Detiding DART$^{®}$ Buoy Data and Extraction of Source Coefficients: A Joint Method

Don Percival

Applied Physics Laboratory
Department of Statistics
University of Washington, Seattle

Overview

- variability in DART® buoy data is dominated by tides
- during a tsunami event, tidal fluctuations – in combination with background noise – can make it difficult to extract a tsunami signal buried within DART® data
- one commonly used approach (with many variations)
  - predict tidal component somehow
  - subtract predictions from DART® data (*detiding* step)
  - obtain tsunami signal from detided data (*extraction* step)
- will look briefly at four variations on this basic idea and then discuss a promising approach that carries out *detiding* and *extraction* steps simultaneously (the *joint method*)

Background on DART® Buoy Data Types

- three types of DART® buoy data:
  - 15-sec stream: bottom pressure (BP) measurements collected every 15 seconds and fully retrieved only during servicing of buoy (once a year or so)
  - 15-min stream: small part (1/60th) of 15-sec stream transmitted in near real time to outside world when buoy is in standard reporting (monitoring) mode
  - 1-min stream: averages of four consecutive BP measurements from 15-sec stream transmitted when buoy is in event reporting mode (triggered by tsunami event)

Detiding: I

- overall goal: use data from DART® buoys to estimate source coefficients needed for tsunami forecast models
- models assume absence of tides and background noise
- need to account for these two components through 'detiding'
- operational model: $\bar{\mathbf{y}} = \mathbf{x} + \alpha_1 \mathbf{g}_1 + \alpha_2 \mathbf{g}_2 + \cdots + \alpha_K \mathbf{g}_K + \mathbf{e}$
  - $\bar{\mathbf{y}}$ is vector with 1-min stream (data from DART® buoy available during a tsunami event)
  - $\mathbf{x}$ is vector representing tidal fluctuations
  - $\alpha_1, \ldots, \alpha_K$ are K source coefficients ($\alpha_k \geq 0$)
  - $\mathbf{g}_1, \ldots, \mathbf{g}_K$ are vectors derived from $K \geq 1$ unit sources that collectively model tsunami signal
  - $\mathbf{e}$ is vector representing background noise

Detiding: II

- five approaches to detiding (there are *many* more!)
  1. harmonic analysis based on 29 days of data prior to event
  2. harmonic analysis based on lots of prior data (300–1000 days)
  3. empirical orthogonal function (EOF) approach
  4. Kalman smoothing (KS) approach
  5. harmonic analysis with joint estimation of source coefficients
- first four methods predict tidal fluctuations using, say, $\hat{\mathbf{x}}$
- predictions are subtracted from $\bar{\mathbf{y}}$ to form detided data:

$$\bar{\mathbf{d}} = \bar{\mathbf{y}} - \hat{\mathbf{x}} = \alpha_1 \mathbf{g}_1 + \alpha_2 \mathbf{g}_2 + \cdots + \alpha_K \mathbf{g}_K + \boldsymbol{\epsilon},$$

  where $\boldsymbol{\epsilon} = \mathbf{e} + \mathbf{x} - \hat{\mathbf{x}}$ is error term (includes background noise and inaccuracies in predicting tidal fluctuations)
- use $\bar{\mathbf{d}}$ to estimate $\alpha_1, \ldots, \alpha_K$ via least squares

Detiding: III

- fifth method handles tidal fluctuations and estimation of source coefficients $\alpha_1, \ldots, \alpha_K$ via joint least squares ('joint method')
- will now briefly describe five methods

Method 1 (29 Day Harmonic Analysis)

- harmonic analysis is standard way to predict tides at coastal stations and assumes tides are sums of sinusoidal constituents
- for detiding DART® data, assume buoy has reported BP measurements every 15 minutes for 29 days prior to tsunami event
- model measurements y_n from 15-min stream as

$$y_n = \mu + \sum_{m=1}^{6} \left[B_m \cos(\omega_m n\,\Delta) + C_m \sin(\omega_m n\,\Delta)\right] + e_n,$$

  where parameters μ, B_m & C_m are estimated via least squares (ω_m's for N2, M2, S2, Q1, O1 & K1 and Δ are known)
- use fitted model to form detided 1-min stream:

$$d_n = \bar{y}_n - \hat{\mu} - \frac{1}{4}\sum_{k=0}^{3}\sum_{m=1}^{6}\left[\widehat{B}_m \cos(\omega_m[n+k]\,\Delta) + \widehat{C}_m \sin(\omega_m[n+k]\,\Delta)\right]$$

Method 2 (Long Harmonic Analysis)

- similar to method 1, but with following key differences
  - model now uses 68 constituents rather than just 6
  - except for mean level μ, data used to fit model are from 15-sec streams retrieved from buoy during regular servicing (totality of streams typically spans from 300 to 1000 days)
  - 29 days from 15-min stream now used just to set μ
- fitting model requires specialized software (G. Mungov, NGDC, provided fits for study discussed later on)
- cannot use method unless buoy has been serviced once in its present location

Linear Time-Invariant (LTI) Filtering

- two best known approaches for detiding are harmonic analysis and LTI high-pass filtering (e.g., Butterworth filtering)
- to isolate tsunami signal without distortion, filter should retain components with periods as long as 2 hours
- potential disadvantages
  - edge effects can significantly distort at least 1-hr sections at beginning and end of filtered series, rendering approach problematic in real-time environment
  - most LTI filters not designed to work with gappy data
- methods 3 and 4 are linear (but not LTI) filters designed to overcome these disadvantages

Method 3 (Empirical Orthogonal Functions)

- premise (Tolkova, 2010): sub-space spanned by leading empirical orthogonal functions (EOFs) of tidally dominated data segments same across all DART® buoys
- EOFs obtained from 250 segments (each spanning one lunar-day) from DART® buoy 46412 in 2007
- available buoy data from 1-min and 15-min streams projected against EOFs $\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_7$ associated with seven largest eigenvalues (along with constant vector $\mathbf{f}_0$) to obtain coefficients c_0, $c_1, \ldots, c_7$
- inverse projection using $c_0, c_1, \ldots, c_7$ yields predicted tides, which are subtracted from buoy data to yield detided data

Method 4 (Kalman Smoothing)

- Kalman smoothing (KS) widely used to 'optimally' smooth a time series, but optimality depends upon adequate model for underlying dynamics
- KS approach here is two-stage procedure
  1. use 29 day harmonic analysis (method 1) to obtain first-stage detided series, say, d_n
  2. use d_n as input to KS based upon local level model (also known as 'random walk plus noise' model') – output from smoother intended to track any tidal component/background noise left over from first-stage detiding
- local level model depends upon just two parameters and estimate of initial state of underlying dynamical system

Method 5 (Joint Method): I

- joint method estimates tidal component along with source coefficients using just 1-min stream arising during tsunami event
- joint method is based on operational model

$$\bar{\mathbf{y}} = \mathbf{x} + \alpha_1 \mathbf{g}_1 + \alpha_2 \mathbf{g}_2 + \cdots + \alpha_K \mathbf{g}_K + \mathbf{e}$$

  but with addition of specific model for tidal fluctuations:

$$\mathbf{x} = \mu \mathbf{1} + \sum_{m=1}^{2} \left(B_m \mathbf{c}_m + C_m \mathbf{s}_m \right),$$

  where $\mathbf{1}$ is a vector of ones; $\mathbf{c}_m$ is a vector with elements $\cos(\omega_m n \Delta)$, where ω_2 is tidal frequency M2 and ω_1 is average of O1 and K1 frequencies ($\Delta = 1$ min); $\mathbf{s}_m$ is analogous to $\mathbf{c}_m$, but with sines replacing cosines; $\mathbf{e}$ is a vector of errors (zero means and a common variance)

Method 5 (Joint Method): II

- unknown parameters $\alpha_1, \alpha_2, \ldots, \alpha_k$ μ, B_1, C_1, B_2 and C_2 estimated jointly via least squares
- can take detided series for this method to be

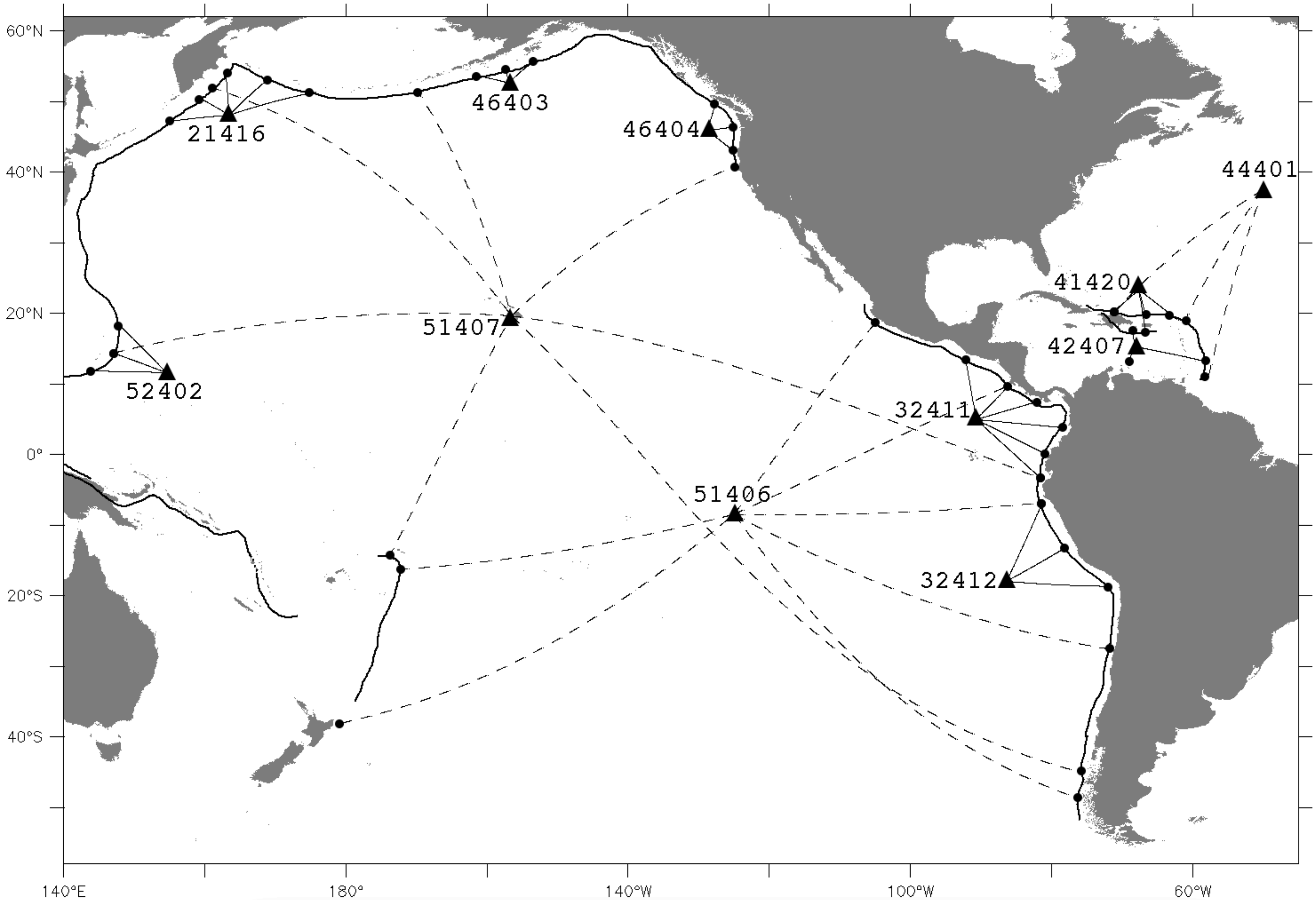
$$\bar{\mathbf{d}} = \bar{\mathbf{y}} - \hat{\mu}\mathbf{1} - \sum_{m=1}^{2} \left(\widehat{B}_m \mathbf{c}_m + \widehat{C}_m \mathbf{s}_m \right),$$

  where $\hat{\mu}$ is least squares estimate of μ etc.
- note that, in contrast to previous four methods, joint method does *not* make use of any data prior to tsunami event – just uses available 1-min stream

Assessing Performance of Five Methods: I

- recall overall goal: use data from DART® buoys to estimate source coefficients of tsunami forecast models
- Q: how do estimated source coefficients compare for five detiding methods?
- to address question, carried out study using archived 15-sec streams from eleven representative DART® buoys (streams ranged in length from 321 to 998 days)

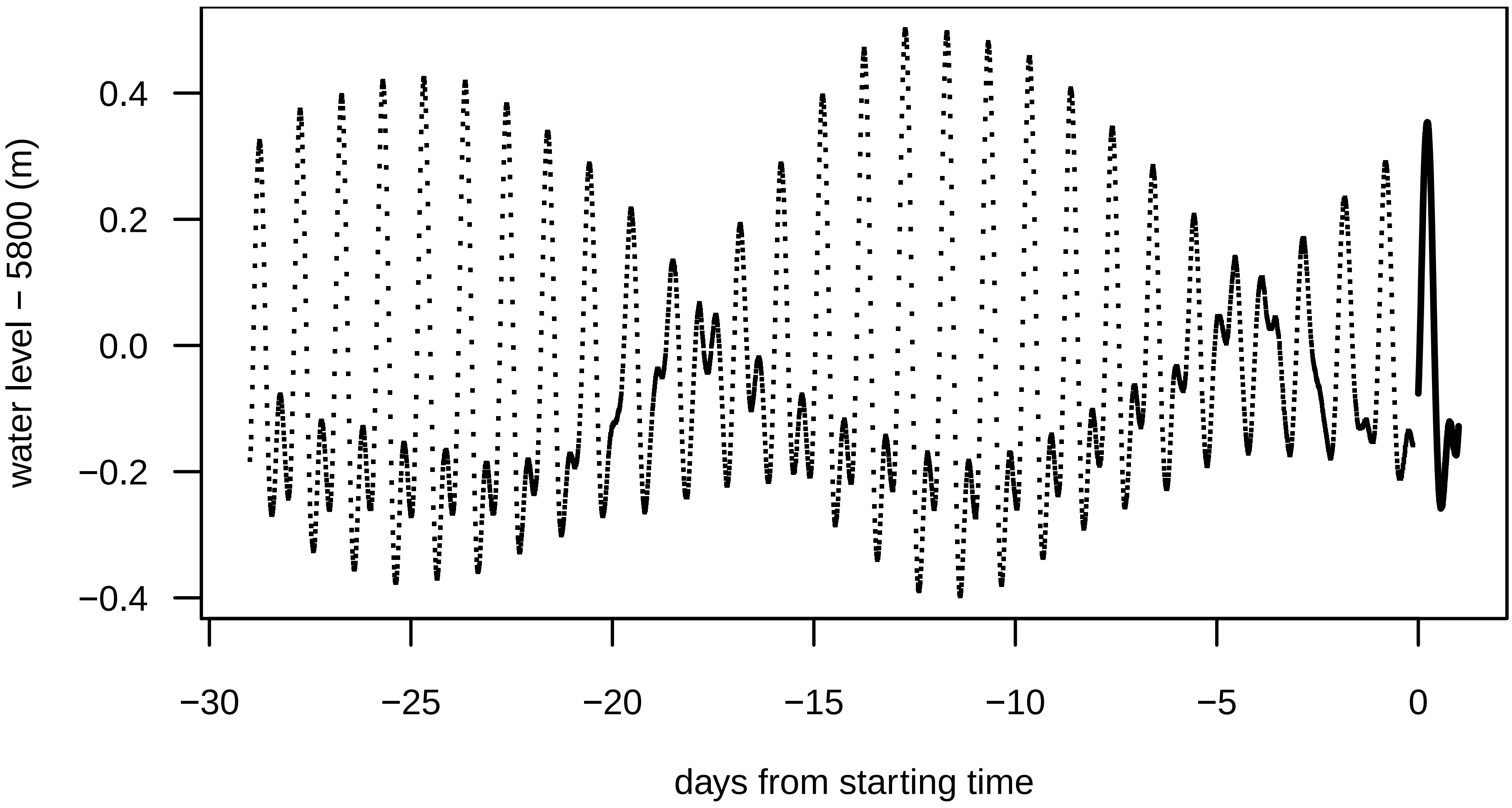
Locations of Eleven DART® Buoys (Triangles)



