Rapid and quantitative uncertainty estimation of coseismic slip distribution for large interplate earthquakes using real-time GNSS data and its application to tsunami inundation prediction

Keitaro Ohno1##*, Yusaku Ohta1,2,3, Ryota Hino1,3, Shunichi Koshimura3,

Akihiro Musa4,5, Takeshi Abe3, Hiroaki Kobayashi6

1Research Center for Prediction of Earthquakes and Volcanic Eruptions, Graduate School of Science, Tohoku University, Sendai, Japan
2Division for the Establishment of Frontier Sciences of Organization for Advanced Studies, Tohoku University, Sendai, Japan
3International Research Institute of Disaster Science, Tohoku University, Sendai, Japan
4Cyberscience Center, Tohoku University, Sendai, Japan
5NEC Corporation, Tokyo, Japan
6Computer Architecture Laboratory, Graduate School of Information Sciences, Tohoku University, Sendai, Japan

#Now at Geospatial Information Authority of Japan, Tsukuba, Japan

*Corresponding author: Keitaro Ohno, Geospatial Information Authority of Japan, 1 Kitasato, Tsukuba 305-0811, Japan, ohno-k96pi@mlit.go.jp
Abstract

This study proposes a new method for the uncertainty estimation of coseismic slip distribution on the plate interface deduced from real-time global navigation satellite system (GNSS) data and explores its application for tsunami inundation prediction. Jointly developed by the Geospatial Information Authority of Japan and Tohoku University, REGARD (REal-time GEONET Analysis system for Rapid Deformation monitoring) estimates coseismic fault models (a single rectangular fault model and slip distribution model) in real time to support tsunami prediction. The estimated results are adopted as part of the Disaster Information System, which is used by the Cabinet Office of the Government of Japan to assess tsunami inundation and damage. However, the REGARD system currently struggles to estimate the quantitative uncertainty of the estimated result, although the obtained result should contain both observation and modeling errors caused by the model settings. Understanding such quantitative uncertainties based on the input data is essential for utilizing this resource for disaster response. We developed an algorithm that estimates the coseismic slip distribution and its uncertainties using Markov chain Monte Carlo methods. We focused on the Nankai Trough of southwest Japan, where megathrust earthquakes have repeatedly occurred, and used simulation data to assume a Hoei-type earthquake. We divided the 2951 rectangular subfaults on the plate interface and designed a multistage sampling flow with stepwise perturbation groups. As a result, we successfully estimated the slip distribution and its uncertainty at the 95% confidence interval of the posterior probability density function. Furthermore, we developed a new visualization procedure that shows the risk of tsunami inundation and the probability on a map. Under the
algorithm, we regarded the Markov chain Monte Carlo samples as individual fault models and clustered them using the k-means approach to obtain different tsunami source scenarios. We then calculated the parallel tsunami inundations and integrated the results on the map. This map, which expresses the uncertainties of tsunami inundation caused by uncertainties in the coseismic fault estimation, offers quantitative and real time insights into possible worst-case scenarios.

**Keywords**
Tsunami inundation, Tsunami prediction, Real-time GNSS, Bayesian inversion, Uncertainty estimation, MCMC, REGARD

**Main Text**

1. Introduction

A rapid understanding of the magnitude and fault areas of large earthquakes is crucial for disaster response. Numerous studies have shown the advantages of the high-rate (typically 1 Hz) global navigation satellite system (GNSS) as a broadband sensor that can directly measure displacement without saturation (e.g., Larson et al. 2003; Ohta et al. 2006; Larson 2009). The use of onshore high-rate GNSS data enable the rapid estimation of offshore finite fault models that are expected to be used for tsunami prediction (e.g., Blewitt et al. 2006 2009; Ohta et al. 2012; Melgar et al. 2012; Tsushima et al., 2014; Tsushima and Ohta 2014; Ohta 2016). The Geospatial Information Authority of Japan (GSI) operates a nationwide continuous GNSS network called GEONET. After the massive tsunami caused by the 2011 Tohoku-Oki earthquake, the
GSI and Tohoku University jointly developed the REGARD system (REal-time GEONET Analysis system for Rapid Deformation monitoring, Kawamoto et al. 2016), which consists of real-time GNSS positioning, automatic detection of coseismic displacement, and quasi-real-time finite fault model inversion routines.

The finite fault models estimated by REGARD are helpful for obtaining the initial sea surface distribution for tsunami forecasting. The consortium led by Tohoku University has developed a real-time damage estimation system for tsunami inundation using the REGARD fault model as an initial tsunami source model (Musa et al. 2018; Ohta et al. 2018). This system is expected to be used for the initial response of the government when a disaster occurs.

For the 2016 Kumamoto earthquake (Mw 7.0), REGARD successfully estimated a single rectangular fault model automatically in real time (Kawamoto et al. 2016). However, estimation using only onshore GNSS data does not necessarily provide accurate estimates of offshore fault areas. From this point of view, in the single rectangular fault model, Ohno et al. (2021) developed a new method to quantitatively evaluate the uncertainties of the estimated results using the Bayesian inversion approach. This study showed a tradeoff between the fault area and slip amount, especially for offshore earthquakes, which affects the initial wave field of the tsunami.

In the slip distribution model of REGARD, quantitative evaluation of the uncertainty is also essential because of the effects of station arrangement, modeling errors, and constraints on slip smoothing. For example, in the case of the 2011 Tohoku-Oki earthquake, it was suggested that there was a large slip around the trench axis based on seafloor geodetic data (e.g., Iinuma et al. 2012), although such slip was not significantly suggested using only onshore GNSS data (e.g., Melgar and Bock 2015).
The quantification of the uncertainty of fault models enables us to consider the accuracy of tsunami predictions calculated based on them. With the underestimation of tsunami wave height in the 2011 Tohoku-Oki earthquake as a lesson, the danger of announcing tsunami predictions as a single value has been proposed. The Japan Meteorological Agency (JMA) introduced the announcement of tsunami warnings with more qualitative expressions (JMA 2012). In addition, several studies have shown a method for adding uncertainty information to tsunami prediction (e.g., Sraj et al. 2014; Fukutani et al. 2015; Dettmer et al. 2016; Goda and Song 2016; Gibbons et al. 2020; Goda et al. 2020; Mulia et al. 2020; Giles et al. 2021). Dettmer et al. (2016) evaluated the uncertainty of the initial sea surface displacement in the 2011 Tohoku-Oki earthquake deduced from tsunami waveforms using a transdimensional algorithm based on a wavelet decomposition of the displacement field. Goda et al. (2020) presented an extensive tsunami hazard assessment for Nankai–Tonankai trough events using 1000 kinematic earthquake rupture models for Monte Carlo tsunami simulations. Giles et al. (2021) proposed a workflow that integrates the entire chain of components from the tsunami source to quantify hazard uncertainties by approximating the functionally complex and computationally expensive high-resolution tsunami simulations with a simple and cheap statistical emulator. On the other hand, in the damage estimation system due to tsunami inundation, it is currently difficult to evaluate the impact of the uncertainty of fault models on tsunami prediction. With this background, the purpose of this study was to add uncertainty information to tsunami inundation estimation by quantifying the uncertainty of the coseismic slip distribution model estimated from onshore real-time GNSS data. We propose a "stepwise partitioning algorithm," for estimating coseismic slip distribution using
Markov chain Monte Carlo methods (MCMC). Furthermore, we propose a "real-time tsunami inundation risk map," which visualizes the probability of risk of tsunami inundation based on the uncertainty of the estimated fault model. We verified the performance of this approach by applying it to simulated data in the Nankai Trough area.

