Accurate source wavelet estimation using Marchenko focusing functions

Journal Article

Author(s):
Mildner, Constantin; Broggini, Filippo (iD); de Vos, Koos; Robertsson, Johan (iD)

TITLE PAGE

Source wavelet estimation using Marchenko focusing functions

(Shortened form:)

Constantin Mildner*, Filippo Broggini*, Koos de Vos[+], Johan Robertsson*

*Institute of Geophysics, Eidgenössische Technische Hochschule Zürich,

Sonneggstrasse 5, CH-8092, Zürich, Switzerland.

[+]Shell Global Solutions International,

Grasweg 31, 1031HW Amsterdam, The Netherlands.

constantin.mildner@erdw.ethz.ch

filippo.broggini@erdw.ethz.ch

Koos.DeVos@shell.com

johan.robertsson@erdw.ethz.ch

Date of submission original paper:

1

INTRODUCTION

Imaging a target zone in the subsurface below a highly scattering overburden can be challenging because waves that are multiply scattered in the overburden potentially interfere with primary reflections and, thus, blur the image of the target zone. Recently, Marchenko redatuming and imaging was introduced (Wapenaar et al., 2014ab; Broggini et al., 2014ab) that allows to create an image of a target zone below a complex overburden that is free of internal multiples. By solving the coupled Marchenko equations for so-called focusing functions (Wapenaar et al., 2014a; van der Neut et al., 2015a), one can calculate the Green's functions decomposed into up and downgoing parts that would be recorded at the chosen focusing (i.e., redatuming) positions according to sources at the surface. Despite one-sided illumination with sources and receivers at the acquisition surface only, these Green's functions contain all orders of internal multiples. Thus, multidimensional deconvolution of the decomposed Green's functions (Wapenaar et al., 2011; van der Neut et al., 2011) allows to create a virtual redatumed reflection response as if both sources and receivers were located at the chosen redatuming level.

The zero-offset zero-time components of the resulting redatumed reflection response can directly be used to image the medium at the redatuming level (Broggini et al. 2014b; Behura et al. 2014). Alternatively, the redatumed data can be used to create an image of the medium below the redatuming level using standard imaging methods, such as reverse-time migration (RTM). Furthermore, the Marchenko method can also be applied to approximate a dataset for the acquisition surface that contains primaries only (Meles et al., 2015ab) and to create target-enclosed extended images (Vasconcelos et al., 2017). Initial field data applications of the Marchenko imaging method show promising results by improving the

2

continuity of reflections in Marchenko images compared to images created from surface data (Ravasi et al., 2015; Ravasi et al., 2016; van der Neut et al., 2015b; Staring et al., 2017).

The Marchenko redatuming algorithm requires two inputs in order to create accurately redatumed data. The first required input for the focusing approach is the first arrival of the inverse of the transmission response between the acquisition surface and the focusing position in the subsurface. Secondly, the redatuming algorithm requires the true reflection response of the medium, i.e. deconvolved by the source wavelet, for densely sampled co-located dipole sources and pressure receivers at the surface (Wapenaar et al., 2014a). Whereas the first input can be approximated for instance by modelling the first arrival in an estimated background velocity model using an arbitrary wavelet, the second input requires accurate knowledge of the source wavelet that is present in the acquired data. Deconvolving an incorrectly scaled source wavelet from the input data leads to artifacts in the redatumed Green's functions (van der Neut et al., 2015b) and in subsequent images created with redatumed data (Jia et al., 2018). So far, different methods were proposed to find the correct scale of the reflection data. Whereas Thomsen (2016) calibrated the reflection data using vertical seismic profile (VSP) data as additional constraint, Ravasi et al. (2016) scaled the reflection data based on the convergence of the solution of the Marchenko equations. Van der Neut et al. (2015b) suggested to optimize a misfit function based on the upgoing part of the redatumed Green's function in order to find the correct scale of the input data. This approach, however, does not always necessarily yield the correct scale of the data as discussed by Brackenhoff (2016).

We propose to optimize a misfit function based on focusing functions, instead of redatumed Green's functions, in order to invert not only for the correct overall scale of the reflection data but also for the amplitude and phase spectrum of the source wavelet contained

in the acquired surface data. Up to now, the effect of an incorrect amplitude and phase spectrum of the estimated source wavelet on the Marchenko focusing functions has not been studied in detail. The sensitivity of Marchenko redatuming for an incorrect source scaling suggests to exploit Marchenko wavefields to invert for the source wavelet. To this end, focusing wavefields can be significantly advantageous compared to redatumed Green's functions, as we will show in the following. Knowing the correct source wavelet does not only enable accurate redatuming and subsequent imaging, but is also of great importance for further applications, such as RTM, full-waveform inversion, and amplitude-versus offset analysis.

The structure of the paper is as follows. We begin with a review of the focusing functions and their connection to the redatumed Green's functions. Then, we analyze the effect of an incorrect amplitude and phase spectrum of the estimated source wavelet on the focusing functions. The findings from those analyses are used to define and demonstrate a workflow to invert for the frequency dependent scaling (amplitude spectrum) and phase of the source wavelet and, hence, of the reflection data. In the following section, we extend the proposed inversion approach enabling the inversion for lateral variations of the source scaling and demonstrate the effect of laterally-varying source scaling errors on images created with redatumed data. Based on obtained results, we discuss and conclude possible limitations as well as extensions of the proposed methodology.

FOCUSING FUNCTIONS

The proposed source wavelet inversion approach relies on the focusing functions f_1 obtained from the Marchenko equations (van der Neut et al., 2015a). The focusing functions relate the surface reflection response R with the redatumed Green's functions $G(x_f, x_0, t)$ between co-located sources and receivers at the surface x_0 and the focusing positions located at x_f (Wapenaar et al., 2014a). In the frequency domain, the decomposed Green's functions G^+ and G^- are related to R through the one-way focusing functions f_1^+ and f_1^-:

$$G^{+*} = f_1^+ - R^* f_1^- \qquad (1)$$

$$G^- = -f_1^- + R f_1^+, \qquad (2)$$

where the superscript $*$ indicates complex conjugation and superscripts $+$ and $-$ denote the down- and upgoing parts of G and f_1, respectively. We construct the decomposed focusing functions by solving the coupled Marchenko equations iteratively (Wapenaar et al., 2014a; Slob et al., 2014). The iterative solution for f_1^+ and f_1^- can be expressed equivalently as a Neumann series expansion (e.g., van der Neut et al., 2015a). The K-th order estimates of f_1^+ and f_1^- are then given by

$$f_1^+ = \sum_{k=0}^{K} (\theta R^* \theta R)^k f_{1d}^+, \qquad (3)$$

$$f_1^- = \theta R \sum_{k=0}^{K} (\theta R^* \theta R)^k f_{1d}^+. \qquad (4)$$

In equations 3 and 4, R denotes the surface reflection response for co-located sources and receivers correctly deconvolved by the source wavelet. The reflection response R needs to be provided as one of the inputs to the Marchenko algorithm. The second required input, f_{1d}^+, is an estimate of the first arrival of the inverse of the transmission response between the surface and the focusing position. It can be extracted directly from the reflection data (Al-Ali

& Verschuur, 2006; van der Neut et al., 2015a). However, we model f_{1d}^+ using an estimated background velocity model. The respective first-arrival times t_d between the focusing positions x_f and the receivers x_0 located at the surface are obtained with an Eikonal solver and convolved with a constant wavelet to create f_{1d}^+. Note that the hereby obtained initial focusing functions f_{1d}^+ assume the absence of short-period interbed multiples in the data. The calculated first arrivals are also used to build the windowing function $\theta(x_f, x_0, t)$. It is constructed such that

$$\theta(x_f, x_0, t) = \begin{cases} 1 & t < t_d^\varepsilon \\ 0 & t \geq t_d^\varepsilon \end{cases}$$

holds. Furthermore, θ is constructed symmetrically in time such that it mutes all events after the direct wave and before the time-reversed direct wave, including the direct wave itself since the small constant ε is subtracted from t_d to calculate t_d^ε.

Once the focusing function are calculated, we can create the redatumed Green's functions from equations 1 and 2. These are defined for the actual medium and contain contributions from reflectors and scatterers above and below the focusing depth. However, the focusing functions (equations 3 and 4) represent solutions of the wave equation with a specific focusing condition in a reference medium. This reference medium is equivalent to the actual medium between the surface and the focusing depth and reflection-free below (Wapenaar et al., 2014a). The downgoing component f_1^+ of the focusing function is defined such that it focuses at the focusing position at $t = 0$ when it is re-injected at the surface into the reference medium. If the reflection response R of a layered medium with n interfaces between the surface and the focusing depth is scaled (i.e., deconvolved) correctly, f_1^+ contains 2^{n-1} events and consists of the initial downgoing focusing function f_{1d}^+ and of subsequent coda events that cancel internal multiples originating in the reference medium (Slob et al., 2014). Thus, when re-injecting f_1^+ into the reference medium, the wavefield focuses at the

6

The redatumed Green's function G is sensitive to an incorrectly scaled reflection response and thus, to an erroneous source wavelet used for deconvolution (e.g., Brackenhoff, 2017). Therefore, one could use G to optimize the source wavelet scaling, as suggested for instance by van der Neut et al. (2015b). However, using the focusing wave field f_1 instead of G allows us to simplify the optimization problem. Contrary to the redatumed Green's functions, f_1 shows only interactions with the medium between the surface and the focusing level. As we will show in the following sections, f_1^- carries all necessary information to invert for the source wavelet. The location of the focusing position can be chosen arbitrarily. Thus, using f_1 for a shallow focusing position leads to fewer events that might interfere with each other compared to f_1 for a deeper focusing position and especially compared to using G.

The advantage of simplified wavefields for the focusing functions compared to the redatumed Green's functions is demonstrated using the velocity model shown in Figure 1. The model, referred to as velocity model from Middle East, consists of a stack of 5 horizontal layers with sharp interfaces (overburden) overlying four layers that are separated by undulating, dipping, and finely layered interfaces. For simplicity, we only show the velocity model here; the density model is characterized by the same structure, where velocity and density values correlate. The deepest horizontal reflector is located at a depth of 1350 m and the layers in the overburden are characterized by alternating impedance contrasts. Additionally, we place

2 Gaussian velocity (and density) anomalies in the third layer of the model introducing lateral variations in the overburden. At $(x, z) = (-1250\ /736)$ m, where x denotes the lateral position and z the depth coordinate, we center a negative anomaly with a maximum of -10%, while a positive anomaly of $+10\%$ is located at $(x, z) = (+1250\ /816)$ m.

