



Ensemble Kalman filter-based data assimilation for three-dimensional multi-phase-field model: Estimation of anisotropic grain boundary properties

Akinori Yamanaka^{a,*}, Yuri Maeda^b, Kengo Sasaki^{b,c}

^a Division of Advanced Mechanical Systems Engineering, Institute of Engineering, Tokyo University of Agriculture and Technology, 2-24-16, Naka-cho, Koganei-shi, Tokyo 184-8588, Japan

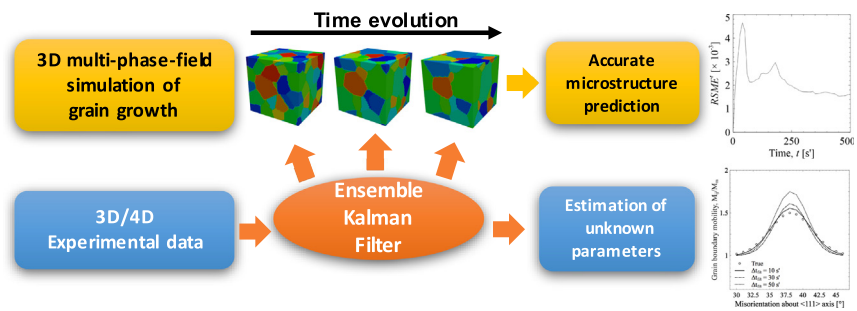
^b Department of Mechanical Systems Engineering, Graduate School of Engineering, Tokyo University of Agriculture and Technology, 2-24-16, Naka-cho, Koganei-shi, Tokyo 184-8588, Japan

^c KOZO KEIKAKU ENGINEERING Inc., 4-38-13, Hon-cho, Nakano-ku, Tokyo 164-0012, Japan

HIGHLIGHTS

- Ensemble Kalman filter-based data assimilation was applied to 3D multi-phase-field model of polycrystalline material.
- Synthetic microstructure data was assimilated to predict energy cusp and mobility peak of $\Sigma 7$ grain boundary.
- Four related model parameters and the assumed shapes of energy cusp and mobility peak were accurately estimated.
- Applications for predicting unknown material properties or unmeasurable microstructure evolution using experimental data.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:

Received 26 October 2018
Received in revised form 26 December 2018
Accepted 29 December 2018
Available online 5 January 2019

Keywords:

Multi-phase-field model
Data assimilation
Ensemble Kalman filter
Grain boundary

ABSTRACT

Data assimilation (DA) has been used as a machine learning approach to estimate a system's state and the unknown parameters in its numerical model by integrating observed data into model predictions. In this paper, we propose using the DA methodology based on the ensemble Kalman filter (EnKF) to improve the accuracy of microstructure prediction using three-dimensional multi-phase-field (3D-MPF) model and estimate the model parameters simultaneously. To demonstrate the applicability of the DA methodology, we performed numerical experiments in which a priori assumed true parameters related to the grain boundary (GB) energy cusp and GB mobility peak of $\Sigma 7$ coincidence site lattice GB were estimated from synthetic data of time-evolving polycrystalline microstructure. Four model parameters related to the $\Sigma 7$ GB properties were successfully estimated by assimilating the synthetic microstructure data to the 3D-MPF model predictions using the EnKF-based DA method. Furthermore, we accurately reproduced the preliminarily assumed true shapes of GB energy cusp and GB mobility peak by using the estimated parameters. The results suggest that implementation of the EnKF-based DA method in the MPF model has great potential for identifying unknown material properties and estimating unmeasurable microstructure evolutions in polycrystalline materials based on real time-series 3D microstructure observation data.

© 2018 Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Polycrystalline grain growth is one of the most important microstructure evolutions in engineering processes. The average grain size

* Corresponding author.

E-mail addresses: a-yamana@cc.tuat.ac.jp (A. Yamanaka), s170868q@st.go.tuat.ac.jp (Y. Maeda), kengo-sasaki@kke.co.jp (K. Sasaki).

and grain size distribution resulting from the grain growth strongly affect mechanical properties of metallic materials such as the yield strength. Therefore, the accurate prediction of grain growth behavior is required to design new materials and improve the performance of practical engineering materials. In order to accelerate the development of new materials, the computation-based material design using various numerical models has been widely studied. In the last decades, many numerical models, e.g. Monte-Carlo Potts model [1–4] and cellular automaton model [5,6], have been developed to predict microstructure evolutions in the grain growth and recrystallization. In particular, the phase-field models have been actively used to simulate various microstructure evolutions on the basis of total free energy of materials [7–9]. Recently, the multi-phase-field (MPF) models [10–12] have been used as the typical numerical models to simulate microstructure evolution in polycrystalline materials. By using massively parallel computing technique with multiple GPUs, very large-scale MPF simulation of polycrystalline grain growth in a system containing thousands of crystal grains has been realized [13].

However, simulation of realistic grain growth and recrystallization behaviors using the MPF models requires accurate data of anisotropic grain boundary (GB) properties and material parameters as input data. In particular, the GB energy and GB mobility influence not only the grain size, but also the crystallographic texture which determines anisotropic mechanical properties of metallic materials [14]. For example, in the recrystallization in aluminum alloys, a major reason for the development of {100}<001>. Cube texture has been attributed to the preferential migration of $\Sigma 7$ coincidence site lattice (CSL) GBs between deformed mother grains and recrystallized grains [15]. The fraction of Cube texture in aluminum alloys influences the deformability [16]. Therefore, there is a strong need to acquire anisotropic GB properties to predict non-uniform GB migration and texture development during recrystallization followed by grain growth [17].

The GB properties have been both measured experimentally and estimated by computational approaches [18–25]. Shibuta et al. [18] calculated the misorientation-dependent GB energies of various tilt GBs in pure iron by molecular dynamics (MD) simulation. Ulomek et al. [19] also investigated the anisotropy of various CSL tilt GB energies in Al, Cu, and γ -Fe by MD simulation. Zhu et al. [20] proposed a computational methodology to estimate the GB energy using the combination of MD simulation and machine learning techniques. The GB mobility in pure iron [21] and the fcc-bcc interface boundary mobility [22] were also studied by MD simulations. On the other hand, Gottstein et al. [23] extensively studied the dependence of GB energy and GB mobility on misorientation using the half loop technique as reviewed in their textbook. Huang et al. [24] revealed the misorientation-dependent GB mobility by measuring the growth of recrystallized grains nucleated at the scratched region in a pure Al crystal. Most recently, Basu et al. extended this technique to measure the mobilities of $\Sigma 5$ and $\Sigma 7$ GBs [25]. Although these conventional experimental methods enable the direct measurement of GB energy and mobility, only a limited number of GB properties can be determined from a single experiment, and so many time-consuming experiments are required to obtain a large amount of such data. Moreover, the GB properties obtained by the conventional methods would not be universal, because they are often determined by observing GBs at the surface of a special sample, e.g. bicrystals and tricrystals, in special environments.

In addition to traditional two-dimensional observations, three-dimensional (3D) and four-dimensional (4D) non-destructive microstructure observation techniques using X-rays have been extensively developed [26–32]. The grain growth behavior has been measured by 3D X-ray diffraction (3DXRD) [26,27], X-ray diffraction constant tomography [29–31], and phase contrast tomography [32]. These in-situ and non-destructive experimental techniques allow the observation of microstructure evolution in bulk material without requiring special samples or environments. Therefore, various direct comparisons between the 3D/4D experiments and phase-field (PF) simulations have been

reported [33–35]. For example, Aagesen et al. [33] estimated the solute diffusion coefficient in a liquid phase during the coarsening of dendrite microstructure by comparing the tomography data with the PF simulation. Zhang et al. [35] proposed an optimization method based on the two-point correlation technique, and estimated the liquid diffusion coefficient and the capillary length in Al-Cu alloy. They determined the material parameters by comparing the dendrite coarsening behavior simulated by the PF model and that measured by X-ray diffraction tomography. Zhang [36] also proposed the application of the two-point correlation technique to estimate the GB energy in pure polycrystalline iron, but the GB mobility has not been estimated.

Contrary to the previous studies, the authors have developed the data assimilation (DA) methodology, which estimates unobserved states, e.g., microstructure and solute concentrations, and parameters involved in the PF and MPF models [37–39]. DA is a computational technique to integrate numerical models and experimental data on the basis of Bayesian inference [40,41]. In DA, we assume that the model state variables follow a probability density function (PDF), and we iteratively update the PDF by incorporating experimental data into the existing PDF based on Bayes' theorem [40]. Comparing to the conventional optimization theories and inverse problem methods, the advantage of DA is that it allows us to evaluate the uncertainties of the estimated states and parameters by analyzing the PDFs. DA methodologies have been applied to improve the accuracy of numerical simulations in various natural sciences, especially in meteorology and oceanography [42]. As explained in the authors' paper [38], DA algorithm is generally classified into non-sequential DA such as the four-dimensional variational method (4DVar), and sequential DA based on a sequential Bayesian filter such as Kalman filter, particle filter [43], and ensemble Kalman filter (EnKF) [44–46]. Ito et al. [37] developed a new DA methodology based on 4DVar and the second-order adjoint method. Their DA method was applied to two-dimensional solidification simulation using the Kobayashi's PF model [7]. Later, they adopted the 4DVar-based DA method to the two-dimensional (2D) MPF model of grain growth [39]. They demonstrated that the DA method enables not only the optimum estimation of the system (i.e. polycrystalline grain structure) and the parameters, but also prediction of their uncertainties by performing numerical experiments called twin experiments. The advantage of the non-sequential DA methods is their small computational cost compared to the particle filter and the EnKF, however, the implementation of the 4DVar and the adjoint method to the simulation models requires complex programming skill. Furthermore, after developing the code for DA calculation of a given model, even a slight change in the numerical model would require extensive modifications of the code. On the other hand, Sasaki et al. developed the EnKF-based DA methodology and implemented it to the one-dimensional PF model of the isothermal γ -to- α transformation in Fe-C-Mn alloy [38]. Through the twin experiments, the EnKF-based DA method was demonstrated to be a powerful tool for estimating unknown parameters involved in the PF model. Although this method generally requires more computational resources than the non-sequential DA methods, it is much more flexible since it needs less programming effort even when the model and the number of parameters to be estimated are changed later.

As described above, some DA methodologies have been already applied to the PF and MPF models to estimate the state and the model parameters. However, to the authors' best knowledge, the implementation of DA method to 3D-PF and 3D-MPF models has not been reported. Since real microstructures always evolve in 3D space, the parameters estimated from 3D/4D microstructure data will be more universal than those estimated from 2D/3D data.

In this study we applied the EnKF-based DA methodology [38] to 3D-MPF model, for not only improving the prediction accuracy of the state, i.e. polycrystalline microstructure, but also for estimating parameters involved in the 3D-MPF model. Although there is no limitation in the kind and number of parameters to be estimated by the proposed DA method, to demonstrate the method's applicability and efficiency for

the 3D MPF model, we performed twin experiments where four parameters related to the GB energy cusp and GB mobility peak of $\Sigma 7$ CSL GB were estimated simultaneously. In the twin experiments, we used a synthetic time-varying microstructure dataset generated by the MPF simulation with a priori determined true GB properties. As a result, our EnKF-based DA method successfully estimated multiple parameters characterizing the GB energy cusp and mobility peak.

