Natural Hazards manuscript No.
(will be inserted by the editor)

Extraction of tsunami source parameters via inversion of DART® buoy data

Donald B. Percival · Donald W. Denbo · Marie C. Eblé · Edison Gica · Harold O. Mofjeld · Michael C. Spillane · Liujuan Tang · Vasily V. Titov

Received: date / Accepted: date

Abstract The ability to accurately forecast potential hazards posed to coastal communities by tsunamis generated seismically in both the near and far field requires knowledge of so-called source parameters, from which the strength of a tsunami can be deduced. Seismic information alone can be used to set the source parameters, but the values so derived reflect the dynamics of movement at or below the seabed and hence might not accurately describe how this motion is manifested in the overlaying water column. We describe here a method for refining source parameter estimates based on

Donald B. Percival
Applied Physics Laboratory, Box 355640, University of Washington, Seattle, WA 98195-5640 USA
Tel.: +1-206-5431368
Fax: +1-206-5436785
E-mail: dbp@apl.washington.edu

Donald W. Denbo · Marie C. Eblé · Edison Gica · Harold O. Mofjeld · Michael C. Spillane · Liujuan Tang · Vasily V. Titov
NOAA/Pacific Marine Environmental Laboratory, 7600 Sand Point Way NE, Seattle, WA 98115 USA
Tel.: +1-206-5266824
Fax: +1-206-5266485

Donald W. Denbo
E-mail: Donald.W.Denbo@noaa.gov

Marie C. Eblé
E-mail: Marie.C.Eble@noaa.gov

Edison Gica
E-mail: Edison.Gica@noaa.gov

Harold O. Mofjeld
E-mail: hmofjeld@u.washington.edu

Michael C. Spillane
E-mail: Mick.Spillane@noaa.gov

Liujuan Tang
E-mail: Liujuan.Tang@noaa.gov

Vasily V. Titov
E-mail: Vasily.Titov@noaa.gov

seismic information by making use of data from Deep-ocean Assessment and Reporting of Tsunamis (DART®) buoys. The method involves using this data to adjust precomputed models via an inversion algorithm so that residuals between the adjusted models and the DART® data are as small as possible in a least squares sense. The inversion algorithm is statistically based and hence has the ability to assess uncertainty in the estimated source parameters. We describe this inversion algorithm in detail and apply it to the November 2006 Kurl Islands event as a case study.

Keywords Tsunami forecasting · Tsunami source estimation · DART® data inversion · 2006 Kuril Islands tsunami

1 Introduction

Tsunamis have been recognized as a potential hazard to United States coastal communities since the mid-twentieth century when multiple destructive tsunamis caused damage to the states of Hawaii, Alaska, California, Oregon, and Washington. The National Oceanic and Atmospheric Administration (NOAA) responded to these disasters with the establishment of two Tsunami Warning Centers responsible for providing warnings to the United States and her territories. In addition, the agency assumed the leadership role in the area of tsunami observations and research and has been measuring tsunamis in the deep ocean for many decades. The scale of destruction and unprecedented loss of life following the December 2004 Sumatra tsunami prompted a strengthening of efforts to address the threats posed by tsunamis, and, on 20 December 2006, the United States Congress passed the "Tsunami Warning and Education Act." Central to the goal of protecting United States coastlines is a "tsunami forecasting capability based on models and measurements, including tsunami inundation models and maps . . . ." To meet this congressionally mandated forecasting capability, the NOAA Center for Tsunami Research has developed the Short-term Inundation Forecast for Tsunamis (SIFT) application (Gica et al. 2009; Titov 2009). This application is designed to rapidly and efficiently forecast tsunami heights at specific coastal communities.

At each community, estimates of tsunami wave arrival time and amplitude are provided by combining real-time tsunami event data with numerical models. Several key components are integrated within SIFT: deep-ocean observations of tsunamis collected in real-time, a basin-wide pre-computed propagation database of water level and flow velocities based on potential seismic unit sources, an inversion algorithm to estimate parameters associated with unit sources based upon the deep-ocean observations, and high-resolution tsunami forecast models developed for specific at-risk coastal communities. As a tsunami wave propagates across the open ocean, Deep-ocean Assessment and Reporting of Tsunamis (DART®) buoys observe the passage of the wave and relay data related to its arrival time and amplitude in real or near-real time for use with SIFT. The SIFT application uses the reported observations to refine an initial assessment of the magnitude of the tsunami that is based purely on seismic information. The refinement is done by comparing observations from the DART® buoys to models in the pre-computed propagation database via an inversion algorithm.

In this article we focus on the inversion algorithm that combines data from DART® buoys with precomputed models, yielding refined estimates of the source parameters. We begin with an overview of the SIFT application (Section 2), after which we describe the data collected by the DART® buoys (Section 3). We then discuss the precomputed models for the DART® data in Section 4 in preparation for a detailed description of the inversion algorithm in Section 5, which also includes an illustration of the use of the algorithm on the 15 November 2006 Kuril Islands event. The inversion algorithm is based upon a statistical model, and hence we are able to assess the uncertainty in the resulting source parameter estimates. In Section 6 we use the parameter estimates and their associated uncertainties to compare the strength of the tsunami as assessed by source parameters set using seismic information and by those estimated using the inversion algorithm. We conclude the main body of the paper with a discussion of potential extensions and refinements to the SIFT application in Section 7 and with our conclusions in Section 8 (Appendix A gives some technical details on assessing the uncertainty in the estimated source parameters).

2 Short-term Inundation Forecast for Tsunamis (SIFT)

The SIFT application exploits the fact that the ocean acts as a low-pass filter, allowing long-period phenomenon such as tsunamis to be detected by measurement of pressure at a fixed point on the seafloor (Meinig et al. 2005). The strategy behind SIFT is to assess the potential effect of a tsunami by combining pressure measurements collected in real time with models, thus refining an initial assessment based purely on seismic data available soon after an earthquake. SIFT is an operational system that must provide its assessments in a timely manner. Given that computations concerning wave generation, propagation, and inundation must be done under time constraints, SIFT makes use of a pre-computed propagation database containing water elevations and flow velocities that are generated by standardized earthquakes located within "unit sources," which are strategically placed along ocean basin subduction zones (Gica et al. 2010). Within SIFT, model time series are extracted from the numerical solution to the propagation of tsunami waves throughout the ocean basin as generated at the unit sources. Dynamics of these tsunami waves in the open ocean allow them to be linearly combined to mimic observed data.

An inversion algorithm is used to extract source parameters that adjust the amplitudes of the pre-computed models from each unit source using deep-ocean measurements collected by DART® buoys. These parameters, once determined by the inversion algorithm, provide the boundary conditions under which previously developed inundation models are run to provide forecasts of incoming tsunami waves at threatened coastal communities. These models are run independent of one another in real-time while a tsunami is propagating across the open ocean. The models provide an estimate of wave arrival time, wave height, and inundation following tsunami generation. Each inundation model has been designed and tested against historical events to perform under very stringent time constraints, given that time is generally the single limiting factor in saving lives and property. A total of seventy-five community inundation models are scheduled for completion at the end of federal fiscal year 2012.

3 Bottom pressure measurements from DART® buoys

A DART® buoy actually consists of two separate units, namely, a surface buoy and a bottom unit with a pressure recorder. These units communicate with each other via acoustic telemetry, and the surface buoy in turn communicates with the outside world via transmissions to a satellite. The bottom pressure recorder internally measures water pressure integrated over nonoverlapping 15-sec time windows, so there are $60 \times 4 = 240$ measurements every hour. We associate each window with an integer-valued time index n. For simplicity we adopt the convention that $n = 0$ corresponds to the 15-sec time window during which a particular tsunami-generating earthquake of interest commenced. The actual time associated with the nth time window is $t_n = a + n\,\Delta$ (in hours), where a is a fixed offset, and $\Delta = 1/240 \doteq 0.004167$ hours. In what follows, it is convenient to set $a = 0$ so that t_n is the elapsed time from the 15-sec window containing the earthquake event. We denote the internal measurements by x_n, where $n < 0$ (or $n > 0$) is the index for a measurement recorded before (or after) the earthquake.

The internally recorded x_n measurements only become fully available when the bottom unit is lifted to the surface for servicing (about once every two years). Normally the buoy operates in a monitoring mode in which the bottom unit packages together one measurement every 15 minutes (a 60 fold reduction in data) over a 6-hour block for transmission up to the surface buoy once every 6 hours. We refer to measurements from this monitoring mode as the "15-min stream." Let n_l, $l = -1, -2, \ldots$, represent the indices associated with the portion of the 15-min stream that occurs just prior to the $n = 0$ measurement (the measurement x_0 itself might or might not be available). Typically we have $n_l - n_{l-1} = 60$, but this need not be true for all l due to data drop-outs. Also note that n_{-1} itself need not be a multiple of 60 since the earthquake can occur anywhere within the 15-min reporting cycle.

