FOCUS ISSUE ARTICLE

Grain growth prediction based on data assimilation by implementing 4DVar on multi-phase-field model

Shin-ichi Ito\textsuperscript{a}, Hiromichi Nagao\textsuperscript{a,b}, Tadashi Kasuya\textsuperscript{c} and Junya Inoue\textsuperscript{c,d}

\textsuperscript{a}Earthquake Research Institute, The University of Tokyo, Tokyo, Japan; \textsuperscript{b}Graduate School of Information Science and Technology, The University of Tokyo, Tokyo, Japan; \textsuperscript{c}Graduate School of Engineering, The University of Tokyo, Tokyo, Japan; \textsuperscript{d}Research Center for Advanced Science and Technology, The University of Tokyo, Tokyo, Japan

ABSTRACT

We propose a method to predict grain growth based on data assimilation by using a four-dimensional variational method (4DVar). When implemented on a multi-phase-field model, the proposed method allows us to calculate the predicted grain structures and uncertainties in them that depend on the quality and quantity of the observational data. We confirm through numerical tests involving synthetic data that the proposed method correctly reproduces the true phase-field assumed in advance. Furthermore, it successfully quantifies uncertainties in the predicted grain structures, where such uncertainty quantifications provide valuable information to optimize the experimental design.

ARTICLE HISTORY

Received 31 May 2017
Revised 9 September 2017
Accepted 9 September 2017

KEYWORDS

Grain growth; data assimilation; Bayesian statistics; prediction method; phase field model; uncertainty quantification

CLASSIFICATION

60 New topics / Others; 404 Materials informatics / Genomics; 403 CALPHAD / Phase field methods; 400 Modeling / Simulations; 500 Characterization

1. Introduction

Controlling the temporal evolution of grain structures in metals is a crucial issue in obtaining materials with desirable mechanical properties \[1–3\]. To control temporal evolution, predicting the dynamics of grain growth is necessary by considering the minimization of interfacial energy among grains with different crystallographic orientations. Phase-field (PF) models \[4–10\] have often been used to simulate such dynamics in materials science. Since PF models allow us to simulate the complex morphological dynamics of grain structures by modeling interactions among grains, simulations using these models help us understand and predict their dynamics with the goal of the efficient development of materials. However, PF models need many phenomenological parameters that are not directly observable in experiments. Thus, establishing a methodology to estimate these parameters from limited observational/experimental data is necessary to predict the dynamics of grain structures. We emphasize here that an evaluation of the uncertainties in the estimated parameters also provides valuable information, since the uncertainties affect the results of the simulated grain structures and, hence, the mechanical properties of metals.

Data assimilation (DA), which integrates simulation models and observational/experimental data on the basis of Bayesian statistics, is a computational technique to estimate unobserved states and/or parameters involved in simulation models \[11\]. It has been applied to many fields of science \[12–19\]. In general, DA can be classified into online and offline DAs. Online DA sequentially estimates states referring to continuously entered observational data, where a sequential Bayesian filter, such as a particle filter \[20,21\] or the ensemble Kalman filter \[22,23\], improves the simulation in accordance with the data. A naïve implementation of online DA is often impractical because of the massive amount of computations needed, proportional to $e^O(N)$, where $N$...
is the dimensionality of the variables. Therefore, online DA does not seem appropriate for most PF models, where $N$ easily increases when a fine grid spacing is adopted for the accurate description of the boundaries of different grains.

On the contrary, an offline DA searches the optimum initial states that best match the observed time series. The representative method for offline DA is a four-dimensional variational method (4DVar) [24,25], which is the main subject of this paper. The number of computations needed to obtain optimum initial states does not depend on $N$ such that 4DVar is applicable to even massive simulation models, such as PF models. Although the conventional 4DVar has a serious disadvantage in that it only provides estimates of initial states but cannot evaluate their uncertainties, Ito et al. [19] solved this problem by establishing a new 4DVar that enables us to obtain not only the optimum, but also the associated uncertainty using a second-order adjoint (SOA) method [26–28]. This new 4DVar is appropriate for DA based on PF models because it can quantify several uncertainties of interest within a practical number of computations. The authors showed the validity of the new 4DVar by applying it to the most fundamental PF model, the Kobayashi PF model [4], which describes the growth of a single grain driven by a chemical reaction between it and its parent metal. However, the validation is inadequate from the perspective of application to more complex PF models because the dynamics of grain growth are mainly driven by interfacial energy, and the chemical reaction is a secondary factor.

In this paper, we investigate whether the new 4DVar works well in the case of a PF model that describes interface-driven dynamics. Moreover, we propose a DA procedure that allows us to investigate the extent to which uncertainties in the estimated parameters influence the prediction of the temporal evolution of grain structures. The proposed DA procedure based on the new 4DVar is expected to provide valuable information concerning the optimization of experimental design, such as regarding the number of experiments or the observation intervals. The remainder of this paper is organized as follows: Section 2 explains the experimental setup used in this paper and the PF model to describe interface-driven grain growth. Section 3 describes the procedure to predict the temporal evolution of the grain structure. Section 4 validates the proposed procedure through numerical tests and Section 5 contains a discussion of the results. Section 6 states the conclusions and directions for future work in the area.

2. Observational data and simulation model

2.1. Observational data

When the grain structure of a metal is evolved in an experiment, a sample of the metal is placed in isother-
where $\Delta$ denotes a Laplacian operator, and $\epsilon$ and $\gamma$ are the parameters to be determined. The parameter $\epsilon$ characterizes the thickness of the grain boundary, where the thickness is proportional to $\epsilon$. Considering that several snapshots of the grain structure can be obtained, $\epsilon$ is a measurable parameter in this model. Thus, $\epsilon$ does not need to be estimated, and we use it as a unit of spacing. On the other hand, the parameter $\gamma$ is an unobservable parameter in this model because we cannot estimate $\gamma$ only by observing the snapshots. Thus, establishing a methodology to estimate $\gamma$ from limited amounts of observable data is needed, and has been an important task in the field of material science.

In order to compute the time evolution of $\{\phi_i(t, x, t)\}$, we discretize Equation (1) in rectangular two-dimensional (2D) space. Let $N_x$ and $N_y$ be the numbers of grid points along the $x$- and $y$-directions, respectively, $N_{tot}$ be the total number of grid points, i.e. $N_{tot} = N_x N_y$, and $h$ be the grid spacing. A periodic boundary condition is imposed on the boundary of the computational domain. Letting $\phi_{i,m}^t$ be the PF variable of grain $i$ at the $m$th grid point and time $t$, Equation (1) can be rewritten as

$$\frac{\partial \phi_{i,m}^t}{\partial t} = \gamma \left( \epsilon^2 \Delta_{mn} \phi_{i,m}^t + 2 (\phi_{i,m}^t)^2 - 2 \phi_{i,m}^t \phi_{i,m}^t \phi_{i,m}^t \right).$$

where the Einstein summation convention has been used for repeated subscripts not appearing on the left-hand side, and $\Delta_{mn}$ denotes a discrete Laplacian in the regular 2D grid. Figures 2(a)–2(d) show the time evolution obtained by solving Equation (2) discretized in time by the Euler scheme with time step 0.1$\tau$, where $\tau$ is the unit of time to be determined by an interval of observations. In this simulation, we used $N_x = N_y = 64$, $n = 10$, $h = \epsilon$, and $\gamma = 1/\tau$. The initial state (Figure 2(a)) was set to be a Voronoi diagram generated from $n$ randomly distributed parent points. The color in Figures 2(a)–2(d) indicates the magnitude of

$$\rho = \sum_{i=1}^{n} \phi_i^2.$$ 

We find that the grain structure evolved to reduce the curvature of the grain boundaries, and larger grains grew to absorb smaller grains.