The remainder of this paper is organized as follows: Sections 2.1 and 2.2 explain the MPF model used in this study and the modelling of anisotropic $\Sigma 7$ GB properties. In Section 2.2, the EnKF is briefly described followed by description of the twin experiment conditions. For details of the theoretical backgrounds and formulation of the EnKF, see the references [38, 44–46]. Section 3 shows the synthetic observation data used for the twin experiments, which is the simulation of polycrystalline grain growth by the MPF model with the true parameters. Section 4 shows the results of the twin experiments, where four parameters related to the GB energy cusp and mobility peak are estimated. In the twin experiments, we investigate the accuracy and uncertainty of the estimated parameters, which depend on the fraction of $\Sigma 7$ GB in the synthetic observation data and the time interval to integrate the observation data to the MPF model. Finally, we state the conclusions of this study and discuss future directions for applying the DA using real experimental data. The Supplementary Table 1 contains the list of symbols used in this paper.

2. Theory

2.1. Multi-phase-field model

The multi-phase-field model proposed by Fan et al. [10] is used in this study. Phase-field variables η_i ($i = 1, 2, \dots, N$) describe the local volume fraction of the i th crystal grain in a system composed of N crystal grains. The value of η_i is unity inside the i th crystal grain and smoothly varies from 1 to 0 inside its boundary region. To simulate the GB migration driven by GB curvature, the total free energy of the system is given as:

$$G = \int_V \left[\left\{ \sum_{i=1}^N \left(-\frac{\alpha}{2} \eta_i^2 + \frac{\beta}{4} \eta_i^4 \right) + \sum_{i=1}^N \sum_{j \neq i}^N \gamma_{ij} \eta_i^2 \eta_j^2 \right\} + \frac{\kappa}{2} \sum_{i=1}^N (\nabla \eta_i)^2 \right] dV \quad (1)$$

Here, the first term of the right-hand side of Eq. (1) describes the bulk free energy that has a minimum value when the constants α , β , and γ_{ij} satisfy $\alpha = \beta > 0$ and $\gamma_{ij} > \beta/2$. γ_{ij} is the constant related to the GB energy between the i th and j th crystal grains. κ is the gradient energy coefficient. Since this study used the following values: $\alpha = 1.0$, $\beta = 1.0$, and $\kappa = 1.0$, the time is treated as a dimensionless unit.

The time evolution of the phase-field variables, η_i , is given by the Allen-Chan equation as:

$$\frac{\partial \eta_i}{\partial t} = -L \frac{\delta G}{\delta \eta_i} = -L \left(\beta \eta_i^3 - \alpha \eta_i + 2 \eta_i \sum_{j \neq i}^n \gamma_{ij} \eta_j^2 - \kappa \nabla^2 \eta_i \right) \quad (2)$$

Here, n is the number of non-zero phase-field variables at an arbitrary position. L is the mobility of phase-field variable, which is given as [47]:

$$L = \frac{\sum_{i \neq j}^n \eta_i^2 \eta_j^2 M_{ij}}{\sum_{i \neq j}^n \eta_i^2 \eta_j^2} \quad (3)$$

where M_{ij} denotes the mobility of the GB between the i th and j th crystal grains. The anisotropies of the GB energy, γ_{ij} , and GB mobility, M_{ij} , are modeled as functions of the misorientation angle, as explained in the next section.

2.2. Modelling of anisotropic grain boundary property

The GB energy and GB mobility are functions of five parameters: three of the misorientation between two adjacent grains and two of GB inclination [48]. Nevertheless, for simplicity we assumed that the GB anisotropy only depends on the misorientation angle about the rotation axis of $\langle 111 \rangle$.

In order to describe the misorientation angle, let us consider a GB between two adjacent crystal grains i and j whose crystal orientations are given by quaternions \mathbf{q}_i and \mathbf{q}_j , respectively. Then, the misorientation angle at the GB is calculated by the following equation [49]:

$$\Delta \theta_{ij} = \min \left\{ 2 \cos^{-1} \left(O_k \mathbf{q}_i \overline{O_l \mathbf{q}_j} \right)_4, 2 \cos^{-1} \left(O_l \mathbf{q}_j \overline{O_k \mathbf{q}_i} \right)_4 \right\} \quad (4)$$

where “ $\overline{\mathbf{q}_i}$ ” describes the inverse rotation with respect to \mathbf{q}_i . O_k ($k = 1, 2, 3, \dots, 24$) represent the 24 independent symmetry operators for a cubic crystal. Using Eq. (4), the deviation angle from the misorientation of the exact $\Sigma 7$ GB about the rotation axis of $\langle 111 \rangle$, $\Delta \theta_{ij, \langle 111 \rangle}$, is given as the following equation:

$$\Delta \theta_{ij, \langle 111 \rangle} = \min \left\{ 2 \cos^{-1} \left(O_k \mathbf{m} \overline{O_l \mathbf{q}_{\Sigma 7}} \right)_4, 2 \cos^{-1} \left(O_l \mathbf{q}_{\Sigma 7} \overline{O_k \mathbf{m}} \right)_4 \right\} \quad (5)$$

where \mathbf{m} represents the unit quaternion which provides the minimum value of Eq. (4). $\mathbf{q}_{\Sigma 7}$ represents the unit quaternion corresponding to the misorientation of the exact $\Sigma 7$ GB about the rotation axis of $\langle 111 \rangle$, and it is calculated by substituting $n_1 = n_2 = n_3 = 1/\sqrt{3}$ and $\chi = 38.21^\circ$ into the following equation:

$$\mathbf{q} = \left(n_1 \sin \frac{\chi}{2}, n_2 \sin \frac{\chi}{2}, n_3 \sin \frac{\chi}{2}, \cos \frac{\chi}{2} \right) \quad (6)$$

In this study, the GB energy, γ_{ij} , is modeled as a function of $\Delta \theta_{ij, \langle 111 \rangle}$ using the Read-Shockley type equation as [50]:

$$\gamma_{ij} = \gamma_m \times \begin{cases} \frac{\Delta \theta_{ij}}{\Delta \theta_m} \left\{ 1 - \log \left(\frac{\Delta \theta_{ij}}{\Delta \theta_m} \right) \right\} & (\Delta \theta_{ij} \leq \Delta \theta_m) \\ 1 - D_1 \left[1 - \frac{\Delta \theta_{ij, \langle 111 \rangle}}{W_1} \left\{ 1 - \log \left(\frac{\Delta \theta_{ij, \langle 111 \rangle}}{W_1} \right) \right\} \right] & (\Delta \theta_{ij, \langle 111 \rangle} < W_1) \\ 1 & (\Delta \theta_{ij} > \Delta \theta_m \text{ and } \Delta \theta_{ij, \langle 111 \rangle} > W_1) \end{cases} \quad (7)$$

where $\Delta \theta_m$ is the minimum misorientation angle of a high-angle grain boundary (HAGB) which is often assumed to have $\Delta \theta_m = 15^\circ$. γ_m is the average grain boundary energy of the HAGB and assumed to be 1.5. D_1 and W_1 are parameters that characterize the depth and width of the GB energy cusp, respectively.

The GB mobility, M_{ij} , is also described as a function of $\Delta \theta_{ij, \langle 111 \rangle}$ using the following equation proposed by Humphrey [51]:

$$M_{ij} = M_m \times \begin{cases} 1 - \exp \left\{ -5 \left(\frac{\Delta \theta_{ij}}{\Delta \theta_m} \right)^4 \right\} & (\Delta \theta_{ij} \leq \Delta \theta_m) \\ 1 - D_2 \exp \left\{ -5 \left(\frac{\Delta \theta_{ij, \langle 111 \rangle}}{W_2} \right)^2 \right\} & (\Delta \theta_{ij, \langle 111 \rangle} < W_2) \\ 1 & (\Delta \theta_{ij} > \Delta \theta_m \text{ and } \Delta \theta_{ij, \langle 111 \rangle} > W_2) \end{cases} \quad (8)$$

where M_m denotes the average GB mobility of the HAGB and assumed to be 1.0. D_2 and W_2 are parameters that characterize the height and width of the GB mobility peak, respectively. Fig. 1 plots the GB energy and GB mobility as functions of $\Delta \theta_{ij, \langle 111 \rangle}$ using $D_1 = 0.5$, $W_1 = 8^\circ$, $D_2 = 0.5$, and $W_2 = 10^\circ$. Here, note that the four parameters, D_1 , W_1 , D_2 , and W_2 are estimated by the EnKF-based DA method in this study.

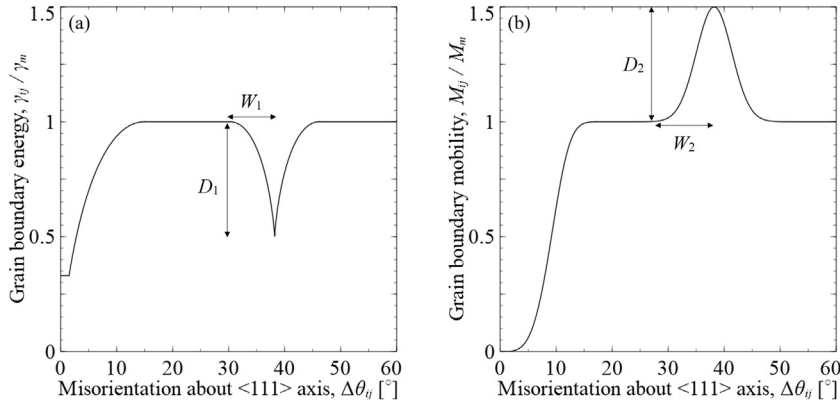


Fig. 1. Variation of grain boundary (GB) (a) energy γ_{ij} and (b) mobility M_{ij} calculated by Eqs. (7) and (8) with the parameters: $D_1 = 0.5$, $W_1 = 8^\circ$, $D_2 = 0.5$, and $W_2 = 10^\circ$. D_1 , D_2 , W_1 , and W_2 , which characterize the depth and width of GB energy cusp and GB mobility peak, are estimated by the EnKF-based DA method.

3. Data assimilation based on ensemble Kalman filter

Similar to our previous work [38], we employ the EnKF with the perturbed observations in this study. Since the detailed formulation of the EnKF was described in the previous paper [38], the EnKF implemented in the MPF model is briefly summarized in this section. For further details of the theoretical background and implementation of the EnKF, see Refs. [40–42, 44–46].

In sequential DA methodologies such as the EnKF, we define a state vector, ξ_t , containing all the state variables at time t . When estimating unknown model parameters based on the sequential DA methodology, the model parameters are included in the state vector, meaning that the parameters are also treated as time-varying state variables. The vector containing both ξ_t and the parameters is called as the augmented state vector, and expressed as:

$$\mathbf{x}_t = \begin{bmatrix} \xi_t \\ \psi_t \end{bmatrix} \quad (9)$$

where ψ_t is a vector containing the model parameters being estimated. As explained in Section 4.1, the augmented state vector contains the phase-field variables, η_i , at each computational grid and the four parameters of D_1 , W_1 , D_2 , and W_2 .

In the EnKF, the system model is defined to describe the time evolution of the augmented state vector as

$$\mathbf{x}_t = f_t(\mathbf{x}_{t-1}) + \mathbf{v}_t \quad (10)$$

where

$$\mathbf{v}_t \sim N(0, \mathbf{Q}_t) \quad (11)$$

Here, f_t is a non-linear operator describing the time evolution of the states from time $t-1$ to t . Therefore, in this study it corresponds to the time evolution equation of the phase-field variables, i.e. Eq. (2). Compared to the real grain growth behavior; the MPF model has various imperfections due to the model uncertainty, unknown parameters, and unknown boundary conditions. Therefore, a system noise, \mathbf{v}_t , is added to represent the model's imperfectness. In Eq. (11), $N(0, \mathbf{Q}_t)$ means that \mathbf{v}_t follows a probability density function (PDF) showing a Gaussian distribution with zero mean and a covariance matrix of \mathbf{Q}_t .

