Doctoral Dissertation

博士論文

Tsunami data assimilation for early warning

（早期警報のための津波データ同化）

A Dissertation Submitted for the Degree of Doctor of Philosophy

July 2021

令和 3 年 7 月博士（理学）申請

Department of Earth and Planetary Science, Graduate School of Science,

The University of Tokyo

東京大学大学院理学系研究科地球惑星科学専攻

Yuchen Wang

王 宇晨

Abstract

The tsunami data assimilation approach has been proposed for tsunami early warning. It estimates the tsunami waveform by assimilating observed offshore data into a numerical simulation without considering the source based on Optimal Interpolation method. However, previous data assimilation approach has a relatively high computational cost as it is necessary to run numerical simulations to obtain the entire tsunami wavefield during the assimilation process. The previous approach also requires a dense offshore observation network, making it difficult to conduct data assimilation in regions with sparse observations.

In this thesis, I first proposed Green's Function-based Tsunami Data Assimilation (GFTDA) to reduce the computation time for assimilation. It can forecast the waveform at Points of Interest (PoIs) by superposing Green's functions between observational stations and PoIs. Unlike the previous assimilation approach, GFTDA does not require the calculation of the tsunami wavefield for the entire region during the assimilation process because Green's functions are calculated in advance. A simple matrix manipulation can allow the calculation of the forecasted waveforms. I applied this method to data from the 2012 Haida Gwaii earthquake, the 2004 off the Kii Peninsula earthquake, and the 2015 Torishima volcanic tsunami earthquake. GFTDA achieved an equivalently high accuracy of tsunami forecasting as the previous approach, while saving sufficient time to achieve an early warning. It also enabled the application of a more complicated tsunami propagation model to data assimilation, such as the linear dispersive (DSP) model, which predicts the tsunami arrival time more accurately.

Next, I proposed a modified tsunami data assimilation method for regions with a sparse observation network. The method uses interpolated waveforms at virtual stations to construct the complete wavefront for tsunami propagation. The artificial tsunami waveforms at the virtual stations between two existing observational stations could be estimated by shifting arrival times with linear interpolation of observed arrival times and by correcting the amplitudes using water depths. This was based on the assumption that tsunamis propagate as a plane wave, and that the wavefront is nearly a straight line. After computing the waveforms for virtual stations, the data assimilation algorithm was applied to real and virtual stations. Its application to the 2004 Sumatra–Andaman earthquake, the

2009 Dusky Sound, New Zealand earthquake, and the 2015 Illapel earthquake revealed that adopting virtual stations greatly improved the tsunami forecasting accuracy for regions without a dense observation network.

Additionally, I proposed a real-time tsunami detection algorithm using Ensemble Empirical Mode Decomposition (EEMD). EEMD adaptively decomposes a time series into a set of Intrinsic Mode Functions (IMFs). The tsunami signals of the Offshore Bottom Pressure Gauge (OBPG) can be automatically separated from the tidal components, seismic waves, and background noise. Unlike traditional tsunami detection methods, the new algorithm does not require tidal predictions. The application to the actual data of cabled OBPGs off the Tohoku coast showed that it successfully detected the tsunami from the 2016 Fukushima earthquake (M 7.4). The method was also applied to the extremely large tsunami from the 2011 Tohoku earthquake (M 9.0) and extremely small tsunami from the 1998 off Sanriku earthquake (M 6.4). The algorithm detected the former, which caused devastating damage, whereas it did not detect the latter micro tsunami, which was not noticed on the coast. The algorithm was also tested for a month-long OBPG data, and no false alarm occurred. Hence, the algorithm could detect tsunami arrival with a short detection delay and accurately characterize the tsunami amplitude.

Furthermore, I combined the tsunami data assimilation approach with the real-time tsunami detection algorithm. The tsunami of the 2016 Fukushima earthquake was recorded by the offshore pressure gauges of the Seafloor Observation Network for Earthquakes and Tsunamis (S-net). I used 28 S-net pressure gauge records for tsunami data assimilation and forecasted the tsunami waveforms at four tide gauges on the Sanriku coast. The S-net raw records were processed using two different methods. In the first method, I removed the tidal components by polynomial fitting and applied a low-pass filter. In the second, I used the real-time tsunami detection algorithm based on EEMD to extract tsunami signals, imitating real-time operations for tsunami early warning. The scores of forecasting accuracy of the two detection methods were 60% and 74%, respectively, for a time window of 35 min, which improved to 89% and 94%, respectively, if stations with imperfect modeling or insufficient offshore observations were omitted. Hence, the proposed tsunami data assimilation approach can be put into practice with the help of the real-time tsunami detection algorithm. Finally, I proposed a tsunami early warning system using data assimilation of offshore data for Japan.

Contents

Figures

Figure 1.1. Three types of tsunami forecasting methods.

Figure 1.2. Offshore observation network composed of OBPGs: (a) DART; (b) DONET; (c) S-net.

Figure 2.1. Illustration of the previous data assimilation approach and GFTDA.

Figure 2.2. Assimilation process of GFTDA.

Figure 2.3. Bathymetry and tsunami waveforms of the observational stations in the Cascadia Initiative Community Experiment.

Figure 2.4. Comparison of the observed and forecasted waveforms at 11 PoIs with an assimilation time window of 24 min.

Figure 2.5. Forecast accuracy and score versus different assimilation time windows.

Figure 2.6. Illustration of the observation network and near-shore PoIs.

Figure 2.7. Distribution of 15 observational stations and waveforms of synthetic tsunamis.

Figure 2.8. Simulated waveforms and waveforms forecasted by data assimilation at nine near-shore PoIs.

Figure 2.9. Forecast accuracy (a) and time lag (b) of two models for various assimilation time windows.

Figure 2.10. Bathymetry map of the 2015 Torishima earthquake.

Figure 2.11. Raw records from 25 observational stations of DONET and DSFO.

Figure 2.12. Processed data from 25 observational stations of DONET and DSFO.

Figure 2.13. Waveform comparison of two PoIs.

Figure 2.14. Spectrum comparison of two PoIs.

Figure 2.15. Relationship between the number of stations and general error function G.

Tables

Table 5.1. Detected tsunami arrival time at each S-net station.

Chapter 1   Introduction

1.1 Tsunami Early Warning

Tsunamis are devastating natural disasters that can cause substantial loss of life and severe property damage. The 2011 Tohoku Earthquake (M 9.0) generated a catastrophic tsunami that killed thousands of people and caused serious damage to the nearby Fukushima Daiichi Nuclear Power Station (Satake, 2015). Tsunami early warning for near-shore communities can help enable timely evacuation. Accurate and rapid forecasting of impending tsunamis is essential for successfully reducing the loss of life and property damage (Tsushima et al., 2009).

Generally, there are three types of tsunami forecasting methods for early warning (Figure 1.1). For most seismogenic tsunamis, the traditional tsunami early warning approach relies on rapid estimates of the earthquake location, depth, and seismic moment (Lomax and Michelini, 2011). For example, the Japan Meteorological Agency (JMA) issues the first tsunami warning for near-field tsunamis about 3 minutes after an earthquake (Kamigaichi, 2015). The earthquake location, depth and magnitude are estimated from seismic data, and then used to search tsunami simulation database that stores tsunami calculation results for many earthquake-fault models. The tsunami arrival times and amplitude on 66 forecasting segments of Japanese coasts are used for tsunami warning. In addition, the Pacific Tsunami Warning Center (PTWC) issues tsunami early warnings based on the results of W-phase source inversion (Duputel et al., 2011). The W phase is a fast, long-period seismic wave used for rapid point source moment tensor inversions (Benavente et al., 2016; Kanamori and Rivera, 2008). W-phase inversion can be completed within 5–10 min when seismic data at regional distances are used. After an earthquake, the point source moment tensor is obtained using the regional seismic data, and forward simulation is then used to calculate the tsunami wavefield (Duputel et al., 2011; Tang et al., 2016; Wang et al., 2012). For the 2015 Illapel earthquake, the PTWC issued a tsunami early warning to the Chilean coast after the earthquake and tsunami warnings or advisories for other regions, such as Easter Island, New Zealand, Fiji, Hawaii,

and the Solomon Islands. It later downgraded the warnings. This approach works quickly but is affected by uncertainties in the rapid seismic source inversion (Cienfuegos et al., 2018). Additionally, some earthquake events have an abnormal focal mechanism (e.g., Compensated Linear Vector Dipole; CLVD), making it difficult to forecast tsunamis from seismic information (Wang et al., 2019b).

Another tsunami forecasting method uses offshore tsunami data to invert the initial condition, such as the initial slip on fault or initial sea surface displacement. It then simulates the propagation process to forecast the tsunami height or inundation on coasts. For example, in Japan, Tsushima et al. (2009) proposed tsunami Forecasting based on Inversion for Sea-surface Height (tFISH). It inverts the tsunami waveform data to estimate the distribution of initial sea-surface height, then derives tsunami waveforms from the estimated source parameters. This method has been successfully applied to retroactive studies of the 1896 Sanriku and 2011 Tohoku earthquakes (Tsushima et al., 2009; 2011). In the United States, the National Oceanic and Atmospheric Administration (NOAA) attempted to build practical tsunami forecasting tools based on offshore tsunami data from the Deep-ocean Assessment and Reporting of Tsunami (DART) system, known as the Short-term Inundation Forecast for Tsunamis (Titov, 2009; Titov et al., 2005). Similar methods include the Near-field Tsunami Inundation Forecasting (Gusman et al., 2014), Time Reverse Imaging (TRI) (Hossen et al., 2015), and multi-index methods (Yamamoto et al., 2016a). However, these methods still must estimate the tsunami source so uncertainties in source estimation affect forecasting. Moreover, both types of methods require the origin time of an earthquake (Hossen et al., 2015; Tsushima et al., 2009), and they do not work for tsunami events caused by non-earthquake origins like landslides or volcanoes. For example, the collapse of the Anak Krakatau Volcano in December 2018 generated a tsunami in the Sunda Strait, Indonesia, with a maximum run-up height of 13.5 m (Muhari et al., 2019). Inversion methods have difficulties with such non-earthquake tsunamis.

Different from these two types of methods, the tsunami data assimilation approach is a new method of tsunami forecasting that completely avoids the complexities and uncertainties in source estimation. Data assimilation was originally used for operational numerical weather forecasts in meteorology and oceanology (Kalnay, 2003; Lynch, 2008), and was first proposed for tsunami forecasting by Maeda et al. (2015). The

tsunami data assimilation approach forecasts the tsunami based on offshore observations without considering the earthquake source parameters or the initial sea surface distribution. The tsunami wavefield, including tsunami height and velocities in two directions, can be estimated based on real-time data from tsunameters (Gusman et al., 2016b; Maeda et al., 2015). A major characteristic of the tsunami data assimilation approach is that it does not rely on source information, and is thus free from the complexities of source characterization (Cienfuegos et al., 2018; Yang et al., 2019). This method has already been successfully applied to retroactive studies of synthetic tsunamis in the Tohoku region of Japan, which were generated by an earthquake similar to the 2011 Tohoku earthquake (Maeda et al., 2015), and real tsunamis recorded by pressure gauges in the Cascadia subduction zone (Gusman et al., 2016b). More details on the tsunami data assimilation approach are provided in Section 1.3.
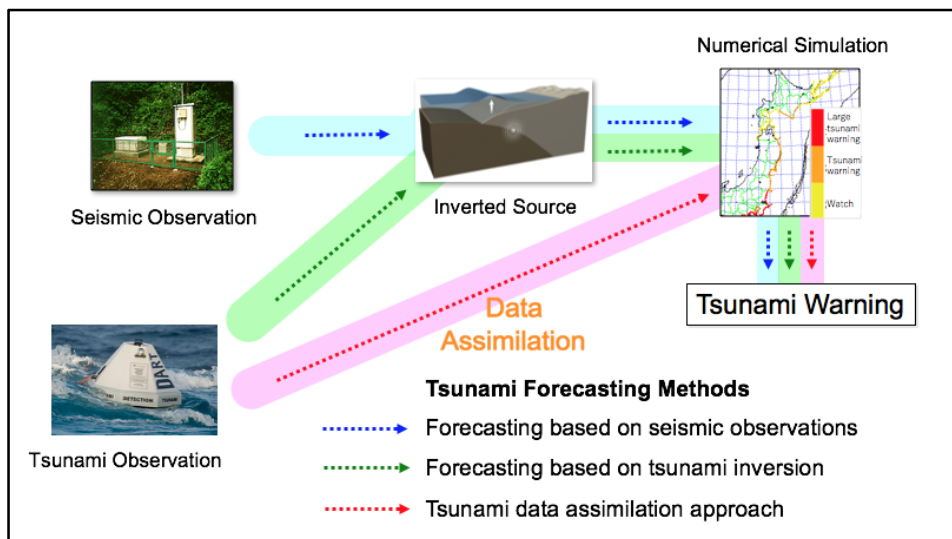
Figure 1.1. Three types of tsunami forecasting methods.

The blue arrow indicates the traditional tsunami early warning method based on seismic observations. The green arrow indicates the method using offshore tsunami data to invert the initial condition. The red arrow indicates the tsunami data assimilation approach.

1.2 Numerical Modeling of Tsunami Propagation

For either type of early warning, it is important to simulate the tsunami

propagation by numerical modeling. In this section, I will introduce the basic governing equations of tsunami propagation. Tsunami is a series of oceanic gravity waves as the gravity is the restoring force. The motion of fluid can be described by the following three-dimensional Navier-Stokes (3-D NS) equations, where we take the z axis in the vertical upward and the x and y axes in horizontal space in Cartesian coordinates (Saito and Furumura, 2009).

$$\frac{\partial V_x}{\partial t} + V_x \frac{\partial V_x}{\partial x} + V_y \frac{\partial V_x}{\partial y} + V_z \frac{\partial V_x}{\partial z} = -\frac{\partial}{\partial x}\left(\frac{p}{\rho}\right) + v\nabla^2 V_x \qquad (1.1)$$

$$\frac{\partial V_y}{\partial t} + V_x \frac{\partial V_y}{\partial x} + V_y \frac{\partial V_y}{\partial y} + V_z \frac{\partial V_y}{\partial z} = -\frac{\partial}{\partial y}\left(\frac{p}{\rho}\right) + v\nabla^2 V_y \qquad (1.2)$$

$$\frac{\partial V_z}{\partial t} + V_x \frac{\partial V_z}{\partial x} + V_y \frac{\partial V_z}{\partial y} + V_z \frac{\partial V_z}{\partial z} == -\frac{\partial}{\partial z}\left(\frac{p}{\rho}\right) + v\nabla^2 V_z - g \qquad (1.3)$$

Here V_x, V_y, and V_z are velocity components in x, y and z directions, respectively, p is the pressure, ρ is the water density, v is the kinematic viscosity coefficient for seawater, ∇^2 is the total Laplace Operator, and g is the gravitational acceleration. For large scale motion of water, the effects of viscosity are negligible. We also assume that the sea water is incompressible. Equations $(1.1) - (1.3)$ can be simplified as:

$$\frac{\partial V_x}{\partial t} + V_x \frac{\partial V_x}{\partial x} + V_y \frac{\partial V_x}{\partial y} + V_z \frac{\partial V_x}{\partial z} = -\frac{1}{\rho}\frac{\partial p}{\partial x} \qquad (1.4)$$

$$\frac{\partial V_y}{\partial t} + V_x \frac{\partial V_y}{\partial x} + V_y \frac{\partial V_y}{\partial y} + V_z \frac{\partial V_y}{\partial z} = -\frac{1}{\rho}\frac{\partial p}{\partial y} \qquad (1.5)$$

$$\frac{\partial V_z}{\partial t} + V_x \frac{\partial V_z}{\partial x} + V_y \frac{\partial V_z}{\partial y} + V_z \frac{\partial V_z}{\partial z} == -\frac{1}{\rho}\frac{\partial p}{\partial z} - g \qquad (1.6)$$

Equations $(1.4) - (1.6)$ are known as the Euler's equations of motion. When the horizontal scale of motion, or the wavelength, is much larger than the water depth, the vertical acceleration of water is negligible compared to gravity. This means that the horizontal motion of water mass is almost uniform from the ocean bottom to the surface. Such a wave is called a shallow-water wave or a long wave (Satake, 2015). In such case, the hydrostatic relation is assumed that the pressure is proportional to tsunami height. The 3-

D equations become a 2-D problem. The advection terms are neglected. For tsunamis, the typical ocean depth is about 4–5 km, and the source scale of large earthquakes is tens to hundreds kilometers. Therefore, tsunami propagation can be approximated by long-wave approximation. If the tsunami height is small compared to the water depth, Equations (1.4) – (1.6) become

$$\frac{\partial V_x}{\partial t} + g\frac{\partial h}{\partial x} = 0 \quad (1.7)$$

$$\frac{\partial V_y}{\partial t} + g\frac{\partial h}{\partial y} = 0 \quad (1.8)$$

Here h is the tsunami height at the sea surface. The equation of continuity, or conservation of mass, can be written as

$$\frac{\partial h}{\partial t} = -d\left(\frac{\partial V_x}{\partial x} + \frac{\partial V_y}{\partial y}\right) \quad (1.9)$$

Here d is the water depth. Equations (1.7) – (1.9) form the linear long-wave (LLW) tsunami propagation model.

However, the long-wave approximation breaks down when the wavelength of the water height distribution is not much greater than the water depth (Saito, 2019). In that case, the dispersive characteristics should be taken into consideration. When the dispersive effects are small ($kh \ll 1$; k is the wavenumber), for the horizontal propagation of waves, the 3-D NS equations can be reduced to a 2-D one by integrating the equations over the water depth (Nwogu, 1993; Roeber et al., 2010). The governing equations of the linear dispersive (DSP) model are derived from the 3-D NS equations, the continuity equation and the kinematic boundary condition at the surface (Saito, 2019; Saito & Furumura, 2009). The continuity equation is the same as Equation (1.9). However, the difference between the linear DSP model and the LLW model lies in the dispersion terms of the equation of motion (Satake, 1989). In the LLW model, the dispersion terms are neglected, so the right-hand sides of Equations (1.7) and (1.8) are zero. In the linear DSP model, the equation of motion becomes:

$$\frac{\partial V_x}{\partial t} + g\frac{\partial h}{\partial x} = \frac{1}{3}h^2\frac{\partial^2}{\partial x\partial t}\left(\frac{\partial V_x}{\partial x} + \frac{\partial V_y}{\partial y}\right) \quad (1.10)$$

$$\frac{\partial V_y}{\partial t} + g\frac{\partial h}{\partial y} = \frac{1}{3}h^2\frac{\partial^2}{\partial y\partial t}\left(\frac{\partial V_x}{\partial x} + \frac{\partial V_y}{\partial y}\right) \quad (1.11)$$

These equations are also known as the linear Boussinesq equations (Satake, 2015). I acknowledge that the linear Boussinesq equations or the linear DSP model will deviate from the exact dispersive relation when the dispersion is large ($kh \gtrsim 1$), but it is applicable to seismogenic tsunamis that have not so large dispersion. The nonlinearity is assumed to be small and negligible in the linear DSP model. Saito & Fumurura (2009) found that the 2-D tsunami simulation can reproduce the results of the 3-D simulation if we take dispersion effect into account. Though we neglect the vertical motion or variation, the 2-D DSP model is almost same as the 3-D model.

There are two types of boundaries considered in the tsunami computation. The first is a land-ocean boundary. In my study, a total reflection of energy on the coast is assumed. However, for the computation of tsunami inundation, a moving land-ocean boundary is needed. Another kind of boundary is the open ocean boundary along the outer area of the computational region. For the radiation conditions, waves are assumed to leave the computational domain and maintain their slopes (Aida et al., 1969).

As tsunami is a gravity wave, the conservation of energy flux can be described as

$$b_0 d_0^{1/2} h_0^2 = b_1 d_1^{1/2} h_1^2 \quad (1.12)$$

Here b is the distance between the rays. This equation is known as the Green's Law. It is a classic linear theory result and it describes the evolution of the wave height of periodic waves on plane beaches (Synolakis, 1990). In previous studies, Green's Law has been used to estimate the tsunami height based on such assumptions (Baba et al., 2004; Sandanbata et al., 2018; Wang et al., 2012). The water depth and tsunami height are defined as average values over a horizontal scale of at least one wavelength of tsunami. For plane waves, the ratio b_1/b_0 is 1. Hence, the tsunami height is inversely proportional to one-fourth of the power of water depth change.

In my study, I correct the tsunami height at tide gauges affected by coastal amplification (shoaling effects), because the bathymetry data used in tsunami numerical simulation may slightly deviate from the realistic bathymetry around the tide gauge.

However, I must comment here that there is a problem with this application of Green's Law. The Green's Law does not include the effects of partial reflection on the continental shelf. Besides, the water depth or tsunami height should be an average value over more than one tsunami wavelength. This requirement may differ from reality. In my study, the tsunami wavelength varies from ~20 to ~100 km. If we want to obtain the tsunami height at coastal tide gauges more accurately, instead of using Green's Law, it is necessary to obtain bathymetry data with higher resolution. In that case, we could use numerical model to calculate the tsunami height strictly, but such calculation requires finer bathymetry data and higher computational cost.

In addition, in the method of virtual station, which will be explained in Section 3.1, the tsunami heights at virtual stations are corrected taking into account the depth difference between the observational stations and interpolation point. The seafloor should be smooth enough to satisfy the assumption that the tsunami propagates as a plane wave and to reduce the effects of partial reflection. However, such correction should still be very cautious because Green's Law requires the average values over a horizontal scale of at least one wavelength.

1.3 Tsunami Data Assimilation Approach

Data assimilation was originally used in meteorology and oceanology. It produces the present state of a system by combining observational data along with an estimated state determined by a forecast model (Bellsky et al., 2014; Ito et al., 2016; Kalnay et al., 2007). Several algorithms are proposed for data assimilation, including Optimal Interpolation, Four-dimensional variational assimilation (4-D Var), Kalman Filter, etc. In this section, I briefly review the methods of 4-D Var and Kalman Filter. Then I explain the Optimal Interpolation method that is applied in this study.

1.3.1 Four-Dimensional Variational Assimilation and Kalman Filter

4-D Var provides a least-squares fit of a model trajectory to the background and the observations in an assimilation window. The propagation of the background x^b from the beginning of the window t_0 to the end of the window t_n using the forecast model is represented as:

$$\underline{x}^b = \underline{F}\{x^b(t_0)\} \quad (1.13)$$

where $\underline{x}^b$ is the four-dimensional representation over the assimilation window (the underscore representing a time-dependent vector). In tsunami science, the forecast model F corresponds to tsunami propagation model. The increment is defined as

$$\delta x(t_0) = x(t_0) - x^b(t_0) \quad (1.14)$$

which minimizes a cost function J (Fairbairn et al., 2014; Tarantola, 1984). In order to improve the conditioning of the cost function, a control variable transform is applied (Lorenc, 2003) so that the increment is defined as a function of v:

$$\delta x(t_0) = Uv \quad (1.15)$$

where U is designed such that $UU^T = B$ gives the climatological background-error covariance matrix. Here B is homogeneous and U consists of a scaling by the square root of the estimated power spectrum and a spectral transform. The increment is propagated using the propagation matrix $\underline{M}$:

$$\delta \underline{x} = \underline{M} \delta x(t_0) \quad (1.16)$$

The cost function obtained is

$$J(v) = \frac{1}{2} v^T v + (\underline{H} \delta \underline{x} - d)^T R^{-1} (\underline{H} \delta \underline{x} - d) \quad (1.17)$$

where R is the observation error covariance matrix, H is the observation matrix that contains only 0 and 1 values. The innovation vector d is defined as:

$$\underline{d} = \underline{y}_n - \underline{H}(\underline{x}^b) \quad (1.18)$$

Kalman Filter is another data assimilation algorithm. It updates an analysis state from a weighted average of a model prediction and observations of the true state (Bellsky et al., 2014). Define the state variable at time step n to be x_n. This corresponds to the true wavefield, which is unknown but can be estimated. At time $n-1$, the best estimate of x_{n-1} is $\hat{x}_{n-1}$. Let F_n represent how the physical process evolves in one time step from $n-1$ to n. Let P_{n-1} be the covariance matrix of $\hat{x}_{n-1}$. The predicted mean of x_n and its covariance matrix are calculated via the following equations, with the

superscript $-$ representing the prior:

$$\widehat{x}_n^- = F_n\widehat{x}_{n-1} \quad (1.19)$$

$$P_n^- = F_n P_{n-1} F_n^T + Q_n \quad (1.20)$$

where Q_n is a covariance matrix of the system noise that follows a normal distribution. The system noise enables the distribution of the state vector x_n to follow observational data. Even though in some situations, the observational data contain phenomena that is difficult to predict, the system noise can still make the simulation follow observations. It can be included in the model of Kalman Filter. However, because it makes the derivation of the adjoint model of 4-D Var difficult, it is hardly included in 4-D Var. To incorporate data, let H be the observation matrix. Then the distribution (assumed to be normal) of the predicted observations from the prior has mean $H\widehat{x}_n^-$ and covariance $HP_n^- H^T$. Let the actual observations be y_n (assumed to be normal) and the covariance of observational error be R_n. To optimally blend the prior estimate and the actual observation, the posterior distribution is calculated as follows:

$$\widehat{x}_{n-1} = \widehat{x}_n^- + K_n(y_n - H\widehat{x}_n^-) \quad (1.21)$$

$$P_n = (I - K_n H)P_n^- \quad (1.22)$$

where the Kalman gain is

$$K_n = P_n^- H^T (HP_n^- H^T + R_n)^{-1} \quad (1.23)$$

Equations (1.19) $-$ (1.23) represent the prediction and update steps in Kalman Filter (Sasaki et al., 2018).

Ensemble Kalman Filter (EnKF) is a stochastic approximation of Kalman Filter, which starts with an initial ensemble with N members. There is an ensemble of predicted means $\widehat{x}_n^- = [\widehat{x}_n^{-(1)}, \dots, \widehat{x}_n^{-(N)}]$ instead of a single estimate at time n. P_n^- is directly calculated as $\mathrm{Cov}(\widehat{x}_n^-)$ instead of applying Equation (1.20). There is also no need to calculate P_n. Instead of storing a full covariance matrix, EnKF presents the same error statistics using an ensemble of model states. N perturbations $v_n^{(i)}$ are added to the observations, which are treated as random variables. For ensemble member i ($i = 1, \dots, N$),

$$\hat{\pmb{x}}_n^{(i)} = \hat{\pmb{x}}_n^{-(i)} + \pmb{K}_n(\pmb{y}_n + \pmb{v}_n^{(i)} - \pmb{H}_n\hat{\pmb{x}}_n^{-(i)}) \quad (1.24)$$

As the ensemble size $N \to \infty$, EnKF converges to Kalman Filter. Hence, it leads to an interpretation of the EnKF as a purely statistical Monte Carlo method. The ensemble of model states evolves in state space with the mean as the best estimate and the spreading of the ensemble as the error variance (Evensen, 2003).

Though 4-D Var and EnKF methods are effective in estimating the present state, they are computationally expensive. For tsunami early warning purpose, it is quite important to adopt a method with low computational cost to save the calculating time.

1.3.2 Optimal Interpolation

The Optimal Interpolation method is usually adopted for tsunami data assimilation because it has a relatively low computational cost (Kalnay et al., 2007; Yang et al., 2019). It is simple and therefore has the advantage of high computational speed for tsunami prediction. It reconstructs the tsunami wavefield from data of offshore tsunami observations by minimizing the total error of all observations (Daley, 1991; Kalnay, 2003). In the numerical simulation, the tsunami wavefield at the n-th time step is represented as a column vector $\pmb{x}_n = (h(n\Delta t, x, y), V_x(n\Delta t, x, y), V_y(n\Delta t, x, y))^T$, where h is tsunami height, and V_x and V_y are velocities in the horizontal directions. The total number of grids is L. The state vector $\pmb{x}_n$ is a $3L \times 1$ column vector, where 1 to L components describe the tsunami height, $L + 1$ to $2L$ components describe the velocity in one direction, and $2L + 1$ to $3L$ components describe the velocity in another direction. The offshore pressure gauges directly provide data of tsunami height, but the velocity components of $\pmb{x}_n$ are reconstructed during the assimilation process.

The data assimilation consists of two steps: a propagation step and an assimilation step. The propagation step is expressed as

$$\pmb{x}_n^f = \pmb{F}\pmb{x}_{n-1}^a \quad (1.25)$$

where $\pmb{F}$ is the tsunami propagation matrix ($3L \times 3L$) such that given the tsunami wavefield at time $t = (n - 1)\Delta t$, the forecasted tsunami wavefield at the next time step $t = n\Delta t$ can be calculated. The propagation matrix $\pmb{F}$ corresponds to the tsunami

propagation model. Previously, the LLW tsunami model was adopted (Gusman et al., 2016b; Maeda et al., 2015). In the assimilation step,

$$x_n^a = x_n^f + W(y_n - Hx_n^f) \quad (1.26)$$

where the observation matrix H ($m \times 3L$) is a sparse linear matrix, which contains only 0 and 1 values. It extracts the tsunami height at m stations from the forecasted wavefield, and the residual is calculated by comparing it with the real-time observed tsunami height y_n, a column vector. Consequently, the residual is multiplied by the weight matrix W ($3L \times m$) to bring the assimilated tsunami wavefield closer to the observed wavefield. The weight matrix W contains the assimilation responses from all stations. Each column represents the assimilation response of one individual station. The weight matrix W is a key controlling factor for the quality of assimilation. It is assumed that the observation matrix H is linear and that the weight matrix W is independent of time. The calculation of weight matrix is based on the correlation distance of the simulation and observation error distributions and the relative magnitude of the observation to simulation errors. We can calculate the weight matrix by minimizing the covariance matrix of $< \varepsilon^a \varepsilon^{aT} >=< (x_n^a - x_n^{True})(x_n^a - x_n^{True})^T >$ as a solution of the linear system:

$$W(R + HP^f H^T) = P^f H^T \quad (1.27)$$

where $P^f =< \varepsilon^f \varepsilon^{fT} >$ and $R =< \varepsilon^O \varepsilon^{OT} >$ are the covariance matrices of the forward numerical simulation and observations, respectively. Here ε^f and ε^O are the error distributions associated with forward numerical simulations and observations, respectively, which are assumed to be described by Gaussian functions, and ε^{fT} and ε^{OT} are their corresponding transposes (Kalnay, 2003). It is assumed that the computational errors are spatially homogeneous on numerical grids, and that the observation errors are not correlated with stations as observations are made independently. These assumptions simplify the matrix R into a diagonal matrix whose components are the standard deviations of the observation errors at each station (Maeda et al., 2015). A simplified form of Equation (1.27) can be obtained:

$$\sum_{j=1}^{m} w_{gj}(\mu_{ij}^b + \delta_{ij}) = \mu_{gi}^b \quad (1.28)$$

where w_{gj} is components of the weight matrix W, μ_{ij}^b is the correlation function of errors in numerical forecasts between two observational points (or computational grids).

A Gaussian-distributed covariance is assumed for the correlation function of errors, with a characteristic distance of 20 km (Maeda et al., 2015; Wang et al., 2019a).

By iterating between the propagation and assimilation steps, the tsunami wavefield is gradually assimilated, and the forecasted waveforms at any location inside the modeling domain can be obtained. The point where the tsunami forecast is made is defined as the Point of Interest (PoI). During or after the assimilation process, the forecasted tsunami waveform at a PoI can be obtained based on the forward tsunami simulation from Equation (1.25) using the present assimilated tsunami wavefield as the initial condition.

1.4 Network of Offshore Bottom Pressure Gauges

Offshore Bottom Pressure Gauges (OBPGs) are installed at water depths ranging from hundreds to thousands of meters to detect tsunamis generated by earthquakes or other causes (Beltrami, 2008; Bernard and Meinig, 2011; Sheehan et al., 2015). They measure the ocean bottom pressure and convert water pressure records into sea surface height (Shinohara et al., 2016). Some data are sent to land in real time, either by satellite or cable. In the United States, the DART system provides an accurate measurement of the tsunami height in the Pacific Ocean through satellite transmission. Offshore of Japan and Canada, there are local dense real-time tsunami observation networks composed of OBPGs, and their data are sent through submarine cables (Barnes et al., 2008; Kaneda, 2010).

Off the Pacific coast of Japan, there are two main dense offshore observation networks. In northern Japan, the Seafloor Observation Network for Earthquakes and Tsunamis (S-net) is installed around the Japan Trench (Kanazawa, 2013; Yamamoto et al., 2016b). It consists of six groups, with up to 150 stations linked by fiber optic cables (Aoi et al., 2019). The observatory spacing is approximately 30 km in the east–west direction, which is almost perpendicular to the Japan Trench and approximately 50–60 km in the north–south direction parallel to the trench. The S-net stations are equipped with OBPGs and Ocean Bottom Seismometers (OBSs). There are a pair of identical pressure gauge sensors (HP1WP, HP2WP) at each station to ensure redundancy for tsunami detection

(Kanazawa, 2013; Kubota et al., 2020a). The S-net pressure gauges can detect tsunamis with amplitudes of less than 1 cm (Kubota et al., 2020b).

In the Nankai region, the Dense Oceanfloor Network System for Earthquakes and Tsunamis (DONET) is a system for real-time continuous observations and monitoring of earthquakes and tsunamis using an observation equipment network installed on the ocean floor (Kawaguchi et al., 2008). It was established by the Japan Agency for Marine-Earth Science and Technology (JAMSTEC) and is now operated by the National Research Institute for Earth Science and Disaster Resilience (NIED). There are two phases of the DONET project. DONET1 is located in the Tonankai earthquake hypocenter area, including the offshore of Kii Peninsula and Kumano-Nada. DONET2 is responsible for the Nankai earthquake hypocenter area, including offshore of Shiono Cape and Muroto Cape. The entire system consists of 12 nodes and up to 51 observation points. High-precision observation equipment, including OBPGs and OBSs, are deployed at each observation point at intervals of 15–20 km (Kaneda, 2010). The OBPGs of DONET can also detect micro tsunamis with amplitudes less than 1 cm (Wang et al., 2019b).
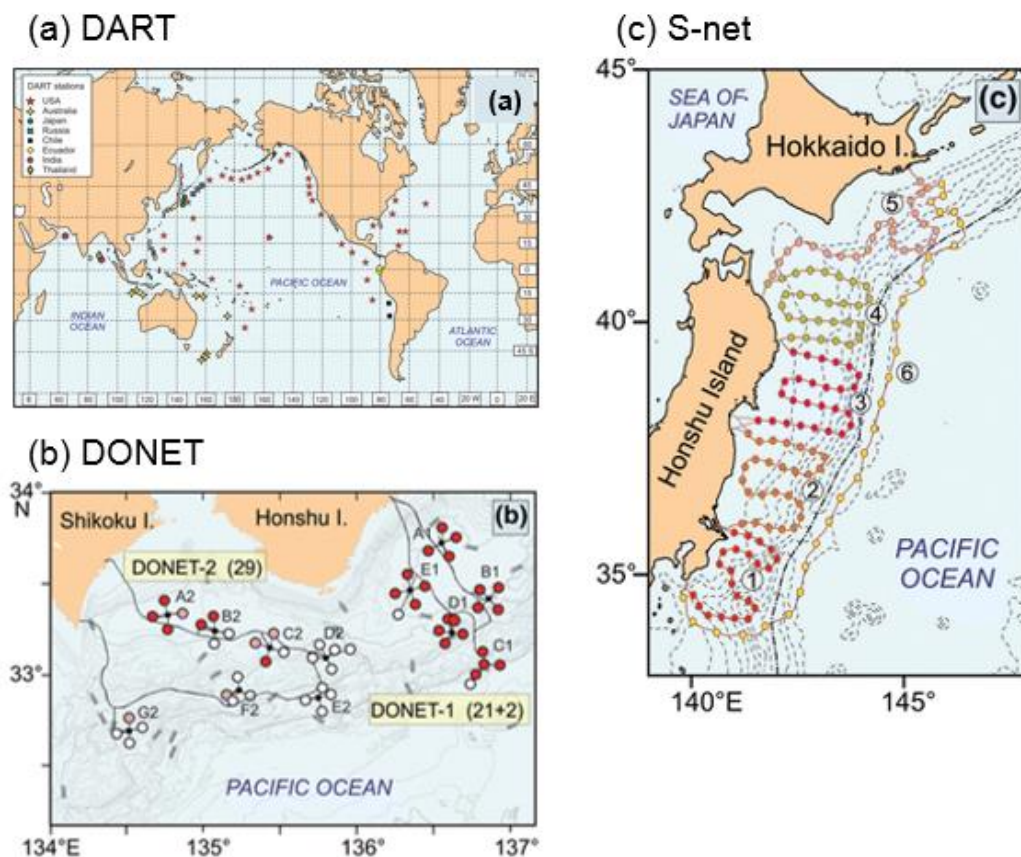
Figure 1.2. Offshore observation network composed of OBPGs: (a) DART; (b) DONET; (c) S-net.

The figures are cited from (Rabinovich and Eblé, 2019).

Besides these two dense observation networks, there are also other distributed OBPGs offshore of Japan. There are two stations of the Deep Sea Floor Observatory (DSFO) off Muroto Cape: PG1 and PG2. JAMSTEC installed them for long-term continuous observations of the phenomena of seafloor changes and seafloor environmental changes. They recorded the tsunamis of the 2015 Torishima earthquake (Wang et al., 2019b). Moreover, Earthquake Research Institute, the University of Tokyo, installed seismic and tsunami observation systems using seafloor optical fibers off the Sanriku region in 1996 (Shinohara et al., 2016). These OBPGs recorded the tsunami events of the 2011 Tohoku earthquake and the 2016 Fukushima earthquake (Gusman et al., 2017; Tsushima et al., 2011). The installment of OBPGs enables us to detect tsunamis in real-time operation and perform tsunami data assimilation. Hence, the OBPGs are very important hardware for tsunami early warning.

1.5 Real-time Tsunami Detection

Real-time tsunami detection plays an important role in tsunami early warning. For the tsunami data assimilation approach, forecasts can be updated continuously as more offshore tsunami data are recorded in real time during the assimilation process. Therefore, real-time tsunami detection is a prerequisite of the tsunami data assimilation approach. Additionally, even for the traditional early warning methods based on seismic observations, the detection of offshore tsunamis is still essential as it helps confirm the actual generation and propagation of a tsunami (Chierici et al., 2017).

The hardware aspect for real-time tsunami detection has advanced significantly in recent years because of the wide installment of OBPGs. As for the software, several algorithms have been proposed for real-time tsunami detection. Although the records of OBPGs are less affected by the complex geometries of bays and harbors, they are usually

influenced by tide and ocean currents (Rabinovich, 1997; Saito, 2019; Sato et al., 2017). Seismic waves and background noise also influence tsunami records. In retroactive studies, the tidal components can be easily removed by subtracting a theoretical tide model, and the tsunami signals can be easily extracted by digital filtering (Heidarzadeh and Satake, 2013; Tsushima et al., 2009). However, traditional digital filters used in such studies are not applicable to real-time operation. The acausal filter that is usually used in retroactive studies requires future data to filter past data, and the causal filter may change the time of the tsunami peak by the group delay of the filter (Boore and Akkar, 2003). Therefore, it is necessary to use a real-time data processing algorithm for automatic tsunami detection. A reliable algorithm for OBPGs should accurately identify the tsunami signals and characterize the tsunami amplitude at a low computational cost and short detection delay.

