Correspondence to:
J. Brackenhoff,
j.a.brackenhoff@tudelft.nl

Virtual Sources and Receivers in the Real Earth: Considerations for Practical Applications

Joeri Brackenhoff[1] , **Jan Thorbecke[1]** , and **Kees Wapenaar[1]**

[1]Department of Geoscience and Engineering, Delft University of Technology, Delft, The Netherlands

Abstract To enhance monitoring of the subsurface, virtual sources and receivers inside the subsurface can be created from seismic reflection data at the surface of the Earth using the Marchenko method. The response between these virtual sources and receivers can be obtained through the use of homogeneous Green's function retrieval. A homogeneous Green's function is a superposition of a Green's function and its time reversal. The main aim of this paper is to obtain accurate homogeneous Green's functions from field data. Classical homogeneous Green's function retrieval requires an unrealistic enclosing recording surface; however, by using a recently proposed single-sided retrieval scheme, this requirement can be avoided. We first demonstrate the principles of using the single-sided representation on synthetic data and show that different source signatures can be taken into account. Because the Marchenko method is sensitive to recording limitations of the reflection data, we study five cases of recording limitations with synthetic data and demonstrate their effects on the final result. Finally, the method is demonstrated on a preprocessed field data set that fulfills the requirements for applying the single-sided Green's function retrieval scheme. The scheme has the potential to be used in future applications, such as source localization.

1. Introduction

Seismic data can be used in a variety of ways to monitor and explore the subsurface of the Earth. Such data are obtained by measuring the wavefield that is propagating through the subsurface at physical receivers. Seismic data can be acquired using an active source at the surface of the Earth, in which case receivers are usually located on the same surface as the source or in a borehole. The receivers measure the full wavefield, that is, both primary and multiply scattered events. These measurements are often used to obtain information about the structure of the subsurface and its properties (Yilmaz, 2001). Alternatively, data can be acquired using a passive source, which is a source of the wavefield that occurs naturally in the subsurface of the Earth. In this setup, the wavefield is recorded by a continuously recording receiver array, usually at the surface of the Earth. These measurements can contain additional information about processes in the subsurface, such as induced seismicity (Grigoli et al., 2017). These types of measurements are receiving more attention because of the potentially damaging effects of induced seismicity in residential areas (Magnani et al., 2017; van Thienen-Visser & Breunese, 2015).

Active measurements can be employed to supplement the passive measurements. Using advanced seismic processing techniques, the wavefield that is measured at the surface of the Earth can be redatumed to locations inside the subsurface. By redatuming receivers from their physical location on the surface to locations at depth, virtual receivers are created. The advantage of such virtual receivers is that, by considering many of them, the evolution of the wavefield through the subsurface over time can be studied, which can provide relevant information about source mechanisms and the locations of scatterers in the subsurface. Similar to receiver redatuming, physical sources at the surface can be redatumed to create virtual sources at any location in the subsurface. Furthermore, the response between any combination of a virtual source and virtual receiver can be retrieved, a process we call homogeneous Green's function retrieval. Whereas a Green's function describes the response of a medium to a Dirac function, a homogeneous Green's function is a Green's function superposed by its time reversal to avoid a source singularity. The classical representation for the homogeneous Green's function retrieval was derived by Porter (1970). This method was further extended for inverse source problems by Porter and Devaney (1982) and inverse scattering methods by Oristaglio (1989). This classical representation has been employed as the theoretical basis in the field of seismic interferometry to create virtual sources (Bakulin & Calvert, 2005; van Manen et al., 2005; Wapenaar, 2004) or virtual

receivers (Curtis et al., 2009). However, in all of these applications, it appeared that a complete enclosing boundary is vital for retrieving a full homogeneous Green's function without artifacts.

Recently, a new single-sided representation for homogeneous Green's function retrieval has been derived. Instead of an enclosing boundary, it uses a single, nonenclosing boundary, typically the Earth's surface (Wapenaar et al., 2017). An example of the application of this method on synthetic data can be found in Wapenaar et al. (2016). In this approach, the data-driven Marchenko method is used to create virtual sources and receivers in the subsurface from reflection data at the Earth's surface. Using the homogeneous Green's function retrieval, the response between one selected virtual source and all virtual receivers is obtained. The Marchenko method, for the purpose of geophysial applications, was first proposed for 1-D by Broggini et al. (2012), based on work by Rose (2001), and was later extended for 2-D and 3-D applications (Wapenaar et al., 2013, 2014). The method uses two types of input. The first is active-source single-sided seismic reflection data measured at the surface of the Earth. The second is an estimation of the first wavefield event, which is called the first arrival, that would be caused by a source from a location in the subsurface to receiver locations at the surface of the Earth (hence, the first arrival of a Green's function between a subsurface location and the surface). The locations of the receivers of the Green's function match the locations of the receivers of the reflection response. The Marchenko method uses these data to create a full waveform Green's function, including all multiple scattering, for a virtual source in the subsurface and receivers at the Earth's surface. To model the first arrival, only a background velocity model is required, which can be estimated by processing the reflection data. A dense array of virtual sources for Green's functions in the subsurface can be created through repeated use of this methodology. Aside from the Green's function, the Marchenko method is also capable of retrieving a focusing function, which is designed to focus from the single-sided surface, where the reflection response is measured, to a focal location in the subsurface without any reverberation artifacts. The single-sided representation uses the focusing function, together with a Green's function, to create the response between a virtual source and receiver. Due to the single-sided focusing properties of the focusing function, the retrieval can be done for a single-sided recording setup without any artifacts.

Employing the Marchenko method on field data for practical applications is challenging due to the sensitivity of the Marchenko method to recording limitations of the reflection response. The sensitivity is partially caused by the fact that in the derivation of the Marchenko method, evanescent waves are ignored, and it is assumed that the medium of interest is lossless. In real media, the wavefield suffers from absorption, which violates the latter assumption. Furthermore, the method requires the reflection response to be well sampled and the aperture to be sufficiently large. The Marchenko method has been successfully applied on field data, by preprocessing the reflection response. Examples for the purpose of imaging can be found in Ravasi et al. (2016) and Staring et al. (2018), who used adaptive corrections in the Marchenko method. Homogeneous Green's function retrieval using the single-sided representation on field data was achieved by Wapenaar et al. (2018) and Brackenhoff et al. (2019). The aim of this paper is to apply the single-sided representation on field data and to consider the influence of recording limitations of the reflection response on the retrieved homogeneous Green's functions. To this end, we consider a 2-D field seismic data set from the Vøring basin off the coast of Norway. Along with the field data, we also consider a subsurface model, that is designed to simulate the subsurface of the area where the actual reflection response is recorded. Using this model, synthetic reflection data are created. First, we use the synthetic data to make a comparison between the results that are obtained when the single-sided representation is used and when the classical representation is used. The results show that the homogeneous Green's function is more accurately retrieved when the single-sided representation is used. The first arrivals that are used in these tests in the Marchenko method are all modeled using a monopole source mechanism. To study the influence of the source mechanism on the final result, the experiment is repeated using first arrivals that were modeled using a double-couple source mechanism, which is more representative for small-scale earthquakes (Aki & Richards, 2002). The homogeneous Green function that is obtained in this way still contains the correct events and has a double-couple signature. Next, we determine the sensitivity of the result to five recording limitations on the reflection data, namely, coarse source-receiver sampling, missing near offsets, small aperture, offsets missing in one direction, and absorption of the reflection data. The results of these numerical experiments are taken into account so that the field reflection data can be preprocessed and the single-sided representation can be applied properly. We employ both the classical and single-sided representation to the field data in order to compare the results. The applications show the potential of the single-sided representation for field data, as well as the possibility of applying the representation to passive field recordings.

2. Theory

In this section, we present an overview of the definitions and equations that are required for homogeneous Green's function retrieval. The Green's function and focusing function are reviewed, followed by the definitions of the classical enclosed boundary and single-sided representations for homogeneous Green's function retrieval. The Marchenko method and its limitations are considered, as well as the double-couple source mechanism.

2.1. Green's Function

The Green's function is defined as the solution of the wave equation to a Dirac point source, which can be written as (Feynman et al., 2011; Morse & Feshbach, 1953)

$$\left(\rho(\mathbf{x})\partial_i \left(\frac{1}{\rho(\mathbf{x})}\partial_i \right) - \frac{1}{c^2(\mathbf{x})}\partial_t^2 \right) G(\mathbf{x}, \mathbf{x}_F, t) = -\rho(\mathbf{x})\delta(\mathbf{x} - \mathbf{x}_F)\partial_t \delta(t), \tag{1}$$

where $G(\mathbf{x}, \mathbf{x}_F, t)$ describes the response of the medium, at time t, at location $\mathbf{x}$ to a source at location $\mathbf{x}_F$. The locations are defined in 3-D such that $\mathbf{x} = (x_1, x_2, x_3)^T$. The symbols ρ and c indicate the density and velocity of the medium, respectively, δ indicates a Dirac delta function, ∂_t a temporal derivative, and ∂_i the partial derivative in the three principal directions. The repeated subscript i follows the Einstein summation convention. Note that the source at the right-hand side is defined with a temporal derivative acting on the Dirac delta function. This choice is made to simulate a volume injection-rate source. According to the reciprocity principle, the source and receiver locations of the Green's functions can be interchanged, $G(\mathbf{x}, \mathbf{x}_F, t) = G(\mathbf{x}_F, \mathbf{x}, t)$.

We also consider the Fourier-transformed Green's function $G(\mathbf{x}, \mathbf{x}_F, \omega)$:

$$G(\mathbf{x}, \mathbf{x}_F, \omega) = \int_{-\infty}^{\infty} G(\mathbf{x}, \mathbf{x}_F, t)e^{i\omega t}\mathrm{d}t, \tag{2}$$

where ω denotes the angular frequency and i the imaginary unit. Note that the sign in the exponential can be reversed, as long as the same is done for the inverse Fourier transform. Using equation (2), equation (1) is transformed to the frequency domain:

$$\left(\rho(\mathbf{x})\partial_i \left(\frac{1}{\rho(\mathbf{x})}\partial_i \right) + \frac{\omega^2}{c^2(\mathbf{x})} \right) G(\mathbf{x}, \mathbf{x}_F, \omega) = i\omega\rho(\mathbf{x})\delta(\mathbf{x} - \mathbf{x}_F). \tag{3}$$

A schematic illustration of the Green's function is shown in Figure 1b, where the receivers are placed at the surface of a medium and its source inside the medium. We assume that the medium has a transparent reflection-free half-space at the top of the medium. In practice, this situation is obtained after the elimination of surface-related multiples. Some possible raypaths, including scattering, have been drawn in the figure. Figure 1a shows a special case of the Green's function, with both source and receivers placed at the surface of the medium. This is called the reflection response $R(\mathbf{x}, \mathbf{x}_S, t)$ and contains all the reflections, both primaries and multiples, of the medium; however, we assume that the direct wave from the sources to the receivers are not present in the reflection response.

