Tomohisa Okazaki[1], Yukitoshi Fukahata[2], and Naonori Ueda[1]

[1]RIKEN Center for Advanced Intelligence Project, Japan.

[2]Disaster Prevention Research Institute, Kyoto University, Japan.

Corresponding author: Tomohisa Okazaki (tomohisa.okazaki@riken.jp)

Key Points:

- Gaussian processes are applied to invert time variation of stress fields from centroid moment tensor data in tractable computational costs.
- The relation between stress and centroid moment tensors is analytically derived as a covariance function based on an observation equation.
- The estimated stress field exhibits small-scale heterogeneity in space, whereas exhibits long-term stability in time.

Abstract

This study presents a new method of centroid moment tensor (CMT) data inversion to estimate time-dependent regional stress fields. A Gaussian process (GP) is applied to resolve a computational difficulty of the existing basis function expansion method in analyzing high-dimensional data including time dependence. A critical step in the formulation is an analytical derivation of the relationship of the covariance function, which is a key ingredient of GP, between CMT data and a model stress field based on an observation equation. Applications to CMT data in and around Japan after the 2011 Tohoku earthquake show the efficiency and validity of the method, which clarifies that the stress field has small-scale heterogeneity in space and long-term stability in time for most regions. Additionally, significant temporal variations are observed around the margin of the focal region of the 2011 event, the sense of which is opposite in the landward side and the oceanward side. GP would be particularly effective in geophysical inversions of high-dimensional data distributed in a broad region.

Plain Language Summary

The stress state in the Earth's crust is important to understand the generation of earthquakes and tectonic processes, and can be estimated from inversion analyses of seismological data. In this study, using the Gaussian process in machine learning, we develop a method for estimating 4-D spatiotemporal stress fields, which is difficult to obtain in an existing method due to high computational costs. The Gaussian process, which is suitable for high-dimensional data, has usually been applied to regression and classification problems. To use the Gaussian process for the stress inversion, we derive a physical relation between stress and seismological data. The developed inversion method is applied to centroid moment tensor data after the 2011 Tohoku earthquake, Japan. The 3-D stress fields obtained by the developed and the existing methods are consistent, which guarantees the validity of the developed method. The 4-D stress field obtained

by the developed method shows significant temporal variations around the margin of the large slip area of the Tohoku earthquake. The Gaussian process would be useful for inverse modeling of various high-dimensional geophysical data.

1 Introduction

The stress state in the Earth's crust is crucial to understand the generation of earthquakes and tectonic processes such as mountain building. Because direct measurements of the stress state are costly and limited to shallow parts of the crust, the stress state is commonly estimated from other sources of information. Stress inversion was initially applied to geological data (e.g., Angelier, 1979), and later extended to seismological data of focal mechanisms (e.g., Gephart & Forsyth, 1984; Michael, 1987; Angelier, 2002) and P-wave first motions (e.g., Rivera & Cisternas, 1990; Horiuchi et al., 1995; Abers & Gephart, 2001).

These methods have usually assumed a uniform (0-D) stress state. Therefore, model regions have been subdivided to discuss spatial and temporal variations of the stress field (e.g., Hardebeck & Hauksson, 2001). However, this may lead to apparent changes of the stress field due to scattering of earthquake data. To rectify it, Hardebeck & Michael (2006) developed a damped least-square method of focal mechanism data, which minimized the difference in the stress state between adjacent subareas. This was further extended to a nonlinear inversion with a fully Bayesian formalism (Carlson et al., 2018). Iwata (2018) also constructed a spatially continuous inversion method of P-wave first motion data. These methods were applied to estimate 1-D temporal and 2-D spatial variations of regional stress fields.

The aforementioned inversion methods are based on the Wallace-Bott hypothesis that a slip on a fault occurs in parallel to the direction of the maximum shear stress on that fault (Wallace, 1951; Bott, 1959), but its validity has been controversial (Pollard et al., 1993; Twiss & Unruh, 1998). On the other hand, Terakawa & Matsu'ura (2008) proposed a stress inversion method using centroid moment tensor (CMT) data that did not assume the Wallace-Bott hypothesis. It instead assumes that stress release due to earthquakes is proportional to stress field surrounding them, which is expected to hold statistically (McKenzie, 1969). The formulation of Terakawa & Matsu'ura (2008) naturally incorporated the spatial extent of earthquake faults, and estimated spatially continuous stress fields. Terakawa & Matsu'ura (2010) applied the method to illuminate the 3-D spatial variation of the stress field in and around Japan. The estimated field clarified large-scale stress states caused by plate interactions as well as characteristic local stress patterns. Modeling the temporal variation of regional stress fields may reveal tectonic loading and stress changes due to large earthquakes. Therefore, this study performs the CMT data inversion to estimate 4-D spatiotemporal variation of stress fields.

However, the basis function expansion (BFE) method used in Terakawa & Matsu'ura (2008, 2010) turns out computationally intractable for 4-D data. In their method, the computational cost scales with the number of model pa-

rameters M: memory consumption $\mathcal{O}(M^2)$ and computational amount $\mathcal{O}(M^3)$. M exponentially increases with data dimensions, because basis functions must be placed in each dimension. As a result, computation is prohibitive for high-dimensional data, which is known as the curse of dimensionality (Bellman, 1957). In regression analysis, the Gaussian process (GP) has been applied to overcome this problem (e.g., Rasmussen and Williams, 2006). GP performs a regression based on Bayesian inference by defining covariance functions on the data space. The computational cost of GP scales with the number of data N irrespective of data dimensions: memory consumption $\mathcal{O}(N^2)$ and computational amount $\mathcal{O}(N^3)$. This approach thus resolves the difficulty in analyzing high-dimensional data.

This study applies GP to perform the CMT data inversion. However, there is a critical difference between regression and inversion analyses. Regression analyses construct a fitting model that interpolates (and extrapolates) observational data, which requires data and a model belong to the same physical space. It is therefore sufficient to specify a covariance function in this physical space. In contrast, inversion analyses infer a model that has a different physical unit from observed data. Data and a model must have some physical relationships (i.e. observation equation), and covariance functions are defined on both the data and model spaces. In this study, by deriving an analytical expression of a covariance function that relates CMT data and a stress field, we make it possible to apply GP to CMT data to estimate a stress field. As we will see, this formulation considerably reduces the computational cost in a 3-D analysis, and realizes a 4-D analysis in a tractable computational cost.

Recent studies utilized deep neural networks to estimate focal mechanisms (Kuang et al., 2021), P-wave first motion polarities (Uchide, 2020), and CMT solutions (Steinberg et al., 2021), which constitute datasets for stress inversions. These networks were applied to waveform data of individual earthquakes. On the other hand, this study utilizes GP to invert regional stress fields from CMT data in target regions.

The remainder of the article is organized as follows. Section 2 describes the formulation of the CMT data inversion including its solution methods. In Section 3, these methods are applied to observational data in and around Japan. Section 4 discusses general features of BFE and GP as inversion methods. Section 5 presents conclusions of this study.

2 Formulation

This section formulates the CMT data inversion including the solution methods. The key concept of CMT data inversion and its solution method using BFE are outlined in Sections 2.1 and 2.2, respectively, following the work of Terakawa & Matsu'ura (2008). Section 2.3 presents a review of GP regression (e.g., Rasmussen and Williams, 2006). In Section 2.4, we develop a solution method of the CMT data inversion using GP, and describe the inclusion of time dependence.

2.1 Centroid moment tensor data inversion

The CMT data inversion estimates a stress field from CMT solutions of earthquakes. Based on a premise that stress release due to earthquakes is proportional to the stress field around them, Terakawa & Matsu'ura (2008) derived a relationship between CMT data and a stress field as

$$M_{\mathrm{ij}}(\mathbf{x}, L) = \frac{M_0}{(2\pi L^2)^{3/2}} \int \tau_{\mathrm{ij}}(\ ) \exp\left(-\frac{1}{2L^2}\ \ |\ -\mathbf{x}|^2\right) d\ + E_{\mathrm{ij}}. \quad (1)$$

Here, $M_{\mathrm{ij}}(\mathbf{x}, L)$ is a CMT solution of an earthquake with a centroid position $\mathbf{x}$ and a fault dimension L, M_0 is a seismic moment, $\tau_{\mathrm{ij}}(\ )$ is a stress field at a position , and E_{ij} is an observational error. L is related with the seismic moment M_0 as $L = cM_0^{1/3}$ with a constant c, and incorporates the effect of finite faults into stress fields.

