NOAA Technical Memorandum OAR PMEL-144

EXTRACTING TSUNAMI SOURCE PARAMETERS VIA INVERSION OF DART® BUOY DATA

Donald D. Percival[1]
Diego Arcas[2,3]
Donald W. Denbo[2,3]
Marie C. Eble[3]
Edison Gica[2,3]
Harold O. Mofjeld[3,4]
Mick C. Spillane[2,3]
Liujuan Tang[2,3]
Vasily V. Titov[2,3]

[1]Applied Physics Laboratory
University of Washington, Seattle, WA

[2]Joint Institute for the Study of the Atmosphere and Ocean (JISAO)
University of Washington, Seattle, WA

[3]NOAA/Pacific Marine Environmental Laboratory
Seattle, WA

[4]Retired from NOAA

Pacific Marine Environmental Laboratory
Seattle, WA
February 2009

UNITED STATES DEPARTMENT OF COMMERCE

Otto J. Wolf
Acting Secretary

NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION

Mary M. Glackin
Acting Undersecretary for Oceans and Atmosphere/Administrator

Office of Oceanic and Atmospheric Research

Richard W. Spinrad
Assistant Administrator

NOTICE from NOAA

Mention of a commercial company or product does not constitute an endorsement by NOAA/OAR. Use of information from this publication concerning proprietary products or the tests of such products for publicity or advertising purposes is not authorized. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Oceanic and Atmospheric Administration.

Contribution No. 3134 from NOAA/Pacific Marine Environmental Laboratory

Also available from the National Technical Information Service (NTIS)
(`http://www.ntis.gov`)

Contents

Extracting tsunami source parameters via inversion of DART® buoy data

D.B. Percival,[1], D. Arcas[2,3], D.W. Denbo[2,3], M.C. Eble[3], E. Gica[2,3], H.O. Mofjeld[2,3,4], M.C. Spillane[2,3], L. Tang[2,3], and V.V Titov[2,3]

Abstract. The ability to accurately forecast the hazards to coastal communities from earthquake-generated tsunamis requires the knowledge of certain tsunami source parameters. The Short-term Inundation Forecast for Tsunamis (SIFT) tool developed by the National Oceanic and Atmospheric Administration estimates these source parameters based upon a combination of observed data from DART® buoys and a set of precomputed models. We give a detailed description of the procedures within SIFT Version 2.0 for estimating the source parameters and for assessing the sampling variability in the estimated parameters. We illustrate these procedures using detided 1-min data collected by two DART® buoys (21414 and 46413) in the northwestern Pacific during the 15 November 2006 Kuril Islands tsunami event, generated by an offshore moment magnitude 8.3 earthquake. We discuss the advantages of using a least squares (L_2) method for fitting model tsunami time series to the data. The model series originate from two model source segments (50 km by 100 km) near the earthquake epicenter. The fitting procedure estimates the source parameters of each segment, both individually and together. Realistic confidence limits on the source parameters are obtained by fitting the residual series (from the data-fitted model) to a first-order autoregressive model to account for the obvious correlation between adjacent residual values. Using data from both buoys improves the source parameter estimates, at an operational cost of some delay in time for the tsunami to reach the farther station. An appropriate fit uses the first full tsunami wave at each station with no improvement found by extending the fitted data segment further. Formulae are derived for estimating the source parameters and computing the confidence limits for multiple DART® buoys and source segments.

1. Introduction

The National Oceanic and Atmospheric Administration (NOAA) is charged with primary responsibility for providing tsunami warnings to the United States and has assumed a leadership role both in research on tsunamis and in gathering data related to tsunamis (see `http://www.tsunami.noaa.gov/` for details on NOAA's tsunami program). The ability to accurately forecast potential dangers to coastal communities due to an earthquake-generated tsunami requires knowledge of tsunami source parameters. The Short-term Inundation Forecast for Tsunamis (SIFT) tool that has been developed at the NOAA Center for Tsunami Research extracts estimates of these parameters based upon two key components. The first is observed data from DART® buoys collected during a tsunami event, and the second is a set of precomputed models for each DART® location. The model series are extracted from the numerical solution to the propagation of tsunami waves throughout the ocean basin from "unit sources" and, because of the dynam-

[1]Applied Physics Laboratory, University of Washington, Box 355640, Seattle, WA 98195-5640

[2]Joint Institute for the Study of the Atmosphere and Ocean (JISAO), University of Washington, Seattle, WA 98195

[3]NOAA/Pacific Marine Environmental Laboratory, Seattle, WA 98115

[4]Retired from NOAA

ics of these waves in the open ocean, can be linearly combined to mimic the observed data. An inversion procedure combines these key components by using the DART® buoy data to extract parameters that adjust the amplitudes of the precomputed models. These parameters, once determined and applied to the basin-wide set of "unit source" solutions, provide the boundary conditions for inundation models that generate high-resolution forecasts of incoming tsunami waves at threatened coastal communities (Gica *et al.*, 2009). The intent of this memorandum is to describe the key steps in inversion procedures in general and to explain the rationale behind the particular procedure implemented in SIFT, which is being installed at the U.S. Tsunami Warning Centers.

The remainder of the memorandum is organized as follows. To motivate our discussion, we use data recorded by two DART® buoys in November 2006 that include a tsunami event originating from an earthquake near the Kuril Islands (see Horrillo *et al.* (2008); Kowalik *et al.* (2008) for details about this event). We describe the data in Section 2, after which we consider precomputed models for the observed data (Section 3). For simplicity, we first focus on a single buoy and a single model and discuss several criteria for fitting the model to the data (Section 4). Use of a least squares criterion has several advantages, not the least of which is the ability to assess the sampling variability in the estimated parameters (Section 5). We then consider the important question of how much data should be used to carry out the inversion (Section 6), after which we expand our methodology to allow for more than one buoy and one model (Section 7). We make some concluding remarks in Section 8.

2. DART® Buoy Data for 2006 Kuril Islands Tsunami Event

The map in Fig. 1 shows the location of the source of the Kuril Islands tsunami event of 15 November 2006, along with the locations of two DART® buoys that recorded the event, namely, 21414 and 46413 (referred to hereafter as buoys 1 and 2). Figure 2 shows plots of the data taken during the event by these two buoys. The data in these plots have been subjected to a detiding procedure. The time associated with each detided data point is in hours since the beginning of the event, which was at 11:14:16 UTC. Note that the event was observed at buoy 2 about half an hour after it occurred at buoy 1, which is consistent with their locations relative to the Kuril Islands shown in Fig. 1. The spacing in time between adjacent values is 1 min, and each data point represents a 1-min average. There are 105 data points for buoy 1 and 114 for buoy 2. In Sections 3 to 6 we concentrate on the data and models for buoy 1 alone, after which we consider both buoys. In what follows, let $\widetilde{\mathcal{T}}$ be the set of times associated with the data from buoy 1, and, for a given $t \in \widetilde{\mathcal{T}}$, let x_t represent the corresponding detided data.

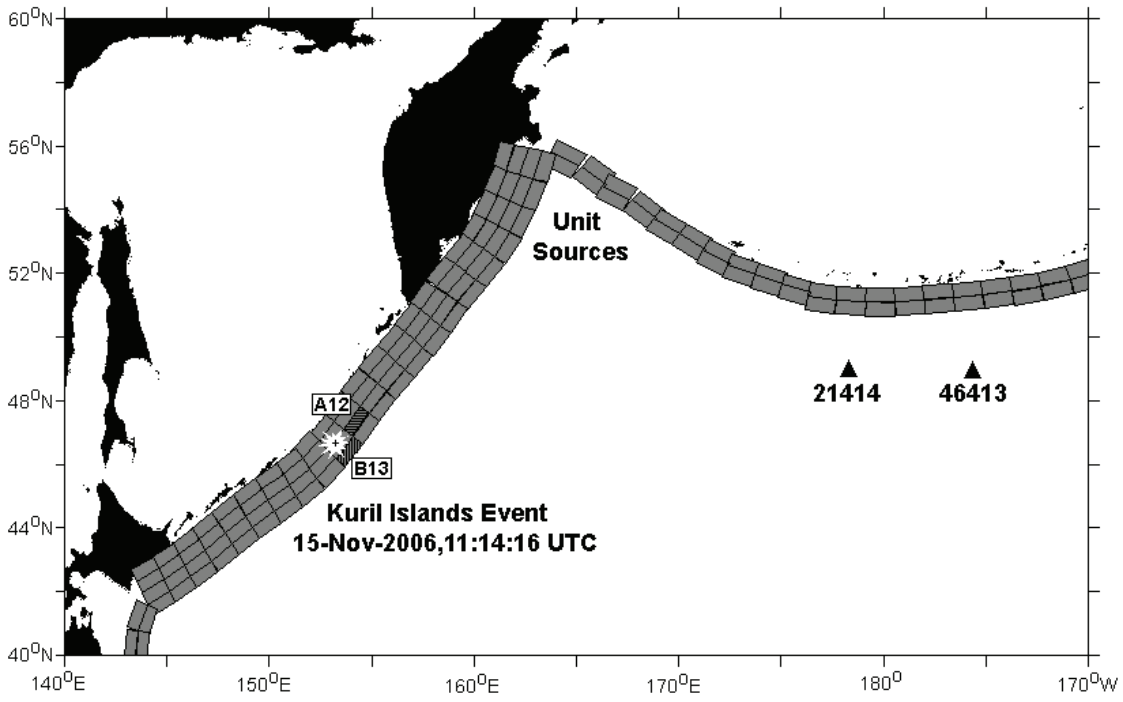

Figure 1: Map showing the approximate origin of the 15 November 2006 Kuril Islands tsunami event, which was generated by a moment magnitude 8.3 earthquake at 11:14:16 UTC. The locations of DART® buoys 21414 and 46413 (henceforth referred to as buoys 1 and 2) are also shown, along with a grid of 50 km by 100 km "unit sources." Two of these unit sources are marked as A12 and B13 (henceforth referred to as sources a and b) and are close to the epicenter of the earthquake (denoted by a star on the map). For both unit sources, models of what would be observed at the two buoys were precomputed under the assumption that the generating event is equivalent to a magnitude 7.5 reverse thrust earthquake located within the unit source.