For the sake of measuring the grain structure quantitatively, we introduce the average grain size $\langle S \rangle$ defined as

$$\langle S \rangle = \frac{\sum_{i=1}^{n} S_i^2}{\sum_{i=1}^{n} S_i},$$

where we define the areal size $S_i$ of grain $i$ as

$$S_i = \frac{1}{N_{tot}} \sum_{m=1}^{N_{tot}} \phi_{i,m}^2.$$ 

Figure 2(e) shows the time evolution of average grain size $\langle S \rangle$, which grows monotonically in accordance with the growth of the grain structure, so that $\langle S \rangle$ is useful for quantitatively evaluating growth.

3. Methodology

We propose a novel methodology for evaluating the grain structure at $t = t_0$ by using observational data based on Bayesian statistics. Since the methodology is based on the DA technique reported in [19], we explain how to apply the technique to the MPF model prior to explaining the methodology. The DA technique allows us to estimate simultaneously the grain structure and the parameter $\gamma$ at an arbitrary time $t_0$, earlier than the time of the first observation $t_1$. Furthermore, the technique, which is based on 4DVar, also allows us to obtain the uncertainty in the parameter $\gamma$. The 4DVar is composed of three procedures: (i) defining a ‘state-space model’ (SSM) and an ‘observation model’, (ii) constructing a probability density function (PDF) called a ‘posterior PDF’ based on Bayesian statistics, and (iii) optimization of the posterior PDF using an ‘adjoint model’. These procedures are explained in Sections 3.1–3.3. Section 3.4 describes the technique for quantifying the uncertainty in the parameter $\gamma$. Finally, we explain our new methodology in Section 3.5.

3.1. The state-space model and the observation model

When implementing 4DVar, we must define an SSM, which is the time evolution of a large vector called the ‘state vector’ that contains variables involved in a given simulation model and, sometimes, the model parameters. The state vector used in this study is given
by a time-dependent vector \( \theta^t \in \mathbb{R}^N \) with elements

\[
\theta^t_{(i,m)} = \phi_{i,m} \quad (m = 1, \ldots, N_{\text{tot}} \text{ and } i = 1, \ldots, n),
\]

where we use the brief notation \((i, m) = m + (i - 1)N_{\text{tot}}\) and

\[
\theta^t_N = \gamma.
\]

\( N \) is the total number of elements in the state vector, i.e. \( N = nN_{\text{tot}} + 1 \). Equation (2) written in terms of \( \theta \) is given by

\[
\frac{\partial \theta^t_{(i,m)}}{\partial t} = \theta_N \left( \epsilon^2 \Delta_{nn'} \theta^t_{(i,m')} + 2\theta^t_{(i,m)} - 2\theta^t_{(i,m')}\theta^t_{(j,m)} \right),
\]

where we abbreviate the superscripts that describe time \( t \) for the sake of simplicity, except when the time must be explicitly specified. Moreover, we can write the time evolution of \( \gamma \) as

\[
\frac{\partial \theta^t_N}{\partial t} = 0,
\]

since the parameter \( \gamma \) is not a time-dependent variable. This technique embeds the parameters involved in the simulation model into a large vector together with the time-dependent variables of the model, and allows us to estimate the parameters and the initial states of the variables simultaneously. The set of Equations (8) and (9) is the SSM considered in this study. Since the SSM indicates that the state vector \( \theta \) evolves autonomously, we are allowed to use a simple notation for the right-hand side, i.e.

\[
\frac{\partial \theta}{\partial t} = F(\theta),
\]

where \( \theta = F : \mathbb{R}^N \rightarrow \mathbb{R}^N \) is a function of \( \theta \). Note that the trajectory of \( \theta^t \) is determined only by controlling the initial state \( \theta^0 = \Theta \), so that our purpose here is to search for the optimum \( \hat{\Theta} \) for \( \Theta \) that best explains the time series of observational data.

In the construction of the methodology for DA, we need an observation model that describes the relation between observational data and theoretical values comparable with them. Supposing that we can detect the spatial distributions of PF variables from snapshots and that the data include noise, we write the relation between PF variables and observational data \( D^t \) as

\[
D^t_{(i,m)} = \theta^t_{(i,m)} + \omega^t (m = 1, \ldots, N_{\text{tot}} \text{ and } i = 1, \ldots, n),
\]

where \( \omega^t \) is the observation noise that occurs when keeping the metal in an isothermal environment. We assume that the observation noise is independent and identically distributed, and follows a normal distribution with mean zero and variance \( \sigma^2 \).

### 3.2. Bayes’ theorem

In order to search for the optimum \( \hat{\Theta} \), the 4DVar measures and maximizes the posterior PDF, which is the conditional PDF of \( \Theta \) with a given \( D \). Bayes’ theorem states that the posterior PDF is given by

\[
p(\Theta | D) \propto p(\Theta)p(D | \Theta),
\]

where \( p(\Theta) \) and \( p(D | \Theta) \) are called the prior PDF and the likelihood function, respectively.

The prior PDF describes prior information provided by experience and intuition. This paper assumes that prior information describes a constraint on \( \Theta \) given by

\[
\Theta^t_L \leq \Theta_i \leq \Theta^t_U \quad (I = 1, \ldots, N),
\]

where \( \Theta^t_L \) and \( \Theta^t_U \) are the lower and upper bounds of \( \Theta \), respectively. These bounds are given by experience, and some of them can be determined by conditions for the MPF model to be physically valid. Since PF variables indicate the existence probabilities of grains, we set the lower and upper bounds of these variables to \( \Theta^t_I = 0 \) and \( \Theta^t_U = 1 \) \((I = 1, \ldots, N - 1)\), respectively. Furthermore, we set the lower bound \( \Theta^t_N \) to zero because \( \gamma \) is a positive parameter. On the other hand, we cannot determine the upper bound \( \Theta^t_S \) on the basis of conditions for the MPF model. In this paper, we set it as a large constant \( \gamma_{\text{max}} \). In computations to obtain estimates, it has been confirmed that the selection of \( \gamma_{\text{max}} \) does not influence the estimation results. Assuming that the prior PDF \( p(\Theta_I) \) for \( \Theta_I \) is uniform within the range given by Equation (13), we describe the joint prior PDF \( p(\Theta) \) as

\[
p(\Theta) = \prod_{I=1}^{N} p(\Theta_I),
\]

where

\[
p(\Theta_I) = \begin{cases} \frac{1}{\Theta^t_U - \Theta^t_L} & (\Theta^t_L \leq \Theta_I \leq \Theta^t_U) \\ 0 & (\text{otherwise}) \end{cases}.
\]

The likelihood function quantifies the difference between observational data and the corresponding theoretical values. The relation defined by Equation (11) allows us to describe the likelihood function as

\[
p(D | \Theta) = \prod_{k=1}^{K} \prod_{s=1}^{N_D} \frac{1}{\sqrt{2\pi \sigma^2}} \exp \left[ -\left( \frac{D^t_{k,s} - \theta^t_{k,s}}{2\sigma^2} \right)^2 \right],
\]

where \( N_D \) denotes the total number of data items, i.e. \( N_D = nN_{\text{tot}} \), and the standard deviation \( \sigma \) is interpreted as a hyper-parameter to be estimated.

