Geophys. J. Int. (2004) **159,** 17–39 doi: 10.1111/j.1365-246X.2004.02383.x

A new approach to time-dependent inversion of geodetic data using a Monte Carlo mixture Kalman filter

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

[1] *Earthquake Research Institute, University of Tokyo, Yayoi* 1-1-1, *Bunkyo-ku, Tokyo* 113-0032, *Japan. E-mail: fukuda@eri.u-tokyo.ac.jp*
[2] *The Institute of Statistical Mathematics,* 4-6-7 *Minami-Azabu, Minato-ku, Tokyo* 106-8569, *Japan*

Accepted 2004 June 8. Received 2004 May 31; in original form 2003 April 24

SUMMARY

In the last decade, continuous Global Positioning System (GPS) networks have observed transient crustal deformation associated with various types of aseismic fault-slip event in many subduction zones. It is important to precisely clarify the entire time history of these events to understand the physical process of earthquake generation. For this purpose, we have developed a new time-dependent inversion method for imaging transient fault slips from geodetic data. Segall & Matthews (1997) presented a time-dependent inversion method to infer the spatiotemporal distribution of fault slip from geodetic data. They modelled a transient crustal deformation associated with fault-slip events using a linear Gaussian state space model and employed a Kalman filter. They introduced a scaling parameter that represents the temporal smoothness of the fault slip, and assumed that the scaling parameter is constant over the observation period. Under this assumption, abrupt changes of slip have been overly smoothed, whereas estimated slips in a 'quiet' steady-state period have been oscillatory. To improve the method, we developed a new filtering technique, a Monte Carlo mixture Kalman filter (MCMKF), and apply it to time-dependent inversion. The MCMKF allows variations of the temporal smoothness of slips by regarding it as a stochastic variable. The MCMKF is based on a Monte Carlo method in which conditional probability density functions of the stochastic variable are estimated recursively. We examine the validity of the introduced MCMKF-based inversion scheme through numerical experiments using simulated displacement time-series. Then, the results are compared with those obtained by a conventional Kalman filter-based scheme. In all cases, MCMKF gives a significantly smaller Akaike information criterion (AIC) values than the Kalman filter. This indicates that MCMKF yields better state estimates than the Kalman filter. We also find that MCMKF is capable of imaging the initiation process of transient slip events in cases with a high signal-to-noise ratio, while the Kalman filter is not. Furthermore, MCMKF is superior to the Kalman filter in detecting small signals from noisy data sets. From all of the results above, we conclude that the new filtering approach introduced here may provide a powerful tool for imaging the time history of fault slips.

Key words: conditional dynamic linear model, fault slip, geodetic data, Monte Carlo mixture Kalman filter, time-dependent inversion, transient crustal deformation.

1 INTRODUCTION

Continuous measurements of surface deformation with dense Global Positioning System (GPS) networks have revealed that transient crustal deformations, the durations of which vary from hours to years, play a very important role in seismic cycles as well as the mechanism of rupture propagation. For example, the 1994 Sanriku-Haruka-Oki earthquake was accompanied by a significant afterslip and almost the same amount of moment was released by the decaying transient slip during the following year (e.g. Heki *et al.* 1997). Besides such post-seismic transient events, we have witnessed other slow slip events in the last decade, which include slow thrust slip events in central Japan (e.g. Ozawa *et al.* 2002), southwest Japan (e.g. Hirose *et al.* 1999) and the Pacific Northwest (e.g. Dragert *et al.* 2001). These transient events may be responsible for the discrepancy between two estimations of seismic coupling, one derived from the seismic slip and recurrence interval (e.g. Pacheco *et al.* 1993) and the other inverted from GPS station velocities (e.g. Ito *et al.* 2000). Yagi (2002) recently showed that the area of seismic rupture, called the asperity, is complementary to that of aseismic slow slip. Heterogeneous frictional and mechanical conditions may be responsible for such variations of rupture modes. Accurate estimates of spatiotemporal characteristics of the subducting plate interface are very important

for understanding how large interplate earthquakes or slow events occur. From a geodetic viewpoint, it is important to investigate the detailed spatiotemporal process of slow events and dense GPS array records provide the most suitable data for this end.

Several studies have tried to image the spatiotemporal variation of transient fault slip. Segall & Matthews (1997) presented one efficient way to retrieve slip distribution. They modelled the space–time history of fault slip using a linear Gaussian state space model (e.g. Kitagawa & Gersch 1996), and employed a Kalman filter (e.g. Kitagawa & Gersch 1996). This method is referred to as the network inversion filter (NIF). The NIF has several advantages over other techniques. First, the NIF employs a non-parametric description of fault-slip evolution. Thus we do not need to know *a priori* the nature of time dependence of the transient slip. Second, the NIF models random benchmark motions (Langbein & Johnson 1997; Wyatt 1982) in a stochastic manner. Accordingly, it can distinguish spatially coherent transient signals from spatially incoherent random benchmark motions. Their Kalman filter-based method has been updated and applied to several cases (e.g. Aoki *et al.* 1999; Segall *et al.* 2000; Ozawa *et al.* 2001, 2002; Bürgmann *et al.* 2001; Miyazaki *et al.* 2003; McGuire & Segall 2003).

In the Segall–Matthews model, a stochastic model is used for the non-parametric description of fault-slip evolution. The temporal smoothness of the fault slip is controlled by an optimized scaling parameter of the stochastic model employed. This scaling parameter, often referred to as a hyperparameter, is optimized with the maximum-likelihood method. In the Kalman filter framework, the scaling parameter is kept fixed over the observation period. However, the temporally invariable scaling parameter could not trace abrupt changes because the optimized scaling parameter is too small to allow such a sudden change of fault slip, and vice versa. As a result, estimated slip evolution is flattened during the event and is oscillatory in the steady-state period; hence, it is hardly possible to identify the initiation of events. One of the clear manifestations of this difficulty can be found in the work of Ozawa *et al.* (2001) and Miyazaki *et al.* (2003). In 1996 and 1997, transient crustal deformation was observed in southwest Japan by the Japanese nationwide GPS network Hirose *et al.* (1999). Hirose *et al.* (1999) attributed the deformation to afterslips of two Hyuganada earthquakes, both $M_w = 6.7$, which occurred on 1996 October 19 and December 2, and a slow slip event, which occurred beneath Bungo Channel in 1997. Ozawa *et al.* (2001) and Miyazaki *et al.* (2003) investigated spatiotemporal slip distribution based on the NIF. Ozawa *et al.* (2001) found that the slow slip event started shortly before or immediately after the first Hyuganada earthquake, whereas Miyazaki *et al.* (2003) found that the event was initiated 1 month after the second Hyuganada earthquake. This difference may stem from the assumption of constancy for the scaling parameter in both cases and a difference in its optimization strategy. The constancy of the scaling parameter seems to obscure the causal relationship among multiple events, and hence motivated us to explore a new approach to time-dependent inversion such that the scaling parameter is variable over time. This generalization may allow us to uncover small transient phenomena hidden in noisy time-series and help us to understand the mechanism of the earthquake generation process.

In the present study, we developed a new filtering algorithm called the Monte Carlo mixture Kalman filter (MCMKF) for imaging time-dependent fault slip from geodetic data. This filter is capable of optimizing the temporal variation of the scaling parameter. In Section 2 we formulate forward models for a transient deformation. The new filtering method is shown in Section 3. In Section 4, the validity of the proposed method is examined through numerical experiments using simulated data.

2 MODELS FOR TRANSIENT DEFORMATION

We employ state space representation (Anderson & Moore 1979; Kitagawa & Gersch 1996) to model transient deformation. State space representation consists of an observation model and a system model. The observation model relates fault slip at depth to surface displacements and the system model describes the time evolution of the fault slip based on system dynamics.

2.1 Observation model

We consider a tangential displacement discontinuity $\mathbf{s}(\boldsymbol{\xi}, t)$ 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 as shown in Fig. 1. Let $\mathbf{u}(\mathbf{x}, t)$ be a cumulative surface displacement as a function of station position $\mathbf{x}$ and time

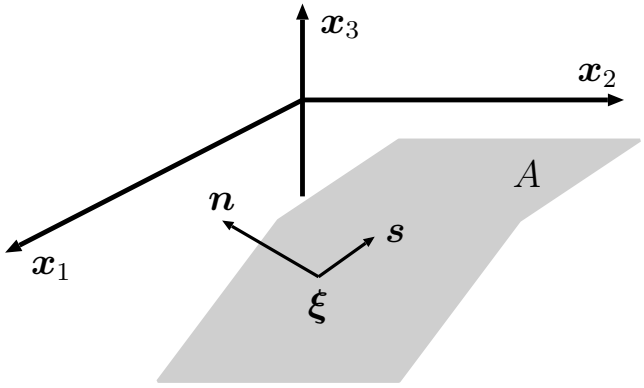

Figure 1. A schematic diagram showing the coordinate system and the geometry of a fault surface. A denotes a fault surface embedded in a homogeneous, isotropic, elastic half-space ($x_3 \leq 0$). **s** is a slip vector and **n** is a unit normal to the fault surface A.

of measurement t. We model the surface displacements at $\mathbf{x} = \mathbf{x}_j$ and $t = t_n$ by

$$u_i(\mathbf{x}_j, t_n) = \mu \sum_{p=1}^{3} \sum_{q=1}^{3} \int_A s_q(\boldsymbol{\xi}, t_n)[G_{ip,q}(\mathbf{x}_j, \boldsymbol{\xi}) + G_{iq,p}(\mathbf{x}_j, \boldsymbol{\xi})]n_p(\boldsymbol{\xi})\, dA(\boldsymbol{\xi}) + \mathcal{L}_i(\mathbf{x}_j, t_n) + e_{ijn} \tag{1}$$

where the left-hand side is the ith component of $\mathbf{u}(\mathbf{x}_j, t_n)$.

The first term on the right-hand side of eq. (1) represents the surface displacement due to cumulative slip $\mathbf{s}(\boldsymbol{\xi}, t)$ on the fault surface $A(\boldsymbol{\xi})$, where $G_{ip,q}(\mathbf{x}, \boldsymbol{\xi})$ is a spatial derivative of Green's tensor $G_{ip}(\mathbf{x}, \boldsymbol{\xi})$ with respect to ξ_q, $n_p(\boldsymbol{\xi})$ is the pth component of a unit normal to the fault surface and μ is the rigidity of the medium. The analytical expressions of $G_{ip,q}$ are given by Okada (1992).

The second term on the right-hand side of eq. (1), $\mathcal{L}_i(\mathbf{x}_j, t_n)$, represents temporally correlated noise in measurement time-series (Langbein & Johnson 1997; Wyatt 1982, 1989). This noise results from small random motions of geodetic monuments, due mainly to interactions between surface soils and weather (Langbein & Johnson 1997). Thus, the effect is expected to be spatially incoherent, whereas tectonic deformation is spatially coherent. Segall & Matthews (1997), who incorporated random benchmark motion into the NIF, succeeded in separating it from tectonic deformation. Therefore, we also incorporate random benchmark motion into our models.

The final term of eq. (1), e_{ijn}, represents the observational 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 hyperparameter. In the case of GPS data, Σ_n is a covariance matrix of GPS site coordinates derived from GPS data analysis. Because the analysis 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.

In the present study, we only consider shear dislocation. The slip vector $\mathbf{s}(\boldsymbol{\xi}, t)$ is represented in terms of two unit basis vectors $\boldsymbol{\nu}_1(\boldsymbol{\xi})$ and $\boldsymbol{\nu}_2(\boldsymbol{\xi})$ on the fault surface as follows:

$$\mathbf{s}(\boldsymbol{\xi}, t) = s_1'(\boldsymbol{\xi}, t)\boldsymbol{\nu}_1(\boldsymbol{\xi}) + s_2'(\boldsymbol{\xi}, t)\boldsymbol{\nu}_2(\boldsymbol{\xi}). \tag{2}$$

At the same time, $\mathbf{s}(\boldsymbol{\xi}, t)$ is written using a set of arbitrarily selected orthonormal basis vectors $\mathbf{e}_i$ $(i = 1, 2, 3)$ as follows:

$$\mathbf{s}(\boldsymbol{\xi}, t) = s_1(\boldsymbol{\xi}, t)\mathbf{e}_1 + s_2(\boldsymbol{\xi}, t)\mathbf{e}_2 + s_3(\boldsymbol{\xi}, t)\mathbf{e}_3. \tag{3}$$

Combining eqs (2) and (3) yields

$$s_q(\boldsymbol{\xi}, t) = \sum_{r=1}^{2} s_r'(\boldsymbol{\xi}, t)\boldsymbol{\nu}_r(\boldsymbol{\xi}) \cdot \mathbf{e}_q, \quad q = 1, 2, 3. \tag{4}$$

We represent the slip distribution $s_r'(\boldsymbol{\xi}, t)$ $(r = 1, 2)$ as a linear combination of spatial basis functions $\mathcal{B}_k^{(r)}(\boldsymbol{\xi})$ with temporally variable coefficients $c_k^{(r)}(t)$

$$s_r'(\boldsymbol{\xi}, t) = \sum_{k=1}^{M_r} c_k^{(r)}(t)\mathcal{B}_k^{(r)}(\boldsymbol{\xi}), \quad r = 1, 2 \tag{5}$$

where $\mathcal{B}_k^{(r)}(\boldsymbol{\xi})$ is the kth basis function for slip in the $\boldsymbol{\nu}_r(\boldsymbol{\xi})$ direction, $c_k^{(r)}(t)$ is the coefficient of $\mathcal{B}_k^{(r)}(\boldsymbol{\xi})$ and M_r is the number of basis function for $s_r'(\boldsymbol{\xi}, t)$.

