Pure and Applied Geophysics manuscript No.
(will be inserted by the editor)

Evaluating the Effectiveness of DART [R] Buoy Networks Based on Inundation Forecast Accuracy

Donald B. Percival · Donald W. Denbo · Edison Gica · Paul Y. Huang · Harold O. Mofjeld · Michael C. Spillane · Vasily V. Titov

Received: date / Accepted: date

Abstract A performance measure for a DART [R] tsunami buoy network has been developed. DART [R] buoys are used to detect tsunamis, but the full potential of

Donald B. Percival

Applied Physics Laboratory, Box 355640, University of Washington, Seattle, WA 98195-5640 USA

E-mail: dbp@apl.washington.edu

Donald B. Percival

Department of Statistics, Box 354322, University of Washington, Seattle, WA 98195-4322 USA

Donald W. Denbo · Edison Gica · Harold O. Mofjeld · Michael C. Spillane · Vasily V. Titov

NOAA/Pacific Marine Environmental Laboratory, 7600 Sand Point Way NE, Seattle, WA 98115 USA

Donald W. Denbo · Edison Gica · Michael C. Spillane

Joint Institute for the Study of the Atmosphere and Ocean, University of Washington, Seattle, WA 98195-5672 USA

Paul Y. Huang

National Tsunami Warning Center, National Weather Service, Palmer, AK, 99645 USA

the data they collect is realized through accurate forecasts of inundations caused by the tsunamis. The performance measure assesses how well the network achieves its full potential through a statistical analysis of simulated forecasts of wave amplitudes outside an impact site and a consideration of how much the forecasts are degraded in accuracy when one or more buoys are inoperative. The analysis uses simulated tsunami amplitude time series collected at each buoy from selected source segments in the Short-term Inundation Forecast for Tsunamis database and involves a set for 1000 forecasts for each buoy/segment pair at sites just offshore of selected impact communities. Random error-producing scatter in the time series is induced by uncertainties in the source location, addition of real oceanic noise, and imperfect tidal removal. Comparison with an error-free standard leads to root-mean-square errors (RMSEs) for DART ^{R} buoys located near a subduction zone. The RMSEs indicate which buoy provides the best forecast (lowest RMSE) for sections of the zone, under a warning-time constraint for the forecasts of 3 hrs. The analysis also shows how the forecasts are degraded (larger minimum RMSE among the remaining buoys) when one or more buoys become inoperative. The RMSEs provide a way to assess array augmentation or redesign such as moving buoys to more optimal locations. Examples are shown for buoys off the Aleutian Islands and off the West Coast of South America for impact sites at Hilo HI and along the U.S. West Coast (Crescent City CA and Port San Luis CA). A simple measure (coded green, yellow or red) of the current status of the network's ability to deliver accurate forecasts is proposed to flag the urgency of buoy repair.

Keywords Aleutian Islands tsunami sources · Buoy network performance measure · Crescent City CA · DART ^{R} data inversion · Hilo HI · Network

assessment · Port San Luis CA · South American tsunami sources · Tsunameter · Tsunami buoys · Tsunami forecasts · Tsunami simulation · Tsunami source estimation

1 Introduction

Tsunamis are potentially devastating disasters for coastal regions worldwide. The loss of life due to tsunamis can be greatly reduced when the propagation time of a tsunami is long enough to allow for timely issuance of warnings. Beginning in the late 1940s soon after a devastating tsunami that hit Hawaii in 1946, the National Oceanic and Atmospheric Administration (NOAA) and its predecessors established warning systems to provide tsunami forecasts to the United States and other nations (Whitmore 2009). The research presented here is part of an ongoing effort to improve the speed and accuracy of such forecasts through enhanced observational capabilities.

Prompted in part by the scale of destruction and unprecedented loss of life following the 2004 Indian Ocean tsunami, NOAA has deployed 39 Deep-ocean Assessment and Reporting of Tsunamis (DART ®) buoys (to date, eight countries other than the United States have set out additional buoys). These buoys can observe the passage of a tsunami and relay data related to its arrival time and amplitude in near-real time. These data are used by the Short-term Inundation Forecast for Tsunamis (SIFT) application developed at the NOAA Center for Tsunami Research (NCTR); for details, see Gica et al. (2008) and Titov (2009). SIFT combines DART ® data with precomputed geophysical models to estimate so-called unit source coe cients (Percival et al. 2011). As outlined in Sect. 2, these

coe cients are used to forecast wave amplitudes at locations in the open ocean outside of impact sites (coastal regions of particular interest, e.g., Hilo HI). In turn forecasts of these wave amplitudes play a critical role in forecasting coastal inundation at impact sites.

In this article we consider the problem of how to evaluate the effectiveness of a DART ^{R} buoy network in providing accurate and timely forecasts of open-ocean wave amplitudes. The purpose of the network is not merely to detect tsunamis (for which instruments much less complex than DART ^{R} buoys could be deployed), but, more importantly, to collect data that can be used to provide accurate inundation forecasts in a timely manner – hence we focus our evaluation of the network on the quality of the forecasts rather than just on the simpler task of detection. For an existing network the goal is to devise a simple measure of effectiveness that can be monitored over time and used to raise an alarm if the network deteriorates. For an alteration to an existing network (typically through adding, decommissioning or relocating a buoy), the goal is to quantify the effect of the change in a simple manner. The challenge in both cases is that the effectiveness of wave amplitude forecasts is predicated on a number of interacting factors. The desire for a simple measure is at odds with the inherent complexity of the forecast process. What is needed is a useful measure that does not oversimplify. Here we propose a measure that makes certain simplifications but nonetheless captures key factors limiting the accuracy of forecasted wave amplitudes. Our approach uses a succession of summaries of simulated forecasts. (An ideal measure would make use of actual rather than simulated forecasts, but there is simply not enough actual data for realistic evaluation of an existing network, much less a proposed network.)

The effort described here is not the first to consider the evaluation of a network designed to contribute to tsunami warnings. Greenslade and Warne (2012) is the first work to evaluate an existing heterogeneous network (i.e., one consisting of both tide gauges and DART® buoys) for its effectiveness for tsunami warnings. Their evaluation is done purely in terms of travel times. Our work focuses on just DART® buoys and uses accuracy of forecasted wave amplitudes as the evaluation criterion while paying nominal attention to travel times. Greenslade and Warne (2012) note that network design is closely related to network evaluation and point to Spillane et al. (2008) on the design of DART® buoy networks. The current effort builds upon the latter in using some of the same tools (e.g., the propagation database described in Gica et al. 2008) but does network evaluation in a manner substantively different from network design.

2 Construction of Open-Ocean Wave Amplitude Forecasts in Practice

Here we give an overview of how open-ocean wave amplitude forecasts are made in practice using SIFT. Generation of these forecasts presumes that a tsunami-causing earthquake originates from within one or more unit sources, which are 100 km by 50 km sections of fault planes in tsunamigenic regions. Figure 1 shows locations of a subset of 74 predesignated unit sources tiling the region where the Pacific plate is being subducted within the Aleutian–Alaska subduction zone. The sources chosen are in 2 rows (`b` and `a`) and 37 columns (`001` to `037`). Row `b` is adjacent to the plate boundary, while row `a` is a continuation of the descending slab to greater depths (we do not make use of additional rows used to represent deeper Aleutian–Alaska earthquakes). The expanded view in the figure shows rows

`b` and `a` along with 3 of the columns (`025`, `026` and `027`). We refer to the unit source in, e.g., row `b` and column `026` as `ac026b`. The figure also shows the locations of three impact sites (Hilo (HIL), Crescent City (CCY) and Port San Luis (PSL)) and the current locations of 14 DART® buoys (labeled clockwise by U, V, ..., X, 1, 2, ..., 8; the expanded view includes buoy 46403, which is labeled as `6`). There are 39 DART® buoys in all, but these 14 are germane for handling events arising from Aleutian–Alaska earthquakes in that, as called for by standard operating procedures, their data could be used as the basis for issuing a warning at least 3 hrs in advance to at least one of the three impact sites (Whitmore et al. 2008).

A geophysical model is precomputed for each pairing of one of the 74 unit sources either with a buoy or with a location on the gridded open-ocean boundary of the inundation model for a particular impact site (see Gica et al. 2008 for details about these models, which are stored in a propagation database maintained at NCTR). The model predicts what would be observed over time at a deep water location if a moment magnitude $M_W = 7.5$ reverse fault earthquake were to originate from within a particular unit source. Figure 2 shows examples of predictions for buoy 46403 due to an earthquake from either unit source `ac026b` or one of five abutting it (for use later on, these predictions are marked by $\boldsymbol{g}_1, \ldots, \boldsymbol{g}_6$). If, as is usually the case, the moment magnitude of the earthquake is different from 7.5, then, relying on the linearity of tsunami waves in deep water, the predictions are adjusted by a nonnegative multiplicative factor α known as a source coe cient (the assumption of a reverse fault earthquake imposes the nonnegativity constraint). Earthquakes can span more than one unit source, in which case the model is taken to be a linear combination of individual unit source models, with the coe cients in the linear combination, say α_k, $k = 1, \ldots, K$, all constrained to be nonnegative.

During the passage of a tsunami, a DART® buoy collects and transmits a time series of bottom pressure (BP) measurements. Each measurement represents the BP averaged over a 1-min span. The spans are non-overlapping, and the measurements occur once per minute, thus forming what we refer to as a 1-min stream of data. Once a sufficient stretch of this stream has been collected by the first buoy that detects the tsunami event, the source coefficients α_k are estimated using constrained least squares (Percival et al. 2014). With the passage of time, more 1-min data from this and potentially other buoys become available, and the estimates $\hat{\alpha}_k$ of the source coefficients can be reevaluated and possibly refined. In operational tsunami forecasting, the estimated source coefficients are applied to a sub-region of the propagation database to provide the boundary forcing fields for the real-time inundation model run with detailed simulation of waves, currents and inundation for the impact site (Gica et al. 2008); however, in the simulation study discussed in the next section, we avoid use of the inundation model by choosing an adjacent deep-water location for each impact site of interest. This approach allows us to efficiently generate a large numbers of simulations by linear combinations of table look-ups.

3 Evaluation of Open-Ocean Wave Amplitude Forecasts via Simulation

In theory the best way to evaluate the effectiveness of a network of DART® buoys would be to compare coastal inundation forecasts formulated during actual tsunami events with the actual inundations. While such comparisons are routinely done as a vital check on current operational procedures, use of these alone is not sufficient for network evaluation. A meaningful assessment would require not

only multiple events, but also ones treated in the same manner (no change in the network of buoys, no equipment upgrades, use of the same methodology for forming the forecasts, etc.). Even if conditions for generating forecasts could be fixed, the amount of time needed to collect data for a reliable evaluation would be inordinate given the rate at which tsunamis occur. In addition, if we were to insist upon using actual events to assess effectiveness, changes to a network could not be assessed until long after the changes had been put into place. For planning purposes, we need the ability to assess the effectiveness of any proposed changes. We thus entertain basing our assessments on simulated forecasts. The simulated forecasts assume that the DART$^{\mathrm{R}}$ buoys are an inherent and vital component of the forecast process (in the Appendix we briefly consider forecasts based on seismic information alone).

As noted in the previous section, the run-up model for generating inundation forecasts at a particular impact site requires its initial conditions be set by wave amplitudes in the open ocean at a nearby location. The quality of the forecasts generated by the run-up model depends critically on these initial conditions, which are provided by forecasts of open-ocean wave amplitudes. Rather than assessing a network via the quality of inundation forecasts directly, we can do so indirectly via the quality of open-ocean wave amplitude forecasts. In the simulation studies discussed below, this indirect approach is computationally attractive because calculation of run-up models is time-consuming. This approach, however, ignores errors in forecasts that might arise from imperfect bathymetry and/or topography.

With the simplification of dealing with open-ocean wave amplitude forecasts rather than inundation forecasts, there are four tasks at hand to simulate and

evaluate the forecasts of interest. Here we glean the key points in our handling of the four tasks, with full details being provided in an appendix (Sect. 7.1 to 7.4).

The first task is to play the role of nature (Sect. 7.1). We must simulate a tsunami signal as it passes by a particular buoy along with what we will regard as the wave amplitudes presumed to occur outside of an impact site of interest. We assume the signal originates within a unit source that we refer to as the central unit source and that is surrounded by five other unit sources (in the expanded view of Fig. 1, the central unit source is `ac026b`). For each of the six unit sources, we have a geophysical model for what would be observed at a particular buoy for a tsunami generating event originating at the center of the unit source. The signal is a weighted average of the six models, with the weights set by picking at random a location within the central unit source. This location is used to define a new unit source that is constructed from pieces of the original six unit sources, with the sizes of the pieces determining the weights. We also have a geophysical model for what would be observed in the open ocean outside of an impact site of interest. We apply the same weights to these models to form the wave amplitudes presumed to occur in the open ocean.

The second task is to play the role of nature's observer (Sect. 7.2). We must simulate the data that a particular DART® buoy would collect during a tsunami event. We do so by adding actual tidal fluctuations and background noise recorded under ambient conditions by the buoy (if archived data are not available for a particular buoy, we use data from a suitable surrogate).

The third task is to use the simulated data to estimate source coe cients, which in turn are used to forecast the wave amplitudes at an impact site (Sect. 7.3). Estimation is accomplished in terms of a linear regression model that, in addition

to the source coe cients, has coe cients designed to compensate for tidal fluctuations. The source coe cients are associated with the geophysical models, but, to mimic what happens in practice, we permit potential mismatches between the tsunami signal and its assumed model. One last consideration is how much data to use for estimating the coe cients. This is settled by determining the location of the first peak in the signal and including data spanning from 1 hr prior to this peak to 21 min after the peak (this assumes that the peak occurs at least 1 hr after the start of the earthquake generating the tsunami signal – if not, then the first data point coincides with the start of the earthquake). We estimate the coe cients in the linear regression model using least squares. We then use the estimated source coe cients to form forecasted wave amplitudes.

The fourth and final task is to evaluate the quality of the forecasted wave amplitudes by comparing them with the presumed amplitudes using an appropriate metric (Sect. 7.4). To do so, we repeat the first three tasks 1000 times, resulting in 1000 forecasted amplitudes to be compared with 1000 presumed amplitudes. We evaluate how well the forecasts do by looking at the difference between the maximum forecasted and presumed amplitudes and using these differences to form a root-mean-square error, which is the basis for the network assessment discussed in the next section.

