Geophysical Research Letters

RESEARCH LETTER
10.1029/2018GL080930

Tsunami Data Assimilation Without a Dense Observation Network

Y. Wang[1,2], T. Maeda[3], K. Satake[1], M. Heidarzadeh[4], H. Su[5], A. F. Sheehan[6], and A. R. Gusman[7]

[1]Earthquake Research Institute, The University of Tokyo, Tokyo, Japan, [2]Department of Earth and Planetary Science, Graduate School of Science, The University of Tokyo, Tokyo, Japan, [3]Graduate School of Science and Technology, Hirosaki University, Aomori, Japan, [4]Department of Civil and Environmental Engineering, Brunel University London, Uxbridge, UK, [5]Department of Earth Sciences, National Central University, Taoyuan City, Taiwan (R.O.C.), [6]Department of Geological Sciences and Cooperative Institute for Research in Environmental Sciences, University of Colorado Boulder, Boulder, CO, USA, [7]GNS Science, Lower Hutt, New Zealand

Key Points:
- A modified tsunami data assimilation approach is proposed for tsunami warning in regions without a dense observation network
- The tsunami waveforms are computed at virtual stations and used for data assimilation between two existing real observation points
- The method is applied to the 2004 Sumatra-Andaman earthquake and the 2009 Dusky Sound earthquake and improves the forecasting accuracy

Supporting Information:
- Supporting Information S1

Correspondence to:
Y. Wang,
ywang@eri.u-tokyo.ac.jp

Citation:
Wang, Y., 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

Received 23 OCT 2018
Accepted 9 FEB 2019
Accepted article online 13 FEB 2019
Published online 21 FEB 2019

Abstract The tsunami data assimilation method enables tsunami forecasting directly from observations, without the need of estimating tsunami sources. However, it requires a dense observation network to produce desirable results. Here we propose a modified method of tsunami data assimilation for regions with a sparse observation network. The method utilizes interpolated waveforms at virtual stations. The tsunami waveforms at the virtual stations between two existing observation stations are estimated by shifting arrival times with the linear interpolation of observed arrival times and by correcting the amplitudes for their water depths. In our new data assimilation approach, we employ the Optimal Interpolation algorithm to both the real observations and virtual stations, in order to construct a complete wavefront of tsunami propagation. The application to the 2004 Sumatra-Andaman earthquake and the 2009 Dusky Sound, New Zealand, earthquake reveals that addition of virtual stations greatly helps improve the tsunami forecasting accuracy.

Plain Language Summary Data assimilation is a method to combine observation and numerical simulation and is widely used in weather forecast. The data assimilation methods have been recently applied for tsunami forecast in North America and Japan where dense observation networks exist. In this study, we proposed a data assimilation method by introducing virtual observation data from neighboring real observations. We applied the method for the Indian Ocean with the 2004 Sumatra-Andaman earthquake tsunami and offshore New Zealand with the 2009 Dusky Sound earthquake tsunami. We found that the method greatly improved the forecasting accuracy and the method could be used for the regions with sparse observation network.

1. Introduction

Tsunami early warning systems play an important role in mitigating the destructive consequences of tsunamis. Various methods have been applied in the past for early tsunami warning, namely, the Method of Splitting Tsunami model (Titov et al., 2005), tsunami Forecasting based on Inversion for initial sea-Surface Height (Tsushima et al., 2009), Near-field Tsunami Inundation Forecasting (Gusman et al., 2014), and Time Reverse Imaging (Hossen et al., 2015). One promising method that can be used for tsunami early warning is the tsunami data assimilation first introduced by Maeda et al. (2015). It forecasts the tsunami heights and arrival times at nearshore points directly from the offshore observations, without the need of considering the earthquake source parameters such as strike, dip, and rake angles of the fault as well as slip values and fault dimensions. This method has already been successfully applied to synthetic tsunamis around Japan based on the Seafloor observation network for earthquakes and tsunamis along the Japan Trench (Maeda et al., 2015) and the real tsunamis recorded by pressure gauges in the Cascadia subduction zone (Gusman et al., 2016). Although a large computational load was an obstacle to apply this data assimilation technique in real time, Wang et al. (2017) accelerated the forecasting process based on the data assimilation by the introduction of Green's Function-based Tsunami Data Assimilation (GFTDA). The forecasted waveforms are superimposed by precalculated Green's functions, without the need of simulating tsunami propagation during the assimilation process.

The tsunami data assimilation method relies on a dense observation network, preferably located offshore in deep ocean. Examples of offshore devices are Deep-Ocean Assessment and Reporting of Tsunamis (DARTs; Gonzalez et al., 2005; Rabinovich & Eblé, 2015) and Ocean Bottom Seismometer (OBS) Pressure Gauges (Heidarzadeh & Gusman, 2018; Sheehan et al., 2015). Such offshore tsunami observation devices usually require a large investment, and many regions with significant tsunami hazard may only afford a few DARTs or OBSs rather than a dense network.

