Appendix B

Data Pre-processing

In order to successfully perform an FWI style of inversion with our dataset, we first need to apply a standard processing flow. An important goal to keep in mind is that we ultimately want to calculate a 'good' residual for use in our inversion. To do this, we need to have some similarity between our synthetically modeled data and the observed data. One feature in our observed data that can be difficult to accurately recreate in our synthetic data is the bubble that follows the initial source injection. For this reason, we choose to remove it from the observed data.

This appendix intends to show that a straight-forward way to simultaneously remove the bubble and increase similarity between our observed and synthetic data is to shape the observed data to the source wavelet we use in our synthetic modeling. To do this, we first perform separation of the up and downgoing components from the hydrophone data using PZ-summation. Next, we extract a representative wavelet from the observed data first-arrival, and then estimate a filter that shapes it to the synthetic data source wavelet. We then apply this filter to the entire dataset, resulting in a debubbled dataset that has a high degree of phase similarity with our synthetic data.

PZ-summation

Since the dataset was recorded on ocean bottom nodes (OBN), the first arrival in the hydrophone component contains an up-going ocean bottom reflection that nearly coincides with the downgoing direct arrival event. This means the wavelet we extract

from the first arrival event will have an ocean bottom event mixed into it. To get an accurate estimated source wavelet, we need to isolate the downgoing direct arrival. For this reason, we first need to perform PZ-summation before we do any source wavelet estimation.

PZ-summation is a technique used to separate the up and down going components of hydrophone data (figure B.1) using the complimentary information found in the vertical component data (figure B.2). We base our application of this processing step on the approach and assumptions used by Biondi and Levin (2014) (originally based on Melbo et al. (2002)), which represents the up and downgoing data with the following equations:

$$P_{up}(f,k) = \frac{1}{2}P(f,k) + a(f)\frac{\rho}{2q(f,k)}Z(f,k), \tag{B.1}$$

$$P_{down}(f,k) = \frac{1}{2}P(f,k) - a(f)\frac{\rho}{2q(f,k)}Z(f,k), \tag{B.2}$$

where P is the designatured pressure data, Z is the designatured vertical data, $a(f)$ is the calibration filter, ρ is the water density, and q is the vertical slowness of the water layer defined as:

$$q(f,k) = \sqrt{c^{-2} - p^2(f,k)}. \tag{B.3}$$

In this case, c is the water velocity at the receiver position and p is the ray parameter.

One event in the data we can leverage is the refraction event, which by definition is an upgoing event. Furthermore, the refraction event is naturally separated in the time domain at far offsets, making it easy to isolate (see figures B.3 and B.4).

We can estimate a filter $\hat{a}(f)$ to apply to the windowed refraction event such that it minimizes its energy. However, the inverted $\hat{a}(f)$ contains the effect of the vertical slowness and water density. We can represent this with equation B.4:

$$\hat{a}(f) = a(f)\frac{\rho}{2q(f,k)}. \tag{B.4}$$

However, we assume the water velocity and density are constant throughout, and

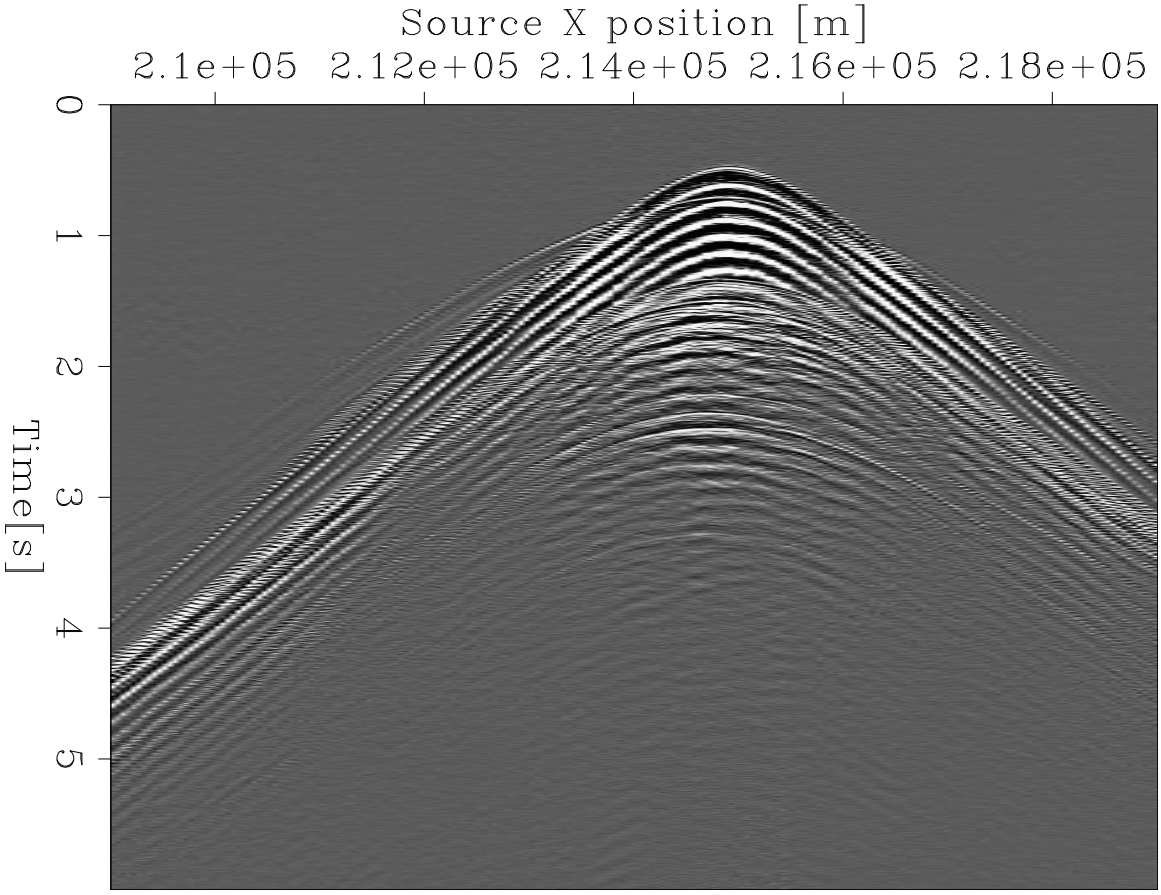

Figure B.1: 2D line receiver gather example from designatured and bandpassed hydrophone data. [**CR**] appendix2/. inputHydro

