

DETERMINATION OF THE INITIAL SEA SURFACE DEFORMATION CAUSED BY A TSUNAMI USING GENETIC ALGORITHM

Iyan E. Mulia¹ and Toshiyuki Asano¹

The characteristics of a tsunami is largely determined by its initial source, which can be defined as a perturbation to the equilibrium state of the ocean hydrodynamic. For tsunamis generated by earthquakes, the initial source (as a form of sea surface deformation) is driven by a coseismic vertical displacement of the seabed. Assuming that the rupture process occurs instantaneously, the sea surface deformation can be inferred from recorded tsunami waveforms. Tsunami waveform inversion using the Green's function technique with least squares optimization is the most widely used method. However, this method can sometimes suffer from ill-posed problems, because of the non-uniqueness of the inversion results. This study proposes a genetic algorithm that helps the traditional least squares method find an optimum solution. We applied the method to an artificial tsunami source with a complex profile. We achieved significant improvements by determining the optimum spatial distribution of the unit sources (the Green's function) prior to the inversion. In contrast with the regular Green's function that has equidistant unit sources, our method generates a random spatial distribution. This random characteristic leads to a better approximation of the initial sea surface deformation.

Keywords: tsunami waveform inversion; genetic algorithm; stochastic optimization

INTRODUCTION

A conservative approach for simulating tsunami sources is to use a simplified elastic dislocation model based on an analytical formula (Okada 1985). However, this method requires an assumption of a uniform slip distribution throughout the rupture area, so the rupture complexity cannot be well represented. A more realistic model of a tsunami source uses tsunami waveform inversion. It considers that the recorded waveforms exhibit the signature of the initial profile of the tsunami, over a specific time period (Satake 1987). Furthermore, tsunami waveform inversion partitions the fault into smaller subfaults, and determines the specific fault parameters for each subfault using the Green's function technique. As a result, we can estimate the heterogeneous slip distribution observed in nature.

Our study is similar to Satake's, but we disregarded the fault model assumptions because we are more interested in estimating the sea surface deformation than a slip on the fault plane. Therefore, we use the term "unit source" instead of subfault. More recently, several studies have used tsunami waveform inversion to estimate the sea surface deformation without fault model assumptions. The basic premise is to replace the fault model by an auxiliary basis function on unit sources (Baba et al. 2005 and Saito et al. 2010). The least squares method (LSM) is the most frequently used technique for fitting synthetic waveforms to the observations. However, the least squares inversion method for tsunami waveforms possesses a limitation, in that the inverse matrix does not always exist because of the non-uniqueness of the solution. In addition to the a large number of unknown parameters, which might produce many local optima on the misfit function measure, the search towards optimality is confined by the uniform distribution of the unit sources used in the regular Green's function.

In this paper, we propose a new approach that helps the LSM find stable and optimal solutions, by determining the optimal position or spatial distribution for the unit sources located around the tsunami source or epicenter. Genetic algorithm (GA) is a global optimization method, which we used to search for these positions prior to the inversion. Because the selected positions are probably located between the initial unit sources, we used interpolations during the optimization to compute the respective waveforms. Therefore, the Green's function dynamically evolves at each generation of the GA. In other words, the GA provides various Green's functions to be evaluated by the LSM in an intelligent manner, and each new Green's function yields a better least squares fit.

MATERIALS AND METHOD

We have conducted numerical experiments using an artificial tsunami source with a complex profile propagated on an actual bathymetry data. We chose an area covering 140–145°E and 35–41°N as the domain of interest. There are eight artificial observation stations associated with the actual location of gauges within the study area. For the initial Green's function, we divided the tsunami source into 28 unit sources that were separated by a uniform distance of approximately 60 km (Fig. 1).