On the other hand, to relate the augmented state vector \mathbf{x}_t to the observation data, we define the observation model as:

$$\mathbf{y}_t = \mathbf{H}_t \mathbf{x}_t + \mathbf{w}_t \quad (12)$$

where

$$\mathbf{w}_t \sim N(0, \mathbf{R}_t) \quad (13)$$

Here, \mathbf{y}_t is a vector containing the experimental or observation data, called as the observation vector. \mathbf{H}_t is a matrix called the observation operator that maps the model state space into the observational space: it extracts from \mathbf{x}_t quantities that are comparable with \mathbf{y}_t . In Eq. (13), \mathbf{w}_t is the observation noise representing errors included in the observation data. In this study, we assumed that \mathbf{w}_t follows a PDF showing a Gaussian distribution with zero mean and a covariance matrix \mathbf{R}_t . As noted in the authors' paper [38], \mathbf{x}_t , \mathbf{v}_t , and \mathbf{w}_t are stochastic variables, because the purpose of DA is not to determine the specific value of each variable but to estimate the PDFs of the stochastic variables.

In the sequential DA algorithms, the augmented state vector is assumed to follow a PDF that is continuously updated by integrating the observation data into the existing PDF based on Bayes' theorem [40]. The EnKF approximates the PDF, $p(\mathbf{x}_t)$, using the Monte-Carlo approximation method by an ensemble of model states as:

$$p(\mathbf{x}_t) \approx \frac{1}{N_{\text{ens}}} \sum_{i=1}^{N_{\text{ens}}} \delta(\mathbf{x}_t - \mathbf{x}_t^{(i)}) \quad (14)$$

where δ is the Dirac delta function. N_{ens} denotes the number of model states called the ensemble size. As explained in Section 4, the ensemble size corresponds to the number of the MPF simulation runs performed in the EnKF-based DA. $\mathbf{x}_t^{(i)}$ represents the i th ($i = 1, 2, 3, \dots, N_{\text{ens}}$) model state (hereafter called the i th ensemble member).

On the basis of the PDF approximated by Eq. (14), we calculate the updates of two PDFs based on the Bayes theorem, i.e. the filter and predictive PDFs, as shown in Fig. 2. The filter PDF at time $t-1$ represents the conditional PDF of the augmented vector \mathbf{x}_{t-1} when the observational data from time $t=1$ to time $t-1$, i.e. $\mathbf{y}_{1:t-1} = \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_{t-1}\}$, are given. The filter PDF at time $t-1$ is written as:

$$p(\mathbf{x}_{t-1} | \mathbf{y}_{1:t-1}) \approx \frac{1}{N_{\text{ens}}} \sum_{i=1}^{N_{\text{ens}}} \delta(\mathbf{x}_{t-1} - \mathbf{x}_{t-1|t-1}^{(i)}) \quad (15)$$

where $\mathbf{x}_{t-1|t-1}^{(i)}$ represents the i th ensemble member which consists of the filter PDF at time $t-1$. Using the system model given by Eq. (10) and $\mathbf{x}_{t-1|t-1}^{(i)}$, the i th ensemble member at time t , i.e. $\mathbf{x}_{t|t-1}^{(i)}$, can be given as:

$$\mathbf{x}_{t|t-1}^{(i)} = f_t(\mathbf{x}_{t-1|t-1}^{(i)}) + \mathbf{v}_t^{(i)} \quad (16)$$

where $\mathbf{v}_t^{(i)}$ represents the system noise for the i th ensemble member. Eq. (16) means that the ensemble of model states is integrated forward

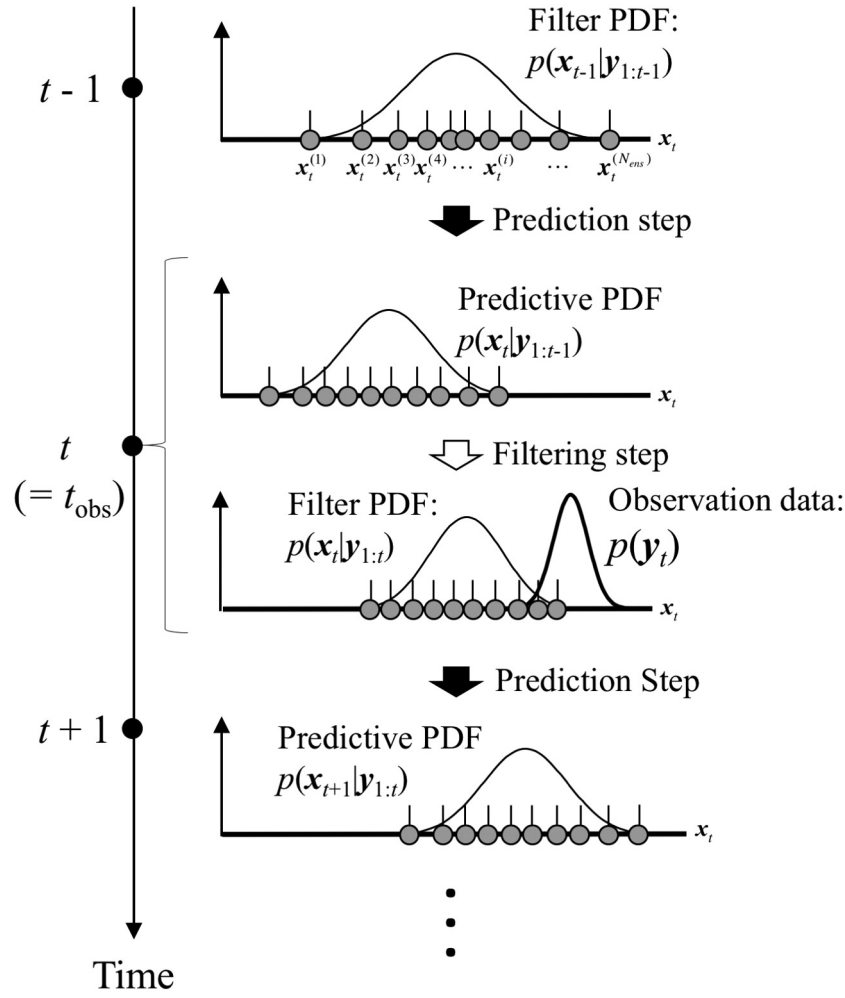


Fig. 2. Schematic procedure to compute the filter and the predictive PDFs based on the EnKF.

in time using the MPF model given as Eq. (2), but subjected to a Gaussian distributed noise which resembles the uncertainty of the MPF model.

Similar to Eq. (15), the predictive PDF at time t is given as:

$$p(\mathbf{x}_t | \mathbf{y}_{1:t-1}) \cong \frac{1}{N_{\text{ens}}} \sum_{i=1}^{N_{\text{ens}}} \delta(\mathbf{x}_{t-1} - \mathbf{x}_{t|t-1}^{(i)}) \quad (17)$$

As shown in Fig. 2, the calculation of the predictive PDF using the ensemble members $\mathbf{x}_{t-1|t-1}^{(i)}$ which approximate the filtered PDF is called as the prediction step.

If the observation data \mathbf{y}^f is available at time t , where the time is denoted as t_{obs} , the observation data is assimilated into the model predictions, and then the filter PDF at time t can be calculated. This process is called as the filtering step. In this step, the calculation of the filter PDF at time t_{obs} is performed based on an extended algorithm of the Kalman filter [53]. The ensemble members that consist of the filter PDF at time t_{obs} ,

denoted as $\mathbf{x}_{t|t}^{(i)}$, are given by the following equation:

$$\mathbf{x}_{t|t}^{(i)} = \mathbf{x}_{t|t-1}^{(i)} + \mathbf{K}_t (\mathbf{y}_t + \tilde{\mathbf{w}}_t^{(i)} - \mathbf{H}_t \mathbf{x}_{t|t-1}^{(i)}) \quad (18)$$

where $\mathbf{y}_t + \tilde{\mathbf{w}}_t^{(i)}$ corresponds to the perturbed observation. \mathbf{K}_t is the ensemble approximation of the Kalman gain (the generalized Kalman gain) at time t and is given as:

$$\mathbf{K}_t = \mathbf{V}_{t|t-1} \mathbf{H}_t^T (\mathbf{H}_t \mathbf{V}_{t|t-1} \mathbf{H}_t^T + \mathbf{R}_t)^{-1} \quad (19)$$

where $\mathbf{V}_{t|t-1}$ is the sample covariance matrix which is calculated as:

$$\mathbf{V}_{t|t-1} = \frac{1}{N_{\text{ens}} - 1} \sum_{i=1}^{N_{\text{ens}}} \tilde{\mathbf{x}}_{t|t-1}^{(i)} (\tilde{\mathbf{x}}_{t|t-1}^{(i)})^T \quad (20)$$

Table 1

Conditions for twin experiments conducted in this study. The impacts of the ensemble size, the initial fraction of $\Sigma 7$ GB, and the filtering interval on the parameter estimation are examined in Tests A, B, and C, respectively. Note that the dimensionless time unit is described as s'.

	Test A	Test B	Test C
Ensemble size: N_{ens}	50, 100, and 200	100	100
Initial fraction of $\Sigma 7$ grain boundary in the observation data: $V_{\Sigma 7}$	38.9%	16.1, 21.4, and 38.9%	38.9%
Filtering interval	10 s'	10 s'	10, 20, 30, and 50 s'

Table 2

Initially estimated true values of the parameters, the mean and the standard deviation (SD) of the initially estimated parameters related to $\Sigma 7$ GB properties, and the SD of the system noise for each parameter. The true values are used to make the synthetic observation data.

Parameter	True value	Mean of initially estimated value	SD of initially estimated value	SD of system noise for the parameter
D_1	0.5	0.1	0.06	0.0025
W_1	8°	1.0°	0.6°	0.04°
D_2	0.5	0.1	0.06	0.0025
W_2	10°	1.0°	0.6°	0.05°

where

$$\tilde{\mathbf{x}}_{t|t-1}^{(i)} = \mathbf{x}_{t|t-1}^{(i)} - \frac{1}{N_{\text{ens}}} \sum_{i=1}^{N_{\text{ens}}} \mathbf{x}_{t|t-1}^{(i)} \quad (21)$$

The time evolution of the sample covariance is given as:

$$\mathbf{V}_{t|t} = \mathbf{V}_{t|t-1} - \mathbf{K}_t \mathbf{H}_t^T \mathbf{V}_{t|t-1} \quad (22)$$

In Eq. (19), \mathbf{R}_t is the observation error covariance matrix given as:

$$\mathbf{R}_t = \frac{1}{N_{\text{ens}} - 1} \sum_{i=1}^{N_{\text{ens}}} \tilde{\mathbf{w}}_t^{(i)} (\tilde{\mathbf{w}}_t^{(i)})^T \quad (23)$$

where

$$\tilde{\mathbf{w}}_t^{(i)} = \mathbf{w}_t^{(i)} - \frac{1}{N_{\text{ens}}} \sum_{i=1}^{N_{\text{ens}}} \mathbf{w}_t^{(i)} \quad (24)$$

Once the ensemble members $\mathbf{x}_{t|t}^{(i)}$ are calculated, the filter PDF at time t , $p(\mathbf{x}_t | \mathbf{y}_{1:t})$, can be obtained by replacing $\mathbf{x}_{t-1|t-1}^{(i)}$ with $\mathbf{x}_{t|t}^{(i)}$ in Eq. (15). In the EnKF, by conducting the above-described filtering and prediction steps iteratively, the model states and parameters are estimated simultaneously.