The homogeneous Green's function is defined as the superposition of the Green's function and its time reversal. Because of the temporal derivative on the Dirac delta function in equation (1), when time reversal is applied, the source term will obtain an opposite sign. As a result, the superposition of the Green's function and its time reversal removes the source term, thereby avoiding a singularity at the source position:

$$G_{\mathrm{h}}(\mathbf{x}, \mathbf{x}_F, t) = G(\mathbf{x}, \mathbf{x}_F, t) + G(\mathbf{x}, \mathbf{x}_F, -t), \tag{4}$$

$$\left(\rho(\mathbf{x})\partial_i \left(\frac{1}{\rho(\mathbf{x})}\partial_i \right) - \frac{1}{c^2(\mathbf{x})}\partial_t^2 \right) G_{\mathrm{h}}(\mathbf{x}, \mathbf{x}_F, t) = 0, \tag{5}$$

and in the frequency domain:

$$G_{\mathrm{h}}(\mathbf{x}, \mathbf{x}_F, \omega) = G(\mathbf{x}, \mathbf{x}_F, \omega) + G^*(\mathbf{x}, \mathbf{x}_F, \omega), \tag{6}$$

$$\left(\rho(\mathbf{x})\partial_i \left(\frac{1}{\rho(\mathbf{x})}\partial_i \right) + \frac{\omega^2}{c^2(\mathbf{x})} \right) G_{\mathrm{h}}(\mathbf{x}, \mathbf{x}_F, \omega) = 0, \tag{7}$$
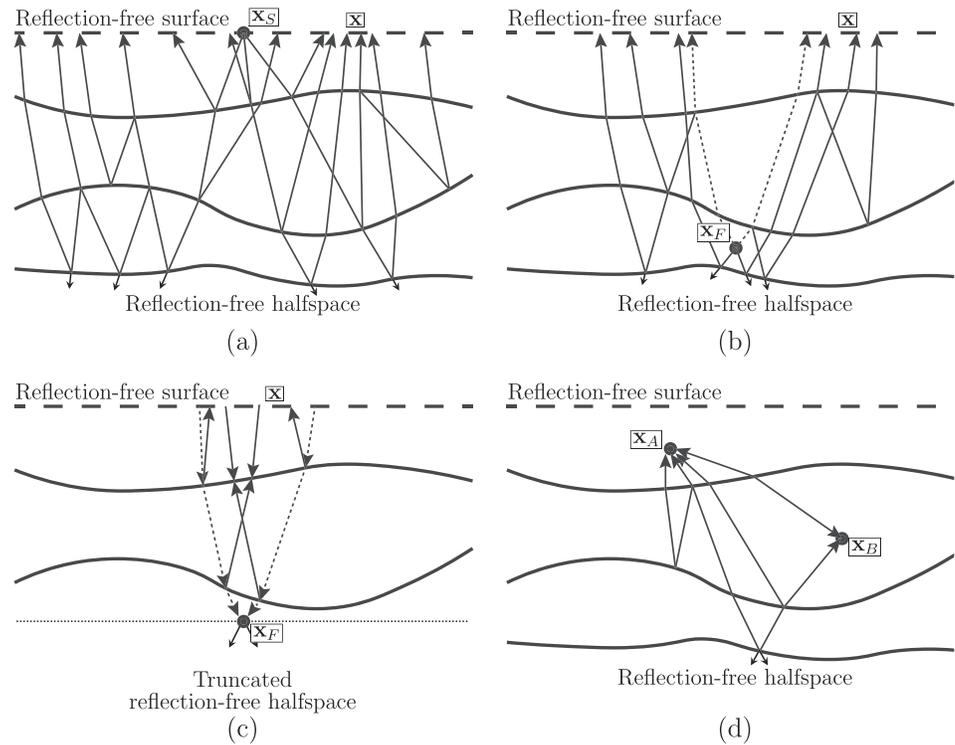
Figure 1. Possible raypaths drawn for (a) a reflection response $R(\mathbf{x}, \mathbf{x}_S, t)$, measured at varying receiver locations $\mathbf{x}$ at the surface, with a source at $\mathbf{x}_S$ also at the surface, (b) a Green's function $G(\mathbf{x}, \mathbf{x}_F, t)$, measured at varying receiver location $\mathbf{x}$ at the surface with a source at $\mathbf{x}_F$ inside the medium, (c) a focusing function $f_1(\mathbf{x}, \mathbf{x}_F, t)$, emitted from the surface at varying locations $\mathbf{x}$, focusing to a focal location $\mathbf{x}_F$ inside a medium that is truncated below $\mathbf{x}_F$ (indicated by the horizontal dotted line), and (d) a homogeneous Green's function $G_{\mathrm{h}}(\mathbf{x}_A, \mathbf{x}_B, t)$, between two locations, $\mathbf{x}_A$ and $\mathbf{x}_B$, inside the medium. The homogeneous Green's function is indicated with two-sided arrows to represent that it is the superposition of Green's function and its time reversal. The dotted arrows in (b) and (c) indicate the first arrival for the Green's function and focusing function. The surfaces at the top of all figures are transparent; hence, there are no free-surface multiples.

where $G_{\mathrm{h}}(\mathbf{x}, \mathbf{x}_F, t)$ and $G_{\mathrm{h}}(\mathbf{x}, \mathbf{x}_F, \omega)$ denote the homogeneous Green's function in the time domain and frequency domain, respectively, and the asterisk indicates complex conjugation. Figure 1d shows a schematic illustration of the homogeneous Green's function, $G_{\mathrm{h}}(\mathbf{x}_A, \mathbf{x}_B, t)$, with some possible raypaths drawn, with both its source and receiver inside the medium. To reflect the superposition of the Green's function and its time reversal, the raypaths are indicated with two-sided arrows.

2.2. Focusing Function

The focusing function $f_1(\mathbf{x}, \mathbf{x}_F, t)$ describes a wavefield, at time t, at location $\mathbf{x}$, that focuses to a focal location $\mathbf{x}_F$ in the subsurface. The focusing function propagates in a medium that is truncated below $\mathbf{x}_F$, which means that there are no reflectors present below the focal location. The focusing function can be decomposed into its upgoing and downgoing parts:

$$f_1(\mathbf{x}, \mathbf{x}_F, t) = f_1^+(\mathbf{x}, \mathbf{x}_F, t) + f_1^-(\mathbf{x}, \mathbf{x}_F, t), \tag{8}$$

where $f_1^+(\mathbf{x}, \mathbf{x}_F, \omega)$ denotes the downgoing focusing function and $f_1^-(\mathbf{x}, \mathbf{x}_F, \omega)$ the upgoing focusing function. The downgoing part of the focusing function is defined as the inverse of the transmission response of the truncated medium (Wapenaar et al., 2014).

The focusing function is schematically illustrated in Figure 1c, where some possible raypaths have been drawn. The first arrival, which is indicated by the dotted raypath, propagates from the surface to the focal location and scatters at the reflectors, creating an upgoing wavefield. In order to ensure that these upgoing waves do not cause additional events arriving after the wavefield has focused, additional downgoing waves are injected from the surface, which cancel out these events. This occurs at the locations where arrowheads meet in Figure 1c. Because of the reflection-free surface at $\mathbf{x}$, there are no events present in the focusing
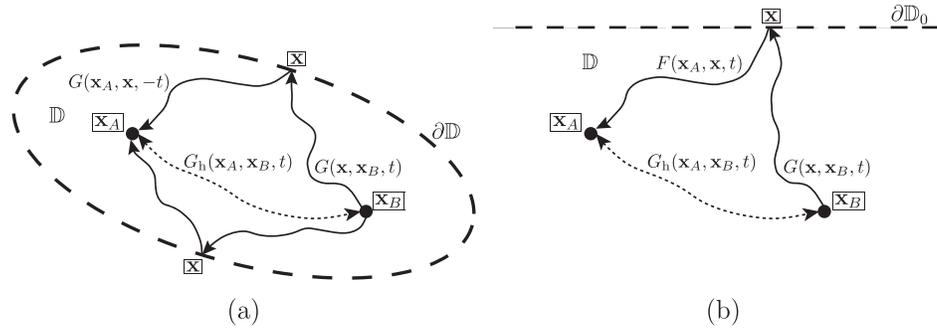
Figure 2. Recording setup for homogeneous Green's function retrieval using (a) the classical representation and (b) the single-sided representation. For both setups, a Green's function, $G(\mathbf{x}, \mathbf{x}_B, t)$, is utilized. For the classical representation, an additional Green's function, $G(\mathbf{x}_A, \mathbf{x}, t)$, is used to create a virtual receiver location. The medium of interest $\mathbb{D}$ is surrounded by an enclosing boundary $\partial\mathbb{D}$. Equation (9) is evaluated over this enclosing boundary. For the single-sided representation, the virtual receiver is not created using a Green's function but rather a focusing function, $F(\mathbf{x}_A, \mathbf{x}, t)$. The medium $\mathbb{D}$ is not enclosed, instead only a single-sided nonenclosing boundary $\partial\mathbb{D}_0$ is present at a single side. Equation (10) is evaluated over this single-sided boundary. The homogeneous Green's function is indicated by the dotted line.

function to account for free-surface multiples. A more detailed description of the focusing function can be found in Slob et al. (2014).

2.3. Homogeneous Green's Function Representation

The classical representation of the homogeneous Green's function states that the response between a source and receiver inside a medium can be retrieved from observations at a boundary. In order to achieve this, an enclosing boundary around the medium of interest, over which the data can be recorded and/or injected, needs to be present (Oristaglio, 1989; Porter, 1970; Porter & Devaney, 1982). The classical representation in the frequency domain can be written as follows:

$$G_{\mathrm{h}}(\mathbf{x}_A, \mathbf{x}_B, \omega) = \oint_{\partial\mathbb{D}} \frac{-1}{i\omega\rho(\mathbf{x})} \{\partial_i G^*(\mathbf{x}_A, \mathbf{x}, \omega)G(\mathbf{x}, \mathbf{x}_B, \omega)$$
$$- G^*(\mathbf{x}_A, \mathbf{x}, \omega)\partial_i G(\mathbf{x}, \mathbf{x}_B, \omega)\}n_i\mathrm{d}^2\mathbf{x}, \tag{9}$$

where n_i indicates the components of the normal vector in the three principal directions. The integral is evaluated over a boundary $\partial\mathbb{D}$ enclosing the medium $\mathbb{D}$. In equation (9), the function $G(\mathbf{x}, \mathbf{x}_B, \omega)$ describes the response of the medium at varying location $\mathbf{x}$ at the boundary to a source at location $\mathbf{x}_B$ inside the medium. The time-reversed function $G^*(\mathbf{x}_A, \mathbf{x}, \omega)$ back propagates the responses from the boundary to the receiver location $\mathbf{x}_A$, thereby creating a virtual receiver at $\mathbf{x}_A$. A schematic overview of the application of this representation is shown in Figure 2a. In practice, the classical representation is often not evaluated correctly, because acquisition on an enclosing boundary is not feasible and only measurements on a single-sided nonenclosing boundary, usually the Earth's surface, are available. As an approximation, equation (9) can be evaluated over the single-sided boundary. Applying the representation in this way causes significant artifacts in the retrieved homogeneous Green's function, however. Due to the fact that few alternatives are available, the method is still widely applied to cases where no closed boundaries are present. An alternative representation that can be employed uses a focusing function instead of a Green's function. This representation is capable of retrieving the full homogeneous Green's function with significantly less artifacts from a single-sided boundary; hence, it is referred to as the single-sided representation. It can be written as (Wapenaar et al., 2017, equation 30)

$$G_{\mathrm{h}}(\mathbf{x}_A, \mathbf{x}_B, \omega) = \Im \int_{\partial\mathbb{D}_0} \frac{4}{\omega\rho(\mathbf{x})} \{\partial_3 F(\mathbf{x}_A, \mathbf{x}, \omega)G(\mathbf{x}, \mathbf{x}_B, \omega)\}\mathrm{d}^2\mathbf{x}, \tag{10}$$

where $\partial\mathbb{D}_0$ denotes the single-sided boundary and $\Im$ the imaginary part of a complex function. The focusing function, $F(\mathbf{x}_A, \mathbf{x}, \omega)$, is defined as

$$F(\mathbf{x}_A, \mathbf{x}, \omega) = f_1^+(\mathbf{x}, \mathbf{x}_A, \omega) - f_1^{-*}(\mathbf{x}, \mathbf{x}_A, \omega). \tag{11}$$

In equation (10), $G(\mathbf{x}, \mathbf{x}_B, \omega)$ serves again as the response to a source location inside the medium, measured at the single-sided boundary $\partial\mathbb{D}_0$. The focusing function $F(\mathbf{x}_A, \mathbf{x}, \omega)$ serves as the back propagator of the

responses from the boundary to the focal location inside the medium. A schematic representation of this procedure is shown in Figure 2b. The two representations in equations (9) and (10) for homogeneous Green's function retrieval are similar in form, as both use a backward propagator on the response measured on the boundary. The main difference is that, for the single-sided representation, the backward propagator is a focusing function instead of a time-reversed Green's function. As one can interpret from Figure 1c, the convergence of the focusing function to the focal location is ensured by the first arrival, whereas the coda of the focusing function removes unwanted reflections caused by the first arrival when it encounters reflectors while propagating to the focal location. The arrival times of the direct wave of the focusing function are the same as the arrival times of the direct wave of the time-reversed Green's function. The difference is that the coda of the focusing function is designed to cancel out the events that are introduced by the direct arrival, whereas the coda of the time-reversed Green's function introduces additional artifacts, in the form of reverberations.

