

# Data assimilation for massive autonomous systems based on second-order adjoint method

Shin-ichi Ito,<sup>1</sup> Hiromichi Nagao,<sup>1,2</sup> Akinori Yamanaka,<sup>3</sup> Yuhki  
Tsukada,<sup>4</sup> Toshiyuki Koyama,<sup>4</sup> Masayuki Kano,<sup>5</sup> and Junya Inoue<sup>6</sup>

<sup>1</sup>*Earthquake Research Institute, The University of Tokyo, Japan*

<sup>2</sup>*Graduate School of Information Science and Technology, The University of Tokyo, Japan*

<sup>3</sup>*Division of Advanced Mechanical Systems Engineering,  
Institute of Engineering, Tokyo University of Agriculture and Technology, Japan*

<sup>4</sup>*Graduate School of Engineering, Nagoya University, Japan*

<sup>5</sup>*Earthquake Research Institute, The University of Tokyo, Japan*

<sup>6</sup>*School of Engineering, The University of Tokyo, Japan*

(Dated: March 15, 2016)

## Abstract

Data assimilation (DA) is a fundamental computational technique that integrates numerical simulation models and observation data on the basis of Bayesian statistics. Originally developed for meteorology, especially weather forecasting, DA is now an accepted technique in various scientific fields. One key issue that remains controversial is the implementation of DA in massive simulation models under limited computation time and resources. In this paper, we propose an adjoint-based DA method for massive autonomous models that produces optimum estimates and their uncertainties within practical computation time and resource constraints. The uncertainties are given as several diagonal components of an inverse Hessian matrix, which is the covariance matrix of a normal distribution that approximates the target posterior probability density function in the neighborhood of the optimum. Conventional algorithms for deriving the inverse Hessian matrix require  $O(CN^2 + N^3)$  computations and  $O(N^2)$  memory, where  $N$  is the number of degrees of freedom of a given autonomous system and  $C$  is the number of computations needed to simulate time series of suitable length. The proposed method using a second-order adjoint method allows us to directly evaluate the diagonal components of the inverse Hessian matrix without computing all of its components. This drastically reduces the number of computations to  $O(C)$  and the amount of memory to  $O(N)$  for each diagonal component. The proposed method is validated through numerical tests using a massive two-dimensional Kobayashi's phase-field model. We confirm that the proposed method correctly reproduces the parameter and initial state assumed in advance, and successfully evaluates the uncertainty of the parameter. Such information regarding uncertainty is valuable, as it can be used to optimize the design of experiments.

## I. INTRODUCTION

Determining the model parameters and initial states of simulation models is an important task in various scientific fields, as it enables the temporal evolution of the target system to be observed. However, in many practical cases, this procedure is somewhat complex, because it is often impossible to observe the parameters and initial states experimentally. In materials engineering, for example, phase-field (PF) models are often used to simulate the evolution of microstructures during the processes of solidification and phase transformation [1–7]. PF models phenomenologically describe the dynamics of phases using field variables that evolve in time depending on the gradient of the total free energy. Since a PF model usually requires a huge number of grid points to discretize the field variables, the computational cost tends to be prohibitive. Nonetheless, PF models are accepted beyond the field of materials engineering, such as in hydrodynamics [8–10], as they can be employed to model phases and their dynamics using mathematical expressions that are easy to manipulate. In PF models, the parameters and initial states perfectly determine the dynamical properties of the phases. However, various limitations in practical experiments sometimes prevent such parameters and initial states from being estimated.

Data assimilation (DA) is a computational technique that integrates numerical simulation models and observational data on the basis of Bayesian statistics. Thus, DA enables the parameters and initial states of PF models to be estimated by systematically extracting as much information as possible from the given observational/experimental data. The process of DA evaluates a probability density function (PDF) (or the “posterior PDF,” to be precise) of the unknown parameters and unobservable states that is conditional on the given observation data [11]. DA was originally developed in the fields of meteorology and oceanography [12–14], but is now applied in areas such as seismology, marketing science, and industrial science [15–18]. Several sequential Bayesian filters and other non-sequential estimation methods have been used in DA. Common sequential Bayesian filters such as the ensemble Kalman filter [19–21] and particle filter [22–24] estimate the target posterior PDF using Bayes’ theorem. This approximation is formed using an ensemble of realizations, meaning that the computational cost is proportional to the number of realizations. The implementation of one sequential Bayesian filter on a given simulation model is not especially complex, and a sufficiently accurate estimate of the posterior PDF can be achieved when

the number of degrees of freedom  $N$  of the simulation model is sufficiently small. However, Bayesian filters become inefficient when applied to massive simulation models, as the number of realizations required to obtain a converged posterior PDF is proportional to  $e^{O(N)}$ . Unlike sequential Bayesian filters, adjoint methods [25–27] directly determine the optimum solution using a gradient method to maximize the target posterior PDF. Although this achieves a drastic reduction in the computational cost, the ordinary adjoint method cannot evaluate the uncertainty in its estimations, something which sequential Bayesian filters obtain in a straightforward manner. Such uncertainties provide valuable information related to both the estimations and the optimum solution. For example, the uncertainties provide feedback for the experimental design that helps to identify the parameters of interest with the required accuracy. The quantification of uncertainty is currently a very important issue in the application of DA to massive simulation models.

This paper describes an adjoint-based DA methodology that simultaneously estimates the optimum solution and its uncertainty, and is capable of being applied to simulation models with a huge number of degrees of freedom. We first construct a method to estimate the parameters and initial states involved in an autonomous system, and then validate our approach using a PF model as a testbed. Section II introduces the formulation of an adjoint-based DA method to simultaneously obtain an estimation and its uncertainty using second-order information of the posterior PDF. Section III describes the formulation of an estimation test using synthetic data generated from the time series given by Kobayashi’s PF model [1]. Section IV presents and discusses the results of estimation tests, and Section V concludes this paper by summarizing the results of this study.

## II. METHOD

### A. State-space model and cost function

DA based on Bayesian statistics always starts by defining a state-space model, which consists of a system model and an observation model. The system model describes how a state vector evolves over time in accordance with a given simulation model. The state vector contains all time-dependent variables used in the simulation model and sometimes the model parameters.

Suppose an autonomous simulation model is given by  $\partial \mathbf{z} / \partial t = \mathbf{A}(\mathbf{z}; \mathbf{a})$ , where  $\mathbf{z}(t) \in \mathbb{R}^{N_z}$  denotes a time-dependent variable and  $\mathbf{A} : \mathbb{R}^{N_z} \rightarrow \mathbb{R}^{N_z}$  is a function of  $\mathbf{z}$  and a time-invariant parameter vector  $\mathbf{a} \in \mathbb{R}^{N_a}$ . The system model describes the time evolution of a state vector consisting of  $\mathbf{z}$  and  $\mathbf{a}$ . Let  $\mathbf{X}(t) = (\mathbf{z}^\top, \mathbf{a}^\top)^\top \in \mathbb{R}^N$  be the state vector, where  $\bullet^\top$  denotes the transpose of  $\bullet$  and  $N = N_z + N_a$ . Since  $\mathbf{a}$  is time-invariant, i.e.,  $\partial \mathbf{a} / \partial t = \mathbf{0}$ , the system model can be represented by

$$\frac{\partial \mathbf{X}}{\partial t} = \mathbf{f}(\mathbf{X}), \quad (1)$$

where  $\mathbf{f} : \mathbb{R}^N \rightarrow \mathbb{R}^N$  is defined as  $f_i = A_i$  for  $1 \leq i \leq N_z$  and  $f_i = 0$  for  $N_z + 1 \leq i \leq N$ .

