Far-field tsunami data assimilation for the 2015 Illapel earthquake

Y. Wang [1,2] K. Satake,[1] R. Cienfuegos,[2,3] M. Quiroz [2,3] and P. Navarrete[2,4]

[1]*Earthquake Research Institute, The University of Tokyo, Tokyo, Japan. E-mail: ywang@eri.u-tokyo.ac.jp*
[2]*Centro de Investigación para la Gestión Integrada del Riesgo de Desastres (CIGIDEN), CONICYT/FONDAP/1511007, Santiago, Chile*
[3]*Departamento de Ingeniería Hidráulica y Ambiental, Escuela de Ingeniería, Pontificia Universidad Católica de Chile, Santiago, Chile*
[4]*Departamento de Ingeniería Civil Industrial Matemática, Escuela de Ingeniería, Pontificia Universidad Católica de Chile, Santiago, Chile*

SUMMARY

The 2015 Illapel earthquake (M_w 8.3) occurred off central Chile on September 16, and generated a tsunami that propagated across the Pacific Ocean. The tsunami was recorded on tide gauges and Deep-ocean Assessment and Reporting of Tsunami (DART) tsunameters in east Pacific. Near-field and far-field tsunami forecasts were issued based on the estimation of seismic source parameters. In this study, we retroactively evaluate the potentiality of forecasting this tsunami in the far field based solely on tsunami data assimilation from DART tsunameters. Since there are limited number of DART buoys, virtual stations are assumed by interpolation to construct a more complete tsunami wavefront for data assimilation. The comparison between forecasted and observed tsunami waveforms suggests that our method accurately forecasts the tsunami amplitudes and arrival time in the east Pacific. This approach could be a complementary method of current tsunami warning systems based on seismic observations.

Key words: Tsunamis; South America; Earthquake hazards; Tsunami warning.

1 INTRODUCTION

The Chilean subduction zone is one of the world's most seismically active areas, which has been struck by great earthquakes. The largest earthquake ever recorded is the 1960 Chile earthquake (M_w 9.5), which generated a catastrophic trans-Pacific tsunami (Barrientos & Ward 1990; Fujii & Satake 2013; Ho *et al.* 2019). In the past 10 yr, great earthquakes have struck frequently, including the 2010 Maule earthquake (M_w 8.8; Fujii & Satake 2013; Yoshimoto *et al.* 2016), the 2014 Iquique earthquake (M_w 8.2; Gusman *et al.* 2015), and the 2015 Illapel earthquake (M_w 8.3; Heidarzadeh *et al.* 2016; Satake & Heidarzadeh 2017). The 16 September 2015 Illapel earthquake occurred at 22:54:33 UTC offshore Illapel, as the result of thrust faulting on the interface between the Nazca and South America Plates in central Chile. The West Coast & Alaska Tsunami Warning Center (WC/ATWC) of National Oceanic and Atmospheric Administration (NOAA) issued the rapid magnitude estimation of M_w 7.2 at 4 min after the earthquake, and the Pacific Tsunami Warning Center (PTWC) estimated the magnitude to be M_w 7.9 at 5 min after the earthquake (Cienfuegos *et al.* 2016). Finally, the United States Geological Survey (USGS) revised the magnitude to be M_w 8.3. The location of epicentre is 31.573°S, 71.674°W, at the depth of 22.4 km. The earthquake caused 15 deaths, 6 missing and large damage to coastal cities. It generated a tsunami that reached the Chilean coastal region within 8 min after the earthquake (Aránguiz *et al.* 2016; An & Meng 2017), and resulted in flooding in the cities along the coast. The tsunami propagated across the Pacific Ocean, and was recorded by tide gauges and Deep-ocean Assessment and Reporting of Tsunami (DART) tsunameters, though the amplitude was much smaller than the 2010 tsunami and slightly smaller than the 2014 tsunami (Satake & Heidarzadeh 2017). The DART tsunameters provided tsunami time series recorded at open ocean, and were used to reconstruct the source model of the earthquake (Heidarzadeh *et al.* 2016; Williamson *et al.* 2017). The characteristics of tsunami waves in the deep water was studied by using 16 records of DART tsunameters of this event (Ren *et al.* 2017).

In order to mitigate the tsunami damage from large earthquakes occurring in the Chilean and other subduction zones, rapid seismic source estimation systems based on real-time seismic observations have been developed and operated. For example, the real-time W phase source inversion algorithm is adopted by PTWC (Duputel *et al.* 2011, 2012; Wang *et al.* 2012). The W phase is a fast, long-period seismic wave that is used for rapid point source moment tensor inversions (Kanamori & Rivera 2008; Benavente *et al.* 2016). The W phase inversion could be completed within 5–10 min when seismic data in regional distances are used. Such characterization is helpful for tsunami early warning. In the event of the 2015 Illapel earthquake, the PTWC issued tsunami early warnings to the Chilean coast after the earthquake, and also issued tsunami warnings or advisories for other regions like Easter Island, New Zealand, Fiji, Hawaii, Solomon Islands, but later it downgraded the warnings (PTWC 2015).