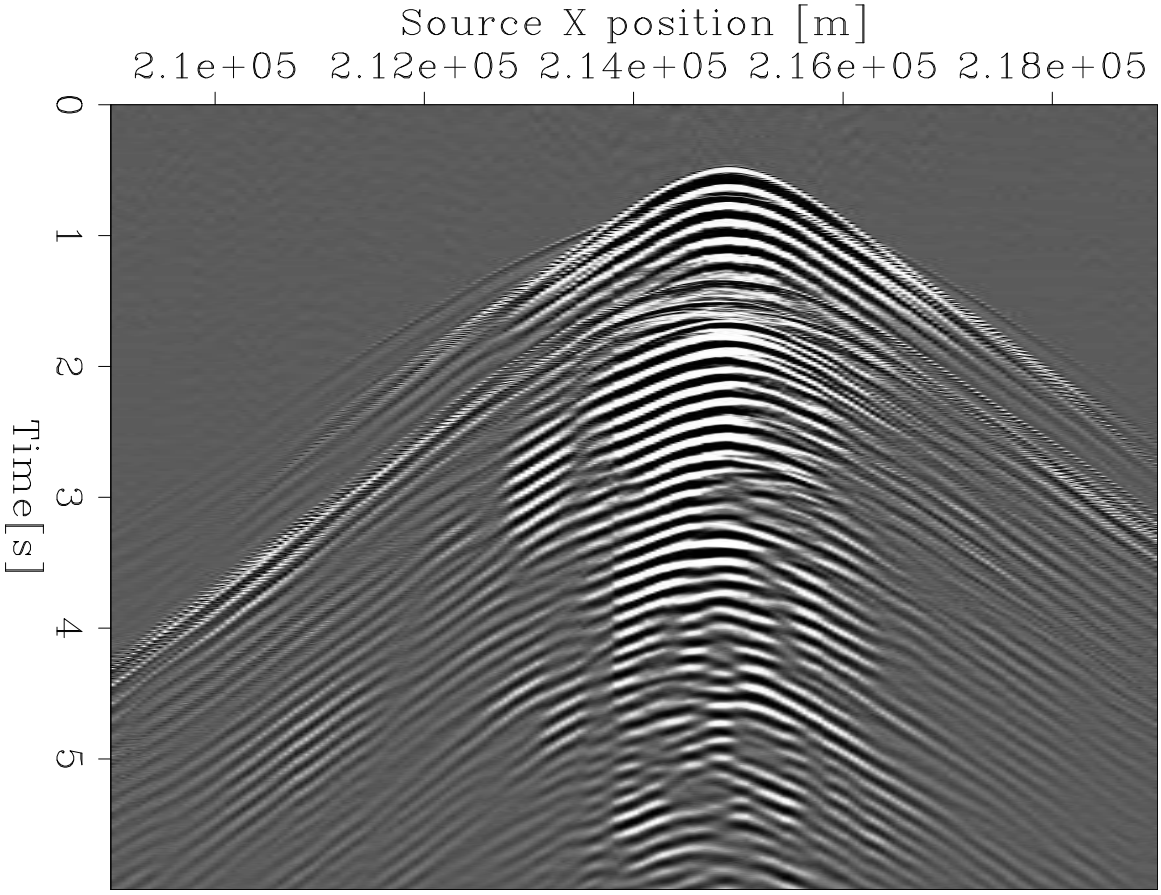

Figure B.2: 2D line receiver gather example from designatured and bandpassed geophone data. [**CR**] appendix2/. inputGeo