The bottom unit switches out of monitoring mode into a rapid reporting mode either automatically if a seismic event is detected by a DART® buoy or when forced to do so by an operator at a tsunami warning center sending a signal via satellite to the surface buoy, which then sends an initiating signal to the bottom unit. When in rapid reporting mode, the bottom unit transmits to the surface buoy either a full reporting of the 15-sec data (the "15-sec stream") or a reporting of 1-min averages, i.e., the average of four consecutive x_n values (the "1-min stream"). The index for a 1-min average is the index associated with the most recent 15-sec time window used in forming the average:

$$\bar{x}_{n_l} = \frac{1}{4} \sum_{k=0}^{3} x_{n_l - k}.$$

Let n_l, $l = 0, 1, \ldots$, represent the indices associated with the data that are available after (and possibly including) the $n = 0$ measurement. Ignoring the occurrence of drop-outs, we have $n_l - n_{l-1} = 1$ when dealing with just the 15-sec stream; by contrast, if both $\bar{x}_{n_{l-1}}$ and $\bar{x}_{n_l}$ are from the 1-min stream, then $n_l - n_{l-1} = 4$. Currently the inversion algorithm uses the 1-min stream primarily, but it can make use of additional measurements from the 15-sec or 15-min streams when available and as needed.

Figure 1 shows an example of the 15-min and 1-min data streams as recorded by DART® buoy 21414 before and after the 15 November 2006 Kuril Islands earthquake. Note that there is a gap between the two streams. This gap is due to drop-outs in the 15-min stream, which disappeared temporarily more than an hour before the earthquake and did not reappear again until more than 12 hours later (well after the tsunami had passed by this buoy). Even if portions of the 15-min stream had not been lost, the data available for use with the inversion algorithm during the critical time period following the earthquake might well have been limited to what is shown in the figure due to the fact that the 15-min stream is transmitted in 6-hour blocks once every 6 hours. Thus, assuming that the last value shown in the figure for the 15-min stream was in a

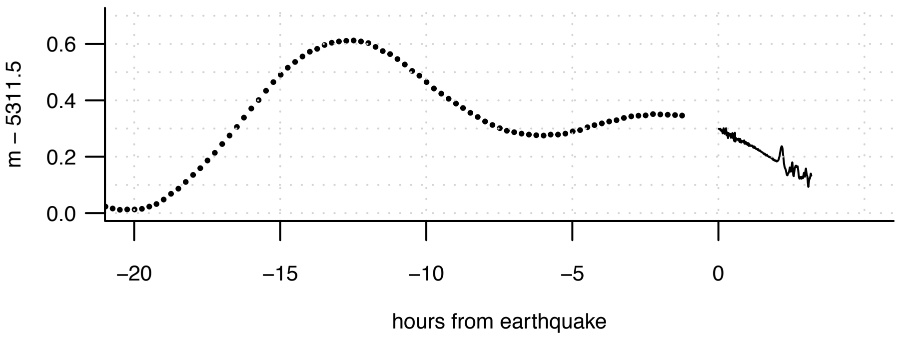

Fig. 1 Data from DART® buoy 21414 recorded before and after 15 November 2006 Kuril Islands earthquake. The black dots show the 15-min stream, while the black curve is the 1-min stream. Time is in hours from the earthquake (a negative/positive time is before/after the earthquake).

6-hour block transmitted soon after it was recorded, the portion of the 15-min stream that would have filled in the gap would not have been scheduled for transmission until almost 5 hours after the earthquake.

The data in Fig. 1 have a prominent tidal component that must be removed prior to use of the inversion algorithm described in Section 5. Detiding must be done nearly in real time and is not a simple matter. We have explored approaches based on harmonic models, Kalman filtering/smoothing, empirical orthogonal functions and low-order polynomial fits (Percival et al. 2010). In what follows, we assume that data x_n from the 15-sec or 15-min streams or data $\bar{x}_n$ from the 1-min stream have been suitably detided. We denote the detided data by d_n and $\bar{d}_n$. (We detided the data using a Kalman filter/smoother in all the examples presented below.)

4 Models for DART® buoy data

The purpose of the inversion algorithm is to use models to estimate the tsunami source strength and associated confidence limits from observed DART® data. Formulation of these models is discussed in detail in Titov et al. (1999) and Gica et al. (2008), from which the following overview is extracted. Seventeen tsunami source regions are defined along portions of the Pacific and Indian Oceans from which earthquake-generated tsunamis are likely to occur (there are also source regions defined for the Atlantic Ocean and Caribbean Sea). Each source region is divided up into a number of “unit sources.” For example, the Aleutian-Alaska-Canada-Cascadia source region consists of 130 unit sources, each of which has an area of 100×50 km^2 (see Fig. 3 below). A database has been constructed containing precomputed adjustable models that predict what would be observed at a given DART® buoy from the beginning of an earthquake event and onwards. This prediction is based under the assumption that the earthquake was located in a particular unit source and was of moment magnitude $M_W = 7.5$ from a reverse thrust of appropriate strike, dip and depth (this corresponds to a coseismic slip of 1 m along the fault in the down-dip direction with a rigidity of 4.0×10^{11} dynes/cm^2; Section 5.2 has more details about the unit sources). The fault movement is assumed to be instantaneous and results in a vertical ground displacement, as computed by the elastic model of Gusiakov (1978) and Okada (1985), that generates the tsunami for the unit source. The database thus has a precomputed model for each pairing of a particular buoy and particular unit source.

Each adjustable model was constructed with a 15-sec time step, but, to save space in the database, was subsampled down to a discrete grid of times with a 1-min spacing. In general, the times used in a precomputed model might or might not correspond to the times at which the DART® buoy data were actually collected relative to the start of the earthquake. To facilitate matching the observed data with an adjustable model, we use cubic splines to interpolate the model. Let $g(t)$ represent the spline-interpolated model at an arbitrary time t for a particular unit source and DART® buoy. The adjustable model value corresponding to a measurement x_{n_l} from that buoy over a 15-sec time window associated with the elapsed time t_{n_l} is just $g(t_{n_l})$. A 1-min average $\bar{x}_{n_l}$ consists of an average of $x_{n_l-3}, x_{n_l-2}, x_{n_l-1}$ and x_{n_l}, so its associated adjustable model is an average of $g(t_{n_l-3}), g(t_{n_l-2}), g(t_{n_l-1})$ and $g(t_{n_l})$.

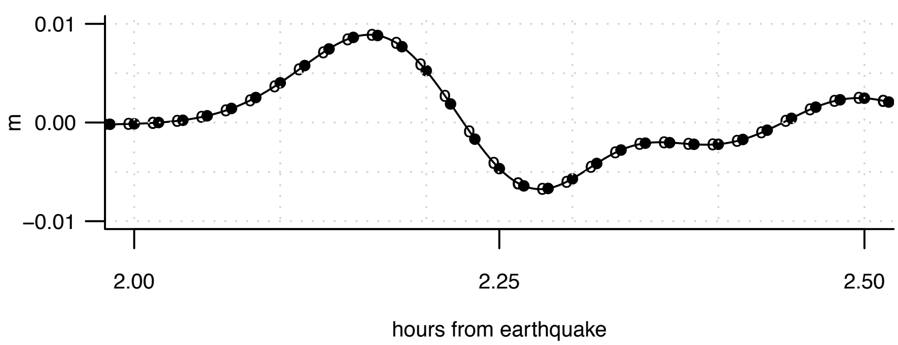

Fig. 2 Adjustable model for DART® buoy 21414 from an earthquake presumed to have originated in unit source a12 in the Kamchatka–Kuril–Japan source region. The black dots indicate the model values stored in the database at 1-min intervals. The black curve is a cubic-spline interpolation of the model outside of the values in the database. The circles show the spline-interpolated values at the times associated with the 1-min stream transmitted from the buoy during the 15 November 2006 Kuril Islands tsunami event. Time is in hours since the earthquake.

Figure 2 shows an example of a spline-interpolated adjustable model $g(t)$ (black curve), which is based upom values precomputed at 1-min intervals and stored in the database (black dots). This model is for DART® buoy 21414 for an earthquake originating from unit source a12, which is in the Kamchatka–Kuril–Japan source region (see Fig. 3 below). During the 15 November 2006 Kuril Islands event, this buoy transmitted a 1-min stream $\bar{x}_{n_l}$ at times t_{n_l}. These times did not coincide exactly with those of the precomputed model. The circles in the plot show the spline-interpolated values $g(t_{n_l})$ versus t_{n_l}, each of which would be the adjustable prediction for a single (unavailable) 15-sec average x_{n_l} (the corresponding prediction for the available $\bar{x}_{n_l}$ would be the average of $g(t_{n_l})$ and three values associated with times occurring 15, 30 and 45 seconds earlier). As the example shows, the cubic spline interpolation provides accurate estimates of the model values at the times of the DART® observations.