We create the surface reflection response using an acoustic 2D finite-difference modelling algorithm (e.g., Moczo et al., 2007). We simulate 901 source locations (vertical force sources) at zero depth using a 24 Hz Ricker wavelet as source wavelet. The acquisition surface (Figure 1, black dashed line) is transparent, as indicated by the additional 500 m above $z = 0$ m, so that we do not observe surface-related multiples, which is a requirement for the Marchenko implementation as it is defined above.
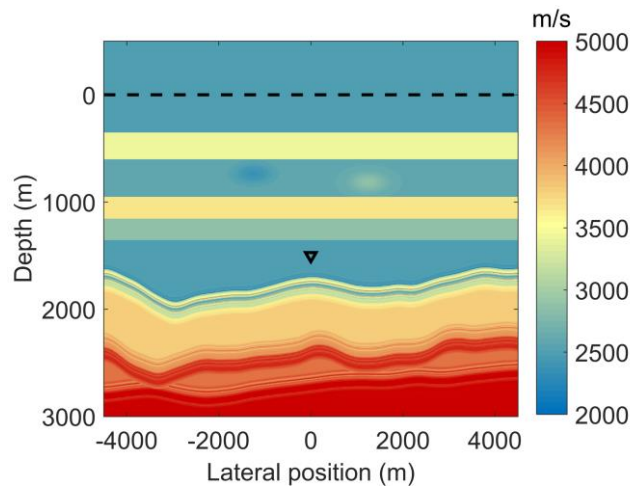

Figure 1 Velocity model used to calculate the surface reflection response. The transparent acquisition surface at a depth of 0 m is indicated by a black dashed line and the focusing position at $(x, z) = (0,1500)$ m by a black triangle.

The sources are co-located with 901 evenly sampled receivers between -4500 m and $+4500$ m. At all receiver locations, we record the response to each source in form of acoustic pressure measurements. This velocity model and source/receiver configuration is used for all examples shown in this study.

We aim to remove the effect of the stack of horizontal layers (overburden) and calculate the focusing functions and the redatumed Green's function between the focusing position at $(x, z) = (0,1500)$ m (Figure 1, black triangle) and the receivers at the surface. In this example, f_{1d}^{+} is modeled as explained above using a 24 Hz Ricker wavelet. The Marchenko redatumed Green's function and the related focusing functions are shown in Figure 2. For the configuration introduced above, the redatumed Green's function (Figure 2a) comprises significantly more events than the downgoing focusing function f_1^{+} (Figure 2b) and the upgoing focusing wavefield component f_1^{-} (Figure 2c). Having 5 interfaces between the surface and the chosen focusing position, we expect $2^{5-1} = 16$ events in both f_1^{+} and f_1^{-}. Clearly, not all expected events are visible in the focusing wavefields either because of their small amplitudes or because of interference and superposition of waves.
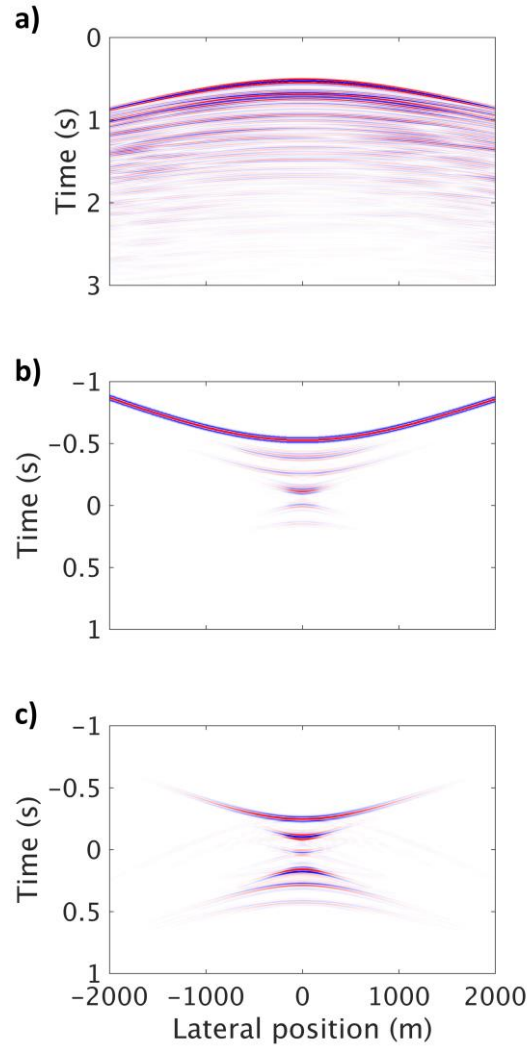
Figure 2 (a) Total Green's function (a), (b) downgoing focusing function f_1^+, and (c) upgoing focusing function f_1^- for the focusing position indicated in Figure 1.

However, to invert for the source wavelet we are not restricted to the target focusing (i.e., redatuming) depth of 1500 m. The choice of a shallower focusing location simplifies the focusing functions by decreasing the number of events and reducing the superposition of waves. Thus, we first analyze the influence of an incorrect source wavelet on the focusing wavefields in detail for a shallower focusing position, followed by the introduction of a criterion used for selecting an optimal focusing positon for the source wavelet estimation.

AMPLITUDE SPECTRUM INVERSION

As described in the previous section, f_1^- consist of primary reflections and subsequent coda reflections resulting from the cancelling events in f_1^+ if the surface reflection response R is deconvolved correctly by the true source wavelet $S(t)$. However, an incorrectly scaled (i.e., deconvolved) R implicates additionally occurring coda events in f_1^- because of the incomplete cancellation of multiples. In the following, these events are referred to as additional or unwanted coda events. The f_1^- for the same velocity model, source-receiver configuration, and focusing position as before ($z = 1500$ m) is shown in Figure 3a. On the other hand, an underestimated source wavelet $S_0(t) = 0.8\,S(t)$ is used for deconvolution in this case. Compared to the correct f_1^- (Figure 2c), the additional coda events are apparent as indicated by the black arrows in Figure 3a, for instance before the second and after the fifth event.

However, the additional coda events that we can expect in case of an incorrectly scaled source wavelet are not clearly separated from primary reflections. In order to show the effect of an incorrect source scaling on f_1^- explicitly, we reduce the number of events in the focusing function and thus simplify it by choosing a shallower focusing position such that there are fewer reflectors between the surface and the focusing depth. Two f_1^- wavefields for a focusing position located in the third layer of the velocity model at $(x, z) = (0,900)$ m are shown in Figure 3 in panel b and c. The same incorrect $S_0(t) = 0.8\,S(t)$ as before is used for deconvolution to create f_1^- shown in panel b, whereas the data R are deconvolved by the true source wavelet $S(t)$ prior to Marchenko focusing for panel c. Having only two reflectors between the surface and the focusing level, the true f_1^- (Figure 3c) comprises only two primary reflections and no coda reflections.

The incorrect f_1^- (Figure 3b), on the contrary, shows one additional event because of the imperfect cancelling of the internal multiple originating in the second layer of the velocity model. Underestimating the source wavelet by a factor of 0.8 and using it to remove the wavelet from R results in too large amplitudes of the reflections in the deconvolved data. Thus, the cancelling impulse in the corresponding f_1^+ wavefield is not scaled correctly and the internal multiple is not suppressed entirely. A similar result is also obtained for an overestimated source wavelet.
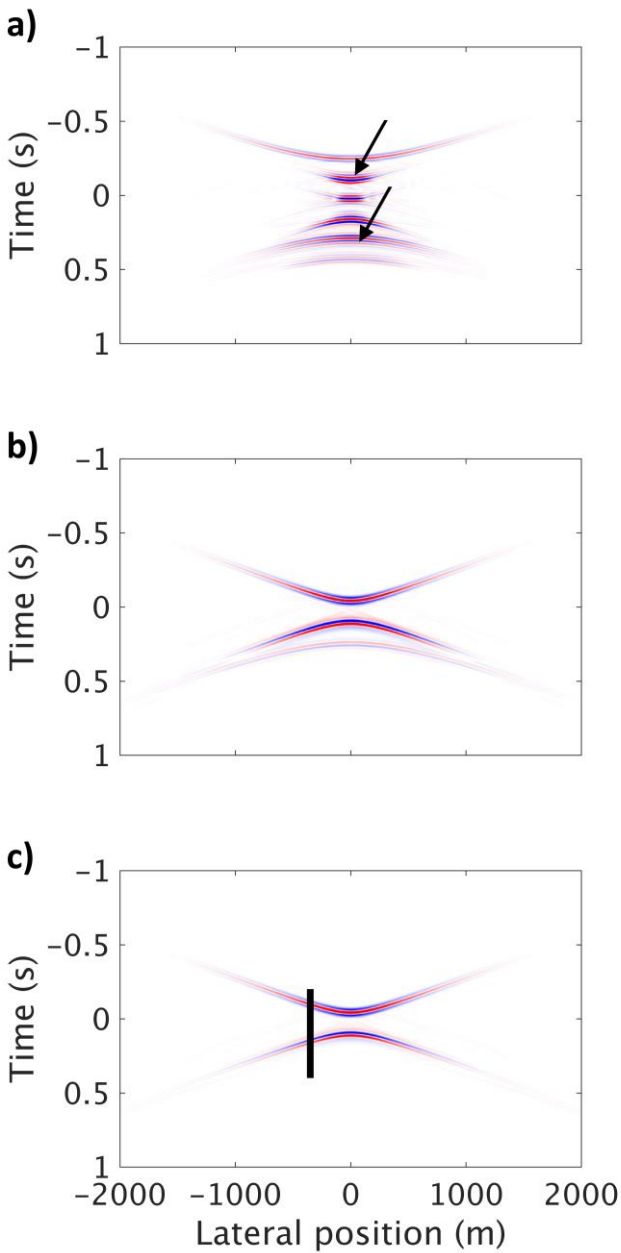
Figure 3 (a) f_1^- for a focusing position at $(x,z) = (0,1500)$ m. An underestimated source wavelet is used for deconvolution prior to Marchenko focusing. Additional coda events due to the incorrectly scaled source wavelet are indicated by black arrows. (b) f_1^- for a focusing position at $(x,z) = (0,900)$ m. An additional third coda event is present due to an incorrect source wavelet. (c) Same as for panel b) but using the true source wavelet for deconvolution. Only the two expected events are present. The black bar denotes a trace segment at $x = -350$ m that is used in the following figures.

For the chosen focusing position at a depth of 900 m, the unwanted coda event is separated distinctly from primary reflections. The additional coda event is only present in the f_1^- if an incorrectly scaled source wavelet is used for deconvolution. Thus, minimizing unwanted coda events in f_1^- allows one to estimate the overall scaling and to invert for the amplitude spectrum of the source wavelet. In case of more than two interfaces between the surface and the focusing level, the focusing wavefield f_1^- calculated using correctly scaled data contains both primaries and coda events. However, also in that case additional unwanted coda events are present if the estimated source wavelet is scaled incorrectly. We only aim to minimize these additional coda events in f_1^- in order to invert for the scaling of the estimated source wavelet. Therefore, we now introduce a selection criterion to distinguish between primaries and unavoidable coda events on the one hand and unwanted additional coda events due to incorrect source scaling on the other hand.