2. Methods

2.1. Stepwise partitioning algorithm

Conventionally, the slip distribution model is estimated with a hyper-parameter governing the intensity of slip smoothing determined with an index such as ABIC (Akaike's Bayesian Information Criterion, Akaike 1980; e.g., Fukahata 2009). The error matrix obtained in this case is essential for interpreting the uncertainty. However, this depends on the smoothing constraint conditions. Recently, many studies have adopted the Bayesian approach to geodetic studies because of its advantages in quantifying uncertainty (Ito et al. 2012; Dettmer et al. 2014; Minson et al. 2014a 2014b; Jiang and Simons 2016; Ohno et al. 2018; Ohno et al. 2021). Owing to its sampling process, MCMC can generate a large number of model samples without smoothing constraints. In this study, we did not use smoothing constraints and regarded the diversity of samples as individual slip distribution models that explain the data to apply to the tsunami inundation calculation.

We adopted simulated data based on the assumed Mw 8.75 1707 Hoei earthquake (Furumura et al. 2011; Todoriki et al. 2013; Inoue et al. 2016; Kawamoto et al. 2017). The assumed slip reached from Suruga Bay to the westernmost part of Shikoku (Figure 1). Let $\mathbf{\theta}$ be a model parameter vector that contains slip amounts on subfaults along
the plate boundary in the Nankai Trough, and \( \mathbf{d} \) be a permanent displacement data vector with three components (horizontal and vertical components) based on real-time GNSS observations. We used the same method as Ohno et al. (2021) for MCMC sampling. The likelihood function measures the degree of fit between the observed data \( \mathbf{d} \), and calculated data \( \mathbf{d}^\ast(\mathbf{\theta}) \). The residuals are given by \( \mathbf{r}(\mathbf{\theta}) = \mathbf{d}(\mathbf{\theta}) - \mathbf{d} \). When the number of observation stations is \( N \), the dimension of \( \mathbf{d} \) is \( 3N(N_{EN} = 2N \) and \( N_U = N \). Assuming that the estimation error follows a normal distribution function and that the standard deviations \( \sigma_{ij} \) (\( i = East \ (E), North \ (N), Up\ Down \ (U) \)) at every GNSS station are identical for each horizontal (EN) and vertical (U) components, the likelihood function is defined as follows:

\[
p(\mathbf{d}|\mathbf{\theta}) = \prod_{i=EN,U} \prod_{j=1}^{N} \frac{1}{\sqrt{2\pi \sigma_{ij}^2}} \exp \left[ -\frac{1}{2\sigma_{ij}^2} \mathbf{r}_i^T \mathbf{r}_i \right] 
\]

where \( \sigma_{ij} \) is an event-dependent value that includes modeling and observation errors. This study assumed them to be \( \sigma_{ij} = \max(0.1d_{ij}, \text{Observation error}_{ij}) \) for a large earthquake. By assuming that \( \sigma_{ij} \) depends on the displacement with the lower bound of the observation error, we can prevent the model from overfitting the data. We used the Metropolis-Hasting method as the MCMC sampler (hereafter, M-H method; Metropolis et al. 1953; Hastings 1970). We adopted parallel tempering with eight parallel chains to improve the search efficiency (Geyer 1991, Jasra et al. 2007). For simplicity of the real-time analysis, we fixed rake angles to 90°; as such, \( \mathbf{\theta} \) contains only slip amounts, the element number of which is the same as that of the subfaults. In addition, we assumed the slip amounts to be non-negative, which is equivalent to assuming that the prior probability density function (PDF) formed a uniform distribution \( U(0, \infty) \) for \( \mathbf{\theta} \). There
were 2951 subfaults on the plate boundary. We adopted Okada (1992) as the Green’s function for rectangular subfaults, where the fault width are approximately 8 km. In the M-H method, we propose new transition candidates by applying random perturbations $\Delta \theta$. However, applying MCMC to such a problem with many unknown parameters takes a long time to converge. Therefore, in addition to parallel tempering, we improved the search efficiency using a multistage approach with stepwise partitioning of perturbation groups. Figure 2 shows an overview of the "stepwise partitioning algorithm". The Markov chain was divided into four stages: stage 1, model 80 ($3 \times 10^6$ steps); stage 2, model 185 ($3 \times 10^6$ steps); stage 3 model 388 ($3 \times 10^6$ steps); and stage 4 model 1451 ($3 \times 10^6$ steps). In each stage, we set perturbation groups in advance because grouping $\Delta \theta$ around subfaults promoted convergence. Thus, we applied the same perturbation $\Delta \theta$ to the same-colored subfaults shown in Figure 2; we refer to these pre-determined areas with the same $\Delta \theta$ as "perturbation groups". Increasing the number of perturbation groups as the stage progresses allows us to estimate the rough-to-detailed features. The utilization of such a grouping is equivalent to reducing the number of unknown parameters. This approach is similar to the trans-dimensional approach (e.g., Dettmer et al. 2014). Our algorithm was designed to force change in one direction to simplify the problem for real-time purposes. It should be noted that we did not change the shapes of the 2951 rectangular subfaults; that is, only the perturbation groups were changed. Thus, the analytical displacements in each step were calculated for 2951 background subfaults using a constant each subfault’s Green's function. As shown in Figure 2, in the transition between stages, we used the median of the posterior PDFs of the previous stage as the initial value for the next stage and used the
95% confidence interval (CI) as the amount of perturbation $\Delta \theta_{\text{max}}$ for the next stage.

By inheriting the uncertainty as the transition amount, we searched for a broader slip amount with more uncertainty in a limited number of samples. Therefore, in the initial 10% of samples for each stage, we adjusted $\Delta \theta_{\text{max}}$ by equal multiplication so that the acceptance rate would be 20%–40% (e.g., Roberts and Rosenthal 2001), and the samples under adjustment were discarded during PDF generation as burn-in.

In the later stage, the partitioning may be excessive for the sensitivity of the data and earthquake magnitude. We calculated the AIC (Akaike's Information Criterion, Akaike 1973) using the following equation to determine the optimum number of perturbation groups:

$$AIC = -2 \log(L) + 2M.$$  \hspace{1cm} (2)

where $L$ indicates the maximum likelihood, and $M$ indicates the number of unknown parameters, which is the number of perturbation groups in each stage.

Thus, the computation time of this approach is within 30 min for each stage ($3 \times 10^6$ steps) using an SX-Aurora TSUBASA Type20B processor with eight parallel chains in the case of 642 GNSS stations. In this study, we decided the number of samples with the highest priority to obtain a sufficient number to evaluate the differences between the stages. For real-time utilization, the convergence judge (e.g., Gelman 1996) should be introduced and must move to the next stage automatically. According to the Markov chain (see Section 3.1), a chain length of about 10% may be sufficient for convergence, and based on the AIC, it is possible to complete the calculation up to stage 2 in about 6 minutes, which is a computation time that may be used in real time in the future.