¹ Department of Ocean and Civil Eng., Kagoshima University, Korimoto 1-21-40, Kagoshima, 890-0065, Japan

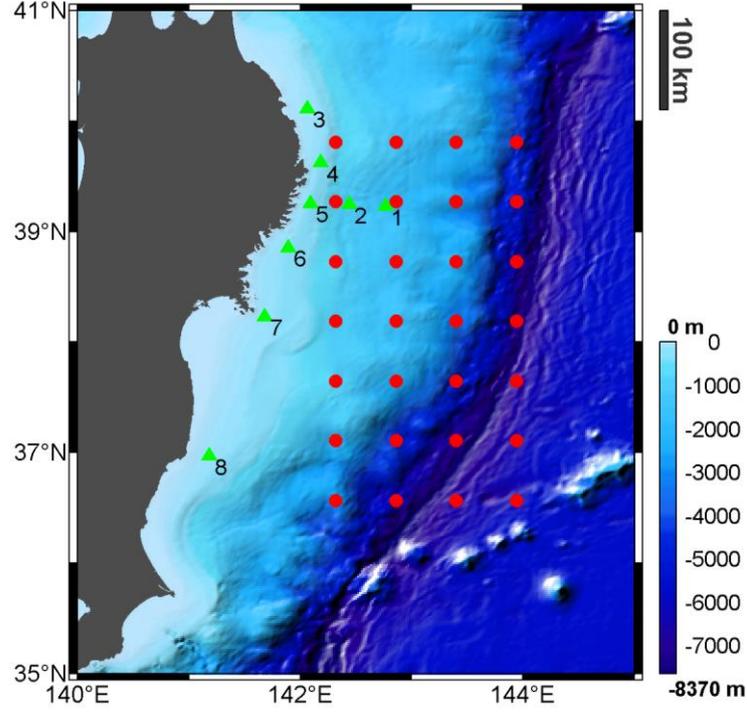


Figure 1. Study area and bathymetry profile. The red dots indicate the locations of unit sources and the green triangles with numbers are the artificial observation stations.

The bathymetry data were obtained from ETOPO1 having a spatial resolution of 1 arc minute. ETOPO1 is a global relief model of the earth's surface, which includes ocean bathymetry. It is available from the National Geophysical Data Center of the National Oceanic and Atmospheric Administration (Amante and Eakins 2009).

Tsunami waveform inversion using the LSM

We used a linear non-dispersive shallow water equation in the forward model to compute the time histories of the sea surface elevation at the specified observation points. That is,

$$\begin{cases} \frac{\partial \mathbf{V}}{\partial t} = -g \nabla \eta \\ \frac{\partial \eta}{\partial t} = -\nabla \cdot \{(d + \eta) \mathbf{V}\} \end{cases}, \quad (1)$$

where η is the water elevation of the tsunami, $\mathbf{V}(u,v)$ is the depth-averaged horizontal fluid velocity vector, d is the water depth, and g is gravitational acceleration. Equation 1 has two boundary conditions,

$$\begin{cases} \mathbf{V} \cdot \mathbf{n} = \frac{g}{c} \eta \\ \mathbf{V} \cdot \mathbf{n} = 0 \end{cases}, \quad (2)$$

where $c = \sqrt{gh}$ is the wave speed, and \mathbf{n} is the unit vector normal to the boundary. The upper and lower expressions in Eq. 2 represent the open and closed boundaries, respectively.

We used a Gaussian shaped basis function with 1-m amplitude and a diameter of approximately 80km as the basis function for each unit source (Liu and Wang 2008). The Green's function $G(x,y,t)$ represents the recorded synthetic waveforms at a location (x,y) , which we then constructed for each i -th unit source. The total observed sea surface fluctuations, ζ , at (x,y,t) can be expressed as

$$\zeta(x, y, t) = \sum_{i=1}^{i=N} w_i G_i(x, y, t), \quad (3)$$

where w_i is the weighting factor for each unit source (the unknown parameters to be determined by the inversion) and N is the number of unit sources. In vector notation, Eq. 3 can be reformulated as