2.4. Marchenko Method

We use the Marchenko method to retrieve the focusing functions and Green's functions required for the representations for homogeneous Green's function retrieval. A more detailed consideration of the method can be found in Wapenaar et al. (2014). Here we only consider the equations and properties of the method relevant for this paper. The Green's function and focusing function of a medium are related via the reflection response:

$$G(\mathbf{x}, \mathbf{x}_B, t) - F(\mathbf{x}_B, \mathbf{x}, -t) = \int_{\partial \mathbb{D}_0} \int_{-\infty}^{\infty} R(\mathbf{x}, \mathbf{x}_S, t') F(\mathbf{x}_B, \mathbf{x}_S, t - t') dt' d^2\mathbf{x}_S, \tag{12}$$

where $\mathbf{x}_B$ is a location inside the medium and $\mathbf{x}_S$ indicates the array of sources that are present on the nonenclosing surface $\partial \mathbb{D}_0$. Equation (12) states that if the reflection response R at a boundary $\partial \mathbb{D}_0$ and a focusing function with a focal location inside medium $\mathbb{D}$ are available, the Green's function with a source at the focal location can be retrieved. The retrieval of the focusing function inside the medium can be achieved using the iterative Marchenko equation:

$$F_{k+1}(\mathbf{x}_B, \mathbf{x}, -t) = D(\mathbf{x}, \mathbf{x}_B, t) - w(\mathbf{x}, \mathbf{x}_B, t) \int_{\partial \mathbb{D}_0} \int_{-\infty}^{\infty} R(\mathbf{x}, \mathbf{x}_S, t') F_k(\mathbf{x}_B, \mathbf{x}_S, t - t') dt' d^2\mathbf{x}_S, \tag{13}$$

where $F_k(\mathbf{x}_B, \mathbf{x}, t)$ is the estimated focusing function after k iterations, $D(\mathbf{x}, \mathbf{x}_B, t)$ is the first arrival of the Green's function, and $w(\mathbf{x}, \mathbf{x}_B, t)$ is a windowing function. The windowing function is used to mute the Green's function completely. When the windowing function is applied to equation (12), the Green's function is removed. The arrival times of the first arrival of the Green's function are the same as the arrival times of the last arrival of the time-reversed focusing function; hence, the windowing function also removes the last arrival of the time-reversed focusing function; however, it will not remove the coda of the time-reversed focusing function. Therefore, in order to obtain the full focusing function in equation (13), $D(\mathbf{x}, \mathbf{x}_B, t)$ needs to be added after the windowing function has been applied. The windowing function $w(\mathbf{x}, \mathbf{x}_B, t)$ can be estimated from $D(\mathbf{x}, \mathbf{x}_B, t)$, as the edge of the muting area is located around the first arrival. In order to use equation (13) and start the iterative scheme, a first estimation of the focusing function is required. The time-reversed first arrival $D(\mathbf{x}_S, \mathbf{x}_B, -t)$ of the Green's function is used as this first estimation of the focusing function $F_0(\mathbf{x}_B, \mathbf{x}_S, t)$. If this arrival is emitted into the medium, it will cause additional reflections that are not cancelled. By using equation (13) iteratively, until convergence is achieved, the coda of the focusing function is retrieved, which will suppress the undesired reflections. The only required components for the iterative scheme are a reflection response measured at the single-sided boundary (i.e., the Earth's surface) and the direct arrival from the focal point to the same boundary. This direct arrival can be modeled using a smooth velocity model. Because only the direct arrival is of interest, the model requires no detailed features. Generally, a monopole point source is used to model these first arrivals. After the focusing function has been retrieved, it can be used in equation (12) to compute the Green's function, $G(\mathbf{x}, \mathbf{x}_B, t)$. Subsequently, this Green's function and a similarly derived focusing function for focal point $\mathbf{x}_A$ are used in equation (10) to retrieve the homogeneous Green's function, $G_h(\mathbf{x}_A, \mathbf{x}_B, \omega)$. All the Green's functions and focusing functions in this paper are retrieved using the Marchenko method to ensure the representations are applied to the field data and the synthetic data in the same way. The Marchenko method has restrictions, particularly when it is applied on field data. An important underlying assumption of the Marchenko method that is considered in this paper is that no free-surface multiples are present in the reflection response. Hence, the

free-surface multiples should be removed prior to applying the Marchenko method, for example, by applying a surface-related multiple elimination scheme (Verschuur et al., 1992). There are ways to incorporate these multiples in the Marchenko method as well, for example, see Singh et al. (2015). Additionally, the reflection response that is used needs to be accurate, as issues with the quality of the recording have strong influences on the final result. An important requirement is that the medium of interest needs to be lossless, which is an unrealistic approximation in real media. Also, the reflection response needs, preferably, to be densely sampled, contain both positive and negative source-receiver offsets, and have sufficient recording length and aperture. The effects of some of the limitations of the Marchenko method are considered in Ravasi et al. (2016), Brackenhoff (2016), and Staring et al. (2018). When synthetic data are used, the reflection response can be modeled without these limitations. However, when field data are recorded, not all of these requirements are fulfilled, and appropriate preprocessing is required.

2.5. Double-Couple Source

As mentioned before, the Green's functions and focusing functions that are retrieved using the Marchenko method usually have a monopole source signature. In the field, especially for passive recordings, there are many different types of source mechanisms. These different source mechanisms can be taken into account for the homogeneous Green's function retrieval as well. We show this by incorporating a double-couple source mechanism in the representation. This mechanism is chosen because it is considered to be representative for an earthquake reponse when the wavelength of the wavefield is larger than the dimensions of the source (Aki & Richards, 2002). The mechanism differs from a monopole source in the sense that the radiation pattern is not homogeneous; rather, the polarity and amplitude vary depending on the radiation angle of the source. There will be four distinct polarity changes along the radiation pattern, at 90°, 180°, 270°, and 360°. The source can be inclined at various angles, for example, aligned along a fault, which changes the orientation of the polarity changes as well. Operator $\mathfrak{D}_B\{\cdot\}$ is introduced, which transforms a monopole source signature to a double-couple source signature. This operator is defined as

$$\mathfrak{D}_B\{\cdot\} = (\theta_i^{\parallel} + \theta_i^{\perp})\partial_{i,B}\{\cdot\}, \tag{14}$$

where $\partial_{i,B}$ is a component of the vector containing the partial derivatives acting on the monopole signal originating from source location $\mathbf{x}_B$, to transform the source signature to a double-couple mechanism, $\theta_i^{\parallel}$ is a component of the unit vector that orients one couple of the signal parallel to the fault plane, and $\theta_i^{\perp}$ is a component of a similar vector that orients the other couple perpendicular to the fault plane. The operator is applied to the representation for homogeneous Green's function retrieval in equation (10), which yields

$$\mathfrak{D}_B\{G_{\mathrm{h}}(\mathbf{x}_A, \mathbf{x}_B, \omega)\} = \mathfrak{I}\int_{\partial\mathbb{D}_0}\frac{4}{\omega\rho(\mathbf{x})}\{\partial_3 F(\mathbf{x}_A, \mathbf{x}, \omega)\mathfrak{D}_B\{G(\mathbf{x}, \mathbf{x}_B, \omega)\}\}\mathrm{d}^2\mathbf{x}, \tag{15}$$

where the subscript B in the double-couple operator indicates that the operator acts on the source in location $\mathbf{x}_B$. In this representation, the double-couple operator is only applied to the Green's function for the virtual source location, $G(\mathbf{x}, \mathbf{x}_B, \omega)$, and not to the focusing function, $F(\mathbf{x}_A, \mathbf{x}, \omega)$, for the virtual receiver location. This is because the focusing function is retrieved using equation (13) and thus has a monopole source signature. For the retrieval of the Green's function, $\mathfrak{D}_B\{G(\mathbf{x}, \mathbf{x}_B, t)\}$, we apply the double-couple operator to equations 12 and (13):

$$\mathfrak{D}_B\{G(\mathbf{x}, \mathbf{x}_B, t)\} - \mathfrak{D}_B\{F(\mathbf{x}_B, \mathbf{x}, -t)\} = \int_{\partial\mathbb{D}_0}\int_{-\infty}^{\infty} R(\mathbf{x}, \mathbf{x}_S, t')\mathfrak{D}_B\{F(\mathbf{x}_B, \mathbf{x}_S, t - t')\}\mathrm{d}t'\mathrm{d}^2\mathbf{x}_S, \tag{16}$$

$$\mathfrak{D}_B\{F_{k+1}(\mathbf{x}_B, \mathbf{x}, -t)\} = \mathfrak{D}_B\{D(\mathbf{x}, \mathbf{x}_B, -t)\} - w(\mathbf{x}, \mathbf{x}_B, t)\int_{\partial\mathbb{D}_0}\int_{-\infty}^{\infty} R(\mathbf{x}, \mathbf{x}_S, t')\mathfrak{D}_B\{F_k(\mathbf{x}_B, \mathbf{x}_S, t - t')\}\mathrm{d}t'\mathrm{d}^2\mathbf{x}_S. \tag{17}$$

Table 1
Acquisition Parameters for the Vøring Data Set

| Parameter	Value
Number of source positions	399
Source spacing	25 m
First source position	5,000 m
Final source position	14,950 m
Number of receiver positions per source	180
Receiver spacing	25 m
Minimum source-receiver offset	150 m
Maximum source-receiver offset	4,625 m
Number of recording samples	2,001
Sampling interval	0.004 s
High-cut frequency	90 Hz

The first arrival is modeled using the double-couple source mechanism instead of a monopole source mechanism and used in equation (17), for $k=0$, as $\mathfrak{D}_B\{F_0(\mathbf{x}_B, \mathbf{x}_S, t)\}$. This results in an estimation of the focusing function $\mathfrak{D}_B\{F(\mathbf{x}_B, \mathbf{x}, t)\}$, with a double-couple source mechanism. This focusing function is then used in equation (16) to obtain the Green's function $\mathfrak{D}_B\{G(\mathbf{x}, \mathbf{x}_B, t)\}$ with a double-couple source signature, which in turn is used in equation (15).

3. Data Sets

We consider both a synthetic data set and a field data set. The synthetic data set is based on the field data set, hence we first consider the parameters of the field recording before the synthetic data set is considered.