Because estimation errors of CMT data are expected to be proportional to M_0, we normalize Eq. (1) as

$$m_{\mathrm{ij}}(\mathbf{x}, L) = \left(\frac{1}{2\pi L^2}\right)^{3/2} \int \tau_{\mathrm{ij}}(\ ) \exp\left(-\frac{1}{2L^2}\ \ |\ -\mathbf{x}|^2\right) d\ + e_{\mathrm{ij}}, \quad (2)$$

where $m_{\mathrm{ij}}(\mathbf{x}, L)$ is a normalized CMT solution. Then, observational errors obey independent and identically distributed Gaussian distribution: $e_{\mathrm{ij}} \sim \mathcal{N}(0, \sigma_n^2)$. Eq. (2) constitutes an observation equation in the CMT data inversion. In the following, we omit the subscripts ij of tensor components and denote as $\tau(\mathbf{x})$ and $m(\mathbf{x}, L)$ for brevity of notation.

2.2 Basis function expansion

Terakawa & Matsu'ura (2008) solved this inverse problem using BFE as formulated in Yabuki & Matsu'ura (1992). The stress field is expressed as a linear combination $\tau(\mathbf{x}) = \ (\mathbf{x})^T \mathbf{a}$ of M fixed basis functions $\ (\mathbf{x}) = (\Phi_1(\mathbf{x}), \ldots, \Phi_M(\mathbf{x}))^T$ with a vector of model parameters $\mathbf{a}$. The observation equation and prior information are represented as probability distributions of $\mathbf{a}$, using hyperparameters σ_n (observational error) and ρ (penalty on the roughness of a model), respectively. By combining the two probability distributions using Bayes' rule, the optimal values of the model parameters $\hat{\mathbf{a}}$ and its covariance matrix $\mathbf{C}[\hat{\mathbf{a}}]$ are obtained as

$$\hat{\mathbf{a}} = \left(\mathbf{H}^T\mathbf{H} + \alpha^2\mathbf{R}\right)^{-1} \mathbf{H}^T\tilde{\mathbf{d}}, \quad (3)$$

$$\mathbf{C}[\hat{\mathbf{a}}] = \sigma_n^2 \left(\mathbf{H}^T\mathbf{H} + \alpha^2\mathbf{R}\right)^{-1}. \quad (4)$$

Here, $\tilde{\mathbf{d}} = (m_1, \ldots, m_N)^T$ is a vector of N observed CMT data, $\mathbf{H}$ is an $N \times M$ matrix that represents the observation equation, $\mathbf{R}$ is an $M \times M$ matrix that represents prior information, and $\alpha = \sigma_n/\rho$. The optimal values of the hyperparameters are determined by minimizing Akaike's Bayesian information criterion (Akaike, 1980), which is equivalent to maximizing the marginal likelihood in this problem setting. The mean and variance of the stress τ_* at an estimation position $\mathbf{x}_*$ are given by

$$\mu_* = \ (\mathbf{x}_*)^T \hat{\mathbf{a}}\ , \quad (5)$$

$$\sigma_*^2 = \ (\mathbf{x}_*)^T \mathbf{C}[\hat{\mathbf{a}}]\ (\mathbf{x}_*). \quad (6)$$

Computational cost scales with the size of the $M \times M$ matrix $(\mathbf{H}^T\mathbf{H} + \alpha^2\mathbf{R})$ in Eqs. (3) and (4).

2.3 Gaussian process regression

This subsection outlines a method of GP regression using the notation of stress fields. GP is a stochastic process that can be regarded as an infinite-dimensional Gaussian distribution over functions. Suppose that an unknown stress field $\tau(\mathbf{x})$ obeys a GP $\tau \sim \mathrm{GP}(\mu, k)$ with a mean function $\mu(\mathbf{x})$ and a covariance function $k\left(\mathbf{x}_i, \mathbf{x}_j\right)$. This means that $\tau(\mathbf{x})$ at arbitrary finite positions $(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ obeys a multivariate Gaussian distribution:

$$\begin{pmatrix} \tau\left(\mathbf{x}_1\right) \\ \vdots \\ \tau\left(\mathbf{x}_n\right) \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu\left(\mathbf{x}_1\right) \\ \vdots \\ \mu\left(\mathbf{x}_n\right) \end{pmatrix}, \begin{pmatrix} k\left(\mathbf{x}_1, \mathbf{x}_1\right) & \cdots & k\left(\mathbf{x}_1, \mathbf{x}_n\right) \\ \vdots & \ddots & \vdots \\ k\left(\mathbf{x}_n, \mathbf{x}_1\right) & \cdots & k\left(\mathbf{x}_n, \mathbf{x}_n\right) \end{pmatrix} \right). \quad (7)$$

The two functions $\mu\left(\mathbf{x}\right)$ and $k\left(\mathbf{x}_i, \mathbf{x}_j\right)$ describe our prior knowledge on the stress field $\tau\left(\mathbf{x}\right)$ without observational data. Let us set

$$\mu\left(\mathbf{x}\right) = 0, \quad (8)$$

$$k\left(\mathbf{x}_i, \mathbf{x}_j\right) = \sigma_s^2 \exp\left(-\tfrac{1}{2\sigma_l^2} \left|\mathbf{x}_i - \mathbf{x}_j\right|^2\right). \quad (9)$$

Eq. (8) states that stress shows no preferred directions *a priori*. The covariance function in Eq. (9) is called a squared exponential kernel, and states that the stress field smoothly varies with characteristic amplitude σ_s and correlation length σ_l. The prior distribution of stress τ_* at a position $\mathbf{x}_*$ is thus given by:

$$\tau_* \sim \mathcal{N}\left(\mu\left(\mathbf{x}_*\right), k\left(\mathbf{x}_*, \mathbf{x}_*\right)\right) = \mathcal{N}(0, \sigma_s^2). \quad (10)$$

Suppose that we have N direct measurements of the crustal stress $\mathrm{d} = \left(\tau_1, \ldots, \tau_N\right)^T$ with an observational error σ_n, and seek an estimate τ_* at an arbitrary position $\mathbf{x}_*$. This corresponds to a regression problem. According to the model assumption (Eq. 7), the joint distribution of $\mathbf{d}$ and τ_* is given by:

$$\begin{pmatrix} \mathbf{d} \\ \tau_* \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mathbf{0} \\ 0 \end{pmatrix}, \begin{pmatrix} \mathbf{K} + \sigma_n^2\mathbf{I} & \mathbf{k}_* \\ \mathbf{k}_*^T & k_{**} \end{pmatrix} \right), \quad (11)$$

where $\mathbf{K}_{\mathrm{ij}} = k(\mathbf{x}_i, \mathbf{x}_j)$, $\left(\mathbf{k}_*\right)_i = k\left(\mathbf{x}_i, \mathbf{x}_*\right)$, and $k_{**} = k\left(\mathbf{x}_*, \mathbf{x}_*\right)$ are covariance functions evaluated at data and estimation positions, and $\mathbf{I}$ is the $N \times N$ unit matrix. Using the formula of a conditional distribution of Gaussian distributions, the posterior distribution of τ_* for given data $\mathbf{d}$ is calculated as

$$\tau_* | \mathbf{d} \sim \mathcal{N}(\mu_*, \sigma_*^2), \quad (12)$$

$$\mu_* = \mathbf{k}_*^T (\mathbf{K} + \sigma_n^2\mathbf{I})^{-1}\mathbf{d}, \quad (13)$$

$$\sigma_*^2 = k_{**} - \mathbf{k}_*^T (\mathbf{K} + \sigma_n^2\mathbf{I})^{-1}\mathbf{k}_*. \quad (14)$$

The optimal values of the hyperparameters σ_s, σ_l and σ_n are determined by maximizing the marginal likelihood (marginalization over the function values; see Rasmussen and Williams, 2006), which is given by

$$L(\sigma_s, \sigma_l, \sigma_n) = (2\pi)^{-N/2} \left|\mathbf{K} + \sigma_n^2 \mathbf{I}\right|^{-1/2} \exp\left[-\tfrac{1}{2}\mathbf{d}^T \left(\mathbf{K} + \sigma_n^2 \mathbf{I}\right)^{-1} \mathbf{d}\right]. \quad (15)$$

Here, the dependence on σ_s and σ_l is contained in the matrix $\mathbf{K}$ (eq. 9).

As shown in Eqs. (13) and (14), computational cost scales with the size of the $N \times N$ matrix $(\mathbf{K} + \sigma_n^2 \mathbf{I})$ in GP. BFE estimates μ_* and σ_* through the estimation of model parameters $\mathbf{a}$ and $\mathbf{C}(\mathbf{a})$, whereas GP directly estimates μ_* and σ_* from observational data $\mathbf{d}$ without introducing model parameters. This leads to a good scalability in high-dimensional data.