Assessing Performance of Five Methods: II

- operational model: $\bar{\mathbf{y}} = \mathbf{x} + \alpha_1 \mathbf{g}_1 + \alpha_2 \mathbf{g}_2 + \cdots + \alpha_K \mathbf{g}_K + \mathbf{e}$
- used archived 15-sec streams to construct 'scenarios' that mimic tidal fluctuations and background noise (i.e., $\mathbf{x} + \mathbf{e}$) present in 15-min & 1-min streams recorded during actual tsunami event
- procedure for constructing one scenario for a particular buoy
  - select random starting time t_0
  - form 29-day segment of 15-min stream by subsampling 15-sec stream prior to t_0 (to mimic operational conditions, create 3-h gap just prior to t_0 in constructed 15-min stream)
  - form 1-day segment of 1-min stream by averaging 4 adjacent values of 15-sec stream after t_0
- constructed 15-min & 1-min streams form one of 1000 scenarios

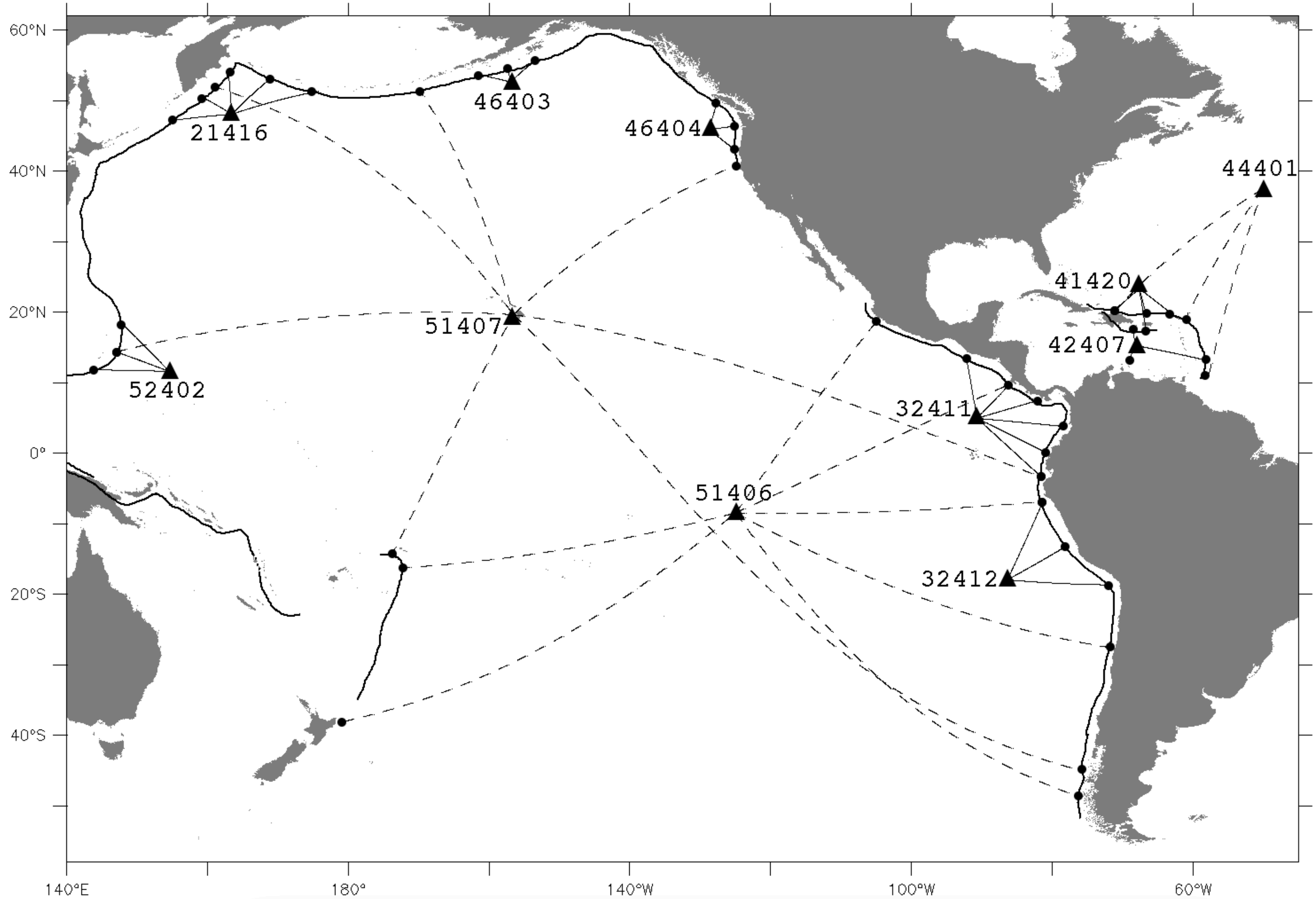
Scenario 943 for Buoy 52402 (t_0 = 9:21:00 UT, 6/27/07)



Assessing Performance of Five Methods: III

- to create an artificial tsunami signal, simplify operational model to have just one unit source: $\bar{\mathbf{y}} = \mathbf{x} + \alpha\mathbf{g} + \mathbf{e}$
- motivated by actual tsunami events, set $\alpha = 6$ as representative source coefficient
- for each buoy, set $\mathbf{g}$ based from three to seven unit sources with different orientations with respect to buoy
  - for example, buoy 52402 is paired with three unit sources (from north to south, ki050b, ki055b and ki060b)
- 42 unit sources in all, five of which were used by two buoys, for a total of 47 pairings of buoys and unit sources (each pairing leads to a different artificial tsunami signal)
- artificial tsunami signal for given buoy/unit source pairing added to each of 1000 scenarios for given buoy

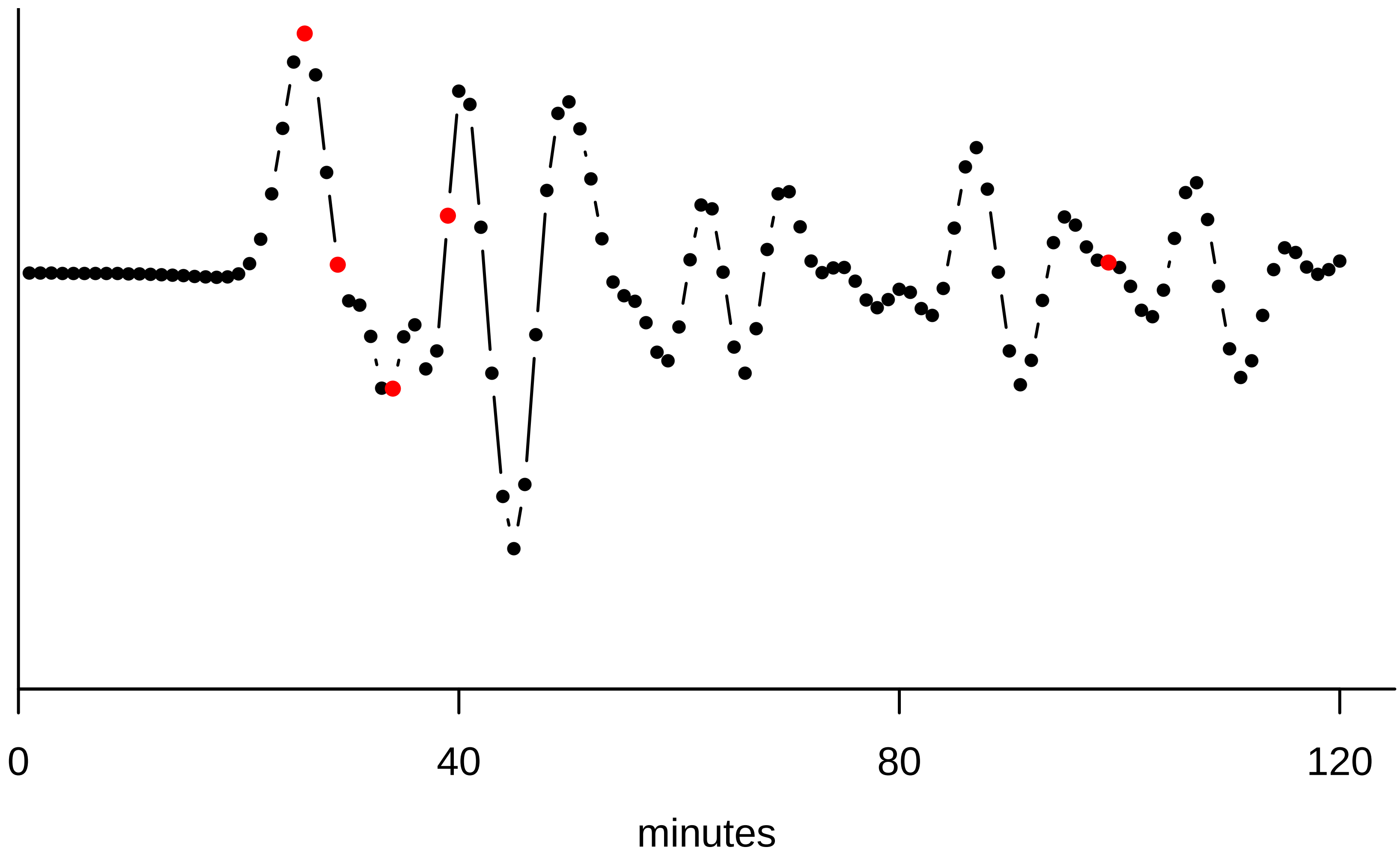
Locations of 11 DART® Buoys and 42 Unit Sources



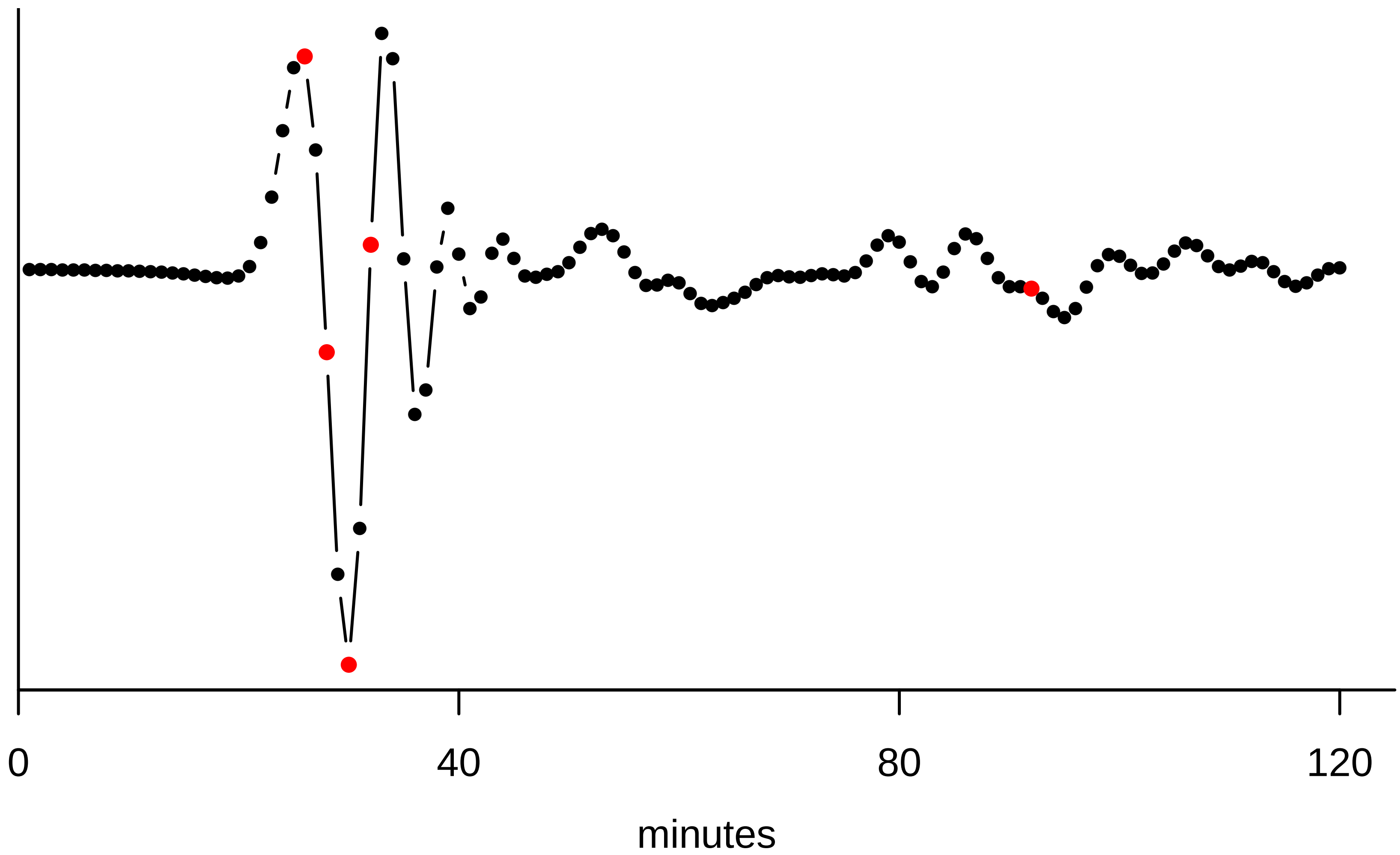
Assessing Performance of Five Methods: IV

- next set of plots show unit sources selected for buoy 52402 (ki050b, ki055b and ki060b)
- five red dots mark arrival times of
  - first quarter wave
  - half wave
  - three-quarters wave
  - first full wave
  - one hour beyond first full wave