Substituting (4) and (5) into (1) yields

$$u_i(\mathbf{x}_j, t_n) = \sum_{r=1}^{2} \sum_{k=1}^{M_r} c_k^{(r)}(t_n)\mathcal{K}_{ik}^{(r)}(\mathbf{x}_j) + \mathcal{L}_i(\mathbf{x}_j, t_n) + e_{ijn} \tag{6}$$

where

$$\mathcal{K}_{ik}^{(r)}(\mathbf{x}) = \mu \int_A \mathcal{B}_k^{(r)}(\boldsymbol{\xi}) \sum_{p=1}^{3} \sum_{q=1}^{3} [G_{ip,q}(\mathbf{x}, \boldsymbol{\xi}) + G_{iq,p}(\mathbf{x}, \boldsymbol{\xi})](\boldsymbol{\nu}_r(\boldsymbol{\xi}) \cdot \mathbf{e}_q)n_p(\boldsymbol{\xi})\, dA(\boldsymbol{\xi}). \tag{7}$$

From (6), the observation model is represented as follows:

$$\mathbf{y}_n = \mathbf{H}\mathbf{x}_n + \mathbf{w}_n, \quad \mathbf{w}_n \sim N(\mathbf{0}, \mathbf{R}_n) \tag{8}$$

where

$$\begin{aligned} \mathbf{y}_n = [&u_1(\mathbf{x}_1, t_n), u_1(\mathbf{x}_2, t_n), \cdots, u_1(\mathbf{x}_{N_1}, t_n), u_2(\mathbf{x}_1, t_n), u_2(\mathbf{x}_2, t_n), \cdots, u_2(\mathbf{x}_{N_1}, t_n), \\ &u_3(\mathbf{x}_1, t_n), u_3(\mathbf{x}_2, t_n), \cdots, u_3(\mathbf{x}_{N_2}, t_n)]^{\mathrm{T}} \end{aligned} \tag{9}$$

$$\begin{aligned} \mathbf{x}_n = \big[&c_1^{(1)}(t_n), c_1^{(1)}(t_{n-1}), c_2^{(1)}(t_n), c_2^{(1)}(t_{n-1}), \cdots, c_{M_1}^{(1)}(t_n), c_{M_1}^{(1)}(t_{n-1}), \\ &c_1^{(2)}(t_n), c_1^{(2)}(t_{n-1}), c_2^{(2)}(t_n), c_2^{(2)}(t_{n-1}), \cdots, c_{M_2}^{(2)}(t_n), c_{M_2}^{(2)}(t_{n-1}), \\ &\mathcal{L}_1(\mathbf{x}_1, t_n), \mathcal{L}_1(\mathbf{x}_2, t_n), \cdots, \mathcal{L}_1(\mathbf{x}_{N_1}, t_n), \\ &\mathcal{L}_2(\mathbf{x}_1, t_n), \mathcal{L}_2(\mathbf{x}_2, t_n), \cdots, \mathcal{L}_2(\mathbf{x}_{N_1}, t_n), \\ &\mathcal{L}_3(\mathbf{x}_1, t_n), \mathcal{L}_3(\mathbf{x}_2, t_n), \cdots, \mathcal{L}_3(\mathbf{x}_{N_2}, t_n)\big]^{\mathrm{T}} \end{aligned} \tag{10}$$

$$\mathbf{w}_n = \big[e_{11n}, e_{12n}, \cdots, e_{1N_1n}, e_{21n}, e_{22n}, \cdots, e_{2N_1n}, e_{31n}, e_{32n}, \cdots, e_{3N_2n}\big]^{\mathrm{T}} \tag{11}$$

$$\mathbf{H} = \left[\mathcal{K}^{(1)}\mathbf{J}^{(1)} \quad \mathcal{K}^{(2)}\mathbf{J}^{(2)} \quad \mathbf{I}_{N_d \times N_d}\right] \tag{12}$$

$$\mathbf{J}^{(r)} = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 & 0 \end{bmatrix}, \quad r = 1, 2 \tag{13}$$

$$\mathbf{R}_n = \sigma^2 \Sigma_n. \tag{14}$$

Here, N_1 and N_2 are the number of stations that observe horizontal and vertical displacement, respectively, and $N_d = 2N_1 + N_2$ is total number of time-series. $\mathbf{y}_n$ is an N_d-dimensional data vector, $\mathbf{w}_n$ is an N_d-dimensional observational error vector and $\mathbf{x}_n$ is a $2M_1 + 2M_2 + N_d = 2M + N_d$-dimensional state vector at time t_n, respectively, where $M = M_1 + M_2$ is the total number of basis functions. $\mathbf{H}$ is an $N_d \times (2M + N_d)$ matrix and $\mathbf{I}_{i\times j}$ denotes an $i \times j$ identity matrix. $\mathcal{K}^{(1)}$ is an $N_d \times M_1$ matrix and $\mathcal{K}^{(2)}$ is an $N_d \times M_2$ matrix. $\mathbf{J}^{(1)}$ is an $M_1 \times 2M_1$ matrix and $\mathbf{J}^{(2)}$ is an $M_2 \times 2M_2$ matrix. $\mathbf{R}_n$ is an $N_d \times N_d$ matrix.

2.2 System model

Next, we introduce a system model to describe the temporal evolution of state variables. For temporal variations of a fault slip, we employ a non-parametric description in which slip acceleration is nearly zero (Segall & Matthews 1997). Discretizing $\ddot{c}_k^{(r)}(t_n) = \delta_{nk}^{(r)}$ yields

$$c_k^{(r)}(t_n) = 2c_k^{(r)}(t_{n-1}) - c_k^{(r)}(t_{n-2}) + \delta_{nk}^{(r)}, \quad \delta_{nk}^{(r)} \sim N\left(0, \alpha_n^2\right). \tag{15}$$

System noise $\delta_{nk}^{(r)}$ is interpreted as a slip acceleration for the basis function $\mathcal{B}_k^{(r)}$ at time t_n, and is assumed to follow a Gaussian distribution with mean *0* and variance α_n^2, where α_n^2 is the temporally variable variance of slip acceleration.

For temporal variations of random benchmark motion we employ Brownian random walk with scale parameter τ (Langbein & Johnson 1997; Wyatt 1982, 1989) following Segall & Matthews (1997):

$$\mathcal{L}_i(\mathbf{x}_j, t_n) = \mathcal{L}_i(\mathbf{x}_j, t_{n-1}) + v_{ijn}, \quad v_{ijn} \sim N(0, \tau^2) \tag{16}$$

where τ is an unknown hyperparameter. In (15) and (16), dependency of α_n and τ on $t_n - t_{n-1}$ is included.

Combining (15) and (16) yields the following system model:

$$\mathbf{x}_n = \mathbf{F}_n \mathbf{x}_{n-1} + \mathbf{v}_n, \quad \mathbf{v}_n \sim N(\mathbf{0}, \mathbf{Q}_n) \tag{17}$$

where

$$\begin{aligned} \mathbf{v}_n = \big[&\delta_{n1}^{(1)}, 0, \delta_{n2}^{(1)}, 0, \cdots, \delta_{nM_1}^{(1)}, 0, \delta_{n1}^{(2)}, 0, \delta_{n2}^{(2)}, 0, \cdots, \delta_{nM_2}^{(2)}, 0, \\ &v_{11n}, v_{12n}, \cdots, v_{1N_1n}, v_{21n}, v_{22n}, \cdots, v_{2N_1n}, v_{31n}, v_{32n}, \cdots, v_{3N_2n}\big]^{\mathrm{T}} \end{aligned} \tag{18}$$

$$\mathbf{F}_n = \begin{bmatrix} \mathbf{F}_n^W & 0 & \cdots & 0 & 0 \\ 0 & \mathbf{F}_n^W & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \mathbf{F}_n^W & 0 \\ 0 & 0 & \cdots & 0 & \mathbf{I}_{N_d \times N_d} \end{bmatrix} \tag{19}$$

$$\mathbf{F}_n^W = \begin{bmatrix} 2 & -1 \\ 1 & 0 \end{bmatrix} \tag{20}$$

$$\mathbf{Q}_n = \begin{bmatrix} \alpha_n^2 \mathbf{Q}_n^W & 0 & \cdots & 0 & 0 \\ 0 & \alpha_n^2 \mathbf{Q}_n^W & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \alpha_n^2 \mathbf{Q}_n^W & 0 \\ 0 & 0 & \cdots & 0 & \tau^2 \mathbf{I}_{N_d \times N_d} \end{bmatrix} \tag{21}$$

$$\mathbf{Q}_n^W = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}. \tag{22}$$

$\mathbf{F}_n$ and $\mathbf{Q}_n$ are $(2M + N_d) \times (2M + N_d)$ matrices and system noise $\mathbf{v}_n$ is a $2M + N_d$-dimensional vector.

3 MONTE CARLO MIXTURE KALMAN FILTER

3.1 Conditional dynamic linear model

The previous studies (e.g. Segall & Matthews 1997) treated α_n/σ as a temporally invariable parameter. To allow for temporal variations of α_n, we treat α_n/σ as a stochastic variable. Note that σ is assumed to be temporally invariable, whereas α_n is temporally variable. We first prepare a finite number of candidate values for α_n/σ, $\{a_{(1)}, \cdots, a_{(m)}\}$, where m denotes the number of candidates. Then, we introduce a discrete stochastic variable $\mathcal{I}_n$, an indicator variable, which takes an integer value among $\{1, \cdots, m\}$. A value of $\mathcal{I}_n$ specifies which of the candidates is selected at $t = t_n$. For example, $\mathcal{I}_n = \lambda_n$ means that the value of α_n/σ at $t = t_n$ is equal to the candidate $a_{(\lambda_n)}$. Therefore, it is sufficient to estimate $\mathcal{I}_n$ to obtain temporal variation of α_n/σ.

Given a realization of $\mathcal{I}_n$, λ_n, we can represent the system model (17) and the observation model (8) as follows:

$$\mathbf{x}_n = \mathbf{F}_n \mathbf{x}_{n-1} + \mathbf{v}_n, \quad \mathbf{v}_n \sim N\left(\mathbf{0}, \sigma^2 \tilde{\mathbf{Q}}_n(\lambda_n)\right) \tag{23}$$

$$\mathbf{y}_n = \mathbf{H}\mathbf{x}_n + \mathbf{w}_n, \quad \mathbf{w}_n \sim N\left(\mathbf{0}, \sigma^2 \tilde{\mathbf{R}}_n\right) \tag{24}$$

where

$$\tilde{\mathbf{Q}}_n(\lambda_n) = \begin{bmatrix} a^2_{(\lambda_n)}\mathbf{Q}^W_n & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & a^2_{(\lambda_n)}\mathbf{Q}^W_n & 0 \\ 0 & 0 & \cdots & 0 & \tau^2/\sigma^2 \mathbf{I}_{N_d \times N_d} \end{bmatrix} \tag{25}$$

$$\tilde{\mathbf{R}}_n = \Sigma_n \tag{26}$$

with $\mathbf{Q}^W_n$ defined in (22). It should be noted that (23) and (24) are not the linear Gaussian state space model (e.g. Kitagawa & Gersch 1996) that is employed by Segall & Matthews (1997). Eqs (23) and (24) are a special form of the conditional dynamic linear model (CDLM) (Chen & Liu 2000). The difference between the CDLM and the linear Gaussian state space model is as follows. In the CDLM, some matrices in state space representation are conditional on an unknown stochastic variable (Chen & Liu 2000). On the other hand, in the linear Gaussian state space model, these matrices do not depend on any stochastic variables. In (23) and (24), $\tilde{\mathbf{Q}}_n$ is conditional on the stochastic variable $\mathcal{I}_n$. Therefore, (23) and (24) are the CDLM. The most important characteristic of the CDLM is that it reduces to the linear Gaussian state space model if a realization of the stochastic variable is given (Chen & Liu 2000). In the case of (23) and (24), they reduce to the linear Gaussian state space model if λ_n, a realization of the stochastic variable $\mathcal{I}_n$, is given. Hence, the conditional density function of the state $\mathbf{x}_n$ given the observation $\mathbf{y}_{1:j} = (\mathbf{y}_1, \ldots, \mathbf{y}_j)$ and realizations of the indicator variable $\lambda_{1:k} = (\lambda_1, \ldots, \lambda_k)$ $(k \geq n)$ follows a Gaussian distribution:

$$p(\mathbf{x}_n|\lambda_{1:k}, \mathbf{y}_{1:j}) \sim N(\mathbf{x}_n(\lambda_n), \mathbf{V}_n(\lambda_n)) \tag{27}$$

where the state vector $\mathbf{x}_n(\lambda_n)$ and its covariance matrix $\mathbf{V}_n(\lambda_n)$ are functions of λ_n. Given λ_n, $\mathbf{x}_n(\lambda_n)$ and $\mathbf{V}_n(\lambda_n)$ are recursively obtained using the Kalman filter and smoother (e.g. Kitagawa & Gersch 1996). The conditional density $p(\mathbf{x}_n|\,\mathbf{y}_{1:j})$ is expressed by

$$p(\mathbf{x}_n|\mathbf{y}_{1:j}) = \int p(\mathbf{x}_n|\lambda_{1:k}, \mathbf{y}_{1:j}) p(\lambda_{1:k}|\mathbf{y}_{1:j})\, d\lambda_{1:k} \tag{28}$$

where $p(\mathbf{x}_n|\lambda_{1:k}, \mathbf{y}_{1:j})$ is the Gaussian distribution given by (27). Because $\mathcal{I}_n$ is the discrete stochastic variable, (28) reduces to weighted sum of the Gaussian distributions.

To estimate time evolution of $\mathcal{I}_n$, it is necessary to specify a stochastic model that describes a time-dependent structure for $\mathcal{I}_n$. In this study, $\mathcal{I}_n$ is assumed to follow a stationary Markov process, i.e.

$$p(\mathcal{I}_n|\mathcal{I}_{1:n-1}) = p(\mathcal{I}_n|\mathcal{I}_{n-1}) \tag{29}$$

where $\mathcal{I}_{i:j} = (\mathcal{I}_i, \mathcal{I}_{i+1}, \cdots, \mathcal{I}_j)$ is a set of indicator variable from time t_i to time t_j. Evolution of $\mathcal{I}_n$ is realized by the Markov switching model with transition probability given by

$$\pi_{ij} = \Pr(\mathcal{I}_n = j|\mathcal{I}_{n-1} = i) \tag{30}$$

where Pr denotes realization probability.

In the following subsections, we present an algorithm that determines the time evolution of $\mathcal{I}_n$ and $\mathbf{x}_n$. We call this algorithm the Monte Carlo mixture Kalman filter (MCMKF).

3.2 Monte Carlo approximation

In contrast to the previous studies in which α_n/σ is regarded as a parameter (e.g. Segall & Matthews 1997), in this study α_n/σ is regarded as a function of the stochastic variable $\mathcal{I}_n$. Therefore, it is necessary to estimate the probability distribution of $\mathcal{I}_n$ rather than its value. The MCMKF algorithm consists of two steps. First, the temporal variation of the probability distribution of the indicator variable $\mathcal{I}_n$ is determined. Second, the temporal variation of the probability distribution of the state vector $\mathbf{x}_n$ is estimated following the history of $\mathcal{I}_n$.

Let $\mathbf{y}_{i:j}$ be a set of data vectors from time t_i to time t_j, i.e. $\mathbf{y}_{i:j} = (\mathbf{y}_i, \mathbf{y}_{i+1}, \cdots, \mathbf{y}_j)$. In the MCMKF, two conditional joint distributions of $\mathcal{I}_{1:n}$: (i) predictive distribution $p(\mathcal{I}_{1:n}|\mathbf{y}_{1:n-1})$ and (ii) filter distribution $p(\mathcal{I}_{1:n}|\mathbf{y}_{1:n})$, are approximated by many 'particles' that can be considered to be independent realizations from each joint distribution. Let $\mathcal{I}^{(j)}_{1:i|k} = (\mathcal{I}^{(j)}_{1|k}, \mathcal{I}^{(j)}_{2|k}, \cdots \mathcal{I}^{(j)}_{i|k})$ be the jth realization (particle) of the conditional joint distribution $p(\mathcal{I}_{1:i}|\mathbf{y}_{1:k})$. Each distribution is approximated by N_p ($N_p \gg 1$) realizations as follows:

$$\left\{\mathcal{I}^{(1)}_{1:n|n-1}, \mathcal{I}^{(2)}_{1:n|n-1}, \cdots, \mathcal{I}^{(N_p)}_{1:n|n-1}\right\} \sim p(\mathcal{I}_{1:n}|\mathbf{y}_{1:n-1}) \tag{31}$$

$$\left\{\mathcal{I}^{(1)}_{1:n|n}, \mathcal{I}^{(2)}_{1:n|n}, \cdots, \mathcal{I}^{(N_p)}_{1:n|n}\right\} \sim p(\mathcal{I}_{1:n}|\mathbf{y}_{1:n}). \tag{32}$$