2.4 Gaussian process inversion

We now formulate the CMT data inversion using GP. Suppose that we have CMT solutions of N earthquakes $\tilde{\mathbf{d}} = (m_1, \ldots, m_N)^T$ instead of stress measurements. Then, the joint distribution of $\tilde{\mathbf{d}}$ and τ_* would be written as:

$$\begin{pmatrix} \tilde{\mathbf{d}} \\ \tau_* \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \tilde{} \\ 0 \end{pmatrix}, \begin{pmatrix} \widetilde{\mathbf{K}} + \sigma_n^2 \mathbf{I} & \tilde{\mathbf{k}}_* \\ \tilde{\mathbf{k}}_*^T & k_{**} \end{pmatrix} \right). \quad (16)$$

A major difference from Eq. (11) is that two physical quantities, $m(\mathbf{x}, L)$ and $\tau(\mathbf{x})$, are involved in a single expression. Two points must be clarified: (i) the mean and covariance functions of CMT solutions, $\tilde{}_i = \tilde{\mu}(\mathbf{x}_i, L_i)$ and $\widetilde{\mathbf{K}}_{\mathrm{ij}} = \tilde{k}\left((\mathbf{x}_i, L_i), (\mathbf{x}_j, L_j)\right)$, must be properly defined; (ii) different kinds of variables $\tau(\mathbf{x})$ and $m(\mathbf{x}, L)$ must be properly treated in the cross term $\tilde{\mathbf{k}}_*$.

For (i), we should note that the mean and covariance functions of CMT solutions cannot be arbitrarily defined; instead, they must be related with those of the stress field (Eqs. 8 and 9) through the observation equation (Eq. 2). The expressions of the mean and covariance can be analytically derived as

$$\tilde{\mu}(\mathbf{x}, L) = 0, \quad (17)$$

$$\tilde{k}\left((\mathbf{x}_i, L_i), (\mathbf{x}_j, L_j)\right) = \sigma_s^2 \left(\frac{\sigma_l^2}{\sigma_l^2 + L_i^2 + L_j^2}\right)^{3/2} \exp\left[-\frac{|\mathbf{x}_i - \mathbf{x}_j|^2}{2(\sigma_l^2 + L_i^2 + L_j^2)}\right]. \quad (18)$$

The derivation is presented in Appendix A. Eq. (17) states that CMT solutions show no preferred direction *a priori*. In Eq. (18), the spatial dependence appears only as the distance between centroid positions $|\mathbf{x}_i - \mathbf{x}_j|$, which indicates no special position or direction. The correlation length is $\left(\sigma_l^2 + L_i^2 + L_j^2\right)^{1/2}$: larger earthquakes affects the stress field over a broader area. The variance at $\mathbf{x}_i = \mathbf{x}_j$ is $\sigma_s^2[\sigma_l^2/(\sigma_l^2 + L_i^2 + L_j^2)]^{3/2}$: larger earthquakes have weaker influence on the stress field at the centroid position. These characteristics satisfy the reciprocal relationship defined in eq. (2), in which the contribution of each earthquake to the stress field is independent of seismic moment.

For (ii), as shown in Eq. (2), a CMT solution $m(\mathbf{x}, L)$ is modeled as a weighted average of the stress field $\tau(\mathbf{x})$ using the Gaussian distribution with a characteristic length L. Because the Gaussian distribution converges to Dirac's delta function as $L \to 0$, we obtain the following relation:

$$m(\mathbf{x}, 0) = \int \tau(\ )\, \delta(\ - \mathbf{x}) d\ = \tau(\mathbf{x})\ . \quad (19)$$

This indicates that the stress field $\tau(\mathbf{x})$ can be regarded as a CMT solution $m(\mathbf{x}, 0)$ of length $L = 0$. In other words, the model space (stress) can be regarded as a subspace of the data space (CMT solution). The consistency is certainly satisfied as $\tilde{k}\left((\mathbf{x}_i, 0), (\mathbf{x}_j, 0)\right) = k\left(\mathbf{x}_i, \mathbf{x}_j\right)$, which can be verified from Eqs. (9) and (18). Therefore, the cross term between the data and the model is calculated from the covariance function $\tilde{k}$ on the data space as

$$\left(\tilde{\mathbf{k}}_*\right)_i = \tilde{k}\left((\mathbf{x}_i, L_i), (\mathbf{x}_*, 0)\right) = s^2 \left(\frac{\sigma_l^2}{\sigma_l^2 + L_i^2}\right)^{3/2} \exp\left[-\frac{|\mathbf{x}_i - \mathbf{x}_*|^2}{2(\sigma_l^2 + L_i^2)}\right]. \quad (20)$$

In this way, the procedure of GP regression (Eqs. 12–14) can be applied to the CMT data inversion, under the modification in Eqs. (16)–(20).

When a time-dependent stress field $\tau(\mathbf{x}, t)$ is modeled, the covariance function can be naturally extended as

$$k((\mathbf{x}_i, t_i), (\mathbf{x}_j, t_j)) = \sigma_s^2 \exp\left(-\frac{1}{2\sigma_l^2} |\mathbf{x}_i - \mathbf{x}_j|^2 - \frac{1}{2\sigma_t^2} |t_i - t_j|^2\right), \quad (21)$$

where σ_t is a hyperparameter corresponding to the correlation time. On the other hand, the observation equation (Eq. 2) remains unchanged, considering that stress release should be determined by the stress state at the timing of earthquake occurrences. Then, the covariance function of CMT data $m(\mathbf{x}, t, L)$ is expressed as

$$\tilde{k}\left((\mathbf{x}_i, t_i, L_i), (\mathbf{x}_j, t_j, L_j)\right) = \sigma_s^2 \left(\frac{\sigma_l^2}{\sigma_l^2 + L_i^2 + L_j^2}\right)^{3/2} \exp\left[-\frac{|\mathbf{x}_i - \mathbf{x}_j|^2}{2(\sigma_l^2 + L_i^2 + L_j^2)} - \frac{1}{2\sigma_t^2} |t_i - t_j|^2\right]. \quad (22)$$

3 Applications

We applied the CMT data inversion methods of BFE and GP to real data in and around Japan. The CMT data from just after (not including) the M9.0 Tohoku earthquake (11 March 2011) until 31 December 2019 were obtained from the F-net broadband seismograph network operated by the National Research Institute for Earth Science and Disaster Resilience (NIED). The events satisfying the seismic moment $M_0 \geq 10^{15}$ [Nm] were selected in the analysis; we confirmed that these events follow the Gutenberg-Righter law of magnitude-frequency distributions. A seismic moment was converted to a fault dimension by

$$L = 4 \times 10^{-5} \left(\frac{M_0[\mathrm{Nm}]}{3\pi}\right)^{1/3} [\mathrm{km}], \quad (23)$$

following the empirical relation used in Terakawa & Matsu'ura (2008).

Here, we summarize the display procedure of inversion results. Stress fields are mostly shown at 10 km depth. The beach balls are colored based on the rake angle to distinguish mechanism types (Table 1). We also use the stress regime A_ϕ (Simpson, 1997) to visualize a continuous change of stress fields. This quantity is based on Anderson's theory that one of the three principal axes should be nearly vertical at the ground surface (Anderson, 1905), but the quantity A_ϕ can be discontinuous where this assumption breaks down. Therefore, light colors

are used where the dip angle of the most vertical axis is less than 60° (less than 50° for T axis) on the basis of the criterion of Frohlich (1992).

The certainty of estimations can be measured by the ratio $r = \| \ _*\| / \sigma_*$ of the norm of a posterior mean to a posterior standard deviation. The quantities in the right hand side are defined by Eqs. (5) and (6) in BFE, and by Eqs. (13) and (14) in GP (the norm is computed for the six components of stress). $r > 1$ indicates that the estimated stress pattern is statistically significant, whereas $r \ll 1$ indicates that the estimated stress pattern is unreliable because of inconsistent or too sparse data points. In the following, beach balls are plotted by large markers for $r > 1$ and by small markers for $1 \geq r > 0.1$, and A_ϕ is plotted for $r > 1$.

Table 1. Classification of mechanism types based on the rake angle.

Mechanism type	Range of rake angle (deg)
Normal	(−120, −60)
Reverse	(60, 120)
Normal-oblique	(−150, −120) or (−60, −30)
Reverse-oblique	(30, 60) or (120, 150)
Strike-slip	Otherwise

3.1 Stress field in a local region

First, the two methods are compared with the same dataset in the Tohoku region. The model area is set as the 400 km × 400 km square with the center at (142°E, 39°N) down to 100 km depth, where earthquake occurrences were strongly activated after the 2011 Tohoku earthquake (Fig. 1a, b). The number of data is $N = 2423$. In BFE, tricubic B-splines are placed with 20 km and 10 km intervals in horizontal and vertical directions, respectively (Terakawa & Matsu'ura, 2010), which results in the number of model parameters $M = 9375$. As a result, GP reduces the memory consumption by ~1/15 and computational amount by ~1/60.

The optimal values of the hyperparameters in GP are listed in Table 2. The observational error $\sigma_n = 0.27$ is identical to that in BFE, which implies the consistency of the two methods. The stress amplitude $\sigma_s = 0.42$ is not much larger than σ_n. This may be caused by scattered orientations of earthquake slips, which utilize existing weak faults (McKenzie, 1969). It should be noted that Eq. (2) does not hold for individual events, but is satisfied statistically. Because σ_n contains this scattering of CMT data, σ_n can be comparable to σ_s. The correlation length $\sigma_l = 14$ km indicates small-scale heterogeneities of the stress field.