5 Inversion algorithm for extracting source parameters

The purpose of the inversion algorithm is to use data collected by DART® buoys (after appropriate detiding) to estimate how large an earthquake-generated tsunami is. As noted in the previous section, the inversion algorithm depends upon a database of precomputed models. There is a model in the database for every pairing of a particular unit source with a particular buoy. This model predicts what would be observed at the buoy if an earthquake with a moment magnitude $M_W = 7.5$ were to originate from a selected unit source. The inversion algorithm adjusts the precomputed model to account for the fact that an actual earthquake rarely has a magnitude exactly equal to 7.5. The adjustment takes the form of a multiplicative factor, which we denote by α and refer to as the source parameter. A value for the parameter α that is greater than unity means that the earthquake has a magnitude greater than 7.5; conversely, a value less than unity indicates an earthquake with a magnitude less than the standard one. The inversion algorithm estimates α by matching the precomputed model and the detided data from the buoy via a least squares procedure. As discussed below, the algorithm takes into account the possibilities that the earthquake might be attributable to more than just a single unit source (so that the adjustments take the form of a vector $\boldsymbol{\alpha}$ of multiplicative factors) and that more than one buoy might have collected data relevant to a particular event. In Section 5.1 we present the inversion algorithm under the simplifying assumptions that we know (1) the unit sources associated with the earthquake and (2) the portions of the detided buoy data that are relevant for assessing the tsunami event (we pay particular attention to assessing the effect of sampling variability on our estimates of $\boldsymbol{\alpha}$). Proper selection of the unit sources and of the relevant data is vital for getting good results from the inversion algorithm. We discuss source selection in Section 5.2 and data selection in 5.3. (For earlier related work on inversion algorithms, see Johnson et al. (1996) and Wei et al. (2003).)

5.1 Estimation of $\boldsymbol{\alpha}$ and assessment of sampling variability

Suppose that we have selected one or more unit sources to explain the tsunami event along with relevant subsets of the detided DART® data. Let $J \geq 1$ represent the number of buoys whose data are to be used in the inversion algorithm, and let $K \geq 1$ be the number of unit sources. Let $\mathbf{d}_j$ be a column vector of length N_j that contains the detided data from the jth buoy, where $j = 1, \ldots, J$ (this can consist of an arbitrary mixture of data from the 15-min, 1-min and 15-sec streams). Let $\mathbf{g}_{j,k}$, $k = 1, \ldots, K$, be a vector of length N_j containing the adjustable model that predicts how the tsunami from a moment magnitude $M_W = 7.5$ earthquake from the kth unit source would be recorded at the jth buoy. The overall model for the data from the jth buoy is taken to be a linear combination of the models associated with the K unit sources; i.e., we write

$$\mathbf{d}_j = \alpha_1 \mathbf{g}_{j,1} + \cdots + \alpha_K \mathbf{g}_{j,K} + \mathbf{e}_j,$$

where α_k is the source parameter for the kth unit source, and $\mathbf{e}_j$ is a vector of N_j error terms that accounts for the mismatch between the idealized overall model and the observed data. We can rewrite the above as

$$\mathbf{d}_j = G_j \boldsymbol{\alpha} + \mathbf{e}_j, \tag{1}$$

where G_j is an $N_j \times K$ matrix whose kth column is $\mathbf{g}_{j,k}$, while $\boldsymbol{\alpha}$ is a column vector of length K containing $\alpha_1, \ldots, \alpha_K$. The models for the data from the individual buoys can be stacked together to form a model for the data from all the buoys, namely,

$$\mathbf{d} = G\boldsymbol{\alpha} + \mathbf{e}, \tag{2}$$

where $\mathbf{d} \equiv \left[\mathbf{d}_1^T, \ldots, \mathbf{d}_J^T\right]^T$ is a column vector of length $N \equiv N_1 + \cdots + N_J$ formed by stacking the individual $\mathbf{d}_j$ on top of one another (here and elsewhere, the superscript "T" denotes the transpose of a vector or a matrix); $G \equiv \left[G_1^T, \ldots, G_J^T\right]^T$ is an $N \times K$ matrix formed in a similar manner by stacking the G_j together; and $\mathbf{e}$ is an N-dimensional column vector of errors whose nth element is e_n.

While the data $\mathbf{d}$ and their models G are known, the K source parameters in $\boldsymbol{\alpha}$ are not, so we need some way of determining them. Typically the amount of data N from all the buoys is much greater than the number of unit sources K. Since there are more equations N in (2) than unknowns K, we must resort to some additional criterion to find an appropriate $\boldsymbol{\alpha}$. One reasonable – and time-honored – criterion is to find the vector such that the sum of squares of the error terms is as small as possible; i.e., we want $\boldsymbol{\alpha}$ to be such that

$$\|\mathbf{e}\|^2 = \|\mathbf{d} - G\boldsymbol{\alpha}\|^2 \text{ is minimized,} \tag{3}$$

where $\|\mathbf{e}\|$ is the Euclidean norm of the vector $\mathbf{e}$:

$$\|\mathbf{e}\|^2 = \sum_{n=1}^{N} e_n^2.$$

This least squares estimator, say $\hat{\boldsymbol{\alpha}}$, is the solution to the so-called normal equations:

$$G^T G\boldsymbol{\alpha} = G^T \mathbf{d}. \tag{4}$$

There are K equations and K unknowns in the above, so we can determine a unique estimator $\hat{\boldsymbol{\alpha}}$ for $\boldsymbol{\alpha}$ as long as $G^T G$ can be inverted. Although $G^T G$ is typically invertible, there is no guarantee that it is such, and numerical problems might prevent a routine that banks upon invertibility from coming up with a stable solution. Because of these considerations, we solve (4) using a singular value decomposition, which, when $G^T G$ is invertible, yields a numerically stable $\hat{\boldsymbol{\alpha}}$ and, when $G^T G$ has a rank lower than K, leads to a solution corresponding to the application of the so-called Moore–Penrose generalized inverse.

A potential complication with the solution to (4) is that the estimated $\hat{\alpha}_k$ in $\hat{\boldsymbol{\alpha}}$ might be a mixture of positive and negative values. This introduces the possibility that prominent random fluctuations in the data that cannot be handled by a model from a single unit source are being matched by a combination of models with $\hat{\alpha}_k$'s that essentially cancel one another out, even though each $|\hat{\alpha}_k|$ might be large. A mixture of positive and negative values for $\hat{\alpha}_k$ is difficult to reconcile with the physics of earthquake generation. To prevent such a mixture, we can alter the least squares criterion such that we seek $\boldsymbol{\alpha}$ such that

$$\|\mathbf{e}\|^2 = \|\mathbf{d} - G\boldsymbol{\alpha}\|^2 \text{ is minimized subject to the constraints } \boldsymbol{\alpha} \geq \mathbf{0}, \tag{5}$$

i.e., $\alpha_k \geq 0$ for $k = 1, \ldots, K$. This minimization problem is a special case of Problem 10.1.1 of Fletcher (1987), and the method we use to solve it is a variation of

Algorithm 10.3.4 in that same reference. Nonnegativity constraints are appropriate for the great majority of tsunamigenic earthquakes in subduction zones, which are the primary source of major trans-ocean tsunamis; however, exceptions do occur, as discussed in Section 7.

The constrained least squares procedure can result in some $\hat{\alpha}_k$ being set to zero, which in effect removes the corresponding unit sources from our model for the data. If we were to entertain a reduced model made up of just the unit sources in our original model for which $\hat{\alpha}_k > 0$, the unconstrained least squares estimate for each unit source in the reduced model will be identical to the corresponding constrained least squares estimate in the original model. Accordingly, if need be, we redefine G by eliminating any unit sources for which $\hat{\alpha}_k = 0$ originally, and redefine K to be the number of remaining unit sources. The end result of the constrained least squares procedure is thus a model that can be fit using unconstrained least squares. The corresponding fitted model is

$$\mathbf{d} = G\hat{\boldsymbol{\alpha}} + \mathbf{r}, \tag{6}$$

where $\hat{\boldsymbol{\alpha}} > 0$, and $\mathbf{r}$ contains the residuals, i.e., the observed errors $\mathbf{r} = \mathbf{d} - G\hat{\boldsymbol{\alpha}}$. Conditional upon the selected model, these residuals can be examined to assess the sampling variability in the estimates $\hat{\boldsymbol{\alpha}}$ using statistical theory, the details of which are given in Appendix A. Since $\boldsymbol{\alpha}$ is of length K, this assessment takes the form of a $K \times K$ covariance matrix Σ for $\hat{\boldsymbol{\alpha}}$. The kth diagonal element of Σ gives us the variance of $\hat{\alpha}_k$, while the (k, l)th off-diagonal element is the covariance between $\hat{\alpha}_k$ and $\hat{\alpha}_l$.

