Tsunami data assimilation of Cascadia seafloor pressure gauge records from the 2012 Haida Gwaii earthquake

Aditya R. Gusman, Anne F. Sheehan, Kenji Satake, Mohammad Heidarzadeh, Iyan Eka Mulia, and Takuto Maeda

1Earthquake Research Institute, University of Tokyo, Tokyo, Japan, 2Department of Geological Sciences and Cooperative Institute for Research in Environmental Sciences, University of Colorado Boulder, Boulder, Colorado, USA, 3Port and Airport Research Institute, Yokosuka, Japan, 4Department of Ocean Civil Engineering, Kagoshima University, Kagoshima, Japan

Abstract We use tsunami waveforms recorded on a dense array of seafloor pressure gauges offshore Oregon and California from the 2012 Haida Gwaii, Canada, earthquake to simulate the performance of two different real-time tsunami-forecasting methods. In the first method, the tsunami source is first estimated by inversion of recorded tsunami waveforms. In the second method, the array data are assimilated to reproduce tsunami wavefields. These estimates can be used for forecasting tsunami on the coast. The dense seafloor array provides critical data for both methods to produce timely (>30 min lead time) and accurate in both timing and amplitude (>94% confidence) tsunami forecasts. Real-time tsunami data on dense arrays and data assimilation can be tested as a possible new generation tsunami warning system.

1. Introduction

Conventional tsunami warning consists of seismological observations and tsunami numerical simulations. Results of tsunami simulations, from either fault slips [Titov et al., 2005; Lorito et al., 2011; Wei et al., 2013; Satake et al., 2013; Melgar et al., 2016] or initial sea surface elevations [Saito et al., 2010; Tsushima et al., 2011; Mulia and Asano, 2015], are usually stored in a database and used for forecasting tsunamis on coasts. The fault slip can be estimated from seismological, geodetic, or tsunami observations, while the sea surface elevation can be directly estimated only from tsunami observations. The Deep-ocean Assessment and Reporting of Tsunamis (DART) buoy systems made a reliable global tsunami warning possible for far-field destinations [Titov, 2009; Tang et al., 2009]. The DART buoys encircle the Pacific Ocean at locations where large tsunamigenic earthquakes might occur, including the Cascadia subduction zone (Figure 1), but the network is sparse with station intervals >300 km. The DART buoy systems are equipped with seafloor pressure gauges, and they are sending the data with sampling interval of 15 and 60 s during the event reporting mode in real time through satellite telemetry [Eblé and Gonzalez, 1991; Gonzalez et al., 1998; Synolakis and Bernard, 2006; Rabinovich and Eblé, 2015].

Recent addition of pressure gauges to ocean bottom seismometers (OBSs) provides alternative data for tsunami forecast simulation. Because OBSs are usually deployed in a dense array, they provide high-density tsunami observations as well. In the southern Cascadia subduction zone, ocean bottom seismometers deployed at nonpermanent locations were equipped with absolute pressure gauges (APGs) or differential pressure gauges (DPGs) [Toomey et al., 2014]. The APGs and DPGs were spaced from 10 to 50 km providing a very dense observation of seafloor pressure (Figure 1). The APG utilizes a quartz transducer with oscillation period related to stress and thus to pressure [Houston and Paros, 1998]. The DPG is a pressure gauge configured to respond to the pressure difference between the ocean and the fluid in the reference chamber [Cox et al., 1984]. The APG and DPG have sampling rates of 40–125 Hz, much higher than that of DART system, but the data are stored in a hard drive without real-time data transmission. Once the pressure data are retrieved, it can then be converted to water column depth.

The 28 October 2012 (03:04:09 UTC) Haida Gwaii earthquake (Mw 7.8) initiated at 52.622°N, 132.103°W, at a depth of 14 km [Kao et al., 2015], and ruptured the Pacific and North American plate boundary all the way up to the trench axis with a thrust fault motion [Lay et al., 2013; Lorito et al., 2016]. Generated by the second largest recorded earthquake offshore British Columbia (Canada), the tsunami ran up to 13 m in the near field [Leonard and Bednarski, 2014; Fine et al., 2015]. The tsunami was recorded on DART stations as well as on the
Cascadia Initiative OBS array. A total of 57 tsunami waveforms were reported, including 8 DARTs, 19 APGs provided by Lamont-Doherty Earth Observatory (LDEO), 9 DPGs provided by Scripps Institution of Oceanography (SIO), and 21 DPGs provided by Woods Hole Oceanographic Institution (WHOI) [Sheehan et al., 2015].