However, real-time estimate for seismic rupture model often has large uncertainties in the characterization of the sources (Cienfuegos *et al.* 2018). In order to directly forecast the tsunamis, other methods have been proposed, for example, tsunami Forecasting based on Inversion for initial sea-Surface Height (tFISH, Tsushima *et al.* 2009; Tsushima & Ohta 2014), Time Reverse Imaging (TRI, Hossen *et al.* 2015a; Hossen *et al.* 2015b), and optimal interpolation (Maeda *et al.* 2015; Gusman *et al.* 2016; Mulia *et al.* 2017; Wang *et al.* 2017), amplification factors method (Baba *et al.* 2014), multi-index method (Yamamoto *et al.* 2016), tsunami prediction system using the Dense Oceanfloor Network System for Earthquakes and Tsunamis (DONET; Takahashi *et al.* 2017). These methods avoid the uncertainties in the seismic source estimation. The application of tFISH and TRI still requires the earthquake origin time, but optimal interpolation and other methods based solely on tsunami data. Optimal interpolation method was first introduced in data assimilation for numerical weather forecasting (Barker 1992; Kalnay 2003), and was applied in tsunami forecasting by Maeda *et al.* (2015). It directly assimilates observed data to forecast the tsunami height and arrival time, and does not need any assumptions on the tsunami source. Previously, a dense observation network was required for optimal interpolation (Mulia *et al.* 2017). Recently, by using interpolation and assuming waveforms at virtual stations, tsunami data on a sparse observation network could also be utilized in tsunami forecasting. This method has been applied to the tsunami data of the 2009 Dusky Sound earthquake off New Zealand recorded at seven Ocean Bottom Seismometer (OBS) equipped with pressure gauges (Wang *et al.* 2019).

In this paper, we focus on far-field tsunami forecasting of Chilean earthquakes. Tsunamis generated by Chilean earthquakes can produce significant damage not only on 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 2015). Recent installation of DART tsunameters off the Chilean coast enables us to apply the tsunami data assimilation approach. The DART stations are: Stations 32 401 and 32 402 owned and maintained by the Hydrographic and Oceanographic Service of the Chilean Navy (SHOA), Stations 32 403, 32 404 and 34 420 owned and maintained by the Cooperative Effort DART 4 G Buoy. They record water levels using ocean bottom pressure sensors and transmit the signal to the land in real time (González *et al.* 2005) for the purpose of real-time tsunami forecasting. In our study, we use tsunami data from the 2015 Illapel earthquake as an application of data assimilation. We retroactively predict the far-field tsunami in the east Pacific by tsunami data assimilation approach (optimal interpolation), and compare modeled results with the observations.

2 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.

2.1 DART data

The DART data of the 2015 Illapel earthquake are available at the website of NOAA: https://www.ngdc.noaa.gov/hazard/dart/2015chile.html. Tsunami signals have already been extracted and processed by the Tsunami Detection Algorithm (Meinig *et al.* 2005). It estimated the amplitudes of the pressure fluctuations within the

tsunami frequency band. Then, the amplitudes were computed by subtracting predicted pressures from the observations, in which the predictions closely match the tides and lower frequency fluctuations. In our retroactive study, in order to remove the high frequency components like seismic signals, we also use a low-pass filter with a cut-off frequency of 0.002 Hz, and apply a zero mask before arriving tsunami, because we only need the waveforms after the tsunami arrival. The tsunami first arrival is defined with a threshold of 0.02 m. In the case of real-time practice, we could remove the tidal signal by subtracting a theoretical tide model (Tsushima *et al.* 2009) or polynomial fitting, and remove the high-frequency components by a moving average. For most tsunamis generated from mega-thrust earthquakes, the first peak of the tsunami at pressure gauges is free of coastal reflections and harbor effects (Williamson & Newman 2018).

At the time of the 2015 Illapel earthquake, only two DART stations 32 401 and 32 402 were operational. Therefore, in addition to these two observed waveforms (blue lines), we build synthetic waveforms (green lines) at stations 32 404 and 34 420 (Fig. 1b) by calculating from the fault model of Ren *et al.* (2017). We only compute synthetic waveforms at two other stations in order to ensure that at least half of the data are real record. Therefore, we skip the station 32 403, which is close to station 32 401. In the research of Ren *et al.* (2017), they assume a focal depth of 25 km, strike angle of 4°, dip angle of 19°, and rake angle of 90°. The fault size (length x width) is 230 km x 100 km. A uniform slip of 4.64 m is adopted. The numerical simulation from this uniform fault matches well with the real observation in near-field and far-field DART stations.

In addition to DART stations near the Chilean coast, we also extract the data of Station 32 412 located off the southwest of Peru. We compare its observations with the forecasted waveform in order to validate our method.

2.2 Tide gauge data

The tide gauge data are available at the website of Sea Level Station Monitoring Facility, Intergovernmental Oceanographic Commission: http://www.ioc-sealevelmonitoring.org/. The tidal signal is removed by polynomial fitting. We use the far-field tsunami record of tide gauges in the east Pacific, including San Felix, Easter, Rikitea, Nuku Hiva Island, Papeete Tahiti, Huahine and Rarotonga. The distance between tide gauges and the South America continent ranges from hundreds to ten thousand kilometres.

3 METHOD

3.1 Tsunami data assimilation

To estimate the tsunami wavefield, we employ the tsunami data assimilation approach (optimal interpolation). The optimal interpolation algorithm (Maeda *et al.* 2015) is composed of a forecasting step and an assimilation step. In the forecasting step:

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

where $\boldsymbol{x}_n^a$ and $\boldsymbol{x}_n^f$ are the assimilated and forecasted tsunami wavefields at the *n*th time step. Their components include tsunami height and tsunami velocities in two directions. $\boldsymbol{F}$ is the propagation matrix depending on the tsunami propagation model. The simplest model is the linear long-wave model based on the shallow-water approximation (Satake 1995), while for far-field tsunamis, the dispersive characteristics is important (Watada *et al.* 2014). In our study, we
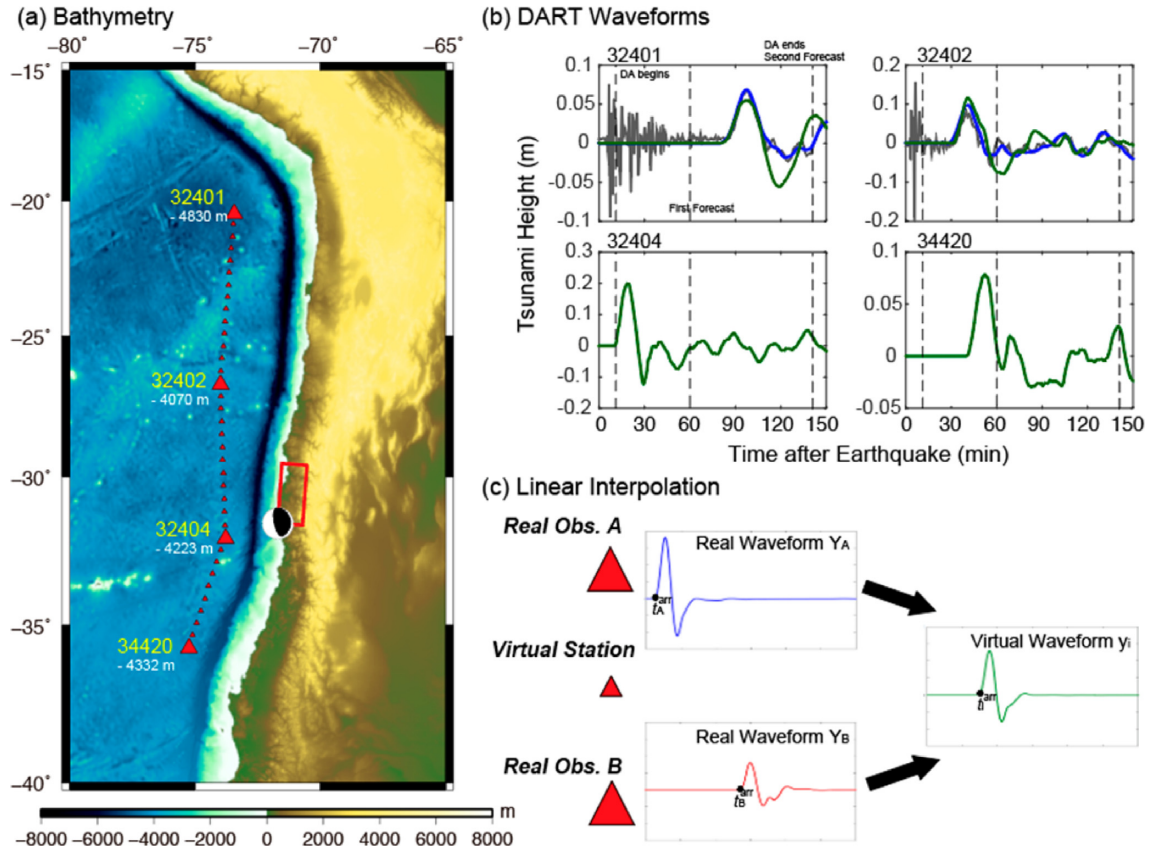
Figure 1. Tsunami data assimilation of DART tsunameters along the Chilean coast. (a) Bathymetry of the source region. We assimilate the data at four real DART stations and 32 virtual stations interpolated between them. The focal mechanism is plotted by using the USGS solution. The fault model of Ren *et al.* (2017) is plotted with a red rectangle. (b) Waveforms at the four DART stations. Grey lines represent the raw data from the website of NOAA, blue lines represent observed waveforms, and green lines represent synthetic waveforms. We use the real observed waveforms of stations 32 401 and 32 402, and synthetic observed waveforms of stations 32 404 and 34 420. The synthetic observed waveforms are calculated from the fault model of Ren *et al.* (2017). We skip the station 32 403. The time for the beginning and ending of data assimilation, and for making the tsunami forecast are marked by the dash lines. (c) The process of linear interpolation to create waveforms at virtual stations (Wang *et al.* 2019). We calculate the virtual waveforms by shifting the arrival times considering the distance and correct the amplitudes considering the water depths at the stations.

apply the linear dispersive model (Saito *et al.* 2010; Wang *et al.* 2018).

In the assimilation step,

$$x_n^a = x_n^f + PH^T\left(R + HPH^T\right)^{-1}(y_n - Hx_n^f), \qquad (2)$$

where $P = \langle \varepsilon^f \varepsilon^{fT} \rangle$ and $R = \langle \varepsilon^O \varepsilon^{OT} \rangle$ are the error covariance matrices of forward numerical simulation and the observations. Here ε^f represents the errors in numerical forecast between two numerical grids, and ε^O represents the observational errors of the tsunameter stations. The errors are assumed to have a Gaussian distribution with characteristic distance of 20 km (Maeda *et al.* 2015; Gusman *et al.* 2016). The selection of characteristic distance affects the forecasting results. It was discussed in detail in Wang *et al.* (2019). In our research we just use an appropriate value of 20 km. H is the observational operator and H^T is its transpose matrix. y_n is the vector of real tsunami observations.

During the data assimilation process, eqs (1) and (2) are simulated alternatively. To speed up the assimilation process, we used the technique of Green's Function-based Tsunami Data Assimilation (GFTDA) proposed by Wang *et al.* (2017).

3.2 Virtual stations

In order to solve the problem of sparse observations, we apply the method of Wang *et al.* (2019), using a linear interpolation of real data to compute artificial waveforms at virtual stations. The employment of virtual stations is based on the Huygens–Fresnel principle, that every point on a wavefront is itself the source of new spherical wavelets. If there are enough source points, the wavefront at the next time step could be equivalently reproduced by these new spherical wavelets.

Similarly, in tsunami data assimilation, the observation stations resemble the points on the optical wavefront. The tsunami wavefront at the next time step is formed by the assimilation of observation stations (Maeda *et al.* 2015). To reproduce the tsunami wavefield, we need a wavefront that is densely sampled (Wang *et al.* 2019). Therefore, we apply linear interpolation to construct virtual waveforms artificially, in order the overcome the problem of sparse real stations. The virtual stations are located along a straight line between the two neighbouring real stations. An interpolation interval of ∼50 km can keep the balance between forecasting accuracy and the computational cost. It is also compatible with the characteristic distance of the tsunami data assimilation algorithm (Wang *et al.* 2019).