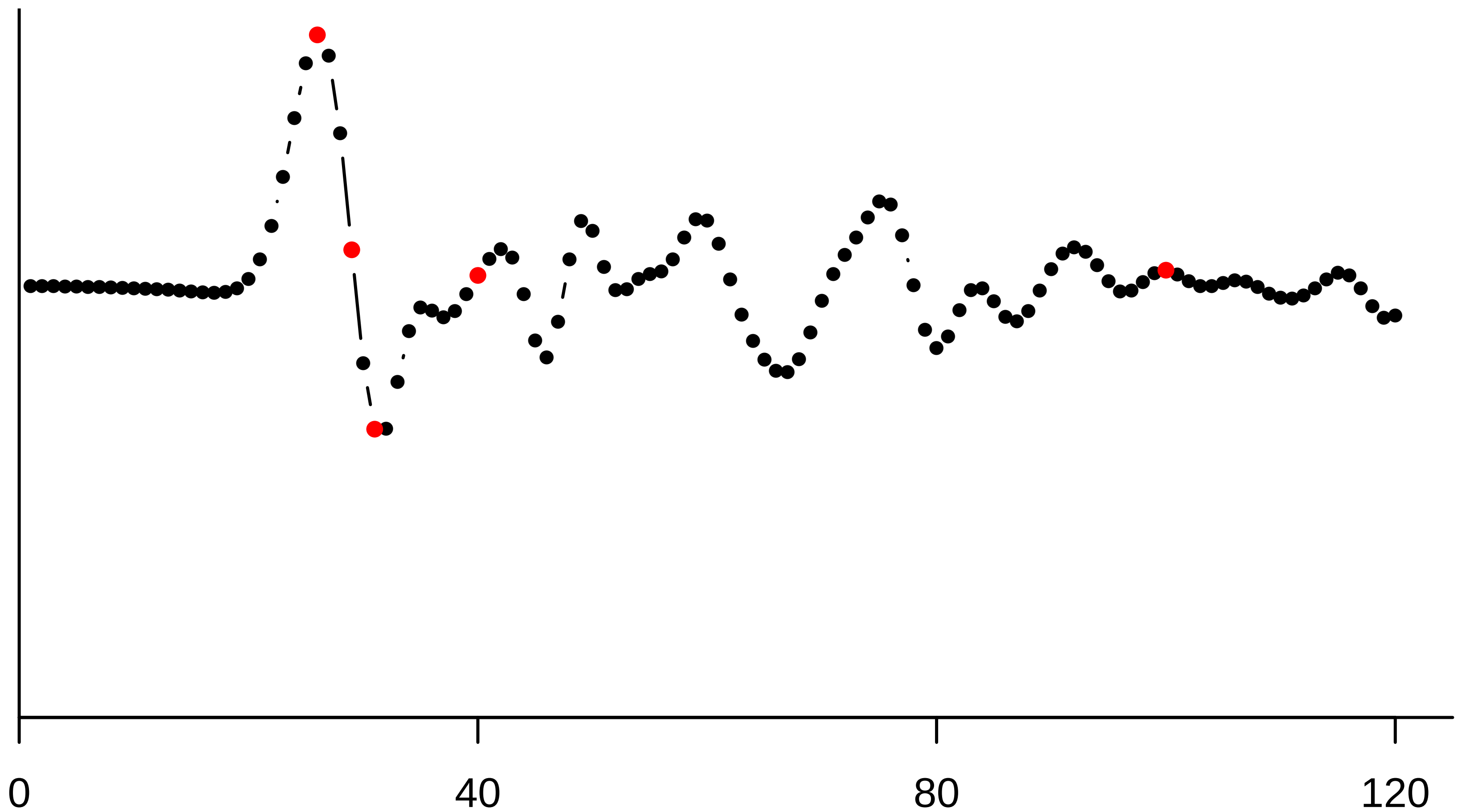
Unit Source ki050b for Buoy 52402



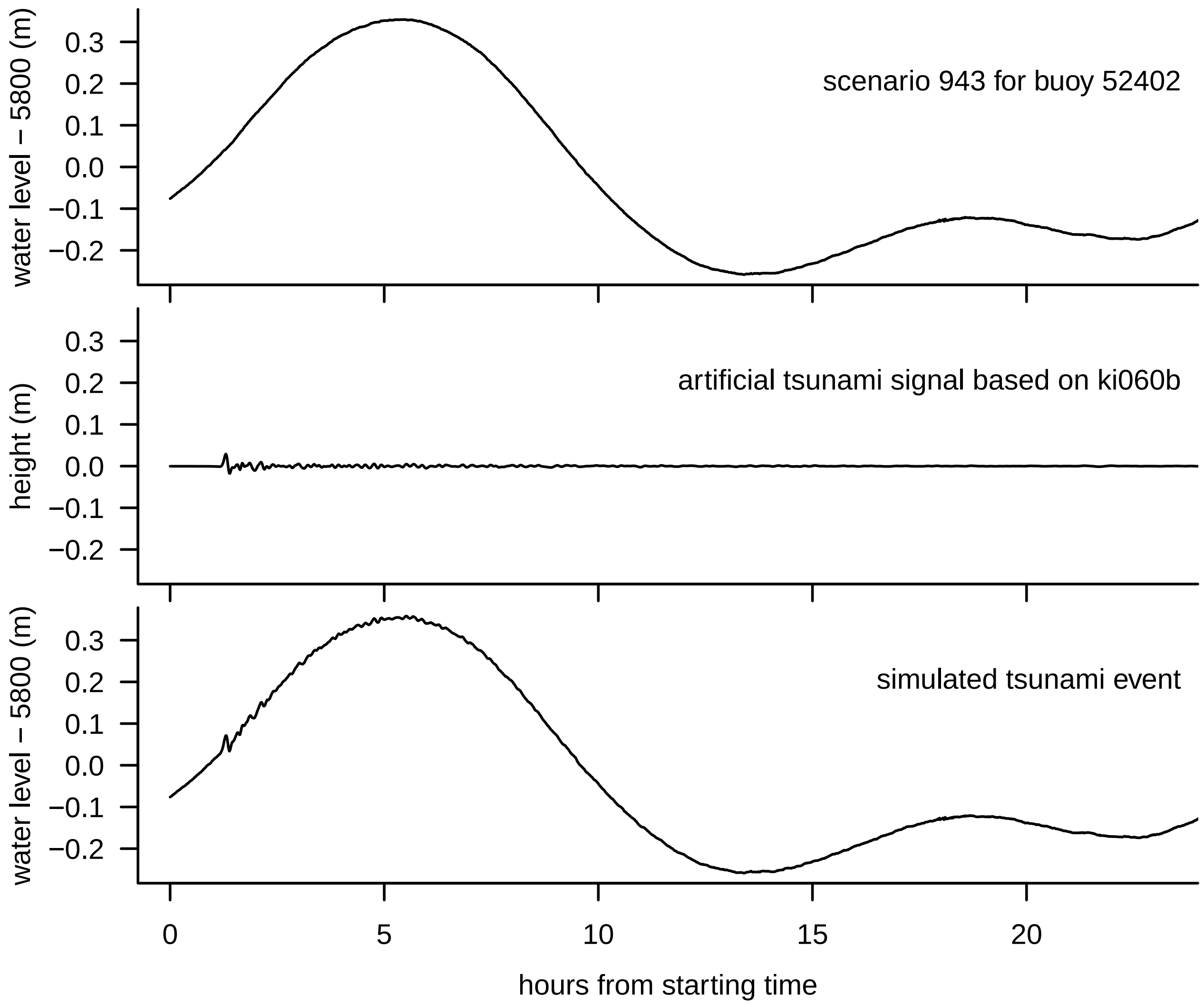
Unit Source ki055b for Buoy 52402



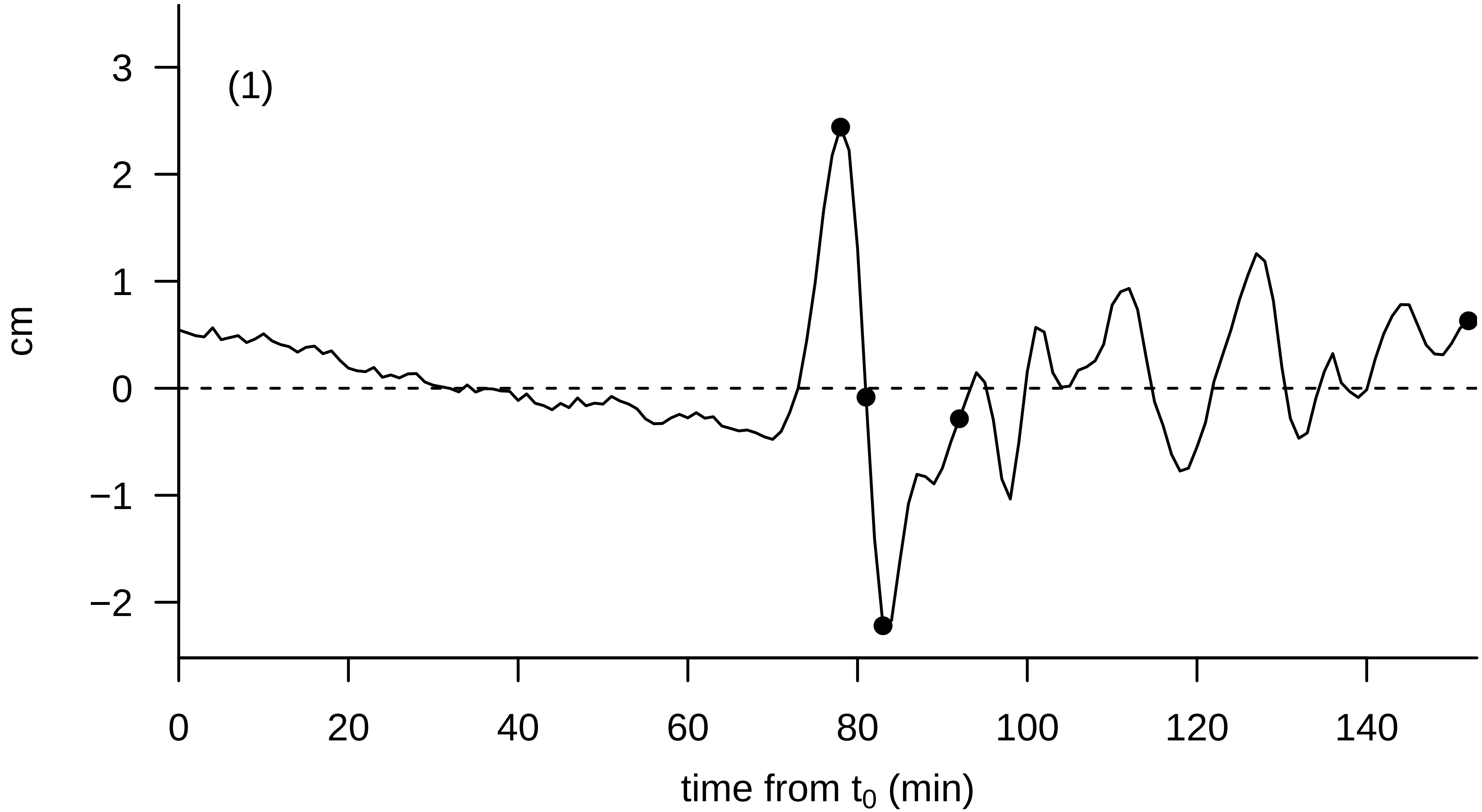
Unit Source ki060b for Buoy 52402



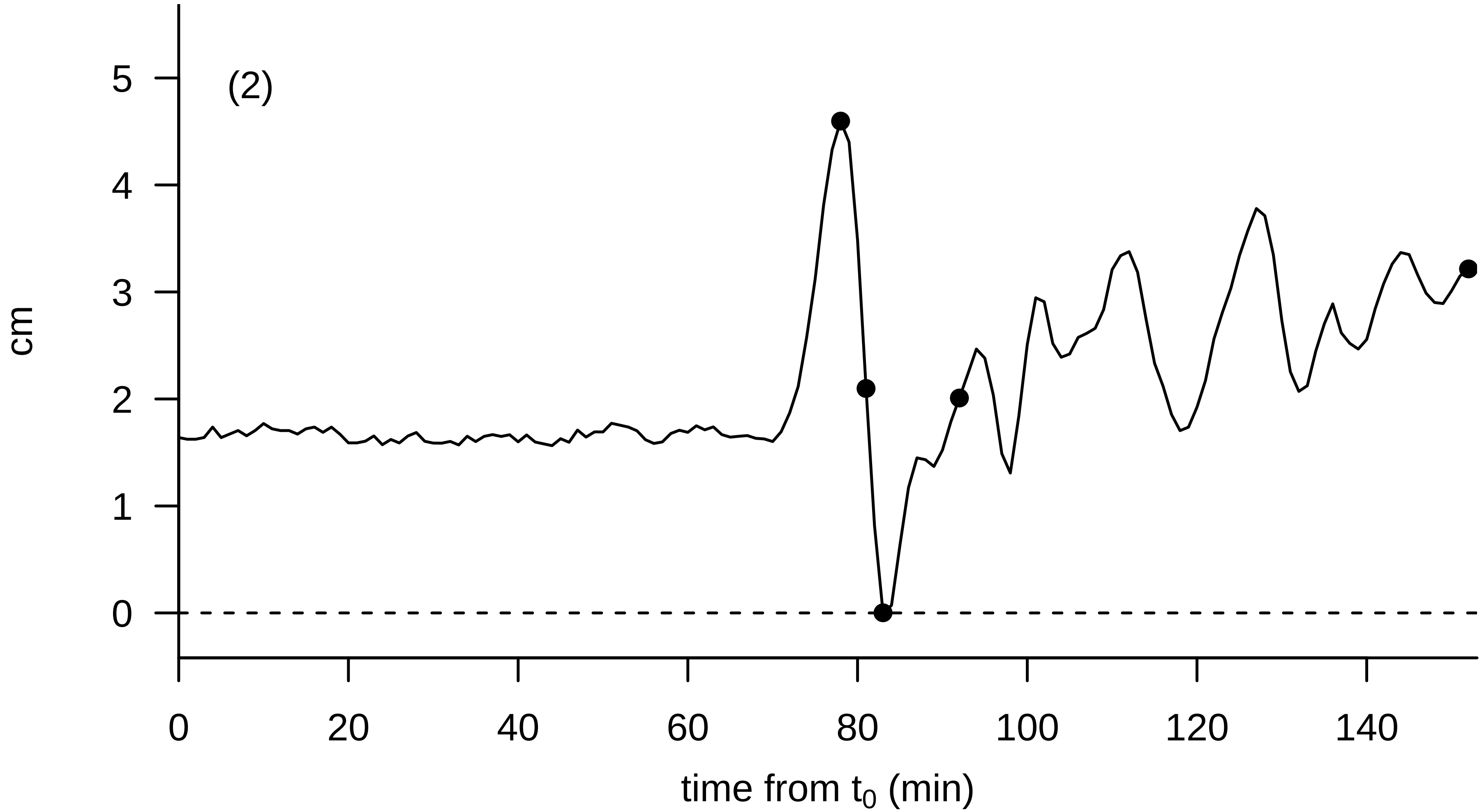
Construction of Simulated Tsunami Event



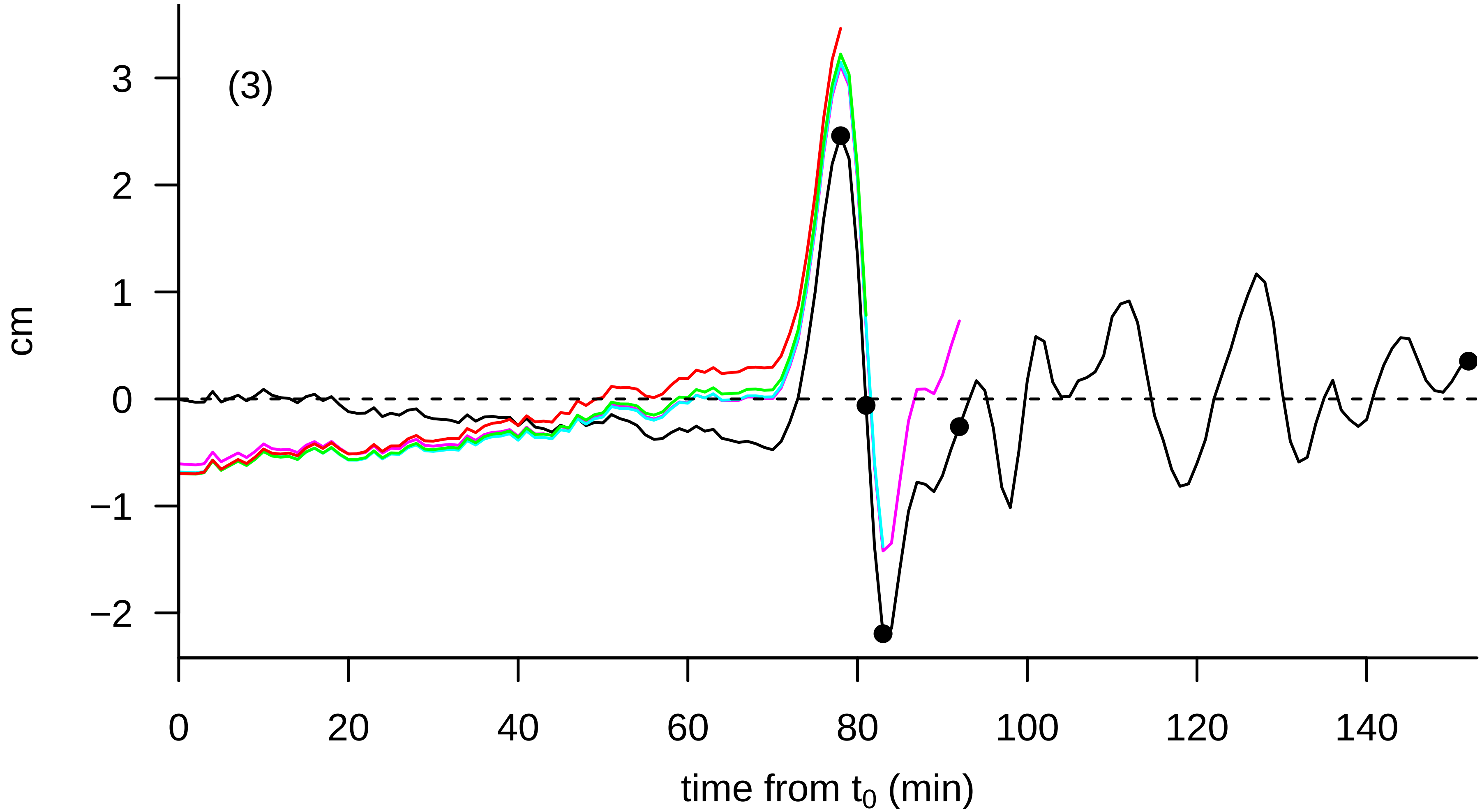
Detiding Using 29 Day Harmonic Analysis



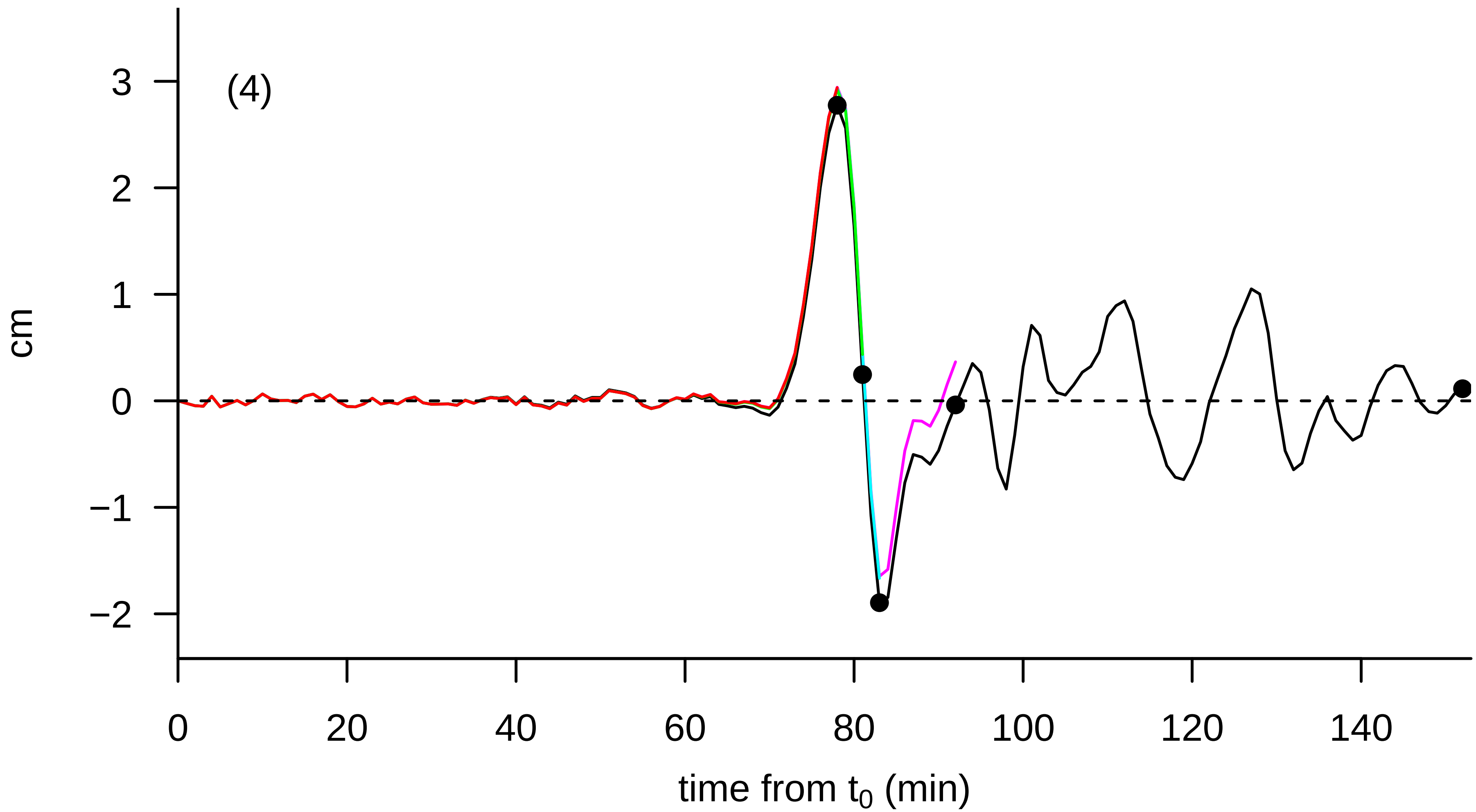
Detiding Using Long Harmonic Analysis



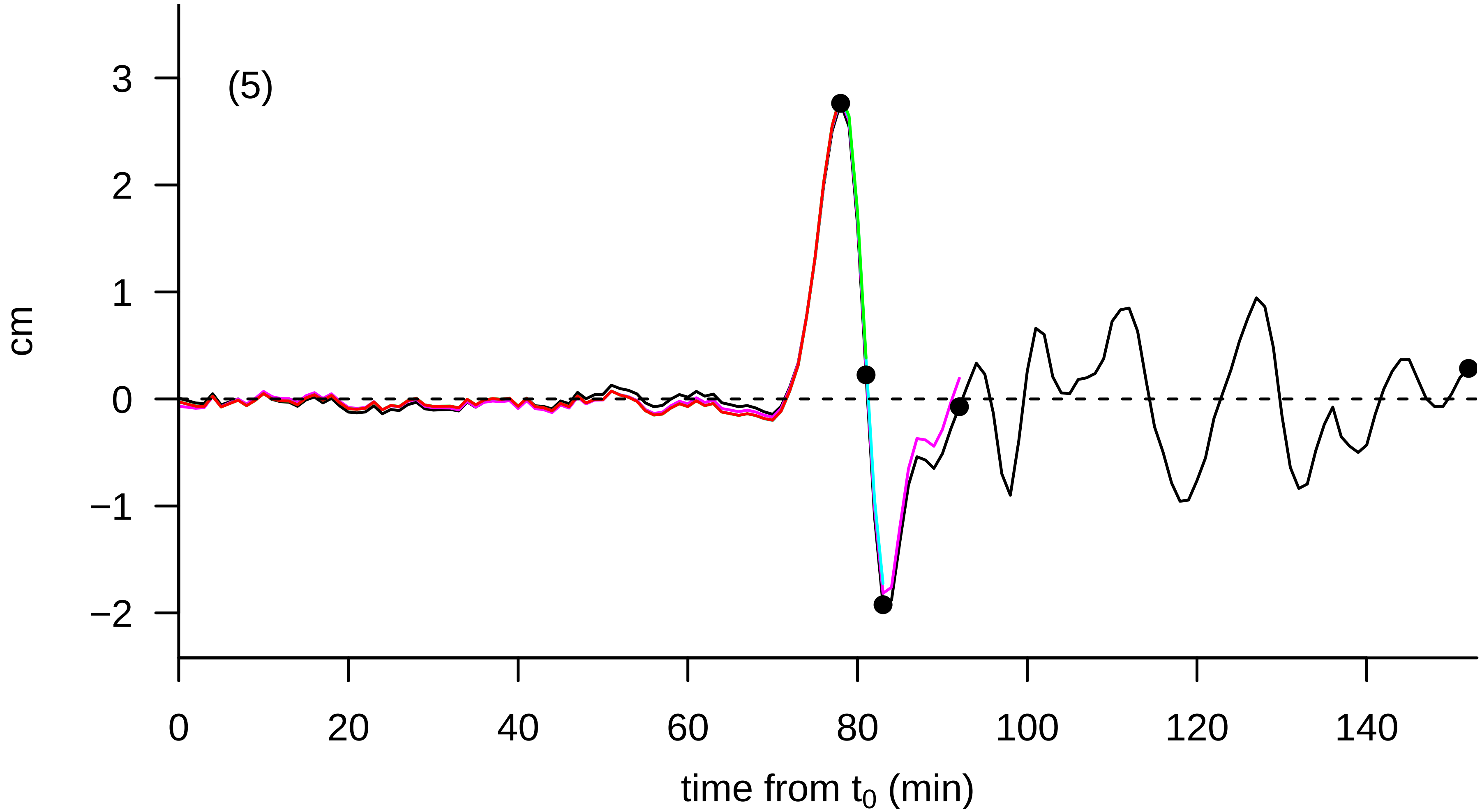
Detiding Using Empirical Orthogonal Functions



Detiding Using Kalman Smoothing