4 Network Assessment

In the previous section, we described our approach for assessing the effectiveness of a particular buoy in forecasting wave amplitudes in the open ocean outside of an impact site of interest when an earthquake-generating tsunami arises from within

a particular central predesignated unit source. The approach takes into account major factors influencing the quality of the forecasts (including tides, background noise, uncertainties in the signal model and amount of available buoy data). In this section we turn our attention to network assessment. To focus our discussion, we consider how well the existing network of DART® buoys does in responding to tsunami events originating first from the Aleutian Islands (Sect. 4.1) and then from South America (Sect. 4.2).

4.1 First Case Study: Aleutian Islands

Figure 1 shows the locations of 74 predesignated unit sources tiling the Aleutian–Alaskan subduction zone, the 14 DART® buoys closest to this zone (8 along the Aleutian Islands (labeled `1` to `8`) and 6 to the west (`U` to `Z`)) and three impact sites (Hilo, Crescent City and Port San Luis – these are labeled as HIL, CCY and PSL). To evaluate this network of buoys, we consider all possible triads involving one buoy, one impact site and one central unit source selected from `ac002b`, ..., `ac036b` (there are thus $14 \times 3 \times 35 = 1470$ triads in all). Figure 3 shows root-mean-square errors (RMSEs) of forecasted maximum wave amplitudes in the open ocean outside of Hilo (see Eq. (6) in the Appendix for the definition of RMSE). The RMSEs are shown only for those buoy/unit source combinations such that, as called for by standard operating procedures, a warning to Hilo could be issued at least 3 hrs in advance of the arrival of the tsunami (when this criterion is applied to all three impact sites, the number of relevant triads shrinks from 1470 down to 835). The RMSEs for the 14 buoys are plotted versus unit source. Let us concentrate on DART® buoy 46403 (also used as an example in Figs. 1 and 2 and throughout

the appendix (Sect. 7)). This buoy's label is `6`, and the only non-black labels in Fig. 3 are the `6`'s. One is colored red, and its location in the figure tells us that, for a $M_W = 7.5$ earthquake originating within unit source `ac026b`, the RMSE over 1000 forecasts at Hilo based on buoy 6 is equal to 5.9×10^{2} cm (this RMSE is also displayed by the `4` in the 21 min category in Fig. 18, which is discussed in Sect. 7.4 of the appendix). The RMSEs for `ac002b`, ..., `ac012b` are not shown because a tsunami originating from one of these unit sources would not arrive at buoy 6 soon enough to issue a timely warning to Hilo.

For assessing the ability of a network of buoys to provide good forecasted maximum wave amplitudes, individual RMSEs are of not as much interest as the best, i.e., lowest, RMSE that can be achieved across all buoys for an event originating from within a particular unit source. The second best RMSE is also of interest because a comparison of this with the best RMSE tells us how much degradation in forecasted amplitude there would be if the buoy with the best RMSE were to become inoperative. Figure 4(a) simplifies Fig. 3 by showing just the first (blue) and second best (red) RMSEs. There is a prominent pinching pattern between the two sets of RMSEs. The unit sources where the pinches occur are ones for which the distances from the source to the two associated buoys are approximately equal. For events originating from pinch locations, if the buoy with the best RMSE were to drop out, there would be little impact on the ability of the network to forecast wave amplitudes outside of Hilo because of the existence of another buoy that performs almost as well. Figures 4(b) and (c) are similar to (a), but now use Crescent City and Port San Luis as impact sites rather than Hilo. Pinching patterns are in evidence again, with the locations of the pinches occurring close to the same unit sources as for Hilo. With some exceptions, the first

and second best RMSEs are also associated with the same buoys. These results indicate that there is some degree of commonality across impact sites that are at transoceanic distances from the unit sources.

In addition to pinching patterns, there are particularly prominent upward trends in Fig. 4(b) and (c) for Crescent City and Port San Luis. A glance at Fig. 1 shows that Hilo is more centrally located with respect to the Aleutian–Alaskan unit sources than the California sites, both of which are off to one side. The travel times from the 35 central unit sources to the California sites both have a nearly monotonic decay proceeding from `ac002b` to `ac036b`, whereas those for Hilo decrease from `ac002b` to `ac021b` and then increase along the remaining 15 unit sources. As travel time increases, the maximum amplitude of a tsunami signal should tend to decrease. Presuming an inverse relationship between travel time and signal amplitude, the prominent upward trends at the California sites suggest that RMSE increases as signal amplitude increases. To explore this suggestion, the right-hand column of Fig. 4 shows the RMSEs of the forecasted maximum wave amplitudes for each of the impact sites versus the maximum wave amplitudes outside of the three impact sites. The maximum wave amplitudes were extracted from a propagation data base of predictions linking the center of each central unit source with the open ocean location outside of the impact site. Crescent City and Port San Luis have prominent upward trends, but Hilo, less so, which is the same set of patterns displayed in the left-hand column. This qualitative comparison offers support for the hypothesis that the upward trends in the left-hand column are due at least in part to a positive relationship between RMSE and signal amplitude.

To quantify how well the existing network of DART® buoys can respond to events emanating from the Aleutian–Alaskan subduction zone, let us focus on

forecasting maximum wave amplitudes based on single buoys and consider what happens if one buoy were to drop out. As a measure of both the contribution of individual buoys to the network and the overall strength of the network, consider the maximum relative increase in RMSE across all 35 central unit sources due to a particular buoy becoming inoperative. Suppose, for example, that buoy `5` becomes inoperative. Focusing on Hilo, Fig. 4(a) indicates that, if a tsunami were to emanate from unit source `ac023b` (corresponding to the dashed vertical line), loss of `5` would result in a 7 fold increase in the best available RMSE because we would then have to rely on buoy `4` (the RMSEs would also increase at the California sites, but less so). For Hilo, buoy `5` also has the best RMSE for three additional unit sources (`ac022b`, `ac024b` and `ac025b`), but the relative increases in RMSE due to `5` being gone are all smaller than that for `ac023b`. For events emanating from any of the remaining 31 unit sources, the loss of `5` would not cause a deterioration in the ability of the network to forecast Hilo's wave amplitudes.

Figure 5 quantifies how the network is affected when a single buoy becomes inoperative. For every buoy that can be used to deliver a warning at least 3 hrs in advance for a particular impact site due to an event arising from at least one of the 35 central unit sources, we examine the ratio of two particular RMSEs for each unit source. The first RMSE is the best available RMSE after elimination of the buoy in question, and the second is the best RMSE when all DART$^{\text{R}}$ buoys are operational. We then concentrate on just the largest such ratio across all unit sources. This ratio reflects a worst case scenario in a clear-cut manner and is arguably preferable to a more complicated measure that attempts to take into account degradations across all unit sources. The left panel of Fig. 5 shows these RMSE ratios for 14 buoys and at Hilo. The smallest possible ratio is unity,

which is indicated by the left-most red dashed line. A ratio of unity occurs when a buoy's RMSEs are never the best at any of the 35 unit sources. Such a buoy does not participate in the blue portion of Fig. 4(a) and typically is not in close proximity to at least one unit source (see Fig. 1). Under a worst case scenario, loss of these buoys would not degrade any forecasts. By this measure, buoys with an RMSE ratio of unity are of secondary importance to Hilo for responding to events arising from the Aleutian Islands (they are primarily intended to handle events arising elsewhere). When the ratio is greater than unity, the network deteriorates at least to some degree when the buoy in question is out of commission. For these buoys, Fig. 5 indicates the unit source linked to the largest proportional increase in RMSE. The largest RMSE ratio occurs when buoy `5` (indicated by a solid blue circle) is not available to handle an event arising from unit source `ac023b`. Thus, assuming a worst case scenario, loss of buoy `5` would cause the largest degradation in the network's ability to forecast maximum wave amplitudes at Hilo. The middle and right panels of Fig. 5 show corresponding results for Crescent City and Port San Luis. For a buoy whose ratios are greater than unity for multiple impact sites, the associated unit sources are either the same or quite close to one another. With the exceptions of buoy `2`, the RMSE ratios for the two California sites are remarkably similar.

Figure 6 explores the effect on network performance when two buoys drop out. As in Fig. 5, there are three panels, one for each impact site. Each panel has 15 red dotted horizontal lines, which breaks the panel up into 14 subplots, one for each of the relevant buoys. Consider the bottommost subplot in the left-hand panel, which is for buoy `U` in combination with Hilo. There are 13 dots in the subplot, each indicating the maximum ratio of two particular RMSEs across the 35 unit

sources. The first RMSE is the best remaining after U drops out along with one of the other 13 buoys, and the second is the best RMSE when all 14 buoys are present (going from top to bottom in the subplot, the dots are associated with buoys 8 to V). The blue dot is the RMSE ratio when both U and 5 are inoperative (the blue dots in the other subplots also indicate ratios in which 5 is involved). Because buoy U never has the best RMSE for any of the 35 unit sources, the ratios merely reflect what happens when the second buoy drops out. As a result, the dots in the subplot have same pattern as the topmost 13 circles in the left-hand panel of Fig. 5. Thus, when examining any of the other 13 subplots, we can compare it to the bottommost subplot to ascertain of how much the leave-out-two pattern deviates from the leave-out-one pattern. The same holds for the middle and right-hand panels for Crescent City and Port San Luis.

The subplots for buoys V to 8 in Fig. 6 are organized in a manner similar to that for U. Consider the subplot for buoy 5 in combination with Hilo. Because any ratio involving 5 is represented by a blue dot, all of the dots in this subplot are blue (in each of the other 13 subplots for Hilo, there is exactly one blue dot). With two exceptions, the ratios all have a value of 7.1. The identical ratios come about because dropping 5 leads to the largest proportional increase in RMSE across all 35 unit sources (see Fig. 4(a)). The two exceptional ratios are 9.7 and 11.2. A dot in a given subplot representing the RMSE ratio for a particular pair of buoys must be replicated in one other subplot. The dot for 9.7 is replicated in the subplot for 6, and the one for 11.2 has a duplicate in the subplot for 4. Hence, under the worst case scenario, the concurrent loss of 5 and 4 would lead to an order of magnitude increase in RMSE. This scenario involves a tsunami-generating earthquake occurring in unit source ac023b, for which the third best

RMSE is for buoy 6. The second worst case scenario would be to lose 5 and 6 for an event in ac028b, for which the third best RMSE is for buoy 4.

The subplots for eight buoys (Z to 7) in the left-hand panel of Fig. 6 all have ratios greater than unity. These same buoys also have ratios greater than unity under the leave-out-one scenario (left-hand panel of Fig. 5). If one of these buoys were to be inoperative, loss of a second buoy from this set of eight could further compromise the network's ability to forecast maximum wave amplitudes at Hilo. The maximum ratios in all eight subpanels are associated with blue dots, which means, once any buoy amongst these eight other than 5 becomes inoperative, the worst case scenario is to lose 5 also (as noted previously, the worst case overall is the loss of 5 and 4 together). If any of the 14 buoys were to becomes inoperative, a study of the corresponding subplot would tell us how much the ability of the network to forecast Hilo wave amplitudes would deteriorate if another buoy were to become inoperative before the first one is returned to operational status.

The middle and right-hand panels of Fig. 6 show leave-out-two RMSE ratios for the two California impact sites. For Crescent City (middle panel), the worst case scenario occurs if buoys 7 and 8 were inoperative during an event originating in either unit source ac034b, ac035b or ac036b (see Fig. 4(b)). These two buoys are the only ones that can provide a warning 3 hrs in advance to Crescent City. Because there is no other buoy to rely on, we can consider the RMSE ratio to be infinite. The second worst case scenario occurs when buoys 2 and 3 drop out for an event from ac012b leading to the RMSE ratio of 9.5, with 1 providing backup. After these worst cases, there are a number of RMSE ratios equal to 7.1, all of which involve buoy 8 and unit source ac012b.

Turning now to Port San Luis (right-hand panel of Fig. 6), the RMSE ratio for the worst case scenario is 13.7, which occurs when 7 and 8 are inoperative and is associated with ac034b (see Fig. 4(c)). With these two buoys gone, buoy 6 has the best RMSE amongst the remaining buoys satisfying the 3 hrs warning time constraint. The second and third largest RMSE ratios are 4.9 and 4.8. The latter is of interest because the two buoys involved are 1 and Z, and the unit source is ac008b (this is in agreement with Fig. 4(c)). With both of these buoys gone, buoy W has the best remaining RMSE. Note that, in this case, the RMSE-based metric for network evaluation differs from one based solely on travel time since the travel times between unit source ac008b and buoys X and Y are both shorter than that for W (see Fig. 1). A possible explanation is the orientation of ac008b.

We can also explore the effect on network performance when three buoys drop out. For Hilo, the worst case scenario is for buoys 4, 3 and 5 to be unavailable when an event arises from unit source ac020b. Buoy 6 then has the best RMSE amongst the remaining buoys satisfying the 3 hrs warning time constraint, and the associated RMSE ratio raises to 23.0 (as compared to 11.2 when two buoys – 5 and 4 – drop out during an event arising from unit source ac023b). For Crescent City, we have already noted that loss of 7 and 8 means that none of the remaining buoys can provide a 3 hrs warning for events from either ac034b, ac035b or ac036b. With the additional loss of 6, events from ac031b, ac032b or ac033b are also problematic because of lack of a fourth buoy that can provide a 3 hrs warning. There are similar concerns for ac029b and ac030b when 5, 6 and 7 are jointly inoperative. Turning to Port San Luis, the worst case scenario involves buoys 7, 8 and 6; the associated unit source is ac034b; and backup comes from buoy 5 with an RMSE ratio of 42.2. The buoys and unit sources match up well with the leave-out-two case.