A typical example of a region with sparse offshore tsunami observations is the Indian Ocean, which suffered from the 2004 Sumatra-Andaman earthquake and tsunami (Fujii & Satake, 2007; Nalbant et al., 2005). Only six ocean bottom gauges are currently available for tsunami detection in the Bay of Bengal (north Indian Ocean): stations 23217, 23218, 23219, 23227, 23223, and 23401. These stations were installed after the 2004 Sumatra-Andaman earthquake.

Another example is the Puysegur subduction zone, south New Zealand, which is the result of subduction of the Australian Plate beneath southwestern New Zealand (Beavan et al., 2010). This subduction zone is believed to be capable of hosting megathrust earthquakes and generating tsunamis that can affect New Zealand, Tasmania, and the southeastern coast of Australia (Hayes & Furlong, 2010; Hayes et al., 2009). As part of the Marine Observations of Anisotropy Near Aotearoa (MOANA) seismic experiment, OBSs were deployed from January 2009 to February 2010 offshore the South Island of New Zealand (Yang et al., 2012). They were equipped with differential pressure gauges (DPGs) with the sampling rates of 40 Hz. Unlike Absolute Pressure Gauges, the amplitude corrections are necessary for DPGs, and such corrections were already provided by Sheehan et al. (2019). However, the MOANA network is not as dense as the Japanese Seafloor observation network for earthquakes and tsunamis along the Japan Trench, and some of the MOANA DPGs did not record the tsunami. Therefore, in the previous study (Sheehan et al., 2019), the W-phase inversion was combined with data assimilation to compensate for the sparse observation. In addition to the Optimal Interpolation scheme introduced by Maeda et al. (2015), the Ensemble Kalman Filter method has been applied to tsunami data assimilation with sparse observation data (Yang et al., 2019), but at the price of relatively high computation cost.

In this study, we introduce a novel methodology to solve the problem of sparse observation at a low cost. We artificially create stations in the area between adjacent real observation points and build virtual dense observation network. We apply our methodology to the synthetic tsunami data of the 2004 Sumatra-Andaman earthquake and the real tsunami data of the 2009 Dusky Sound earthquake.

2. Methodology

2.1. Linear Interpolation With Huygens-Fresnel Principle

Our approach is based on a linear interpolation of real data for computing the artificial waveforms for virtual stations. The principle of this method is similar to the Huygens-Fresnel principle in optics. According to the Huygens-Fresnel principle (e.g., Hadamar, 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 tsunami wave, the observation stations resemble the points on the wavefront, and we need a wavefront that is densely sampled. Therefore, we can apply linear interpolation to construct artificial waveforms at the location of virtual points (or stations).

Virtual stations do not exist, and their "waveforms" are only used for building the assimilation wavefield. We consider several virtual stations between two real stations (Figure 1). The virtual stations were located with equal distances along a straight line between the real stations. The effects of the number of stations, or the distance between the virtual stations, will be discussed in the supporting information. More generally, the interpolation scheme can be applied to three or more real stations (e.g., virtual stations inside a triangle), but we only consider the two stations for the sake of simplicity. The network of stations (both real and virtual) could form a wavefront together as the Huygens-Fresnel principle.

The first task for constructing the virtual waveform is a linear interpolation of two real arrival times to estimate the tsunami arrival time at the virtual station. We define a threshold for tsunami arrival in each station. In real practice, the tsunami waveform should be obtained after the removal of high-frequency seismic

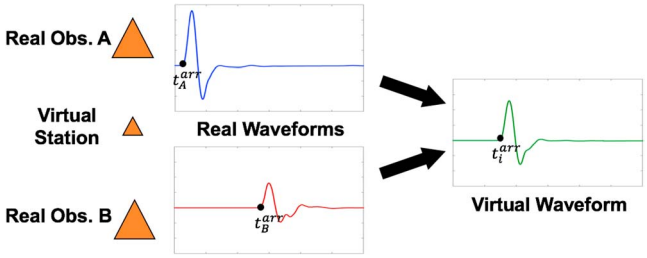

Figure 1. Illustration of the linear interpolation process. We first find the tsunami arrival time of two real stations and calculate the arrival time of virtual station(s) by weighted average. Then, we calculate the virtual waveform(s) by shifting the arrival time with correction to water depth. The virtual waveform(s) will be adopted in Optimal Interpolation along with real waveforms.