Eqs (31) and (32) mean that the predictive and the filter distributions are approximated by the following probability density functions:

$$p(\mathcal{I}_{1:n}|\mathbf{y}_{1:n-1}) = \frac{1}{N_p}\sum_{j=1}^{N_p}\delta\left(\mathcal{I}_{1:n} - \mathcal{I}^{(j)}_{1:n|n-1}\right) \tag{33}$$

$$p(\mathcal{I}_{1:n}|\mathbf{y}_{1:n}) = \frac{1}{N_p}\sum_{j=1}^{N_p}\delta\left(\mathcal{I}_{1:n} - \mathcal{I}^{(j)}_{1:n|n}\right) \tag{34}$$

where $\delta(x)$ is the delta function for the discrete variable x defined by $\delta(x) = 1$ if $x = 0$ and $\delta(x) = 0$ otherwise. Eqs (33) and (34) are equivalent to the probabilities

$$\Pr\left(\mathcal{I}_{1:n} = \mathcal{I}^{(j)}_{1:n|n-1}|\mathbf{y}_{1:n-1}\right) = \frac{1}{N_p} \tag{35}$$

$$\Pr\left(\mathcal{I}_{1:n} = \mathcal{I}^{(j)}_{1:n|n}|\mathbf{y}_{1:n}\right) = \frac{1}{N_p} \tag{36}$$

which mean all the particles have the same probability, or they are equally weighted. In this study, we refer to

$$\left\{\mathcal{I}^{(1)}_{1:n|n-1}, \mathcal{I}^{(2)}_{1:n|n-1}, \cdots, \mathcal{I}^{(N_p)}_{1:n|n-1}\right\} \quad \text{and} \quad \left\{\mathcal{I}^{(1)}_{1:n|n}, \mathcal{I}^{(2)}_{1:n|n}, \cdots, \mathcal{I}^{(N_p)}_{1:n|n}\right\}$$

as the 'approximated predictive distribution' and 'approximated filter distribution', respectively. Because $\mathcal{I}^{(j)}_n = \mathcal{I}^{(j)}_{n|k}$ ($k = n-1$ or n) is a realization of the distribution of $\mathcal{I}_n$, the CDLM defined by (23) and (24) is satisfied for each $\mathcal{I}^{(j)}_n (j = 1, \cdots, N_p)$:

$$\mathbf{x}^{(j)}_n = \mathbf{F}_n\mathbf{x}^{(j)}_{n-1} + \mathbf{v}^{(j)}_n, \quad \mathbf{v}^{(j)}_n \sim N\left(\mathbf{0}, \sigma^{(j)2}\tilde{\mathbf{Q}}_n\left(\mathcal{I}^{(j)}_n\right)\right) \tag{37}$$

$$\mathbf{y}_n = \mathbf{H}\mathbf{x}^{(j)}_n + \mathbf{w}^{(j)}_n, \quad \mathbf{w}^{(j)}_n \sim N\left(\mathbf{0}, \sigma^{(j)2}\tilde{\mathbf{R}}_n\right) \tag{38}$$

where $\tilde{\mathbf{Q}}_n(\mathcal{I}^{(j)}_n)$ is given by (25) and $\sigma^{(j)}$ is the scaling factor for the data covariance for the jth particle. Starting from $\{\mathcal{I}^{(1)}_{0|0}, \ldots, \mathcal{I}^{(N_p)}_{0|0}\}$ which are realizations of an initial distribution for $\mathcal{I}_n$, $p(\mathcal{I}_0)$, we show that a set of particles approximating the predictive distribution and the filter distribution are obtained recursively in two steps:

$$\text{Prediction:} \quad \left\{\mathcal{I}^{(1)}_{1:n-1|n-1}, \cdots, \mathcal{I}^{(N_p)}_{1:n-1|n-1}\right\} \longrightarrow \left\{\mathcal{I}^{(1)}_{1:n|n-1}, \cdots, \mathcal{I}^{(N_p)}_{1:n|n-1}\right\} \tag{39}$$

$$\text{Filtering:} \quad \left\{\mathcal{I}^{(1)}_{1:n|n-1}, \cdots, \mathcal{I}^{(N_p)}_{1:n|n-1}\right\} \longrightarrow \left\{\mathcal{I}^{(1)}_{1:n|n}, \cdots, \mathcal{I}^{(N_p)}_{1:n|n}\right\}. \tag{40}$$

3.3 Prediction

In this subsection, we show that an approximated predictive distribution at time t_n, $\{\mathcal{I}^{(1)}_{1:n|n-1}, \cdots, \mathcal{I}^{(N_p)}_{1:n|n-1}\}$ is obtained from an approximated filter distribution at time t_{n-1}, $\{\mathcal{I}^{(1)}_{1:n-1|n-1}, \cdots, \mathcal{I}^{(N_p)}_{1:n-1|n-1}\}$. We assume that $\{\mathcal{I}^{(1)}_{1:n-1|n-1}, \cdots, \mathcal{I}^{(N_p)}_{1:n-1|n-1}\}$ and $\mathbf{y}_{1:n-1}$ are given. Then the predictive distribution $p(\mathcal{I}_{1:n}|\mathbf{y}_{1:n-1})$ is expressed by

$$\begin{aligned}
p(\mathcal{I}_{1:n}|\mathbf{y}_{1:n-1}) &= p(\mathcal{I}_n, \mathcal{I}_{1:n-1}|\mathbf{y}_{1:n-1}) \\
&= p(\mathcal{I}_n|\mathcal{I}_{1:n-1}, \mathbf{y}_{1:n-1})p(\mathcal{I}_{1:n-1}|\mathbf{y}_{1:n-1}) \\
&= p(\mathcal{I}_n|\mathcal{I}_{1:n-1})p(\mathcal{I}_{1:n-1}|\mathbf{y}_{1:n-1}) \\
&= p(\mathcal{I}_n|\mathcal{I}_{n-1})\frac{1}{N_p}\sum_{j=1}^{N_p}\delta\left(\mathcal{I}_{1:n-1} - \mathcal{I}^{(j)}_{1:n-1|n-1}\right) \\
&= \frac{1}{N_p}\sum_{j=1}^{N_p} p\left(\mathcal{I}_n|\mathcal{I}_{n-1} = \mathcal{I}^{(j)}_{n-1|n-1}\right).
\end{aligned} \tag{41}$$

Here, we used the Markovian assumption in the third and the fourth equalities, (34) in the fourth equality, and the property of the delta function in the last equality. Eq. (41) indicates that $\mathcal{I}^{(j)}_{1:n|n-1}$, the jth particle of the approximated predictive distribution, is obtained by generating a realization $\mathcal{I}^{(j)}_{n|n-1}$ from the distribution $p(\mathcal{I}_n|\mathcal{I}_{n-1} = \mathcal{I}^{(j)}_{n-1|n-1})$, and setting $\mathcal{I}^{(j)}_{1:n|n-1} = (\mathcal{I}^{(j)}_{1:n-1|n-1}, \mathcal{I}^{(j)}_{n|n-1})$. Note that $p(\mathcal{I}_n|\mathcal{I}_{n-1} = \mathcal{I}^{(j)}_{n-1|n-1})$ is given by the Markovian transition probability defined by (30). More specifically, N_p particles $\mathcal{I}^{(j)}_{1:n|n-1}$ ($j = 1, \ldots, N_p$) can be obtained by the following algorithm:

For $j = 1, \ldots, N_p$, repeat the following steps (i)–(iii).

(i) Generate a uniform random number $u_n^{(j)} \in U(0, 1)$ where $U(0, 1)$ denotes uniform distribution over the interval [0,1).
(ii) Find i that satisfies $\sum_{k=1}^{i-1} \Pr(\mathcal{I}_n = k|\mathcal{I}_{n-1} = \mathcal{I}_{1:n-1|n-1}^{(j)}) \le u_n^{(j)} < \sum_{k=1}^{i} \Pr(\mathcal{I}_n = k|\mathcal{I}_{n-1} = \mathcal{I}_{1:n-1|n-1}^{(j)})$.
(iii) Obtain $\mathcal{I}_{1:n|n-1}^{(j)}$ by $\mathcal{I}_{1:n|n-1}^{(j)} = i$.

3.4 Filtering

In this subsection, we show that an approximated filter distribution at time t_n, $\{\mathcal{I}_{1:n|n}^{(1)}, \cdots, \mathcal{I}_{1:n|n}^{(N_p)}\}$ is obtained from an approximated predictive distribution at time t_n, $\{\mathcal{I}_{1:n|n-1}^{(1)}, \cdots, \mathcal{I}_{1:n|n-1}^{(N_p)}\}$. Though all the particles approximating the predictive distribution have the same probability, or they are equally weighted (see eq. 35), they do not have the same probability any longer after the given data $\mathbf{y}_n$. Given the observation $\mathbf{y}_n$, the posterior probability of the particle $\mathcal{I}_{1:n|n-1}^{(j)}$ is expressed by

$$\begin{aligned}
&\Pr\left(\mathcal{I}_{1:n} = \mathcal{I}_{1:n|n-1}^{(j)}|\mathbf{y}_{1:n}\right) \\
&\quad = \Pr\left(\mathcal{I}_{1:n} = \mathcal{I}_{1:n|n-1}^{(j)}|\mathbf{y}_{1:n-1}, \mathbf{y}_n\right) \\
&\quad = \frac{p\left(\mathbf{y}_n|\mathcal{I}_{1:n} = \mathcal{I}_{1:n|n-1}^{(j)}, \mathbf{y}_{1:n-1}\right) \Pr\left(\mathcal{I}_{1:n} = \mathcal{I}_{1:n|n-1}^{(j)}|\mathbf{y}_{1:n-1}\right)}{\sum_{i=1}^{N_p} p\left(\mathbf{y}_n|\mathcal{I}_{1:n} = \mathcal{I}_{1:n|n-1}^{(i)}, \mathbf{y}_{1:n-1}\right) \Pr\left(\mathcal{I}_{1:n} = \mathcal{I}_{1:n|n-1}^{(i)}|\mathbf{y}_{1:n-1}\right)} \\
&\quad = \frac{w_n^{(j)}(1/N_p)}{\sum_{i=1}^{N_p} w_n^{(i)}(1/N_p)} \\
&\quad = \frac{w_n^{(j)}}{\sum_{i=1}^{N_p} w_n^{(i)}}
\end{aligned} \tag{42}$$

where $w_n^{(j)}$ can be regarded as the importance weight of the jth particle and is given by

$$w_n^{(j)} = p\left(\mathbf{y}_n|\mathcal{I}_{1:n} = \mathcal{I}_{1:n|n-1}^{(j)}, \mathbf{y}_{1:n-1}\right). \tag{43}$$

We used Bayes' theorem in the second equality of (42) and (35) in the third equality of (42). From (42), the probability density function associated with $\Pr(\mathcal{I}_{1:n} = \mathcal{I}_{1:n|n-1}^{(j)}|\mathbf{y}_{1:n})$ is expressed by

$$\frac{1}{\sum_{i=1}^{N_p} w_n^{(i)}} \sum_{j=1}^{N_p} w_n^{(j)} \delta\left(\mathcal{I}_{1:n} - \mathcal{I}_{1:n|n-1}^{(j)}\right). \tag{44}$$

This means that the particle $\mathcal{I}_{1:n|n-1}^{(j)}$ with probability or weight $w_n^{(j)} / \sum_{i=1}^{N_p} w_n^{(i)}$ can be considered to be a realization of the filter distribution $p(\mathcal{I}_{1:n}|\mathbf{y}_{1:n})$. Because we have assumed in (36) that all the particles approximating the filter distribution are equally weighted, it is necessary to represent (44) by the probability density function of the form (34). This is achieved by generating N_p particles $\{\mathcal{I}_{1:n|n}^{(1)}, \cdots, \mathcal{I}_{1:n|n}^{(N_p)}\}$ by the resampling (sampling with replacement, i.e. a particular particle can be sampled more than once, others not at all) of $\{\mathcal{I}_{1:n|n-1}^{(1)}, \cdots, \mathcal{I}_{1:n|n-1}^{(N_p)}\}$ with sampling probabilities

$$\Pr\left(\mathcal{I}_{1:n} = \mathcal{I}_{1:n|n-1}^{(j)}|\mathbf{y}_{1:n}\right) = \frac{w_n^{(j)}}{\sum_{i=1}^{N_p} w_n^{(i)}}, \quad \text{for} \quad i = 1, \cdots, N_p. \tag{45}$$

More specifically, the resampling algorithm proceeds as follows:

For $j = 1, \ldots, N_p$, repeat the following steps (i)–(iii).

(i) Generate a uniform random number $u_n^{(j)} \in U(0, 1)$ where $U(0, 1)$ denotes uniform distribution over the interval [0,1).
(ii) Find i that satisfies $\frac{1}{C}\sum_{l=1}^{i-1} w_n^{(l)} \le u_n^{(j)} < \frac{1}{C}\sum_{l=1}^{i} w_n^{(l)}$ where $C = \sum_{l=1}^{N_p} w_n^{(l)}$.
(iii) Multiply $\mathcal{I}_{1:n|n-1}^{(i)}$ to obtain $\mathcal{I}_{1:n|n}^{(j)}$ by $\mathcal{I}_{1:n|n}^{(j)} = \mathcal{I}_{1:n|n-1}^{(i)}$.

Intuitively, the resampling procedure can be considered as a 'selection procedure' multiplying/discarding particles $\mathcal{I}_{1:n|n-1}^{(j)}$ with high/low importance weights $w_n^{(j)}$ to obtain N_p particles $\mathcal{I}_{1:n|n}^{(j)}$. The importance weight $w_n^{(j)}$ for each particle can be obtained using the Kalman filter. The derivation of $w_n^{(j)}$ is given in Appendix A.

A summary of one cycle of the prediction and the filtering steps is graphically illustrated in Fig. 2. Using the prediction and the filtering algorithms recursively, we finally obtain N_p particles $\{\mathcal{I}_{1:N_e|N_e}^{(1)}, \cdots, \mathcal{I}_{1:N_e|N_e}^{(N_p)}\}$ that approximate $p(\mathcal{I}_{1:N_e}|\mathbf{y}_{1:N_e})$, the posterior distribution of $\mathcal{I}_{1:N_e}$ conditioned on all of the available data. Here, N_e is the number of observation epochs. $p(\mathcal{I}_{1:N_e}|\mathbf{y}_{1:N_e})$ is called a smoother distribution of $\mathcal{I}_{1:N_e}$. A sequence of each particle, $\mathcal{I}_{1:N_e|N_e}^{(j)} = [\mathcal{I}_{1|N_e}^{(j)}, \mathcal{I}_{2|N_e}^{(j)}, \cdots, \mathcal{I}_{N_e|N_e}^{(j)}]$, is called the trajectory.

This filtering algorithm is conceptually similar to the storing state vector algorithm in the Monte Carlo filter proposed by Kitagawa (1996). He applied the Monte Carlo approximation directly to the distribution of the state, whereas we apply the approximation to the distribution of the indicator variable. He showed that in the Monte Carlo filter the repetition of resampling gradually decreases the number of different realizations of the state vector as time passes, because the number of realizations is finite. Therefore, the shape of the distribution of the state

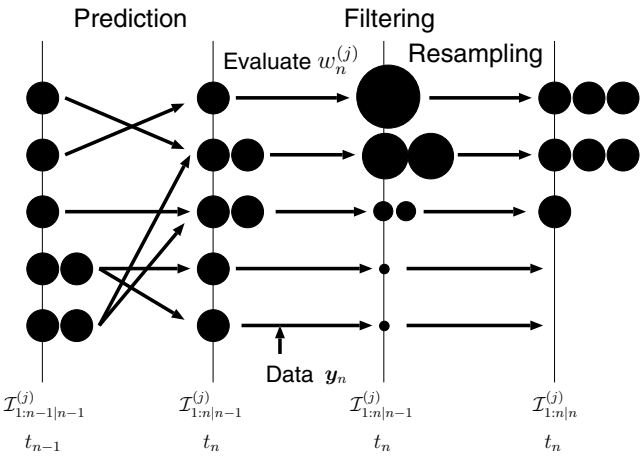

Figure 2. Graphical representation of the prediction and the filtering algorithms for one time step. The vertical axis denotes value of the indicator variable. The solid circles denote the particles and the area of the particles is proportional to their weights. Starting with the equally weighted particles at time t_{n-1} (the first column), a new set of particles is generated at time t_n (the second column). Then we evaluate the importance weight of each particles (the third column). Finally, we multiply/discard particles with high/low importance weights to obtain equally weighted particles at time t_n (the fourth column).

deteriorates as time passes. Kitagawa (1996) showed that this difficulty can be eliminated by employing fixed L-lag smoother rather than a fixed-interval smoother. Although we apply the Monte Carlo approximation to the indicator variable instead of the state, this situation also applies to the MCMKF. Thus, following Kitagawa (1996), we modify the MCMKF filtering algorithm as follows:

For fixed L, generate N_p particles $\{\mathcal{I}_{n-L:n|n}^{(1)}, \mathcal{I}_{n-L:n|n}^{(2)}, \cdots, \mathcal{I}_{n-L:n|n}^{(N_p)}\}$ by the resampling of $\{\mathcal{I}_{n-L:n|n-1}^{(1)}, \mathcal{I}_{n-L:n|n-1}^{(2)}, \cdots, \mathcal{I}_{n-L:n|n-1}^{(N_p)}\}$ with the sampling probability defined by (45).