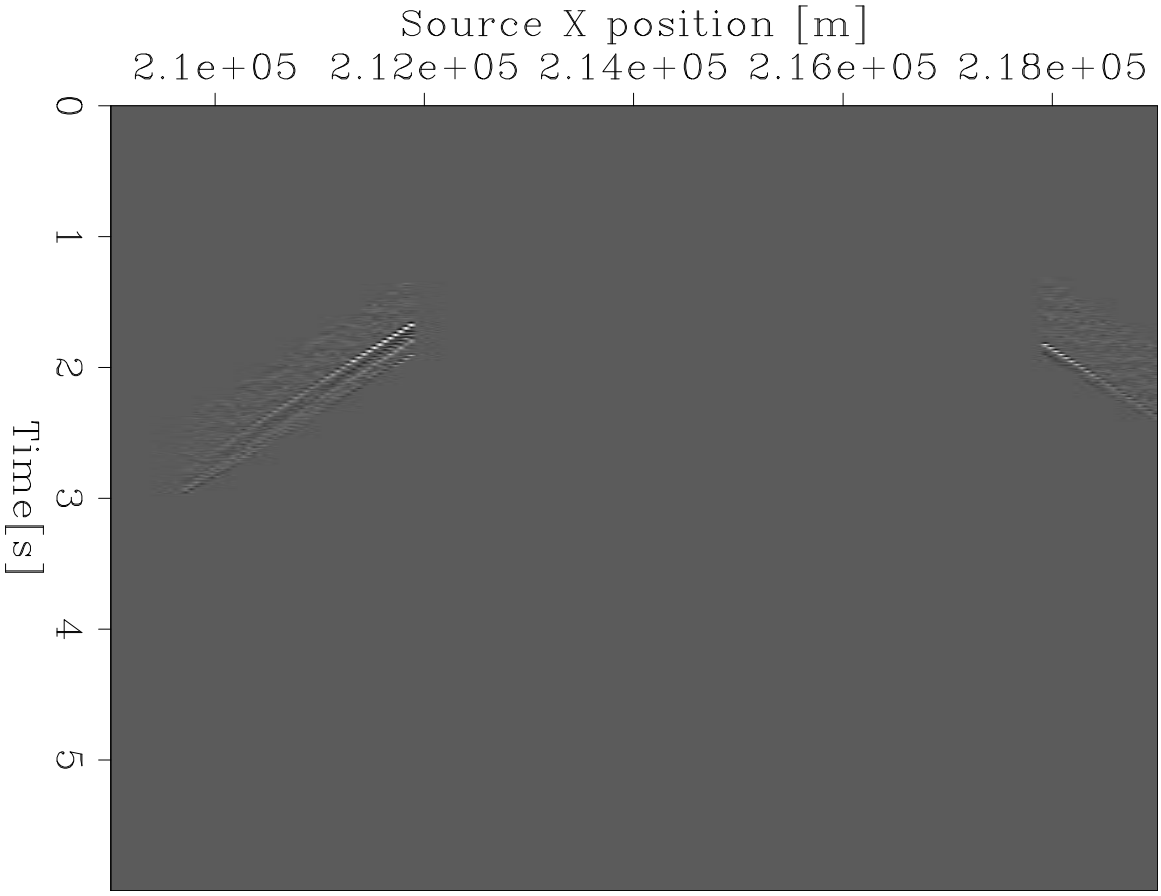

Figure B.3: Hydrophone data window (isolating the refraction events) used for estimating PZ-summation filter. [**CR**] appendix2/. Pwindow

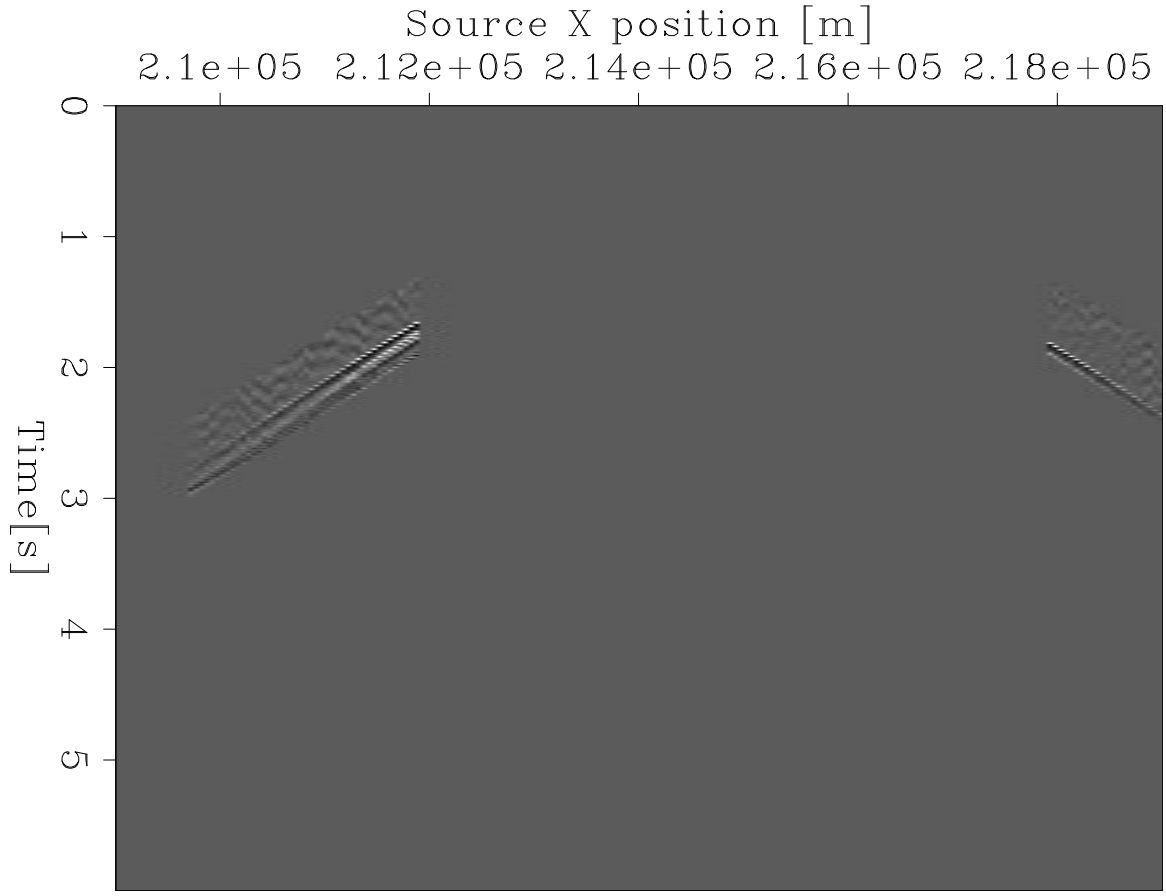

Figure B.4: Geophone data window (isolating the refraction events) used for estimating PZ-summation filter. [**CR**] appendix2/. Zwindow

that the ray parameter is constant in the window we minimize. We can then remove this effect from $\hat{a}(f)$ to get $a(f)$. We can estimate the ray parameter p in the rest of the data which allows us to reuse $a(f)$ to separate the upgoing (figure B.5) and downgoing (figure B.6) components using equations B.1 and B.2.