<u>Coda determination</u>

Minimizing additional coda events in f_1^- requires that we are able to distinguish between the unwanted coda events and primary reflections or coda events, which we do not aim to minimize. This is now demonstrated using, for clarity, only trace segments of f_1^- as

indicated by the black bar in Figure 3c, located at $x = -350$ m. The separation of such coda events from primary reflections is also taken as selection criterion for the focusing position that is used for the source wavelet inversion. Trace segments of f_1^- for the focusing position at $(x, z) = (0,900)$, as described above, are shown in Figure 4a. The f_1^- wavefield using the true source wavelet $S(t)$ for deconvolution prior to Marchenko focusing (Figure 3c) shows two primary reflections that appear at -0.1 and 0.15 s for $x = -350$ m. However, using an incorrectly scaled starting source wavelet $S_0(t)$ for deconvolution leads to an additional coda event in the f_1^- focusing function as demonstrated above.

To distinguish the primary reflections from the unwanted coda events, we create an ensemble of f_1^-, which is illustrated by colored traces in Figure 4a. All traces are normalized with respect to the second primary reflection.
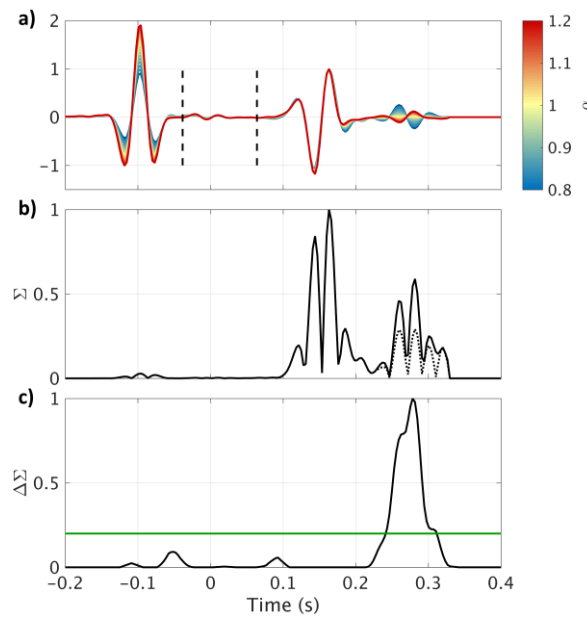

Figure 4 (a) Trace segment of f_1^- for a focusing position at $(x, z) = (0,900)$ m, as indicated in Figure 3c, using different source wavelets for deconvolution which are scaled by a factor α. A window without any events is marked by black dashed lines. This window is later used to define a stopping criterion of the inversion as explained in the text. (b) Stack of absolute values (black solid line) and absolute value of the stack (dashed black line) of the traces shown in

panel a). Only for the coda events the two stacks differ significantly. For clarity, a gain of $\exp(10t)$ is applied within the time window. (c) Normalized, smoothed difference of the two stacks from panel b) in black. Only for the coda event, the difference is above the threshold (green line) set to 20% of the maximum in this example.

To calculate the i-th member ($f_{1,i}^-$) of this ensemble, we use a scaled version of $S_0(t)$ for deconvolution given by $S_{0,i}(t) = S_0(t) \cdot \alpha_i$, with the frequency independent scaling coefficient α_i. Here, the source wavelet $S(t)$ is scaled by the factor α_i going from 0.8 (blue trace) to 1.2 (red trace). Considering the ensemble of f_1^-, we can observe that the primary reflection at -0.1 s scales with α_i. The additional coda event appearing at 0.28 s, however, shows a different behavior. The polarity of this additional coda event changes depending if we underestimate (blue to yellow colors) or overestimate (yellow to red colors) the source wavelet because the necessary cancelling impulse is either too strong or too weak. Primary reflections in $f_{1,i}^-$ do not show a polarity change but simply scale with α_i. This property holds for deeper focusing positions for all events in f_1^- that are present if the source wavelet is scaled correctly, i.e. for primaries and subsequent coda.

Therefore, only for additional unwanted coda events there is a difference between the stack of the absolute values and the absolute value of the stack of all $f_{1,i}^-$ and, thus, we can define a selection criterion to detect additional coda events:

$$\sum_i \left| f_{1,i}^-(x,t) \right| - \left| \sum_i f_{1,i}^-(x,t) \right| > \varepsilon. \tag{5}$$

The time samples of f_1^- are classified as additional coda when the difference between the stack of the absolute values and the absolute value of the stack of all $f_{1,i}^-$ is greater than a specific threshold ε. Figure 4b shows the minuend (solid black line) and the subtrahend (dashed black line) of equation 5 normalized to their maximum with a gain function of

$\exp(10t)$ within the time window. Only in the case of unwanted coda events, the absolute value of the stack is significantly smaller than the stack of the absolute values of $f_{1,i}^-$ because amplitudes cancel out when the polarity changes. The normalized left-hand side of the selection criterion (equation 5) is shown in black in Figure 4c, where a Gaussian smoothing filter is applied to the difference of the stacks to compensate the nodal points of the coda event wavelet. Only for the coda events, the selection criterion lies above the threshold that is set to 20% of the maximum of the left-hand side of equation 5 in this example (Figure 4c, green line). Thus, Equation 5 provides a criterion for identifying the additional coda events that we aim to minimize in order to estimate the source wavelet scaling. To identify the optimal focusing position for the source wavelet estimation, we first scan through various focusing positions in the subsurface and chose the focusing position where the left-hand side of equation 5 is maximal. Having found the optimized focusing position, we can extract the additional coda events from f_1^- using a tapered window based on the result of the selection criterion.

The importance of minimizing only the energy of unwanted coda events and not of the entire f_1^- focusing function is illustrated in Figure 5. The energy, defined as the sum of the squared amplitudes, of the coda window and the one of the entire f_1^- are shown in blue and red, respectively, for different α values. Again, α represents the (frequency independent) scaling coefficient of the source wavelet used for deconvolution in relation to the true source wavelet scaling.
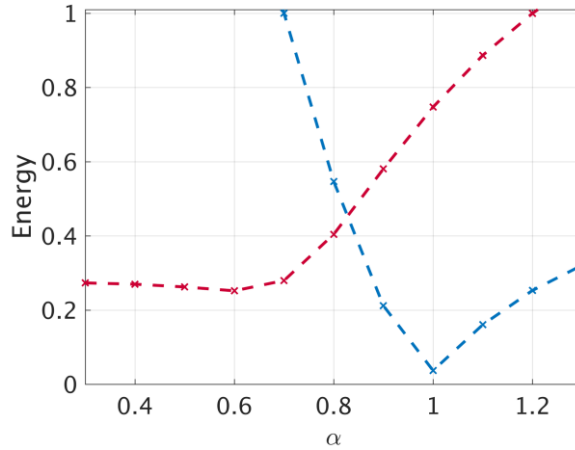
Figure 5 Energy of the entire trace of f_1^- (red) and of the windowed coda only (blue) for different relative source wavelet scaling values α. Only for the energy of the coda, the minimum is located at the correct relative source scaling of $\alpha = 1$.

We observe minimal energy in case of the coda window (blue trace) for the true source scaling of $\alpha = 1$. However, the energy of the entire f_1^- (red trace) is not minimal for the true source scaling, but it is minimal for $\alpha \approx 0.6$ because of the scaling (due to α) of the primary reflections. Thus, minimizing the energy of the complete f_1^- in order to invert for the source wavelet will lead to an incorrect solution. The analysis of the amplitude spectrum of additional coda events allows one to invert not only for the overall scale but for the amplitude spectrum of the source wavelet.

<u>Characterization of misfit function</u>

As shown above numerically, using an incorrectly scaled source wavelet for deconvolution prior to Marchenko focusing creates additional coda events in f_1^- that change polarity depending on an under- or overestimation of the source wavelet. Using an initial $S_0(\omega)$ for deconvolution instead of the true source wavelet $S(\omega)$, we can define a frequency dependent error $\hat{\alpha}(\omega)$ such that

$$[S_0(\omega)]^{-1}S(\omega)R = \hat{\alpha}(\omega)R \qquad (6)$$

holds. The parameter $\hat{\alpha}$ is realted to the previously used α by $\hat{\alpha} = \alpha^{-1}$. Introducing this error in equation 4 means replacing R by $\hat{\alpha}(\omega)R$. Expanding the sum in equation 4, we can formulate two equations for the upgoing focusing function, namely a correct version f_1^- and an erroneous version $f_{1\alpha}^-$ carrying the error given by equation 6 as

$$f_1^- = \theta R f_{1d}^+ + \theta R[\theta R^*\theta R]\, f_{1d}^+ + \theta R[\theta R^*\theta R]^2\, f_{1d}^+ + \cdots \qquad (7)$$

$$f_{1\alpha}^- = \theta\hat{\alpha}R f_{1d}^+ + \theta\hat{\alpha}R[\theta(\hat{\alpha}R)^*\theta\hat{\alpha}R]\, f_{1d}^+ + \theta\hat{\alpha}R[\theta(\hat{\alpha}R)^*\theta\hat{\alpha}R]^2\, f_{1d}^+ + \cdots. \qquad (8)$$

We drop the frequency dependency (ω) for clarity and assume that the phase spectrum of the source wavelet is correct such that $|\hat{\alpha}(\omega)| = \hat{\alpha}(\omega)$ holds. The effect of an incorrect phase spectrum is analyzed in the next section separately. Assuming that the order of $\hat{\alpha}$ and θ can be changed, which is valid for events away from the boundaries of the windowing function θ such that events do not interfere with the windowing, we can summarize the $i\text{-}th$ summand in equation 7 with c_i and thus simplify equations 7 and 8 to equations 9 and 10:

$$f_1^- = c_1 + c_2 + c_3 + \cdots \qquad (9)$$

$$f_{1\alpha}^- \approx \hat{\alpha}c_1 + \hat{\alpha}^3 c_2 + \hat{\alpha}^5 c_3 + \cdots. \qquad (10)$$

As seen before, additional coda events appear in f_1^- if the source wavelet is scaled incorrectly. Thus, such events can be expressed as the difference between f_1^- and $f_{1\alpha}^-$ denoted as Δf_1^-:

$$\Delta f_1^- = f_1^- - f_{1\alpha}^- = c_1 - \hat{\alpha}c_1 + c_2 - \hat{\alpha}^3 c_2 + c_3 - \hat{\alpha}^5 c_3 + \cdots$$

$$= c_1(1 - \hat{\alpha}) + c_2(1 - \hat{\alpha}^3) + c_3(1 - \hat{\alpha}^5) + \cdots$$

$$= (1 - \hat{\alpha})\left[c_1 + c_2 \sum_{j=0}^{2} \hat{\alpha}^j + c_3 \sum_{j=0}^{4} \hat{\alpha}^j + \cdots\right] \qquad (11)$$

The polarity changing behavior of additional coda events described earlier (Figure 4a) can be directly inferred from equation 11. The difference Δf_1^- changes its sign around $\hat{\alpha} = 1$

due to the factor of $(1 - \hat{\alpha})$. Primaries and subsequent related coda are present in both f_1^- and $f_{1\alpha}^-$ such that an incorrect source scale (represented by $\hat{\alpha}$) leads to increased or decreased amplitudes of these events. The additional unwanted coda events, however, are only present for the focusing function using the incorrectly scaled source wavelet, i.e. for $f_{1\alpha}^-$. Therefore, only these events change their sign for overestimating ($\hat{\alpha} < 1$) or underestimating ($\hat{\alpha} > 1$) the scale of the source wavelet. As expected, $\Delta f_1^- = 0$ for the true scaling of $\hat{\alpha} = 1$ and no additional coda events appear in this case.