The observation model describes how  $\mathbf{X}(t)$  relates to a time series of observation data  $\mathbf{D}(t) \in \mathbb{R}^K$ , where  $K$  denotes the dimension of the observations. Considering that the data include noise, the observation model can be described as

$$\mathbf{D} = \check{\mathbf{h}}(\mathbf{X}) + \mathbf{\Omega}, \quad (2)$$

where  $\check{\mathbf{h}} : \mathbb{R}^N \rightarrow \mathbb{R}^K$  is an observation operator that outputs quantities from  $\mathbf{X}$  comparable with the data and  $\mathbf{\Omega}(t)$  denotes observation noise. In this paper, we assume that  $\mathbf{f}$  and  $\check{\mathbf{h}}$  are nonlinear functions, and  $\mathbf{\Omega}$  is white noise that follows a normal distribution with a diagonal covariance matrix. Our purpose is to obtain the optimum initial state  $\mathbf{X}(0)$  together with the uncertainties of the variables of interest.

In consideration of PF models, we also assume that  $\mathbf{X}(0)$  is constrained by

$$X_i^{\text{Lower}} < X_i(0) < X_i^{\text{Upper}} \quad (i = 1, \dots, N), \quad (3)$$

where  $\mathbf{X}^{\text{Lower}} \in \mathbb{R}^N$  and  $\mathbf{X}^{\text{Upper}} \in \mathbb{R}^N$  denote the lower and upper bounds of  $\mathbf{X}(0)$ , respectively.

To simplify our formulation, we normalize  $\mathbf{X}$  as

$$\theta_i(t) = \frac{X_i(t) - X_i^{\text{Lower}}}{X_i^{\text{Upper}} - X_i^{\text{Lower}}} \quad (i = 1, \dots, N). \quad (4)$$

This leads to the following  $\boldsymbol{\theta}$ -dependent forms of Eqs. (1)-(3):

$$\frac{\partial \boldsymbol{\theta}}{\partial t} = \mathbf{F}(\boldsymbol{\theta}), \quad (5)$$

$$\mathbf{D} = \mathbf{h}(\boldsymbol{\theta}) + \mathbf{\Omega}, \quad (6)$$

$$0 < \Theta_i < 1 \quad (i = 1, \dots, N), \quad (7)$$

where  $F_i(\boldsymbol{\theta}) = f_i(\mathbf{X}) / (X_i^{\text{Upper}} - X_i^{\text{Lower}})$ ,  $\mathbf{h}(\boldsymbol{\theta})$  is an observation operator after transforming  $\mathbf{X}$  to  $\boldsymbol{\theta}$ , and  $\boldsymbol{\Theta} = \boldsymbol{\theta}(0)$ .

Bayes' theorem states that a conditional PDF  $p(\boldsymbol{\Theta}|\mathbf{D})$ , which is called the posterior PDF, can be described as

$$p(\boldsymbol{\Theta}|\mathbf{D}) = \frac{p(\boldsymbol{\Theta})p(\mathbf{D}|\boldsymbol{\Theta})}{p(\mathbf{D})}, \quad (8)$$

where  $p(\boldsymbol{\Theta})$  and  $p(\mathbf{D}|\boldsymbol{\Theta})$  are called the prior PDF and likelihood, respectively. Note that  $p(\mathbf{D})$  is constant, since  $\mathbf{D}$  is a definite vector. Thus, Eq. (8) implies that  $p(\boldsymbol{\Theta}|\mathbf{D})$  is proportional to a product of the prior PDF and the likelihood.

The prior PDF contains prior information provided by experience and intuition. If we suppose that this prior information is the constraint condition given by Eq. (7) and that  $\Theta_i$  is independent for each  $i$ ,  $p(\boldsymbol{\Theta})$  is given by a product of prior PDFs of  $\Theta_i$ ,

$$p(\boldsymbol{\Theta}) = \prod_{i=1}^N p(\Theta_i), \quad (9)$$

where

$$p(\Theta_i) = \begin{cases} 1 & \text{for } 0 < \Theta_i < 1 \\ 0 & \text{otherwise.} \end{cases} \quad (10)$$

When observation data are obtained at  $t = t_1, t_2, \dots, t_n$ ,  $p(\mathbf{D}|\boldsymbol{\Theta})$  can be written as

$$p(\mathbf{D}|\boldsymbol{\Theta}) = \prod_{s=1}^n \prod_{k=1}^K p(\Omega_k(t_s)), \quad (11)$$

where

$$p(\Omega_k(t_s)) = \frac{1}{\sqrt{2\pi\sigma_k^2}} \exp \left[ -\frac{\{D_k(t_s) - h_k(\boldsymbol{\theta}(t_s))\}^2}{2\sigma_k^2} \right], \quad (12)$$

and  $\sigma_k$  is the standard deviation of  $\Omega_k$  ( $k = 1, \dots, K$ ). We consider  $\sigma_k$  to be a hyperparameter.

This paper defines the optimum solution  $\hat{\boldsymbol{\Theta}}$  to be the  $\boldsymbol{\Theta}$  that maximizes the posterior PDF  $p(\boldsymbol{\Theta}|\mathbf{D})$ . For the convenience of numerical computation, we aim to minimize a cost function

$$J = \sum_{s=1}^n \sum_{k=1}^K \left[ \frac{\log(2\pi\sigma_k^2)}{2} + \frac{\{D_k(t_s) - h_k(\boldsymbol{\theta}(t_s))\}^2}{2\sigma_k^2} \right] \text{ subject to } 0 < \Theta_i < 1, \quad (13)$$

which comes from a negative logarithmic posterior PDF, i.e.,  $p(\boldsymbol{\Theta}|\mathbf{D}) \propto e^{-J}$ , to find  $\hat{\boldsymbol{\Theta}}$ , rather than maximizing  $p(\boldsymbol{\Theta}|\mathbf{D})$ . The constraint in Eq. (13) arises from the term  $-\log p(\boldsymbol{\Theta})$ , which appears when calculating  $-\log p(\boldsymbol{\Theta}|\mathbf{D})$ .

An optimum solution  $\hat{\sigma}_k$  for  $\sigma_k$  can be determined as follows. By letting  $\partial J/\partial\sigma_k = 0$ , we obtain

$$\sigma_k = \sqrt{\frac{1}{n} \sum_{s=1}^n [D_k(t_s) - h_k(\boldsymbol{\theta}(t_s))]^2}. \quad (14)$$

Then,  $\hat{\sigma}_k$  is obtained by substituting  $\boldsymbol{\theta}$ , which is calculated by Eq. (5) using  $\boldsymbol{\theta}(0) = \hat{\boldsymbol{\Theta}}$ , into Eq. (14).