3.1. Vøring Data

The considered field data were recorded in a marine setting over the Vøring basin in offshore Norway by SAGA Petroleum A.S., which is currently part of Equinor. The data consist of a reflection response acquired along a 2-D line. For each source location, the receivers were moved along with the source position. The parameters of the acquisition can be found in Table 1. An example of a single common-source record is shown in Figure 3a, where the wavelet on the data have been reshaped to a 30 Hz Ricker wavelet for display purposes (a Ricker wavelet is defined as minus the second time derivative of a Gaussian function). There are several events present in the common-source record; however, it should be noted that the near offsets are missing. This is a common situation for marine acquisition, because receivers cannot be placed too close to active sources. The sources and receivers are located inside the water, and because S waves cannot propagate in water, only P waves are measured by the receivers. There are conversions from P to S waves and back in the subsurface below the water, so there are P waves present in the reflection data that were converted from S waves. The data also contain free-surface multiples, which are indicated by the black arrows.
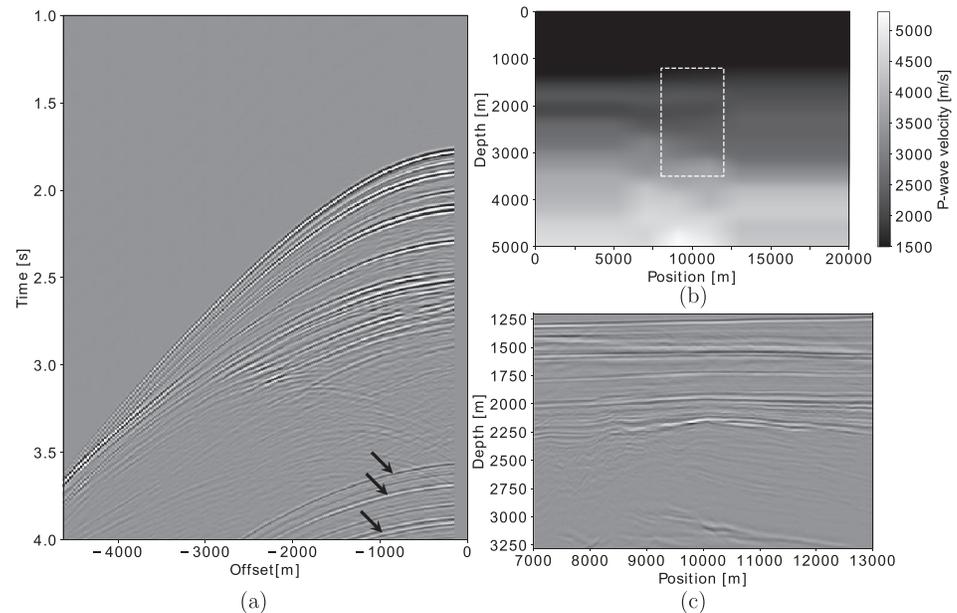
Figure 3. (a) Unprocessed common-source record of the reflection response recorded in the Vøring basin. The source is located at zero offset, where along with other near offsets no data could be recorded. Free-surface multiples are present in the data at late times, indicated by the black arrows. (b) Estimated P wave velocity model in m/s of the subsurface region where the data in the Vøring basin were recorded. The white dashed box represents the region of interest. (c) Image of the region of interest, indicated by the white dashed box in (b). The wavelet in the data in (a) and (c) has been reshaped into a 30 Hz Ricker wavelet for display purposes.
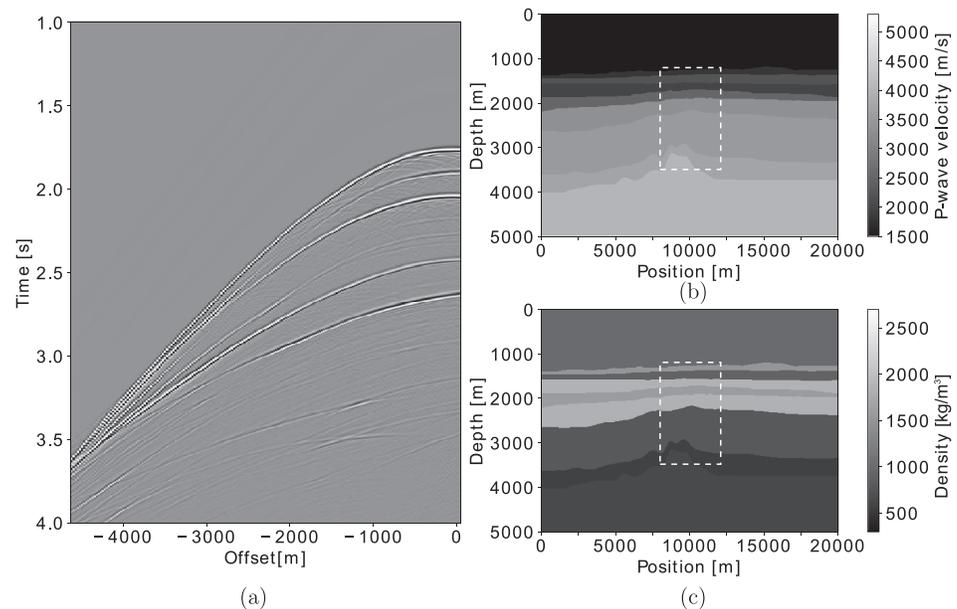
Figure 4. (a) Common-source record of the acoustic reflection response, modeled using finite-difference modeling in the *P* wave velocity model from (b) and density model from (c), at the same location as the common-source record in Figure 3a. The data have been convolved with a 30 Hz Ricker wavelet for display purposes. (b) Synthetic *P* wave velocity model in m/s based on the smooth velocity model from Figure 3b and the image from Figure 3c. (c) Synthetic density model in kg/m³ based on the image from Figure 3c. The white boxes in (b) and (c) indicate the region of interest.

Aside from the reflection data, a smooth *P* wave velocity model is provided by Equinor and displayed in Figure 3b. This model is used to create the first arrivals required for the Marchenko method. The dashed white box in Figure 3b indicates the region of interest where the homogeneous Green's functions are retrieved in this paper. An image of the region of interest was constructed, which is shown in Figure 3c, using the reflection data, the velocity model, and a one-way recursive depth migration based on Thorbecke et al. (2004). Imaging is not the main subject of this paper, so the details of the construction of this image are not discussed; however, it should be noted that free-surface multiples were removed prior to imaging. Also note that the retrieval of the homogeneous Green's function (discussed in the sections on synthetic data and field data) and the construction of the image were done independently of each other. The image is used to construct a subsurface model and to validate the homogeneous Green's function retrieval on the field data.

3.2. Synthetic Data

The synthetic models that we use in this paper are based on the field data set. Because there are no direct measurements of the subsurface available, an acoustic model is interpreted based on the smooth *P* wave velocity model in Figure 3b and the image in Figure 3c. This is done, because there is no *S* wave velocity information of the subsurface available; hence, we cannot construct an elastodynamic model. It is possible to use an elastodynamic representation for homogeneous Green's function retrieval (see Reinicke & Wapenaar, 2019); however, for the Vøring basin, we do not have the required multicomponent data to do so. To construct the model, we use the image to determine the locations of geological layer boundaries and determine the *P* wave velocities by calculating the interval velocities in the smooth *P* wave velocity model between the contrasts. The interpreted velocity model is displayed in Figure 4b, which shows hard boundaries. Notice that below the area of interest, there are no reflectors present in the medium. It is not possible to achieve reliable imaging in this area, therefore no structures are interpreted. Features outside the region of interest were extrapolated to create a full model. A density model is also constructed in order to ensure strong amplitudes in the reflection data. Because no direct measurements of density in the subsurface are available, it is not possible to create a density model that represents the subsurface accurately. Instead, the densities are chosen based on realistic ranges, that ensure strong contrasts between the layers. Figure 4c displays the density model.

The finite-difference wavefield modeling code by Thorbecke and Draganov (2011) is used to create the single-sided reflection data as input for the Marchenko method. The reflection response of the interpreted

model is modeled using the same measurement parameters as for the real data set, shown in Table 1, with pressure receivers recording the modeled wavefield at the surface of the model, in response to force sources that are located at the same surface. However, all offsets are included, no absorption is modeled, and free-surface multiples are ignored. When comparing Figure 3a and Figure 4a, there are similar events; however, fewer events are present in the synthetic data. Not all the reflectors in the subsurface can be properly imaged and interpreted; therefore, only the major features are present. The field data is more complex than the synthetic data. Because the exact *P* wave velocity and density information of the actual medium are not available, there is an amplitude mismatch. The converted waves due to elastic interactions from the actual recording are also not taken into account. Furthermore, no free-surface multiples are present in the modeled reflection data.
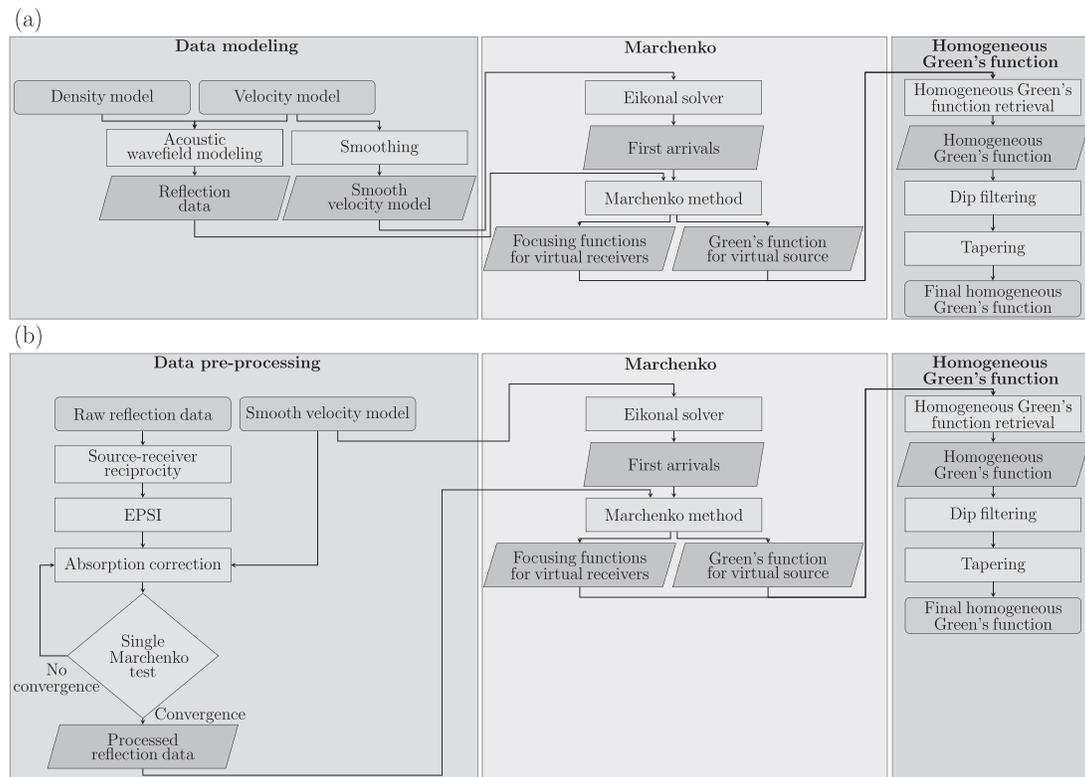
4. Synthetic Data

The synthetic data are used to retrieve the homogeneous Green's function. The single-sided reflection data and the velocity model are the only information employed. Furthermore, the velocity model is smoothed to create a background velocity model, without any detailed features. This smooth velocity model is used in an eikonal solver that is based on the method by Vidale (1988) to determine the arrival times from every location in the region of interest, that is indicated in Figures 4b and 4c, to the single-sided surface. These arrival times are combined with amplitude estimations using the method by Spetzler and Angelov (2005) and convolved with a wavelet to create a band-limited pressure wavefield with a monopole source signature. Note that only the smooth velocity model is used for this approach, and no density information is used. The first arrivals and the reflection data are used as input for the iterative Marchenko equation, as described by equation (13), using the code by Thorbecke et al. (2017). The windowing function is created by using the arrival times from the eikonal solver. Through use of the iterative code, Green's functions and focusing functions are obtained.