3. Models for DART® Buoy Data

Here we consider models for the detided DART® data from buoy 1 (see Gica *et al.*, 2008, for details about how these models were formulated). The earthquake that generated the Kuril Islands tsunami event is presumed to have originated from one or more unit sources, which are defined by nominal 50 km by 100 km rectangles located along a fault line. We will consider initially just one of these unit sources marked on Fig. 1 as "A12" and henceforth referred to as unit source a. The circles in Fig. 3 show a precomputed model for what we should expect to observe at buoy 1 under the assumption that there was a magnitude 7.5 reverse thrust earthquake of appropriate strike, dip, and depth originating from unit source a. The model was run with a 15-sec time step, but, to save space on the operational database, was subsampled down

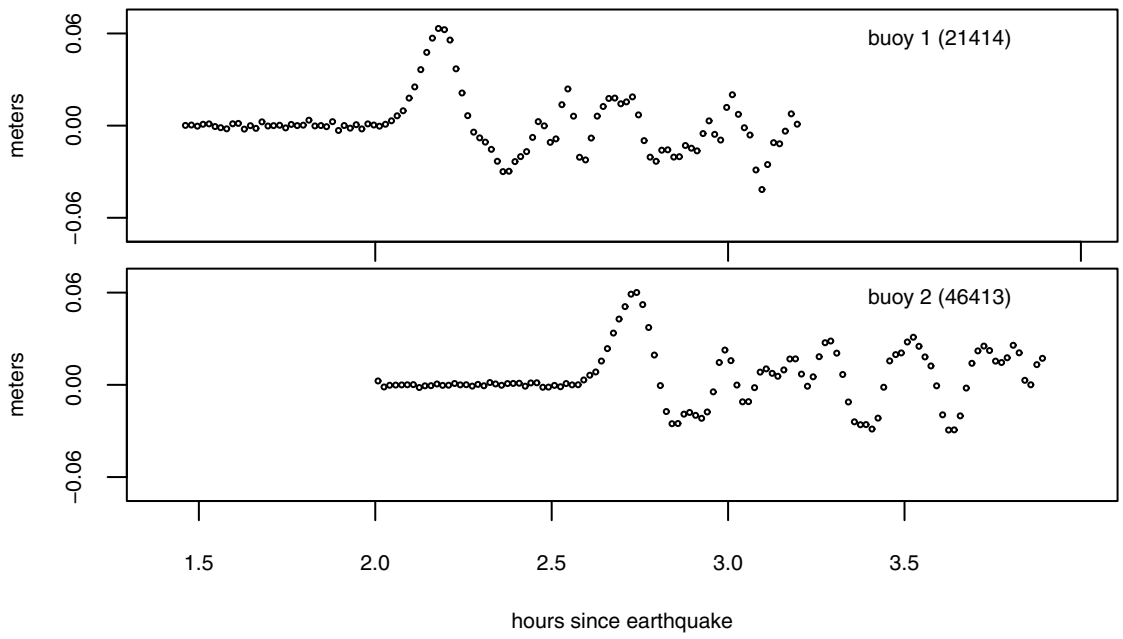

Figure 2: Detided data from 15 November 2006 Kuril Islands tsunami event as derived from data recorded by two DART® buoys.

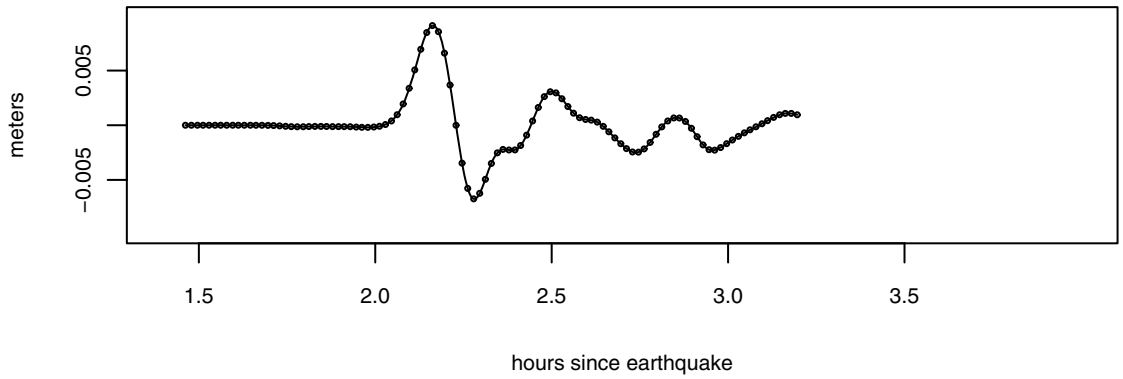

Figure 3: Model for data observed at buoy 1 generated assuming a magnitude 7.5 reverse thrust earthquake originating from unit source a (i.e., A12).

to a discrete grid of times with a spacing of 1 min. 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. To facilitate matching up the model and the observed data, we use a cubic spline to interpolate the model, the result of which is shown by the solid curve passing through the circles in Fig. 3. We let $g(t)$ represent the spline-interpolated model in what follows.

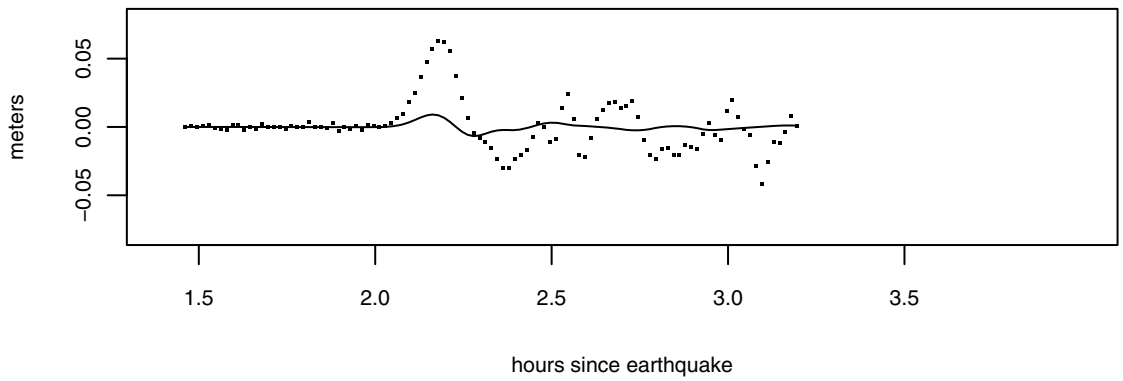

Figure 4: Comparison of detided data from buoy 1 (dots) and its model from unit source a (solid curve, as also shown in Fig. 3). In the inversion process the model will be rescaled to match the observed amplitude of the detided data.

4. Fitting a Single Model to Data from a Single Buoy

Figure 4 compares the precomputed model $g(t)$ from unit source a (the solid curve) and the detided data x_t for buoy 1 (the dots). The model and the data look different because the model was computed for a magnitude 7.5 reverse thrust earthquake. We assume that an earthquake with an arbitrary magnitude can be handled by multiplying $g(t)$ by a parameter α. Accordingly we entertain the model

$$x_t = \alpha g(t) + e_t, \qquad t \in \widetilde{\mathcal{T}},$$

where e_t is a residual term that represents any remaining mismatch between the data and the amplitude-adjusted model.

What criterion should we use to set the parameter α? For any proposed α, we can compute a corresponding set of residuals $e_t = x_t - \alpha g(t)$ for $t \in \widetilde{\mathcal{T}}$. As examples, let's compute residuals corresponding to the choices $\alpha = 2$, 4. and 8. The adjusted models $\alpha g(t)$ and data are shown in the left-hand column of Fig. 5, while the corresponding residuals e_t are displayed in the right-hand column (the residual with the largest absolute value is marked with a circle). Arguably the "best" choice for α is such that the residuals e_t are "small" by some measure. There are many possible measures, but three common ones are to

1. make the sum of the squared residuals as small as possible:

$$\sum_{t\in\widetilde{\mathcal{T}}} e_t^2 = \sum_{t\in\widetilde{\mathcal{T}}} [x_t - \alpha g(t)]^2 \equiv f_2(\alpha);$$