Detiding Using Joint Method



Assessing Performance of Five Methods: V

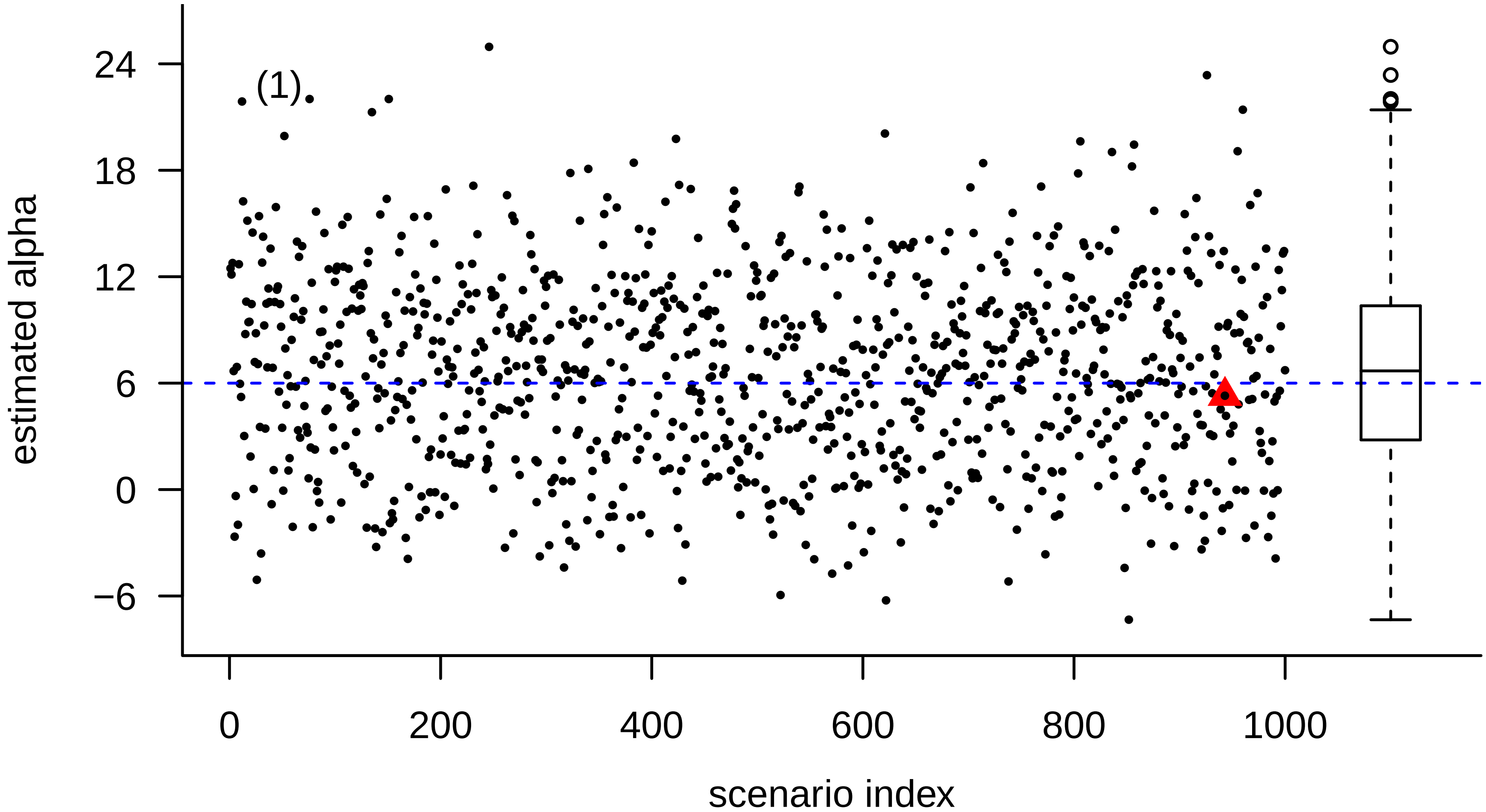
- here are $\hat{\alpha}$'s for five methods using different amount of data from scenario 943 for buoy 52402 with artificial tsunami based on unit source ki060b (recall that true α is 6)

method	1/4	1/2	3/4	full	full+1 hour
29 day HA	5.04	4.95	5.28	5.71	5.69
Long HA	9.25	9.77	8.55	6.90	6.28
EOF	7.72	7.16	6.70	6.24	5.77
KS	6.29	6.22	6.18	6.06	5.88
Joint Method	6.03	6.02	5.98	5.99	6.02

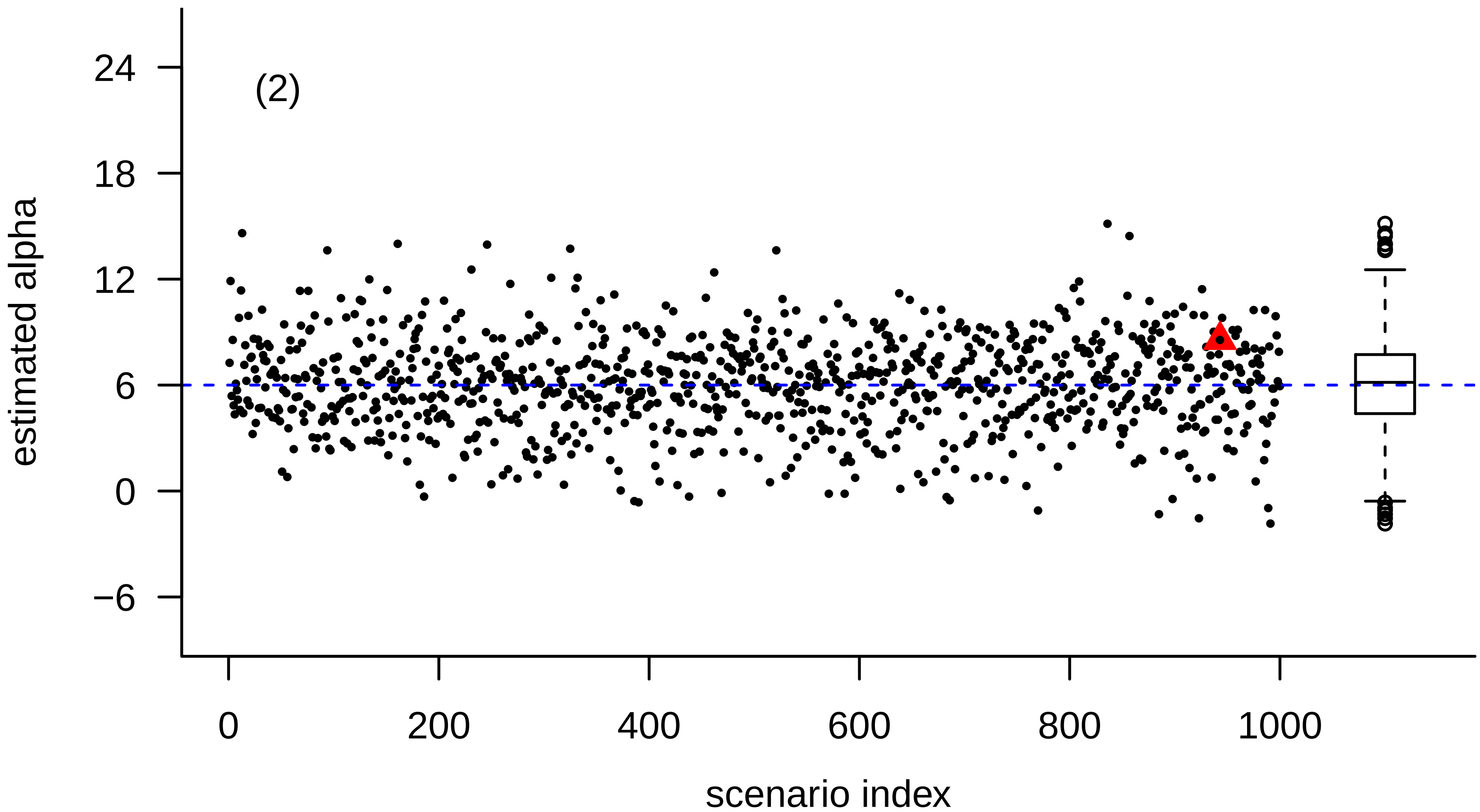
Assessing Performance of Five Methods: VI