To help assess the effect of one, two or three buoys becoming inoperative, we introduce a simple indicator of the current status of the network reflecting its ability to forecast maximum wave amplitudes at a given impact site. The indicator has three levels:

1. if the RMSE ratio is less than 2, we award the network a 'green' status (the network is healthy);
2. a ratio between 2 and 4 indicates a 'yellow' status (the network has deteriorated somewhat); and
3. a ratio greater than 4 is a 'red' status (there is serious deterioration)

(the values 2 and 4 are placeholders for our demonstration, but are arbitrary and subject to change when used in practice). In Fig. 5, there are vertical dotted lines at 2 and 4 in each of the plots – these delineate the boundaries between the three levels. Considering the plot for Hilo, we see that the network would still have a green status even if any one of 6 buoys (`U` to `Y` or `8`) out of a total of 14 (i.e., 43%) were to drop out by itself; on other hand, loss of `5` would result in a red status (7%), while losing any of the remaining 7 buoys would give a yellow status (50%). Figure 7(a) depicts these percentages, and plots (b) and (c) show corresponding results for the two California sites. The middle and bottom rows of Fig. 7 shows the percentages when two or three buoys drop out together. Not surprisingly, as more buoys drop out, the percentage of cases with red status increases for a given impact site. Note that three buoys can drop out but there can still be a green rating at all three impact sites for events arising from Aleutian Islands (but presumably not elsewhere). On the other hand, dropout of the single buoy `7` can raise a red alarm for Port San Luis. Once a particular buoy has dropped out, we can assess the

effect of loss of one of the other buoys, thus giving an indication of the urgency of returning the buoy to normal operation. This exercise would address the question of how close the array is to falling into a red status.

As a concluding example, the National Data Buoy Center (`ndbc.noaa.gov`) reported that six of the 14 buoys we have been considering were inoperative at the time of writing (early May 2017), namely, `U`, `V`, `W`, `X`, `Y` and `1`. Loss of `1` alone is enough to raise a yellow alarm for all three impact sites (see Fig. 5). This loss coupled with the loss of any one of the other five buoys does not increase the alarm level for the three sites (see Fig. 6). With `1` gone, the alarm status for Hilo would switch from yellow to red only if buoy `5` were to become inoperative; for Crescent City, if `Z` or `2` become inoperative; and for Port San Luis, if `Z` or `7` become inoperative (again, see Fig. 6).

4.2 Second Case Study: South America

In the preceding case study we concentrated on evaluating an existing network of DART® buoys. Here we illustrate another use for our proposed methodology, namely, to assess the impact of extending an existing network by adding a new buoy.

Figure 8 is a map of the West Coast of South America showing the locations of 130 predesignated unit sources, nine operational DART® buoys (labeled `1` to `9`) and one decommissioned buoy (`X`). These buoys can provide forecasts both for Southern Hemisphere impact sites and for Northern ones such as Hilo (indicated by HIL on the map; we do not report on results for Crescent City and Port San Luis because they are quite similar to those for Hilo). Merely to exercise our proposed

methodology, we presume buoys 1 to 8 to be a preexisting network and consider the benefit to Hilo of adding buoy 9. In addition we explore a hypothetical question: what would have been the effect on Hilo if, rather than augmenting the network by adding buoy 9 (a coastal buoy), buoy X (an open ocean buoy) had been reactivated instead?

Figure 9 shows RMSEs of forecasted maximum wave amplitudes at Hilo for tsunami events arising from South America. The requirement that a warning to Hilo must be issued at least 3 hrs in advance of the arrival of the tsunami is satisfied for all buoy/unit source combinations with two exceptions: buoy 1 with cs109b, and the same buoy with cs114b. The RMSEs for buoys 9 and X are colored, respectively, red and blue. For the network consisting of buoys 1 to 8 (black symbols in Fig. 9), there is an increase in RMSE of approximately an order of magnitude going from the northern-most central unit sources (cs052b to cs080b) to the southern-most (cs100b to cs114b). This pattern is in contrast to Fig. 3 for the Aleutian–Alaskan subduction zone, where the RMSEs are not strikingly different across all unit sources. Augmenting the network of eight buoys with coastal buoy 9 leads to a decrease in best RMSE by a factor to two or more for unit sources cs087b to cs094b (the unit source closest to 9 is cs090b – see Fig. 8); however, although there is some improvement in best RMSE from unit source cs085b up to cs097b, there is none elsewhere. The augmented network thus does not improve forecasts of wave amplitudes at Hilo for events arising from the southern-most unit sources.

Suppose now that we augment the 1 to 8 network by adding the open ocean buoy X rather than buoy 9. We would now have a factor of two or more decrease in best RMSE for unit sources cs093b to cs114b (cs100b is an exception – there is still a noticeable decrease, but less than a factor of two; in addition, there are

small improvements in best RMSEs for unit sources cs091b and cs092b). Thus, in contrast to augmenting the network with buoy 9, augmenting with X improves the response of the network in handling events events from the southern-most unit sources. The X augmentation improves the best RMSE at 24 unit sources, while the 9 augmentation, at 12 unit sources; however, the drastic improvement offered by the 9 augmentation at unit sources close to cs090 is not matched anywhere by the X augmentation.

As is evident from Fig. 8, a quicker warning to Hilo from an event originating from a unit source near the tip of South America is possible using data from buoy 9 rather than X. Consider unit source cs109b as an example. As indicated by Fig. 9, the three smallest RMSEs for cs109b are associated with buoys X, 5 and 9 and have values of, respectively, 0.14, 0.49 and 0.56 cm. The corresponding travel times from cs109b to these buoys are 9.8, 5.9 and 3.0 hrs, whereas the travel time from cs109b to Hilo is 15.4 hrs. Use of data from any of the three buoys should result in a warning issued well within the 3 hrs called for by standard operating procedures. Buoys 5 and 9 yield forecasts of approximately the same quality, so adding 9 to the 1 to 8 network offers no improvement in RMSE. The cleaner signal appearing at X offers a decrease in RMSE by a factor of 3.5 over the one at 5. If we assume a wait of 1 hr after the arrival of the signal to capture data 21 min past the first peak, warnings based on 5 and 9 could be issued 8.5 and 11.4 hrs in advance, whereas the corresponding time for buoy X is 4.6 hrs. Adding X to the existing network improves the RMSE (and hence the quality of the warning) at the expense of decreasing the warning time by 3.9 hrs; on the other hand, adding 9 offers no improvement in RMSE but increases the warning time by 2.9 hrs. If we think of 5 as providing an initial assessment of the event su ciently early, adding

`X` opens up the possibility of a later – but still timely – higher-quality assessment whereas adding `9` does not.

In terms of improving forecasts for Hilo, augmenting the `1` to `8` network by reactivating the open ocean buoy `X` is arguably preferable to going with coastal buoy `9`; however, there are many other impact sites of importance, each of which would need to be evaluated in a similar manner, and the individual evaluations would need to be combined to come up with an objective evaluation of which buoy is the preferable augmentation (how best to combine the evaluations is a subject for future research). Our intent is not to take issue with the current placement of South American buoys, but rather to demonstrate how our methodology can be used to evaluate variations of an existing network.

5 Discussion

Four aspects of our method for evaluating buoy networks merit additional discussion. First, as described in Sect. 3 and discussed in more detail in the appendix (Sect. 7), our approach is based on certain simplifications, including (but not limited to)

1. tsunami-generating events that come from within a single unit source;
2. mismatches between tsunami events and their associated models solely due to uncertainty in source location; and
3. use of data from a single DART® buoy to forecast wave amplitudes at an impact site even though tsunami events are routinely observed at multiple buoys (i.e., we do not combine together data from several buoys to generate forecasts as is done during actual tsunami events).

With regard to 1, we are in effect assuming that highly localized tsunami events are useful for assessing the effectiveness of a network of buoys even though events originating from regions spanning more than a single unit source occur more typically. A valid concern with this simplification is that use of tsunami events that are large, complex and/or spread out over an area larger than that trapped by a single unit source might lead to a substantially different network evaluation. While we hypothesize that this is not the case, it is a hypothesis that should be addressed in future research. With regard to 2, while there are other potentially important causes for mismatches besides the one we have chosen to focus on, we are assuming that, while including these additional causes might lead to increased RMSEs, the pattern of RMSEs across unit sources would not change to such a degree that network assessment would be substantially different. With regard to 3, we are assuming that forecasts based on each buoy separately can assess the effectiveness of the overall network and that using two or more buoys simultaneously won't lead to a significantly different evaluation even though use of data from multiple buoys for generating forecasts is more typical. Going beyond these three simplifying assumptions introduces levels of complexity that would complicate the simulation procedures needed for the network evaluation. In addition, assumption 3 provides a way of evaluating the contribution to the network of each buoy individually in a simple manner.

Second, our performance measure is predicated on tsunami events coming from a specific subduction zone (e.g., Aleutian–Alaskan) and heading toward a specific impact site (e.g., Hilo). We have not addressed the question of how to take the measures for different subduction zones and different impact sites and combine them together. For example, combining measures for multiple subduction zones

and Hilo would lead to an evaluation of how well Hilo is serviced by the global buoy network (and would likely reveal that, while buoys `U`, `V` and `W` in Fig. 1 proved to be of limited use to Hilo for events arising from the Aleutian–Alaskan subduction zone, they are vital for events arising in the Western Pacific such as the disastrous March 2011 Japan event). On the other hand, combining measures for a particular subduction zone and multiple impact sites would tell us how well the buoy network offers protection globally for events arising from that particular zone. Combined measures such as these are certainly of interest, but how best to do so is outside the scope of this article.

Third, we have concentrated on impact sites that are distant from the unit sources generating the tsunami events. For these sites insisting on a 3 hrs warning time is appropriate. For sites that are close to a particular unit source (of which there are many along the West Coast of South America), a more stringent warning time (e.g., 30 min) must unfortunately be entertained. A mixture of impact sites – some distant and some close to a given unit source – adds a degree of complexity (beyond the scope of this article) for assessing how well a tsunami event arising from such a source is handled by a buoy network.

Fourth, our work on the Aleutian–Alaska subduction zone points to an interesting potential practical method for quantifying uncertainties in forecasts. Figure 10 shows RMSEs at the three impact sites versus maximum wave amplitude at the buoys. For a given buoy, Fig. 10 shows that the lowest RMSEs are associated with the highest maximum wave amplitudes. That is, buoys that are positioned well to sample the strongest portion of the main tsunami beam yield the best forecasts in terms of RMSEs. The relationship between the maximum wave amplitudes at the buoys and the RMSEs is roughly linear on a log/log scale, with a fairly strong

sample correlation of 0.88. This suggests that the RMSE in forecasts at impact sites can be usefully assessed from the maximum wave amplitude at the buoys. This opens up the possibility of assigning a quality measure to forecasts at impact sites based on a simple summary of the data collected at the buoy.

6 Summary and Conclusions

We have developed a performance measure that quantifies the impact on forecasts of open-ocean wave amplitudes near an impact site due to loss of one or more buoys in a network of DART® buoys. The measure is based on a simulation procedure that takes into account key factors contributing to uncertainty in the forecasts. One factor is the model for the tsunami signal, which is never known perfectly in practice, in part due to uncertainty in the location of the tsunami-generating earthquake. We mimic this uncertainty by taking a predetermined central unit source and relocating it through random selection of a new location for its center. This relocation leads to perturbations both in the tsunami signal as it appears at DART® buoys and in the open-ocean wave amplitudes. The perturbations are codified by source coe cients, which serve as weights for forming the perturbed signal and wave amplitudes through linear combinations of signals and wave amplitudes associated with the central unit source and ones abutting it.

Tidal fluctuations and background noise recorded at the buoys are two additional factors contributing to forecast uncertainty. We mimic these through the use of archived DART® data recorded under ambient conditions either by the buoy of interest or by a surrogate operating under similar oceanographic conditions. The simulated tsunami signal and randomly selected archived data are added together

to form simulated buoy data. These data are the response in a regression model for estimating source coe cients, which in turn lead to estimates of wave amplitudes. Discrepancies between the estimated and presumed coe cients translate into errors in forecasted wave amplitudes. These estimation errors are due in part to tidal fluctuations and background noise in the simulated buoy data, but are also influenced by the amount of available data and by a mismatch between the regression model and the presumed model for the signal, both of which we mimic in our simulations.

A thousand replications of our simulation procedure for a given unit source/buoy combination gives a thousand forecasts of open-ocean wave amplitudes at as many impact sites as desired. To summarize the quality of the forecasts, we compare the thousand maximum values of the forecasted amplitude to the corresponding presumed maximum values by forming a root-mean-square error (RMSE). The RMSEs for a network of buoys in conjunction with a set of unit sources relevant for a particular impact site form the basis for network evaluation. The evaluation consists of dropping one or more buoys from the network and determining what change there is in the minimum RMSE for each unit source. We propose to use the maximum change in minimum RMSE as a metric for measuring the health of the buoy network as one or more buoys become inoperative. We also propose a simple green/yellow/red indicator of network health as a management tool.

Currently we have prototyped our procedure for evaluating the effectiveness of DART® buoy networks in the `R` language (Ihaka and Gentleman 1996; R Development Core Team 2010). Using the `R` code as guidance, current plans are to encapsulate the methodology we have described into a tool that would serve multiple purposes. We propose to use the tool to go beyond our Aleutian Island and

South American case studies to do a comprehensive study of the existing network of 39 buoys. This study might identify unit sources for which the current network configuration is particularly sensitive at certain impact sites to the dropout of buoys. We could also use the tool to assess the impact of adjusting the network by relocating, adding or removing buoys. Subject to additional research, it should be possible to augment the tool to investigate the use of tide gauges to help mitigate the loss of buoys. Despite the simplifications we have made, we contend that our proposed methodology is an effective and realistic way to assess network effectiveness and that a tool built around this methodology would be valuable for managing networks of DART® and other types of tsunami buoys.

Acknowledgements This work was funded by the Joint Institute for the Study of the Atmosphere and Ocean (JISAO) under NOAA Cooperative Agreement No. NA15OAR4320063 and is JISAO Contribution No. 2714. This work is also Contribution No. 4507 from NOAA/Pacific Marine Environmental Laboratory. The authors thank Peter Dahl for discussion on a running example.

7 Appendix: Details on Simulations and Evalution of Forecasts

Here we give details about how we have chosen to simulate and evaluate open-ocean wave amplitude forecasts. This material is summarized in Sect. 3, where we noted four tasks to be carried out, which are described in Sect. 7.1 to 7.4. We illustrate the four tasks by focusing on a representative triad consisting of one predesignated central unit source (`ac026b`), one DART® buoy (46403) and one impact site (Hilo HI).