Frequency dependency of coda

Minimizing the energy of additional coda events in f_1^- overall allows us to invert and determine the correct source scaling (i.e. amplitude spectrum) in case of a frequency independent error of the source wavelet, as shown before (Figure 5). However, in realistic applications, the estimated (initial) source wavelet is often not only a scaled version of the true source wavelet, but carries a frequency dependent error. Prior to an inversion example for a frequency dependent error, we demonstrate the effect of such an error on the additional coda events in f_1^-.

If only distinct frequencies (or frequency bands) ω_i of the source wavelet are scaled incorrectly, the spectra of coda events in Δf_1^- given by equation 11 are dominated by the same frequencies ω_i. Energy of correctly scaled frequencies on the other hand is theoretically not present in the spectrum of the additional coda events because $\hat{\alpha}(\omega) = 1$ and $\Delta f_1^-(\omega) = 0$ in that case, considering the definitions for $\hat{\alpha}(\omega)$ and $\Delta f_1^-(\omega)$ given by equations 6 and 11. This holds, if we can neglect leakage effects to neighboring frequencies due to the windowing of the coda event. Thus, the amplitude spectra of unwanted coda events allow us to draw conclusions on the true amplitude spectrum $S(\omega)$. Erroneous frequencies of the estimated or

initial source wavelet $S_0(\omega)$ are reflected in additional coda events in f_1^- dominated by the same frequencies.

The connection between the source wavelet error $\hat{\alpha}(\omega)$ and the spectrum of the coda event in $\Delta f_1^-(\omega)$ is shown numerically in Figure 6. The spectrum of the true source wavelet (24 Hz Ricker wavelet) $S(\omega)$ is shown in black in Figure 6a together with the spectra of three different scenarios shown in blue, red and yellow. For these three scenarios, we overestimate the true source wavelet at peak frequencies of 17.66 Hz (blue), 24 Hz (red), and 30.46 Hz (yellow) and taper the error to zero using a Hann window. The wavelets obtained in this way are used for deconvolution, introducing an error in the deconvolved reflection response according to the erroneous frequencies. Subsequently, we calculate f_1^- for the same configuration and focusing positon $(x, z) = (0,900)$ as before for each of the three scenarios and analyze the amplitude spectrum of the additional coda event identified previously. Figure 6b depicts the normalized amplitude spectra for the three different cases using the same colors as for the source wavelet spectra in Figure 6a. For each of the three scenarios, maximal coda energy is observed at the respective peak frequency of the overestimation of the source wavelet indicated by colored squares in Figure 6b. Coda energy is also observed at neighboring frequencies because of the tapering of the overestimation and because of leakage effects due to the windowing of the coda event.
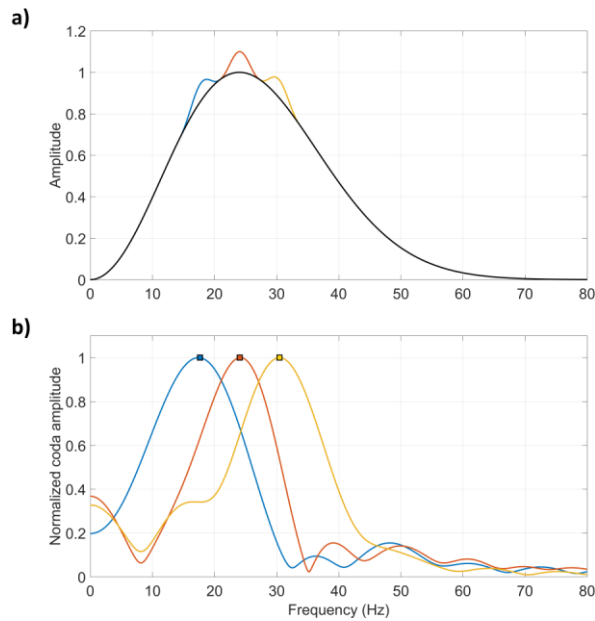
Figure 6 (a) Amplitude spectra of true source wavelet (black) and 3 perturbed (incorrect) source wavelets where frequencies at 17.66 Hz (blue), 24 Hz (red), and 30.46 Hz (yellow) are overestimated. (b) Resulting normalized amplitude spectra of observed coda using the incorrect source wavelets from panel a) for deconvolution. The colors correspond to the three scenarios from panel a). The corresponding frequencies of maximal perturbation are marked as colored squares and coincide with the respective spectral maximum of the observed coda.

<u>Inversion example</u>

An error at a specific frequency in the source wavelet leads to unwanted coda events in f_1^- with the same peak frequency, as analyzed and demonstrated above. This relates a frequency dependent error of the source wavelet to the spectrum of an observed coda event. Thus, we can invert for the amplitude spectrum of the source wavelet by minimizing the amplitude spectrum of additional coda events in f_1^-. To this end, we use the following iterative workflow:

i.) identification of optimal focusing position and determination of coda event with ensemble $f_{1,i}^-$;

ii.) windowing of coda from each $f_{1,i}^-$ and calculation of the amplitude spectrum $A(\omega, \alpha_i)$ of coda events for every $f_{1,i}^-$;

iii.) for each frequency ω, determination of α_i where A is minimal at selected ω resulting in frequency dependent scaling coefficient $\alpha(\omega)$;

iv.) update of the initial source wavelet by $S_{updated}(\omega) = S_0(\omega) \cdot \alpha(\omega)$;

v.) initiation of next iteration at ii.) with new ensemble $f_{1,i}^-$ created with $S_{updated}(\omega)$.

We can set the stopping criterion and estimate the uncertainty of the inversion using for example the noise level in the acquired data or, ideally, in f_1^-. We chose a window with the same length as the window used in the inversion such that neither coda nor primary reflections are present. Such a window is for instance depicted in Figure 4a by dashed black lines. We calculate the amplitude spectrum of this window and stop the update of the source wavelet amplitude spectrum when the observed coda for the current iteration is smaller or equal to the observed noise amplitude at each frequency. Depending on the noise level, this criterion defines a natural limit of this approach because further updates would potentially minimize only noise and not actual additional coda events. Thus, the smallest and largest possible scaling values $\alpha(\omega)$ that create a coda event having the same spectral amplitude as the noise build the lower and upper bound of the inverted source wavelet amplitude spectrum and provide information about the uncertainty of the inversion. In practice, however, it may be hard to identify such a window in order to define the stopping criterion as described. In that case, we stop the iterations when scaling updates are small compared to the previous iteration.
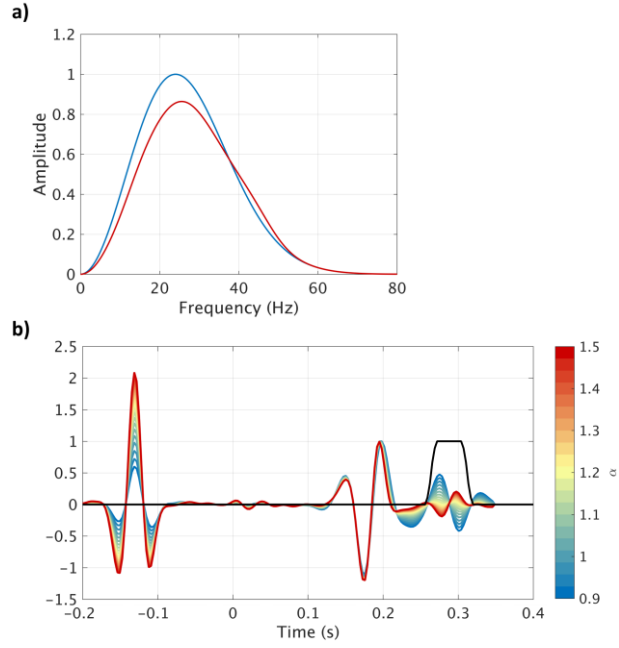
Figure 7 (a) True (blue) and initial (red) source wavelet amplitude spectrum. (b) Ensemble of f_1^- for coda identification for $0.9 \leq \alpha_i \leq 1.5$. The identified (tapered) coda window is shown in black.

By iteratively minimizing the amplitude spectrum of the determined coda event, the inversion converges to the amplitude spectrum of the true source wavelet as shown in the following example. We use the focusing function at $(x, z) = (0,900)$ m and initiate the Marchenko focusing with a 24 Hz Ricker wavelet for f_{1d}^+. The amplitude spectrum of the true source wavelet $S(\omega)$ (24 Hz Ricker wavelet) is shown in blue together with the initial wavelet $S_0(\omega)$ in red in Figure 7a. In this example, we underestimate the amplitudes between 0 and 38 Hz by a variable factor going from 0.7 to 1.0 and overestimate the amplitudes between 38 and 55 Hz by a variable factor between 1.0 and 1.2 with a maximum at 48 Hz. The first ensemble $f_{1,i}^-$, with α_i going from 0.9 to 1.5 (Figure 7b, blue to red colors), shows the polarity changing behavior of the additional coda event at 0.28 s. Thus, a tapered window, shown as solid black line, can be defined as described above.

23

After the first update, the inverted source spectrum is closer to true solution as one can see from the spectra in Figure 8a ($S(\omega)$ in blue and $S_1(\omega)$ in red). Here, the true wavelet is underestimated between 0 and 20 Hz and overestimated between 20 and 55 Hz. The amplitude spectra for all members $f_{1,i}^-$ of the ensemble of the second iteration (for $0.9 < \alpha_i < 1.1$) are shown in black and white in Figure 8b where the red line depicts the line of minimal energy for each frequency. As expected, minima of $A(\omega, \alpha_i)$ are found for $\alpha_i > 1$ where $S_1(\omega)$ is underestimated (0 to 20 Hz) and for $\alpha_i < 1$ where $S_1(\omega)$ is overestimated (20 to 55 Hz).
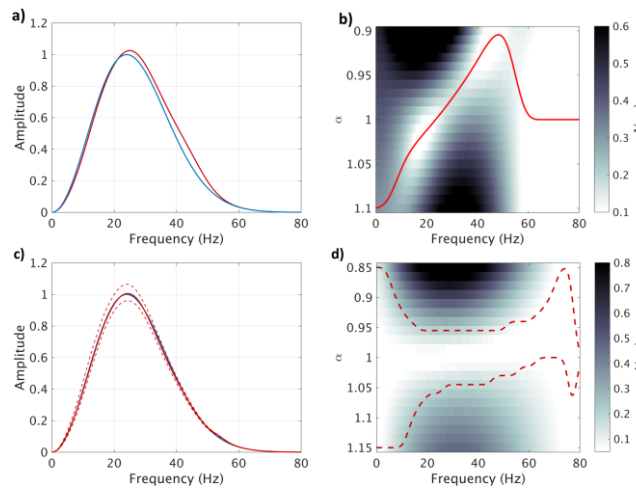

Figure 8 (a) Amplitude spectra of the true source wavelet (blue) and the result after the first iteration (red). (b) Objective function $A(\omega, \alpha_i)$ for the source wavelet scaling after the first iteration (panel a, red). Minimal energy (red line) is observed for scaling values α_i corresponding to the frequency dependent scaling error of the amplitude spectrum. (c) and (d) same as (a) and (b) now for the final iteration. Lower and upper bounds of scaling (according to the noise level in f_1^-) are depicted by red dashed lines.