Kitagawa (1996) recommended not to make L too large (say, 10 or 20 at the largest 50). We adopt $L = 20$ in our application study shown in Section 4.

3.5 Estimation of the state

We present here an algorithm to estimate the state using all of the N_p trajectories $\{\mathcal{I}_{1:N_e|N_e}^{(1)}, \cdots, \mathcal{I}_{1:N_e|N_e}^{(N_p)}\}$. Given $\{\mathcal{I}_{1:N_e|N_e}^{(1)}, \cdots, \mathcal{I}_{1:N_e|N_e}^{(N_p)}\}$, $\tilde{\mathbf{Q}}_n(\mathcal{I}_n^{(j)}) = \tilde{\mathbf{Q}}_n(\mathcal{I}_{n|N_e}^{(j)})(j = 1, \cdots, N_p)$ in (37) reduce to sets of known matrices, which have different time evolutions corresponding to the trajectories. In this case, the CDLM defined by (37) and (38) reduces to N_p linear Gaussian state space models ($j = 1, \ldots, N_p$):

$$\mathbf{x}_n^{(j)} = \mathbf{F}_n \mathbf{x}_{n-1}^{(j)} + \mathbf{v}_n^{(j)}, \quad \mathbf{v}_n^{(j)} \sim N\big(\mathbf{0}, \sigma^{(j)2}\tilde{\mathbf{Q}}_n\big(\mathcal{I}_{n|N_e}^{(j)}\big)\big) \tag{46}$$

$$\mathbf{y}_n = \mathbf{H}\mathbf{x}_n^{(j)} + \mathbf{w}_n^{(j)}, \quad \mathbf{w}_n^{(j)} \sim N\big(\mathbf{0}, \sigma^{(j)2}\tilde{\mathbf{R}}_n\big). \tag{47}$$

Let

$$\mathbf{x}_{i|k}^{(j)} = \mathrm{E}\big(\mathbf{x}_i | \mathbf{y}_{1:k}, \mathcal{I}_{1:N_e} = \mathcal{I}_{1:N_e|N_e}^{(j)}\big) \tag{48}$$

and

$$\sigma^{(j)2} V_{i|k}^{(j)} = \mathrm{Cov}\big(\mathbf{x}_i | \mathbf{y}_{1:k}, \mathcal{I}_{1:N_e} = \mathcal{I}_{1:N_e|N_e}^{(j)}\big) \tag{49}$$

be the conditional mean and the covariance matrix of the state at time t_i given data $\mathbf{y}_{1:k}$ for the jth trajectory. Then

$$\big\{\mathbf{x}_{n+1|n}^{(j)}, V_{n+1|n}^{(j)}\big\}_{j=1}^{N_p}, \quad \big\{\mathbf{x}_{n|n}^{(j)}, V_{n|n}^{(j)}\big\}_{j=1}^{N_p} \quad \text{and} \quad \big\{\mathbf{x}_{n|N_e}^{(j)}, V_{n|N_e}^{(j)}\big\}_{j=1}^{N_p}$$

are recursively obtained by applying the Kalman filter and smoother algorithm (e.g. Kitagawa & Gersch 1996) to the jth state space model. The unknown hyperparameter $\sigma^{(j)}$ is estimated by maximizing likelihood for the jth trajectory (see Appendix B for derivation of $\sigma^{(j)}$). Given $\{\mathbf{x}_{n|N_e}^{(j)}, \sigma^{(j)2} V_{n|N_e}^{(j)}\}_{j=1}^{N_p}$, the distribution of the final estimate for $\mathbf{x}_n$, $p(\mathbf{x}_n|\mathbf{y}_{1:N_e})$, is written as

$$\begin{aligned} p(\mathbf{x}_n|\mathbf{y}_{1:N_e}) &= \sum_{j=1}^{N_p} p\big(\mathbf{x}_n, \mathcal{I}_{1:N_e} = \mathcal{I}_{1:N_e|N_e}^{(j)} | \mathbf{y}_{1:N_e}\big) \\ &= \sum_{j=1}^{N_p} p\big(\mathbf{x}_n | \mathcal{I}_{1:N_e} = \mathcal{I}_{1:N_e|N_e}^{(j)}, \mathbf{y}_{1:N_e}\big) \Pr\big(\mathcal{I}_{1:N_e} = \mathcal{I}_{1:N_e|N_e}^{(j)} | \mathbf{y}_{1:N_e}\big) \\ &= \frac{1}{N_p} \sum_{j=1}^{N_p} N\big(\mathbf{x}_{n|N_e}^{(j)}, \sigma^{(j)2} V_{n|N_e}^{(j)}\big). \end{aligned} \tag{50}$$

Initialization

- For $j = 1, \ldots, N_p$, generate a realization $\mathcal{I}_{0|0}^{(j)}$ from $p(\mathcal{I}_0)$. Here, $p(\mathcal{I}_0)$ is an initial distribution for $\mathcal{I}_n$.
- For $j = 1, \ldots, N_p$, set initial state $(\tilde{\boldsymbol{x}}_{1|0}^{(j)}, \tilde{V}_{1|0}^{(j)})$.

Repeat the prediction and the filtering steps for $n = 1, \ldots, N_e$.

1. *Prediction* (Section 3.3)
   - For $j = 1, \ldots, N_p$, generate a realization $\mathcal{I}_{n|n-1}^{(j)}$ from $p(\mathcal{I}_n | \mathcal{I}_{n-1} = \mathcal{I}_{n-1|n-1}^{(j)})$.
   - For $j = 1, \ldots, N_p$, set $\mathcal{I}_{1:n|n-1}^{(j)} = (\mathcal{I}_{1:n-1|n-1}^{(j)}, \mathcal{I}_{n|n-1}^{(j)})$.
2. *Filtering* (Section 3.4, Appendix A)
   - For $j = 1, \ldots, N_p$, update $(\bar{\boldsymbol{x}}_{n-1|n-1}^{(j)}, \bar{V}_{n-1|n-1}^{(j)})$ to obtain $(\tilde{\boldsymbol{x}}_{n|n-1}^{(j)}, \tilde{V}_{n|n-1}^{(j)})$ and $(\tilde{\boldsymbol{x}}_{n|n}^{(j)}, \tilde{V}_{n|n}^{(j)})$ according to (A9), (A10), (A16), and (A17).
   - For $j = 1, \ldots, N_p$, evaluate the importance weight $w_n^{(j)}$ according to (A14).
   - Obtain $\left\{\mathcal{I}_{1:n|n}^{(j)}, \bar{\boldsymbol{x}}_{n|n}^{(j)}, \bar{V}_{n|n}^{(j)}\right\}_{j=1}^{N_p}$ by the resampling of $\left\{\mathcal{I}_{1:n|n-1}^{(j)}, \tilde{\boldsymbol{x}}_{n|n}^{(j)}, \tilde{V}_{n|n}^{(j)}\right\}_{j=1}^{N_p}$ with sampling probability $w_n^{(j)} / \sum_{i=1}^{N_p} w_n^{(i)}$.

State estimation (Section 3.5)

- Obtain $\left\{\boldsymbol{x}_{n|N_e}^{(j)}, \sigma^{(j)2} V_{n|N_e}^{(j)}\right\}_{j=1}^{N_p}$ by using Kalman filter and smoother.
- Obtain the density $p(\boldsymbol{x}_n | \boldsymbol{y}_{1:N_e})$ according to (50).

Figure 3. Summary of the MCMKF algorithm.

In the third equality, (36), (48) and (49) are used. Therefore, $p(\mathbf{x}_n | \mathbf{y}_{1:N_e})$ is a non-Gaussian distribution with mean

$$\mathbf{x}_{n|N_e} = \frac{1}{N_p} \sum_{j=1}^{N_p} \mathbf{x}_{n|N_e}^{(j)}. \tag{51}$$

The procedure for the state estimation using all trajectories described above is computationally massive, both in calculation time and in memory. A more efficient algorithm is obtained by reducing the number of trajectories to which the Kalman filter is applied. This is done by sampling N'_p ($N'_p < N_p$) trajectories randomly from N_p trajectories $\{\mathcal{I}_{1:N_e|N_e}^{(1)}, \cdots, \mathcal{I}_{1:N_e|N_e}^{(N_p)}\}$. Once N'_p trajectories are selected, the procedure for state estimation is identical to the case using all N_p trajectories. The distribution of the final estimate for $\mathbf{x}_n$, $p(\mathbf{x}_n | \mathbf{y}_{1:N_e})$, and its mean vector, $\mathbf{x}_{n|N_e}$, are obtained by replacing N_p in (50) and (51) with N'_p, respectively.

The MCMKF algorithm described in Section 3.3, 3.4 and 3.5 is summarized in Fig. 3.

4 VERIFICATION OF OUR ALGORITHM AND DISCUSSION

In this section we apply the proposed method to three simple examples to demonstrate the validity of the method. We consider an infinitely long strike-slip fault that slips from depth d_2 to d_1 embedded in a homogeneous, isotropic, elastic half-space as shown in Fig. 4. We set axis x_1 in the direction perpendicular to the strike direction, axis x_2 upwards and axis x_3 in the strike direction. In this problem, only non-zero displacement is $u_3 = u_3(x_1, x_2, t)$. Slip is assumed to be spatially uniform for simplicity. Thus, slip is a function only of time. For a spatially uniform slip, it is well known that the cumulative surface displacement is written as

$$u_3(x_1, x_2 = 0, t) = -\frac{s(t)}{\pi} \left[\tan^{-1}\left(\frac{x_1}{d_1}\right) - \tan^{-1}\left(\frac{x_1}{d_2}\right) \right]. \tag{52}$$

We set $d_1 = 15$ km, $d_2 = 5$ km and place 10 equally spaced stations in the range -100 km $\leq x_1 \leq 100$ km. In all cases, simulated data of length $N_e = 100$, sampled regularly, are generated by calculating surface displacements due to true slip history $s(t)$ and adding observational error (white noise) and random benchmark motion (random walk). The standard deviations (SD) of observational error and random benchmark motion are set at $\sigma = 4$ mm and $\tau = 0.04$ mm, respectively. The data covariance matrix, Σ_n, in (14) is set to be an identity matrix for simplicity.

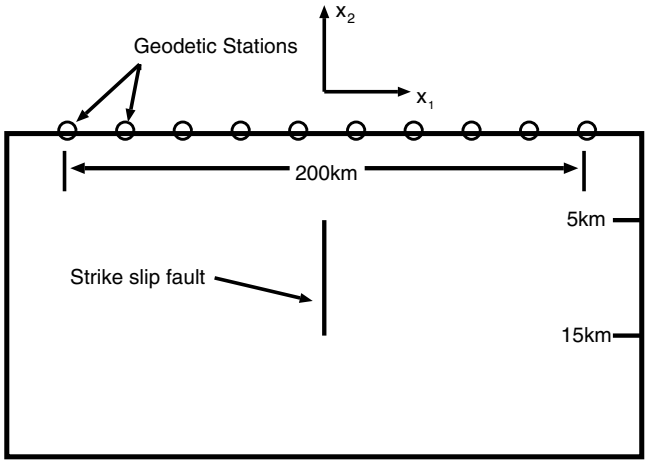

Figure 4. Geometry of the infinitely long strike-slip fault embedded in a homogeneous, isotropic, elastic half-space. The fault slips from depth $d_2 = 5$ km to $d_1 = 15$ km. Ten equally spaced geodetic stations are placed in the range -100 km $\leq x_1 \leq 100$ km.

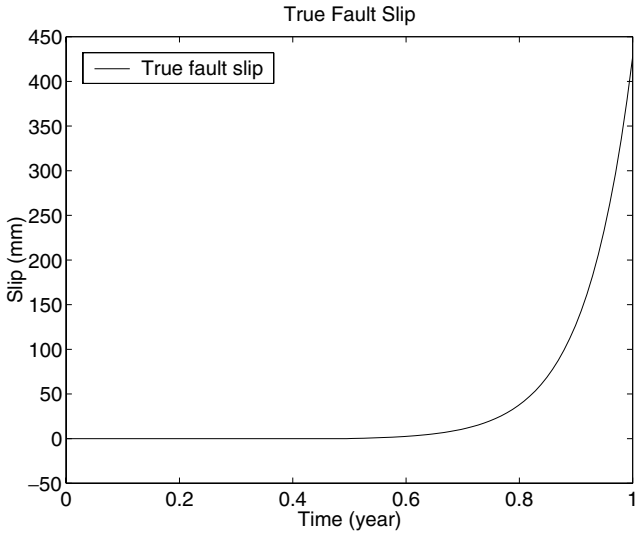

Figure 5. True slip history given on the fault surface.

4.1 Accelerating slip

We invert 10 displacement time-series generated by accelerating slip on the fault surface. Here, we use all N_p trajectories for the estimation of state. The true input signal is shown in Fig. 5. Simulated data, which are the sum of contributions from fault slip, random benchmark motion and observational error, are shown in Fig. 6. We consider $m = 13$ candidate values for α_n/σ:

$$a_{(i)} = 10^{i-10}, \quad \text{for } \mathcal{I}_n = i, \quad i = 1, \cdots, 13. \tag{53}$$

$N_p = 5000$ particles are used for this problem. The initial distribution of the indicator variable is assumed to have a uniform distribution: $\Pr(\mathcal{I}_{0|0} = i) = 1/m (i = 1, 2, \cdots, m)$. Initial states $\mathbf{x}_{1|0}^{(j)}$ and $V_{1|0}^{(j)}$ are given by $\mathbf{x}_{1|0}^{(j)} = [0, 0, \cdots, 0]^{\mathrm{T}}$ and $V_{1|0}{}^{(j)} = 16\mathbf{l}_{(2M+N_d)\times(2M+N_d)}$, respectively. The lag is set at $L = 20$. The hyperparameter for random benchmark motion, τ, is assumed to be known and is fixed to its correct value in this analysis, whereas we estimate the hyper-parameter for observational error, $\sigma^{(j)}$, using (B8). We consider 11 candidate transition probabilities:

$$\pi_{ij} = \begin{cases} p_{\mathrm{tr}} & i = j \\ (1 - p_{\mathrm{tr}})/(m-1) & i \neq j \end{cases} \tag{54}$$

where p_{tr}=0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.95 and 0.98 and $m = 13$. The transition probability is optimized by maximizing the log-likelihood. Derivation of the log-likelihood is shown in Appendix C. The log-likelihood values for 11 candidate transition probabilities are shown in Table 1. The p_{tr} value that maximizes the log-likelihood is $p_{\mathrm{tr}} = 0.95$.