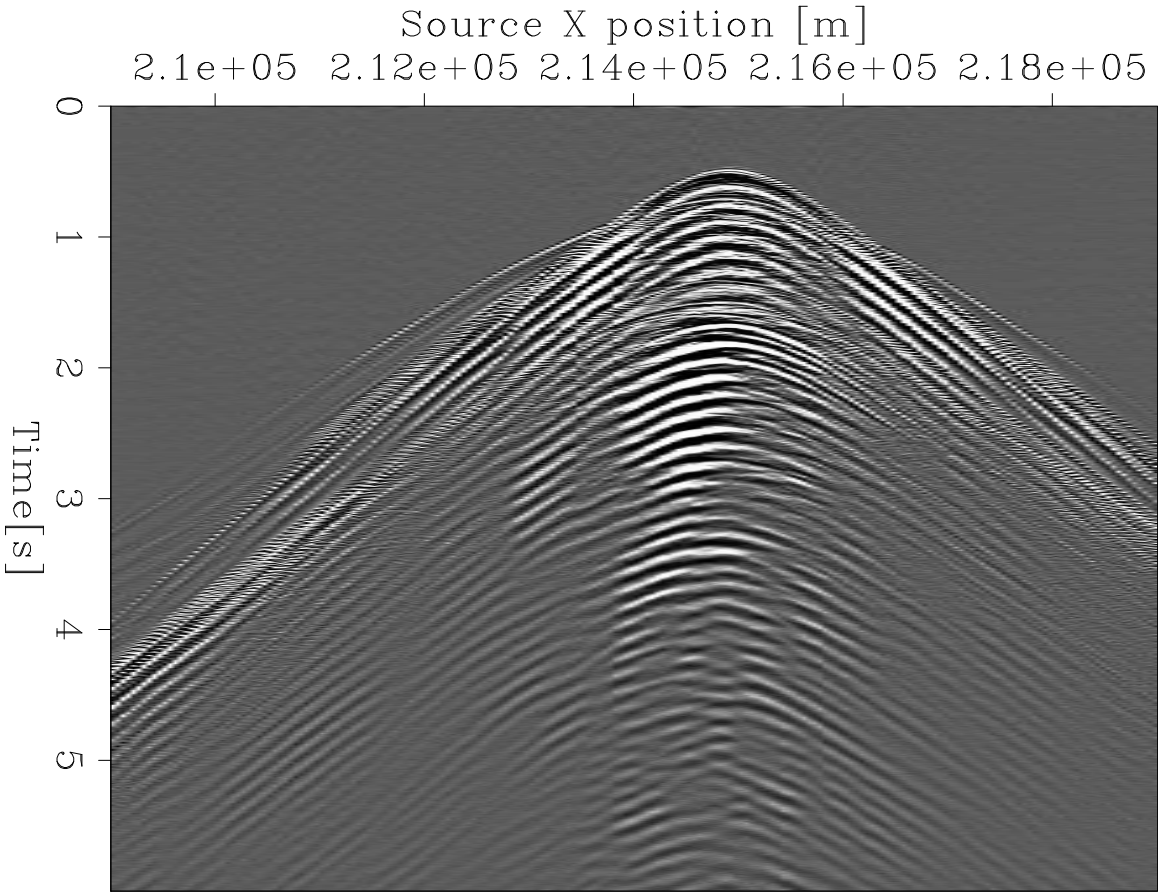

Figure B.5: Upgoing component output from PZ-summation. [**CR**] appendix2/. upgoing

Debubble and wavelet shaping

Picking the data wavelet

Once we have the downgoing component of the hydrophone data separated, we can estimate a source wavelet from a near-offset subset of it (figure B.7). First, we perform hyperbolic moveout (HMO) for each node gather to align the first arrival event across all offsets (figure B.8). We can then stack across all offsets to find a single wavelet