Due to the convenience of numerical computation, we search for the optimum \( \hat{\Theta} \) by minimizing cost func-
tion $J$:

$$J = \frac{K N_D}{2} \log(2 \pi \sigma^2) + \frac{1}{2 \sigma^2} \sum_{k=1}^{K} \sum_{s=1}^{N_D} (D^s_k - \hat{\theta}^s_k)^2$$

s.t. $\Theta^I_J \leq \Theta \leq \Theta^U_J$ (I = 1, ..., N),

(17)

which is from a negative logarithm of $p(\Theta|D)$, i.e. $p(\Theta|D) \propto e^{-J}$, instead of maximizing $p(\Theta|D)$. The constraint in Equation (17) arises from the negative logarithm of the prior PDF $p(\Theta)$, which appears when calculating $- \log p(\Theta|D)$. The optimum standard deviation $\hat{\sigma}$ of $\sigma$ is determined by optimizing $J$ with respect to $\sigma$. By solving $\partial J / \partial \sigma = 0$, we obtain

$$\hat{\sigma} = \frac{1}{K N_D} \sum_{k=1}^{K} \sum_{s=1}^{N_D} (D^s_k - \hat{\theta}^s_k)^2,$$

(18)

where $\hat{\theta}$ is the $\theta$ calculated by Equation (10) using $\hat{\theta}$. In practical computations, searching for $\hat{\theta}$ and $\hat{\sigma}$ simultaneously is complex as $\sigma$ might make $J$ singular. To avoid this difficulty, we decompose the optimization problem into two steps: (i) optimization of another cost function given by

$$J' = \frac{1}{2} \sum_{k=1}^{K} \sum_{s=1}^{N_D} (D^s_k - \theta^s_k)^2$$

s.t. $\Theta^I_J \leq \Theta \leq \Theta^U_J$ (I = 1, ..., N);

(19)

and (ii) the calculation of $\hat{\sigma}$ via Equation (18) after optimizing $J'$. The optimization of the cost function $J'$ provides us the same solution as obtained by the optimization of $J$. For the definition of $J'$, the posterior PDF into which the optimum $\hat{\sigma}$ is substituted satisfies $p(\Theta|D) \propto e^{-J'/\hat{\sigma}^2}$.

### 3.3. Optimization utilizing an adjoint model

When searching for $\hat{\theta}$ by optimizing the cost function $J'$ with the constraints, we should use gradient-based optimization, such as the L-BFGS-B method [29], because the number of dimensions of $\Theta$ is large. However, the gradient $\partial J' / \partial \Theta$ needed in gradient-based optimization cannot be calculated by directly differentiating $J'$ with respect to $\Theta$, since $J'$ does not include $\Theta$ explicitly as seen in Equation (19). In 4DVar, we utilize an adjoint method to calculate the objective gradient. This method involves constructing the following three equations with respect to a time-dependent vector $\lambda^t \in \mathbb{R}^N$:

$$-\frac{d\lambda^t}{dt} = \left( \frac{\partial F}{\partial \Theta} \right)^\top \lambda + \frac{\partial J'}{\partial \Theta},$$

(20)

$$\lambda^t(0) = \frac{\partial J'}{\partial \Theta}$$

(21)

where $\frac{\partial J'}{\partial \Theta}$ denotes the transpose of $\theta$, and $t_f$ is an arbitrary time later than the time of the last observation $t_k$. $J'$ is a time-dependent function satisfying

$$J'(\theta) = \frac{1}{2} \sum_{k=1}^{K} (t - t_k) \sum_{s=1}^{N_D} (D^s_k - \theta^s_k)^2,$$

(23)

where $\delta(\bullet)$ stands for the Dirac delta function. The details of the derivation of these equations can be found in [24–26]. The set of Equations (20)–(22), which is called the adjoint model, states that we can obtain the objective gradient $\partial J'/\partial \Theta$ as the initial condition of $\lambda$ (Equation 22) by solving Equation (20) backward in time, with the final condition given by Equation (21). By utilizing the gradient obtained by solving the adjoint model in the iteration of gradient-based optimization, we can search for $\hat{\theta}$ quickly even when the dimensionality of $\Theta$ is large. The number of computations needed to solve the adjoint model is proportional to the same order of the computations needed to solve the simulation model (Equation 10), which is denoted by C. Thus, searching for the optimum $\hat{\theta}$ can be dealt with by a number of computations proportional to $O(C)$. In Appendix A.1, we present a tangible description of the adjoint model where $F$ is not used.

### 3.4. Uncertainty quantification utilizing a second-order adjoint method

The conventional 4DVar described in Section 3.1 allows us to estimate the optimum $\hat{\theta}$ quickly even when the dimensionality of $\Theta$ is large. However, it does not provide us information about the uncertainty in $\hat{\theta}$. To extract this information, another procedure must be implemented on the conventional 4DVar. Ito et al. [19] recently developed a methodology to solve this problem by combining Laplace’s method and the SOA method [26–28] that efficiently evaluates the second derivative of the cost function. The methodology allows us to extract an uncertainty of interest more efficiently than conventional approaches, and hence is appropriate for our purposes, i.e. evaluating the uncertainty in the parameter $\gamma$.

This section briefly explains the essence of the methodology. Having obtained the optimum $\hat{\theta}$ by optimization using the conventional 4DVar, we assess the posterior PDF $p(\Theta|D)$ in the neighborhood of $\Theta = \hat{\theta}$. Given that gradient $\partial J/\partial \Theta|_{\Theta=\hat{\Theta}} = 0$, an approximation of $p(\Theta|D)$ can be obtained by Laplace’s
method as a multivariate normal distribution:

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

where \(|\bullet|\) denotes the determinant of \(\bullet\), and \(H\) is a Hessian matrix of \(\Theta\) given by

\[
H_{ij} = \left. \frac{\partial^2 f}{\partial \Theta_i \partial \Theta_j} \right|_{\Theta = \hat{\Theta}} (i, j = 1, \ldots, N),
\]

or, alternatively,

\[
H = \frac{1}{\sigma^2} H',
\]

for the definition of \(J\). \(H'\) is a Hessian matrix of \(J\), i.e.

\[
H_{ij} = \left. \frac{\partial^2 J}{\partial \Theta_i \partial \Theta_j} \right|_{\Theta = \hat{\Theta}} (i, j = 1, \ldots, N).
\]