2.2 Real-time tsunami inundation risk map
In general, because the tsunami inundation calculation is a nonlinear problem, it is difficult to evaluate the uncertainty of the slip distribution based on a single model. Therefore, to evaluate the uncertainty of tsunami inundation, we need to prepare multiple coseismic fault slip models to identify that which best explains the data. However, it is not realistic to calculate tsunami inundation for all MCMC samples, even if calculation speed were to improve in the future. Therefore, our algorithm aims to efficiently classify MCMC samples and integrate multiple tsunami inundation scenarios to probabilistically present tsunami inundation risk. Here, assuming that multiple calculations of $10^2$ orders will be possible within a time of several minutes in the future, we added "real-time" to the flow name in this study.

Figure 3 shows a flow chart of the real-time tsunami inundation risk map process.

i. Obtain sufficient MCMC samples to evaluate the uncertainty of the slip distribution model (Section 2.1). Samples from which Variance Reduction (VR) was extracted are equivalent to a representative value based on the posterior PDF.

ii. Classify these samples into small clusters (K) using the k-means approach using $\theta$ as the feature value.

iii. Generate K representative slip distributions (using the median value of each subfault), and use them as inputs to calculate individual K tsunami inundations.

iv. Count the number of inundations on the map.

The resulting map shows the possible tsunami arrival rate in each computation grid based on the uncertainty of fault slip estimated from real-time GNSS data.

The main feature of our algorithm is the clustering of MCMC samples, generally utilize posterior PDFs (see Section 3.1). As discussed, the nonlinearity of tsunami inundation
calculation is a problem; thus, we need to obtain sufficient fault models to evaluate the tsunami risk. In contrast, samples using the M-H method are not necessarily independent, because the M-H method proposes samples by perturbing the previous one. Thus, we could obtain a set of different fault slip distributions to explain the observed data well. We adopted the k-means method to classify the samples into a predetermined number, with the number of clusters K determined by considering the computational cost of tsunami inundation, the availability of denominators for probability display, and the maintenance of the diversity of MCMC samples. The feature value used for clustering is the slip vector \( \mathbf{\theta} \) because the clustering phase does not depend on the specific target region of the tsunami calculation. Samples with the same VR were used to ensure fairness in terms of reproducibility of the observed data. Here, the target of the tsunami inundation calculation was a 1,563 km\(^2\) area between Tosa City and Aki City in Kochi Prefecture (red box in Figure S1). Table S1 lists the conditions of the tsunami inundation calculation. The grid size in the target area was 30 \( \times \) 30 m, and the number of grids was 1,736,388 (2,082 grids along the east–west direction and 834 grids along the north–south direction). Tsunami propagation and inundation were calculated using the TUNAMI code (Tohoku University's numerical analysis model for investigating tsunami), which numerically solves the nonlinear shallow water equation using the staggered leap-flog finite difference method (Imamura 1995). The tsunami simulation was carried out from the time of the earthquake to 6 h later. The fracture propagation of the earthquake was not considered, and the slip was assumed to be instantaneous. We calculated seafloor displacement using the analytical result of Okada (1992), and the uplift of the seawater due to the horizontal movement of the seafloor was taken into account (Tanioka and Satake 1996). Coastal structures, such
as breakwaters, were input as line data every 30 m and were not breached by solid
earthquakes or tsunamis; buildings were not considered. We used the maximum
inundation depth up to 6 h after each calculation grid as the inundated depth.

3. Results

3.1 Uncertainty of slip distribution

We applied the stepwise partitioning algorithm (Section 2.1) to the simulated data—that
is, displacement data at 642 GEONET stations calculated using the analytical result of
Okada (1992)—and added Gaussian noise (standard deviation of 2 cm horizontally and
5 cm vertically).

Figure 4 shows the estimation results (median of the posterior PDFs with the 95% CI),
which were calculated from the difference between samples located at 2.5% and those
at 97.5% for four stages. For all stages, the estimated fault model based on the median
value well reproduced the observed data. However, the 95% CI was large, especially
near the trough axis, and increased from near the land toward the offshore. This
suggests that slip was well estimated in regions with large slip, while that offshore had
low estimated resolution.

As the number of perturbation groups increased with the progress of the stage, the
obtained slip distribution consisted of slips with shorter wavelengths (Figure 4). In
particular, in model 1451, a large amount of local slip occurred on certain small faults.

Figure 5 shows the Markov chains and PDFs of Mw and VR for all samples except the
burn-in. The results show that the four-stage sampling successfully estimated models
with high VR values without smoothing constraints. VR increased with an increase in
the freedom of the model. However, the amount of increase was small from model 388
to model 1451. These results suggest that the spatial scale of the slip depended on that of the perturbation groups, because they were estimated without the smoothing constraint. The bottom chains in Figure 5(a) show the calculated AIC values for each sample using Eq. (2). The mean values (± standard deviation) of AIC in each stage were 5502 ± 35, 5384 ± 47, 5455 ± 66, and 7572 ± 78, respectively. The AIC values of models 185 and 388 were lower than those of model 1451, which objectively suggests the possibility of over-division. Based on this, we mainly use the samples from model 185 for further analysis.

Comparing the input (Figure 1), the median slip is located more landward at all stages (Figure 4). The Mw and VR values calculated based on the median models (inset in Figure 4) were slightly different from those calculated based on each sample (Figure 5). This indicates that the model obtained for each sample does not completely agree with the slip distribution by representative values because of the diversity of the slip in the near-trench area. The diversity of offshore slip is more pronounced owing to the lack of real-time seafloor observation points. We discuss the tradeoff of slip in the fault dip direction in Section 4.1.

Figure 6 shows eight examples of posterior PDFs from model 185 for the area off Cape Ashizuri, where the estimated slip was large and arranged in the fault dip direction. The 95% CI at point A (close to land) was small, while the value at point B (farther offshore) was large. At point F, located further offshore, the search extends to approximately 10 m, while the maximum frequency value remains at zero. These tendencies were also observed for points C and G. In the case of point E, which corresponds to a region where the assumed slip gradually decays, the PDF includes the assumed slip (red line in the figure). These results show that the MCMC can estimate
significant slip distributions even without smoothing constraints, and that the variability
of the uncertainty based on data can be quantified as posterior PDFs.

3.2 Uncertainty of tsunami inundation

We applied the "real-time tsunami inundation risk map (Section 2.2)" to the MCMC
samples of model 185, where had the lowest AIC (see Section 3.1). First, among the
$3 \times 10^6$ MCMC samples of model 185, we extracted 319,184 samples with VR of
99.52%, which is the representative value of the PDF (Figure 3i). We then classified the
extracted samples into 100 clusters using the k-means method. In each cluster, we
generated one scenario using the median values for each subfault belonging to the
cluster (Figure 3ii). Using these scenarios as inputs, we calculated 100 tsunami
inundation calculations for the area of Kochi Prefecture (Figure 3iii). Finally, we
counted the number of inundated points on each grid on the map (Figure 3iv). The
threshold of the inundated point was 1 cm or more during the 6 h since the earthquake.
We did not apply any weights to obtain the frequencies.