4. Conditions for twin experiments and synthetic observation data

We conduct three twin experiments termed Tests A, B, and C, where the initially assumed true values of the parameters are estimated with the conditions listed in Table 1. As explained in Section 3, the EnKF updates the ensemble of model states which approximates the predictive and filter PDFs. When the ensemble size N_{ens} increases, the errors of the ensemble approximations of the PDFs decrease proportional to $\sqrt{N_{\text{ens}}}$ [45]. Therefore, adequate approximation of the PDFs generally requires a large ensemble size, which entails high computational costs, i.e. huge computer memory and computational time. Thus, for a given problem we need to decide on a reasonable ensemble size by considering the trade-off between the computational requirements and the accuracy of the ensemble approximation of the PDFs. In Test A, we investigate the impact of the ensemble size on the parameter estimation by performing the twin experiments with different ensemble sizes. Then, we discuss a reasonable ensemble size used for Tests B and C.

The EnKF calculation with the observation data including a large amount of data of the target GB (i.e. $\Sigma 7$ GB in this study) may provide a good estimation of the state and the parameters. However, real experiments for a given system size and number of grains would not always result in data including many target GBs with fine time resolution, because the observation data strongly depend on the sample preparation, observation methods, and experimental environments. Conversely, it is useful to indicate a critical amount and time resolution of the observation data for the state and parameter estimations by conducting twin experiments. Therefore, in Test B, we investigate the effect of the initial fraction of $\Sigma 7$ GB included in the synthetic observation data on the parameter estimation. Finally, Test C clarifies the influence of time interval between the filtering steps on the accuracy of parameter estimation.

4.1. State and observation vectors for twin experiments

We describe the augmented state and the observation vectors used in Tests A, B and C. The time evolution equation of the phase-field variables given by Eq. (2) is solved using the first-order Euler finite difference method (FDM) for the time integration and the second-order

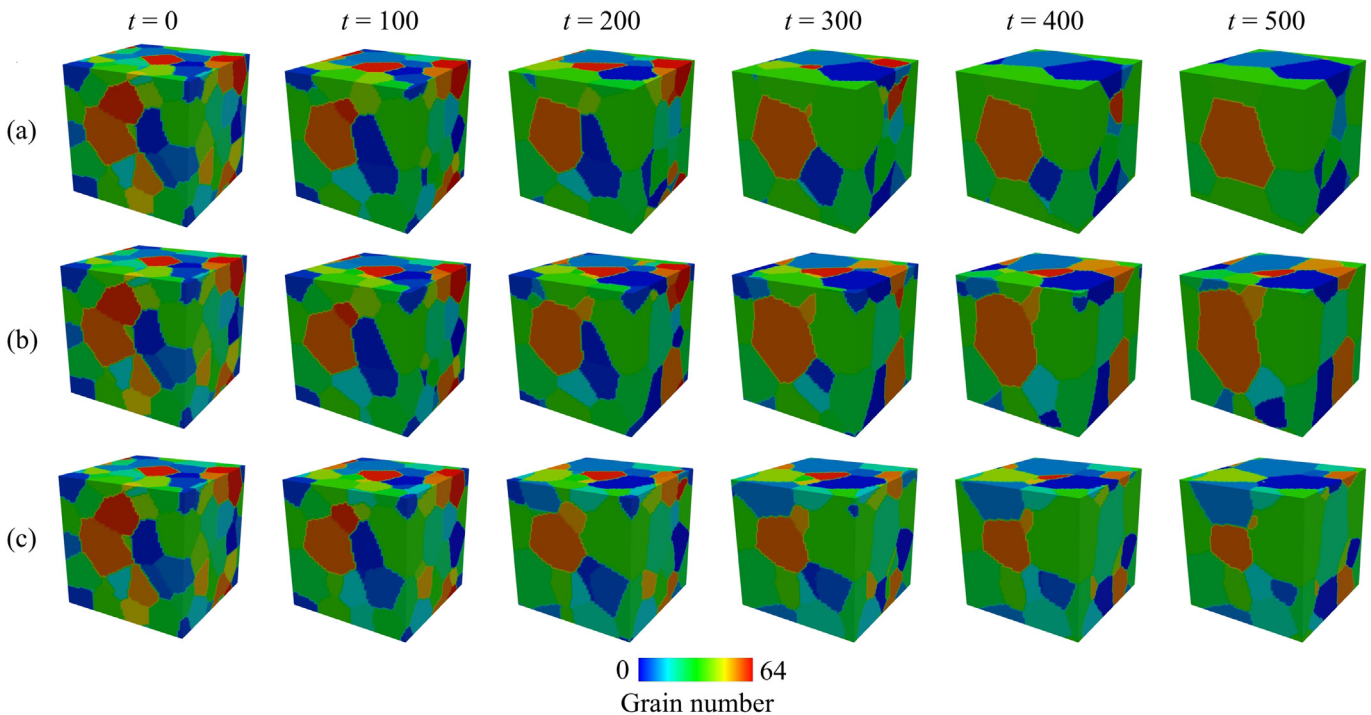


Fig. 3. Snapshots of grain growth in systems with the initial fraction of $\Sigma 7$ GB ($V_{\Sigma 7}$) being (a) 38.9%, (b) 21.4%, and (c) 16.1%. The same initial polycrystalline microstructure is used for all values of $V_{\Sigma 7}$. The phase-field variables that correspond to time evolutions of the grain growth are used as the synthetic observation data in the twin experiments.

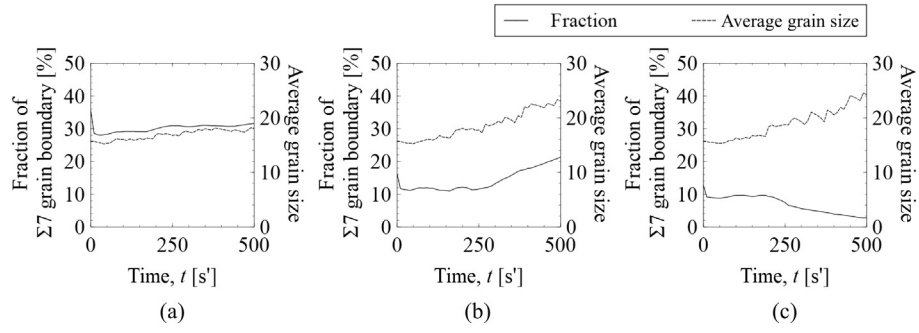


Fig. 4. Time variations of $\Sigma 7$ GB fraction contained in the system for different initial $V_{\Sigma 7}$ values of (a) 38.9%, (b) 21.4%, and (c) 16.1%. The fraction of $\Sigma 7$ GB is calculated by dividing the number of finite difference grids where $|\Delta\theta_{ij, \{111\}}| < 8^\circ$ by the total number of finite different grids.

central FDM for the spatial discretization. Thus, the augmented state vector included in the i th ensemble member, $\mathbf{x}_t^{(i)}$, is given as:

$$\mathbf{x}_t^{(i)} = \begin{bmatrix} \boldsymbol{\xi}_t^{(i)} \\ \boldsymbol{\psi}_t^{(i)} \end{bmatrix} \quad (25)$$

where $\boldsymbol{\xi}_t^{(i)}$ is given as,

$$\boldsymbol{\xi}_t^{(i)} = [\eta_{1,1}^{(i)} \ \eta_{1,2}^{(i)} \ \dots \ \eta_{1,N_{\text{grid}}}^{(i)} \ \eta_{2,1}^{(i)} \ \eta_{2,2}^{(i)} \ \dots \ \eta_{2,N_{\text{grid}}}^{(i)} \ \dots \ \eta_{N,1}^{(i)} \ \eta_{N,2}^{(i)} \ \dots \ \eta_{N,N_{\text{grid}}}^{(i)}]^T \quad (26)$$

where $\eta_{lm}^{(k)}$ denotes the phase-field variable which corresponds to an l th grain on a finite difference grid m ($m = 1, 2, 3, \dots, N_{\text{grid}}$) at time t in the k th ($k = 1, 2, 3, \dots, N_{\text{ens}}$) ensemble member. Since the twin experiments in this study aim to estimate four parameters related to the GB energy cusp and GB mobility peak of $\Sigma 7$ GB, i.e. D_1, D_2, W_1 and W_2 , therefore $\boldsymbol{\psi}_t^{(i)}$ in Eq. (25) is given as:

$$\boldsymbol{\psi}_t^{(i)} = [D_1^{(i)} \ D_2^{(i)} \ W_1^{(i)} \ W_2^{(i)}]^T \quad (27)$$

Here, it should be noted that the parameters are assumed to be updated with time. Their time evolutions can be expressed as follows:

$$D_k^{(i)} = D_k^{t-1(i)} + \mathbf{v}_t \quad \mathbf{v}_t \sim \mathcal{N}(0, \sigma_{\text{sys}}^{D_k}) \quad (28)$$

and

$$W_k^{(i)} = W_k^{t-1(i)} + \mathbf{v}_t \quad \mathbf{v}_t \sim \mathcal{N}(0, \sigma_{\text{sys}}^{W_k}) \quad (29)$$

Here, \mathbf{v}_t denotes the system noise. $\sigma_{\text{sys}}^{D_i}$ and $\sigma_{\text{sys}}^{W_i}$ denote covariance for each noise.

The initial PDF, $p(\mathbf{x}_0|\mathbf{y}_0)$, is assumed to follow a Gaussian distribution with a mean vector \mathbf{x}_0 and a covariance matrix \mathbf{V}_0 . The dimension of \mathbf{V}_0 is $(N \times N_{\text{grid}} + 4) \times (N \times N_{\text{grid}} + 4)$, and its last four diagonal elements are σ_0 , the initial covariance value, and the others zero. It is noted that the identical initial ensemble is used for all the tests.

The synthetic observation data are results obtained by the MPF simulation of grain growth using a priori assumed true values of the parameters. Therefore, \mathbf{y}_t is given as,

$$\mathbf{y}_t = [\eta_{1,1}^t \ \eta_{1,2}^t \ \dots \ \eta_{1,N_{\text{grid}}}^t \ \eta_{2,1}^t \ \eta_{2,2}^t \ \dots \ \eta_{2,N_{\text{grid}}}^t \ \dots \ \eta_{N,1}^t \ \eta_{N,2}^t \ \dots \ \eta_{N,N_{\text{grid}}}^t]^T \quad (30)$$

where η_{ij}^t denotes the phase-field variable describing the local volume fraction of the i th grain on the finite difference grid j ($j = 1, 2, 3, \dots, N_{\text{grid}}$) at time t , which is obtained by MPF simulation with the true parameters. Therefore, the observation matrix H_t used in this study has a dimension of $1 \times (N \times N_{\text{grid}} + 4)$, and its first $N \times N_{\text{grid}}$ diagonal components are 1 and the others are zeros.