### 4.1. Homogeneous Green's Function Retrieval

To demonstrate the advantage of the single-sided representation, we compare different ways of retrieving the homogeneous Green's function. We aim to replicate a realistic situation as much as possible; hence, the only data that are used are the smooth velocity model and the modeled reflection response. First, to create a benchmark for the homogeneous Green's function, we model the wavefield directly inside the medium by placing a volume injection-rate source and pressure receivers inside the region of interest. The wavefield is time reversed and added to the original wavefield conforming to equation (4) to create the homogeneous Green's function. Three snapshots of this result are shown in Figures 6a–6c at 0, 200, and 400 ms, respectively. This is the result to which the homogeneous Green's functions that are obtained using the representations in equations 9 and (10) are compared to. In Figure 6, the wavefield is convolved with a 15 Hz Ricker wavelet, and dotted black lines are shown for reference where we expect scattering to take place. This type of visualization is used for all the snapshots we produce for the synthetic data.

Now, we assume that we do not know the exact model and use the Marchenko method to retrieve Green's functions in the region of interest in the subsurface, from the reflection data at the single-sided surface at the top of the model. These Green's functions are used to evaluate equation (9) at the single-sided surface, not at an enclosing surface. The location $\mathbf{x}_B$ of one response, $G(\mathbf{x}, \mathbf{x}_B, t)$, is kept constant as the virtual source position, while the location $\mathbf{x}_A$ of the other response, $G(\mathbf{x}_A, \mathbf{x}, t)$, varies to serve as the virtual receiver positions. The positions of the virtual source and receivers are exactly the same as the respective source and receiver positions of the modeled wavefield. The retrieved homogeneous Green's function is shown in Figures 6d–6f at 0, 200, and 400 ms, respectively. In Figure 6d, for zero time, there is noise present in the entire snapshot, and while there is a focus of the wavefield to the source position, there are strong artifacts present directly above the source position. Furthermore, coherent artifacts with weaker amplitudes are present throughout the snapshot. The snapshot at 200 ms in Figure 6e shows that the downgoing primary wavefield is retrieved; however, the upgoing primary wavefield is missing, instead forming into a weakening event with incorrect arrival times. Furthermore, while the coda of the wavefield has been retrieved with the correct upgoing reflections, the downgoing reflections are missing, and downgoing artifacts from above the source location are present in the snapshot. Aside from these problems, the snapshot is contaminated by artifacts. The final snapshot at 400 ms in Figure 6f shows similar problems. For comparsion, we repeat the experiment; however, instead of using a full Green's function $G(\mathbf{x}, \mathbf{x}_B, t)$ for the virtual source, we only use its first arrival to reduce the number of artifacts. The results are shown in Figures 6g–6i at 0, 200, and

Figure 5. Flowchart for the retrieval of the homogeneous Green's function for (a) the synthetic data and (b) the field data.

400 ms, respectively. Compared to the previous experiment, the number of artifacts decreases, although not all are removed. The strong source artifacts at time zero remain present, and the upgoing primary wavefield and downgoing coda are not restored. Next we apply the single-sided representation using equation (10). For the virtual source location $\mathbf{x}_S$, the same Green's function, $G(\mathbf{x}, \mathbf{x}_B, t)$, is used as in the previous two experiments. However, the Green's function that was used to create the virtual receiver, $G(\mathbf{x}_A, \mathbf{x}, t)$, is replaced by a focusing function, $F(\mathbf{x}_A, \mathbf{x}, t)$. Figures 6j–6l shows the result at 0, 200, and 400 ms, respectively. The improvement is noticeable, as artifacts are removed from the homogeneous Green's function. Aside from this, the primary wavefield is fully reconstructed, as is the coda of the downgoing wavefield. When comparing this result to the benchmark, it shows that the events are retrieved at the correct locations and times, although an amplitude mismatch is present. This is due to the fact that the amplitude of the first arrival is not exact, because we assume that we cannot model the first arrival in the real medium. When the amplitude of the first arrival is incorrect, it scales the retrieved focusing functions and Green's function, although the correct relative amplitude, arrival times and events are obtained (Brackenhoff, 2016; Brackenhoff et al., 2019). Some of the events are not reconstructed, especially when the angle of the reflection is high. This is because the single-sided boundary is assumed to be infinite, while in reality, the aperture is limited. The reflection response lacks certain angles of reflection, so these angles cannot be reconstructed for the homogeneous Green's function. At zero time, the snapshot contains less artifacts; however, some remain, which contaminate the homogeneous Green's function at later times. An analysis of the wave number-frequency spectrum of the homogeneous Green's function retrieved using equation (10) revealed noise at locations corresponding to high angles of the wavefield. As mentioned before, the aperture of the reflection response is limited, so at these angles, no events can be reconstructed. However, this also introduces noise into the final result as the Marchenko method tries to reconstruct these angles. To remove the noise at these high angles, dip filtering is applied to the homogeneous Green's function, which can be applied because no physical events are reconstructed for these angles. This creates some small artifacts around the first arrival of the homogeneous Green's function, which in turn are removed by applying a time-dependent taper. The improved result is shown in Figures 6m–6o at 0, 200, and 400 ms, respectively. None of the desired events have been removed; however, the artifacts around the source position are gone. When this result is compared to the modeled
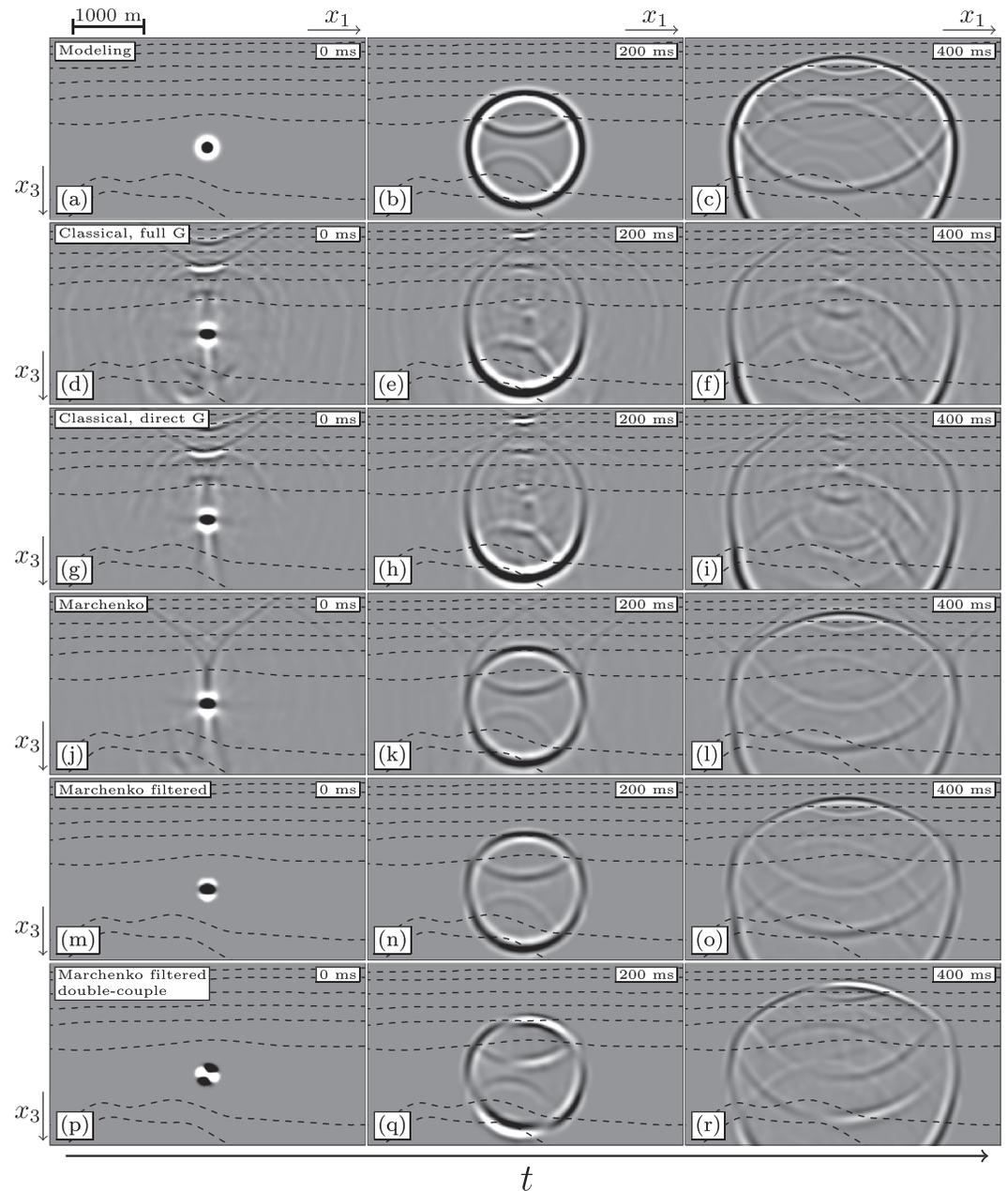
Figure 6. Snapshots of the wavefield at different times. Column 1 indicates $t = 0$ ms, Column 2 $t = 200$ ms, and Column 3 $t = 400$ ms. All wavefields have been convolved with a 15 Hz Ricker wavelet for display purposes. The black dotted lines indicate the locations of geological layer interfaces. (a–c) Directly modeled homogeneous Green's function in the subsurface used as a reference. (d–f) Homogeneous Green's function retrieval using equation (9) with full Green's functions for the virtual source and receiver positions. (g–i) Idem, however with a full Green's function for the virtual receiver position and a direct arrival as the Green's function for the virtual source position. (j–l) Homogeneous Green's function retrieval using equation (10), a Green's function for the virtual source position and a focusing function for the virtual receiver position, without filtering. (m–o) Idem, with dip filtering and tapering applied. (p–r) Idem, using a Green's function with a double-couple source signature inclined at −20°.

response, we find the match of the arrival times quite satisfactory, while the amplitude mismatch remains, and the improvement over the homogeneous Green's function obtained using the classical representation is significant. When comparing the result of this retrieval to the modeled result from Figures 6a–6c, the match between desired events and their arrival times are strong. A notable difference occurs in the shape of the source at zero time. When the wavefield is modeled directly, the source radiates uniformly in all directions; however, when the wavefield is retrieved with the Marchenko method, this is not possible. This is caused by the limited aperture, which cannot capture the parts of the wavefield that propagate horizontally at this depth. Additionally, the dip filtering that is applied will also affect the source radiation. Even with these limitations, this type of workflow yields the optimal result and is used for the retrieval of the homogeneous Green's function from here onward. The workflow is summarized in Figure 5a. Finally, we consider the workflow in Figure 5a with a double-couple source mechanism. Instead of using equation (10) for the homogeneous Green's function retrieval, we use equation (15). This means that we use a Green's function, with a double-couple source signature, obtained with the Marchenko method, to investigate whether the source signature has a strong effect on the retrieved homogeneous Green's function. A first arrival caused by a double-couple source, needed to initiate the Marchenko method, cannot be retrieved using the eikonal solver; therefore, we model the required first arrival using the finite-difference code. The double-couple source mechanism is incorporated into the code using the moment tensor approach for finite-difference modeling, as described by Li et al. (2014). The double-couple source mechanism is an elastic mechanism; however, it is assumed that all the data are acoustic. To model an acoustic response to the double-couple source, we use the smoothed P wave velocity model and a homogeneous density model of 1,000 kg/m^3, as well as a homogenous S wave velocity model of 1,000 m/s, except for the top layer where the S wave velocity is set to zero, simulating a marine setting. This means that no S waves will arrive at the receiver locations. The coda of this modeling will be incorrect; however, as we only use the first arrival, this has no consequences for our results, aside from the scaling of the first arrival. The first arrival will be a pure P wave, as the velocities of the P wave model are larger than those of the S wave model, which is a realistic situation.