## B. Optimization via an adjoint method

Typically,  $J$  is optimized using a gradient method such as steepest gradient descent, the nonlinear conjugate gradient method, or the Limited-memory Broyden-Fletcher-Goldfarb-Shanno (LBFGS) method [28]. Gradient methods require  $\partial J/\partial\boldsymbol{\Theta}$  to update  $J$ , but it is difficult to calculate this quantity because  $J$  does not explicitly include  $\boldsymbol{\Theta}$ , as seen in Eq. (13). Generally,  $J$  is fully determined by setting  $\boldsymbol{\Theta} = \boldsymbol{\theta}(0)$  through Eq. (5), so that  $J$  must be written as a function of  $\boldsymbol{\Theta}$ , i.e.,  $J(\boldsymbol{\Theta})$ . According to Eq. (13),  $J$  is also a function of  $\boldsymbol{\theta}(t_s)$  ( $s = 1, \dots, n$ ) that satisfies Eq. (5), not including  $\boldsymbol{\Theta}$ . Summarizing these two expressions for  $J$ :

$$J(\boldsymbol{\Theta}) = \int_0^{t_f} dt \mathcal{J}(\boldsymbol{\theta}) + \int_0^{t_f} dt \boldsymbol{\lambda}^\top \left( \mathbf{F} - \frac{\partial \boldsymbol{\theta}}{\partial t} \right), \quad (15)$$

where  $t_f$  is an arbitrary time later than  $t_n$ , and  $\boldsymbol{\lambda}(t) \in \mathbb{R}^N$  denotes a vector of the Lagrange multipliers that impose Eq.(5) as the constraint condition of  $\boldsymbol{\theta}$ .  $\mathcal{J}$  is the time-dependent function

$$\mathcal{J}(\boldsymbol{\theta}) = \sum_{s=1}^n \delta(t - t_s) \sum_{k=1}^K \left[ \frac{\log(2\pi\sigma_k^2)}{2} + \frac{\{D_k(t) - h_k(\boldsymbol{\theta}(t))\}^2}{2\sigma_k^2} \right], \quad (16)$$

that satisfies  $J = \int_0^{t_f} dt \mathcal{J}$ , where  $\delta(t)$  denotes the Dirac delta function. Taking a variation of Eq. (15), we have a time evolution equation for  $\boldsymbol{\lambda}$ :

$$\frac{\partial \boldsymbol{\lambda}}{\partial t} + \left( \frac{\partial \mathbf{F}}{\partial \boldsymbol{\theta}} \right)^\top \boldsymbol{\lambda} + \frac{\partial \mathcal{J}}{\partial \boldsymbol{\theta}} = \mathbf{0}, \quad (17)$$

where

$$\boldsymbol{\lambda}(0) = \frac{\partial J}{\partial \boldsymbol{\Theta}}, \quad (18)$$

$$\boldsymbol{\lambda}(t_f) = \mathbf{0}. \quad (19)$$

Details of the derivation can be found in [26, 29]. Solving Eq. (17) backwardly in time with the condition in Eq. (19), we obtain the objective  $\partial J/\partial\boldsymbol{\Theta}$  as  $\boldsymbol{\lambda}(0)$ . Such a procedure to

obtain the gradient of the cost function using the adjoint equation (Eq. (17)) is called the adjoint method.

When we apply the adjoint method to our problem, a variable transformation is needed in the process of updating  $\Theta$  based on a gradient method, since  $\Theta$  has the constraint shown in Eq. (7). The variable transformation

$$\Psi_i = \log \Theta_i - \log (1 - \Theta_i) \quad (i = 1, \dots, N) \quad (20)$$

converts the constrained optimization problem of  $\Theta$  into an unconstrained one with respect to  $\Psi$ . The update procedure is as follows. After obtaining  $\partial J/\partial\Theta$  by the adjoint method based on Eqs. (17)-(19), we convert  $\Theta$  to  $\Psi$  using Eq. (20) and  $\partial J/\partial\Theta$  to  $\partial J/\partial\Psi$  as

$$\frac{\partial J}{\partial\Psi_i} = \Theta_i (1 - \Theta_i) \frac{\partial J}{\partial\Theta_i} \quad (i = 1, \dots, N). \quad (21)$$

Using this formulation to update  $\Psi$ , we can obtain an updated  $\Theta$  from the inverse transformation of Eq. (20):

$$\Theta_i = \frac{1}{1 + \exp(-\Psi_i)} \quad (i = 1, \dots, N). \quad (22)$$

Although this update procedure does not allow  $\Theta_i$  to be exactly 0 or 1, owing to the definition of Eq. (22),  $\Theta_i$  can be sufficiently close to 0 or 1 to pose no problem in practical cases.

The adjoint method calculates  $\partial J/\partial\Theta$  for a fixed  $\sigma_k$ , so that an optimization of  $\sigma_k$  is to be done at the same time as  $\Theta$  by substituting Eq. (14) into Eq. (16) every time  $\Theta$  is updated.

The advantages of the adjoint method over sequential Bayesian filters are that only  $O(C)$  computations and  $O(N)$  memory are required to find  $\hat{\Theta}$ , where  $C$  is the number of computations needed to run the given simulation model from  $t = 0$  to  $t = t_f$ .

### C. Evaluation of uncertainty via a second-order adjoint method

The adjoint method described in Section II B gives the optimum solution  $\hat{\Theta}$  that maximizes  $p(\Theta|\mathbf{D})$ . However, it does not provide information about the behavior of  $p(\Theta|\mathbf{D})$  in the neighborhood of  $\Theta = \hat{\Theta}$ , which reflects the uncertainty in the estimation of  $\hat{\Theta}$ . To extract such information, another procedure must be implemented on the adjoint method. Considering that  $\partial J/\partial\Theta|_{\Theta=\hat{\Theta}} = \mathbf{0}$ , the Taylor expansion of  $J$  with respect to  $\Theta - \hat{\Theta}$  is

$$J(\Theta) \sim J(\hat{\Theta}) + \frac{1}{2}(\Theta - \hat{\Theta})^\top \mathbf{H}(\Theta - \hat{\Theta}), \quad (23)$$



where terms of order higher than three have been neglected, and  $\mathbf{H}$  is a Hessian matrix given by

$$H_{i,j} = \frac{\partial^2 J}{\partial \Theta_i \partial \Theta_j} \Big|_{\Theta = \hat{\Theta}} \quad (i, j = 1, \dots, N). \quad (24)$$

We normalize  $p(\Theta|\mathbf{D}) \propto e^{-J}$  into which Eq. (23) is substituted as

$$p(\Theta|\mathbf{D}) \sim \frac{\exp \left[ -\frac{1}{2}(\Theta - \hat{\Theta})^\top \mathbf{H}(\Theta - \hat{\Theta}) \right]}{(2\pi)^{N/2} |\mathbf{H}^{-1}|^{1/2}}, \quad (25)$$

where  $\mathbf{H}^{-1}$  is the inverse of  $\mathbf{H}$  and  $|\bullet|$  denotes the determinant of  $\bullet$ . Equation (25) indicates that, in the neighborhood of  $\Theta = \hat{\Theta}$ ,  $p(\Theta|\mathbf{D})$  can be approximated by a multivariate normal distribution with mean vector  $\hat{\Theta}$  and covariance matrix  $\mathbf{H}^{-1}$ . Let  $\Theta_l$  ( $1 \leq l \leq N$ ) be a component of interest in  $\Theta$ . Integrating Eq. (25) over all variables except for  $\Theta_l$ , the marginal distribution with respect to  $\Theta_l$  is the normal distribution with mean  $\hat{\Theta}_l$  and variance  $(H^{-1})_{l,l}$ , which is the  $l$ -th diagonal element of  $\mathbf{H}^{-1}$ . This means that the uncertainty of  $\Theta_l$  is given by  $(H^{-1})_{l,l}$ . When  $N \gg 1$ , it is unrealistic to obtain  $\mathbf{H}^{-1}$  directly by numerically differentiating  $\mathbf{H}$ , which is generally dense, as this would require  $O(CN^2 + N^3)$  computations and  $O(N^2)$  memory. In practical cases, it is not necessary to evaluate all elements of  $\mathbf{H}^{-1}$ , since the number of elements of interest is usually much smaller than  $N$ . Therefore, we propose to use a second-order adjoint method [30, 31] to efficiently obtain such uncertainties in massive autonomous systems. The following procedure to obtain the uncertainties requires  $O(C)$  computations and  $O(N)$  memory for each uncertainty. When evaluating the uncertainty of  $\Theta_l$ , we consider a linear equation of  $\mathbf{r} \in \mathbb{R}^N$ :