- next set of plots show $\hat{\alpha}$'s for five methods using
  - data up to 3/4 of first full wave
  - all 1000 scenarios for buoy 52402
  - artificial tsunami based on unit source ki060b
- triangles mark scenario 943
- box plots summarize distribution of $\hat{\alpha}$'s
  - central box depicts lower quartile, median and upper quartile
  - upper/lower hinges indicate values closest to (but not more extreme than) 1.5 times interquartile distance (upper quartile minus lower quartile)
  - points more extreme than hinges indicated by circles

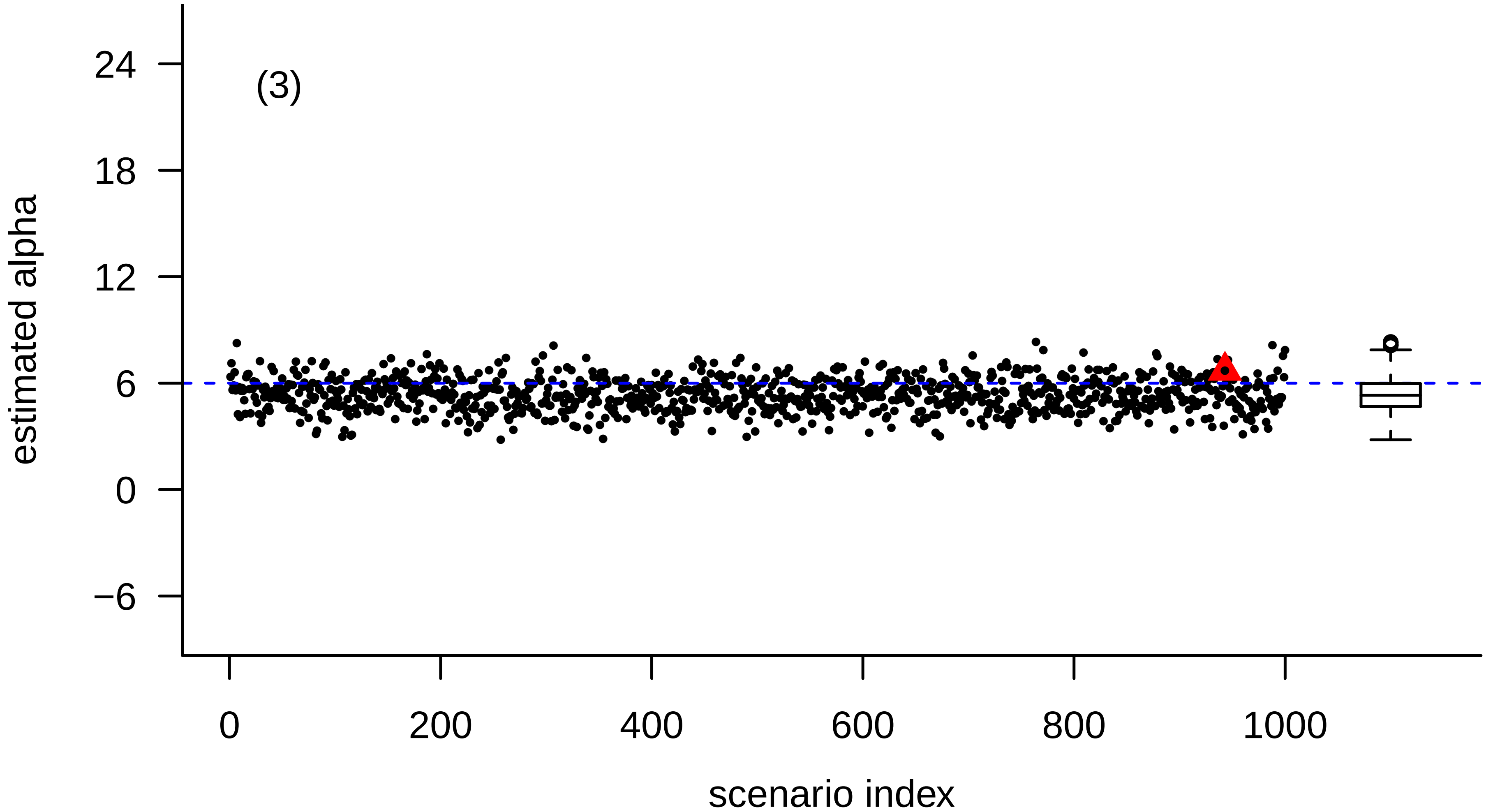
$\hat{\alpha}$'s Using 29 Day Harmonic Analysis



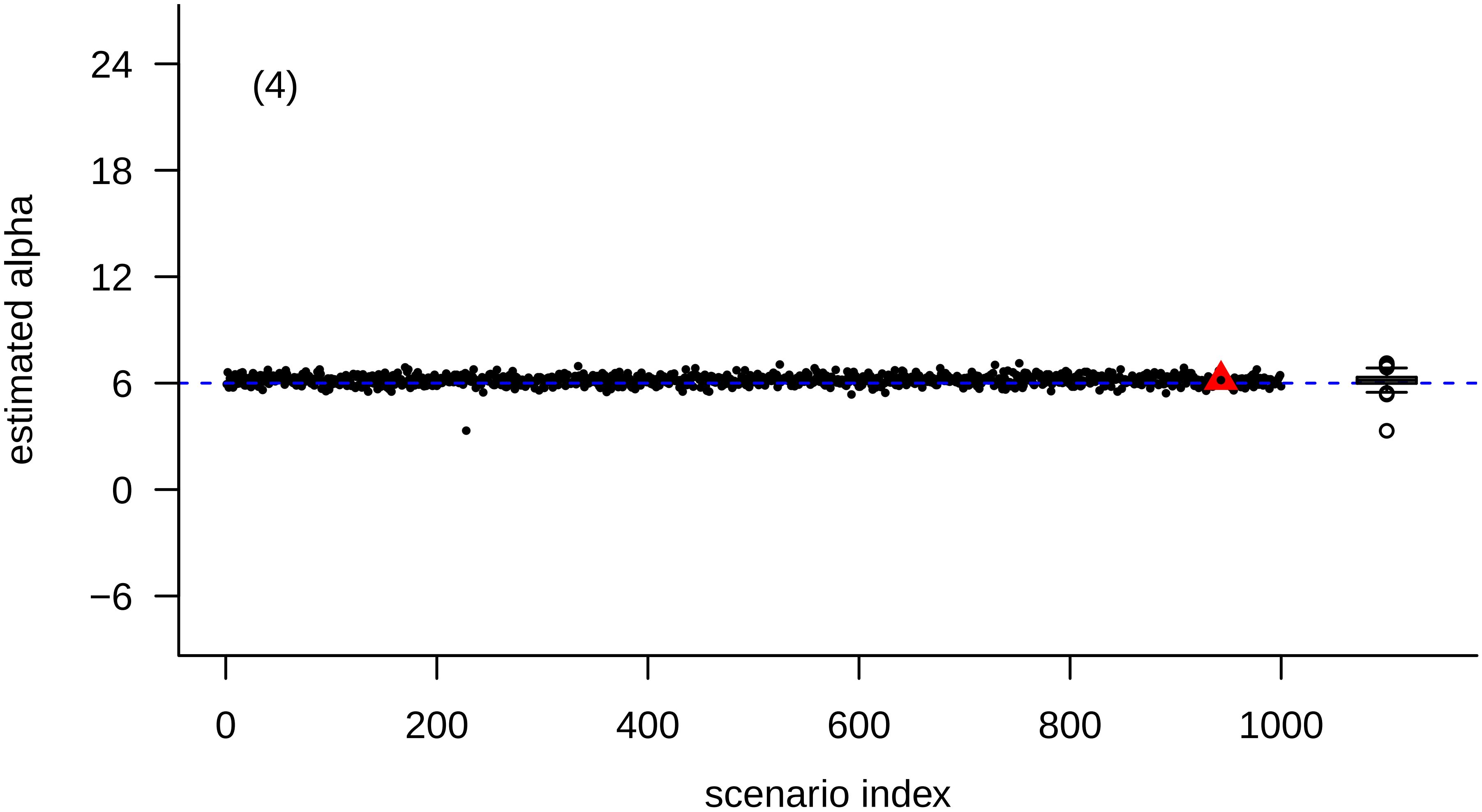
$\hat{\alpha}$'s Using Long Harmonic Analysis



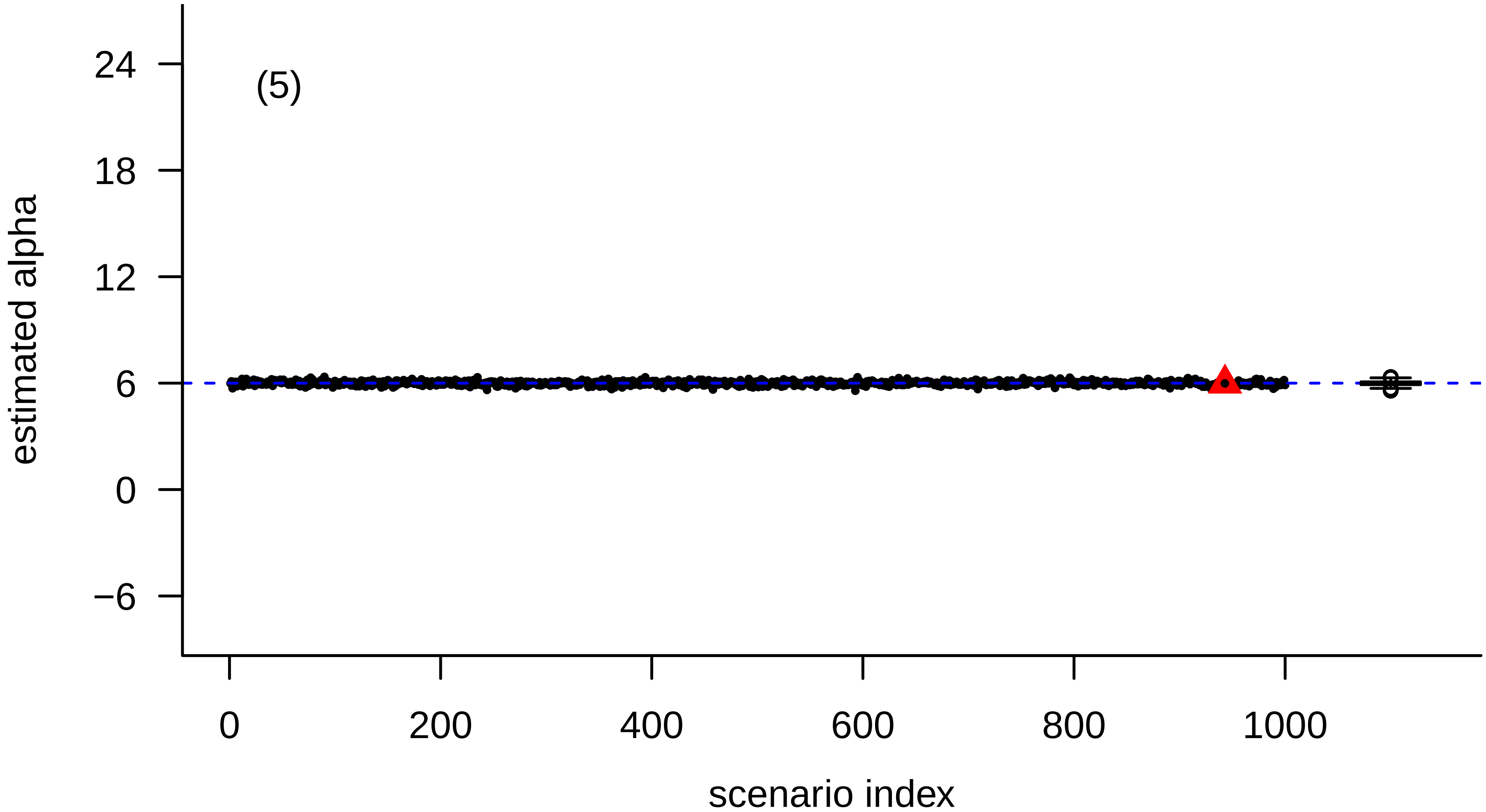
$\hat{\alpha}$'s Using Empirical Orthogonal Functions