The estimated stress fields are shown in Figure 1c–f. BFE (Figure 1c, e) and GP (Figure 1d, f) produce similar spatial changes of stress patterns to one another where the estimation is statistically reliable ($r > 1$). This suggests the validity

of GP as a CMT data inversion method. On the other hand, in inland areas with few events, amplitudes of the stress estimated by GP, which is often $r \leq 0.1$, is significantly smaller than that by BEF, which is $r > 0.1$ for the whole area (Figure 1c, d). To see this more clearly, the norm of mean $\| \ _*\|$ and uncertainty σ_* of the estimated stress fields are shown in Figure 2. For regions with many earthquakes, stress fields of a large mean with a small standard deviation are obtained by both methods. For regions with few earthquakes, however, small but finite mean values are estimated by BFE, whereas almost 0 mean values are estimated by GP. This difference results from the difference in prior information, which will be discussed in Section 4.4.

Table 2. List of the optimal values of the hyperparameters.

Section	Region	Number of data	Time dependence	σ_s	σ_n	σ_l (km)	σ_t (year)
	Tohoku		No				–
	Japan		No				–
	Japan		Yes				
	Northeast Japan		Yes				
	Southwest Japan		Yes				

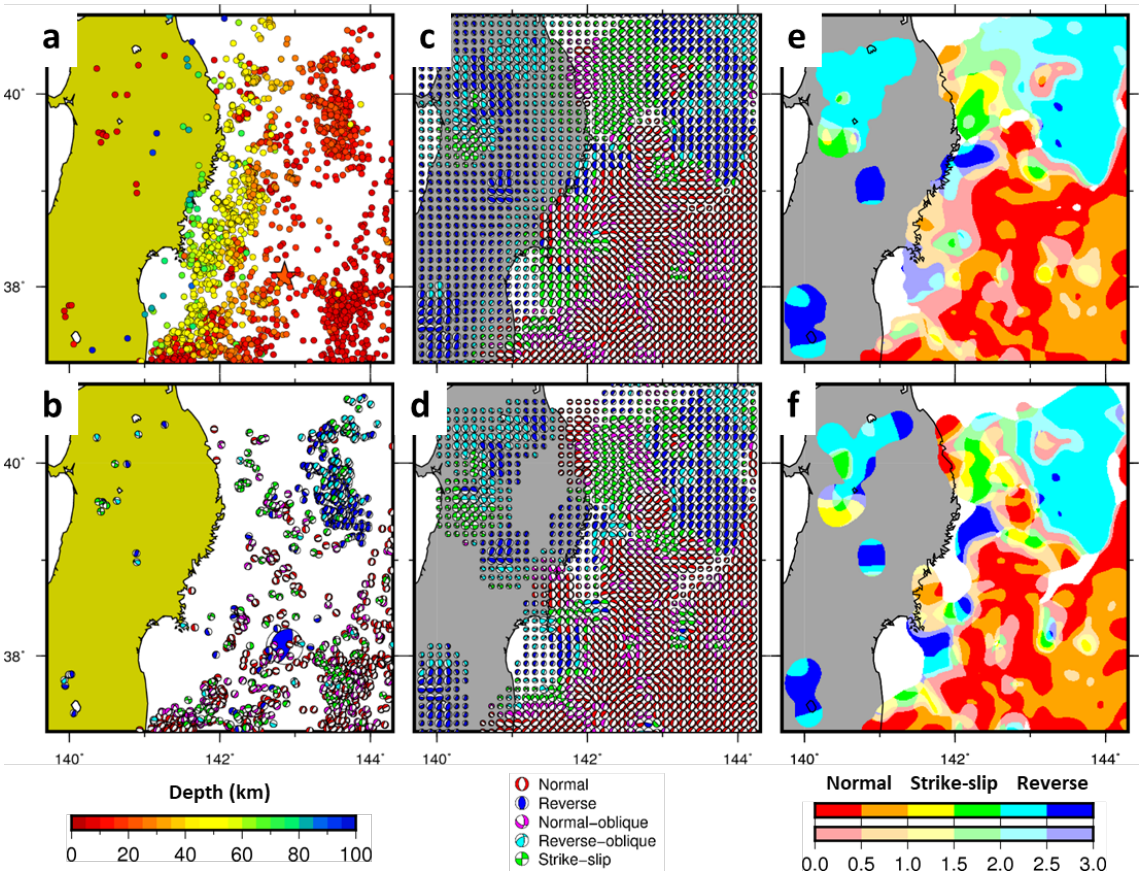

Figure 1. Comparison of the CMT data inversion methods of BFE and GP. (a) Distribution of earthquakes in the Tohoku region. Colors represent focal depths. A large star denotes the centroid location of the 2011 Tohoku earthquake. (b) Focal mechanisms of earthquakes with depths less than 20 km. Colors represent mechanism types. A large beach ball denotes the focal mechanism of the 2011 Tohoku earthquake. (c, d) Beach ball representation of the stress fields estimated by BFE (c) and GP (d). Beach balls are plotted by large markers for the region with $r > 1$, by small markers for $1 \geq r > 0.1$, and no marker for $r \leq 0.1$, where r represents the estimation certainty (see the text). (e, f) Stress regime A_ϕ estimated by BEF (e) and GP (f). A_ϕ is plotted for the region with $r > 1$.

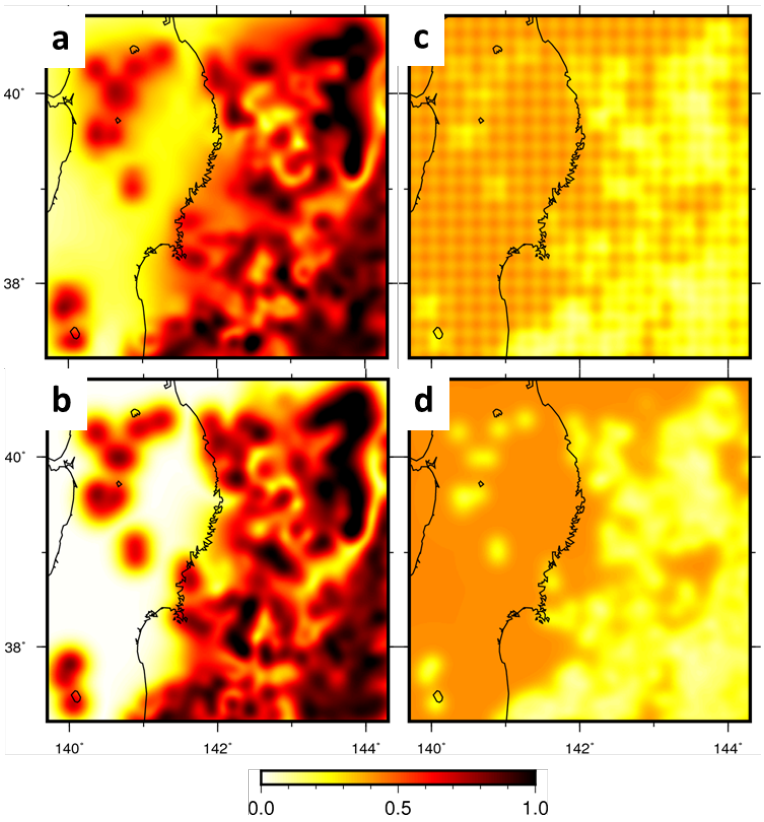

Figure 2. Mean (left) and uncertainty (right) of the stress fields in the Tohoku region estimated by BFE (top) and GP (bottom). (a, c) Norm $\| \ _*\|$of the posterior mean. (b, d) Posterior standard deviation σ_*.

3.2 Stress field in and around Japan

Next, the stress state in and around Japan is estimated only by GP. The number of data is $N = 9965$ (Figure 3a, b), while the model region constructed by Terakawa & Matsu'ura (2010) has $M = 146,848$ basis functions. As a result, GP reduces the memory consumption by ~1/200 and computational amount by ~1/3200. In fact, the model region was divided to nine subareas to reduce the computational cost in Terakawa & Matsu'ura (2010). GP enables us to perform the inversion analysis for the whole area at one time. The locations of seismic events span wider than 20° in latitude and longitude, and so the spherical geometry of the Earth cannot be ignored. Therefore, the spatial distance, $|\mathbf{x}_i - \mathbf{x}_j|$, in covariance functions is calculated according to the geographic coordinate system.

The optimal values of the hyperparameters are listed in Table 2. The correlation length $\sigma_l = 21$ km is larger than that in the Tohoku region. This suggests that the stress state in the Tohoku region has smaller-scale heterogeneities than in

whole Japan. On the other hand, σ_s and σ_n have similar values, which suggests similar scattering of individual focal mechanisms. The estimated stress field is shown in Figure 3c and d. General features are similar to previous studies that analyzed earthquakes before the 2011 great Tohoku event (Terakawa & Matsu'ura, 2010; Yukutake et al., 2015) except for the focal area of the 2011 event where the stress state is a normal fault type in this study. In the Pacific Ocean, the stress state is a reverse fault type along the trench axes and a normal fault type in the outerrise. In the Japanese Islands, the stress state is a reverse fault type in Hokkaido and northeast Japan and a strike-slip fault type in central and southwest Japan.