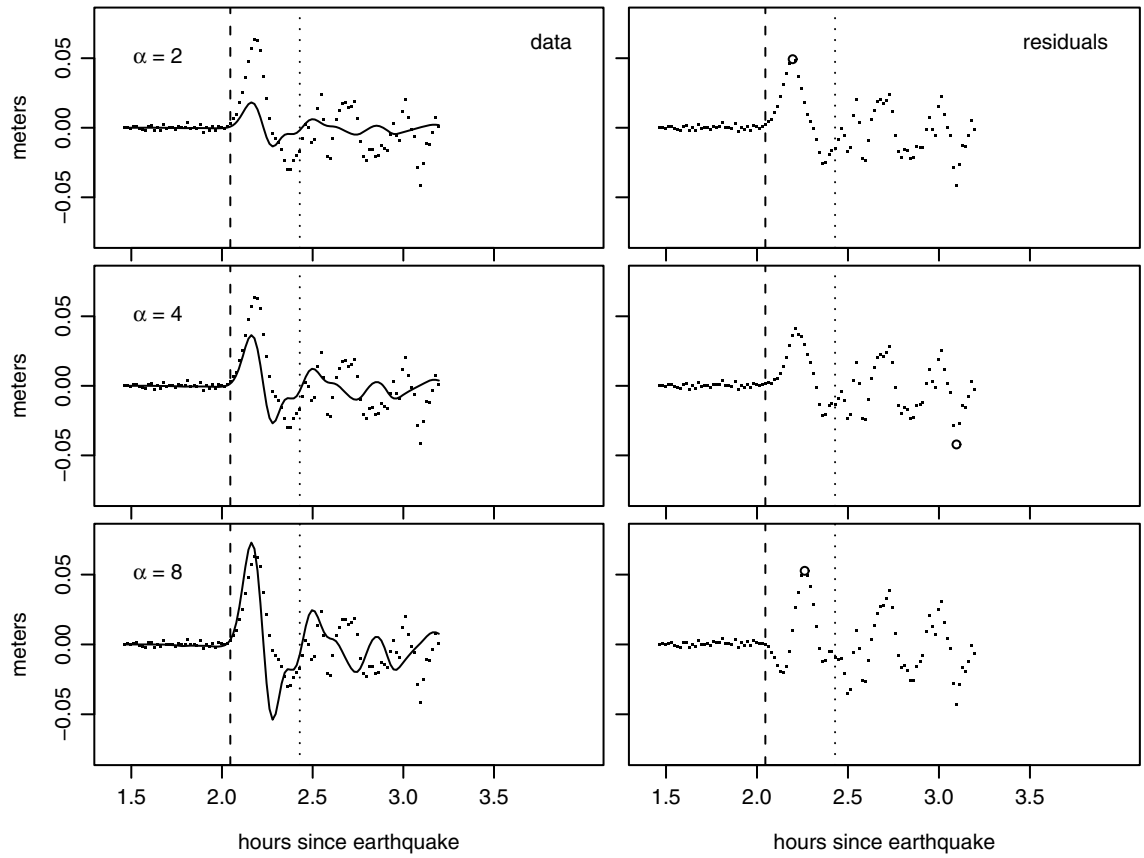

Figure 5: Adjusted model $\alpha g(t)$ with $\alpha = 2$, 4, and 8 (solid curves in left-hand column of plots), detided data from buoy 1 (dots in left-hand column) and corresponding residuals e_t (dots in right-hand column). The circle in each right-hand plot indicates the residual with the largest magnitude. The vertical dashed and dotted lines delineate the beginning and end of the first complete wave from unit source a.

2. make the sum of the magnitudes of the residuals as small as possible:

$$\sum_{t \in \widetilde{\mathcal{T}}} |e_t| = \sum_{t \in \widetilde{\mathcal{T}}} |x_t - \alpha g(t)| \equiv f_1(\alpha);$$

3. and make the largest magnitude of the residuals as small as possible:

$$\max_{t \in \widetilde{\mathcal{T}}} |e_t| = \max_{t \in \widetilde{\mathcal{T}}} |x_t - \alpha g(t)| \equiv f_\infty(\alpha).$$

These measures correspond to what are known in the literature as the L_2, L_1, and L_∞ norms. A specialized measure of some interest is to

4. make the sum of the squared residuals at the model peak and trough as small as possible:

$$e_{t_0}^2 + e_{t_1}^2 = [x_{t_0} - \alpha g(t_0)]^2 + [x_{t_1} - \alpha g(t_1)]^2 \equiv f_{\mathrm{pt}}(\alpha),$$

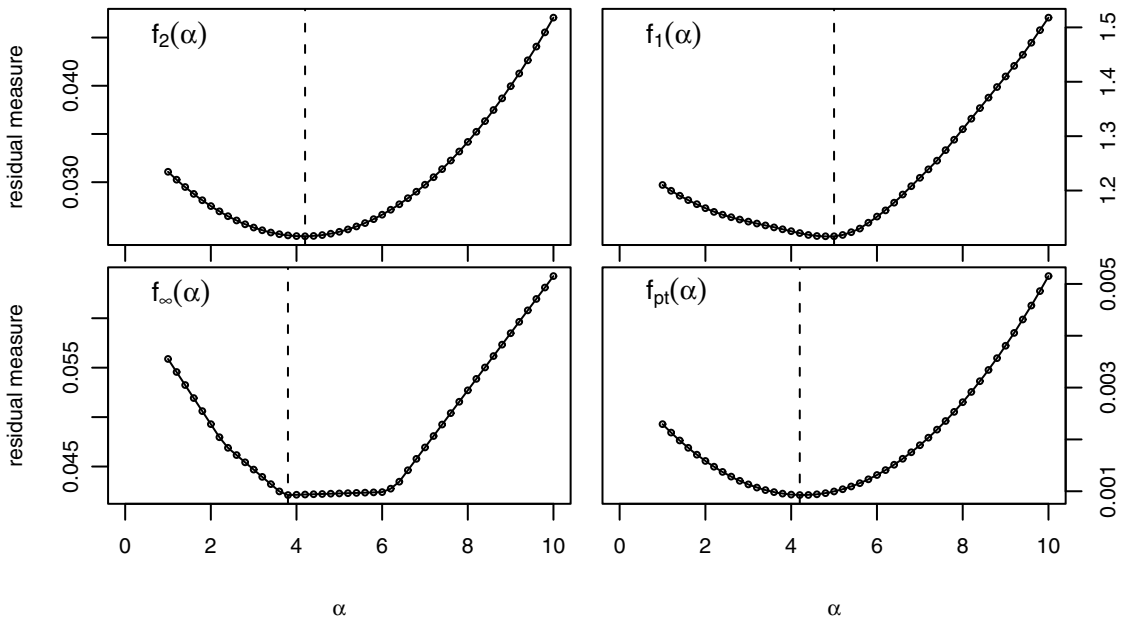

Figure 6: Residual measures $f_2(\alpha)$, $f_1(\alpha)$, $f_\infty(\alpha)$ and $f_{\rm pt}(\alpha)$ versus parameter α. The vertical dashed lines show the location of the minima for each measure.

where $t_0 \in \widetilde{\mathcal{T}}$ and $t_1 \in \widetilde{\mathcal{T}}$ are such that $g(t_0) = \max_{t\in\widetilde{\mathcal{T}}}\{g(t)\}$ and $g(t_1) = \min_{t\in\widetilde{\mathcal{T}}}\{g(t)\}$. The above can be expressed more compactly as

$$f_{\rm pt}(\alpha) = \sum_{t\in\{t_0,t_1\}} e_t^2 = \sum_{t\in\{t_0,t_1\}} \left[x_t - \alpha g(t)\right]^2 ,$$

which points out the similarity between this measure and $f_2(\alpha)$, both of which are based on squared residuals and are commonly referred to as least squares criteria. In what follows, it will be convenient to let $f_{\rm ls}(\alpha)$ stand for either $f_2(\alpha)$ or $f_{\rm pt}(\alpha)$ and to write

$$f_{\rm ls}(\alpha) = \sum_{t} e_t^2 = \sum_{t} \left[x_t - \alpha g(t)\right]^2 ,$$

where it is understood that t takes on values in the set $\widetilde{\mathcal{T}}$ for $f_2(\alpha)$ and in $\{t_0, t_1\}$ for $f_{\rm pt}(\alpha)$.

Figure 6 shows plots of $f(\alpha)$ versus α for the four measures, where we vary α from 1 to 10 in steps of 0.2, a total of 46 different settings in all. For any given measure, the estimated parameter is the value of α among the 46 settings for which $f(\alpha)$ is minimized. The estimated parameters corresponding to $f_2(\alpha)$, $f_1(\alpha)$, $f_\infty(\alpha)$ and $f_{\rm pt}(\alpha)$ are $\hat{\alpha}_2 = 4.2$, $\hat{\alpha}_1 = 5.0$, $\hat{\alpha}_\infty = 3.8$ and $\hat{\alpha}_{\rm pt} = 4.2$. These estimates are indicated on Fig. 6 by vertical dashed lines.

While all four measures give estimated parameters that are roughly the same, the two measures based on least squares criteria have certain advantages. First, the value of α minimizing $f_{\rm ls}(\alpha)$ is given by a simple formula,

namely,

$$\hat{\alpha}_{\mathrm{ls}} = \frac{\sum_t x_t g(t)}{\sum_t g^2(t)} \tag{1}$$

(this follows by setting the derivative of $f_{\mathrm{ls}}(\alpha)$ with respect to α to zero and solving the resulting equation for α). This formula yields the more refined estimates $\hat{\alpha}_2 \doteq 4.17$ and $\hat{\alpha}_{\mathrm{pt}} \doteq 4.26$, without the need to locate the minimizing α using an exhaustive grid search. Second, as we show in the next section, we can quantify the statistical variation in $\hat{\alpha}_{\mathrm{ls}}$ readily once we make some assumptions about the nature of the residuals e_t. Similar assessments for either $\hat{\alpha}_1$ or $\hat{\alpha}_\infty$ are not as easy to come by. Third, the $f_1(\alpha)$ measure is known to be less influenced by large data values than least squares measures. De-emphasizing large data values is a disadvantage in the current context because they are arguably the best indicators in the measured data of the magnitude of the tsunami event. Fourth, while $f_\infty(\alpha)$ is more influenced by large data values than least squares measures, it can be unduly influenced by a single rogue data point. Given these considerations, we will focus on the least squares measures $f_2(\alpha)$ and $f_{\mathrm{pt}}(\alpha)$ henceforth.