We use a double-couple source that was inclined at $-20°$ to model the first arrival that it is used in the Marchenko method to create a Green's function. This Green's function replaces the one created from a monopole source that was used in the previous examples and the resulting homogeneous Green's function is shown in Figures 6p-6r. The arrival times of the events are the same as the ones in the previous experiment where only monopole signatures were employed. The main differences are found in the polarity of the events. Due to the double-couple source, the amplitude and polarity of the wavefield changes depending on the angle. Because the source we used to model the first arrival for the Green's function is inclined at $-20°$, these polarity changes are not occuring at 90°, 180°, 270°, and 360° but, shifted by $-20°$, at approximately 70°, 160°, 250°, and 340°. All retrieved events, not just the first arrival, are affected by this inclination, without introducing any additional artifacts. This shows that different source signatures can be incorporated into the single-sided representation for homogeneous Green's function retrieval.

4.2. Limitations of Reflection Data

The reflection response that we use to retrieve the results in Figure 6 is nearly ideal, due to the recording setup and the absence of absorption. In the following experiment, we perform the homogeneous Green's function retrieval using the workflow in Figure 5a, with five different types of acquisition limitations applied to the reflection response. In all five cases, we perform the entire process that is outlined in the workflow, starting with the Marchenko method to retrieve the focusing function and Green's function from the reflection response, to which one of the limitations is applied, followed by applying equation (10) to obtain the homogeneous Green's function. This demonstrates the effects of the limitations of the reflection response on the homogeneous Green's function that is obtained. The results of these tests are shown in Figure 7, where (a)–(c) show the result from Figures 6(f), 6(i), and 6(o), respectively, which are used as a reference. All snapshots in Figure 7 are shown at 400 ms. The results shown in the rows below the first one are achieved in the same way as the result from Figure 7c, with different types of limitations applied to the reflection response. Each column shows a varying value of the limitation to indicate the sensitivity of the method to these limitations.

To determine the Nyquist spatial sampling interval for the data, we use the sampling criterion:

$$\Delta x < \frac{c_1}{2 f_{max} sin(\theta_{max})}, \tag{18}$$
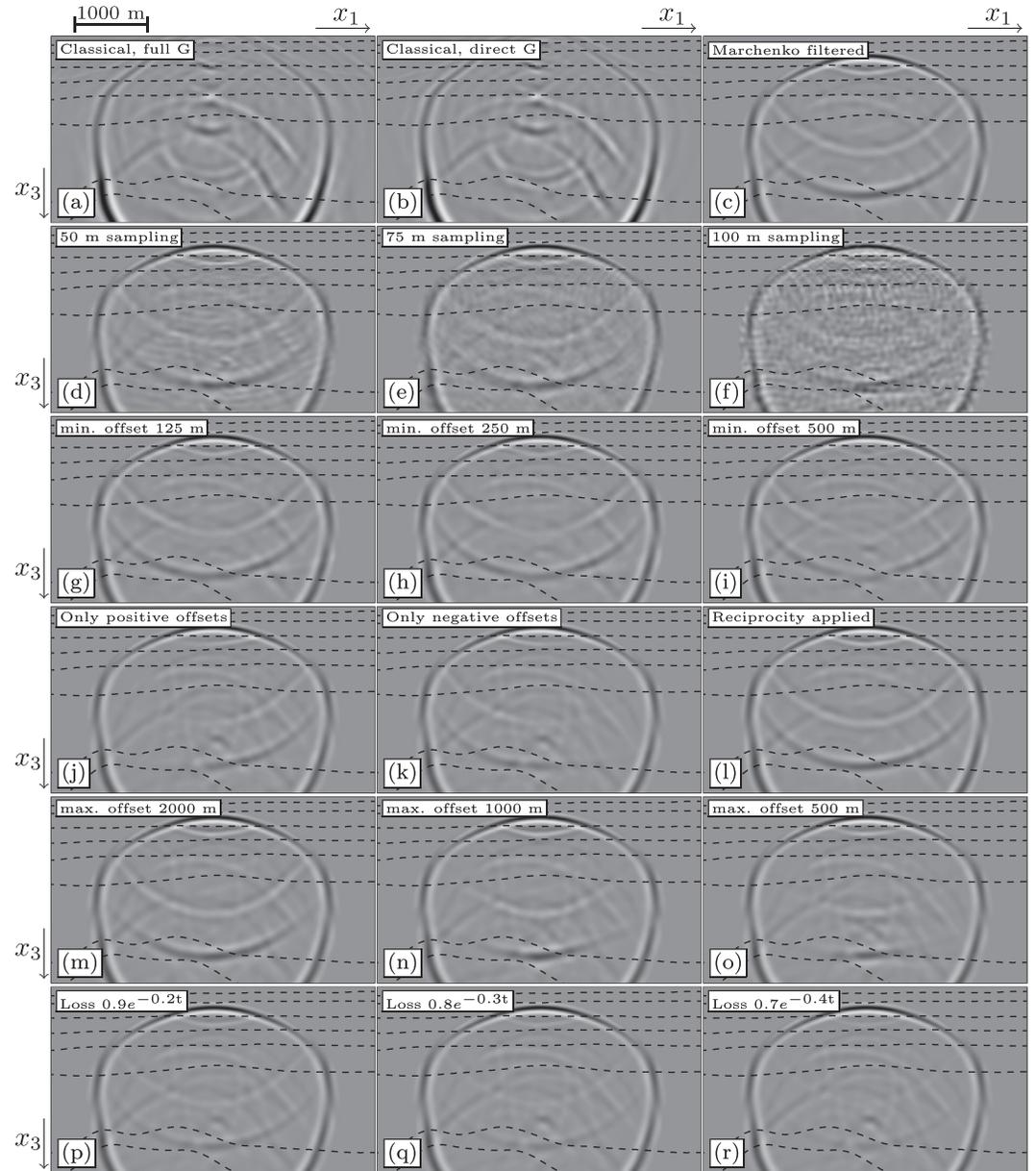
Figure 7. Snapshots of the homogeneous Green's function at $t = 400$ ms, retrieved using varying limitations of the reflection response, following the workflow in Figure 5a. All wavefields have been convolved with a 15 Hz Ricker wavelet. (a) Result from Figure 6f, (b) result from Figures 6i, and (c) result from Figure 6o , the latter used as a benchmark for the other results. (second row) Homogeneous Green's function obtained, when the reflection response has a source and receiver spacing of (d) 50, (e) 75, and (f) 100 m. (third row) Homogeneous Green's function obtained when the reflection response is missing the small source-receiver offsets up to a distance of (g) 125, (h) 250, and (i) 500 m. (fourth row) Homogeneous Green's function obtained when the reflection response contains (j) only positive source-receiver offsets, (k) only negative source-receiver offsets, and (l) has all source-receiver offsets restored using source-receiver reciprocity. (fifth row) Homogeneous Green's function obtained when the reflection response has an aperture limited to (m) 2,000, (n) 1,000, and (o) 500 m. (sixth row) Homogeneous Green's function obtained when the reflection response has a loss applied to it of (p) $0.9e^{-0.2t}$, (q) $0.8e^{-0.3t}$, and (r) $0.7e^{-0.4t}$.

where Δx is the spatial sampling, c_1 is the velocity of the top layer of the subsurface, f_{max} is the maximum frequency of the wavelet, and θ_{max} is the maximum angle of the waves in the top layer. To determine the sampling, we use the top layer velocity of the data, which corresponds to 1,500 m/s, a "maximum" frequency of 30 Hz for the 15 Hz Ricker wavelet and a maximum angle of 60°. This results in a recommended spatial sampling of about 29 m. The current sampling of 25 m should therefore be sufficient for our applications, which is supported by the quality of the retrieved homogeneous Green's function that have been obtained in the previous section. In Figures 7d–7f, we display the result retrieved using a reflection response that has a coarse source-receiver sampling. The sampling values for the receiver and source spacing are 50, 75, and 100 m, which are double, triple, and quadruple the original spacing and all exceed the Nyquist sampling. When the spacing is doubled, noise is introduced into the homogeneous Green's function due to spatial aliasing. The physical events are distorted by this noise, and background artifacts are present. This issue is worsened when the spacing distance is tripled. Some events are obscured, and strong noise is present. Quadrupling the spacing produces a result that is unusable. It consists almost entirely of noise, and the physical events cannot be distinguished. Hence, for successful use of the method, the source-receiver sampling of the reflection response must not be larger than the Nyquist sampling. Next, we consider the influence of missing small source-receiver offsets. The result is shown in Figures 7g–7i, where, respectively, the first 125, 250, and 500 m of the offsets are removed from the reflection response, for both positive and negative offsets, and replaced with empty traces. When 125 m of offsets are missing, the homogeneous Green's function is still comparable to the ideal situation. There is a degradation in quality, and a few artifacts are present. Removing 250 m of near offsets produces a less accurate homogeneous Green's function, with a stronger degradation in quality. When 500 m of near offsets are removed, the low angle reflections are not reconstructed properly, similarly to the high angle noise that is caused by the limited aperture of the data. The events below and above the virtual source position are missing, and strong artifacts are present. These artifacts can be partially removed by adjusting the dip filtering to remove the low angle events; however, this procedure will also remove part of the physical events and is therefore not reccomendable to apply. The near offsets do have an impact on the final result and ideally should be reconstructed, if possible, before applying the Marchenko method. Aside from missing near offsets, a reflection response can also be recorded exclusively in one direction. Figures 7j shows the result using only positive source-receiver offsets in the reflection response, and Figure 7k does the same for negative source-receiver offsets. In both cases, unwanted artifacts are present, and depending on the direction of the source-receiver offsets, large parts of the physical events are missing. This issue can be avoided by applying source-receiver reciprocity. We perform source-receiver reciprocity on the reflection response containing only the negative offsets and retrieve the result shown in Figure 7l. This homogeneous Green's function is similar to the one produced in the ideal situation. We retrieve a similar result when we apply source-receiver reciprocity on the reflection response containing only positive offsets. The final acquisition limitation that we review is the absence of large source-receiver offsets or aperture limitation of the data. In Figures 7m–7o, we show the homogeneous Green's function when the largest source-receiver offset is, respectively, 2,000, 1,000, and 500 m. When the aperture is 2,000 m, the homogeneous Green's function is comparable to the one that is retrieved in the ideal situation, with some artifacts introduced. When the aperture is limited to 1,000 m, a homogeneous Green's function is produced that contains more artifacts and is missing a larger portion of the angles of the desired events. This is once again due to the fact that the angles of this part of the wavefield are not present in the reflection response. If only 500 m of aperture are available, a large portion of the angles of events are missing. This is clear when Figure 7o is compared to Figure 7i. The parts of the events that are missing due to the lack of near offsets are present in the case of limited aperture and vice versa. By applying a stronger dip filter, the artifacts can be suppressed; however, as mentioned before, this also removes part of the physical events. Finally, we consider the case of absorption, which is a factor that cannot directly be influenced during the acquistion of the reflection response. Even if the recording setup is ideal, absorption of the wavefield is present and can degrade the result. This is demonstrated in Figures 7p–7r, where the loss is simulated by applying time-dependent absorption functions to the reflection data of $0.9e^{-0.2t}$, $0.8e^{-0.3t}$, and $0.7e^{-0.4t}$, respectively. In case of low absorption, the homogeneous Green's function still contains the physical events, although they have a lower amplitude. The artifacts are present with a low amplitude. If the absorption is increased, the physical events start to vanish, and the artifacts are more pronounced. In case of high absorption, the physical events have very low amplitude, and there are strong artifacts present.
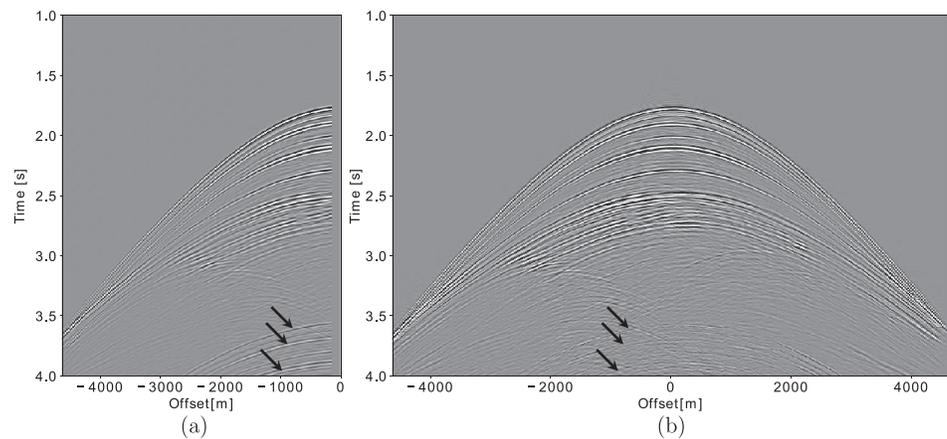
Figure 8. (a) Common-source record from Figure 3a before any processing is applied and (b) common-source record from (a) with source-receiver reciprocity, EPSI, and an exponential gain of $1.73e^{1.3t}$ applied. Both common-source records have their wavelets reshaped to a 30 Hz Ricker wavelet. Note the removal of the free-surface multiples at late times, as indicated by the black arrows.