The arrival time and the tsunami waveform of the ith station are computed by the following equations:

$$t_i^{arr} = w_{iA} \cdot t_A^{arr} + w_{iB} \cdot t_B^{arr} \tag{3}$$

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

where t_A^{arr} and t_B^{arr} are arrival times of two neighbouring real stations, respectively. w_{iA} and w_{iB} are weight parameters for linear interpolation, depending on the relative distance between the virtual and real stations. d_A, d_B and d_i are the water depth of two real stations and the ith virtual station, respectively. The tsunami arrival time and tsunami waveform are calculated separately. Instead of calculating a simple average of tsunami waveform, we should firstly compute the arrival time by linear interpolation. That is to say, we calculate the virtual waveforms by shifting the arrival times considering the distance (Wang *et al.* 2019). In eq. (4), it is also important to correct the tsunami height considering the water depths at the stations. According to the Green's Law, the tsunami amplitude is inversely proportional to the fourth root of water depth (Satake 2015). Finally, both the real observed waveforms, the synthetic observed waveforms and the computed waveforms of virtual stations are assimilated in the tsunami data assimilation approach. The linear interpolation depends on the relative distance between the virtual station and real observation. The virtual waveforms after interpolation could be used for data assimilation with any tsunami propagation model. We do not add new information of tsunami data to the assimilation algorithm. Instead, we use the existing data to generate virtual waveforms in order to improve the forecasting accuracy. In our research, station 32 404 is close to the earthquake source. We acknowledge that the virtual waveform calculated by our linear interpolation method may not be exactly the same as the real observations, especially if the earthquake source is near the bisection of two real stations; this method may not capture the largest amplitude, but it can still improve the performance of data assimilation.

In our study, the waveforms at the four DART tsunameters are adopted as the input for the tsunami data assimilation (Fig. 1b). We interpolate them at 32 virtual stations between the neighboring DART tsunameters (Fig. 1a). To prove the effectiveness of our method, we also compare the data assimilation results with and without virtual stations.

3.3 Simulation settings

In our simulation, the Green's functions are calculated by the JAGURS tsunami code (Bata *et al.* 2014) with a linear dispersive tsunami propagation model on spherical coordinates. To ensure the linearity, nonlinear effects such as bottom friction were not considered in the numerical simulation. Because most part of tsunami propagation path is deep ocean, in order to avoid huge computational cost, we select the grid spacing of 2 arcmin. And the time step is 1 s, which meets the requirement of the stability condition. The computation is done on the computer system of the Earthquake Information Center (EIC) at the University of Tokyo. For far-field tsunamis, most of the propagating path is 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 General Bathymetric Chart of the Ocean released in 2014 (GEBCO_2014; Weatherall *et al.* 2015). 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, we moved the gauge location around and put it on the first available point with negative altitude.

We forecast the tsunami amplitudes and arrival times at points of interest, including DART tsunameters and tide gauges (Fig. 2a). Tide gauge data are more affected by local topography and harbor effects, and thus have larger amplitudes and more complicated waveforms and spectra than tsunameter records (Rabinovich 1997; Rabinovich & Eblè 2015).

3.4 Accuracy analysis

We take the ratio of observed and calculated amplitudes at points of interest for evaluation of accuracy (Aida 1978; Gusman *et al.* 2016). The geometric mean K value of the amplitude ratios and the forecasting accuracy are calculated as:

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

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

where A_i^{obs} and A_i^{pred} stand for the maximum amplitude of the observed and forecasted waveforms, and N is the number of points of interest. However, if some ratios are very large and some ratios are very small, the accuracy will still be abnormally high (Wang *et al.* 2017). Thus, we also evaluate the geometric standard deviation σ of the amplitude ratios as the eq. (7). The geometric standard deviation σ is always larger than 1. A small value indicates a narrow distribution of the amplitude ratios.

$$\log (\sigma) = \sqrt{\frac{\sum_{i=1}^{N} \left(\log \left(\frac{A_i^{obs} / A_i^{pred}}{K} \right) \right)^2}{N}} \tag{7}$$

In addition to the tsunami amplitudes, we also evaluate the forecasting accuracy of tsunami arrival times. The time lag of the ith tide gauge station is defined as (Tsushima *et al.* 2012):

$$\Delta T_i = \left| t_i^{pred} - t_i^{obs} \right|, \tag{8}$$

where t_i^{obs} and t_i^{pred} are the arrival time of the observed and forecasted tsunamis at the ith station. We use the absolute value because both early and late arrival time prediction are regarded as forecasting errors. A small value of time lag indicates accurate forecasting of the arrival time. In our research, we mainly focus on the first tsunami peak, and we use the above two index for evaluation. For a more rigorous evaluation of forecasted waveforms, we could adopt the least-square measurement for the tsunami waveforms.

4 RESULTS

We define the time of origin $t = 0$ when the 2015 Illapel earthquake occurred, used for reference. The tsunami arrival time is defined by a threshold of 0.02 m. This value is used to identify the tsunami arrival in both assimilation stations and points of interest. The tsunami arrives at the station 32 404 at $t = 12$ min. Then, the assimilation