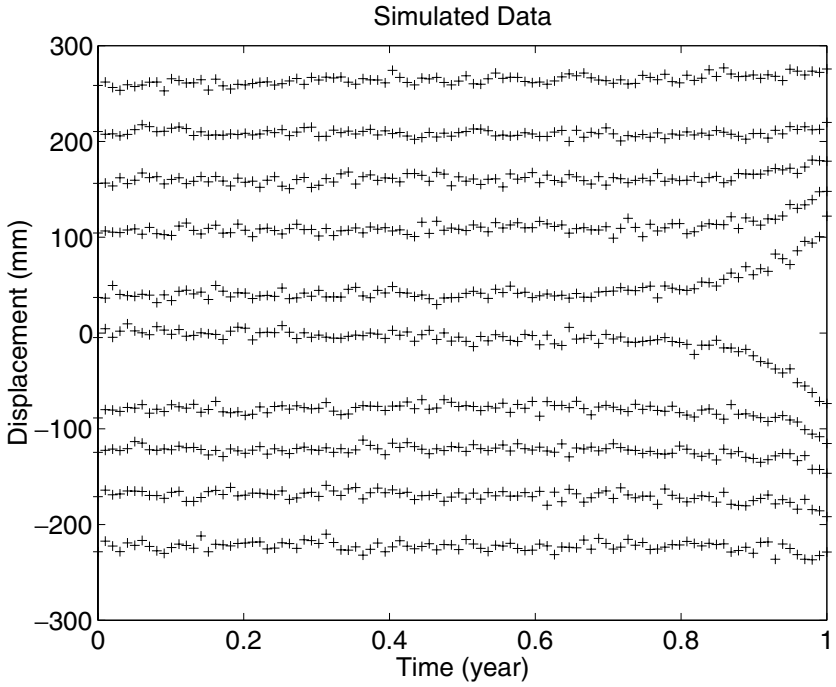

Figure 6. Simulated displacement time-series at the 10 stations due to fault slip (Fig. 5), random benchmark motion and observational error. The SDs of random benchmark motion and observational error are set at $\tau = 0.04$ mm and $\sigma = 4$ mm, respectively. Curves are offset vertically for clarity.

Table 1. The log-likelihood values for 11 candidate transition probabilities defined by (54).

$p_{\rm tr}$	Log-likelihood
0.1	−2815.20
0.2	−2814.11
0.3	−2813.12
0.4	−2811.50
0.5	−2810.45
0.6	−2809.44
0.7	−2808.99
0.8	−2808.46
0.9	−2808.29
0.95	−2808.11
0.98	−2808.41

Fig. 7 shows estimated temporal variation of the smoother distribution of $\mathcal{I}_n$, $p(\mathcal{I}_n|\mathbf{y}_{1:N_e})$, for $p_{\rm tr} = 0.95$. The five curves drawn in Fig. 7 are defined by

$$\sum_{i=1}^{2k} \Pr\left(\mathcal{I}_n = i|\mathbf{y}_{1:N_e}\right), \quad k = 1, \cdots, 5 \tag{55}$$

where k increases from 1 to 5 in order from the lower curve to the upper curve. It should be noted that we can calculate the probability $\Pr(\mathcal{I}_n = i|\mathbf{y}_{1:N_e})$ by

$$\Pr(\mathcal{I}_n = i|\mathbf{y}_{1:N_e}) = \frac{1}{N_p}\left(\text{number of particles with } \mathcal{I}^{(j)}_{n|N_e} = i\right). \tag{56}$$

Fig. 7 shows that the realization probability of $\mathcal{I}_n$, $\Pr(\mathcal{I}_n = i|\mathbf{y}_{1:N_e})$, corresponding to larger α_n/σ values increases after $t = 0.5$, reflecting increasing slip acceleration after $t = 0.5$. Fig. 8 shows the mean value of the estimated slip history with error bounds of 1 SD. It is clear that the estimated slip history shows a good agreement with the true slip history. The mean of the estimated hyperparameters for observational error, $\hat{\sigma} = \sqrt{\frac{1}{N_p}\sum_{j=1}^{N_p} \hat{\sigma}^{(j)2}}$, is $\hat{\sigma} = 3.9149$. Fig. 9(a) compares simulated and calculated displacement time-series for the station located at $x_1 = -33.3$ km. The estimated state reproduces the simulated ground displacements well. It should be noted that the simulated data are sums of contributions from fault slip, random benchmark motion and observational error. Figs 9(b) and (c) show fault slip and random benchmark motion components of the calculated displacement time-series (bold lines), respectively, together with corresponding components of the simulated data (crosses). As can be seen from this figure, the MCMKF can separate spatially coherent fault-slip signals from spatially incoherent random benchmark motion as NIF can (Segall & Matthews 1997).

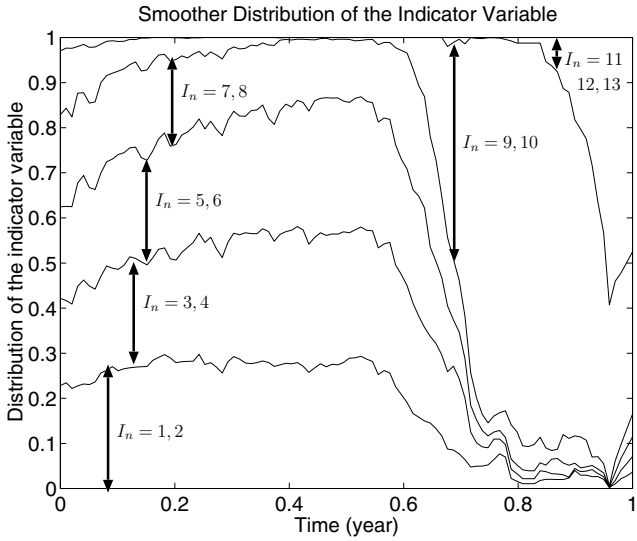

Figure 7. Estimated temporal variation of the smoother distribution of the indicator variable, $p(\mathcal{I}_n|\mathbf{y}_{1:N_e})$, for the maximum-likelihood estimate of the transition probability, $p_{\rm tr} = 0.95$. The five curves are defined by (55).

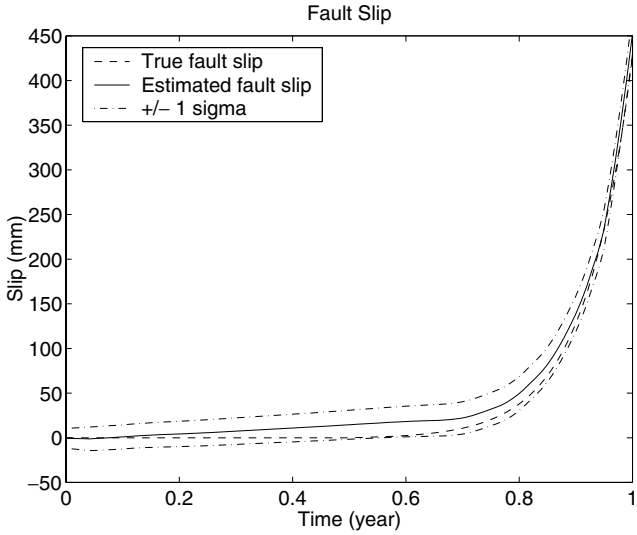

Figure 8. Comparison of the fault-slip history estimated by the MCMKF-based method with the true input signal. The solid curve shows the mean value of the slip history estimated by the MCMKF for the maximum-likelihood estimate of the transition probability, $p_{\rm tr} = 0.95$. The dash-dotted curves denote its SD error bounds. The broken curve shows the true input signal.

To illustrate an advantage of the MCMKF algorithm over a Kalman filter, we estimate state evolution from observation model (8) and system model (17) with the temporally invariable hyperparameter for slip acceleration (i.e. $\alpha_n = \alpha$ in eq. 21), following Segall & Matthews (1997), using a Kalman filter. The hyperparameter τ for random benchmark motion is again fixed to its correct value, and α and σ are estimated by the maximum-likelihood method. α and σ are optimized as $\hat{\alpha} = 3.0661$ and $\hat{\sigma} = 3.9071$, respectively. Fig. 10 shows the estimated slip history with 1 SD error bounds. The slip history estimated by the Kalman filter (Fig. 10) is oscillatory compared with the true slip history in the period $t < 0.8$. On the other hand, the slip history estimated by the MCMKF (Fig. 8) reproduces true slip more accurately throughout the entire period. We compare the goodness of the models using the Akaike information criterion (AIC) (Akaike 1974) defined by (C4). Unknown parameters for the MCMKF are $\mathbf{x}_{1|0}$, σ, $\mathcal{I}_{0|0}$ and $p_{\rm tr}$. Thus number of unknown parameters is $2M + N_d + 3$ ($2M + N_d$ is the dimension of the state vector). On the other hand, unknown parameters in the case of the Kalman filter are $\mathbf{x}_{1|0}$, σ and α. Therefore, the number of unknown parameters is $2M + N_d + 2$. From (C4), the AIC value for the MCMKF is 5646.22, whereas that for the Kalman filter is 5793.84. This result indicates that we can obtain a better state estimate using the MCMKF algorithm than the Kalman filter.

In the example shown here, we use all N_p trajectories for state estimation. This procedure requires a heavy computational burden because we need to run the Kalman filter and smoother N_p times to obtain $\{\mathbf{x}^{(j)}_{n|N_e}, V^{(j)}_{n|N_e}\}^{N_p}_{j=1}$. However, as described in Section 3.5, it is possible to reduce this computational burden by decreasing the number of trajectories to obtain the state. Here, we examine the effects of the number of trajectories. We use $N'_p (N'_p < N_p)$ trajectories that are randomly sampled from $N_p = 5000$ trajectories. Then, the log-likelihood values for

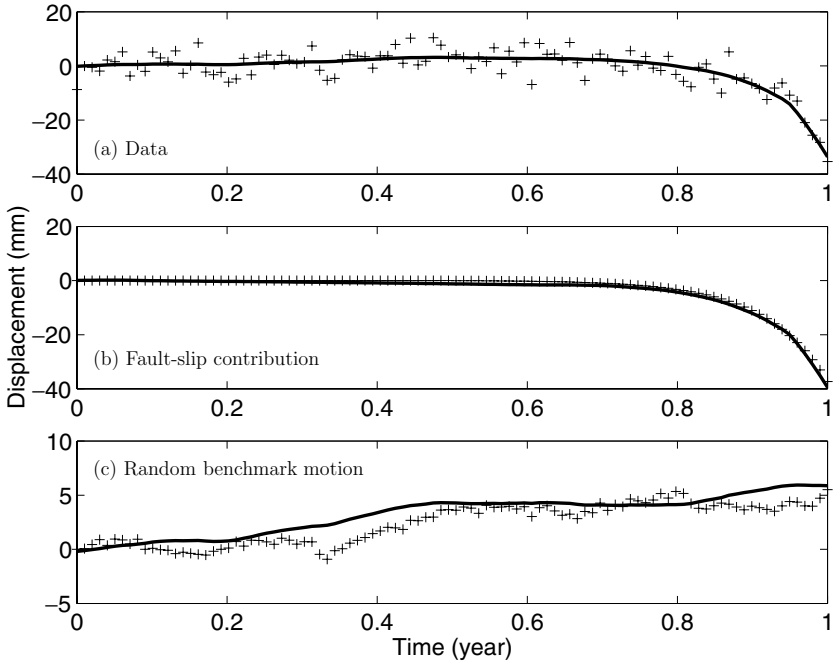

Figure 9. (a) Simulated data (crosses) and calculated displacements from the estimated state (bold line) for the station located at $x_1 = -33.3$ km. Simulated data are sums of contributions from fault slip, random benchmark motion and observational error. Calculated displacements contains the fault-slip contribution and the random benchmark motion contribution. (b) Displacements due to true fault slip (crosses) and estimated fault-slip evolution (bold line) for the same station. (c) True (crosses) and estimated (bold line) random benchmark motion for the same station.

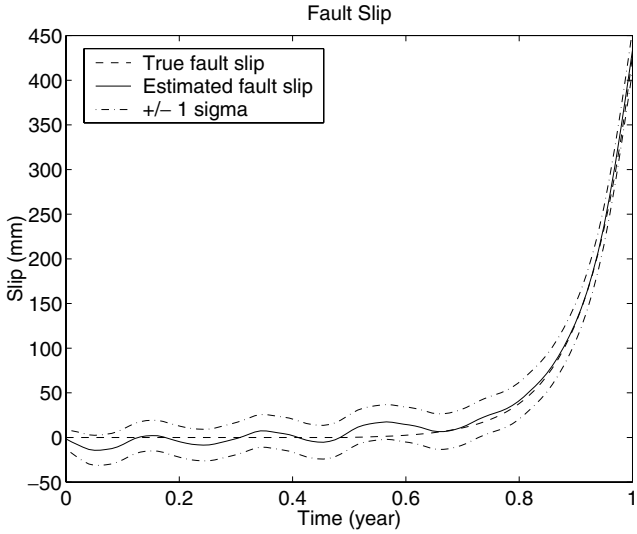

Figure 10. Comparison of the fault-slip history estimated by the Kalman filter-based method with the true input signal. The solid curve shows the mean value of the slip history estimated by the Kalman filter with the dash-dotted curves denoting its SD error bounds. The broken curve shows the true input signal.

each case are computed and compared with one another. Table 2 lists log-likelihood values for various N'_p values. One can see that for the cases of $N'_p \geq 4$, the log-likelihood values are rather insensitive to the number of trajectories and the differences of the log-likelihood values are at most 0.8. Estimated fault slip and random benchmark motion are visually indistinguishable for values $N'_p \geq 4$. This result suggests that the computational burden can be significantly reduced by decreasing the number of trajectories.

4.2 High signal-to-noise ratio case

The previous example assumed that the fault-slip signal started smoothly. In this subsection, we consider the case in which the transient event starts more rapidly. The true slip history used in this subsection is shown in Fig. 11. Rapid changes of slip rate are seen at $t = 0.45$ and $t = 0.55$. Because we confirmed in the previous subsection that the number of trajectories used for state estimation can be reduced without significant effects on state estimate, we use $N'_p = 20$ trajectories. Simulated data that contain fault-slip signal, random benchmark motion and

Table 2. The log-likelihood values for various number of trajectories to which the Kalman filter is applied. The trajectories are sampled randomly from $N_p = 5000$ trajectories.

N'_p	Log-likelihood
5000	−2808.11
2000	−2808.04
1000	−2808.10
500	−2808.09
200	−2808.05
100	−2807.94
80	−2807.80
60	−2807.69
40	−2807.41
20	−2807.81
10	−2807.73
9	−2807.82
8	−2807.59
7	−2807.41
6	−2807.54
5	−2807.79
4	−2808.19
3	−2808.93
2	−2809.30
1	−2811.33

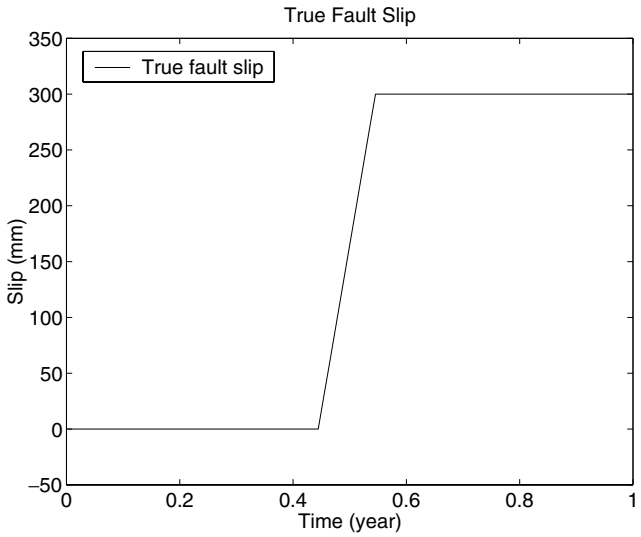

Figure 11. True slip history given on the fault surface.

observational error are shown in Fig. 12. We employ $m = 13$ candidate values for α_n/σ defined by (53). The number of particles, the initial distribution of the indicator variable, the initial state and the lag are the same as in Section 4.1. The hyperparameter for random benchmark motion, τ, is again assumed to be known and is fixed to its correct value and $\sigma^{(j)}$ is determined using (B8). We employ nine candidate transition probabilities, $p_{\rm tr} = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9$ where $p_{\rm tr}$ is defined by (54). The best transition probability is determined by maximizing the log-likelihood defined by (C3) with N_p replaced by N'_p. The log-likelihood values for the nine transition probabilities are listed in Table 3. The optimal value of $p_{\rm tr}$ is 0.8.