Here we use the tsunami waveforms recorded on the Cascadia OBS array [Sheehan et al., 2015] to demonstrate two different approaches for tsunami forecast: (1) estimation of the fault slip distribution of the 2012 Haida Gwaii earthquake by tsunami waveform inversion and then forecasting the coastal tsunami heights by numerical forward modeling and (2) progressive assimilation of the tsunami waveforms recorded in the array, reproduction of wavefields in the vicinity of the array, and then forecasting of wavefields by numerical modeling. Despite the fact that a progressive tsunami data inversion is often adopted in the first method [Titov et al., 2005; Wei et al., 2013; Tang et al., 2016], we use the final tsunami waveforms to accurately estimate the tsunami source for comparison with the data assimilation method. We will show that tsunami observations from a dense array can be continuously assimilated to update the wavefield at each time step and to forecast the coastal tsunami in the vicinity of the array, without assuming a tsunami source [Maeda et al., 2015].

Figure 1. Station map of the tsunami dense array. Distribution of DART, APG, and DPG stations in the Cascadia (CSZ) and Aleutian subduction zones (ASZ). Light blue star represents the earthquake's epicenter.
2. Data and Methodology

2.1. Observed Tsunami Waveforms

The tsunami waveforms used in this research were initially processed as described in Sheehan et al. [2015]. The tsunami amplitudes recorded by the DPGs are less reliable than those recorded by the DARTs and APGs. Therefore, only 27 tsunami waveforms at DART and APG stations are used for waveform inversion (Figure 2b). Because the DPG waveforms are accurate in terms of the tsunami arrival times and wave periods [Sheehan et al., 2015; Lin et al., 2015], we correct the observed amplitudes of 30 DPG records based on tsunami simulations from the source model. This correction makes it possible to use DPG data together with the DART and APG data to fulfill the requirement on the data’s spatial coverage for the tsunami data assimilation method.

2.2. Fault Slip Inversion

For tsunami waveform inversion [Satake et al., 2013; Gusman et al., 2015], we arrange 11 × 4 subfaults with size of 15 km × 15 km and the total fault length and width of 165 km and 60 km (Figure 3a). The focal mechanism of the 2012 Haida Gwaii earthquake based on W phase centroid moment tensor solution, strike = 317°, dip = 18.5°, and rake = 103.3° [Lay et al., 2013], is assumed. The total seafloor displacement for each subfault is calculated from the vertical displacement from the faulting [Okada, 1985] and additional vertical movement due to the horizontal displacement of the seafloor slope (bathymetric slope displacement effect) [Tanioka and Satake, 1996]. The seafloor displacement is converted to sea surface displacement by the equations in Kajjura [1963]. To produce synthetic tsunami waveforms from each subfault, or the Green’s function for the inversion, the tsunami waveforms computed from linear long wave simulation on the spherical coordinate system [Satake, 1995] are corrected by using a frequency-dependent phase correction method [Watada et al., 2014]. The usage of the relatively new phase correction method to build tsunami Green’s function for fault slip inversion was also demonstrated in Gusman et al. [2015]. The bathymetry data for the tsunami simulation have a grid spacing of 1 arc min, resampled from the 30 arc sec bathymetry grid of GEBCO-08 digital atlas [Intergovernmental Oceanographic Commission et al., 2003].

2.3. Tsunami Data Assimilation Method

To estimate the tsunami wavefield, we use the observed tsunami waveforms in a tsunami data assimilation technique which is based on the optimal interpolation method and an assumption of a linear system.

Figure 2. Tsunami simulation result. (a) Plots of simulated against observed peak amplitudes at SIO and WHOI stations. The ratios between the simulated and observed peak amplitudes are used for correction of DPG waveforms. (b) Comparison of observed/corrected (gray lines) and simulated tsunami waveforms from the estimated slip distribution (red lines). Station names at which tsunami waveforms are used in the inversion are written in blue.
Progressive data assimilation has been used for many years in weather forecasting [e.g., Miller et al., 1994; Xue et al., 2003]. At every time step of 1 s, first the wavefield for the current time step \( \mathbf{X}_t \) is simulated by numerically solving the shallow water equations using the wavefield in the previous time step (\( \mathbf{X}_t \equiv \mathfrak{F} \mathbf{X}_a/n \)). From the observed tsunami amplitude at the station \( y_n \), the residual at the current time step from the simulated wavefield is calculated as (\( y_n - \mathcal{H} \mathbf{X}_f \)). This residual is used to correct for the assimilated wavefield \( \mathbf{X}_a \) through a smoothing matrix \( \mathbf{W} \) as

\[
\mathbf{X}_a = \mathbf{X}_f + \mathbf{W} [y_n - \mathcal{H} \mathbf{X}_f].
\]

This smoothing matrix is an important controlling factor as it transmits the information of tsunami amplitude from the station to the surrounding area. The smoothing matrix does not change with time and depends only on the station distribution. We assume that the smoothing matrix has a cutoff distance of 10 km from the station. More details on the computation steps and how to construct the smoothing matrix can be seen in Maeda et al. [2015].