(a) Points of Interest

(b) Waveform Comparison
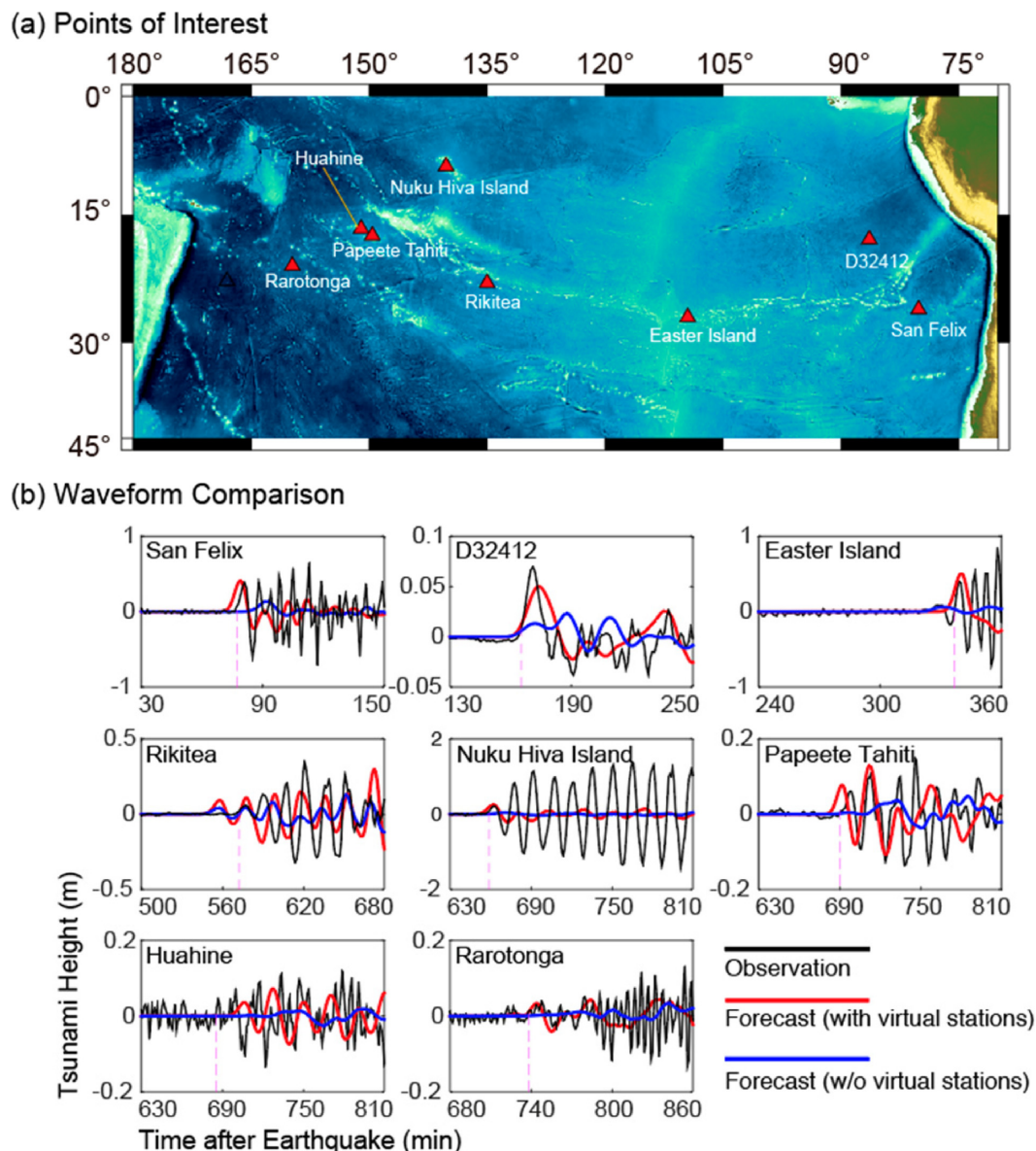
Figure 2. (a) Illustration of eight points of interest. We forecast the far-field tsunami in east Pacific, from the ocean off the Chilean coast to Oceania regions. (b) Comparison of the observed and forecasted waveforms at points of interest. The black lines represent the observed waveforms. The red lines represent the forecasted waveforms by data assimilation with virtual stations. The blue lines represent the forecasted waveforms by data assimilation without virtual stations. The arrival time of observed tsunamis is marked by dash lines.

process begins. It is important to mention that the waveforms at virtual stations will not be calculated until tsunami arrives at both of the neighboring stations. It may cause some delay in forecasting, but the effects are small when considering far-field tsunamis. The first tsunami forecast is made at $t = 1$ hr, when the first tsunami peak has passed stations 32 402, 32 404 and 34 420, as well as the virtual stations interpolated between them, and the tsunami waveform at San Felix is synthesized. Then, as time passes by, more observed data are used in the assimilation, and the forecasting accuracy is improved. The whole assimilation process ends at $t = 2$ hr 20 min, after the first tsunami peak has thoroughly passed through the four DART tsunameters. The second tsunami forecast is made for other points of interest.

We compare the forecasted and observed tsunami waveforms at points of interest (Fig. 2b; Table 1). The tsunami height at tide

Table 1. Observed and forecasted tsunami arrival time and first-peak amplitudes at points of interest. The forecasted results are calculated with the help of virtual stations.

Points of interest	Arrival time (hour:min)		First-peak amplitude (m)	
	Observation	Forecast	Observation	Forecast
San Felix	1:17	1:13	0.39	0.41
D32412	2:45	2:45	0.05	0.07
Easter Island	5:36	5:34	0.40	0.51
Rikitea	9:30	9:12	0.06	0.09
Nuku Hiva Island	10:57	10:54	0.22	0.26
Papeete Tahiti	11:30	11:25	0.06	0.08
Huahine	11:266	11:40	0.05	0.04
Rarotonga	12:18	12:18	0.04	0.03

gauges have already been corrected by Green's Law, assuming that the coastal depth is ∼ 1 m. (Synolakis & Skjelbreia 1993; Baba *et al.* 2004; Wang *et al.* 2012). In addition, the waveforms of tide gauges are affected by local topography and harbor effects that may not be fully modelled by our simulation using 2-arcmin grid. Thus, we mainly focus on the amplitude and arrival times of the first tsunami peak at tide gauges. From the disaster mitigation point of view, the first tsunami arrival is the largest threat to coastal communities (Tsushima *et al.* 2012). Therefore, the forecasting of arrival time and amplitude of the first tsunami peak is important. For comparison, we also plot the waveforms forecasted from the fault model of Ren *et al.* (2017) in Fig. S1.

The tsunami waveforms forecasted by data assimilation without virtual stations (blue lines) 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 lines), the results are substantially improved.

