

OPEN

A universal Bayesian inference framework for complicated creep constitutive equations

Yoh-ichi Mototake¹, Hitoshi Izuno², Kenji Nagata², Masahiko Demura²✉ & Masato Okada³

Evaluating the creep deformation process of heat-resistant steels is important for improving the energy efficiency of power plants by increasing the operating temperature. There is an analysis framework that estimates the rupture time of this process by regressing the strain–time relationship of the creep process using a regression model called the creep constitutive equation. Because many creep constitutive equations have been proposed, it is important to construct a framework to determine which one is best for the creep processes of different steel types at various temperatures and stresses. A Bayesian model selection framework is one of the best frameworks for evaluating the constitutive equations. In previous studies, approximate-expression methods such as the Laplace approximation were used to develop the Bayesian model selection frameworks for creep. Such frameworks are not applicable to creep constitutive equations or data that violate the assumption of the approximation. In this study, we propose a universal Bayesian model selection framework for creep that is applicable to the evaluation of various types of creep constitutive equations. Using the replica exchange Monte Carlo method, we develop a Bayesian model selection framework for creep without an approximate-expression method. To assess the effectiveness of the proposed framework, we applied it to the evaluation of a creep constitutive equation called the Kimura model, which is difficult to evaluate by existing frameworks. Through a model evaluation using the creep measurement data of Grade 91 steel, we confirmed that our proposed framework gives a more reasonable evaluation of the Kimura model than existing frameworks. Investigating the posterior distribution obtained by the proposed framework, we also found a model candidate that could improve the Kimura model.

Evaluating the creep deformation process of heat-resistant steels is important for improving the energy efficiency of power plants by increasing the operating temperature^{1,2}. One of the most effective frameworks for estimating this process is the creep constitutive equation approach. On the basis of the regression result of the strain–time relationship using a regression model called the creep constitutive equation^{1–9}, the approach predicts the transition of a creep deformation process obeying the alternation of temperature and stress. It has been reported that the creep constitutive equation approach can predict the quantities required to evaluate creep phenomena, such as rupture time, with high accuracy. Because many creep constitutive equations have been proposed, it is important to construct a framework to determine which one is best for the creep processes of different steel types at various temperatures and stresses. Hereinafter, this problem is called creep model selection for simplicity.

Bayesian model selection is one of the best frameworks for evaluating such creep constitutive equations because it can select the model (=equation) that minimizes the fitting error while avoiding model complexity¹⁰ (see Appendix A). Evaluating the model on the basis of only the fitting error tends to result in the selection of a more complex model, which leads to overfitting¹⁰. Overfitting the creep measurement data under specific temperature and stress conditions would make it difficult to model the transition of a creep deformation process obeying the alternation of temperature and stress. Bayesian model selection has another advantage, that is, the model parameters are estimated as probabilistic variables rather than point values. The maximum-likelihood methods, such as the least-squares method, estimate model parameters as point values. The probabilistic treatment of parameters introduces both prior and posterior distributions of parameters in the model selection framework. The prior distributions model the prior knowledge of parameters. For example, we can add knowledge, such as

¹The Institute of Statistical Mathematics, Tachikawa, Tokyo, 190-8562, Japan. ²Research and Services Division of Materials Data and Integrated System, National Institute for Materials Science, Namiki 1-1, Tsukuba, Ibaraki, 305-0044, Japan. ³Graduate School of Frontier Sciences, University of Tokyo, Kashiwa, Chiba, 277-8561, Japan. ✉e-mail: DEMURA.Masahiko@nims.go.jp

the range of a parameter, into the model. The posterior distributions are the estimation results of parameters based on measurement data. From the posterior distributions, we can obtain the estimation reliability of the parameters or obtain the properties of the model, such as the correlations among the parameters. These cannot be obtained by a maximum-likelihood method. Such knowledge helps to design a measurement plan or comprehend the model^{11,12}.

In a previous study¹³, we applied a Bayesian model selection framework for the evaluation of creep constitutive equations constructed from linear sums of basis functions. We computed two simple creep constitutive equations with and without a steady-state term, and found that the equation with a steady-state term was selected over a wide temperature and stress range for Grade 91 (Gr.91) steel¹⁴, a high-Cr, ferritic, heat-resistant steel. For a more accurate prediction of creep phenomena, it is necessary to apply a more sophisticated constitutive equation. There are several constitutive equations that have been proposed so far. Recently, Kimura *et al.* have proposed a novel constitutive equation^{1,2} and showed that it reproduces the experimental creep curve very well. The creep constitutive equation is formulated as

$$\varepsilon(t; \theta_{\text{kimura}}) = \varepsilon_0 + a_1 t^{b_1} + a_2 t^{b_2} + c_1 \exp(d_1 t) + c_2 \exp(d_2 t), \quad (1)$$

where ε is the strain, t is the time, and θ_{kimura} is the parameter set $\{\varepsilon_0, a_1, a_2, b_1, b_2, c_1, c_2, d_1, d_2\}$ of the creep constitutive equation. We refer to this creep constitutive equation as the Kimura model. Thus, the Kimura model consists of a linear sum of the four basis functions that depend on time t . This is one of the creep constitutive equations with the largest number of parameters among the equations that have been proposed so far^{15,16}. Our previous framework is not applicable to such sophisticated or complicated creep constitutive equations. This is because the parameters of the exponential part of each basis function are not treated as probabilistic variables in the previous framework: the parameters are optimized by a grid search and point estimation based on the empirical Bayes method¹⁰. If the number of grids per axis is G and the number of parameters is d , the computational cost to execute Bayesian model selection is G^d . Therefore, it is not realistic to apply the previous framework to a model with many basis functions, such as the Kimura model. Treating the parameters of a basis function as deterministic parameters is an approximation-expression method in which the posterior distribution of the parameters is a delta function, which is called as the empirical Bayes approximation. If the parameters of the basis function are not well-determined, the empirical Bayes approximation generates a bias in the model selection result. The exchange symmetry of parameter pairs, such as $a_1 - a_2, b_1 - b_2, c_1 - c_2$, and $d_1 - d_2$ in the Kimura model, is an example of parameters that are not well-determined. Keitel *et al.* also applied a Bayesian model selection framework for the model selection of creep constitutive equations of concrete¹⁷, where they approximated the posterior distribution of parameters as a Gaussian. Such the approximate-expression method is called the Laplace approximation. The Laplace approximation also leads to a false conclusion if the model parameters are not well-determined¹⁰. To evaluate the complex creep constitutive equations, it is necessary to perform Bayesian creep model selection without an approximate-expression method. There is another difficulty in Bayesian creep model selection. In the measurement of the creep deformation process in steel, it is difficult to estimate the measurement noise intensity correctly; instead, a noise range is given. In creep model selection, it is important to set the noise intensity correctly, because the noise intensity is related to a criterion that determines whether the signal is to be regressed or considered as noise. To set the noise intensity as a range, it is necessary to set the noise intensity as a probabilistic variable and set its range as a prior distribution. However, in the existing creep model selection frameworks, it is difficult to treat the noise intensity as a probabilistic value. Thus, there is no Bayesian creep model selection framework that can be used to evaluate all types of creep constitutive equations correctly.

In this study, using the replica exchange Monte Carlo method¹⁸, we propose a universal Bayesian creep model selection framework that can be applied to various types of creep constitutive equations without the approximate-expression method or the ability to set the range of the measurement noise intensity as a prior distribution. By applying the proposed framework to the evaluation of the Kimura model using the creep measurement data of Gr.91 steel¹⁴, we confirmed that our proposed framework gives a more reasonable evaluation of the Kimura model than existing frameworks. From the accurate posterior distribution obtained by the proposed framework, we also found a way to simplify the Kimura model without losing the model likelihood.

Methods

In the proposed framework, the criterion to achieve a Bayesian creep model selection is obtained by numerical integration. The numerical integration consists of a sequential run of numerical integration using the REMC method and the Riemann sum. The flowchart of the framework is shown in Fig. 1.

Bayesian creep model selection. We evaluate K creep constitutive equations M_k in terms of their ability to represent the creep deformation data $\mathbf{D} = \{\mathbf{t}, \varepsilon\} = \{(t_1, t_2, \dots, t_N), (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_N)\}$, where ε is the strain and t is the time. The likelihood of a given creep constitutive equation $M_k (k = 1, \dots, K)$ for the creep deformation data \mathbf{D} is

$$P(M_k|\mathbf{D}) = \frac{P(\mathbf{D}|M_k)P(M_k)}{P(\mathbf{D})} \propto P(\mathbf{D}|M_k)P(M_k), \quad (2)$$

where $P(\mathbf{D})$ is a normalization constant. In this study, we assume that there is no prior knowledge about the likelihood of the model. Thus, we set the prior probability $P(M_k)$ as a uniform distribution; in this study, it is equal to $\frac{1}{K}$. We also assume that \mathbf{t} in the dataset $\mathbf{D} = \{\mathbf{t}, \varepsilon\}$ is given deterministically, that is, non-probabilistically. Then, on the basis of Bayes' theorem, the likelihood of the model is transformed as

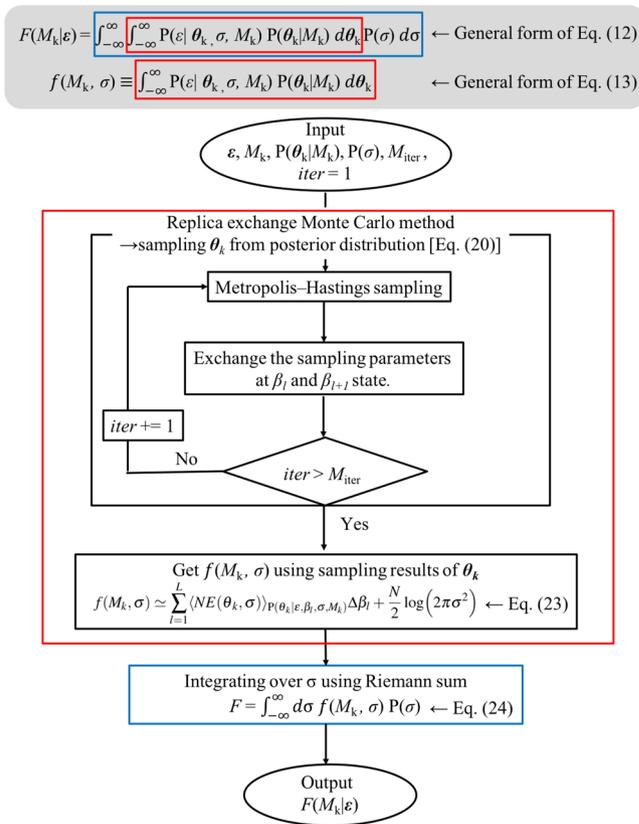


Figure 1. Flowchart of proposed framework.

$$P(M_k|\mathbf{D}) = P(M_k|\varepsilon, \mathbf{t}) = P(M_k|\varepsilon) \propto P(\varepsilon|M_k) \tag{3}$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P(\varepsilon|\theta_k, \sigma, M_k) P(\theta_k|M_k) P(\sigma) d\theta_k d\sigma, \tag{4}$$

where θ_k is the parameter set of the creep constitutive equation M_k and σ is the noise intensity.