NOAA developed the tsunami detection algorithm for DART (Mofjeld, 1997). It first estimates the amplitudes of the pressure fluctuations within the tsunami frequency band and then tests these amplitudes against a 3 cm equivalent sea-level fluctuation threshold value. In NOAA's algorithm, the amplitudes are computed by subtracting predicted pressures from the observations, where the predictions are made using a cubic polynomial fitted to bottom pressure observations over the past 3 h (Mofjeld, 1997). Additionally, Beltrami (2008) used an Artificial Neural Network (ANN) to design a tsunami detection algorithm. The ANN algorithm is based on a two adaptive-weight layer network characterized by logistic sigmoid and linear activation functions. The adaptive weights are obtained from the network's supervised learning. It has a better performance than NOAA's algorithm when the tidal range at the location of interest is higher and the background sea noise is lower (Beltrami, 2008; Beltrami, 2011). Both algorithms are effective in predicting and filtering out tidal components. However, as the prediction of tides depends on the observation average, an actual propagating tsunami will inevitably influence the predictions. This can affect their performance in characterizing the tsunami amplitude (Beltrami, 2011). Moreover, pure predictive algorithms usually fail in reducing the high-frequency components of environmental noise (Chierici et al., 2017).

1.6 Objectives

In this thesis, I mainly focus on four aspects. First, I propose Green's Function-based Tsunami Data Assimilation (GFTDA) to improve the assimilation speed. Then, I adopt virtual stations to perform data assimilation for regions without a dense observation network. Moreover, I propose a new real-time tsunami detection algorithm based on Ensemble Empirical Mode Decomposition (EEMD). Finally, I combine the tsunami data assimilation approach with the real-time tsunami detection algorithm and apply this to S-net pressure gauge records.

The previous tsunami data assimilation approach has a relatively high computational cost because it is necessary to run numerical simulations to obtain the tsunami wavefield. However, the tsunami wavefield of the entire modeling domain needs to be repeatedly calculated during the assimilation process (Gusman et al., 2016b), which requires substantial calculation time. Moreover, because of the high computational cost, the adoption of other more complex but realistic models such as the DSP tsunami model is not practical. It may limit the accuracy of tsunami forecasting in some cases where the tsunami dispersive characteristics are evident (Saito et al., 2010). In such cases, the long-wave approximation breaks down, and the previous data assimilation approach using the LLW model may not be able to forecast tsunamis with high accuracy. For early warning, the waveforms at the PoIs should be predicted quickly and accurately. In Chapter 2, I propose GFTDA for solving this problem.

Additionally, the previous tsunami data assimilation approach requires enough offshore observed tsunami data to reconstruct the tsunami wavefield (Gusman et al., 2016b; Mulia et al., 2017b). The neighboring stations of an observation network should not be far from each other. Past studies proposed an empirical distance criterion of 50 km to perform tsunami data assimilation (Wang et al., 2019a; Yang et al., 2019). However, an offshore network usually requires a large investment, and many regions with significant tsunami hazards may only afford a few OBPGs rather than a dense network. For example, the Indian Ocean suffered from the 2004 Sumatra−Andaman earthquake and tsunami (Fujii and Satake, 2007; Nalbant et al., 2005). However, unlike the Pacific Ocean, there are only sparse tsunami observations in the Indian Ocean. Only six OBPGs are currently available for tsunami detection in the Bay of Bengal (north Indian Ocean). In Chapter 3, I introduce virtual stations between two real observational stations to complement sparse observations.

To put tsunami data assimilation into practice, offshore tsunami data should be provided as the input in real time. In Chapter 4, I describe a novel method of real-time tsunami detection. Unlike the previous predictive algorithms, the method detects the tsunami arrival based on the EEMD without tidal prediction. The tsunami signals are automatically extracted from the original time series of OBPG records. I apply the method to the records of the cabled OBPGs off the Tohoku coast and examine its performance for three earthquakes with different sizes to confirm the validity of the algorithm: the 2016 Fukushima earthquake (M 7.4), the 2011 Tohoku earthquake (M 9.0), and the 1998 off Sanriku earthquake (M 6.4).

In Chapter 5, I combine the tsunami data assimilation approach with the real-time tsunami detection algorithm. Using the 2016 Fukushima earthquake as an example, I forecast the tsunami waveforms at tide gauges in the Sanriku region by assimilating the S-net pressure gauge records. Then, I compare the forecasted waveforms with the real observations at tide gauges to validate the method. Although this study is conducted retroactively, I use the real-time tsunami detection algorithm based on EEMD to process the raw records as if it is operating in real time. In Chapter 6, I discuss the practical implementation and future improvements of the tsunami early warning system.

Chapter 2 Green's Function-based Tsunami Data Assimilation (GFTDA)

2.1 Principles of GFTDA

GFTDA is proposed to improve the assimilation speed. Green's functions have been applied to tsunami waveform inversion for fault estimation (Satake, 1989) and TRI for far-field tsunami forecasting (Hossen et al., 2015; Hossen et al., 2017; Korolev, 2011). Here I adopt Green's functions in data assimilation for tsunami forecasting. In the Optimal Interpolation algorithm (Maeda et al., 2015), if the residual between the observed and forecasted tsunami height is non-zero, the station will have an assimilation response, which can be near Gaussian shape if the distance between stations is larger than the characteristic distance of matrix $\boldsymbol{P}^f$ and $\boldsymbol{R}$. I then define the residual vector as $\boldsymbol{x}_n^r \equiv \boldsymbol{W}(\boldsymbol{y}_n - \boldsymbol{H}\boldsymbol{x}_n^f)$, which represents the residual of each station at time $t = n\Delta t$. This can be rewritten as a linear combination of the unit column vector $\boldsymbol{e}_i^T$ multiplied by the corresponding residual of tsunami height at the i-th station r^i:

$$\boldsymbol{x}_n^r = \boldsymbol{W} \sum_i r_n^i \boldsymbol{e}_i^T \quad (2.1)$$

Mathematically, the assimilation response of the i-th station can be represented as $r_n^i \boldsymbol{W}\boldsymbol{e}_i^T$. The assimilation response will propagate across the region under consideration, following the tsunami propagation model. This will result in changes in the tsunami height and velocity at other grid points. When the tsunami wave propagation model is linear, the assimilation response of different steps and different stations can be superposed. I define the waveform at the j-th grid point resulting from the propagation of the i-th station's assimilation response as the Green's function $G_{i,j}$. In Figure 2.1, plots of Green's functions $G_{1,s}$, $G_{2,s}$, and $G_{3,s}$ represent the waveforms at PoIs resulting from the propagation of the assimilation responses of the first, second, and third stations, respectively.

Therefore, by using GFTDA, it is not necessary to simulate the tsunami propagation of the whole region during the assimilation process. Instead, the waveforms

at PoIs are synthesized by Green's functions directly. The propagation of the assimilation response can be calculated with any linear tsunami propagation model, not only the LLW model. The preparation of Green's functions is time-consuming but can be done in advance of the assimilation process.

Storing the output of Green's functions requires computer memory, but the amount of memory required can be reduced by limiting the number of PoIs. For a given region with a tsunami observation network and PoIs, the number of Green's functions to be calculated is $m * (N_s + m)$, where m is the number of observational stations, and N_s is the total number of PoIs. The PoIs should include near-shore points because forecasting the tsunami height and velocity at these points is of the greatest importance to residents living nearby. Conversely, forecasting the tsunami at points that are far away from the coast is not useful for coastal communities. Therefore, the computation of Green's functions at points far away from the coast can be omitted to obtain a manageable number of Green's functions. The Green's functions between observational stations themselves are also important because their waveforms should be updated continuously during the assimilation process.
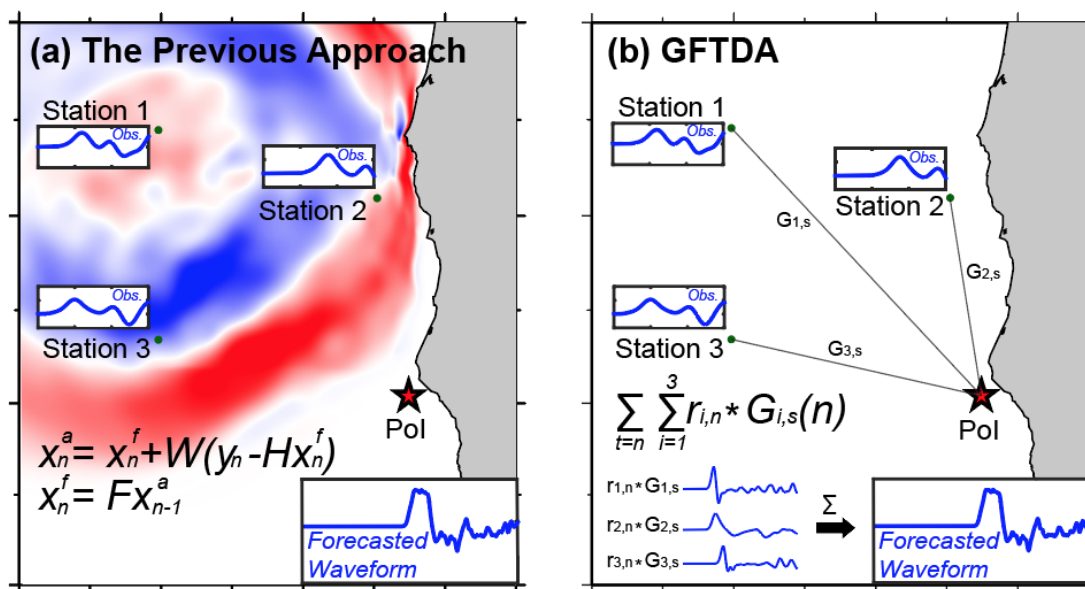

Figure 2.1. Illustration of the previous data assimilation approach and GFTDA.

For simplification, I only use three offshore stations to forecast the waveform at the PoI (red star). In the previous approach, the observed tsunami height of stations is assimilated,

and the wavefield of the entire modeling domain is computed. Then, the forecasted waveform at the PoI can be recorded in the wavefield. In GFTDA, the Green's functions are computed in advance, representing the waveform at the PoI resulting from the propagation of each station's assimilation response. The forecasted waveform is directly synthesized with Green's functions by matrix manipulation. Both approaches give the same result for the forecasted waveform.

2.2 Assimilation Process and Mathematical Equivalence

The assimilation process of GFTDA is demonstrated in Figure 2.2. The Green's functions are calculated by computers or supercomputers in advance. When a tsunami arrives at a region of interest, similar to the previous data assimilation approach, the residual between the real-time observed tsunami height and the forecasted tsunami height is calculated (initial condition is zero) after necessary data processing such as filtering. Then, the residual of each station is multiplied with the corresponding Green's functions to forecast the tsunami heights at the observational stations and PoIs for this time step. In the next time step, the observation data at each station are updated, and the residuals are calculated and multiplied with the Green's functions corresponding to that time step. These steps are repeated continuously during the assimilation process. The forecast is made using the tsunami data of observational stations and pre-calculated Green's functions. As the tsunami propagates, more observational stations from the network detect the tsunami height, and more Green's functions are involved in the data assimilation. Consequently, the accuracy of the forecasted waveforms gradually increases with time.

The forecasted waveforms at PoIs are directly obtained after the assimilation process. The tsunami warning level can be decided according to the forecasted waveforms before the tsunami reaches the shore.

If linearity is assumed, GFTDA is mathematically equivalent to the previous data assimilation approach. This can be proved as follows.

$$x_n^a = x_n^f + W\left(y_n - H x_n^f\right) = x_n^f + x_n^r \ (2.2)$$

31

$$x_n^f + x_n^r = Fx_{n-1}^a + x_n^r = F\left(x_{n-1}^f + x_{n-1}^r\right) + x_n^r = F\left(F(x_{n-2}^f + x_{n-2}^r) + x_{n-1}^r\right) +$$
$$x_n^r = \cdots = \sum_{t=1}^{n} F^{n-t} x_t^r \quad (2.3)$$

$$\sum_{t=1}^{n} F^{n-t} x_t^r = \sum_{t=1}^{n} F^{n-t} \left(\sum_{i=1}^{m} r_t^i W e_i^T\right) = \sum_{i=1}^{m} \left(\sum_{t=1}^{n} r_t^i F^{n-t} W e_i^T\right) \quad (2.4)$$

As defined earlier, Equation (2.2) describes the assimilation step of the previous approach. In a linear system, the propagation of the tsunami wavefield can be written as Equation (2.3). The final assimilated tsunami waveforms are related to the residuals of each station at each time step. The residual vector is then rewritten as Equation (2.4), where We_i^T is the unit assimilation response of the i-th station. Here, $F^{n-t} We_i^T$ represents the propagation of the assimilation response, i.e., the Green's functions between the i-th station and all other grid points. Therefore, I only need to consider the grid points of the stations and PoIs during the assimilation process.
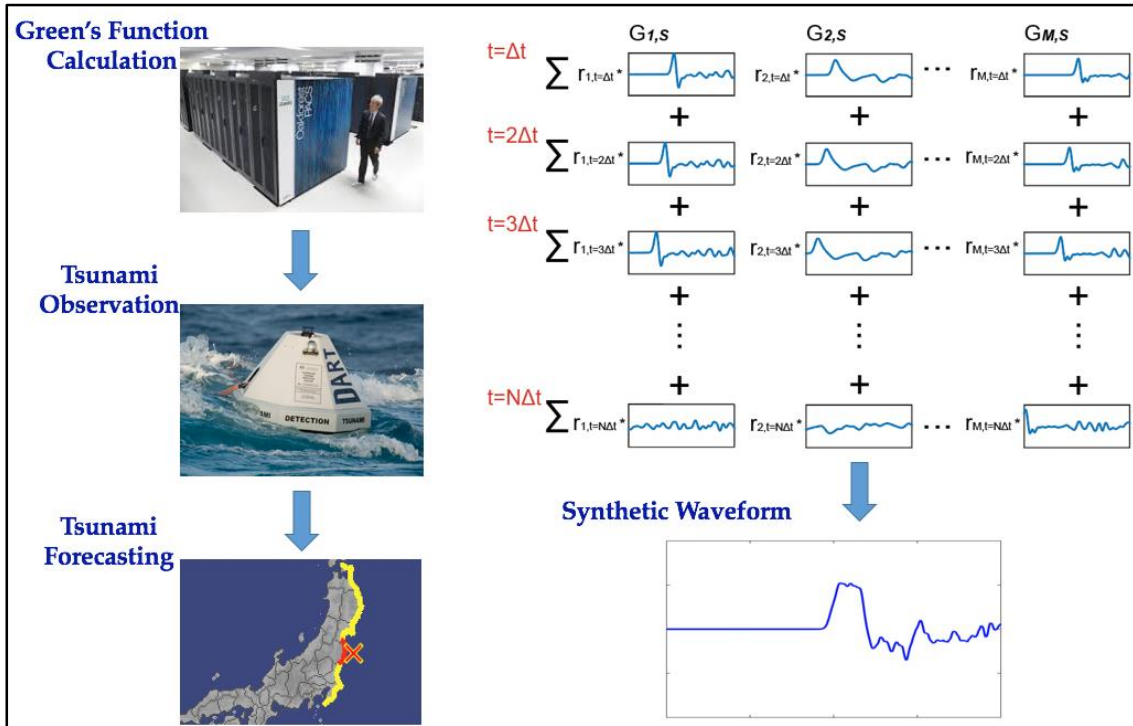

Figure 2.2. Assimilation process of GFTDA.

Step 1: The Green's functions are calculated by a computer and stored in advance. Step 2: Offshore stations observe the tsunami height. The residual between the observed and forecasted height of each station is calculated. The assimilated waveforms at stations and

PoIs are repeatedly synthesized with Green's functions. Step 3: The forecasted waveforms at PoIs are obtained after the assimilation process.

2.3 Validation Test—2012 Haida Gwaii Earthquake

The 2012 Haida Gwaii earthquake (M 7.8) occurred at 52.622 °N, 132.103 °W, off the coast of Haida Gwaii, British Columbia, Canada, on October 28, 2012. The seafloor uplift generated a tsunami with a run-up height greater than 3 m. Fortunately, no casualties or major injuries were recorded from the earthquake and tsunami, likely because of the sparse population of the region. The tsunami generated by the earthquake was recorded in the near field at tide gauges and an OBS array with bottom pressure gauges in the Cascadia subduction zone during the Cascadia Initiative Community Experiment (Gusman et al., 2016a; 2016b; Sheehan et al., 2015). Although this network did not send data in real time, the recorded tsunami data can be used to perform tsunami data assimilation retroactively.

2.3.1 Observed Tsunami Data

I use the tsunami observation data for the 2012 Haida Gwaii earthquake recorded at 26 stations of the Cascadia OBS array (Sheehan et al., 2015). Gusman et al. (2016b) applied the previous data assimilation approach to the same case study, while I apply GFTDA to forecast the tsunami retroactively. A band-pass filter with cut-off frequencies of 0.0002 and 0.005 Hz (5000 and 200 s) was applied to the raw data to remove the influence of seismic waves and ocean tides. The pressure data were converted to water height, and the instrument response was removed from the data. Details on the processing of seafloor pressure gauge data of the Cascadia OBS array were provided in previous studies (Gusman et al., 2016a; 2016b; Sheehan et al., 2015). I divide the tsunami observational stations into two groups: a group for data assimilation of the 15 offshore stations and a group for validation of 11 near-shore stations (PoIs). The locations of these stations are shown in Figure 2.3a.
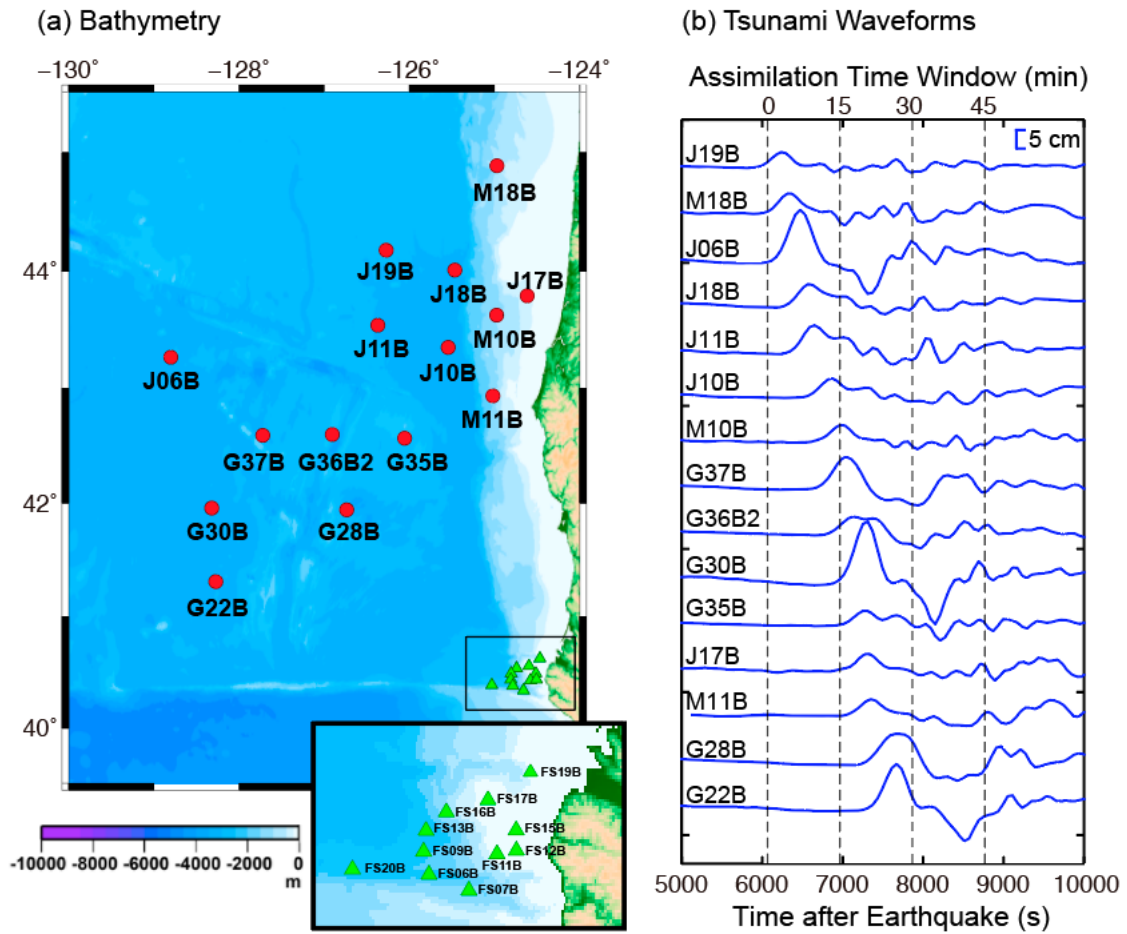
Figure 2.3. Bathymetry and tsunami waveforms of the observational stations in the Cascadia Initiative Community Experiment.

(a) The 15 offshore stations that provide tsunami observation for data assimilation are marked by red circles. Other 11 near-shore stations (PoIs) are indicated with green triangles, and I compare their data with the forecasted waveforms calculated by GFTDA. (b) The observed waveforms at the 15 offshore stations are plotted on the right. The data assimilation starts when the tsunami arrives at the first station.

2.3.2 Assimilation Setting

I synthesize the tsunami waveforms at the stations and PoIs by GFTDA and compare the forecasted waveforms with the observed ones. To compute Green's functions, I use the *JAGURS* LLW model (Baba et al., 2015).

The initial condition for computing Green's function $G_{i,j}$ is the assimilation response of the i-th station calculated from $\boldsymbol{W}$; the velocities in two directions are zero. The grid size is 0.5 arcmin, and the bathymetry grid data is derived from the General Bathymetric Chart of the Ocean released in 2014 (GEBCO_2014) (Weatherall et al., 2015).

The target area is 35 °N–56 °N, 122 °W–136 °W. The time step of Green's functions is 1 s, the same as that of the processed observation data. There are 15 offshore stations for data assimilation and 11 PoIs for waveform comparison. Therefore, the number of Green's functions computed is $15 \times (11 + 15) = 390$, which does not require a large amount of computer memory.

I set the earthquake origin time as $t = 0$. The tsunami generated by the earthquake propagates across the Cascadia subduction zone, arriving at the first observational station (J19B) at $t = 6066$ s (Figure 2.3b). Then, the assimilation process begins. The *assimilation time window* is defined as the period during which the observed data are used for assimilation. The waveforms at PoIs are synthesized with the observation data and the Green's functions during the assimilation time window. For example, if the assimilation time window is set to 30 min (1800 s), the assimilation process will end at $t = 7866$ s. The tsunami height at stations is forecasted after $t = 7866$ s.

2.3.3 Results

I apply GFTDA with different time windows from 2 to 50 min, at intervals of 2 min. The assimilation process takes less than 10 s on the computer system of the Earthquake Information Center (EIC) at the University of Tokyo. This computation speed is much faster than that of the previous data assimilation approach that takes approximately 1,000 s. Hence, GFTDA reduces 99% of the calculation time during the assimilation process.

Figure 2.4 presents the comparison between the observed and assimilated waveforms at 11 PoIs for an assimilation time window of 24 min. The assimilated waveforms generally match the observations. As shown in Figure 2.3, the first tsunami

peak has already passed most stations at the end of the assimilation time window. That is, the data from most of the stations have been assimilated. For most PoIs, the tsunami height is forecasted accurately, but for stations FS06B and FS09B, the maximum heights of the forecasted waveforms are smaller than the observations, and for station FS19B, the maximum height of the forecasted waveform is slightly higher. Nevertheless, the overestimation or underestimation is quite small, so it is fair to say that GFTDA provides accurate tsunami forecasting at the near-shore PoIs.
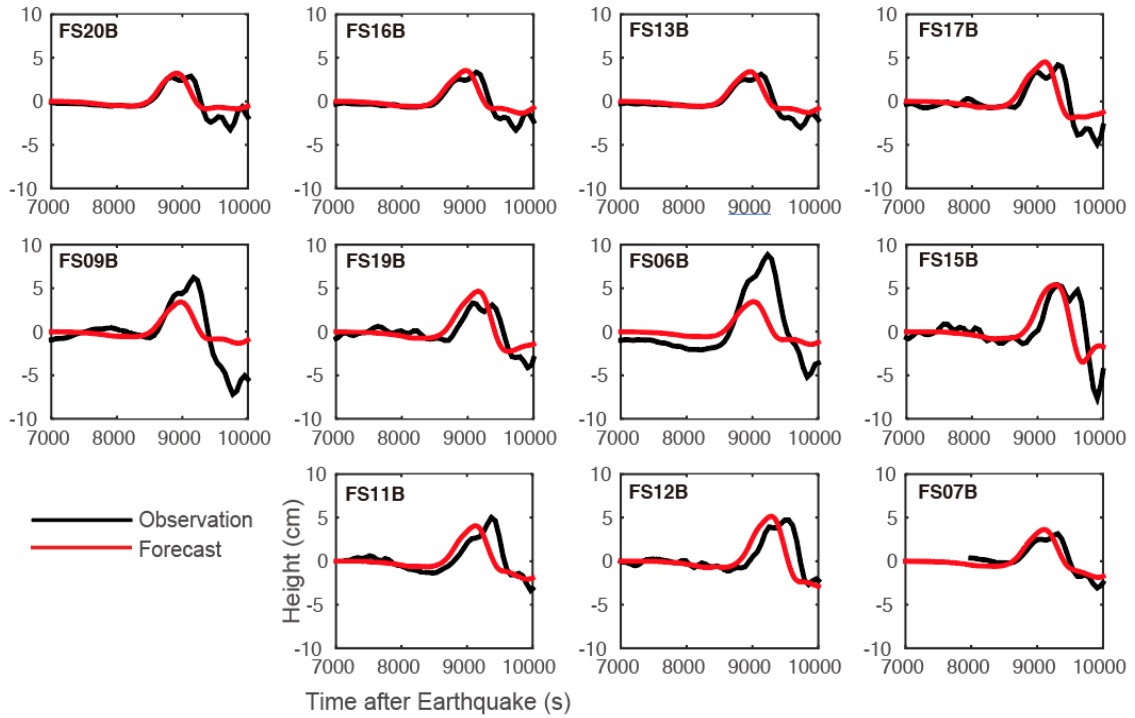
Figure 2.4. Comparison of the observed and forecasted waveforms at 11 PoIs with an assimilation time window of 24 min.

To quantitatively evaluate the performance of GFTDA, I calculate the forecast accuracy (Gusman et al., 2016a) and score (Tsushima et al., 2011). The forecast accuracy is calculated by the geometric mean ratio (K) of the observed (H_o^{arr}) and forecasted (H_f^{arr}) first-peak amplitude (Aida, 1978):

$$\log(K) = \frac{1}{N_s}\sum \log\left(\frac{H_o^{\mathrm{arr}}}{H_f^{\mathrm{arr}}}\right) \quad (2.5)$$

$$Accuracy(\%) = \frac{1}{K} \times 100\% \ (K \geq 1) \ or \ K \times 100\% \ (K < 1) \ (2.6)$$

where N_s is the number of PoIs, and the score is calculated as

$$Score(\%) = \left[1 - \frac{\Sigma(H_f^{arr} - H_o^{arr})^2}{\Sigma(H_o^{arr})^2} \right] \times 100\% \ (2.7)$$
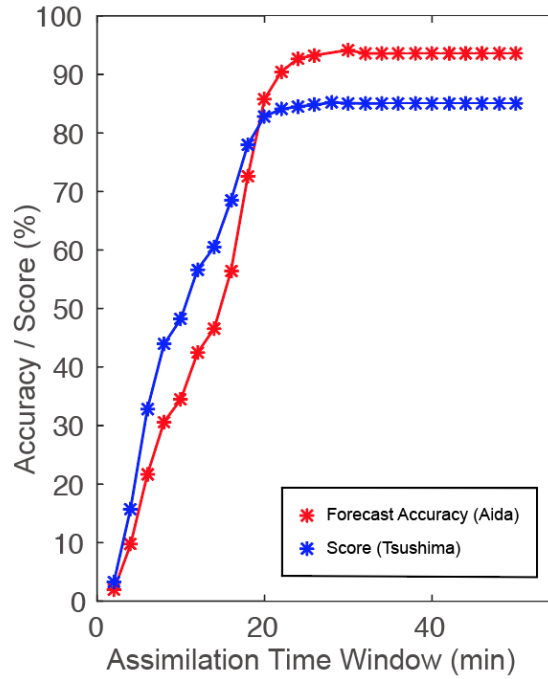
Figure 2.5. Forecast accuracy and score versus different assimilation time windows.

The accuracy (Aida's method) and score (Tsushima's method) show a sharp increase initially, and tend to saturate after an assimilation time window of 30 min.

Figure 2.5 shows the forecast accuracy and score versus different assimilation time windows. Initially, the forecast accuracy and score are both very low. This is because the data length used for the assimilation is too short to provide accurate forecasting, and the first tsunami peak has not passed many stations. Then, as the time window becomes longer, there is a sharp increase in the forecast accuracy and score. For the forecast made with a time window of 24 min, a high accuracy of over 90% and a high score of over 80% are achieved. After 24 min, the increase in forecast accuracy and score becomes smaller,

and after 30 min there is almost no improvement. As the first tsunami peak has already passed all 15 offshore observational stations at this point, the forecast accuracy and score are saturated.

The accuracy of tsunami forecasting depends on the length of the assimilation time window. If the assimilation time window is too short, the tsunami can be forecasted quickly, but the accuracy is low, and the confidence of early warning remains doubtful. If the time window is too long, although the tsunami forecasting will be more accurate, people living along the coast may not have enough time to evacuate. Therefore, to produce reliable and useful tsunami forecasting, we need to select a suitable assimilation time window to strike a balance between forecast accuracy and computation time. In this case study, a time window of 24 min is a practical choice. After a reliable tsunami warning, the residents along the coast would have approximately 25 min to evacuate. Moreover, it is important to repeat the forecast over time to improve accuracy. Because GFTDA is able to forecast the tsunami instantaneously, it is very effective for repetitive forecast.

For the 2012 Haida Gwaii tsunami, the score proposed by Tsushima et al. (2011) tends to evaluate the performance better than the forecast accuracy (Aida's criterion). I use the sum of the logarithm as in Equation (2.5) to calculate the forecast accuracy. When some observations are higher than the forecasts, and some other observations are lower, the positive and negative logarithm terms cancel out. Hence, $\log(K)$ can still be very small, and the accuracy can be high even though the residuals between the forecasts and observations at PoIs are large. In Figure 2.5, I remove a point (28 min) where the calculated forecast accuracy is abnormally high. Conversely, the score calculation includes the squared error (Equation 2.7), which means that the errors of different PoIs cannot be canceled out. Here we compare two criteria to evaluate the performance thoroughly. As a general practice of accuracy analysis, it is important to select an appropriate criterion to evaluate the performance fairly.

2.4 Adoption of Linear Dispersive Model—2004 off the Kii Peninsula Earthquake

This section applies GFTDA to synthetic tsunami data of the 2004 off the Kii Peninsula Earthquake. I calculate Green's functions with both the LLW and linear DSP

models and assimilate the tsunami data with both models.

2.4.1 2004 off the Kii Peninsula Earthquake

Great earthquakes have occurred along the fault between the Philippine Sea Plate and Eurasian Plate along the Nankai Trough (Furumura et al., 2011; Saito et al., 2010). Such great earthquakes in recent years include the 1944 Tonankai (M 8.0) and the 1946 Nankai (M 7.9) earthquakes. Additionally, a large (M 7.4) earthquake occurred off the Kii Peninsula within the subducting Philippine Sea Plate in 2004.

The 2004 off the Kii Peninsula earthquake occurred on September 5, 2004. It was an intraplate event occurring in the outer rise of the Philippine Sea Plate. The main shock at 14:57 UTC followed the foreshock at 10:07 UTC. The tsunami generated by the mainshock showed evident dispersive characteristics (Saito et al., 2010). One of the important features of tsunami generation is that dispersive waves have a strong directional dependence with respect to the fault strike (Saito et al., 2010). The tsunami dispersion is strong in the short-axis direction of the source fault where the tsunami wavelength is short, and it is also strong in the direction along the deep sea where the tsunami travels fast.

GFTDA enables us to use a more complicated tsunami propagation model for data assimilation, as long as the model is linear. I adopt the linear DSP model for data assimilation and compare it with the LLW model.

I compute Green's functions between observational stations and PoIs with both models. The propagation matrix $\boldsymbol{F}$ is more complicated when using the linear DSP model. Although the computation of Green's functions is quite time-consuming, this step is done in advance. This does not affect the efficiency of the data assimilation process. Then, during the assimilation process, the forecasted tsunami waveforms can be directly synthesized by multiplying the residual with corresponding Green's functions.

2.4.2 Synthetic Tsunami Data

As DONET1 was completed in 2011 and DONET2 started operation in 2015, real tsunami records for the 2004 off the Kii Peninsula earthquake were not available. To assess the ability of the data assimilation approach, I use synthetic tsunami waveforms

from the 2004 earthquake source model.

First, I use a source model similar to the mainshock of the 2004 off the Kii Peninsula earthquake for simulation. The fault parameters are set according to the unpublished results of Yamanaka (data available at http://wwweic.eri.u-tokyo.ac.jp/sanchu/Seismo_Note/2004/EIC153.html) by analysis of the teleseismic body waves. The epicenter is 137.142 °E, 33.143 °N, and the depth is 10.0 km. The fault direction is perpendicular to the trough axis with a strike of 135 °. The dip angle is 40 ° and the rake angle is 123 °. The length and width of the rectangular fault are 50.0 and 30.0 km, respectively. The fault slip is 6.5 m, which is consistent with the magnitude of the mainshock (M 7.4).

I use Okada's model to calculate the initial sea surface elevation in an elastic half-space (Okada, 1985), as the initial condition for numerical simulation. Then, I use the *JAGURS* model to calculate the tsunami propagation. A total reflection is assumed at the coastline. Sea-bottom friction is not considered in my simulation. To make the tsunami propagation closer to the real situation, I apply the linear DSP model. The synthetic waveforms at observational stations and PoIs are recorded during the numerical simulation.

For observation points, DONET has 12 nodes, and each node is linked with several observation points. To build an evenly-distributed observation network for data assimilation, I take one point for each node except for Node C, for which I take two points. Additionally, I use the submarine stations PG1 and PG2 belonging to the DSFO off Muroto Cape. In total, 15 observational stations are used for data assimilation (Figure 2.6).

Additionally, nine PoIs are selected near population centers on Shikoku and Kyushu Islands (Figure 2.6), under the potential threat of tsunami disasters. They are used to compare simulated waveforms and waveforms predicted by data assimilation.
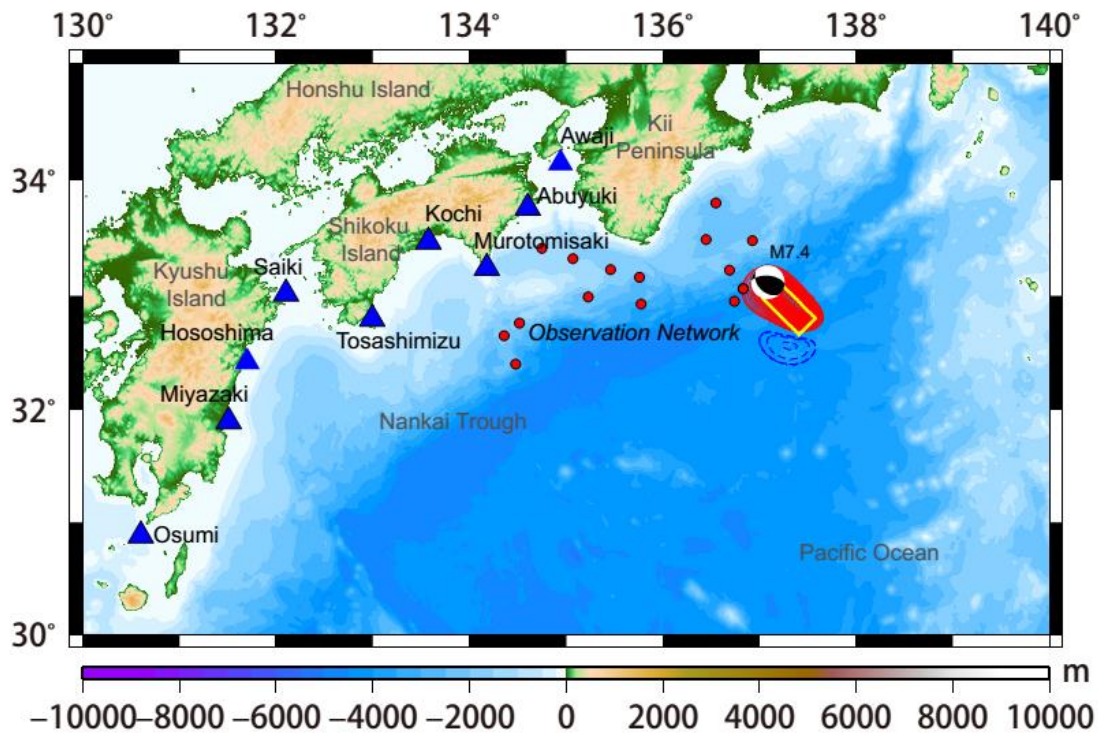
Figure 2.6. Illustration of the observation network and near-shore PoIs.

The observation network contains 13 DONET stations, PG1, and PG2, which are marked with red circles. Nine PoIs near Shikoku Island and Kyushu Island are marked with blue triangles. The focal mechanism and source model are plotted according to the unpublished results of Yamanaka (2004). Red and blue curves are contours of uplift and subsidence, respectively. The contour interval is 0.05 m. The long axis (length) is in the NW-SE direction, whereas the short axis (width) is in the NE-SW direction.

2.4.3 Assimilation Setting

The bathymetry and topography dataset for computing Green's functions are derived from GEBCO_2014. The finite difference method with an implicit scheme is employed for numerical simulations. I use a grid spacing of 30 arcsec and a time step of 1 s. The target area 30°N–35°N, 130°E–140°E and the total grid number is $600 \times 1200 = 720000$.

The *JAGURS* tsunami code (Baba et al., 2015) is used to compute Green's

functions between the observation points and PoIs. The same parameters of *JAGURS* used for the forward tsunami simulation are adopted. Meanwhile, Green's functions between different observation points are also computed for data assimilation as the residuals should be calculated during the iterative process. So, the total number of Green's functions is $15 \times (9 + 15) = 360$, which does not take up much computer memory. The Optimal Interpolation algorithm parameters are the same as those in the work of Maeda et al. (2015). *JAGURS* enable the use of both the linear DSP and LLW models.