At the San Felix tide gauge off the Chilean coast, the actual tsunami arrives at 1 hr 17 min after the earthquake with an amplitude of 0.39 m. Our results show that the forecasted tsunami arrival time is 1 hr 13 min after the earthquake according to the threshold. And its first-peak amplitude is 0.41 m. Because the tsunami arrives at San Felix before the end of data assimilation process, we used the waveform synthesized at $t = 1$ hr. The DART station 32 412 records the tsunami arrival at 2 hr 45 min after the earthquake according to the threshold. And its first-peak amplitude is 0.07 m. The forecasted waveform has an arrival time of 2 hr 45 min, and an amplitude of 0.05 m. At both stations, there is a high consistency between the observed and forecasted waveforms.

In Easter Island, the tsunami arrives at 5 hr 36 min after the earthquake, and the amplitude of the first tsunami peak is 0.40 m. The comparison indicates that our method forecasts the tsunami arrival time (5 hr 34 min) precisely. And the forecasted amplitude is 0.51 m, which is higher than the observed first peak. The amplitude of following waveforms are larger than the first peak, but they are not reproduced by tsunami data assimilation.

In Oceania, because the 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 well predicted in most tide gauges including Nuku Hiva Island, Papeete Tahiti and Rarotonga. In Rikitea, the forecasted tsunami arrives approximately 18 min earlier than real observation. In Huahine, the forecasted tsunami arrives approximately 14 min later than the real observation. As for the tsunami amplitude of the first tsunami peak, in Rikitea, Nuku Hiva Island and Papeete Tahiti, the forecasts underestimate the real observations, while in Huahine and Rarotonga the amplitude of the first tsunami peak is slightly overestimated. Although we mainly focus on the first tsunami peak, in most stations, the following waveforms after the first peak are still consistent with the real observations. However, in Nuku Hiva Island, the following waveform showed strong oscillation characteristics which was not well captured by modelled results.

Overall, the forecasting accuracy is 87.5 per cent for all the stations using eqs (5) and (6), with a small geometric standard deviation of 1.24. Without virtual stations, the forecasting accuracy is only 13.2 per cent, substantially lower than the results with virtual stations, with the geometric standard deviation (eq. 7) of 3.15. We also calculate the average value of time lag as 5.75 min using eq. (8). Considering the long traveltime of several hours, the arrival time forecast is judged as accurate. We compare the forecasted and
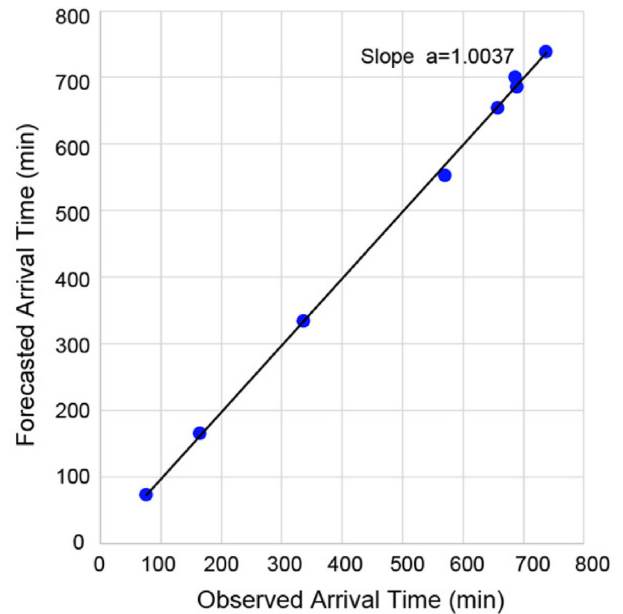
Figure 3. Comparison of the observed and forecasted tsunami arrival time at the eight points of interest.

observed arrival times at each point of interest (Fig. 3). The slope of the regression line is 1.0037, with the *R*-square value of 0.999. It indicates that our method does not have a bias to forecast the arrival time earlier or later. For the forecasting without virtual stations, it was very difficult to identify the forecasted arrival time. Thus, we do not calculate the time lag.

5 DISCUSSION AND CONCLUSION

In our retroactive data assimilation experiment, the far-field tsunami forecasting is first issued at 1 hr after the earthquake. The forecasting is gradually updated by assimilating new data as time passes, but leaving enough time left before the tsunami arrives at the points of interest. Thus, there will be enough evacuation time if the forecast is disseminated to local residents. Unlike the previous application to S-net (Maeda *et al.* 2015) or DONET (Wang *et al.* 2018), the current experiment focus on the far-field points of interest. Even the nearest station (San Felix) has a distance of more than 500 km from the source.

In the current experiment, the forecasted waveforms generally match well with the real observations in the east Pacific. It proves the ability of our method to predict the tsunami amplitude and arrival times at thousands of kilometres far from the source. The adoption of virtual stations enables us to conduct tsunami data assimilation with a few real stations. The tsunami forecasting for ocean bottom gauges has much better performance than that for tide gauges because they are free from nearshore topography or harbor effects. In some coastal tide gauges, our method can only predict the first tsunami peak. For example, the tide gauge Nuku Hiva Island is located inside a bay. The local topographic features caused a remarkable effect creating harbor oscillations (Yamazaki & Cheung 2011; Allgeyer *et al.* 2013; Calisto *et al.* 2016) that we could not reproduce. Achieving higher accuracy would require finer grid and higher computational costs.

It is also important to point out that the data assimilation method does not need information on the source. In the retroactive study, we computed synthetic data at DART stations 32 404 and 34 420

from a tsunami source model to compensate for the sparse real observation. Currently, the DART tsunameters along the Chilean subduction zone, together with virtual stations, can form a dense network with enough spatial coverage. Introduction of these virtual stations can reduce the investment for tsunami mitigation.