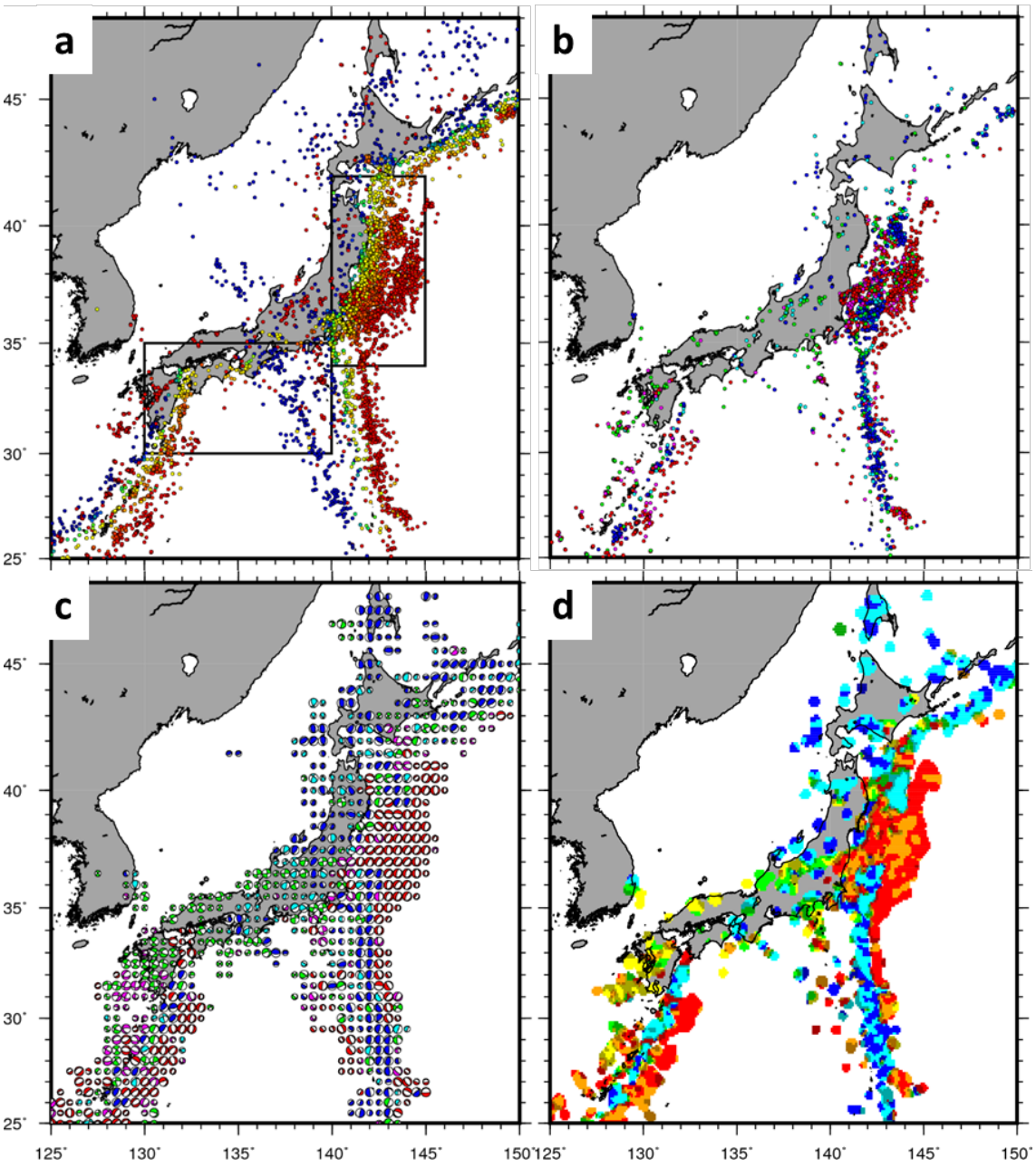

Figure 3. (a) Distribution of earthquakes in and around Japan used in the analysis of Section 3.2. Colors represent focal depths. Two boxes indicate the regions analyzed in Section 4.4. (b) Mechanism types of earthquakes with depths less than 20 km. Colors represent mechanism types. (c) Beach ball representation of stress fields. (d) Stress regime A_ϕ. See Figure 1 for color

representations.

3.3 Time variation of the stress field in and around Japan

Finally, the 4-D spatiotemporal variation of the stress state is estimated in and around Japan. The dataset is the same as in Section 3.2. Therefore, the computational cost of the GP inversion is almost identical to the time-independent modeling in Section 3.2. In contrast, if we perform the BFE inversion, the number of basis functions M is much larger, because the number of basis functions in the time axis is multiplied, which leads to practically too heavy computational cost.

The estimated optimal values of the hyperparameters are listed in Table 2. The optimal values of σ_s, σ_n, and σ_l are almost equal to those in the time-independent modeling, though the optimal value of σ_l slightly increases because σ_l has a negative correlation with σ_t. The optimal value $\sigma_t = 15$ years is longer than the data period (~8.8 years), which suggests that the stress state is stable for a long term. To examine the dependence of the result on the hyperparameters, the log marginal likelihood is plotted against each hyperparameter in Figure 4. σ_s, σ_n, and σ_l are well constrained. As for the correlation time σ_t, a short-term variation ($\sigma_t < 10$ years) is clearly disfavored, whereas a solution with long-term stability ($\sigma_t \to \infty$) is acceptable. This indicates that time variation is basically insignificant during the analysis period.

Figure 5 shows time variation of the estimated stress fields during the analysis period. By setting t_1 at 1 June 2011 and t_2 at 1 June 2019, the change of the stress field and its uncertainty are calculated as

$$\Delta\, {}_*(\mathbf{x}) = {}_*(\mathbf{x}, t_2) - {}_*(\mathbf{x}, t_1), \quad (23)$$

$$\Delta\sigma_*^2(\mathbf{x}) = \sigma_*^2(\mathbf{x}, t_1) + \sigma_*^2(\mathbf{x}, t_2). \quad (24)$$

We should note that this does not represent a change before and after the 2011 event, but represents a temporal change after the 2011 event. The result shows insignificant changes ($r \leq 0.1$) in most regions. Small changes ($r > 0.1$) appear along the trench axes in the Pacific Ocean, and around Niigata and Kumamoto on land. They well correspond to the areas with many seismic events during the analysis period (Figure 3a, b). A significant change ($r > 1$) is observed around the focal area of the 2011 Tohoku earthquake. For comparison, the coseismic slip distribution estimated from teleseismic P-wave data (Yagi & Fukahata, 2011) are plotted in Figure 5c and d. The regions with significant stress change coincide with the margin of large slip areas (~1–10 m). The inner, land-side areas show a change to normal fault types, while the outer, trench-side areas show a change to reverse fault types. Let us focus on a southern part (36–37°N). This area exhibited reverse fault types before the 2011 event (Terakawa & Matsu'ura, 2010), but a significant number of normal fault type aftershocks occurred after the 2011 event (e.g., Imanishi et al., 2012). The estimated result suggests that the stress state has gradually returned to reverse fault types in the off Ibaraki region, whereas remains in normal fault types in the off Fukushima

region, as illustrated by the 2016 M_W 7.0 off Fukushima earthquake. This may indicate a long-lived influence of the stress change in this area.

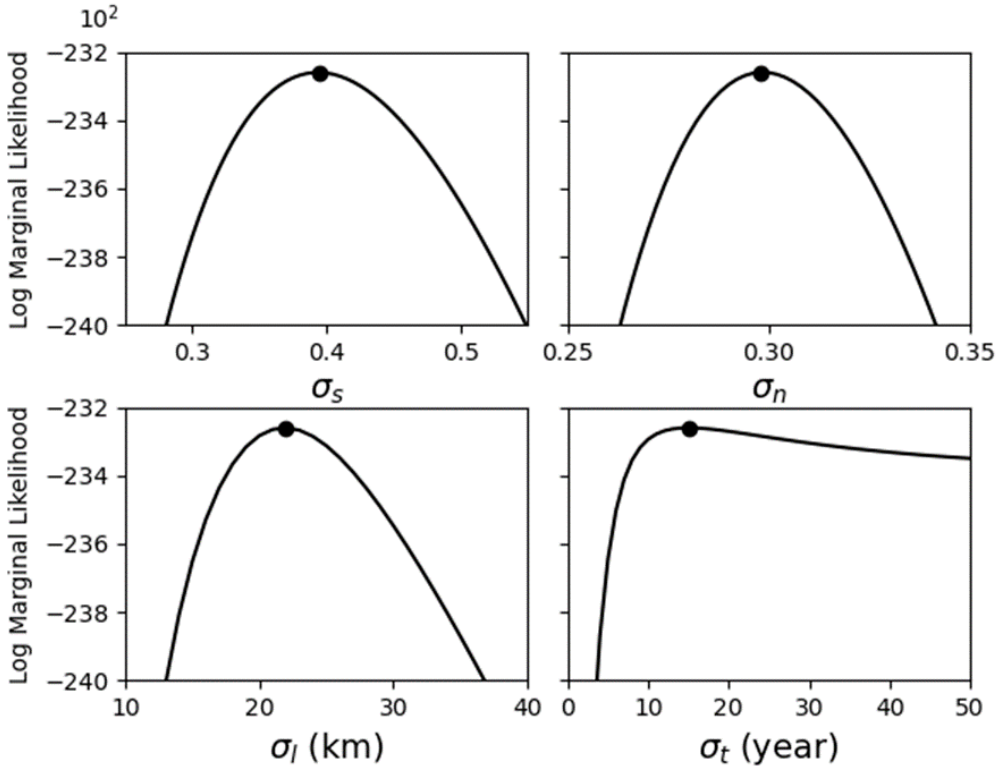

Figure 4. Relation between the log marginal likelihood and the hyperparameters. Dependence on one hyperparameter is plotted by fixing the other hyperparameters at the optimal values given in Table 2.