5. Assessing Sampling Variability in Fitted Model

Here we consider how we can quantify the sampling variability in a least squares estimator of the parameter α. To facilitate our discussion, let us reformulate the model $x_t = \alpha g(t) + e_t$ in vector notation as

$$\tilde{\mathbf{x}} = \alpha \tilde{\mathbf{g}} + \tilde{\mathbf{e}},$$

where $\tilde{\mathbf{x}}$, $\tilde{\mathbf{g}}$, and $\tilde{\mathbf{e}}$ are column vectors whose tth elements are, respectively, x_t, $g(t)$, and e_t. The dimension $\widetilde{N}$ of these vectors is the number of elements in $\widetilde{\mathcal{T}}$ for $f_2(\alpha)$ and is two for $f_{\mathrm{pt}}(\alpha)$. The least squares estimator of α_{ls} in (1) can be rewritten as $\hat{\alpha}_{\mathrm{ls}} = \tilde{\mathbf{g}}^T\tilde{\mathbf{x}}/\tilde{\mathbf{g}}^T\tilde{\mathbf{g}}$, where $\tilde{\mathbf{g}}^T$ denotes the transpose of $\tilde{\mathbf{g}}$.

In order to quantify the sampling variability in $\hat{\alpha}_{\mathrm{ls}}$, we need to consider the statistical properties of the residuals e_t. A common assumption is that the residuals are a sample from a collection of independent and identically distributed (IID) random variables from a Gaussian (normal) distribution with mean zero and variance σ_e^2. With this assumption, a standard statistical result says that $\hat{\alpha}_{\mathrm{ls}}$ is Gaussian distributed with mean α and with a variance given by $\sigma_e^2/\tilde{\mathbf{g}}^T\tilde{\mathbf{g}}$ (see, e.g., Draper and Smith, 1998). Let $\hat{\mathbf{e}} \equiv \tilde{\mathbf{x}} - \hat{\alpha}_{\mathrm{ls}}\tilde{\mathbf{g}}$ be a vector containing the observed residuals. An unbiased estimator of σ_e^2 is

$$\hat{\sigma}_e^2 = \frac{\hat{\mathbf{e}}^T\hat{\mathbf{e}}}{\widetilde{N}-1}.$$

The estimated standard deviation for $\hat{\alpha}_{\mathrm{ls}}$ is the square root of $\hat{\sigma}_e^2/\tilde{\mathbf{g}}^T\tilde{\mathbf{g}}$. Using this estimate, we can express the uncertainty in the least squares estimates

as $\hat{\alpha}_2 = 4.17 \pm 0.59$ and $\hat{\alpha}_{\rm pt} = 4.26 \pm 2.69$, with the interpretation that the corresponding intervals $[3.58, 4.76]$ and $[1.57, 6.95]$ represent 68% confidence intervals for the unknown parameter α. Note that the sampling variability in $\hat{\alpha}_{\rm pt}$ is considerably larger than that of $\hat{\alpha}_2$, presumably a reflection of the fact that $\hat{\alpha}_2$ is based on significantly more data than is $\hat{\alpha}_{\rm pt}$ (105 data points versus 2).

Unfortunately the IID assumption is highly questionable for the data used to estimate $\hat{\alpha}_2$. A plot of the observed residuals $\hat{\mathbf{e}}$ associated with this estimate looks virtually the same as the $\alpha = 4$ case shown in Fig. 5 (middle plot, right-hand column). The IID assumption dictates that the residuals have the same variance, but note that those associated with the data prior to the tsunami arrival clearly have less variability than elsewhere. Data prior to the wave arrival are associated with portions of model for which $g(t)$ is practically zero. As can be seen from (1), these data values do not influence the estimate $\hat{\alpha}_2$ since the products $x_t g(t)$ are also close to zero. For values of t corresponding to the actual tsunami, the observed residuals appear to have the same variability to a viable approximation, but they cannot reasonably be assumed to be independent: residuals that are adjacent to one another are quite similar, exhibiting a significant degree of autocorrelation.

We can adjust our model to account for the obvious violations of the IID assumption by considering the reduced model

$$\mathbf{x} = \alpha \mathbf{g} + \mathbf{e}, \tag{2}$$

where the vector $\mathbf{x}$ of dimension $N = 70$ is the portion of $\tilde{\mathbf{x}}$ that contains just the data starting from the earliest time at which $g(t)$ differs substantively from zero, with $\mathbf{g}$ and $\mathbf{e}$ being similarly extracted from $\tilde{\mathbf{g}}$ and $\tilde{\mathbf{e}}$. The least squares estimator of α is now expressed as $\hat{\alpha}_2 = \mathbf{g}^T\mathbf{x}/\mathbf{g}^T\mathbf{g}$. To account for the apparent autocorrelation in the observed residuals, we assume a simple model, namely, that the residuals obey a first-order stationary autoregressive process $e_t = \phi e_{t-1} + w_t$, where $|\phi| < 1$, and the w_t variables are IID Gaussian deviates with mean zero and a common variance. This model implies that the correlation between residuals e_t and $e_{t+\tau}$ separated by τ minutes is $\phi^{|\tau|}$. Let us redefine the vector of observed residuals to be $\hat{\mathbf{e}} \equiv \mathbf{x} - \hat{\alpha}_2 \mathbf{g}$. Since ϕ is the correlation between adjacent residuals e_t and e_{t+1}, we can estimate it via

$$\hat{\phi} \equiv \frac{\hat{\mathbf{e}}_{(f)}^T \hat{\mathbf{e}}_{(l)}}{\hat{\mathbf{e}}^T \hat{\mathbf{e}}} \doteq 0.86, \tag{3}$$

where $\hat{\mathbf{e}}_{(f)}$ consists of all of $\hat{\mathbf{e}}$ except for its first element, and $\hat{\mathbf{e}}_{(l)}$ has everything but the last element. Statistical theory says $\hat{\alpha}_2$ is Gaussian distributed with a mean of α and a variance given by $\sigma_e^2\, \mathbf{g}^T V \mathbf{g}/(\mathbf{g}^T\mathbf{g})^2$, where V is a matrix whose (j,k)th element is $\phi^{|j-k|}$. An approximately unbiased estimator of σ_e^2 is given by

$$\hat{\sigma}_e^2 = \frac{(1-\phi)^2 \hat{\mathbf{e}}^T \hat{\mathbf{e}}}{N(1-\phi)^2 - (1-\phi^2) + 2\phi(1-\phi^N)N^{-1}}. \tag{4}$$

We can thus estimate the standard deviation of $\hat{\alpha}_2$ by taking the square root of $\hat{\sigma}_e^2\, \mathbf{g}^T V \mathbf{g}/(\mathbf{g}^T\mathbf{g})^2$ after replacing ϕ by $\hat{\phi}$. We now have $\hat{\alpha}_2 = 4.17 \pm$

1.50, which reports a larger uncertainty than under the questionable IID assumption, namely, $\hat{\alpha}_2 = 4.17 \pm 0.59$. (The procedure for getting a more realistic estimate of the uncertainty in $\hat{\alpha}_2$ breaks down if we attempt to adjust it to work with $\hat{\alpha}_{\mathrm{pt}}$. This estimator is based on just two data points, and hence getting a good estimate of ϕ is problematic.)

6. Selecting Amount of Data to Use for Estimating α

The inversion algorithm in Version 2.0 of SIFT uses the least squares estimator $\hat{\alpha}_2$ to estimate the parameter α based upon available DART® buoy data. In the example in the previous section, we used all of the data relevant to the Kuril Islands tsunami event observed by buoy 1. Using the estimated parameter to forecast the potential dangers of a tsunami to coastal communities leads naturally to the question of how this estimate would be affected if we were to use less data than what eventually became available. In theory, using more data should yield a better estimate of the parameter, but this must be balanced off with the necessity of obtaining an estimate as quickly as possible. Here we use the example presented in the last section to illustrate how estimates of the parameter change as we use smaller amounts of data.

The top plot of Fig. 7 shows the data (the dots) for the Kuril Islands tsunami event recorded at buoy 1, along with the model (shown as a curve) from unit source a after multiplication by the parameter $\hat{\alpha}_2 = 4.17$, which is based on all the available data. There are two vertical dashed lines on the plot, which mark the times $t_0 = 2.046$ and $t_1 = 2.196$. The first of these is the earliest time at which the model $g(t)$ is measurably different from zero; the second, the time just following the crest of the first full tsunami wave as measured at the buoy (i.e., just after the so-called quarter-wave point). There are 10 data points in all between (and including) t_0 and t_1; similarly, there are 70 data points between t_0 and the last observed value ($t_e = 3.196$). Various results related to estimating α using just the first 10 points and all 70 points are shown by, respectively, the left and right ends of the curves in the bottom four plots. The rest of the curves shows results using $11, 12, \ldots, 69$ points. From top to bottom, the second plot in the figure shows the least squares estimates $\hat{\alpha}_2$; the third, the estimated standard deviations for $\hat{\alpha}_2$; the fourth, the root mean square (RMS) errors of the residuals over just the data used to determine $\hat{\alpha}_2$ (solid curve) and over all 70 data points (dashed curve, which necessarily corresponds exactly to the solid curve at the very end); and the fifth, two versions of the so-called R^2 statistic, which is the percentage of the sample variance of the data explained by the model (the solid curve is the squared correlation—expressed as a percentage—between the observed data and the fitted model, while the dashed curve shows the correlations under the assumptions that the true means of both the data and the model are zero).