$$\mathbf{H}\mathbf{r} = \mathbf{q}, \quad (26)$$

where  $\mathbf{q} \in \mathbb{R}^N$  is a vector with elements  $q_l = 1$  and  $q_{i \neq l} = 0$ . The solution  $\hat{\mathbf{r}}$  obviously includes  $(H^{-1})_{l,l}$  as  $\hat{r}_l = \sum_{j=1}^N (H^{-1})_{l,j} q_j = (H^{-1})_{l,l}$ . Note that  $\mathbf{H}$  is a constant matrix that requires complex computations because of its large dimension. We must obtain  $\hat{\mathbf{r}}$  from an initial guess via an iterative technique such as the conjugate gradient method or conjugate residual method. The iterative method needs, in the way of the iteration, to compute each of the Hessian-vector products  $\mathbf{H}\boldsymbol{\gamma}$  for a vector  $\boldsymbol{\gamma}$ . The second-order adjoint method enables us to compute such Hessian-vector products.

Let  $\boldsymbol{\xi}(t) \in \mathbb{R}^N$  and  $\boldsymbol{\zeta}(t) \in \mathbb{R}^N$  be perturbations of  $\hat{\boldsymbol{\theta}}$  and  $\hat{\boldsymbol{\lambda}}$ , which respectively correspond

to  $\boldsymbol{\theta}$  and  $\boldsymbol{\lambda}$  when  $\boldsymbol{\Theta} = \hat{\boldsymbol{\Theta}}$ . Their time evolutions are given by

$$\frac{\partial \boldsymbol{\xi}}{\partial t} = \left. \frac{\partial \mathbf{F}}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \boldsymbol{\xi}, \quad (27)$$

$$\frac{\partial \boldsymbol{\zeta}}{\partial t} + \left( \frac{\partial \mathbf{F}}{\partial \boldsymbol{\theta}} \right)^\top \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \boldsymbol{\zeta} + \left( \frac{\partial^2 \mathbf{F}}{\partial \boldsymbol{\theta}^2} \boldsymbol{\xi} \right)^\top \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \hat{\boldsymbol{\lambda}} + \left. \frac{\partial^2 \mathcal{J}}{\partial \boldsymbol{\theta}^2} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \boldsymbol{\xi} = \mathbf{0}. \quad (28)$$

The combination of Eqs. (27) and (28), which are called the tangent linear model and second-order adjoint model, respectively, gives the Hessian-vector product for an arbitrary vector. Solving Eq. (27) forwardly in time for a given vector  $\boldsymbol{\xi}(0) = \boldsymbol{\gamma}$ , we obtain the time series of  $\boldsymbol{\xi}$ . Then, solving Eq. (28) backwardly with the given  $\boldsymbol{\zeta}(t_f) = \mathbf{0}$  and the time series  $\boldsymbol{\xi}$ , we obtain the objective Hessian-vector product  $\boldsymbol{\zeta}(0) = \mathbf{H}\boldsymbol{\gamma}$ . The detailed derivation is given in [30].

### III. TWIN EXPERIMENT

#### A. Kobayashi's phase-field model

The accuracy of the proposed method is verified through numerical simulations termed ‘‘twin experiments,’’ details of which are given in Section III B. We choose a two-dimensional Kobayashi's PF model as the testbed in the twin experiments. Kobayashi's PF model describes the fundamental growth dynamics of two phases, such as in solidification or a phase transformation. The time evolution of one of the phases is described by

$$\tau \frac{\partial \phi}{\partial t} = \epsilon^2 \Delta \phi + \phi(1 - \phi) \left( \phi + m - \frac{1}{2} \right), \quad -\frac{1}{2} < m < \frac{1}{2} \quad (29)$$

where the PF variable  $\phi(\mathbf{x}, t)$  denotes the existence probability of the relevant phase, e.g., solid or liquid. The parameters  $\tau$  and  $\epsilon$  non-dimensionalize time and space, respectively, and  $m$  characterizes the velocity of the interface between the two phases. We assume that these parameters are time-invariant constants. We know that  $\phi(\mathbf{x}, t)$  should be constrained in  $0 \leq \phi(\mathbf{x}, t) \leq 1$ , as it describes a probability. This condition is automatically satisfied by setting the initial phase to  $0 \leq \phi(\mathbf{x}, 0) \leq 1$ , since  $\phi = 0$  and  $\phi = 1$  are the fixed points of Eq. (29).

Kobayashi's PF model underlies various PF models that describe physical phenomena such as dendrite growth [1, 5], crack propagation [32, 33], and interface-driven pattern

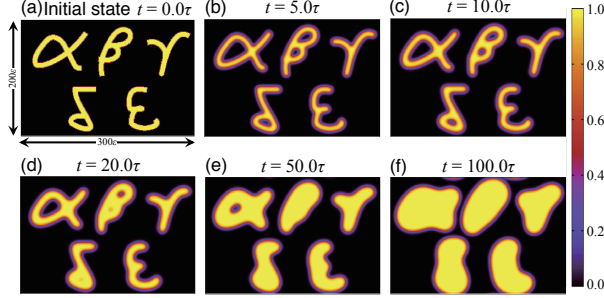


FIG. 1. Time evolution of phase field  $\phi$  starting from the initial state shown in (a) for  $m = 0.1$ . (b)-(f) show  $\phi$  at  $t = 5.0\tau$ ,  $10.0\tau$ ,  $20.0\tau$ ,  $50.0\tau$ , and  $100.0\tau$ , respectively. The color indicates the magnitude of  $\phi$ .

formation [34]. Therefore, Kobayashi's PF model is a good choice for verifying whether the proposed DA method works well, and is a first step towards future applications in more complex PF models.

## B. Synthetic data

Twin experiments are often conducted in the field of DA to verify a newly developed method on the basis of synthetic data. The synthetic data are usually generated using the given simulation model, in which the true parameters and initial state are pre-determined. Verification then proceeds by checking whether the DA method applied to the synthetic data reproduces the true parameters and initial state. In our case, the synthetic dataset is a time series of  $\phi$  that is numerically calculated by Kobayashi's PF model with a true initial state and parameter  $m$ . The synthetic data are then contaminated by observation noise that follows a normal distribution with mean zero and variance  $\sigma^2$ . The twin experiments are intended to confirm that the proposed method estimates the initial state and parameter with the associated uncertainty.

Let  $n_x$  and  $n_y$  be the numbers of grid points in the  $x$ - and  $y$ -directions, respectively,  $M$  be the total number of grid points, i.e.,  $M = n_x n_y$ , and  $h$  be the grid spacing. A periodic boundary condition is imposed on the boundary of the computational domain. Letting  $\phi_i(t)$  be the phase at the  $i$ -th grid point, Eq. (29) can be rewritten as

$$\tau \frac{\partial \phi_i}{\partial t} = \epsilon^2 \Delta_i \phi_i + \phi_i \left( 1 - \phi_i \right) \left( \phi_i + m - \frac{1}{2} \right), \quad (30)$$

where  $\Delta_i$  denotes a second-order difference operator acting on the four nearest neighbors of the  $i$ -th grid point  $\mathcal{S}_i$ , i.e.,  $\Delta_i\phi_i = \sum_{j \in \mathcal{S}_i} (\phi_j - \phi_i) / h^2$ .