I set the earthquake origin time as $t = 0$. The observational stations of the assimilation network are not far from the epicenter of the 2004 off the Kii Peninsula earthquake. Hence, the tsunami arrives at the nearby stations of KMC09, KMC21, and KMD13 soon after the earthquake (Figure 2.7). The data assimilation process begins at $t = 0$, i.e., at the origin time of the earthquake. During the assimilation time window, the waveforms at PoIs are synthesized with Green's functions.

I apply GFTDA with both the LLW and the linear DSP models. The length of assimilation time windows is set from 2 to 24 min, at intervals of 2 min. The calculation time for the data assimilation process is less than 10 s, almost negligible on the EIC computer system at Earthquake Research Institute, the University of Tokyo.
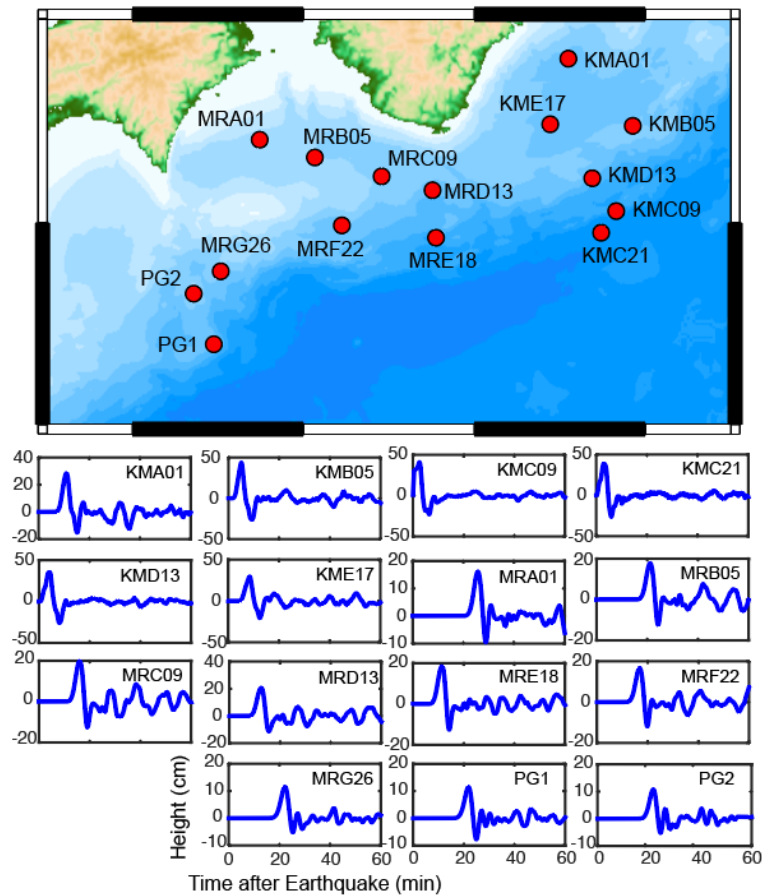
Figure 2.7. Distribution of 15 observational stations and waveforms of synthetic tsunamis.

The tsunami arrives at the KMC09, KMC21, and KMD13 stations soon after the earthquake. The data assimilation process begins at the origin time of the earthquake.

2.4.4 Results

Figure 2.8 demonstrates a comparison between simulated waveforms and waveforms predicted with an assimilation time window of 20 min. The waveforms predicted using both the LLW model and the linear DSP model agree well with the simulated waveforms. This proves the validity of data assimilation based on the observation network of DONET, PG1, and PG2. The LLW and linear DSP models have similar performances for the forecast of the first tsunami peak amplitude.

At coastal PoIs, the discrepancies of amplitude forecasted by the LLW and the linear DSP models are quite small. The main difference between two models lies in the

arrival time. At almost every PoI, the predicted waveform using the linear DSP model has a more accurate arrival time of the first tsunami peak. The performance of two models are also related to station locations. For example, the Osumi station, located in the direction of short axis (NE-SW direction), shows highly dispersive tsunami, and the linear DSP model has an evidently better performance than the LLW model. To the contrary, at Awaji or Abuyuki, which are in the long-axis direction (NW-SE direction), the predicted waveforms of two models are almost same.
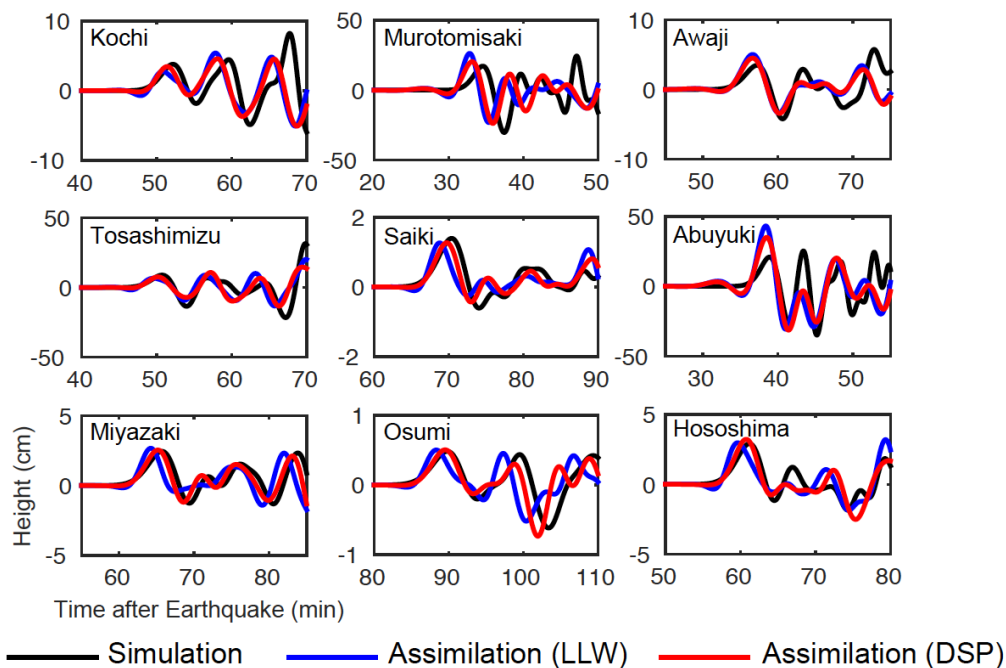
Figure 2.8. Simulated waveforms and waveforms forecasted by data assimilation at nine near-shore PoIs.

The simulated waveforms (black curves) are used as "observations" for comparison. The forecasted waveforms are calculated by GFTDA with the LLW model (blue curves) and the linear DSP model (red curves). The time window of data assimilation is 20 min.

To quantitatively evaluate the performance of two models in data assimilation, I calculate the tsunami forecast accuracy (Aida, 1978; Gusman et al., 2016b) for different assimilation time windows using Equations (2.5) and (2.6).

The accuracy for various assimilation time windows is plotted in Figure 2.9a.

The shapes of the forecast accuracy curves for both the LLW and linear DSP models are quite similar. In the beginning, when the time window is only 2 min, the accuracies of both models are very low. As the first tsunami peak has not passed any observational station of the network, the data length used for assimilation is too short to provide accurate forecasting. Then, at 4 min, there is a sharp increase in the accuracy curves for both models, with the forecast accuracies exceeding 85%. After that, the forecast accuracies vary slightly but exhibit a rising trend in general. There is not a large difference in forecast accuracy between the LLW model and the linear DSP model. After a time window of 20 min, the first tsunami peak has already passed all observational stations. The forecast accuracies become saturated and stop increasing. Here, the forecast accuracy of both models is similar, with of the linear DSP model being slightly higher.

The difference in arrival time of the first tsunami peak between the two models is more evident. To quantitatively analyze the accuracy of the forecasted arrival time, I calculate the time lag as proposed by Tsushima et al. (2012). The time lag of the i-th coastal station (PoI) is defined as:

$$\Delta T_i = t_i^S - t_i^O \quad (2.8)$$

where t_i^O is the arrival time of the maximum amplitude of the first tsunami peak at the i-th station, and t_i^S is the arrival time forecasted by data assimilation. A negative lag time indicates that the forecasted arrival time is earlier than the observation. A small absolute value of time lag indicates accurate forecasting of the arrival time. I calculate the time lag of all PoIs and calculate the average value.

In Figure 2.9b, it is clear that the time lags calculated by the two models are negative, which means that both the LLW and linear DSP models forecast the tsunami arrival time earlier. Moreover, as the assimilation time window increases, the time lag becomes closer to zero. The shape of the time lag curve is similar to that of the accuracy curve (Figure 2.9a). As more observed data are used in data assimilation, the absolute value of time lag decreases quickly from the 2 min time window to the 4 min time window. Then, it decreases slowly. After the time window of 20 min, the discrepancy becomes very small. It is important to note that the difference between the LLW model and the linear DSP model is noticeable in the figure. The linear DSP model leads to a much smaller time lag than the LLW model, indicating that the linear DSP model performs

better in forecasting the tsunami arrival time.

The results suggest that the tsunami propagation model could affect the accuracy of tsunami forecasting by data assimilation. For the maximum amplitude of the first tsunami peak, the two models perform similarly. However, with respect to the arrival time, the linear DSP model has a much better accuracy than the LLW model. The average time lag calculated by the linear DSP model is smaller. For individual stations, if the station is near Shikoku Island, close to the observation network, the lag time difference is not large. However, if it is located near Kyushu Island, which is approximately 200 km away from the observation network, the difference in lag time becomes noticeable. The limitation of long-wave approximation causes this. For the Kii Peninsula earthquake considered in my research, the large dip results in short-wavelength components of a tsunami, and the long-wave approximation overestimates the velocity of tsunami propagation (Saito et al., 2010). Thus, the arrival time forecasted by the LLW model is quite earlier than that of the DSP model, especially for PoIs far from the observation network, as a longer propagation distance exaggerates such errors.
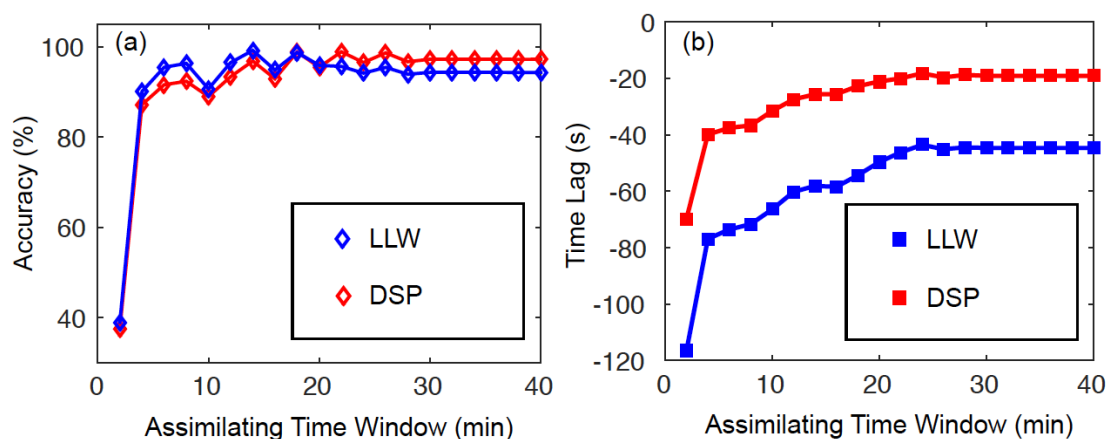

Figure 2.9. Forecast accuracy (a) and time lag (b) of two models for various assimilation time windows.

The forecast accuracy is used to evaluate the forecasted maximum amplitude of the first tsunami peak (Aida, 1978). The time lag is used to examine the accuracy of forecasted arrival time (Tsushima et al., 2012).

2.5 Application to Real-time Data—2015 Torishima Volcanic Tsunami Earthquake

Near Torishima, a volcanic island along the Izu–Bonin arc, unusual earthquakes repeatedly occurred in 1984, 1996, 2006, and 2015 (Fukao et al., 2018; Satake and Kanamori, 1991; Sugioka et al., 2000). They had a seismic magnitude of M ~ 6 but generated anomalously large tsunamis, and hence were characterized as "tsunami earthquakes" (Fukao et al., 2018; Sandanbata et al., 2018). The focal mechanism of these earthquakes deviated from a double-couple source and showed a CLVD type. Abnormal source mechanisms with volcanic origins have been proposed, such as hydrofracturing driven by supercritical water heated by injected magma (Kanamori et al., 1993), ring-fault rupture (Ekström, 1994; Shuler et al., 2013), or horizontal tensile crack opening (Fukao et al., 2018). Through teleseismic body-wave analysis and examination of the frequency content, it was found that these events had longer source durations than tectonic earthquakes of similar magnitude (Shuler et al., 2013). Fukao et al. (2018) estimated large seafloor uplift around a submarine volcanic caldera and successfully reproduced tsunami waves observed at temporal OBPGs and a tide gauge. However, the complex source mechanisms are still controversial, making it difficult to relate tsunami amplitudes to seismic parameters quantitatively.

The 2015 Torishima earthquake occurred on May 2, 2015 (UTC), and a tsunami was recorded on the southern coasts of Japan (Figure 2.10). According to a JMA report, the tsunami reached Hachijo Island with an amplitude of 0.5 m and Kozu Island with an amplitude of 0.3 m. The tsunami was also recorded by the offshore OBPGs, including the DONET and the DSFO off Muroto Cape (Figure 2.10; Kubota, 2018). In this section, I assimilate the actual tsunami data recorded at OBPGs from the 2015 Torishima earthquake and retroactively forecast the coastal tsunami arrival time and amplitude.
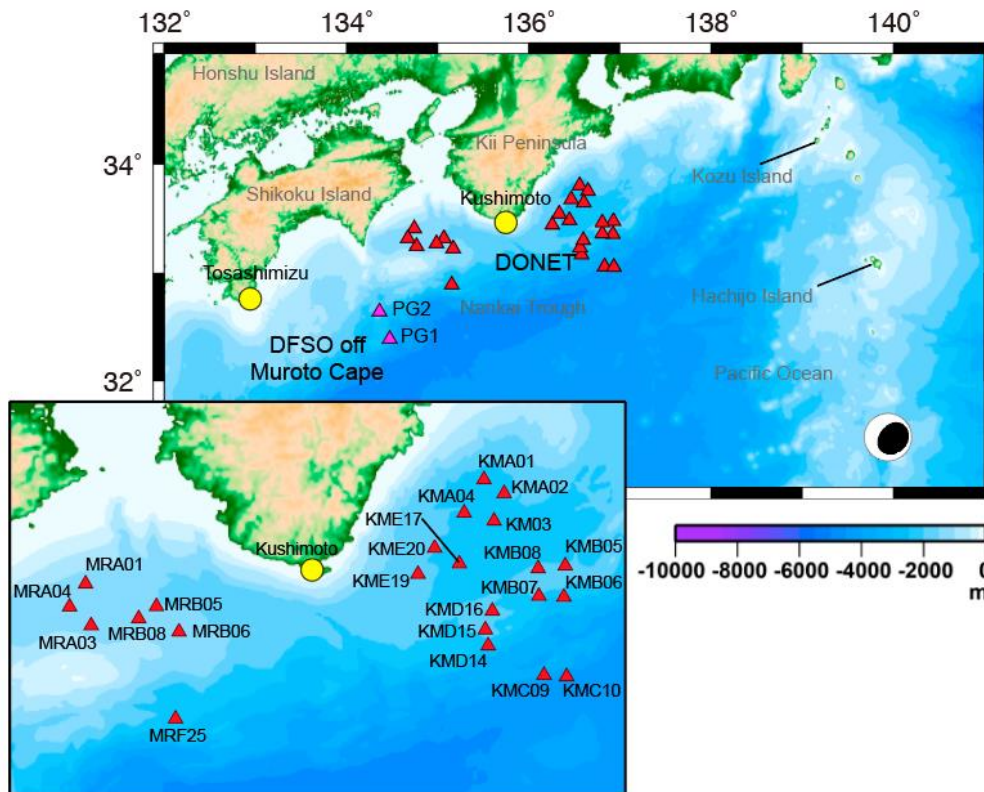
Figure 2.10. Bathymetry map of the 2015 Torishima earthquake.

The observation network is composed of 23 DONET stations (red triangles) and two DSFO stations off Muroto Cape. They are far from the earthquake epicenter, as shown by the focal mechanism solution of the United States Geological Survey (USGS). Two tide gauges (PoIs), Kushimoto and Tosashimizu, are indicated with yellow circles.

2.5.1 Observed Tsunami Data

I assimilate the tsunami records of DONET and the DSFO off Muroto Cape. There are 23 DONET stations and two DSFO stations: PG1 and PG2, with available tsunami records for the 2015 Torishima earthquake. The data are obtained from NIED and JAMSTEC. DONET records have a sampling rate of 0.1 s, and DSFO records have a sampling rate of 1 s. The raw records are plotted in Figure 2.11.

I first examine the records to ensure no data gaps, spikes, or repeating values in the time series and convert the pressure to water height by assuming the hydrostatic

pressure. Then, I remove the tidal components by fitting them with trigonometric functions. Finally, I remove the high-frequency components and extract tsunami signals. I apply a band-pass filter with cut-off frequencies of 0.002 and 0.01 Hz (500 and 100 s). The band-pass filter is the same as that used by Kubota (2018) for extracting tsunami signals of the 2015 and 2018 Torishima earthquakes. In this application, the tsunami signals are not affected by the seismic waves as the tsunami source is not close to the observation network. The processed data are plotted in Figure 2.12.

The tide gauge data are used for comparison to validate the method. As this tsunami event is quite small, there were tsunami records only at two tide gauges: Kushimoto, located in the Kii Peninsula of Honshu Island, and Tosashimizu at Shikoku Island (Figure 2.10). The tide gauge data are downloaded from the website of Sea Level Station Monitoring Facility, Intergovernmental Oceanographic Commission (IOC): http://www.ioc-sealevelmonitoring.org/. They have continuous records with a sampling rate of 60 s. The data processing steps are similar to those of OBPGs, but I apply a band-pass filter (0.002–0.00833 Hz) to tide gauge data as the Nyquist frequency was 0.00833 Hz.
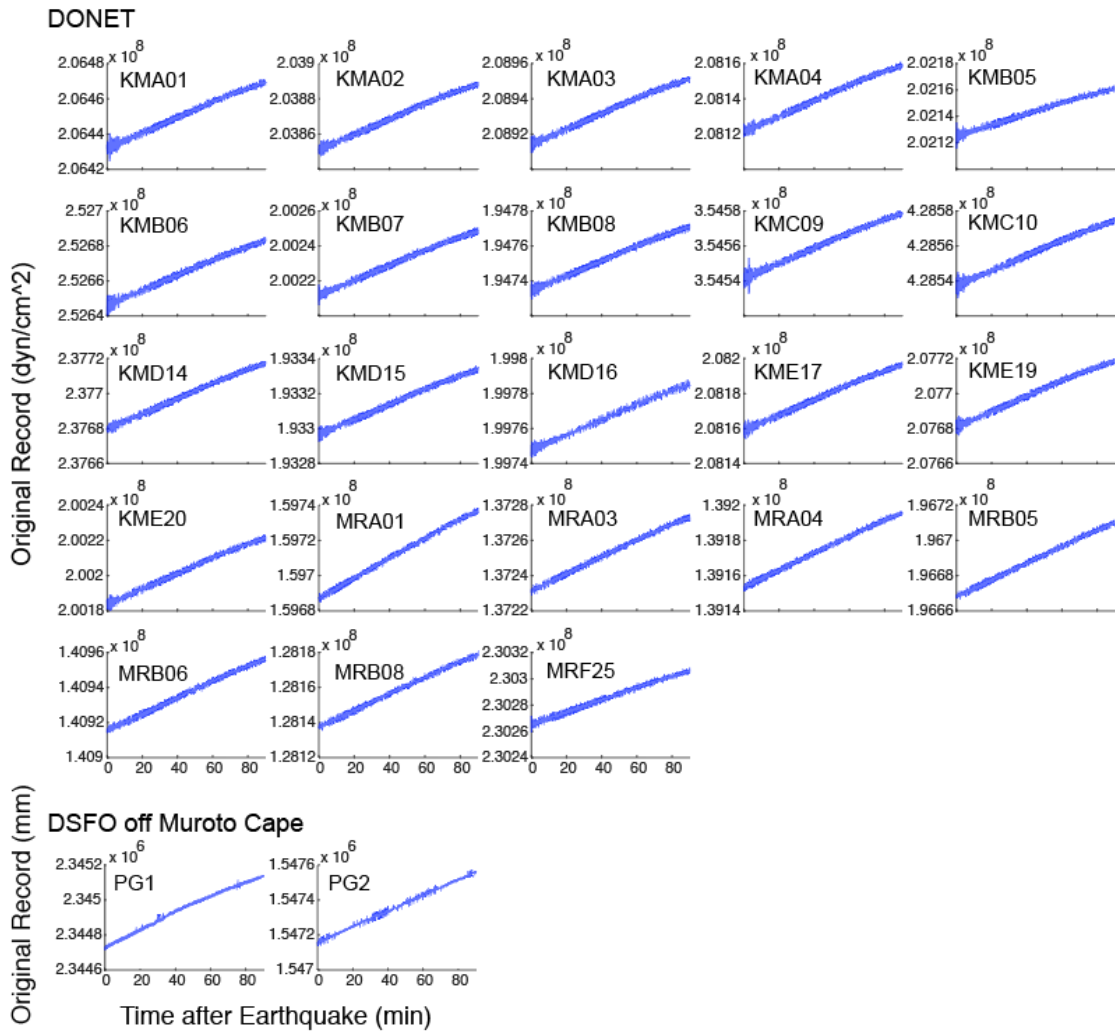
Figure 2.11. Raw records from 25 observational stations of DONET and DSFO.
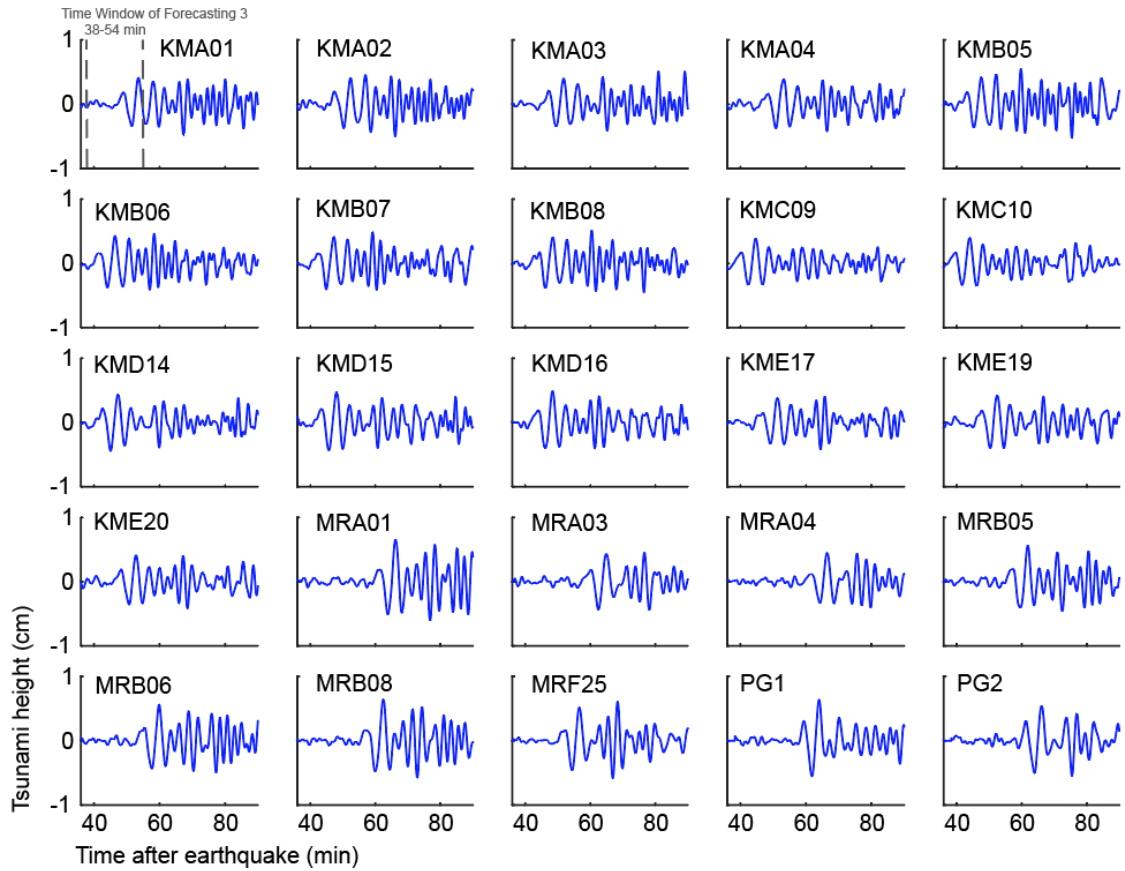
Figure 2.12. Processed data from 25 observational stations of DONET and DSFO.

Here I plot the time window of Forecasting 3 (38–54 min; see Section 2.5.2) as an example.

2.5.2 Assimilation Setting

The tsunami arrives at the observation network (KMC10) 38 min after the earthquake, which is defined by a threshold of 0.15 cm, at which point the data assimilation process begins. The first tsunami forecast is made 40 min after the earthquake. As time passes by, the records of more stations are assimilated, and the forecasting results are continuously updated at two PoIs, Kushimoto and Tosashimizu. The forecasted and observed waveforms with different assimilation time windows are then compared. A longer time window indicates that more offshore data are used in the assimilation process.

The bathymetry and topography data are derived from GEBCO_2014. The numerical grids of two tide gauges are located on land when using the current bathymetry data resolution. Therefore, I shift the original station locations to the nearest wet grid. As the shifted grid of Tosashimizu has a depth of more than 10 m, I correct its forecasted heights by the Green's Law, assuming that the coastal depth is 1 m (Baba et al., 2004; Wang et al., 2012).

The *JAGURS* model calculates Green's functions between the 25 observational stations and two PoIs with the linear DSP model (Baba et al., 2015). A time step of 1 s and a grid spacing of 0.5 arcmin are employed.

2.5.3 Results

For the tide gauge at Kushimoto (Figure 2.13a), the tsunami arrives 60 min after the earthquake, with a first-peak amplitude of 1.3 cm. The maximum amplitude is 2.7 cm. I compare the forecasted waveforms with time windows of 38–40 min and 38–45 min (i.e., 20 and 15 min before the arrival, respectively). At 40 min, the forecasting has an acceptable performance. The arrival time is predicted to be 61 min, slightly later than the observation. The forecasted first-peak and the maximum amplitudes are 0.9 and 1.4 cm. At 45 or 15 min before the tsunami arrival, the forecasted waveform also matches well with the observation. The forecasted first-peak amplitude is almost the same as the one at 40 min, but the forecasted maximum amplitude is 2.2 cm.

For the tide gauge at Tosashimizu (Figure 2.13b), the first-peak amplitude is 1.2 cm, and the maximum amplitude is 6.6 cm. I compare the forecasted waveforms with time windows of 38–54 min and 38–74 min (i.e., 35 and 15 min before the arrival, respectively). At 54 min, the forecasted amplitude is much lower than the observation; thus, it is difficult to determine the arrival time from the forecasted waveform. At 74 or 15 min before the tsunami arrival, the forecasted arrival time is 89 min, matching well with the observation. The forecasted first-peak and maximum amplitudes are 0.8 and 3.7 cm.
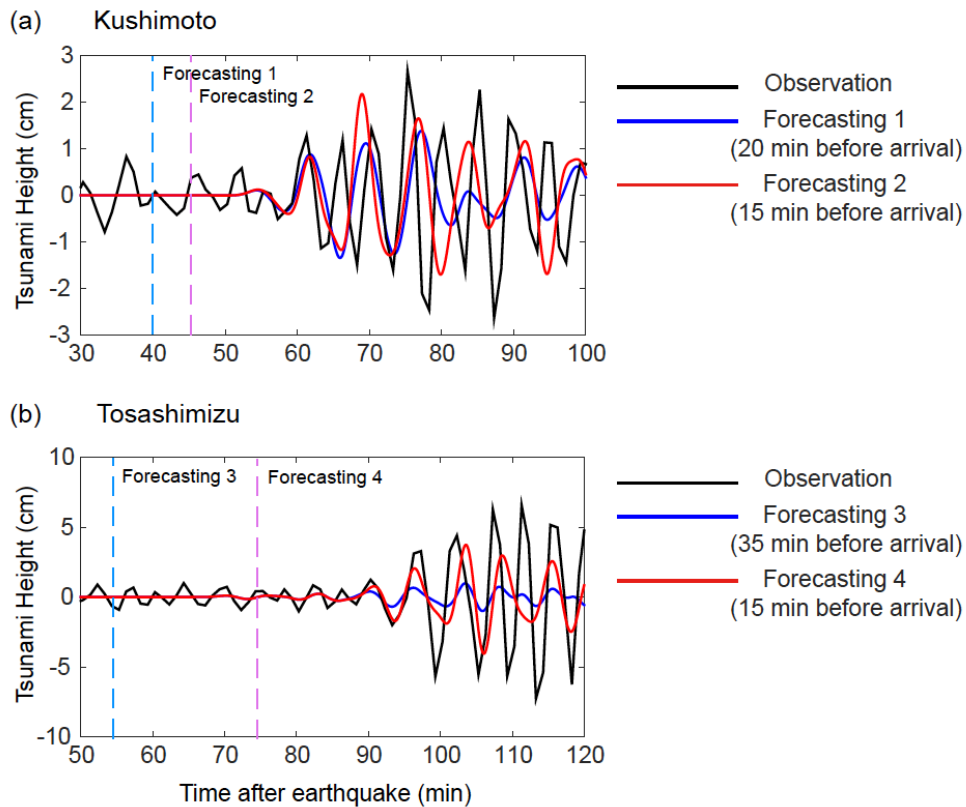
Figure 2.13. Waveform comparison of two PoIs.

I compare the observed and forecasted waveforms at Kushimoto (a) and Tosashimizu (b) with different time windows. The black curves indicate the observation. The blue curves indicate the forecasted waveforms computed at 20 min (Kushimoto) or 35 min (Tosashimizu) before tsunami arrival. The red curves indicate the forecasted waveforms computed 15 min before tsunami arrival. The dashed blue lines and dashed red lines represent the different timings before the tsunami arrival for obtaining the tsunami forecast.

I also compare the spectrum of the observation and the forecasted waveforms (Figure 2.14). I use the forecasted waveforms that are computed 15 min before the tsunami arrival at each station. The tsunami spectrum at Kushimoto has a large peak at approximately 0.0033 Hz (5.00 min) and small peaks at approximately 0.0021 Hz (7.93 min) and 0.0054 Hz (3.08 min). The spectrum of the forecasted waveform is smoother than that observed. It has a large peak at approximately 0.0022 Hz (7.58 min) and a small

peak near 0.0033 Hz (5.00 min). The tsunami spectrum at Tosashimizu has the greatest energy near 0.004 Hz (4.17 min) and another prominent peak at approximately 0.0025 Hz (6.67 min). In general, the spectrum of forecasted waveforms has a similar shape as that of the observation, except that the amplitude is lower for some frequencies.

At the two PoIs, the tsunami data assimilation approach forecasts the tsunami arrival time and first-peak amplitude accurately. The following waveform is also well predicted at the Kushimoto tide gauge. On the one hand, a longer time window leads to better forecasting as more data is assimilated. On the other hand, the tsunami early warning should be issued promptly, leaving enough time before the tsunami arrival.

The maximum amplitude of Kushimoto is accurately forecasted, but for Tosashimizu, the forecasting is not satisfactory. The reason for this is that Tosashimizu is located inside a harbor that lacks a dedicated local bathymetry. Although I correct its amplitude by Green's Law, the harbor effects could not be well simulated. This could also explain why the frequency spectrum of forecasted waveforms does not match well with observations. The spectral amplitudes of forecasted waveforms are generally smaller than those of observations likely caused by the resolution issue of bathymetry. The spectrum of tsunami waveforms greatly depends on the local bathymetry (Rabinovich, 1997; Saito, 2019). It contains components of the complex geometries of bays and harbors that are not well simulated. To solve this problem, bathymetric grids with higher resolution are necessary to simulate the effect of radiation damping through the mouth of the bay/harbor. Additionally, as I shift the station locations to the nearest wet grid, the arrival time estimation may be slightly inaccurate.
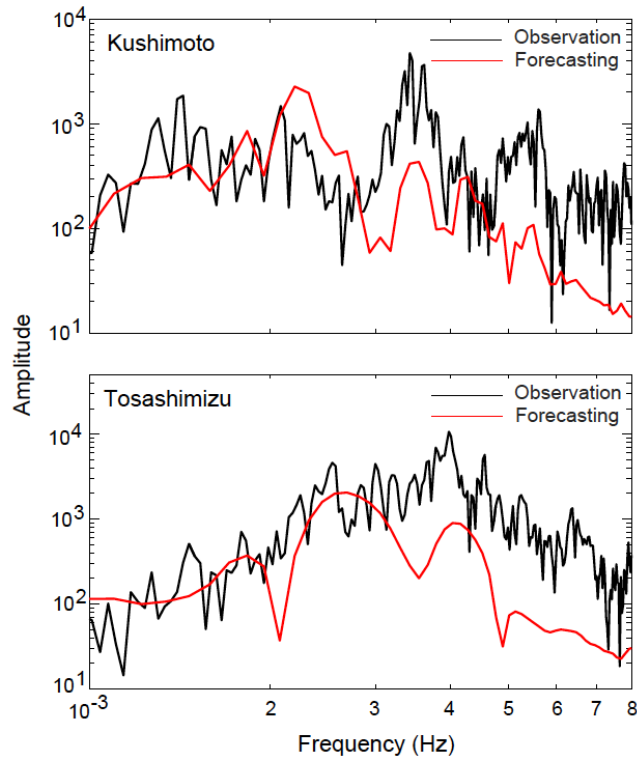
Figure 2.14. Spectrum comparison of two PoIs.

The spectrums of observed (black curves) and forecasted (red curves) waveforms are shown. The forecasted waveforms are computed 15 min before the tsunami arrival.

2.5.4 Accuracy and Number of Stations

For the 2015 Torishima volcanic tsunami earthquake, I utilize tsunami data recorded at the stations of a dense observation network in the Nankai region. As some stations are located quite close, the information may be redundant. In future operations, some stations may also encounter technical problems and may not be able to provide real-time data transmission. Therefore, I investigate the relationship between the number of observational stations and the forecast accuracy. A similar type of study has previously been proposed (e.g., Heidarzadeh et al., 2019; Hossen et al., 2018; Mulia et al., 2017a; Navarrete et al., 2020). Here, I discuss which stations are more important for tsunami data assimilation. To evaluate the forecasting performance in terms of tsunami amplitude and arrival time, I use a general error function G (Navarrete et al., 2020) as follows.

$$G = \alpha(1 - P) + \beta Q \quad (2.9)$$

$$P = 1 - \frac{\sum(H_f^{\max} - H_o^{\max})^2 + \sum(H_f^{\mathrm{arr}} - H_o^{\mathrm{arr}})^2}{\sum(H_o^{\max})^2 + \sum(H_o^{\mathrm{arr}})^2} \quad (2.10)$$

$$Q = \frac{1}{N_s}\sum\frac{\left|t_f^{\mathrm{arr}} - t_o^{\mathrm{arr}}\right|}{T_{\mathrm{eva}}} \quad (2.11)$$

The subscripts f and o represent the forecasted and observed tsunami waveforms, respectively. Equation (2.10) characterizes the forecast accuracy of tsunami amplitude, developed from the method of Tsushima et al. (2012). P is a normalized index (ranging from 0 to 1), $H^{\max}$ is the maximum amplitude of the tsunami waveform, and H^{arr} is the first-peak amplitude. A higher (close to 1) P value indicates better forecasting. Using Equation (2.11), I evaluate the forecast accuracy of tsunami arrival time. Both early and late arrival time predictions are regarded as forecasting errors. N_s is the total number of PoIs. The t^{arr} represents the arrival time of the first tsunami peak. Here I use the absolute value to avoid canceling out of the positive and negative time lags. The average time shift Q is the ratio of absolute time lag and evacuation time, meaning that a smaller Q is more accurate. The evacuation time is the interval between the forecasting time and the tsunami arrival time. The index α and β in Equation (2.9) are two positive weight parameters determined as per forecasting requirements. Their relative values represent the assumption of whether the accurate forecasting of arrival time or amplitude is more important. Hence, I set these two parameters in such a way that $\alpha + \beta = 1.0$. Overall, general error function G evaluates the tsunami forecast accuracy quite well. As the absolute value of G largely depends on the absolute value of weight parameters, I normalize the G for comparison. A relatively smaller value of G indicates a higher accuracy.

I start from the entire number of observational stations and gradually decrease the number. The general error function G varies with a changing station distribution. For each number, I perform a search of the best design that could produce the least general error function $minG$. I use the gradient descent algorithm for searching for the optimal solution (Garg et al., 2012). The results represent the best distribution of a certain number of stations. This evaluation method was first adopted in the optimal design of OBPGs off

the Chilean coast (Navarrete et al., 2020). In this study, the goal is to evaluate the importance of existing stations in tsunami data assimilation.
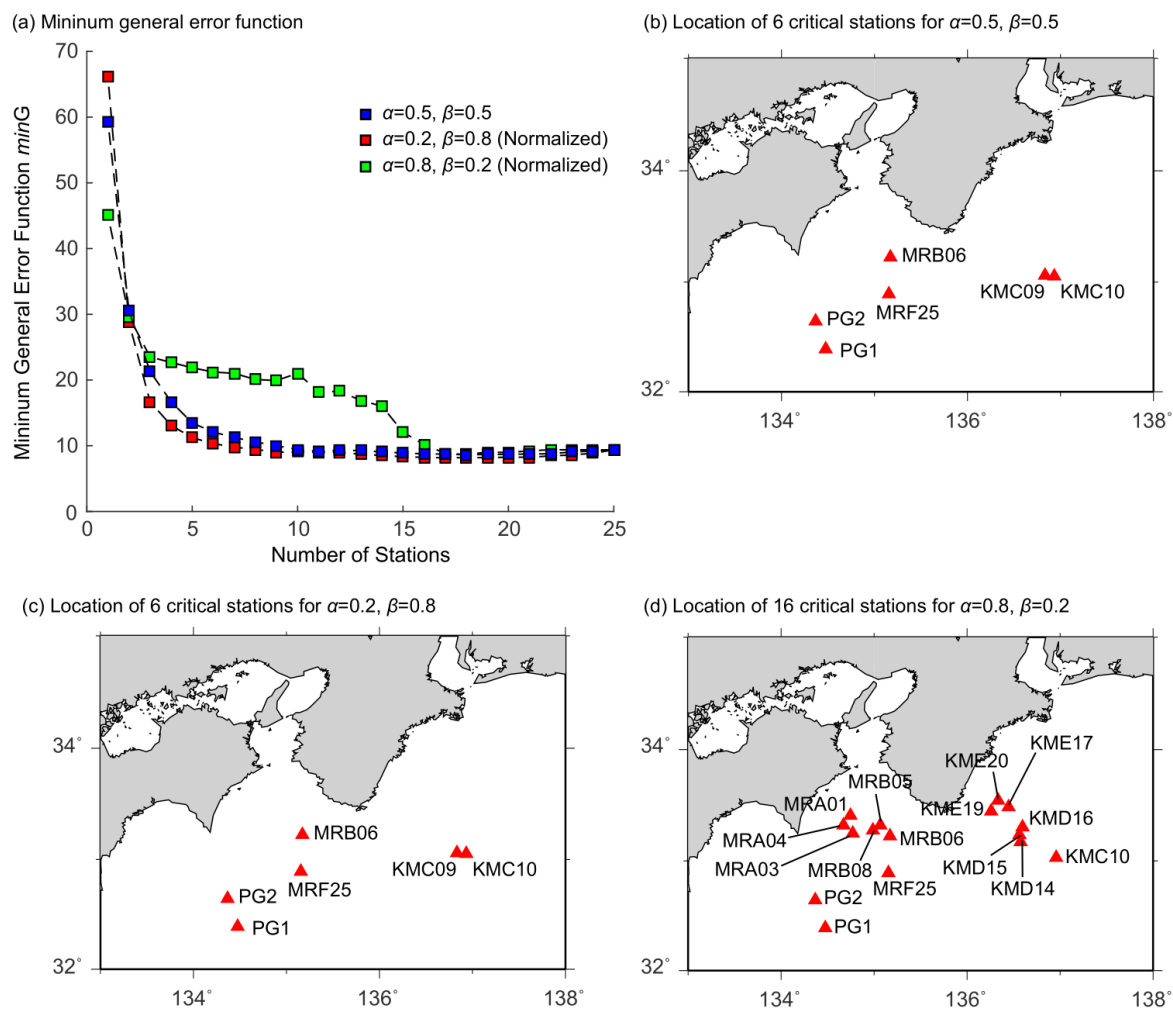
Figure 2.15. Minimum general error function and location of critical stations.