Figure 7(a) shows the generated real-time tsunami inundation risk map. The envelope of
the maximum inundation area and the arrival probability (risk level) are shown. The
area where the risk of tsunami inundation is extremely high in all 100 cases (pink
coloring) is similar to the inundation area due to the assumed slip (Figure 8 bottom).
The areas with the highest probability of tsunami inundation are the Uranouchi Bay,
Monobe River, and Yasu River. In addition, the southern half of Kochi Airport also has
a high risk of inundation. These results show that tsunami inundation is strongly related
to land topography and rivers. Tsunamis would likely run northward along large rivers
(e.g., Niyodo River, Monobe River), and extend their inundation area around rivers and
waterways (e.g., Shimoda River). The limit of tsunami run-up on land was defined in some cases by the fact that the kinetic energy was zero, and in other cases, because it could not overcome the difference in elevation of the terrain (rivers, channels, and embankments). For comparison, Figure 7(b) shows an elevation map of the same area. High inundation risk is generally distributed in areas with elevations of 10 m or less. On the other hand, tsunami inundation does not spread to Urado Bay, despite the low elevation, owing to the structural conditions in the tsunami calculation. The elevation of area A in Figure 7(b) varies gently (from 0 to 8 m), and Figure 7(a) shows a clear gradation of 0%–100%. Area B has an embankment of more than 1 m and this becomes the inundation boundary. Area C has an elevation of more than 10 m parallel to the coast; in about 50% of the scenarios inundation proceeded from the east side, and in a few scenarios, inundation overcame from the south to the north. Therefore, the results of inundation calculations vary depending on small differences in topographical data and arrival tsunami scenarios.

Figure 7(a) shows that there were few areas where the probability of inundation was 50%–90%. This indicates that the tendency of inundation is different between always inundated and rarely inundated areas. Figure 8 shows the cases with large inundation areas. The inundation area of the largest two cases was more than 40 km², which is clearly larger than the assumed slip (16.784 km²). In both cases, the initial wave fields had uplift/subsidence gaps in the offshore area of Kochi Prefecture. If these two cases were removed from the count, the appearance of the figure changed (Figure S2). In addition, samples with large vertical seafloor deformation for the entire slip area did not necessarily have large tsunami inundation areas (Figure S3) because tsunami inundation was evaluated on a local scale.
It is also important to consider the method of editing the real-time tsunami inundation risk maps. Figure S4 shows editing examples where (a) the map is weighted by the label frequency distribution in Figure 9(b) and (b) where the threshold value of the inundation depth considered to be inundation was changed from 1 cm to 1 m. In Figure S4(a), there was almost no change in the frequency tendency of the inundation area because there was no strong correlation between the frequency and the inundation area. However, as shown in Figure S4(b), there was a decrease in the inundation probabilities because inundation of less than 1 m was not counted.

These real-time tsunami inundation risk maps provide quantitative probabilistic information. As the concept of the target area is not included in the scenario classification method (Figure 3i, ii), the same map can be generated for any region using the same MCMC samples.

4. Discussions

4.1 Setting optimization in stepwise partition algorithm

4.1.1. Optimization of $\sigma_{ij}$ in the likelihood function

In this study, we assumed $\sigma_{ij} = \max(0.1d_{ij}, \text{Observation error}_{ij})$ [see Eq. (1)]. In contrast, Ohno et al. (2021) automatically adjusted $\sigma_{ij}$ according to the event using the maximum likelihood estimation scheme. Here, this adjustment was not applied to the slip distribution model because the large number of parameters easily explained the observed data, resulting in a very small estimate of $\sigma_{ij}$. This dynamic adjustment is more effective for a single rectangular fault model; however, modeling errors exist in the slip distribution model owing to the uncertainty of Green's function. Therefore, in
this study, we assumed $\sigma_{ij}$, which depends on the displacements and observation errors.

4.1.2. Optimization of a priori information

Ohno et al. (2021) used earthquake early warning values (EEW: latitude, longitude, depth, M) as a priori information. While our algorithm does not use such a priori information, it is important to appropriately utilize such information to achieve accurate and rapid estimation for real-time estimation. For example, the algorithm can use the EEW and independent estimated models, such as the single rectangular fault model and the slip distribution by REGARD, as a priori information.

4.1.3. Optimization of perturbation groups and fault settings for each stage

As discussed in Section 3.1, there is a tradeoff of slip in the dip direction. We applied a stepwise partitioning algorithm to various assumed slip distributions and found a similar tendency (Figure S5–10). The position and wavelength of the slip in the dip direction affected the tsunami calculation. Optimization of the perturbation group settings is especially important because the empirical geophysical law (slip continuity) is not considered when a smoothing constraint is not used. For example, it is desirable to evaluate the spatial resolution depending on the location of the observation stations and to determine the division of the perturbation group accordingly. Kimura et al. (2019) developed an algorithm for automatically setting up the division of subfaults according to the arrangement of observation stations. It may be effective to use this algorithm for setting subfaults (smallest division) and arranging the perturbation group in each stage. The required size varies depending on the magnitude of the earthquake, and so we should determine the optimum stage objectively (i.e., using the AIC) and judge the
4.2 Optimum number of clusters

The optimum number of clusters \( K \) for k-means (Section 3.2) should be determined considering the cost of real-time computation, the availability of denominators for probability display, and the availability of a sufficient number of clusters that do not distort the diversity of all samples. Here, we consider the validity of \( K = 100 \) in comparison with \( K = 300 \).

Figure 9 shows the results of clustering using k-means. According to the elbow method graph (Figure 9a), a cluster number of \( K = 50 \)–100 seems to be appropriate. On the other hand, according to the frequency distribution of the clusters (Figure 9b), even if we increased the number of clusters from \( K = 100 \) to \( K = 300 \), there was no particular change in the tendency. Figure 9(c) shows the correlation between the frequency of each cluster and the variability of the slip distribution, which is a scalar sum of the ranges of slip amounts on each subfault. According to Figure 9(c), both the frequency (horizontal axis) and variability (vertical axis) are smaller for \( K = 300 \) than for \( K = 100 \). This may indicate that by increasing the number of clusters, clusters were formed among the slip distribution models that have the closest features to each other. Figure 9(d) shows the variability of the VRs for the \( K \) scenarios (the median model in each cluster). Although the used samples have the same VR (99.52%), those of the classified scenarios vary in the order of the second decimal place. Note that the direction of variation was the direction of the improvement in VR, probably because the short-wavelength slip of each sample was smoothed by taking the median of the cluster, resulting in a model that better explains the data.
Figure 10 shows the correlation between the frequency of each cluster and the inundation areas for $K = 100$ and $K = 300$. The frequency distribution of the tsunami inundation area (vertical axis) is asymmetric, with a peak near the tsunami inundation area due to the assumed slip ($16.784 \text{ km}^2$) and long tail on the side of the large tsunami inundation. The maximum and minimum inundation areas for $K = 100$ and $K = 300$ are similar. On the other hand, 300 cases of $K = 300$ are distributed so that they filled in the gaps between the 100 cases of $K = 100$. The maximum inundation area for $K = 300$ is slightly larger than that for $K = 100$ (and the maps look slightly different; Figure S11). These characteristics suggest that the larger the number of $K$, the more the tsunami scenario is subdivided, and the degree of smoothing of extreme slip decreases. However, for the present trial, $K = 100$ was sufficient to obtain the range of the inundation area.