Figure 1 shows the time evolution of  $\phi$  in two-dimensional space, where  $n_x = 300$ ,  $n_y = 200$ ,  $h = \epsilon$ , and the time increment in the Euler method is  $0.1\tau$ . The assumed initial state is shown in Fig. 1(a), and the true value for the parameter  $m$  is assumed to be 0.1. Figures 1(b)-(f) show snapshots indicating that the interface between the phases  $\phi = 0$  and  $\phi = 1$  migrates, expanding the area of  $\phi = 1$ . Motivated by the fact that such snapshots are sometimes obtained as observation data in practical experiments, we use snapshots such as in Figs. 1(b)-(f) with added observation noise as the synthetic data for the twin experiments. The synthetic data are given by

$$\phi_i^{\text{obs}}(t) = \phi_i(t) + \omega_i(t) \quad (i = 1, \dots, M), \quad (31)$$

where  $\omega_i(t)$  is normally-distributed observation noise with mean zero and variance  $\sigma^2$ .

When DA is applied to Kobayashi's PF model, the time evolution equation with respect to  $m$  is needed to construct a system model within the state-space model (Section II A). The time evolution equation can be written as

$$\tau \frac{\partial b}{\partial t} = 0, \quad (32)$$

where  $b = m + 1/2$ , which denotes the normalization of  $m$ , i.e.,  $0 < b < 1$ .

### C. Cost function

We consider the synthetic observation data to be the snapshots of  $\phi$  obtained from  $t = T_{\min}$  to  $t = T_{\max}$  with time interval  $\Delta T$ . Let  $\mathcal{T}$  be the set of observation times and  $n$  be the number of observations. Combining Eqs. (13) and (31),  $J$  can be rewritten as

$$J = \frac{nM}{2} \log(2\pi\sigma^2) + \frac{1}{2\sigma^2} \sum_{t_s \in \mathcal{T}} \sum_{i=1}^M (\phi_i^{\text{obs}}(t_s) - \phi_i(t_s))^2. \quad (33)$$

The values of  $\phi_i(0)$  and  $b(0)$  that minimize Eq. (33) also minimize

$$J' = \int_0^T dt \mathcal{J}', \quad (34)$$

where

$$\mathcal{J}' = \frac{1}{2} \sum_{t_s \in \mathcal{T}} \delta(t - t_s) \sum_{i=1}^M (\phi_i^{\text{obs}}(t) - \phi_i(t))^2, \quad (35)$$

since  $\sigma$  is independent of  $\phi_i(0)$  and  $b(0)$ . When the optimum  $\hat{\phi}_i(0)$  for  $\phi_i(0)$  and  $\hat{b}(0)$  for  $b(0)$  are obtained by minimizing  $J'$ , the optimum  $\sigma$  can be obtained as

$$\hat{\sigma} = \sqrt{\frac{1}{nM} \sum_{t_s \in \mathcal{T}} \sum_{i=1}^M \left( \phi_i^{\text{obs}}(t_s) - \hat{\phi}_i(t_s) \right)^2}, \quad (36)$$

where  $\hat{\phi}_i(t)$  denotes  $\phi_i(t)$  simulated using  $\hat{\phi}_i(0)$  and  $\hat{b}(0)$ .

#### D. Procedures

Prior to applying the proposed method to the PF model, the constraint for the initial state  $0 \leq \phi_i(0) \leq 1$  is to be changed to  $0 < \phi_i(0) < 1$  to satisfy the domain of the variable transformation Eq. (20). The state variables  $\boldsymbol{\theta}(t) \in \mathbb{R}^{M+1}$  and  $\boldsymbol{\Theta} \in \mathbb{R}^{M+1}$  can be defined as

$$\begin{aligned} \boldsymbol{\theta} &= (\phi_1, \dots, \phi_M, b)^\top \\ \boldsymbol{\Theta} &= (\phi_1(0), \dots, \phi_M(0), b(0))^\top. \end{aligned} \quad (37)$$

The constraint for  $\boldsymbol{\Theta}$  becomes  $0 < \Theta_i < 1$  ( $i = 1, \dots, M+1$ ). The system models of Eqs. (30) and (32) are rewritten in terms of  $\boldsymbol{\theta}$  as

$$\tau \frac{\partial \theta_i}{\partial t} = \begin{cases} \epsilon^2 \Delta_i \theta_i + \theta_i (1 - \theta_i) (\theta_i + \theta_{M+1} - 1) & \text{for } i = 1, \dots, M, \\ 0 & \text{otherwise.} \end{cases} \quad (38)$$

Substituting the right-hand side of this equation for  $\mathbf{F}$  in Eq. (17), replacing  $\mathcal{J}$  (Eq. (17)) with  $\mathcal{J}'$  (Eq. (35)), and replacing  $J$  (Eq. (18)) with  $J'$  (Eq. (34)), the adjoint method described by Eqs. (17)-(19) is rewritten as

$$-\tau \frac{\partial \lambda_i}{\partial t} = \begin{cases} \epsilon^2 \Delta_i \lambda_i + \{-3\theta_i^2 + (4 - 2\theta_{M+1})\theta_i + \theta_{M+1} - 1\} \lambda_i + \frac{\partial \mathcal{J}'}{\partial \theta_i} & \text{for } i = 1, \dots, M, \\ \sum_{j=1}^M \theta_j (1 - \theta_j) \lambda_j & \text{otherwise,} \end{cases} \quad (39)$$

$$\boldsymbol{\lambda}(0) = \frac{\partial J'}{\partial \boldsymbol{\Theta}}, \quad (40)$$

$$\boldsymbol{\lambda}(t_f) = \mathbf{0}. \quad (41)$$

We adopt the LBFGS technique [28] as the gradient method for optimizing  $\boldsymbol{\Theta}$ . Starting from an initial guess, the LBFGS method updates  $\boldsymbol{\Theta}$  by satisfying  $0 < \Theta_i < 1$  for all  $i$  owing to the variable transformation mentioned in Section II B. To tune the LBFGS method, we

set the tolerance to  $10^{-8}$  and determine the step length by Armijo's rule [35]. Once the optimum  $\hat{\Theta}$  has been obtained, the optimum standard deviation  $\hat{\sigma}$  can be estimated by Eq. (36) and the optimum  $\hat{m}$  for  $m$  is given by  $\hat{\Theta}_{M+1} - 1/2$  or  $\hat{b}(0) - 1/2$ .