4.2. Synthetic observation data

The computational domain and the initial distribution of crystal grains used in all twin experiments are described below. The same

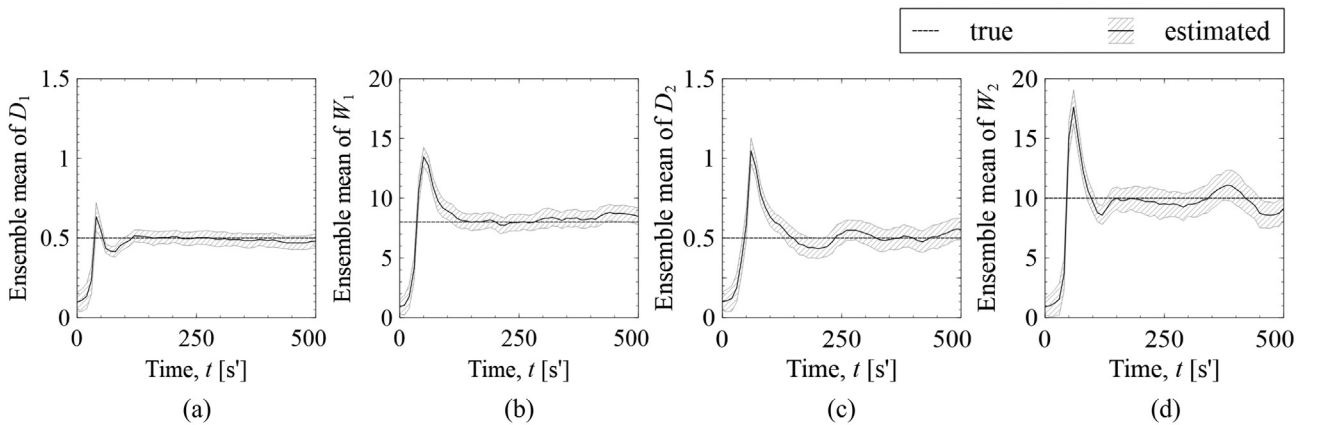


Fig. 5. Time variations of mean and standard deviation (SD) of the estimated parameters, (a) D_1 , (b) W_1 , (c) D_2 , and (d) W_2 , estimated using the ensemble size $N_{\text{ens}} = 100$. Solid and dotted lines indicate the ensemble mean of the estimated parameters and their true values, respectively. The shaded area around the solid line indicates the SD of the estimated parameters in the ensemble members, which represents the uncertainty of each estimated parameter.

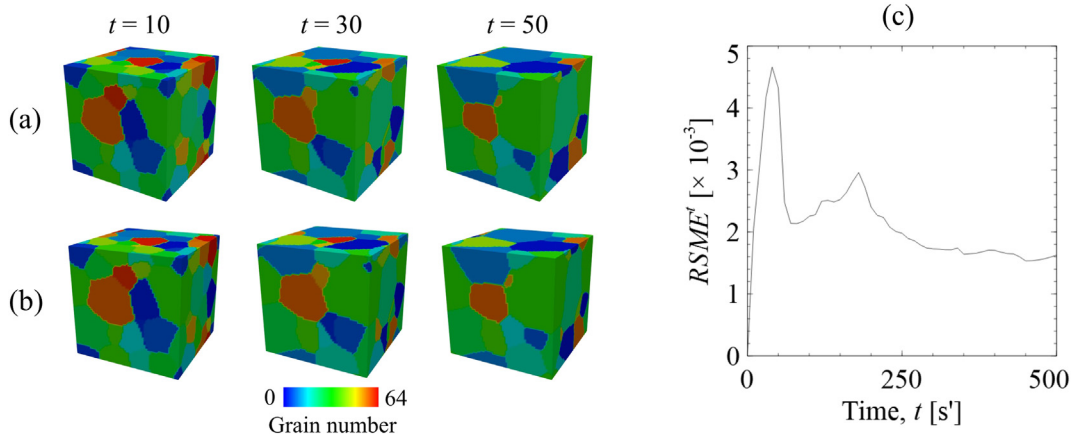


Fig. 6. Snapshots of (a) the true state (polycrystalline structure) used as the synthetic observation data and (b) the estimated state using $N_{ens} = 100$ at $t = 10, 30$, and 50 s'. The time variation of the root mean square error (RMSE), which describes the error of the state estimation, is shown in (c).

conditions are also used to obtain the synthetic observation data. The dimensionless size of the computational domain is $64 \times 64 \times 64$ and divided by $64 \times 64 \times 64$ uniform finite difference grids. Thus, the spacing between the grids is 1.0. Periodic boundary condition is used for all directions. The time increment is 0.05 s' where s' denotes the dimensionless time unit. Table 2 shows the a priori assumed true values of the parameters, as well as the mean and standard deviation of the true parameters. The true parameters are the same as those used to plot Fig. 1. The standard deviation of the system noise for each parameter, which is the hyper parameter for the EnKF with the perturbed observations, is shown in Table 2. The tuning of the hyper parameter is essential for a better DA calculation. In this paper, we tuned the hyper parameters by performing preliminary twin experiments. Although further optimization of hyper parameters could be performed by the maximum-likelihood method as reported in [53], the hyper parameters tuned in this study would be useful information for DA using real observation data.

The initial distribution of crystal grains is generated by performing the MPF simulation of grain growth with random nucleation of the initial crystal grains. The number of crystal grains (i.e. N), in the initial microstructure is 64.

In Test B, the initial fraction of $\Sigma 7$ GB in the synthetic observation data, $V_{\Sigma 7}$, is changed by using the following procedures. A GB between two adjacent grains, whose crystal orientations are $(\phi_1, \phi, \phi_2) = (95^\circ, 92^\circ, 94^\circ)$ with a near Cube texture and $(\phi_1, \phi, \phi_2) = (58^\circ, 35^\circ, 60^\circ)$ near S texture, corresponds to an exact $\Sigma 7$ GB. Therefore, the initial crystal orientations of grains are created by the following steps:

Step 1 : Set the initial orientations of N/k crystal grains, so that the initial crystal orientations satisfy $|\phi_1 - 95^\circ| \leq 4^\circ$, $|\phi - 92^\circ| \leq 4^\circ$, and $|\phi_2 - 94^\circ| \leq 4^\circ$.

Step 2 : Set the initial orientations of another N/k crystal grains, so that the initial crystal orientations satisfy $|\phi_1 - 58^\circ| \leq 4^\circ$, $|\phi - 35^\circ| \leq 4^\circ$, and $|\phi_2 - 60^\circ| \leq 4^\circ$.

Step 3 : For the other crystal grains, the initial crystal orientations are chosen to satisfy $0 \leq \phi_1 \leq 360^\circ$, $0 \leq \phi \leq 180^\circ$, and $0 \leq \phi_2 \leq 360^\circ$.

By following these steps, $V_{\Sigma 7}$ can be controlled by changing the value of k . In Test B, we determine $V_{\Sigma 7}$ as 16.1%, 21.4%, and 38.9% using $k = 8, 5$, and 3, respectively. Note that $V_{\Sigma 7}$ is calculated by dividing the number of finite difference grids where $|\Delta\theta_{ij, \langle 111 \rangle}| < 8^\circ$ by the total number of finite difference grids, N_{grids} . The SD of the perturbed observation is set to be 0.5.

Fig. 3 shows snapshots of grain growth behaviors simulated by the MPF model for different initial values of $V_{\Sigma 7}$. This figure corresponds to the visualization of three sets of synthetic observation data. The initial distributions of crystal grains are the same for all cases, but their grain growth behaviors are completely different from each other because of the different $V_{\Sigma 7}$: the rapid grain coarsening proceeds for the highest $V_{\Sigma 7} = 38.9\%$. Fig. 4 shows the variation of $V_{\Sigma 7}$ and the average grain size during the grain growth for each case. The average grain size increases with time, while the variation of $\Sigma 7$ GB fraction depends on the value of $V_{\Sigma 7}$.

5. Results and discussions

5.1. Impact of ensemble size

In Test A, we aim to investigate the impact of ensemble size, N_{ens} , on the accuracy of parameter estimation, and determine the reasonable

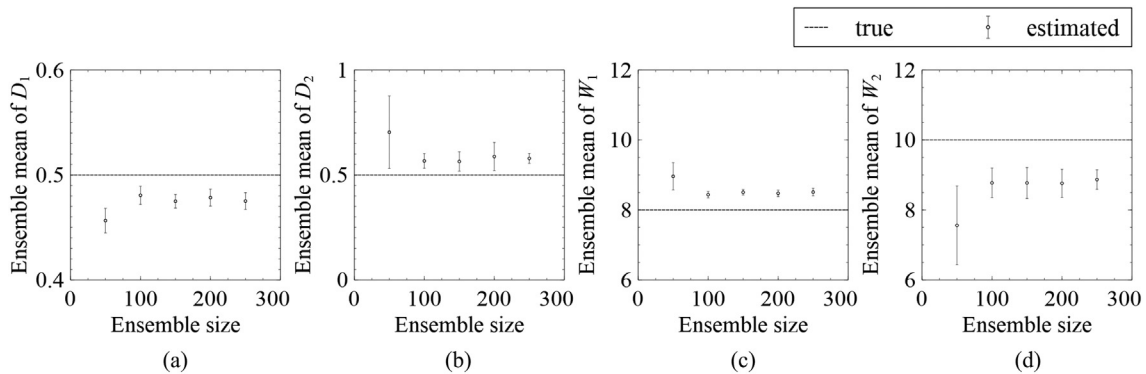


Fig. 7. Effect of ensemble size on the estimated values of parameters (a) D_1 , (b) D_2 , (c) W_1 , and (d) W_2 . The circle symbols and the error bars indicate the ensemble mean and spread of the estimated parameters, respectively. The error bars are calculated by performing five twin experiments using different random numbers.

ensemble size for Tests B and C. In order to explain the typical result obtained in Test A, we show the result of the parameter estimation using $N_{\text{ens}} = 100$.

Fig. 5 shows the time variations of the mean and the standard deviation (SD) of the estimated parameters. Here, the mean value of the parameters is evaluated by the ensemble mean as:

$$\bar{\psi}_t = \frac{1}{N_{\text{ens}}} \sum_{i=1}^{N_{\text{ens}}} \psi_t^{(i)} = [\bar{D}_1^t \quad \bar{W}_1^t \quad \bar{D}_2^t \quad \bar{W}_2^t]^T \quad (31)$$

where the overline denotes average over the ensemble members. The SD corresponds to the uncertainty of the ensemble mean of the estimated parameters. As shown in Fig. 5 (c) and (d), the initially estimated values of parameters related to the $\Sigma 7$ GB mobility peak, i.e. D_2 and W_2 ,

are smaller than the true values. It implies that the GB properties of HAGB are almost isotropic at the beginning of the simulation, so that the predicted migration of $\Sigma 7$ GB is slower than that in the true grain growth. Therefore, in the very early stage (within 25 s'), the EnKF increases the GB mobility peak by increasing D_2 and W_2 , and deepens the GB energy cusp by increasing D_1 and W_1 , so that the estimated state (polycrystalline microstructure) gets closer to the true state. These parameter modifications accelerate the migration of $\Sigma 7$ GB and the grain coarsening. In the early stage up to 100 s', as shown in Fig. 5, the mean of all estimated parameters exceed the true ones. Subsequently, since the estimated state gets closer to the true one whenever observation data is assimilated to the MPF simulation, the means of all estimated parameters decrease with time and converge to the corresponding true values in <200 s'. After the middle stage ($t > 250$), the means of estimated parameters oscillate around