In the case of $K = 100$ (Figure 10), there is a negative correlation between the inundation area and the frequency of clustering. We colored the clusters with an inundation area of $40 \text{ km}^2$ or more, as shown in Figure 9(b) and found no relationship between the degree and flooded area in Kochi City. This may be because the frequency shows a tendency for the entire slip distribution, whereas inundation was evaluated by focusing on a target area, which is a natural result. On the other hand, samples with extreme slip patterns are classified into clusters with small frequencies, and the tendency of these samples to cause maximum inundation in some other areas cannot be excluded.

4.3 Utilization of real-time tsunami inundation risk maps
The risk of tsunami inundation is expressed by multiplying hazard (e.g., tsunami inundation) by exposure (e.g., population and buildings) and fragility (e.g., buildings), then dividing it by resiliency (e.g., Wood 2011). The real-time tsunami inundation risk maps evaluate the probability of tsunami inundation for each area, which corresponds to the probabilistic representation of hazards in the above equation. For example, if we reflect the difference in day/night population (i.e., exposure), we can evaluate the difference in disaster risk depending on the time of day, and if we reflect the difference in resiliency, we can quantify the disaster risk according to the time scale we want to evaluate. In addition, to maximize the benefit of evaluating disaster triggers as probabilities in this study, it is easier to evaluate the risk level if other factors can be treated as probabilities in the same dimension. For easy-to-understand expression, qualitative evaluation may also be useful (e.g., three risk levels of large, medium, and small).

As mentioned in Section 1, the real-time tsunami inundation damage estimation system is expected to be utilized as a function of the disaster information system (DIS) of the Cabinet Office for the initial response of the government when a disaster occurs. Our developed framework for the quantitative evaluation of uncertainty in tsunami inundation prediction has the potential to improve this system. The risk level can be expressed as a probability instead of a single inundation prediction. This system has the potential to improve understanding of high disaster-risk regions and to support the decision-making process for the initial response in such areas.

5. Conclusions

In this study, we developed a new coseismic fault model estimation algorithm using
Bayesian statistics to quantitatively evaluate the estimation uncertainty in the real-time estimation of slip distribution using real-time GNSS data. In addition, as an application of the quantified uncertainty, we investigated how the uncertainty of the fault model estimation may affect tsunami inundation prediction and proposed a new method to extract the uncertainty of tsunami inundation prediction.

One of the major problems in the estimation of slip distribution models with MCMC is that they take a long time to converge. To overcome this problem for real-time purposes, we developed a stepwise partitioning algorithm. We divided the entire Markov chain into four stages with different perturbing groups, and used the 95% CI in the previous stage as the amount of perturbation in the next stage. This enabled us to extract multiple models with different spatial patterns that explained the observed data well. We applied the algorithm to the numerical simulation data of the Nankai Trough region (assumed to be the Mw 8.75 1707 Hoëi earthquake). The 95% C.I. of the obtained posterior PDFs increased with distance from the coast to the offshore, indicating the uncertainty in estimating plate boundary slip from onshore GNSS.

We developed a real-time tsunami inundation risk map process to fully utilize the advantages of the stepwise partitioning algorithm, wherein the uncertainty of slip distribution is quantified as a large number of samples. Specifically, we classified the slip distribution models into 100 clusters determined by the k-means method after extracting samples that have the same VR. We calculated the individual tsunami inundation using 100 scenarios. Furthermore, we counted the number of inundations in each grid on the map and displayed them probabilistically. We showed the envelope of the maximum inundation area and the probability of reaching each computational grid in Kochi City, Kochi Prefecture. However, in order to truly utilize the obtained tsunami
inundation risk, it may be necessary to examine the kind of information needed not only from the viewpoint of disaster triggers but also from the viewpoint of society as a recipient.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

List of abbreviations

ABIC: Akaike's Bayesian Information Criterion
AIC: Akaike's Information Criterion
CI: Confidence intervals
DIS: Disaster information system
EEW: earthquake early warning
GEONET: continuous GNSS network
GNSS: Global Navigation Satellite System
GSI: Geospatial Information Authority of Japan
JMA: Japan Meteorological Agency
MCMC: Markov chain Monte Carlo methods
PDF: Probability density function
 REGARD: REal-time GEONET Analysis system for Rapid Deformation monitoring
TUNAMI: Tohoku University's numerical analysis model for investigating tsunami
VR: Variance Reduction

**Availability of data and materials**

The analyzed data for the current study are available from the corresponding author upon reasonable request.

**Competing interests**

The authors declare that they have no competing interests.

**Funding**

This study was supported by the Japan Society for the Promotion of Science Grant-in-Aid for Scientific Research (KAKENHI; grant no. 17H06108) and by the Toray Science Foundation (Toray Science and Technology Grant). This work was also supported by the JST FOREST Program (grant number: JPMJFR202P, Japan). In addition, this research was supported in part by the Ministry of Education, Culture, Sports, Science, and Technology's (MEXT) "Earthquake Observation Research Program to contribute to disaster mitigation". This work was also supported by the Next Generation High-Performance Computing Infrastructures and Applications R&D Program by MEXT. This work was also supported by the Research Project for Disaster Prevention on the great Earthquakes along the Nankai trough by MEXT.
Authors’ contributions

K.O. and Y.O. designed the study, developed the programs, and analyzed the data. S.K. and T.A. performed the tsunami inundation calculations. A.M. and H.K. contributed to speeding up the programs developed. Y.O., S.K., R.H., and H.K. assisted in the interpretation of the results. All authors read and approved the final manuscript.

Acknowledgements

Not applicable

Authors’ information

K.O. was a graduate student at the Graduate School of Science, Tohoku University (now at the Geospatial Information Authority of Japan). Y.O. and R.H. are associate professor and professor, respectively, at the Graduate School of Science, Tohoku University. S.K. and T.A. are professor and research fellow, respectively, at the International Research Institute for Disaster Science, Tohoku University. A.M. and H.K. are visiting professors and professors, respectively, at the Graduate School of Information Sciences, Tohoku University.

References

Akaike H (1973) Information theory and the maximum likelihood principle. In: Petrov BN, and Csáki F (eds) 2nd International Symposium on Information Theory. Akademiai Kiadó, Budapest.
Akaike H (1980) Likelihood and the Bayes procedure. In: Bernardo JM, DeGroot MH, Lindley DV, Smith AFM (eds) Bayesian Statistics. University Press, Valencia, pp 143–166.

Blewitt G, Kreemer C, Hammond WC, Plag H-P, Stein S, Okal E (2006) Rapid determination of earthquake magnitude using GPS for tsunami warning systems. Geophys Res Lett 33:L11309. doi:10.1029/2006GL026145.

Blewitt G, Hammond WC, Kreemer C, Plag H-P, Stein S, Okal E (2009) GPS for real-time earthquake source determination and tsunami warning systems. J Geod 83:335–343. doi:10.1007/s00190-008-0262-5.

Dettmer J, Benavente R, Cummins PR, Sambridge M (2014) Trans-dimensional finite-fault inversion. Geophys J Int 199:735–751.