7.1 Simulation of tsunami signal and open-ocean wave amplitudes

Our first task is to simulate a tsunami event, which manifests itself as a tsunami signal that passes by a particular DART® buoy, say, 46403 (Fig. 1 shows the location of this buoy). We denote this signal by $\boldsymbol{t}$. A simple way to specify $\boldsymbol{t}$ would be to use one of the precomputed models predicting what 46403 should see if the tsunami event were to originate from within a particular predesignated unit source, say, `ac026b` (shown in Fig. 1). We would thus just set $\boldsymbol{t}$ equal to $\boldsymbol{g}_5$ depicted in Fig. 2; however, we have chosen *not* to adopt this simple approach because an important factor impacting open-ocean wave amplitude forecasts is modeling of the tsunami signal $\boldsymbol{t}$ (this is the subject of Sect. 7.3). Because we cannot model the signal perfectly in an actual tsunami event, there is a danger of generating unrealistic simulated forecasts if we were to assume the signal to be identical to its model. One of many sources of mismatch is uncertainty in the location of the source within the subduction zone. Figure 11 illustrates the effect of relocating unit source `ac026b`, producing a relocated 100 km by 50 km source that overlaps (in this example) with three additional predesignated sources. To create the relocated source, we picked a random point (the red diamond) within `ac026b` and used this point as the center for the relocated unit source (blue rectangle). We can now compute a model predicting what would be observed at buoy 46403 if an earthquake were to originate from the randomly relocated unit source. This new model will differ from the precomputed models shown in Fig. 2. If we take this new model to be our tsunami signal $\boldsymbol{t}$ and if we were then to entertain modeling $\boldsymbol{t}$ using a subset of the models shown in Fig. 2, then the constructed signal need not be exactly equal to its model. Randomly choosing a set of locations within a

central unit source (`ac026b` in this example) then leads to an ensemble of relocated sources that reflects the uncertainty in source location.

Because it is time consuming to generate geophysical models, we make an assumption of linearity and construct the tsunami signal $\boldsymbol{t}$ using a linear combination of precomputed models. For the example shown in Fig. 11, the signal would be a linear combination of the models for unit sources `ac025a`, `ac026a`, `ac025b` and `ac026b` (the central unit source). These models are labeled in Fig. 2 as $\boldsymbol{g}_1$, $\boldsymbol{g}_2$, $\boldsymbol{g}_4$ and $\boldsymbol{g}_5$. The weights w_k in the linear combination are based on geometry and are dictated by the degree of intersection of the relocated source with the predesignated sources – in this example, $\boldsymbol{g}_5$ (corresponding to `ac026b`) gets the most weight because it intersects the most, while $\boldsymbol{g}_1$ (corresponding to `ac025a`) gets the least nonzero weight. Figure 12 illustrates the construction of the N-dimensional column vector $\boldsymbol{t}$ in terms of the six vectors $\boldsymbol{g}_k$ shown in Fig. 2:

$$\boldsymbol{t} = \sum_{k=1}^{6} w_k \boldsymbol{g}_k \tag{1}$$

(black curve, lower middle plot). Note that, since neither $\boldsymbol{g}_3$ nor $\boldsymbol{g}_6$ are involved in constructing $\boldsymbol{t}$, the weights w_3 and w_6 are zero. Past experience with actual tsunami events suggests that the weights should not be arbitrarily set if we want to regard the simulated event as originating from a 100 km by 50 km region (Hanks and Kanamori 1979; Papazachos et al. 2004). We use the normalization $\sum_k w_k = 4$ in all of our simulations because a larger setting than this would result in a tsunami signal whose magnitude would be more realistically associated with multiple unit sources. For the example shown in Fig. 11, we have $w_1 \doteq 0.28$, $w_2 \doteq 1.04$, $w_4 \doteq 0.57$ and $w_5 \doteq 2.11$.

We can construct additional tsunami signals by selecting other points at random within the rectangle for the central unit source `ac026b`. Figure 13 breaks the rectangle for `ac026b` up into four quadrants of equal size (labeled as I, II, III and IV). The specific models $\boldsymbol{g}_k$ that are used to construct the tsunami signal depend upon which quadrant the randomly selected point falls in – these are listed in Fig. 13 for the four quadrants. The random pick in Fig. 11 falls in quadrant II, and hence, as previously noted, the signal is constructed using $\boldsymbol{g}_1$, $\boldsymbol{g}_2$, $\boldsymbol{g}_4$ and $\boldsymbol{g}_5$. No matter into which quadrant the random pick falls, the model $\boldsymbol{g}_5$ is always included with a weight w_5 satisfying $1 \leq w_5 \leq 4$, and this weight is close to 4 when the random pick is close to the center of the rectangle. The upper two quadrants involve three models in addition to $\boldsymbol{g}_5$, whereas the lower quadrants involve just one (either $\boldsymbol{g}_4$ or $\boldsymbol{g}_6$). The characteristics of the constructed tsunami signals will thus depend upon the randomly selected quadrant and where the randomly selected point occurs within the quadrant. No matter where the random pick ends up, we can write the resulting constructed tsunami signal as per Eq. (1), with the understanding that either two or four of the w_k's will be zero.

We can also construct tsunami signals by focusing on a unit source other than `ac026b`; however, for simplicity, we will use as a central unit source only those having a setup similar to `ac026b`, namely, that they are in row b and are surrounded by five other unit sources. This restriction for central unit sources simplifies computer code somewhat and arguably should not adversely impact the evaluation of the effectiveness of a buoy network (a hypothesis that is yet to be investigated).

Given a relocated unit source, there are two methods for generating the corresponding open-ocean wave amplitudes $\boldsymbol{h}$ outside of an impact site of interest. The more accurate method is via high-resolution model runs, which are time con-

suming; the less accurate makes use of lower-resolution runs already available in a precomputed propagation data base, which has the advantage of being easy to extract. We tested both methods and found that results based on the propagation data base differed little from those based on high-resolution model runs. We have thus elected to use the propagation data base approach, for which we take the presumed wave amplitudes to be

$$\boldsymbol{h} = \sum_{k=1}^{6} w_k \boldsymbol{h}_k, \tag{2}$$

where the weights w_k are identical to those in Eq. 1, while $\boldsymbol{h}_k$ is the forecast of the wave amplitudes in the open ocean that would occur from an event occurring in the same predesignated unit source associated with $\boldsymbol{g}_k$ (note, however, that, while $\boldsymbol{h}_k$ depends upon the location of the unit source and the location in the open ocean, it does not depend upon *any* of the buoy locations).

As an example, Fig. 14 shows wave amplitudes $\boldsymbol{h}_k$ at an open-ocean location outside of Hilo associated with the same six predesignated unit sources considered in Figs. 1 and 2. The left-hand plot of Fig. 15 shows $\boldsymbol{h}$ formed using the weights stated in the caption to Fig. 12 (the right-hand plot is explained in Sect. 7.3).

7.2 Simulation of 1-min stream

Our second task is to simulate data as it would be recorded at a DART$^{\text{R}}$ buoy. Due both to the complexity of recorded tsunami data and the way in which these data are used to create wave amplitude forecasts, simulations of observations and forecasts that are fully realistic are di cult and time consuming to create. The approach we take is to impose simplifications that nonetheless retain the salient factors limiting the ability of a buoy network to contribute to accurate forecasts.

As a starting point, we assume the 1-min stream recorded at a particular buoy is given by

$$\bar{\boldsymbol{y}} = \boldsymbol{x} + \boldsymbol{t} + \boldsymbol{\epsilon}, \tag{3}$$

where $\bar{\boldsymbol{y}}$ is a vector containing N consecutive values from the 1-min stream; $\boldsymbol{x}$ represents tidal fluctuations; $\boldsymbol{t}$ is the tsunami signal constructed as per Eq. (1); and $\boldsymbol{\epsilon}$ is background noise. Tidal fluctuations and background noise in the DART ᴿ data are two factors that can adversely impact inundation forecasts, so it is important to handle these realistically. Rather than simulating these factors, we can use archived DART ᴿ data that were recorded under ambient conditions and retrieved during routine servicing of the buoy. The tsunami signal in Eq. (3) is missing during ambient conditions, so random samples from historical data can serve to generate $\boldsymbol{x} + \boldsymbol{\epsilon}$ (for details about the sampling procedure, see Sect. 3, Percival et al. 2015). A complication is that not all currently deployed buoys have associated archived data (and proposed buoys certainly don't). To handle such buoys, we use data from a surrogate buoy with good matching oceanographic conditions.

Figure 16 shows a simulated 1-min stream $\bar{\boldsymbol{y}}$ constructed as per Eqs. (1) and (3). The top plot shows the tsunami signal $\boldsymbol{t}$ (this is the same as the black curve in middle bottom plot of Fig. 12). The middle plot shows the sum of tidal fluctuations $\boldsymbol{x}$ and background noise $\boldsymbol{\epsilon}$, which were obtained from archived data for DART ᴿ buoy 46403. The bottom plot shows $\bar{\boldsymbol{y}}$, which is the sum of the time series in the top and middle plots.

7.3 Estimation of source coe cients and forecasting of open-ocean wave amplitudes

With simulated 1-min stream $\bar{\boldsymbol{y}}$ in hand, the third task is to estimate the source coe cients α_k. In practice factors limiting the accurate estimation of α_k include tidal fluctuations, background noise, an insu cient amount of data and imperfect knowledge about the underlying tsunami signal. We assume a linear regression model of the form

$$\bar{\boldsymbol{y}} = \mu\mathbf{1} + \quad {}_1\boldsymbol{c} + \quad {}_2\boldsymbol{s} + \sum_{k\in\mathcal{K}} \alpha_k \boldsymbol{g}_k + \boldsymbol{e}, \tag{4}$$

where $\mathbf{1}$ is an N-dimensional vector of ones associated with the regression coe - cient μ; $\boldsymbol{c}$ is a vector with elements $\cos(\omega n \quad)$, $n = 0, 1, \ldots, N \quad 1$, and is associated with $\quad_1$ – here ω is the tidal frequency M2 and $\quad = 1$ min; $\boldsymbol{s}$ is associated with $\quad_2$ and is similar to $\boldsymbol{c}$ but its elements are given by $\sin(\omega n \quad)$; $\mathcal{K}$ is a subset of $\{1, 2, \ldots, 6\}$ that specifies the $\boldsymbol{g}_k$ to be used to model the tsunami signal; and $\boldsymbol{e}$ is a vector of stochastic errors assumed to have zero mean (note that this vector is *not* taken to be the same as $\boldsymbol{\epsilon}$ in Eq. (3)). As discussed in Percival et al. (2015), the coe cients μ, $\quad_1$ and $\quad_2$ and their associated vectors serve to model the tidal fluctuations in Eq. (3) in a simple – but statistically e cient – manner that is superior to other methods for handling the tides including filtering. Modeling of the signal $\boldsymbol{t}$ is handled through specification of $\mathcal{K}$. We consider four protocols for setting $\mathcal{K}$, which are of interest because they make different assumptions about what is known about the random pick and because they lead to different degrees of mismatch between the constructed signal and its model.

a. The *single unit source protocol* ('1 protocol' for brevity) sets $\mathcal{K}$ to be $\{5\}$; i.e., the model for the signal is $\alpha_5 \boldsymbol{g}_5$. With probability one, this protocol yields an

incorrect signal model in the sense that the constructed signal consists of $\boldsymbol{g}_5$ combined with either one or three additional $\boldsymbol{g}_k$'s, whereas our model for it involves just $\boldsymbol{g}_5$. This protocol in effect assumes we only know within which predesignated unit source the earthquake occurred (i.e., we know nothing about the random pick).

b. The *two unit sources protocol* ('2 protocol') assumes we have limited information about the random pick, namely, whether the earthquake originates from the left-hand or right-hand side of the rectangle (the random pick in Fig. 11 is left-handed). For a left-hand pick, we set $\mathcal{K}$ to $\{4, 5\}$, and we set it to $\{5, 6\}$ for a right-hand pick. For both settings, there are two $\boldsymbol{g}_k$'s in the model. With this protocol, there is a 50% chance of having a mismatch between the constructed signal and its model, and the nature of the mismatch is that the model has two too few unit sources.

c. The *four unit sources protocol* ('4 protocol') also assumes we know the left- or right-hand location of the earthquake. Now we set $\mathcal{K}$ to $\{1, 2, 4, 5\}$ for a left-hand pick and to $\{2, 3, 5, 6\}$ otherwise. There is again a 50% chance of having a mismatch, but now the nature of the mismatch is that there are two too many unit sources.

d. The *matched protocol* sets $\mathcal{K}$ so that it contains *exactly* the indices used in forming the constructed signal; i.e., we use the same set of $\boldsymbol{g}_k$'s both to form the signal and to model it so there is no mismatch. This protocol essentially presumes knowledge of the quadrant in which the random pick falls.

After selection of the protocol $\mathcal{K}$, the regression model of Eq. (3) is now fully specified. The model involves three coe cients for the tidal model (μ, $_1$ and $_2$)

and either one, two or four source coe cients α_k for the signal model. We estimate the regression coe cients using constrained least squares; i.e., the estimated coe cients $\hat{\mu}$, $\hat{\ }_1$, $\hat{\ }_2$ and $\hat{\alpha}_k$ are those minimizing

$$\bar{\boldsymbol{y}} \quad \mu\mathbf{1} \quad {}_1\boldsymbol{c} \quad {}_2\boldsymbol{s} \quad \sum_{k\in\mathcal{K}} \alpha_k \boldsymbol{g}_k \quad {}_2^2 \text{ subject to } \alpha_k \quad 0 \text{ for all } k,$$

where $\|\boldsymbol{x}\|_2^2 = \sum_n x_n^2$ is the squared Euclidean norm of a vector $\boldsymbol{x}$ with elements x_n. The non-negativity constraint on each source coe cient is critical because, without it, there is nothing to prevent the least square estimate of α_k from being negative, which would render it inconsistent with the type of generating event we are assuming (a reverse fault earthquake).

Another important factor that impacts the quality of the inundation forecasts is the amount of data N available for estimating the source coe cients. In general, more data in $\bar{\boldsymbol{y}}$ improves the quality of the estimated coe cients, which, in turn, improves inundation forecasts. To study the effect of N, we consider three settings intended to loosely mimic what would be available during different stages of an ongoing tsunami event. These settings are based on the location of the first peak of the constructed signal $\boldsymbol{t}$ (alternatively we could use the first peak of the noisy data $\bar{\boldsymbol{y}}$, but, to avoid issues that arise in automatic detection of a peak possibly distorted by noise, we have chosen to use the constructed signal instead). The first setting is 1 min past the first peak in the signal, and the second and third settings are 11 and 21 min past the peak. The three settings are illustrated in Figs. 16(a) and (c) using red circles. For this particular example, the settings correspond to 24, 34 and 44 min after the start of the earthquake, and we would estimate the coe cients based on placing the corresponding first $N = 25$, 35 or 45 simulated buoy measurements in (c) into $\bar{\boldsymbol{y}}$. For tsunami signals other than this example, we

identify the first peak and the same related three points, but only use a maximum of 60 min worth of data prior to the first peak in cases where this peak occurs more than 1 hr after the start of the earthquake (this mimics the amount of the 1-min stream available during an actual event).