After 7 iterations, both spectra match very well (Figure 8c) and for all frequencies, the minimal energy is found close to a scaling of 1 such that an update would not change the source wavelet significantly (Figure 8d). Additionally, we can define the level of uncertainty as indicated by red dashed lines in Figure 8c and 8d based on the window as defined before

PHASE SPECTRUM INVERSION

In the previous analyses and examples, we only consider an error in the amplitude spectrum of the source wavelet and assume that the phase spectrum is correct. However, as shown in the following, the upgoing focusing function f_1^- can also be used to invert for the phase spectrum of the source wavelet independently of the amplitude spectrum inversion. If we define $X(\omega)$ as the ratio of the incorrect version of the focusing function $f_{1\alpha}^-$ (equation 10) and the true f_1^- (equation 9), $X(\omega)$ reads as

$$X(\omega) = \frac{f_{1\alpha}^-}{f_1^-} = \alpha(\omega) \left\{ \frac{c_1 + \alpha(\omega)^*\alpha(\omega)c_2 + [\alpha(\omega)^*\alpha(\omega)]^2 c_3 + \cdots}{c_1 + c_2 + c_3 + \cdots} \right\}. \quad (12)$$

If the amplitude spectrum of the estimated source wavelet is correct (i.e., $|\alpha(\omega)| = 1$), then $\alpha(\omega)^*\alpha(\omega) = 1$, such that $X(\omega) = \alpha(\omega)$ and

$$f_{1\alpha}^- = \alpha(\omega)f_1^-. \quad (13)$$

Having only an error in the phase spectrum of the source wavelet, we can derive from equations 12 and 13 that f_1^- is distorted by an overall phase error $\alpha(\omega)$. Figure 9 shows different versions of the trace at $x = 0$ m of f_1^- for the focusing position $(x, z) = (0,900)$. These traces are calculated using phase shifted (correct amplitude spectrum) versions of the true source wavelet for deconvolution prior to Marchenko focusing. The source phase shift φ ranges from $-90°$ (blue to yellow traces) to $+90°$ (yellow to red traces). As before, we initiate the Marchenko scheme with a 24 Hz Ricker wavelet for f_{1d}^+. If the phase spectrum of the estimated source wavelet is correct, we expect all events in f_1^- to be time shifted versions of f_{1d}^+ or 90° phase shifted versions of f_{1d}^+, if a signal in f_1^- crosses a caustic (Barash, 1968; Bakker, 1998). As one can see from the traces in Figure 9, the source phase shift φ shifts the entire f_1^- trace according to the introduced error. Only for the $\varphi = 0°$ trace depicted in black, the two events correspond to a Ricker wavelet (or a 90° phase shifted version).
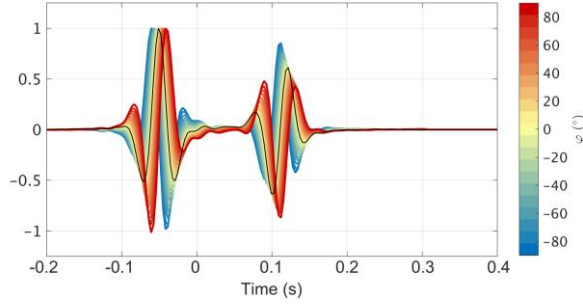
Figure 9 Trace segment of f_1^- for a focusing position at $(x, z) = (0,900)$ m using altered versions of the true source wavelet for deconvolution that are phase shifted frequency independently by φ_i for $-90° \leq \varphi_i \leq +90°$. No additional coda event is observed and f_1^- is phase shifted by the phase error of the source wavelet. The true solution ($\varphi_i = 0°$) is shown in black.

Contrary to an incorrectly scaled amplitude spectrum, an erroneous phase spectrum does not create additional coda events, but distorts all primaries in f_1^- by the phase difference between the true and (phase shifted) estimated source wavelet. This confirms the analytic findings derived from equations 12 and 13. The independent effects of an erroneous amplitude and phase spectrum on f_1^- allow us to invert for these spectra independently from each other. To invert for the phase spectrum, we exploit the phase difference between primary reflections in f_1^- and the wavelet used to create f_{1d}^+, since the difference should be zero if the phase spectrum of the source wavelet is estimated correctly. Thus, we can define a similar misfit function and workflow as for the amplitude spectrum inversion. As before, we create an ensemble of f_1^- where every member $f_{1,i}^-$ is calculated using, for deconvolution, an altered version of $S_0(\omega)$ that is phase shifted by φ_i. Such an ensemble is shown in Figure 9 for $-90° \leq \varphi_i \leq +90°$. From each member of that ensemble, we extract and window one primary reflection and align it with the f_{1d}^+ wavelet. In the frequency domain, we calculate the misfit function $P(\omega, \varphi_i)$ as the unwrapped phase difference between f_{1d}^+ and the extracted

26

primaries of every $f_{1,i}^-$. As before, we can determine a frequency dependent phase correction such that we minimize the phase difference P for each frequency ω.
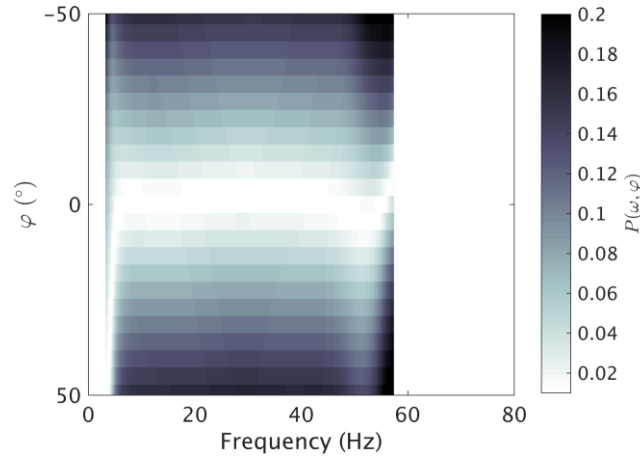

Figure 10 Objective function $P(\omega, \varphi_i)$ defined as the phase difference between primaries in f_1^- and the wavelet used to create f_{1d}^+ using phase shifted versions of the true source wavelet for deconvolution. Minimal energy is observed at around $\varphi = 0°$ for the dominant frequency range.

The misfit $P(\omega, \varphi_i)$ for the ensemble shown in Figure 9 is displayed in black and white in Figure 10. For stabilization, $P(\omega, \varphi_i)$ is set to 0 for frequencies where the amplitude of f_{1d}^+ is smaller than 5% of the maximal amplitude. In the dominant frequency range, small misfit values are found for all frequencies close to $\varphi = 0°$. As expected, the phase difference between the primary in f_1^- and the f_{1d}^+ wavelet is minimal for the $\varphi(\omega) = 0°$ member of the ensemble of f_1^-, which is created based on the correct source wavelet without a phase error. However, if there is a phase error in the estimated or initial source wavelet, the line of minima in $P(\omega, \varphi_i)$ is not located at $\varphi(\omega) = 0°$ but it will follow $-\Delta\varphi_S(\omega)$, where $\Delta\varphi_S(\omega)$ is the phase difference between the (incorrectly) estimated source wavelet and the true source wavelet. Similar to the amplitude spectrum update, we correct the phase spectrum of the initial source wavelet with the determined $-\Delta\varphi_S(\omega)$.

27

COMBINED EXAMPLE

The findings from the previous two sections are now combined to invert for both the amplitude and phase spectrum of a source wavelet. Using the same source-receiver configuration as before, we apply the amplitude and phase inversion algorithm to a f_1^- focusing function for $(x, z) = (0,900)$ m. The true and the initially estimated source wavelet are shown in the time domain in Figure 11a in blue and red, respectively. We create the true source wavelet $S(\omega)$ in the time domain by differentiating a Gaussian pulse and applying an arbitrary amplitude variation to its spectrum in the frequency domain (Figure 11b, blue). We estimate the initial source wavelet $S_0(\omega)$ as a Ricker wavelet with a central frequency of 36 Hz (Figure 11b, red). Thus, we overestimate the amplitude spectrum with regard to the true spectrum by a factor varying between 1 and 1.6, as shown in Figure 11b in red and blue, respectively. Using a Ricker wavelet as the estimated source wavelet instead of the derivative of a Gaussian pulse introduces a phase shift of $\Delta\varphi_S(\omega) \approx -72°$. Again, the Marchenko scheme is initiated with a 24 Hz Ricker wavelet for f_{1d}^+.
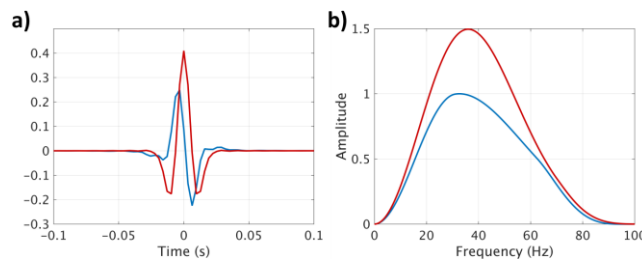

Figure 11 (a) True (blue) and initial (red) source wavelet having an incorrect amplitude and phase spectrum. (b) Amplitude spectra of wavelets shown in panel a).