One of the most remarkable features of the proposed method is its evaluation of the uncertainties. These uncertainties can provide important information that is beneficial to updating the experimental design. In accordance with the procedure mentioned in Section II C, we consider a linear equation  $\mathbf{H}'\mathbf{r} = \mathbf{q}$ , where  $\mathbf{H}' = \partial^2 J' / \partial \Theta^2 |_{\Theta = \hat{\Theta}}$  is a Hessian matrix,  $\mathbf{r} \in \mathbb{R}^{M+1}$  is a vector to be determined, and  $\mathbf{q} \in \mathbb{R}^{M+1}$  is a vector containing the elements  $q_{M+1} = 1$  and  $q_{i \neq M+1} = 0$ . The uncertainty  $\delta \hat{m}$  can be computed from the solution  $\hat{\mathbf{r}}$  as

$$\delta \hat{m} = \hat{\sigma} \sqrt{\hat{r}_{M+1}}. \quad (42)$$

The conjugate residual method, in which the tolerance is set to  $10^{-8}$ , is adopted to solve the linear equation. The second-order adjoint method computes each of the Hessian-vector products  $\mathbf{H}'\boldsymbol{\gamma}$  that appears in the optimization process of the conjugate residual method. Substituting the right-hand side of Eq. (38) for  $\mathbf{F}$  in Eq. (27), the tangent linear model can be rewritten as

$$\tau \frac{\partial \xi_i}{\partial t} = \begin{cases} \epsilon^2 \Delta_i \xi_i + \hat{\theta}_i \left(1 - \hat{\theta}_i\right) \xi_{M+1} + \left\{-3\hat{\theta}_i^2 + \left(4 - 2\hat{\theta}_{M+1}\right) \hat{\theta}_i + \hat{\theta}_{M+1} - 1\right\} \xi_i \\ \text{for } i = 1, \dots, M, \\ 0 \text{ otherwise,} \end{cases} \quad (43)$$

with the initial condition  $\boldsymbol{\xi}(0) = \boldsymbol{\gamma}$ , where  $\hat{\boldsymbol{\theta}}$  denotes the state vector corresponding to  $\hat{\Theta}$ . Substituting the right-hand side of Eq. (38) for  $\mathbf{F}$  in Eq. (28) and replacing  $\mathcal{J}$  (Eq. (28)) with  $\mathcal{J}'$  (Eq. (35)), the second-order adjoint model in Eq. (28) becomes

$$-\tau \frac{\partial \zeta_i}{\partial t} = \begin{cases} \epsilon^2 \Delta_i \zeta_i \\ + \left\{-3\hat{\theta}_i^2 + \left(4 - 2\hat{\theta}_{M+1}\right) \hat{\theta}_i + \hat{\theta}_{M+1} - 1\right\} \zeta_i \\ - \left(6\hat{\theta}_i + 2\hat{\theta}_{M+1} - 4\right) \hat{\lambda}_i \xi_i - \left(2\hat{\theta}_i - 1\right) \hat{\lambda}_i \xi_{M+1} \\ + \sum_{j=1}^M \frac{\partial^2 \mathcal{J}'}{\partial \theta_i \partial \theta_j} \Big|_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}} \xi_j \text{ for } i = 1, \dots, M, \\ \sum_{j=1}^M \left[\hat{\theta}_j \left(1 - \hat{\theta}_j\right) \zeta_j - \left(2\hat{\theta}_j - 1\right) \hat{\lambda}_j \xi_j\right] \text{ otherwise,} \end{cases} \quad (44)$$

where  $\hat{\lambda}$  is the perturbation of  $\hat{\theta}$ . Solving Eq. (44) with the condition  $\zeta(t_f) = \mathbf{0}$ , we obtain the objective Hessian-vector product as  $\zeta(0) = \mathbf{H}'\gamma$ .

## IV. RESULTS AND DISCUSSION

The proposed method is verified through three twin experiments: (I) estimation of the parameter  $m$  conditional on the true initial state  $\phi_i^{\text{true}}(0)$  (Section IV A), (II) simultaneous estimation of the parameter  $m$  and the initial state  $\phi_i(0)$  (Section IV B), and (III) estimation of the initial state  $\phi_i(0)$  conditional on the true parameter  $m^{\text{true}}$  (Section IV C). Twin experiment I investigates how the estimation depends on observation data, twin experiment II verifies whether the proposed method outputs correct estimations, even for massive simulation models, and twin experiment III validates the unknown phenomena that appear in the results of experiment II.

### A. Twin experiment I: Parameter estimation

Twin experiment I investigates the influences of three parameters related to the observation data: (i) the length of the observation time  $T_{\text{max}}$ , (ii) the time interval of the observations  $\Delta T$ , and (iii) the standard deviation of the observation noise  $\sigma$ . Table I summarizes the parameter values used in the experiments. The first observation is assumed to occur at  $T_{\text{min}} = 0.1\tau$ , and the true value of  $m$  is assumed to be  $m^{\text{true}} = 0.1$ . The initial guess used in the LBFGS method is  $m = -0.1$ . In this experiment, the true phase field at  $t = 0$ , i.e.,  $\phi_i^{\text{true}}(0)$  (see Fig. 1(a)) is given as the initial state. The results reported here are the average values of  $\hat{m}$  and  $\delta\hat{m}$  over twenty trials with different random seeds.

Figure 2 shows the results of Test I-(i). Figure 2 (a) indicates that the parameter estimation is successful, because the true parameter is included in the range  $\hat{m} - \delta\hat{m} < m^{\text{true}} < \hat{m} + \delta\hat{m}$ . The estimation of the uncertainty  $\delta\hat{m}$  fails when  $T_{\text{max}}$  is at its minimum, i.e.,  $T_{\text{max}} = 0.2\tau$ , because insufficient data cause  $\hat{r}_{M+1}$  to become negative. Figure 2(b) indicates that  $\delta\hat{m}$  is proportional to  $T_{\text{max}}^{-1.5}$  in this range. The reason that the decrease is more rapid than the law of large numbers would suggest is related to the nonlinearity of Kobayashi's PF model. A theoretical evaluation actually indicates that  $\delta\hat{m}$  is proportional to  $T_{\text{max}}^{-2.5}$  when  $T_{\text{max}} \ll \tau$ , and will converge with a constant value as  $T_{\text{max}}$  increases. This is because

	$T_{\max}$	$\Delta T$	$\sigma$
Test I-(i)	$0.2\tau - 102.4\tau$	$0.1\tau$	0.01
Test I-(ii)	$102.5\tau$	$0.1\tau - 51.2\tau$	0.01
Test I-(iii)	$102.4\tau$	$0.1\tau$	$10^{-5} - 1.0$

TABLE I. Length of the observation time  $T_{\max}$ , time interval of observations  $\Delta T$ , and standard deviation of observation noise  $\sigma$  used in twin experiment I, where  $\tau$  is the unit of time in the simulation. Test I-(i), (ii) and (iii) investigate how the estimation depends on  $T_{\max}$ ,  $\Delta T$  and  $\sigma$ , respectively.

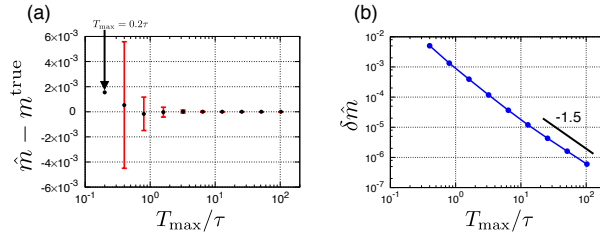


FIG. 2. Results of twin experiment I-(i). The length of each error bar for the optimum parameter  $\hat{m}$  in (a) corresponds to the estimated uncertainty  $\delta \hat{m}$  in (b). The uncertainty cannot be determined when  $T_{\max} = 0.2\tau$ . The black solid line in (b) indicates a power function of order  $-1.5$ .

no additional information is included in the observation data after  $\phi(\mathbf{x}, t)$  becomes almost uniform across the entire domain. Figure 3(a) indicates that the error bars for the optimum parameter  $\hat{m}$  increase as  $\Delta T$  increases, and Fig. 3(b) shows that the estimated uncertainty  $\delta \hat{m}$  increases with  $\Delta T$ , following a power law of order 0.5.