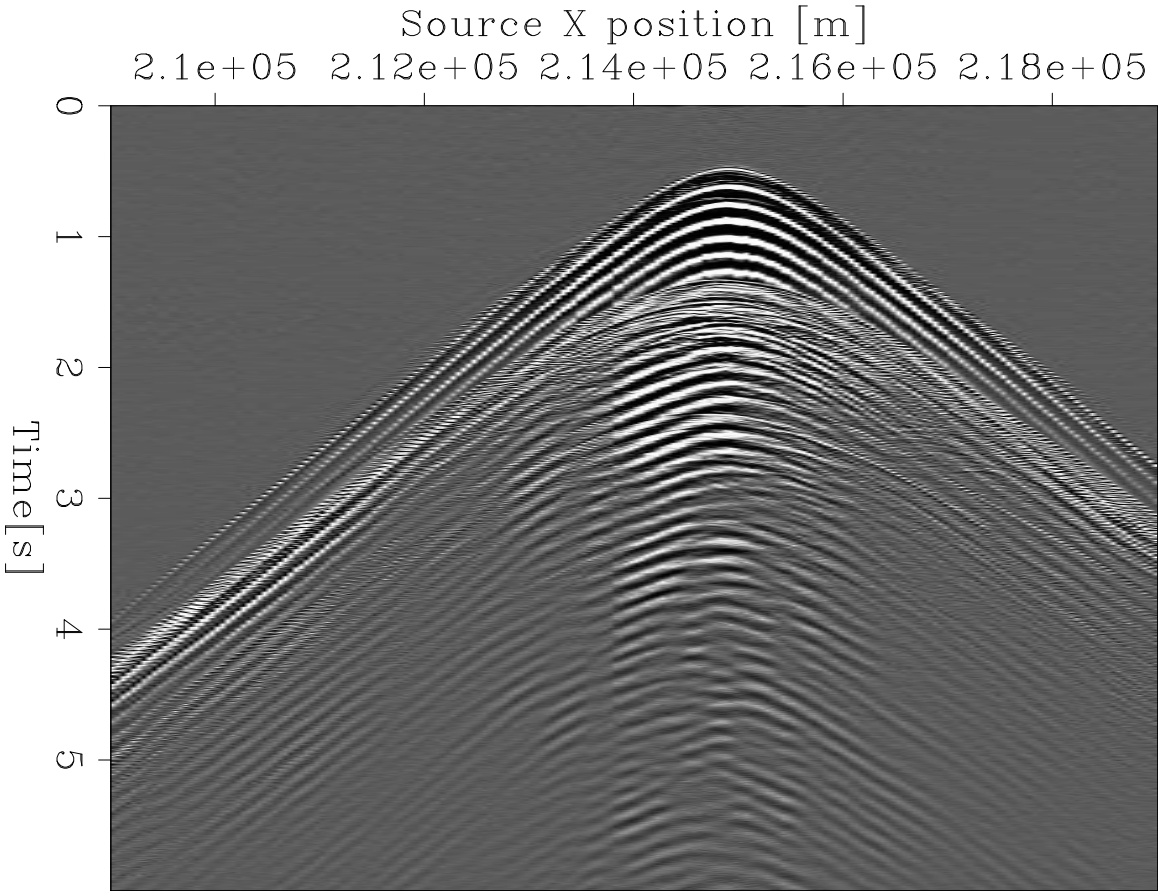

Figure B.6: Downgoing component output from PZ-summation. [**CR**] appendix2/. downgoing

representative of that node gather (figure B.10). To increase our averaging further, we can stack across all node gathers (figure B.11). To do this however, we need to align the first arrival event in each node gather to a common time position ($t_0 = 0$ for example). This means shifting based on the relative depth of each of the nodes respectively. We perform this shifting and stacking across nodes to find a final wavelet representative of the average of the sources actually used. In this wavelet we notice the bubble signature as periodic, fading pulses.

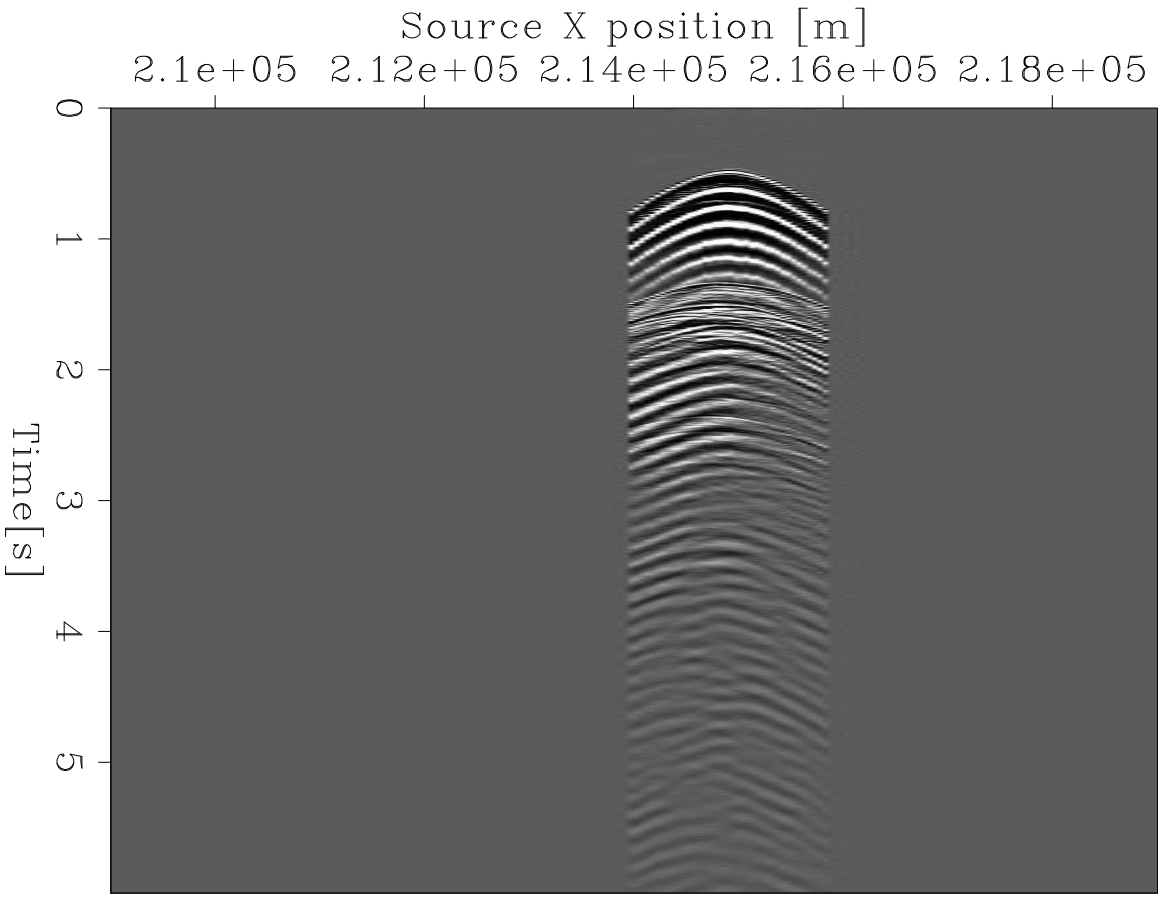

Figure B.7: Near offset ($<$ 1000m) subset of figure B.6 chosen for estimating observed source wavelet. [**CR**] appendix2/. UncorrectedNear

Finding the shaping filter

In our synthetic data, we model using a simple 8[Hz] central frequency Ricker wavelet (see figure B.9). When we compare against the averaged observed data wavelet (figure B.11), we can see that there are some significant differences that warrant the use of a