As shown before, amplitude and phase inversion can be treated independently from each other. Therefore, we first invert for the amplitude spectrum minimizing the present determined coda event in f_1^-, followed by the phase spectrum inversion. The final result after 15 iterations for the amplitude spectrum inversion and subsequent phase spectrum

estimation is shown in Figure 12a and 12b in the time and frequency domain. The final inverted (red) and true source wavelet (blue) match very well in the time domain indicating that the phase spectrum is estimated correctly. An accurate match can also be observed in the frequency domain. The amplitude spectrum of the true source wavelet lies within the bounds of uncertainty of the inversion result in the dominant frequency band (up to 70 Hz) indicated by red dashed lines in Figure 12b. This example shows that we can successfully invert for a source wavelet using a starting source wavelet with an incorrect amplitude and phase spectrum. The two step approach with separate misfit functions $A(\omega, \alpha_i)$ (for the amplitude spectrum) and $P(\omega, \varphi_i)$ (for the phase spectrum) simplifies the inversion and might also be beneficial in combination with other source wavelet inversion methods, e.g. the ones proposed by Velis and Ulrych (1996) or Misra and Sacchi (2007).
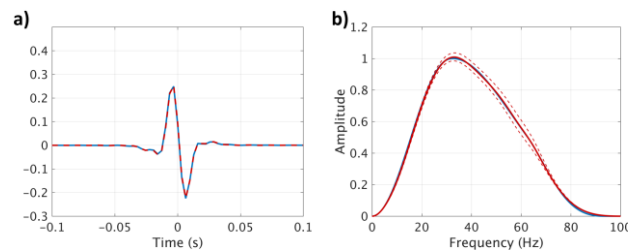

Figure 12 (a) True (blue) and final inverted (red dashed) source wavelet. (b) Amplitude spectra of true (blue) and final inverted (red) source wavelet. Lower and upper scaling bounds are shown in red dashed lines.

LATERALLY-VARYING SOURCE SCALING

In the previous examples, we demonstrate that we are able to invert successfully for an incorrect source wavelet having a frequency dependent amplitude and phase error. However, in the given examples the error is the same for all 901 sources at the surface. Now, we extend the approach to invert for laterally-varying scaling errors of sources at the surface by minimizing the energy of additional coda events for multiple, laterally-spread focusing positions in the subsurface. As a reference source wavelet, we model the far-field response of an air gun array (Landrø, 2017 get reference for code from Johan) as shown in Figure 13a. The modelled air gun array is based on the configuration of the M/V Kondor air gun array, used in 2003 with a total volume of 2590 cubic inches distributed over 28 air guns (DeRuiter et al., 2006). In Figure 13a, an air gun is denoted by a red circle, where the diameter of the circle is proportional to the air gun volume, varying between 20 and 250 cubic inches. The far-field response in the time and frequency domain is shown in Figure 13 b and c, respectively. Maximal amplitudes are observed between 30 Hz and 40 Hz.

As shown before, a phase error of the source wavelet only introduces an overall phase error in f_1^- and therefore also in the redatumed Green's functions. The accuracy of the Marchenko focusing, however, does not decrease because of phase errors in the source wavelet. In case of amplitude errors of the source wavelet, on the contrary, we observe additional events in f_1^-, resulting in imperfect focusing and thus also in artifacts in redatumed Green's functions and images created with redatumed data. Ultimately, the entire Marchenko focusing is unstable if the true source wavelet is underestimated significantly. Therefore, we focus here on laterally-varying scaling errors of sources at the surface, while the phase spectra are assumed to be correct.
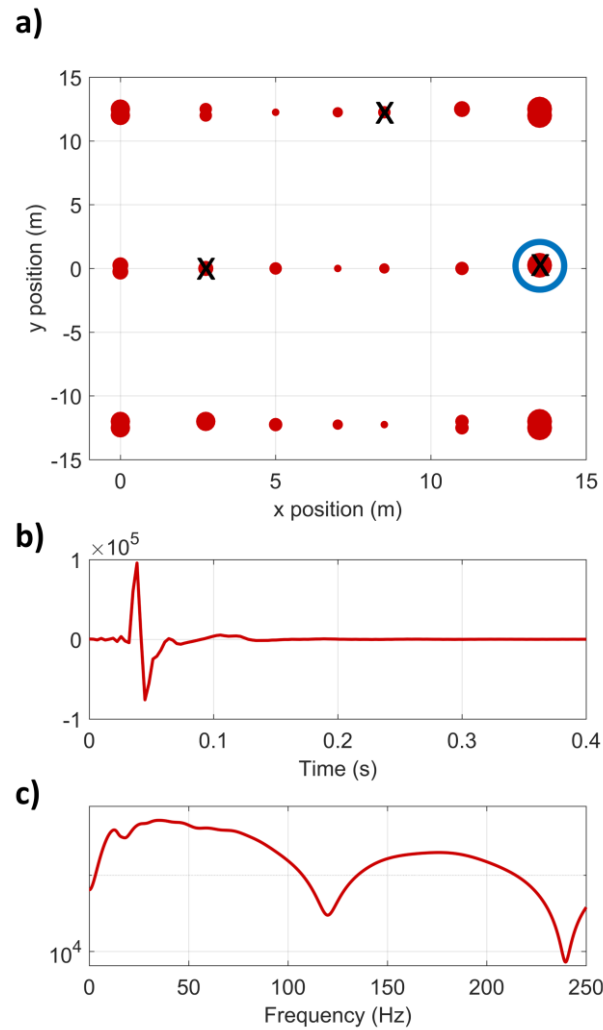
a)

b)

c)

Figure 13 (a) Air gun array geometry. The diameters of red circles correspond to the volume of the individual air guns varying between 20 and 250 cubic inches. The blue circle and the black crosses mark air guns that fail for specific shots for the second example given in this section. (b) Far-field response of the complete array. (c) Amplitude spectrum of the far-field response shown in panel b).

The inversion algorithm is explained using the same velocity model as before with 901 sources and receivers at the surface. Defining the previous uniform source scaling as 1, the (true) relative scaling of sources between the lateral positions of -1400 m and $+1400$ m varies now laterally, smoothly undulating between 0.8 and 1.15 as shown in black in Figure

14a. As a starting model, we use a homogenous relative scaling of 1 for all sources as shown in red in Figure 14a. The identically scaled source wavelets are used for deconvolution prior to the Marchenko focusing for the first inversion iteration. In contrast to previous examples, we now aim to minimize the coda that appears due to the incorrect source scaling for 21 laterally spread focusing positions located at a depth of 900 m between – 1000 and +1000 m with a spatial sampling of 100 m (Figure 14b, black triangles). We invert for the relative scaling of the individual sources using a Gauss-Newton based inversion scheme (e.g., Nuber et al., 2015).
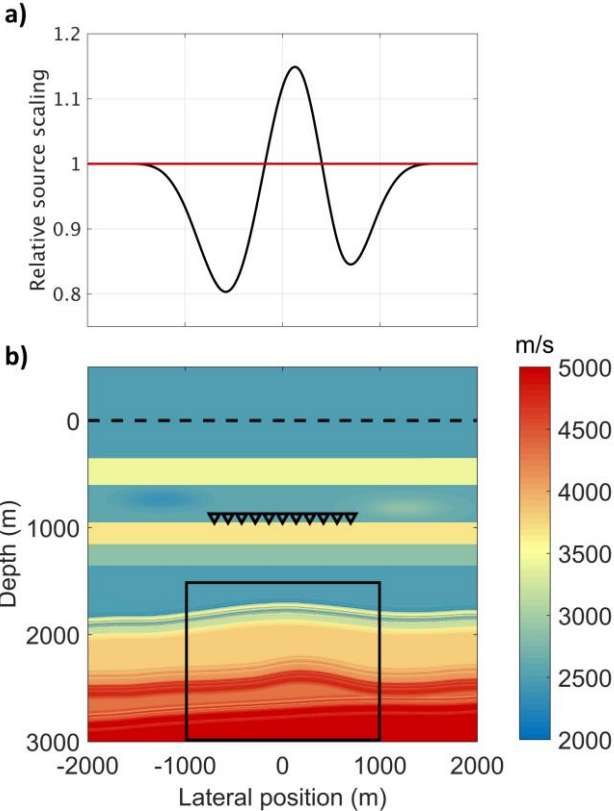

Figure 14 (a) True (black) and initial (red) relative source scaling as a function of the lateral position. (b) Same velocity model as in Figure 1, now with multiple focusing positions at a depth of 900 m indicated with black triangles. Images of the target zone denoted by the black box are used to evaluate the performance of the source scaling inversion.

The update of the model parameters m_i (i.e., the relative scaling of the individual sources) of the i-th iteration is given by

$$m_{i+1} = H^{-1}\{J^T[Jm_i - A] + \varepsilon^2 Im_i\},$$

where the Hessian matrix H is approximated using the Jacobian matrix J:

$$H = J^T J + \varepsilon^2 I.$$

The matrix J contains the sensitivity of the observed unwanted coda for all focusing positions with respect to the model parameters:

Here, I is the identity matrix and the data misfit is represented by the matrix A that contains the energy of the additional coda events in f_1^- for the 21 focusing positions. For our application, the objective function does not describe the difference between observed and predicted data, but the energy of the observed additional coda events in f_1^- for multiple focusing positions collected in the matrix A. We aim to optimize the model parameters such that we minimize the additional coda events, i.e. our objective function.

The sensitivity of the coda at the focusing positions i with respect to the scaling of the source j can be expressed as

$$J_{ij} = \frac{\partial A_i}{\partial m_j}$$

$$\approx \frac{\Delta A_i}{\Delta m_j}.$$

For every iteration, we approximate the sensitivity matrix J numerically. To this end, we perturb a subset of model parameters by Δm (corresponding to ∂m_j), compute f_1^- at every focusing position for the perturbed state, and calculate the difference of the additional coda for each focusing position with respect to the unperturbed state ($\partial A \approx A(m + \Delta m) - A(m)$). We move the window containing the perturbed subset of model parameters along the line of sources at the surface, where the perturbed windows overlap such that a single source is

perturbed multiple times, depending on the chosen overlap. The final sensitivity of A at one focusing position with respect to a single source scaling is then obtained by averaging ∂A for all the f_1^- computations where the respective source is part of the perturbed subset of sources. Thereby, we obtain a smooth Jacobian matrix, where the smoothing depends on the window length of the perturbed subset of model parameters and on the chosen overlap. We stabilize the inversion using the damping parameter ε, which minimizes the deviation of the model parameters from the ones of the previous iteration (Nuber et al., 2015). Further smoothing constraints are not needed for the inversion because of the inherent smoothing resulting from the Jacobian matrix calculation. At each iteration, ε is found by a line search minimizing the observed coda for 3 chosen focusing positions only, to reduce the computational cost. In this example, we set the perturbation window for calculating J to 10 sources with an overlap of 80% and we calculate J for sources between -1500 and $+1500$ m. In other words, we also invert only for the scaling of sources at the surface being in that interval. As before, we stop the iterative inversion when the coda energy at every focusing position is within the noise level in f_1^- or when no significant source scaling updates occur for the next iteration.

The final inversion result for 8 iterations is shown in Figure 15. The match of the actual true source scaling in black and the final inverted source scaling in red indicates that the proposed inversion approach enables to invert for laterally-varying scaling errors correctly.
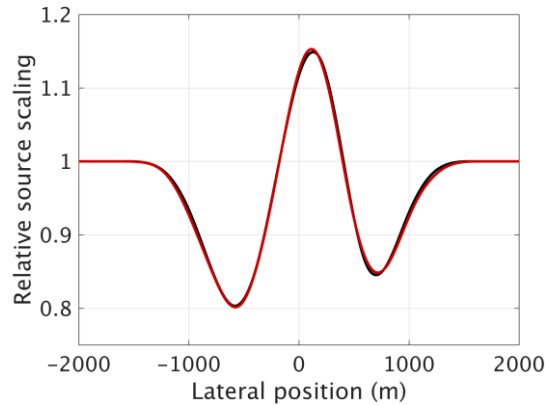
Figure 15 True (black) and final inverted (red) relative source scaling as a function of the lateral position.