(a) Relationship between the number of stations and general error function G. The minimum general error function $minG$ for each station is searched by a gradient descent algorithm. For a better comparison, the results of the three settings are plotted by adjusting the second and third settings so that they all have the same values with 25 stations. (b–d) Location of critical stations of different settings. According to the value of weight parameters α and β, the number and location of critical stations are different.

I choose the assimilation time window of 7 min (38–45 min) for Kushimoto and 36 min (38–74 min) for Tosashimizu. This means that for both PoIs, the forecast is made 15 min before tsunami arrival. Therefore, the evacuation time T_{eva} is 15 min for both stations.

First, I set the two weight parameters α and β to be equal at 0.5, which indicates that I assume that the accurate forecasting of arrival time and amplitude are equally important, and calculate the minimum general error function $minG$ for each number of stations (Figure 2.15a). I start from the current number 25. In the beginning, as the number of stations decreases, $minG$ increases slowly with small variations. In this situation, even though I remove some "unimportant" stations from the observation network, there is not a substantial influence on the forecast accuracy as the remaining network is dense enough. This trend continues until there are ten remaining stations. After that, as the number of stations decrease, $minG$ increases dramatically. That is to say, all stations in this state are important for tsunami data assimilation. If one of the stations has technical problems, the overall forecast accuracy would be seriously affected, and the tsunami data assimilation approach would no longer be reliable. All in all, in the first set with $\alpha = 0.5$, $\beta = 0.5$, six is the critical number of stations. This could achieve a fair forecast accuracy ($minG = 9.37$) with the least number of observational stations. At this number, the network is composed of KMC09, KMC10, MRB06, MRF25, PG1, and PG2 (Figure 2.15b). These stations play a significant role in tsunami data assimilation because of their locations. For data assimilation purposes, more attention should be paid to their operation and maintenance.

Furthermore, I test two other settings: $\alpha = 0.2$, $\beta = 0.8$; $\alpha = 0.8$, $\beta = 0.2$. The former indicates a higher requirement on the forecasting of arrival time as I set a larger value of weight parameter β. In other words, I assume that the accurate forecasting of arrival time is more important. In contrast, the latter indicates a higher requirement on the forecast accuracy of amplitude. I normalize the value of $minG$ of these two settings, assuming that their values are equal to those of first setting, at 25 stations. It is found that they have similar shapes as that of the first setting. The critical points are 6 and 16 for the second and third setting, respectively (Figure 2.15c; d). Different settings represent a different emphasis on the forecast accuracy of tsunami amplitude or arrival time. Furthermore, the determination of "critical points" is arbitrary, depending on the tolerance

of forecasting errors. Here, the results of critical stations are calculated from a single tsunami event. To analyze the necessary design of OBPGs thoroughly, we can use Empirical Orthogonal Function analysis to find the place where the most energetic wave dynamics occur (Navarrete et al., 2020).

To summarize, by using the evaluation method based on the general error function G, I could determine which stations are more important in tsunami data assimilation. Therefore, more efforts could be devoted to the operation and maintenance of these stations.

2.6 Discussion

The previous tsunami data assimilation approach of Maeda et al. (2015) and Gusman et al. (2016b) can provide accurate tsunami forecasting for coastal regions, even though the computational cost is high. To provide accurate tsunami forecasting more rapidly, I develop a new method based on Green's functions. In the previous approach, the observed data are repeatedly assimilated over the time window, and in the remaining time, the tsunami propagation model is run for the assimilated tsunami wavefield. In contrast, GFTDA does not need to calculate the whole tsunami wavefield during the data assimilation process. Instead, I directly synthesize the waveforms at PoIs by a simple matrix manipulation.

The 2012 Haida Gwaii earthquake application indicates that GFTDA has a similar accuracy as that of the previous approach, with a much shorter assimilation time. The previous method usually requires supercomputers for the assimilation process, and the high computational complexity inevitably results in time-consuming calculations. GFTDA simplifies the assimilation process because Green's functions are calculated and stored in advance. The assimilation process is so simple that it can even be run on personal computers. For early tsunami warnings, saving time means saving lives. The reduction of the assimilation time is thus essential for an effective tsunami early warning.

The application to the 2004 off the Kii Peninsula earthquake shows that the simple assimilation process of GFTDA makes it possible to apply more complicated models, apart from the LLW model. For potential tsunamis with more dispersive

characteristics or tsunamis that propagate over a longer distance from the observation network, the linear DSP model can make improvements in forecasting the arrival time accurately.

Moreover, the application to the 2015 Torishima earthquake is the first time that uses the records of a real-time observation network to forecast a tsunami by the data assimilation approach. The forecasted waveforms computed 15 min before the tsunami arrival have a good consistency with real observations. The tsunamis at two tide gauges, Kushimoto and Tosashimizu, are predicted merely based on the offshore observed data, avoiding the complexities in the source characterization of the volcanic tsunami earthquake. Although the tsunami is small, retroactive forecasting is very meaningful in tsunami disaster mitigation. If a tsunami is larger, by combining the GFTDA with nonlinear tsunami inundation models on coastal regions of interest (Liu et al., 2009), inundation forecasting will also be possible.

Chapter 3   *Tsunami Data Assimilation with Interpolated Virtual Stations*

3.1 Linear Interpolation with Huygens–Fresnel Principle

In this chapter, I adopt virtual stations to perform tsunami data assimilation for regions without a dense observation network. The artificial waveforms of virtual stations are computed by the linear interpolation of real data. The principle of this method is similar to the Huygens–Fresnel principle in optics. According to the Huygens–Fresnel principle (e.g., Hadamard, 1924), every point on a wavefront is itself the source of spherical wavelets. The resulting amplitude at any position in the scattered field will be the vector sum of the amplitudes of all the individual waves. For the data assimilation of a tsunami wave, the observational stations resemble the points on the wavefront. If there is a wavefront consisting of dense observational stations, the wavefield could be reproduced using the data assimilation approach. Otherwise, some information can be lost during the assimilation process. Hence, to compensate for the lost information because of sparse observations, I artificially construct waveforms at virtual stations.

Virtual stations do not exist, and their "waveforms" are only used for building the assimilation wavefield. I consider several virtual stations between two real stations (Figure 3.1). The virtual stations are located at equal distances along a straight line between real stations. The effects of the number of stations, or the distance between the virtual stations, will be discussed in Section 3.2.3. More generally, the interpolation scheme can be applied to three or more real stations (e.g., virtual stations inside a triangle), but I only consider two stations for the sake of simplicity. The network of stations (both real and virtual) could form a wavefront as in the Huygens–Fresnel principle.

The first task for constructing a virtual waveform is a linear interpolation of two real arrival times to estimate the tsunami arrival time at a virtual station. I define a threshold for tsunami arrival in each station, using t_A^{arr} and t_B^{arr} to represent the arrival

times of two real stations. Then I calculate the arrival time of the *i*-th virtual station between two real stations as:

$$t_i^{arr} = w_{iA} \cdot t_A^{arr} + w_{iB} \cdot t_B^{arr} \quad (3.1)$$

where w_{iA} and w_{iB} are two weight parameters for linear interpolation. The subscripts A and B represent the two neighboring real stations. Practically, w_{iA} and w_{iB} are the relative distances between the virtual and real stations if the tsunami velocity around the two stations is assumed to be constant. Their sum equals one. For instance, if the *i*-th virtual station is located at the middle point of two real stations, the value of the two weight parameters are both 0.5. After obtaining the arrival time of the virtual station, the amplitudes of the two real observations are interpolated to obtain that of the artificial waveform by taking the water depths at the stations into consideration. The tsunami waveforms after the arrival time of two real stations are represented as $Y_A(t - t_A^{arr})$ and $Y_B(t - t_B^{arr})$. The tsunami waveform at the *i*-th virtual station is calculated as follows.

$$y_i(t - t_i^{arr}) = \left[\frac{w_{iA} \cdot Y_A(t - t_A^{arr})}{d_A^{-1/4}} + \frac{w_{iB} \cdot Y_B(t - t_B^{arr})}{d_B^{-1/4}}\right] \cdot d_i^{-1/4} \quad (3.2)$$

where w_{iA} and w_{iB} are weight parameters, and d_A, d_B and d_i are the water depths of two real stations and the *i*-th virtual station. The correction to water depth follows Green's law that suggests that the tsunami amplitude is inversely proportional to the fourth root of water depth (Satake, 2015).

The main characteristic of the linear interpolation method is that the virtual waveforms are calculated by shifting the arrival times considering the distance and the amplitudes are corrected considering the water depths at the stations. The arrival times are linearly interpolated by assuming constant velocity because we do not know the direction of wave arrival. On the contrary, the correction of amplitudes depends only on the water depths at the known stations, based on the assumption that the tsunami propagates as a plane wave. I validate my method with real data in Section 3.3.2. The virtual stations help construct a complete tsunami wavefront. As the interpolation depends on the waveform information of two adjacent real stations, the virtual waveforms cannot be computed until the tsunami arrives at both real stations. During the assimilation process, the tsunami height at both real and virtual stations is used to forecast the tsunami
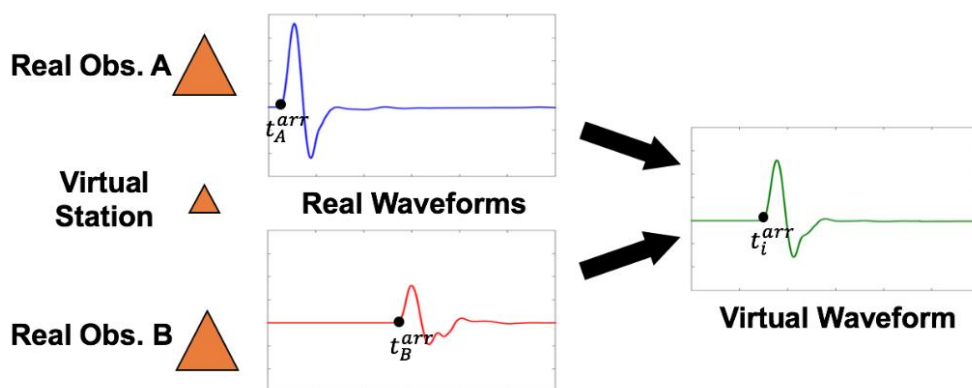
waveforms.

Figure 3.1. Illustration of the linear interpolation process.

I first find the tsunami arrival time of two real stations and calculate the arrival time of virtual station(s) by the weighted average. Then, I calculate the virtual waveform(s) by shifting the arrival time with a correction for water depth. The virtual waveform(s) will be adopted in data assimilation along with real waveforms.

3.2 Test with Synthetic Data—2004 Sumatra–Andaman Earthquake

To test the effectiveness of the improved tsunami data assimilation method, I perform a synthetic experiment of the 2004 Sumatra–Andaman earthquake. The earthquake occurred at 00:58:53 UTC on December 26, 2004, and generated a tsunami that caused more than 283,000 deaths (Lay et al., 2005). The tsunami propagated across the Bay of Bengal (Figure 3.2) and arrived at India and Sri Lanka's coasts about 2 h after the earthquake (Fujii and Satake, 2007). As no OBPGs were installed near the location of this tsunami event, I use synthetic data in the experiment.

3.2.1 Synthetic Tsunami Data

The fault model and seismic parameters are based on the results of Fujii and Satake (2007). They estimate the slip distributions by inverting the tide gauge and satellite altimeter data, assuming a rupture velocity of 1.5 km/s. I calculate the seafloor

63

displacement from the faulting (Okada, 1985) and use it as the initial condition for tsunami propagation (Figure 3.2). The LLW model is employed in the numerical simulation. The bathymetry grid data is derived from GEBCO_2014, with a grid size of 2 arcmin. The computation domain for Green's functions is 70 °E–100 °E, 0 °–25 °N, with the total grid number of 675,000. In the numerical simulation, the time step is 1 s, which satisfies the Courant–Friedrichs–Lewy condition, necessary for a stable simulation. I store the simulated tsunami waveforms at six OBPGs in the Bay of Bengal as the synthetic observations: Stations 23217, 23218, 23219, 23227, 23223, and 23401. These stations were installed after the 2004 Sumatra–Andaman earthquake, and they are currently the only available stations for tsunami detection in the Bay of Bengal (north Indian Ocean). Then, I use linear interpolation to compute the waveforms at 25 virtual stations with an average interval of approximately 50 km (Figure 3.2).

To validate the method of tsunami forecasting, I compare the tsunami waveforms at seven near-shore PoIs along the coasts of India, Sri Lanka, and Maldives (Figure 3.2). As the tsunami waveforms computed from the source model by Fujii and Satake (2007) do not match well with the real observations, probably because of inaccurate bathymetry data near the stations, I use the synthetic waveforms. Green's functions of observations (real and virtual stations) and near-shore PoIs are computed and stored in advance. The characteristic distance of the Optimal Interpolation algorithm is 20 km.
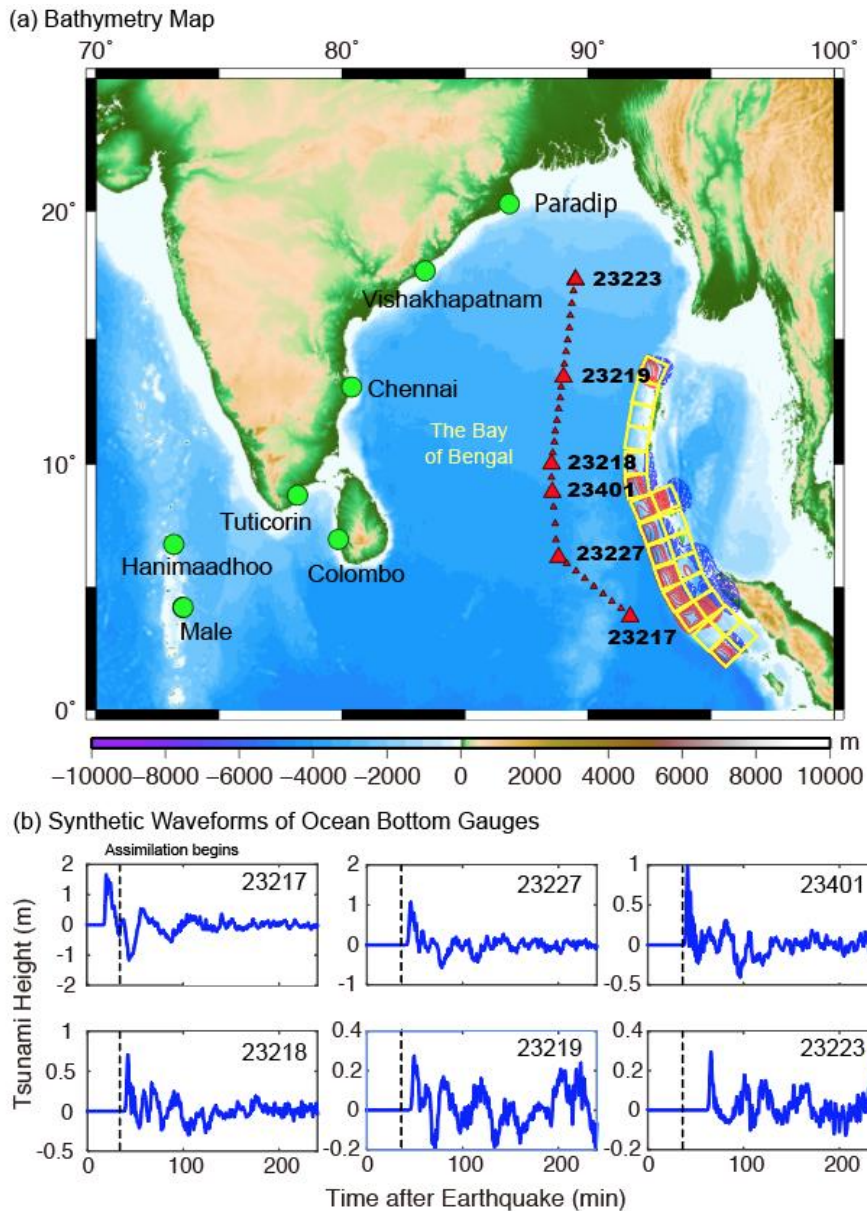
Figure 3.2. Bathymetry map of the Bay of Bengal and synthetic tsunami waveforms.

(a) The six OBPGs used in the numerical simulation are indicated with large red triangles. The 25 virtual stations indicated with small red triangles are interpolated between OBPGs. The seven near-shore points (green circles) record the tsunami waveform, and I compare them with the forecasted waveforms calculated by tsunami data assimilation. (b) Synthetic tsunami waveforms at six OBPGs. The assimilation begins 30 min after the earthquake.

3.2.2 Assimilation Setting and Results

I set the earthquake origin time as $t = 0$. When the propagating tsunami reaches the OBPG Station 23217, at around 30 min after the earthquake, the data assimilation process begins. After this time, the tsunami height and arrival time are forecasted by the superposition of Green's functions.
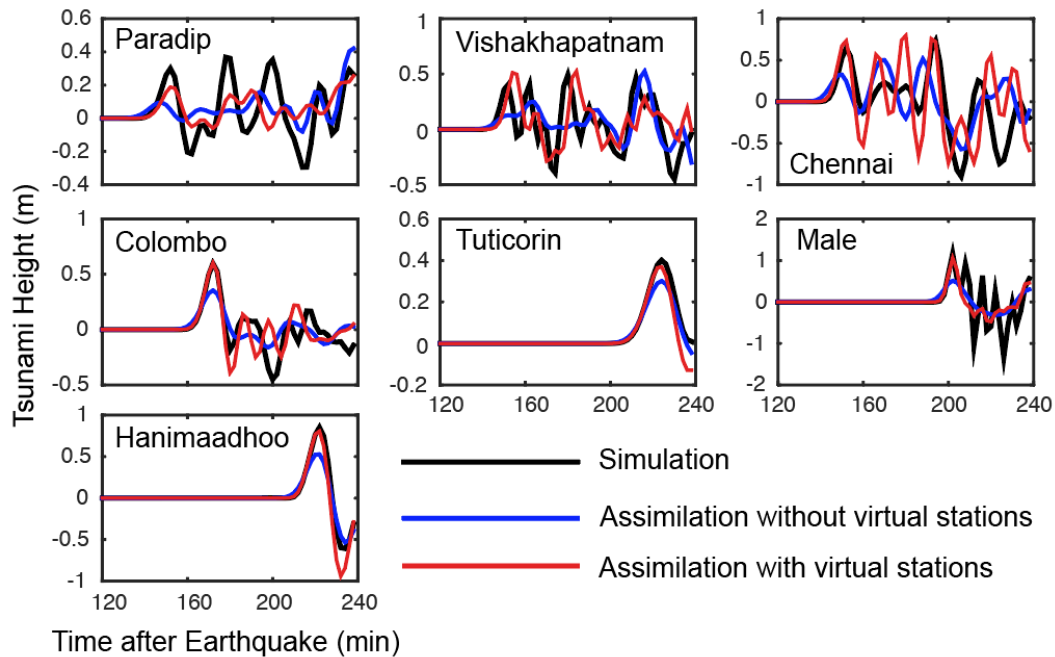
Figure 3.3. Comparison of the observed and forecasted waveforms at seven near-shore points.

The black curves represent the simulated waveforms. The blue curves represent the assimilated waveforms without virtual stations, whereas the red curves represent the assimilated waveforms with virtual stations. The assimilation time window is 60 min for both cases.

Figure 3.3 compares the simulated and forecasted tsunami waveforms at seven near-shore stations, with an assimilation time window of 60 min. The forecasted waveforms generally match with the simulated waveforms at all stations. However, without virtual stations, the forecasted waveforms underestimate the maximum amplitude of the first tsunami peak. On the contrary, the assimilation with virtual stations has better

performance, improving the forecasting of tsunami amplitude and period. For example, at the near-shore PoI of Male, the simulated maximum amplitude of the first tsunami peak is 1.16 m. The assimilated amplitudes without and with virtual stations are 1.06 and 0.51 m, respectively, indicating a significant improvement of the results by including virtual stations. Overall, the forecast accuracy calculated by Equations (2.5) and (2.6) increases from 51.4% to 73.1% with the help of virtual stations.

It is important to note that the tsunami forecast is made at 60 min (90 min after the earthquake) when the first tsunami peak has passed all OBPGs. As the tsunami arrives at the Indian coast around 150 min after the earthquake and arrives at Sri Lanka and the Maldives even later, there is enough time to conduct the data assimilation process and transfer appropriate warning messages to the public at risk.

3.2.3 Effects of Interpolation Interval

In this application, I use synthetic waveforms at six OBPGs to artificially build a dense network. Different interpolation intervals may affect forecast accuracy. Meanwhile, in the Optimal Interpolation algorithm, the covariance matrices of the forward numerical simulation and the observation are assumed to have a Gaussian correlation with a certain characteristic distance (Kalney, 2003; Maeda et al., 2015). In the last section, a Gaussian-distributed covariance is assumed for the correlation function of errors, with a characteristic distance of 20 km (Equation 1.28). The relationship between the characteristic distance and the interpolation interval remains to be discussed. According to the recent study of Yang et al. (2019), the Optimal Interpolation algorithm may have a poor performance if the interval between the neighboring stations is greater than ~50 km.

To study these effects, I use synthetic 2004 Sumatra–Andaman earthquake data and test four interpolation scenarios of virtual stations at the Bay of Bengal (Figure 3.4). In Case 1, only four virtual stations are assumed to be present, with station intervals of ~150 km. In Case 2, 10 stations are assumed at intervals of ~100 km. Case 3 is adopted in Section 3.2.2, with 25 virtual stations at ~50 km intervals. In Case 4, 40 virtual stations are assumed with the smallest interval of ~30 km. For each case, I test different

characteristic distances for the Optimal Interpolation algorithm, from 20 to 80 km, and calculate their forecast accuracies.
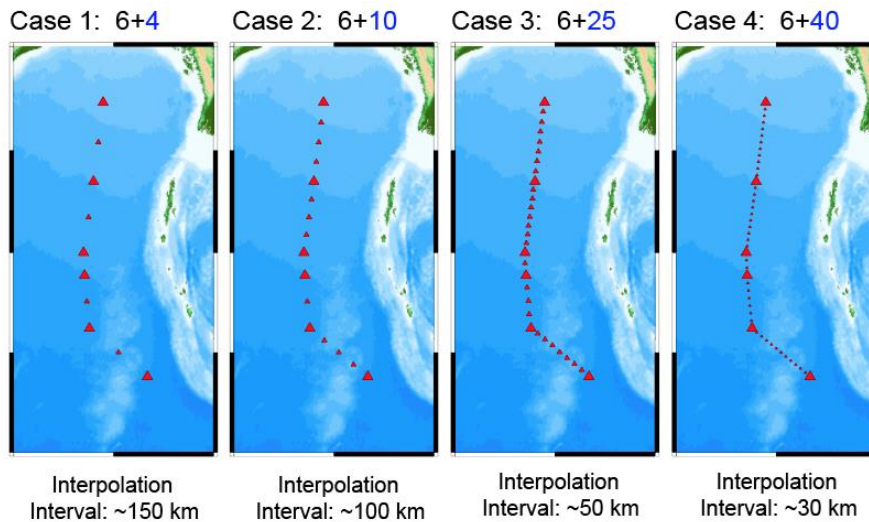
Figure 3.4. Four interpolation scenarios of virtual stations at the Bay of Bengal.

The number and location of real OBPGs are fixed, but the number and interpolation interval of virtual stations are different. From Case 1 to Case 4, the interpolation becomes denser.

After computing the synthetic tsunamis, I calculate the forecast accuracy by Equations (2.5) and (2.6) for each case with different characteristic distances (Figure 3.5). Case 1 has the lowest accuracy for all characteristic distances, and Case 2 has a better accuracy. Cases 3 and 4 have the best performance, and their results are close to each other. It can be concluded that a dense interpolation of virtual stations may help improve the forecasting up to a certain level, but beyond that, the accuracy tends to saturate. On the other hand, the number of Green's functions will also increase by adding virtual stations, which may affect the forecasting efficiency. Therefore, it is important to balance the number of virtual stations and the desired forecast accuracy.

It is also observed that the characteristics distance affects the optimum point of forecast accuracy. The smallest characteristic distance of 20 km results in the highest accuracy for Cases 3 and 4, which have interpolation intervals of ~50 and ~30 km, respectively. Case 2 with an interpolation interval of ~100 km requires a 30 km

characteristic distance, and Case 1 with an interpolation interval of ~150 km requires a 40 km characteristic distance to produce a high forecast accuracy.

As the interpolation becomes denser, the interval between the two stations becomes smaller. A dense interpolation is helpful to improve the forecast accuracy. In this study, the station interval should be less than ~ 50 km to guarantee a high accuracy over 75%. This empirical value is consistent with the distance criterion for Optimal Interpolation (Yang et al., 2019). A characteristic distance of 20 km is appropriate in such interpolation scenario.
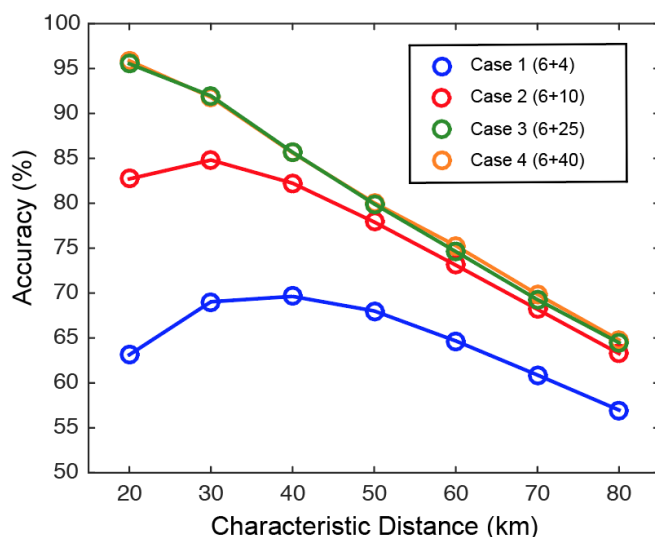
Figure 3.5. Forecast accuracies of four scenarios versus different characteristic distances. The characteristics distances range from 20 to 80 km. The accuracy is calculated by Equations (2.5) and (2.6).

3.3 Test with Real Data—2009 Dusky Sound Earthquake

The Dusky Sound earthquake (M 7.8) occurred near the southwestern coast of New Zealand at 09:22:29 UTC on July 15, 2009 (Beavan et al., 2010; Fry et al., 2010; Heidarzadeh and Gusman, 2019). It was the largest event ever recorded at the Puysegur subduction zone. The earthquake generated a tsunami that was recorded by tide gauges around the southwestern South Island and by DART gauges in the south Pacific (Clark et

al., 2011). The Marine Observations of Anisotropy near Aotearoa (MOANA) OBS network also detected the signal of the tsunami (Sheehan et al., 2019).

3.3.1 Observed Tsunami Data

At the time of the earthquake, up to 30 OBSs were equipped with OBPG in the west and east of South Island. The raw data are available at the Ocean Bottom Seismograph Instrument Pool (website: http://www.obsip.org). The tsunami signals are extracted by removing the tidal components and bandpass filtering at 0.0002–0.005 Hz with a fourth-order Butterworth filter. Then, the instrument responses are de-convolved, following the method of Sheehan et al. (2019). As the tsunami was recorded by differential pressure gauges, the tsunami amplitude at each station is corrected by the ratio between the observed and simulated peak amplitude at each station (Beavan et al., 2010).

Tide gauge data are used for waveform comparison to validate the method. As the region of interest is the western coast of South Island, I only use the tide gauge observation of Charleston (Figure 3.6a). The tsunami amplitude of tide gauges is usually affected by local bathymetry and harbor effects (Baba et al., 2004; Heidarzadeh et al., 2016a; Kontar et al., 2013; Leonard, 2006).
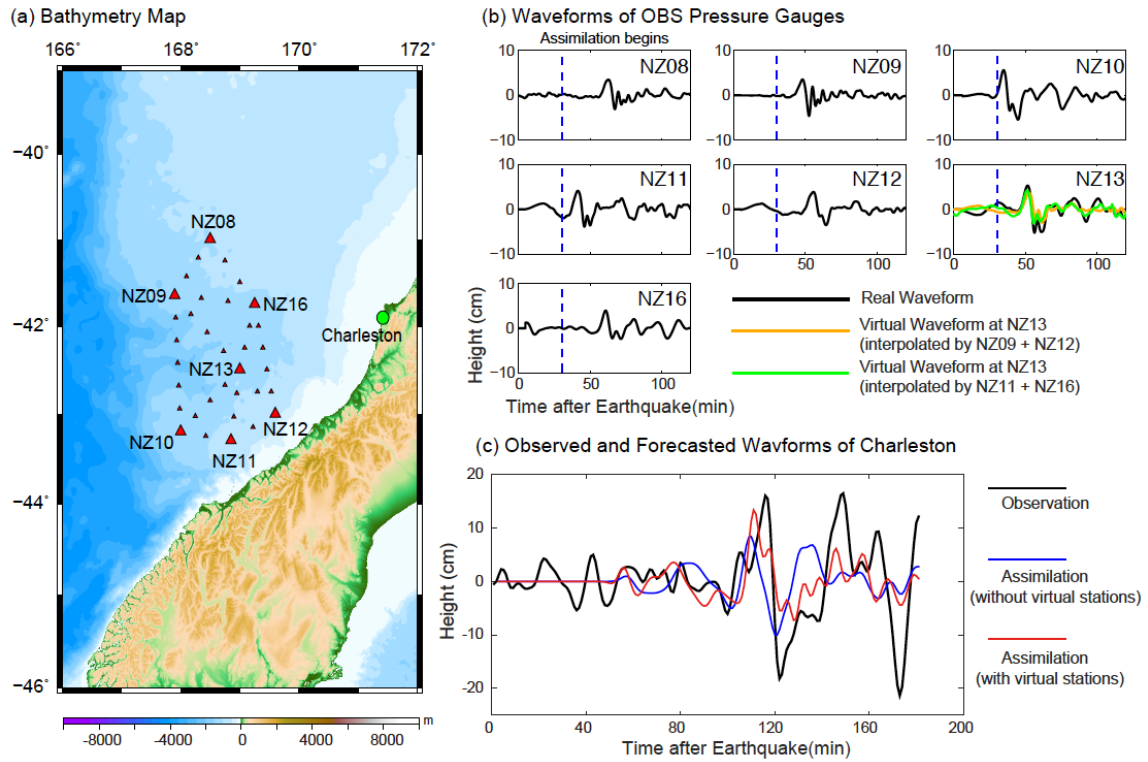
Figure 3.6. Tsunami data assimilation for the 2009 Dusky Sound earthquake.

(a) Distribution of OBPGs and the Charleston tide gauge. The seven OBPGs indicated with large red triangles are used for data assimilation. The 28 virtual stations indicated with small red triangles are interpolated between OBPGs. (b) The observed tsunami waveforms at seven OBPGs. The assimilation begins 30 min after the earthquake. The real (black curve) and virtual (yellow and green curves) waveforms at NZ13 are compared. (c) Comparison of the observed and forecasted waveforms at the Charleston tide gauge with an assimilation time window of 40 min.

3.3.2 Assimilation Setting

Though 30 OBSs are installed during the MOANA project, data from only a few stations are available for data assimilation. Among the stations in the west, NZ15 was trawled up by a fishing vessel, and NZ17 was not recovered (Yang et al., 2012). NZ01 and NZ02 waveforms were clipped (Sheehan et al., 2019). Additionally, some stations are too far from the Charleston tide gauge, so the tsunami arrives at these even later than

at the tide gauge. Therefore, there are only seven stations for tsunami data assimilation: NZ08, NZ09, NZ10, NZ11, NZ12, NZ13, and NZ16.

I use 28 virtual stations interpolated between real stations, with an average interpolation interval of approximately 50 km. I choose neighboring station pairs so that virtual stations do not overlap with each other. As station NZ13 is nearly located in the line between NZ09 and NZ12 and the line between NZ11 and NZ16, it provides an opportunity to validate the linear interpolation method of virtual stations. I interpolate a virtual waveform at NZ13 using data from NZ09 and NZ12 (yellow curve in Figure 3.6b) and another virtual waveform at NZ13 using the data from NZ11 and NZ16 (green curve in Figure 3.6b). Then I compare this with the real waveform (black curve).

The characteristic distance of the Optimal Interpolation algorithm is 20 km. The bathymetry grid data is derived from GEBCO_2014 with a grid size of 0.5 arcmin. The computation domain for Green's functions is 163°E–175°E, 49°S–37°S, with the total grid number of 2,073,600. The time step of the numerical simulation is 1 s. Unlike the approach of Sheehan et al. (2019), I only use the real tsunami data as the input and interpolate virtual stations to overcome the problem of sparse observations.

3.3.3 Results

In the test of NZ13, the arrival time and amplitude of the first tsunami peak are very similar between the real observation and the two virtual waveforms (Figure 3.6b). Hence, it validates the linear interpolation method of virtual stations. Although there are slight discrepancies, virtual waveforms could supplement real observation in tsunami data assimilation and compensate for the lost information because of sparse stations.

The data assimilation process begins 30 min after the earthquake when the tsunami arrives at the first OBPG (NZ10). Figure 3.6c compares observed and forecasted waveforms at the Charleston tide gauge at the time window of 40 min (i.e., 40 min after the tsunami arrival at NZ10). The forecasted waveform matches the observed waveform reasonably, and their periods are similar. However, without virtual stations, the amplitude of the forecasted waveform (8.4 cm) is smaller than that of the observation (16.0 cm). The accuracy is only 52.5%. The method with virtual stations gives a better forecast of

the tsunami amplitude (13.4 cm), with an accuracy of 83.8%. As the tsunami arrives at the Charleston tide gauge at 112 min after the earthquake, the tsunami forecast is made at approximately 32 min before arrival.

3.4 Application to Far-field Event—2015 Illapel Earthquake

The 2015 Illapel earthquake occurred on September 16, 2015, at 22:54:33 (UTC) offshore Illapel, Chile, because of thrust faulting on the interface between the Nazca and South America Plates. The West Coast & Alaska Tsunami Warning Center of NOAA issued a rapid magnitude estimation of M 7.2 at 4 min after the earthquake, and the PTWC estimated the magnitude to be M 7.9 at 5 min after the earthquake (Cienfuegos et al., 2016). Finally, the USGS revised the magnitude to M 8.3. The location of the epicenter is 31.573 °S, 71.674 °W, at a depth of 22.4 km. The earthquake caused 15 deaths, six missing people, and considerable damage to coastal cities. It generated a tsunami that reached the Chilean coastal region within 8 min after the earthquake (An and Meng, 2017; Aránguiz et al., 2016) and resulted in flooding in the cities along the coast. The tsunami propagated across the Pacific Ocean and was recorded by tide gauges and DART tsunameters, though the amplitude was much smaller than that of the 2010 tsunami and slightly smaller than that of the 2014 tsunami (Satake and Heidarzadeh, 2017). The DART stations provide tsunami time series recorded at the open ocean and can be used to reconstruct the source model of the earthquake (Heidarzadeh et al., 2016b; Williamson et al., 2017). For example, Ren et al. (2017) studied the characteristics of tsunami waves in the deep ocean by using the records of 16 DART tsunameters of this event.

In this research, I focus on far-field tsunami forecasting of Chilean earthquakes. Tsunamis generated by Chilean earthquakes produce significant damage not only to the Chilean coast but also in areas far away from the source, including Easter Island (Fritz et al., 2011), Oceania (Hébert et al., 2001), and even Japan (Satake et al., 2020). The recent installation of DART tsunameters off the Chilean coast enables the application of the tsunami data assimilation approach. The DART stations are 32401 and 32402 owned and maintained by the Hydrographic and Oceanographic Service of the Chilean Navy, and stations 32403, 32404, and 34420 owned and maintained by the Cooperative Effort DART 4G Buoy (Figure 3.7a). They record water levels and transmit a signal to the land

in real time (González et al., 2005). I use the tsunami data from the 2015 Illapel earthquake as an application of far-field tsunami data assimilation. I retroactively predict the tsunami in the east Pacific using the tsunami data assimilation approach and compare the forecasted results with the observations.

3.4.1 Observed and Synthetic Tsunami Data

Tsunami waveform data from DART tsunameters and tide gauges are considered for this study. Four DART tsunameters are used for data assimilation, and forecasted waveforms are compared with the observed data on tide gauges and other DART tsunameters.

The DART data of the 2015 Illapel earthquake are available at NOAA's website (https://www.ngdc.noaa.gov/hazard/dart/2015chile.html). Tsunami signals have already been extracted and processed by the Tsunami Detection Algorithm that estimates the amplitude of the pressure fluctuations within the tsunami frequency band (Meinig et al., 2005). Then, the amplitude is computed by subtracting predicted pressures from the observations, in which the predictions closely match the tides and lower frequency fluctuations. In my retroactive study, to remove the high-frequency components like seismic waves, I also use a low-pass filter with a cut-off frequency of 0.002 Hz and apply a zero mask before the arriving tsunami because I only need the waveforms after the tsunami arrival. For most tsunamis generated from mega-thrust earthquakes, the first peak of the tsunami at OBPGs is free of coastal reflections and harbor effects (Williamson and Newman, 2018).