Last but not least, our method is not aimed at replacing the current tsunami warning systems based on seismic observations. For example, the W phase inversion estimates the source parameters in a short time. It helps issue the warning for the Chilean coast for potential earthquakes along Peru-Chile Trench. Thus, our method is not practical for near-field tsunami forecasting along the Chilean coast. Instead, we focus on the Pacific side. It provides an alternative approach for far-field tsunami forecasting in east Pacific, including the Chilean territory San Felix Island, Easter Island, etc. It successfully avoids the complexities and uncertainties of estimating the tsunamis source, and becomes applicable to all kinds of tsunamis, including the tsunamis generated by earthquake, landslide tsunamis, and volcanic tsunamis. A combination of our method with the current PTWC approach will improve the reliability of tsunami forecast. It could be implemented for far-field tsunami warning systems for disaster mitigation in Chile and east Pacific.

ACKNOWLEDGEMENTS

REFERENCES

Aida, I., 1978. Reliability of a tsunami source model derived from fault parameter, *J. Phy. Earth,* **26**(1), 57–73.

Allgeyer, S., Hébert, H. & Madariaga, R., 2013. Modelling the tsunami free oscillations in the Marquesas (French Polynesia), *Geophy. J. Int.,* **193**(3), 1447–1459.

An, C. & Meng, L., 2017. Time reversal imaging of the 2015 Illapel tsunami source, *Geophys. Res. Lett.,* **44**, 1732–1739.

Aránguiz, R. *et al.*, 2016. The 16 September 2015 Chile tsunami from the post-tsunami survey and numerical modeling perspectives, *Pure appl. Geophys.,* **173**, 333–348.

Baba, T., Hirata, K. & Kaneda, Y., 2004. Tsunami magnitudes determined from ocean-bottom pressure gauge data around Japan, *Geophys. Res. Lett.,* **31**, L08303, doi:10.1029/2003GL019397.

Baba, T., Takahashi, N. & Kaneda, Y., 2014. Near-field tsunami amplification factors in the Kii Peninsula, Japan for Dense Oceanfloor Network for Earthquakes and Tsunamis (DONET), *Mar. Geophys. Res.,* **35**, 319–325.

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 appl. Geophys.,* **172**, 3433–3472.

Baker, N., 1992. Design of the Navy's multivariate optimum interpolation analysis system, *Weather Forecast,* **7**, 220–231.

Barrientos, S.E. & Ward, S.N., 1990. The 1960 Chile earthquake: inversion for slip distribution from surface deformation, *Geophys. J. Int.,* **103**, 589–598.

Benavente, R., Cummins, P.R. & Dettmer, J., 2016. Rapid automated W-phase slip inversion for the Illapel great earthquake (2015, Mw = 8.3), *Geophys. Res. Lett.,* **43**, 1910–1917.

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 appl. Geophys.,* **173**, 1051–1061.

Cienfuegos, R., Catalán, P.A., León, J., Gonzalez, G., Repetto, P., Urrutia, A., Tomita, T. & Orellana, V., 2016. Lessons for tsunami risk mitigation from recent events occurred in Chile: research findings for alerting and evacuation from inter disciplinary perspectives, in *AGU Fall Meeting 2016,* NH52A-07.

Cienfuegos, R., Catalán, P.A., Urrutia, A., Benavente, R., Aránguiz, R. & González, G., 2018. What can we do to forecast tsunami hazards in the near field given large epistemic uncertainty in rapid seismic source inversion, *Geophys. Res. Lett.,* **45**, 1–12.

Duputel, Z., Rivera, L., Kanamori, H. & Hayes, G., 2012. W phase source inversion for moderate to large earthquakes (1990-2000), *Geophys. J. Int.,* **189**, 1125–1147.

Duputel, Z., Rivera, L., Kanamori, H., Hayes, G.P., Hirshorn, B. & Weinstein, S., 2011. Real-time W phase inversion during the 2011 off the Pacific coast of Tohoku Earthquake, *Earth Planet Sp.,* **63**, 5, doi:10.5047/eps.2011.05.032.

Fritz, H.M. *et al.*, 2011. Field survey of the 27 February 2010 Chile tsunami, *Pure appl. Geophys.,* **168**, 1989–2010.

Fujii, Y. & Satake, K., 2013. Slip distribution and seismic moment of the 2010 and 1960 Chilean earthquakes inferred from tsunami waveforms and coastal geodetic data, *Pure appl. Geophys.,* **170**, 1493–1509.

González, F.I., Bernard, E.N., Meinig, C., Eble, M.C., Mofjeld, H.O. & Stalin, S., 2005. The NTHMP tsunameter network, *Nat. Hazards,* **35**, 25–39.

Gusman, A.R., Murotani, S. & Satake, K.et al., 2015. Fault slip disrtibution of the 2014 Iquique, Chile, earthquake estimated from coean-wide tsunami waveforms and GPS data: The 2014 Iquique earthquake, *Geophys. Res. Lett.,* **42**, 1053–1060.

Gusman, A.R., Sheehan, A.F., Satake, K., Heidarzadeh, M., Mulia, I.E. & Maeda, T., 2016. 2016. Tsunami data assimilation of Cascadia seafloor pressure gauge records from the 2012 Haida Gwaii earthquake, *Geophys. Res. Lett.,* **43**, 4189–4196.

Heidarzadeh, M., Murotani, S., Satake, K., Ishibe, T. & Gusman, A.R., 2016. Source model of the 16 September 2015 Illapel, Chile, Mw 8.4 earthquake based on teleseismic and tsunami data, *Geophys. Res. Lett.,* **43**, 643–650.

Hossen, M.J., Cummins, P.R., Dettmer, J. & Bata, T., 2015a. Time reverse imaging for far-field tsunami forecasting: 2011 Tohoku earthquake case study, *Geophys. Res. Lett.,* **42**, 9906–9915.

Hossen, M.J., Cummins, P.R., Dettmer, J. & Bata, T., 2015b. Tsunami waveform inversion for sea surface displacement following the 2011 Tohoku earthquake: importance of dispersion and source kinematics, *J. geophys. Res.: Solid Earth,* **120**, 6452–6473.