Additionally, we demonstrate the importance of the correct source scaling with images created with redatumed data. We define the medium below a depth of 1500 m as target zone covering three laminated undulating or dipping interfaces. The target zone is depicted in Figure 14b with a black box.

Having the inverted source scaling from the previous inversion (Figure 15), we define the following workflow in order to image the target zone:

1) we deconvolve the surface reflection response with the initial incorrect source scaling (Fig 16a, red line).

2) For the redatuming of sources to the target depth just above the target zone, we calculate the redatumed decomposed (into up- and downgoing wavefields) Green's function for 201 virtual sources (VS) from -1000 to 1000 m at a depth of 1500 m solving for the coupled Marchenko equations (equations 1 to 4).

3) Create a redatumed reflection response for virtual sources and receivers at the redatuming level by multi-dimensional deconvolution (see van der Neut et al., 2011, Wapenaar et al., 2011, and Broggini et al., 2014a for details). Thus, we obtain a

redatumed reflection response for 201 virtual sources and receivers at the target depth of 1500 m where, ideally, the internal multiples originating in the overburden are removed.

4) We use the redatumed data to create images from the target zone using reverse-time migration (RTM) and create the final image by stacking the individual images created for all virtual sources.

We repeat the same workflow for the inverted source scaling, i.e., we only alter the first step and deconvolve the surface reflection response with the final inverted source scaling (Figure 15, red line). All other steps remain unchanged. The final images using the initial and the final inverted source scaling for deconvolution are shown in Figure 16a and 16b, respectively.
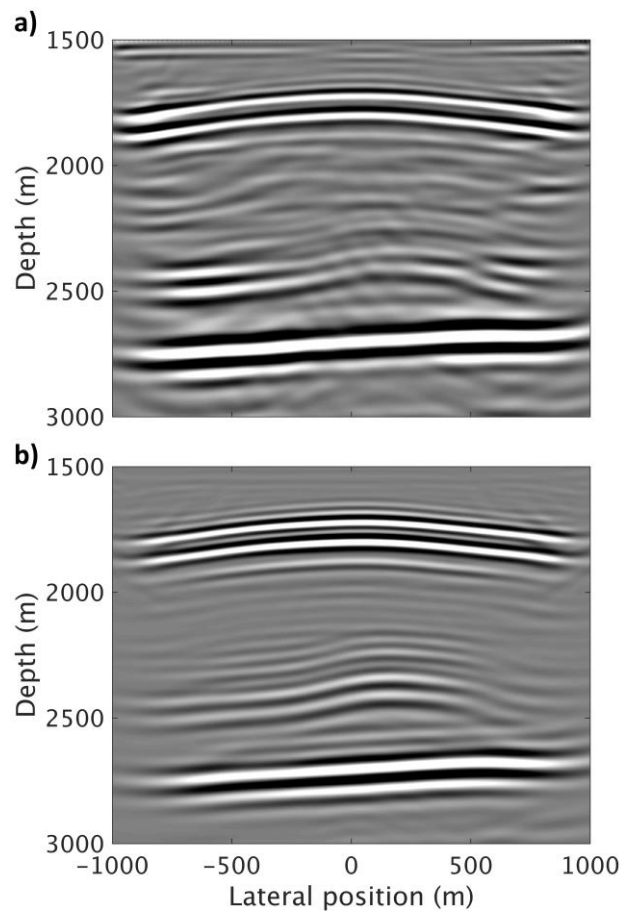
Figure 16 Images of the target zone (Figure 14b, black box) created with Marchenko redatumed data and RTM using the initial (panel a) and the final inverted (panel b) source wavelet scaling for deconvolution prior to Marchenko redatuming.

Using the inverted source scaling for deconvolution (Figure 16b) clearly improves the image compared to the image where the initial source scaling is used (Figure 16a). Internal multiples appearing between 2000 m and 2400 m in Figure 16a are suppressed in 16b, reflectors (especially the second reflector at 2500 m) are more consistent, and more details are visible, such as a weak laminated reflector (e.g., at $(x, z) = (0,2300)$) above the second strong reflector at 2500 m.

In a second example of laterally-varying source scaling inversion, we demonstrate the effect of sharp transitions of source scaling. We create the reflection data at the surface using the far-field response of the same air gun array as before (Figure 13a) as source wavelet. However, for sources between $-1000 \leq x \leq +1000$ m, we model the data using the far-field response for an altered version of the full array. For sources between -1000 m and 0 m one air gun marked by a blue circle in Figure 13a, fails. For $0 < x \leq +1000$ m, we use the far-field response for a version of the array where 3 air guns marked by black crosses in Figure 13a fail. By simulating varying air gun failures, the amplitude of the far-field response decreases because of the overall reduced volume of the air gun array. This amplitude decrease is nearly frequency independent, such that we can invert for the lateral source scaling as in the previous example.

The true laterally-varying source scaling resulting from the described air gun failures is shown in Figure 17 (black line). For the intervals with one or three failing air guns, the source scaling relative to the full array response is about 0.95 and 0.88, respectively. As initial source

scaling that is used for deconvolution at the first iteration, we assume that the full array was working for all shots, i.e. the relative initial scaling equals 1 for all shots (Figure 17, red line). Thus, we overestimate the source scaling for sources between -1000 m and $+1000$ m due to the failing air guns. The inversion result after 12 iterations using the Gauss-Newton based approach, as described previously is shown in blue in Figure 17. Overall, the true scaling is inverted correctly. However, the sharp transitions at -1000 m, 0 m and $+1000$ m are not recovered correctly. The inherent smoothing resulting from the Jacobian calculations leads to a smoothed inversion result. Choosing smaller perturbation windows with less overlap when calculating J would help to reduce the smoothing. In the ultimate case, one would perturb every source individually in order to calculate J for each source. However, this approach increases the computational cost significantly. In general, the proposed inversion algorithm is able to invert for the laterally-varying source scaling effectively, where the accuracy depends on the complexity of the lateral variation.
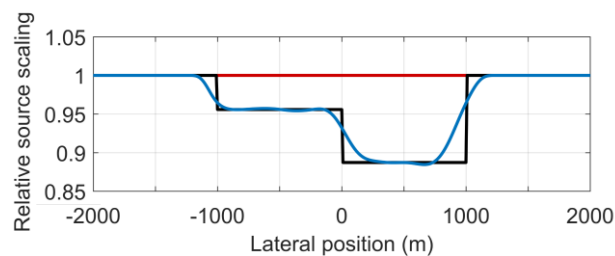

Figure 17 The true relative source scaling resulting from failing air guns is shown in black together with the initial (red) and final inverted (blue) relative source scaling as function of the lateral position.

DISCUSSION

One crucial requirement for a successful Marchenko redatuming is the correct deconvolution of the source wavelet from the surface reflection response, such that the deconvolved data represents the true reflection response of the medium below the surface. The methodology proposed in this paper allows to invert for both the amplitude and the phase spectrum of the source wavelet using the upgoing focusing function f_1^- obtained by the Marchenko method.

If only the phase of the estimated source wavelet, and thus of the reflection data after deconvolution, is incorrect, the focusing functions and the redatumed Green's functions obtained by the Marchenko method are phase shifted by the inherent phase error. However, the Marchenko focusing, Green's function retrieval, and subsequent imaging are not corrupted by an incorrectly estimated phase spectrum. An erroneous amplitude spectrum of the source wavelet, on the other hand, implies an incorrect scaling of the reflection data. In that case the quality of the Marchenko focusing is reduced and there is residual multiple energy left in the f_1^- wavefields. Consequently, images created with the corresponding redatumed data still contain internal multiples from the overburden that are not present if the image is created with correctly redatumed data. Thus, knowing the amplitude spectrum of the source wavelet, i.e. the scale of the reflection data, is essential for successful Marchenko imaging as also discussed by van der Neut et al. (2015b) and Brackenhoff (2016).

The key step of the scaling inversion algorithm presented here is the detection and identification of additional coda events in f_1^- that we aim to minimize in order to invert for the amplitude spectrum of the source wavelet. As presented here, our approach requires isolated unwanted coda events in f_1^- that do not interfere with primary reflections. However, the focusing depth used for the source wavelet inversion and the actual target redatuming

depth do not have to coincide. Finding a single focusing position where such an event can be identified suffices for the amplitude spectrum estimation. In contrast to the redatumed Green's functions, the focusing functions only contain interactions with the medium between the surface and the chosen focusing level (Wapenaar et al., 2014a). Thus, the focusing wavefield can be simplified by choosing an optimal shallow focusing positon, reducing the possible interference of coda events with primary reflections. In order to find such an optimal focusing position, the left-hand side of the selection criterion (equation 5) is evaluated for different (shallow) focusing positions. Eventually, that focusing position is chosen for the amplitude spectrum inversion, where we observe the maximum value of the coda selection criterion. Once the identified coda event is minimized, i.e. the optimal reflection data scaling is found, the data can be redatumed to the actual target depth, followed by possible further imaging or analyses steps as shown in the previous section.

Minimizing the energy of an additional coda event in f_1^- will not yield the correct source scale if this event interferes with an actual primary reflection. However, in that case amplitudes of the interfering primary are still reduced or increased for over- or underestimating the source scaling. Thus, we currently investigate the possibility to extend the proposed methodology such that an isolated coda event is not required. However, this will include another misfit function based, for instance, on the similarity of the individual members of the f_1^- ensemble.

The examples for amplitude spectrum inversion presented here consider smoothly varying errors in the amplitude spectrum of the initially estimated source wavelet. As shown by the analysis of the effect of individual mis-scaled frequencies, the amplitude spectrum of the additional coda event is related to the corresponding erroneous frequencies. However, we observe also leakage to neighboring frequencies for instance because of the temporal

truncation (windowing). Using the proposed grid search algorithm to invert for the optimal scale as a function of frequency might be challenging and might take many iterations in case of a strongly oscillating amplitude spectrum because of the described spectral leakage. A more sophisticated inversion approach (e.g., the conjugate gradients method) might improve the inversion algorithm. Furthermore, the analysis and minimization of wavelet amplitude spectra of identified coda events might reduce the issue of leakage because wavelets are limited in time by definition and no temporal windowing would be needed.

Besides the incorrect deconvolution of the source wavelet, we do not consider any other potential error source in the presented study. Possible other factors that might affect the focusing functions and also the presence of unwanted coda events might be mis-located sources or receivers at the surface and incorrectly estimated attenuation in the medium. The applied Marchenko algorithm is derived for a lossless medium. However, if the surface reflection data is not corrected for attenuation, this might have similar effects on the focusing functions as an incorrectly scaled source wavelet (Alkhimenkov, 2017).