The conditional probability $P(\varepsilon|\theta_k, \sigma, M_k)$ of Eq. (4) is a stochastic generative model of the creep constitutive equation M_k . When the measurement noise of the creep deformation data is given as an independent and identically distributed Gaussian with average 0 and standard deviation σ , the conditional probability can be expressed as

$$P(\varepsilon|\theta_k, \sigma, M_k) = \prod_{i=1}^N P(\varepsilon_i|\theta_k, \sigma, M_k) \tag{5}$$

$$= \left(\frac{1}{2\pi\sigma^2}\right)^{N/2} \prod_{i=1}^N \exp\left[-\frac{1}{2\sigma^2}(\varepsilon_i - \varepsilon_k(t_i; \theta_k))^2\right] \tag{6}$$

$$= \left(\frac{1}{2\pi\sigma^2}\right)^{N/2} \exp\left\{-\sum_{i=1}^N \left[\frac{1}{2\sigma^2}(\varepsilon_i - \varepsilon_k(t_i; \theta_k))^2\right]\right\}, \tag{7}$$

where $\varepsilon_k(t_i; \theta_k)$ is the regression function of the creep constitutive equation M_k as described in the introduction. The probabilities $P(\theta_k|M_k)$ and $P(\sigma)$ in Eq. (4) respectively simulate the prior knowledge about the model parameters θ_k and the noise intensity σ as probability distributions. By substituting Eq. (7) into Eq. (4), we obtain

$$P(\varepsilon|M_k) = \int_{-\infty}^{\infty} d\sigma \left(\frac{1}{2\pi\sigma^2} \right)^{N/2} \int_{-\infty}^{\infty} d\theta_k \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^N (\varepsilon_i - \varepsilon_k(t_i; \theta_k))^2 \right\} P(\theta_k|M_k) P(\sigma) \quad (8)$$

$$= \int_{-\infty}^{\infty} d\sigma \left(\frac{1}{2\pi\sigma^2} \right)^{N/2} \int_{-\infty}^{\infty} d\theta_k \exp[-NE(\theta_k, \sigma)] P(\theta_k|M_k) P(\sigma) \quad (9)$$

$$= \int_{-\infty}^{\infty} d\sigma \exp[-f(M_k, \sigma)] P(\sigma), \quad (10)$$

where

$$E(\theta_k, \sigma) = \frac{1}{2N\sigma^2} \sum_{i=1}^N (\varepsilon_i - \varepsilon_k(t_i; \theta_k))^2, \quad (11)$$

$$f(M_k, \sigma) = -\log \left\{ \left(\frac{1}{2\pi\sigma^2} \right)^{N/2} \int_{-\infty}^{\infty} d\theta_k \exp[-NE(\theta_k, \sigma)] P(\theta_k|M_k) \right\}. \quad (12)$$

The probability $P(\varepsilon|M_k)$ is often referred to as the marginal likelihood and is proportional to the likelihood of the recognition model M_k . The negative log-likelihood

$$F(M_k) = -\log P(\varepsilon|M_k) \quad (13)$$

is often referred to as the Bayesian free energy. In this way, $F(M_k)$ is proportional to the negative log-likelihood of the recognition model M_k . Therefore, the creep constitutive equation M_k with the smallest $F(M_k)$ value represents the best model.

Replica exchange Monte Carlo sampling method. To obtain the value of $F(M_k)$, we need to execute the integration in Eq. (4). However, it is difficult to analytically execute the integration owing to the complicated relationship between $\varepsilon_k(t_i; \theta_k)$ and θ_k . We overcame this difficulty by numerical integration. The numerical integration was performed in two steps. The first step is integration with respect to the model parameter set θ_k to obtain the value of $f(M_k, \sigma)$ with σ [Eq. (12)], and the second step is, by using the obtained $f(M_k, \sigma)$ value, integration with respect to the noise intensity σ [Eq. (10)] to obtain $F(M_k)$. Since step 2 is a one-variable integral with respect to σ [Eq. (10)], the numerical integration can be performed as a Riemann sum. On the other hand, since the θ_k integral is high-dimensional, we carried out the integration by the sampling method. In this section, we explain how to calculate $f(M_k, \sigma)$ with a given noise intensity σ by the sampling method.

Markov chain Monte Carlo (MCMC) methods¹⁹ are efficient sampling methods for estimating the expectation value of a probability distribution in a high-dimensional space. $f(M_k, \sigma)$ is given using an auxiliary variable β ;

$$f(M_k, \sigma) = -\log \int_{-\infty}^{\infty} \exp[-NE(\theta_k, \sigma)] P(\theta_k|M_k) d\theta_k + \frac{N}{2} \log(2\pi\sigma^2) \quad (14)$$

$$= \int_0^1 \frac{\partial}{\partial \beta} \left\{ -\log \left[\int_{-\infty}^{\infty} \exp(-\beta NE(\theta_k, \sigma)) P(\theta_k|M_k) d\theta_k \right] \right\} d\beta + \frac{N}{2} \log(2\pi\sigma^2) \quad (15)$$

$$= \int_0^1 \int_{-\infty}^{\infty} NE(\theta_k, \sigma) P(\theta_k|\varepsilon, \beta, \sigma, M_k) d\theta_k d\beta + \frac{N}{2} \log(2\pi\sigma^2) \quad (16)$$

$$= \int_0^1 \langle NE(\theta_k, \sigma) \rangle_{P(\theta_k|\varepsilon, \beta, \sigma, M_k)} d\beta + \frac{N}{2} \log(2\pi\sigma^2), \quad (17)$$

where $\langle \cdot \rangle$ represents an expectation and

$$P(\theta_k|\varepsilon, \beta, \sigma, M_k) = \frac{\exp[-\beta NE(\theta_k, \sigma)] P(\theta_k|M_k)}{\int_{-\infty}^{\infty} \exp[-\beta NE(\theta_k, \sigma)] P(\theta_k|M_k) d\theta_k}. \quad (18)$$

When $NE(\theta_k, \sigma)$ is regarded as energy, Eq. (18) suggests that $P(\theta_k|\varepsilon, \beta, \sigma, M_k)$ and β correspond to the Boltzmann distribution and the inverse temperature in statistical physics, respectively. Equation (17) is approximated by a Riemann sum,

$$f(M_k, \sigma) \simeq \sum_{l=1}^L \langle NE(\theta_k, \sigma) \rangle_{P(\theta_k|\epsilon, \beta_l, \sigma, M_k)} \Delta\beta_l + \frac{N}{2} \log(2\pi\sigma^2), \tag{19}$$

where β_l is a sequence of inverse temperatures $0 = \beta_1 < \beta_2 < \dots < \beta_L = 1$ obtained by dividing the region from $\beta = 0$ to $\beta = 1$ into L pieces in some manner, and each $\langle NE(\theta_k, \sigma) \rangle_{P(\theta_k|\epsilon, \beta_l, \sigma, M_k)}$ is obtained by performing MCMC sampling independently at each inverse temperature β_l . However, MCMC sampling often results in trapping at local minima.

The replica exchange Monte Carlo (REMC) method is an algorithm of an MCMC method used to avoid trapping at local minima. The REMC method takes samples from the joint density

$$P(\theta_k^1, \theta_k^2 \dots \theta_k^L | \epsilon, \sigma, M_k) = \prod_{l=1}^L P(\theta_k^l | \epsilon, \beta_l, \sigma, M_k), \tag{20}$$

where the probability density $P(\theta_k^l | \epsilon, \beta_l, \sigma, M_k)$ is defined in Eq. (18). The REMC method performs sampling from the joint density $P(\theta_k^1, \theta_k^2 \dots \theta_k^L | \epsilon, \sigma, M_k)$ on the basis of the following updates.

- 1 **Sampling from each density** $P(\theta_k^l | \epsilon, \beta_l, \sigma, M_k)$
Sampling θ_k^l from $P(\theta_k^l | \epsilon, \beta_l, \sigma, M_k)$ by a conventional MCMC method, such as the Metropolis–Hastings algorithm²⁰.
- 2 **Exchange process between two densities corresponding to adjacent inverse temperatures**

The exchanges between the configurations θ_k^l and θ_k^{l+1} correspond to adjacent inverse temperatures following the probability $R = \min(1, r)$, where

$$\begin{aligned} r &= \frac{P(\theta_k^1, \dots, \theta_k^{l+1}, \theta_k^l, \dots, \theta_k^L | \epsilon, \sigma, M_k)}{P(\theta_k^1, \dots, \theta_k^l, \theta_k^{l+1}, \dots, \theta_k^L | \epsilon, \sigma, M_k)} \\ &= \frac{P(\theta_k^{l+1} | \epsilon, \beta_l, \sigma, M_k) P(\theta_k^l | \epsilon, \beta_{l+1}, \sigma, M_k)}{P(\theta_k^l | \epsilon, \beta_l, \sigma, M_k) P(\theta_k^{l+1} | \epsilon, \beta_{l+1}, \sigma, M_k)} \\ &= \exp\{N[\beta_{l+1} - \beta_l][E(\theta_k^{l+1}, \sigma) - E(\theta_k^l, \sigma)]\}. \end{aligned}$$

Sampling from a distribution with a smaller β corresponds to sampling from a distribution with a larger intensity of noise; thus, the distribution tends not to have a local minimum. Hence, sampling from the joint density $P(\theta_k^1, \theta_k^2 \dots \theta_k^L | \epsilon, \sigma, M_k)$ overcomes the local minima in distributions with large β and enables the rapid convergence of sampling.

Using the sampling result of the $\beta = 1$ state, we can obtain the posterior distribution of the parameter $P(\theta_k | \epsilon, \sigma, M_k) := P(\theta_k | \epsilon, \beta = 1, \sigma, M_k)$ [Eq. (18)] for the noise intensity σ . From the posterior distribution of θ_k , we can estimate the model parameters θ_k of M_k and the related information, such as the estimation accuracy.