Ho, T., Satake, K., Watada, S. & Fujii, Y., 2019. Source estimate for the 1960 Chile earthquake for joint inversion of geodetic and transoceanic tsunami data, *J. geophys. Res.: Solid Earth.,* doi:10.1029/2018JB016996.

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), *J. geophys. Res.,* **106**, 9161–9177.

Kalnay, E., 2003. *Atmospheric Modeling, Data Assimilation and Predictability,* Cambridge Univ. Press.

Kanamori, H. & Rivera, L., 2008. Source inversion of W phase: speeding up seismic tsunami warning, *Geophys. J. Int.,* **175**(1), 222–238.

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, *Geophys. Res. Lett.,* **42**, 7923–7932.

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 *Proc. IEEE OCEANS 2005,* 2, pp. 1673–1679.

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. *J. Geophys. Res. Oceans,* **122,** 7924–7937.

Pacific Tsunami Warning Center (PTWC), 2015. PTWC Tsunami Threat Message, http://ptwc.weather.gov/text.php?id=pacific.TSUPAC.2015.09.16.2347. Accessed 16 Sep 2015.

Rabinovich, A.B., 1997. Spectral analysis of tsunami waves: Separation of source and topography effects, *J. geophys. Res.,* **102**(12), 663–612, 676.

Rabinovich, A.B. & Eblé, M.C., 2015. Deep-ocean measurements of tsunami waves, *Pure appl. Geophys.,* **172,** 3281–3312.

Ren, Z., Yuan, Y., Wang, P., Fan, T., Wang, J. & Hou, J., 2017. The September 16, 2015 Mw 8.3 Illapel, Chile Earthquake: characteristic of tsunami wave from near-field to far-field, *Acta Oceanol. Sin.,* **36,** 73–82.

Saito, T., Satake, K. & Furumura, T., 2010. Tsunami waveform inversion including dispersive waves: the 2004 earthquake off Kii Peninsula, Japan, *J. geophys. Res.,* **115,** B06363, doi:10.1029/2009JB006884.

Satake, K., 1995. Linear and nonlinear computations of the 1992 Nicaragua earthquake tsunami, *Pure appl. Geophys.,* **145,** 455–470.

Satake, K., 2015. *Treatise on Geophysics,* 2nd edn, Vol **4,** Elsevier,pp. 447–504.

Satake, K. & Heidarzadeh, M., 2017. A review of source models of the 2015 Illapel, Chile earthquake and insights from tsunami data, in *The Chile-2015 (Illapel) Earthquake and Tsunami,* eds. Braitenberg, C. & Rabinovich, A., Pageoph Topical Volumes, Birkhauser, Cham.

Synolakis, C.E. & Skjelbreia, J.E., 1993. Evolution of maximum amplitude of solitary waves on plane beaches, *J. Waterway, Port, Coastal, Ocean Eng.,* **119**(3), 323–342.

Takahashi, N., Imai, K., Ishibashi, M. & Sueki, K., 2017. Real-time tsunami prediction system using DONET, *J. Disaster Res.,* **12**(4), 766–774.

Tsushima, H., Hino, R., Fujimoto, H., Tanioka, Y. & Imamura, F., 2009. Near-field tsunami forecasting from cabled ocean bottom pressure data, *J. geophys. Res.,* **114,** B06309, doi:10.1029/2008JB005988.

Tsushima, H., Hino, R. & Tanioka, Y., 2012. Tsunami waveform inversion incorporating permanent seafloor deformation and its application to tsunami forecasting, *J. geophys. Res.,* **117,** B03311, doi:10.1029/2011JB008877.

Tsushima, H. & Ohta, Y., 2014. Review on near-field tsunami forecasting from offshore tsunami data and onshore GNSS data for tsunami early warning, *J. Disaster Res.,* **9**(3), 339–357.

Wang, D. *et al.*, 2012. Real-time forecasting of the April 11, 2012 Sumatra tsunami, *Geophys. Res. Lett.,* **39**(19), L19601, doi:10.1029/2012GL053081.

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, *Geophys. Res. Lett.,* **46,** 2045–2053.

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, *Geophys. Res. Lett.,* **44**(10), 282–210, 289.

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

Watada, S., Kusumoto, S. & Satake, K., 2014. Traveltime delay and initial phase reversal of distant tsunamis coupled with the self-gravitating elastic Earth, *J. geophys. Res.: Solid Earth,* **119,** 4287–4310.

Weatherall, P. *et al.*, 2015. A new digital bathymetric model of the world's oceans. *Earth Space Sci.,* **2**(8), 331–345.

Williamson, A. & Newman, A., 2018. Limitations of the resolvability of finite-fault models using static land-based geodesy and open-ocean tsunami waveforms, *J. geophys. Res.: Solid Earth,* **123,** 9033–9048.

Williamson, A., Newman, A. & Cummins, P., 2017. Reconstruction of coseismic slip from the 2015 Illapel earthquake using combined geodetic and tsunami waveform data, *J. geophys. Res.: Solid Earth,* **122,** 2119–2130.

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

Yamazaki, Y. & Cheung, K.F., 2011. Shelf resonance and impact of near-field tsunami generated by the 2010 Chile earthquake, *Geophys. Res. Lett.,* **38,** L12605, doi:10.1029/2011GL047508.

Yoshimoto, M., Watada, S., Fujii, Y. & Satake, K., 2016. Source estimate and tsunami forecast from far-field deep-ocean tsunami waveforms – the 27 February 2010 Mw 8.8 Maule earthquake, *Geophys. Res. Lett.,* **43**(2), 659–665.

SUPPORTING INFORMATION

Supplementary data are available at *GJI* online.

Figure S1. Comparison of the observed and forecasted waveforms at points of interest. The black lines represent the observed waveforms. The red lines represent the forecasted waveforms by data assimilation with virtual stations. The dark green lines represent the forecasted waveforms by forwarding simulation from the fault model of Ren *et al.* (2017).

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the paper.