.,* **69,** 132–137.

Kalman, R.E., 1960. A new approach to linear filtering and prediction problems, *J. Basic Eng.,* **82,** 35–45.

Kitagawa, G., 1984. A smoothness priors-state space modeling of time series with trend and seasonality, *J. Am. Statist. Assoc.,* **79,** 378–389.

Kitagawa, G., 1987. Non-Gaussian state-space modeling of nonstationary time series, *J. Am. Statist. Assoc.,* **82,** 1032–1041.

Kitagawa, G., 1996. Monte Carlo filter and smoother for non-Gaussian nonlinear state space models, *J. Comput. Graph. Statist.,* **5,** 1–25.

Kitagawa, G. & Gersch, W., 1996. *Smoothness Priors Analysis of Time Series,* Springer-Verlag, New York.

Kostoglodov, V., Singh, S.K., Santiago, J.A., Franco, S.I., Larson, K.M., Lowry, A.R. & Bilham, R., 2003. A large silent earthquake in the Guerrero seismic gap, Mexico, *Geophys. Res. Lett.,* **30**(15), 1807, doi:10.1029/2003GL017219.

Langbein, J. & Johnson, H., 1997. Correlated errors in geodetic time series: Implications for time-dependent deformation, *J. geophys. Res.,* **102,** 591–603.

Linde, A.T., Gladwin, M.T., Johnston, M.J.S., Gwyther, R.L. & Bilham, R.G., 1996. A slow earthquake sequence on the San Andreas fault, *Nature,* **383,** 65–68.

MacKay, D.J.C., 1992. Bayesian interpolation, *Neural Comput.,* **4,** 415–447.

MacKay, D.J.C., 2003. *Information Theory, Inference, and Learning Algorithms,* Cambridge University Press, Cambridge.

McGuire, J.J. & Segall, P., 2003. Imaging of aseismic fault slip transients recorded by dense geodetic networks, *Geophys. J. Int.,* **155,** 778–788.

Miyazaki, S., Saito, T., Sasaki, M., Hatanaka, Y. & Iimura, Y., 1997. Expansion of the GSI's nation wide GPS array, *Bull. Geogr. Surv. Inst.,* **43,** 23–34.

Miyazaki, S., McGuire, J.J. & Segall, P., 2003. A transient subduction zone slip episode in southwest Japan observed by the nationwide GPS array, *J. geophys. Res.,* **108**(B2), 2087, doi:10.1029/2001JB000456.

Miyazaki, S., Segall, P., Fukuda, J. & Kato, T., 2004. Space time distribution of afterslip following the 2003 Tokachi-oki earthquake: implications for variations in fault zone frictional properties, *Geophys. Res. Lett.,* **31,** L06623, doi:10.1029/2003GL019410.

Miyazaki, S., Segall, P., McGuire, J.J., Kato, T. & Hatanaka, Y., 2006a. Spatial and temporal evolution of stress and slip rate during the 2000 Tokai slow earthquake, *J. geophys. Res.,* **111,** B03409, doi:10.1029/2004JB003426.

Miyazaki, S., Segall, P., Fukuda, J., Johnson, K.M. & Kato, T., 2006b. Postseismic deformation following two thrust earthquakes at Kurile-Japan Trench: The 2003 Tokachi-oki and the 2005 Miyagi-oki earthquakes, *EOS Trans. AGU,* **87**(52), Fall Meet. Suppl., Abstract G31A-05.

Murakami, M., Suito, H., Ozawa, S. & Kaidzu, M., 2006. Earthquake triggering by migrating slow slip initiated by M8 earthquake along Kuril Trench, Japan, *Geophys. Res. Lett.,* **33,** L09306, doi:10.1029/2006GL025967.

Murray, J.R. & Segall, P., 2005. Spatiotemporal evolution of a transient slip event on the San Andreas fault near Parkfield, California, *J. geophys. Res.,* **110,** B09407, doi:10.1029/2005JB003651.

National Research Institute for Earth Science and Disaster Prevention, 2003. Earthquake swarm activity east off Boso peninsula in Oct and Nov, 2002 (in Japanese), *Rep. Coord. Comm. Earthquake Predict.,* **69,** 127–131.

Obara, K., Hirose, H., Yamamizu, F. & Kasahara, K., 2004. Episodic slow slip events accompanied by non-volcanic tremors in southwest Japan subduction zone, *Geophys. Res. Lett.,* **31,** L23602, doi:10.1029/2004GL020848.

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

Ozawa, S., Murakami, M. & Tada, T., 2001. Time-dependent inversion study of the slow thrust event in the Nankai trough subduction zone, southwest Japan, *J. geophys. Res.,* **106,** 787–802.