5. Field Data

5.1. Preprocessing

The raw seismic field reflection data cannot directly be used with the Marchenko method, because when the method uses these data, it does not converge to a solution. The reflection data need to be preprocessed to compensate for the limitations (Staring et al., 2018). As was shown by the results in Figure 7, there are multiple effects that need to be taken into account. The reflection data are missing small source-receiver offsets, only have offsets in a negative direction, may be subsampled, possibly have insufficient aperture and are affected by absorption. Aside from this, the 2-D line was recorded in a 3-D setting, which can cause complications due to out of plane effects. The effects of having geometric spreading in a 3-D setting while the Marchenko scheme that is applied is 2-D are corrected by applying a time-dependent gain on the data. This gain is equal to $\sqrt{t}$, as the geometrical spreading in 3-D can be approximated as being $\frac{1}{t}$ and in 2-D as $\frac{1}{\sqrt{t}}$ (Berkhout, 1987). Next, source-receiver reciprocity is applied to the reflection data to create offsets in the positive direction. Having offsets in both positive and negative directions is vital for the next step, where the Estimation of Primaries through Sparse Inversion (EPSI) method is applied. The EPSI method estimates the primaries in the data, through the use of an inversion process, and estimates the free-surface multiples separately. This allows for the retrieval of a data set without the free-surface multiples. Furthermore, in this process, the information about the subsurface that is contained in the free-surface multiples is used to reconstruct the missing near offsets of the primaries. Simultaneously, an estimation of the source wavelet is made, which can be used to deconvolve the reflection data. Note that the EPSI method only removes the free-surface multiples. The internal multiples remain in the data after the application of the EPSI method. See van Groenestijn and Verschuur (2009) for a detailed overview of the EPSI method. The effects of absorption on the data are adaptively corrected for by applying exponential time-gain and a scaling factor. Note that the exponential time-gain is applied in addition to the $\sqrt{t}$ gain that was applied to account for the geometrical spreading factor. This second gain is intended to counteract the effects of the absorption on the data and therefore is estimated separately from the first gain. The first estimations of the time-gain are based on the velocity model conforming to the method discussed by Draganov et al. (2010). The scaling factor is estimated by minimizing the cost functions proposed by Brackenhoff (2016). After these processing steps are applied, the reflection data are used to retrieve a Green's function for one location in the subsurface. If the method does not converge to a solution, where the artifacts are minimal, the exponential gain and scaling factor are adjusted, and the test is run again. After a few iterations, we found that applying a gain of $1.73e^{1.3t}$ to the data resulted in a solution that converged with significant removal of artifacts. Interpolation was also tested by interpolating the source-receiver spacing on the reflection data to smaller values; however, we found that this did not significantly improve our results. Another aspect we cannot improve is the limited aperture of the data. The final workflow for the field data can be found in Figure 5b. An example of a common-source record before and after the preprocessing is shown in Figure 8. Note the improvement
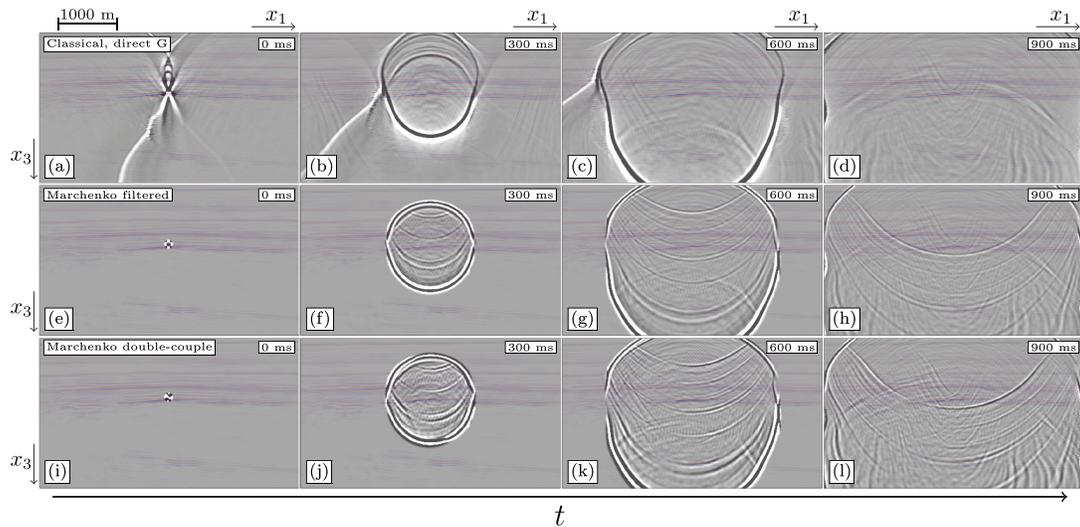
Figure 9. Snapshots of the homogeneous Green's function in the subsurface of the Vøring basin, obtained using the classical representation of equation (9) for a monopole source at (a) 0, (b) 300, (c) 600, and (d) 900 ms. Idem, using the single-sided representation of equation (10) for a monopole source at (e) 0, (f) 300, (g) 600, and (h) 900 ms. Idem, using the single-sided representation of equation (15) for a double-couple source inclined at −20° at (i) 0, (j) 300, (k) 600, and (l) 900 ms. All data contain a 30 Hz Ricker wavelet.

of the common-source record, as the coverage of the data has been increased and the free-surface multiples have been weakened, as indicated by the black arrows.

5.2. Homogeneous Green's Function Retrieval

After applying all the corrections, the Marchenko method is utilized to retrieve the required Green's functions and focusing functions in the region of interest that is indicated in Figure 3b. These functions are retrieved using only the preprocessed single-sided reflection response and a smooth velocity model. Next, we use these functions to retrieve the homogeneous Green's function using the single-sided boundary representation of equation (10). Snapshots of the obtained result are shown in Figures 9e–9h, for 0, 300, 600, and 900 ms, respectively. For comparison, the homogeneous Green's function obtained using the classical representation and using only the first arrival for the Green's function for the virtual source is shown in Figures 9a–9d, for 0, 300, 600, and 900 ms, respectively. In both cases, a monopole source was used to model all the first arrivals.

We also test the creation of a virtual source that has a double-couple source signature on the field data. We use a double-couple source inclined at −20° to create the first arrival that is used to obtain the Green's function for the virtual source using the Marchenko method. Similar to our approach on the synthetic data, we use the smooth P wave velocity model, a homogeneous density model of 1,000 kg/m^3, and a constant S wave velocity model of 1,000 m/s, with the top layer of water, where the S wave velocity is zero. We select the P wave first arrival and use the Marchenko method to create the Green's function for the virtual source. The homogeneous Green's function is obtained through equation (15) and the workflow in Figure 5b. Snapshots of the result are shown in Figures 9i–9l for 0, 300, 600,0 and 900 ms, respectively. For all images, a transparent overlay of the image from Figure 3c is used to indicate locations where scattering is expected. This image is only used for verification and was not used for the retrieval of the homogeneous Green's function. The results of the single-sided representation for the monopole source from Figures 9e–9h were previously shown in Wapenaar et al. (2018), and the results of the single-sided representation for the double-couple source from Figures 9i–9l were previously shown in Brackenhoff et al. (2019). The details about the preprocessing of the reflection data were not discussed before. The snapshots of the homogeneous Green's functions obtained using the single-sided representation for both types of sources show multiple events, both upgoing and downgoing. The locations of the scattering and the layer interfaces in the image overlay have a strong match, and aside from the primary reflections, the multiple reflections can also be seen. All of these events are completely absent when the classical representation is used. Strong artifacts are present in this case, and the coda of the downgoing wavefield is missing entirely. The primary downgoing wavefield is present; however, the upgoing primary wavefield is absent, which is similar to the result that was obtained on the synthetic

data. This shows that, as expected, the single-sided representation is an improvement over the classical representation on field data as well. The results for the double-couple source signature are similar to the ones that were obtained on the synthetic data. The polarity changes are present both on the primary wavefield and on the coda. Because of these changes, there are a few events in the coda where the amplitude is low, which makes it harder to distinguish these events. The results in Figure 9 are different from those that were obtained in Figure 6. As stated in section 3, this is because the synthetic data do not contain all the events that are present in the field data. The results on the field data contain more events, which could not be predicted from the image. This is an advantage of the data-driven approach of the Marchenko method. We find the overall results in Figure 9 encouraging. However, even the homogeneous Green's functions obtained using the single-sided representation are not perfect as there are still artifacts present. This is partially due to the presence of background noise in the data set, which distorts the final result. More coherent events are also present, which do not correlate with the primary wavefield and scattering locations from the image. Because we cannot be sure whether the reflection response has been preprocessed perfectly, there may be some low amplitude artifacts present that are nor fully compensated for or which are created by the Marchenko method. The Marchenko method that we applied was intended for 2-D acoustic media; however, the true medium is 3-D and elastic. As the geological layering for this region appears to be close to horizontal, the out-of-plane effects are assumed to be small. Due to conversion from P to S waves and back, some events are present in the reflection response that would not be present if the medium was purely acoustic and these are not handled correctly by our acoustic Marchenko implementation.