From the sampling result of Eq. (20), $f(M_k, \sigma_s)$ can be obtained, where $\sigma_s := \sqrt{\frac{1}{\beta_s}} \sigma$. Here, we describe how to obtain $f(M_k, \sigma_s)$. $f(M_k, \sigma_s)$ can be rewritten by using σ as

$$f(M_k, \sigma_s) = \int_0^1 \langle NE(\theta_k, \sigma_s) \rangle_{P(\theta_k|\epsilon, \beta', \sigma_s, M_k)} d\beta' + \frac{N}{2} \log(2\pi\sigma_s^2) \tag{21}$$

$$= \int_0^{\frac{\sigma_s^2}{\sigma^2}} \langle NE(\theta_k, \sigma) \rangle_{P(\theta_k|\epsilon, \beta, \sigma, M_k)} \frac{\sigma_s^2}{\sigma^2} d\beta + \frac{N}{2} \log(2\pi\sigma_s^2), \tag{22}$$

where $\beta' = \frac{\sigma_s^2}{\sigma^2} \beta$. Then, $f(M_k, \sigma_s)$ can be obtained from a Riemann sum as

$$f(M_k, \sigma_s) \simeq \frac{\sigma_s^2}{\sigma^2} \sum_{l=0}^s \langle NE(\theta_k, \sigma) \rangle_{P(\theta_k|\epsilon, \beta_l, \sigma, M_k)} \Delta\beta_l + \frac{N}{2} \log(2\pi\sigma_s^2). \tag{23}$$

Thus, it is possible to calculate an approximate value of $f(M_k, \sigma_s)$ using the obtained expectation values $\langle NE(\theta_k, \sigma) \rangle_{P(\theta_k|\epsilon, \beta_l, \sigma, M_k)}$ by sampling at σ . Using the obtained integral value set $\{f(M_k, \sigma_s)\}_{s=1}^S$, the Bayesian free energy $F(M_k)$ can be calculated as

$$F(M_k) = -\log \int_{-\infty}^{\infty} d\sigma P(\epsilon | \sigma, M_k) P(\sigma) \tag{24}$$

$$= -\log \int_{-\infty}^{\infty} d\sigma \exp[-f(M_k, \sigma)] P(\sigma) \tag{25}$$

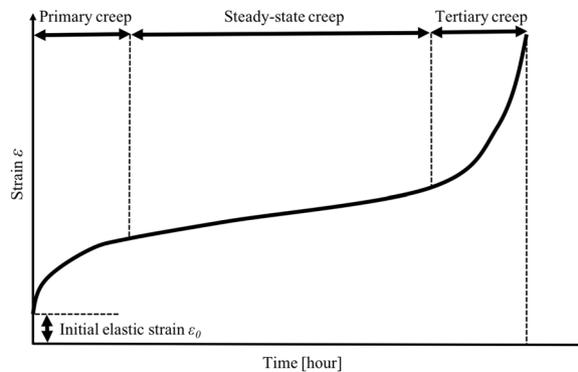


Figure 2. Creep curves and three time domains. Primary creep represents the zone of creep rate deceleration, steady-state creep represents the creep rate zone of constant velocity, and tertiary creep represents the final acceleration zone.

	C	Si	Mn	P	S	Cu	Ni	Cr	Mo	V	Nb	Al	N
(mass%)	0.10	0.25	0.43	0.006	0.002	0.012	0.06	8.87	0.93	0.19	0.07	0.014	0.06

Table 1. Chemical composition of the Gr.91 steel (9Cr-1Mo-Nb-V)¹⁴.

param	min	max	param	min	max
a_1	1.0×10^{-4}	2.0×10^{-3}	b_1	$1.0 \times 10^{-1.0}$	$1.0 \times 10^{0.0}$
a_2	1.0×10^{-7}	2.0×10^{-5}	b_2	0.5	1.9
c_1	1.0×10^{-5}	3.0×10^{-4}	d_1	$1.0 \times 10^{-3.0}$	$1.0 \times 10^{-2.5}$
c_2	1.0×10^{-22}	5.0×10^{-15}	d_2	$1.0 \times 10^{-3.0}$	$1.0 \times 10^{-1.9}$
ε_0	1.0×10^{-4}	1.0×10^{-3}	σ	3.0×10^{-5} to 3.3×10^{-3}	3.0×10^{-4} to 3.3×10^{-2}

Table 2. Prior distributions of Kimura model and modified Kimura model. All prior distributions were uniform, and their upper and lower limits were set as follows. The prior distribution of b_2 was used only with the Kimura model.

$$\simeq -\log \sum_{s=1}^L \exp[-f(M_k, \sigma_s)] P(\sigma_s) \Delta \sigma_s. \quad (26)$$

Application example

Here, as an application example to verify the effectiveness of the proposed framework, we evaluate the Kimura model, which is one of the most successful creep constitutive equations, and the modified Kimura model, which was created as a model for comparison with the Kimura model. These models are difficult to evaluate using the existing framework.

Creep constitution models and material. The modified Kimura model was set on the basis of the following assumptions. There are a number of creep constitutive models, which are roughly classified into two types according to the way the steady state, where the deformation rate is constant, is modeled (Fig. 2). One is a steady-state creep model, in which a linear region with a constant deformation rate is represented as an independent linear term. The other is the unsteady creep model, in which the steady state is generated by the balance between the deceleration of the primary creep and the acceleration of the tertiary creep. The Kimura model is formulated as Eq. (1). This model is an unsteady creep constitutive equation in which $a_1 t^{b_1} + a_2 t^{b_2}$ represents the primary creep and $c_1 \exp(d_1 t) + c_2 \exp(d_2 t)$ represents the tertiary creep, and the linear region is represented by its balance. By regression analysis using the creep data set of Gr.91 steel, it was previously found that b_2 takes a value close to 1¹. This implies that the Kimura model models the steady state as a linear term rather than a balance. In our previous study¹³, it was also found that, using the measurement data¹⁴ of Gr.91 steel, the likelihood of another type of creep constitutive model, the theta projection model⁴, was improved by adding a linear term. On the other hand, it was also reported by Kimura *et al.* that b_2 can take a value larger than 1 by changing the fitting method². To determine whether a steady state is required, we designed the following steady-state creep model by replacing b_2 in the Kimura model with 1.

$$\varepsilon = \varepsilon_0 + a_1 t^{b_1} + a_2 t + c_1 \exp(d_1 t) + c_2 \exp(d_2 t) \quad (27)$$

We call this regression model the modified Kimura model. Using the proposed framework, we determined whether the modified Kimura model or the Kimura model is more likely for Gr.91 creep data (Table 1) obtained at 650 °C and 80 MPa.

In this verification, on the basis of the regression results shown in refs. ¹ and ², the prior distributions of the Kimura model and the modified Kimura model were set as shown in Table 2. In Bayesian inference, priors are an important part of the model. To examine the effect of the prior distribution in Bayesian model selection, several different prior distributions of the noise intensity σ were prepared. The relationship between the prior and the model selection result was verified. Concretely, the prior distribution of the noise intensity σ was set as a uniform distribution with a certain value as the lower limit and 10 times the lower limit as the upper limit. Then, 100 different prior distributions were prepared with values of the lower limit from 3.0×10^{-5} to 3.3×10^{-3} at equal intervals in logarithmic space. We applied the proposed framework under these conditions.

When we set the prior distributions of $a_1, a_2, b_1, b_2, c_1, c_2, d_1$, and d_2 as described in Table 2, the exchange symmetry of parameter pairs in the Kimura model disappears. The exchange symmetry is one of the reasons why the Kimura model is an undetermined model. On the other hand, the Kimura model continues to have an undetermined structure around the parameters estimated as the maximum-likelihood solution for the following reasons. To model the transition of a creep curve under a change in stress and temperature conditions, Kimura *et al.* estimated the stress and temperature dependences of parameters. They reported that the distribution of c_2 in stress and temperature space has a significantly wider dispersion than the other parameters^{1,2}. This behavior can be understood from the fact that the estimation of parameter c_2 is unstable, and, therefore, the parameters of the Kimura model are not well-determined around the maximum-likelihood solution.

In the REMC sampling, we adopted the Metropolis–Hastings algorithm²⁰ to sample each state of the inverse temperature. The states of the inverse temperature were determined using the following exponential function²¹:

$$\beta_l = \begin{cases} 0.0 & (l = 1) \\ \gamma^{l-L} & (l \geq 13) \end{cases}, \quad (28)$$

where $L = 311$ and $\gamma = 1.05$. The approximate error Er_l of the numerical integration, i.e., the Riemann sum, between β_l and β_{l+1} is given as

$$Er_l \sim \frac{1}{2} \left[\langle NE(\theta_k, \sigma) \rangle_{P(\theta_k | \epsilon, \beta_l, \sigma, M_k)} - \langle NE(\theta_k, \sigma) \rangle_{P(\theta_k | \epsilon, \beta_{l+1}, \sigma, M_k)} \right] \Delta\beta_l \quad (29)$$

$$= \frac{1}{2} \Delta \langle E \rangle_l \Delta\beta_l. \quad (30)$$

Therefore, it is necessary to make the division $\Delta\beta_l$ sufficiently smaller than $\Delta \langle E \rangle_l$. In the Kimura model, which is a linear sum of exponential functions and power functions, slight parameter variations cause large functional changes. In particular, at $l = 1$, where sampling is performed over the entire range of the prior distribution, the $l = 1$ state takes a very large average energy $\langle E_1 \rangle$ because of such the nature of this model. On the other hand, in the region of $l > 1$, the parameter region with a small squared error is sampled. Therefore, the energy in the $l = 2$ state becomes $\langle E_2 \rangle \ll \langle E_1 \rangle$. As a result, $\Delta \langle E \rangle_l$ becomes very large and the approximation error Er_l becomes non-negligible. Therefore, in this study, to reduce the approximation error Er_l , the following temperature states are inserted between β_1 and β_{13} so that $\Delta\beta_1 < \Delta \langle E \rangle_l$.

$$\beta_l = \begin{cases} 10^{-28} & (l = 2) \\ 10^{-25} & (l = 3) \\ 10^{-22} & (l = 4) \\ 10^{-20} & (l = 5) \\ 10^{-18} & (l = 6) \\ 10^{-16} & (l = 7) \\ 10^{-14} & (l = 8) \\ 10^{-12} & (l = 9) \\ 10^{-10} & (l = 10) \\ 10^{-8} & (l = 11) \\ 10^{-7} & (l = 12) \end{cases} \quad (31)$$

In the sampling, we abandoned the first 100,000 steps and sampled the next 100,000 steps.