As discussed before, the correct scaling of the amplitude spectrum is crucial for an accurate Marchenko redatuming and can be achieved by optimizing f_1^- without requiring more information than conventional Marchenko algorithm. The accuracy of the proposed phase spectrum inversion method, however, might be influenced by the character of reflectors in the medium, for example if a reflector in the subsurface alters the phase of a primary reflection in f_1^-. However, it has been applied successfully to laboratory data, e.g. in Mildner et al. (2017), to correct for the source transfer function of a speaker in a laboratory 1D Marchenko application. Additionally, we have shown that the phase spectrum inversion and the amplitude spectrum estimation can be treated independently of each other.

Therefore, any more sophisticated phase spectrum inversion algorithm can be combined with the proposed amplitude spectrum estimation.

The extension to the inversion for laterally-varying scaling errors relies on the accurate calculation of the sensitivity of the misfit (coda amplitude in f_1^-) with respect to the model parameters (source scaling). Naturally, the calculated sensitivities depend on the analyzed medium and, thus, have to be evaluated for each model independently. In the flat layered example as shown here, we observe clear maxima of sensitivities with respect to sources located vertically above the focusing positon. This might not be the case for more complex overburden models. Therefore, a more detailed analysis for more complex media and with respect to possible errors in the velocity model of the overburden could provide further insights. An additional extension of the methodology is also the implementation and testing for a Marchenko algorithm that is capable of handling free surface multiples (Singh et al., 2017).

Finally, it is noteworthy that there are also Marchenko algorithms that require only limited or no information of the source scaling at all. These algorithms, however, are either adaptive implementations of the Marchenko method (Staring et al., 2017) or require the acquisition of multi component one-way reflection data (Ravasi, 2017; Slob and Wapenaar, 2017). The application of the methodology as presented in this paper on a field dataset is ongoing research and currently prepared for publication.

CONCLUSION

The proposed approach of source wavelet inversion using Marchenko focusing functions enables the correct estimation of the source wavelet inherent in the surface reflection response. Deconvolving the correctly scaled inverted source wavelet from the surface data allows accurate Marchenko redatuming and subsequent target-oriented imaging of the subsurface. The inversion of the amplitude and the phase spectrum can be considered independently from each other because they have different effects on the upgoing focusing function f_1^- used for the source wavelet inversion. Using f_1^- for data scaling inversion proves to be advantageous compared to redatumed Green's functions since the focusing functions are defined for a medium that is reflection free below the focusing level.

An incorrectly scaled amplitude spectrum results in additional unwanted coda events in f_1^- because incompletely cancelled internal multiple energy is present in the focusing wavefields. Depending on over- or underestimating the scale of the source wavelet, the additional coda events change polarity because the cancelling impulse is either over- or underestimated. This property enables to differentiate the unwanted coda events from the actual primary reflections and by minimizing the identified coda events the correct scale of the source wavelet can be determined. Additionally, the amplitude spectra of determined coda events are related to the incorrectly scaled frequencies of the source wavelet. Therefore, a frequency dependent minimization of additional coda events allows the inversion of the source wavelet amplitude spectrum. An incorrect phase spectrum of the source wavelet, however, does not create additional events in f_1^- but distorts the focusing function by the inherent phase error.

Minimizing the phase difference between primaries in f_1^- and the wavelet used to create the initial downgoing focusing function f_{1d}^+ therefore enables to estimate the phase spectrum of the source wavelet.

Additionally, the proposed source wavelet scaling inversion can be extended to laterally-varying scaling errors. Using incorrectly scaled sources for deconvolution prior to redatuming leads to residual internal multiples in images created with inaccurately redatumed data. Optimizing f_1^-, i.e. minimizing identified coda events, for laterally spread focusing positions in the subsurface allows to invert for the source scaling as a function of lateral position. Deconvolving the source wavelet from the surface data scaled according to the inverted source scaling results in accurately redatumed data. Compared to target-oriented images using the inaccurately redatumed data due to mis-scaled sources, the subsequent images are clarified and free of internal multiples. The accurately inverted source wavelet does not only enable precise Marchenko redatuming but might also be beneficial for further analyses such as RTM or AVO analysis.

References

Al-Ali, M.N. and D. J. Verschuur, 2006. An integrated method for resolving the seismic complex near-surface problem: Geophysical Prospecting, **54**, 739–750.

Alkhimenkov, Y., 2017, Redatuming and quantifying attenuation from reflection data using the Marchenko equation: M.S. thesis, Delft University of Technology.

Behura, J., K. Wapenaar, and R. Snieder, 2014, Autofocus imaging: Image reconstruction based on inverse scattering theory: Geophysics, **79**, no. 3, A19-A26.

Brackenhoff, J. A., 2016, Rescaling of incorrect source strength using Marchenko Redatuming: M.S. thesis, Delft University of Technology.

Broggini, F., R. Snieder, and K. Wapenaar, 2014a, Data-driven wavefield focusing and imaging with multidimensional deconvolution: Numerical examples for reflection data with internal multiples: Geophysics, **79**, no. 3, WA107-WA115.

Broggini, F., K. Wapenaar, J. van der Neut, and R. Snieder, 2014b, Data-driven Green's function retrieval and application to imaging with multidimensional deconvolution: Journal of Geophysical Research: Solid Earth, **119**, 425-441.

DeRuiter, S. L., P. L. Tyack, Y-T. Lin, A. E. Newhall, J. F. Lynch, and P. J. O. Miller, 2006, Modeling acoustic propagation of airgun array pulses recorded on tagged sperm whales (Physeter microcephalus): The Journal of the Acoustical Society of America, **120**, no. 6, 4100-4114.

Jia, X., A. Guitton, and R. Snieder, 2018, A practical implementation of subsalt Marchenko imaging with a Gulf of Mexico dataset: Geophysics, XXX, XXXX-XXXX

Meles, G. A., K. Löer, M. Ravasi, A. Curtis, and C. A. da Costa Filho, 2015a, Internal multiple prediction and removal using Marchenko autofocusing and seismic interferometry: Geophysics, **80**, no. 1, A7–A11.

Meles, G. A., K. Wapenaar, and A. Curtis, 2015b, Synthesising primary reflections by Marchenko redatuming and convolutional interferometry: Geophysics, **81**, no. 2, Q25–Q26.

Mildner, C., T. S. Becker, K. de Vos, F. Broggini, and J. O. A. Robertsson, 2017, Source wavelet estimation using Marchenko focusing functions: Theory and laboratory data example: 87th Annual International Meeting, SEG, Expanded Abstracts, 5521–5525.

Misra, S., and M. D. Sacchi, 2007, Non-minimum phase wavelet estimation by non-linear optimization of all-pass operators: Geophysical Prospecting, **55**, 223-234.

Moczo, P., J. O. A. Robertsson, and L. Eisner, 2007, The finite-difference time-domain method for modeling of seismic wave propagation: Advances in Geophysics, **48**, 421-516.

Nuber, A., E. Manukyan, and H. Maurer, 2015, Enhancement of near-surface elastic full waveform inversion results in regions of low sensitivities: Journal of Applied Geophysics, **122**, 192-201

Ravasi M., I. Vasconcelos, A. Kritski, A. Curtis, C. A. da Costa Filho, and G.A. Meles, 2015, Marchenko Imaging of Volve Field, North Sea: 77th Conference and Exhibition, EAGE, Extended Abstracts, N10603.

Ravasi, M., I. Vasconcelos, A. Kritski, A. Curtis, C. A. da Costa Filho, and G. A. Meles, 2016, Target-oriented Marchenko imaging of a North Sea field: Geophysical Journal International, **205**, 99–104.

Ravasi, M., 2017, Rayleigh-Marchenko redatuming for target-oriented, true-amplitude imaging: Geophysics, **82**, no. 6, S439–S452.

Singh, S., R. Snieder, J. Behura, J. van der Neut, K. Wapenaar, and E. Slob, 2015, Marchenko imaging: Imaging with primaries, internal multiples, and free-surface multiples: Geophysics, **80**, no. 5, S165–S174.

Slob, E., and K. Wapenaar, 2017, Theory for Marchenko imaging of marine seismic data with free surface multiple elimination: 79th Annual International Conference and Exhibition, EAGE, Extended Abstracts.

Slob, E., K. Wapenaar, F. Broggini, and R. Snieder, 2014, Seismic reflector imaging using internal multiples with Marchenko-type equations: Geophysics, **79**, no. 2, S63-S76.

Staring, M., R. Pereira, H. Douma, J. van der Neut, and C. Wapenaar, 2017, Adaptive double-focusing method for source-receiver Marchenko redatuming on field data: 87th Annual International Meeting, SEG, Expanded Abstracts, 4808–4812.

Thomsen, H. R., 2016, Investigating the robustness of Green's function retrieval via Marchenko focusing and Seismic Interferometry: M.S. thesis, ETH Zürich

van der Neut, J., J. Thorbecke, K. Mehta, E. Slob, and K. Wapenaar, 2011, Controlled-source interferometric redatuming by crosscorrelation and multidimensional deconvolution in elastic media: Geophysics, **76**, no. 4, S63-S76

van der Neut, J., I. Vasconcelos, and K. Wapenaar, 2015a, On Green's function retrieval by iterative substitution of the coupled Marchenko equations: Geophysical Journal International, **203**, 792-813

van der Neut, J., K. Wapenaar, J. Thorbecke, and E. Slob, 2015b, Practical challenges in adaptive Marchenko imaging: 85th Annual International Meeting, SEG, Expanded Abstracts, 4505–4509.

Vasconcelos, I., M. Ravasi, J. van der Neut, A. Kritski, and T. Cui, 2017, Retrieving reservoir-only reflection and transmission responses from target- enclosing extended images: 79th Annual International Conference and Exhibition, EAGE, Extended Abstracts.

Velis, D. R., and T. J. Ulrych, 1996, Simulated annealing wavelet estimation via fourth-order cumulant matching: Geophysics, 61, no. 6, 1939-1948.

Wapenaar, K., J. van der Neut, E. Ruigrok, D. Draganov, J. Hunziker, E. Slob, J. Thorbecke, and R. Snieder, 2011, Seismic interferometry by crosscorrelation and by multi-dimensional deconvolution: A systematic comparison: Geophysical Journal International, **185**, 1335-1364

Wapenaar, K., J. Thorbecke, J. van der Neut, F. Broggini, E. Slob, and R. Snieder, 2014a, Marchenko imaging: Geophysics, **79**, no.3, WA39-WA57.

Wapenaar, K., J. Thorbecke, J. van der Neut, F. Broggini, E. Slob, and R. Snieder, 2014b, Green's function retrieval from reflection data, in absence of a receiver at the virtual source position: Journal of the Acoustical Society of America, **135**, 2847-2861.