$$\zeta = \mathbf{wG} . \tag{4}$$

Based on the LSM, we can determine the vector of unknown parameters \mathbf{w} by solving the inverse equation

$$\mathbf{w} = (\mathbf{G}^T \mathbf{G})^{-1} \mathbf{G}^T \zeta . \tag{5}$$

Genetic algorithm

A GA is a global optimization method that search for an optimal value of a complex function, and is based on natural evolution (Goldberg 1989). There are three basic operators in a GA: selection, crossover, and mutation. Briefly, the algorithm starts by generating random populations of chromosomes. These are the candidate solutions for a given problem. Then, we evaluate the fitness function of each chromosome to determine the selection probabilities. In this study, we used a combination of the root mean square error and Pearson correlation coefficient as the fitness function. Pairs of chromosomes are selected and combined using the crossover operation. The components of the chromosome (the genes) are altered at randomly selected positions. This last genetic manipulation is called mutation. The aim of these three genetic operations is to generate better offspring. The offspring produced by the genetic manipulation processes are then the next population to be evaluated. The optimization is terminated when a user-defined converging criterion is satisfied. A schematic illustration of a GA is shown in Fig. 2.

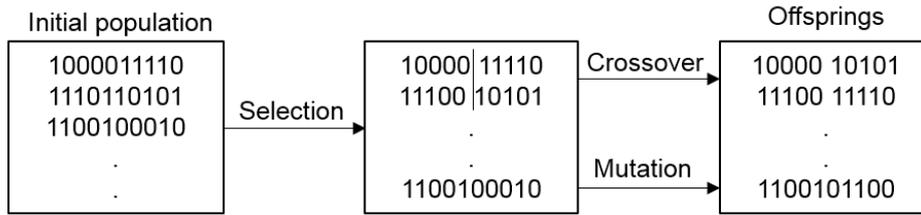


Figure 2. Schematic illustration of GA.

We used a GA to search for the optimal locations of unit sources, and calculated the initial water elevation using the LSM. The area is bounded according to the spatial distribution of the initial unit sources, or the inverse region used to construct the Green’s function. We identified the computational grids of the tsunami propagation model located inside the inverse region using unique integer numbers. This is to avoid using longitude and latitude, which would double the number of model parameters. The GA searched for the best possible location of unit sources based on the identification number. Therefore, in our study, the GA was used to solve an integer programming problem.

During optimization, the selected source locations may not exactly be in the initial 28 unit sources. Therefore, we needed to use interpolation to produce the sea surface fluctuations at the observation stations. As a result, the Green’s function evolved in each iteration of the GA. It is difficult to deterministically calculate a perfect Green’s function (Yagi and Fukuhata 2011). Therefore, in this study, we used a stochastic approach with the GA to overcome this issue.

Waveform interpolation

The waveform interpolation had two stages. We first interpolated the arrival time, and then the waveform amplitude. In the first stage, we interpolated the arrival time of the GA-selected unit sources based on the four nearest unit sources in the initial Green’s function. We used nearest neighbor weighted interpolation to estimate the properties of interest. The general form of the method is given by

$$P = \begin{cases} \frac{\sum_{i=1}^N \frac{P_i}{\sqrt{(x_i - x)^2 + (y_i - y)^2}}}{\sum_{i=1}^N \frac{1}{\sqrt{(x_i - x)^2 + (y_i - y)^2}}} & x_i \neq x \text{ or } y_i \neq y \\ P_i & x_i = x \text{ and } y_i = y \end{cases} , \tag{6}$$

where P is the property of interest (in our case, either the arrival time or wave amplitude), i is the index of the nearest data points (which has a total of $N=4$), and x and y represent the position of the unit source. After computing the arrival time of the GA-selected unit source, the waveforms of the four nearest unit sources were shifted according to the estimated arrival time. As a result, the waveforms of the four nearest unit sources then had the same phase. We then interpolated the wave amplitude. The phase-shifting process was necessary because it is difficult to perform a direct interpolation of waveforms with different phases.