signal and low-frequency tidal signals. We use t_A^{arr} and t_B^{arr} to represent the arrival times of two real stations, respectively. Then we calculate the arrival time of the ith virtual station between two real stations as

$$t_i^{arr} = w_{iA} \cdot t_A^{arr} + w_{iB} \cdot t_B^{arr}, \tag{1}$$

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 distance between the virtual and real stations, if the tsunami velocity around the two stations are assumed to be constant. The sum of them equals to one. For instance, if the ith virtual station is located in the middle point of two real stations, the value of weight parameters will be 0.5. After obtaining the arrival time of the virtual station, we interpolate the amplitudes of the two real observations to obtain that of the artificial waveform by taking the water depths at the stations into consideration. We record the tsunami waveforms after the arrival time of two real stations and represent them as $Y_A\left(t-t_A^{arr}\right)$ and $Y_B\left(t-t_B^{arr}\right)$, respectively. The tsunami waveform of the ith virtual station is calculated as follows.

$$y_i\left(t-t_i^{arr}\right) = \left[\frac{w_{iA} \cdot Y_A\left(t-t_A^{arr}\right)}{d_A{}^{-\frac{1}{4}}} + \frac{w_{iB} \cdot Y_B\left(t-t_B^{arr}\right)}{d_B{}^{-\frac{1}{4}}}\right] \cdot d_i{}^{-\frac{1}{4}}, \tag{2}$$

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

The main characteristic of our linear interpolation method is that we calculate the virtual waveforms by shifting the arrival times considering the distance and correct the amplitudes considering the water depths at the stations. The arrival times are linearly interpolated by assuming constant velocity, or water depth, because we do not know the direction of wave arrival. On the contrary, the corrections of amplitudes depend only on the water depths at the stations, which are known. We acknowledge that the virtual waveform calculated by our linear interpolation method may not be exactly the same as the real observations, especially in some places with abrupt changes in bathymetry, but it can still improve the performance of data assimilation. We validate our method with real data in part 4. The virtual stations help construct a more complete tsunami waveform. On the other hand, because the interpolation depends on the waveform information of the two adjacent real stations, the virtual waveforms cannot be computed until the tsunami arrives at both real stations.

2.2. Optimal Interpolation

We assimilate the tsunami height at both real and virtual observation to forecast the tsunami waveforms. The tsunami wavefield at the nth time step is represented as $\boldsymbol{x_n}(\eta(n\Delta t, x, y), M(n\Delta t, x, y), N(n\Delta t, x, y))$, where η is tsunami height, M and N are velocities in two directions, Δt is the time step, and x and y are the spatial coordinates. The Optimal Interpolation method (Gusman et al., 2016; Kalnay, 2003; Maeda et al., 2015; Mulia et al., 2017) is performed as the following equations.

$$\boldsymbol{x}_n^f = \boldsymbol{F}\boldsymbol{x}_{n-1}^a, \tag{3}$$

$$\boldsymbol{x}_n^a = \boldsymbol{x}_n^f + \boldsymbol{P}\boldsymbol{H}^T\left(\boldsymbol{R} + \boldsymbol{H}\boldsymbol{P}\boldsymbol{H}^T\right)^{-1}\left(\boldsymbol{y}_n - \boldsymbol{H}\boldsymbol{x}_n^f\right). \tag{4}$$

At each time step, the forecasted tsunami wavefield $\boldsymbol{x}_n^f$ is simulated by solving the tsunami propagation equations using the assimilated wavefield in the last time step $\boldsymbol{x}_{n-1}^a$. The propagation matrix $\boldsymbol{F}$ corresponds to the tsunami propagation model. $\boldsymbol{H}$ is the observational operator, and $\boldsymbol{y}_n$ is the vector of tsunami observation. Equation (4) is used to bring the forecasted tsunami wavefield closer to the observed tsunami wavefield, where $\boldsymbol{P} = <\varepsilon^f \varepsilon^{fT}>$ and $\boldsymbol{R} = <\varepsilon^O \varepsilon^{OT}>$ are the covariance matrices of the forward numerical simulation and the observations, respectively. The standard error between grids is assumed to be homogeneous in

space, and the correlation between two points merely depends on a Gaussian correlation with a specific characteristic distance (Kalnay, 2003; Maeda et al., 2015).

2.3. Green's Function-Based Tsunami Data Assimilation

To further reduce the computational cost during the data assimilation process, we use the GFTDA (Wang et al., 2017). The location of real and virtual stations is fixed. We calculate the Green's functions between stations and nearshore points by using the well-validated tsunami code JAGURS (Baba et al., 2015). GFTDA requires the linearity of the tsunami propagation model, and we adopt the linear long-wave model based on long-wave approximation (Satake, 1995) in our computation of Green's functions. The number of the Green's functions is $n \times (n + m)$, where n is the number of stations (real and virtual) and m is the number of nearshore points. Although the calculation of Green's functions is time-consuming, once it is calculated and stored in advance, the data assimilation process will be very quick.