The computation domain for tsunami data assimilation is on the Cartesian coordinate system that includes a geographical area from 39° to 46°N and from 123° to 131°W, with a grid spacing of 2000 m. This grid spacing is selected to be close to the 1 arc min grid spacing used in the computation of the Green’s function for tsunami waveform inversion.

### 3. Results and Discussion

#### 3.1. Source Model of the 2012 Haida Gwaii Earthquake

The fault slip distribution estimated by tsunami waveform inversion shows largest slip (5.5 m) near the trench axis and moderate slip (~3 m) on the plate interface southeast of the epicenter and beneath the Queen Charlotte Fault (QCf) (Figure 3a and supporting information Table S1). The seismic moment from the estimated fault slip distribution is calculated, assuming rigidity of \( 4 \times 10^{10} \text{ N/m}^2 \), as \( 5.1 \times 10^{20} \text{ N m} \) (\( \text{M}_w 7.8 \)), which is consistent with the Global CMT (centroid moment tensor) solution (\( 5.2 \times 10^{20} \text{ N m} \)). The estimated slip distribution produces large (up to 1.8 m) sea surface uplift near the trench, with the total uplifted area of ~140 km long and ~30 km wide (Figure 3b). Maximum subsidence of ~0.3 m on Haida Gwaii islands is much smaller than the maximum uplift. The estimated slip southeast of the epicenter below the QCf is significantly larger than previous estimations from a combination of seismic and DART data [Lay et al., 2013] and from GPS data [Nykolaishen et al., 2015]. Although we use only tsunami data, the computed displacements from our model at the GPS stations match with the observations (see supporting information Figure S1).
The horizontal coseismic movement of the steep slope [Tanioka and Satake, 1996] also contributed to the tsunami generation. In particular, the sea surface displacement near the west coast of Haida Gwaii is almost entirely from the horizontal motion of the steep slope, i.e., bathymetric slope displacement effect, rather than vertical displacement from faulting (see supporting information Figure S2c). The potential energy [Satake and Kanamori, 1991] calculated from the bathymetric slope displacement effect is $0.12 \times 10^{13}$ J, while that from pure faulting is $2.20 \times 10^{13}$ J; about 5% of the total potential energy is due to the bathymetric slope displacement effect. From the estimated sea surface displacement, we simulate tsunami waveforms at the DPG stations. Plots of simulated versus observed peak amplitudes for the SIO-DPG and WHOI-DPG instruments (Figure 2a) suggest systematic error in the DPG amplitudes. By assuming a flat gain of instrumental response within the tsunami frequency band, the ratio between simulated and observed peak amplitudes at each DPG station is defined as the amplitude correction coefficient (Figure 2a). The waveforms are then corrected by applying those coefficients to the original waveforms (Figure 2b). Once corrected, we can use all 57 tsunami waveforms for tsunami forecast by data assimilation approach.

### 3.2. Data-Assimilated Waveform for Tsunami Forecast

The initial tsunami phase reached the northernmost D46404 (DART) station (Figure 4a) of the Cascadia array approximately 70 min after the earthquake’s origin time. We use tsunami data assimilation [Maeda et al., 2015] to estimate wavefields at every time step (1 s) in the vicinity of the dense array. For tsunami forecast by the data assimilation method, we do not use any information about the tsunami source and no tsunami energy is transmitted through the modeling boundaries. Therefore, we cannot expect accurate wavefields at the beginning of the data assimilation process until the first cycles of tsunami passes through several stations. The estimated wavefields in Figure 4a show that as more data are assimilated, more realistic wavefields emerge. After the first tsunami cycle passes through five stations in the north of the modeling domain at $t = 110$ min, the general pattern of a realistic tsunami wavefield begins to emerge (Figure 4a and Movie S1). The timing and amplitude of the approaching tsunami toward the coast from the data assimilation method using 130 min of data (Figure 4a) are similar to the ones simulated from the estimated initial tsunami source (Figure 4b). However, the later tsunami waves with much shorter wave periods could not be precisely reproduced by the data assimilation method because the spacing of the stations is relatively sparse compared to the wavelength of the later phases. The snapshots of wavefields also show how denser stations between 40° and 44°N improve the results from the data assimilation method (Figures 4a and 4b).