Dettmer J, Hawkins R, Cummins PR, Hossen J, Sambridge M, Hino R, Inazu D (2016) Tsunami source uncertainty estimation: The 2011 Japan tsunami. J Geophys Res Solid Earth, 121(6):4483–4505. doi:10.1002/2015JB012764.

Fukahata Y (2009) Development of Study of Inversion Analyses Using ABIC in Seismology. J Seismol Soc Japan 61:S103–S113. doi:10.4294/zisin.61.103 (in Japanese).

Fukutani Y, Suppasri A, Imamura F (2015) Stochastic analysis and uncertainty assessment of tsunami wave height using a random source parameter model that targets a Tohoku-type earthquake fault, Stoch. Environ Res Risk Assess 29(7):1763–1779. doi:10.1007/s00477-014-0966-4.

Furumura T, Imai K, Maeda T (2011) A revised tsunami source model for the 1707 Hoei earthquake and simulation of tsunami inundation of Ryujin Lake, Kyushu, Japan. J Geophys Res Solid Earth 116(2):1–17. doi:10.1029/2010JB007918.
Gelman A (1996) Interface and monitoring convergence. In Gilks WR, Richardson S, Spiegelhalter DJ (eds) Markov chain Monte Carlo in practice. Chapman and Hall/CRC, London, pp 131–143.

Geyer CJ (1991) Markov Chain Monte Carlo Maximum Likelihood. In Computing Science and Statistics. Proceedings of the 23rd Symposium on the Interface, 156–163.

Gibbons SJ, Lorito S, Macias J, Løvholt F, Selva J, Volpe M, Sánchez-Linares C, Babeyko A, Brizuela B, Cirella A, Castro MJ, de la Asunción M, Lanucara P, Glimsdal S, Lorenzino MC, Nazaria M, Pizzimenti L, Romano F, Scala A, Tonini R, González Vida JM, Vöge M (2020) Probabilistic Tsunami Hazard Analysis: High Performance Computing for Massive Scale Inundation Simulations. Front Earth Sci 8(December):1–20. doi:10.3389/feart.2020.591549.

Giles D, Gopinathan D, Guillas S, Dias F (2021) Faster Than Real Time Tsunami Warning with Associated Hazard Uncertainties. Front Earth Sci 8(January):1–16. doi:10.3389/feart.2020.597865.

Goda K, and Song J (2016) Uncertainty modeling and visualization for tsunami hazard and risk mapping: a case study for the 2011 Tohoku earthquake. Stoch Environ Res Risk Assess 30(8):2271–2285. doi:10.1007/s00477-015-1146-x.

Goda K, Yasuda T, Mori N, Muhammad A, de Risi R, de Luca F (2020) Uncertainty quantification of tsunami inundation in Kuroshio, Kochi Prefecture, Japan, using the Nankai-Tonankai megathrust rupture scenarios. Nat Hazards Earth Syst Sci 20(11):3039–3056. doi:10.5194/nhess-20-3039-2020.

Hastings WK (1970) Monte Carlo Sampling Methods Using Markov Chains and Their
Inuma T, Hino R, Kido M, Inazu D, Osada Y, Ito Y, Ohzono M, Tsushima H, Suzuki S, Fujimoto H, Miura S (2012) Coseismic slip distribution of the 2011 off the Pacific coast of Tohoku earthquake (M9.0) refined by means of seafloor geodetic data. J Geophys Res 117:B07409. doi:10.1029/2012JB009186.

Inoue T, Ohta Y, Koshimura S, Hino R, Kawamoto S, Hiyama Y, Douke T (2016) A study on methods for applying fault model rapidly estimated using real-time GNSS to tsunami simulation. Japan Soc Civil Eng 72(2):355–360. doi:10.2208/kaigan.72.I_355 (in Japanese).

Ito T, Gunawan E, Kimata F, Tabei T, Simons M, Meilano I, Agustan A, Ohta Y, Nurdin I, Sugiyanto D (2012) Isolating along-strike variations in the depth extent of shallow creep and fault locking on the northern Great Sumatran Fault. J Geophys Res 117:B06409. doi:10.1029/2011JB008940.

Jiang J, and Simons M (2016) Probabilistic imaging of tsunamigenic seafloor deformation during the 2011 Tohoku-Oki Earthquake. J Geophys Res Sol Earth 12:9050–9076. doi:10.1002/2016JB013760.

JMA (2012): Work for improving tsunami warning, https://www.jma.go.jp/jma/kishou/books/hakusho/2012/HN2012sp2.pdf (in Japanese).

Kawamoto S, Hiyama Y, Ohta Y, Nishimura T (2016) First result from the GEONET real-time analysis system ( REGARD): the case of the 2016 Kumamoto earthquakes. Earth Plane Space 68(190):1–12. doi:10.1186/s40623-016-0564-4.
Kawamoto S, Ohta Y, Hiyama Y, Todoriki M, Nishimura T, Furuya T, Sato Y, Yahagi T, Miyagawa K (2017) REGARD: A new GNSS-based real-time finite fault modeling system for GEONET. J Geophys Res Solid Earth 122:1324–1349. doi:10.1002/2016JB013485.

Kimura H, Tadokoro K, Ito T (2019) Interplate coupling distribution along the Nankai Trough in southwest Japan estimated from the block motion model based on onshore GNSS and seafloor GNSS/A observations. J Geophys Res Solid Earth 124:6140–6164. doi:10.1029/2018JB016159.

Larson KM, Bodin P, Gomberg J (2003) Using 1-Hz GPS data to measure deformations caused by the Denali fault earthquake. Science 300:1421–1424. doi:10.1126/science.1084531.

Larson KM (2009) GPS seismology. J Geod 83(3–4):227–233. doi:10.1007/s00190-008-0233-x.

Melgar D, Bock Y, Crowell BW (2012) Real-time centroid moment tensor determination for large earthquakes from local and regional displacement records. Geophys J Int 188(2):703–718.

Melgar D, and Bock Y (2015) Kinematic earthquake source inversion and tsunami runup prediction with regional geophysical data. J Geophys Res Solid Earth 120:3324–3349. doi:10.1002/2014JB011832.

Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E (1953) Equations of State Calculations by Fast Computing Machines. J Chem Phys 21(6):1087–1092.

Minson SE, Simons M, Beck JL, Ortega F, Jiang J, Owen SE, Moore AW, Inbal A, Sladen A (2014a) Bayesian inversion for finite fault earthquake source
models—II: The 2011 great Tohoku-Oki, Japan earthquake. Geophys J Int 198(2):922–940. doi:10.1093/gji/ggu170.

Minson SE, Murray JR, Langbein JO, Gomberg JS (2014b) Real-time inversions for finite fault slip models and rupture geometry based on high-rate GPS data. J Geophys Res Solid Earth 119:3201–3231. doi:10.1002/2013JB010622.

Mulia IE, Ishibe T, Satake K, Gusman AR, Murotani S (2020) Regional probabilistic tsunami hazard assessment associated with active faults along the eastern margin of the Sea of Japan. Earth Planet Sp 72(1). doi:10.1186/s40623-020-01256-5.