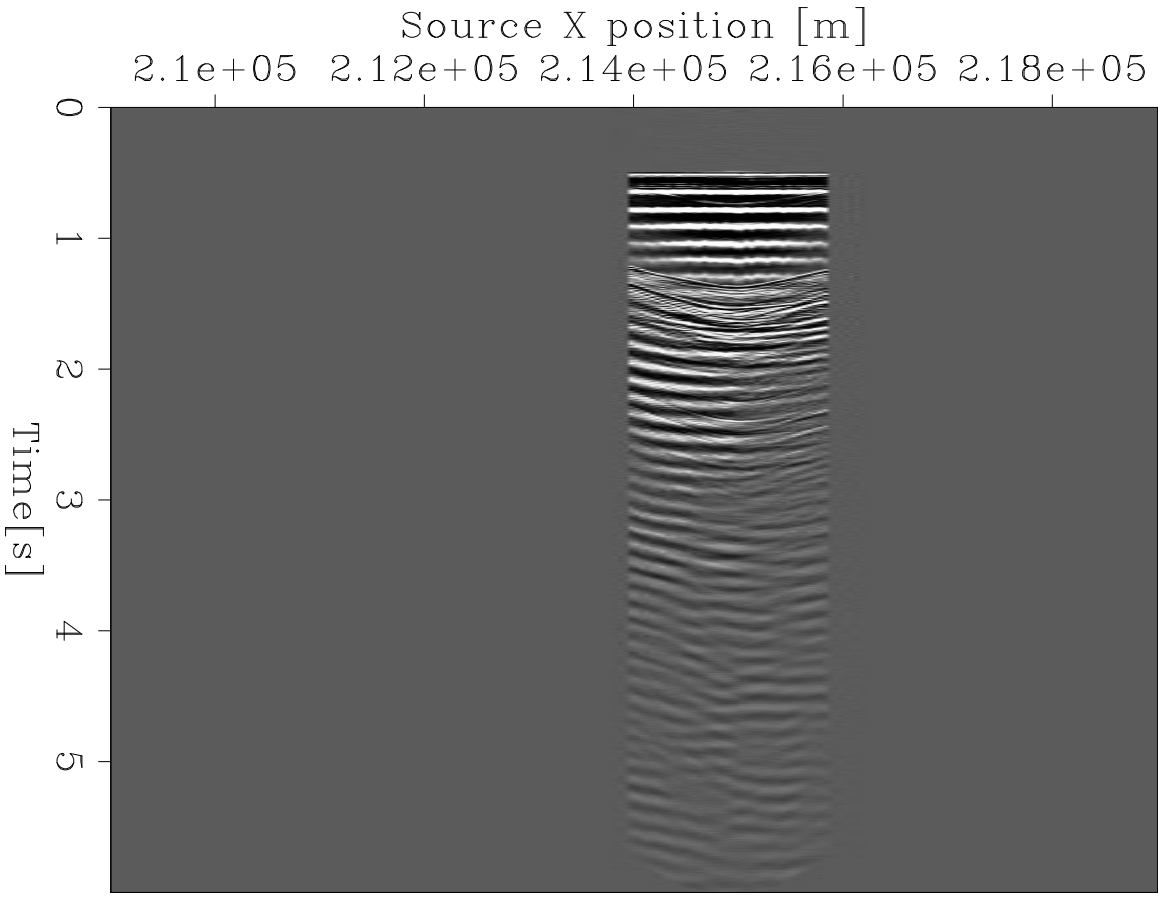

Figure B.8: 1500 m/s HMO correction applied to data in figure B.7. [**CR**] appendix2/. HMOcorrected

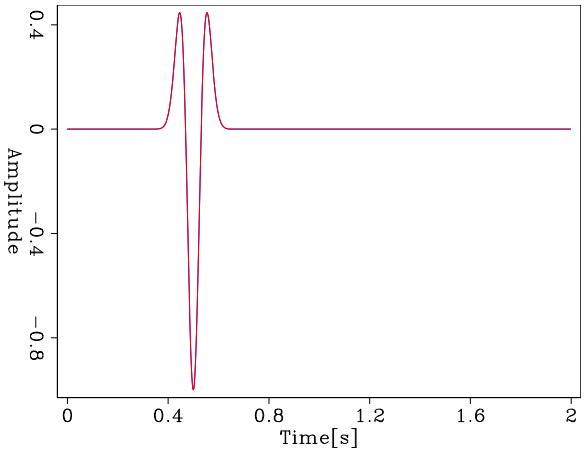

Figure B.9: Wavelet used in synthetic data modeling. [**CR**] appendix2/. synWavelet

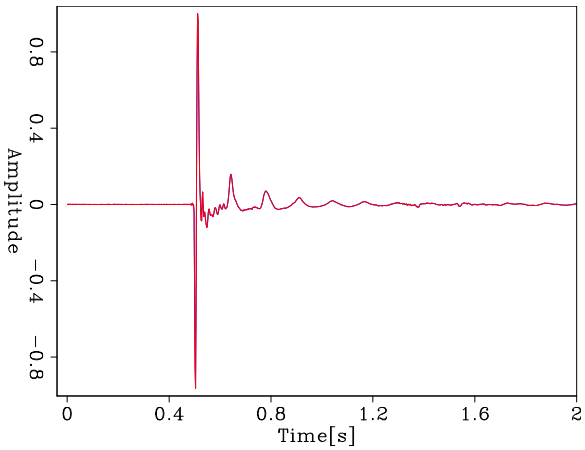

Figure B.10: Wavelet produced from stacking across traces in figure B.8. [**CR**] appendix2/. Node3StackedWavelet