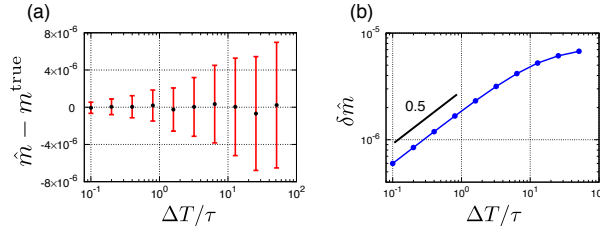


FIG. 3. Results of twin experiment I-(ii). The length of each error bar for optimum parameter  $\hat{m}$  in (a) corresponds to the estimated uncertainty  $\delta \hat{m}$  in (b). The black solid line in (b) indicates a power function of order of 0.5.



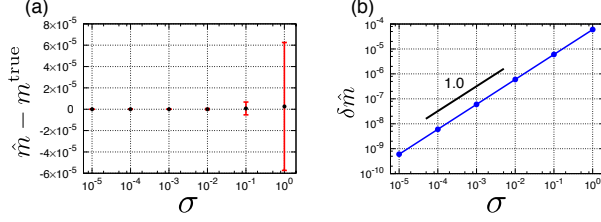


FIG. 4. Results of twin experiment I-(iii). The length of each error bar for optimum parameter  $\hat{m}$  in (a) corresponds to the estimated uncertainty  $\delta\hat{m}$  in (b). The black solid line in (b) indicates a linear function.

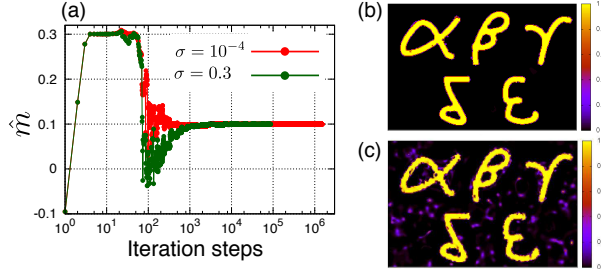


FIG. 5. Results of twin experiment II. (a) How the LBFGS method updates  $m$  when the observation data noise has a small (red) and large (green) standard deviation. (b) The optimum initial state of the phase field  $\phi_i(0)$  in the case of small noise, and (c) that in the case of large noise.

the parameter estimation is again successful, and Fig. 4(b) shows that  $\delta\hat{m}$  is proportional to  $\sigma$ . In summary, the results of Test I demonstrate that the proposed method is capable of estimating the true parameter and the associated uncertainty.

## B. Twin experiment II: Simultaneous estimation

Twin experiment II investigates the influence of the observation noise in two cases: (i) when the noise has a small standard deviation ( $\sigma = 10^{-4}$ ) and (ii) when the noise has a large standard deviation ( $\sigma = 0.3$ ). The true parameter is assumed to be  $m^{\text{true}} = 0.1$ , and the true initial state  $\phi_i^{\text{true}}(0)$  is assumed to be the phase field shown in Fig. 1(a). The other observational conditions are  $T_{\min} = 5.0\tau$ ,  $T_{\max} = 30.0\tau$ ,  $\Delta T = 0.1\tau$ , and the initial guesses are  $\phi_i(0) = 0.2$  and  $m = -0.2$ .

Figure 5 shows the results of Test II. Figure 5(a) indicates how each iteration of the LBFGS method updates the estimation of  $m$ . It is clear that each estimation converges

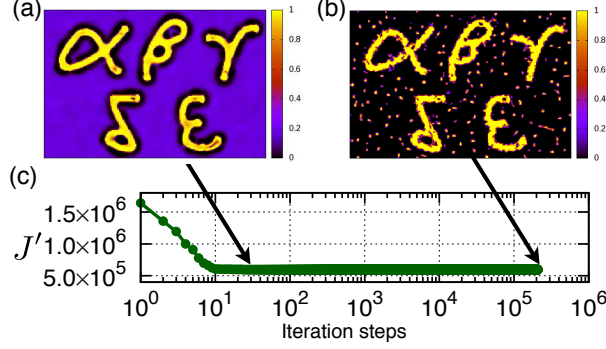


FIG. 6. Results of twin experiment III. Estimated initial states of the phase field  $\phi_i(0)$  (a) after the 31st step and (b) after the final step in the iteration of the LBFGS method. (c) Improvement in the cost function  $J'$ .

with  $m^{\text{true}}$ . Figures 5(b) and (c) indicate the estimated initial states  $\hat{\phi}_i(0)$  in Test II-(i) and (ii), respectively. These results appear to be almost consistent with the true initial states, although “spot-like” pattern appears in Fig. 5(c). This spot-like pattern would be conspicuous if the observation noise was large or if the time of the first observation  $T_{\min}$  was far from  $t = 0$ . Additionally, the spot-like pattern does not disappear under lower tolerance levels.

### C. Twin experiment III: Estimation of initial state

Twin experiment III confirms whether the estimation of  $m$  affects the generation of the spot-like pattern found in twin experiment II-(ii). Therefore, twin experiment III is set up to estimate only the initial state  $\phi_i(0)$  with a fixed parameter  $m = 0.1$ . The true initial state  $\phi_i^{\text{true}}(0)$  is assumed to be the phase field shown in Fig. 1(a). The other observational conditions are  $T_{\min} = 8.0\tau$ ,  $T_{\max} = 30.0\tau$ ,  $\Delta T = 0.1\tau$  and  $\sigma = 0.3$ , and the initial guess is  $\phi_i(0) = 0.2$ .

Figures 6(a) and (b) show the estimated initial states after the 31st and after the final iterations, respectively, and Fig. 6(c) shows how the cost function  $J'$  varies with the iteration. A spot-like pattern again appears in the estimated initial state (Fig. 6(b)) as the number of iterations increases. Note that the cost function  $J'$  is almost the same after the 31st step and after the final step, although Figs. 6(a) and (b) are much different.

This is caused by a feature inherent in the two-dimensional Kobayashi’s PF model. When

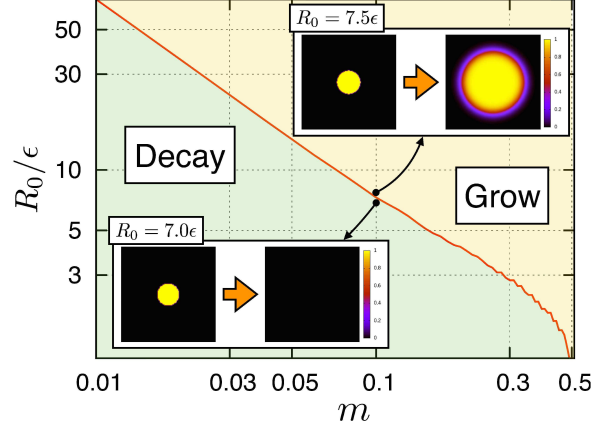


FIG. 7. Phase diagram for an axisymmetric two-dimensional Kobayashi's PF model. The red line indicates the critical radius as a function of the parameter  $m$ . The region above or below the critical line corresponds to a spot growing or decaying with time, respectively.

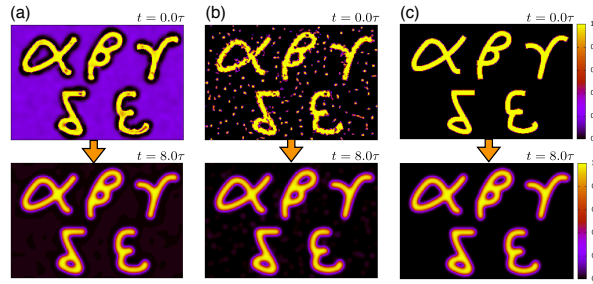


FIG. 8. Time evolutions of the phase fields starting from different initial states  $\phi_i(0)$ : (a) the estimated initial state obtained after the 31st step (Fig. 6(a)), (b) that after the final step (Fig. 6(b)), and (c) the true initial state (Fig. 1(a)).

a spot of radius  $R_0$  evolves with time based on the PF model, whether it grows or decays depends on the relation between  $m$  and  $R_0$ . Figure 7 shows the phase diagram obtained by the two-dimensional Kobayashi's PF model under the assumption of the axial symmetry. The destiny of a given spot depends on whether the radius is above or below the critical line, which means the critical radius is approximately inversely proportional to  $m$  [36]. The radius of each spot in Fig. 6(b) is actually smaller than the critical radius, which is approximately  $7.3\epsilon$  in the case of  $m = 0.1$ .