$\hat{\alpha}$'s Using Kalman Smoothing



$\hat{\alpha}$'s Using Joint Method



Assessing Performance of Five Methods: VII

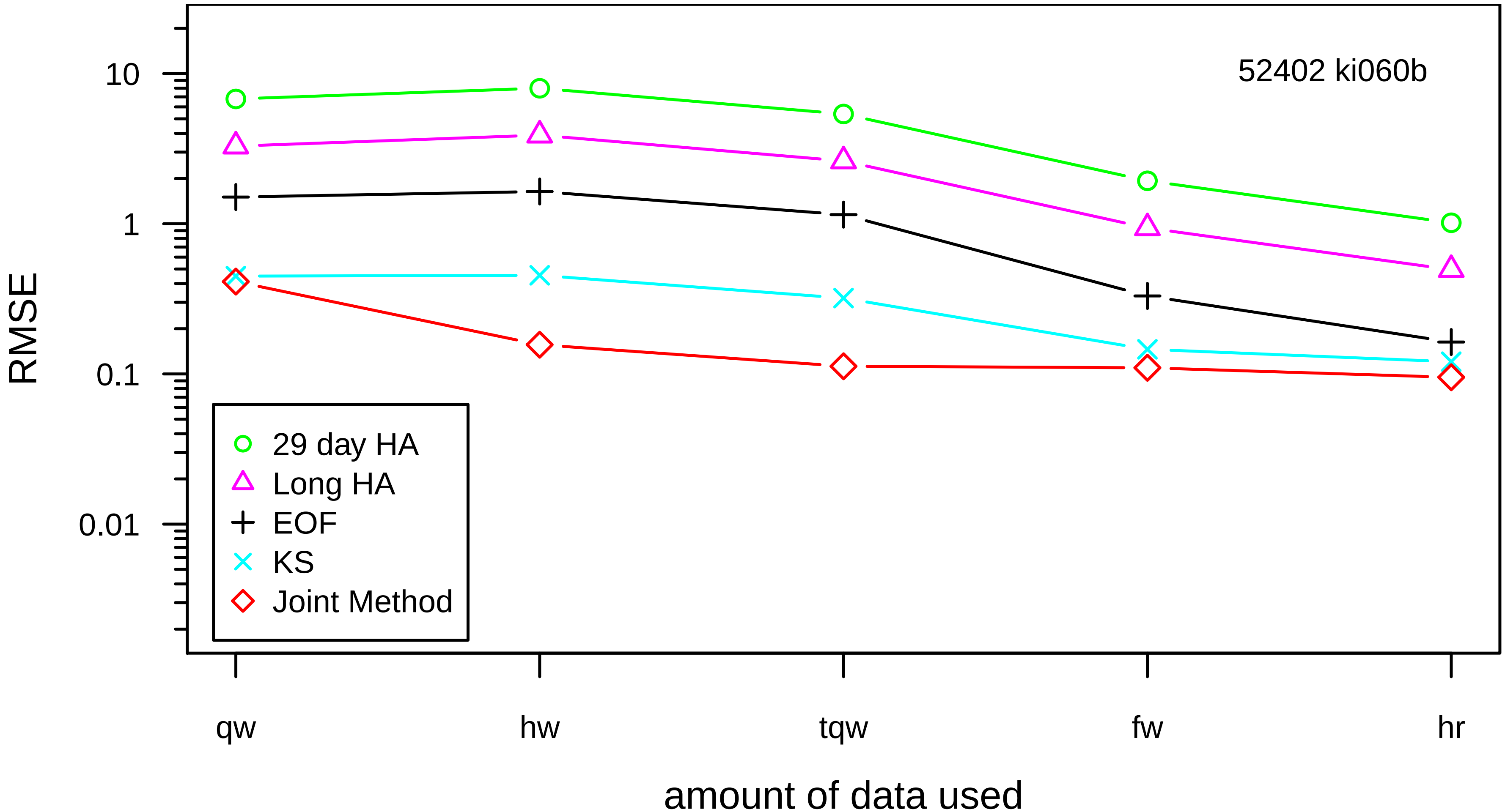
- rather than summarizing distribution of $\hat{\alpha}$'s using box plots, will now concentrate on root-mean-square errors (RMSEs) as a single measure:

$$\text{RMSE} = \sqrt{\frac{1}{1000} \sum_{n=1}^{1000} (\hat{\alpha}_n - 6)^2}$$

  where $\hat{\alpha}_n$ is estimate of α obtained from nth scenario (recall that true α is 6)

- next plot shows RMSEs for all five detiding methods versus amount of data utilized for pairing of buoy 52402 with artificial tsunami based on unit source ki060b

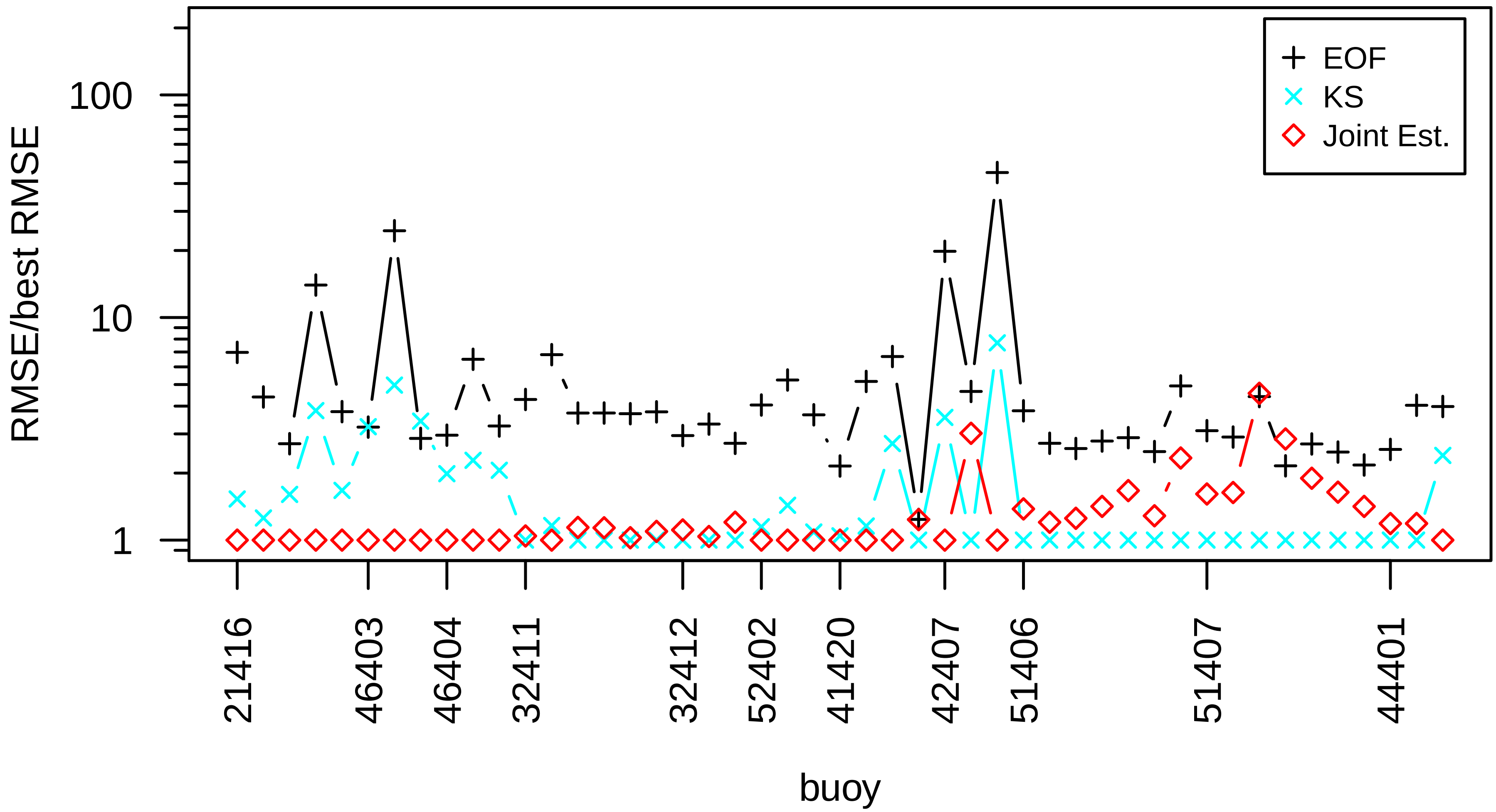
Root-Mean-Square Errors for 1000 $\hat{\alpha}$ Estimates



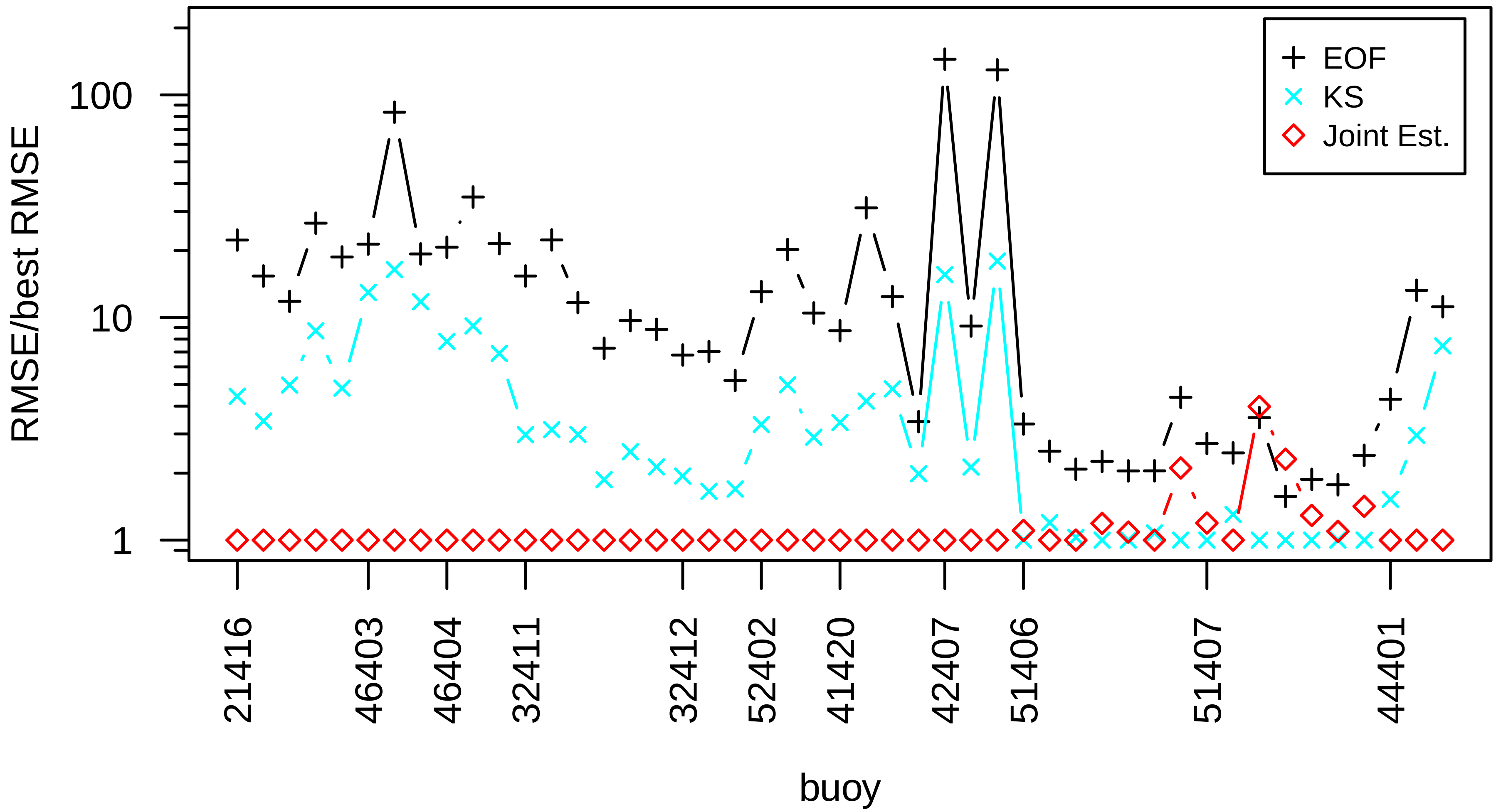
Assessing Performance of Five Methods: VIII

- comments about plot
  - note use of logarithmic scale
  - RMSEs generally decrease as amount of data increases
  - substantial drop in RMSEs going from first quarter wave to first full wave
  - RMSEs for best & worst methods (29 day harmonic analysis & joint method) differ by about an order of magnitude
  - first two methods (29 day and long harmonic analyses) not competitive with other three methods
- next set of plots look at ratios of RMSEs to best RMSE for EOF, KS and joint methods for three amounts of data and for all 47 pairings of buoys and unit sources