Ozawa, S., Murakami, M., Kaidzu, M., Tada, T., Sagiya, T., Hatanaka, Y., Yarai, H. & Nishimura, T., 2002. Detection and monitoring of ongoing aseismic slip in the Tokai region, central Japan, *Science,* **298,** 1009–1012.

Ozawa, S., Miyazaki, S., Hatanaka, Y., Imakiire, T., Kaidzu, M. & Murakami, M., 2003. Characteristic silent earthquakes in the eastern part of the Boso peninsula, central Japan, *Geophys. Res. Lett.,* **30**(6), 1283, doi:10.1029/2002GL016665.

Perfettini, H. & Avouac, J.-P., 2004. Postseismic relaxation driven by brittle creep: a possible mechanism to reconcile geodetic measurements and the decay rate of aftershocks, application to the Chi-Chi earthquake, Taiwan, *J. geophys. Res.,* **109,** B02304, doi:10.1029/2003JB002488.

Perfettini, H., Avouac, J.-P. & Ruegg, J.-C., 2005. Geodetic displacements and aftershocks following the 2001 $M_W = 8.4$ Peru earthquake: Implications for the mechanics of the earthquake cycle along subduction zones, *J. geophys. Res.,* **110,** B09404, doi:10.1029/2004JB003522.

Pritchard, M.E. & Simons, M., 2006. An aseismic slip pulse in northern Chile and along-strike variations in seismogenic behavior, *J. geophys. Res.,* **111,** B08405, doi:10.1029/2006JB004258.

Rogers, G. & Dragert, H., 2003. Episodic tremor and slip on the Cascadia subduction zone: the chatter of silent slip, *Science,* **300,** 1942–1943.

Sagiya, T., 2004. Interplate coupling in the Kanto district, central Japan, and the Boso peninsula silent earthquake in May 1996, *Pure appl. Geophys.,* **161,** 2327–2342.

Schaff, D.P., Beroza, G.C. & Shaw, B.E., 1998. Postseismic response of repeating aftershocks, *Geophys. Res. Lett.,* **25,** 4549–4552.

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

Segall, P., Bürgmann, R. & Matthews, M., 2000. Time dependent triggered afterslip following the 1989 Loma Prieta earthquake, *J. geophys. Res.,* **105,** 5615–5634.

Segall, P., Desmarais, E.K., Shelly, D., Miklius, A. & Cervelli, P., 2006. Earthquakes triggered by silent slip events on Kilauea volcano, Hawaii, *Nature,* **442,** 71–74.

Wdowinski, S., Bock, Y., Zhang, J., Fang, P. & Genrich, J., 1997. Southern California permanent GPS geodetic array: Spatial filtering of daily positions for estimating coseismic and postseismic displacements induced by the 1992 Landers earthquake, *J. geophys. Res.,* **102,** 18 057–18 070.

Wessel, P. & Smith, W.H.F., 1998. New, improved version of Generic Mapping Tools released, *EOS, Trans. Am. geophys. Un.,* **79,** 579.

Zumberge, J.F., Heflin, M.B., Jefferson, D.C., Watkins, M.M. & Webb, F.H., 1997. Precise point positioning for the efficient and robust analysis of GPS data from large network, *J. geophys. Res.,* **102,** 5005–5017.

APPENDIX A: OBSERVATION MODEL

In this appendix, we give an explicit expression of the state space representation of the observation model. We divide the fault surface $A(\boldsymbol{\xi})$ into M_f rectangular subfaults and assume that slip is spatially uniform in each subfault. In this case, the observation model (1) can be expressed as

$$u_i(\mathbf{r}_j, t_n) = \sum_{k=1}^{2}\sum_{l=1}^{M_f} a_{kl}(t_n)\mathcal{G}_{ikl}(\mathbf{r}_j) + \mathcal{L}_i(\mathbf{r}_j, t_n) + f_i(t_n) + e_{ijn}, \tag{A1}$$

where $a_{kl}(t_n)$ is kth component of slip on lth subfault at time t_n, and $\mathcal{G}_{ikl}(\mathbf{r}_j)$ relates slip $a_{kl}(t_n)$ to surface displacement $u_i(\mathbf{r}_j, t_n)$ and is computed by the analytical expressions given by Okada (1985). From (A1), the observation model at time $t = t_n$ is represented in vector form using a data vector $\mathbf{y}_n$, slip distribution $\mathbf{a}_n$, random benchmark motion $\mathbf{l}_n$, common mode error $\mathbf{f}_n$ and observation noise $\mathbf{w}_n$ as follows:

$$\mathbf{y}_n = \Gamma\mathbf{a}_n + \mathbf{l}_n + \mathbf{f}_n + \mathbf{w}_n, \quad \mathbf{w}_n \sim N(\mathbf{0}, \mathbf{R}_n), \tag{A2}$$

where

$$\mathbf{y}_n = [u_1(\mathbf{r}_1, t_n), u_2(\mathbf{r}_1, t_n), u_3(\mathbf{r}_1, t_n), u_1(\mathbf{r}_2, t_n), u_2(\mathbf{r}_2, t_n), u_3(\mathbf{r}_2, t_n), \ldots, u_1(\mathbf{r}_{N_s}, t_n), u_2(\mathbf{r}_{N_s}, t_n), u_3(\mathbf{r}_{N_s}, t_n)]^T, \tag{A3}$$

$$\mathbf{a}_n = [a_{11}(t_n), a_{21}(t_n), a_{12}(t_n), a_{22}(t_n), \ldots, a_{1M_f}(t_n), a_{2M_f}(t_n)]^T, \tag{A4}$$

$$\mathbf{l}_n = [\mathcal{L}_1(\mathbf{r}_1, t_n), \mathcal{L}_2(\mathbf{r}_1, t_n), \mathcal{L}_3(\mathbf{r}_1, t_n), \mathcal{L}_1(\mathbf{r}_2, t_n), \mathcal{L}_2(\mathbf{r}_2, t_n), \mathcal{L}_3(\mathbf{r}_2, t_n), \ldots, \mathcal{L}_1(\mathbf{r}_{N_s}, t_n), \mathcal{L}_2(\mathbf{r}_{N_s}, t_n), \mathcal{L}_3(\mathbf{r}_{N_s}, t_n)]^T, \tag{A5}$$

$$\mathbf{f}_n = [f_1(t_n), f_2(t_n), f_3(t_n), f_1(t_n), f_2(t_n), f_3(t_n), \ldots, f_1(t_n), f_2(t_n), f_3(t_n)]^T, \tag{A6}$$

$$\mathbf{w}_n = [e_{11n}, e_{21n}, e_{31n}, e_{12n}, e_{22n}, e_{32n}, \ldots, e_{1N_sn}, e_{2N_sn}, e_{3N_sn}]^T, \tag{A7}$$

and

$$\mathbf{R}_n = \sigma^2\Sigma_n. \tag{A8}$$

Here N_s is the number of observation sites. $\boldsymbol{\Gamma}$ is a $3N_s \times 2M_f$ matrix whose pq component is defined by

$$\Gamma_{pq} = \mathcal{G}_{ikl}(\mathbf{r}_j) \tag{A9}$$

with

$$p = 3(j-1) + i \tag{A10}$$

$$q = 2(l-1) + k. \tag{A11}$$

We represent slip distribution $\mathbf{a}_n$ using the spatial basis functions as in eq. (2). In what follows, we present an explicit expression of the spatial basis functions (Segall *et al.* 2000) following Harris & Segall (1987) and Segall *et al.* (2000), and then rewrite the observation model (A2) in state space form. Harris & Segall (1987) introduced a transformation of $\boldsymbol{\Gamma}$,

$$\tilde{\Gamma} = \Gamma\mathbf{T}^{-1}, \tag{A12}$$

where $\mathbf{T}$ is a $2M_f \times 2M_f$ matrix and is taken to be the finite difference approximation to the Laplacian operator, and applied the SVD to $\tilde{\Gamma}$, $\tilde{\Gamma} = \mathbf{U}\Lambda\mathbf{V}^T$, where $\mathbf{U}$ is a $3N_s \times 3N_s$ matrix of eigenvectors that span the data space, $\mathbf{V}$ is a $2M_f \times 2M_f$ matrix of eigenvectors spanning the model space, and $\boldsymbol{\Lambda}$ is a $3N_s \times 2M_f$ matrix containing the singular values $\lambda_1, \lambda_2, \ldots, \lambda_{3N_s}$ as diagonal elements. Here the singular values are assumed to be ranked as $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_{3N_s}$. In this case, the observation model (A2) reduces to

$$\mathbf{y}_n = \tilde{\Gamma}\mathbf{T}\mathbf{a}_n + \mathbf{l}_n + \mathbf{f}_n + \mathbf{w}_n, \quad \mathbf{w}_n \sim N(\mathbf{0}, \mathbf{R}_n). \tag{A13}$$