As an example, we consider the Kuril Islands event of 15 November 2006 (see Horrollo et al., 2008, and Kowalik et al., 2008, for additional analyses of this event). Portions of the data received from $J = 11$ buoys during the event are shown (after detiding) as gray circles in the bottom panel of Fig. 3. The locations of the buoys are shown in the upper panel. The displayed data for four of the buoys (21414, 46413, 46408 and 46402) were fit to a model involving $K = 3$ unit sources (denoted as a12, a13 and a14 – the rectangles representing their locations are shaded in dark gray in the insert in the upper panel). The curves in the bottom panel depict the fitted models at all eleven buoys. The fitted models and data are in reasonably good agreement, which demonstrates the efficacy of the procedure in modeling this event over a rather large geographic area. The estimated source parameters and their covariance matrix are

$$\hat{\boldsymbol{\alpha}} \doteq \begin{bmatrix} 5.88 \\ 4.23 \\ 2.29 \end{bmatrix} \quad \text{and} \quad \Sigma \doteq \begin{bmatrix} 0.188 & 0.137 & 0.165 \\ 0.137 & 0.253 & 0.256 \\ 0.165 & 0.256 & 0.597 \end{bmatrix}. \tag{7}$$

The square roots of the diagonal elements of Σ give the standard errors of the corresponding elements of $\hat{\boldsymbol{\alpha}}$. We can form approximate 95% confidence intervals (CIs) for the unknown source parameters $\boldsymbol{\alpha}$ by multipling the standard errors by 1.96 and then adding and subtracting the resulting products from the estimates $\hat{\boldsymbol{\alpha}}$. This procedure yields 95% CIs of $[5.03, 6.73]$ for α_1 (source a12), $[3.25, 5.22]$ for α_2 (a13) and $[0.78, 3.81]$ for α_3 (a14). Note that none of these CIs traps zero. Had the kth such interval done so, we would be unable to reject the null hypothesis that the unknown α_k is equal to zero at the 5% level of significance. Since the CIs indicate that none of the unknown α_k's are likely to be zero, we can deem all three source parameters to be significantly different from zero with level of significance of 0.05.

The results shown in Fig. 3 are based upon using data from the first four DART® buoys to observe the Kuril Islands event. Figure 4 shows the effect on the estimated

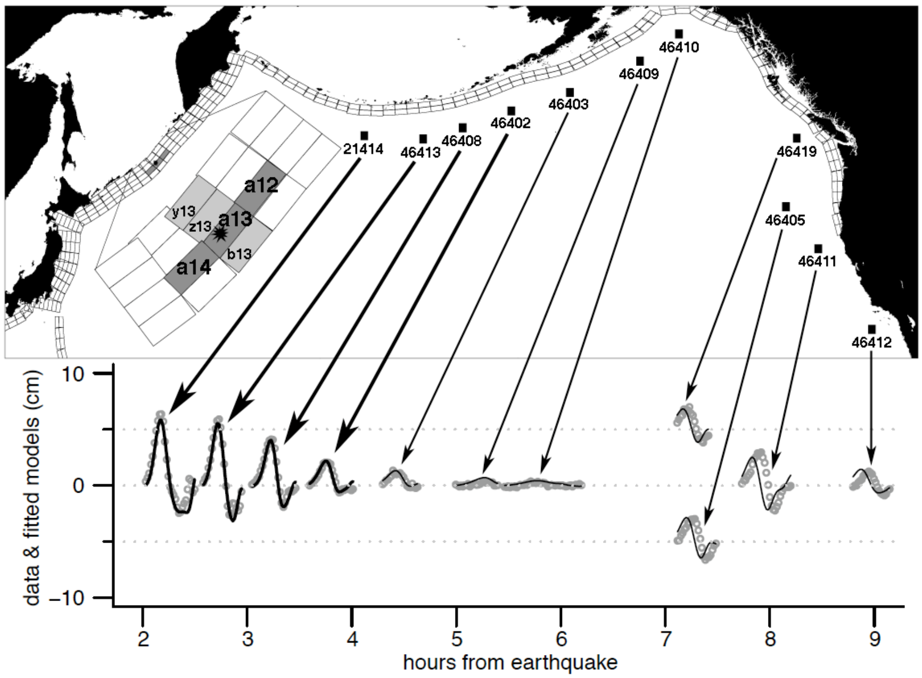

Fig. 3 Locations (upper panel) of six unit sources (a12, a13, a14, b13, z13 and y13) in the Kamchatka–Kuril–Japan (KKJ) source region and eleven DART® buoys (21414, 46413, ..., 46412), along with detided data from these buoys (gray circles, lower panel) and fitted models (solid curves) for the Kuril Islands event of 15 November 2006. The fitted models are based upon unit sources a12, a13 and a14 and data from four buoys (21414, 46413, 46408 and 46402). The unit sources b13, a13, z13 and y13 were selected using seismic information only, while a12, a13 and a14 were selected through a trial-and-error process involving an examination of sums of squares of residuals for various combinations of unit sources. Because the data for buoys 46419 and 46405 arrived at approximately the same time, the data and models for these have been displaced displaced 5 cm upwards (for 46419) and downwards (for 46405). (Although the earthquake emanated from the KKJ source region, ten of the eleven DART® buoys were positioned close to the Aleutian–Alaska–Canada–Cascadia source region. Additional buoys have been deployed since 2006, some now close to the KKJ source region. The reader can go to `http://www.ndbc.noaa.gov/dart.shtml` to see where buoys are currently deployed.)

source parameters caused by using a differing number of buoys. We start by using data from the first buoy to see the tsunami event (21414) and then add in one buoy at a time in the order dictated by the arrival times of the tsunami event. The estimated α_k are fairly stable across time, with the width of the 95% CIs decreasing markedly upon addition of the next two buoys (46413 and 46408) and then gradually after that, up until the addition of the last four buoys (46419, 46405, 46411 and 46412). The fact that the CIs increase upon adding these final buoys can be traced to a misalignment in time between the models and observed data, as is evident in the bottom panel of Fig. 3. For a variety of reasons (including inadequate bottom depth (bathymetry) information, assumed wave dynamics, limited spatial resolution in the model and issues related to finite difference approximations to the equations of motion), any mismatch in propagation time between actual and modeled tsunamis will tend to increase with distance from the unit source. The recent deployment of addition DART® buoys ensures that

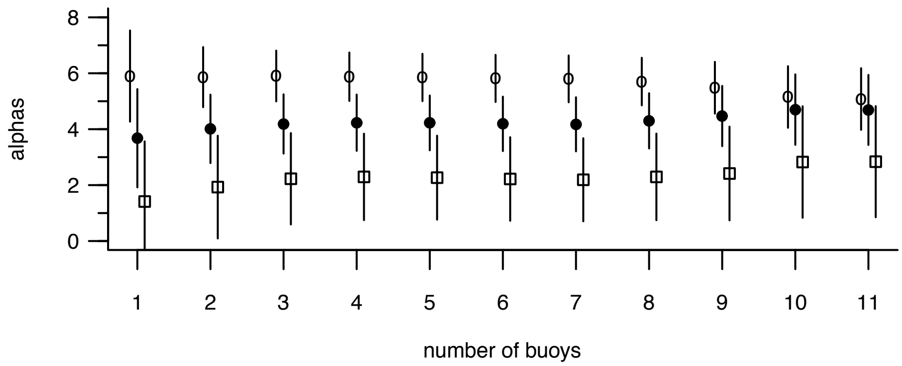

Fig. 4 Estimated source parameters $\hat{\boldsymbol{\alpha}}$ and associated 95% confidence intervals (CIs) versus number of buoys used in the least squares fit. The source parameters $\hat{\alpha}_1$, $\hat{\alpha}_2$ and $\hat{\alpha}_3$ (corresponding to unit sources a12, a13 and a14) are depicted by, respectively, open circles, solid circles and open squares.

most earthquake-generated tsunamis will be observable with near-field buoys, thus reserving far-field buoys more for confirmatory use rather than actual determination of the source parameters during a tsunami event, as we have done in this example. Figure 4 also suggests that use of two or three near-field buoys suffices to get good α_k estimates. Adding more buoys does not lead to a marked improvement in the statistical properties of $\hat{\alpha}_k$, so there is no operational need to wait for additional data to arrive before proceeding with use of the estimated source parameters to drive inundation models for coastal communities.

5.2 Selection of sources

Selection of the unit sources to be used in the inversion algorithm is an iterative process commencing with preliminary estimates of the epicenter and moment magnitude M_W for an earthquake with the potential for generating a tsunami. These estimates are provided by seismic networks and are available shortly after the occurrence of an earthquake and prior to the arrival of any relevant DART® data. We use these estimates to predetermine unit sources and their associated source parameters α_k as follows. First, we select K based upon the size of M_W as indicated by Table 1. Second, we select K unit sources using an algorithm that picks sources close to the epicenter, but in a pattern suggested by studies of past events. Third, we equate expressions for the seismic moment M_0 (which depends upon the unitless M_W) and an inversion-derived tsunami magnitude T_M (which depends upon the coseismic slip S_k, measured in cm):

$$M_0 = 10^{1.5(M_W+10.7)} \text{ and } T_M = \mu LW \sum_{k=1}^{K} S_k, \tag{8}$$

where μ is the earth's rigidity (taken to be 4.0×10^{11} dynes/cm^2); L and W are the length and width of each unit source measured in cm (the unit sources represent an area of 100×50 km^2, so $L = 10^7$ cm and $W = L/2$); and both M_0 and T_M have units

criteria	K
$M_W \leq 7.8$	1
$7.8 < M_W \leq 8.1$	2
$8.1 < M_W \leq 8.3$	4
$8.3 < M_W \leq 8.5$	6
$M_W > 8.5$	8

Table 1 Initial determination of number K of unit sources contributing to an earthquake event based solely on the moment magnitude M_W.

of dynes $\cdot$ cm. If we assume $S_1 = S_2 = \cdots = S_K$ and equate T_M with M_0, we obtain

$$S_k = \frac{M_0}{\mu K L W} \text{ for all } k. \tag{9}$$

Finally, we set the dimensionless α_k by dividing S_k by a reference value S_0:

$$\alpha_k = \frac{S_k}{S_0}. \tag{10}$$

For the unit source dimensions and rigidity chosen above, a reference value of $S_0 =$ 100 cm corresponds to $M_W = 7.5$.