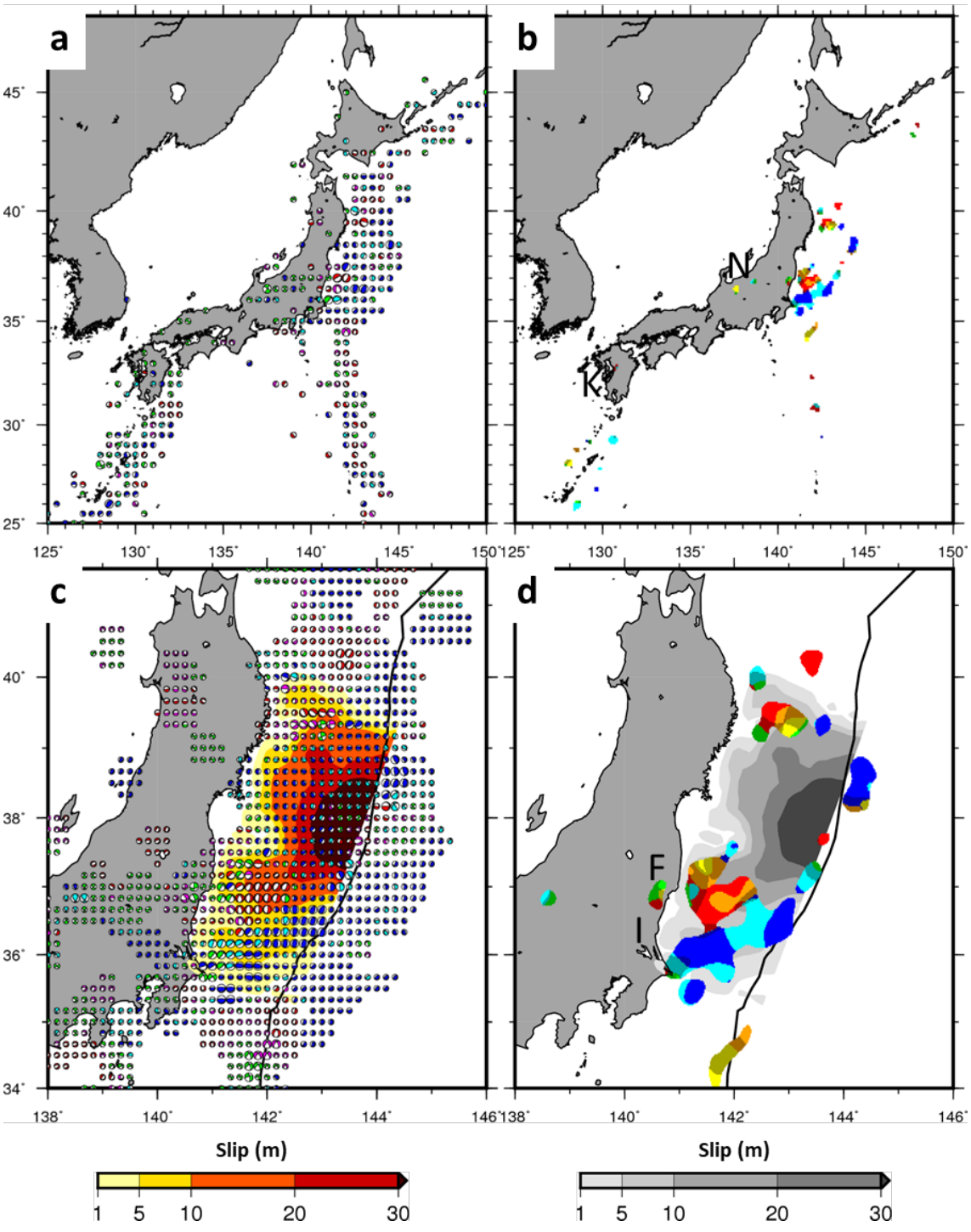
a
b
c
d
N
K
F
I
45°
40°
35°
30°
25°
125°
130°
135°
140°
145°
150°
38°
36°
34°
138°
142°
144°
146°
Slip (m)
1
5
10
20
30
Slip (m)
1
5
10
20
30

Figure 5. Time variation of the estimated stress fields from 1 June 2011 to 1 June 2019. (a) Beach ball representation of the change of the stress fields. (b) Stress regime A_ϕ of the change of the stress fields. (c) Enlarged map of (a) in the Tohoku region. (d) Enlarged map of (b) in the Tohoku region. The trench axis (Bird, 2003) and the estimated slip distribution of the 2011 Tohoku earthquake (Yagi & Fukahata, 2011) are plotted in (c) and (d). See Figure 1 for color representations. F, Fukushima; I, Ibaraki; K, Kumamoto; N, Niigata.

4 Discussion

In this section, we systematically compare general features of BFE and GP as inversion methods.

4.1 Computational cost

A primary motivation of developing a CMT data inversion method with GP is an issue of computational costs. The computational bottleneck of both inversion methods lies in the calculation of inverse matrices. The size of matrices is equal to the number of model parameters M in BFE whereas to the number of data N in GP. The number of data per basis function (N/M) controls the relative efficiency of the two methods. Computational costs and suitable problems of the two methods are summarized in Table 3.

GP is more efficient for $N/M < 1$ (i.e., underdetermined problems). This inequality holds well for problems with high-dimensional, relatively sparse data for large model regions. CMT solutions are high-dimensional (3-D or 4-D) and usually determined only for large earthquakes, and the CMT data inversion is usually suitable for the estimation of background stress fields in a wider area (Iwata et al., 2019). Therefore, the CMT data inversion is a typical problem for which GP is effective. In general, modeling of time dependence adds one additional dimension, which leads to a larger number of M, whereas N is unchanged. Therefore, GP is particularly effective for time-dependent modeling. On the other hand, BFE is more efficient for $N/M > 1$ (i.e., overdetermined problems). For instance, slip inversion on a 2-D fault using dense InSAR data (e.g., Fukahata & Wright, 2008) falls into this category. A suitable inversion method depends on individual problems.

Table 3. Computational costs and suitable problems.

Item	Basis function expansion	Gaussian process
Controlling factor	Number of model parameters M	Number of data N
Memory consumption	$\mathcal{O}(M^2)$	$\mathcal{O}(N^2)$
Computational amount	$\mathcal{O}(M^3)$	$\mathcal{O}(N^3)$
High-dimensional data		
Large model region		
Dense data		

4.2 Model region

In BFE, a set of basis functions are defined before an analysis. Basis functions should be placed densely enough to resolve data variation, and a model region should be taken large enough to mitigate the artifact caused by model boundaries. These requirements must be balanced with an increasing computational cost. In addition, some problems require a careful treatment of basis functions at the model boundaries (Fukahata & Wright, 2008; Okazaki et al., 2021). When continental-scale data are analyzed, basis functions must be set to conform a spherical Earth (Tape et al., 2009). A careful design is necessary to suppress modeling errors arising from the setting of model regions.

On the other hand, GP does not require to set a model region. Prior information and data distributions automatically determine the range of reliable estimations. The sphericity of the Earth can be easily incorporated by computing the geographical distance between two positions appropriately (Section 3.2). GP is free from modeling errors related with the setting of model regions.

4.3 Hyperparameters

Both methods introduce hyperparameters in modeling. A common parameter σ_n represents observational errors in data. The other hyperparameters are different from one another. ρ in BFE is an abstract parameter that controls the relative importance of data fitting and model smoothness. This type of parameter has been used in other stress inversions (Hardebeck & Michael, 2006; Iwata, 2018). In contrast, covariance functions of GP define physical hyperparameters: characteristic amplitude σ_s, correlation length σ_l, and correlation time σ_t. Therefore, optimal values of them provide direct insights on the nature of inverted fields, as described in Section 3.