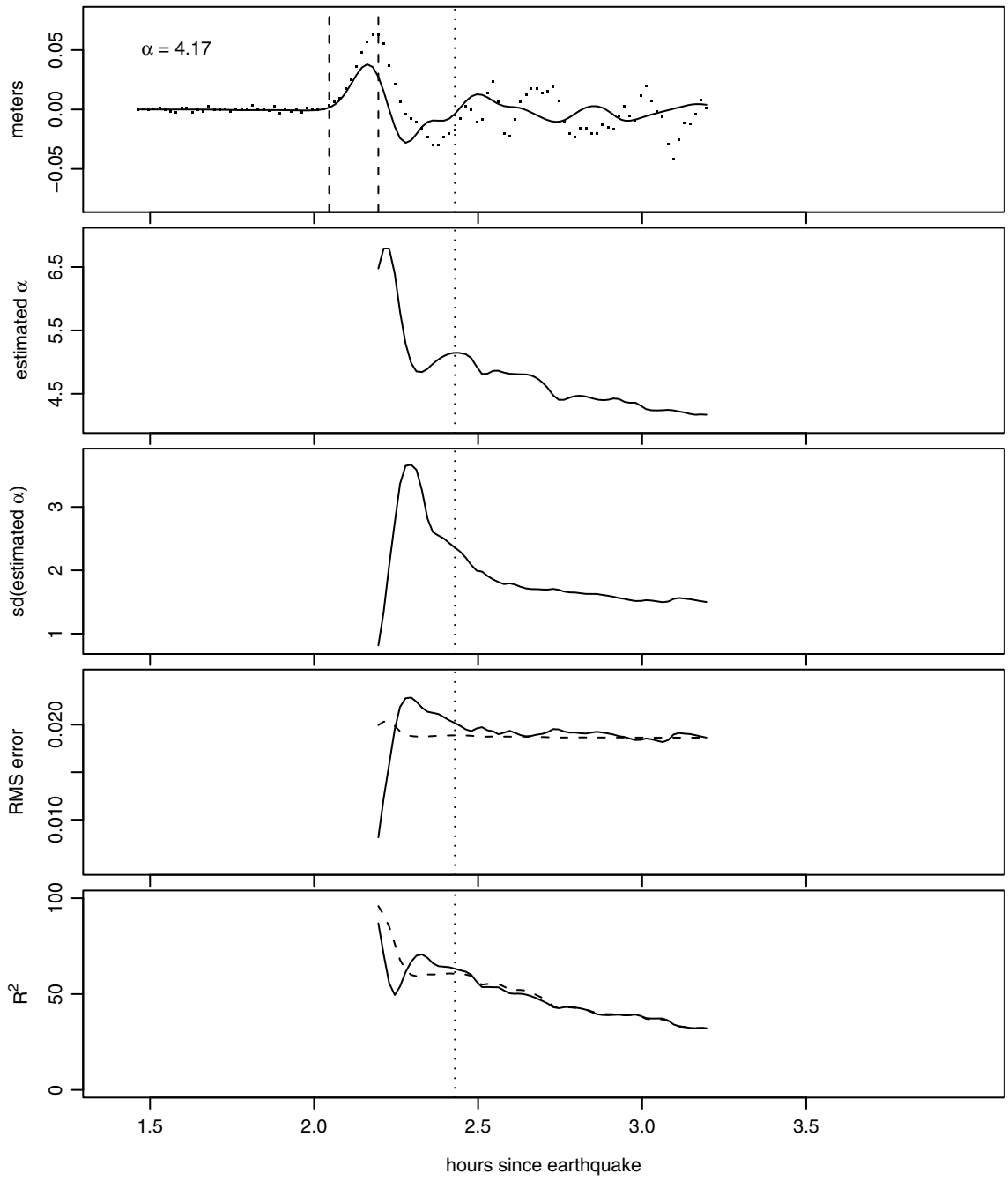

Figure 7: Detided data from buoy 1 (dots, top plot) and model $\alpha g(t)$ from unit source a, with α set to the least squares estimate $\hat{\alpha}_2 = 4.17$ using all the data (solid curve, top plot). The bottom four plots summarize the effect of using increasing amounts of data to estimate α, with the smallest amount used being indicated by the vertical dashed lines in the top plot. From top to bottom, these plots show $\hat{\alpha}$, the estimated standard deviation for $\hat{\alpha}$, the RMS error over the selected data (solid curve) and all the data (dashed) and two variations on the R^2 statistic. The left-most dashed line and the dotted line in the top plot delineate the 24 data points recorded during the first complete wave from unit source a; the dotted lines in the other plots mark values corresponding to an estimate of α based on these 24 values.

Some interesting patterns emerge as the amount of data used to estimate α increases from 10 to 70. The second plot shows that the estimated parameters $\hat{\alpha}_2$ roughly decrease with increasing amounts of data, so the estimated parameter does depend on the amount of data used. The third indicates that the associated standard deviation of the estimator increases initially, but then decreases as more data are used, as intuition would suggest. The fourth shows that the root mean square errors over the data used in the fit and over all the data are relatively constant after about $t = 2.4$. The last plot shows that the percent variance explained is initially fairly high (around 90%) when just 10 data points are used, but then decreases monotonically after about $t = 2.5$ with the use of more data, eventually ending up at around 32%. This indicates that the model has little explanatory power beyond the first complete wave observed at the buoy.

This Kuril Islands tsunami event is one in which the very first wave is the largest when observed at buoy 1. This is because there is an unobstructed path for the tsunami to propagate from the source of the event to this buoy (see Fig. 1, where buoy 1 is marked as 21414). For this case, we are thus better off just using the data up to the first complete wave (as defined by model a) to estimate α since the data and the model disagree substantially beyond that point. Doing so yields $\hat{\alpha}_2 = 5.15 \pm 2.36$ based upon $N = 24$ data points. 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 parameter α. How best to select the amount of data to use in the inversion procedure is an open question requiring additional research. Version 2.0 of SIFT allows the user to manually select the portions of DART® data to use from each buoy, but there are provisional plans for future versions to offer guidance based upon statistical tests.

7. Handling Multiple Unit Sources and Buoys

So far we have considered an inversion procedure that is based upon a model from a single unit source for the data received at a single DART® buoy. In practice, certain tsunami events are potentially better described by considering multiple source models. This possibility is illustrated in Fig. 1, where sea floor disturbances are highlighted in several unit source rectangles that might have contributed to the tsunami. In addition the same tsunami event can be recorded by multiple buoys. Figure 2 reflects the fact that the Kuril Islands event was clearly detected at at least two DART® buoys, thereby providing additional observations for use in source determination. In the subsections that follow, we describe extensions to our inversion procedure that handle multiple source models and/or multiple buoys. We use unit sources a and b and buoys 1 and 2 to illustrate how these extensions work. For $j = 1$ or 2, we let $x_{j,t}, t \in \mathcal{T}_j$, represent the relevant data from buoy j, and we let $\mathbf{x}_j$ be the corresponding vector of length N_j, where N_j is the

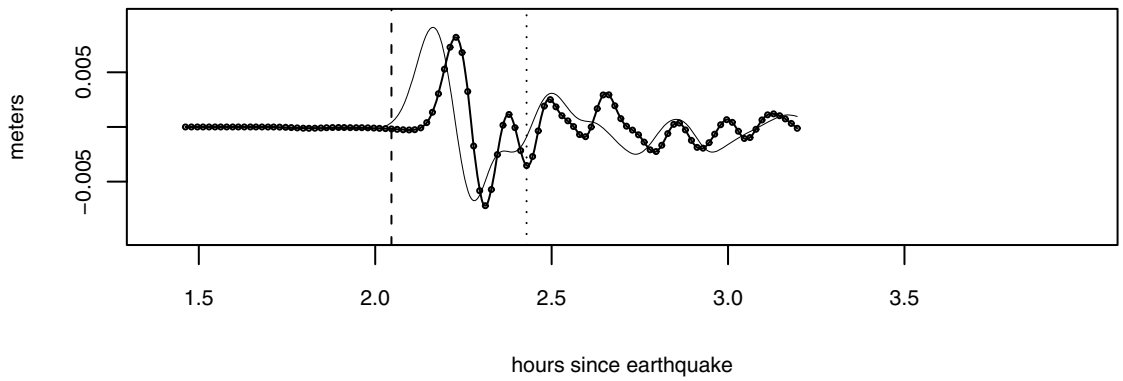

Figure 8: Model series for buoy 1 from unit sources a (thin curve, which is the same as the curve in Fig. 3) and b (thick curve with dots) generated under the assumption that a magnitude 7.5 reverse thrust earthquake originated from each source alone. The dashed and dotted lines mark the beginning and end of the first complete wave from source a.

number of elements in the set of indices $\mathcal{T}_j$. With this new notation, we can re-express the reduced model of (2) for buoy 1 as

$$\mathbf{x}_1 = \alpha_a \mathbf{g}_{1;a} + \mathbf{e}_1, \tag{5}$$

where the subscript on α_a says that this parameter is associated with the model from unit source a; the subscripts on $\mathbf{g}_{1;a}$ indicate that this vector contains the model for buoy 1 from source a; and the subscript on $\mathbf{e}_1$ says that these are the residuals associated with buoy 1. We can now express the least squares estimator of α_a as

$$\hat{\alpha}_{1;a} = \mathbf{g}_{1;a}^T \mathbf{x}_1 / \mathbf{g}_{1;a}^T \mathbf{g}_{1;a} = \left(\mathbf{g}_{1;a}^T \mathbf{g}_{1;a}\right)^{-1} \mathbf{g}_{1;a}^T \mathbf{x}_1 \tag{6}$$

(cf. Equation (1)), where the subscripts on $\hat{\alpha}_{1;a}$ emphasize that this estimate of the parameter for unit source a is based on data from buoy 1 (the right-most expression above is of interest to compare with generalizations to come).