Let \(l\) \((1 \leq l \leq N)\) be an index of the component of interest in \(\Theta\). Integrating Equation (24) over all variables except \(\Theta_l\), we obtain a marginal PDF with respect to \(\Theta_l\) as

\[
p(\Theta_l|D) = \frac{1}{\sqrt{2\pi H_{ll}^{-1}}} \exp \left[ -\frac{1}{2} H_{ll}^{-1} \left( \Theta_l - \hat{\Theta}_l \right)^2 \right],
\]

where \(H_{ll}^{-1}\) is the \(l\)th diagonal element in the inverse of the Hessian matrix \(H^{-1}\). This equation means that the variance with respect to \(\Theta_l\) can be obtained by extracting \(H_{ll}^{-1}\) from \(H^{-1}\). However, when \(N\) is massive, it is unrealistic to calculate \(H^{-1}\) directly, as this requires \(O(N^3 + N^2)\) computations and \(O(N^2)\) memory. To save on the computational cost of obtaining \(H_{ll}^{-1}\), the methodology considers the following linear equation of \(r \in \mathbb{R}^N\):

\[
H' r = b,
\]

where \(b \in \mathbb{R}^N\) is a vector with elements \(b_l = 1\) and \(b_{l \neq i} = 0\). The solution \(\hat{r}\) includes \(H_{ll}^{-1}\) as \(\hat{r}_l = \sum_{j=1}^N H_{lj}^{-1} b_j = H_{ll}^{-1}\), and the objective standard deviation \(\hat{H}_{ll}^{-1}\) is calculated by \(\hat{H}_{ll}^{-1} = \hat{\sigma} \sqrt{H_{ll}^{-1}}\). The linear equation shown in Equation (29) can be solved by an iterative technique utilizing the Krylov subspace, such as the conjugate gradient method or the conjugate residual method. The iterative method utilizing the Krylov subspace needs to compute many Hessian-vector products. The SOA method can compute these Hessian-vector products avoiding the computations of all elements of the matrix. The SOA method considers the time evolutions of two time-dependent vectors \(\xi^t \in \mathbb{R}^N\) and \(\zeta^t \in \mathbb{R}^N\), given by

\[
\frac{d\xi}{dt} = \frac{\partial F}{\partial \theta} \xi,
\]

and

\[
\frac{d\zeta}{dt} = \left( \frac{\partial F}{\partial \theta} \right)^\top \zeta + \left( \frac{\partial^2 F}{\partial^2 \theta} \right) \hat{\lambda} + \frac{\partial^2 J}{\partial \theta^2} \zeta.
\]

where \(\hat{\lambda}\) corresponds to \(\lambda\) obtained by solving Equation (20) with \(\theta = \hat{\theta}\), and the \(\theta\)-derivatives represent the \(\theta\)-derivatives into which \(\theta = \hat{\theta}\) is substituted. A combination of Equations (30) and (31), which are called the ‘tangent linear model’ (TL) and the ‘SOA model’, respectively, allows us to compute the Hessian-vector product for an arbitrary vector. Solving Equation (30) in the forward direction of time with a given vector \(\xi^{t_0} = v\), we obtain the time series of \(\xi\). Then, solving Equation (31) backward in time with \(\xi^{t_f} = 0\) and the time series of \(\xi\), we obtain the objective Hessian-vector product as \(\xi^{t_f} = H' v\). See [27] for the detailed derivation of Equations (30) and (31). The number of computations used to solve these models is proportional to \(O(C)\), so that obtaining an uncertainty in \(\Theta\) by solving Equation (29) with an iterative method is proportional to \(O(C)\).

This paper aims to obtain the uncertainty \(\delta \hat{\gamma}\) in the estimated parameter \(\hat{\gamma}\), and hence we associate \(l\) with the index of \(\Theta\) that corresponds to \(\gamma\), i.e. \(l = N\). The procedure used to obtain the uncertainty \(\delta \hat{\gamma}\) can be summarized as follows.

Step 1. Prepare a vector \(b\) with elements \(b_N = 1\) and \(b_{l \neq N} = 0\).

Step 2. Solve Equation (29) via an iterative method utilizing a Krylov subspace. The Hessian-vector products needed in the Krylov subspace method are computed by solving Equations (30) and (31).

Step 3. Extract \(r_N\) from the solution \(r\) of Equation (29) and compute the uncertainty by \(\delta \hat{\gamma} = \hat{\sigma} \sqrt{H_{ll}^{-1}}\), where \(\hat{\sigma}\) can be computed by Equation (18).

Appendix A.2 presents the derivations of the TL and the SOA models for the MPF model.

### 3.5. Prediction of grain structure

The 4DVar-based methodology presented in Sections 3.1–3.4 enables us to determine the optimum estimates that maximize the posterior PDF together with their uncertainties. This section presents how the predictive problem is solved through the obtained optimum estimates and the uncertainties.

Let \(\phi = (\phi_1, \ldots, \phi_{N-1})^\top\) be a vector describing the grain structure at time \(t\). In the prediction methodology, we first evaluate the optimum \(\hat{\gamma}\) for \(\gamma\) and its uncertainty \(\delta \hat{\gamma}\) by the 4DVar-based methodology, where the initial time \(t_0\) is set to \(t_1\). Since we know the initial grain structure accurately in advance of placing the sample in the heating furnace, we utilize the grain structure...
as the ‘true’ structure $\phi^{t_1}$, i.e. we fix $\phi^{t_1}$ to $\hat{\phi}^{t_1}$ in the optimization of the 4DVaR methodology.

The simplest way to obtain the predicted grain structure $\phi^{t_p}$ is to solve the MPF model until $t = t_p$ with the optimum parameter $\gamma$, starting from $\hat{\phi}^{t_1}$. However, it is not adequate from the perspective of the evaluation of $\phi^{t_p}$ because the error in the parameter influences the forecast of $\phi^{t_p}$. Thus, we calculate the change in the forecast of $\phi^{t_p}$ when using a value of $\gamma = \tilde{\gamma}$ slightly deviating from $\gamma$, by associating the change with the time evolution $p(\phi^t | \tilde{\gamma}, D)$, which is a conditional PDF of $\phi^t$ for a given $\gamma = \tilde{\gamma}$ and observational data $D$. Although the time evolution of the conditional PDF, in principle, can be computed by using a sampling method, such as a particle filter, it is unrealistic as the number of dimensions of $\phi^t$ is large. Thus, we employ a strategy whereby we track only the evolution of the maximizer of $p(\phi^t | \gamma, D)$, which is the solution that maximizes the PDF, instead of directly calculating $p(\phi^t | \tilde{\gamma}, D)$. This strategy can help save a large number of computations, as the complexity of obtaining the time evolution of the maximizer is $O(C)$, as described below.

The strategy requires the construction of the time evolution equation of the maximizer from that of $p(\phi^t | \gamma, D)$. To evolve $p(\phi^t | \gamma, D)$, we first need to define the initial PDF $p(\phi^{t_1} | \gamma, D)$. It is linked to the initial joint PDF $p(\theta^{t_1} | D)$ and described as

$$p(\phi^{t_1} | \gamma, D) \propto p(\phi^{t_1}, \gamma | D) = p(\theta^{t_1} | D) |_{\phi^{t_1} = \gamma},$$

(32)

where the initial joint PDF is equivalent to $p(\theta | D)$ shown in Equation (24). However, employing the right-hand side of Equation (32) as the initial PDF makes the computation difficult as we must completely calculate the time evolution of $p(\theta^{t_1} | D)$. Thus, we consider approximating $p(\theta^{t_1} | D)$ by a multivariate normal distribution $q(\theta^{t_1})$ with mean vector $\tilde{\theta}^{t_1} = (\hat{\phi}^{t_1}_1, \hat{\phi}^{t_1}_2, \ldots, \hat{\phi}^{t_1}_{N_{\Theta_1}})$ and covariance matrix $H^{-1}$. Since $q(\theta^{t_1})$ satisfactorily approximates $p(\theta^{t_1} | D)$ when $\gamma \simeq \tilde{\gamma}$, we consider the time evolution of $q(\theta^{t_1})$, rather than that of $p(\theta^{t_1} | D)$, by an integral form using a small time step $\Delta t$ as

$$q(\theta^t) = \int d\theta^{t-\Delta t} w(\theta^t | \theta^{t-\Delta t}) q(\theta^{t-\Delta t}),$$