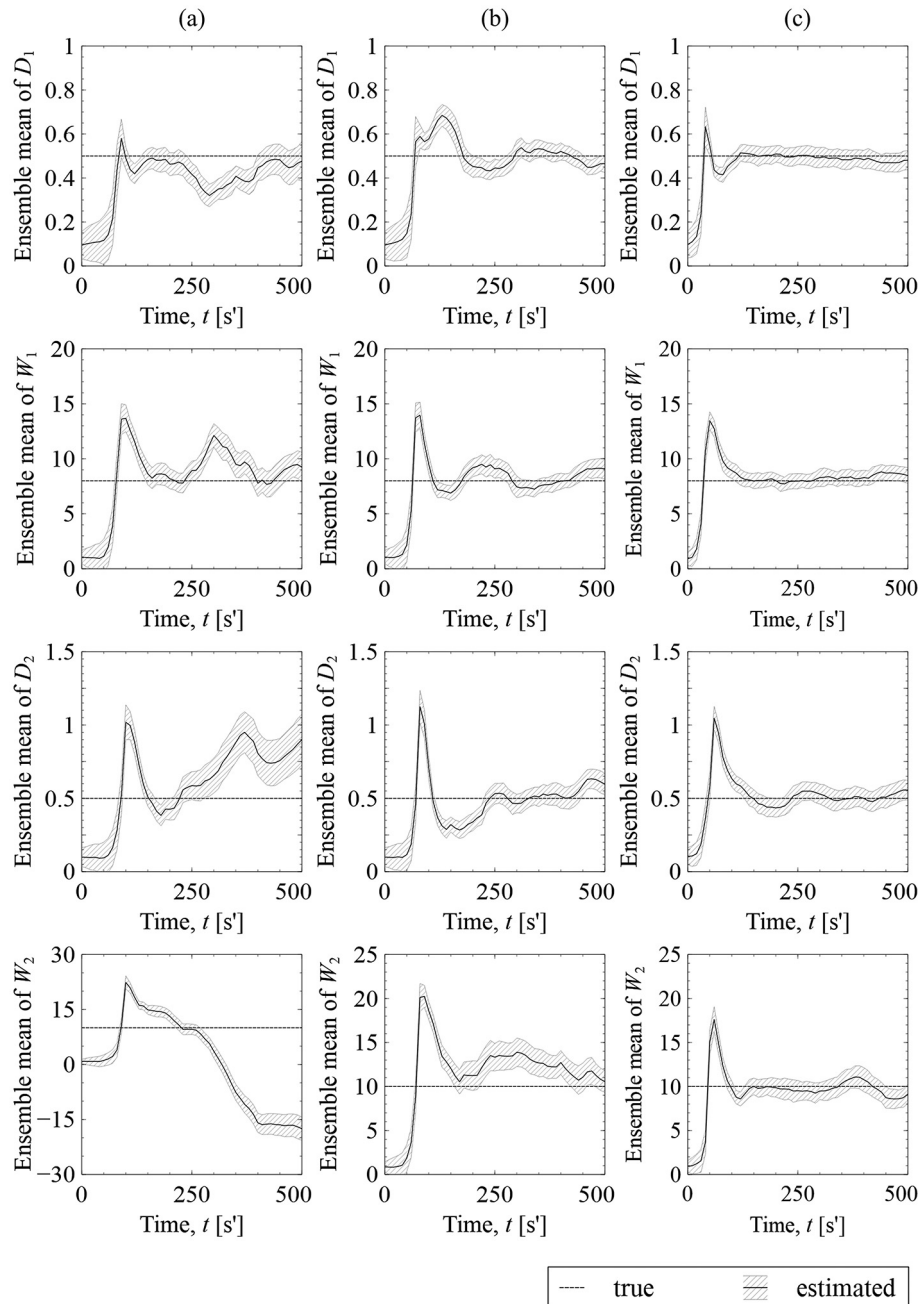


Fig. 8. Variations of the mean and SD for the estimated parameters D_1 , W_1 , D_2 , and W_2 for different initial fractions of $\Sigma 7$ grain boundary ($V_{\Sigma 7}$): (a) 16.1%, (b) 21.4%, and (c) 38.9%. Solid and dotted lines indicate the mean of the estimated parameters and the true values, respectively. The shaded region shows the SD of the estimated parameters that represents the uncertainty.

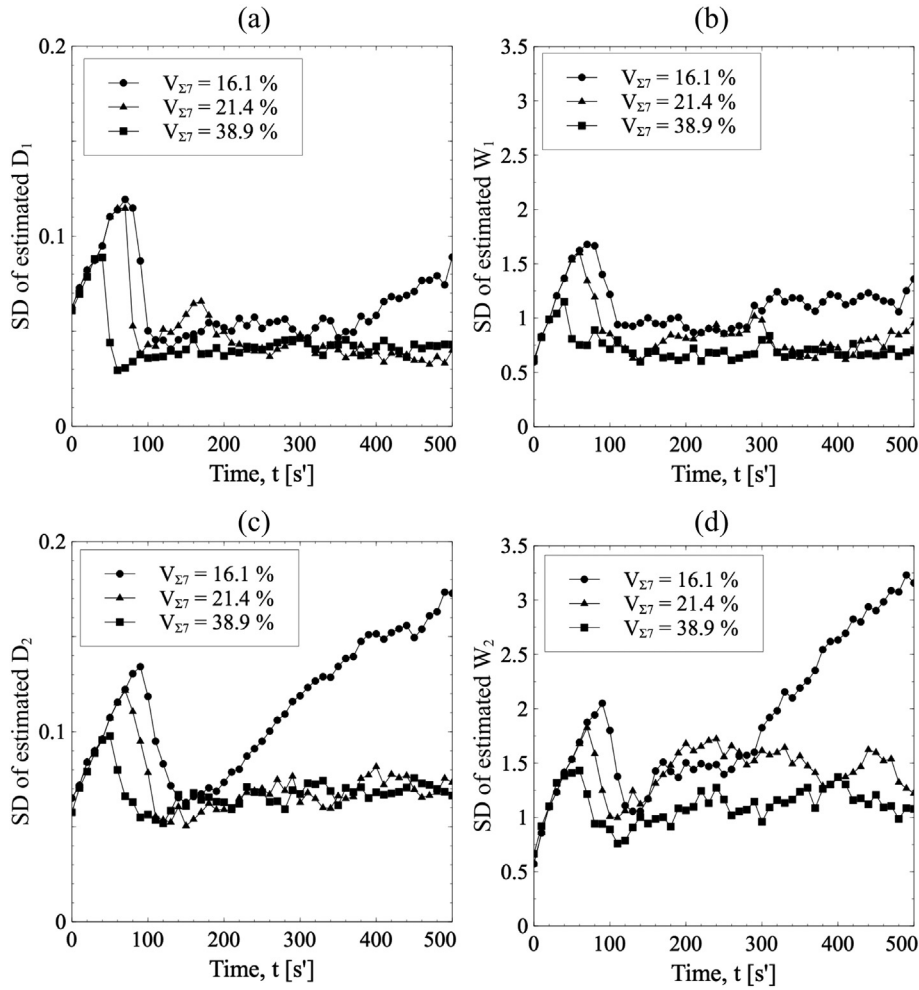


Fig. 9. Variations of SD of the estimated parameters (a) D_1 , (b) W_1 , (c) D_2 , and (d) W_2 for different initial fractions of $V_{\Sigma 7}$.

the true values. These oscillations reflect the uncertainty of the estimated state as well as the perturbed noise. We find that the estimated parameters always deviate slightly from the true values. However, the shaded area in Fig. 5 (which describes the SD of the estimated parameters) covers the true value, demonstrating that the four parameters related to the $\Sigma 7$ GB properties are successfully estimated by DA based on the EnKF.

We examine the effect of parameter estimation on the state estimation error. In order to compare the true state with the estimated state, their snapshots are shown in Fig. 6(a) and (b), respectively. In the early stage up to 50 s there is some difference between the two, because the estimated parameters deviate from the true parameters as shown in Fig. 5, as the $\Sigma 7$ GB migrates slower than that of the true state. After the parameters converge to the true ones, however, the estimated state is in

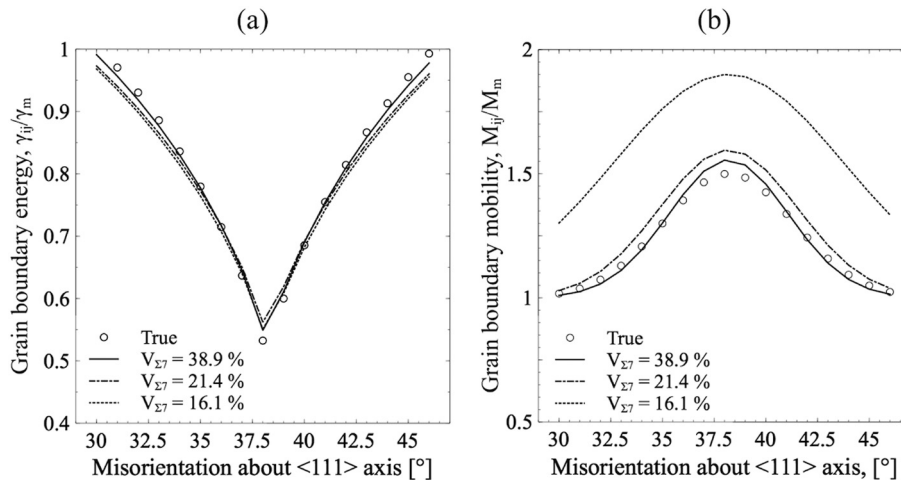


Fig. 10. Estimated GB (a) energy cusp and (b) mobility peak for different initial fractions of $V_{\Sigma 7}$. Solid curves show the estimated cusp and peaks, which are calculated by substituting the parameters estimated at $t = 500$ s into Eqs. (7) and (8).

good agreement with the true one. We evaluate the error in state estimation using the root mean squared error (RMSE) given as:

$$\text{RMSE} = \sqrt{\frac{1}{N_{\text{grid}}N} \sum_{j=1}^{N_{\text{grid}}} \sum_{i=1}^N \left(\frac{1}{N_{\text{ens}}} \sum_{k=1}^{N_{\text{ens}}} \eta_{i,j}^{t(k)} - (\eta_{i,j}^t)_{\text{true}} \right)^2} \quad (32)$$

where $(\eta_{i,j}^t)_{\text{true}}$ means the phase-field variable of the i th grain used as the synthetic observation data on the finite difference grid j at time t . A small RMSE means that the microstructure estimated by the MPF simulation with the EnKF successfully reproduces the true ones. The time variation of the RMSE is shown in Fig. 6(c). It increases in the first 50 s' when the estimated parameters deviated from the true ones. After that, the estimated parameters converge to the true values as the RMSE decreases with time, indicating that the accuracy of the state

estimation is improved as the synthetic observation data is integrated into the MPF simulation.