Results

In this study, model parameters were estimated from the posterior distributions using the maximum a posteriori (MAP) method. The MAP method estimates parameters on the basis of the following equations:

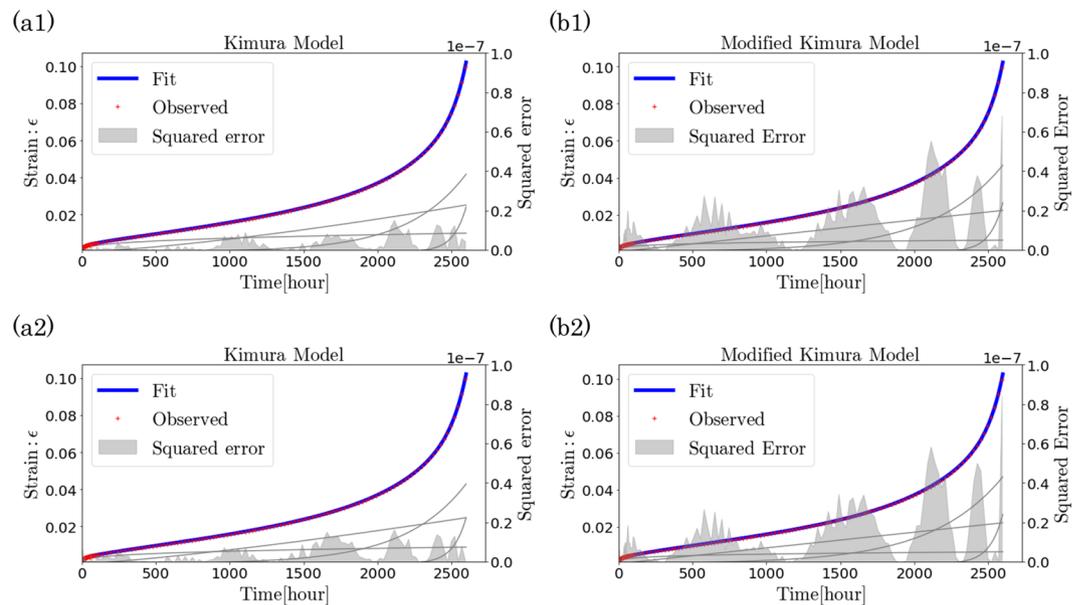


Figure 3. Fitting results (blue curve) of creep strain data (red points). **(a1,b1)** Fitting results of Kimura model and modified Kimura model **(b1)** with noise prior set as a uniform distribution from 4.0×10^{-5} to 4.0×10^{-4} . **(a2,b2)** Fitting results of Kimura model **(a2)** and modified Kimura model **(b2)** with noise prior set as a uniform distribution from 4.0×10^{-4} to 4.0×10^{-3} . The gray curve represents the component of each term included in the Kimura and modified Kimura models, and the gray area represents the squared error between the experimental data and the regression curve at each time.

$$\begin{aligned}\sigma^{MAP} &= \arg \max_{\sigma} P(\sigma | \epsilon, M_k), \\ \theta_k^{MAP} &= \arg \max_{\theta_k} P(\theta_k | \epsilon, \sigma^{MAP}, M_k).\end{aligned}\quad (32)$$

We examined the fitting results by the MAP solution with two types of noise prior. One was a small-noise-intensity prior, which was set as a uniform distribution from 4.0×10^{-5} to 4.0×10^{-4} , and the other was a large-noise-intensity prior, which was set as a uniform distribution from 4.0×10^{-4} to 4.0×10^{-3} . From the regression results (Fig. 3), it was confirmed that both the Kimura model and the modified Kimura model provided a good fitting regardless of the prior distribution of noise. To examine the regression results more precisely, we compared the mean square error (MSE) for the fitting results:

$$MSE = \frac{1}{N} \sum_{i=1}^N (\epsilon_i - \epsilon_k(t_i; \theta_k))^2. \quad (33)$$

For the small-noise-intensity prior distribution, the MSE of the Kimura model was 1.65×10^{-9} and that of the modified Kimura model was 8.31×10^{-9} . For the large-noise-intensity prior distribution, the MSE of the Kimura model was 2.69×10^{-9} and that of the modified Kimura model was 8.44×10^{-9} . Thus, the Kimura model always has a smaller MSE regardless of the noise intensity of the prior distribution. Moreover, the MSE did not change significantly with the noise intensity of the prior distribution. It was also confirmed that the squared error between the experimental data and the regression curve at each time (gray area of Fig. 3) has the same error distribution in each model regardless of the noise intensity. The creep rate curve was calculated by differentiating this regression curve. The estimated creep rate curves closely fit the creep rate data points. The creep rate data points was estimated as the difference in the adjacent points of strain data (black dots in Fig. 4). The value of the coefficient a_2 of the modified Kimura model [Eq. (27)], representing stationary creep, was similar to the minimum creep rate estimated by the experimenter as the creep rate in the steady area¹³. This result is consistent with the assumption that the second term of the modified Kimura model models the steady-state creep.

Using the proposed framework, it was confirmed that the selected model switched from the Kimura model to the modified Kimura model at the point where the prior distribution of the noise intensity σ is set as a uniform distribution from 3.88×10^{-4} to 3.88×10^{-3} (Fig. 5(a)). On the other hand, it was also confirmed that the MSE of the Kimura model was always smaller than that of the modified Kimura model regardless of the noise prior distribution (Fig. 5(b)). For comparison with the existing framework¹⁷, model evaluation based on the existing framework was performed. In the existing framework¹⁷, the Bayesian free energy is calculated on the basis of the Laplace approximation (see Appendix A for a more general theory of the Laplace approximation including σ). Using the MAP solution θ_k^{MAP} , the Laplace approximation is given by

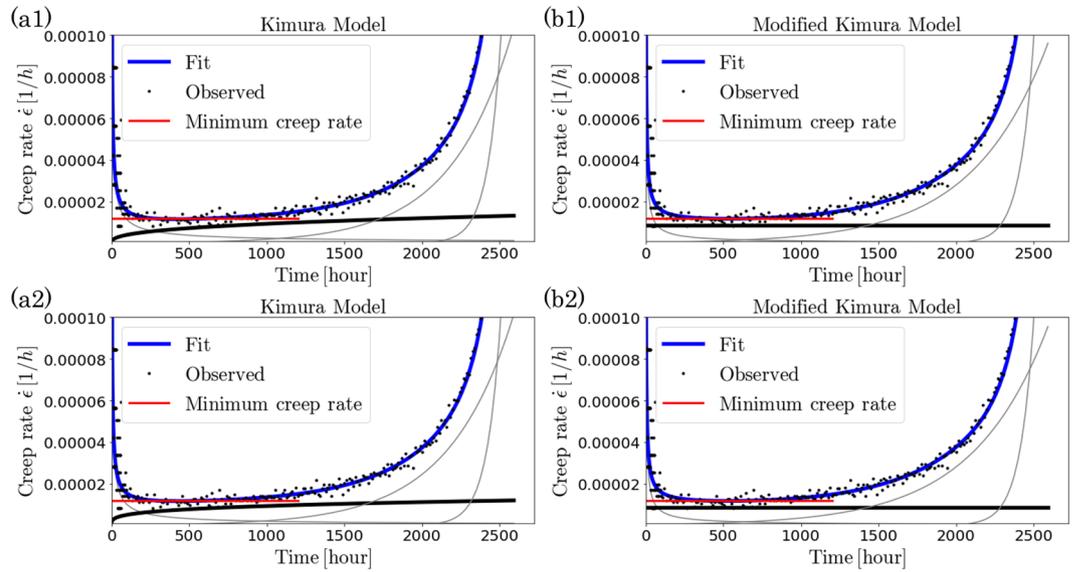


Figure 4. Creep rate curves obtained by differentiating the regression strain curve. **(a1,b1)** Creep rate curve obtained by fitting strain curve by Kimura model **(a1)** and modified Kimura model **(b1)** with noise prior set as uniform distribution from 4.0×10^{-5} to 4.0×10^{-4} . **(a2,b2)** Creep rate curve obtained by fitting strain curve by Kimura model **(a2)** and modified Kimura model **(b2)** with noise prior set as uniform distribution from 4.0×10^{-4} to 4.0×10^{-3} . The creep rate data points (black dots) were calculated as the slope of the adjacent points of strain measurement data. The red line represents the minimum creep rate estimated by the creep measurement experimenter¹³. The gray and black lines represent the components of each term in the model. In particular, the black line represents the component corresponding to the second term of the model.

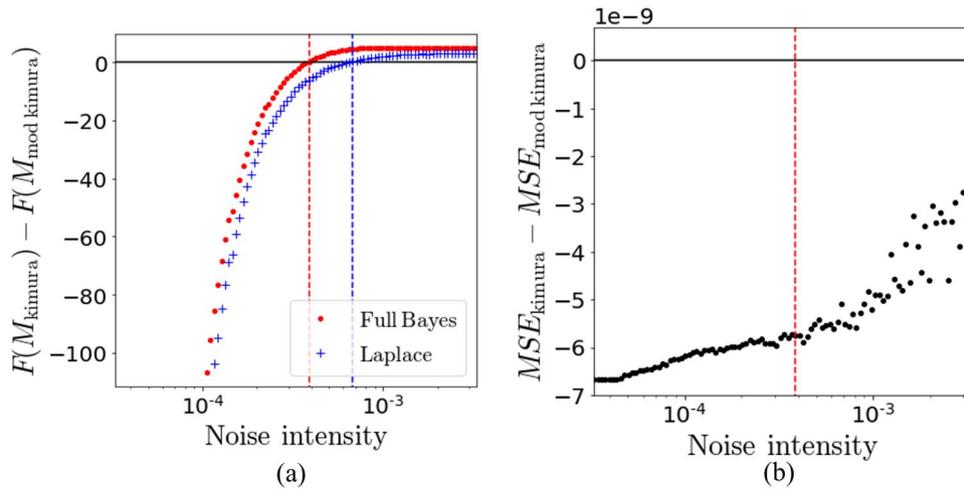


Figure 5. **(a)** Model selection result. Relationship between prior distribution of noise intensity σ and model selection results of Kimura model and modified Kimura model. The x-axis represents the lower bound of the prior distribution of noise intensity. The prior distribution was a uniform distribution from this value to 10 times this value. “Full Bayes” is the result of the proposed framework and “Laplace” is the result of the existing framework proposed by Keitel *et al.*¹⁷. **(b)** Comparison of MSE of Kimura model and modified Kimura model. The x-axis is the same as in **(a)**.

$$\begin{aligned}
 F(M_k) \sim & -\log P(\epsilon|\theta_k^{MAP}, \sigma^{MAP}, M_k)P(\theta_k^{MAP}|M_k)P(\sigma^{MAP}|M_k) \\
 & - \frac{d+1}{2}\log(2\pi) + \frac{d+1}{2}\log(N) + \frac{1}{2}\log[\det(\mathbf{H})],
 \end{aligned}
 \tag{34}$$

where H is the Hessian matrix: $\mathbf{H}^{ij} = \sum_{l=1}^{d+1} \sum_{j=1}^{d+1} \frac{\partial^2 g(\theta_l)}{\partial \theta_i^l \partial \theta_j^l} \Big|_{\theta_k = \theta_k^{MAP}}$. In the existing framework¹⁷, H is replaced by the inverse of the variance-covariance matrix \mathbf{Cov}^{-1} of θ_k to reduce the computational cost,