As an example of this procedure, suppose we take the epicenter for the Kuril Islands event to be 46.592° N and 153.266° E and M_W to be 8.3, as is currently listed on a USGS Web page [23]. Table 1 says to set $K = 4$, and the rule $\alpha_k = S_k/S_0$ with $S_0 = 100$ cm in conjunction with Equation (9) yields $\alpha_k \doteq 3.95$ for $k = 1, 2, 3$ and 4. The epicenter of the earthquake is in unit source a13, and the algorithm picks b13, z13 and y13 in addition to a13 as the four unit sources (see Fig. 3, where the epicenter is indicated by an asterisk in the rectangle representing a13, with the rectangles for b13, z13 and y13 being shaded in light gray). The solid curve in Fig. 5(a) shows the resulting model for what would be observed at buoy 21414 (Fig. 3 shows the location of this buoy). This model at time t takes the form

$$3.95 \left(g_{1,1}(t) + g_{1,2}(t) + g_{1,3}(t) + g_{1,4}(t) \right),$$

where $g_{1,k}(t)$ is the spline-interpolated model for the kth unit source and buoy 21414 (indexed as $j = 1$). In principle this model would have been available soon after the earthquake and prior to the arrival of the tsunami at any of the DART® buoys. The actual data recorded at buoy 21414 are indicated by circles and asterisks.

Once sufficient DART® data become available, we can use the inversion algorithm with the initial selection of unit sources to obtain estimates $\hat{\boldsymbol{\alpha}}$ of the source parameters – these estimates are refinements of the initial determination based on seismic information alone. Because of the nonnegativity constraints, it is possible that some, say K', of the source parameters will be set to zero, so that only $K - K'$ unit sources are retained in the model. An examination of the CIs for the remaining coefficients might recommend dropping additional unit sources whose corresponding $\hat{\alpha}_k$'s are not significantly different from zero. The solid curve in Fig. 5(b) is the model that results from using the subset of data from buoy 21414 (indicated by the gray circles in the bottom panel of Fig. 3) to obtain the least squares estimates $\hat{\alpha}_k$. Here $K' = 2$ of the coefficients were set to zero, thus eliminating unit sources z13 and y13 from the model, while retaining b13 and a13; however, the 95% CIs for the α_k's corresponding to b13

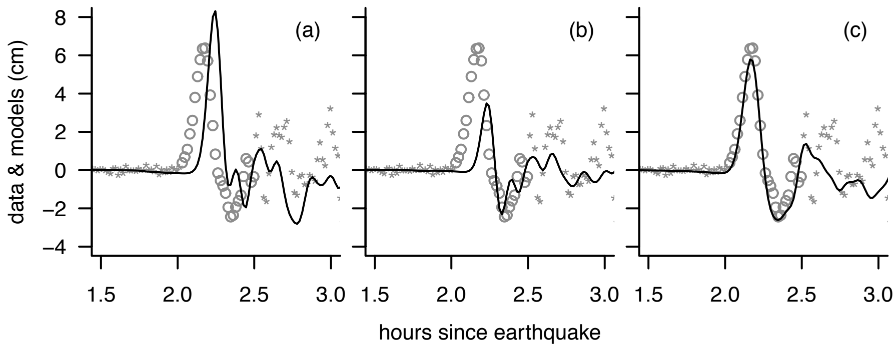

Fig. 5 Three models for 15 November 2006 Kuril Islands tsunami at the location for buoy 21414 (solid curves), along with actual observations (gray circles and asterisks). In plot (a), the source parameters α_k for the model are based on Equations (9) and (10), which require knowledge of the epicenter and the moment magnitude of the earthquake; in plot (b), the same unit sources are used as in (a), namely, b13, a13, z13 and y13, but the α_k's are now determined by a least squares fit with nonnegativity constraints to the portion of the data observed at buoy 21414 indicated by the circles; and in plot (c), a different set of unit sources is used, namely, a12, a13 and a14, with the α_k's set in the same manner as for plot (b). The asterisks in each plot indicate data observed at the buoy, but not used in the least squares fits.

and a13 indicate that neither coefficient is significantly different from zero. The match between the models and the observed data is poor in both Fig. 5(a) and (b). By contrast, Fig. 5(c) shows that we can get a much better match with a different set of unit sources, namely, a12, a13 and a14 (as used in Figs. 3 and 4). This set was selected by trial and error from amongst all the units sources close to the epicenter. The interface for the SIFT application is designed to make it easy for an operator to experiment with various models by facilitating the addition (or removal) of unit sources.

5.3 Selection of data from buoys

The purpose of the inversion algorithm is to estimate the source parameters $\boldsymbol{\alpha}$, which are part of the input needed to forecast the potential dangers of a tsunami to coastal communities. Once we have an appropriate selection of unit sources, the inversion algorithm estimates $\boldsymbol{\alpha}$ based upon whatever selection of DART® buoy data we hand to it. It might seem obvious we would want to use as much data as possible since statistical theory would seem to suggest that, as the amount of data increases, the variability in $\hat{\boldsymbol{\alpha}}$ should tend to decrease, leading to a better estimate of $\boldsymbol{\alpha}$. There are, however, at least two reasons for entertaining smaller amounts of data. First, warnings to coastal communities must be provided in a timely manner – there is no luxury during a tsunami event of waiting for all possible relevant data to arrive from a DART® buoy. Second, empirical evidence suggests that models and data are not equally well matched across time. The quality of the match is time-dependent, suggesting that we focus on particular segments of the data for purposes of fitting the model. Here we illustrate these points by considering the effect of data selection on the estimation of $\boldsymbol{\alpha}$ for the 15 November 2006 Kuril Islands tsunami.

Figure 6(a) shows the detided data (as circles and asterisks) obtained from buoy 21414 during the Kuril Islands tsunami event. The (subjectively determined) beginning

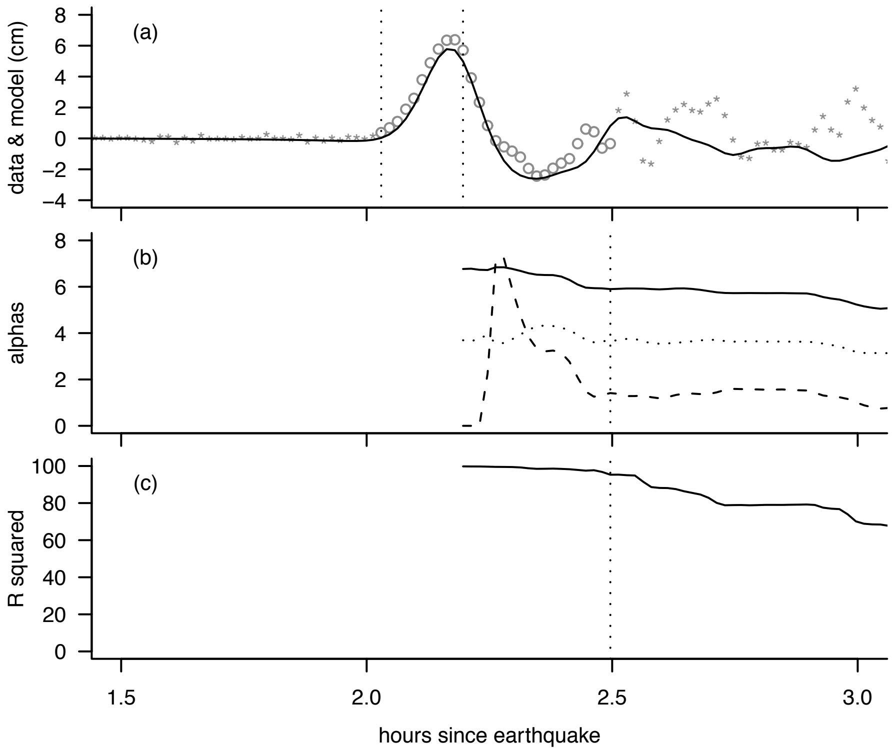

Fig. 6 Effect of estimating $\boldsymbol{\alpha}$ using different segments of data. Plot (a) shows detided data from buoy 21414 observed during the 15 November 2006 Kuril Islands event (circles and asterisks, with the former indicating data comprising the first full wave), along with a fitted model involving unit sources a12, a13 and a14 and the data indicated by the circles. The curves in plot (b) shows the parameter estimates $\hat{\boldsymbol{\alpha}}$ as we increase the amount of data that the estimates are based upon by one data value at a time (the estimates for a12, a13 and a14 are given by, respectively, the solid, dotted and dashed lines). The two vertical dotted lines in (a) indicate the smallest segment of data used to estimate $\boldsymbol{\alpha}$ (eleven data values in all) – these estimates are indicated by the left-most portions of the curves in (b). The curve in plot (c) shows the R^2 statistic, which is the percentage of the sample variance of the data explained by the model. The vertical dotted lines in plots (b) and (c) indicate the $\hat{\boldsymbol{\alpha}}$ and R^2 values associated with the fit involving the first full wave.