7.1 Multiple Unit Sources and a Single Buoy

The event that generates a tsunami is sometimes better described as originating from more than just one unit source. When this is the case, we assume that the overall effect of the generating event is a linear combination of the individual models from each source. As an example, the thin curve in Fig. 8 shows the model from unit source a for buoy 1 (shown previously in Fig. 3), while the thick curve with superimposed dots shows the model from source b. Note that the first quarter wave from b arrives later than that from a, but that the duration of the first complete wave is shorter for b than for a. If we compare the data from buoy 1 with the fitted model from a in the top plot of Fig. 7, we see that the peak in the data occurs after the peak in the model, and hence the peak location might be better represented

by the model coming from b; on the other hand, the duration of the first wave in the data is more in keeping with that exhibited by a than by b. We might get a better overall fit if we postulate an overall model for the data from buoy 1 as being a blend of the models coming from a and b.

Accordingly we now augment the model of Equation (5) to include an additional term:

$$\mathbf{x}_1 = \alpha_a \mathbf{g}_{1;a} + \alpha_b \mathbf{g}_{1;b} + \mathbf{e}_1,$$

where α_b is the parameter for unit source b, and $\mathbf{g}_{1;b}$ is a vector of dimension N_1 containing the unadjusted model for buoy 1 from source b. As before, we assume that the residual term $\mathbf{e}_1$ is multivariate Gaussian with a mean vector of zeros and a covariance matrix given by $\sigma_1^2 V_1$, where the (j,k)th element of the $N_1 \times N_1$ matrix V_1 is given by $\phi_1^{|j-k|}$ for some $|\phi_1| < 1$. We can express this model more compactly as

$$\mathbf{x}_1 = G_1 \boldsymbol{\alpha} + \mathbf{e}_1, \tag{7}$$

where G_1 is an $N_1 \times 2$ dimensional array whose first and second columns are $\mathbf{g}_{1;a}$ and $\mathbf{g}_{1;b}$, while $\boldsymbol{\alpha} = [\alpha_a, \alpha_b]^T$. The least squares estimator $\hat{\boldsymbol{\alpha}}_1$ of $\boldsymbol{\alpha}$ satisfies the equation

$$G_1^T G_1 \hat{\boldsymbol{\alpha}}_1 = G_1^T \mathbf{x}_1, \tag{8}$$

where the subscript on $\hat{\boldsymbol{\alpha}}_1$ emphasizes that the parameters are being estimated based upon the data from buoy 1. Under the assumption that the 2×2 matrix $G_1^T G_1$ is invertible, the least squares estimator is given by

$$\hat{\boldsymbol{\alpha}}_1 = \left(G_1^T G_1\right)^{-1} G_1^T \mathbf{x}_1,$$

which is similar in form to the estimator in Equation (6) when using unit source a alone. Standard least squares theory says that $\hat{\boldsymbol{\alpha}}_1$ is an unbiased estimator of $\boldsymbol{\alpha}$ and obeys a Gaussian distribution with a covariance matrix given by

$$\Sigma_1 \equiv \sigma_1^2 \left(G_1^T G_1\right)^{-1} G_1^T V_1 G_1 \left(G_1^T G_1\right)^{-1} \tag{9}$$

(see, e.g., Draper and Smith, 1998 for details). Once we have determined $\hat{\boldsymbol{\alpha}}_1$, we can compute the observed residuals

$$\hat{\mathbf{e}}_1 = \mathbf{x}_1 - G_1 \hat{\boldsymbol{\alpha}}_1$$

and use these to form an estimator of ϕ_1 via the obvious analog of Equation (3). Conditional upon this estimate of ϕ_1, we can obtain an approximately unbiased estimator of σ_1^2 using the analog of Equation (4), following which we have all the pieces needed to determine the elements of the covariance matrix Σ_1.

Using the $N_1 = 24$ data values collected by buoy 1 during the first complete wave from source a, the least squares estimate of $\boldsymbol{\alpha}$ is

$$\hat{\boldsymbol{\alpha}}_1 = \begin{bmatrix} \hat{\alpha}_{1;a} \\ \hat{\alpha}_{1;b} \end{bmatrix} \doteq \begin{bmatrix} 4.24 \\ 3.00 \end{bmatrix}. \tag{10}$$

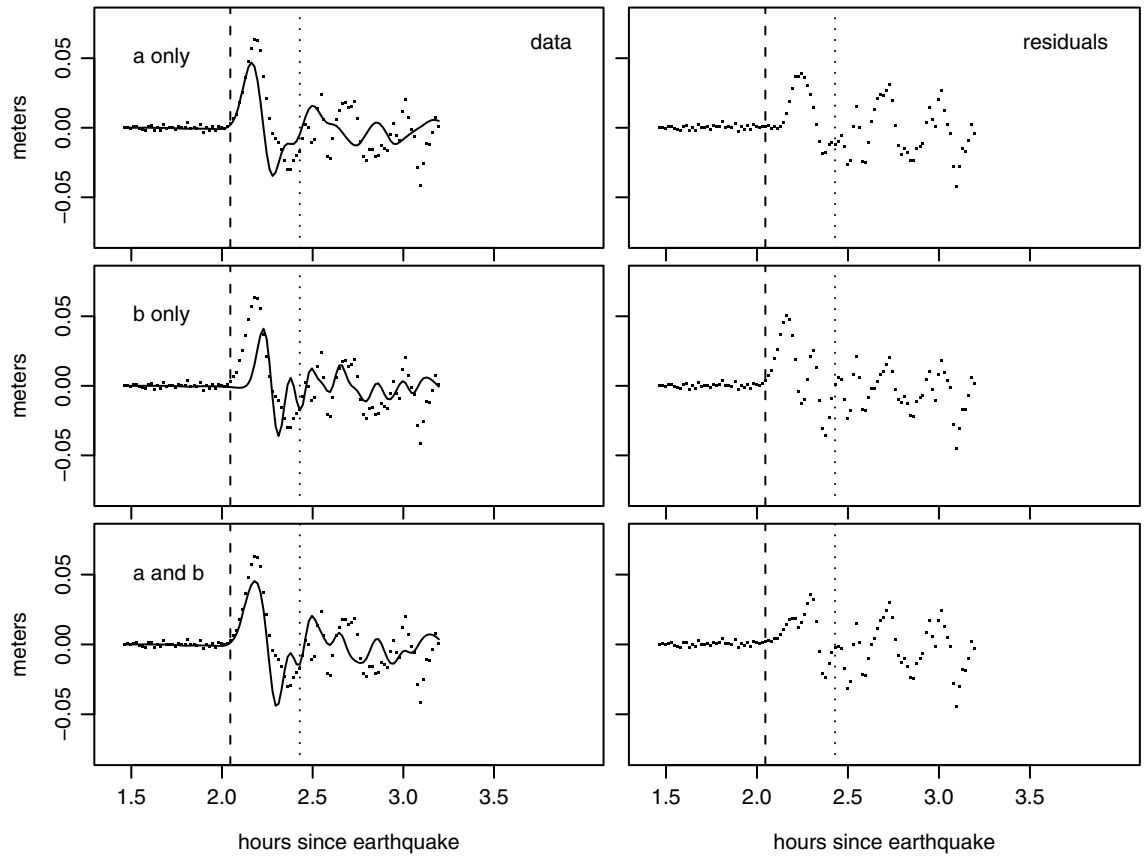

Figure 9: Comparison of fitted models involving unit sources a and b with data from buoy 1. The left-hand plots show the data (dots) along with the models after adjustment using the estimated source parameters (solid curves). The right-hand plots show the corresponding observed residuals. The vertical dashed and dotted lines mark the stretch of 24 data points used to estimate the parameters. In the first row, only unit source a is used; in the second, only b; and the bottom row uses both sources together.

The corresponding estimates of ϕ_1 and σ_1^2 are 0.89 and 0.00061, based upon which we calculate

$$\hat{\Sigma}_1 \doteq \begin{bmatrix} 3.43 & -0.11 \\ -0.11 & 2.19 \end{bmatrix}. \tag{11}$$

The estimated parameters, along with their standard errors, for the two sources are thus $\hat{\alpha}_{1;a} \doteq 4.24 \pm 1.85$ and $\hat{\alpha}_{1;b} \doteq 3.00 \pm 1.48$.