Model development

We created an artificial tsunami source generated from a superposition of small unit sources with random amplitudes and positions inside the inverse region (Fig. 3a). We expect that the randomness exhibited by the artificial source mimics actual tsunami sources that are often associated with complexities and uncertainties (Geist 2005). The proposed model aims to address such difficulties in tsunami waveform inversion, which are sometimes not accessible by deterministic approaches. After computing the recorded waveforms that originated from the artificial source, we prepared the model. The model development process can be summarized as follows.

1. Construct the initial Green's function.
2. Initialize the GA model by randomly distributing the unit source locations. The search for the optimal location is bounded by the area of the inverse region.
3. Perform interpolation and update the Green's function.
4. Evaluate the fitness using least squares inversion.
5. After reaching the stopping criteria, run the forward numerical model for each of the optimized unit sources to avoid errors generated from the interpolation results. Then perform one final inversion.

RESULTS AND DISCUSSION

Our inversion results for the sea surface deformation using the LSM and GA-LSM can be seen in Fig. 3b and 3c, respectively. The artificial tsunami source used as the target to be approximated is also shown in Fig. 3a, to aid qualitative assessments on the model performance. Figure 3b clearly shows unexpected patterns at the sea surface profile resulting from the LSM. The profile deviates from the target source at several locations. This is probably caused by the coarse spatial resolution of the unit sources, which cannot capture the details of the target source profile. A finer resolution should increase the accuracy (Baba et al. 2005). However, such an approach is difficult to achieve without introducing a proper constraint for maintaining the stability of the inversion. The number of model parameters is proportional to the degrees of freedom in the optimization, so a large number of parameters may lead to a local optimum problem. A smoothing constraint should be enforced if the spatial resolution of the unit source increases. This maintains the stability of the inversion, and also avoids violating the long wave assumption that the surface profile should remain smooth with a wavelength of approximately 20 times the average water depth.

The smoothing constraint, however, could restrict the exploration through the feasible search space, and make it more difficult to discover an optimum solution. Therefore, in this study, we preferred not to use such a constraint. Instead, we used the GA to help the LSM find the optimum and stable solution, by providing it the best Green's function obtained through stochastic processes. At each iteration of the GA, the Green's function was updated according to the selected unit source locations. To avoid excessive computational efforts, we used interpolation when the selected locations were between the initial Green's functions. As explained in the previous section, we then recalculated the optimized unit source locations to reduce the errors caused by the interpolation process. This is because, in least squares inversion, small errors in the solution can be significantly amplified as a result of the ill-posed problem.

Another plausible explanation for the unsatisfactory results of the LSM is that the uniform distances between the unit sources limited the search. This is demonstrated by the results of the GA-LSM model (Fig. 3c). Using the proposed method, the same number of model parameters (28 unit sources) with random locations yielded a much better approximation than the LSM. The optimum locations of the unit sources calculated using the GA appear random (see the gray dots in Fig. 3c). This is the likely cause for the more precise approximations of the target source. This also indicates that optimization using LSM with the regular Green's function is limited by the equidistant unit sources. The GA-LS model improved the inversion accuracy without increasing the spatial resolution. Instead, it altered the location of the initial unit sources. Thus, we do not need a smoothing constraint in this case, because the number of model parameters is the same as in the initial Green's function (28). However, we may need this constraint in real applications, especially for a large tsunami that can only accurately reconstructed using hundreds of unit sources.