Time evolutions starting from three different initial states shown in Fig. 6. These results indicate that the phase fields at the time of the first observation, i.e.,  $t = 8.0\tau$ , are completely coincident (Fig. 8).

## V. CONCLUSIONS

This paper has described an adjoint-based DA method for massive autonomous models that not only determines the optimum estimates but also gives their uncertainties within a practical computation time and reasonable resource requirements. The uncertainties can be obtained as several diagonal components of the inverse Hessian matrix, which is the covariance matrix of the normal distribution that approximates the posterior PDF in the neighborhood of the optimum estimates. This proposed approach provides a new methodology for evaluating uncertainties using a second-order adjoint method that obtains Hessian-vector products in the process of the computation. Twin experiments using a two-dimensional Kobayashi PF model demonstrated the validity of the proposed method.

The uncertainties associated with physical quantities of interest depend on the quality and amount of data. Thus, conducting twin experiments prior to practical experiments allows us to determine how many observations are required to obtain the physical quantities of interest to the desired accuracy. Such feedback to practical experiments is already possible in systems with only a few degrees of freedom, but the proposed method makes this possible for massive simulation models.

The proposed method is not only applicable to PF models, but to various models described by autonomous systems, e.g., shallow water equations, Navier equations for elastic materials, and Boltzmann equations. The proposed method is of great utility for evaluating the uncertainties of model parameters through DA, which is important in various fields of science, even when using massive models.

## ACKNOWLEDGMENTS

This study is supported by the Cross-ministerial Strategic Innovation Promotion Program (SIP). The authors are grateful to Prof. Munekazu Ohno, Prof. Peter XK Song and Dr. Jonggyu Baek for useful discussions.

---

[1] R. Kobayashi, *Physica D* **63**, 410 (1993).

- [2] W. J. Boettinger, J. A. Warren, C. Beckermann, and A. Karma, [Annu. Rev. Mater. Res. \*\*32\*\*, 163 \(2002\)](#).
- [3] L.-Q. Chen, [Annu. Rev. Mater. Res. \*\*32\*\*, 113 \(2002\)](#).
- [4] Y. Tsukada, Y. Murata, T. Koyama, and M. Morinaga, [Mater. Trans. \*\*49\*\*, 484 \(2008\)](#).
- [5] T. Shimokawabe, T. Aoki, T. Takaki, T. Endo, A. Yamanaka, N. Maruyama, A. Nukada, and S. Matsuoka, in [Proceedings of 2011 International Conference for High Performance Computing, Networking, Storage and Analysis](#), SC '11 (ACM, New York, NY, USA, 2011) pp. 3:1–3:11.
- [6] M. Ohno, T. Yamaguchi, D. Sato, and K. Matsuura, [Comp. Mater. Sci. \*\*69\*\*, 7 \(2013\)](#).
- [7] T. Takaki, [ISIJ Int. \*\*54\*\*, 437 \(2014\)](#).
- [8] D. Jacqmin, [J. Comput. Phys. \*\*155\*\*, 96 \(1999\)](#).
- [9] M. De Menech, [Phys. Rev. E \*\*73\*\*, 031505 \(2006\)](#).
- [10] H. G. Lee, K. Kim, and J. Kim, [Int. J. Numer. Meth. Eng. \*\*85\*\*, 1633 \(2011\)](#).
- [11] S. Reich and C. Cotter, [Probabilistic Forecasting and Bayesian Data Assimilation](#) (Cambridge University Press, 2015).
- [12] E. Kalnay, [Atmospheric Modeling, Data Assimilation and Predictability](#) (Cambridge University Press, 2003).
- [13] T. Tuyuki and T. Miyoshi, [J. Meteorol. Soc. Jpn. Ser. II \*\*85B\*\*, 331 \(2007\)](#).
- [14] M. Ghil and P. Malanotte-Rizzoli, [Advances in Geophysics](#), Vol. 33 (Elsevier, 1991) pp. 141 – 266.
- [15] M. Kano, S. Miyazaki, Y. Ishikawa, Y. Hiyoshi, K. Ito, and K. Hirahara, [Geophys. J. Int. \*\*203\*\*, 646 \(2015\)](#).
- [16] T. Maeda, K. Obara, M. Shinohara, T. Kanazawa, and K. Uehira, [Geophys. Res. Lett. \*\*42\*\*, 7923 \(2015\)](#).
- [17] E. Motohashi, N. Isozaki, H. Nagao, and T. Higuchi, [J. Oper. Res. Soc. Jpn. \*\*57\*\*, 574 \(2012\)](#).
- [18] K. Sasaki, A. Yamanaka, S.-i. Ito, and H. Nagao (unpublished)
- [19] G. Evensen, [Ocean Dynam. \*\*53\*\*, 343 \(2003\)](#).
- [20] P. L. Houtekamer and H. L. Mitchell, [Mon. Weather Rev. \*\*126\*\*, 796 \(1998\)](#).
- [21] G. Ueno, T. Higuchi, T. Kagimoto, and N. Hirose, [SOLA \*\*3\*\*, 5 \(2007\)](#).
- [22] G. Kitagawa, [Introduction to Time Series Modeling](#), Chapman & Hall/CRC Monographs on Statistics & Applied Probability (CRC Press, 2010).

- [23] A. Doucet, S. Godsill, and C. Andrieu, *Stat. Comput.* **10**, 197 (2000).
- [24] H. Nagao, T. Higuchi, S. Miura, and D. Inazu, *Comput. J.* **56**, 355 (2013).
- [25] J. M. Lewis and J. C. Derber, *Tellus A* **37A**, 309 (1985).
- [26] F.-X. Le Dimet and O. Talagrand, *Tellus A* **38A**, 97 (1986).
- [27] M. Iri and K. Kubota, *JSIAM* **1**, 17 (1991).
- [28] J. Nocedal, *Math. Comput.* **35**, 773 (1980).
- [29] Z. Wang, I. Navon, F. Le Dimet, and X. Zou, *Meteorol. Atmos. Phys.* **50**, 3 (1992).
- [30] Z. Wang, K. Droegemeier, and L. White, *Comput. Optim. Appl.* **10**, 283 (1998).
- [31] F.-X. Le Dimet, I. Navon, and D. N. Daescu, *Mon. Weather Rev.* **130**, 629 (2002).
- [32] I. S. Aranson, V. A. Kalatsky, and V. M. Vinokur, *Phys. Rev. Lett.* **85**, 118 (2000).
- [33] R. Spatschek, M. Hartmann, E. Brener, H. Müller-Krumbhaar, and K. Kassner, *Phys. Rev. Lett.* **96**, 015502 (2006).
- [34] S. Komura and Y. Yamazaki, *J. Phys. Soc. Jpn.* **76**, 083801 (2007).
- [35] L. Armijo, *Pacific J. Math.* **16**, 1 (1966).
- [36] M. Castro, *Phys. Rev. B* **67**, 035412 (2003).