2.4. Accuracy of Data Assimilation

To evaluate the performance of data assimilation method quantitatively, we adopt the geometric mean ratio K and accuracy between the synthetic observation and forecasting using the following equations (Aida, 1978; Gusman et al., 2016):

$$\log(K) = \frac{1}{N}\sum_{i=1}^{N} \log\left(\frac{A_j^{obs}}{A_j^{pred}}\right), \tag{5}$$

$$Accuracy(\%) = \frac{1}{K}\times 100\%(K\geq 1) \text{ or } K\times 100\%(K<1), \tag{6}$$

where A_j^{obs} and A_j^{pred} stand for the maximum amplitude of the computed and forecasted waveforms and N is the number of nearshore points. Although at some time the positive and negative logarithm terms may cancel each other and the geometric mean ratio becomes anomalously high (Wang et al., 2017), a mean value close to 1 generally indicate accurate forecasting. The overall forecasting accuracy is then calculated by equation (6).

3. Application to Synthetic Tsunami Simulation

3.1. The 2004 Sumatra-Andaman Earthquake

To test the effectiveness of our improved tsunami data assimilation method, we performed a numerical simulation of the 2004 Sumatra-Andaman earthquake. The earthquake occurred at 00:58:53 UTC on 26 December 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 2) and arrived at the coasts of India and Sri Lanka about 2 hr after the earthquake (Fujii & Satake, 2007). Because no ocean bottom gauges were installed at the time of the 2004 Indian Ocean tsunami, we use synthetic data in our assimilation experiment.

3.2. Simulation Procedure

The fault models and seismic parameters are based on the source model by Fujii and Satake (2007). They estimated the slip distributions by inverting the tide gauge and satellite altimeter data assuming a rupture velocity of 1.5 km/s. We calculated the seafloor displacement from the faulting (Okada, 1985) and used it as the initial condition for tsunami propagation (Figure 2). The linear long-wave model was employed in numerical simulation. The bathymetry grid data are derived from the General Bathymetric Chart of the Ocean with a grid size of 2 arc minute. The computation domain for Green's functions is 70–100°E, 0–25°S, with the total grid number of 675,000. In numerical simulation, the time step is 1 s, which satisfies the Courant-Friedrichs-Lewy condition, a necessary condition for stability. We stored the simulated tsunami waveforms at six ocean bottom gauges in the Bay of Bengal as the synthetic observation. Then, we used linear interpolation to compute the waveforms at 25 virtual stations with an average interval of around 50 km (Figure 2). The effects of interpolation intervals and the characteristic distances of OI are examined in the supporting information.

To validate our method of tsunami forecasting, we compared the tsunami waveforms of seven nearshore points along the coasts of India, Sri Lanka, and Maldives (Figure 2). Because the tsunami waveforms