(33)

where the kernel function $w(\theta^t | \theta^{t-\Delta t})$ provides a mapping from $\theta^{t-\Delta t}$ to $\theta^t$. Since the mapping is given by the discretized Equation (10),

$$\theta^t = \theta^{t-\Delta t} + \Delta t F(\theta^{t-\Delta t}),$$

(34)

the kernel function becomes

$$w(\theta^t | \theta^{t-\Delta t}) = \delta(\theta^t - \theta^{t-\Delta t} - \Delta t F(\theta^{t-\Delta t})).$$

(35)

The combination of Equations (33) and (35) provides a transformation from $q(\theta^{t-\Delta t})$ to $q(\theta^t)$, meaning that $q(\theta^t)$ for any $t$ can be calculated in principle by recursively using Equations (33) and (35), starting at the initial PDF. However, the direct computation of $q(\theta^t)$ is difficult due to the large number of dimensions of $\theta^t$ and the nonlinearity of $F$. Thus, we extract the time evolution equation of the maximizer from that of $q(\theta^t)$ by linearizing Equation (34). Let $\hat{\theta}^t$ be the maximizer of $q(\theta^t)$, and $\theta$ be the state vector obtained by solving Equation (34) with given $\gamma = \tilde{\gamma}$ and initial state $\hat{\phi}^{t_1}$. A Taylor expansion of Equation (34) in the neighborhood of $\theta = \hat{\theta}^t$ is given by

$$\theta^t \simeq \theta^{t-\Delta t} + \Delta t \left[ F(\theta^{t-\Delta t}) + \frac{\partial F}{\partial \theta} |_{\theta = \hat{\theta}^t} (\theta^{t-\Delta t} - \hat{\theta}^{t-\Delta t}) \right]$$

(36)

where both matrices $\hat{A}^{t-\Delta t}$ and $\hat{B}^{t-\Delta t}$ are functions of $\hat{\theta}^{t-\Delta t}$. The use of Equation (36) instead of Equation (34) provides an approximation of the kernel function as

$$w(\theta^t | \theta^{t-\Delta t}) \simeq \delta(\theta^t - \hat{A}^{t-\Delta t} \theta^{t-\Delta t} - \hat{B}^{t-\Delta t}).$$

(37)

The kernel function that describes a linear mapping ensures that when $q(\theta^{t-\Delta t})$ is a multivariate normal distribution, $q(\theta^t)$ is another multivariate normal distribution. Since $q(\theta^{t_1})$ is a multivariate normal distribution, the time evolution equation of the maximizer $\hat{\theta}^t$ is obtained by substituting Equation (37) by Equation (33) and then calculating its expectation,

$$\hat{\theta}^t = \hat{A}^{t-\Delta t} \hat{\theta}^{t-\Delta t} + \hat{B}^{t-\Delta t},$$

(38)

where the initial condition is given by $\hat{\theta}^{t_1} = \hat{\theta}^{t_1}$. Note that we need to solve Equations (34) and (38) simultaneously to obtain the time evolution of $\hat{\theta}^t$. The computational complexity is $O(C)$ in this case, so that we can save a large number of computations compared with the direct computation of the time evolution of $q(\theta^t)$, the complexity of which is proportional to $O(C) \times e^{O(N)}$.

In the next step, we associate $\hat{\theta}^t$ with the maximizer of $p(\phi^t | \gamma, D)$. Let $\hat{\phi}^t$ be the maximizer of $p(\phi^t | \gamma, D)$. Considering that $p(\theta^t | D)$ is approximated by $q(\theta^t)$, $p(\phi^t | \gamma, D)$ is described as

$$p(\phi^t | \gamma, D) \propto p(\theta^t | D) \simeq q(\theta^t).$$

(39)
The linear mapping ensures that \( q(\theta^t) \) is a multivariate normal distribution
\[
q(\theta^t) = \frac{1}{(2\pi)^{N/2}|U|^{1/2}} \exp \left[ -\frac{1}{2} \Delta \theta^\top U^{-1} \Delta \theta \right],
\]
where \( U \in \mathbb{R}^{N \times N} \) is a covariance matrix, and \( \Delta \theta \in \mathbb{R}^N \) is a residual vector with elements \( \Delta \theta_i = \theta^t_i - \tilde{\theta}^t_i \) 
\((i = 1, \ldots, N - 1)\) and \( \Delta \theta_N = \theta^t_N - \tilde{\theta}^t_N = \gamma - \tilde{\gamma} \).
Substituting Equation (40) and \( \gamma = \tilde{\gamma} \) by Equation (39) and subsequently calculating its expectation, \( \bar{\phi}^t_i \) becomes consistent with the elements of \( \tilde{\theta}^t_i \) i.e. \( \bar{\phi}^t_i = \tilde{\theta}^t_i \) 
\((i = 1, \ldots, N - 1)\). Note that the calculation of the expectation does not require the covariance matrix \( U \).
In summary, the procedure to obtain the maximizer of \( p(\phi^{tp}|\tilde{\gamma}, D) \) consists of the following three steps.

Step 1. Set \( \tilde{\theta}^{t_0} = (\tilde{\phi}^{t_0}_1, \tilde{\phi}^{t_0}_2, \ldots, \tilde{\phi}^{t_0}_{N-1}, \tilde{\gamma})^\top \) and \( \theta^{t_0} = \tilde{\theta}^{t_0} \).
Step 2. Evolve \( \tilde{\theta} \) and \( \theta \) simultaneously from \( t = t_{p} \) to \( t = t_{p} \) by using Equation (34) for \( \tilde{\theta} \) and Equation (38) for \( \theta \).
Step 3. Extract \( \phi^{tp} \) from \( \theta^{tp} \).

The use of this procedure together with the method in Section 3.4 enables us to evaluate the uncertainty in prediction from parameter uncertainties. This paper evaluates \( p(\phi^{tp}|\tilde{\gamma}, D) \) by giving \( \tilde{\gamma} \) and \( \delta \tilde{\gamma} \), i.e. it calculates three vectors \( \phi^{tp}_1, \phi^{tp}_2, \) and \( \phi^{tp}_3 \), which are the maximizers of \( p(\phi^{tp}|\tilde{\gamma}, D), p(\phi^{tp}|\tilde{\gamma} + 3\delta \tilde{\gamma}, D) \), and \( p(\phi^{tp}|\tilde{\gamma} - 3\delta \tilde{\gamma}, D) \), respectively.