At the time of the 2015 Illapel earthquake, only two DART stations, 32401 and 32402, were operational. Therefore, in addition to these two observed waveforms (blue curves), I build synthetic waveforms (green curves) at stations 32404 and 34420 (Figure 3.7b) by using the fault model of Ren et al. (2017). I only compute synthetic waveforms at two other stations to ensure that at least half of the data are real records. Therefore, I skip the station 32403, which is close to station 32401. Ren et al. (2017) assumed a focal depth of 25 km, strike angle of 4°, dip angle of 19°, and rake angle of 90°. The fault size (length $\times$ width) is 230 km $\times$ 100 km and a uniform slip of 4.64 m is adopted. The

numerical simulation from this uniform fault matches well with the real observations from near-field and far-field DART stations. In addition to DART stations near the Chilean coast, I also extract the data of DART station 32412, located off the southwest of Peru. I compare its observations with the forecasted waveform to evaluate the accuracy of the method.

The tide gauge data are available at Sea Level Station Monitoring Facility, IOC (http://www.ioc-sealevelmonitoring.org/). The tidal components are removed by polynomial fitting. I use the far-field tsunami records of tide gauges in the east Pacific, including those of San Felix, Easter, Rikitea, Nuku Hiva Island, Papeete Tahiti, Huahine, and Rarotonga (Figure 3.8a). The distance between tide gauges and South America ranges from hundreds to ten thousand kilometers.
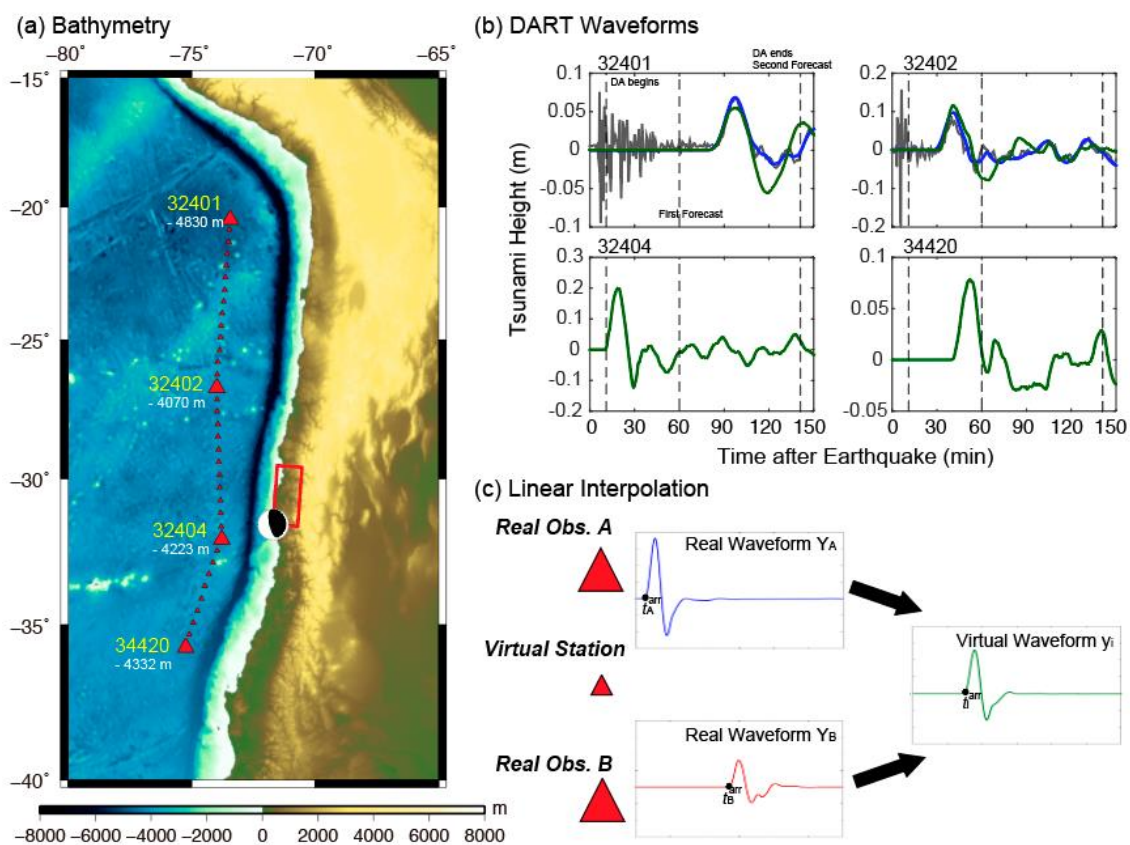

Figure 3.7. Tsunami data assimilation of DART tsunameters along the Chilean coast.

(a) Bathymetry of the source region. I assimilate the data at four real DART stations and 32 virtual stations interpolated between them. The focal mechanism is plotted by using the USGS solution. The fault model of Ren et al. (2017) is plotted with a red rectangle.

(b) Waveforms at the four DART stations. Grey curves represent the raw data from the NOAA website, blue curves represent observed waveforms, and green curves represent synthetic waveforms. Observed waveforms at stations 32401 and 32402 and synthetic waveforms at stations 32404 and 34420 are considered as the "observations." The synthetic waveforms are calculated from the fault model of Ren et al. (2017). Station 32403 is not considered. The time for the beginning and end of data assimilation and tsunami forecasting are marked by the dashed lines. (c) The process of linear interpolation to create waveforms at virtual stations. The virtual waveforms are calculated by shifting the arrival times considering the distance and correcting the amplitudes considering the water depths at the stations.

3.4.2 Assimilation Setting

The waveforms at the four DART tsunameters are adopted as the input for the tsunami data assimilation (Figure 3.7b). I interpolate 32 virtual stations between the neighboring DART tsunameters (Figure 3.7a). To prove the effectiveness of the method, I compare the data assimilation results with and without virtual stations.

The Green's functions are calculated by the *JAGURS* tsunami code (Bata et al., 2015) with a linear DSP model on spherical coordinates. To ensure linearity, nonlinear effects such as bottom friction are not considered in the numerical simulation. As most of the tsunami propagation path is in the deep ocean, I select the grid spacing of 2 arcmin to avoid a large computational cost, and the time step is 1 s, which meets the requirements of the stability condition. For far-field tsunamis, most of the propagating path is in the wide Pacific Ocean, for which the bathymetry or tsunami velocity is well known.

The bathymetry data are derived from the 30 arcsec gridded data of GEBCO_2014. Because of the grid size, the exact location of tide gauges at the coast cannot be well represented. The altitude of the corresponding point sometimes becomes positive (i.e., above the sea surface). To avoid this problem, I moved the gauge location and shift it to the nearest wet grid.

I forecast the tsunami amplitudes and arrival times at PoIs, including DART tsunameters and tide gauges (Figure 3.8a). Tide gauge data are more affected by local

topography and harbor effects and thus have larger amplitudes and more complicated waveforms and spectrums than OBPG records (Rabinovich, 1997; Rabinovich et al., 2015).

3.4.3 Results

I define the time of origin $t = 0$ when the 2015 Illapel earthquake occurred. The tsunami arrival time is defined by a threshold of 0.02 m. The tsunami arrives at station 32404 at $t = 12$ min, after which the assimilation process begins. It is important to mention that the waveforms at virtual stations are not calculated until the tsunami arrives at both neighboring stations. This may cause some delay in forecasting, but the effects are small when considering far-field tsunamis. The first tsunami forecasting is made at $t = 1$ h when the first tsunami peak has passed stations 32402 and 32404, as well as the virtual stations interpolated between them, and the tsunami waveforms at PoIs (e.g., San Felix) are synthesized. As time passes, more observed data are used in the assimilation, and the forecast accuracy improves. The whole assimilation process ends at $t = 2$ h 20 min when the first tsunami peak has completely passed the four DART tsunameters.

I compare the forecasted and observed tsunami waveforms at PoIs (Figure 3.8b). The tsunami heights at tide gauges have already been corrected by Green's Law, assuming that the coastal depth is ~1 m. (Baba et al., 2004; Synolakis and Skjelbreia, 1993; Wang et al., 2012). Additionally, the waveforms at tide gauges are affected by local topography and harbor effects that may not be fully modeled by the simulation using the 2 arcmin grid. Thus, I mainly focus on the amplitude and arrival times of the first tsunami peak at tide gauges as the first peak is less affected (Calisto et al., 2016; Yamazaki and Cheung, 2011).

The tsunami waveforms forecasted by data assimilation without virtual stations (blue curves) show poor results when compared with the observations. They underestimate the tsunami amplitudes and do not predict the arrival time accurately. On the contrary, with the help of virtual stations (red curves), the results improve substantially.
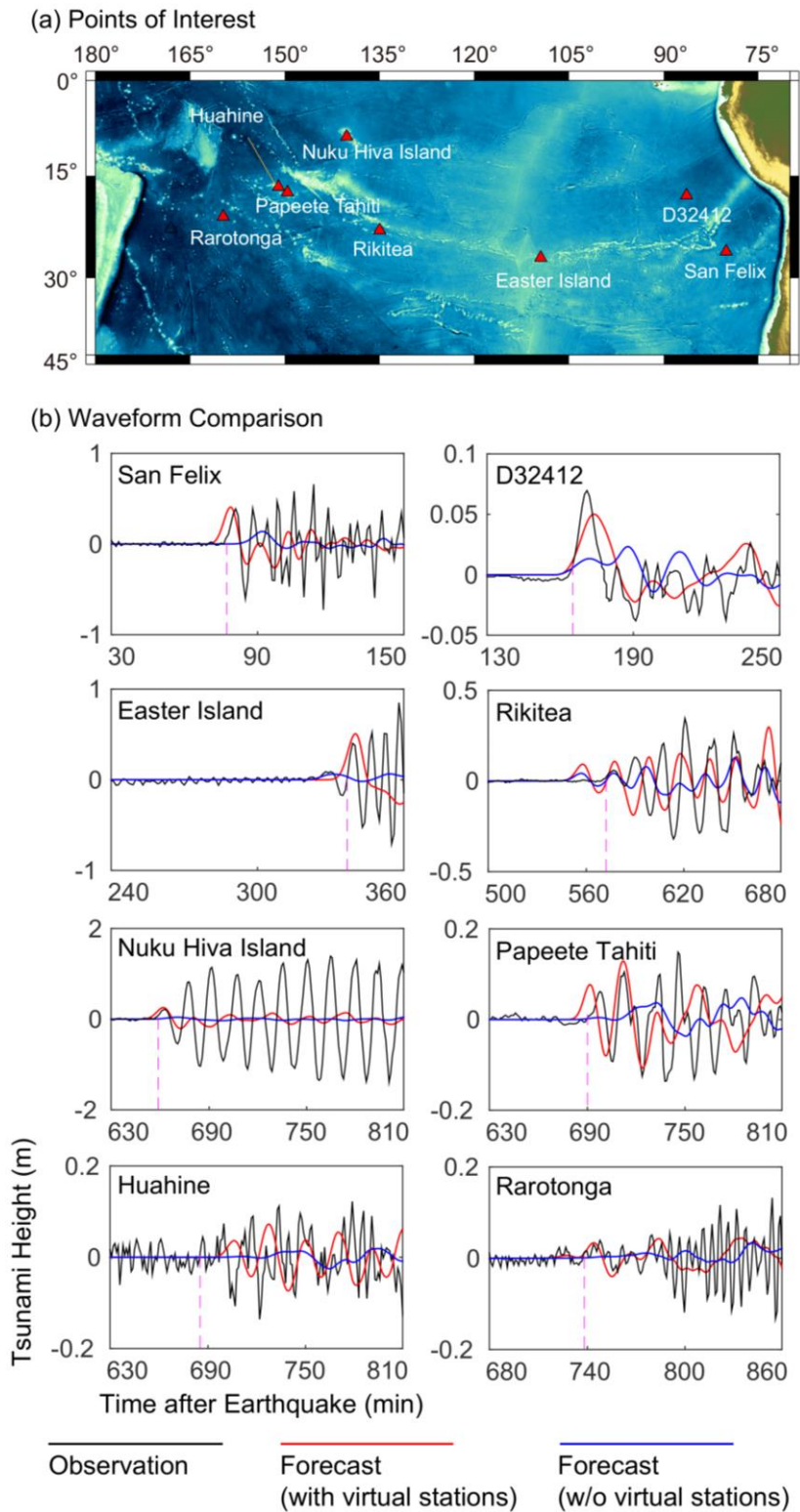
Figure 3.8. Locations of and waveforms at PoIs.

(a) Locations of eight PoIs. The far-field tsunami in the east Pacific is forecast from the

ocean off the Chilean coast to regions in Oceania. (b) Comparison of the observed and forecasted waveforms at PoIs. The black curves represent the observed waveforms. The red curves represent the forecasted waveforms by data assimilation with virtual stations. The blue curves represent the forecasted waveforms by data assimilation without virtual stations.

At the San Felix tide gauge near the Chilean coast, the actual tsunami arrives at 1 h 17 min after the earthquake with an amplitude of 0.39 m. The results show that the forecasted tsunami arrival time is 1 h 13 min after the earthquake, and the amplitude is 0.41 m. As the tsunami arrives at San Felix before the end of the data assimilation process, I use the results forecasted at $t = $ 1 h 6 min. DART station 32412 records the tsunami arrival at 2 h 45 min after the earthquake, and the amplitude is 0.07 m. The forecasted waveform has an arrival time of 2 h 45 min and an amplitude of 0.05 m. At both stations, there is a high consistency between the observed and forecasted waveforms.

At Easter Island, the tsunami arrives at 5 h 36 min after the earthquake, and the amplitude of the first tsunami peak is 0.40 m. The comparison indicates that the method forecasts the tsunami arrival time (5 h 34 min) precisely. Additionally, the forecasted amplitude is 0.51 m, which is higher than the first observed peak, but at the same level as the following observed peaks.

In Oceania, as tide gauges are quite far from the source, the tsunami takes a very long time to propagate. The forecasted waveforms still generally match the observation. The tsunami arrival times of the first tsunami peak are predicted well at most tide gauges, including Nuku Hiva Island, Papeete Tahiti, and Rarotonga. At Rikitea, the forecasted tsunami arrives approximately 18 min earlier than real observation. At Huahine, the forecasted tsunami arrives approximately 14 min later than the real observation. As for the tsunami amplitude of the first tsunami peak, the forecasted results underestimate the observed values at Rikitea, Nuku Hiva Island, and Papeete Tahiti, while at Huahine and Rarotonga, the amplitude of the first tsunami peak is slightly overestimated. Although I mainly focus on the first tsunami peak, the following waveforms after the first peak are consistent with the real observations at most stations. However, at Nuku Hiva Island, the following waveform shows strong oscillation

characteristics, which are not captured well by the forecasted results.

Overall, the forecast accuracy is 87.5% for all the stations as per Equations (2.5) and (2.6). I also calculate the average value of absolute time lag as 5.75 min using Equation (2.8). Considering the long travel time of several hours, the arrival time forecasting is judged to be accurate. I compare the forecasted and observed arrival times at each PoI (Figure 3.9). The slope of the regression line is 1.0037, with an R-square value of 0.999. This indicates that the method does not have a bias in forecasting the arrival time.
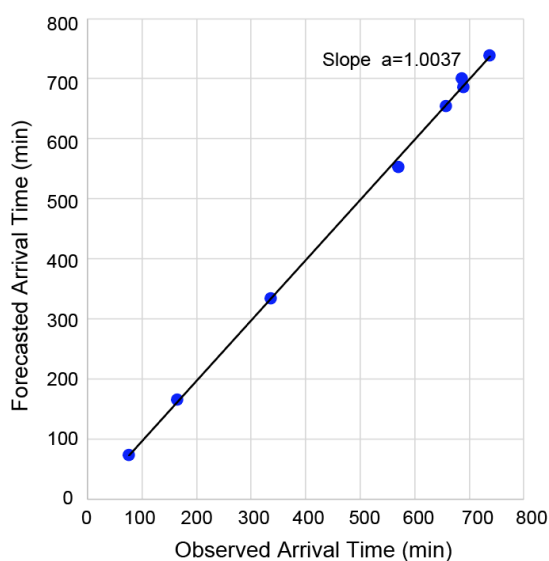
Figure 3.9. Comparison of the observed and forecasted tsunami arrival time at eight PoIs.

3.5 Discussion

The adoption of virtual stations enables us to perform tsunami data assimilation in regions without a dense observation network, such as the Bay of Bengal. The new method does not introduce any new information except the water depth. It is based on the assumption that a tsunami propagates as a plane wave, and that the wavefront is a nearly straight line. Hence, the tsunami amplitude in the open sea depends on the water depth in most cases. The data assimilation approach resembles the Huygens–Fresnel Principle. If there is no dense observation network, the previous assimilation procedure can lose some

information of the wave as there are not enough "secondary sources" on the wavefront. Therefore, the interpolated virtual stations are an approximation of the lost information because of sparse stations. They help reconstruct the wavefront and prevent information loss to some extent. Of course, if there is enough budget, it will be better to build sufficient real stations to prevent such information loss instead of using interpolated virtual stations as an approximation.

As this method is based on the assumption that the tsunami travels as a plane wave, I acknowledge that it will have some limitations when this assumption cannot be satisfied. Generally, there are mainly two scenarios where the method of virtual stations that cannot be applied. In the first scenario, the source is close to the observational stations (Figure 3.10a), and the tsunami wavefront is more like a spherical wave instead of a plane wave. Therefore, it is not applicable to create waveforms at virtual stations. However, Tanioka (2018) proposed a tsunami data assimilation method based on OBPG data in the earthquake source region. Inoue et al. (2019) successfully applied this method to synthetic studies of the 1952 Tokachi-oki earthquake (M 8.2) and the 1968 Tokachi-oki earthquake (M 8.1). Although this method is different from the Optimal Interpolation algorithm, it can help in cases where the virtual-station method cannot be applied. In the second scenario, the bathymetry varies very quickly. For example, when there is a seamount between two real observational stations (Figure 3.10b), the tsunami wavefront will be bent and no longer be a straight light. Because the tsunami velocity depends on water depth, the bathymetry change should be smooth to satisfy the plane-wave assumption. Except for these two scenarios, the virtual waveforms calculated by linear interpolation could compensate for the lack of real observations, which has been verified by the case study on the 2009 Dusky Sound earthquake. Hence, it can improve tsunami data assimilation performance and can make it affordable for more countries and regions.
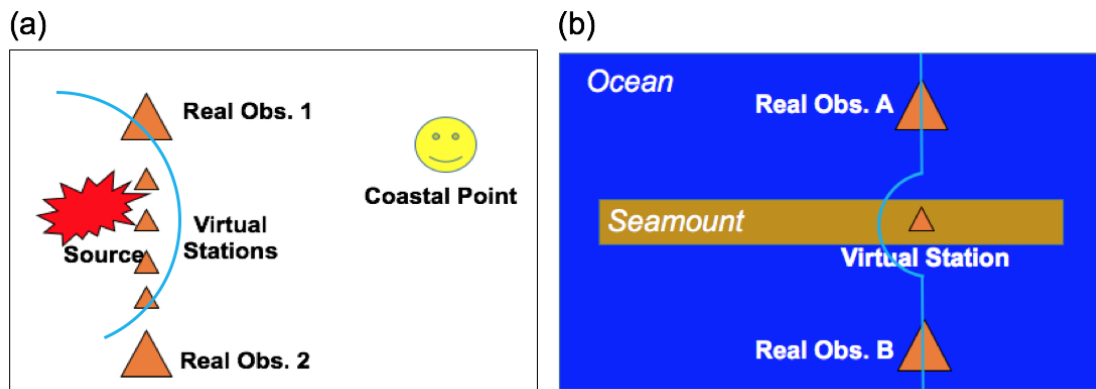
Figure 3.10. Two scenarios where the virtual-station method cannot be adopted.

(a) The scenario where the source is close to the observational stations (marked by orange triangles). The blue curve indicates the tsunami wavefront. (b) The scenario where the bathymetry varies very quickly.

Chapter 4   Real-Time Tsunami Detection based on Ensemble Empirical Mode Decomposition (EEMD)

4.1 EEMD

Time–frequency analysis can characterize and manipulate signals in time and frequency domains. The EEMD is an adaptive time–frequency analysis method that is widely adopted in the atmospheric and oceanic sciences, for investigating parameters such as the variation of sea level pressure index, irregular coastal water waves, and the spatial–temporal change in air pollutants (Dätig and Schlurmann, 2004; Liu et al., 2018; Wu and Huang, 2009). The decomposition assumes that at any given time, the original data has many coexisting oscillatory modes of significantly different frequencies superimposed on each other.

The Empirical Mode Decomposition (EMD) is the predecessor of EEMD. It decomposes the time series into a finite number of basis, defined as Intrinsic Mode Functions (IMFs). IMFs represent the natural oscillatory modes embedded in the signal and work as the basis functions. The number of IMFs is decided by the length of the data (Wu and Huang, 2009). The IMFs need to satisfy the following two conditions: 1) The number of extrema and zero crossings must either be equal or differ at most by one. 2) The mean value of the upper envelope and lower envelope should be zero (Huang et al., 1998; 1999). Further details of EMD can be found in Huang et al. (1998). Here, I briefly describe the decomposition process as the following steps (Figure 4.1):

(1) Find the local maxima and minima of the original time series.

(2) Connect all the local maxima (minima) by a cubic spline line as the upper (lower) envelope.

(3) Subtract the mean value of two envelopes from the original time series.

(4) Iterate these processes until the extracted signals satisfy the two conditions of IMFs (i.e., if the two conditions are satisfied, the residue becomes the raw IMF. Otherwise, it enters the next iteration and repeats Steps 1–3 again to the residue).

After obtaining the 1st raw IMF, I subtract it from the original input signal and repeat the iteration above to extract the 2nd raw IMF and 3rd raw IMF. The number of raw IMFs, n_r, depends on the length of data: $n_r = \log_2 N_L$, where N_L is the length of the input signal (Wu and Huang, 2009). The raw IMFs are listed in order of descending frequency. Short-period oscillation tends to appear in the lower mode IMFs, whereas higher-mode IMFs mostly contain long-period signals. The dominant period of each mode is not fixed and depends on the characteristics of the input signal.

To make it more robust to noise, Wu and Huang (2009) modified the EMD by introducing the noise-assisted data analysis and proposed the EEMD. They used the mean of an ensemble of EMD trials, each consisting of the signal plus white noise of finite amplitude.

Compared with other time–frequency analysis methods like Fast Fourier transform or wavelet analysis, the most pronounced characteristic of EEMD is that the decomposition is determined by the characteristics of the original time series itself rather than a pre-determined basis (e.g., trigonometric functions). Therefore, it is an adaptive signal processing method that can be applied to nonlinear and non-stationary processes (Huang and Wu, 2008). Moreover, unlike causal filters, EEMD does not cause any change of the tsunami peak by group delay. These advantages make it suitable for real-time tsunami detection.
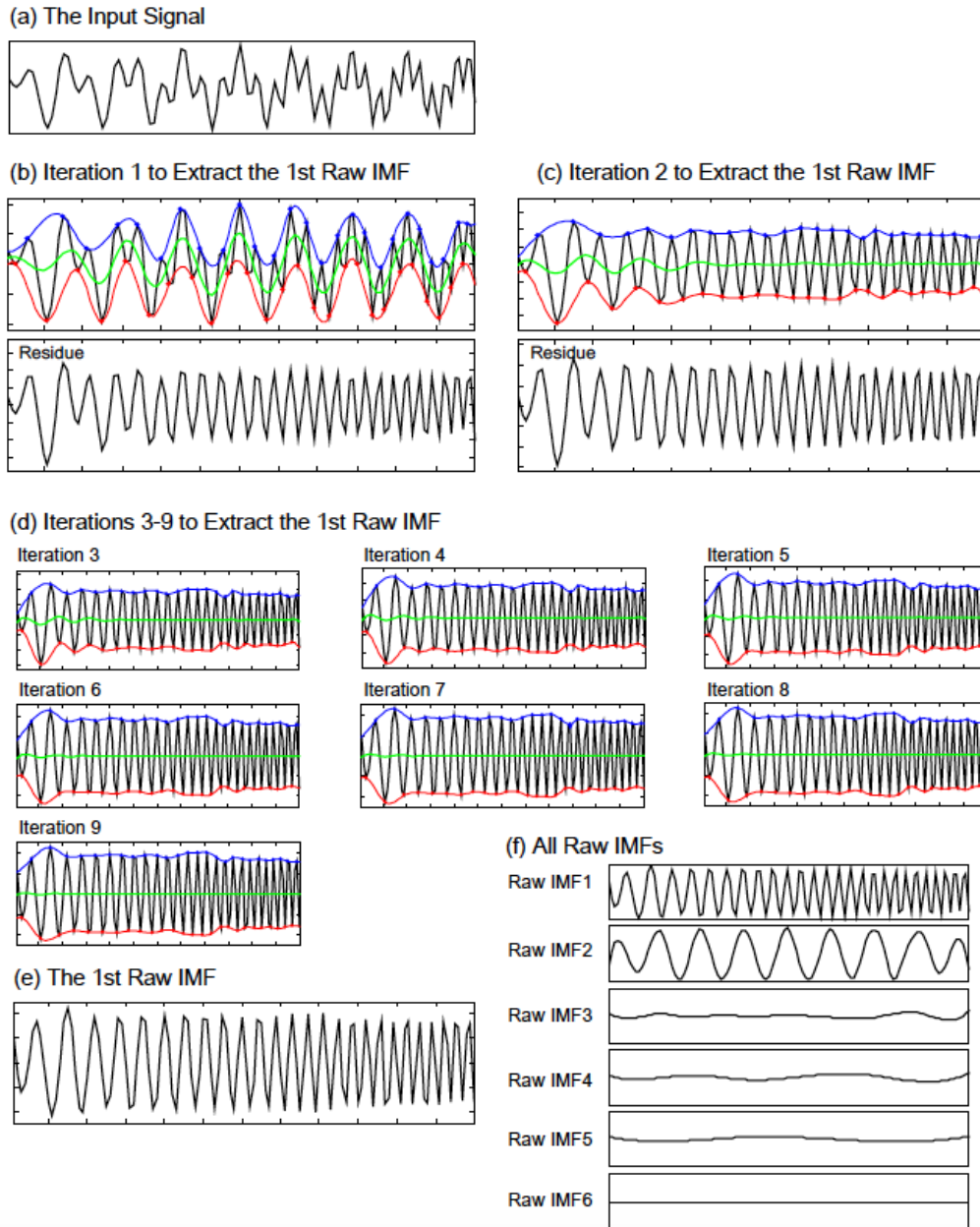
Figure 4.1. EMD of the input signals.

The decomposition of an input signal (a) is iterated nine times until the residue satisfies the two conditions of IMFs (b–e). Blue and red dots represent local maxima and minima. The blue curve represents the upper envelope that connects all the local maxima, the red curve represents the lower envelope that connects all the local minima, and the green curve represents the mean value of two envelopes. Then it becomes the 1st raw IMF (e). After that, the 1st raw IMF is subtracted from the input signal and the iteration is repeated to extract other raw IMFs (f).

To employ EEMD in a real-time tsunami detection algorithm, I follow several steps. The first step is the quality control of real-time data. I examine if there are any technical problems with the observational devices. Next, the time series in the past 3 h are selected: the segment from $y(t - 3h)$ to $y(t)$, where y is the value of tsunami height, and t is the time. I use the same segment length as NOAA's algorithm (Mofjeld, 1997). These processes can be done in real time. Then, I conduct EEMD to the 3 h segment and obtain the corresponding raw IMFs. Each of the raw IMF has different frequency characteristics. However, it is important to note that these raw IMFs are not orthogonal, and therefore some neighboring raw IMFs with similar frequency characteristics should be combined artificially (Liu et al., 2018; Wu and Huang, 2009). The input 3 h time series has 180 data points, and hence the decomposition of the 3 h time series generates seven raw IMFs, listed in descending order of frequency. I use the 1st raw IMF as IMF1, combine the 2nd to 4th raw IMFs as IMF2, and combine the 5th to 7th raw IMFs as IMF3. The combination of raw IMFs is automatic and common for different events. In Appendix A1, I demonstrate the combination of raw IMFs. Finally, I compare the IMF amplitude at the last moment with a threshold value. If the Last-Moment IMF (LM-IMF) exceeds the threshold, the algorithm will declare the tsunami arrival. Otherwise, these steps are repeated for the next interval, with the new segment from $y(t - 3h + \Delta t)$ to $y(t + \Delta t)$, where Δt is the sampling interval. The IMFs are recalculated, and the amplitudes of LM-IMF are compared again, similar to the automatic phase detection method of seismograms, STA/LTA (Allen, 1978).

4.2 Validation Test—2016 Fukushima Earthquake

To validate the real-time tsunami detection algorithm based on EEMD, I use the data of the 2016 Fukushima earthquake (M 7.4) as an example. The earthquake occurred on November 21, 2016, at 20:59:47 UTC, offshore Fukushima Prefecture, Japan. Based on the JMA earthquake catalog, the hypocenter of the earthquake was at 37.355 °N, 141.604 °E, at a depth of 25 km (Gusman et al., 2017; Figure 4.2a). The earthquake generated a moderate tsunami, with the maximum tsunami amplitude of 1.4 m measured at Sendai Port. It was the largest tsunami event since the 2011 Tohoku earthquake (Suppasri et al., 2017).
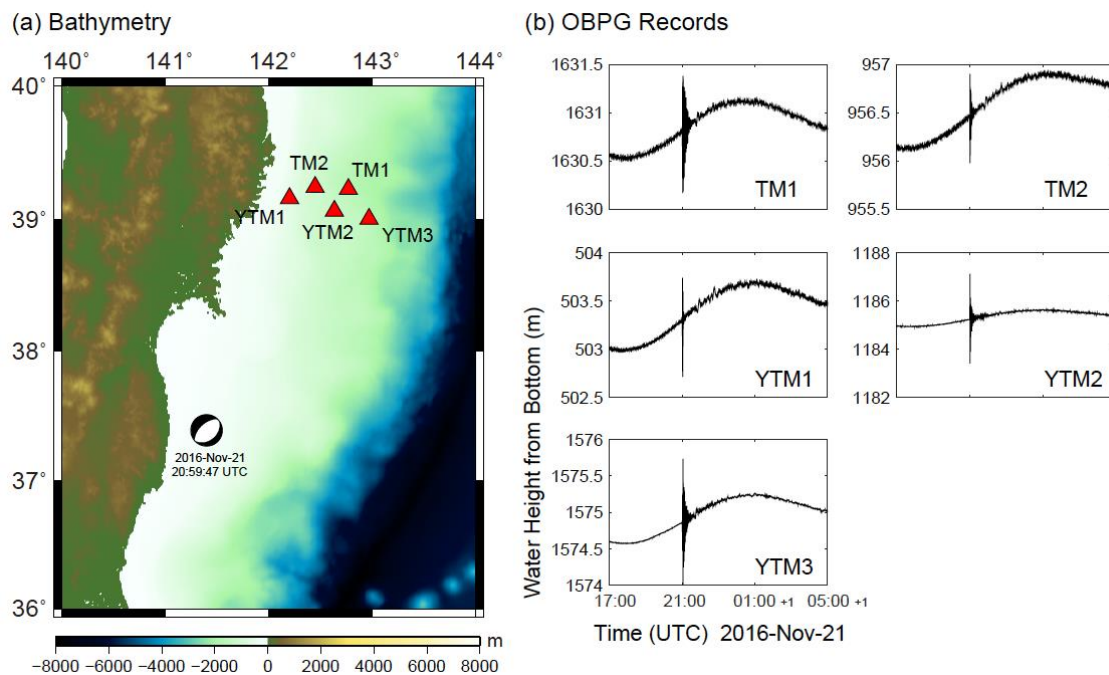
Figure 4.2. Bathymetry and the OBPG records of the 2016 Fukushima earthquake.

(a) Map of the epicenter of the 2016 Fukushima earthquake and OBPGs that recorded the tsunami. The locations of OBPGs are marked by red triangles. (b) The OBPG records of the event. These records have already been converted from water pressure to water height from the bottom. The tsunami signals are mixed up with tidal components and seismic waves.

4.2.1 Observed Tsunami Data

The tsunami is clearly recorded by OBPGs offshore Iwate Prefecture: TM1, TM2, YTM1, YTM2, and YTM3 (Figure 4.2b). The sampling rates are 1 s for the two TM gauges and 0.1 s for the three YTM gauges.

For all five OBPGs, I convert the water pressure to water height by assuming hydrostatic pressure at the OBPG depth (Gusman et al., 2017). I resample the data by calculating the 1 min average for each OBPG. Then, I select a 3 h segment for EEMD. I test the algorithm as if it is processed in "real time" and assume no priori information

about the earthquake event. I start the EEMD process approximately 1 h before the earthquake, with the first segment from 17:00 to 20:00 UTC. I obtain raw IMFs and combine the neighboring ones. Then, I compare the amplitude of LM-IMF2 with a threshold of 2.0 cm to decide whether the tsunami has arrived or not.

4.2.2 Results

In Figure 4.3a, I plot the EEMD results of TM1 at 20:00 UTC, before the earthquake. At this moment, there is no seismic wave nor tsunami waves. The IMFs have a very small amplitude. IMF1 represents high-frequency components, including the background noise and the white noise artificially added for ensembling. The IMF3 represents the tide, oceanic, and meteorological signals. Unlike the traditional detection algorithm, the new algorithm does not need a theoretical model to predict tides. Instead, these signals are removed automatically by EEMD. In Figure 4.3b, I plot the EEMD results of TM1 at 21:30 UTC. At this moment, the earthquake has already occurred, but it is before the tsunami arrival. The results suggest that the seismic waves are detected by OBPGs and represented in IMF1. IMF2, which represents the tsunami signal, still has a very small amplitude. Although the recorded amplitude is very large in the original time series because of ground motion caused by seismic waves, the algorithm does not confirm the tsunami arrival as the amplitude of IMF2 does not exceed the threshold. Therefore, it successfully avoids false tsunami alarms caused by ground motion. In Figure 4.3c, I plot the EEMD results of TM1 at 21:46 UTC. This is the first moment when the amplitude of IMF2 exceeds the negative threshold (-2 cm). Hence, I determine that the tsunami arrives at this time. In Figure 4.3d, I plot the EEMD results at 22:00 UTC, 14 min after the tsunami arrival. The dominant period bands of each IMF is adaptively determined by the characteristic of the input signal. It is different from digital filtering that has a specific period band. In this retroactive study, the dominant period of IMF1 is less than 100 s, the dominant period of IMF2 is 100–800 s, and the dominant period of IMF3 is above 1000 s.
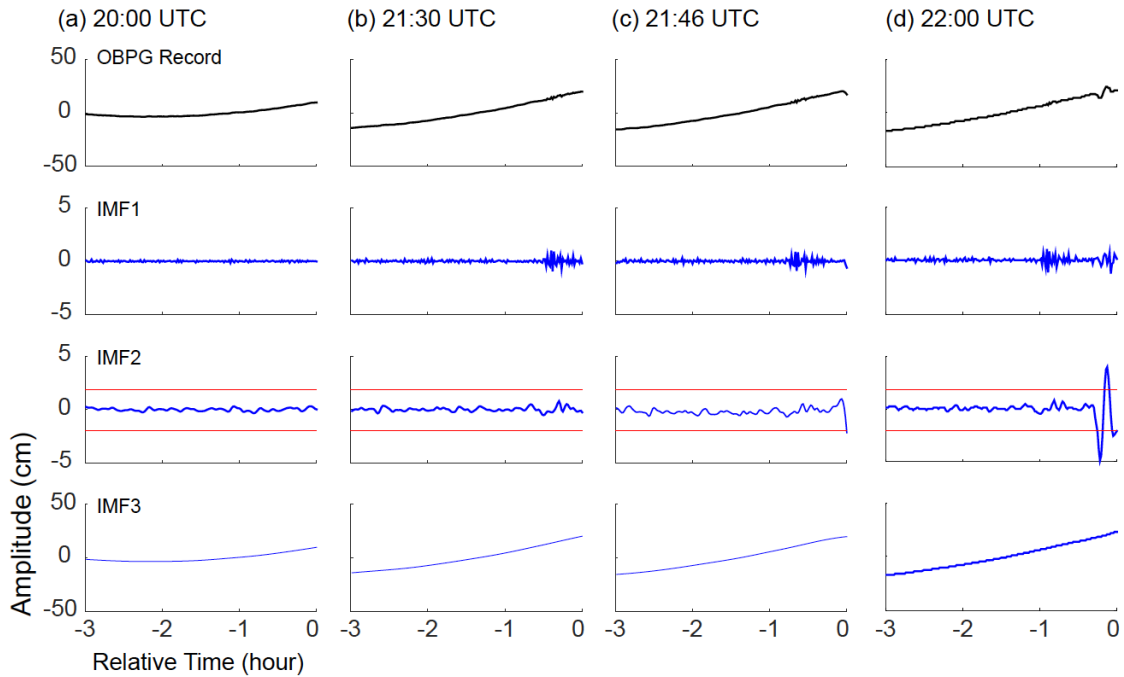
Figure 4.3. Demonstration of the real-time tsunami detection algorithm.

The OBPG records and EEMD results of TM1 at 20:00 UTC (a), 21:30 (b), 21:46 (c), and 22:00 (d) are plotted, representing the moment before an earthquake, after an earthquake, and at and after tsunami arrival, respectively. IMF2 represents the tsunami signal. In real-time operation, the EEMD results are updated continuously. When the amplitude of LM-IMF2 exceeds the threshold, tsunami arrival is confirmed.

4.2.3 Evaluation of Detection: Comparison with Post-Processed Waveforms

I also compare the real-time detected results with post-processed waveforms. The series of real-time detected results of EEMD are LM-IMF2 at different times. They are used to determine the tsunami arrival and characterize the tsunami amplitude as if in "real time". For the post-processed waveforms, I use polynomial fitting to remove the tidal components, following Gusman et al. (2017) and Heidarzadeh et al. (2020). Additionally, I use a band-pass filter to process the time series, with the corner frequencies at 0.000667 and 0.0083 Hz (1500 and 120 s). The post-processed waveforms can characterize the tsunami arrival time and amplitude accurately, but they cannot be applied in real time. In TM1, the post-processed waveform shows that the tsunami arrives at 21:45

UTC, which is slightly earlier than the real-time detected results (< 1 min; Figure 4.4). Hence, the EEMD algorithm could detect the tsunami arrival at TM1 in real time, with a very short detection delay. For comparison, I also use a casual filter within the same frequency band to process the time series as if in "real time". Though the causal filters can be applied in real time, the detection delay (~ 3 min) is larger than that of EEMD and the tsunami amplitude is underestimated.