As an example of how the sample size influences source coe cient estimation, let us focus on the simulated DART R buoy data $\bar{\boldsymbol{y}}$ shown in Fig. 16(c) and set $\mathcal{K}$ as per the 2 protocol. Because the random pick is left-handed, the model for the tsunami signal becomes $\alpha_4 \boldsymbol{g}_4 + \alpha_5 \boldsymbol{g}_5$. There is thus a mismatch between the model and signal since the latter makes use of $\boldsymbol{g}_1$ and $\boldsymbol{g}_2$ in addition to $\boldsymbol{g}_4$ and $\boldsymbol{g}_4$. Table 1 shows the estimated source coe cients for the three sample sizes along with the presumed weights w_4 and w_5. The estimates improve with increasing N in the sense that $\hat{\alpha}_k$ gets closer to w_k.

The forecasted wave amplitudes are

$$\hat{\boldsymbol{h}} = \sum_{k \in \mathcal{K}} \hat{\alpha}_k \boldsymbol{h}_k. \tag{5}$$

For the 2 protocol with a left-handed pick, these amplitudes are given by

$$\hat{\boldsymbol{h}} = \hat{\alpha}_4 \boldsymbol{h}_4 + \hat{\alpha}_5 \boldsymbol{h}_5,$$

where $\boldsymbol{h}_4$ and $\boldsymbol{h}_5$ are depicted in Fig. 14. Use of the estimates for α_k corresponding to $N = 45$ in Tab. 1 leads to the forecasted wave amplitudes shown in the right-hand plot of Fig. 15.

To summarize, estimates of the source coe cients α_k are imperfect in practice due to major factors including, inter alia,

1. an imperfect tidal model;
2. background noise;

3. a limited amount of data;
4. a mismatch between the assumed model and the actual tsunami signal; and
5. seismic noise, which is potentially important, but, because of timing, often does not come into play.

Our simulation study takes into account 1–4, but ignores seismic noise (which the next generation of DART $^{\text{R}}$ buoys is designed to suppress entirely).

7.4 Evaluation of forecasted open-ocean wave amplitudes

Our final task is to quantify how well the forecasted wave amplitudes $\hat{\boldsymbol{h}}$ match the presumed amplitudes $\boldsymbol{h}$. We considered three metrics. The first is the maximum cross-correlation between $\hat{\boldsymbol{h}}$ and either $\boldsymbol{h}$ or a lagged versions thereof; the second is the squared difference between the maximum amplitudes in $\hat{\boldsymbol{h}}$ and $\boldsymbol{h}$; and the third is the squared difference between the maximum 4-hrs 'energies' in $\hat{\boldsymbol{h}}$ and $\boldsymbol{h}$, where the energies in question are the sum of squares of data over all possible 4-hrs stretches contained within $\hat{\boldsymbol{h}}$ or $\boldsymbol{h}$. The latter two metrics are of more operational interest than maximum cross-correlation, which is also the most sensitive of the three to innocuous misalignments in time between $\hat{\boldsymbol{h}}$ and $\boldsymbol{h}$. In tests to date, we have found that use of either maximum wave amplitudes or maximum energies leads to evaluations of network effectiveness that are qualitatively similar. Since maximum wave amplitudes are less time consuming to compute, we stick with these in all that follows. For the example shown in Fig. 15, we have $\max\{\boldsymbol{h}\} \doteq 2.3$ cm and $\max\{\hat{\boldsymbol{h}}\} \doteq 3.9$ cm, and the metric is $(\max\{\hat{\boldsymbol{h}}\} \quad \max\{\boldsymbol{h}\})^2 \doteq 2.6$ cm^2.

A more thorough assessment of how well we can forecast wave amplitudes outside of Hilo requires repeating what is set forth in the previous three subsections.

We do so by generating 1000 different tsunami signals $\boldsymbol{t}$ as per Eq. 1. The signals are associated with 1000 independent random picks from within unit source `ac026b`. Each pick serves as a center for a relocated unit source whose location dictates the weights w_k used to form both $\boldsymbol{t}$ and the presumed open-ocean wave amplitudes $\boldsymbol{h}$ of Eq. 2. Each signal $\boldsymbol{t}$ is then added to an independently drawn random sample of data recorded under ambient conditions by an appropriate DART® buoy. This addition forms the simulated 1-min stream $\bar{\boldsymbol{y}}$ as per Eq. 3. For each realization, we then compute twelve sets of estimated source coe cients $\hat{\alpha}_k$ corresponding to twelve formulations of the linear model of Eq. 4. The formulations involve the four protocols for setting $\mathcal{K}$ in combination with the three ways of determining the amount of data N to be placed in $\bar{\boldsymbol{y}}$. For a given formulation, we use the source coe cient estimates $\hat{\alpha}_k$ to generate – as per Eq. 5 – forecasted wave amplitudes $\hat{\boldsymbol{h}}$ outside of the chosen impact site (Hilo in our representative triad). We then compare the peak values in $\boldsymbol{h}$ and $\hat{\boldsymbol{h}}$.

Figure 17 shows forecasted maximum amplitudes versus corresponding presumed amplitudes for 1000 realizations of our representative triad (central unit source `ac026b`, DART® buoy 46403 and a location in the open ocean outside of the impact site Hilo). Each realization corresponds to a single point in each of the four plots. In addition to the 1000 points, each plot has a diagonal line indicating where a point would fall if the forecasted and presumed amplitudes were in perfect agreement. For all four plots, we set N such that 21 min worth of data past the peak value in the signal $\boldsymbol{t}$ is used to estimate the source coe cients and other parameters in the regression model (Eq. (4)). Plots (a) to (d) correspond to, respectively, the 1, 2, 4 and matched protocols $\mathcal{K}$ for fully specifying the model.

The colors of the points in Fig. 17 indicate the quadrant in which the random pick for a particular realization fell. Of the 1000 picks, 245 were in quadrant I (colored red); 247, in II (green); 268, in III (blue); and 240, in IV (black). This distribution is not inconsistent with each pick being equally likely to fall in any of the four quadrants. As noted in Fig. 13, picks in quadrants I and II involve linear combinations of four $\boldsymbol{g}_k$'s. For the 1 and 2 protocols, forecasted amplitudes are thus based on mismatched models. The mismatch is due to relevant $\boldsymbol{g}_k$'s being left out of the regression model. As a result, the red and green points in plots (a) and (b) have prominent scatter off the diagonal line. Plots (c) and (d) indicate that the forecasts are markedly better when there is no model mismatch, as occurs in the 4 and matched protocols (note that the red and green points in (c) and (d) are necessarily identical). In the absence of model mismatch, the remaining inaccuracies in the forecasts are mainly due to background noise and the tidal component.

By contrast, picks in quadrants III and IV (blue and black points) correspond to linear combinations of two $\boldsymbol{g}_k$'s. There is a model mismatch with the 1 and 4 protocols, but not for the 2 and matched protocols (note that the blue and black points in (b) and (d) are necessarily identical). Interestingly, when the mismatch is due to one less $\boldsymbol{g}_k$ in the 1 protocol, there are highly structured deviations from the diagonal line. On the other hand, when the mismatch is due to two extraneous $\boldsymbol{g}_k$'s, as happens with the 4 protocol, the scatter about the diagonal line visually does not increase much over that of the 2 protocol (no mismatch).

We can summarize how well forecasted maximum wave amplitudes match up with the presumed amplitudes by computing a root-mean-square error (RMSE) for each scatterplot. Letting $\max\{\hat{\boldsymbol{h}}_l\}$ and $\max\{\boldsymbol{h}_l\}$ represent the forecasted and

presumed amplitudes for the lth realization, this error is defined as

$$\text{RMSE} = \left[\frac{1}{1000} \sum_{l=1}^{1000} \left(\max\{\hat{\boldsymbol{h}}_l\} - \max\{\boldsymbol{h}_l\} \right)^2 \right]^{1/2}. \tag{6}$$

Figure 18 shows the RMSEs for the four protocols in combination with three settings N for the amount of data used to estimate the source coe cients (dictated by up to either 1, 11 or 21 min past the first peak in the signal $\boldsymbol{t}$). The four RMSEs corresponding to the scatterplots in Fig. 17 are shown above the '21' label on the horizontal axis. In addition this figure has a dashed line showing the RMSE for a so-called seismic solution, for which a forecast is generated using $4\boldsymbol{g}_5$. This procedure assumes idealized seismic information about the tsunami event, namely, that the generating earthquake occurs within the central unit source `ac026b` and that the magnitude of the earthquake suggests setting the source coe cient to 4, which corresponds to our standard normalization $\sum_k \alpha_k = 4$ for the relocated unit source. The seismic-based forecast is of interest because it is arguably the best that can achieved without any information supplied by data from DART® buoy 46403.

For all four protocols for formulating the regression model, the RMSEs decrease as more data are collected by the DART® buoy. The decrease is minimal for the 1 protocol, but substantial for the other three protocols (in particular, there is a drop of more than two orders of magnitude for the 4 and matched protocols). The 1 protocol always involves a mismatch between the signal and its model. For the other protocols, there is at least a 50% chance of no mismatch. The poor performance of the 1 protocol points out the need for an adequate regression model. On the other hand, there is little difference in the RMSEs for the 4 and matched protocols even though the former involves mismatches in about 50% of

the realizations, whereas the latter involves none. In contrast to the 1 protocol, the nature of the mismatch with the 4 protocol is too many $\boldsymbol{g}_k$'s rather than too few. This result suggests that having extraneous $\boldsymbol{g}_k$'s in the model (overfitting) does not significantly impact the forecasted wave amplitudes.

The seismic-based forecast outperforms the buoy-based forecasts either when there is an insu cient amount of data or when the regression model is always inadequate due to missing $\boldsymbol{g}_k$'s (the 1 protocol). When there is too little data, forecasts deteriorate considerably as the complexity of the regression model increases (i.e., more coe cients must be estimated). For our representative triad, there is some value to be gained by waiting the extra 10 min from 11 min past the first peak to 21 min (the RMSEs drop by 60% to 70%). For other triads involving different central unit sources and different buoys, the gain can be more substantial.

For the network assessment discussed in Sect. 4, we concentrate on the 4 protocol with N selected as dictated by 21 min after the first peak in the signal. This choice is close to the best RMSE for most – but not all – triads and should reflect the contribution of a particular buoy to wave amplitude forecasts at an impact site once enough data have been collected and once an adequate regression model has been identified.

References

Gica, E., Spillane, M.C., Titov, V.V., Chamberlin, C.D., and Newman, J.C. (2008), Development of the forecast propagation database for NOAA's Short-term Inundation Forecast for Tsunamis (SIFT). NOAA Technical Memorandum OAR PMEL–139, 89 pp. http://nctr.pmel.noaa.gov/pubs.html

Greenslade, D.J.M., and Warne, J.O. (2012), Assessment of the effectiveness of a sea-level observing network for tsunami warning. *Journal of Waterway, Port, Coastal, and Ocean Engineering*, 138(3), 246–255.

Hanks, T.C., and Kanamori, H. (1979), A moment-magnitude scale. *Journal of Geophysical Research*, 84(B5), 2348–2350.

Ihaka, R., and Gentleman, R. (1996), R: a language for data analysis and graphics. *Journal of Computational and Graphical Statistics* 5, 299–314.

Papazachos, B.C., Scordilis, E.M., Panagiotopoulos, D.G., Papazachos, C.B., and Karakaisis, G.F. (2004), Global relations between seismic fault parameters and moment magnitude of earthquakes, *Bulletin of the Geological Society of Greece*, 36, 1482–1489.

Percival, D.B., Denbo, D.W., Eblé, M.C., Gica, E., Huang, P.Y., Mofjeld, H.O., Spillane, M.C., Titov, V.V., and Tolkova, E.I. (2015), Detiding DART® buoy data for real-time extraction of source coe cients for operational tsunami forecasting, *Pure and Applied Geophysics*, 172(6), 1653–1678.

Percival, D.B., Denbo, D.W., Elbé, M.C., Gica, E., Mofjeld, H.O., Spillane, M.C., Tang, L., and Titov, V.V. (2011), Extraction of tsunami source parameters via inversion of DART® buoy data, *Natural Hazards*, 58(1), 567–590.

Percival, D.M., Percival, D.B., Denbo, D.W., Gica, E., Huang, P.Y., Mofjeld, H.O., and Spillane, M.C. (2014), Automated tsunami source modeling using the sweeping window positive elastic net, *Journal of the American Statistical Association*, 109(506), 491–499.

R Development Core Team (2010) http://www.r-project.org/

Spillane, M.C., Gica, E., Titov, V.V., and Mofjeld, H.O. (2008), Tsunameter network design for the U.S. DART® arrays in the Pacific and Atlantic Oceans.

NOAA Technical Memorandum OAR PMEL–143, 165 pp. http://nctr.pmel.noaa.gov/pubs.html

Titov, V.V. (2009), Tsunami forecasting. In E.N. Bernard and A.R. Robinson (Eds.), *The sea, volume 15: tsunamis* (pp. 371–400). Cambridge, MA: Harvard University Press.

Whitmore, P.M. (2009), Tsunami warning systems. In E.N. Bernard and A.R. Robinson (Eds.), *The sea, volume 15: tsunamis* (pp. 401–442). Cambridge, MA: Harvard University Press.

Whitmore, P., Benz, H., Bolton, M., Crawford, G., Dengler, L., Fryer, G., Goltz, J., Hansen, R., Kryzanowski, K., Malone, S., Oppenheimer, D., Petty, E., Rogers, G., and Wilson, J. (2008), NOAA/West Coast and Alaska Tsunami Warning Center Pacific Ocean response criteria. *Science of Tsunami Hazards*, 27(2), 1–21.