Harris & Segall (1987) showed that slip distribution that minimize both data misfit, $\|\mathbf{y}_n - \Gamma\mathbf{a}_n\|^2$, and roughness of spatial slip distribution, $\|\mathbf{T}\mathbf{a}_n\|^2$, is obtained by replacing $\tilde{\Gamma} = \mathbf{U}\Lambda\mathbf{V}^T$ in (A13) with

$$\tilde{\Gamma}_M = \mathbf{U}_M\Lambda_M\mathbf{V}_M^T, \tag{A14}$$

where $\mathbf{U}_M$ is a $3N_s \times M$ matrix which consists of first M column of the matrix $\mathbf{U}$, $\mathbf{V}_M$ is a $2M_f \times M$ matrix which consists of first M column of the matrix $\mathbf{V}$, and Λ_M is given by $\Lambda_M = \mathrm{diag}[\lambda_1, \lambda_2, \ldots, \lambda_M]$. Based on this solution, Segall *et al.* (2000) introduced the spatial basis functions

$$\mathbf{B} = \mathbf{T}^{-1}\mathbf{V}_M\Lambda_M^{-1}. \tag{A15}$$

They showed that the squared roughness norm of the slip distribution, $\|\mathbf{T}\mathbf{a}_n\|^2$, is expressed by

$$\|\mathbf{T}\mathbf{a}_n\|^2 = \mathbf{c}_n^T\Lambda_M^{-2}\mathbf{c}_n. \tag{A16}$$

This means that increasing the number of singular values, M, in the spatial basis functions increases the squared roughness norm of the slip distribution. Therefore, spatial smoothing can be implemented by ignoring large singular values.

In this study, we employ the spatial basis set (A15) in order to obtain a solution that is spatially smooth to some degree. Using (2), (A12), (A14) and (A15), $\Gamma\mathbf{a}_n$ in (A2) can be expressed by $\Gamma\mathbf{a}_n = \mathbf{U}_M\mathbf{c}_n$. Substituting this relation into the observation model (A2) yields

$$\mathbf{y}_n = \mathbf{U}_M\mathbf{c}_n + \mathbf{l}_n + \mathbf{f}_n + \mathbf{w}_n, \quad \mathbf{w}_n \sim N(\mathbf{0}, \mathbf{R}_n). \tag{A17}$$

The observation model (A17) can be expressed in state space form as follows:

$$\mathbf{y}_n = \mathbf{H}_n\mathbf{x}_n + \mathbf{w}_n, \quad \mathbf{w}_n \sim N(\mathbf{0}, \mathbf{R}_n), \tag{A18}$$

where

$$\begin{aligned}\mathbf{x}_n = [&c_1(t_n), \dot{c}_1(t_n), c_2(t_n), \dot{c}_2(t_n), \ldots, c_M(t_n), \dot{c}_M(t_n),\\ &\mathcal{L}_1(\mathbf{r}_1, t_n), \mathcal{L}_2(\mathbf{r}_1, t_n), \mathcal{L}_3(\mathbf{r}_1, t_n), \ldots, \mathcal{L}_1(\mathbf{r}_{N_s}, t_n), \mathcal{L}_2(\mathbf{r}_{N_s}, t_n), \mathcal{L}_3(\mathbf{r}_{N_s}, t_n),\\ &f_1(t_n), f_2(t_n), f_3(t_n)]^T\end{aligned} \tag{A19}$$

$$\mathbf{H}_n = \left[\mathbf{U}_M\mathbf{J}\ \mathbf{I}_{3N_s\times 3N_s}\ \mathbf{L}\right] \tag{A20}$$

$$\mathbf{J} = \begin{bmatrix} 1 & 0 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 0 & 1 & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \ldots & 1 & 0 \end{bmatrix} \tag{A21}$$

$$\mathbf{L} = \left[\mathbf{I}_{3\times 3}\ \mathbf{I}_{3\times 3}\ \ldots\ \mathbf{I}_{3\times 3}\right]^T. \tag{A22}$$

Here $c_k(t_n)$ is kth component of $\mathbf{c}_n$ and $\dot{c}_k(t_n)$ is the first derivative of $c_k(t_n)$ with respect to t. $\mathbf{H}_n$, $\mathbf{J}$, and $\mathbf{L}$ is $3N_s \times (2M + 3N_s + 3)$, $M \times 2M$, and $3N_s \times 3$ matrices, respectively. $\mathbf{y}_n$, $\mathbf{w}_n$, and $\mathbf{R}_n$ are defined in (A3), (A7) and (A8), respectively. $\mathbf{I}_{3N_s\times 3N_s}$ and $\mathbf{I}_{3\times 3}$ are $3N_s \times 3N_s$ and 3×3 identity matrices, respectively.