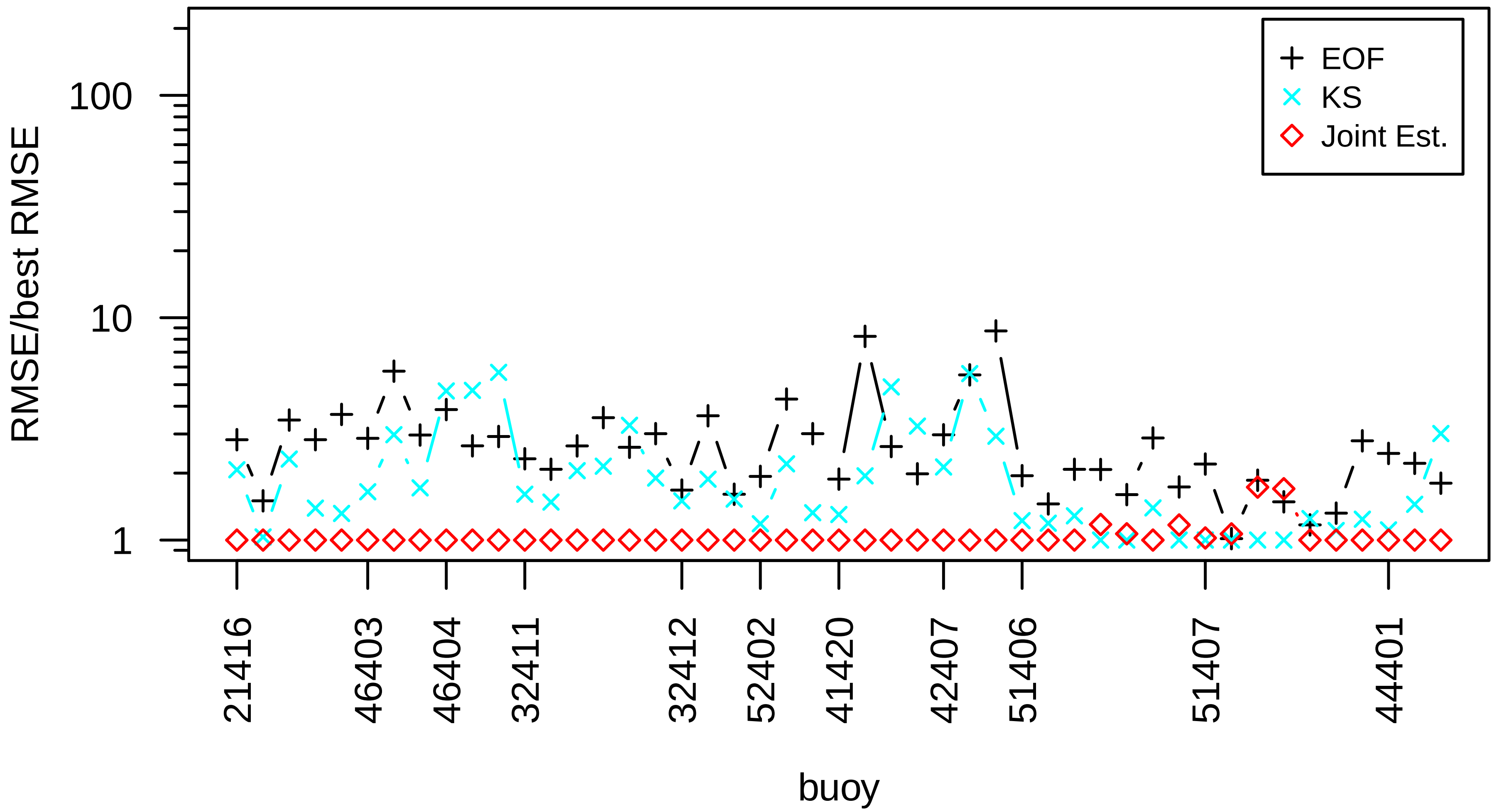
RMSE to Best RMSE Ratio Using 1/4 Wave



RMSE to Best RMSE Ratio Using 1/2 Wave



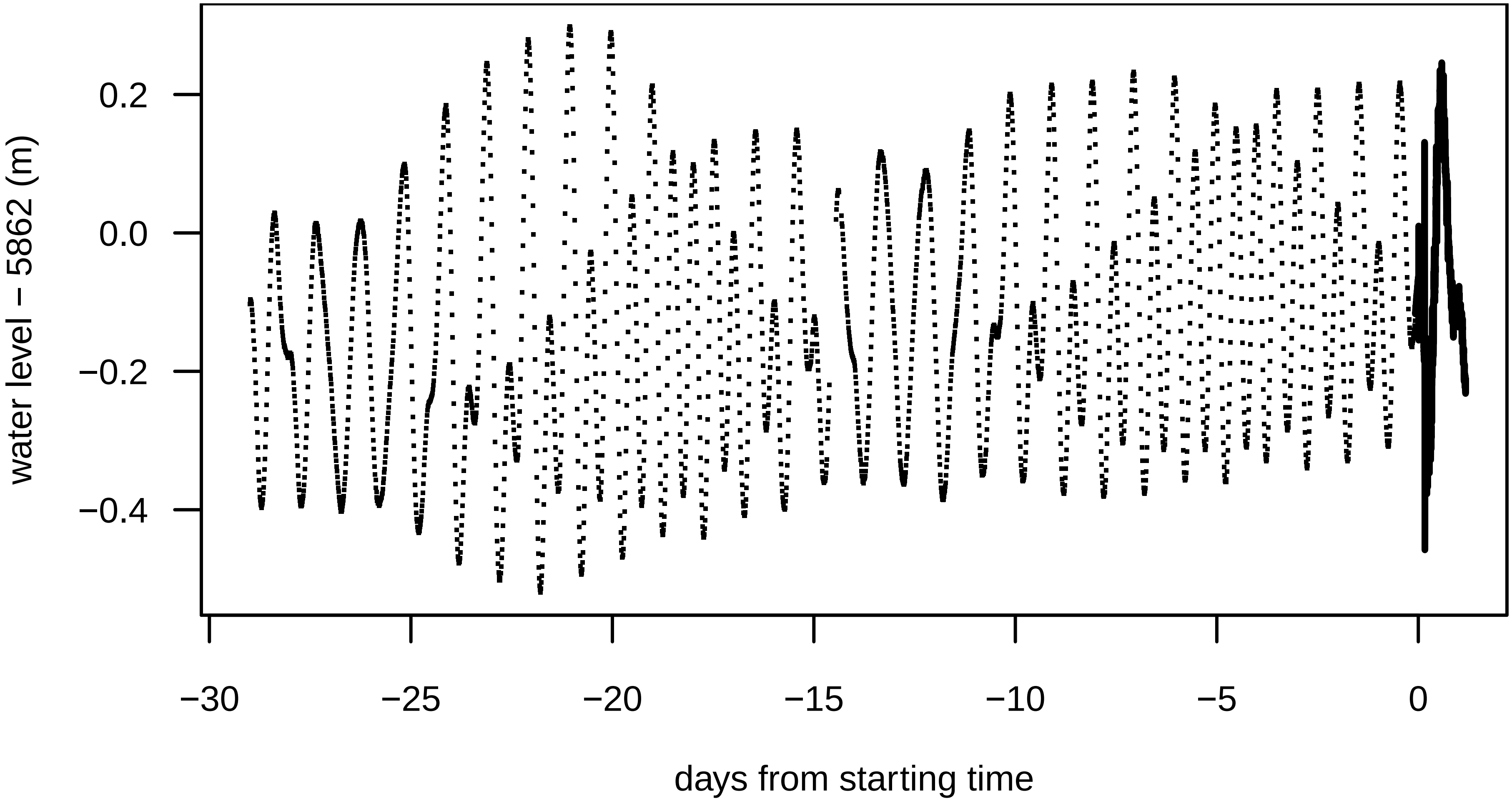
RMSE to Best RMSE Ratio Using Full Wave



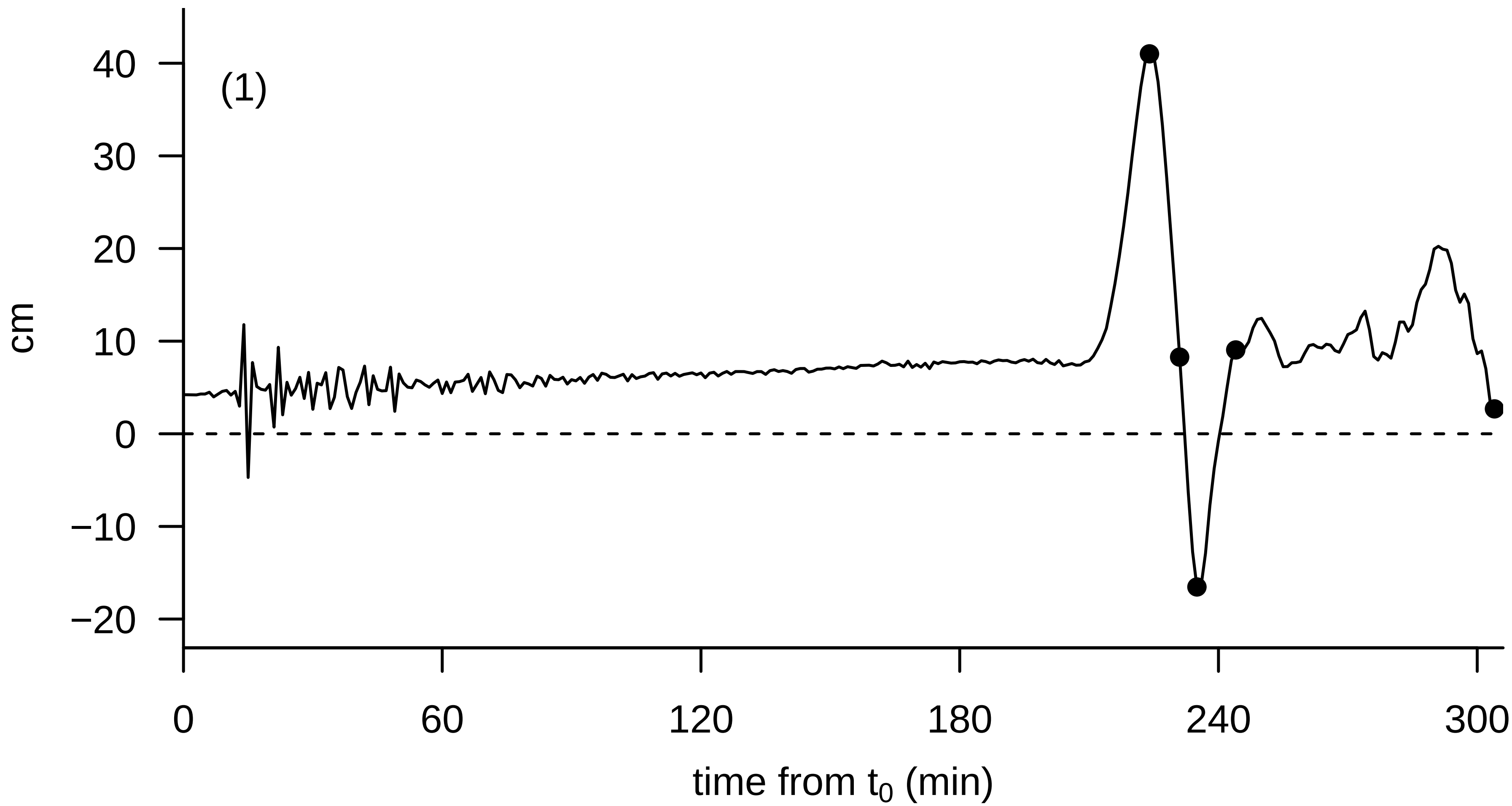
Assessing Performance of Five Methods: IX

- overall conclusion: joint method works best
  - note: pairings for which joint method is bested tend to have long lead-up to tsunami signal – joint method might be improved with appropriate windowing of data
- time permitting: consider real-world example based on March 2011 Japan tsunami

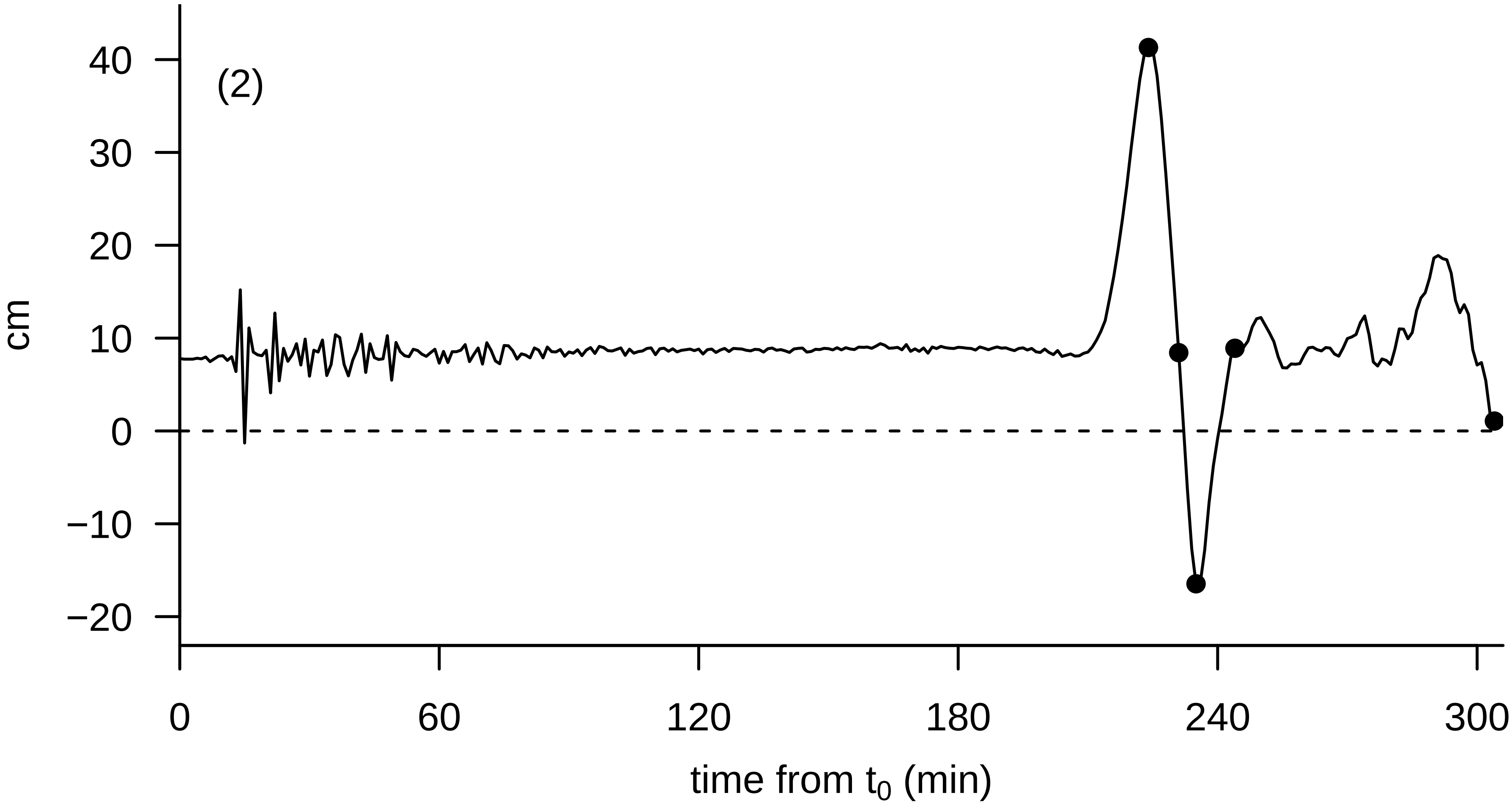
Data from Buoy 52402 during March 2011 Japan Tsunami (t_0 = 5:46:23 UT, 3/11/11)