		$\hat{\alpha}_k$		
k	w_k	$N = 25$	$N = 35$	$N = 45$
4	0.57	88.92	1.63	1.33
5	2.11	2.51	2.34	2.10

Table 1 Estimated source coe cient $\hat{\alpha}_4$ and $\hat{\alpha}_5$ for three sample sizes N along with presumed weights w_4 and w_5

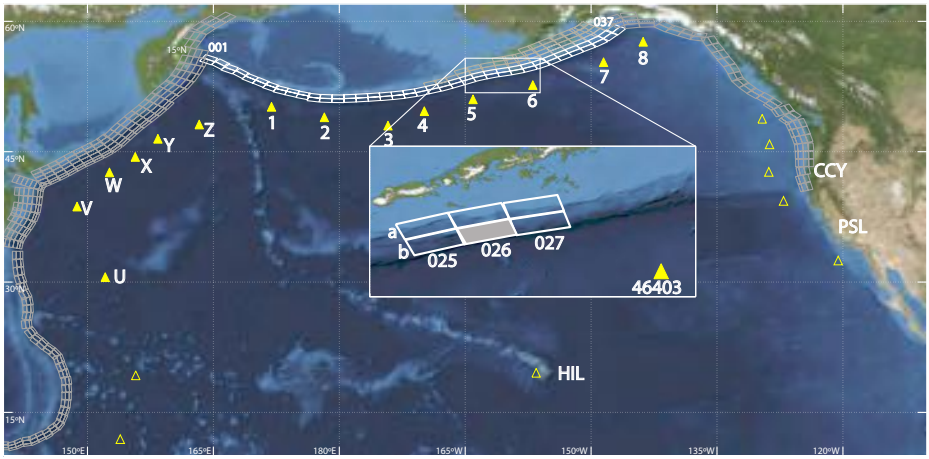

Fig. 1 Locations of 74 unit sources in Aleutian–Alaskan subduction zone (small white rectangles, with an expanded view of six of them), DART® buoys in the Northern Pacific (triangles, 14 of which are labeled by `U` to `Z` and `1` to `8`) and three impact sites (HIL marks Hilo, while CCY and PSL show where Crescent City and Port San Luis are). The unit sources are arranged in two rows denoted by `a` (upper row) and `b` (lower) and in 37 columns labelled as, from left to right, `001` to `037`. Going clockwise, the World Meteorological Organization designations for the 14 buoys are (`U`) 21413, (`V`) 21418, (`W`) 21401, (`X`) 21419, (`Y`) 21402, (`Z`) 21416, (`1`) 21415, (`2`) 21414, (`3`) 46413, (`4`) 46408, (`5`) 46402, (`6`) 46403, (`7`) 46409 and (`8`) 46410. The expanded view shows central unit source `ac026b` and the five unit sources abutting it, along with the location of the nearby buoy 46403

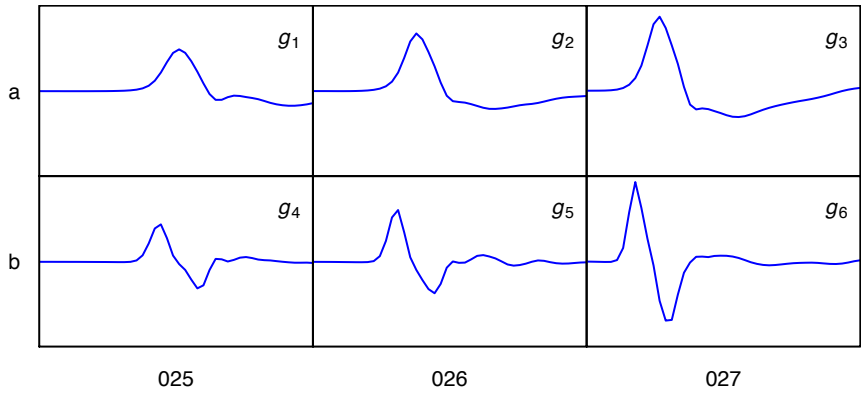

Fig. 2 Models for what would be observed at DART® buoy 46403 from an earthquake originating at the center of one of six unit sources in the Aleutian–Alaskan subduction zone (the 100 km by 50 km sections of the ocean corresponding to these unit sources are highlighted in Fig. 1). The top row shows the models for unit sources `ac025a`, `ac026a` and `ac027a` (labeled as $\boldsymbol{g}_1$, $\boldsymbol{g}_2$ and $\boldsymbol{g}_3$); the bottom, for `ac025b`, `ac026b` and `ac027b` ($\boldsymbol{g}_4$, $\boldsymbol{g}_5$ and $\boldsymbol{g}_6$). The time span for each displayed model is 45 min, and the span starts 10 min after the start of the earthquake. The vertical axes all range from −4.1 cm to 4.1 cm

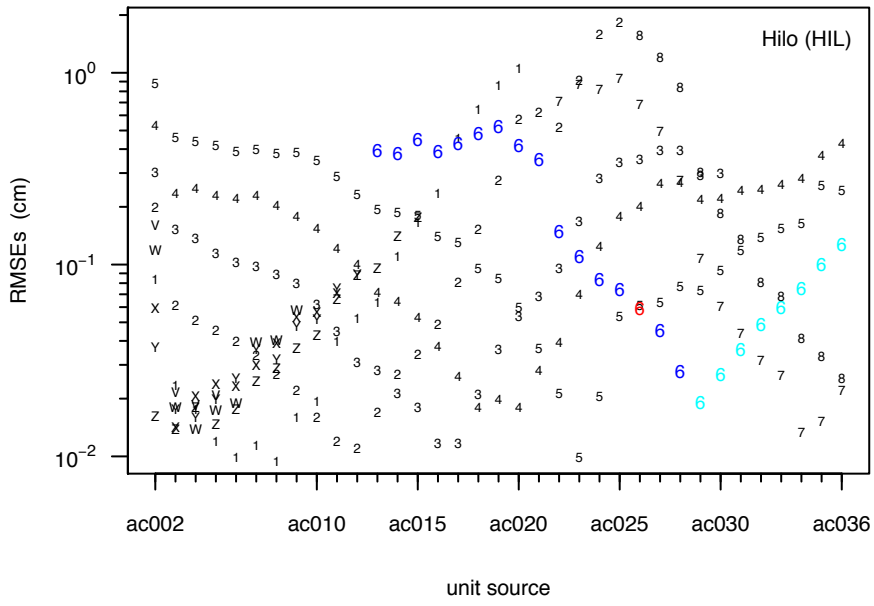

Fig. 3 RMSEs for forecasted maximum wave amplitudes in the open ocean outside of Hilo based on 14 buoys (labeled as indicated in Fig. 1) in combination with central unit sources `ac002b` to `ac036b` from the Aleutian–Alaskan subduction zone. The RMSEs for buoy 46403 (labeled as `6`) are colored, with red indicating its pairing with `ac026b`

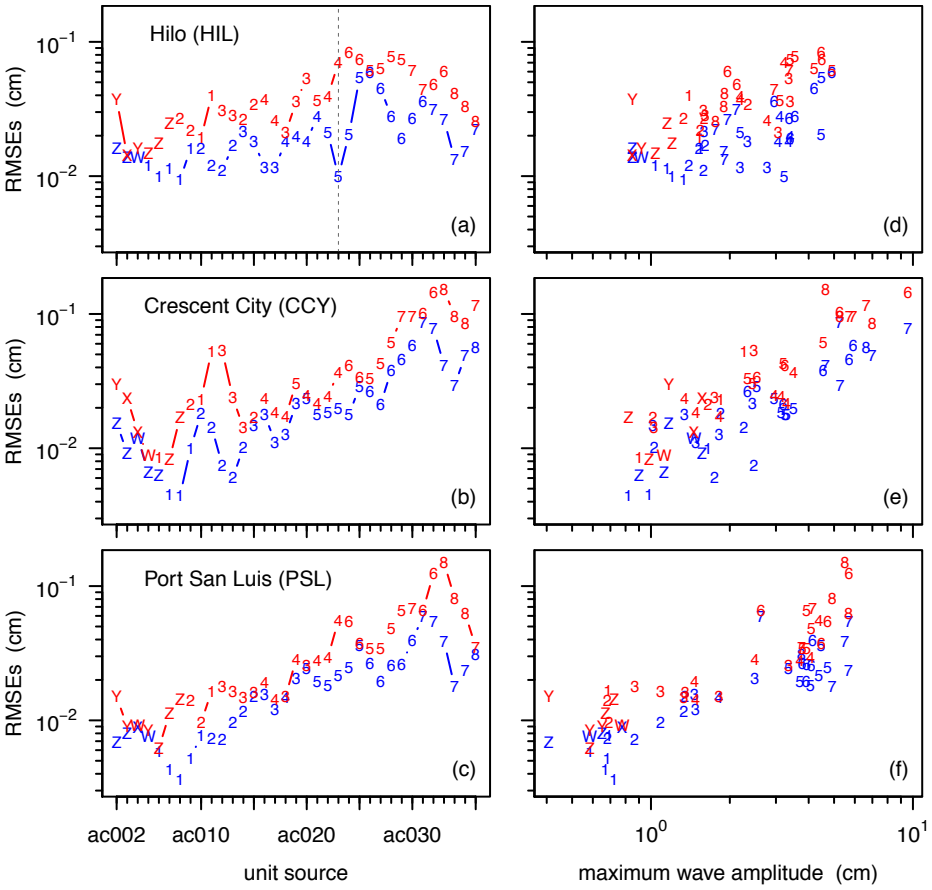

Fig. 4 Best (blue) and second best (red) RMSEs for forecasted maximum wave amplitudes in the open ocean outside of, from top to bottom rows, Hilo, Crescent City and Port San Luis. The RMSEs are plotted versus central unit sources `ac002b`, . . . , `ac036b` in the left-hand column and, in the right-hand, versus maximum wave amplitudes. Plot (a) shows RMSEs extracted from Fig. 3, and the vertical dashed line crosses the RMSEs for unit source `ac023b` in combination with buoys `5` and `4`

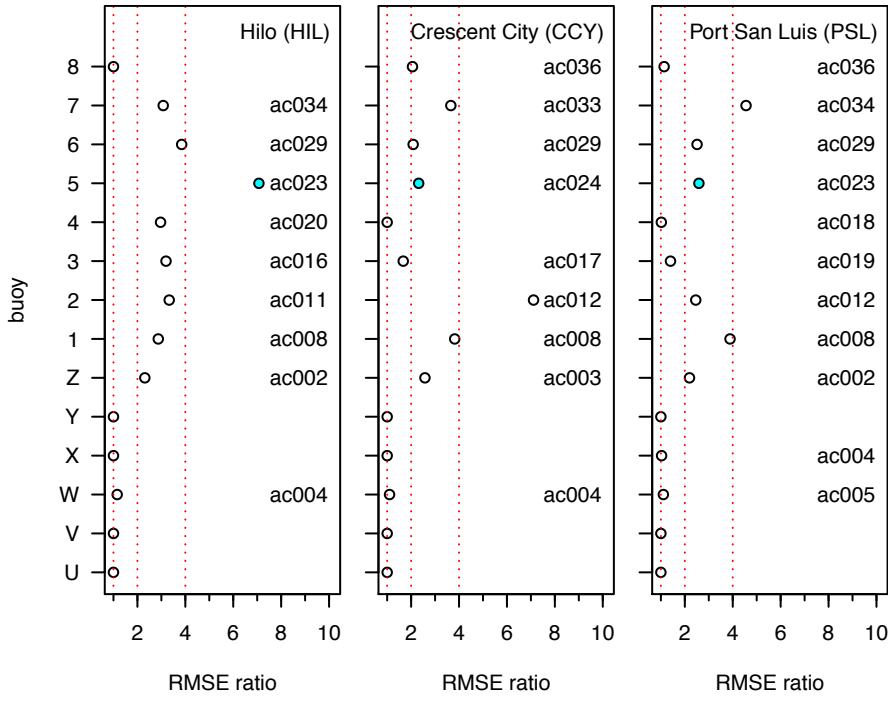

Fig. 5 Effect on performance of DART® buoy network across 35 unit sources and at three impact sites due to one-by-one removal of the 14 buoys depicted in Fig. 1. For each buoy, the largest increase in RMSE over the 35 unit sources is shown relative to the best available RMSE when that buoy is dropped as compared to when the full network of buoys is available. The case where buoy 5 is dropped is highlighted with colored circles. The one corresponding to Hilo is the change in RMSE indicated by the vertical dashed line in Fig. 4(a). The three red dashed lines mark RMSE ratios of 1, 2 and 4. These ratios are used to define a green, yellow and red network status – see Fig. 7 and the discussion pertaining to it

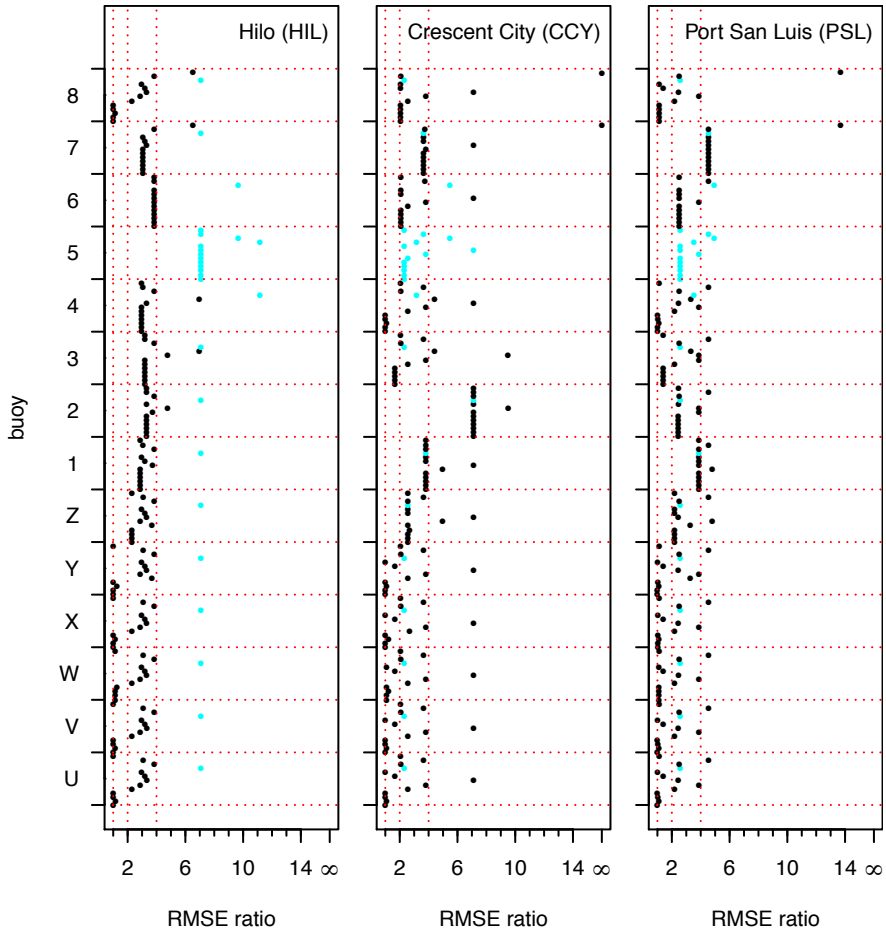

Fig. 6 Effect on performance of DART® buoy network across 35 unit sources and at three impact sites due to concurrent dropout of two of 14 buoys depicted in Fig. 1. For each pair of buoys, the largest increase in RMSE over the 35 unit sources is shown relative to the best available RMSE when both buoys are dropped as compared to when the full network of buoys is available. The RMSE ratios are shown as dots. The blue dots show cases for which buoy 5 is one of the dropped buoys. The three red vertical dashed lines mark RMSE ratios of 1, 2 and 4. These ratios are used to define a green, yellow and red network status – see Fig. 7 and the discussion pertaining to it