In both methods, optimal values of the hyperparameters are determined by maximizing the marginal likelihood. In a time-independent model, BFE has two hyperparameters. In fact, the optimization can be analytically solved for one variable (Yabuki & Matsu'ura, 1992), and the grid search is needed for only one variable. GP has three hyperparameters, and the grid search requires more explorations. Nevertheless, the gradient of the marginal likelihood can be analytically computed for GP, and the gradient ascent method can be applied for a model with many hyperparameters.

4.4 Prior information

Both methods are based on Bayesian inference, and prior information is essential for stable estimations. In BFE, prior information is represented as a probability distribution of model parameters. Yabuki & Matsu'ura (1992) imposed a penalty on the roughness of solutions, which has been widely adopted in later applications. These constraints are called indirect prior information (Matsu'ura et al., 2007), and results in the interpolation and extrapolation of solutions into no data areas. This produces a plausible result in a whole model region, but may yield a poor-constrained model where data are sparse. In GP, on the other

hand, prior information is represented as mean and covariance functions on the data space. It is known that GP regression can be regarded as Bayesian linear regression using (possibly an infinite number of) basis functions whose prior distribution penalizes the magnitude of model parameters (Rasmussen & Williams, 2006). This corresponds to direct prior information in Matsu'ura et al. (2007). A posterior mean μ_* converges to a prior mean μ (typically zero) beyond the range of correlation from data points. Therefore, a nontrivial posterior mean suggests the significance of estimates, but no information is obtained where data are sparse. Direct prior information can be easily introduced in BFE, whereas it cannot be easily removed in GP because the convergence to a prior mean is a fundamental property.

As shown in Table 2, the optimal value of the correlation length σ_l is significantly different between the Tohoku region (Section 3.1) and whole Japan (Section 3.2). This suggests a spatial dependence of the characteristic length on regions. As an experiment, optimal values of the hyperparameters are searched for northeast Japan (many earthquakes) and southwest Japan (fewer earthquakes) shown in Figure 3a. The obtained correlation length σ_l is approximately three times longer in southwest Japan than in northeast Japan (Table 2). This indicates that a single hyperparameter value cannot correctly capture the variation of stress fields with different spatial scales. Spatial (and temporal) uniformity of prior information is a common limitation on the present modeling in both methods. In BFE, Fukuda et al. (2004, 2008) introduced a time-varying smoothing regularization in the inversions of transient crustal deformation, although its application to real data still seems to be not easy (Kano et al., 2018; Sakaue et al., 2019). In GP, the limitation results from the use of a stationary covariance function that depends only on the difference $(\mathbf{x}_i - \mathbf{x}_j)$ and $(t_i - t_j)$. Introducing a nonstationary covariance functions (e.g. Paciorek & Schervish, 2004) may resolve this limitation.

5 Conclusions

A continuous spatiotemporal variation of the seismogenic stress field was estimated from the CMT data inversion. Because an inversion with BFE is computationally prohibitive for 4-D data, an inversion method based on GP was developed to perform inversions in a tractable computational cost. An essential step in the formulation is the analytical derivation of the covariance function of CMT solutions (Eq. 18) from that of stress (Eq. 9) and the observation equation (Eq. 2).

The analysis in a local region (Section 3.1) showed that the two methods produced consistent estimations where data were dense, whereas a significant difference in the estimation certainty was found where data were sparse. This resulted from the difference in prior information (Section 4.4). GP enabled us to perform an inversion analysis for whole Japan at one time (Section 3.2). The estimated stress field exhibited a small-scale (~20 km) heterogeneity. GP also enabled a 4-D time-dependent modeling in a similar computational cost to a 3-D time-independent modeling (Section 3.3). The estimated stress field exhib-

ited a longer term stability than the data period, although a significant time variation was found around the margin of the large slip area of the 2011 Tohoku earthquake.

As an inversion method, GP does not require the design of model regions, and its hyperparameters has clear physical meanings. GP instead requires more hyperparameters to be optimized than the BFE method, and estimated fields inevitably converge to a prior mean where data are sparse. The inversion with GP is computationally efficient in a high-dimensional modeling. This approach would therefore be effective to estimate 3-D and/or 4-D field variations in various geophysical data inversions.

Appendix A

In this appendix, we derive the mean and covariance functions $\tilde{\mu}$ and $\tilde{k}$ (Eqs. 17 and 18) of CMT solutions $m(\mathbf{x}, L)$ from those of stress $\tau(\mathbf{x})$ (Eqs. 8 and 9). First, we consider the mean function $\tilde{\mu}$. Because $m(\mathbf{x}, L)$ is linearly related to $\tau(\mathbf{x})$ (Eq. 2) that obeys a GP, $m(\mathbf{x}, L)$ also obeys a GP. Hence, the mean function is calculated from the observation equation (Eq. 2) as

$\tilde{\mu}(\mathbf{x}, L) = \mathbb{E}[m(\mathbf{x}, L)] = \mathbb{E}\left[\left(\frac{1}{2\pi L^2}\right)^{\frac{3}{2}} \int \tau(\ ) \exp\left(-\frac{1}{2L^2} \ | \ -\mathbf{x}|^2\right) d \ \right]$. (A1). Because the expectation value $\mathbb{E}[\cdot]$ operates only on a random variable $\tau(\ )$, we obtain

$\tilde{\mu}(\mathbf{x}, L) = \left(\frac{1}{2\pi L^2}\right)^{\frac{3}{2}} \int \mathbb{E}[\tau(\ )] \exp\left(-\frac{1}{2L^2} \ | \ -\mathbf{x}|^2\right) d \ = 0$, (A2)

For the second equal sign, we used Eq. (8): $\mathbb{E}[\tau(\ )] = \mu(\mathbf{x}) = 0$. Similarly, the covariance function is calculated using Eq. (2) as

$\tilde{k}\left((\mathbf{x}_i, L_i), (\mathbf{x}_j, L_j)\right) = \mathbb{E}\left[m(\mathbf{x}_i, L_i) m(\mathbf{x}_j, L_j)\right] = \left(\frac{1}{2\pi L_i L_j}\right)^3 \iint \mathbb{E}[\tau(\ _i)\, \tau(\ _j)] \exp\left(-\frac{1}{2L_i^2} \ |\ _1 - \mathbf{x}_1|^2 - \frac{1}{2L_j^2} \ |\ _2 - \mathbf{x}\right.$ (A3)

Substituting $\mathbb{E}\left[\tau(\ _i)\, \tau(\ _j)\right] = k(\ _i, \ _j) = \sigma_s^2 \exp\left(-\frac{1}{2\sigma_l^2} \ |\ _i - \ _j|^2\right)$ (Eq. 9) for Eq. (A3), we obtain

$\tilde{k}\left((\mathbf{x}_i, L_i), (\mathbf{x}_j, L_j)\right) = \sigma_s^2 \left(\frac{1}{2\pi L_i L_j}\right)^3 \iint \exp\left(-\frac{1}{2\sigma_l^2} \ |\ _i - \ _j|^2 - \frac{1}{2L_i^2} \ |\ _i - \mathbf{x}_i|^2 - \frac{1}{2L_j^2} \ |\ _j - \mathbf{x}_j|^2\right) d\ _i d\ _j$. (A4)

Finally, integration with respect to $\ _i$ and $\ _j$ by completing the square in the exponential function gives the desired result (Eq. 18).

Acknowledgments

The authors declare no competing interests. Y. Yagi provided the result of a coseismic slip distribution caused by the 2011 Tohoku earthquake.

Open Research

The CMT data were obtained from the National Research Institute for Earth Science and Disaster Resilience (NIED) broadband seismograph network F-net (https://www.fnet.bosai.go.jp, last accessed March 2022).

References

Abers, G. A., & Gephart, J. (2001). Direct inversion of earthquake first motions for both the stress tensor and focal mechanisms and application to southern California. *Journal of Geophysical Research*, 106, 26,523–26,540. https://doi.org/10.1029/2001JB000437

Akaike, H. (1980). Likelihood and Bayes procedure. In J. M. Bernardo, M. H. DeGroot, D. V. Lindley, & A. F. M. Smith (Eds.), *Bayesian Statistics* (pp. 143–166), University Press, Valencia.

Anderson, E. M. (1905). The dynamics of faulting, *Transactions of the Edinburgh Geological Society*, *8*, 387–402, https://doi.org/10.1144/transed.8.3.387

Angelier, J. (1979). Determination of the mean principal directions of stresses for a given fault population. *Tectonophysics, 56*(3–4), T17–T26. https://doi.org/10.1016/0040-1951(79)90081-7

Angelier, J. (2002). Inversion of earthquake focal mechanisms to obtain the seismotectonic stress IV—a new method free of choice among nodal planes. *Geophysical Journal International, 150*, 588–609. https://doi.org/10.1046/j.1365-246X.2002.01713.x

Bellman, R. (1957). *Dynamic programming*, Princeton University Press.