I also plot the detected results at the other four OBPGs and compare them with the post-processed waveforms. For most OBPGs, there is only a very short detection delay. The average tsunami detection delay is less than 1 min, whereas the tsunami wave period is approximately 10 min in this event. Compared with the average detection delay of one wave period of the NOAA's algorithm (Chierici et al., 2017; Mofjeld, 1997), the algorithm can detect the tsunami arrival with a shorter delay. As for tsunami amplitude characterization, the real-time detected results of EEMD slightly underestimate the amplitude of the negative phase of the first tsunami peak, but they slightly overestimate the amplitude of the positive phase. Generally, they match quite well with the post-processed waveforms, with a mean-square error of 0.9308. Whereas, the real-time detected results of causal filter are less consistent with the post-processed waveforms, with a mean-square error of 2.2227. The amplitude of the positive phase is underestimated and there exist evident discrepancies in the following peaks.
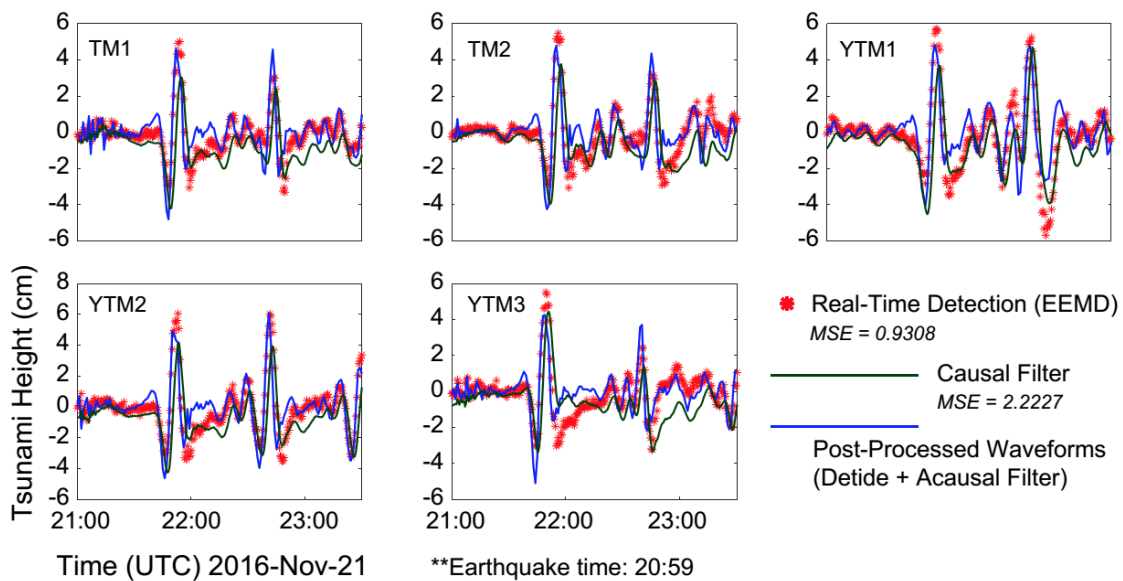

Figure 4.4. Comparison of the real-time detected results and post-processed waveforms.

The red stars represent the amplitude of LM-IMF2, calculated by EEMD. The dark green curves present the tsunami waveforms that are processed with a causal filter as if operated in "real time". The blue curves represent the tsunami waveforms that are post-processed with an acausal filter.

4.3 Discussion

4.3.1 Applications to Extreme Cases

To test the applicable range of the real-time tsunami detection algorithm, I apply it to extreme tsunami cases. First, I adopt the 2011 Tohoku earthquake (M 9.0) as an example of extremely large tsunami. The earthquake occurred at 05:46:18 on March 11, 2011 (UTC), at 38.103 °N, 142.860 °E, at a depth of 24 km, according to JMA. At the time of this extremely large event, only stations TM1 and TM2 were installed offshore of the Tohoku region. The two OBPGs recorded the first peak of the tsunami (Figure 4.5). However, at approximately 30 min after the earthquake, the devastating tsunami damaged the transmission station on land, and the two OBPGs lost connection. In the short records, the pressure data of TM1 and TM2 include the effects of seismic waves, tsunami, ocean tides, the vertical deformation of the seafloor, and atmospheric and oceanographic disturbances (Hayashi et al., 2011; Tsushima et al., 2011). After removing the high-frequency components, both records show a two-stage tsunami, an initial gradual increase for ~10 min, followed by a rapid increase within a few minutes (Fujii et al., 2011; Maeda et al., 2011; Satake et al., 2013; Tsushima et al., 2011).
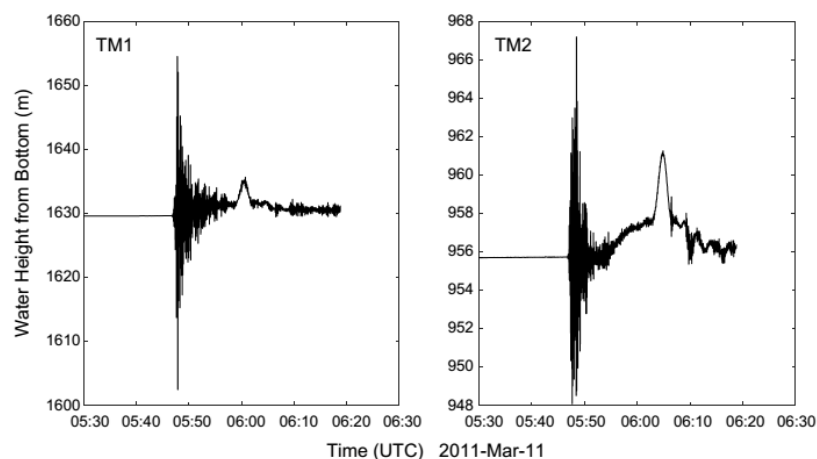
Figure 4.5. Records of stations TM1 and TM2 during the 2011 Tohoku earthquake.

Similar to the application to the 2016 Fukushima earthquake, I apply the algorithm to the records of the two OBPGs, as if in "real time." Figure 4.6 demonstrates the EEMD results at 06:16:00 UTC (i.e., approximately 30 min after the earthquake origin time). IMF1 represents the high-frequency components, including the seismic waves and the first part of the two-stage tsunami (Fujii et al., 2011; Maeda et al., 2011; Tsushima et al., 2011). EEMD determines the dominant period bands of each IMF adaptively. The strong seismic signals are partly filtered out in advance because of the data resampling (averaging over 1 min). IMF2 represents the main tsunami signals. IMF3 is the low-frequency component that represents the tidal components and the sea level increase owing to the vertical deformation of the seafloor. Though the IMF2 is affected by the long-period ground motion for this extremely large earthquake, it still characterizes the tsunami arrival time and amplitude. Therefore, it can confirm the tsunami arrival and provide information for tsunami early warning of such a large event.
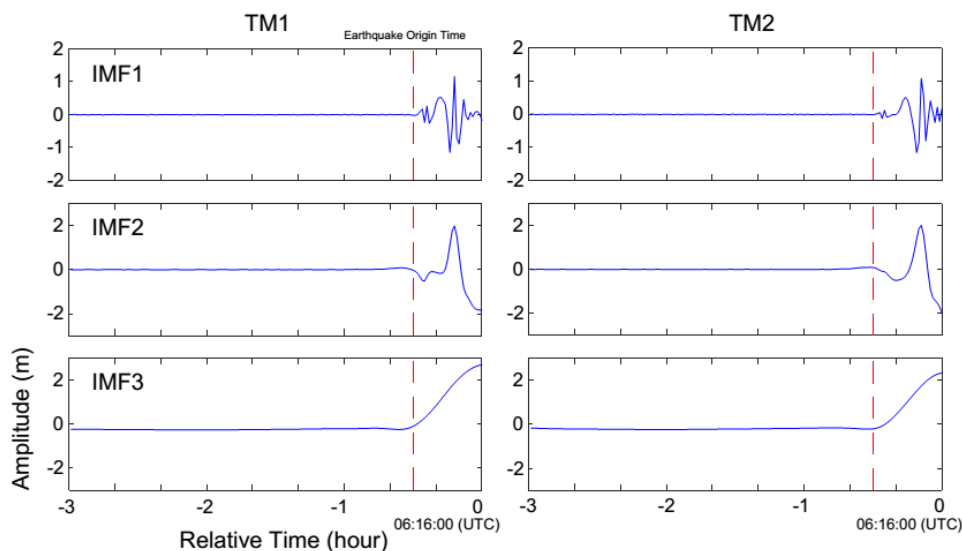

Figure 4.6. EEMD results of stations TM1 and TM2.

The EEMD results are calculated at 06:16:00 (UTC), approximately 30 min after the 2011 Tohoku earthquake origin time.

The algorithm is also tested for a micro tsunami event by considering the 1998 off Sanriku earthquake (M 6.4). The earthquake occurred off the Sanriku region at 18:18:09 on May 30, 1998 (UTC) at 39.028 ̊N, 143.847 ̊E, at a depth of 11 km according to JMA (Figure 4.7a). A micro tsunami was generated and observed by OBPGs TM1 and TM2 (Figure 4.7b). Hino et al. (2001) conducted a retroactive study and extracted the tsunami signals from the records of the OBPGs. The maximum peak-to-peak amplitude observed is approximately 1.5 cm (Hino et al., 2001), and it is in good agreement with the numerical simulation. I apply the real-time tsunami detection algorithm to the OBPG records and extract tsunami signals from the original time series.
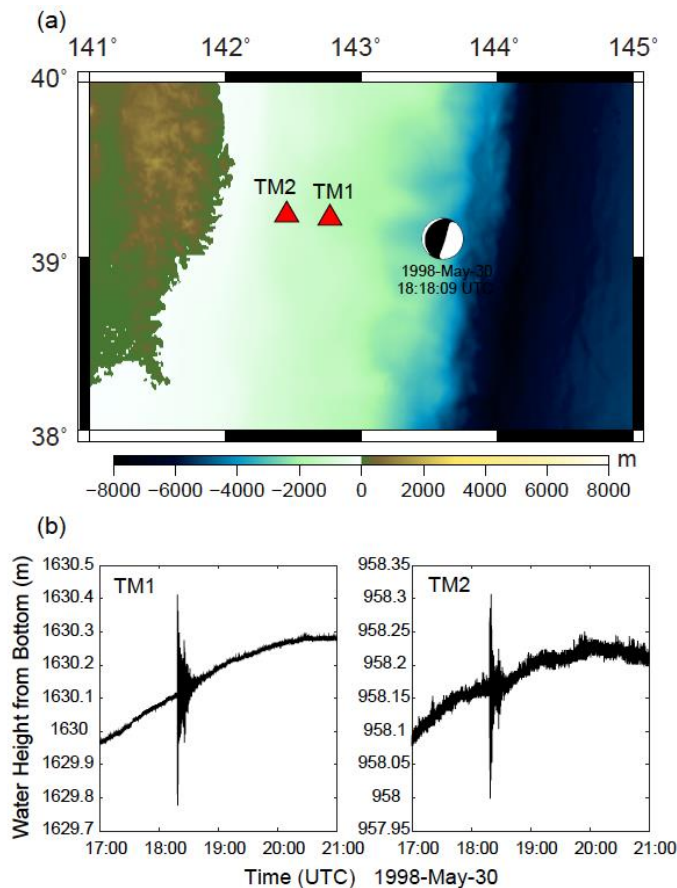

Figure 4.7. Bathymetry and the OBPG records of the 1998 off Sanriku earthquake.

(a) Map of the epicenter of the 1998 off Sanriku earthquake and OBPGs that recorded the tsunami. The locations of OBPGs are marked by red triangles. (b) The records of stations TM1 and TM2 during the 1998 off Sanriku earthquake.

Figure 4.8 demonstrates the EEMD results at 21:00:00 UTC. The IMF1 represents the high-frequency components. Seismic waves could be observed after the earthquake, but the strong seismic signals are partly filtered out because of the 1 min data resampling. IMF3 is the low-frequency component, which shows the tidal signals. IMF2 is supposed to represent the tsunami signals. However, the tsunami cannot be observed in IMF2. The absolute value of the IMF amplitude never exceeds 0.01 m. It thus does not reach the threshold of 2 cm during the event, so I could not detect the tsunami arrival by the algorithm. If I set a smaller threshold, there will be too many false alarms to give reliable tsunami detection. Therefore, for such small events with a magnitude of approximately M 6.4, which are unlikely to cause tsunami damage, the algorithm does not work for real-time tsunami detection.
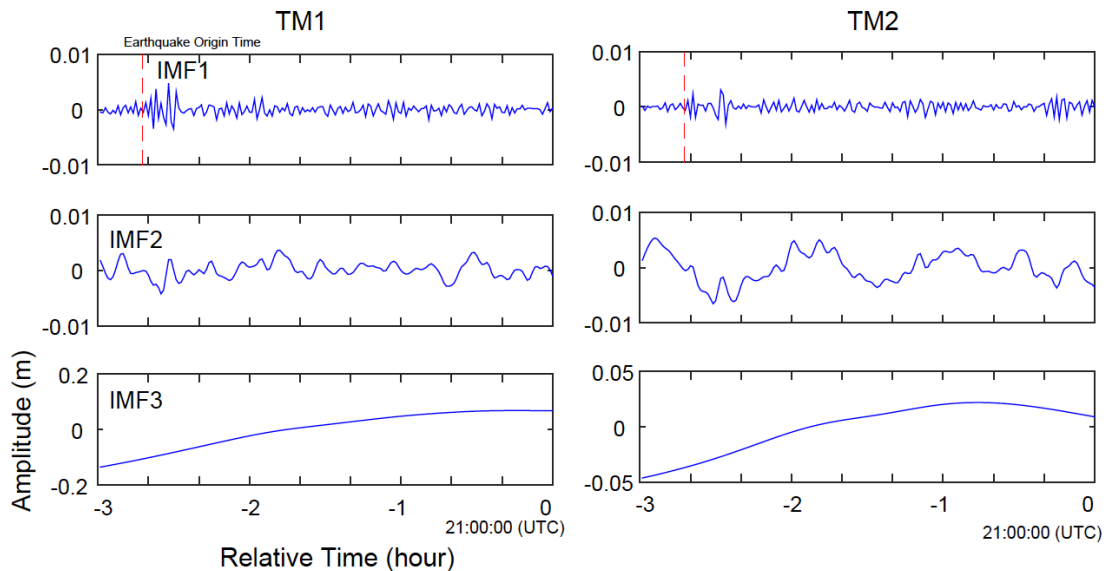
Figure 4.8. EEMD results of stations TM1 and TM2.

The EEMD results are calculated at 21:00:00 (UTC), around 2 h 30 min after the 1998 off Sanriku earthquake.

4.3.2 False Alarms and Missed Alarms

The EEMD algorithm performance is estimated by evaluating the detection probability, detection delay, and occurrence rate of false alarms (Chierici et al., 2017). It is very important to choose an appropriate threshold for the tsunami detection algorithm.

In the application to the 2016 Fukushima earthquake, I use a threshold of 2.0 cm, which is slightly smaller than the threshold of NOAA's method (3.0 cm; Mofjeld, 1997). A larger threshold decreases the occurrence rate of false alarms, but it also makes the algorithm less sensitive to moderate tsunamis.

To find a proper threshold of the real-time tsunami detection algorithm, I calculate the LM-IMF2 for one month before the 2016 Fukushima earthquake. I use the observational records of the same five OBPGs. Starting from the data of October 22, 2016, I repeat the EEMD process as if in "real-time operation" and calculate the LM-IMF2 of each trial (Figure 4.9).
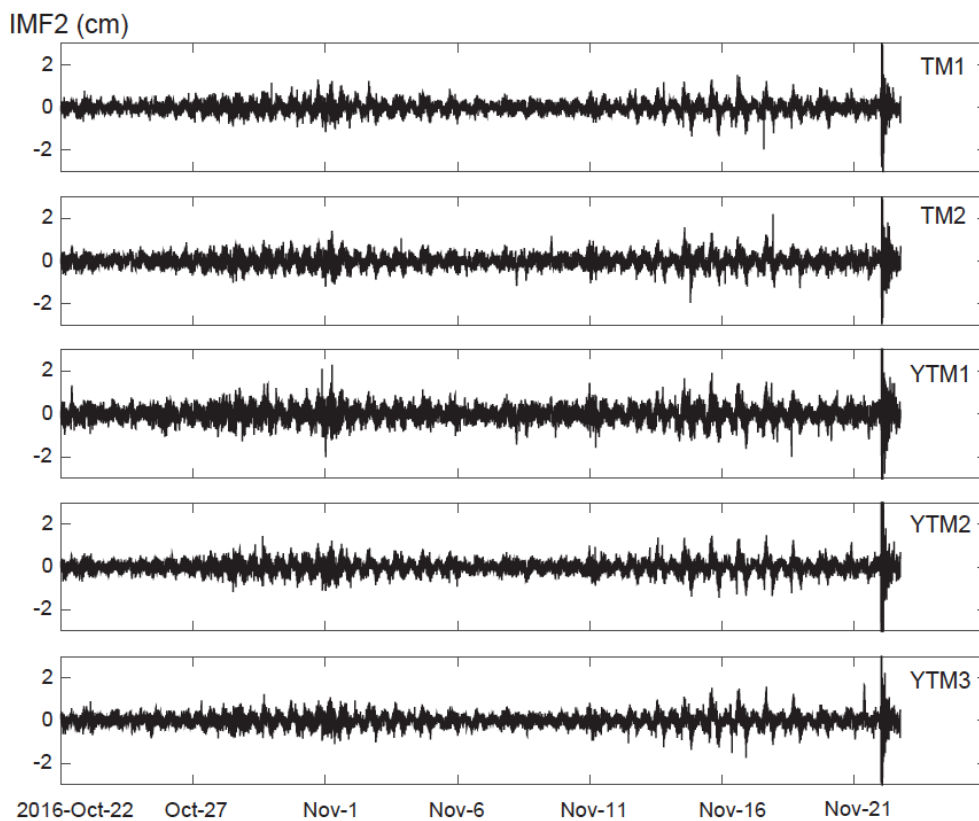

Figure 4.9. LM-IMF2 of five OBPGs for one month before the 2016 Fukushima earthquake.

During the one-month period before the earthquake, the value of LM-IMF2 has a very small variation, within the range of -3 to 3 cm. When the tsunami arrives after the

earthquake, the value of LM-IMF2 immediately increases, and the algorithm detects the tsunami. As the LM-IMF2 never exceeds the absolute value of 3 cm in the one-month records, a threshold of 3 cm will never cause a false alarm in this period. This setting is similar to the threshold of NOAA's algorithm. Though there may be some missed alarms for very small tsunamis, it could effectively detect tsunamis that could result in potential flooding. On the other hand, if a stricter threshold is set (i.e., 2 cm in this study), the possibility of missed alarms will decrease, but false alarms will increase. For example, a threshold of 2 cm will lead to false alarms: three times at YTM1 and once at TM2. The false alarms of non-tsunami perturbation may result from oceanographic or meteorological phenomena like storm surges (Bernard and Robinson, 2009; Hayashi, 2008). If their frequency characteristics are similar to tsunamis, it is likely that these signals are also extracted in IMF2, leading to a false alarm. Therefore, it is important to select a threshold according to the requirements of the early warning system.

Furthermore, stations YTM1 and TM2 are the closest to the coast, and their depths are lower than those of the other three OBPGs (Figure 4.2a). The amplitudes of tsunamis and other non-tsunami perturbations become larger according to Green's Law (Satake, 2015; Wang et al., 2012). As a result, stations with a smaller depth are more sensitive to perturbations and are more likely to give a false alarm. Therefore, it is also meaningful to consider different thresholds for stations at different depths.

Chapter 5   Real-time Tsunami Data Assimilation of S-net Pressure Gauge Records during the 2016 Fukushima Earthquake

5.1 Introduction

In previous chapters, I improve the tsunami data assimilation approach and propose a real-time tsunami detection algorithm using EEMD. In this chapter, I combine the real-time tsunami detection algorithm and data assimilation approach to forecast tsunamis on the coast.

The 2016 Fukushima earthquake is used as an example. In this event, the S-net pressure gauges recorded the sea-surface variations (Figure 5.1). Although the observation systems suffered mechanical issues at a few OBPGs, the tsunami signals at many stations were extracted by Kubota et al. (2020a). I apply the tsunami data assimilation approach to S-net pressure gauge records and forecast the tsunami waveforms at tide gauges in the Sanriku region. Although this study is still conducted retroactively, I use the real-time tsunami detection algorithm to process the raw records as if operating in real time. Then, I compare the forecasted waveforms with the real observations at tide gauges to validate the method.
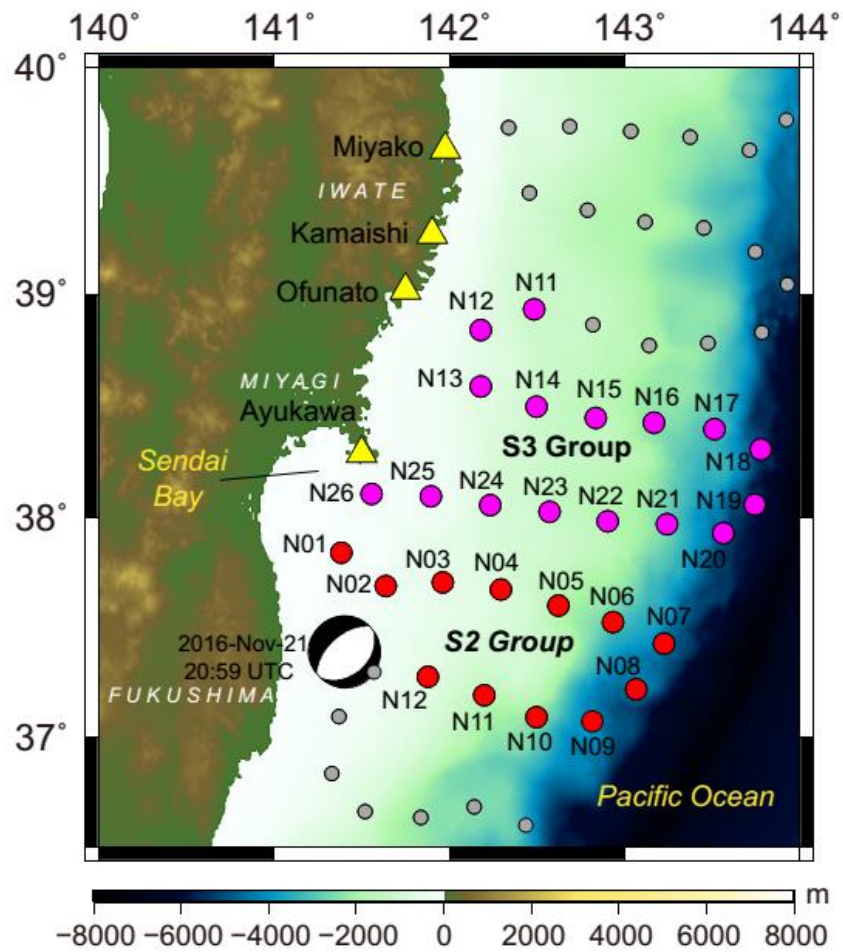
Figure 5.1. Bathymetry map around the S-net pressure gauges and tide gauges.

The red and pink circles indicate the OBPGs used for data assimilation, belonging to groups S2 and S3, respectively. The grey circles indicate other OBPGs that are not used in this study. Yellow triangles represent the tide gauges used for waveform comparison. The focal mechanism of the 2016 Fukushima earthquake is shown.

5.2 Data and Methods

5.2.1 Tsunami Records

The raw S-net pressure gauge records are obtained from NIED. The raw data are water pressure records, with a sampling rate of 0.1 s (Figure 5.2). I use the tsunami records from 28 stations: S2N01–12 and S3N11–26 (Figures 5.1 and 5.2). Here, the prefixes S2

and S3 refer to the groups of S-net stations. At each station, the records of the two pressure gauges are almost identical (Kubota et al., 2020a). I use the records of the second sensor (HP2WP) because there are more broken first sensors (HP1WP) than second sensors.

The tide gauge records are used for waveform comparison to validate the method. I use the records of four tide gauges in the Sanriku region: Miyako, Kamaishi, Ofunato, and Ayukawa (Figure 5.1). The tide gauge records of Miyako are digitized from a JMA report (https://www.jishin.go.jp/main/chousa/16dec/p16-e.htm). The records of Kamaishi are obtained from the Japan Oceanographic Data Center, with a sampling rate of 30 s. The records of Ofunato are obtained from IOC, with a sampling rate of 60 s. The digital records for Ayukawa are obtained from JMA, with a sampling rate of 60 s.

Although the S-net pressure gauges are also located to the south of the source region and recorded the 2016 tsunami waveforms, few coastal tide gauge stations to the south recorded the tsunami (Gusman et al., 2017). Therefore, I limit the assimilation area to the north of the source area, where a large number of observed tsunami waveforms are available, both offshore and on the coast.
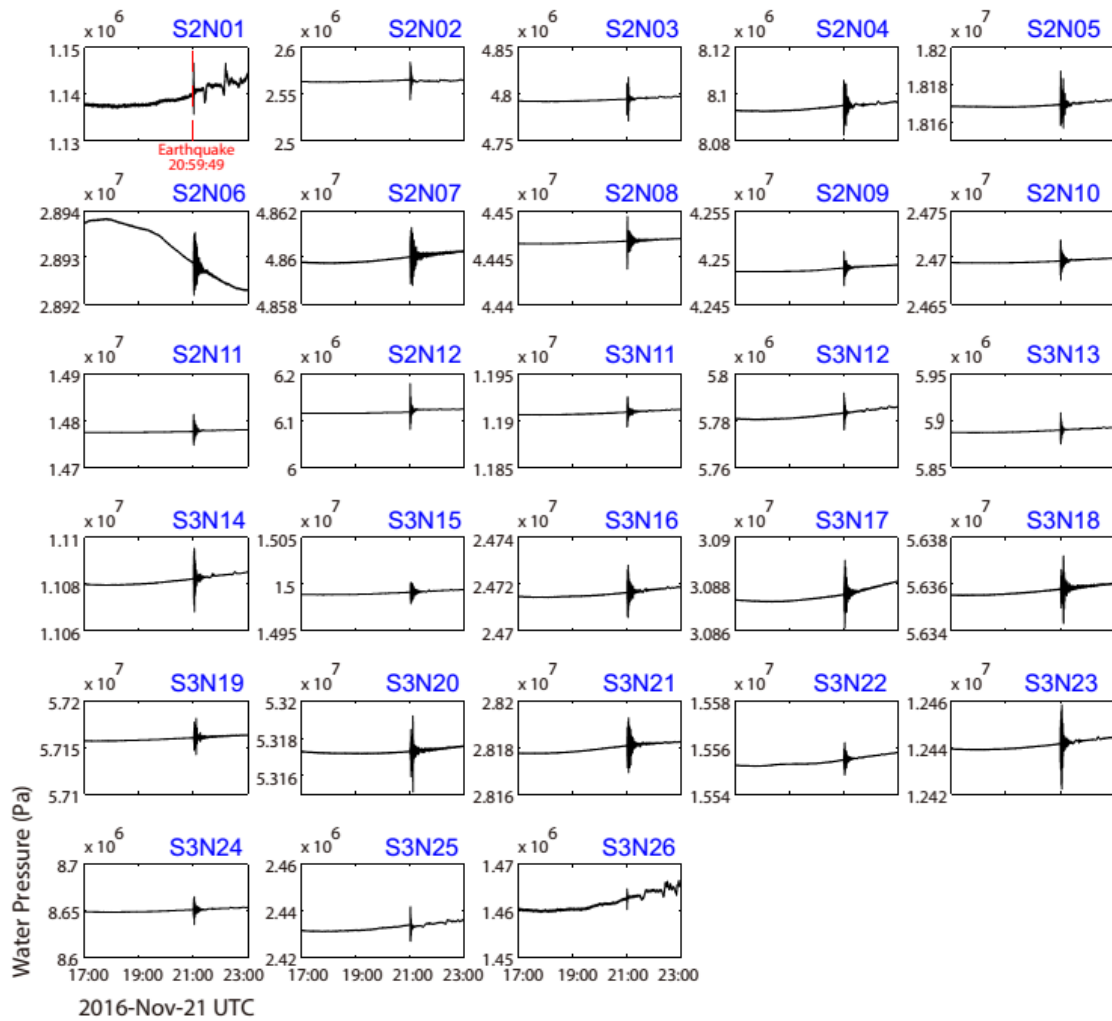
Figure 5.2. Raw records from S-net pressure gauges during the 2016 Fukushima earthquake.

Each station is equipped with two identical pressure gauge sensors to maintain redundancy. I use the records of the second sensor (HP2WP) in this study.

5.2.2 Waveform Processing

The raw S-net records contain tidal components and seismic signals. Tsunami signals should be extracted before being used as an input for data assimilation. I compare two different methods of waveform processing: 1) the traditional post-processing method, including de-tiding and filtering (Heidarzadeh et al., 2016a), and 2) the real-time tsunami

detection algorithm based on EEMD. For both methods, I first convert the water pressure records to water height using the hydrostatic equation $p = \rho gh$, where p is the water pressure, and ρ is the density of seawater (1013 kg/m^3). For some short-wavelength tsunamis, the pressure on OBPGs may deviate from the hydrostatic relation when the wavelength of the sea-bottom deformation is not much greater than the sea depth (Saito and Kubota, 2020).

The traditional post-processing method is widely adopted in retroactive studies to extract tsunami signals from OBPG sea-level records or tide gauges. First, the tidal components are removed by polynomial fitting. Then, a low-pass filter with a corner frequency of 0.01667 Hz (60 s) is applied to remove high-frequency components such as seismic waves (Kubota et al., 2020a). However, as mentioned in Section 1.5, this method is not applicable in real-time operations because of the features of digital filters. On the contrary, EEMD can extract tsunami signals in real time. Over time, the decomposition is repeated continuously, and the time series of the LM-IMF2 can be used as the input for tsunami data assimilation, following the method in Section 4.1.

5.2.3 Tsunami Data Assimilation

I apply the tsunami data assimilation approach to forecast the waveforms at tide gauges based on offshore observations. In the study of the 2016 Fukushima earthquake, the assimilation begins at 21:00:30 UTC, when the tsunami is detected at the first OBPG (S2N12). In Table 5.1, I list the tsunami detection time at each S-net station determined by the algorithm. The waveforms at tide gauges are forecasted and updated continuously within the assimilation time window, at a time interval of 1 s. I test the time windows of 20 and 35 min for tsunami early warning and compare their performance.

To improve the assimilation speed, I adopt GFTDA, in which the Green's functions between the OBPGs and the tide gauges are computed in advance. Then, the forecasted waveforms are directly synthesized by the superposition of Green's functions. For the numerical simulation of tsunamis, the linear long-wave propagation model of *JAGURS* (Baba et al., 2015) is adopted. Nested grids with four layers are used for the numerical simulation. Layer 1 covers the entire region of the OBPGs and tide gauges,

with a grid size of 18 arcsec (~ 555 m), and is derived from the bathymetry data of GEBCO_2014. Layers 2–4 are finer grids that cover the tide gauges, with a grid size of 6 arcsec (~ 185 m), 2 arcsec (~ 60 m), and 0.667 arcsec (~ 20 m), respectively. They are derived from the M7000 Digital Bathymetric Chart from the Japan Hydrographic Association. The time step for calculating Green's functions is 0.25 s. The simulation is conducted on the EIC supercomputer at the Earthquake Research Institute, the University of Tokyo.

5.3 Results

5.3.1 Extracted Tsunami Signals

Figure 5.3 shows the tsunami signals extracted by the traditional post-processing method and real-time tsunami detection algorithm. In the traditional post-processing method, the S-net raw records are processed using polynomial fitting and low-pass filtering. In the real-time detection algorithm based on EEMD, the amplitude of LM-IMF2 represents the tsunami height characterized in real-time operations. For most OBPGs (e.g., S2N05, S3N11, and S3N26), the waveforms of the real-time tsunami detection algorithm match well with the waveforms of the traditional post-processing method in terms of both tsunami height and arrival time. These results are consistent with Section 4.2.3, where the raw records of five OBPGs owned by Earthquake Research Institute, the University of Tokyo, were processed using two similar methods. Hence, I confirm that the algorithm based on EEMD can detect the tsunami arrival and can accurately characterize the amplitude in real time.

In Table 5.1, I list the detected tsunami arrival time at each S-net station. The arrival time is determined by the real-time tsunami detection algorithm based on EEMD. After decomposition, I compare the value of last-moment IMF2 (LM-IMF2) with a threshold of 3.0 cm. The time that this threshold is exceeded is recorded as the detected tsunami trigger time. The data assimilation process begins after the first OBPG is triggered. As the OBPGs are near the source region, the tsunami is detected by OBPGs soon after the earthquake happens. At most stations (e.g., S3N11, S3N14), the detected arrival time is the moment when tsunami actually arrives, indicating that the real-time

tsunami detection algorithm accurately detects the tsunami arrival.

Table 5.1. Detected tsunami arrival time at each S-net station.

The trigger time is in UTC.

Station	Trigger Time	Station	Trigger Time	Station	Trigger Time	Station	Trigger Time
S2N01	21:01:00	S2N08	21:28:00	S3N13	21:45:30	S3N20	21:12:30
S2N02	21:05:00	S2N09	21:03:00	S3N14	21:41:30	S3N21	21:35:00
S2N03	21:01:00	S2N10	21:01:00	S3N15	21:35:30	S3N22	N/A
S2N04	21:01:30	S2N11	21:01:00	S3N16	21:40:00	S3N23	21:28:00
S2N05	21:22:30	S2N12	21:00:30	S3N17	21:05:30	S3N24	21:27:00
S2N06	21:29:00	S3N11	21:48:30	S3N18	21:10:00	S3N25	21:01:30
S2N07	21:30:30	S3N12	21:50:30	S3N19	21:05:00	S3N26	21:03:30

However, at some stations close to the epicenter, there are some large drift components (e.g., S2N06) or abrupt steps (e.g., S2N11) at the earthquake origin time. According to the analysis of Kubota et al. (2020a), these drifts or steps are possibly caused by the mechanical issue of the observation system as the trend of the drifts or steps is completely identical in the pair of pressure sensors (i.e., HP1WP, HP2WP) at each S-net station. Similar steps were also found in the records of a much smaller tsunami (< 1 cm) caused by the 2016 off Sanriku earthquake (Mw 6.0; Kubota et al., 2020b). The traditional post-processing method cannot completely remove these unexpected signals, and they also appear in the waveforms of the real-time detection algorithm at some stations (Figure 5.3). The detected arrival time is affected by such unexpected signals at these stations, which may cause false tsunami alarms. Nevertheless, the real-time tsunami detection algorithm accurately characterizes the waveforms after the earthquake origin time, especially that of the first tsunami peak, which is very important for data assimilation. In this study, the waveforms processed with the two methods are used separately as the input for tsunami data assimilation.
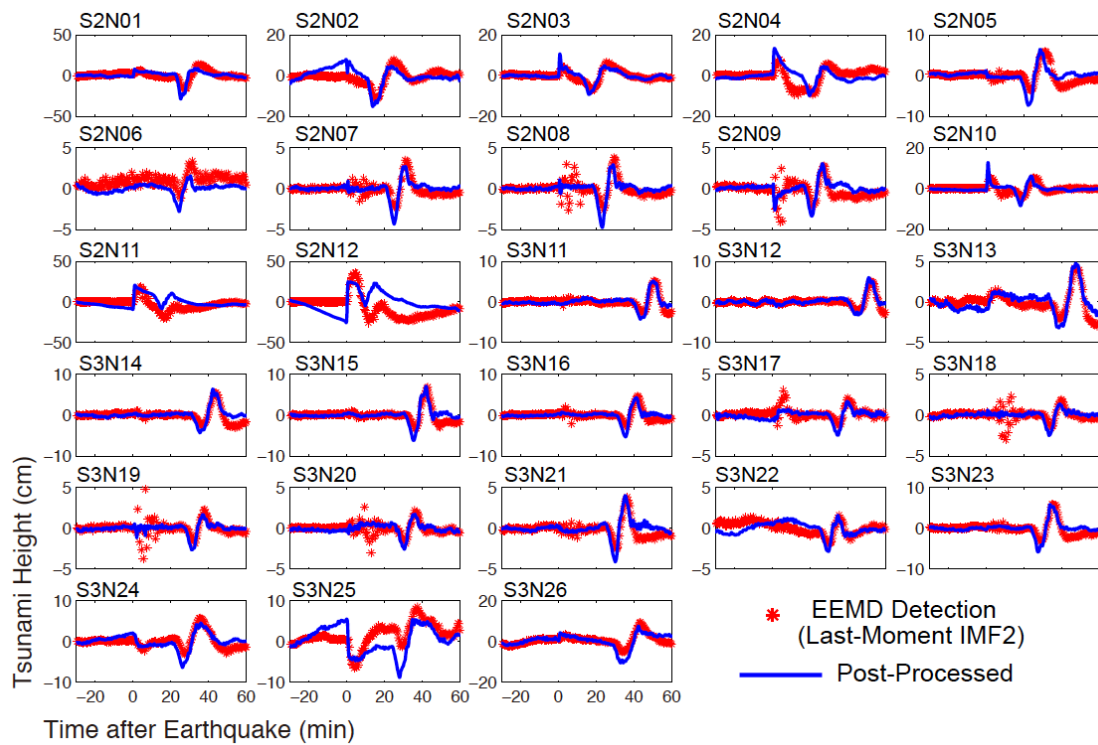
Figure 5.3. Tsunami signals extracted by the traditional post-processing method (blue curves) and the real-time tsunami detection algorithm based on EEMD (red stars).

5.3.2 Forecasted Tsunami Waveforms at Tide Gauges

Figure 5.4 shows the comparison between the forecasted tsunami waveforms and the real observations. The waveforms are forecasted by tsunami data assimilation for time windows of 20 min (a) and 35 min (b). The actual tsunami arrives first at Ayukawa (57 min after the earthquake), and then at Ofunato (68 min), Kamaishi (77 min), and Miyako (80 min). At most tide gauges, the waveforms forecasted using the traditional post-processing method and real-time tsunami detection algorithm match well with real observations.

At Miyako, the data assimilation approach forecasts the arrival time of the first tsunami peak accurately. The observed first-peak amplitude is 26 cm, though the assimilation result with a 20 min time window may underestimate the first-peak amplitude. The 35 min time window has a better performance in forecasting the amplitude,

especially for the results with input time series obtained using the real-time detection algorithm. Its forecasted first-peak amplitude is 18 cm. Moreover, the forecasted waveform is also consistent with the real observation in another peak of the later phase (102 min). At Kamaishi, the forecasted waveforms generally match real observations, but the amplitude of the first tsunami peak is overestimated. This issue is discussed in Section 5.4.

At Ofunato, the forecasted waveforms match quite well with the real observations regarding the arrival time and amplitude. Both input time series lead to accurate forecasting by data assimilation. For example, for a 35 min time window, the observed amplitude of the first tsunami peak is 26 cm, and the forecasted amplitudes are 29 cm (by post-processing method) and 23 cm (by real-time detection algorithm). The forecast could be accurately made more than 30 min before the tsunami arrival. Nevertheless, the forecasted waveforms do not match well with the real observations at Ayukawa. The tsunami arrival time is not well predicted, and the forecasted results overestimate the amplitude of the first tsunami peak.