of the event as observed at this buoy is indicated by the left-hand vertical dotted line. The data increase monotonically for a while, but then start to decrease. The right-hand line marks the time just following the crest of the first full tsunami wave (i.e., just after the so-called quarter-wave point). The two vertical lines delineate eleven data values. The left-most portions of the curves in Figs. 6(b) and 6(c) show results obtained by using the inversion algorithm with these eleven values to fit a model based upon unit sources a12, a13 and a14, while the remaining parts of the curves show what happens when we increase the amount of data going into the algorithm one value at a time. The solid curve in Fig. 6(b) indicates the estimate $\hat{\alpha}_1$ for a12, whereas the dotted and dashed curves show, respectively, the estimates $\hat{\alpha}_2$ for a13 and $\hat{\alpha}_3$ for a14. The curve in Fig. 6(c) shows the so-called R^2 statistic, which is the percentage of the sample variance of the data explained by the model (this is the squared correlation – expressed as a percentage – between the observed data and the fitted model). As we keep giving the algorithm one more data value to work with, the change in $\hat{\boldsymbol{\alpha}}$ caused by addition of a new value tends to become smaller, indicating that, after a certain point in time, adding more data doesn't drastically change $\hat{\boldsymbol{\alpha}}$. The amount of variance explained by the model is relatively stable at the beginning, after which it starts to decrease markedly. The vertical dotted lines in Figs. 6(b) and 6(c) indicate the values for a fit involving the entire first full wave, as determined subjectively from an examination of the data (the circles in Fig. 6(a) denote this first full wave). Use of data from the first full wave gives us estimates of $\boldsymbol{\alpha}$ that do not differ markedly from those obtained with more data, with an associated R^2 statistic that is close to the maximum value.

This Kuril Islands tsunami event is one in which the very first wave is the largest when observed at buoy 21414. This is because there is an unobstructed path for the tsunami to propagate from the source of the event to this buoy (see Fig. 3). For this case, we are thus better off just using the data up to the first complete wave to estimate $\boldsymbol{\alpha}$ since the data and the model disagree substantially beyond that point; i.e., the explanatory power of the model decreases beyond the first complete wave observed at the buoy. There are other situations in which later waves can be larger, in which case it would be desirable to use a longer stretch of the data for estimating the parameters $\boldsymbol{\alpha}$. The interface for the SIFT application makes it easy for an operator to select the data to be used in the inversion algorithm.

6 Confidence limits for inversion-derived tsunami and moment magnitudes

The primary use for the estimated source parameters $\hat{\boldsymbol{\alpha}}$ is to provide boundary conditions for inundation models for coastal regions (Bernard et al. 2006; Bernard and Titov 2007; Tang et al. 2009). The $\hat{\boldsymbol{\alpha}}$ can also be used for other purposes, one of which is to provide a check that the size of the tsunami event is consistent with the moment magnitude. This check is made by backing out an estimate of the moment magnitude from $\hat{\boldsymbol{\alpha}}$ and comparing it to the seismically determined M_W, per the following procedure.

In view of Equations (8) and (10), a natural estimator of the inversion-derived tsunami magnitude T_M is

$$\widehat{T}_M = \mu L W S_0 \sum_{k=1}^{K} \hat{\alpha}_k.$$

Equating the seismic moment M_0 with this tsunami magnitude in turn leads to an estimator of the moment magnitude M_W based on $\widehat{T}_M$:

$$\widehat{M}_W = \frac{2}{3} \log_{10}\left(\widehat{T}_M\right) - 10.7.$$

Standard least squares theory (Draper and Smith 1998) says that the variance of $\widehat{T}_M$ is given by

$$\operatorname{var}\{\widehat{T}_M\} = \mu^2 L^2 W^2 S_0^2 \mathbf{1}^T \Sigma \mathbf{1},$$

where $\mathbf{1}$ is a vector of length K, all of whose elements are unity, while, as before, Σ is the $K \times K$ covariance matrix for $\hat{\boldsymbol{\alpha}}$. A Taylor series expansion of $\log_{10}(\widehat{T}_M)$ about the true T_M leads to the approximation

$$\operatorname{var}\{\widehat{M}_W\} = \log_{10}^2(e) \frac{4 \operatorname{var}\{\widehat{T}_M\}}{9 T_M^2}.$$

Substitution of $\widehat{T}_M^2$ for T_M^2 in the denominator gives us a means of assessing the sampling variability in $\widehat{T}_M$, which is needed to determine if any observed difference between $\widehat{M}_W$ and M_W is statistically significant, given the sampling variability in $\widehat{M}_W$. Assuming a normal distribution, we can express this variability in terms of an approximate 95% confidence interval:

$$\left[\widehat{M}_W - 1.96\sqrt{\mathrm{var}\{\widehat{M}_W\}}, \widehat{M}_W + 1.96\sqrt{\mathrm{var}\{\widehat{M}_W\}}\right].$$

As an example, let us reconsider the 15 November 2006 Kuril Islands tsunami, for which $M_W = 8.3$. Using $\hat{\boldsymbol{\alpha}}$ and Σ as given in Equation (7), we obtain $\widehat{M}_W \doteq 8.23$, while the square root of $\mathrm{var}\{\widehat{M}_W\}$ is 0.034. An approximate 95% CI is thus $[8.16, 8.3]$, which just barely traps the value $M_W = 8.3$; however, if we also take into account potential rounding error in the latter, there is little evidence that $\widehat{M}_W$ is significantly different from M_W.

While the inversion-derived $\widehat{M}_W$ and the seismically determined M_W are consistent in this example, in other cases $\widehat{M}_W$ could be significantly smaller or larger than M_W. There are at least three explanations for the case $\widehat{M}_W > M_W$. First, the initial estimate of M_W might have been low because it was based on too short a set of seismic waves (e.g., slower – but more energetic – waves arrived after M_W had been determined). Second, a slowly rupturing earthquake can produce less energetic seismic waves for the same vertical ground displacement that generates the tsunami, resulting in a tsunami earthquake, as defined by Kanamori (1972; see Okal 2009). Third, a coseismic landslide can occur that generates an additional tsunami near the earthquake. On the other hand, if $\widehat{M}_W < M_W$, a rare possibility is that M_W was overestimated from a short set of seismic waves, but a more likely explanation is that the earthquake mechanism (e.g., strike-slip) is different from the one assumed in SIFT (a reverse thrust fault event), producing a smaller vertical ground displacement and hence a smaller tsunami. This case is of practical importance because of the potential need to cancel an initial warning that was issued based on the seismically determined M_W.

7 Discussion

While the current version of the SIFT application is fully functional, here we discuss some possible extensions to the software that might impact upcoming versions.

The SIFT application is capable of estimating tsunami source parameters in near real time, but there is a need to provide operators with help in its use during a tsunami event. As discussed above, two critical elements in successful use of the SIFT application are choosing a set of appropriate unit sources and selecting appropriate subsets of DART® buoy data. Currently these choices depend upon experienced operators, but, for operators with limited experience and as potential guidance for experienced operators under time constraints during a tsunami event, it is desirable to look for ways to automate the selection procedures. The problem of selecting unit sources is closely related to the topic of variable selection in linear regression analysis, for which there is a considerable literature that we can draw upon for ideas. Complicating factors are the dynamic nature in which the data arrive, the potential desire to have spatially coherent unit sources, the correlated nature of the errors and the possible interplay between selecting unit sources and subsetting the DART® data. The problem of selecting appropriate subsets of DART® buoy data is related to the topic of isolating transients, for which wavelets and other techniques for extracting a signal from a time series with nonstationary behaviour can be looked to for guidance. How best to automatically select unit sources and to subset the DART® data are subjects of ongoing research.

A complicating factor we have not discussed is contamination of the DART® buoy data from seismic noise. While the November 2006 Kuril Islands tsunami event and many others are relatively free from such noise, there are cases where seismic noise is co-located in time with the tsunami event itself in the DART® data. How best to eliminate this noise is also the subject of ongoing research.

Another complication is that a tsunami can arise from an earthquake whose epicenter falls outside the set of all predefined unit sources. This happened with the 29 September 2009 Samoa event, which – after the event – prompted the addition of new unit sources to the database. In such a case, the current strategy within SIFT is to pick unit sources whose distance from the epicenter is as small as possible. For the Samoa event, it was possible to get good fits to data from individual buoys using this strategy, but not to data from combination of buoys. This occurrence points out the need for a fail-safe option within the SIFT application for an operator to be able to set up new unit sources on the fly. Currently implementation of this option faces substantial technical challenges due to the amount of time needed to compute the models.