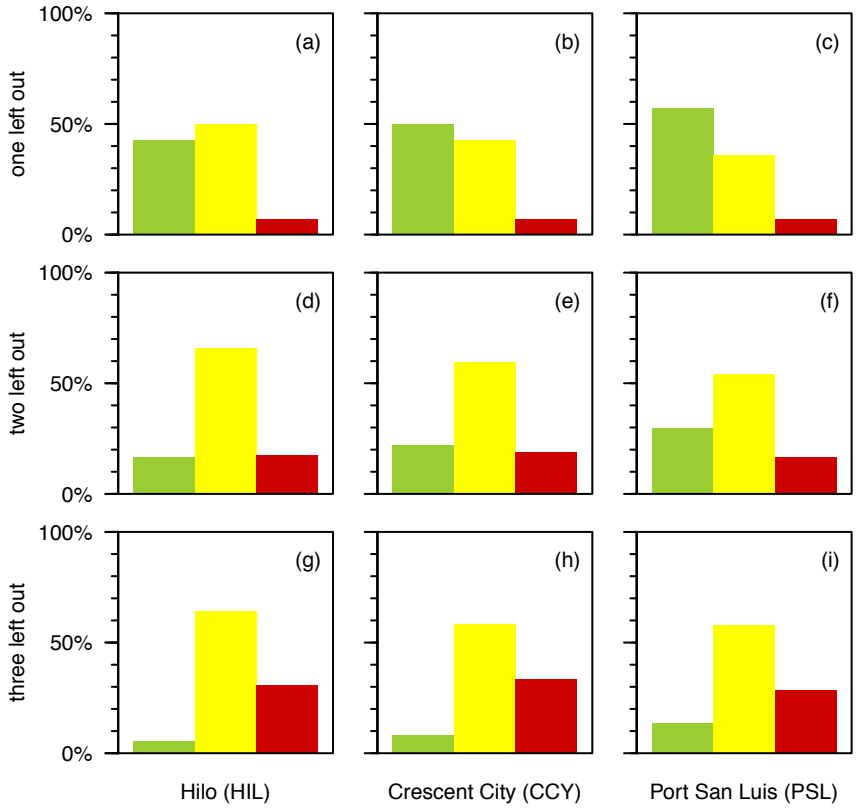

Fig. 7 Percentages of cases that would result in a green, yellow or red network status for Hilo, Crescent City and Port San Luis (left to right columns) due to 1, 2 or 3 of 14 DART® buoys becoming inoperative (top to bottom rows, respectively)

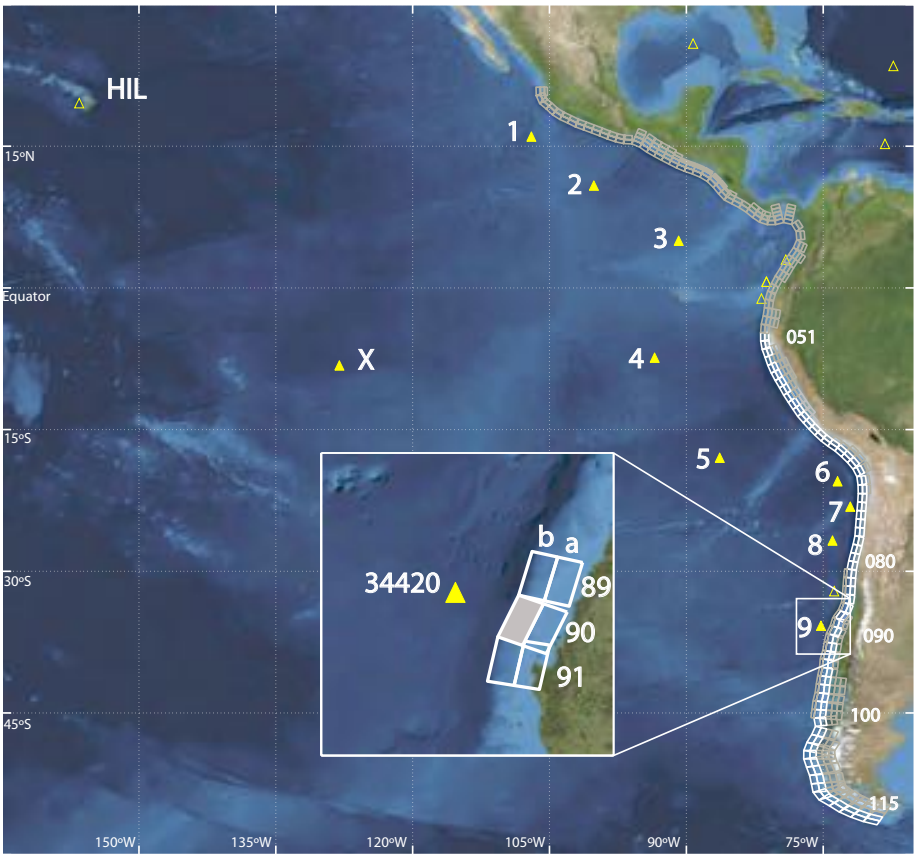

Fig. 8 Locations of 130 unit sources along the West Coast of South America, DART® buoys (triangles, ten of which are labeled by `1` to `9` and `X`) and the Hilo impact site (HIL). The unit sources are arranged in two rows denoted by `a` (eastern) and `b` (western) and in 65 columns denoted by, from top to bottom, `051` to `115`. Going clockwise from the top, the World Meteorological Organization designations for the ten buoys are (`1`) 43412, (`2`) 43413, (`3`) 32411, (`4`) 32413, (`5`) 32412, (`6`) 32401, (`7`) 32403, (`8`) 32402, (`9`) 34420 and (`X`) 51406. The expanded view shows central unit source `cs090b` and the five unit sources surrounding it, along with the nearby buoy 34420

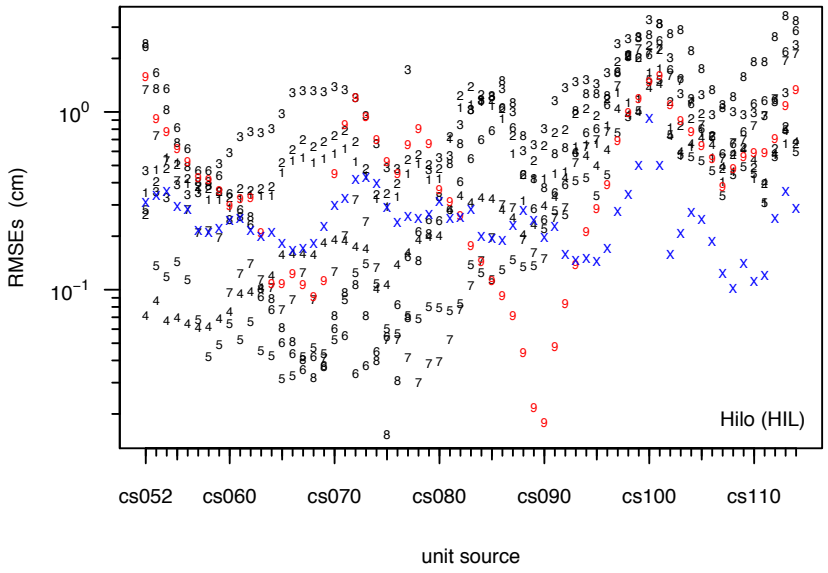

Fig. 9 RMSEs for forecasted maximum wave amplitudes in the open ocean outside of Hilo based on ten buoys (labeled as indicated in Fig. 8) in combination with central unit sources `cs052b` to `cs114b` along the West Coast of South America

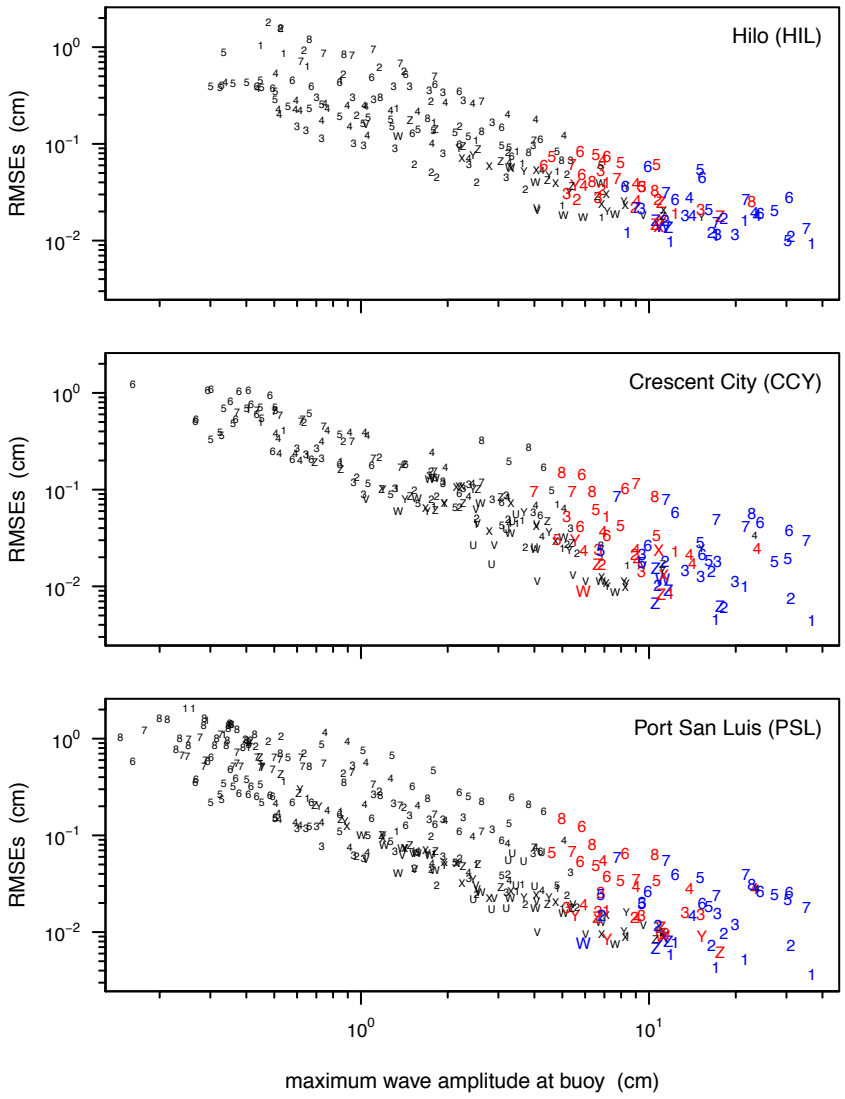

Fig. 10 As in the right-hand column of Fig. 4, but now with the RMSEs of the maximum wave amplitudes forecasted at the open ocean outside of the impact sites plotted versus maximum wave amplitudes at the DART® buoys rather than plotted versus maximum wave amplitudes in the open ocean outside of the impact sites. While only the best RMSEs (colored by blue symbols) and second best (red) are shown in Fig. 4, both these and all remaining RMSEs (black) are shown here

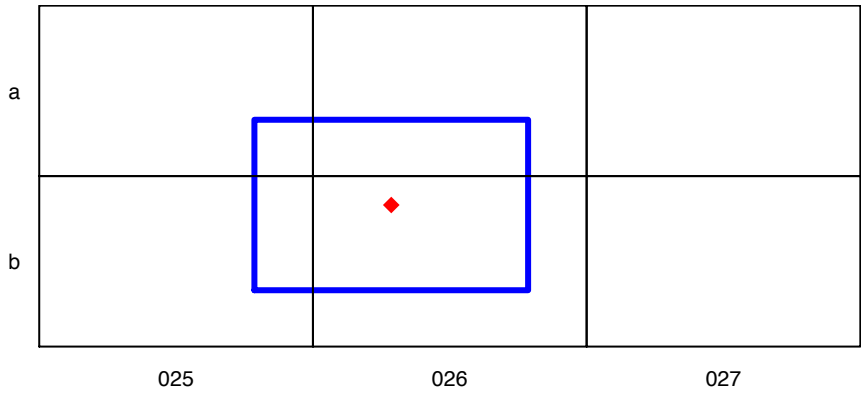

Fig. 11 Six rectangles (black lines) representing 100 km by 50 km sections of ocean associated with predesignated unit sources `ac025a`, `ac026a` and `ac027a` (top row) and `ac025b`, `ac026b` and `ac027b` (bottom) in the Aleutian–Alaskan subduction zone (these are highlighted in Fig. 1). The red diamond within the rectangle for the central unit source `ac026b` defines the center of a rectangle for a relocated unit source (blue lines). The relocated rectangle intersects the rectangles associated with `ac025a`, `ac026a`, `ac025b` and `ac026b` (these are associated with models $\boldsymbol{g}_1$, $\boldsymbol{g}_2$, $\boldsymbol{g}_4$, $\boldsymbol{g}_5$ shown in Fig. 2).