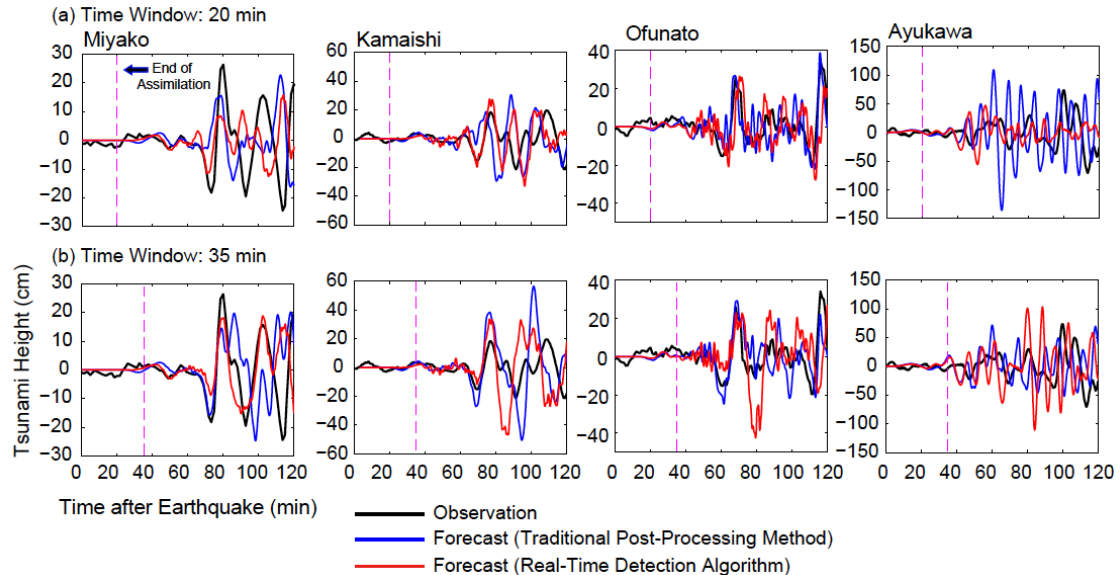**Figure 5.4.** Comparison of observed and forecasted tsunami waveforms.

The assimilation time windows are 20 min (a) and 35 min (b). The black curves indicate the real observations. The blue curves represent the results forecasted by data assimilation with input time series obtained by the traditional post-processing method. The red curves represent the results forecasted by data assimilation with the input time series obtained

using the real-time tsunami detection algorithm.

To evaluate the forecast accuracy quantitatively, I use Equation (2.7). For a 20 min time window, the scores are 47% (by post-processing method) and 41% (by real-time detection algorithm) when considering all stations. For a 35 min time window, the scores are 60% and 74%, respectively. If I do not consider the Kamaishi and Ayukawa stations, the scores are 91% (by post-processing method) and 77% (by real-time detection algorithm) for a 20 min time window and they become 89% and 94%, respectively, for a 35 min time window.

5.4 Discussion

5.4.1 Adoption of a Real-time Detection Algorithm

Compared with previous studies on tsunami data assimilation (e.g., Sheehan et al., 2019; Wang et al., 2019b), the most notable feature of this work is the adoption of a real-time tsunami detection algorithm for waveform processing. I process the raw OBPG records as if operating in real time and use the processed waveforms as the input for data assimilation. The assimilation results with input obtained by the real-time tsunami detection algorithm perform well in tsunami forecasting. Hence, my experiment proves the feasibility of establishing a real-time tsunami early warning system based on the data assimilation approach. Although there are some unexpected signals at the earthquake origin time, as shown in Section 5.3.1, they do not significantly affect the performance of tsunami data assimilation as data assimilation is a self-correcting process. Over time, the effects of such unexpected signals are gradually corrected by assimilating more input data into the model, and the forecasted wavefield matches the real observations more accurately (Maeda et al., 2015).

5.4.2 Results of the Synthetic Experiment

To further investigate the forecasting performance, I conduct a synthetic

experiment using numerical simulations. Data assimilation for synthetic data can facilitate the separation of two factors of assimilation performance: imperfect modeling and insufficient observations. If there is a discrepancy between the observed and synthetic waveforms, and if the data assimilation performs better for synthetic data, the modeling is not perfect. On the other hand, if the observed and synthetic waveforms are similar, and the data assimilation does not perform well for synthetic data, then the offshore observations are insufficient for tsunami data assimilation.

I adopt the source model of Gusman et al. (2017), which is inverted from OBPG and tide gauge records. The bathymetry files are the same as those used to compute Green's functions. First, I run the tsunami propagation model *JAGURS* to compute the synthetic waveforms at S-net OBPGs and tide gauges. I compare the synthetic waveforms at four tide gauges with the real observations in Figure 5.5. Next, I use the synthetic waveforms at the OBPGs as the input for tsunami data assimilation and forecast the waveforms at tide gauges. The forecasted waveforms and the synthetic results at tide gauges are compared in Figure 5.6.
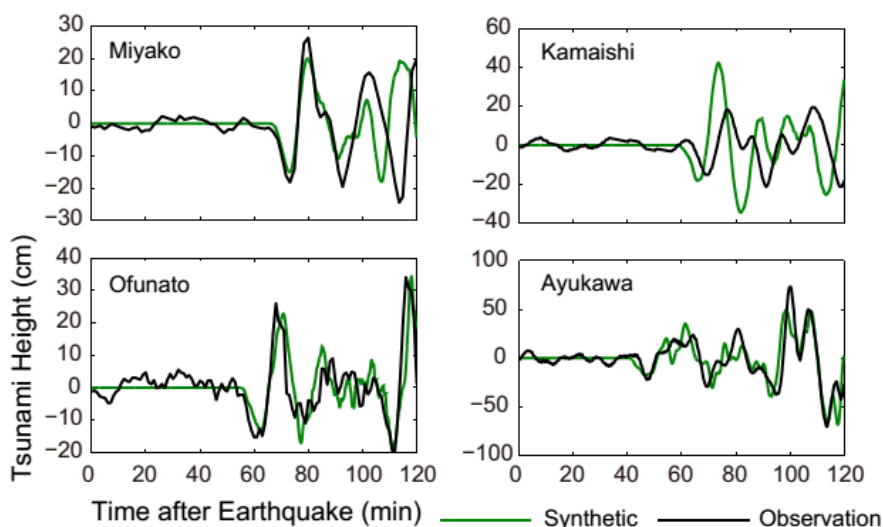

Figure 5.5. Comparison of synthetic tsunami waveforms and real observations.

The green curves represent synthetic tsunami waveforms, and the black curves represent real observations. The synthetic waveforms are generally consistent with the real observations at Miyako, Ofunato, and Ayukawa, but overestimate the tsunami amplitude at Kamaishi.

In Figure 5.5, the synthetic first-peak amplitude at Kamaishi (42 cm) is much higher than the real observation (18 cm). However, the data assimilation with synthetic data works perfectly at Kamaishi (Figure 5.6). This indicates that the discrepancies between the forecasted and observed waveforms at Kamaishi may have resulted from imperfect modeling because of bathymetry data and the source model. The Kamaishi tide gauge is located deep inside the harbor. Local bathymetry (e.g., breakwater and wave attenuator) has a substantial effect on tsunami simulation, difficult to model even with the 0.667 arcsec grid (~ 20 m). Green's functions between the OBPGs and the Kamaishi tide gauge are also computed using these grid files; hence, the forecasted waveform synthesized by Green's functions overestimates the amplitude systematically. For a 20 min time window, there are fewer offshore data used for assimilation. This counteracts such systematic errors. For a 35 min time window, there are sufficient offshore data to make the systematic error evident. If we could obtain finer grids that represent the detailed bathymetry around the Kamaishi tide gauge, the accuracy of data assimilation could be improved. Furthermore, in my study, a Gaussian-distributed covariance is assumed for the correlation functions of both observation and computational errors when calculating the weight matrix W (Equation 1.28). If the parameters of correlation functions are adjusted according to different data quality, the accuracy may also be improved.
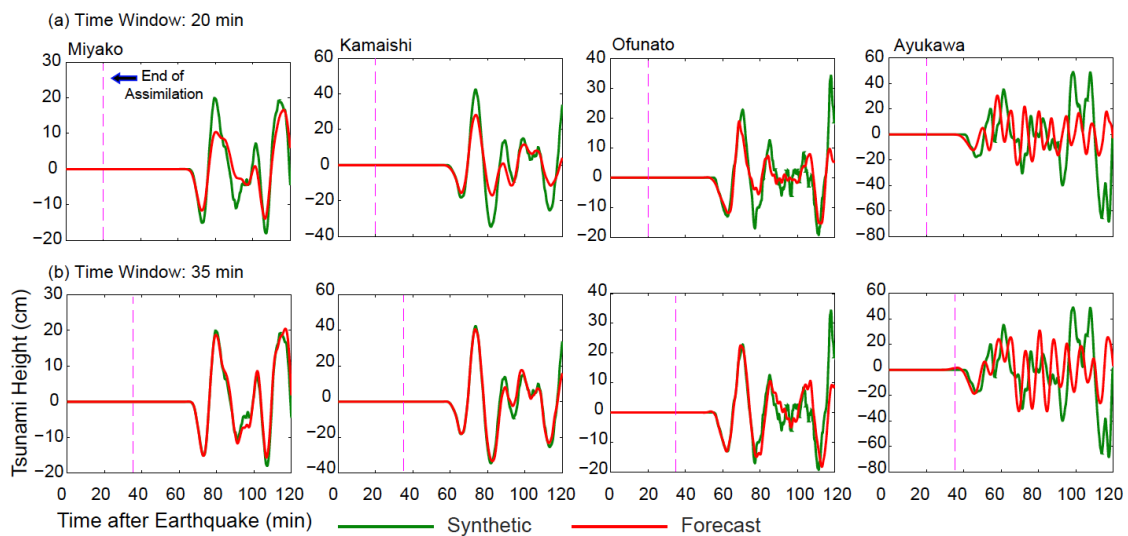

Figure 5.6. Comparison of synthetic and forecasted tsunami waveforms.

The green curves represent synthetic tsunami waveforms, and the red curves represent forecasted tsunami waveforms. The assimilation time windows are 20 min (a) and 35 min (b). The synthetic waveforms at tide gauges are used as the "observations" for comparison.

However, the synthetic experiment still has unsatisfactory assimilation results at the Ayukawa tide gauge. The first tsunami peak and the following waveforms are not well predicted at both time windows (Figure 5.6). The location of Ayukawa is quite close to the epicenter (Figure 5.1). There are very few OBPGs between the epicenter and the tide gauge. Although data assimilation does not require source information, it requires sufficient offshore data to reconstruct the tsunami wavefield and facilitate the forecasting. Hence, the discrepancies at Ayukawa are caused by insufficient offshore observations. The installation of some OBPGs inside Sendai Bay is necessary for improving the forecasting ability.

5.4.3 Data Assimilation at Ayukawa with Assumed OBPGs

To prove the effectiveness of additional OBPGs, I add two assumed OBPGs inside Sendai Bay and use data assimilation to forecast the tsunami waveforms at the Ayukawa tide gauge (Figure 5.7). The locations of two assumed OBPGs are V01 (38.20 °N, 141.25 °E) and V02 (38.00 °N, 141.25 °E). The synthetic waveforms at the OBPGs are used as the input for tsunami data assimilation. The other conditions are the same as those in the synthetic experiment.
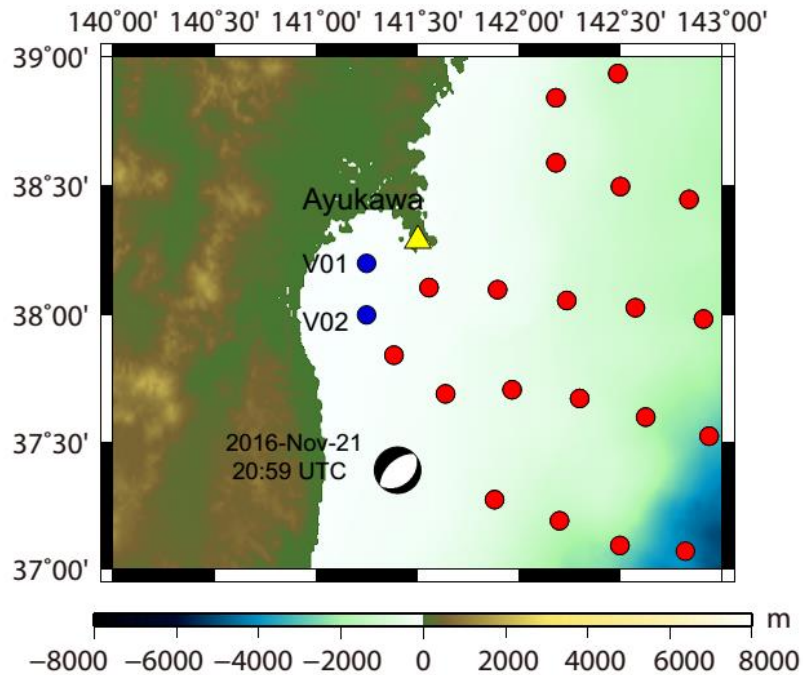
Figure 5.7. Bathymetry map around Ayukawa tide gauge.

The red circles indicate the real S-net OBPGs used for data assimilation. The blue circles indicate two hypothetical OBPGs: V01 and V02. The yellow triangle represents the Ayukawa tide gauge. The focal mechanism of the 2016 Fukushima earthquake is shown.

      Figure 5.8 shows the comparison between the synthetic and forecasted tsunami waveforms. Here, the synthetic waveforms at tide gauges are used as the "observations" for comparison. I compare the forecasted results with and without two hypothetical OBPGs, for assimilation time windows of 20 and 35 min. For a 20 min assimilation time window, additional OBPGs inside Sendai Bay do not significantly improve the forecast accuracy, though the prediction of the first tsunami peak is better. However, for a 35 min assimilation window, the results with two hypothetical OBPGs have better performance. The first tsunami peak can be predicted better, and it works particularly well for the following waveform at 40–90 min after the earthquake. Hence, it is confirmed that installing OBPGs inside Sendai Bay is necessary to improve forecasting.
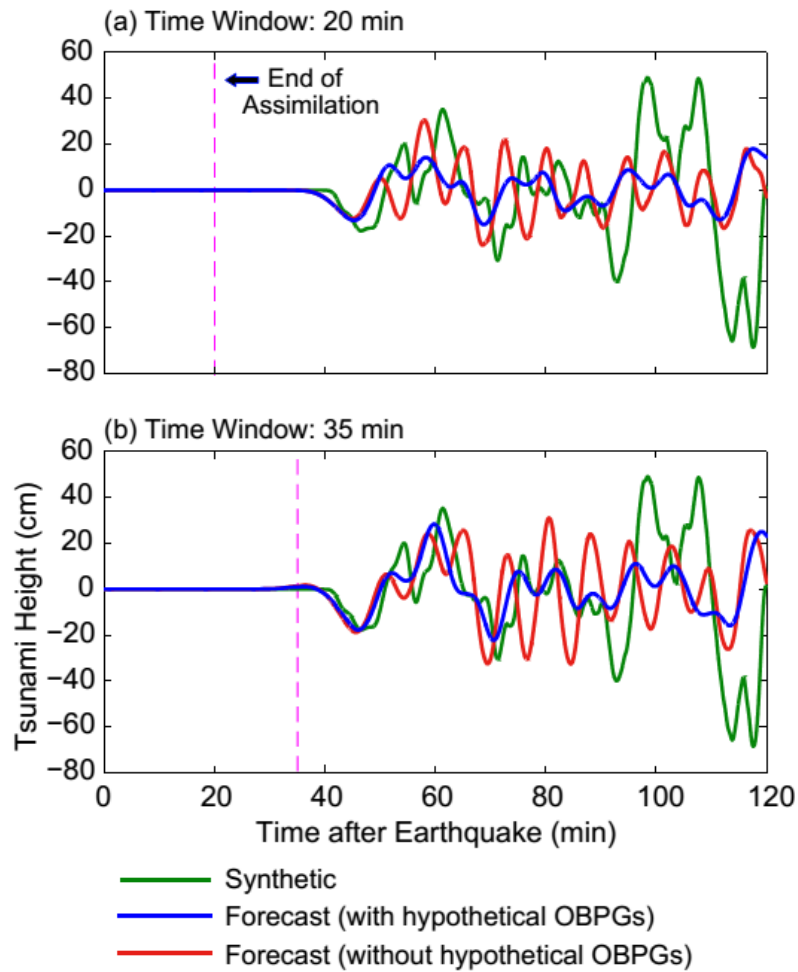
Figure 5.8. Comparison between the synthetic and forecasted tsunami waveforms.

The green curves represent synthetic tsunami waveforms, the blue curves represent forecasted tsunami waveforms with hypothetical OBPGs, and the red curves represent forecasted tsunami waveforms without hypothetical OBPGs at the Ayukawa tide gauge. The assimilation time windows are 20 min (a) and 35 min (b). The synthetic waveform at Ayukawa is used as the "observation" for comparison.

Chapter 6   Tsunami Early Warning System Using Data Assimilation of Offshore Data

6.1 Practical Implementation

In Chapter 5, the retroactive study of the 2016 Fukushima earthquake confirmed the applicability of real-time tsunami forecasting independent of the source. Therefore, a tsunami early warning system can be designed for Japan based on data assimilation.

The main target regions are the Tohoku and Nankai regions. These two regions face potential tsunami hazards, and there are dense OBPG networks with real-time data transmission: DONET and S-net. These networks represent reliable hardware for tsunami early warning. In general, four steps need to be taken to establish the tsunami early warning system using data assimilation of offshore data.

First, Green's functions between OBPGs and the coastal points (PoIs) are calculated and stored in advance. GFTDA is adopted to improve the efficiency of the system. Although the calculation of Green's functions is time-consuming, it does not affect the speed of the assimilation process.

Next, the real-time tsunami detection algorithm based on EEMD is used. It can detect the tsunami arrival without a priori earthquake information, and is thus suitable for tsunami data assimilation. The algorithm is performed continuously, and the LM-IMF2 is compared with the pre-defined threshold at each moment. Before the earthquake (or landslide or volcano eruption), the LM-IMF2 should remain below the threshold, and the system is in the *General Mode* (Figure 6.1). The assimilation process does not begin. After an earthquake (or landslide or volcano eruption) event, as long as the tsunami does not arrive at OBPGs, the system remains in the General Mode.

When a tsunami arrives at OBPGs, the LM-IMF2 calculated by the EEMD algorithm will exceed the threshold. The early warning system is triggered and enters the

Assimilation Mode. The pre-calculated Green's functions are superposed by matrix manipulation to predict the waveforms at the coast (Figure 6.1). As time passes, more offshore tsunami data is observed and assimilated. The forecasted waveforms are updated continuously. Based on the forecasted results, we can decide whether to issue a tsunami warning or tsunami advisory to specific regions.

Finally, after the tsunami arrives at the coast, the performance of the tsunami early warning system is evaluated by comparing the forecasted results with real observations.
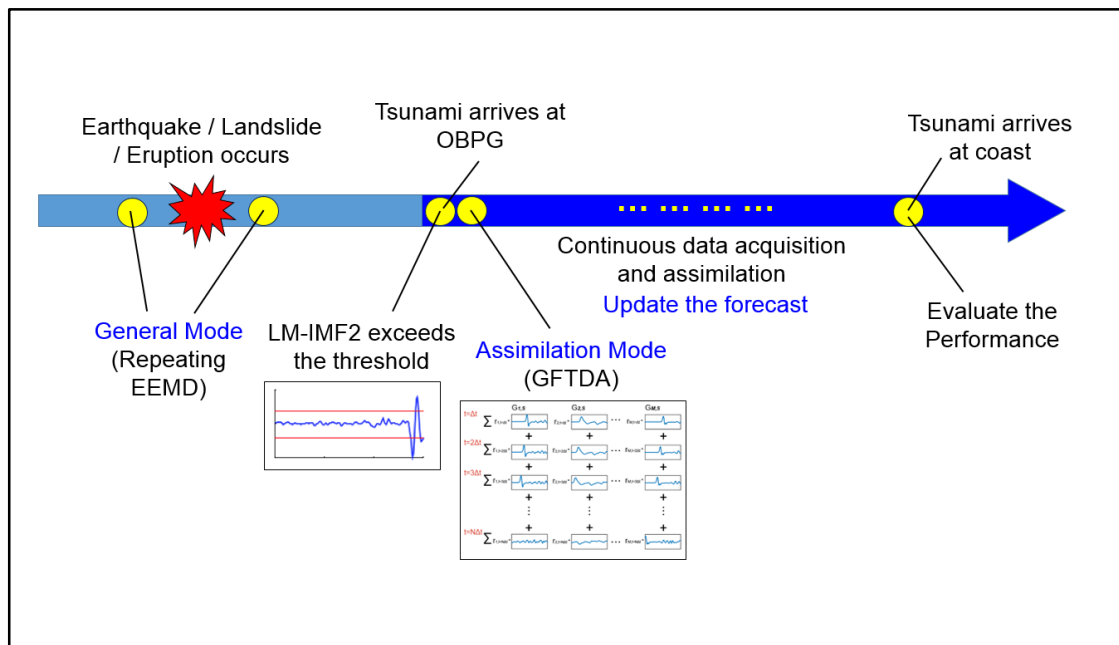
Figure 6.1. Flow chart for the proposed tsunami early warning system based on the data assimilation approach.

To summarize, the tsunami early warning system based on the data assimilation approach can provide early warning for coastal regions, regardless of the tsunami source. However, it is important to note that the goal of such a system is not to replace the traditional tsunami early warning system based on seismic observations (Duputel et al., 2011). Instead, it can be an important supplement to traditional early warning systems. If a tsunamigenic earthquake occurs, the traditional method can take advantage of the rapid

availability of seismic data and make an initial warning quickly. When OBPGs detect the tsunami, the new system can start the data assimilation process and update the tsunami forecast by assimilating the offshore observations continuously. Additionally, it can also help with forecasting methods based on tsunami waveform inversion, such as the tFISH method operated by JMA (Tsushima et al., 2011). Using the EEMD algorithm, the tsunami waveforms for inversion could be extracted directly without the need of a theoretical tide model, and the tFISH method can become more efficient.

Although this system is designed for Japan, it can be used as a reference in other tsunamigenic zones, especially in regions with tsunami events caused by non-earthquake origins like landslides or volcanoes. For example, the collapse of the Anak Krakatau Volcano in December 2018 generated a tsunami in the Sunda Strait, Indonesia, with a maximum run-up height of 13.5 m (Muhari et al., 2019). And the September 2018 Palu tsunami is widely believed to be of a dual landslide–earthquake source (Carvajal et al., 2019; Heidarzadeh et al., 2018). Traditional tsunami early warning systems have difficulties in forecasting such non-earthquake tsunamis. To the contrary, the early warning system based on data assimilation does not require the construction of seismic/tsunami source model, nor even the detection of ground motion. Thus, it is very effective in non-seismic tsunamis. Additionally, the method of tsunami data assimilation with virtual stations can reduce the engineering cost of installing OBPGs in these regions.

6.2 Future Improvements

In the future, the tsunami early warning system based on data assimilation can be further improved by adopting deep learning techniques.

As mentioned in Chapter 4, the EEMD algorithm can detect the tsunami signals by comparing the LM-IMF2 with a pre-defined threshold. However, the threshold is set artificially according to experience. A higher threshold can cause more missed alarms, and a lower threshold increases the rate of false alarms. It is challenging to select an appropriate threshold. Therefore, we could use the deep learning technique to improve the real-time tsunami detection algorithm and to avoid setting a threshold artificially. Deep learning greatly improves the development of neural networks and has been widely

used in earth science (Kuyuk et al., 2018; Zhang et al., 2018). After being trained, the deep learning model can detect the tsunami arrival and characterize the amplitude automatically instead of artificially setting a threshold for comparison. With the help of deep learning, the input for data assimilation can be more accurate, which would lead to better forecasts of the tsunami early warning system.

Chapter 7   Summary

In this thesis, I improve the tsunami data assimilation approach and propose a real-time tsunami detection algorithm.

First, I propose GFTDA to increase the assimilation speed. Green's functions between the offshore observational stations and coastal points are calculated and stored in advance. During the assimilation process, tsunami waveforms are forecasted by the superposition of Green's functions without calculating the tsunami wavefield of the entire region. Hence, it greatly reduces the real-time tsunami forecasting time and enables the use of more complicated linear models. The application to the 2012 Haida Gwaii earthquake reveals that GFTDA achieves the same accuracy as the previous data assimilation approach while reducing the time required to issue a valid tsunami warning. The application to the 2004 off the Kii Peninsula earthquake shows that tsunami data assimilation based on the linear DSP model could predict the tsunami arrival time more accurately than the LLW model. It is also successfully applied to the records of the real-time observation network of the 2015 Torishima volcanic tsunami earthquake. The relationship between the accuracy and station number is discussed. Depending on the requirement of forecast accuracy, more attention should be paid to some specific observational stations that play an important role in tsunami data assimilation.

Second, I introduce the concept of virtual stations to conduct tsunami data assimilation for regions with sparse observations. I produce artificial waveforms at virtual stations by interpolating real data of neighboring real stations. I demonstrate the use of this method for the synthetic 2004 Sumatra–Andaman earthquake. The tsunami waveforms at the coasts of India, Sri Lanka, and the Maldives can be forecasted with greater than 70% accuracy. The application of this method to the 2009 Dusky Sound, New Zealand earthquake suggests that the method overcomes the problem of insufficient observations and improves the accuracy from 52.5% to 83.8%. Moreover, it can forecast the far-field tsunamis in the east Pacific generated by the earthquakes in the Chilean subduction zone. The DART tsunameters along the Chilean subduction zone, together with virtual stations, can form a dense network with a substantial spatial coverage. The

application of this method to the 2015 Illapel earthquake achieves an accuracy of 87.5% in the east Pacific. The adoption of virtual stations can reduce the investment for tsunami disaster mitigation.

Third, I propose a real-time tsunami detection algorithm based on EEMD to detect tsunamis without tidal prediction. In the application of the method to the 2016 Fukushima earthquake (M 7.4), the algorithm can detect a tsunami in real time with a detection delay of less than 1 min (i.e., less than 1/10 of the tsunami wave period). The algorithm could also detect the tsunami signals for the extremely large tsunami from the 2011 Tohoku earthquake (M 9.0) but fails to detect a micro-tsunami from the 1998 off Sanriku earthquake (M 6.4). Whereas the Tohoku earthquake generated a devastating tsunami, the off Sanriku earthquake did not. Hence, the algorithm can greatly improve the efficiency of tsunami early warning system, especially when OBPGs are located close to the coasts of interest. Although there are slight differences between the amplitudes of real-time detected results and the post-processed waveforms, the algorithm can sufficiently determine the tsunami arrival, especially for tsunamis with the potential for substantial inundation. Additionally, the algorithm does not require any source information, and can thus be applied to the tsunami data assimilation approach for providing an early warning for non-earthquake tsunamis.

Finally, I combine the tsunami data assimilation approach with the real-time tsunami detection algorithm. I assimilate the S-net pressure gauge records of the 2016 Fukushima earthquake to forecast the tsunami waveforms at tide gauges in the Sanriku region. To process the raw records for assimilation, I adopt the real-time tsunami detection algorithm based on EEMD to imitate real-time operations. I also compare this method with the traditional post-processing method. The assimilation results of both input time series forecast the tsunami waveforms accurately. The forecast scores of the post-processing method and real-time detection algorithm are 60% and 74%, respectively, for a 35 min assimilation time window; these values improve to 89% and 94%, respectively, on neglecting the stations with imperfect modeling or insufficient offshore observations. Therefore, the combination facilitates a satisfactory tsunami forecast and enables the establishment of a real-time tsunami early warning system. I design a tsunami early warning system based on the data assimilation approach for Japan. It can forecast the tsunami independent of the source, and it can be an important supplement of the

traditional tsunami early warning system. Apart from tsunami early warning, in the future, the data assimilation approach can be applied to understand the tsunami propagation processes. The tsunami source characteristics can also be studied through wavefield reconstructing. For example, after an earthquake/landslide occurs, data assimilation may help reconstruct the tsunami wavefield by assimilating offshore data. Then, we can use tsunami inversion method to estimate the tsunami source geometry. It can help us understand the source characteristics (i.e., length, width, slip).

Acknowledgements

References

Aida, I. (1969). Numerical experiments for the tsunami propagation-the 1964 Niigata tsunami and the 1968 Tokachi-oki tsunami. *Bulletin of the Earthquake Research Institute*, **47**, 673–700.

Aida, I. (1978). Reliability of a tsunami source model derived from fault parameter. *Journal of Physics of the Earth*, **26**(1), 57–73. https://doi.org/10.4294/jpe1952.26.57.

Allen, R. V. (1978). Automatic earthquake recognition and timing from single traces. *Bulletin of the Seismological Society of America*, **68**(5), 1521–1532.

An, C., & Meng, L. (2017). Time reversal imaging of the 2015 Illapel tsunami source. *Geophysical Research Letters*, **44**, 1732–1739. https://doi.org/10.1002/2016GL071304.

Aoi, S., Suzuki, W., Chikasada, N. Y., Miyoshi, T., Arikawa, T., & Seki, K. (2019). Development and utilization of real-time tsunami inundation forecast system using S-net data. *Journal of Disaster Research*, **14**(2), 212–224. https://doi.org/10.20965/jdr.2019.p0212.

Aránguiz, R., González, G., González, J., et al. (2016). The 16 September 2015 Chile Tsunami from the Post-Tsunami Survey and Numerical Modeling Perspectives. *Pure and Applied Geophysics*, **173**, 333–348. https://doi.org/10.1007/s00024-015-1225-4.

Baba, T., Hirata, K., & Kaneda, Y. (2004). Tsunami magnitudes determined from ocean-bottom pressure gauge data around Japan. *Geophysical Research Letters*, **31**, L08303. https://doi.org/10.1029/2003GL019397.

Baba, T., Takahashi, N., Kaneda, Y., Ando, K., Matsuoka, D., & Kato, T. (2015). Parallel implementation of dispersive tsunami wave modeling with a nesting algorithm for the 2011 Tohoku Tsunami. *Pure and Applied Geophysics*, **172**, 3433–3472. https://doi.org/10.1007/s00024-015-1049-2.

Barnes, C., Best, M., Johnson, F., Phibbs, P., & Pirenne, B. (2008). Transforming the

ocean science through cabled observatories. *Marine Technology Reporter*, 30–36. https://doi.org/10.1109/AERO.2009.4839295.

Beavan, J., Samsonov, S., Denys, P., Sutherland, R., Palmer, N., & Denham, M. (2010). Oblique slip on the Puysegur subduction interface in the 2009 July MW 7.8 Dusky Sound earthquake from GPS and InSAR observations: implications for the tectonics of southwestern New Zealand. *Geophysical Journal International*, **183**, 1265–1286. https://doi.org/10.1111/j.1365-246X.2010.04798.x.

Bellsky, T., Kostelich, E. J., & Mahalov, A. (2014). Kalman filter data assimilation: Targeting observations and parameter estimation. *Chaos*, **24**, 024406. https://doi.org/10.1063/1.4871916.

Beltrami, G. M. (2008). An ANN algorithm for automatic, real-time tsunami detection in deep-sea level measurements. *Ocean Engineering*, **35**, 572–587. https://doi.org/10.1016/j.oceaneng.2007.11.009.

Beltrami, G. M. (2011). Automatic, real-time detection and characterization of tsunamis in deep-sea level measurements. *Ocean Engineering*, **38**, 1677–1685. https://doi.org/10.1016/j.oceaneng.2011.07.016.

Benavente, R., Cummins, P. R., & Dettmer, J. (2016). Rapid automated W-phase slip inversion for the Illapel great earthquake (2015, Mw=8.3). *Geophysical Research Letters*, **43**, 1910–1917. https://doi.org/10.1002/2015GL067418.

Bernard, E. N., & Meinig, C. (2011). History and future of deep-ocean tsunami measurements. *OCEANS'11 MTS/IEEE KONA,* Waikoloa, Hawaii, pp. 1–7. https://doi.org/10.23919/OCEANS.2011.6106894.

Bernard, E. N., & Robinson A. R. (2009). *Tsunamis*, 1st edition, Vol 15. Cambridge: Harvard University Press, p. 28.

Boore, D. M., & Akkar, S. (2003). Effect of causal and acausal filters on elastic and inelastic response spectra. *Earthquake Engineering and Structural Dynamics*, **32**, 1729–1748. https://doi.org/10.1002/eqe.2009.

Calisto, I., Miller, M., & Constanzo, I. (2016). Comparison between tsunami signals

generated by different source models and the observed data of the Illapel 2015 earthquake. *Pure and Applied Geophysics*, **173**, 1051–1061. https://doi.org/10.1007/s00024-016-1253-8.

Carvajal, M., Araya-Cornejo, C., Sepúlveda, I., et al. (2019). Nearly instantaneous tsunamis following the Mw 7.5 2018 Palu earthquake. *Geophysical Research Letters*, **46**, 5117–5126. https://doi.org/10.1029/2019GL082578.

Chierici, F., Embriaco, D., & Pignagnoli, L. (2017). A new real-time tsunami detection algorithm. *Journal of Geophysical Research: Oceans*, **122**, 636–652. https://doi.org/10.1002/2016JC012170.

Cienfuegos, R., Catalán, P. A., León, J. et al. (2016). Lessons for tsunami risk mitigation from recent events occurred in Chile: Research findings for alerting and evacuation from inter disciplinary perspectives. *AGU Fall Meeting 2016*, NH52A-07, San Francisco.

Cienfuegos, R., Catalán, P. A., Urrutia, A. et al. (2018). What can we do to forecast tsunami hazards in the near field given large epistemic uncertainty in rapid seismic source inversion. *Geophysical Research Letters*, **45**, 1–12. https://doi.org/10.1029/2018GL076998.

Clark, K. J., Johnson, P. N., Turnbull, I. M., & Litchfield, N. J. (2011). The 2009 Mw 7.8 earthquake on the Puysegur subduction zone produced minimal geological effects around Dusky Sound, New Zealand. *New Zealand Journal of Geology and Geophysics*, **54**, 237–247. https://doi.org/10.1080/00288306.2010.543690.

Daley, R. (1991). *Atmospheric data analysis*, Cambridge University Press, pp. 98–149.

Dätig, M., & Schlurmann, T. (2004). Performance and limitations of the Hilbert-Huang transformation (HHT) with an application to irregular water waves. *Ocean Engineering*, **2004**, 1783–1834. https://doi.org/10.1016/j.oceaneng.2004.03.007.

Duputel, Z., Rivera, L., Kanamori, H., et al. (2011). Real-time W phase inversion during the 2011 off the Pacific coast of Tohoku Earthquake. *Earth, Planets and Space*, **63**, 5. https://doi.org/10.5047/eps.2011.05.032.

Ekström, G. (1994). Anomalous earthquakes on volcano ring-fault structures. *Earth and Planetary Science Letters*, **128**, 707–712. https://doi.org/10.1016/0012-821X(94)90184-8.

Evensen, G. (2003). The Ensemble Kalman Fitler: theoretical formulation and practical implementation. *Ocean Dynamics*, **53**, 343–367. https://doi.org/10.1007/s10236-003-0036-9.

Fairbairn, D., Pring, S. R., Lorenc, A. C., & Roulstone, I. (2014). A comparison of 4DVar with ensemble data assimilation methods. *Quarterly Journal of the Royal Meteorological Society*, **140**, 281–294. https://doi.org/10.1002/qj.2135.

Fujii, Y., & Satake, K. (2007). Tsunami Source of the 2004 Sumatra-Andaman Earthquake Inferred from Tide Gauge and Satellite Data. *Bulletin of the Seismological Society of America*, **97**(1A), S192–S207. https://doi.org/10.1785/0120050613.

Fujii, Y., Satake, K., Sakai, S., Shinohara, M., & Kanazawa, T. (2011). Tsunami source of the 2011 off the Pacific coast of Tohoku Earthquake. *Earth, Planets and Space*, **63**, 55. https://doi.org/10.5047/eps.2011.06.010.

Fukao, Y., Sandanbata, O., Sugioka, H., Ito, A., Shiobara, H., Watada, S., & Satake, K. (2018). Mechanism of the 2015 volcanic tsunami earthquake near Torishima, Japan. *Science Advances*, **4**, eaao0219. https://doi.org/10.1126/sciadv.aao0219.

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. *Journal of Geophysical Research: Solid Earth,* **116**, B02308. https://doi.org.10.1029/2010JB007918.

Fritz, H. M., Petroff, C. M., Catalán, P. A., et al. (2011). Field survey of the 27 February 2010 Chile tsunami. *Pure and Applied Geophysics*, **168**, 1989–2010. https://doi.org/10.1007/s00024-011-0283-5.

Fry, B., Bannister, S., Beavan, J., Bland, L., et al., & GeoNet team. (2010). The Mw 7.6 Dusky Sound earthquake of 2009: Preliminary report. *Bulletin of the New Zealand Society for Earthquake Engineering*, **43**(1), 24–40.

https://doi.org/10.5459/bnzsee.43.1.24-40.

Garg, R., Varna, A. L., & Wu, M. (2012). An efficient gradient descent approach to secure localization in resource constrained wireless sensor networks. *IEEE Transactions on Information Forensics and Security*, **7**(2), 717–730. https://doi.org/10.1109/TIFS.2012.2184094.

González, F. I., Bernard, E. N., Meinig C., et al. (2005). The NTHMP tsunameter network. *Natural Hazards*, **35**, 25–39. https://doi.org/10.1007/s11069-004-2402-4.

Gusman, A. R., Mulia, I. E., Satake, K., Watada, S., Heidarzadeh, M., & Sheehan, A. F. (2016a). Estimate of tsunami source using optimized unit sources and including dispersion effects during tsunami propagation: The 2012 Haida Gwaii Earthquake. *Geophysical Research Letters*, **43**, 9819–9828. https://doi.org/10.1022/2016GL070140.

Gusman, A. R., Satake, K., Shinohara, M., Sakai, S., & Tanioka, Y. (2017). Fault slip distribution of the 2016 Fukushima earthquake estimated from tsunami waveforms. *Pure and Applied Geophysics*, **174**, 2925–2943. https://doi.org/10.1007/s00024-017-1590-2.

Gusman, A. R., Sheehan, A. F., Satake, K., Heidarzadeh, M., Mulia, I. E., & Maeda, T. (2016b). Tsunami data assimilation of Cascadia seafloor pressure gauge records from the 2012 Haida Gwaii earthquake. *Geophysical Research Letters*, **43**, 4189–4196. https://doi.org/10.1002/2016GL068368.

Gusman, A. R., Tanioka, Y., Maclnnes, B. T., & Tsushima, H. (2014). A methodology for near-field tsunami inundation forecasting: Application to the 2011 Tohoku tsunami. *Journal of Geophysical Research: Solid Earth*, **119**, 8186–8206. https://doi.org/10.1002/2014JB010958.