As discussed in Section 5.1, there is a need to impose nonnegativity constraints on the estimated source parameters as per Equation (5). These constraints prevent a mixture of positive and negative estimates, which would be difficult to interpret physically. The assumption behind these constraints is a reverse thrust mechanism for the earthquake, which is the most common occurrence for major subduction zones. An earthquake can, however, be caused by a normal, or thrust, mechanism, for which we would then want to entertain nonpositivity constraints; i.e., in contrast to Equation (5), we now seek $\boldsymbol{\alpha}$ such that

$$\|\mathbf{e}\|^2 = \|\mathbf{d} - G\boldsymbol{\alpha}\|^2 \text{ is minimized subject to the constraints } \boldsymbol{\alpha} \leq \mathbf{0}.$$

The SIFT application currently gives the operator the option of imposing nonnegativity, nonpositivity or no constraints, with nonnegativity constraints being the default. It might be possible to provide guidance in selecting between nonnegativity and nonpositivity constraints from an analysis of the initial seismic waves from an earthquake, but the feasibility of doing so needs further research.

The primary use for the source parameters that the inversion algorithm produces is to provide initial conditions for models that forecast inundation in particular coastal regions. Currently the forecasts of wave heights and runup in areas likely to be impacted by a tsunami do not take into account the uncertainty in the source parameter estimates. Research is needed to determine how this uncertainty impacts these forecasts and how best to present this uncertainty to managers in charge of issuing warnings to coastal communities.

Finally we note that the automated system in the current version of the SIFT application is intended to handle events with moment magnitudes M_W at or below 8.5. For larger magnitude earthquakes (the 26 December 2004 Indian Ocean tsunami being a prime example), an operator at a tsunami warning system can manually match a set of sources to the DART® data that extend over large distances along a fault zone. Development is underway to enhance the capability of the automated system to work for these larger events, but there are a number of technical issues to overcome (e.g., the timing of the contribution from the sources needs to be adjusted when the sources are widely spread apart).

8 Summary and Conclusions

The SIFT application is a tool developed at the NOAA Center for Tsunami Research to provide a capability for estimating source parameters during an on-going tsunami event. These source parameters are needed in a timely manner as input to inundation models that can forecast the effect of tsunamis at various coastal communities. While the source parameters can be set initially based solely on seismic information, experience has shown that these initial settings can be improved upon substantially by estimating the parameters based upon DART® buoy data collected during the on-going event. The SIFT application is designed to compute these refined estimates soon after the DART® data become available by making use of a database of precomputed models. These geophysically-based models predict what would be observed at each buoy given a standardized earthquake emanating from a set of unit sources. The refined estimates of the source parameters are computed within SIFT via an inversion algorithm, which relates the data to the geophysically-based models via a linear regression model. With suitable nonnegativity or nonpositivity constraints, this statistical model allows for physically interpretable source parameter estimates, along with an assessment of their sampling variability. The model is formulated in a manner flexible enough to allow for arbitrary combinations of the different types of data reported by the DART® buoys (either pressure measurements integrated over 15-sec time windows or 1-min averages of four such measurements).

We demonstated the efficacy of the inversion algorithm by applying it to data from the 15 November 2006 Kuril Islands event. This example shows that estimates of the source parameters based upon data from a single buoy produces a much better match to the observed DART® buoy data than what is provided by parameters set using just seismic information (see Fig. 5). These refined estimates in principle would have been available no more than 2.5 hours after the occurrence of the earthquake generating the tsunami. Use of data from an additional one to three buoys (available within 3 to 4 hours after the earthquake) produces estimates of the source parameters with sampling variabilities that are not substantially improved upon by using data from distantly located buoys (see Fig. 4). Operationally this finding suggests that there is no need to wait for the tsunami to pass by more than a couple of buoys in the hope of getting better estimates of the source parameters. Models for the data fit using four buoys were able to predict quite well the pattern – but not the exact timing – of the tsunami as it passed by distantly located buoys, demonstrating the ability to model tsunami events on ocean-wide scales based on just four freely adjustable source parameters (see Fig. 3). Finally the fact that we can assess the sampling variability in the estimated source parameters allows us to say whether the strength of the generating event as determined by the inversion algorithm is significantly different statistically from the seismically determined moment magnitude M_W (see Section 6, where we showed that, for the Kuril Islands event, the two ways of assessing the strength gave comparable results when sampling variability was taken into account).

While work is in progress to add more functionality to the SIFT application, it has already proven to be a valuable tool for assessing the potential hazards of tsunamis to coastal communities, in part due to the inversion algorithm that is the focus of this article. Pending successful completion of the ongoing research described in Section 7, future versions of the SIFT application will have features that should make it easier for operators to specify the input required for successful use of the inversion algorithm.

Appendix A

Here we assess the statistical properties of the least squares estimator for $\boldsymbol{\alpha}$. To do so, let us reconsider the model (1) for the data from a single buoy. The vector $\mathbf{d}_j$ potentially consists of a mixture of 15-sec and 1-min measurements. To deal with this possibility, we can create a vector $\tilde{\mathbf{d}}_j$ by conceptually replacing each 1-min measurement with the four unobserved 15-sec values that were averaged to form it. For example, if $\mathbf{d}_j = [x_{15}, \bar{x}_{20}, \bar{x}_{24}]^T$, then $\tilde{\mathbf{d}}_j = [x_{15}, x_{17}, x_{18}, x_{19}, x_{20}, x_{21}, x_{22}, x_{23}, x_{24}]^T$, and $\mathbf{d}_j$ and $\tilde{\mathbf{d}}_j$ are related by $\mathbf{d}_j = \mathit{\Gamma}_j \tilde{\mathbf{d}}_j$, where

$$\mathit{\Gamma}_j = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{bmatrix}.$$

Let $\widetilde{N}_j$ be the length of the vector $\tilde{\mathbf{d}}_j$. In a similar manner, augment G_j to obtain $\widetilde{G}_j$, and consider the model

$$\tilde{\mathbf{d}}_j = \widetilde{G}_j \boldsymbol{\alpha} + \tilde{\mathbf{e}}_j,$$

where $\tilde{\mathbf{e}}_j$ is a vector of random variables (RVs) that obeys a multivariate normal (MVN) distribution with zero mean and a covariance matrix given by, say, V_j. Under the assumption that the variance for all the error terms is the same, say, σ_j^2, we can express V_j as $\sigma_j^2 \mathit{\Phi}_j$, where $\mathit{\Phi}_j$ is an $\widetilde{N}_j \times \widetilde{N}_j$ matrix whose diagonal elements are all unity. Then we have

$$\mathbf{d}_j = \mathit{\Gamma}_j \tilde{\mathbf{d}}_j = \mathit{\Gamma}_j \widetilde{G}_j \boldsymbol{\alpha} + \mathit{\Gamma}_j \tilde{\mathbf{e}}_j = G_j \boldsymbol{\alpha} + \mathbf{e}_j$$

where $\mathbf{e}_j = \mathit{\Gamma}_j \tilde{\mathbf{e}}_j$ is MVN with zero mean and a covariance matrix given by $\mathit{\Gamma}_j \mathit{\Phi}_j \mathit{\Gamma}_j^T \sigma_j^2$ (cf. Equation (1)). Standard least squares theory (Draper and Smith 1998) says that the least squares estimator for $\boldsymbol{\alpha}$ in the model above has a covariance matrix given by

$$\mathit{\Sigma}_j = (G_j^T G_j)^{-1} G_j^T \mathit{\Gamma}_j \mathit{\Phi}_j \mathit{\Gamma}_j^T G_j (G_j^T G_j)^{-1} \sigma_j^2.$$

By stacking together the models for $\tilde{\mathbf{d}}_j$ for $j = 1, \ldots, J$, we obtain

$$\tilde{\mathbf{d}} = \widetilde{G} \boldsymbol{\alpha} + \tilde{\mathbf{e}}.$$

Here $\tilde{\mathbf{d}}$ is related to $\mathbf{d}$ of Equation (2) via $\mathbf{d} = \mathit{\Gamma} \tilde{\mathbf{d}}$, where

$$\mathit{\Gamma} = \begin{bmatrix} \mathit{\Gamma}_1 & 0 & \cdots & 0 \\ 0 & \mathit{\Gamma}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathit{\Gamma}_J \end{bmatrix}$$

(here and elsewhere, the zeros need to be interpreted as matrices of appropriate dimensions, all of whose elements are zero). Pending more research, we make the simplifying assumption that the errors associated with two different buoys are uncorrelated so that the covariance matrix for $\tilde{\mathbf{e}}$ is given by

$$V = \begin{bmatrix} \sigma_1^2 \Phi_1 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 \Phi_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_J^2 \Phi_J \end{bmatrix}. \tag{11}$$