Before showing the results of Tests B and C, we discuss the impact of ensemble size on the accuracy of the parameter estimation and determine the reasonable ensemble size for Tests B and C. If the ensemble size is not large enough for a given parameter estimation problem, one can fail to estimate the true parameters because of the filter divergence, in which the ensemble variance significantly underestimates the true variance [46]. Otherwise, the estimated parameter tends to converge to the true value with a sufficient spread of the ensemble. Therefore, in order to examine the influence of ensemble size on the parameter estimation, we perform the twin experiments using different ensemble sizes, $N_{\text{ens}} = 50, 100, 150$, and 200 , where each twin experiment was performed for five times with different seeds used to generate random numbers. Fig. 7 shows the variations of the mean of parameters

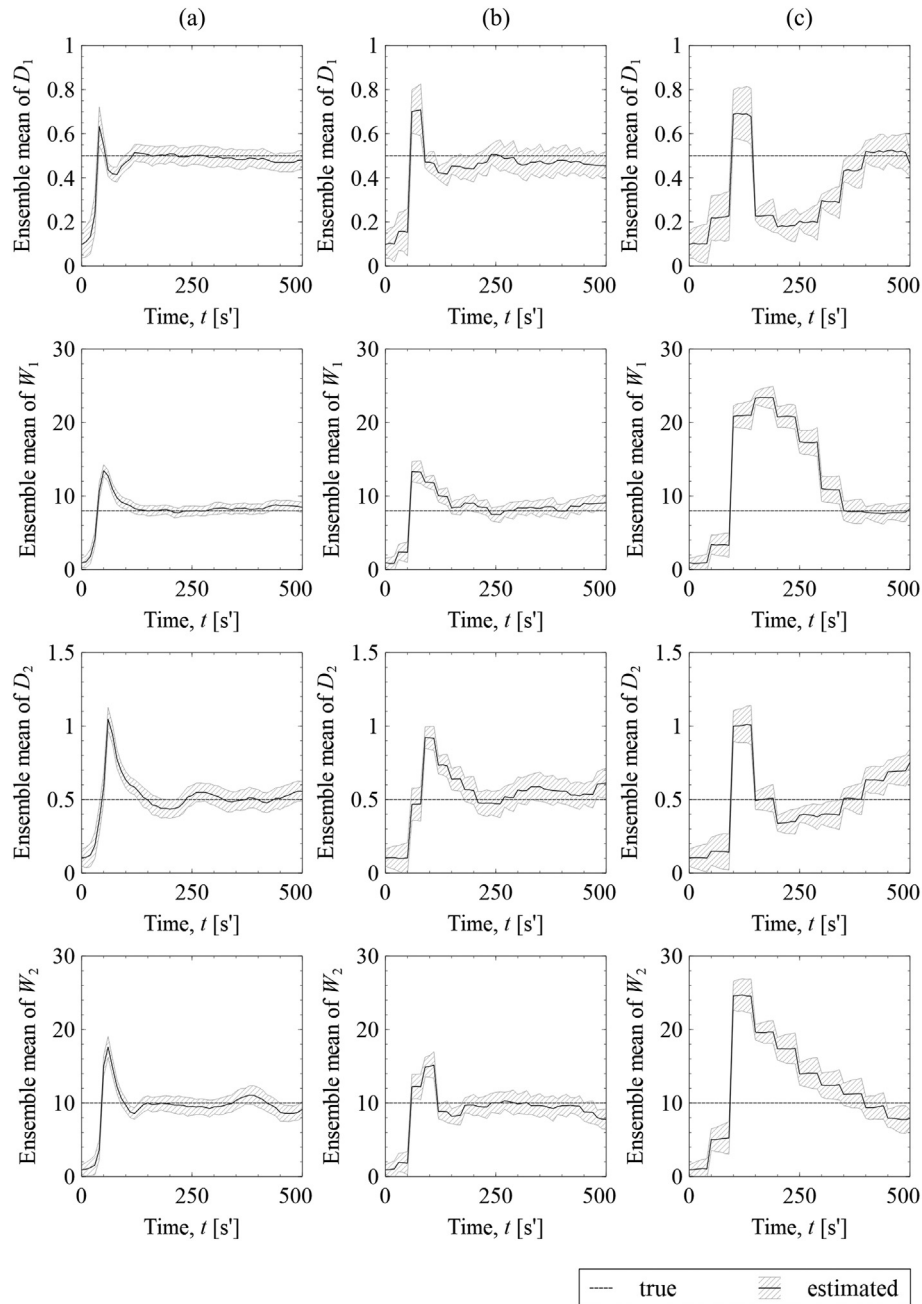


Fig. 11. Variations of the mean and SD of the estimated parameters, D_1 , W_1 , D_2 , and W_2 , for different filtering rates (Δt_{fit}): (a) 10 s', (b) 30 s', and (c) 50 s'. Solid and dotted lines indicate the mean estimated parameter and the true values, respectively. The shaded region shows the SD of the estimated parameter corresponding to its uncertainty.

estimated at the time of 500 s' for different ensemble sizes. In the case of $N_{ens} = 50$, the means of estimated parameters deviate more from the true ones compared to the other cases with larger ensemble sizes. As the ensemble size increases, the ensemble means of the estimated parameter get closer to their true values and those spreads decrease. According to the results shown in Fig. 7, we determined that $N_{ens} > 50$ is required to accurately estimate the parameters in this study. As a result, we used $N_{ens} = 100$ for Tests B and C.

5.2. Effect of $\Sigma 7$ grain boundary fraction

In the twin experiment named as Test B, we investigate the influence of the fraction of targeted GB, i.e. $\Sigma 7$ GB in this study, included in the observation data on the accuracy of the parameter estimation. This twin experiment would be a benchmark to indicate how much of the targeted GB should be included in the experimental data when we estimate poorly known parameters from a real experimental dataset by the EnKF-based DA method.

In Fig. 8, we compare the time variation of the mean and the SD of the parameters estimated using different initial fractions of $\Sigma 7$ GB, i.e. $V_{\Sigma 7}$. As shown in Fig. 8(a), for the smallest $V_{\Sigma 7} = 16.1\%$, the estimated parameters related to the GB energy cusp, i.e. D_1 and W_1 , approach the true values as the observation data is assimilated, but the parameters related to the GB mobility peak, D_2 and W_2 , gradually deviate from the true values after the time of 200 s'. On the other hand, as shown in Fig. 8(b) and (c), where a larger fraction of $\Sigma 7$ GB is included in the observation data than that in Fig. 8(a), we can get better estimates of all four parameters with reasonable accuracy.

In order to evaluate the estimation errors corresponding to the uncertainty of the estimated parameters when using different $V_{\Sigma 7}$ values, Fig. 9 displays the time-varying SD of the estimated parameters that are shown by the shaded area in Fig. 8. For $V_{\Sigma 7} = 16.1\%$, the SD with respect to all four parameters increases with time after 150 s', indicating that the parameter estimation would have large errors if the observation data contains only a small amount of the target GB. In the other cases ($V_{\Sigma 7} = 21.4\%$ and 38.9%), the SD keeps almost constant after the time of 100 s', but is smaller for $V_{\Sigma 7} = 38.9\%$ than for $V_{\Sigma 7} = 21.4\%$. Although the specific value of $V_{\Sigma 7}$ cannot be determined, these results demonstrate that the EnKF allows us to estimate the parameters with a low uncertainty, if we assimilate the observation data containing a sufficient amount of $\Sigma 7$ GB.

Fig. 10 shows the GB energy cusp and the GB mobility peak around $\Sigma 7$ GB calculated using the mean values of the estimated parameters at the time of 500 s'. As shown in Figs. 8 and 9, the estimation accuracy of the GB energy cusp and mobility peak can be improved dynamically

as $V_{\Sigma 7}$ increases. The estimated GB energy cusp and mobility peak for $V_{\Sigma 7} = 38.9\%$ show very good agreement with the true GB properties, but these properties are poorly estimated for the smaller $V_{\Sigma 7}$. Therefore, from Test B we can determine the critical fraction of the targeted GB required to adequately estimate the parameters using real observation data by performing designed twin experiments.

5.3. Effect of filtering interval

In Test C, we investigate the influence of filtering interval, which is the time interval between the assimilations of the observation data, on the parameter estimation. Fig. 11 shows the time variations of the mean and SD of the filtered PDF related to the parameters for different filtering intervals: $\Delta t_{filt} = 10, 30$, and 50 s'. It is found that the time required to estimate the true values of the parameters increases with increasing the filtering interval. Therefore, the more observation data assimilated into the MPF model, the more rapidly the accurate parameters can be estimated. Nevertheless, comparing the results shown in Fig. 11(a) with those in Fig. 11(b), we find that the accuracy of the estimated parameters is not degraded very much when Δt_{filt} increases from 10 s' to 30 s'. As shown in Fig. 11(c), even if we use the longest filtering interval, the estimated parameters converge to the true values at around 500 s', but the uncertainty of the estimated parameters becomes the largest.

Finally, Fig. 12 shows the shapes of GB energy cusp and GB mobility peak calculated using the mean of the estimated parameters at 500 s'. Similar to results obtained in Test B, the GB energy cusp can be estimated more accurately for all the filtering intervals, compared to the estimation of the GB mobility peak. On the other hand, although the width of the GB mobility peak is successfully estimated, its height is overestimated as the filtering interval increases. Nevertheless, even at the long filtering interval, the GB mobility peak may be estimated by assimilating a larger amount of observation data into the MPF model.

6. Conclusion

This paper presents the first application of the DA methodology based on the EnKF to a 3D MPF model, in order to improve the prediction of system state (i.e. the microstructure) and to estimate unknown parameters included in the MPF model. We performed numerical experiments, called the twin experiments, where the anisotropic $\Sigma 7$ GB energy cusp and mobility peak were estimated simultaneously by assimilating the synthetic time-evolving polycrystalline microstructure data to the MPF simulation of grain growth. Through the twin experiments, we have demonstrated that the EnKF-based DA method

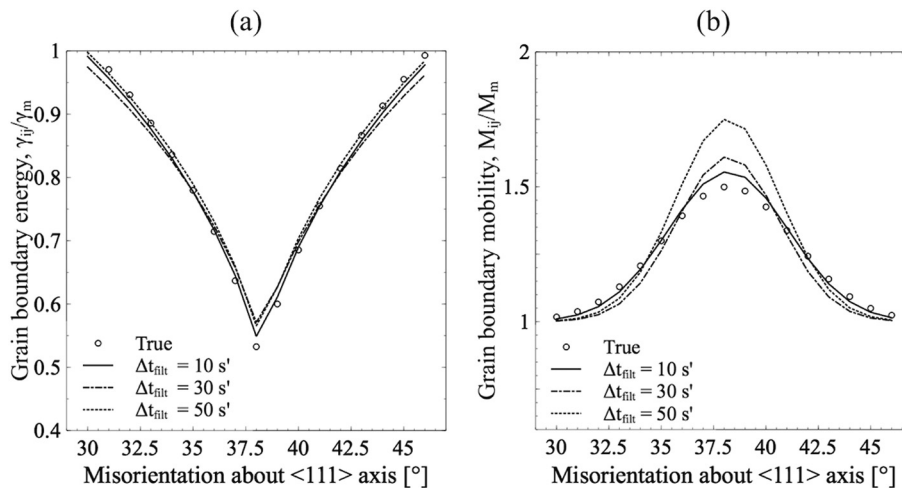


Fig. 12. Estimated cusps of the GB (a) energy and (b) mobility for different filtering intervals. Open circles correspond to the true values, and solid lines show the estimations calculated by Eqs. (7) and (8) with the mean estimated parameter at $t = 500$ s'.

successfully and simultaneously estimated the four parameters that characterize the $\Sigma 7$ GB energy cusp and mobility peak. It was also clearly shown that the estimation accuracy was improved by increasing the amount of observation data assimilated to the MPF simulation. Furthermore, the shapes of the GB energy cusp and mobility peak of $\Sigma 7$ GB were successfully reproduced using the estimated parameters.

Although we have shown the results of twin experiments using only synthetic observation data, the EnKF-DA method proposed in this paper can be directly applied to estimate anisotropic GB properties using real experimental data. The results obtained by the twin experiments indicate great potential of the EnKF-DA methodology for state estimation (polycrystalline microstructure) and parameter estimation using real experimental data. An interesting topic in future works will be DA calculation using experimental data obtained with state-of-art 3D/4D microstructure observation techniques, such as 3DXRD and diffraction contrast tomography (DCT).

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.matdes.2018.107577>.