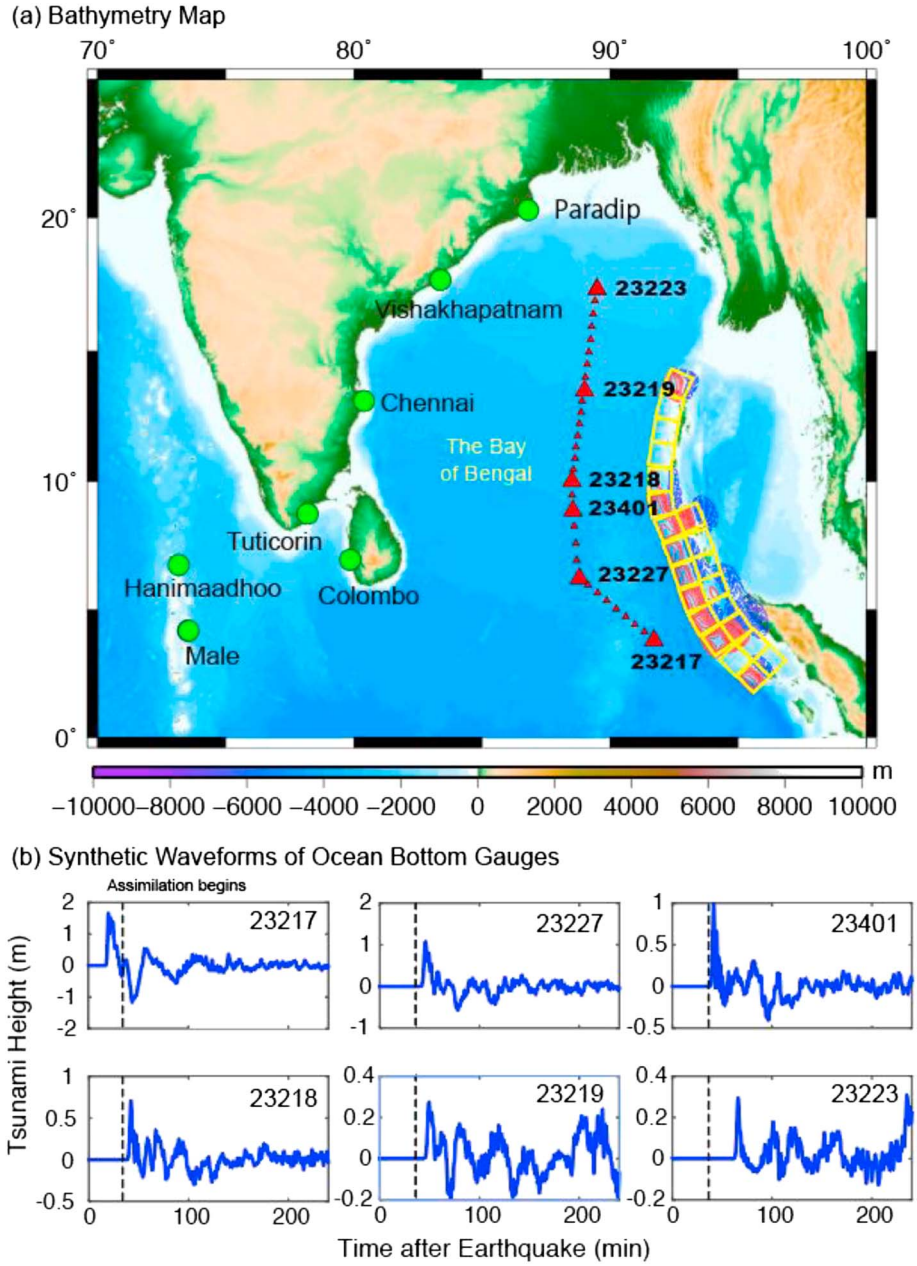

Figure 2. (a) The bathymetry map of the Bay of Bengal. The six ocean bottom gauge locations used in our numerical simulation are indicated with red large triangles. The 25 virtual stations indicated with red small triangles are interpolated between ocean bottom gauges. The nine nearshore points (green circles) record the tsunami waveform, and we compare them with the forecasted waveforms calculated by tsunami data assimilation. (b) Synthetic tsunami waveforms of six ocean bottom gauges. The assimilation begins at 30 min after the earthquake.

computed from the source model by Fujii and Satake (2007) do not match well with the real observation, probably due to inaccurate bathymetry data near the stations, we used the synthetic waveforms. The Green's functions of observations (real and virtual stations) and nearshore points were computed and stored in advance. The characteristic distance of Optimal Interpolation is 20 km.

3.3. Data Assimilation

We set the earthquake origin time as $t = 0$. When the propagating tsunami reaches the real ocean bottom gauges, in this example around 30 min at ocean bottom gauge 23217, the data assimilation process begins.

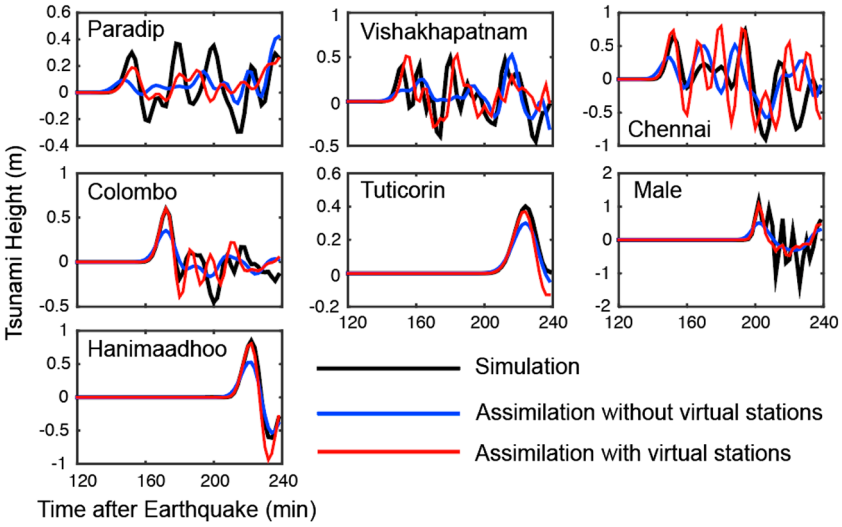

Figure 3. Comparison of the observed and forecasted waveforms at seven nearshore points. The black lines represent the simulated waveforms. The blue lines represent the assimilated waveforms without virtual stations, while the red lines represent the assimilated waveforms with virtual stations. The assimilation time window is 60 min for both cases.

The time window is defined as the period during which the observed data are used for assimilation (Wang et al., 2017). After the time window, the tsunami height and arrival time can be forecasted by the superposition of Green's functions.

Figure 3 compares the simulated and forecasted tsunami waveforms of seven nearshore stations, with an assimilation time of 60 min. The forecasted waveforms generally match with the simulated waveforms of all stations. However, without virtual stations, the forecasted waveforms underestimate the maximum amplitude of the first tsunami peak. To the contrary, the assimilation with virtual stations has a better performance, improving both amplitudes and periods of the tsunami waves. For example, in the nearshore point of Male, the simulated maximum amplitude of the first tsunami peak is 1.16 m. The amplitudes of assimilation without and with virtual stations are 1.06 and 0.51 m, respectively, indicating significant improvement of the results by inclusion of virtual stations. Overall, the forecast accuracy increases from 51.4% to 73.1% with the help of virtual stations.