4. Setup for estimation tests

We apply the proposed methodology to the MPF model to assess its performance by using estimation tests involving a time series of synthetic data. The synthetic data are created by the following procedures: (i) we prepare a time series of \( \phi \) with a given initial state \( \phi^0 = \hat{\phi} \) and parameter \( \gamma = \tilde{\gamma} \) (we call this time series the ‘true trajectory’); (ii) we extract \( \phi^k \) at time \( t_k \) \((k = 1, \ldots, K)\) from the calculated time series; and (iii) we then add noise that follows a normal distribution with mean zero and variance \( \sigma^2 \). Our methodology is validated if it can reproduce \( \tilde{\gamma} \) and accurately estimate \( \phi^{tp} \).

We call this type of estimation test a ‘twin experiment’. Throughout the experiments reported in this paper, we use the grain structure shown in Figure 2(a) as \( \phi^0 \) and \( \gamma = 1/\tau \). When performing the twin experiments, we need to calculate the temporal evolutions of the adjoint model, the TL model, and the SOA model, by discretizing them. For simplicity, we discretize them in time with the Euler scheme, and use the same time step and space resolution as the ones used in Section 2.2. To ensure that each element in \( \Theta \) satisfies the constraint (Equation 13), \( J' \) is optimized by using the L-BFGS-B technique [29]. The upper bound \( \gamma_{\max} \) of the constraint on \( \gamma \) described in Section 3.2 is set to \( \gamma_{\max} = 2/\tau \). When we quantify the uncertainty in \( \gamma \), we use the conjugate residual method as the Krylov subspace method.

5. Results and discussion

The proposed method is validated through two types of twin experiments: (I) estimation of parameter \( \gamma \) conditional on fixing \( \phi^0 = \hat{\phi} \), and (II) evaluation of \( \phi^{tp} \). Test I determines whether the 4DVar-based methodology works well by investigating how the estimates depend on observational data. Test II investigates and quantifies how the prediction of the grain structure depends on the estimations.

5.1. Test I: Parameter estimation

Test I investigates the influences of three parameters related to the observational data: (i) the interval of observation \( \Delta T \); (ii) the standard deviation of observation noise \( \sigma \); and (iii) the length of observation time \( T_{\max} \). In this test, we assume that the observations are obtained at equal intervals \( \Delta T \), i.e. the number of observations \( K \) is given by \( K = T_{\max}/\Delta T \), and the observation times are given by \( t_k = k\Delta T \) \((k = 1, \ldots, K)\). The observational parameters used in Test I are shown in Table 1.

Figure 3 shows the results of Test I(i). Figure 3(a) indicates that parameter estimation is successful as the estimated parameter is in the range \( \hat{\gamma} - \delta \hat{\gamma} < \hat{\gamma} < \hat{\gamma} + \delta \hat{\gamma} \). Figure 3(b) indicates that \( \delta \hat{\gamma} \) is a monotonically increasing function of \( \Delta T \), and is proportional to \( \sqrt{\Delta T} \) for \( \Delta T < \tau \).

The proposed method is validated through two types of twin experiments: (I) estimation of parameter \( \gamma \) conditional on fixing \( \phi^0 = \hat{\phi} \), and (II) evaluation of \( \phi^{tp} \). Test I determines whether the 4DVar-based methodology works well by investigating how the estimates depend on observational data. Test II investigates and quantifies how the prediction of the grain structure depends on the estimations.

5.1. Test I: Parameter estimation

Test I investigates the influences of three parameters related to the observational data: (i) the interval of observation \( \Delta T \); (ii) the standard deviation of observation noise \( \sigma \); and (iii) the length of observation time \( T_{\max} \). In this test, we assume that the observations are obtained at equal intervals \( \Delta T \), i.e. the number of observations \( K \) is given by \( K = T_{\max}/\Delta T \), and the observation times are given by \( t_k = k\Delta T \) \((k = 1, \ldots, K)\). The observational parameters used in Test I are shown in Table 1.

Figure 3 shows the results of Test I(i). Figure 3(a) indicates that parameter estimation is successful as the estimated parameter is in the range \( \hat{\gamma} - \delta \hat{\gamma} < \hat{\gamma} < \hat{\gamma} + \delta \hat{\gamma} \). Figure 3(b) indicates that \( \delta \hat{\gamma} \) is a monotonically increasing function of \( \Delta T \), and is proportional to \( \sqrt{\Delta T} \) for \( \Delta T < \tau \).

| Test I(i) | Test I(ii) | Test I(iii) |
|-----------|-----------|------------|
| \( \Delta T \) | 0.1 \tau – 204.8 \tau | 0.1 \tau | 0.1 \tau |
| \( \sigma \) | 10^{-2} | 10^{-6} – 10^{-2} | 10^{-2} |
| \( T_{\max} \) | 409.6 \tau | 409.6 \tau | 12.8 \tau – 409.6 \tau |

Figure 3. Results of Test I(i). The length of each error bar for optimum parameter \( \hat{\gamma} \) in (a) corresponds to the estimated uncertainty \( \delta \hat{\gamma} \) in (b). The black solid line in (b) indicates a square root function.
ure5(b) shows that parameter estimation is satisfactory. Figures 3(b) and 4(b) become fine in accordance with increasing number of observations (i.e. $1/\Delta T \to \infty$) or decreasing roughness of data (i.e. $\sigma \to 0$).

Figure 5 shows the results of Test I(iii). Figure 5(a) indicates that parameter estimation is satisfactory. Figure 5(b) shows that $\delta \hat{\gamma}$ is a monotonically decreasing function of $T_{\text{max}}$ that does not obey any power laws in the range, and its functional form appears to saturate to a constant when $T_{\text{max}} \to \infty$.

These features of the uncertainty $\delta \hat{\gamma}$ can be explained by the following theoretical consideration of the Hessian matrix. Having obtained the optimum $\hat{\Theta}$ by optimization using the conventional 4DVar described in Section 3.3, the Hessian matrix at the optimum $\hat{\Theta}$ can be described by a formal solution given by

$$H = \frac{1}{\hat{\sigma}^2} \sum_{k=1}^{K} R^k,$$

where $\hat{\sigma}$ is obtained by Equation (18). The time-dependent matrix $R \in \mathbb{R}^{N \times N}$ is given by

$$R^k = \sum_{s=1}^{N_D} \left[ \frac{\partial \hat{\sigma}^k}{\partial \hat{\Theta}} \otimes \frac{\partial \hat{\sigma}^k}{\partial \hat{\Theta}} + (\hat{\sigma}^k - D^k) \frac{\partial^2 \hat{\sigma}^k}{\partial \hat{\Theta}^2} \right],$$

where $\otimes$ denotes a direct product. Letting the number of observations be infinity with a fixed length of observation time, i.e. $\Delta T \to 0$ and $K \to \infty$ such that $\Delta TK = T_{\text{max}}$ = constant, the limitation of the Hessian matrix becomes

$$H \to \frac{1}{\hat{\sigma}^2 \Delta T} Q,$$

where a matrix $Q \in \mathbb{R}^{N \times N}$ is given by

$$Q = \int_0^{T_{\text{max}}} \frac{dt}{\sigma^2} R^t = \sum_{s=1}^{N_D} \int_0^{T_{\text{max}}} \frac{dt}{\sigma^2} \left[ \frac{\partial \hat{\sigma}}{\partial \hat{\Theta}} \otimes \frac{\partial \hat{\sigma}}{\partial \hat{\Theta}} + (\hat{\sigma} - D^t) \frac{\partial^2 \hat{\sigma}}{\partial \hat{\Theta}^2} \right].$$