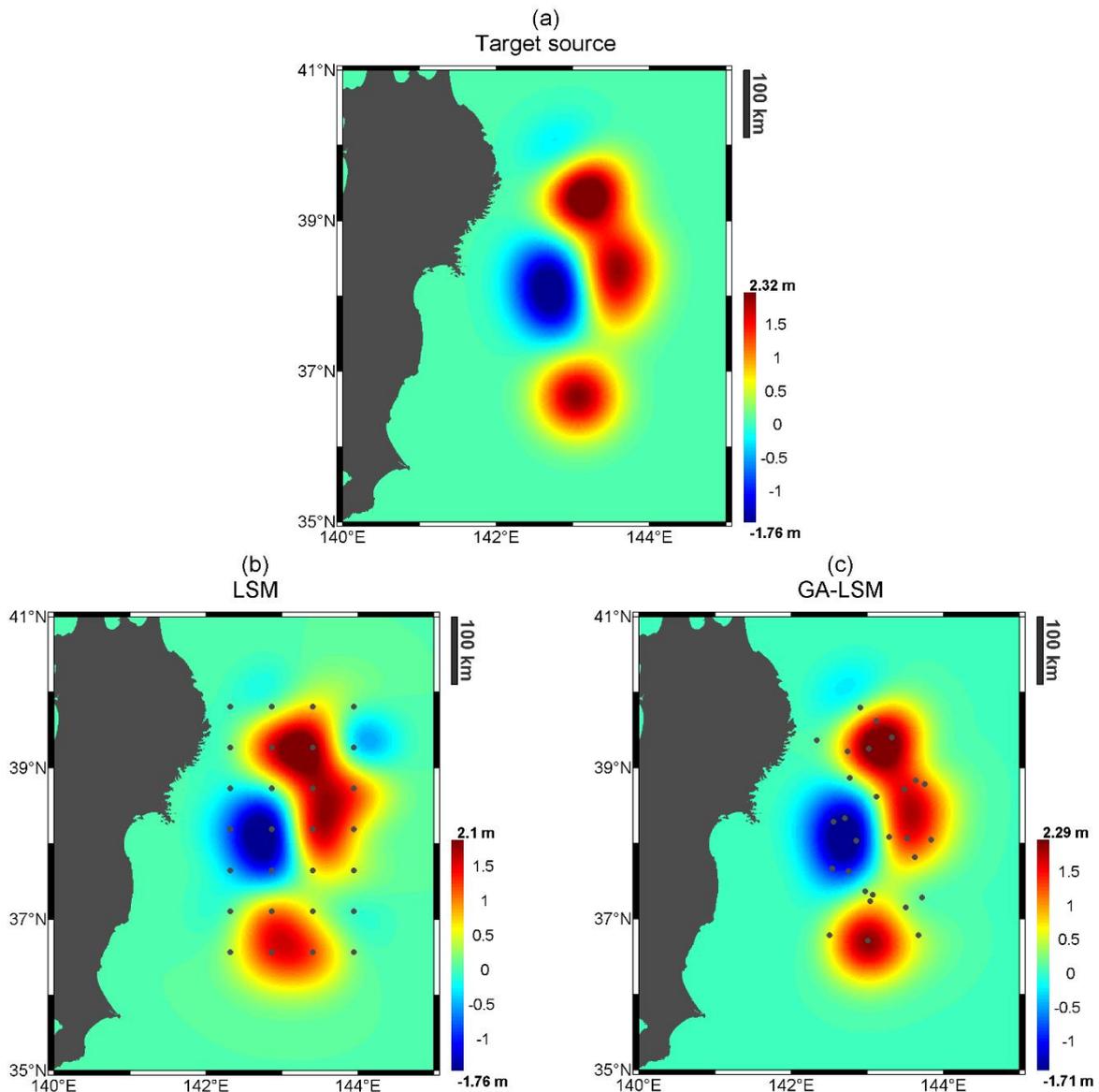


Figure 3. Sea surface deformation. The gray dots indicate the centroids of the unit sources. (a) Target source, (b) inverted source using LSM, (c) inverted source using GA-LSM.