Musa A, Watanabe O, Matsuoka H, Hokari H, Inoue T, Murashima Y, Ohta Y, Hino R, Koshimura S, Kobayashi H (2018) Real-time tsunami inundation forecast system for tsunami disaster prevention and mitigation. J Supercomput 74(7):3093–3113. doi:10.1007/s11227-018-2363-0.

Ohno K, Ohta Y (2018) Trial of quantitative uncertainty estimation for the single rectangular coseismic fault model deduced from real-time GNSS data using the MCMC approach: a case study of the 2011 Tohoku-Oki earthquake. J Geod Soc Japan 64:39–50 (in Japanese).

Ohno K, Ohta Y, Kawamoto S, Abe S, Hino R, Koshimura S, Musa A, Kobayashi H (2021) Real-time automatic uncertainty estimation of coseismic single rectangular fault model using GNSS data. Earth Planet Space 73:127.

Ohta Y, Meilano I, Sagiya T, Kimata F, Hirahara K (2006) Large surface wave of the 2004 Sumatra-Andaman earthquake captured by the very long baseline kinematic analysis of 1-Hz GPS data. Earth Planet Space 58(2):153–157. doi:10.1186/BF03353372.
Ohta Y, Kobayashi T, Tsushima H, Miura S, Hino R, Takasu T, Fujimoto H, Iinuma T, Tachibana K, Demachi T, Sato T, Ohzono M, Umino N (2012) Quasi real-time fault model estimation for near-field tsunami forecasting based on RTK-GPS analysis: Application to the 2011 Tohoku-Oki earthquake (Mw 9.0). J Geophys Res 117:B02311. doi:10.1029/2011JB008750.

Ohta Y (2016) Study on the improvement of real-time kinematic GNSS data analysis and development of rapid coseismic fault determination method based on those data. J Geod Soc Japan 62(1):119. doi:10.11366/sokuchi.62.1 (in Japanese).

Ohta Y, Inoue T, Koshimura S, Kawamoto S, Hino R (2018) Role of real-time GNSS in near-field tsunami forecasting. J Disaster Res 13(3):453–459. doi:10.20965/jdr.2018.p0453.

Okada Y (1992) Internal deformation due to shear and tensile faults in a half-space. Bull Seismol Soc Am 82(2):1018–1040.

Roberts GO, Rosenthal JS (2001) Optimal scaling for various Metropolis-Hastings algorithms. Statistical Science 16(4):351–367.

Sraj I, Mandli KT, Knio OM, Dawson CN, Hoteit I (2014) Uncertainty quantification and inference of Manning's friction coefficients using DART buoy data during the Tōhoku tsunami. Ocean Model 83:82–97. doi:10.1016/j.ocemod.2014.09.001.

Tanioka Y, Satake K, Ruff L (1995) Total analysis of the 1993 Hokkaido Nansei-oki earthquake using seismic wave, tsunami, and geodetic data. Geophys Res Lett 22(1):9–12. doi:10.1029/94GL02787.

Tsushima H, and Ohta Y (2014) Review on near-field tsunami forecasting from offshore tsunami data and onshore GNSS data for tsunami early warning. J Dis Res
Tsushima, H., R. Hino, Y. Ohta, T. Iinuma, and S. Miura (2014), tFISH/RAPiD: Rapid improvement of near-field tsunami forecasting based on offshore tsunami data by incorporating onshore GNSS data, *Geophys. Res. Lett.*, 41, 3390–3397, doi:10.1002/2014GL059863.

Todoriki M, Furumura T, Maeda T (2013) 3D FDM simulation of seismic wave propagation for Nankai Trough earthquake: Effects of topography and seawater. AGU Fall Meeting, San Francisco, USA, 9–13 December, 2013.

Wood N (2011) Understanding Risk and Resilience to Natural Hazards. U.S. Geological Survey Fact Sheet 2011-3008, p 2.

**Figure legends**

**Figure 1.** Inputted slip distribution for the 1707 Hoei earthquake in southwestern Japan (subdivision model of Inoue et al. 2016). The Mw is 8.75 and the color scale indicates the slip amount along each subfault. White dots indicate the 642 continuous Global Navigation Satellite System (GNSS) network (GEONET) stations, with displacements calculated using Okada (1992). The width of rectangular subfaults is approximately 8 km.

**Figure 2.** “Stepwise partitioning algorithm” overview, depicting transition from stage \( n \) to stage \( n + 1 \) (\( n = 1,2,3 \)). Colors indicate the perturbation groups of \( \Delta \theta \). Perturbation groups that differ from stage to stage enable search from the outline to the details without smoothing constraints. The number of perturbations \( \Delta \theta_{\text{max}} \) that differ for each perturbation group enables a flexible parameter search with large estimation uncertainty.
Figure 3. Overview of the “real-time tsunami inundation risk map” process using model 185. (i) Obtain enough Markov chain Monte Carlo (MCMC) samples to evaluate the uncertainty of the slip distribution model (Section 2.1). Then, extract the samples in which variance reduction (VR) is equivalent to the representative value based on the posterior probability density function (PDF). (ii) Classify them into small clusters (K) by the k-means approach using $\boldsymbol{\theta}$ as the feature value. (iii) Generate K representative slip distributions (using the median value in each subfault), and using them as inputs, calculate K individual tsunami inundations. (iv) Count the number of inundations on the map.

Figure 4. Estimation results for four stages. Left column: median model of posterior probability density functions (PDFs) of each subfault. Inserted values indicate the Mw and variance reduction (VR) of the median model. Black and white vectors indicate inputted and calculated horizontal displacements. Right column: 95% confidence intervals (CI) of posterior PDFs of each subfault, as calculated from the difference between the samples located at 2.5% and 97.5%.

Figure 5. Markov chains and posterior probability density functions (PDFs). We divided the Markov chain into four stages: stage 1, model 80 ($3 \times 10^6$ steps); stage 2, model 185 ($3 \times 10^6$ steps); stage 3 model 388 ($3 \times 10^6$ steps); and stage 4 model 1451 ($3 \times 10^6$ steps). a: Markov chains of Mw, variance reduction (VR), and Akaike's Information Criterion (AIC). In dotted lines, $\Delta \boldsymbol{\theta}_{\text{max}}$ is adjusted for $5 \times 10^4$ steps as burn-in, which samples are not shown in this figure. b: Posterior PDFs of Mw and VR. The range on the horizontal axis is same with the range on the vertical axis of (a). Inserted values indicate mean, median, and mode (from top to bottom).

Figure 6. Examples of posterior probability density functions (PDFs) in model 185. A–
G: slip amounts of eight perturbation groups. The range on the horizontal axis is normalized to 0–20 m. Solid and dotted black lines show the median and the 95% confidence intervals (CI). The solid red lines show the assumed slip amount in the background subfault of the perturbation group. Inserted values indicate mean, median, and mode (from the top to bottom). A, B, and F shows the area off Cape Ashizuri where the estimated slip was large, arranged in the fault dip direction. These tendencies can also be confirmed for C and G. The case of E corresponds to a region where the assumed slip gradually decayed.

**Figure 7.** Real-time tsunami inundation risk maps for Kochi city. a: Real-time tsunami inundation risk map for Kochi city (K = 100) overlaid on a shadow topography map (Geospatial Information Authority of Japan). Colors indicate normalized frequency of inundation from the 100 tsunami inundation calculations. b: Elevation map (Geospatial Information Authority of Japan). Dotted rectangles indicate areas mentioned in the text.