As $Q$ does not depend on $\Delta T$ or $\hat{\sigma}$, we immediately find, from the functional form of Equation (43), that the uncertainty is proportional to $\sigma \sqrt{\Delta T}$ on the assumption that $\hat{\sigma}$ converges with $\sigma$ when $K \to \infty$. This fact qualitatively explains the power laws associated with $\Delta T$ and $\sigma$ shown in Figures 3(b) and 4(b), respectively. Furthermore, as $Q$ includes all information pertaining to the simulation model, the power laws are naturally independent of the details of the model. The $T_{\text{max}}$ dependence of the uncertainty can be described by $Q^{-1}$. Although it is difficult to evaluate $Q^{-1}$ analytically, its long-term behavior can be discussed. As seen in Equation (44), the diagonal elements in term (a) are always positive, whereas the off-diagonal elements in term (a) and all elements in term (b) can be either positive or negative. When $T_{\text{max}}$ is significantly large, the diagonal elements of $Q$ grow to positive values larger than the off-diagonal elements, and this growth continues until $\phi^t$ is uniform in the computational domain. That is, $Q^{-1}$ approaches a constant diagonal matrix when $T_{\text{max}} \to \infty$, and this causes the saturation of the uncertainty. For this reason, we consider that the uncertainty shown in Figure 5(b) approaches a constant value when $T_{\text{max}} \to \infty$.

5.2. Test II: Prediction of grain structure.

In Test II, we use three types of synthetic datasets parameterized by interval $\Delta T$. The synthetic dataset is assumed to be composed of two snapshots of grain structures (i.e. $K = 2$) at times $t_1 = 25.6\tau$ and $t_2 = t_1 + \Delta T$, where the list of $\Delta T$ used in this test is summarized in Table 2. The standard deviation $\sigma$ for generating the synthetic data is set to 0.1. Using the synthetic dataset, we estimate the parameter $\gamma$ together with its uncertainty, and then apply our prediction method. In the three tests, we confirmed that the uncertainties were significantly small compared with the estimated parameters.

Figure 6 shows the results of Test II. To quantify the differences between Tests II(i)–II(iii), we investigate...
They show a small grain in the average grain size. In particular, the large reductions of the deviations appear at $\tau_{m}$ and $\tau_{s}$, respectively. Figures 6(a)–6(c) indicate the trajectories of the normalized variations in Tests II(i)–II(iii). The estimation of the grain structure is successful because the trajectory of $d_{m}$ is almost zero. Furthermore, we find that the trajectories of $d_{+}$ and $d_{-}$ enclose that of $d_{m}$, meaning that our methodology is successful in providing the upper/lower realizations of the grain structure. The deviations in $d_{+}$ and $d_{-}$ from $d_{m}$ tend to grow over time, and their growth rates are smaller in accordance with the magnitude of $\Delta T$. These results imply that using a larger $\Delta T$ provides better estimation of grain structure because $\gamma$ characterizes the speed of the grain boundary, meaning that using snapshots of structures that are significantly different is advantageous for estimating $\gamma$. Furthermore, $d_{+}$ and $d_{-}$ fluctuate in their trajectories, indicating that the normalized variation is sensitive to that in the average grain size. In particular, the large reductions of the deviations appear at $t \sim 200\tau$ in Figures 6(a)–6(c). These phenomena can be explained by considering the dynamical properties of the MPF model. Figures 6(d) and 6(e) show the predicted grain structures $\rho_{+}$ and $\rho_{-}$, respectively, at $t = 200\tau$ in Test II(i), which are $\rho$ calculated by $\phi_{+}$ and $\phi_{-}$, respectively. They show a small grain in $\rho_{-}$ that is not observed in $\rho_{+}$. In general, once small grains have disappeared, the grain boundaries around multiple points rapidly vary to their equilibrium forms (i.e. triplet points with a 120-degree structure). Once the 120-degree structure has been formed, the growth/decay of grains slows down. This implies that the grain structures $\rho_{+}$ and $\rho_{-}$, once the small grains have disappeared, tend to be synchronized with each other. The large reduction in the deviation at $t \sim 200\tau$ is due to this synchronization.

### 6. Conclusions

This paper proposed a 4DVar-based DA procedure applicable to the MPF model that describes purely interface-driven grain growth. We adopted the 4DVar technique proposed by Ito et al. [19] that allows us to obtain the optimum estimates together with their uncertainties within realistic computational time and resources, even in the case of massive simulation models like the MPF model. The procedure properly estimated parameters of the MPF model together with their uncertainties. Moreover, we devised a method to determine how a perturbation of the parameter influences the prediction of the temporal evolution of grain structure. The method allows us to quantify the uncertainty propagation of the parameter by combining the 4DVar technique proposed by Ito et al. We tested the method through numerical tests and quantified how the observational conditions influenced the prediction of the temporal evolution of grain structure. These results are expected to provide valuable information related to the mechanical properties of metals.

Further investigation into the uncertainty in the prediction can improve the design of the observational setup. For example, we can optimize a distribution of sensors and/or observation frequencies that minimizes prediction uncertainties. Such an investigation can allow us to attain a cost-effective development of materials within limited observational time and resources. Furthermore, the proposed prediction method is applicable not only to the interface-driven MPF model, but also to other PF models that involve more complex physics, such as various types of chemical reactions, and thermal and elastic conditions. Through such applications, we can predict the future states of the key factors related to mechanical properties, such as the spatial distributions of chemical compositions and the residual stress in the metal, as well as the grain structure. The 4DVar technique can also estimate past states by shifting the time of origin $t_{O}$ to before the first observational time $t_{1}$. This enables us to uncover the histories of the mechanical properties and then pursue the causality that explains the given state of the metal. We believe that such applications will contribute to the efficient development of materials in the future.

### Disclosure statement

No potential conflict of interest was reported by the authors.
**Funding**

This work was supported by the Council for Science, Technology and Innovation (CSTI), the Cross-ministerial Strategic Innovation Promotion Program (SIP) 'Structural Materials for Innovation’ (Funding agency: JST).

**References**

[1] Manohar PA, Dunne DP, Chandra T, et al. Grain growth predictions in microalloyed steels. ISIJ Int. 1996;36(2):194–200. DOI:10.2355/isijinternational.36.194.

[2] Uhm S, Moon J, Lee C, et al. Prediction model for the austenite grain size in the coarse grained heat affected zone of Fe-C-Mn steels: considering the effect of initial grain size on isothermal growth behavior. ISIJ Int. 2004;44(7):1230–1237. DOI:10.2355/isijinternational.44.1230.

[3] Lee SJ, Lee YK. Prediction of austenite grain growth during austenitization of low alloy steels. Mater Design. 2008;29(9):1840–1844. DOI:10.1016/j.matdes.2008.03.009.

[4] Kobayashi R. Modeling and numerical simulations of dendritic crystal growth. Physica D. 1993;63(3):410–423. DOI:10.1016/0167-2789(93)90120-P.

[5] Steinbach I, Pezzolla F, Nestler B, et al. Phase-field simulation of solidification. Ann Rev Mater Res. 2002;32(1):163–194. DOI:10.1146/annurev.matsci.32.101901.155803.

[6] Chen LQ. Phase-field models for microstructure evolution. Ann Rev Mater Res. 2002;32(1):163–194. DOI:10.1146/annurev.matsci.32.101901.155803.