Fig. 13 shows estimated temporal variations of the smoother distribution of $\mathcal{I}_n$, $p(\mathcal{I}_n|\mathbf{y}_{1:N_e})$, for $p_{\rm tr} = 0.8$. The five curves drawn in Fig. 13 are defined by (55). From Fig. 13, we can see that the realization probability of $\mathcal{I}_n$ corresponding to larger α_n/σ values suddenly increases at time points of slip-rate discontinuities. Fig. 14 shows the estimated slip history with 1 SD error bounds. The estimated slip history closely reproduces the true slip history denoted by the broken line in Fig. 14. The mean of the estimated hyperparameters for observational error, $\hat{\sigma} = \sqrt{\frac{1}{N'_p}\sum_{j=1}^{N'_p}\hat{\sigma}^{(j)2}}$, is $\hat{\sigma} = 3.9040$. Fig. 15(a) compares simulated and calculated displacement time-series for the station located at $x_1 = 100.0$ km. The estimated state closely reproduces the simulated ground displacements. Figs 15(b) and (c) show fault slip and random benchmark motion components of the calculated displacement time-series (bold lines), respectively, together with corresponding components

Table 3. The log-likelihood values for nine candidate transition probabilities defined by (54).

$p_{\rm tr}$	Log-likelihood
0.1	−2817.97
0.2	−2815.14
0.3	−2813.60
0.4	−2813.13
0.5	−2813.57
0.6	−2811.08
0.7	−2811.12
0.8	−2809.86
0.9	−2810.23

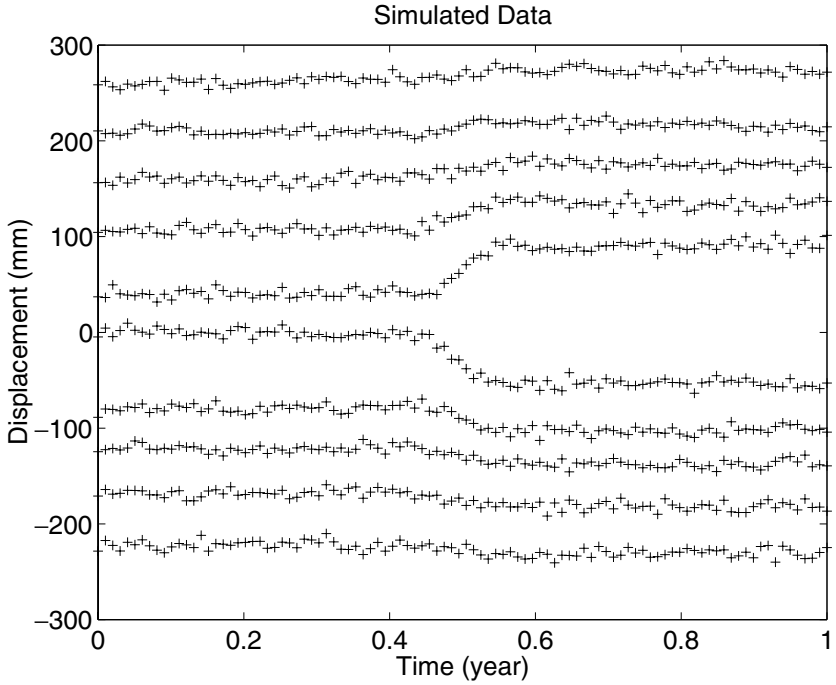

Figure 12. Simulated displacement time-series at the 10 stations due to fault slip (Fig. 11), random benchmark motion and observational error. The SDs of random benchmark motion and observational error are set at $\tau = 0.04$ mm and $\sigma = 4$ mm, respectively. Curves are offset vertically for clarity.

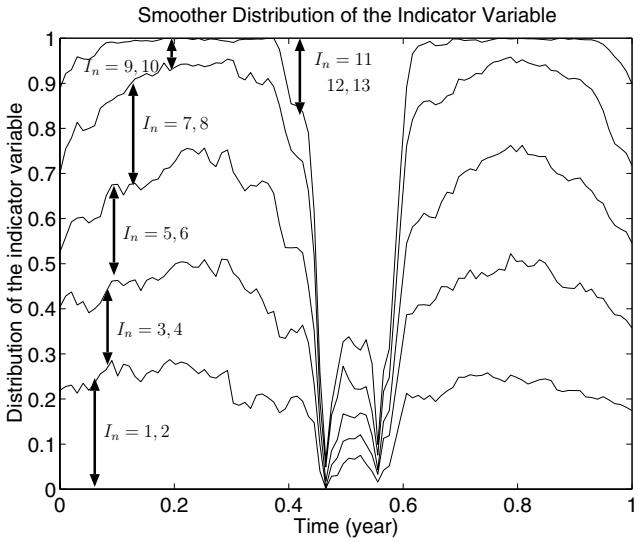

Figure 13. Estimated temporal variation of the smoother distribution of the indicator variable, $p(\mathcal{I}_n|\mathbf{y}_{1:N_{\rm e}})$, for the maximum-likelihood estimate of the transition probability, $p_{\rm tr} = 0.8$. The five curves are defined by (55).

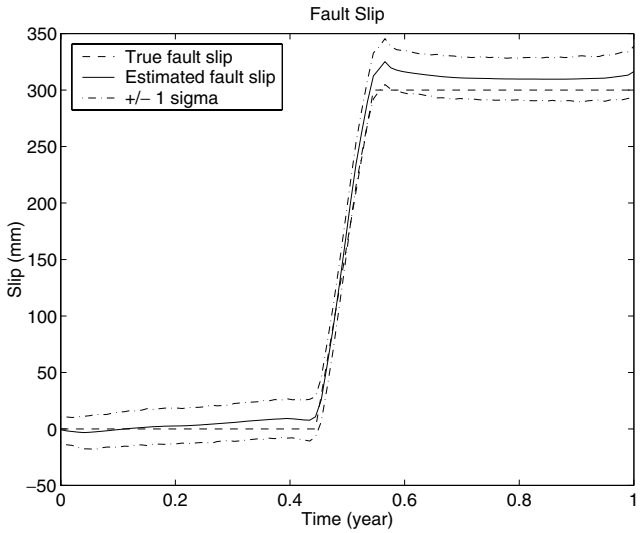

Figure 14. Comparison of the fault-slip history estimated by the MCMKF-based method with the true input signal. The solid curve shows the mean value of the slip history estimated by the MCMKF for the maximum-likelihood estimate of the transition probability, $p_{\mathrm{tr}} = 0.8$. The dash-dotted curves denote its SD error bounds. The broken curve shows the true input signal.

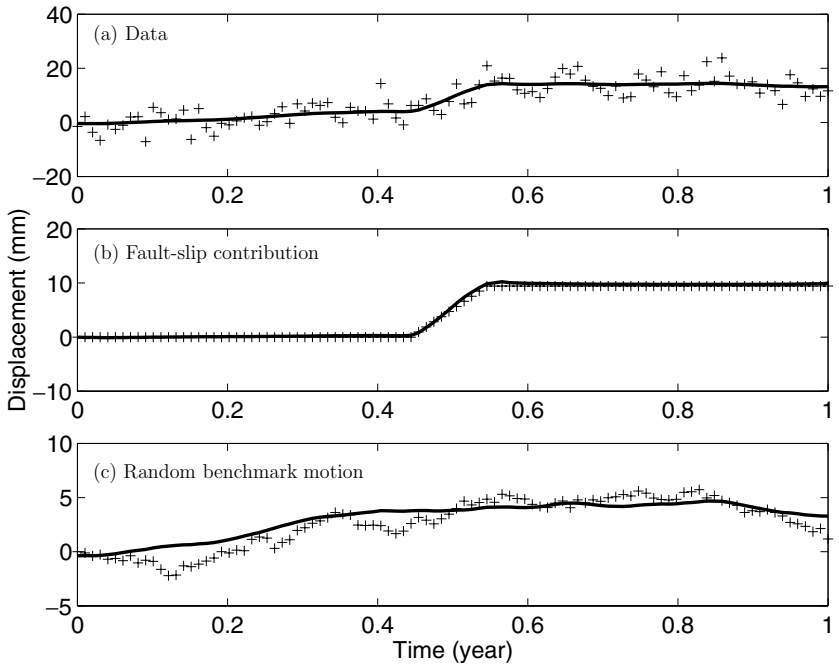

Figure 15. (a) Simulated data (crosses) and calculated displacements from the estimated state (bold line) for the station located at $x_1 = 100.0$ km. Simulated data are sums of contributions from fault slip, random benchmark motion and observational error. Calculated displacements contains the fault-slip contribution and the random benchmark motion contribution. (b) Displacements due to true fault slip (crosses) and estimated fault-slip evolution (bold line) for the same station. (c) True (crosses) and estimated (bold line) random benchmark motion for the same station.

of the simulated data (crosses). In this example, we again see that the MCMKF is capable of separating transient signal from random benchmark motion.

Next, we estimate state evolution with the temporally invariable hyperparameter for slip acceleration, α, using the Kalman filter (Segall & Matthews 1997). In this case, maximum-likelihood estimates of α and σ are $\hat{\alpha} = 5.6026$ and $\hat{\sigma} = 3.8949$, respectively. Fig. 16 shows estimated slip history with 1 SD error bounds. The estimated slip history is oscillatory in quiet periods, whereas slip-rate discontinuities are overly smoothed. Therefore, the Kalman filter-based method cannot image the initial process of the transient event accurately, nor can it infer precisely when the event started. On the other hand, the slip history estimated by the MCMKF (Fig. 14) provides a better reproduction of both sudden changes of slip rate and quiet portions of the signal. Fig. 14 indicates that initial process of the event can be imaged more accurately using the MCMKF-based method. Furthermore, the start time of the event can be inferred more precisely based on the estimated slip history (Fig. 14) and the temporal variation of the distribution of the indicator variable (Fig. 13). The AIC value for the MCMKF is 5649.72, which is

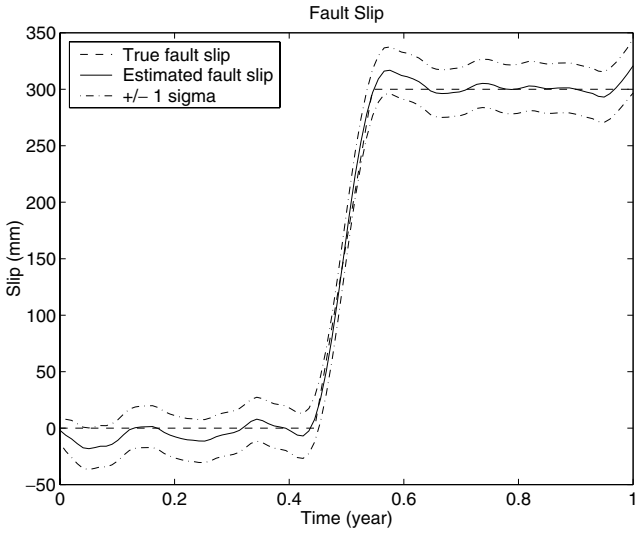

Figure 16. Comparison of the fault-slip history estimated by the Kalman filter-based method with the true input signal. The solid curve shows the mean value of the slip history estimated by the Kalman filter with the dash-dotted curves denoting its SD error bounds. The broken curve shows the true input signal.

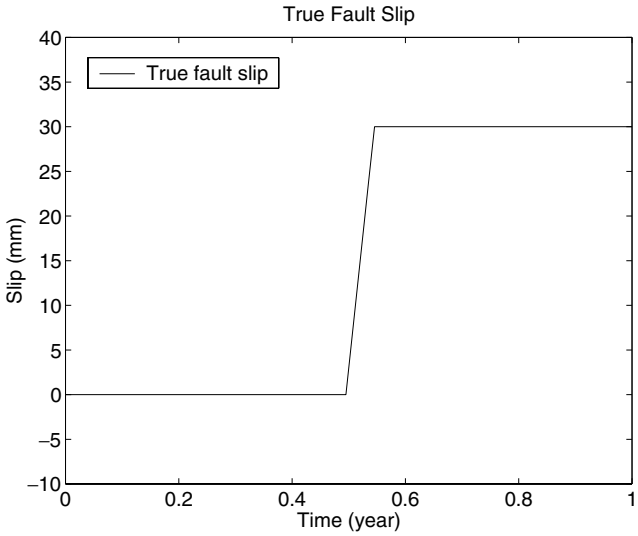

Figure 17. True slip history given on the fault surface.

significantly smaller than the value obtained by the Kalman filter, 5810.60. This result again confirms that we can obtain a better state estimate using the MCMKF than with the Kalman filter.

4.3 Low signal-to-noise ratio case

One of our motivations for developing the new time-dependent inversion method is to elucidate small slow slip events that are buried in noisy time-series. In this subsection, we examine whether our new scheme is capable of detecting such 'invisible' signals. We use the true fault-slip signal shown in Fig. 17. Sudden changes of slip rate are seen at $t = 0.50$ and $t = 0.55$. The cumulative slip is only 30 mm, which is one-tenth of the previous example. Simulated data, which consist of fault-slip signal, random benchmark motion, and observational error, are shown in Fig. 18. Note that the transient signal is invisible at all stations. We employ $m = 13$ candidate values for α_n/σ defined by (53), as in the previous examples. For the state estimation, $N'_p = 20$ trajectories are used. The number of particles, the initial distribution of the indicator variable, the initial state and the lag are kept the same as in the previous examples. The hyperparameter for random benchmark motion, τ, is again assumed to be known and is fixed at its correct value and $\sigma^{(j)}$ is determined using (B8). We employ 11 candidate transition probabilities, $p_{\rm tr} = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.95$ and 0.98 where $p_{\rm tr}$ is defined by (54). The transition probability is optimized by maximizing the log-likelihood defined by (C3) with N_p replaced by N'_p. The log-likelihood values for 11 candidate transition probabilities are shown in Table 4. The $p_{\rm tr}$ value that maximizes the log-likelihood is $p_{\rm tr} = 0.95$.