We further evaluated the model results by comparing the time series of the waveforms recorded on the gauges (Fig. 4). We also performed statistical evaluations to quantify the model performances. The time series comparisons of the inverted waveforms from both methods confirmed the results of the inverted sea surface deformations. The waveforms inverted using the GA-LSM were a much better fit than those from the LSM. A clear example can be seen in Gauge 5; the LSM method failed to reconstruct the first leading wave, even though it was located within the inversion time range. The fitness levels of all the gauges with respect to the waveforms originating from the target source are presented as scatter plots in Fig. 5. The correlation coefficient (r) of the GA-LSM model was 0.9987, and was 0.9908 for the LSM. The correlation coefficient is sensitive to phase matching between the compared series (Barnston 1992). Additionally, the root mean square error (RMSE) is typically used to assess the fitness level in terms of the amplitude. The RMSE for the GA-LSM was 0.0239, and was 0.0622 for the LSM. This means that the GA-LSM model produced waveforms that were approximately 2.5 times more accurate in terms of amplitude matching, when compared to the LSM.

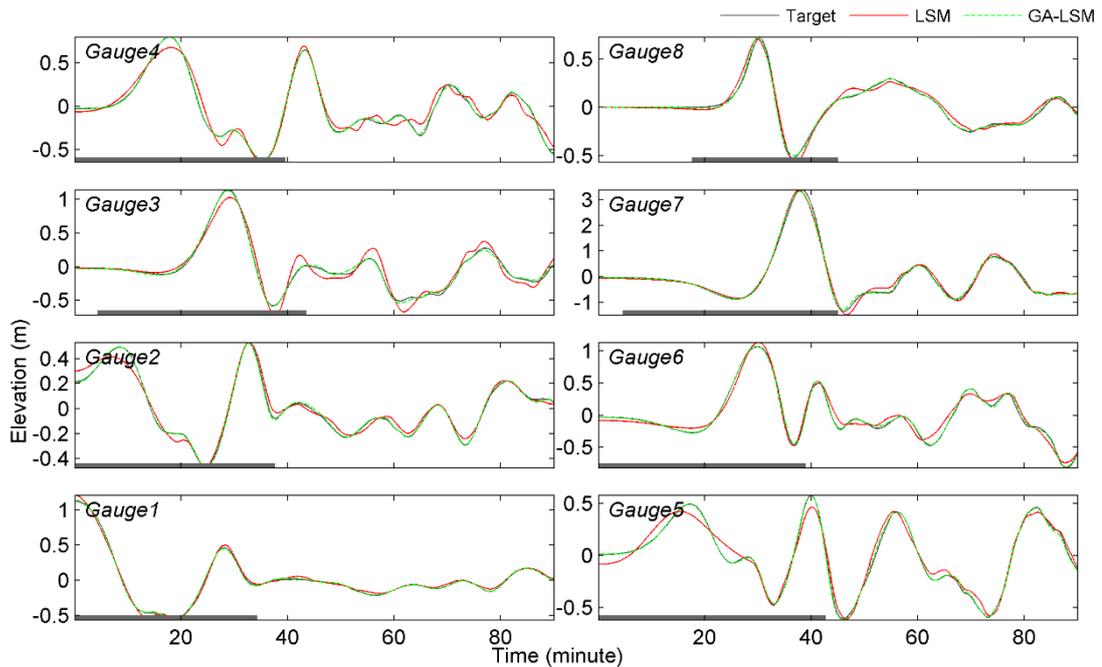


Figure 4. Comparison of the waveforms at gauges. The gray bar above the time axis indicates the time range for the inversion.