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

4. Application to Real Tsunami Data

4.1. The 2009 Dusky Sound Earthquake

The Mw 7.8 Dusky Sound earthquake occurred near the southwestern coast of New Zealand, at 09:22:29 UTC on 15 July 2009 (Beavan et al., 2010; Fry et al., 2010; Heidarzadeh & Gusman, 2018; Sheehan et al., 2012). 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 (Berezina, 2017; Clark et al., 2011) and the DART gauges in the south Pacific. The OBS network in this region also detected the signal of the tsunami.

4.2. Data and Assimilation Setting

At the time of the earthquake, there were up to 30 OBSs in the west and east of the South Island (Figure 3). 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). In addition, some OBSs are too

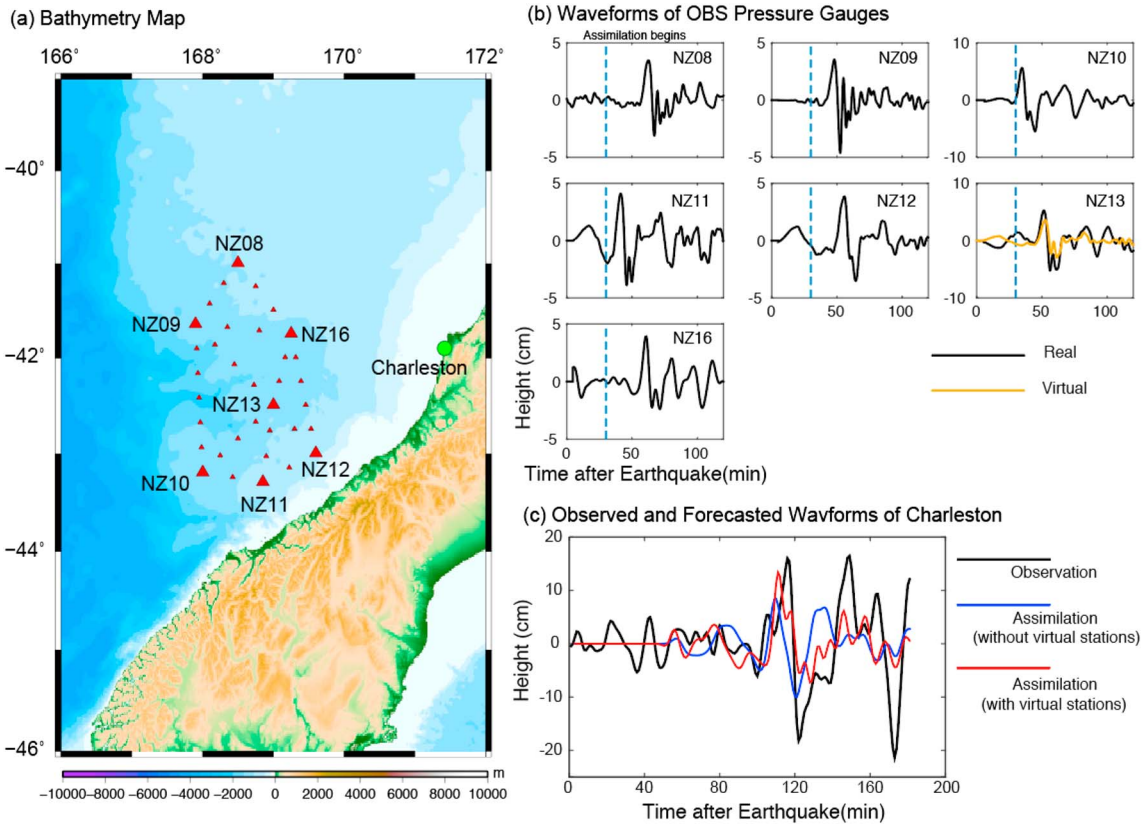

Figure 4. Tsunami data assimilation of 2009 Dusky Sound earthquake. (a) Distribution of Ocean Bottom Seismometers (OBSs) and the tide gauge Charleston. The seven OBS pressure gauges indicated with red large triangles are used for data assimilation. The 28 virtual stations indicated with red small triangles are interpolated between ocean bottom gauges. (b) The real observed tsunami waveforms of seven OBS pressure gauges. The assimilation begins at 30 min after the earthquake. The real (black line) and virtual (yellow line) waveforms of NZ13 are compared. (c) Comparison of the observed and forecasted waveforms at the tide gauge Charleston with an assimilation time window of 40 min.

far from the tide gauge, so the tsunami arrived even later than the tide gauge. In our study, we used seven OBSs for tsunami data assimilation: NZ08, NZ09, NZ10, NZ11, NZ12, NZ13, and NZ16. All OBS data are available at the Ocean Bottom Seismograph Instrument Pool website (http://www.obsip.org). The data processing is similar to the work of Sheehan et al. (2019). To extract the tsunami signals from OBSs, we first reduced aliasing of tide signals and processed the records by removing trend, removing mean, decimating, applying 5% Hann Window, and band-pass filtering from 0.0002 to 0.005 Hz using a fourth-order Butterworth filter. Then we tapered again and deconvolved the instrument responses. Because the tsunami was recorded by DPGs, finally, the tsunami amplitude at each station was corrected by the ratio between the observed and simulated peak amplitude (Beavan et al., 2010) at the station.