[7] Chen LQ. Phase-field models for microstructure evolution. Ann Rev Mater Res. 2002;32(1):163–194. DOI:10.1146/annurev.matsci.32.101901.155803.

[8] Ohno M, Yamaguchi T, Sato D, et al. Existence or nonexistence of thermal pinning effect in grain growth under temperature gradient. Comput Mater Sci. 2013;69:7–13. DOI:10.1016/j.commatsci.2012.11.017.

[9] Kitagawa G. Introduction to time series modeling. Chapman & Hall/CRC Monographs on Statistics & Applied Probability, Boca Raton, FL: CRC Press; 2010.

[10] Doucet A, Godsill S, Andrieu C. On sequential Monte Carlo sampling methods for Bayesian filtering. Stat Comput. 2001;10(3):197–208. DOI:10.1023/A:1008935410038.

[11] Le Dimet FX, Talagrand O. Variational algorithms for data assimilation of meteorological observations: theoretical aspects. Tellus A. 1986;38A(2):97–110. DOI:10.1111/j.1600-0870.1986.tb00459.x.

[12] Talagrand O, Courtier P. Variational assimilation of meteorological observations with the adjoint vorticity equation. I: theory. Q J R Meteorol Soc. 1987;113(478):1311–1328. DOI:10.1002/qj.49711347812.

[13] Wang Z, Navon I, Le Dimet F, et al. The second order adjoint analysis: theory and applications. Meteorol Atmos Phys. 1992;50(1–3):3–20. DOI:10.1007/BF01025501.

[14] Wang Z, Droegemeier K, White L. The adjoint Newton algorithm for large-scale unconstrained optimization in meteorology applications. Comput Optim Appl. 1998;10(3):283–320. DOI:10.1023/A:1018321307393.

[15] Kitagawa G. Introduction to time series modeling. Chapman & Hall/CRC Monographs on Statistics & Applied Probability, Boca Raton, FL: CRC Press; 2010.

[16] Doucet A, Godsill S, Andrieu C. On sequential Monte Carlo sampling methods for Bayesian filtering. Stat Comput. 2001;10(3):197–208. DOI:10.1023/A:1008935410038.

[17] Le Dimet FX, Talagrand O. Variational algorithms for data assimilation of meteorological observations: theoretical aspects. Tellus A. 1986;38A(2):97–110. DOI:10.1111/j.1600-0870.1986.tb00459.x.

[18] Talagrand O, Courtier P. Variational assimilation of meteorological observations with the adjoint vorticity equation. I: theory. Q J R Meteorol Soc. 1987;113(478):1311–1328. DOI:10.1002/qj.49711347812.

[19] Wang Z, Navon I, Le Dimet F, et al. The second order adjoint analysis: theory and applications. Meteorol Atmos Phys. 1992;50(1–3):3–20. DOI:10.1007/BF01025501.

[20] Wang Z, Droegemeier K, White L. The adjoint Newton algorithm for large-scale unconstrained optimization in meteorology applications. Comput Optim Appl. 1998;10(3):283–320. DOI:10.1023/A:1018321307393.

[21] Wang Z, Droegemeier K, White L. The adjoint Newton algorithm for large-scale unconstrained optimization in meteorology applications. Comput Optim Appl. 1998;10(3):283–320. DOI:10.1023/A:1018321307393.

[22] Le Dimet FX, Navon I, Daescu DN. Second-order information in data assimilation*.* Mon Weather Rev. 2002;130(3):629–648. DOI:10.1175/1520-0493(2002)130<0629:SOID>2.0.CO;2.

[23] Byrd RH, Lu P, Nocedal J, et al. A limited memory algorithm for bound constrained optimization. SIAM J Sci Comput. 1995;16(5):1190–1208. DOI:10.1137/0916069.

**Appendix 1. Models for data assimilation**

This appendix presents the tangible descriptions of the adjoint model, the TL model, and the SOA model used for the construction of our methodology in the main body of the paper.
A.1. Derivation of the adjoint model

Substituting the right-hand sides shown in Equations (8) and (9) for \( F \) as shown in Equation (20), the adjoint model can be given as

\[
- \frac{d\lambda_I}{dt} = \begin{cases} 
\theta_N \left( \epsilon^2 \Delta m_{mm'} \lambda_{i,m'} + 4 \lambda_{i,m'} \theta_{i,m} \right) \\
- 2 \theta_N \theta_{i,m} \left( \lambda_{i,m} \theta_{i,m} + 2 \theta_{i,m} \lambda_{i,m} \right) \\
- \sum_{k=1}^{K} \delta(t - t_k) \left( D'_{i,m} - \theta_{i,m} \right) \\
\lambda_{i,m} \left( \epsilon^2 \Delta m_{mm'} \theta_{i,m'} + 2 \theta_{i,m}^2 - 2 \theta_{i,m} \theta_{i,m} \right) \\
(I = N), \\
\lambda_{I^0}^I = 0 \quad (I = 1, \ldots, N), \\
\lambda_{I^0}^I = \frac{\partial f^I}{\partial \theta_I} \quad (I = 1, \ldots, N). 
\end{cases} 
\]  

\[ \lambda_{I^0}^I = 0 \quad (I = 1, \ldots, N), \]  

\[ \lambda_{I^0}^I = \frac{\partial f^I}{\partial \theta_I} \quad (I = 1, \ldots, N). \]  

A.2. Derivation of the tangent linear model and the second-order adjoint model

The TL model and the SOA model are derived in the same manner as the adjoint model shown in Appendix A.1. The TL model is given by

\[
- \frac{d\xi_I}{dt} = \begin{cases} 
\hat{\theta}_N \left( \epsilon^2 \Delta m_{mm'} \xi_{i,m'} + 4 \xi_{i,m'} \hat{\theta}_{i,m} \right) \\
- 2 \hat{\theta}_N \hat{\theta}_{i,m} \left( \xi_{i,m} \hat{\theta}_{i,m} + 2 \hat{\theta}_{i,m} \xi_{i,m} \right) \\
+ \xi_N \left( \epsilon^2 \Delta m_{mm'} \hat{\theta}_{i,m'} + 2 \hat{\theta}_{i,m}^2 - 2 \hat{\theta}_{i,m} \hat{\theta}_{i,m} \right) \\
(I = i, m, m = 1, \ldots, N_{\text{tot}} \text{ and } i = 1, \ldots, n), \\
0 \quad (I = N), 
\end{cases} 
\]

and the SOA model by

\[
- \frac{d\xi_I}{dt} = \begin{cases} 
\hat{\theta}_N \left( \epsilon^2 \Delta m_{mm'} \xi_{i,m'} + 4 \xi_{i,m'} \hat{\theta}_{i,m} \right) \\
- 2 \hat{\theta}_N \hat{\theta}_{i,m} \left( \xi_{i,m} \hat{\theta}_{i,m} + 2 \hat{\theta}_{i,m} \xi_{i,m} \right) \\
+ \xi_N \left( \epsilon^2 \Delta m_{mm'} \hat{\theta}_{i,m'} + 4 \hat{\theta}_{i,m} \hat{\theta}_{i,m} \right) \\
(I = i, m, m = 1, \ldots, N_{\text{tot}} \text{ and } i = 1, \ldots, n), \\
0 \quad (I = N), 
\end{cases} 
\]