Then we have

$$\mathbf{d} = \Gamma \tilde{\mathbf{d}} = \Gamma \widetilde{G} \boldsymbol{\alpha} + \Gamma \tilde{\mathbf{e}} = G \boldsymbol{\alpha} + \mathbf{e},$$

where $\mathbf{e} = \Gamma \tilde{\mathbf{e}}$ is MVN with zero mean and a covariance matrix given by $\Gamma V \Gamma^T$ (cf. Equation (2)). The least squares estimator $\hat{\boldsymbol{\alpha}}$ for $\boldsymbol{\alpha}$ in the model above has a covariance matrix given by

$$\Sigma = (G^T G)^{-1} G^T \Gamma V \Gamma^T G (G^T G)^{-1}.$$

Since G and Γ are known, we need only specify V to be able to compute the desired Σ; i.e., we need to set $V_j = \sigma_j^2 \Phi_j$ for each j. To do so, we regard the RVs in $\tilde{\mathbf{e}}_j$ as being extracted from a stationary first-order autoregressive AR(1) process with variance σ_j^2 and AR(1) parameter ϕ_j satisfying $|\phi_j| < 1$. If we take the indices associated with $\tilde{\mathbf{d}}_j$ to be $\tilde{n}_1, \tilde{n}_2, \ldots, \tilde{n}_{\widetilde{N}_j}$, then the (p,q)th element of Φ_j is $\phi_j^{|\tilde{n}_p - \tilde{n}_q|}$. In practice we can estimate σ_j^2 and ϕ_j based upon the residuals $\mathbf{r}_j = \mathbf{d}_j - G_j \hat{\boldsymbol{\alpha}}$ from the least squares fit. One approach for doing so is to make the simplifying assumption that $\mathbf{r}_j$ obeys the same multivariate normal distribution as $\Gamma \tilde{\mathbf{e}}_j$ and to estimate σ_j^2 and ϕ_j using the maximum likelihood (ML) method. The likelihood function for a given σ_j^2 and ϕ_j can be evaluated using a state-space formulation (Jones 1980; Durbin and Koopman 2001), in which the state equation is dictated by an AR(1) process, while the observation equation handles the underlying AR(1) process and what is observable from a mixture of 15-sec and 1-min measurements. The ML estimators are obtained by embedding the evalution of the likelihood function in a nonlinear optimatization proceduce.

The above approach for determining the covariance matrix for the least squares estimator $\hat{\boldsymbol{\alpha}}$ works for an estimator based on an arbitrary mixture of 15-sec and 1-min measurements, but it can be simplified considerably if we assume that each N_j-dimensional vector $\mathbf{d}_j$ consists of stretch of 1-min averages with no missing values, as is the case occurring most often in practice (Percival et al. 2009). With this additional assumption, we can dispense with the Γ_j matrices and formulate a statistical model directly in terms of the model $\mathbf{d} = G\boldsymbol{\alpha} + \mathbf{e}$. The least squares estimator $\hat{\boldsymbol{\alpha}}$ for $\boldsymbol{\alpha}$ now has a covariance matrix given by

$$\Sigma = (G^T G)^{-1} G^T V G (G^T G)^{-1},$$

where V has a structure analogous to Equation (11), but with the (p,q)th element of the $N_j \times N_j$ matrix Φ_j being given by $\phi_j^{|p-q|}$. We can then estimate ϕ_j via

$$\hat{\phi}_j = \frac{\mathbf{r}_{j,(f)}^T \mathbf{r}_{j,(l)}}{\mathbf{r}_j^T \mathbf{r}_j},$$

where $\mathbf{r}_{j,(f)}$ consists of all of $\mathbf{r}_j$ except for its first element, and $\mathbf{r}_{j,(l)}$ has everything but the last element. An approximately unbiased estimator of σ_j^2 is given by

$$\hat{\sigma}_j^2 = \frac{N_j(1-\hat{\phi}_j)^2\mathbf{r}_j^T\mathbf{r}_j/(N_j-1)}{N_j(1-\hat{\phi}_j)^2 - 1 + \phi_j^2 + 2\phi_j(1-\phi_j^{N_j})N_j^{-1}}.$$

We use this simplified approach to obtain the statistical properties of $\hat{\boldsymbol{\alpha}}$ reported in this paper.

Acknowledgements This work was funded by the Joint Institute for the Study of the Atmosphere and Ocean (JISAO) under NOAA Cooperative Agreement No. NA17RJ1232 and is JISAO Contribution No. 1783. This work is also Contribution No. 3557 from NOAA/Pacific Marine Environmental Laboratory. The authors would like to thank Bradley Bell for computer code and discussions on constrained least squares.

References

1. Bernard EN, Mofjeld HO, Titov VV, Synolakis CE, González FI (2006) Tsunami: Scientific frontiers, mitigation, forecasting, and policy implications. Proc Roy Soc Lond A 364(1845): 1989–2007. doi:10.1098/rsta.2006.1809
2. Bernard EN, Titov VV (2007) Improving tsunami forecast skill using deep ocean observations. Mar Technol Soc J 40(4): 23–26
3. Draper NR, Smith H (1998) Applied regression analysis (3rd ed.). John Wiley & Sons, New York
4. Durbin J, Koopman SJ (2001) Time series analysis by state space methods. Oxford University Press, Oxford
5. Fletcher R (1987) Practical methods of optimization (2nd ed.). John Wiley & Sons, Chichester & New York
6. Gica E, Spillane MC, Titov VV, Chamberlin CD, Newman JC (2008) Development of the forecast propagation database for NOAA's Short-term Inundation Forecast for Tsunamis (SIFT). NOAA Technical Memorandum OAR PMEL–139, 89 p. http://nctr.pmel.noaa.gov/pubs.html
7. Gica E, Spillane MC, Arcas D, Percival DB, Titov VV (2010) Development of NOAA's Short-term Inundation Forecast for Tsunamis (SIFT) and a case study of the 15 November 2006 Kuril tsunami. To be submitted.
8. Gusiakov VK (1978) Static displacement on the surface of an elastic space. In Ill-posed problems of mathematical physics and interpretation of geophysical data, publication of the Computing Center of Soviet Academy of Sciences (VC SOAN SSSR), Novosibirsk, Russia (Soviet Union), 23–51 (in Russian)
9. Horrillo J, Knight W, Kowalik Z (2008) Kuril Islands tsunami of November 2006: 2. Impact at Crescent City by local enhancement. J Geophys Res 113:C01021. doi:10.1029/2007JC004404
10. Johnson JM, Satake K, Holdahl SR, Sauber J (1996) The 1964 Prince William earthqake: Joint inversion of tsunami and geodetic data. J Geophys Res 101(B1): 523–532
11. Jones RH (1980) Maximum likelihood fitting of ARMA models to time series with missing observations. Technometrics 22(3): 389–395
12. Kanamori H (1972) Mechanism of tsunami earthquakes. Phys Earth Planet Inter 6: 346–359
13. Kowalik Z, Horrillo J, Knight W, Logan T (2008). Kuril Islands tsunami of November 2006: 1. Impact at Crescent City by distant scattering. J Geophys Res 113:C01020. doi:10.1029/2007JC004402
14. Meinig C, Stalin SE, Nakamura AI, González F, Milburn HB (2005) Technology developments in real-time tsunami measuring, monitoring and forecasting. In Oceans 2005 MTS/IEEE, 19–23 September 2005, Washington, DC
15. Okada Y (1985) Surface deformation due to shear and tensile faults in a half-space. Bull Seismol Soc Am 75: 1135–1154

16. Okal E (2009) Excitation of tsunamis by earthquakes. In: Bernard EN, Robinson AR (ed) The sea (vol 15). Harvard University Press, Cambridge, pp 137–177
17. Percival DB, Arcas D, Denbo DW, Elbé MC, Gica E, Mofjeld HO, Spillane MC, Tang L, Titov VV (2009) Estimating tsunami source parameters via inversion of DART® buoy data. NOAA Technical Memorandum OAR PMEL–144, 22 p. http://nctr.pmel.noaa.gov/pubs.html
18. Percival, D.B., Denbo, D.W., Elbé, M.C., Gica, E., Mofjeld, H.O., Spillane, M.C., Tolkova, E.I. and Titov, V.V. (2010). Near real-time methods for detiding DART® buoy data, to be submitted
19. Tang L, Titov VV, Chamberlin CD (2009) Development, testing, and applications of site-specific tsunami inundation models for real-time forecasting. J Geophys Res 6. doi:10.1029/2009JC005476
20. Titov VV (2009) Tsunami forecasting. In: Bernard EN, Robinson AR (ed) The sea (vol 15). Harvard University Press, Cambridge, pp 371–400
21. Titov VV, Mofjeld HO, González FI, Newman JC (1999) Offshore forecasting of Hawaiian tsunamis generated in Alaska–Aleutian Subduction Zone. NOAA Technical Memorandum ERL PMEL–114, 22 p. http://nctr.pmel.noaa.gov/pubs.html
22. Wei Y, Cheung KF, Curtis GD, McCreery CS (2003) Inverse algorithm for tsunami forecasts. J Waterw Port Coast Ocean Eng 129(3): 60–69
23. http://earthquake.usgs.gov/earthquakes/eqarchives/significant/sig_2006.php