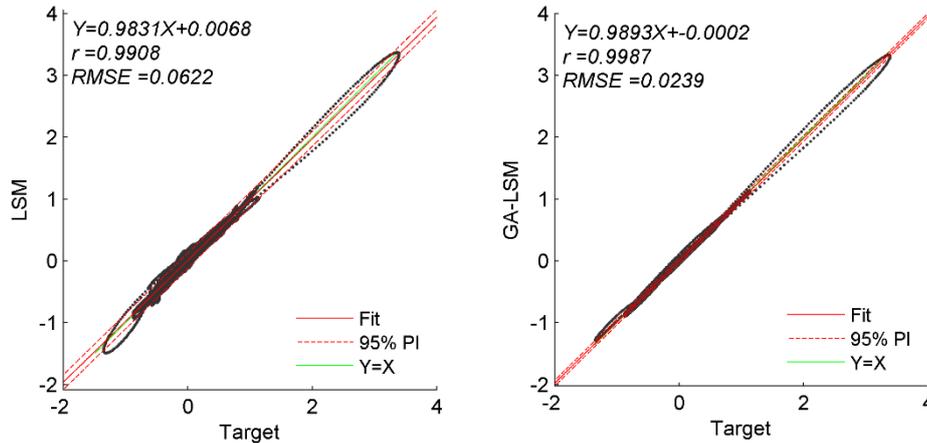


Figure 5. Scatter plots of the inverted waveforms with respect to the waveforms of the target source, for all gauges.

CONCLUSIONS

We conducted numerical experiments to determine the initial sea surface deformation caused by a tsunami. Our proposed method uses a combination of a GA and the LSM, and yielded a better approximation of the initial sea surface deformation than conventional tsunami waveform inversion using only the LSM. We concluded that the inversion results using the regular Green's function are limited by the equidistant unit sources. This restricted the search for optimal solutions. In this study, we used a GA to locate the optimum locations of unit sources prior to the inversion using the LSM with a stochastic process. The final spatial distribution of the unit sources appeared random, which helped the LSM find a unique and optimum solution. However, it is not straightforward to evaluate the actual physics of the stochastic optimization results, when compared to deterministic results. Therefore, in real applications, we need prior knowledge on the physical phenomena of the events of interest, so that we can better

interpret the model results. Further extensions of this study with additional optimization options can be found in Mulia and Asano (2014).

REFERENCES

- Amante, C. and Eakins, W. 2009. ETOPO1 1 arc minute global relief model: procedures, data sources and analysis, *NOAA technical memorandum NESDIM NGDC-24*, 19 pp.
- Baba, T., Cummins, P. R., and Hori, T. 2005. Compound fault rupture during the 2004 off the Kii Peninsula earthquake (M 7.4) inferred from highly resolved coseismic sea-surface deformation, *Earth Planets Space*, 57, this issue, 167–172.
- Barnston, A. G. 1992. Correspondence among the Correlation, RMSE, and Heidke Forecast Verification Measures; Refinement of the Heidke Score, *Wea. Forecasting*, 7, 699–709.
- Geist, E. L. 2005. Local Tsunami Hazards in the Pacific Northwest from Cascadia Subduction Zone Earthquakes, *U.S. Geological Survey Professional Paper*, 1661-B, 17 pp.
- Goldberg, D. E. 1989. *Genetic algorithms in search, optimization and machine learning*, Longman Publishing Co., New York.
- Liu, P. L. F. and Wang, X. 2008. Tsunami source region parameter identification and tsunami forecasting, *Journal of Earthquake and Tsunami*, Vol. 2, No. 2, 87-106.
- Mulia, I. E. and Asano, T. 2014. Randomly distributed unit sources to enhance optimization in tsunami waveform inversion, *Nat. Hazard Earth Syst. Sci. Discuss.*, 2, 3659-3682.
- Okada, Y. 1985. Surface deformation due to shear and tensile faults in a half-space, *Bull. Seism. Soc. Am.*, 75, 1135-1154.
- Satake, K. 1987. Inversion of tsunami waveforms for the estimation of fault heterogeneity: Method and numerical experiments, *J. Phys. Earth*, 35, 241-254.
- Saito, T., Satake, K., and Furumura, T. 2010. Tsunami waveform inversion including dispersive waves: the 2004 earthquake off Kii Peninsula, Japan, *J. Geophys. Res.*, 115, B06303, doi: 10.1029/2009JB006884.
- Yagi, Y. and Fukuhata, Y. 2011. Introduction of uncertainty of Green's function into waveform inversion for seismic source processes, *Geophys. J. Int.*, 186, 711-720.