The left-hand plot in the bottom row of Fig. 9 shows the fitted model $G_1\hat{\boldsymbol{\alpha}}_1$ over $\mathcal{T}_1$ and its extension outside of $\mathcal{T}_1$ (solid curve), where the vertical dashed and dotted lines indicate the earliest and latest times in $\mathcal{T}_1$. The data from buoy 1 that were used to determine $\hat{\boldsymbol{\alpha}}_1$ are shown as dots between the vertical lines, and the remaining data, as the dots before and after these lines. The right-hand plot shows the observed residuals $\hat{\mathbf{e}}_1$ (dots between the vertical lines) and the differences between the observed data and the extended fitted model (dots outside the vertical lines). The top row in

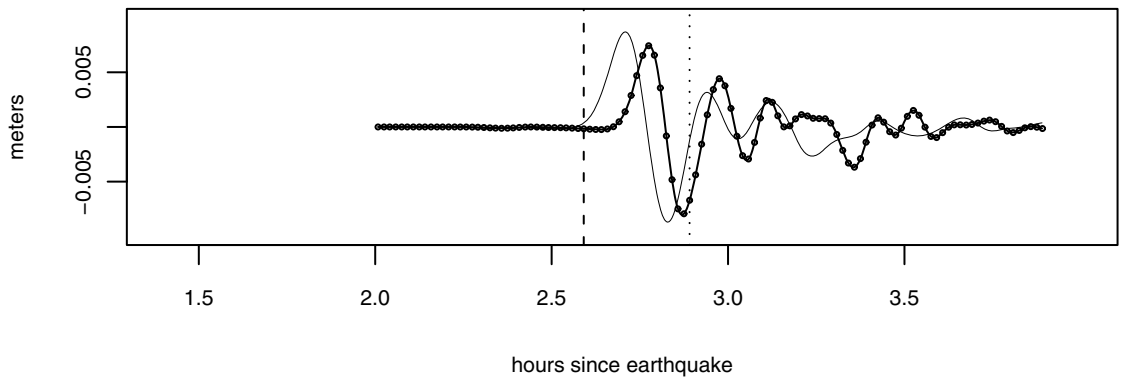

Figure 10: As in Fig. 8, but now for buoy 2.

Fig. 9 shows corresponding plots when the data are modeled using just unit source a, and the middle row, just source b. We see that the use of both unit sources gives somewhat smaller observed residuals over the first quarter wave of the observed data.

We can readily adapt the above procedure to handle more than two unit sources. If we are interested in N_s sources in all, then G_1 becomes an $N_1 \times N_s$ dimensional array whose kth column contains the unadjusted model for buoy 1 from the kth source, while $\boldsymbol{\alpha}$ becomes an N_s dimensional column vector. Two problems, however, can crop up when dealing with multiple unit sources. The first is that, due to dependencies amongst the models from different unit sources, the matrix $G_1^T G_1$ might not be of full rank or might be close enough to being rank deficient so that computing the inverse of $G_1^T G_1$ becomes numerically unstable. We deal with this potential problem by using a singular value decomposition in solving Equation (8) for $\boldsymbol{\alpha}$ to determine $\hat{\boldsymbol{\alpha}}_1$. The second problem is that some of the estimated parameters might turn out to be positive, while others are negative. A mixture of positive and negative parameters is difficult to reconcile with the physics of earthquake generation. While use of such a mixture might achieve a closer fit between model and observations, the application of these parameters throughout the ocean basin can lead to unrealistic predictions for inundations of coastal regions. We can deal with this problem using a constrained least squares procedure, whereby, given $\mathbf{x}_1$ and G_1, we minimize

$$\|\mathbf{x}_1 - G_1\boldsymbol{\alpha}\|^2 \text{ over } \boldsymbol{\alpha} \text{ subject to either the condition } \boldsymbol{\alpha} \geq \mathbf{0} \text{ or } \boldsymbol{\alpha} \leq \mathbf{0},$$

i.e., all of the elements of $\boldsymbol{\alpha}$ are either nonnegative or nonpositive (in the above $\|\cdot\|$ denotes the Euclidean norm, and $\mathbf{0}$ is an N_s dimensional column vector of zeros). This constrained least squares procedure can result in some parameters being set to zero, which in effect removes the corresponding unit sources from our model for the data. The above 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 of Fletcher (1987).

7.2 Multiple Unit Sources and Multiple Buoys

We now consider the data collected by buoy 2 (i.e., 46413), which is shown in the bottom plot of Fig. 2. Figure 10 shows the unadjusted models for these data from unit sources a and b. We can make use of these additional data and associated models in three ways. First, we can use the least squares estimates of the parameters from Equation (10) based upon the data from buoy 1 to see how well it matches the data observed by buoy 2 (i.e., "cross-validation"). The first row of Fig. 11 shows the data for buoy 2, its adjusted model using unit sources a and b and the associated residuals (the layout of the plots follows that of Fig. 9). The agreement between the model and the data is relatively good, thus lending some additional credence to the estimated parameters. Second, we can formulate a model that is analogous to what we used for buoy 1 and estimate $\boldsymbol{\alpha}$ based upon just the data from buoy 2. Let $\mathbf{x}_2$ be a vector of dimension $N_2 = 19$ containing the data from buoy 2 that was sampled over the first complete wave in the model from unit source a (this region is marked by the dashed and dotted vertical lines in Fig. 10 and in the middle two rows of Fig. 11). The model parallels that of Equation (7) and states that

$$\mathbf{x}_2 = G_2\boldsymbol{\alpha} + \mathbf{e}_2, \tag{12}$$

where G_2 is an $N_2 \times 2$ dimensional array whose first and second columns are $\mathbf{g}_{2;a}$ and $\mathbf{g}_{2;b}$ (vectors of dimension N_2 containing the unadjusted models for buoy 2 from sources a and b), and $\mathbf{e}_2$ is an N_2 dimensional vector of error terms that is multivariate Gaussian with zero mean and a covariance matrix given by $\sigma_2^2 V_2$ (the (j,k)th element of the $N_2 \times N_2$ matrix V_2 is given by $\phi_2^{|j-k|}$ for some $|\phi_2| < 1$). The least squares estimate of $\boldsymbol{\alpha}$ is

$$\hat{\boldsymbol{\alpha}}_2 = \begin{bmatrix} \hat{\alpha}_{2;a} \\ \hat{\alpha}_{2;b} \end{bmatrix} \doteq \begin{bmatrix} 3.43 \\ 3.63 \end{bmatrix}.$$

The estimates of ϕ_2 and σ_2^2 are 0.95 and 0.00098, based upon which we calculate the analog of $\hat{\Sigma}_1$ in Equation (11), namely,

$$\hat{\Sigma}_2 \doteq \begin{bmatrix} 2.49 & 0.19 \\ 0.19 & 1.94 \end{bmatrix}.$$

The estimated parameters are $\hat{\alpha}_{2;a} \doteq 3.43 \pm 1.58$ and $\hat{\alpha}_{2;b} \doteq 3.63 \pm 1.39$, which, when sampling variability is taken into account, are in reasonably good agreement with what we obtained using just the data from buoy 1. The second row of Fig. 11 has plots for this model that parallel those in the top row.

A third use of the additional data from buoy 2 is to estimate $\boldsymbol{\alpha}$ using a model involving data $\mathbf{x}_1$ and $\mathbf{x}_2$ from both buoys. This joint model can be formulated by "stacking" together the models given by Equations (7) and (12):