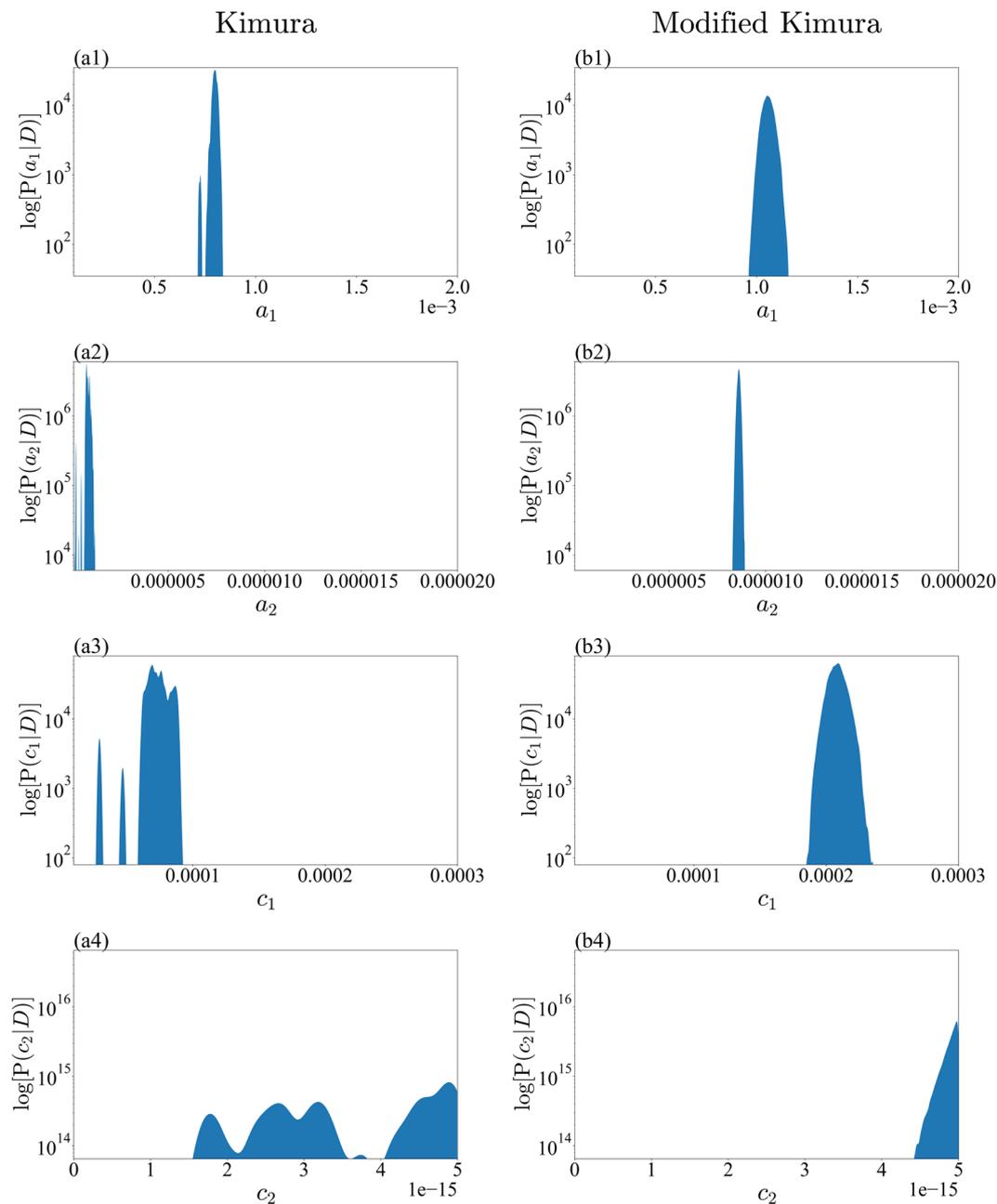


Figure 6. Marginalized posterior distributions for model parameters $\{a_1, a_2, c_1, c_2\}$ of Kimura model and modified Kimura model with small-noise-intensity prior. **(a1)** a_1 of Kimura model. **(b1)** a_1 of modified Kimura model. **(a2)** a_2 of Kimura model. **(b2)** a_2 of modified Kimura model. **(a3)** c_1 of Kimura model. **(b3)** c_1 of modified Kimura model. **(a4)** c_2 of Kimura model. **(b4)** c_2 of modified Kimura model.

$$\begin{aligned}
 F(M_k) \sim & -\log P(\epsilon | \theta_k^{MAP}, \sigma^{MAP}, M_k) P(\theta_k^{MAP} | M_k) P(\sigma^{MAP} | M_k) \\
 & - \frac{d+1}{2} \log(2\pi) + \frac{d+1}{2} \log(N) - \frac{1}{2} \log[\det(\mathbf{Cov})].
 \end{aligned} \tag{35}$$

Model selection was performed using this approximated free energy. In the model selection results using this approximated free energy, selected models are switched at a higher noise intensity than the switched noise intensity of the proposed framework (Fig. 5(a)). This means that the existing framework more often evaluates the Kimura model as the more likely model than the proposed framework.

To visualize a posterior distribution with three or more dimensions, we calculated the following marginal posterior distribution, which marginalizes the posterior distribution of the parameters θ_k^m except for the parameter of interest θ_k^m :

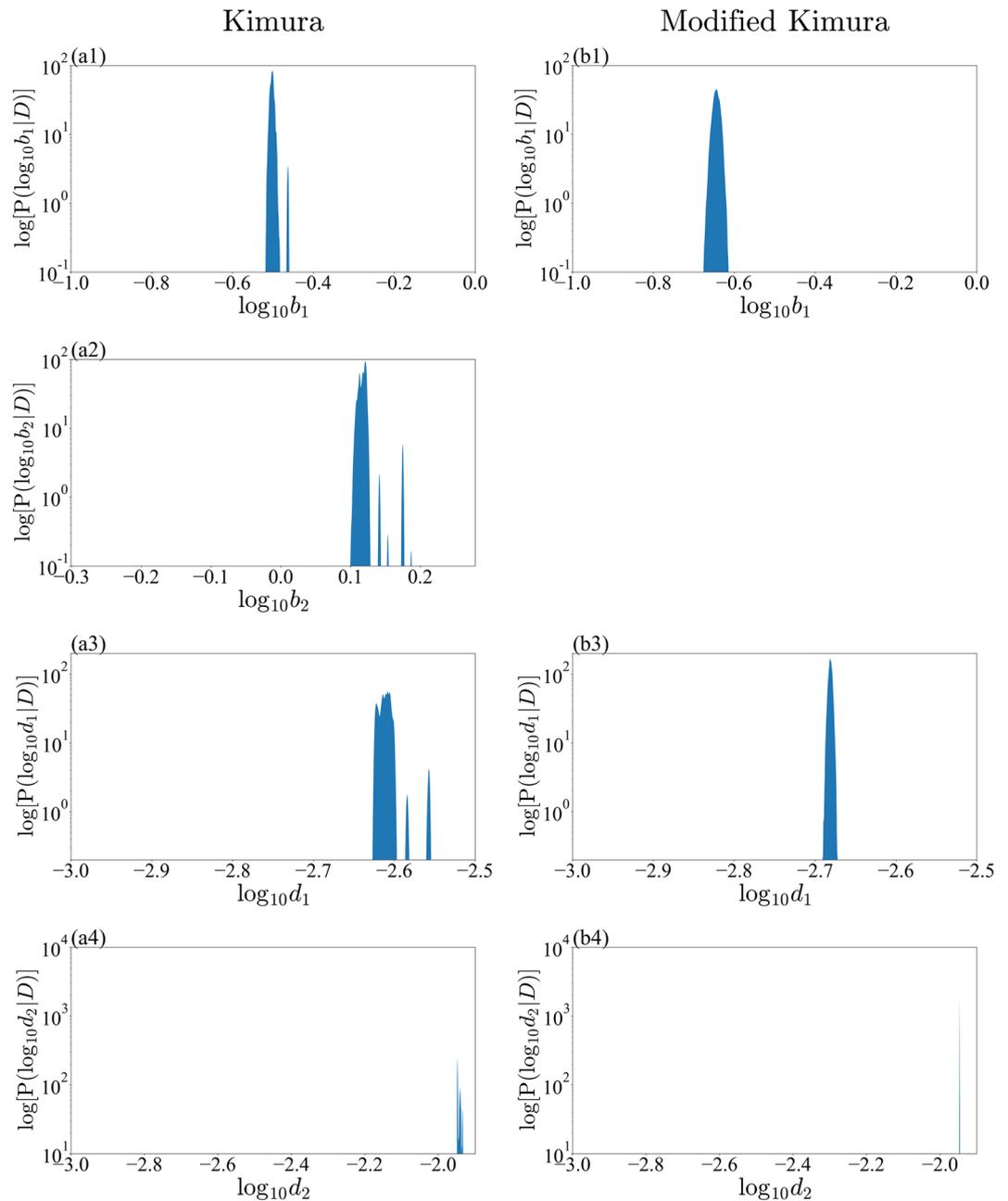


Figure 7. Marginalized posterior distributions for model parameters $\{b_1, b_2, d_1, d_2\}$ of Kimura model and modified Kimura model with small-noise-intensity prior. **(a1)** b_1 of Kimura model. **(b1)** b_1 of modified Kimura model. **(a2)** b_2 of Kimura model. **(a3)** d_1 of Kimura model. **(b3)** d_1 of modified Kimura model. **(a4)** d_2 of Kimura model. **(a4)** d_2 of modified Kimura model.

$$P(\theta_k^m | \varepsilon, \sigma^{MAP}, M_k) = \int_{-\infty}^{\infty} P(\theta_k | \varepsilon, \sigma^{MAP}, M_k) P(\theta_k, M_k) d\theta_k^{-m}. \quad (36)$$

This integration can be approximated from the sum of the sampling result of the REMC method. The density function of the marginal posterior can be estimated by the kernel density estimation method using Gaussian kernels. We determined the bandwidth of the Gaussian kernels using Scott's rule²². We compared the marginalized posterior distribution focusing on one parameter for the small-noise-intensity prior, where the Kimura model was selected (Figs. 6 and 7), and the large-noise-intensity prior, where the modified Kimura model was selected (Figs. 8 and 9). From the visualization results of the marginalized posterior distribution for the small-noise-intensity prior (Figs. 6 and 7), it was confirmed that the parameters' posterior distribution of the Kimura model has a multimodal and complex structure, which is different from the posterior distribution of the modified Kimura model. The posterior distribution of c_2 with a small-noise-intensity prior has a particularly multimodal and broadened distribution (Fig. 6(a4)). Whereas, in the large-noise-intensity prior, it was observed