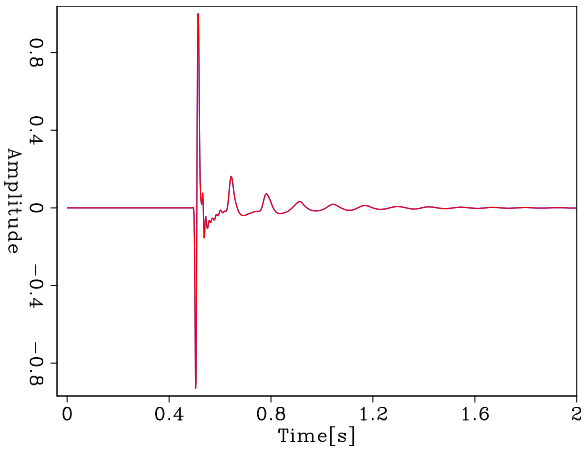

Figure B.11: Observed source wavelet built from average of wavelets extracted from 381 node gathers (just as in figure B.10). [**CR**] appendix2/. AverageStackedWavelet

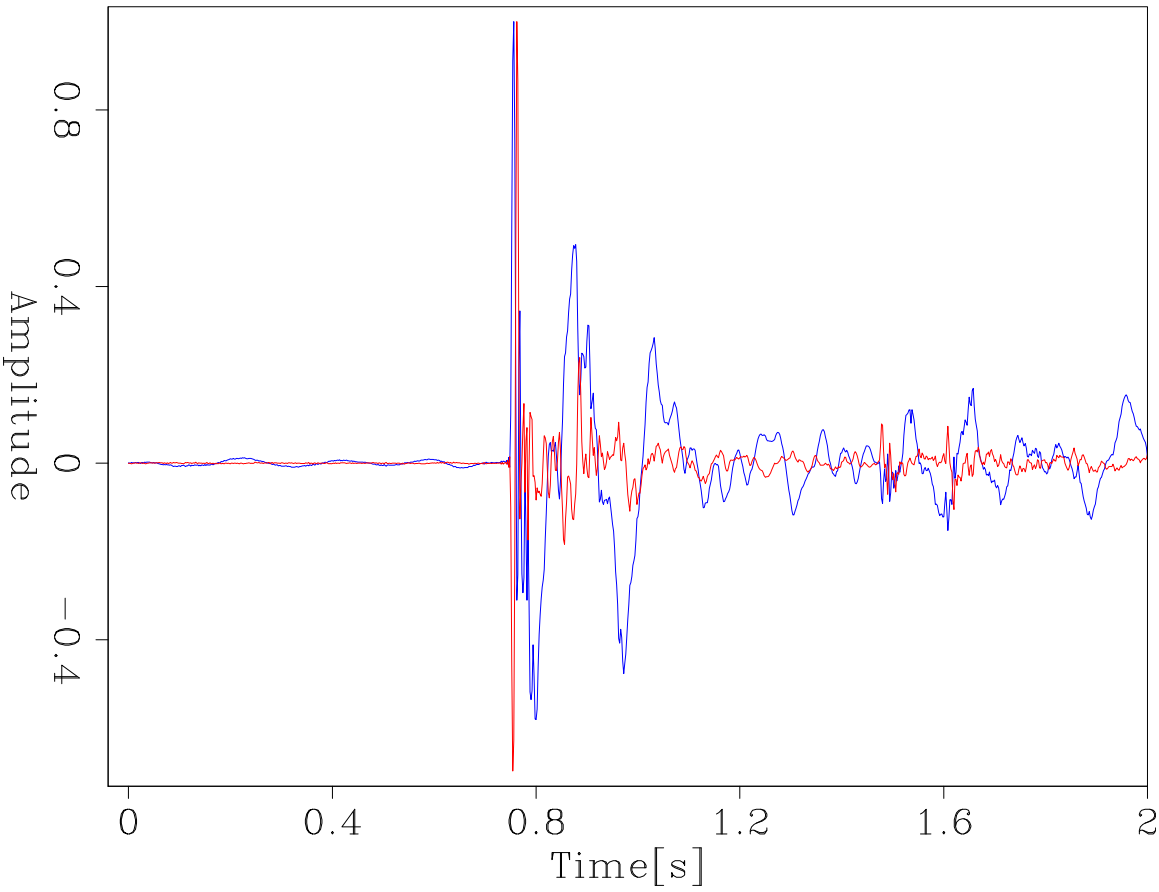

Figure B.12: Example trace before (red) and after (blue) shaping filter applied. [**CR**] appendix2/. ShapedTraceCompare

shaping filter. We estimate the filter that shapes the averaged observed data wavelet to the synthetic wavelet (see Yilmaz (1987)), and then apply to the full dataset (figures B.13 and B.14). When we compare figure B.6 with figure B.13, we can see that the bubble removal is effective, and that the frequency and phase content has become more similar to what we would expect given the lower frequency Ricker wavelet to which we match.