Tide gauge data was used for waveform comparison in order to validate our method. To remove the tide signal and high-frequency component, we applied a fourth-order band-pass filter with the frequency band of 0.000167 to 0.00333 Hz. Because our region of interest is the western coast of South Island, we only used the tide gauge observation of Charleston (Figure 4a). The tsunami amplitude of tide gauges is usually affected by local bathymetry and harbor effects (Baba et al., 2004; Heidarzadeh et al., 2016; Kontar et al., 2013; Leonard, 2006).

We used 28 virtual stations interpolated between seven real OBS pressure gauges, with the average interpolation interval of around 50 km. We chose neighboring station pairs so that virtual stations do not overlap to each other. These data were assimilated in order to forecast the tsunami waveform at the tide gauge Charleston. Because the station NZ13 is nearly located in the line between NZ09 and NZ12, it provides us with an opportunity to validate our linear interpolation method (part 2.1). We interpolated a virtual waveform of NZ13 using the data of NZ09 and NZ12 (yellow line in Figure 4b) and compared it with the real

waveform (black line). The arrival time and amplitude of the first tsunami peak are very similar between two waveforms. This validates our linear interpolation method based on Huygens-Fresnel principle. Though there are some discrepancies, virtual waveforms could be used as the supplement of real observation in tsunami data assimilation.

The characteristic distance of Optimal Interpolation is 20 km. The bathymetry grid data are also derived from the General Bathymetric Chart of the Ocean with a grid size of 0.5 arc minute. The computation domain for Green's functions is 163–175°E, 49–37°S, with the total grid number of 2,073,600. The time step of numerical simulation is 1 s. Unlike the results of Sheehan et al. (2019), we used only real OBS tsunami data as inputs. We interpolated virtual stations to overcome the problem of sparse observations.

4.3. Data Assimilation

The data assimilation process begins at 30 min after the earthquake, when the tsunami arrives at the first OBS (NZ10). Figure 4c shows the comparison between observed and forecasted waveforms of the tide gauge Charleston, 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 amplitudes of the forecasted waveform (8.4 cm) are smaller than those of the observations (16.0 cm). The accuracy is only 52.5%. Our new method with virtual stations gives a better forecasting of tsunami amplitudes (13.4 cm), with an accuracy of 83.8%. Because the tsunami arrives at Charleston 112 min after the earthquake, the tsunami forecasting is made around 32 min before arrival. By combining the tsunami data assimilation with nonlinear tsunami inundation models on coastal regions of interest (Liu et al., 2009), the inundation forecasts will also be possible.

5. Conclusions

We proposed a new method for tsunami data assimilation for regions with sparse deep-ocean tsunami observation network. In our method, we produced artificial waveforms at virtual stations by interpolating real data of available real stations. Although no new information other than water depth was adopted, we used the existing information in order to construct a relatively dense observation network. We demonstrated that for synthetic 2004 Sumatra-Andaman earthquake, our method forecasted the tsunami waveforms at the coasts of India, Sri Lanka, and Maldives, with a forecasting accuracy of more than 70%. The application to 2009 Dusky Sound, New Zealand, earthquake suggested that our method overcame the insufficient number of observation and improved the accuracy from 52.5% to 83.8%. In this study, Charleston is the only tide gauge stations on the targeted coast. Other tide gauges were either too close to be useful (tsunami arrived even before the OBS stations) or quite far in Australia. This method could be implemented for future tsunami warning systems in those regions without a dense observation network. The assimilation process costs less than 10 s by using GFTDA. Although at the present time the system does not have real-time transmission, the recorded tsunami data can be used to evaluate the performance of tsunami forecasting methods.

Acknowledgments
We thank the U.S. Ocean Bottom Seismograph Instrument Pool (OBSIP) for providing tsunami data of 2009 Dusky Sound earthquake (http://www.obsip.org), the Land and Information of New Zealand for providing the tide gauge record (https://www.linz.govt.nz/sea/tides/sea-level-data), and the General Bathymetric Chart of the Ocean (GEBCO) for the bathymetric data of the ocean floor. Y. W. thanks the Earthquake Research Institute for providing funding for his Master and PhD's Program at the University of Tokyo (Global Science Graduate Course). This work was partially supported by KAKENHI (16H01838) and KAKENHI (15K16306). We used tsunami simulation code JAGURS (Baba et al., 2015; available at https://github.com/jagurs-admin/jagurs) and the data assimilation code TDAC (Gusman et al., 2016; Maeda et al., 2015; Wang et al., 2017; available at https://github.com/takuto-maeda/tdac).