Hadamard, J. (1924). Le principe de Huygens (In French). *Bulletin de la Société Mathématique de France*, **52**, 610–640.

Hayashi, Y. (2008). Extracting the 2004 Indian Ocean tsunami signals from sea surface height data observed by satellite altimetry. *Journal of Geophysical Research: Oceans*, **113**, C01001. https://doi.org/10.1029/2007JC004177.

Hayashi, Y., Tsushima, H., Hirata, K., Kimura, K., & Maeda, K. (2011). Tsunami source area of the 2011 off the Pacific coast of Tohoku Earthquake determined from tsunami arrival times at offshore observation stations. *Earth, Planets and Space*, **63**, 809–813. https://doi.org/10.5047/eps.2011.06.042.

Hébert, H., Heinrich, P., Schindelé, F., & Piatanesi A. (2001). Far-field simulation of tsunami propagation in the Pacific Ocean: Impact on the Marquesas Islands (French Polynesia). *Journal of Geophysical Research: Oceans*, **106**, 9161–9177. https://doi.org/10.1029/2000JC000552.

Heidarzadeh, M., & Gusman, A. R. (2019). Application of dense offshore tsunami observations from Ocean Bottom Pressure Gauges (OBPGs) for tsunami research and early warnings. In: *Geological Disaster Monitoring Based on Sensor Networks*, Springer, Natural Hazards. Springer, Singapore. https://doi.org/10.1007/978-981-13-0992-2_2.

Heidarzadeh, M., Harada, T., Satake, K., Ishibe, T., & Gusman, A. R. (2016a). Comparative study of two tsunamigenic earthquakes in the Solomon Islands: 2015 Mw 7.0 normal-fault and 2013 Santa Cruz Mw 8.0 megathrust earthquakes. *Geophysical Research Letters*, **43**(9), 4340–4349. https://doi.org/10.1002/2016GL068601.

Heidarzadeh, M., Ishibe, T., Sandanbata, O., Muhari, A., & Wijanarto, A. B. (2020). Numerical modeling of the subaerial landslide source of the 22 December 2018 Anak Krakatoa volcanic tsunami, Indonesia. *Ocean Engineering*, **195**, 106733. https://doi.org/10.1016/j.oceaneng.2019.106733.

Heidarzadeh, M., Muhari, A., & Wijanarto, A. B. (2018). Insights on the source of the 28 September 2018 Sulawesi tsunami, Indonesia based on spectral analyses and numerical simulations. *Pure and Applied Geophysics*, **176**, 25–43. https://doi.org/10.1007/s00024-018-2065-9.

Heidarzadeh, M., Murotani, S., Satake, K., Ishibe, T., & Gusman, A. R. (2016b). Source model of the 16 September 2015 Illapel, Chile, Mw 8.4 earthquake based on teleseismic and tsunami data. *Geophysical Research Letters*, **43**(2), 643–650. https://doi.org/10.1002/2015GL067297.

Heidarzadeh, M., & Satake, K. (2013). Waveform and spectral analyses of the 2011 Japan tsunami records on tide gauges and DART stations across the Pacific Ocean. *Pure and Applied Geophysics*, **170**, 1275–1293. https://10.1007/s00024-012-0558-5.

Heidarzadeh, M., Wang, Y., Satake, K., & Mulia, I. E. (2019). Potential deployment of offshore bottom pressure gauges and adoption of data assimilation for tsunami warning system in the western Mediterranean Sea. *Geoscience Letters*, **6**, 19. https://doi.org/10.1186/s40562-019-0149-8.

Hino, R., Tanioka, Y., Kanazawa, T., et al. (2001). Micro-tsunami from a local interpolate earthquake detected by cabled offshore tsunami observation in northeastern Japan. *Geophysical Research Letters*, **28**, 3533–3536. https://doi.org/10.1029/2001GL013297.

Hossen, M. J., Cummins, P. R., Dettmer, J., & Baba, T. (2015). Time reverse imaging for far-field tsunami forecasting: 2011 Tohoku earthquake case study. *Geophysical Research Letters*, **42**, 9906–9915. https://doi.org/10.1002/2015GL065868.

Hossen, M. J., Cummins, P. R., & Satake, K. (2017). Complete implementation of the Green's Function Based Time Reverse Imaging and sensitivity analysis of reversed-time tsunami source inversion. *Geophysical Research Letters*, **44**, 9844–9855. https://doi.org/10.1002/2017GL074528.

Hossen, M. J., Gusman, A., Satake, K., & Cummins, P. R. (2018). An adjoint sensitivity method applied to time reverse imaging of tsunami source for the 2009 Samoa earthquake. *Geophysical Research Letters*, **45**, 627–636. https://doi.org/10.1002/2017GL076031.

Huang, N. E., Shen, Z., & Long, S. R. (1999). A new view of nonlinear water waves: The Hilbert Spectrum. *Annual Review of Fluid Mechanics*, **31**, 417–457. https://doi.org/10.1146/annurev.fluid.31.1.417.

Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., et al. (1998). The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society A*, **454**, 903–995. https://doi.org/10.1098/rspa.1998.0193.

Huang, N. E., & Wu, Z. (2008). A review on Hilbert-Huang Transform: Method and its applications to geophysical studies. *Review of Geophysics*, **46**, RG2006. https://doi.org/10.1029/2007RG000228.

Inoue, M., Tanioka, Y., & Yamanaka, Y. (2019). Method for near-real time estimation of tsunami sources using ocean bottom pressure sensor network (S-Net). *Geosciences*, **9**(7), 310. https://doi.org/10.3390/geosciences9070310.

Ito, S., Nagao, H., Yamanaka, A., Tsukada, Y., Koyama, T., Kano, M., & Inoue, J. (2016). Data assimilation for massive autonomous systems based on a second-order adjoint method. *Physical Review E*, **94**(4-1), 043307. https://doi.org/10.1103/PhysRevE.94.043307.

Kalnay, E. (2003). *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge Univ. Press, Cambridge. https://doi.org/10.1017/CBO9780511802270.

Kalnay, E., Li, H., Miyoshi, T., Yang, S.-C., & Ballabrera, J. (2007). 4-D-Var or Ensemble Kalman Filter. *Tellus A: Dynamic Meteorology and Oceanography*, **59**, 758–773. https://doi.org/10.1111/j.1600-0870.2007.00261.x.

Kamigaichi, O. (2015). Tsunami forecasting and warning. In: Meyers R. A. (eds) Encyclopedia of Complexity and Systems Science. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-27737-5_568-3.

Kanamori, H., Ekström, G., Dziewonski, A., Barker, J. S., & Sipkin, S. A. (1993). Seismic radiation by magma injection: An anomalous seismic event near Tori Shima, Japan. *Journal of Geophysical Research: Solid Earth*, **98**(B4), 6511–6522. https://doi.org/10.1029/92JB02867.

Kanamori, H., & Rivera, L. (2008). Source inversion of W phase: speeding up seismic tsunami warning. *Geophysical Journal International*, **175**(1), 222–238. https://doi.org/10.1111/j.1365-246X.2008.03887.x.

Kanazawa, T. (2013). Japan Trench earthquake and tsunami monitoring network of cable-linked 150 ocean bottom observatories and its impact to earth disaster science. *2013 IEEE International Underwater Technology Symposium (UT)*, Tokyo, pp. 1–5. https://doi.org/10.1109/UT.2013.6519911.

Kaneda, Y. (2010). The advanced ocean floor real time monitoring system for mega thrust earthquakes and tsunami — application of DONET and DONET2 data to seismological research and disaster mitigation. *OCEAN 2010*. https://doi.org/10.1109/OCEANS.2010.5664309.

Kawaguchi, K., Kaneda, Y., & Araki, E. (2008). The DONET: A real-time seafloor research infrastructure for the precise earthquake and tsunami monitoring. *OCEANS 2008 — MTS/IEEE Kobe Techno-Ocean*, Kobe, pp. 1–4. https://doi.org/10.1109/OCEANSKOBE.2008.4530918.

Kontar, Y. A., Santiago-Fandiño, V., & Takahashi, T. (2013). *Tsunami Events and Lessons Learned: Environmental and Societal Significance*, Springer, Berlin. https://doi.org/10.1007/978-94-007-7269-4.

Korolev, Y. P. (2011). An approximate method of short-term tsunami forecast and the hindcasting of some recent events. *Natural Hazards and Earth System Sciences*, **11**, 3081–3091. https://doi.org/10.5194/nhess-11-3081-2011.

Kubota, T. (2018). The propagation characteristic of ocean waves and tsunamis according to the analysis of pressure gauge array (In Japanese). *The 8th Conference on Giant Tsunami Disasters*, Tokyo.

Kubota, T., Chikasada, N. Y., Tsushima, H., Suzuki, W., Nakamura, T., & Kubo, H. (2020a). Tsunami analysis using the S-net pressure gauge records during the Mw 7.0 Off-Fukushima earthquake on 22 November 2016 to reduce the effects of tsunami-irrelevant pressure components. *JpGU-AGU Joint Meeting 2020*, HDS08-11, Online.

Kubota, T., Saito, T., & Suzuki, W. (2020b). Millimeter-scale tsunami detected by a wide and dense observation array in the deep ocean: Fault modeling of an Mw 6.0 interplate earthquake off Sanriku, NE Japan. *Geophysical Research Letters*, **47**, e2019GL085842. doi:10.1029/2019GL085842.

Kuyuk, H. S., & Susumu, O. (2018). Real-time classification of earthquake using deep learning. *Procedia Computer Science*, **140**, 298–305.

https://doi.org/10.1016/j.procs.2018.10.316.

Lay, T., Kanamori, H., Ammon, C. J., Nettles, M., et al. (2005). The great Sumatra-Andaman earthquake of 26 December 2004. *Science*, **308**, 1127–1133. https://doi.org/10.1126/science.1112250.

Leonard, R. (2006). Analysis of tide gauge records from the December 2004 Indian Ocean tsunami. *Geophysical Research Letters*, **33**, L17602. https://doi.org/10.1029/2006GL026552.

Liu, M., Lin, J., Wang, Y., Sun, Y., et al. (2018). Spatiotemporal variability of NO2 and PM2.5 over Eastern China: observational and model analyses with a novel statistical method. *Atmospheric Chemistry and Physics*, **18**, 12,933–12,952. https://doi.org/10.5194/acp-18-12933-2018.

Liu, P. L.-F., Wang, X., & Salisbury, A. J. (2009). Tsunami hazard and early warning system in South China Sea. *Journal of Asian Earth Sciences*, **36**, 2–12. https://doi.org/10.1016/j.jseaes.2008.12.010.

Lomax, A., & Michelini, A. (2011). Tsunami early warning using earthquake rupture duration and P-wave dominant period: the importance of length and depth of faulting. *Geophysical Journal International*, **185**(1), 283–291. https://doi.org/10/1111/j.1365-246X.2010.04916.x.

Lorenc, A. C. (2003). Modelling of error covariances by four-dimensional variational data assimilation. *Quarterly Journal of the Royal Meteorological Society*, **129**, 3167–3182. https://doi.org/10.1256/qj.02.131.

Lynch, P. (2008). The origins of computer weather prediction and climate modeling. *Journal of Computational Physics*, **227**, 3431–3444. https://doi.org/10.1016/j.jcp2007.02.034.

Maeda, T., Furumura, T., Sakai, S., & Shinohara, M. (2011) Significant tsunami observed at ocean-bottom pressure gauges during the 2011 off the Pacific coast of Tohoku Earthquake. *Earth, Planets and Space*, **63**, 53. https://doi.org/10.5047/eps.2011.06.005.

Maeda, T., Obara, K., Shinohara, M., Kanazawa, T., & Uehira, K. (2015). Successive estimation of a tsunami wavefield without earthquake source data: A data assimilation approach toward real-time tsunami forecasting. *Geophysical Research Letters*, **42**, 7923–7932, https://doi.org/10.1002/2015GL065588.

Meinig, C., Stalin, S. E., Nakamura, A. I., González, F., & Milburn, H. B. (2005). Technology developments in real-time tsunami measuring, monitoring and forecasting. In: *Proceedings of OCEANS 2005 MTS/IEEE*, **2**, 1673–1679. https://doi.org/10.1109/OCEANS.2005.1639996.

Mofjeld, H. O. (1997). Tsunami Detection Algorithm, Not published paper. [Available at http://nctr.pmel.noaa.gov/tda_documentation.html.]

Muhari, A., Heidarzadeh, M., Susmoro, H., et al. (2019). The December 2018 Anak Krakatau Volcano Tsunami as Inferred from Post-Tsunami Field Surveys and Spectral Analysis. *Pure and Applied Geophysics*, **176**, 5219–5233. https://doi.org/10.1007/s00024-019-02358-2.

Mulia, I. E., Gusman, A. R., & Satake, K. (2017a). Optimal design for placements of tsunami observing systems to accurately characterize the inducing earthquake. *Geophysical Research Letters*, **44**, 12,106–12,115. https://doi.org/10.1002/2017GL075791.

Mulia, I. E., Inazu, D., Waseda, T., & Gusman, A. R. (2017b). Preparing for the future Nankai Trough tsunami—a data assimilation and inversion analysis from various observational systems. *Journal of Geophysical Research: Oceans*, **122**, 7924–7937. https://doi.org/10.1002/2017JC012695.

Nalbant, S. S., Steacy, S., Sieh, K., Natawidjaja, D., & McCloskey, J. (2005). Earthquake risk on the Sunda trench. *Nature*, **435**, 756–757. https://doi.org/10.1038/nature435756a.

Navarrete, P., Cienfuegos, R., Satake, K., Wang, Y., et al. (2020). Sea surface network optimization for tsunami forecasting in the near field: application to the 2015 Illapel earthquake. *Geophysical Journal International*, **221**, 1640–1650. https://doi.org/10.1093/gji/ggaa098.

Nwogu, O. (1993). Alternative form of Boussinesq equations for nearshore wave propagation. *Journal of Waterway, Port, Coastal, and Ocean Engineering*, **119**, 618–638. https://doi.org/10.1061/(ASCE)0733-950X(1993)119:6(618).

Okada, Y. (1985). Surface deformation due to shear and tensile faults in a half-space. *Bulletin of the Seismological Society of America*, **75**(4), 1135–1154.

Rabinovich, A. B. (1997). Spectral analysis of tsunami waves: Separation of source and topography effects. *Journal of Geophysical Research: Oceans*, **102**(C6), 12,663–12,676. https://doi.org/10.1029/97JC00479.

Rabinovich, A. B., & Eblé, M. C. (2015). Deep-ocean measurements of tsunami waves. *Pure and Applied Geophysics*, **172**, 3281–3312. https://doi.org/10.1007/s00024-015-1058-1.

Ren, Z., Yuan, Y., Wang, P., et al. (2017). The September 16, 2015 Mw 8.3 Illapel, Chile Earthquake: characteristic of tsunami wave from near-field to far-field. *Acta Oceanologica Sinica*, **36**, 73–82. https://doi.org/10.1007/s13131-017-1005-3.

Roeber, V., Cheung, K. F., and Kobayashi, M. H. (2010). Shock-capturing Boussinesq-type model for nearshore wave processes. *Coastal Engineering*, 57, 407–423. https://doi.org/10.1016/j.coastaleng.2009.11.007.

Saito, T. (2019). *Tsunami Generation and Propagation* (pp. 1–16). Tokyo: Springer Geophysics. https://doi.org/10.1007/978-4-431-56850-6.

Saito, T., & Furumura, T. (2009). Three-dimensional simulation of tsunami generation and propagation: Application to intraplate events. *Journal of Geophysical Research: Solid Earth*, **114**, B02307. https://doi.org/10.1029/2007JB005523.

Saito, T., & Kubota, T. (2020). Tsunami modeling for the deep sea and inside focal areas. *Annual Review of Earth and Planetary Sciences*, **48**, 121–145. https://doi.org/10.1146/annurev-earth-071719-054845.

Saito, T., Satake, K., & Furumura, T. (2010). Tsunami waveform inversion including dispersive waves: the 2004 earthquake off Kii Peninsula, Japan. *Journal of Geophysical Research: Solid Earth*, **115**, B06303.

https://doi.org/10.1029/2009JB006884.

Sandanbata, O., Watada, S., Satake, K., Fukao, Y., Sugioka, H., Ito, A., & Shiobara, H. (2018). Ray tracing for dispersive tsunamis and source amplitude estimation based on Green's law: Application to the 2015 volcanic tsunami earthquake near Torishima, south of Japan. *Pure and Applied Geophysics*, **175**(4), 1371–1385. https://doi.org/10.1007/s00024-017-1746-0.

Sasaki, K., Yamanaka, A., Ito, S., Nagao, H. (2018). Data assimilation for phase-field models based on the ensemble Kalman filter. *Computational Materials Science*, **141**, 141–152. https://doi.org/10.1016/j.commatsci.2017.09.025.

Satake, K. (1989). Inversion of tsunami waveforms for the estimation of heterogeneous fault motion of large submarine earthquakes: The 1968 Tokachi-oki and 1983 Japan Sea Earthquakes. *Journal of Geophysical Research: Solid Earth*, **94**(B5), 5627–5636. https://doi.org/10.1029/JB094iB05p05627.

Satake, K. (2015). *Treatise on Geophysics* (2nd ed., Vol. 4, pp. 447–504). Oxford: Elsevier.

Satake, K., Fujii, Y., Harada, T., & Namegaya, Y. (2013). Time and space distribution of coseismic slip of the 2011 Tohoku earthquake as inferred from tsunami waveform data. *Bulletin of the Seismological Society of America*, **103**, 1473–1492. https://doi.org/10.1785/0120120122.

Satake, K. & Heidarzadeh, M. (2017). A review of source models of the 2015 Illapel, Chile earthquake and insights from tsunami data. *Pure and Applied Geophysics*, **174**, 1–9. https://doi.org/10.1007/s00024-016-1450-5.

Satake, K., Heidarzadeh, M., Qurioz, M., & Cienfuegos, R. (2020). History and features of trans-oceanic tsunamis and implications for paleo-tsunami studies. *Earth-Science Reviews*, **202**, 103112. https://doi.org/10.1016/j.earscirev.2020.103112.

Satake, K., & Kanamori, H. (1991). Abnormal tsunamis caused by the June 13, 1984, Torishima, Japan, earthquake. *Journal of Geophysical Research: Solid Earth*, **96**, 19,933–19,939. https://doi.org/10.1029/91JB01903.

Sato, T., Hasegawa, S., Kono, A., Shiobara, H., Yagi, T., Yamada, T., Shinohara, M., &

Usui, N. (2017). Detection of vertical motion during a slow-slip event off the Boso Peninsula, Japan, by ocean bottom pressure gauges. *Geophysical Research Letters*, **44**, 2710–2715. https://doi.org/10.1002/2017GL072838.

Sheehan, A. F., Gusman, A. R., Heidarzadeh, M., & Satake, K. (2015). Array observations of the 2012 Haida Gwaii tsunami using Cascadia Initiative absolute and differential seafloor pressure gauges. *Seismological Research Letters*, **86**(5), 1278–1286. https://doi.org/10.1785/0220150108.

Sheehan, A. F., Gusman, A. R., & Satake, K. (2019). Improving Forecast Accuracy with Tsunami Data Assimilation: The 2009 Dusky Sound New Zealand Tsunami. *Journal of Geophysical Research: Solid Earth*, **124**, 566–577. https://doi.org/10.1029/2018JB016575.

Shinohara, M., Yamada, T., Sakai, S., Shiobara, H., & Kanazawa, T. (2016). Development and installation of new seafloor cabled seismic and tsunami observation system using ICT. In: *OCEANS 2016 MTS/IEEE Monterey*, pp. 1–4. https://doi.org/10.1109/OCEANS.2016.7761350.

Shuler, A. E., Ekström, G., & Nettles, M. K. (2013). Physical mechanisms for vertical-CLVD earthquakes at active volcanoes. *Journal of Geophysical Research: Solid Earth*, **118**, 1569–1586. https://doi.org/10.1002/jgrb.50131.

Sugioka, H., Fukao, Y., Kanazawa, T., & Kanjo, K. (2000). Volcanic events associated with an enigmatic submarine earthquake. *Geophysical Journal International*, **142**(2), 361–370. https://doi.org/10.1046/j.1365-246x.2000.00153.x.

Suppasri, A., Leelawat, N., Latcharote, P., Roeber, V., et al. (2017). The 2016 Fukushima earthquake and tsunami: Local tsunami behavior and recommendations for tsunami disaster risk reduction. *International Journal of Disaster Risk Reduction*, **21**, 323–330. https://doi.org/10.1016/j.ijdrr.2016.12.016.

Synolakis, C. E. (1990). Green's law and the evolution of solitary waves. *Physics of Fluids A Fluid Dynamics*, **3**(3), 490–491. https://doi.org/10.1063/1.858107.

Synolakis, C. E., & Skjelbreia, J. E. (1993). Evolution of maximum amplitude of solitary waves on plane beaches. *Journal of Waterway, Port, Coastal, and Ocean*

Engineering, **119**(3), 323–342. https://doi.org/10.1061/(ASCE)0733-950X(1993)119:3(323).

Tang, L., Titov, V. V., Moore, C., & Wei, Y. (2016). Real-time assessment of the 16 September 2015 Chile tsunami and implications for near-field forecast. *Pure and Applied Geophysics*, **173**, 369–387. https://doi.org/10.1007/s00024-015-1226-3.

Tanioka, Y. (2018). Tsunami simulation method assimilating ocean bottom pressure data near a tsunami source region. *Pure and Applied Geophysics*, **175**, 721–729. https://doi.org/10.1007/s00024-017-1697-5.

Tarantola, A. (1984). Linearized inversion of seismic reflection data. *Geophysical Prospecting*, **32**, 998–1015. https://doi.org/10.1111/j.1365-2478.1984.tb00751.x.

Titov, V. V. (2009). Tsunami forecasting. In: Bernard EN, Robinson AR (eds) *In the sea, Ideas and Observations on Progress in the Study of the Sea*, 15. Harvard University Press, Cambridge, 371–400.

Titov, V. V., Gonzalez, F. I., Bernard, E. N., Eble, M. C., et al. (2005). Real-Time Tsunami Forecasting: Challenges and Solutions. *Natural Hazards*, **35**, 35–41. https://doi.org/10.1007/s11069-004-2403-3.

Tsushima, H., Hino, R., Fujimoto, H., Tanioka, Y., & Imamura, F. (2009). Near-field tsunami forecasting from cabled ocean bottom pressure data. *Journal of Geophysical Research: Solid Earth*, **114**, B06309. https://doi.org/10.1029/2008JB005988.

Tsushima, H., Hino, R., Tanioka, Y., Furumura, F., & Fujimoto, H. (2012). Tsunami waveform inversion incorporating permanent seafloor deformation and its application to tsunami forecasting. *Journal of Geophysical Research: Solid Earth*, **117**, B03311. https://doi.org/10.1029/2011JB008877.

Tsushima, H., Hirata, K., Hayashi, Y., Tanioka, Y., Kimura, K., et al. (2011). Near-field tsunami forecasting using offshore tsunami data from the 2011 off the Pacific coast of Tohoku Earthquake. *Earth, Planets and Space*, **63**, 821–826. https://doi.org/10.5047/eps.2011.06.052.

Wang, D., Becker, N. C., Walsh, D., et al. (2012). Real-time forecasting of the April 11,

2012 Sumatra tsunami. *Geophysical Research Letters*, **39**(19), L19601. https://doi.org/10.1029/2012GL053081.

Wang, Y., Maeda, T., Satake, K., Heidarzadeh, M., Su, H., Sheehan, A. F., & Gusman, A. R. (2019a). Tsunami data assimilation without a dense observation network. *Geophysical Research Letters*, **46**(4), 2045–2053. https://doi.org/10.1029/2018GL080930.

Wang, Y., Satake, K., Sandabata, O., Maeda, T., & Su, H. (2019b). Tsunami data assimilation of cabled ocean bottom pressure records for the 2015 Torishima volcanic tsunami earthquake. *Journal of Geophysical Research: Solid Earth*, **124**, 10,413–10,422. https://doi.org/10.1029/2019JB018056.

Weatherall, P., Marks, K. M., Jakobsson, M., et al. (2015). A new digital bathymetric model of the world's oceans. *Earth and Space Science*, **2**(8), 331–345. https://doi.org/10.1002/2015EA000107.

Williamson, A., & Newman, A. (2018). Limitations of the resolvability of finite-fault models using static land-based geodesy and open-ocean tsunami waveforms. *Journal of Geophysical Research: Solid Earth*, **123**, 9033–9048. https://doi.org/10.1029/2018JB016091.

Williamson, A., Newman, A., & Cummins, P. (2017). Reconstruction of coseismic slip from the 2015 Illapel earthquake using combined geodetic and tsunami waveform data. *Journal of Geophysical Research: Solid Earth*, **122**, 2119–2130. https://doi.org/10.1002/2016JB013883.

Wu, Z., & Huang, N. E. (2009). Ensemble Empirical Mode Decomposition: A noise-assisted data analysis method. *Advances in Adaptive Data Analysis*, **1**, 1–41. https://doi.org/10.1142/S1793536909000047.

Yang, Y., Dunham, E. M., Barnier, G., & Almquistj, M. (2019). Tsunami wavefield reconstruction and forecasting using the Ensemble Kalman Filter. *Geophysical Research Letters*, **46**, 853–860. https://doi.org/10.1029/2018GL080644.

Yang, Z., Sheehan, A. F., Collins, J. A., & Laske, G. (2012). The character of seafloor ambient noise recorded offshore New Zealand: Results from the MOANA ocean

bottom seismic experiment. *Geochemistry, Geophysics, Geosystems*, **13**, Q10011. https://doi.org/10.1029/2012GC004201.

Yamamoto, N., Aoi, S., Hirata, K., Suzuki, W., Kunugi, T., & Nakamura, H. (2016a). Multi-index method using offshore ocean-bottom pressure data for real-time tsunami forecast. *Earth, Planets and Space*, **68**, 128. https://doi.org/10.1186/s40623-016-0500-7.

Yamamoto, N., Hirata, K., Aoi, S., Suzuki, W., Nakamura, H., & Kunugi, T. (2016b). Rapid estimation of tsunami source centroid location using a dense offshore observation network. *Geophysical Research Letters*, **43**, 4263–4269. https://doi.org/10.1002/2016GL068169.

Yamanaka, Y. (2004). EIC Seismology Note. http://wwweic.eri.u-tokyo.ac.jp/sanchu/Seismo_Note/2004/EIC153.html.

Yamazaki, Y., & Cheung, K. F. (2011). Shelf resonance and impact of near-field tsunami generated by the 2010 Chile earthquake. *Geophysical Research Letters*, **38**, L12605. https://doi.org/10.1029/2011GL047508.

Zhang, X., Zhang, Q., Zhang, G., Nie, Z., et al. (2018). A novel hybrid data-driven model for daily land surface temperature forecasting using Long Short-Term Memory neural network based on Ensemble Empirical Mode Decomposition. *International Journal of Environmental Research and Public Health*, **15**, 1032. https://doi.org/10.3390/ijerph15051032.

Data Resources

The bathymetric data of ocean floor is provided by the General Bathymetric Chart of the Ocean (https://www.gebco.net), and the Japan Hydrographic Association (https://www.jha.or.jp/en/jha/). I use tsunami simulation code *JAGURS* (Baba et al., 2015; available at https://github.com/jagurs-admin/jagurs) and data assimilation code *TDAC* (Maeda et al., 2015; Gusman et al., 2016b; available at https://github.com/takuto-maeda/tdac). The open tsunami data are obtained from Intergovernmental Oceanographic Commission (http://www.ioc-sealevelmonitoring.org), Japan Agency for Marine-Earth Science and Technology (http://www.jamstec.go.jp/j/), Japan Meteorological Agency (https://www.jma.go.jp/jma/index.html), National Research Institute for Earth Science and Disaster Resilience (https://hinetwww11.bosai.go.jp/auth/oc/) and National Oceanic and Atmospheric Administration (https://nctr.pmel.noaa.gov/Dart/).

Appendix

A1 Combination of Raw IMFs

In this section, I demonstrate the combination of raw IMFs. I use the tsunami records of the 2016 Fukushima earthquake as an example. The input signal is a three-hour record of station TM1 from 19:00 to 22:00 (UTC), November 21, 2016. The decomposition generates seven raw IMFs, listed in a descending order of frequency. Short-period oscillation tends to appear in the lower mode of IMFs, whereas the higher-mode IMFs mostly contain long-period signals (Figure A1, left column). I plot their amplitude spectrum in the middle column to show the frequency characteristics. The 1st raw IMF has a dominant frequency of over 0.01 Hz (< 100 s). The 2nd to 4th raw IMFs have dominant frequencies of 0.002–0.01 Hz (500–100 s). The 5th to 7th raw IMFs have dominant frequencies of approximately 0.001 Hz (1000 s).

Hence, I use the 1st raw IMF as IMF1, combine the 2nd to 4th raw IMFs as IMF2, and combine the 5th to 7th raw IMFs as IMF3. Each IMF represents different components. The tsunami signal is presented in IMF2.
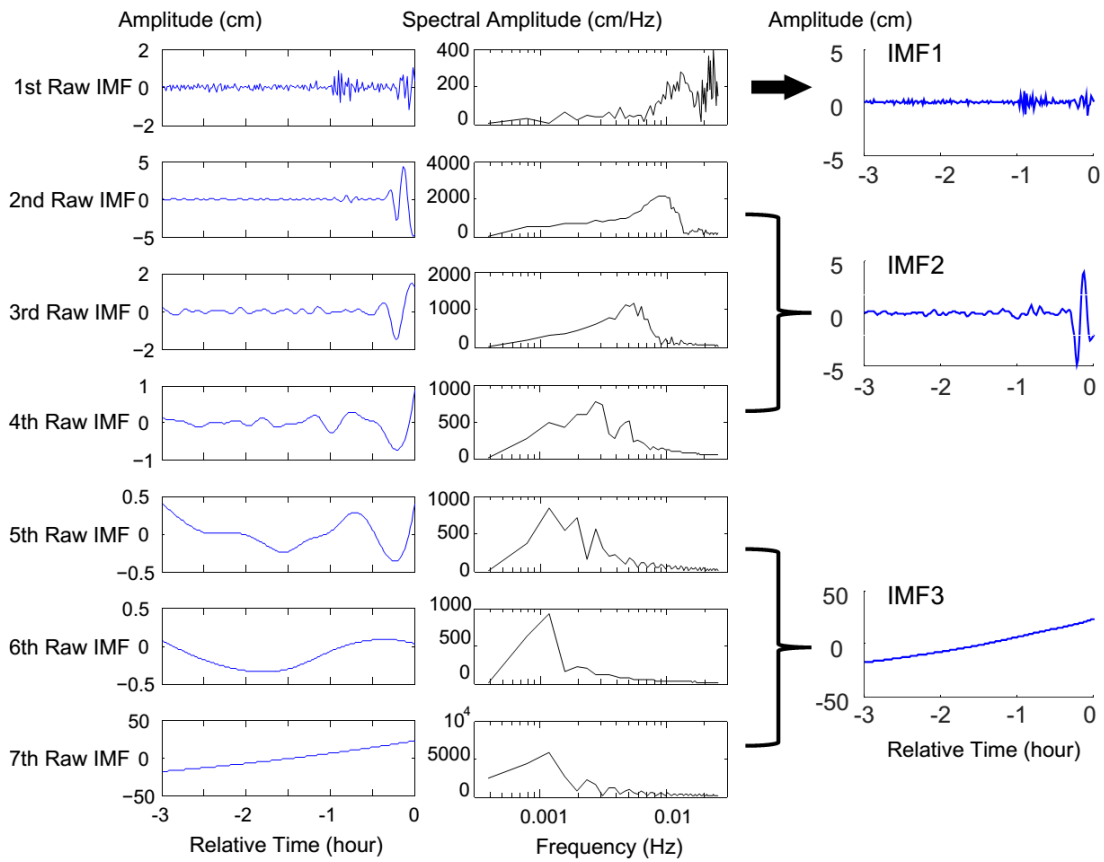
Figure A1. Combination of raw IMFs.

The input signal is the sea-level record of station TM1 from 19:00 to 22:00 (UTC), November 21, 2016. It is decomposed into seven raw IMFs (left column). The frequency characteristics of raw IMFs are plotted in the middle column. I combine the raw IMFs with similar frequency characteristics artificially and obtain three IMFs (right column).

A2 Abbreviations

ANN: Artificial Neural Network

CLVD: Compensated Linear Vector Dipole

DART: Deep-ocean Assessment and Reporting of Tsunami

DONET: Dense Oceanfloor Network System for Earthquakes and Tsunamis

DSFO: Deep Sea Floor Observatory

DSP: Dispersive

EEMD: Ensemble Empirical Mode Decomposition

EIC: Earthquake Information Center

EMD: Empirical Mode Decomposition

EnKF: Ensemble Kalman Filter

4-D Var: Four-dimensional variational assimilation

GEBCO_2014: General Bathymetric Chart of the Ocean released in 2014

GFTDA: Green's Function-based Tsunami Data Assimilation

IMF: Intrinsic Mode Function

IOC: Intergovernmental Oceanographic Commission

JAMSTEC: Japan Agency for Marine-Earth Science and Technology

JMA: Japan Meteorological Agency

LLW: Linear Long-Wave

LM-IMF: Last-Moment IMF

MOANA: Marine Observations of Anisotropy near Aotearoa

NIED: National Research Institute for Earth Science and Disaster Resilience

NOAA: National Oceanic and Atmospheric Administration

OBPG: Offshore Bottom Pressure Gauge

OBS: Ocean Bottom Seismometer

PoI: Point of Interest

PTWC: Pacific Tsunami Warning Center

S-net: Seafloor Observation Network for Earthquakes and Tsunamis

3-D NS: Three-dimensional Navier-Stokes

tFISH: tsunami Forecasting based on Inversion for sea-Surface Height

TRI: Time Reverse Imaging

USGS: United States Geological Survey

A3 Publications

[1] **<u>Wang, Y.</u>**, & Satake, K. (2021). Real-time tsunami data assimilation of S-net pressure gauge records during the 2016 Fukushima earthquake. *Seismological Research Letters*, **92**, 2145–2155. https://doi.org/10.1785/0220200447.

[2] **<u>Wang, Y.</u>**, Zamora, N., Quiroz, M., Satake, K., & Cienfuegos, R. (2021). Tsunami resonance characterization in Japan due to trans-Pacific sources: Response on the bay and continental shelf. *Journal of Geophysical Research: Oceans*, **126**, e2020JC017037. https://doi.org/10.1029/2020JC017037.

[3] **<u>Wang, Y.</u>**, Heidarzadeh, M., Satake, K., Mulia, I. E., & Yamada, M. (2020). A tsunami early warning system based on offshore bottom pressure gauges and data assimilation for Crete Island in the eastern Mediterranean basin. *Journal of Geophysical Research: Solid Earth*, **125**, e2020JB20293. https://doi.org/10.1029/2020JB020293.

[4] **<u>Wang, Y.</u>**, Satake, K., Maeda, T., Shinohara, M., & Sakai, S. (2020). A method of real-time tsunami detection using Ensemble Empirical Mode Decomposition. *Seismological Research Letters*, **91**, 2851–2861. https://doi.org/10.1785/0220200115.

[5] **<u>Wang, Y.</u>**, Satake, K., Sandabata, O., Maeda, T., & Su, H. (2019). Tsunami data assimilation of cabled ocean bottom pressure records for the 2015 Torishima volcanic tsunami earthquake. *Journal of Geophysical Research: Solid Earth*, **124**, 10,413–10,422. https://doi.org/10.1029/2019JB018056.

[6] **<u>Wang, Y.</u>**, Satake, K., Cienfuegos, R., Quiroz, M., & Navarrete, P. (2019). Far-field tsunami data assimilation for the 2015 Illapel earthquake. *Geophysical Journal International*, **209**, 514–521. https://doi.org/10.1093/gji/ggz309.

[7] **<u>Wang, Y.</u>**, Maeda, T., Satake, K., Heidarzadeh, M., Su, H., Sheehan, A. F., & Gusman, A. R. (2019). Tsunami data assimilation without a dense observation network. *Geophysical Research Letters*, **46**, 2045–2053. https://doi.org/10.1029/2018GL080930.

[8] **Wang, Y.**, Satake, K., Maeda, T., & Gusman, A. R. (2018). Data assimilation with dispersive tsunami model: a test for the Nankai Trough. *Earth, Planets and Space*, **70**, 131. https://doi.org/10.1186/s40623-018-0905-6.

[9] **Wang, Y.**, Satake, K., Maeda, T., & Gusman, A. R. (2017). Green's Function-based Tsunami Data Assimilation (GFTDA): A fast data assimilation approach toward tsunami early warning. *Geophysical Research Letters*, **44**, 10282–10289. https://doi.org/10.1002/2017GL075307.

[10] Ren, Z., Hou, J., Wang, P., **Wang, Y.** (2021). Tsunami resonance and standing waves in Hangzhou Bay. *Physics of Fluids*, **33**, 081702. https://doi.org/10.1063/5.0059383.

[11] Mulia, I. E., Watada, S., Ho, T. C., Satake, K., **Wang, Y.**, et al. (2020). Simulation of the 2018 tsunami due to the flank failure of Anak Krakatau volcano and implication for future observing systems. *Geophysical Research Letters*, **47**, e2020GL087334. https://doi.org/10.1029/2020GL087334.

[12] Navarrete, P., Cienfuegos, R., Satake, K., **Wang, Y.**, et al. (2020). Sea surface network optimization for tsunami forecasting in the near field: application to the 2015 Illapel earthquake. *Geophysical Journal International*, **221**, 1640–1650. https://doi.org/10.1093/gji/ggaa098.

[13] Ren, Z., **Wang, Y.**, Wang, P., Hou, J., et al. (2020). Numerical study of the triggering mechanism of the 2018 Anak Krakatau tsunami: eruption of collapsed landslide? *Natural Hazards*, **102**, 1–13. https://doi.org/10.1007/s11069-020-03907-y.

[14] Hu, Y., Xia, Y., Liu, Z., **Wang, Y.**, et al. (2020). Distorted Pacific-North American teleconnection at the Last Glacial Maximum. *Climate of the Past*, **16**, 199–209. https://doi.org/10.5194/cp-16-199-2020.

[15] Heidarzadeh, M., **Wang, Y.**, Satake, K., & Mulia, I. E. (2019). Potential deployment of offshore bottom pressure gauges and adoption of data assimilation for tsunami warning system in the western Mediterranean Sea. *Geoscience Letters*, **6**, 19. https://doi.org/10.1186/s40562-019-0149-8.

[16] Liu, M., Lin, J., **Wang, Y.**, Sun, Y., et al. (2018). Spatiotemporal variability of NO2 and PM2.5 over Eastern China: observational and model analyses with a novel

statistical method. *Atmospheric Chemistry and Physics*, **18**, 12933–12952. https://doi.org/10.5194/acp-18-12933-2018.