### 3.3. Tsunami Forecast

To evaluate the performance of the forecast algorithm using the data assimilation method, the simulated tsunami amplitudes are compared with the observations. We use the geometric mean ratio ($K$) of observation ($O$) and simulation ($S$) approach [Aida, 1978] (equation (2)) to calculate the forecast accuracy (equation (3)).

\[
\log(K) = \frac{1}{N} \sum_{i=1}^{N} \log \left( \frac{O_i}{S_i} \right)
\]

\[
\text{Accuracy(\%)} = \begin{cases} 
100 \times K, & K \geq 1 \\
100 \times \frac{1}{K}, & K < 1
\end{cases}
\]

We use the resulting tsunami wavefield at every 10 min from 70 to 150 min after the earthquake origin time as an input for tsunami simulation. The forecast accuracy versus the length of data used for assimilation is shown in Figures 4c and 4d. High accuracies of more than 94% on average are produced from data assimilation wavefield at stations near the shoreline (Figure 4e and supporting information Figure S3). As an example, using the 130 min data-assimilated wavefield, the tsunami amplitudes at station FS12B (Figure 4c) are forecasted with an accuracy of 98% about 30 min in advance. At this time, however, the computation accuracies at stations in the northern part of the modeling domain near the boundary are much lower, decreasing the overall accuracy down to 76% (Figure 4d). Because our aim is to provide a reliable tsunami warning for coastal areas, only the accuracy of the predicted tsunami at stations close to the coast is important (Figure 4e). The data-assimilated wavefield gives a good prediction of initial tsunami phase, including the observed small initial negative depressions (Figure 4c and supporting information Figure S3). These results imply that the current distribution of offshore pressure gauge stations (Figure 1) is enough to accurately forecast the tsunami along the shore (Figure 4).
Moreover, the forecasted tsunami waveforms can be used as input for a forecasting system that employs a precomputed tsunami database to produce high-resolution inundation maps in a couple of minutes [e.g., Gusman et al., 2014]. Forward numerical simulation with supercomputers can also be used to produce high-resolution tsunami inundation maps [Oishi et al., 2015].

The accuracy of the tsunami forecast strongly depends on the spatial distribution of the stations. A denser array would predict the wavefield between the stations both accurately and quickly, which in turn widens the lead time of an accurate forecast. Although the APG and DPG data during the 2012 Haida Gwaii tsunami were not transmitted in real time, our retrospective data assimilation demonstrates the capability of such a dense tsunami array to forecast an incoming tsunami. Real-time tsunami observation technologies, such as the cabled offshore dense tsunami array of S-net (about 150 stations spaced at 30–50 km intervals) that is being deployed in the Japan subduction zone [Saito, 2013; Maeda et al., 2015], would provide data required for real-time tsunami forecasts using the methods presented in this paper.

4. Conclusions

A tsunami waveform inversion of the 2012 Haida Gwaii earthquake using 27 offshore tsunami waveforms gives a precise initial sea surface elevation. Large slip patches are detected on the plate interface near the
Haida Gwaii trench and also beneath the Queen Charlotte Fault (QCF). We also estimated that there is no significant slip on the deep plate interface north of QCF. Such a detailed fault slip distribution model can still be obtained even though no near-field tsunami observation is used in the inversion.

We demonstrate that tsunami records from the 2012 Haida Gwaii earthquake on a dense pressure gauge array in southern Cascadia can deliver both timely and accurate tsunami forecasts in the nearby coast. The tsunami forecast from the tsunami data assimilation method produces similar results as those from the traditional tsunami-forecasting method which starts from a fault model. The tsunami data assimilation method that we present can be run continuously in real time and does not require a tsunami source model. The method can be tested further for various configurations of tsunami source and coast to be implemented for future tsunami warning systems.

Acknowledgments
The sea bottom pressure data at the Ocean Bottom Seismograph stations used in this research were provided by instruments from the Ocean Bottom Seismograph Instrument Pool (https://www.obsip.org) which is funded by the U.S. National Science Foundation. The authors are grateful to the Cascadia Initiative Expedition Team for acquiring the Amphibious Array Ocean Bottom Seismograph data and appreciate the open data policy that made the data available shortly after they were acquired. The facilities of the IRIS Data Management System were used to access the data used in this study. The IRIS DMS is funded through the U.S. National Science Foundation under cooperative agreement EAR-0552316. The sea level observations at the DART stations used in this research were provided by the National Oceanic and Atmospheric Administration and publicly available from http://ncdc.noaa.gov/IRIS/ Dart/. Some of the figures were made using the public domain Generic Mapping Tools [Wessel and Smith, 1998]. We thank the Editor, Andrew V. Newman, an anonymous reviewer, and Stefano Lorito for their constructive comments and suggestions to this article.

References