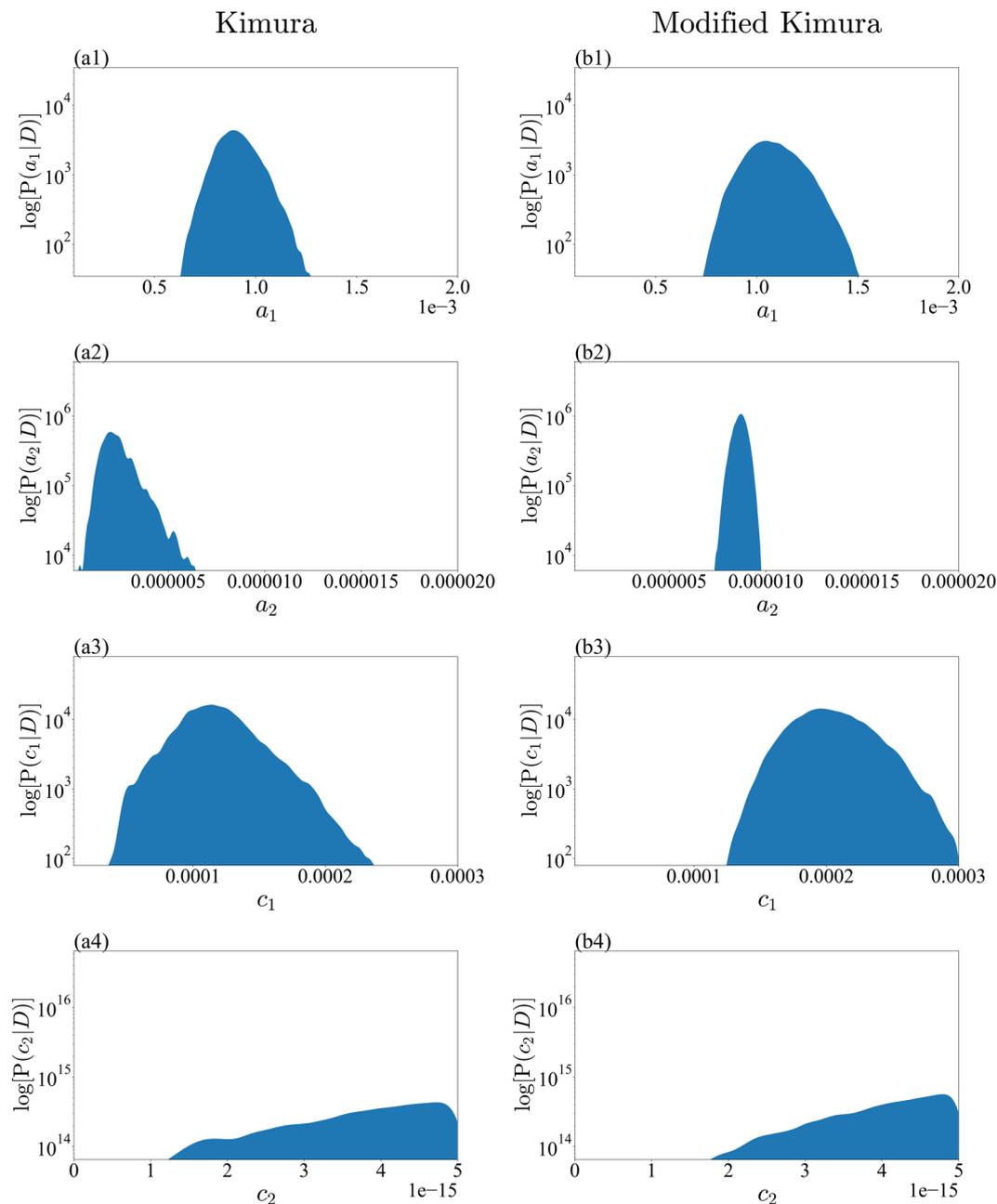


Figure 8. Marginalized posterior distributions for model parameters $\{a_1, a_2, c_1, c_2\}$ of Kimura model and modified Kimura model with large-noise-intensity prior. **(a1)** a_1 of Kimura model. **(b1)** a_1 of modified Kimura model. **(a2)** a_2 of Kimura model. **(b2)** a_2 of modified Kimura model. **(a3)** c_1 of Kimura model. **(b3)** c_1 of modified Kimura model. **(a4)** c_2 of Kimura model. **(b4)** c_2 of modified Kimura model.

that the posterior distribution of the Kimura model also became unimodal distribution (Figs. 8 and 9). Next, we examined the relationship between two parameters from the marginalized posterior distribution. Many marginalized posterior distributions were unstructured isotropic Gaussian distributions such as the marginalized posterior distributions of c_1 and ϵ_0 (Figs. 10(a1,b1)). On the other hand, the marginalized posterior distributions between parameters of the exponential part and the coefficient part of the basis function have structured distributions, such as c_2 and $\log_{10} d_2$ (Figs. 10(a2,b2)).

Summary and Discussion

The results of the application of the proposed framework to the Kimura model show that the model selection results vary depending on the noise intensity range. Furthermore, it is clarified for the first time that the posterior distribution of the Kimura model is multimodal in the small-noise-intensity region. In this section, after discussing the validity of the model selection results and their properties, we compare the Kimura model and the modified Kimura model in terms of multimodality of their posterior distributions.

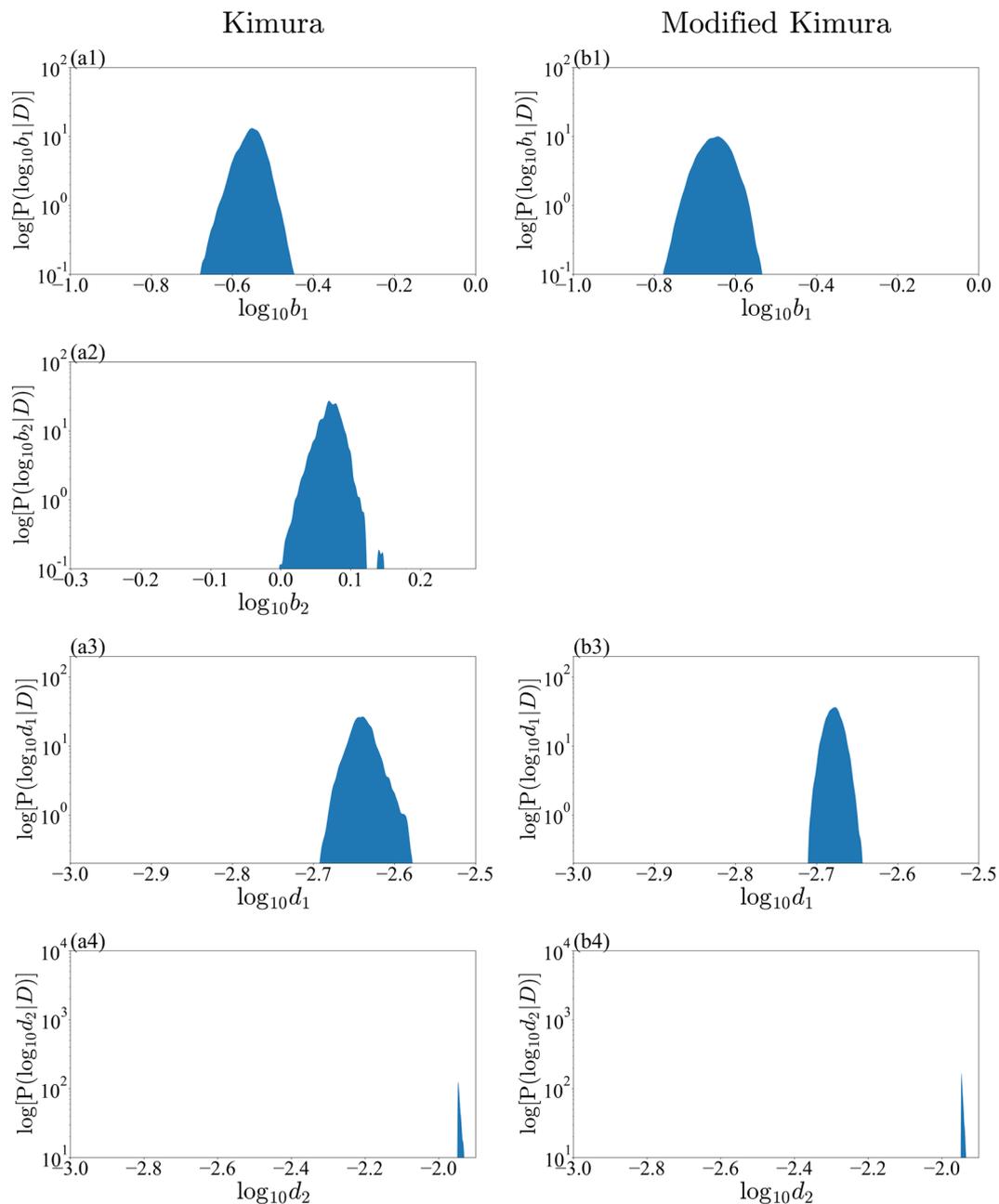


Figure 9. Marginalized posterior distributions for model parameters $\{b_1, b_2, d_1, d_2\}$ of Kimura model and modified Kimura model with large-noise-intensity prior. **(a1)** b_1 of Kimura model. **(b1)** b_1 of modified Kimura model. **(a2)** b_2 of Kimura model. **(a3)** d_1 of Kimura model. **(b3)** d_1 of modified Kimura model. **(a4)** d_2 of Kimura model. **(a4)** d_2 of modified Kimura model.

Kimura *et al.* achieved high-accuracy fitting to creep measurement data using the Kimura model^{1,2}. Similarly to their work^{1,2}, our proposed framework achieved data fitting with high accuracy. This suggests that our framework performed under the same conditions as the Kimura model. In addition, the closeness of the minimum creep rate and the slope a_2 of the second term of the modified Kimura model (Eq. (27)) suggests that the steady state was modeled by the linear term $a_2 t$, which was assumed when we set the modified Kimura model. Thus, it was confirmed that the regression results using the MAP solution of the proposed framework were reasonable.

The MSE of the Kimura model was smaller than that of the modified Kimura model, even in the large-noise-intensity prior distribution, in which the modified Kimura model was selected. This result indicates that the proposed framework gives model selection results that do not depend solely on the magnitude of the error. The results are explained using the following approximated Bayesian free energy formula. By extracting terms higher than order $\log(N)$ from the result of the Laplace approximation, the following equation is obtained (see Appendix A):

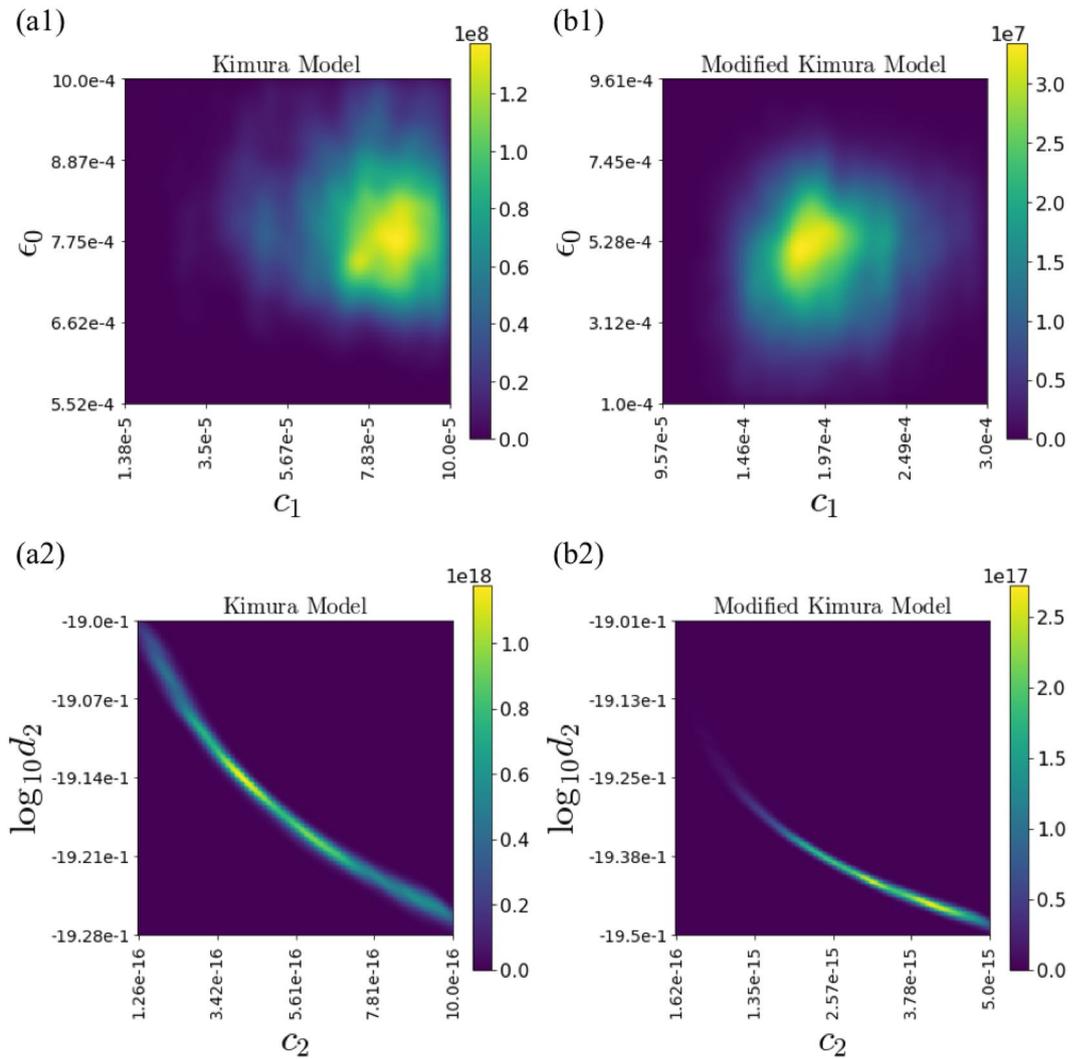


Figure 10. Marginalized posterior distributions for two parameters.