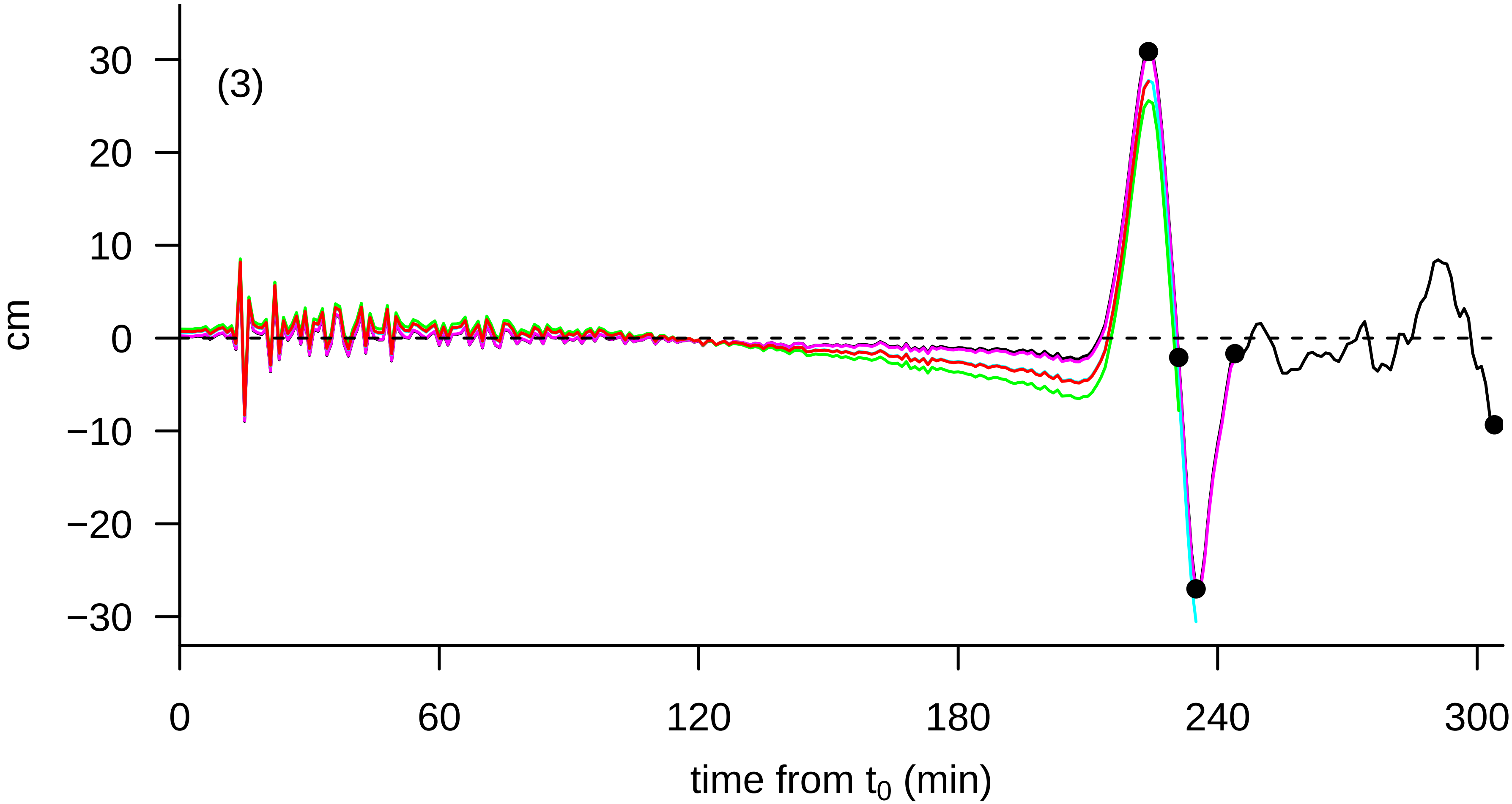
Detiding Using 29 Day Harmonic Analysis



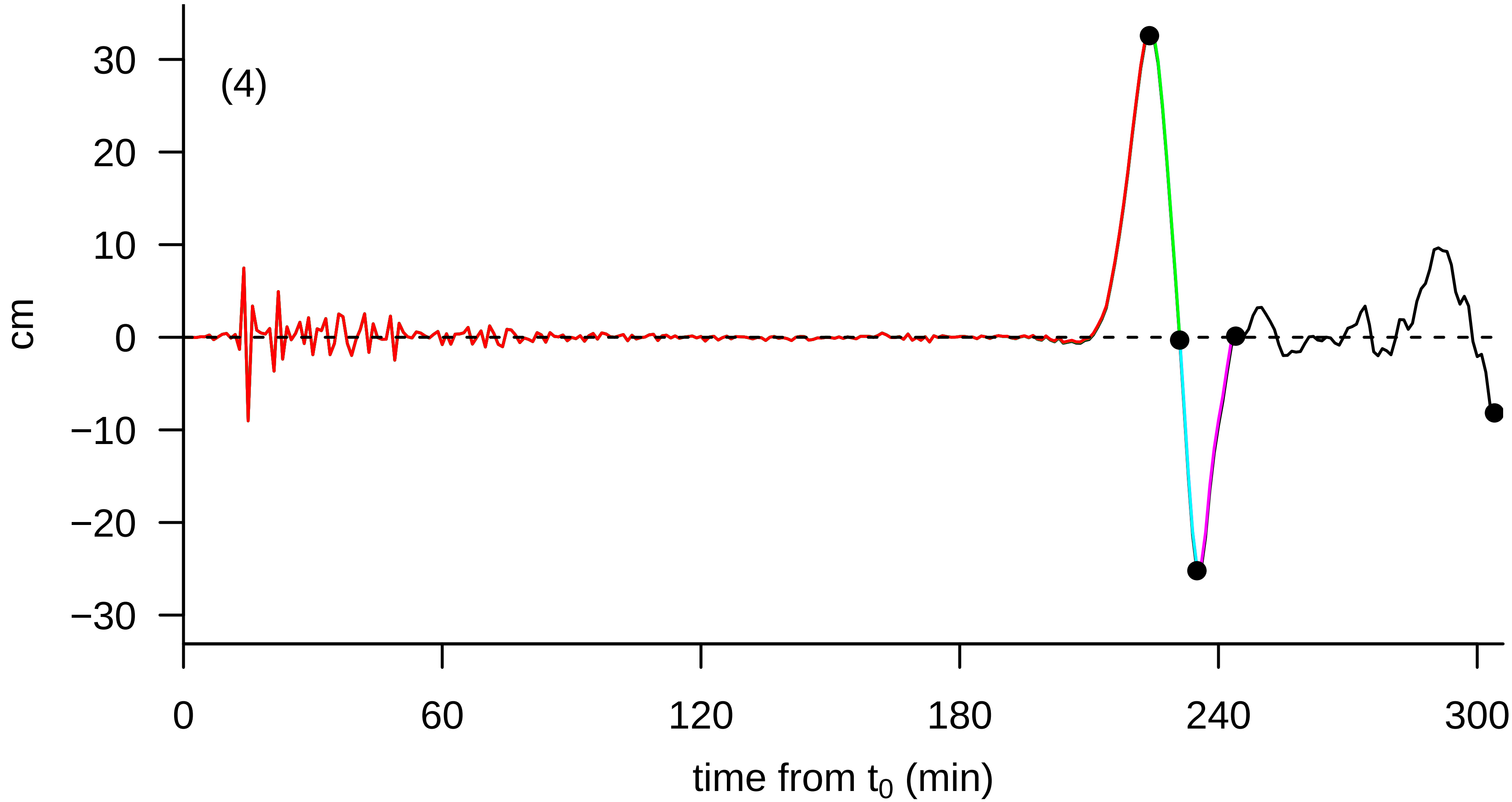
Detiding Using Long Harmonic Analysis



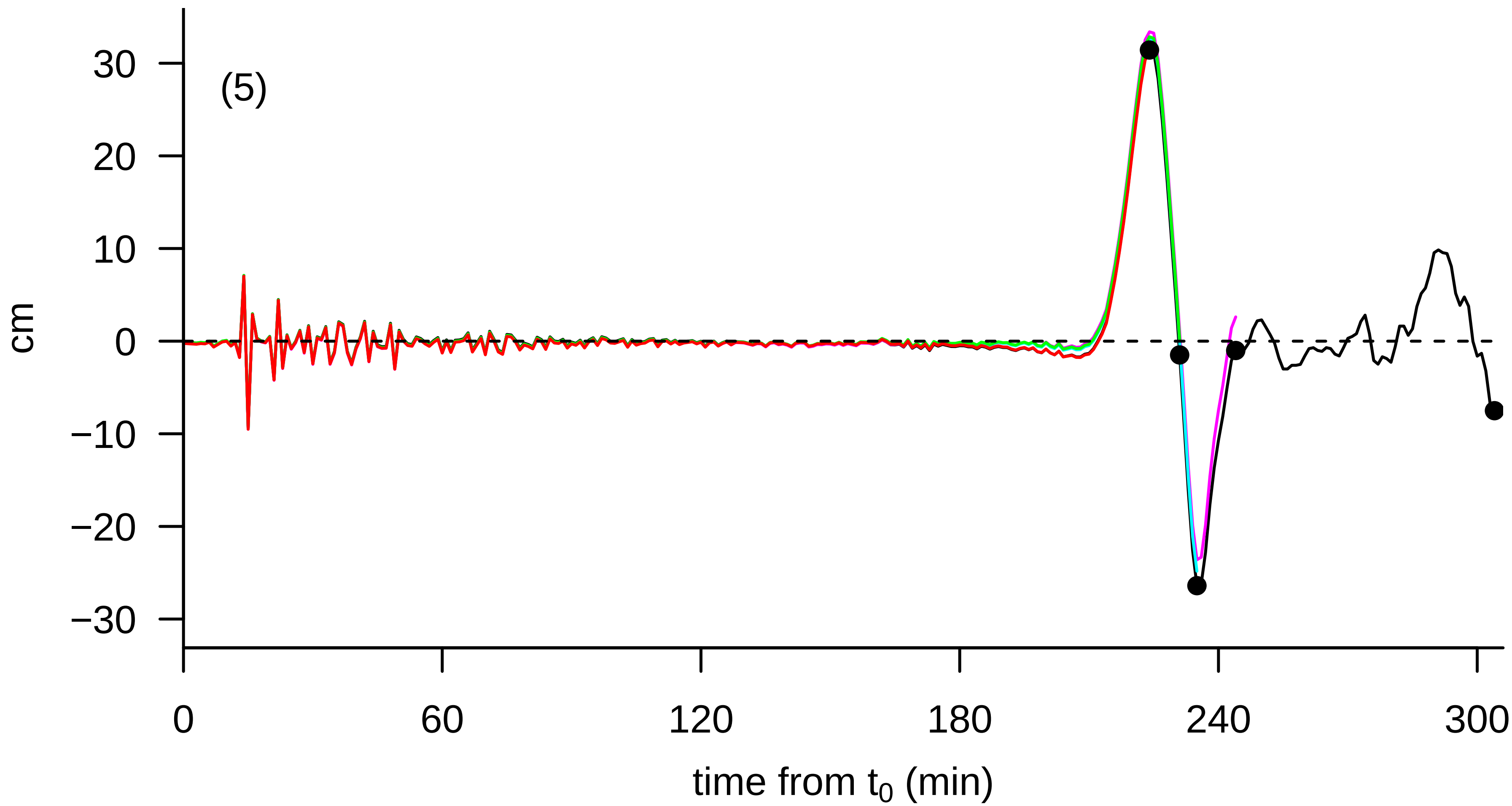
Detiding Using Empirical Orthogonal Functions



Detiding Using Kalman Smoothing



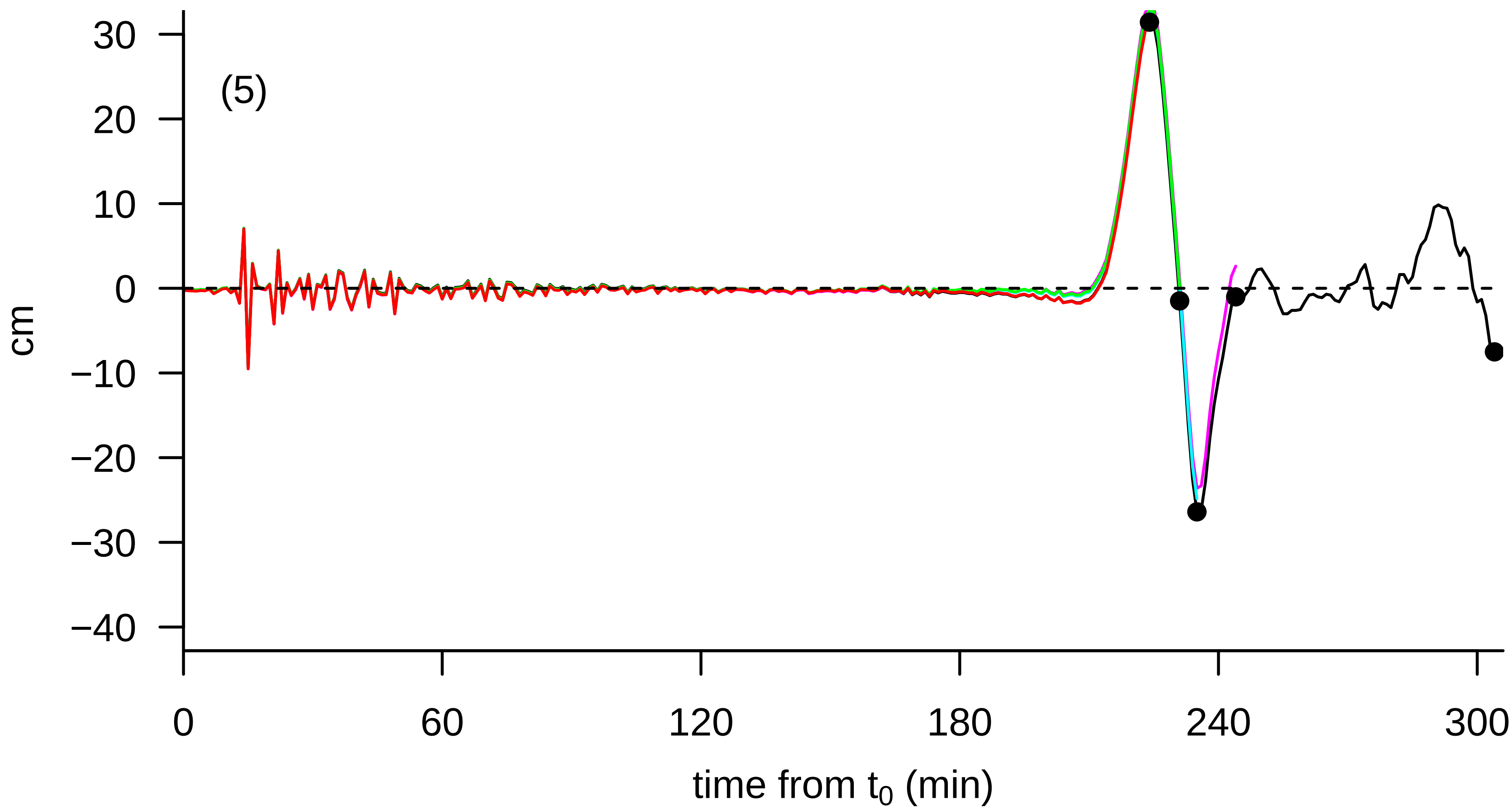
Detiding Using Joint Method



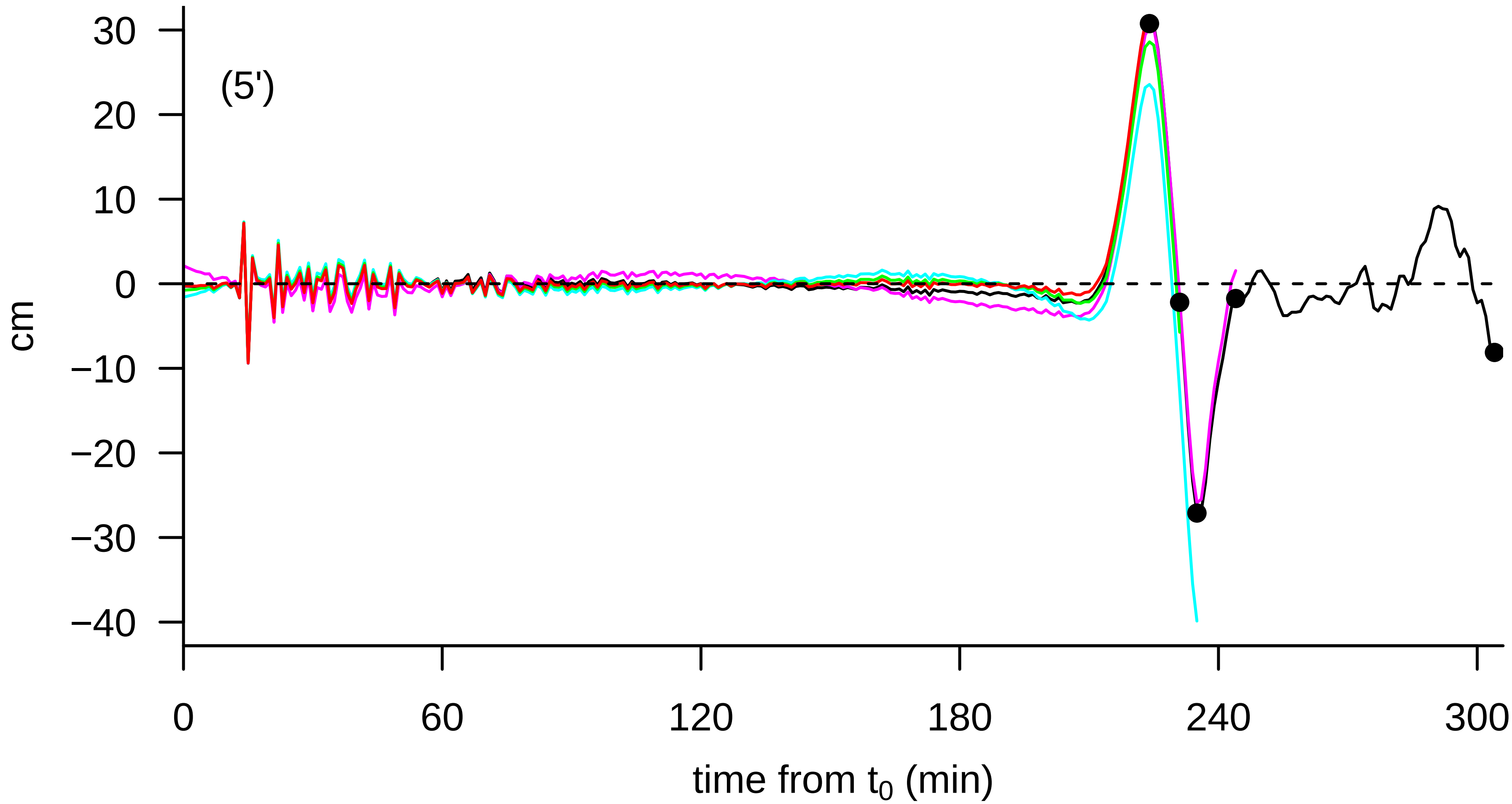
Assessing Performance of Five Methods: X

- joint method depends upon model for tsunami signal
- model used here picked out by automatic method based upon data from three other DART® buoys (21418, 21401 and 21413), each closer to location of generating earthquake than 52402
- model consists of seven units sources (ki24a, ki24b, ki25b, ki26a, ki26b, ki27a and ki27b)
- to see effect of using an inappropriate model, consider model consisting of a single unit source (ki26b)

Detiding Using Joint Method (Seven Unit Sources)



Detiding Using Joint Method (One Unit Source)



Concluding Comments

- why does joint method work well?
  - really only need a simple model for tidal fluctuations when dealing with short stretches of time
  - least squares theory suggests that simultaneous estimation of model parameters is preferable to stage-wise estimation if predictors are correlated
- why might joint method fail?
  - method needs an appropriate model for tsunami signal, but recently developed automatic procedure for model selection offers a promising solution to problem
  - long stretches of time problematic since simple model for tidal fluctuations can deteriorate

Reference

- 'Detiding DART® Buoy Data for Real-Time Extraction of Source Coefficients for Operational Tsunami Forecasting', by D.B. Percival, D.W. Denbo, M.C. Eblé, E. Gica, P.Y. Huang, H.O. Mofjeld, M.C. Spillane, V.V. Titov and E.I. Tolkova, *Pure and Applied Geophysics*, 2015, to appear