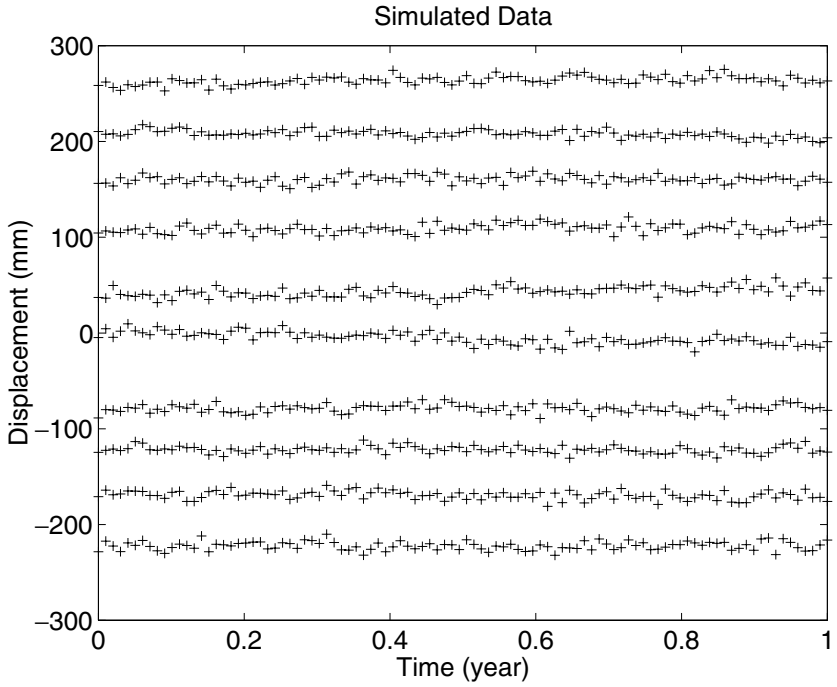

Figure 18. Simulated displacement time-series at the 10 stations due to fault slip (Fig. 17), random benchmark motion and observational error. The SDs of random benchmark motion and observational error are set at $\tau = 0.04$ mm and $\sigma = 4$ mm, respectively. Curves are offset vertically for clarity.

Table 4. The log-likelihood values for 11 candidate transition probabilities defined by (54).

$p_{\rm tr}$	Log-likelihood
0.1	−2813.62
0.2	−2812.64
0.3	−2809.53
0.4	−2810.27
0.5	−2808.09
0.6	−2808.60
0.7	−2807.60
0.8	−2806.58
0.9	−2806.54
0.95	−2805.01
0.98	−2805.34

Fig. 19 shows the estimated temporal variation of the smoother distribution of $\mathcal{I}_n$, $p(\mathcal{I}_n|\mathbf{y}_{1:N_e})$, for $p_{\rm tr} = 0.95$. The five curves drawn in Fig. 19 are defined by (55). From Fig. 19, the realization probability of $\mathcal{I}_n$ corresponding to larger α_n/σ values is highest at $t = 0.5$ when the slip rate suddenly increases. However, the distribution of $\mathcal{I}_n$ is excessively smoothed, reflecting a low signal-to-noise ratio. Fig. 20 shows estimated slip history with 1 SD error bounds. Although the estimated slip history is smoother compared with the true slip history, transient fault slip is clearly detected. The mean of the estimated hyperparameters for observational error, $\hat{\sigma} = \sqrt{\frac{1}{N'_p}\sum_{j=1}^{N'_p}\hat{\sigma}^{(j)2}}$, is $\hat{\sigma} = 3.9166$. Fig. 21(a) compares simulated and calculated displacement time-series for the station located at $x_1 = 77.8$ km. Figs 21(b) and (c) show fault slip and random benchmark motion components of the calculated displacement time-series (bold lines), respectively, together with corresponding components of the simulated data (crosses). Even in this noisy example, the MCMKF is capable of distinguishing the fault-slip signal from random benchmark motion.

For the comparison, we estimate state evolution with the temporally invariable hyperparameter for slip acceleration, α, using the Kalman filter (Segall & Matthews 1997). In this case, maximum-likelihood estimates of α and σ are $\hat{\alpha} = 0.5009$ and $\hat{\sigma} = 3.9313$, respectively. Fig. 22 shows the estimated slip history with 1 SD error bounds. The estimated slip history is excessively smoothed throughout the entire period. It is obvious that the MCMKF-based method does a better job than the Kalman filter-based method (see Figs 20 and 22). The AIC value for the MCMKF is 5640.02, which is significantly smaller than the value obtained by the Kalman filter, 5770.52. This result suggests that the MCMKF is superior to the Kalman filter in detecting a small fault-slip signal from a noisy data set.

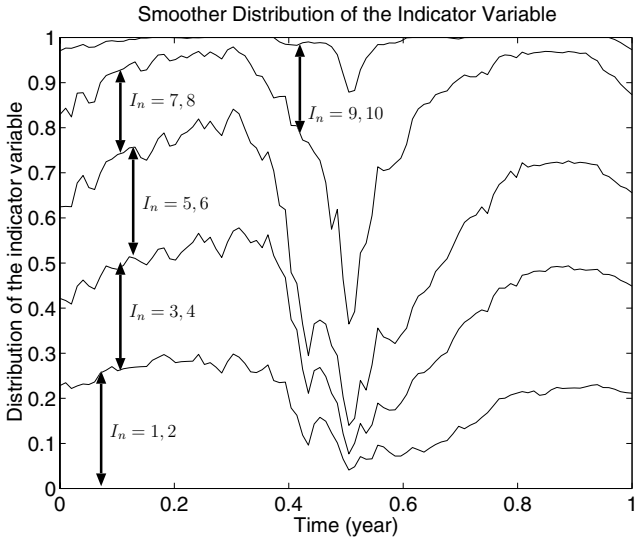

Figure 19. Estimated temporal variation of the smoother distribution of the indicator variable, $p(\mathcal{I}_n|\mathbf{y}_{1:N_e})$, for the maximum-likelihood estimate of the transition probability, $p_{tr} = 0.95$. The five curves are defined by (55).

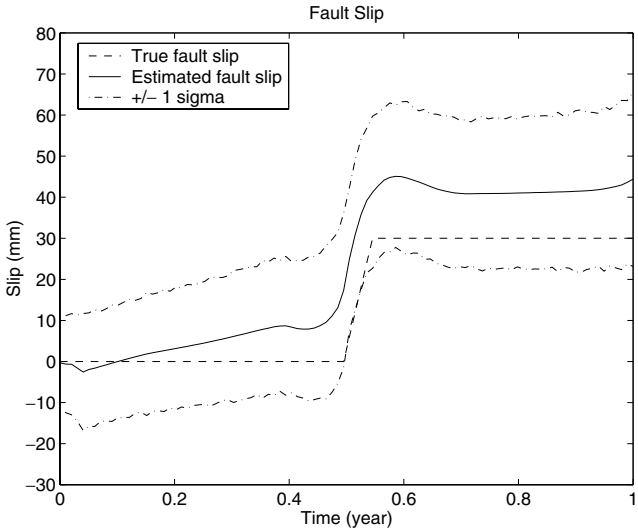

Figure 20. Comparison of the fault-slip history estimated by the MCMKF-based method with the true input signal. The solid curve shows the mean value of the slip history estimated by the MCMKF for the maximum-likelihood estimate of the transition probability, $p_{tr} = 0.8$. The dash-dotted curves denote its SD error bounds. The broken curve shows the true input signal.

5 CONCLUSION

We have developed a new inversion method for the imaging of time-dependent fault slip during a transient slip event from surface deformation data. We modelled crustal deformation resulting from the transient fault slip using state space representation. To improve the temporal resolution of the fault slip, we assumed that α_n/σ, which had been assumed to be a temporally invariable hyperparameter in the previous studies, was temporally variable. To realize the time dependency of α_n/σ, we introduced a stochastic variable $\mathcal{I}_n$, an indicator variable, which specified the value of α_n/σ at time t_n. To estimate the indicator variable and the state, we developed a new filtering method, the Monte Carlo mixture Kalman filter (MCMKF). In the MCMKF, the predictive and the filter distributions of the indicator variable are approximated by independent realizations (particles) from each distribution and approximated predictive and filter distributions are calculated recursively by using the prediction and the filtering algorithms. Once all the trajectories for the smoother distribution of $\mathcal{I}_n$ are obtained, the evolution of the state for each trajectory is estimated by running the Kalman filter and smoother.

We demonstrated the validity of the new MCMKF-based inversion method through three numerical experiments using simulated surface displacement time-series. The results were compared with those obtained using the conventional Kalman filter-based method. The MCMKF clearly reproduces the true slip history better than the Kalman filter. For the high signal-to-noise ratio case, the MCMKF is better capable of

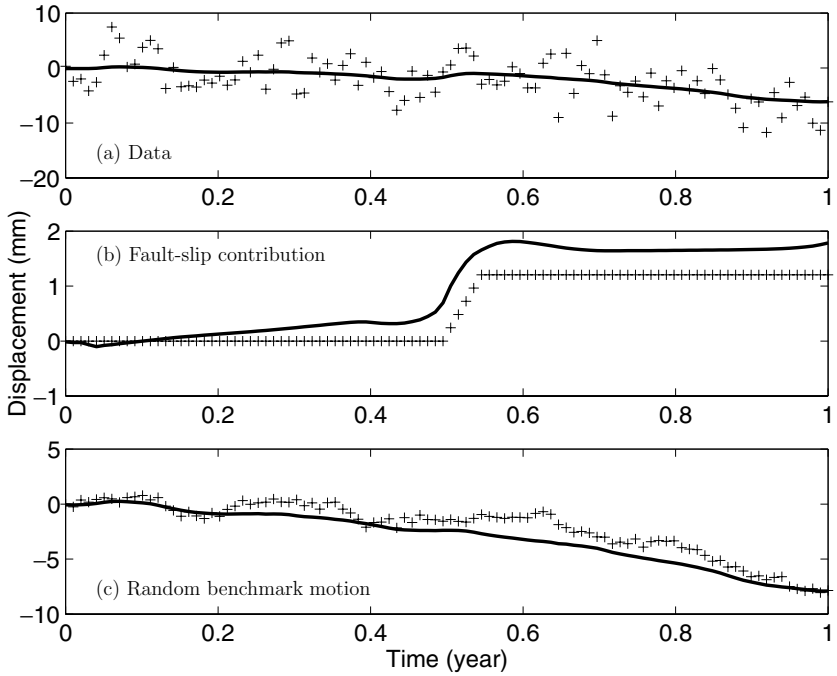

Figure 21. (a) Simulated data (crosses) and calculated displacements from the estimated state (bold line) for the station located at $x_1 = 77.8$ km. Simulated data are sums of contributions from fault slip, random benchmark motion and observational error. Calculated displacements contains the fault-slip contribution and the random benchmark motion contribution. (b) Displacements due to true fault slip (crosses) and estimated fault-slip evolution (bold line) for the same station. (c) True (crosses) and estimated (bold line) random benchmark motion for the same station.

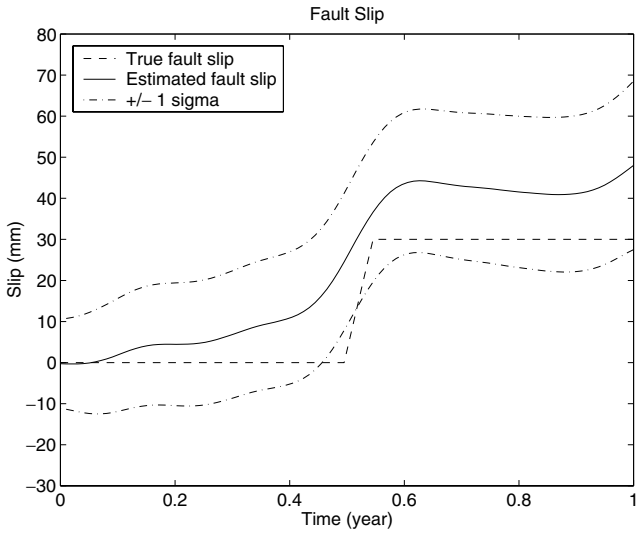

Figure 22. Comparison of the fault-slip history estimated by the Kalman filter-based method with the true input signal. The solid curve shows the mean value of the slip history estimated by the Kalman filter with the dash-dotted curves denoting its SD error bounds. The broken curve shows the true input signal.

imaging the initiation process of a transient slip event than the Kalman filter. For the low signal-to-noise ratio case, we found that the MCMKF is superior to the Kalman filter in detecting small fault-slip signals from a noisy data set. Furthermore, the following results are obtained:

(1) The MCMKF gives significantly smaller AIC values than the Kalman filter. This indicates that the MCMKF yields a better state estimate than the Kalman filter.

(2) The computational burden of the MCMKF can be reduced by decreasing the number of trajectories to which the Kalman filter is applied. Our numerical experiments confirm that a small number of trajectories (say 10 or 20) is enough to estimate state evolution.

(3) The MCMKF can distinguish a spatially coherent fault-slip signal from a spatially incoherent random benchmark motion even in the low signal-to-noise ratio case.

ACKNOWLEDGMENTS

We thank Yosuke Aoki for helpful comments. Reviews by Malcolm Sambridge, James C. Savage and an anonymous reviewer improved the manuscript. This work was partly supported by Grant-in-Aid for Scientific Research (A) 14208025, from Ministry of Education, Culture, Sports, Science and Technology, Japan.

REFERENCES

Akaike, H., 1974. A new look at the statistical model identification, *IEEE Trans. Autom. Control,* **AC-19,** ,716–723.

Anderson, B.D.O. & Moore, J.B., 1979. *Optimal Filtering,* Prentice Hall, Englewood Cliffs, NJ.

Aoki, Y., Segall, P., Kato, T., Cervelli, P. & Shimada, S., 1999. Imaging magma transport during the 1997 seismic swarm off the Izu Peninsula, Japan, *Science,* **286,** 927–930.

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.

Chen, R. & Liu, J.S., 2000. Mixture Kalman filters, *J. R. Statist. Soc. B,* **62,** 493–508.

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

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.

Ito, T., Yoshioka, S. & Miyazaki, S., 2000. Interplate coupling in northeast Japan deduced from inversion analysis of GPS data, *Earth planet. Sci. Lett.,* **176,** 117–130.

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, New York.

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

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., 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.

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

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.

Pacheco, J.F., Sykes, L.R. & Scholtz, C.H., 1993. Nature of seismic coupling along simple plate boundaries of the subduction type, *J. geophys. Res.,* **98,** 14 133–14 159.

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.

Wyatt, F., 1982. Displacement of surface monuments: horizontal motion, *J. geophys. Res.,* **87,** 979–989.

Wyatt, F., 1989. Displacement of surface monuments: vertical motion, *J. geophys. Res.,* **94,** 1655–1664.

Yagi, Y., 2002. Complementary relation of seismic and aseismic slip in Hyganada and Sanriku region, Japan, *PhD Thesis,* University of Tokyo, Tokyo.