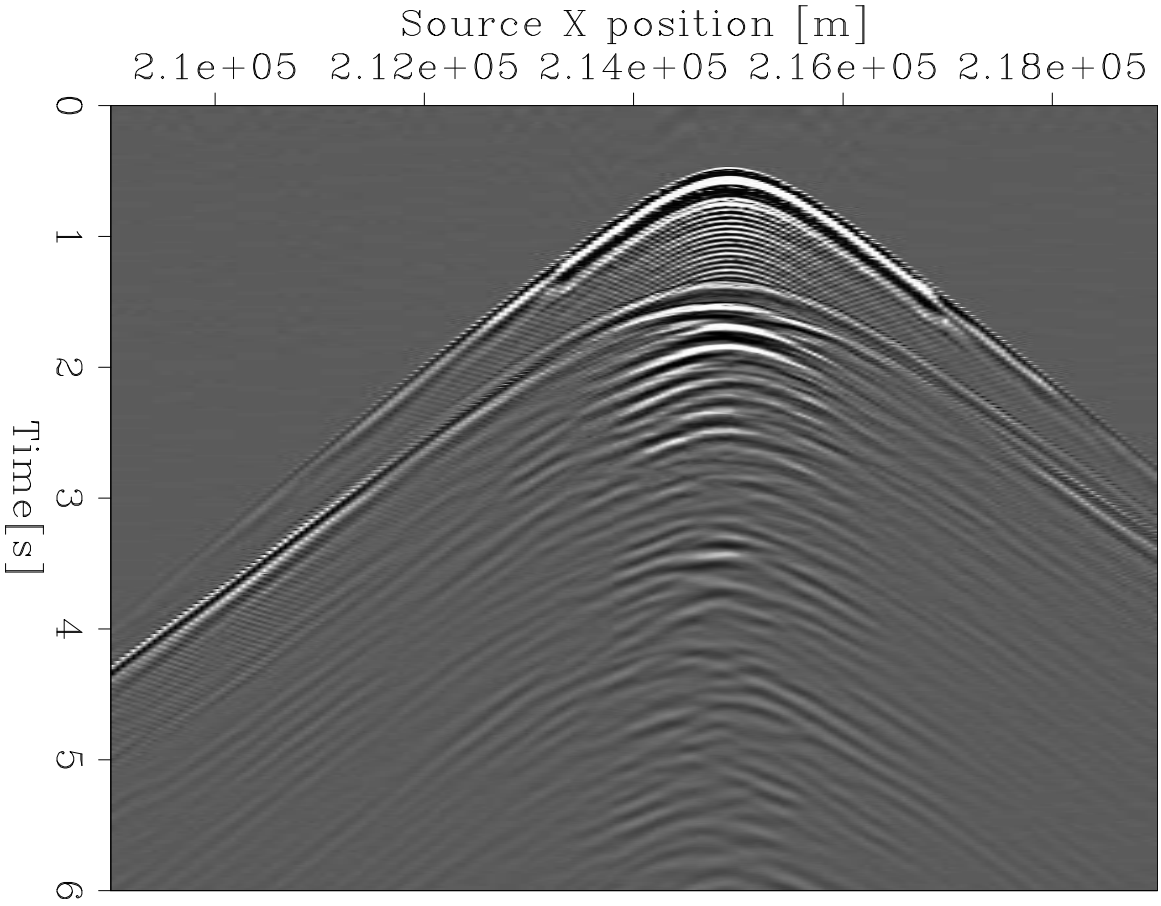

Figure B.13: Downgoing hydrophone data after shaping filter applied. [**CR**] appendix2/. ShapedDOWNData

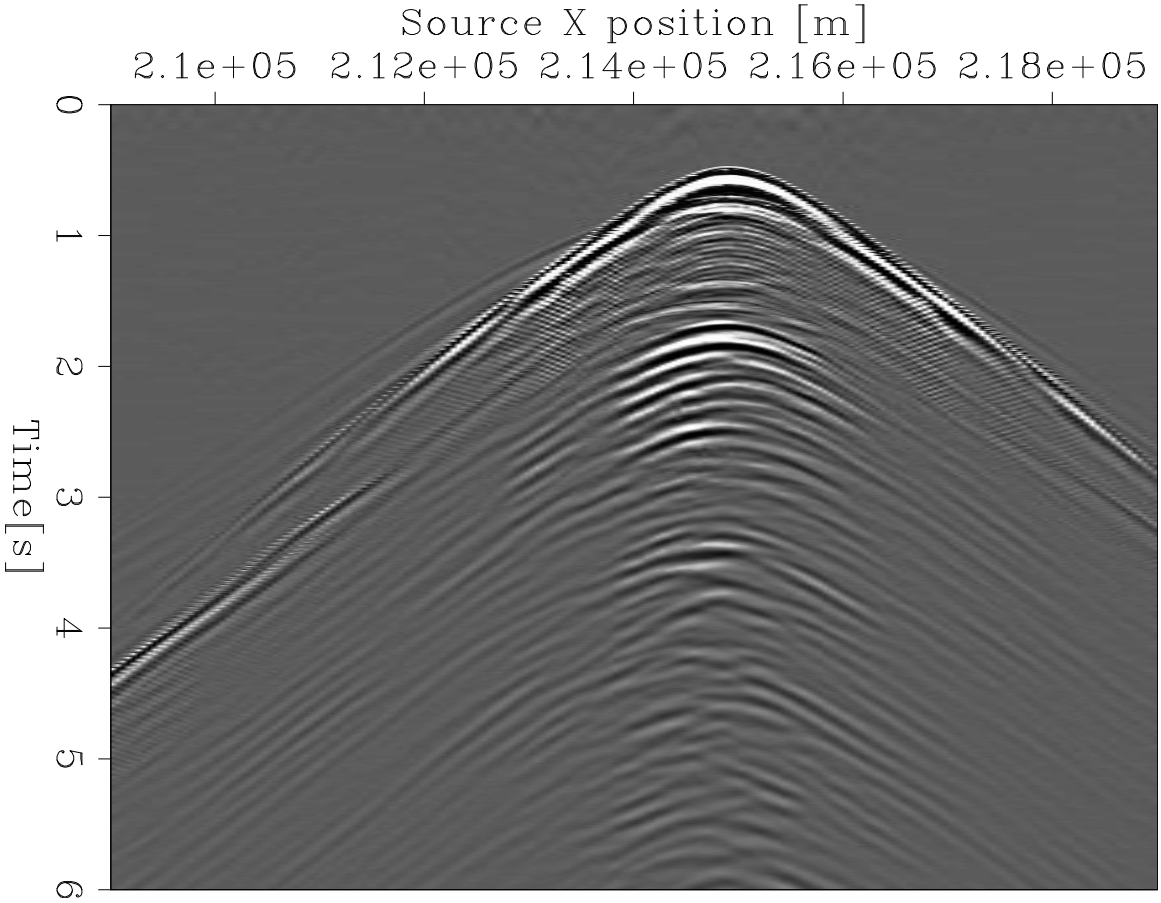

Figure B.14: Upgoing hydrophone data after shaping filter applied. [**CR**] appendix2/. ShapedUPData