CRedit authorship contribution statement

Akinori Yamanaka: Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Writing - original draft, Writing - review & editing. **Yuri Maeda:** Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Writing - original draft, Writing - review & editing. **Kengo Sasaki:** Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Writing - original draft, Writing - review & editing.

Acknowledgement

The authors thank Professor Hiromichi Nagao and Dr. Shin-ichi Ito at Earthquake Research Institute, The University of Tokyo for helpful discussions on data assimilation based on the EnKF. This work was supported by the Council for Science, Technology and Innovation (CSTI), Cross-ministerial Strategic Innovation Promotion Program (SIP), "Structural Materials for Innovation" (Funding agency: JST, Japan).

References

- [1] D. Zöllner, P. Streitenberger, Three-dimensional normal grain growth: Monte Carlo Potts model simulation and analytical mean field theory, *Scr. Mater.* 54 (2006) 1697–1702.
- [2] D.J. Srolovitz, M.P. Anderson, P.S. Sahni, G.S. Grest, Computer simulation of grain growth-II. Grain size distribution, topology and local dynamics, *Acta Metall.* 32 (1984) 793–802.
- [3] D.J. Srolovitz, M.P. Anderson, G.S. Grest, P.S. Sahni, Computer simulation of grain growth-II. Influence of a particle dispersion, *Acta Metall.* 32 (1984) 1429–1438.
- [4] A.D. Tutcuoglu, A. Vidyasagar, K. Bhattacharya, D.M. Kochmann, Stochastic modeling of discontinuous dynamic recrystallization at finite strains in hcp metals, *J. Mech. Phys. Solids* 122 (2019) 590–612.
- [5] M. Kühbach, G. Gottstein, L.A. Barrales-Mora, A statistical ensemble cellular automaton microstructure model for primary recrystallization, *Acta Mater.* 107 (2016) 366–367.
- [6] L. Wang, G. Fang, L. Qian, Modeling of dynamic recrystallization of magnesium alloy using cellular automata considering initial topology of grains, *Mater. Sci. Eng. A* 711 (2018) 268–283.
- [7] R. Kobayashi, Modeling and numerical simulations of dendritic crystal growth, *Physica D* 63 (1993) 410–423.
- [8] A.A. Wheeler, W.J. Boettinger, G.B. McFadden, Phase-field model for isothermal phase transitions in binary alloys, *Phys. Rev. A* 45 (1992) 7424.
- [9] A. Karma, W.J. Rappel, Quantitative phase-field modeling of dendritic growth in two and three dimensions, *Phys. Rev. E* 57 (1998) 4323.
- [10] D. Fan, L.Q. Chen, Computer simulation of grain growth using a continuum field model, *Acta Mater.* 45 (1997) 611–622.
- [11] I. Steinbach, F. Pezzolla, A generalized field method for multiphase transformations using interface fields, *Physica D* 134 (1999) 385–393.
- [12] N. Moelans, A quantitative and thermodynamically consistent phase-field interpolation function for multi-phase systems, *Acta Mater.* 59 (2011) 1077–1086.
- [13] E. Miyoshi, T. Takaki, M. Ohno, Y. Shibuta, S. Sakane, T. Shimokawabe, Ultra-large-scale phase-field simulation study of ideal grain growth, *npj Comput. Mater.* 3 (2017) 25.
- [14] F.J. Humphreys, M. Hartherly, *Recrystallization and Related Annealing Phenomena*, Pergamon Press, 2004.
- [15] J.R. Gatti, P.P. Bhattacharjee, Nucleation behavior and formation of recrystallization texture in pre-recovery treated heavily cold and warm-rolled Al-2.5 wt.% Mg alloy, *Mater. Charact.* 106 (2015) 141–151.
- [16] P.D. Wu, S.R. MacEwen, D.J. Lloyd, K.W. Neale, Effect of cube texture on sheet metal formability, *Mater. Sci. Eng. A* 364 (2004) 182–187.
- [17] M.M. Miszczyk, H. Paul, J.H. Driver, J. Poplewska, The influence of deformation texture on nucleation and growth of cube grains during primary recrystallization of AA1050 alloy, *Acta Mater.* 129 (2017) 378–387.
- [18] Y. Shibuta, S. Takamoto, T. Suzuki, A molecular dynamics study of the energy and structure of the symmetric tilt boundary of iron, *ISIJ Int.* 48 (2008) 1582–1591.
- [19] F. Ulomek, V. Mohles, Molecular dynamics simulations of grain boundary mobility in Al, Cu and γ -Fe using a symmetric driving force, *Model. Simul. Mater. Sci. Eng.* 22 (2014), 055011.
- [20] Q. Zhu, A. Samanta, B. Li, R.E. Rudd, T. Frolov, Predicting phase behavior of grain boundaries with evolutionary search and machine learning, *Nat. Commun.* 9 (2018) 467.
- [21] K.G.F. Janssens, D. Olmsted, A. Holm, S.M. Foils, S.J. Pilmpton, P.M. Derlet, Computing the mobility of grain boundaries, *Nat. Mater.* 5 (2006) 124–127.
- [22] J.J. Hoyt, Atomistic simulations of grain and interphase boundary mobility, *Model. Simul. Mater. Sci. Eng.* 22 (2014), 033001.
- [23] G. Gottstein, L.S. Shvindlerman, *Grain Boundary Migration in Metals*, CRC Press, 1999.
- [24] Y. Huang, F.J. Humphreys, Measurements of grain boundary mobility during recrystallization of a single-phase aluminum alloy, *Acta Mater.* 47 (1999) 2259–2268.
- [25] I. Basu, M. Chen, M. Loeck, T. Al-Samman, D.A. Molodov, Determination of grain boundary mobility during recrystallization by statistical evaluation of electron backscatter diffraction measurements, *Mater. Charact.* 117 (2016) 99–112.
- [26] S. Schmidt, U.L. Olsen, H.F. Poulsen, H.O. Sørensen, E.M. Lauridsen, L. Margulies, C. Maurice, D.J. Jensen, Direct observation of 3-D grain growth in Al-0.1%Mn, *Scr. Mater.* 59 (2008) 491–494.
- [27] E.M. Lauridsen, S. Schmidt, S.F. Nielsen, L. Margulies, H.F. Poulsen, D. Juul Jensen, Non-destructive characterization of recrystallization kinetics using three-dimensional X-ray diffraction microscopy, *Scr. Mater.* 55 (2006) 51–56.
- [28] B. Lin, Y. Jin, C.M. Hefferan, S.F. Li, J. Lind, R.M. Suter, M. Bernacki, N. Bozzolo, A.D. Rollet, G.S. Rohrer, Observation of annealing twin nucleation at triple lines in nickel during grain growth, *Acta Mater.* 99 (2015) 63–68.
- [29] C.M. Hefferan, J. Lind, S.F. Li, U. Lienert, A.D. Rollet, R.M. Suter, Observation of recovery and recrystallization in high-purity aluminum measured with forward modeling analysis of high-energy diffraction microscopy, *Acta Mater.* 60 (2012) 4311–4318.
- [30] M. Syha, W. Rheinheimer, M. Bäurer, E.M. Lauridsen, W. Ludwig, D. Weygand, P. Gumbsch, Three-dimensional grain structure of sintered bulk strontium titanate from X-ray diffraction contrast tomography, *Scr. Mater.* 66 (2012) 1–4.
- [31] W. Ludwig, A. King, P. Reischig, M. Herbig, E.M. Lauridsen, S. Schmidt, H. Proudhon, S. Forest, P. Cloetens, S. Rolland du Roscoat, J.Y. Buffiere, T.J. Marrow, H.F. Poulsen, New opportunities for 3D materials science of polycrystalline materials at the micrometer lengthscale by combined use of X-ray diffraction and X-ray imaging, *Mater. Sci. Eng. A* 524 (2009) 69–76.
- [32] I.M. McKenna, S.O. Poulsen, E.M. Lauridsen, W. Ludwig, P.W. Voorhees, Grain growth in four dimensions: a comparison between simulation and experiment, *Acta Mater.* 78 (2014) 125–134.
- [33] L.K. Aagesen, J.L. Fife, E.M. Lauridsen, P.W. Voorhees, The evolution of interfacial morphology during coarsening: a comparison between 4D experiments and phase-field simulations, *Scr. Mater.* 64 (2011) 394–397.
- [34] P. Steinmetz, Y.C. Yabansu, J. Hotzer, M. Jainta, B. Nestler, S.R. Kalidindi, Analytics for microstructure datasets produced by phase-field simulations, *Acta Mater.* 103 (2016) 192–203.
- [35] J. Zhang, S.O. Poulsen, J.W. Gibbs, P.W. Voorhees, H.F. Poulsen, Determining material parameters using phase-field simulations and experiments, *Acta Mater.* 129 (2017) 229–238.
- [36] J. Zhang, PhD Thesis, Technical University of Denmark, 2018.
- [37] S. Ito, H. Nagao, A. Yamanaka, Y. Tsukada, T. Koyama, M. Kano, J. Inoue, Data assimilation for massive autonomous systems based on a second-order adjoint method, *Phys. Rev. E* 94 (2016), 043307.
- [38] K. Sasaki, A. Yamanaka, S. Ito, H. Nagao, Data assimilation for phase-field models based on the ensemble Kalman filter, *Comput. Mater. Sci.* 141 (2018) 141–152.
- [39] S. Ito, H. Nagao, T. Kasoya, J. Inoue, Grain growth prediction based on data assimilation by implementing 4DVar on multi-phase-field model, *Sci. Technol. Adv. Mater.* 18 (2017) 857–869.
- [40] P.J. van Leeuwen, Y. Cheng, S. Reich, *Nonlinear Data Assimilation*, Springer, 2015.
- [41] W. Lahoz, B. Khattatov, R. Ménard, *Data Assimilation, Making Sense of Observations*, Springer, 2010.
- [42] M. Ghill, P.M. Rizzoli, *Data Assimilation in meteorology and oceanography*, *Adv. Geophys.* 33 (1991) 141–266.
- [43] A. Doucet, S. Godsill, C. Andrieu, On sequential Monte Carlo sampling methods for Bayesian filtering, *Stat. Comput.* 10 (2000) 197–208.
- [44] G. Evensen, Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, *J. Geophys. Res.* 99 (1994), 10143.
- [45] G. Evensen, The ensemble Kalman filter: theoretical formulation and practical implementation, *Ocean Dyn.* 53 (2003) 343–367.
- [46] G. Evensen, *Data Assimilation, The Ensemble Kalman Filter*, 2nd edition Springer, 2009.
- [47] N. Moelans, B. Blanpain, P. Wollants, Quantitative analysis of grain boundary properties in a generalized phase field model for grain growth in anisotropic system, *Phys. Rev. B* 78 (2008), 024113.

- [48] H. Zhang, M.I. Mendelev, D.J. Srolovitz, Mobility of $\Sigma 5$ tilt grain boundaries: inclination dependence, *Scr. Mater.* 52 (2005) 1193–1198.
- [49] A.P. Sutton, R.W. Balluffi, *Interfaces in Crystalline Materials*, Oxford University Press, 1995.
- [50] W.T. Read, W. Shockley, Dislocation models of crystal grain boundaries, *Phys. Rev.* 78 (1950) 275–289.
- [51] F.J. Humphreys, A unified theory of recovery, recrystallization and grain growth, based on the stability and growth of cellular microstructures-I. The basic model, *Acta Mater.* 45 (1997) 4231–4240.
- [53] A. Nakabayashi, G. Ueno, An extension of the ensemble Kalman filter for estimating the observation error covariance matrix based on the variational Bayes's method, *Mon. Weather Rev.* 145 (2017) 199–213.