References

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

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*(12), 3455–3472. https://doi.org/10.1007/s00024-015-1049-2

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

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*(3), 1265–1286. https://doi.org/10.1111/j.1365-246X.2010.04798.x

Berezina, P. (2017). Tsunami simulation of 2009 Dusky Sound earthquake in New Zealand, UTRIP 2017 Report, The University of Tokyo.

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*(2), 237–247. https://doi.org/10.1080/00288306.2010.543690

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

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

Gonzalez, F. I., Bernard, E. N., Meinig, C., Eble, M. C., Mofjeld, H. O., & Stalin, S. (2005). The NTHMP tsunameter network. *Natural Hazards*, *35*(1), 25–39. https://doi.org/10.1007/s11069-004-2402-4

Gusman, A. R., Sheehan, A. F., Satake, K., Heidarzadeh, M., Mulia, I. E., & Maeda, T. (2016). 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., & Tsunami, H. (2014). A methodology for near-field tsunami indunation forecasting: Application to the 2011 Tohoku tsunami. *Journal of Geophysical Research: Solid Earth*, *119*, 8186–8206. https://doi.org/10.1002/2014JB010958

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

Hayes, G. P., & Furlong, K. P. (2010). Quantifying potential tsunami hazard in the Puysegur subduction zone, south of New Zealand. *Geophysical Journal International*, *183*(3), 1512–1524. https://doi.org/10.1111/j.1365-246X.2010.04808.x

Hayes, G. P., Furlong, K. P., & Ammon, C. J. (2009). Intraplate deformation adjacent to the Macquarie Ridge south of New Zealand—The tectonic evolution of a complex plate boundary. *Tectonophysics*, *463*(1-4), 1–14. https://doi.org/10.1016/j.tecto.2008.09.024

Heidarzadeh, M., & Gusman, A. R. (2018). 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* (Chap. 2, pp. 1–16). Singapore: Springer. https://doi.org/10.1007/978-981-13-0992-2_2

Heidarzadeh, M., Harada, T., Satake, K., Ishibe, T., & Gusman, A. R. (2016). 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*, 4340–4349. https://doi.org/10.1002/2016GL068601

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

Kalnay, E. (2003). *Atmospheric modeling, data assimilation and predictability*. Cambridge: Cambridge University Press.

Kontar, Y. A., Santiago-Fandiño, V., & Takahashi, T. (2013). *Tsunami events and lessons learned: Environmental and societal significance*. Berlin: Springer.

Lay, T., Kanamori, H., Ammon, C. J., Nettles, M., Ward, S. N., Aster, R. C., et al. (2005). The great Sumatra-Andaman earthquake of 26 December 2004. *Science*, *308*(5725), 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, 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*(1), 2–12. https://doi.org/10.1016/j.jseaes.2008.12.010

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 forecating. *Geophysical Research Letters*, *42*, 7923–7932. https://doi.org/10.1002/2015GL065588

Mulia, I. E., Inazu, D., Waseda, T., & Gusman, A. R. (2017). 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*(7043), 756–757. https://doi.org/10.1038/nature435756a

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., & Eblé, M. C. (2015). Deep-ocean measurements of tsunami waves. *Pure and Applied Geophysics*, *172*(12), 3281–3312. https://doi.org/10.1007/s00024-015-1058-1

Satake, K. (1995). Linear and nonlinear computations of the 1992 Nicaragua earthquake tsunami. *Pure and Applied Geophysics*, *144*(3-4), 455–470. https://doi.org/10.1007/BF00874378

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

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*. https://doi.org/10.1029/2018JB016575

Sheehan, A. F., Yang, Z., Nicolsky, D., Mungov, G., & Eakins, B. (2012). Exploring tsunamis with non-traditional dataset: Array recordings from temporary ocean-bottom seismic experiment, American Geophysical Union Fall Meeting 2012, NH33A-1562.

Titov, V. V., Gonzalez, F. I., Bernard, E. N., Eble, M. C., Mofjeld, H. O., Newman, J. C., & Venturato, A. J. (2005). Real-time tsunami forecasting: Challenges and solutions. *Natural Hazards*, *35*(1), 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*, *114*, B06309. https://doi.org/10.1029/2008JB005988

Wang, Y., Satake, K., Maeda, T., & Gusman, A. R. (2017). Green's function-based tsunami data assimilation: A fast data assimilation approach toward tsunami early warning. *Geophysical Research Letters*, *44*, 10,282–10,289. https://doi.org/10.1002/2017GL075307

Yang, Y. E. M., Dunham, G. B., & Almquist, 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