$$F(M_k) \sim -\log P(\epsilon | \theta_k^{MAP}, \sigma^{MAP}, M_k) + \frac{d+1}{2} \log(N) \tag{37}$$

$$= NE(\theta_k^{MAP}, \sigma) + \frac{d+1}{2} \log(N), \tag{38}$$

where d is the number of model parameters θ_k . The first term in Eq. (38) represents the regression error of the model and the second term represents the complexity of the model. In general, the more complex the model, the smaller the value of the first term; however, overfitting occurs in more complex models. The presence of the second term in the Bayesian free energy makes it possible to evaluate the model while preventing overfitting.

Thus, the Laplace approximation is useful for a rough clarification of the property of Bayesian free energy. However, the Laplace approximation is based on the hypothesis that the posterior distribution of the parameters is a unimodal Gaussian distribution. On the other hand, the proposed method selects the model by assuming that the posterior distributions of the Kimura model with small-noise-intensity prior are multimodal. The difference in the model selection results between the frameworks (Fig. 5(a)) is due to this difference in the property of the Laplace approximation and the proposed framework. This presumption is validated from the behavior of $\Delta F = F(M_{kimura}) - F(M_{mod\ kimura})$ in the Laplace approximation and the proposed framework (Fig. 5(a)), which tends to be small in a large-noise-intensity prior situation where the posterior distribution of the Kimura model is estimated as unimodal in the proposed framework (Figs. 8(a1–4) and 9(a1–4)). This might be the mechanism by which the switching noise intensity of the selected model differs between the proposed framework and the Laplace approximation (Fig. 5(a)).

Using the proposed framework, it is possible to set the prior knowledge of the measurement noise range as the prior distribution. In this research, we developed an efficient marginalization method of the noise intensity σ

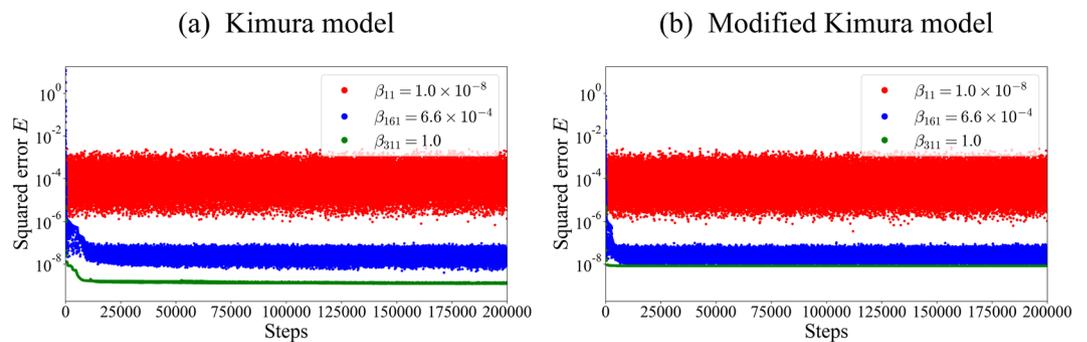


Figure 11. Convergence of REMC sampling. **(a)** Sampling series of squared error $E(\theta_k, \sigma)$ of the Kimura model. Red dots represent the sampling series at low β_p , blue is intermediate, and green is large. **(b)** Sampling series of squared error $E(\theta_k, \sigma)$ of the modified Kimura model. Red dots represent the sampling series at low β_p , blue is intermediate, and green is large.

using the sampling result of the REMC method. Such a method is novel not only in creep model selection but also in general Bayesian model selection. As a result of the model selection with this prior noise distribution, it was confirmed that the model selection results of the creep constitutive equation markedly change with the prior distribution of the noise. The measurement noise intensity of the creep data in this study was estimated to be more than 1.4×10^{-4} from the resolution of the linear gauge (see Appendix B). Since the influences of temperature and humidity fluctuations are added to this noise, the lower bound of the noise-intensity estimate is exist around the region where the model selection results are switched (Fig. 5(a)). This result shows the importance of specifying the measurement noise intensity in advance, even if the measurement noise intensity is only estimated as a range.

Our analysis revealed that for the Kimura model the posterior distributions of all parameters are multimodal with a small-noise-intensity prior (Figs. 6 and 7). The multimodality should result in unstable optimization when the parameters are estimated by the least-squares method, in which the parameters are optimized at a zero-noise limit. The posterior distribution of c_2 with a small-noise-intensity prior has a particularly multimodal and broadened distribution (Fig. 6(a4)). In fact, according to the work of Kimura *et al.*^{1,2}, the estimation of c_2 is particularly unstable, as we explained in the previous section. Considering the present result, it would be very difficult to stably estimate not only c_2 but also the other parameters by the least-squares method. By contrast, for the modified Kimura model, the posterior distributions of all the parameters were unimodal regardless of the noise-intensity prior. The modified Kimura model was simplified by omitting only one parameter b_2 . The posterior distribution of b_2 with small-noise-intensity prior (Fig. 7(a2)) shows that b_2 has a value greater than 1. This suggests that the secondary term of the Kimura model contributes to the tertiary creep as well as the secondary creep. Inevitably, the Kimura model represents the tertiary creep by three basis functions. Perhaps this is the reason for the multi modality of the posterior distribution of the Kimura model with the small-noise-intensity prior. This demonstrates that the present simplification is efficient for modifying the Kimura model toward a well-determined one. The obtained unimodality of the posterior distributions should allow us to stably estimate the parameters even by the least-squares method. Furthermore, the stability of the parameter estimation would yield a better regression of the parameters in terms of the test temperature and applied stress. Consequently, we consider that the modified Kimura model has the potential to improve the estimation of creep curves under a wide range of test conditions. These results are due to the ability of the proposed framework to capture the complex relationships among the parameters. The previous frameworks for creep, such as the framework using the Laplace approximation^{13,17}, could not capture such complex relationships among the parameters.

From the posterior distributions of two parameters, we can also obtain other information about the model properties. The posterior distributions of the parameter of the exponent part and the coefficient parameter are correlated distributions (Figs. 10(a2,b2)). Mathematically, if the parameter of the exponent part increases or decreases, the coefficient parameter must decrease or increase, respectively, to maintain the same trend of the basis function. Such an analysis is useful for quantitative analysis to understand the property of the creep model.

Lastly, we discuss the applicability of the proposed framework to other types of creep constituent equation. The framework is computationally expensive, especially for sampling by the REMC method. In the proposed framework, the samples should be sufficiently representative such that the approximation using Eq. (26), i.e., reusing the samples for estimating the different $f(M_k, \sigma)$, would be accurate. The REMC method, which is one of the generalized-ensemble algorithms, allows us to efficiently perform the marginalization of the multi modal probabilistic distribution compared with the other sampling methods²³. We have now conservatively sampled more than 200,000 steps, but we can probably reduce it to 50,000, as shown in Fig. 11 of Appendix C. Nevertheless, applying the proposed framework to a more complex model than the Kimura model will be difficult in terms of sampling convergence. Fortunately, as we mentioned in Introduction, the Kimura model is one of the most complex creep constitutive equations^{15,16}. Thus, one can apply the proposed framework to a wide range of creep constitutive equations that have been proposed so far.

Another issue is how this method applies to the case where the constituent equations are given by simultaneous differential equations. Such types of constitutive equation have been proposed on the basis of damage mechanics, such as the Hayhurst model²⁴, K-R model²⁵, L-M model²⁶, and W-T model²⁷. In principle, the

proposed method is applicable to these models because the strain ϵ can be computed for a given time t . However, it is necessary to run a numerical simulation in each sampling step to calculate the squared error $E(\theta_k, \sigma)$ (Eq. (11)), which takes a long computational time.

In this paper, we proposed a universal Bayesian creep model selection framework that can evaluate any type of creep constitutive equation. We also showed the effectiveness of our proposed framework using the measurement data of Gr.91 steel, a high-Cr, ferritic, heat-resistant steel. Through this evaluation, we found that the modified Kimura model could be a candidate model to improve the estimation stability of creep curves compared with the Kimura model. This was achieved by accurate estimation of the posterior distribution obtained only using the proposed framework. By applying the proposed method to different steel types at various temperatures and stresses, in-depth knowledge could be obtained about the creep deformation process and how to improve its prediction.