5.3. Discussion

The results on the field data show promise for further applications of the single-sided representation, for example, by combining them with passive measurements. In this case, the Green's function for the virtual source, which was constructed through the use of the Marchenko method, would be replaced with a passive recording. The accurately retrieved propagation and scattering of the wavefield in the inhomogeneous medium also holds much promise, despite the presence of some artifacts. If we wanted to track the paths of the wavefronts emitted, for example, by an induced seismic source, a result like this could provide substantial insight. However, there are several limitations. In this paper, we exclusively considered a 2-D data set that has few out-of-plane effects. For complex 3-D media, where the out-of-plane effects are much more severe, 3-D reflection data would be required, with sufficient coverage, as well as a 3-D smooth P wave velocity model. In this case, the issues with offset, sampling, and aperture that were present for the 2-D data would be more complex to deal with and will be present in two directions. Furthermore, passive measurement setups are usually sparser than for active measurements, which could yield additional complications if one wants to use a passive recording for the source of the homogeneous Green's function. These aspects are subject of ongoing research.

6. Conclusion

We demonstrated the data-driven generation of virtual receivers and a virtual source, which have the potential to improve, for example, the monitoring of the subsurface and the prediction of the complex response of different source mechanisms. We did this by utilizing a single-sided representation to retrieve the homogeneous Green's function in the subsurface. To this end, we applied the Marchenko method, which only requires a single-sided reflection response and a smooth velocity model. We showed that the single-sided representation produces significantly more accurate results than the classical representation when the reflection data are only available on a single-sided nonenclosing boundary, typically the Earth's surface. The sensitivity of the Marchenko method to limitations of the reflection data was investigated by manipulating synthetic reflection data. This showed that preprocessing of the reflection response to compensate for coarse source-receiver sampling, missing offsets, and absorption is vital for the succesful application of the Marchenko method. We considered a data set from the Vøring basin (offshore Norway), which was affected by some of such limitations and processed the data using geometric spreading correction, source-receiver reciprocity, the EPSI method, and applying a time-gain and scaling factor. The processed reflection response was used to obtain the focusing functions and Green's functions, needed to apply the representations for homogeneous Green's function retrieval. The homogeneous Green's function obtained using the single-sided representation shows potential in our opinion for wavefield monitoring in the subsurface, as the complete coda of the wavefield is recovered, which is not the case when the classical representation is used. The scattering occurs at locations that correlate with possible reflector locations.

Monopole source and double-couple source signatures can both be used in the Marchenko method and in the single-sided representation to obtain homogeneous Green's functions with the same source signature. To further explore this potential, more complex source mechanisms should be considered, such as dynamic fault planes, that are active over an extended area and time period. This includes taking into account the effects caused by elastic media instead of acoustic media.

References

Aki, K., & Richards, P. G. (2002). *Quantitative seismology*. San Fransisco: W.H. Freeman and Company.

Bakulin, A., & Calvert, R. (2005). Virtual shear source: A new method for shear-wave seismic surveys. In *SEG technical program expanded abstracts 2005*, Society of Exploration Geophysicists, Houston, Texas. https://doi.org/10.1190/1.2148265

Berkhout, A. J. (1987). *Applied seismic wave theory*. Amsterdam: Elsevier.

Brackenhoff, J. A. (2016). Rescaling of incorrect source strength using Marchenko redatuming (M.Sc. thesis), Delft, Zuid-Holland, the Netherlands. repository.tudelft.nl, http://resolver.tudelft.nl/uuid:0f0ce3d0-088f-4306-b884-12054c39d5da

Brackenhoff, J., Thorbecke, J., & Wapenaar, K. (2019). Monitoring induced distributed double-couple sources using Marchenko-based virtual receivers. *Solid Earth Discussions*, *2019*, 1–25. https://www.solid-earth-discuss.net/se-2018-142/

Broggini, F., Snieder, R., & Wapenaar, K. (2012). Focusing the wavefield inside an unknown 1D medium: Beyond seismic interferometry. *Geophysics*, *77*(5), A25–A28. https://doi.org/10.1190/geo2012-0060.1

Curtis, A., Nicolson, H., Halliday, D., Trampert, J., & Baptie, B. (2009). Virtual seismometers in the subsurface of the Earth from seismic interferometry. *Nature Geoscience*, *2*(10), 700.

Draganov, D., Ghose, R., Ruigrok, E., Thorbecke, J., & Wapenaar, K. (2010). Seismic interferometry, intrinsic losses and Q-estimation. *Geophysical Prospecting*, *58 (3)*, 361–373. https://doi.org/10.1111/j.1365-2478.2009.00828.x

Feynman, R. P., Leighton, R. B., & Sands, M. (2011). *The Feynman lectures on physics, Vol. I: The new millennium edition: Mainly mechanics, radiation, and heat* (Vol. 1): Basic books.

Grigoli, F., Cesca, S., Priolo, E., Rinaldi, A. P., Clinton, J. F., Stabile, T. A., et al. (2017). Current challenges in monitoring, discrimination, and management of induced seismicity related to underground industrial activities: A European perspective. *Reviews of Geophysics*, *55*, 310–340. https://doi.org/10.1002/2016RG000542

Li, D., Helmberger, D., Clayton, R. W., & Sun, D. (2014). Global synthetic seismograms using a 2-D finite-difference method. *Geophysical Journal International*, *197*(2), 1166–1183. https://doi.org/10.1093/gji/ggu050

Magnani, M. B., Blanpied, M. L., DeShon, H. R., & Hornbach, M. J. (2017). Discriminating between natural versus induced seismicity from long-term deformation history of intraplate faults. *Science Advances*, *3*(11), e1701593. https://doi.org/10.1126/sciadv.1701593

Morse, P. M., & Feshbach, H. (1953). *Methods of theoretical physics* (Vol. I). New York: McGraw-Hill Book Company Inc.

Oristaglio, M. L. (1989). An inverse scattering formula that uses all the data. *Inverse Problems*, *5*(6), 1097–1105. https://doi.org/10.1088/0266-5611/5/6/015

Porter, R. P. (1970). Diffraction-limited, scalar image formation with holograms of arbitrary shape. *Journal of the Optical Society of America*, *60*(8), 1051–1059. https://doi.org/10.1364/JOSA.60.001051

Porter, R. P., & Devaney, A. J. (1982). Holography and the inverse source problem. *Journal of the Optical Society of America*, *72*(3), 327–330. https://doi.org/10.1364/JOSA.72.000327

Ravasi, M., Vasconcelos, I., Kritski, A., Curtis, A., da Costa Filho, C. A., & Meles, G. A. (2016). Target-oriented Marchenko imaging of a North Sea field. *Geophysical Journal International*, *205*(1), 99–104. https://doi.org/10.1093/gji/ggv528

Reinicke, C., & Wapenaar, K. (2019). Elastodynamic single-sided homogeneous Green's function representation: Theory and numerical examples. *Wave Motion*, *89*, 245–264. https://doi.org/10.3997/2214-4609.201701220

Rose, J. H. (2001). "Single-sided" focusing of the time-dependent Schrödinger equation. *Physical Review A*, *65*(1), 012707. https://doi.org/10.1103/PhysRevA.65.012707

Singh, S., Snieder, R., Behura, J., van der Neut, J., Wapenaar, K., & Slob, E. (2015). Marchenko imaging: Imaging with primaries, internal multiples, and free-surface multiples. *Geophysics*, *80*(5), S165–S174. https://doi.org/10.1190/geo2014-0494.1

Slob, E., Wapenaar, K., Broggini, F., & Snieder, R. (2014). Seismic reflector imaging using internal multiples with Marchenko-type equations. *Geophysics*, *79*(2), S63–S76.

Spetzler, J., & Angelov, P. (2005). Ray perturbation theory for traveltime and amplitude attributes in 3D heterogeneous media with weak anisotropy, *75th SEG annual meeting*.

Staring, M., Pereira, R., Douma, H., van der Neut, J., & Wapenaar, K. (2018). Source-receiver Marchenko redatuming on field data using an adaptive double-focusing method. *Geophysics*, *83*(6), 1–48.

Thorbecke, J., & Brackenhoff, J. (2019). *Open source*. Delft, The Netherlands: Zenodo. https://doi.org/10.5281/zenodo.3374728

Thorbecke, J., & Draganov, D. (2011). Finite-difference modeling experiments for seismic interferometry. *Geophysics*, *76*(6), H1–H18. https://doi.org/10.1190/geo2010-0039.1

Thorbecke, J., Slob, E., Brackenhoff, J., van der Neut, J., & Wapenaar, K. (2017). Implementation of the Marchenko method. *Geophysics*, *82*(6), WB29–WB45. https://doi.org/10.1190/geo2017-0108.1

Thorbecke, J. W., Wapenaar, K., & Swinnen, G. (2004). Design of one-way wavefield extrapolation operators, using smooth functions in WLSQ optimization. *Geophysics*, *69*(4), 1037–1045. https://doi.org/10.1190/1.1778246

van Groenestijn, G. J., & Verschuur, D. J. (2009). Estimating primaries by sparse inversion and application to near-offset data reconstruction. *Geophysics*, *74*(3), A23–A28. https://doi.org/10.1190/1.3111115

van Manen, D.-J., Robertsson, J. O. A., & Curtis, A. (2005). Modeling of wave propagation in inhomogeneous media. *Physical Review Letters*, *94*(16), 164301. https://doi.org/10.1103/PhysRevLett.94.164301

van Thienen-Visser, K., & Breunese, J. N. (2015). Induced seismicity of the Groningen gas field: History and recent developments. *The Leading Edge*, *34*(6), 664–671. https://doi.org/10.1190/tle34060664.1

Verschuur, D. J., Berkhout, A. J., & Wapenaar, C. P. A. (1992). Adaptive surface-related multiple elimination. *Geophysics*, *57*(9), 1166–1177. https://doi.org/10.1190/1.1443330

Vidale, J. (1988). Finite-difference calculation of travel times. *Bulletin of the Seismological Society of America*, *78*(6), 2062–2076.

Wapenaar, K. (2004). Retrieving the elastodynamic Green's function of an arbitrary inhomogeneous medium by cross correlation. *Physical Review Letters*, *93*, 254301. https://doi.org/10.1103/PhysRevLett.93.254301

Wapenaar, K., Brackenhoff, J., Thorbecke, J., van der Neut, J., Slob, E., & Verschuur, E. (2018). Virtual acoustics in inhomogeneous media with single-sided access. *Scientific Reports*, *8*(1), 2497. https://doi.org/10.1038/s41598-018-20924-x

Wapenaar, K., Broggini, F., Slob, E., & Snieder, R. (2013). Three-dimensional single-sided Marchenko inverse scattering, data-driven focusing, Green's function retrieval, and their mutual relations. *Physical Review Letters*, *110*, 084301. https://doi.org/10.1103/PhysRevLett.110.084301

Wapenaar, K., Thorbecke, J., van Der Neut, J., Broggini, F., Slob, E., & Snieder, R. (2014). Marchenko imaging. *Geophysics*, *79*(3), WA39–WA57. https://doi.org/10.1190/geo2013-0302.1

Wapenaar, K., Thorbecke, J., & van der Neut, J. (2016). A single-sided homogeneous Green's function representation for holographic imaging, inverse scattering, time-reversal acoustics and interferometric Green's function retrieval. *Geophysical Journal International*, *205*(1), 531–535. https://doi.org/10.1093/gji/ggw023

Wapenaar, K., Thorbecke, J., van der Neut, J., Slob, E., & Snieder, R. (2017). Virtual sources and their responses, part II: Data-driven single-sided focusing. *Geophysical Prospecting*, *65*(6), 1430–1451.

Yilmaz, Ö. (2001). *Seismic data analysis* (Vol. 1). Tulsa, OK: Society of Exploration Geophysicists.