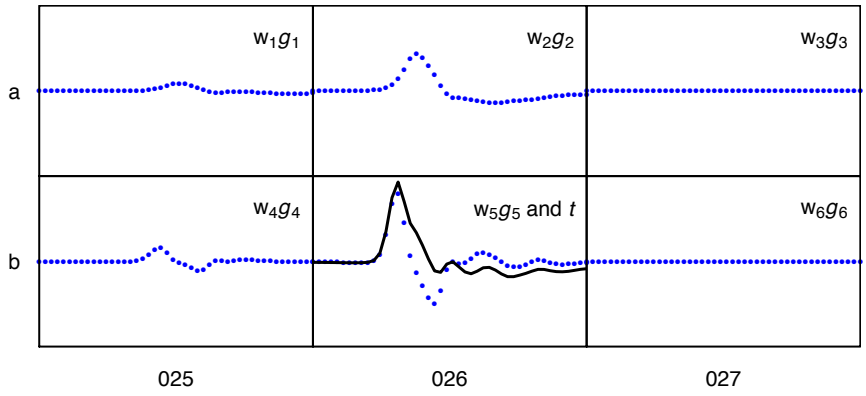

Fig. 12 Tsunami signal $\boldsymbol{t}$ (black curve, middle bottom plot) formed from linear combination of models $\boldsymbol{g}_1, \ldots, \boldsymbol{g}_6$ as per Eq. 1 with $w_1 \doteq 0.28$, $w_2 \doteq 1.04$, $w_3 = 0$, $w_4 \doteq 0.57$, $w_5 \doteq 2.11$ and $w_6 = 0$ (the $\boldsymbol{g}_k$ models are shown in Fig. 2). The blue dotted curves show $w_k \boldsymbol{g}_k$, and the sum of all six of these is equal to $\boldsymbol{t}$ (because $w_3 = w_6 = 0$, models $\boldsymbol{g}_3$ and $\boldsymbol{g}_6$ do not contribute to $\boldsymbol{t}$). The time spans are the same as in Fig. 2, but now the vertical axes range from −6.6 cm to 6.6 cm

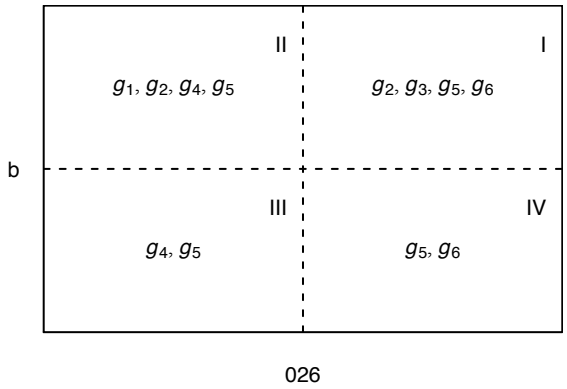

Fig. 13 Division of rectangle for central unit source `ac026b` into four quadrants (I, II, III and IV). A point chosen at random within this rectangle is equally likely to fall in each of the four quadrants. The models $\boldsymbol{g}_k$ used to construct the tsunami signal $\boldsymbol{t}$ depend upon which quadrant the random pick falls in – these models are listed in the middles of the quadrants. The model $\boldsymbol{g}_5$ for `ac026b` is always part of the constructed tsunami signal. The weight w_k assigned to each model depends upon the location of the random pick within the quadrant, with the weight for $\boldsymbol{g}_5$ increasing as the random pick gets closer to the center of the rectangle

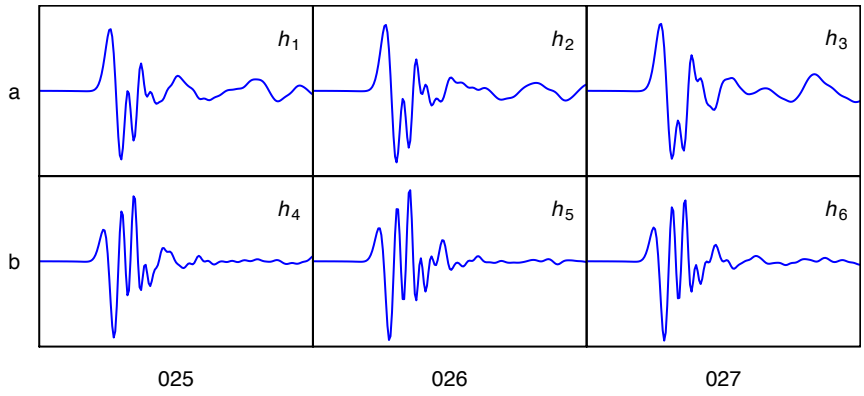

Fig. 14 Models $\boldsymbol{h}_1, \ldots, \boldsymbol{h}_6$ for the wave amplitudes in the open ocean outside of Hilo from an earthquake originating from either the predesignated central unit source `ac026b` (corresponding to $\boldsymbol{h}_5$) or one of five adjacent unit sources `ac025a` (corresponding to $\boldsymbol{h}_1$), `ac026a` ($\boldsymbol{h}_2$), `ac027a` ($\boldsymbol{h}_3$), `ac025b` ($\boldsymbol{h}_4$) or `ac027b` ($\boldsymbol{h}_6$) located in the Aleutian–Alaskan subduction zone (Fig. 1 shows the locations of the 100 km by 50 km sections of the ocean associated with these unit sources). The time span for each displayed model is 3 hrs, and the span starts at 4 hrs from the beginning of the earthquake. The vertical axes all range from −1.5 cm to 1.5 cm

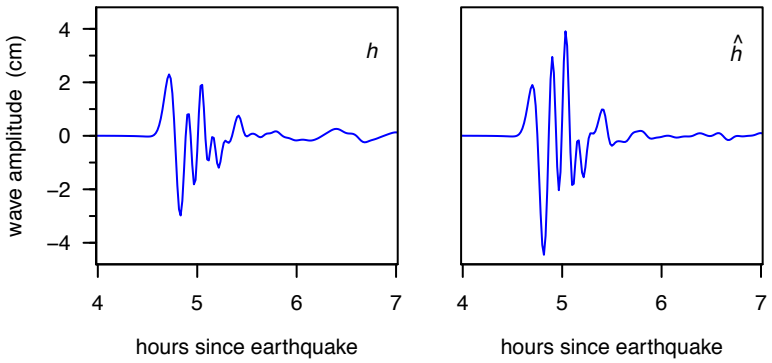

Fig. 15 Open-ocean wave amplitudes $\boldsymbol{h}$ outside of Hilo as dictated by relocated unit source of Fig. 11 (left-hand plot) and associated forecasted wave amplitudes $\hat{\boldsymbol{h}}$ (right-hand). The amplitudes $\boldsymbol{h}$ are given as per Eq. 2 with the same setting for w_k as listed in the caption to Fig. 12. The amplitudes $\hat{\boldsymbol{h}}$ are given as per Eq. 5 with $\mathcal{K} = \{4, 5\}$, $\hat{\alpha}_4 \doteq 1.33$ and $\hat{\alpha}_5 \doteq 2.10$

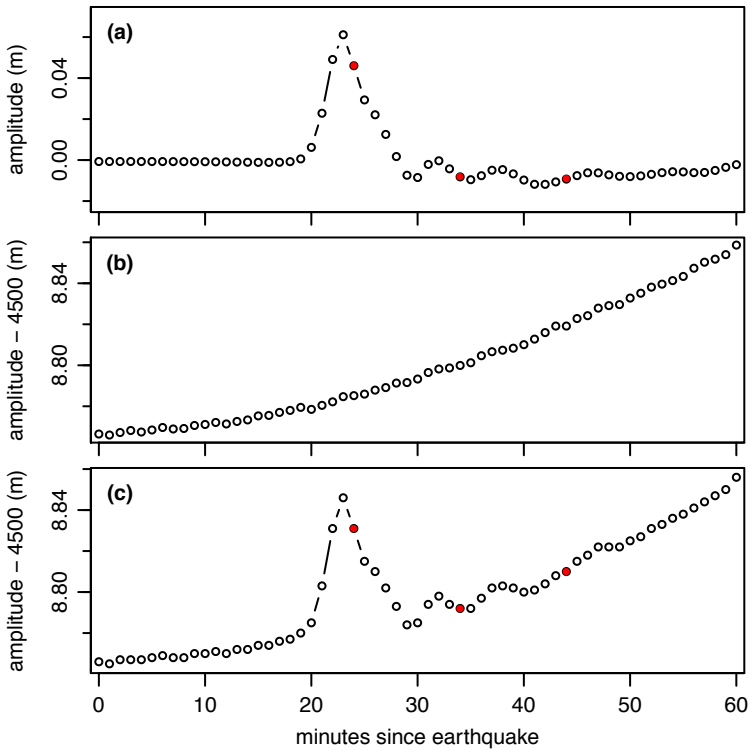

Fig. 16 Construction of a simulated 1-min stream $\bar{\boldsymbol{y}}$. Plot **(a)** is the simulated tsunami signal – this is the same as black curve in middle bottom plot of Fig. 12. The three red circles indicate data occurring 1, 11 and 21 min after the signal's peak value. Plot **(b)** shows tidal fluctuations and background noise, i.e., $\boldsymbol{x} + \boldsymbol{\epsilon}$, recorded by DART® buoy 46403 under ambient conditions. Plot **(c)** is $\bar{\boldsymbol{y}}$, which is the sum of the time series in plots **(b)** and **(c)**

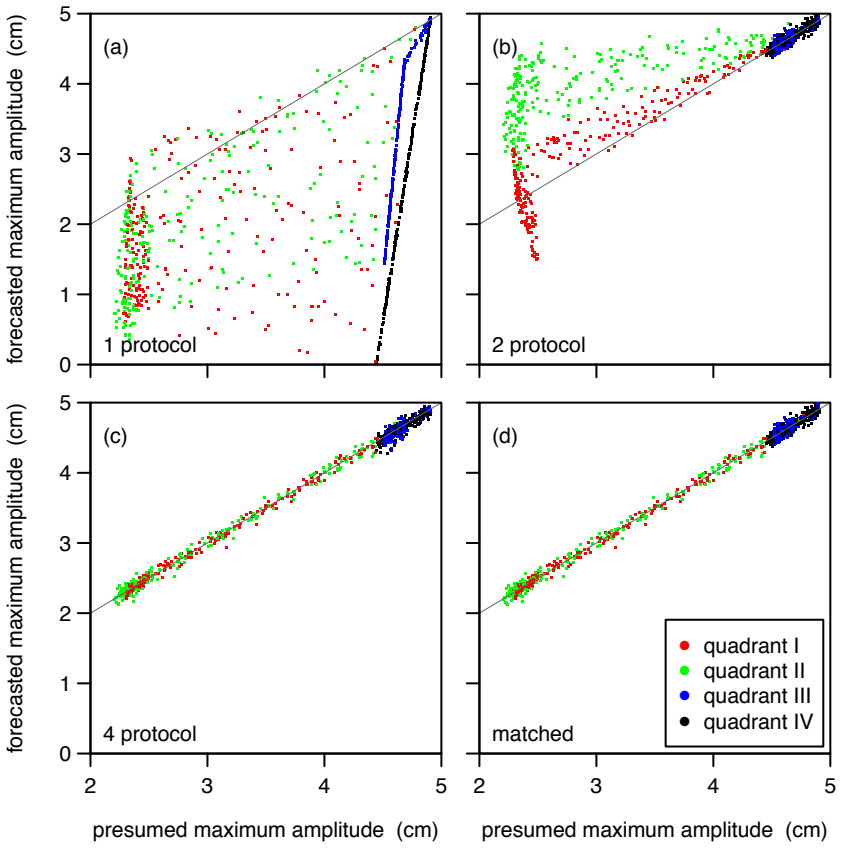

Fig. 17 Scatterplots of forecasted versus presumed maximum wave amplitudes for 1000 realizations of representative triad (central unit source `ac026b`, buoy 46403 and impact site Hilo). Plots (a), (b), (c) and (d) correspond to the 1, 2, 4 and matched protocols for fully specifying the regression model of Eq. (4). Each realization is based on a random pick that has an equal chance of falling within one of the four quadrants shown in Fig. 13, with the quadrant determining which models $\boldsymbol{g}_k$ receive nonzero weights w_k in constructing the signal $\boldsymbol{t}$ via Eq. (1). As per the legend on plot (d), the colors of the points in the scatterplots indicate into which quadrant the random pick fell. The diagonal line in each plot shows where a point would fall if the forecasted and presumed amplitudes were in perfect agreement. The amount of data N used in the regression model is dictated by 21 min past the first peak in the constructed signal $\boldsymbol{t}$

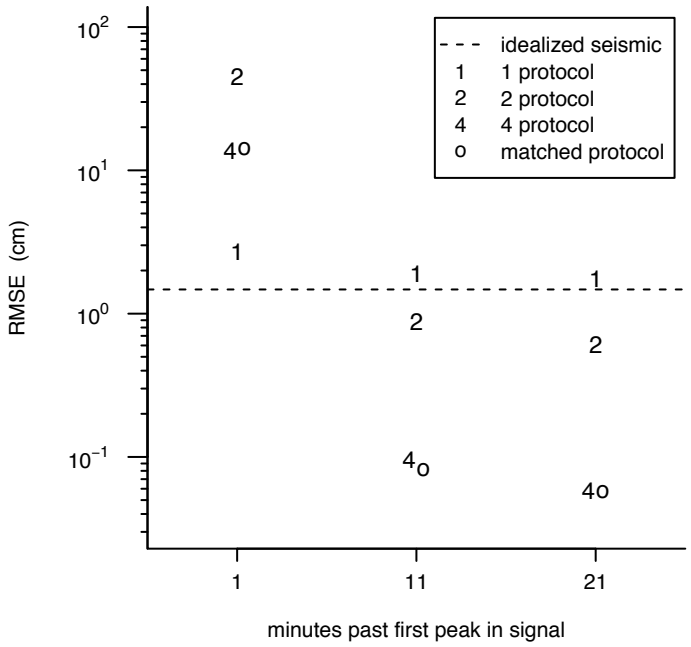

Fig. 18 Root-mean-square errors (RMSEs) for forecasted versus presumed maximum wave amplitudes for 1000 realizations of representative triad (central unit source `ac026b`, buoy 46403 and impact site Hilo). There are twelve RMSEs in all (each computed as per Eq. (6)) corresponding to selecting one of the four protocols (the 1, 2, 4 and matched protocols) in combination with three different amounts of data (dictated by 1, 11 and 21 min past the first peak in each constructed signal $\boldsymbol{t}$) to be placed into $\bar{\boldsymbol{y}}$ in Eqs. (3) and (4). The RMSEs corresponding to the four scatterplots of Fig. 17 are plotted above the '21' label on the horizontal axis. The horizontal dashed line indicates the RMSE for forecasted amplitudes based on idealized seismic information only (and not any simulated data collected by buoy 46403)