Appendix A: Laplace approximation of Bayesian free energy

The log marginal likelihood function is defined as

$$\begin{aligned} g(\theta_k, \sigma) &= -\frac{1}{N} \log P(\epsilon | M_k) \\ &= -\frac{1}{N} \log [P(\epsilon | \theta_k, \sigma, M_k) P(\theta_k | M_k) P(\sigma | M_k)]. \end{aligned} \tag{39}$$

If the Hessian matrix of the log marginal likelihood function about θ_k and σ , defined as $\mathbf{H}^{ij} = \sum_{i=1}^{d+1} \sum_{j=1}^{d+1} \frac{\partial^2 g(\theta_k)}{\partial \Theta_k^i \partial \Theta_k^j} \Big|_{\Theta_k = \hat{\Theta}_k}$, does not degenerate, the Taylor expansion around the MAP solution $(\hat{\theta}_k, \hat{\sigma}) := \arg \max_{\theta_k, \sigma} g(\theta_k, \sigma)$ becomes

$$g(\theta_k) = g(\hat{\Theta}_k) + \frac{1}{2} \sum_{i=1}^d \sum_{j=1}^d \frac{\partial^2 g(\theta_k)}{\partial \Theta_k^i \partial \Theta_k^j} \Big|_{\Theta_k = \hat{\Theta}_k} (\Theta_k^i - \hat{\Theta}_k^i)(\Theta_k^j - \hat{\Theta}_k^j) + \mathcal{O}(\Theta^3), \tag{40}$$

where $\Theta_k := (\theta_k, \sigma)$ and we use the fact that the first term of the Taylor expansion becomes 0 from the definition of $(\hat{\theta}_k, \hat{\sigma})$:

$$\frac{\partial g(\theta_k)}{\partial \Theta_k^i} \Big|_{\Theta_k = \hat{\Theta}_k} = 0 \quad (i = 1 - d + 1). \tag{41}$$

By using this Taylor expansion, the saddle point approximation of $F(M_k)$ around the MAP solution is obtained as

$$F(M_k) = -\log \left[\int_{-\infty}^{\infty} d\theta_k d\sigma P(\epsilon | \theta_k, \sigma, M_k) P(\theta_k | M_k) P(\sigma | M_k) \right] \tag{42}$$

$$= -\log \left\{ \int_{-\infty}^{\infty} d\theta_k d\sigma \exp[-Ng(\theta_k, \sigma)] \right\} \tag{43}$$

$$\sim -\log \left\{ \exp[-Ng(\hat{\Theta}_k)] \int_{-\infty}^{\infty} d\Theta_k \exp \left[-\frac{N}{2} \sum_{i=1}^{d+1} \sum_{j=1}^{d+1} (\Theta_k^i - \hat{\Theta}_k^i) \mathbf{H}^{ij} (\Theta_k^j - \hat{\Theta}_k^j) \right] \right\} \tag{44}$$

$$\begin{aligned} &= -\log P(\epsilon | \hat{\theta}_k, \hat{\sigma}, M_k) P(\hat{\theta}_k | M_k) P(\hat{\sigma} | M_k) \\ &\quad - \frac{d+1}{2} \log(2\pi) + \frac{d+1}{2} \log(N) + \frac{1}{2} \log[\det(\mathbf{H})], \end{aligned} \tag{45}$$

where H is the Hessian matrix. This approximation is called the Laplace approximation. Furthermore, if we take terms whose order is higher than $\log n$, Eq. (45) becomes

$$F(M_k) \sim -\log P(\epsilon | \hat{\theta}_k, \hat{\sigma}, M_k) + \frac{d+1}{2} \log(N) \tag{46}$$

$$= NE(\theta_k, \sigma) + \frac{d+1}{2} \log(N), \tag{47}$$

where we use the fact that $\log P(\epsilon | \theta_k, \sigma, M_k) = \sum_{i=1}^N \log P(\epsilon_i | \theta_k, \sigma, M_k) \equiv \mathcal{O}(N)$. If there is no effect on the log-likelihood $g(\theta_k, \sigma)$ of the prior distribution, such as in the limit of $n \rightarrow \infty$, then $(\hat{\theta}_k, \hat{\sigma})$ matches the maximum-likelihood estimator. Equation (47) is referred to as the Bayesian information criterion (BIC).

Appendix B: Estimation of measurement error

By using the parameters of the strain gauge, i.e., the initial length l_0 and the length l_t at time t , the strain ϵ is expressed as,

$$\epsilon = \frac{l_t - l_0}{l_0} (= \hat{\epsilon} \pm \delta\epsilon), \quad (48)$$

$$l_0 = \hat{l}_0 \pm \delta l_0, \quad l_t = \hat{l}_t \pm \delta l_t, \quad (49)$$

where $\hat{\epsilon}$, \hat{l}_0 and \hat{l}_t are the true values and $\delta\epsilon$, δl_0 , and δl_t are the measurement errors of each value. Because we assume that creep measurement noise $\delta\epsilon$ is added to measurement data independently and identically, the measurement error $\delta\epsilon$ propagates from δl_t and δl_0 as follows.

$$\delta\epsilon = \left| \frac{\partial\epsilon}{\partial l_0} \right| \delta l_0 + \left| \frac{\partial\epsilon}{\partial l_t} \right| \delta l_t = \left| \frac{-l_t}{l_0^2} \right| \delta l_0 + \left| \frac{1}{l_0} \right| \delta l_t \quad (50)$$

In the measurement data used in this study, $\hat{l}_0 = 3 \times 10^{-2}$ [m]. From the measurement accuracy of the linear gauge of the measurement system, we find that 2×10^{-6} [m] $\leq \delta l_0, \delta l_t$, no matter how small the estimate. Consequently, the lower limit of the measurement error of ϵ is obtained as

$$\delta\epsilon \geq \frac{l_t + l_0}{l_0^2} \times 2 \times 10^{-6} = (2 + \epsilon) \times \frac{2 \times 10^{-6}}{l_0} \quad (51)$$

$$\sim 2 \times 7 \times 10^{-5} = 1.4 \times 10^{-4}, \quad (52)$$

where we assume $\epsilon \ll 2$.

Appendix C: Convergence of REMC sampling

In this study, the samples should be sufficiently representative such that the approximation using Eq. (26), i.e., reusing the samples for estimating the different $f(M_k, \sigma_k)$, would be accurate. Therefore, it is important to confirm the convergence of sampling at all β_i states. We checked their sampling series at all β_i in Figs. 11(a,b). In this study, accuracy rather than computational cost was emphasized, and conservatively long sampling, 200,000 steps, was performed. However, the sampling series of the squared error $E(\theta_k, \sigma)$ (Eq. (11)) described in Fig. 11 shows that the sampling converges at about 50,000 steps (Figs. 11(a,b)).

Data availability

The data that support the findings of this study are available from Dr. M. Demura but restrictions apply to the availability of the data used under license for the current study, which are not publicly available. Data are, however, available from the authors upon reasonable request and with the permission of Dr. M. Demura.

Received: 29 February 2020; Accepted: 12 May 2020;

Published online: 26 June 2020

References

- Kimura, K., Sawada, K. & Kushima, H. Creep deformation analysis of grade 91 steels and prediction of creep strength properties. In ASME 2014 Pressure Vessels and Piping Conference, PVP2014-28674 (American Society of Mechanical Engineers Digital Collection, 2014).
- Kimura, K., Sawada, K. & Kushima, H. Evaluation of creep deformation property of grade 91 steels. In ASME 2015 Pressure Vessels and Piping Conference, PVP2015-45405 (American Society of Mechanical Engineers Digital Collection, 2015).
- Garofalo, F. *Fundamentals of creep and creep-rupture in metals* (Macmillan, 1965).
- Evans, R. W. *An Extrapolation Procedure for Long Term Creep-Strain and Creep Life Prediction* (Pineridge Press, 1982).
- Evans, R. W. & Wilshire, B. *Creep of metals and alloys* (IMM North American Pub. Center, 1985).
- Maruyama, K., Harada, C. & Oikawa, H. Formulation of creep curves and rupture lives for long-term creep property prediction with special reference to a 12 Cr (H46) steel. *Trans. of the Iron and Steel Institute of Japan* **26**, 212–218 (1986).
- Bartsch, H. A new creep equation for ferritic and martensitic steels. *Steel Research* **66**, 384–388 (1995).
- Prager, M. Development of the MPC Omega method for life assessment in the creep range. *J. Pressure Vessel Technology* **117**, 95–103 (1995).
- Granacher, J., Moehlig, H., Schwienheer, M. & Berger, C. Sa-12-1 (004) creep equations for high temperature materials (inelastic modeling & analysis 2). In *Creep: Proceedings of the International Conference on Creep and Fatigue at Elevated Temperatures*, 1, 609–616 (The Japan Society of Mechanical Engineers, 2001).
- Bishop, C. M. *Pattern recognition and machine learning* (Springer, 2006).
- Nagata, K., Muraoka, R., Mototake, Y., Sasaki, T. & Okada, M. Bayesian spectral deconvolution based on Poisson distribution: Bayesian measurement and virtual measurement analytics (VMA). *Journal of the Physical Society of Japan* **88**, 044003 (2019).
- Mototake, Y., Mizumaki, M., Akai, I. & Okada, M. Bayesian hamiltonian selection in x-ray photoelectron spectroscopy. *Journal of the Physical Society of Japan* **88**, 034004 (2019).
- Izuno, H., Demura, M., Tabuchi, M., Ichi Mototake, Y. & Okada, M. Data-based selection of creep constitutive models for high-creep heat-resistant steel. *Science and Technology of Advanced Materials* **21**, 219–228 (2020).
- Hongo, H., Tabuchi, M. & Watanabe, T. Type IV creep damage behavior in Gr. 91 steel welded joints. *Metallurgical and Materials Trans. A* **43**, 1163–1173 (2012).

15. Holdsworth, S. *et al.* Factors influencing creep model equation selection. *International Journal of Pressure Vessels and Piping* **85**, 80–88 (2008).
16. Kimura, K. Creep rupture life prediction of creep resistant steels. *J. Japan Inst. Metals* **73**, 323–333 (2009).
17. Keitel, H., Dimmig-Osburg, A., Vandewalle, L. & Schueremans, L. Selecting creep models using Bayesian methods. *Materials and Structures* **45**, 1513–1533 (2012).
18. Hukushima, K. & Nemoto, K. Exchange Monte Carlo method and application to spin glass simulations. *Journal of the Physical Society of Japan* **65**, 1604–1608 (1996).
19. Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics* **21**, 1087–1092 (1953).
20. Hastings, W. K. *Monte Carlo sampling methods using Markov chains and their applications* (Oxford University Press, 1970).
21. Nagata, K. & Watanabe, S. Asymptotic behavior of exchange ratio in exchange monte carlo method. *Neural Networks* **21**, 980–988 (2008).
22. Scott, D. W. *Multivariate density estimation: theory, practice, and visualization* (John Wiley & Sons, 2015).
23. Nagata, K., Sugita, S. & Okada, M. Bayesian spectral deconvolution with the exchange monte carlo method. *Neural Networks* **28**, 82–89 (2012).
24. Hayhurst, D. Cdm mechanisms-based modelling of tertiary creep: ability to predict the life of engineering components. *Archives of Mechanics* **57**, 103–132 (2005).
25. Kachanov, L. M. Rupture time under creep conditions. *International Journal of Fracture* **97**, 11–18 (1999).
26. Liu, Y. & Murakami, S. Damage localization of conventional creep damage models and proposition of a new model for creep damage analysis. *JSME International Journal Series A* **41**, 57–65 (1998).
27. Wen, J.-F. & Tu, S.-T. A multiaxial creep-damage model for creep crack growth considering cavity growth and microcrack interaction. *Engineering fracture mechanics* **123**, 197–210 (2014).

Acknowledgements

We would like to thank Dr. M. Yamazaki, Dr. H. Onodera, Mr. J. Sakurai, Dr. K. Koiwa, and Prof. J. Inoue for useful discussions. We are also grateful to Dr. M. Tabuchi for providing the measurement data¹⁴. This work was supported by the Council for Science, Technology and Innovation (CSTI), Cross-Ministerial Strategic Innovation Promotion Program (SIP), “Structural Materials for Innovation” and “Materials Integration for Revolutionary Design System of Structural Materials” (funding agency: JST).

Author contributions

Y.M. developed the machine learning procedures, performed the data analyses, and wrote the manuscript. H.I., K.N., M.D., and M.O. supervised this research. M.D. and K.N. also edited the manuscript. All authors have given approval to the final version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to M.D.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020