APPENDIX A: DERIVATION OF THE IMPORTANCE WEIGHT

In this appendix we show that the importance weight $w_n^{(j)}$ defined in (43) can be calculated by using the Kalman filter. Consider the conditional density functions

$$p\left(\mathbf{x}_{n-1}|\mathcal{I}_{1:n-1}=\mathcal{I}_{1:n-1|n-1}^{(j)},\mathbf{y}_{1:n-1}\right) \tag{A1}$$

$$p\left(\mathbf{x}_{n}|\mathcal{I}_{1:n}=\mathcal{I}_{1:n|n-1}^{(j)},\mathbf{y}_{1:n-1}\right) \tag{A2}$$

$$p\left(\mathbf{x}_{n}|\mathcal{I}_{1:n}=\mathcal{I}_{1:n|n-1}^{(j)},\mathbf{y}_{1:n}\right) \tag{A3}$$

$$p\left(\mathbf{x}_{n}|\mathcal{I}_{1:n}=\mathcal{I}_{1:n|n}^{(j)},\mathbf{y}_{1:n}\right). \tag{A4}$$

As described in Section 3.1, these density functions become Gaussian distributions:

$$p\left(\mathbf{x}_{n-1}|\mathcal{I}_{1:n-1}=\mathcal{I}_{1:n-1|n-1}^{(j)},\mathbf{y}_{1:n-1}\right)\sim N\left(\bar{\mathbf{x}}_{n-1|n-1}^{(j)},\sigma^{(j)2}\bar{V}_{n-1|n-1}^{(j)}\right) \tag{A5}$$

$$p\left(\mathbf{x}_{n}|\mathcal{I}_{1:n}=\mathcal{I}_{1:n|n-1}^{(j)},\mathbf{y}_{1:n-1}\right)\sim N\left(\tilde{\mathbf{x}}_{n|n-1}^{(j)},\sigma^{(j)2}\tilde{V}_{n|n-1}^{(j)}\right) \tag{A6}$$

$$p\left(\mathbf{x}_{n}|\mathcal{I}_{1:n}=\mathcal{I}_{1:n|n-1}^{(j)},\mathbf{y}_{1:n}\right)\sim N\left(\tilde{\mathbf{x}}_{n|n}^{(j)},\sigma^{(j)2}\tilde{V}_{n|n}^{(j)}\right) \tag{A7}$$

$$p\left(\mathbf{x}_{n}|\mathcal{I}_{1:n}=\mathcal{I}_{1:n|n}^{(j)},\mathbf{y}_{1:n}\right)\sim N\left(\bar{\mathbf{x}}_{n|n}^{(j)},\sigma^{(j)2}\bar{V}_{n|n}^{(j)}\right). \tag{A8}$$

Assume that $(\bar{\mathbf{x}}_{n-1|n-1}^{(j)},\ \bar{V}_{n-1|n-1}^{(j)})$ and $\mathcal{I}_{1:n-1|n-1}^{(j)}(j=1,\ldots,N_p)$ are given. First, we generate $\mathcal{I}_{1:n|n-1}^{(j)}$ by the prediction scheme of the MCMKF (Section 3.3), and then obtain $(\tilde{\mathbf{x}}_{n-1|n-1}^{(j)},\ \tilde{V}_{n-1|n-1}^{(j)})$ using the prediction scheme of the Kalman filter (e.g. Kitagawa & Gersch 1996) as follows:

$$\tilde{\mathbf{x}}_{n|n-1}^{(j)}=\mathbf{F}_n\bar{\mathbf{x}}_{n-1|n-1}^{(j)} \tag{A9}$$

$$\tilde{V}^{(j)}_{n|n-1} = \mathbf{F}_n \bar{V}^{(j)}_{n-1|n-1}\mathbf{F}_n^{\mathrm{T}} + \tilde{\mathbf{Q}}_n\big(\mathcal{I}^{(j)}_{n|n-1}\big). \tag{A10}$$

From (38), the predictive distribution of data defined by $p(\mathbf{y}_n|\mathcal{I}_{1:n} = \mathcal{I}^{(j)}_{1:n|n-1}, \mathbf{y}_{1:n-1})$ also becomes a Gaussian distribution:

$$p\big(\mathbf{y}_n|\mathcal{I}_{1:n} = \mathcal{I}^{(j)}_{1:n|n-1}, \mathbf{y}_{1:n-1}\big) \sim N\big(\tilde{\mathbf{y}}^{(j)}_{n|n-1}, \sigma^{(j)2}\tilde{W}^{(j)}_{n|n-1}\big) \tag{A11}$$

where $\tilde{\mathbf{y}}^{(j)}_{n|n-1}$ and $\tilde{W}^{(j)}_{n|n-1}$ are the mean vector and the covariance matrix of the predictive distribution of $\mathbf{y}_n$, respectively, and are defined by

$$\tilde{\mathbf{y}}^{(j)}_{n|n-1} = \mathbf{H}\tilde{\mathbf{x}}^{(j)}_{n|n-1} \tag{A12}$$

$$\tilde{W}^{(j)}_{n|n-1} = \mathbf{H}\tilde{V}^{(j)}_{n|n-1}\mathbf{H}^{\mathrm{T}} + \tilde{\mathbf{R}}_n. \tag{A13}$$

The left side of (A11) is identical with the importance weight $w^{(j)}_n$ defined in (43). Thus $w^{(j)}_n$ follows the Gaussian distribution with mean $\tilde{\mathbf{y}}^{(j)}_{n|n-1}$ and covariance matrix $\tilde{W}^{(j)}_{n|n-1}$ as follows:

$$w^{(j)}_n = \big(2\pi\sigma^{(j)2}\big)^{-N_d/2}\big|\tilde{W}^{(j)}_{n|n-1}\big|^{-1/2}\exp\left[-\frac{1}{2\sigma^{(j)2}}\big(\mathbf{y}_n - \tilde{\mathbf{y}}^{(j)}_{n|n-1}\big)^{\mathrm{T}}W^{(j)-1}_{n|n-1}\big(\mathbf{y}_n - \tilde{\mathbf{y}}^{(j)}_{n|n-1}\big)\right] \tag{A14}$$

where $|\tilde{W}^{(j)}_{n|n-1}|$ is the absolute value of the determinant of $\tilde{W}^{(j)}_{n|n-1}$. The unknown parameter $\sigma^{(j)2}$ is estimated by maximizing $w^{(j)}_n$. The result is given by

$$\sigma^{(j)2} = \frac{1}{N_d}\big(\mathbf{y}_n - \tilde{\mathbf{y}}^{(j)}_{n|n-1}\big)^{T}\tilde{W}^{(j)-1}_{n|n-1}\big(\mathbf{y}_n - \tilde{\mathbf{y}}^{(j)}_{n|n-1}\big). \tag{A15}$$

It is necessary to obtain $(\bar{\mathbf{x}}^{(j)}_{n|n}, \bar{V}^{(j)}_{n|n})$ to calculate the importance weight at the next epoch, $w^{(j)}_{n+1}$. For this purpose, first, $(\tilde{\mathbf{x}}^{(j)}_{n|n}, \tilde{V}^{(j)}_{n|n})$ are obtained using the Kalman filter (e.g. Kitagawa & Gersch 1996):

$$\tilde{\mathbf{x}}^{(j)}_{n|n} = \tilde{\mathbf{x}}^{(j)}_{n|n-1} + K^{(j)}_n\big(\mathbf{y}_n - \mathbf{H}\tilde{\mathbf{x}}^{(j)}_{n|n-1}\big) \tag{A16}$$

$$\tilde{V}^{(j)}_{n|n} = \big(\mathbf{I} - K^{(j)}_n\mathbf{H}\big)\tilde{V}^{(j)}_{n|n-1} \tag{A17}$$

where

$$K^{(j)}_n = \tilde{V}^{(j)}_{n|n-1}\mathbf{H}^{\mathrm{T}}\big(\mathbf{H}\tilde{V}^{(j)}_{n|n-1}\mathbf{H}^{\mathrm{T}} + \tilde{\mathbf{R}}_n\big)^{-1}. \tag{A18}$$

It should be noted that whereas $(\tilde{\mathbf{x}}^{(j)}_{n|n}, \tilde{V}^{(j)}_{n|n})$ is obtained using the particle $\mathcal{I}^{(j)}_{1:n|n-1}$, $(\bar{\mathbf{x}}^{(j)}_{n|n}, \bar{V}^{(j)}_{n|n})$ is conditional on the particle $\mathcal{I}^{(j)}_{1:n|n}$. Because $\{\mathcal{I}^{(1)}_{1:n|n}, \cdots, \mathcal{I}^{(N_p)}_{1:n|n}\}$ are obtained by the resampling of $\{\mathcal{I}^{(1)}_{1:n|n-1}, \cdots, \mathcal{I}^{(N_p)}_{1:n|n-1}\}$ with sampling probabilities (45) (Section 3.4), $\{\bar{\mathbf{x}}^{(j)}_{n|n}, \bar{V}^{(j)}_{n|n}\}^{N_p}_{j=1}$ can be obtained by modifying the resampling algorithm as follows:

Generate $\{\mathcal{I}^{(j)}_{1:n|n}, \bar{\mathbf{x}}^{(j)}_{n|n}, \bar{V}^{(j)}_{n|n}\}^{N_p}_{j=1}$ by the resampling of $\{\mathcal{I}^{(j)}_{1:n|n-1}, \tilde{\mathbf{x}}^{(j)}_{n|n}, \tilde{V}^{(j)}_{n|n}\}^{N_p}_{j=1}$ with sampling probabilities (45).

APPENDIX B: ESTIMATION OF THE HYPERPARAMETER σ

The hyperparameter $\sigma^{(j)}$ is estimated by maximizing the likelihood for jth trajectory. Assume that the hyperparameter for random benchmark motion, τ, is known by analysing each geodetic time-series separately (e.g. Langbein & Johnson 1997). In this case, the likelihood for the jth trajectory, which is a function of $\sigma^{(j)}$, is expressed as follows (e.g. Kitagawa & Gersch 1996):

$$\begin{aligned} L^{(j)}\big(\sigma^{(j)}\big) &= p\big(\mathbf{y}_{1:N_\mathrm{e}}|\mathcal{I}_{1:N_\mathrm{e}} = \mathcal{I}^{(j)}_{1:N_\mathrm{e}|N_\mathrm{e}}, \tau, \sigma^{(j)}\big) \\ &= \prod_{n=1}^{N_\mathrm{e}} p\big(\mathbf{y}_n|\mathbf{y}_{1:n-1}, \mathcal{I}_{1:n} = \mathcal{I}^{(j)}_{1:n|N_\mathrm{e}}, \tau, \sigma^{(j)}\big) \\ &= \prod_{n=1}^{N_\mathrm{e}} \tilde{w}^{(j)}_n \end{aligned} \tag{B1}$$

where

$$\begin{aligned} \tilde{w}^{(j)}_n &= p\big(\mathbf{y}_n|\mathbf{y}_{1:n-1}, \mathcal{I}_{1:n} = \mathcal{I}^{(j)}_{1:n|N_\mathrm{e}}, \tau, \sigma^{(j)}\big) \\ &= \big(2\pi\sigma^{(j)2}\big)^{-N_d/2}\big|W^{(j)}_{n|n-1}\big|^{-1/2}\exp\left[-\frac{1}{2\sigma^{(j)2}}\big(\mathbf{y}_n - \mathbf{y}^{(j)}_{n|n-1}\big)^{\mathrm{T}}W^{(j)-1}_{n|n-1}\big(\mathbf{y}_n - \mathbf{y}^{(j)}_{n|n-1}\big)\right] \end{aligned} \tag{B2}$$

$$\mathbf{y}^{(j)}_{n|n-1} = \mathrm{E}\big(\mathbf{y}_n|\mathbf{y}_{1:n-1}, \mathcal{I}_{1:n} = \mathcal{I}^{(j)}_{1:n|N_\mathrm{e}}, \tau, \sigma^{(j)}\big) \tag{B3}$$

$$\sigma^{(j)2}W^{(j)}_{n|n-1} = \mathrm{Cov}\big(\mathbf{y}_n|\mathbf{y}_{1:n-1}, \mathcal{I}_{1:n} = \mathcal{I}^{(j)}_{1:n|N_\mathrm{e}}, \tau, \sigma^{(j)}\big). \tag{B4}$$

$\mathbf{y}^{(j)}_{n|n-1}$ and $W^{(j)}_{n|n-1}$ are calculated by

$$\mathbf{y}^{(j)}_{n|n-1} = \mathbf{H}\mathbf{x}^{(j)}_{n|n-1} \tag{B5}$$

$$W^{(j)}_{n|n-1} = \mathbf{H}V^{(j)}_{n|n-1}\mathbf{H}^{\mathrm{T}} + \tilde{\mathbf{R}}_n. \tag{B6}$$

Substituting (B2) into (B1) yields the log-likelihood for jth trajectory:

$$l^{(j)}\left(\sigma^{(j)}\right) = \log L^{(j)}\left(\sigma^{(j)}\right)$$
$$= -\frac{1}{2}\left[N_e N_d \log\left(2\pi\sigma^{(j)2}\right) + \sum_{n=1}^{N_e} \log\left|W_{n|n-1}^{(j)}\right| + \frac{1}{\sigma^{(j)2}} \sum_{n=1}^{N_e} \left(\mathbf{y}_n - \mathbf{y}_{n|n-1}^{(j)}\right)^{\mathrm{T}} W_{n|n-1}^{(j)-1}\left(\mathbf{y}_n - \mathbf{y}_{n|n-1}^{(j)}\right)\right]. \tag{B7}$$

Therefore, the maximum likelihood estimate of $\sigma^{(j)2}$ is obtained by

$$\hat{\sigma}^{(j)2} = \frac{1}{N_e N_d} \sum_{n=1}^{N_e} \left(\mathbf{y}_n - \mathbf{y}_{n|n-1}^{(j)}\right)^{\mathrm{T}} W_{n|n-1}^{(j)-1}\left(\mathbf{y}_n - \mathbf{y}_{n|n-1}^{(j)}\right). \tag{B8}$$

APPENDIX C: LIKELIHOOD OF THE MODEL

In this appendix we present a formula for the log-likelihood of the model. Let $\boldsymbol{\theta} = (\tau, \sigma)$ be a vector that contains hyperparameters. Given $\boldsymbol{\theta}$, the likelihood of the model is expressed by

$$L(\boldsymbol{\theta}) = p\left(\mathbf{y}_{1:N_e}|\boldsymbol{\theta}\right)$$
$$= \prod_{n=1}^{N_e} p(\mathbf{y}_n|\mathbf{y}_{1:n-1}, \boldsymbol{\theta}). \tag{C1}$$

If we use all N_p trajectories for the state estimation, $p(\mathbf{y}_n|\ \mathbf{y}_{1:n-1}, \boldsymbol{\theta})$ in (C1) is given by

$$p(\mathbf{y}_n|\mathbf{y}_{1:n-1}, \boldsymbol{\theta}) = \sum_{j=1}^{N_p} p\left(\mathbf{y}_n, \mathcal{I}_{1:n} = \mathcal{I}_{1:n|N_e}^{(j)}\middle|\mathbf{y}_{1:n-1}, \tau, \sigma^{(j)}\right)$$
$$= \sum_{j=1}^{N_p} p\left(\mathbf{y}_n|\mathbf{y}_{1:n-1}, \mathcal{I}_{1:n} = \mathcal{I}_{1:n|N_e}^{(j)}, \tau, \sigma^{(j)}\right) \Pr\left(\mathcal{I}_{1:n} = \mathcal{I}_{1:n|N_e}^{(j)}\middle|\mathbf{y}_{1:n-1}, \tau, \sigma^{(j)}\right)$$
$$= \frac{1}{N_p} \sum_{j=1}^{N_p} \tilde{w}_n^{(j)} \tag{C2}$$

where $\tilde{w}_n^{(j)}$ is given by (B2). In the third equality, we use the fact that all trajectories are equally weighted. Combining (C1) and (C2) yields the following formula for the log-likelihood of the model:

$$l(\boldsymbol{\theta}) = \sum_{n=1}^{N_e} \log p(\mathbf{y}_n|\mathbf{y}_{1:n-1}, \boldsymbol{\theta})$$
$$= \sum_{n=1}^{N_e} \log\left(\sum_{j=1}^{N_p} \tilde{w}_n^{(j)}\right) - N_e \log N_p. \tag{C3}$$

If we use N_p' trajectories that are randomly sampled from N_p trajectories, the log-likelihood of the model is obtained by replacing N_p in (C3) with N_p'.

The goodness of the model is evaluated by the Akaike information criterion (AIC) (Akaike 1974). The AIC is defined as

$$\mathrm{AIC} = -2l(\boldsymbol{\theta}) + 2 \times (\text{number of unknown parameters}). \tag{C4}$$