**Figure 8.** Slip distribution, seafloor vertical deformation (initial wave field), and tsunami inundation scenarios (K = 100). a: The case of flooding across the largest area. b: The case of flooding across the second largest area. c: The case of assumed model (Figure 1). Left: Slip distribution (tsunami scenario), for which the slip amount of each subfault is the median of the cluster. Inserted values indicate the Mw and frequency of the cluster and variance reduction (VR) with horizontal and vertical components. Black rectangles indicate the area of seafloor vertical deformation shown in the middle panels. Middle: seafloor vertical deformation calculated from the left slip distribution using Okada (1992). Inserted values indicate the maximum and minimum vertical deformation. Black rectangles indicate the area shown in the tsunami inundation map to the right. Right: Tsunami inundation, where colors indicate tsunami height. Inserted
values indicate the inundation area.

Figure 9. Clustering using k-means with the slip vector as the feature value. a: Elbow method when the number of clusters is changed. b: Frequency distribution of each cluster when $K = 100$ and $K = 300$. Red and light blue bins indicate clusters with inundation areas of 60 km$^2$ or more, and 40 to 60 km$^2$, respectively. c: Correlation between the frequency of each cluster and the variation of the slip distribution, which is the scalar value obtained by adding the slip ranges of all perturbation faults for each cluster. d: Correlation between Mw and variance reduction (VR) of the median model (tsunami scenario) of each cluster.

Figure 10. Tsunami inundation calculation results for multiple tsunami scenarios. Correlation between the frequency of each cluster and the tsunami inundation area. The dotted black line indicates the inundation area of assumed model (Figure 1). The vertical and horizontal axes on the inserted diagrams indicate the frequency distributions. Inserted vectors indicate the scenarios of Fig. 9 (a, b).
Table S1. Tsunami inundation calculation parameters

| Contents                                      | Details                                      |
|----------------------------------------------|----------------------------------------------|
| Governing equation                           | Nonlinear longwave equation                  |
| Calculation method                           | Finite difference method                     |
| Boundary condition                           | Open boundary conditions for free transmission; run up along a coastline |
| Time to calculate                            | 6 h                                          |
| Grid size                                    | 30 m, 90 m, 270 m, 810 m                    |
| Time interval                                | 0.5 s                                        |
| The still water level                        | 0 m                                          |
| Top of the run-up wave                       | 0.01 m                                       |
| Manning’s roughness                          | n = 0.025                                    |
| Structure                                    | Breakwaters (buildings not taken into account) |
| Wave source                                  | Slip distribution model                      |
Figure S1. Area settings for the tsunami inundation calculation. There are four grid sizes (black frame: 810 m grid, green frame: 270 m grid; blue frame: 90 m grid, red frame: 30 m grid). The red frame is the target area for tsunami inundation.
Figure S2. Editing attempt for a real-time tsunami inundation risk map (K = 100). a: Case where an inundation area of 60 km$^2$ or more is excluded. b: Case where three inundation areas of 40 km$^2$ or more are excluded. Before data removal, the map was as shown in Fig. 7(a).
Figure S3. Relationship between vertical seafloor deformation and inundation area. a: Modeling grid, where the red rectangle indicates the target area of tsunami inundation calculation and the black rectangle indicates the area where we calculated vertical seafloor deformation. b: Vertical seafloor deformation. c: Correlations between sum of absolute vertical seafloor deformation at 0.5 degree intervals and inundation area for models 185 and 388. The dotted black line indicates the inundation area of assumed model (Figure 1). In each stage, we extracted 300 cases from the sample with the representative value of variance reduction (VR), in descending order of vertical seafloor deformation, and calculated tsunami inundation. A sample with a large vertical seafloor deformation did not necessarily have a large tsunami inundation area. This is because we estimated vertical seafloor deformation for the entire region, whereas tsunami inundation area was evaluated for the local region.
Figure S4. Editing attempt for a real-time tsunami inundation risk map (K = 100). a: Map weighted using the frequencies of the clusters. b: Map where inundation depth of 1 m or more is regarded as "inundation". In (a), there is almost no change in the frequency tendency of the inundation area because there was no strong correlation between the frequency and the inundation area. In (b), there is a change in the decrease in the inundation probability because inundation of less than 1 m was not counted.
Figure S5. Five slip distribution scenarios. All scenarios consist of 2, 5, and 10 m slips. The inserted value indicates the Mw of the assumed slip. a: (1) Large slip along the trench axis of the entire Nankai Trough. b: (2) Slip on the western side of the Kii Peninsula. c: (3) Slip on the eastern side of the Kii Peninsula. d: (4) Slip in the Tonankai area. e: (5) Slip in the Tokai area.
Figure S6. (1) Slip distribution when large slip occurs along the trench axis of the entire Nankai Trough. Left column: median model of posterior probability density functions (PDFs) of each subfault. Inserted values indicate the Mw and variance reduction (VR) of the median model. Black and white vectors indicate inputted and calculated horizontal displacements. Right column: 95% confidence intervals (CI) of posterior PDFs for each subfault, as calculated from the difference between the samples located at 2.5% and 97.5%.
Figure S7. (2) Slip distribution when slip occurs on the west side of the Kii Peninsula.

Left column: median model of posterior probability density functions (PDFs) of each subfault. Inserted values indicate the Mw and variance reduction (VR) of the median model. Black and white vectors indicate inputted and calculated horizontal displacements. Right column: 95% confidence intervals (CI) of posterior PDFs for each subfault, as calculated from the difference between the samples located at 2.5% and 97.5%.
Figure S8. (3) Slip distribution when slip occurs on the east side of the Kii Peninsula.

Left column: median model of posterior probability density functions (PDFs) of each subfault. Inserted values indicate the Mw and variance reduction (VR) of the median model. Black and white vectors indicate inputted and calculated horizontal displacements. Right column: 95% confidence intervals (CI) of posterior PDFs for each subfault, as calculated from the difference between the samples located at 2.5% and 97.5%.
Figure S9. (4) Slip distribution when slip occurs in the Tonankai area. Left column: median model of posterior probability density functions (PDFs) of each subfault. Inserted values indicate the Mw and variance reduction (VR) of the median model. Black and white vectors indicate inputted and calculated horizontal displacements. Right column: 95% confidence intervals (CI) of posterior PDFs for each subfault, as calculated from the difference between the samples located at 2.5% and 97.5%.
Figure S10. (5) Slip distribution when slip occurs in the Tokai area. Left column: median model of posterior probability density functions (PDFs) of each subfault. Inserted values indicate the Mw and variance reduction (VR) of the median model. Black and white vectors indicate inputted and calculated horizontal displacements. Right column: 95% confidence intervals (CI) of posterior PDFs for each subfault, as calculated from the difference between the samples located at 2.5% and 97.5%.
Figure S11. Real-time tsunami inundation risk map for Kochi city (K = 300) overlaid on a shadow topography map (Geospatial Information Authority of Japan). Colors indicate normalized frequency of inundation for 300 tsunami inundation calculations. The dotted circle indicates an area exhibiting a slight difference compared with K = 100.