$$\mathbf{x}_{1:2} = G_{1:2}\boldsymbol{\alpha} + \mathbf{e}_{1:2}, \tag{13}$$

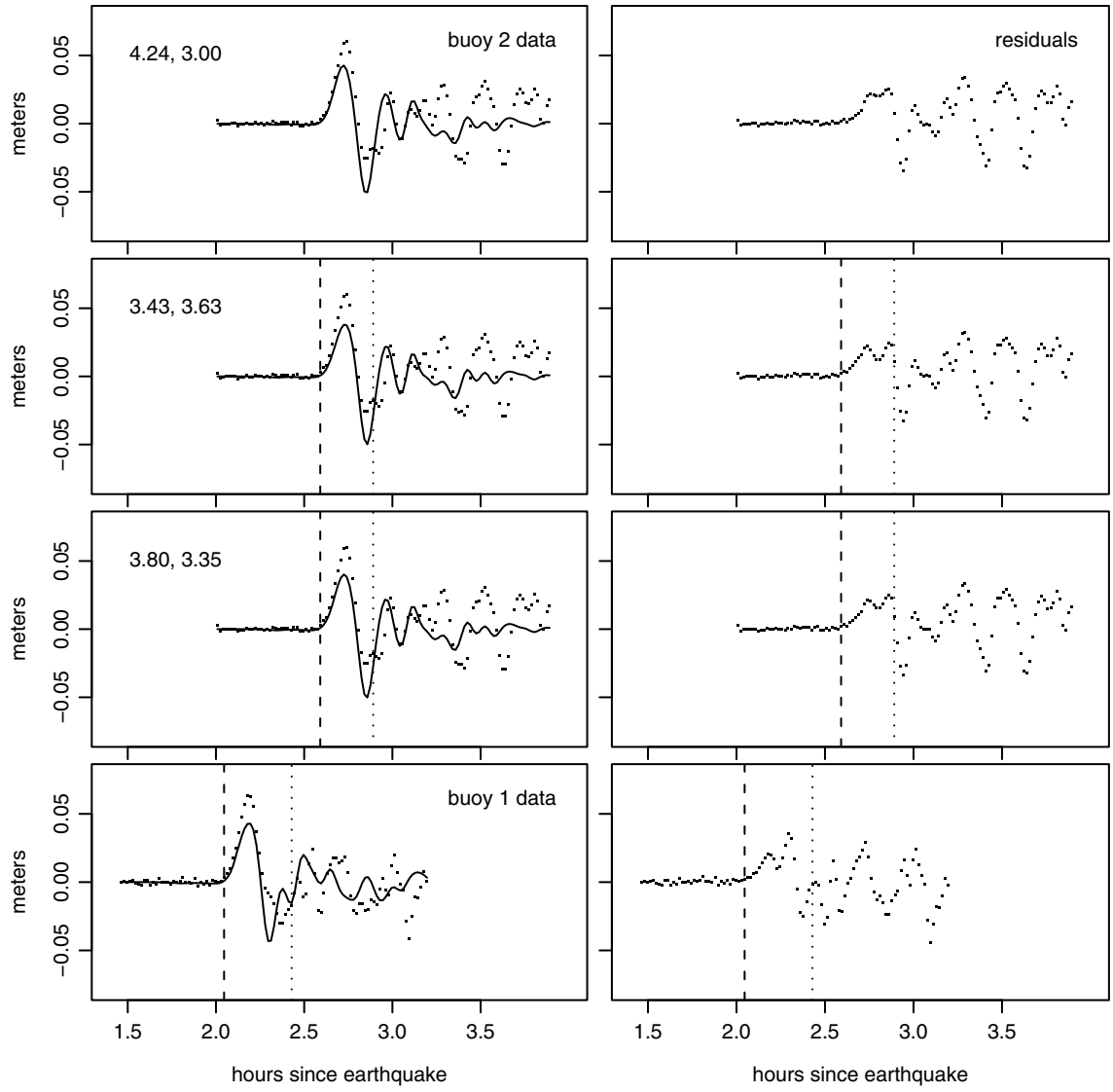

Figure 11: As in Fig. 9, with upper three rows comparing the data from buoy 2 to models based upon unit sources a and b with parameters determined by data from buoy 1 alone (upper row), buoy 2 alone (second row) and both buoys (third). The two values given in the upper left-hand corners are the estimated parameters involving unit sources a and b. The bottom plot compares the data from buoy 1 to the model using the parameters determined by data from both buoys (i.e., the same as in the third row). In the left-hand plots the vertical dashed and dotted lines indicate the starts and ends of the data used to estimate the parameters (these lines are replicated in the right-hand plots).

where $\mathbf{x}_{1:2}$ is an $N_1 + N_2$ dimensional vector obtained by stacking $\mathbf{x}_1$ on top of $\mathbf{x}_2$; $G_{1:2}$ is an $(N_1 + N_2) \times 2$ dimensional array obtained by stacking G_1 on top of G_2; and $\mathbf{e}_{1:2}$ is the residual vector; thus

$$\mathbf{x}_{1:2} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} \text{ and } G_{1:2} = \begin{bmatrix} G_1 \\ G_2 \end{bmatrix}.$$

We assume that $\mathbf{e}_{1:2}$ is multivariate Gaussian with a mean vector of zeros and an $(N_1 + N_2) \times (N_1 + N_2)$ covariance matrix given by

$$V_{1:2} = \begin{bmatrix} \sigma_1^2 V_1 & 0 \\ 0^T & \sigma_2^2 V_2 \end{bmatrix},$$

where the (j,k)th element of the $N_1 \times N_1$ dimensional submatrix V_1 is given by $\phi_1^{|j-k|}$ while that for the $N_2 \times N_2$ dimensional submatrix V_2 is $\phi_2^{|j-k|}$, where $|\phi_1| < 1$ and $|\phi_2| < 1$ (in the above 0 and 0^T are taken to represent an $N_1 \times N_2$ matrix of zeros and its transpose). This model assumes that any two error terms are uncorrelated if they are associated with different buoys (this assumption is a reasonable starting point, but is subject to further research pending examination of residuals from operational use of this model). The least squares estimator $\hat{\boldsymbol{\alpha}}_{1:2}$ of $\boldsymbol{\alpha}$ satisfies the equation

$$G_{1:2}^T G_{1:2} \hat{\boldsymbol{\alpha}}_{1:2} = G_{1:2}^T \mathbf{x}_{1:2}, \tag{14}$$

which is analogous to Equation (8). As with the models involving data from a single buoy, $\hat{\boldsymbol{\alpha}}_{1:2}$ is an unbiased estimator of $\boldsymbol{\alpha}$ and follows a Gaussian distribution with a covariance matrix given by

$$\Sigma_{1:2} \equiv \left(G_{1:2}^T G_{1:2}\right)^{-1} G_{1:2}^T V_{1:2} G_{1:2} \left(G_{1:2}^T G_{1:2}\right)^{-1}$$

(this is analogous to Equation (9)). The least squares estimate of $\boldsymbol{\alpha}$ is

$$\hat{\boldsymbol{\alpha}}_{1:2} = \begin{bmatrix} \hat{\alpha}_{1:2;a} \\ \hat{\alpha}_{1:2;b} \end{bmatrix} \doteq \begin{bmatrix} 3.80 \\ 3.35 \end{bmatrix}.$$

The corresponding estimates of ϕ_1, ϕ_2, σ_1^2, and σ_2^2 are 0.89, 0.96, 0.00059 and 0.00113, based upon which we calculate

$$\hat{\Sigma}_{1:2} \doteq \begin{bmatrix} 1.48 & 0.04 \\ 0.04 & 1.04 \end{bmatrix}.$$

The estimated parameters using data from both buoys are thus $\hat{\alpha}_{1:2;a} \doteq 3.80 \pm 1.22$ and $\hat{\alpha}_{1:2;b} \doteq 3.35 \pm 1.02$. When we take into account sampling variability, these estimates are consistent with what we got using the data from each buoy by itself. The standard errors associated with these estimated parameters are smaller than those obtained when we used data from just one buoy. The last two rows of Fig. 11 compare the observations, model fits, and residuals at buoy 2 and buoy 1 when sources a and b and data from the first complete wave at each buoy are considered.

8. Summary and Discussion

We have described in some detail the SIFT inversion procedure by which data from DART® buoys are used to estimate parameters that adjust precomputed models for the data arising from one or more unit sources. The models are computed under the assumption that a magnitude 7.5 reverse thrust earthquake occurred within each unit source. The parameters α essentially adjust the models to take into account earthquakes whose magnitude differs from 7.5, and does not coincide with, or is of greater spatial extent than, a single unit source. We advocate a least square criterion for estimating the parameters. This criterion has appeal for picking out appropriate estimates for the problem at hand and, in addition, leads to quick computation and realistic assessment of the effects of sampling variability.

Version 2.0 of SIFT allows the user to select which unit sources are to provide models for the observed data. For upcoming versions we plan to implement ideas from statistical model selection to help the selection process. Given the now completed U.S. array of DART® buoys in the Pacific and Atlantic Oceans and given the extensive database of precomputed models from over 1200 unit sources providing coverage of potential tsunami sources in these oceans, a statistical approach to selecting appropriate combinations of models and data can relieve some of the burden on users, particularly when using SIFT during an ongoing tsunami event. Two simple approaches involve step-up or step-down strategies. Suppose, for example, that we entertain eight unit sources in all. A step-up approach would start with the unit source whose model gives the best individual fit to the data (as measured by residual variance) and would then use statistical tests to add models from other unit sources one at a time, stopping when using an additional model does not significantly improve the match with the data. A step-down approach would start with a model that involves all eight unit sources and then would use statistical tests to remove sources that do not appear to be important contributors to the tsunami event.

Other areas of on-going research that might impact future versions of SIFT are (1) the use of statistical criteria to select the amount of data to be used to estimate the parameters; (2) allowing the precomputed models to be shifted in time a bounded amount to compensate for uncertainties in propagation times between the unit sources and the buoys; (3) real-time detiding of the DART® buoy data; and (4) use of a Kalman filtering formulation to allow efficient updating of the estimated parameters as new data arrives.

Finally we note that an important topic that we have left unaddressed is the connection between the estimated source parameters α and coastal inundation forecasts that use these parameters as inputs. The quality of these forecasts depends on how well the source parameters can be used to successfully predict the tsunami event at locations far removed from the DART® stations whose data formed the basis for the parameter estimates. For ex-

ample, in the context of the Kuril Islands event focused on here, a further test would be to see how well we can use the estimated α's to predict what was actually observed at far-field DART® stations along the West Coast of the United States. An additional test could be based on tide gage data at Crescent City and Hilo using nearshore standby inundation model (SIM) forecasts (see Gica *et al.*, 2009 for details). These additional tests need to be done in order to fully validate the inversion procedure. In addition, work is needed to determine how uncertainty in the estimated source parameters translates into uncertainty in the SIM forecasts themselves.

9. Acknowledgments

This work is 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. 1627. This work is also Contribution No. 3134 from NOAA/Pacific Marine Environmental Laboratory. The authors would also like to thank R.L. Whitney for his help in putting this document together.

10. References

Draper, N.R., and H. Smith (1998): *Applied Regression Analysis*. 3rd edition, John Wiley and Sons, New York, 736 pp.

Fletcher, R. (1987): *Practical Methods of Optimization*. 2nd edition, John Wiley and Sons, Chichester and New York, 450 pp.

Gica, E., M. Spillane, D. Arcas, D.B. Percival, and V.V. Titov (2009): Development of NOAA's Short-term Inundation Forecast for Tsunamis (SIFT) and a case study of the 15 November 2006 Kuril tsunami. NOAA Tech. Memo. OAR PMEL, NOAA/Pacific Marine Environmental Laboratory, Seattle, WA, to be submitted.

Gica, E., M. Spillane, V.V. Titov, C.D. Chamberlin, and J.C. Newman (2008): Development of the forecast propagation database for NOAA's Short-term Inundation Forecast for Tsunamis (SIFT). NOAA Tech. Memo. OAR PMEL-139, NOAA/Pacific Marine Environmental Laboratory, Seattle, WA, 89 pp.

Horrillo, J., W. Knight, and Z. Kowalik (2008): Kuril Islands tsunami of November 2006: 2. Impact at Crescent City by local enhancement. *J. Geophys. Res.*, *113*, c01021, doi: 10.1029/2007JC004404.

Kowalik, Z., J. Horrillo, W. Knight, and T. Logan (2008): Kuril Islands tsunami of November 2006: 1. Impact at Crescent City by distant scattering. *J. Geophys. Res.*, *113*, c01020, doi: 10.1029/2007JC004402.