Bird, P. (2003). An updated digital model of plate boundaries. *Geochemistry. Geophysics, Geosystems, 4*(3), 1027. https://doi.org/10.1029/2001GC000252

Bott, M. H. P. (1959). The mechanics of oblique slip faulting. *Geological Magagine, 96*, 109–117.

Carlson, G., Johnson, K., Chuang, R., & Rupp, J. (2018). Spatially varying stress state in the central U.S. from Bayesian inversion of focal mechanism and in situ maximum horizontal stress orientation data. *Journal of Geophysical Research: Solid Earth, 123*, 3871–3890. https://doi.org/10.1002/2017JB015158

Frohlich, C. (1992). Triangle diagrams: ternary graphs to display similarity and diversity of earthquake focal mechanisms. *Physics of the Earth and Planetary Interiors, 75*(1–3), 193–198. https://doi.org/10.1016/0031-9201(92)90130-N

Fukahata, Y. & Wright, T. J. (2008). A non-linear geodetic data inversion using ABIC for slip distribution on a fault with an unknown dip angle, *Geophysical Journal International, 173*(2), 353–364. https://doi.org/10.1111/j.1365-246X.2007.03713.x

Fukuda, J., Higuchi, T., Miyazaki, S., Kato, T. (2004). A new approach to time-dependent inversion of geodetic data using a Monte Carlo mixture Kalman filter, *Geophysical Journal International*, 159(1), 17–39. https://doi.org/10.1111/j.1365-246X.2004.02383.x

Fukuda, J., Miyazaki, S., Higuchi, T., Kato, T. (2008). Geodetic inversion for space-time distribution of fault slip with time-varying smooth-

ing regularization, *Geophysical Journal International*, 173(1), 25–48. https://doi.org/10.1111/j.1365-246X.2007.03722.x

Gephart, J., & Forsyth, D. W. (1984). An improved method for determining the regional stress tensor using earthquake focal mechanism data: Application to the San Fernando Earthquake Sequence. *Journal of Geophysical Research, 89*, 9305–9320. https://doi.org/10.1029/JB089iB11p09305

Hardebeck, J. L., & Hauksson, E. (2001). Crustal stress field in southern California and its implications for fault mechanics. *Journal of Geophysical Research, 106*, 21,859–21,882. https://doi.org/10.1029/2001JB000292

Hardebeck, J. L., & Michael, A. J. (2006). Damped regional-scale stress inversions: Methodology and examples for southern California and the Coalinga aftershock sequence. *Journal of Geophysical Research, 111*, B11310. https://doi.org/10.1029/2005JB004144

Horiuchi, S., Rocco, G., & Hasegawa, A. (1995). Discrimination of fault planes from auxiliary planes based on simultaneous determination of stress tensor and a large number of fault plane solutions. *Journal of Geophysical Research*, *100*, 8327–8338. https://doi.org/10.1029/94JB03284

Imanishi, K., Ando, R., & Kuwahara, Y. (2012). Unusual shallow normal-faulting earthquake sequence in compressional northeast Japan activated after the 2011 off the Pacific coast of Tohoku earthquake, *Geophysical Research Letters, 39*, L09306. https://doi.org/10.1029/2012GL051491

Iwata, T. (2018). A Bayesian approach to estimating a spatial stress pattern from P wave first-motions. *Journal of Geophysical Research: Solid Earth, 123*, 4841–4858. https://doi.org/10.1002/2017JB015359

Iwata, T., Yoshida, K., & Fukahata, Y. (2019). Stress Tensor Inversion Using Seismological Data. *Journal of Geography, 128*(5), 797-811. https://doi.org/10.5026/jgeography.128.797 (in Japanese with English abstract)

Kano, M., Fukuda, J., Miyazaki, S., & Nakamura, M. (2018). Spatiotemporal evolution of recurrent slow slip events along the southern Ryukyu subduction zone, Japan, from 2010 to 2013. *Journal of Geophysical Research: Solid Earth, 123*, 7090–7107. https://doi.org/10.1029/2018JB016072

Kuang, W., Yuan, C., & Zhang, J. (2021). Real-time determination of earthquake focal mechanism via deep learning. *Nature Communications*, *12*(1), 1432. https://doi.org/10.1038/s41467-021-21670-x

Matsu'ura, M., Noda, A., & Fukahata, Y. (2007). Geodetic data inversion based on Bayesian formulation with direct and indirect prior information, *Geophysical Journal International, 171*(3), 1342–1351. https://doi.org/10.1111/j.1365-246X.2007.03578.x

McKenzie, D. P. (1969). The relation between fault plane solutions for earthquakes and the directions of the principal stresses. *Bulletin of the Seismological Society of America, 59*, 591–601.

Michael, A. J. (1987). Use of focal mechanisms to determine stress: A control study. *Journal of Geophysical Research, 92*, 357–368. https://doi.org/10.1029/JB092iB01p00357

Okazaki, T., Fukahata, Y., & Nishimura, T. (2021). Consistent estimation of strain-rate fields from GNSS velocity data using basis function expansion with ABIC. *Earth, Planets and Space, 73*, 153. https://doi.org/10.1186/s40623-021-01474-5

Paciorek, C. J., & Schervish, M. J. (2004). Nonstationary Covariance Functions for Gaussian Process Regression. In *Advances in Neural Information Processing Systems, 16*, 273–280.

Pollard, D. D., Saltzer, S. D., & Rubin, A. M. (1993). Stress inversion methods: are they based on faulty assumptions?. *Journal of Structural Geology*, *15*(8), 1045–1054. https://doi.org/10.1016/0191-8141(93)90176-B

Rasmussen, C., & Williams, C. (2006). *Gaussian Processes for Machine Learning*, MIT Press.

Rivera, L., & Cisternas, A. (1990). Stress tensor and fault plane solutions for a population of earthquakes. *Bulletin of the Seismological Society of America, 80*, 600–614. https://doi.org/10.1785/BSSA0800030600

Sakaue, H., Nishimura, T., Fukuda, J., & Kato, T. (2019). Spatiotemporal evolution of long- and short-term slow slip events in the Tokai region, central Japan, estimated from a very dense GNSS network during 2013–2016. *Journal of Geophysical Research: Solid Earth, 124*, 13207–13226. https://doi.org/10.1029/2019JB018650

Simpson, R. W. (1997). Quantifying Anderson's fault types. *Journal of Geophysical Research, 102*(B8), 17909–17919. https://doi.org/10.1029/97JB01274

Steinberg, A., Vasyura-Bathke, H., Gaebler, P., Ohrnberger, M., & Ceranna, L. (2021). Estimation of seismic moment tensors using variational inference machine learning. *Journal of Geophysical Research: Solid Earth*, 126, e2021JB022685. https://doi.org/10.1029/2021JB022685

Tape, C., Musé, P., Simons, M., Dong, D., Webb, F. (2009). Multiscale estimation of GPS velocity fields. *Geophysical Journal International, 179*(2), 945–971. https://doi.org/10.1111/j.1365-246X.2009.04337.x

Terakawa, T., & Matsu'ura, M. (2008). CMT data inversion using a Bayesian information criterion to estimate seismogenic stress fields. *Geophysical Journal International*, *172*, 674–685. https://doi.org/10.1111/j.1365-246X.2007.03656.x

Terakawa, T., & Matsu'ura, M. (2010). The 3-D tectonic stress fields in and around Japan inverted from centroid moment tensor data of seismic events. *Tectonics*, *29*, TC6008. https://doi.org/10.1029/2009TC002626.

Twiss, R. J., & Unruh, J. R. (1998). Analysis of fault slip inversions: do they constrain stress or strain rate?. *Journal of Geophysical Research*, *103*(B6), 12205–12222. https://doi.org/10.1029/98JB00612

Uchide, T. (2020). Focal mechanisms of small earthquakes beneath the Japanese islands based on first-motion polarities picked using deep learning. *Geophysical Journal International*, 223(3), 1658–1671. https://doi.org/10.1093/gji/ggaa401

Wallace, R. E. (1951). Geometry of shearing stress and relation to faulting. *Journal of Geology, 59*, 118–130.

Yabuki, T. & Matsu'ura M. (1992). Geodetic data inversion using a Bayesian information criterion for spatial distribution of fault slip, *Geophysical Journal International, 109*, 363–375. https://doi.org/10.1111/j.1365-246X.1992.tb00102.x

Yagi, Y., & Fukahata, Y. (2011). Rupture process of the 2011 Tohoku-oki earthquake and absolute elastic strain release, *Geophysical Research Letters, 38*, L19307, https://doi.org/10.1029/2011GL048701

Yukutake, Y., Takeda, T., & Yoshida, A. (2015). The applicability of frictional reactivation theory to active faults in Japan based on slip tendency analysis. *Earth and Planetary Science Letters, 411*, 188–198. https://doi.org/10.1016/j.epsl.2014.12.005