

Estimation of primaries by sparse inversion with scattering-based multiple predictions for data with large gaps

Tim T.Y. Lin¹ and Felix J. Herrmann¹

¹Seismic Laboratory for Imaging and Modeling (SLIM), Department of Earth, Ocean & Atmospheric Sciences, University of British Columbia

ABSTRACT

We propose to solve the Estimation of Primaries by Sparse Inversion problem without any explicit data reconstruction, from a seismic record with missing near-offsets and large holes in the acquisition grid which does not produce aliasing but otherwise causes large errors in the multiple prediction. Exclusion of the unknown data as an inversion variable from the process is desirable, since it sidesteps possible issues arising from fitting observations that are also unknowns in the same inversion problem. Instead, we simulate their multiple contributions by augmenting the forward prediction model for the total wavefield with a scattering series that mimics the action of the free surface reflector within the confines of the unobserved trace locations. Each term in this scattering series simply involves convolution of the predicted wavefield once more with the current estimated surface-free Green's function at these unobserved locations. We investigate the necessary modifications to the primary estimation algorithm to account for the resulting nonlinearity in the modeling operator, and also demonstrate that just a few scattering terms are enough to satisfactorily mitigate the effects of near-offset data gaps during the inversion process. Numerical experiments on synthetic data show that the final derived method can significantly outperform explicit data reconstruction for large near-offset gaps. This is achieved with a similar computational cost and better memory efficiency compared to explicit data reconstruction. We also show on real data that our scheme outperforms pre-interpolation of the near-offset gap.

INTRODUCTION

Surface-related multiple events are among the largest causes of errors in seismic subsurface imaging, and its removal from survey records is an important aspect of seismic processing. It is generally accepted that the most accurate model of surface-related multiples are derived from wave equation-based, data-driven approaches, such as ones exploiting the Rayleigh-Helmholtz reciprocity relationship between the field-measured wavefield data and its multiple-free version (Fokkema and van den Berg, 1993; Frijlink et al., 2011). The most well-known method of this type is what came to be known as Surface-Related Multiple Elimination (SRME, Verschuur, 1992; Berkhout and Verschuur, 1997; Verschuur and Berkhout, 1997). There are also many other contemporary methods that also rely on the same physical principle but with a modified problem structure (Ziolkowski et al., 1999; Amundsen, 2001; Majdański et al., 2011).

Perhaps the biggest technical challenges of this family of methods is the requirement for recorded wavefields to be completely sampled over the survey area. This is because

primaries in the measured wavefield are used as the source functions for the multiple wavefield. Primaries that are not measured causes gaps in the subsequent ray paths that reflect back down from the free surface, and consequently can lead to serious errors in the final surface multiple prediction. This characteristic behavior poses serious budget and technical challenges for survey design, and is one of the main motivations behind the decades-long effort to improve near-offset acquisition, since near-offset primaries are known to be the strongest contributors of surface multiples. While it is now not uncommon for deep-water surveys to use a combination of cables in over/under configurations and off-azimuth recordings to reconstruct the ideal near-offset traces, this is still nearly impossible for shallow-water and on-shore surveys.

Therefore, current practice calls for intricate, on-the-fly trace interpolations, by inserting moveout-corrected versions of kinematically similar traces into non-sampled locations as a preprocessing step of the SRME multiple prediction. However, this paradigm is now being challenged by the increasing ubiquity of simultaneous or blended source acquisition designs, where overlapping wavefronts of conflicting dip and velocity are the norm. Alternative interpolation methods that are more immune to this situation but coming at a higher computational cost are generally transform-based, such as FK-domain reconstruction (sometimes with weighted matching-pursuit), the venerable parabolic Radon-based interpolation (Kabir and Verschuur, 1995), as well more contemporary methods based on advanced transforms such as curvelets (Herrmann and Hennenfent, 2008; Hennenfent et al., 2010) or exploiting low-rank structures (Ma, 2013; Kreimer et al., 2013; Kumar et al., 2015). Unfortunately, these methods are all at their weakest when it comes to reconstructing parts of the wavefront with high curvature and amplitude variation, which is typically characteristic of near-offset traces in seismic data.

Some recent trends have developed where these data are explicitly reconstructed by looking at the multiple information. A method that attempts to formulate multiple removal from the view of mapping multiple arrivals back to primaries is the Estimation of Primaries by Sparse Inversion (EPSI, van Groenestijn and Verschuur, 2009). This approach involves finding the correct surface-free Green’s function that will unambiguously relate all wavefronts in the observed data under a primary-multiple relationship through the action of the free surface. In the original formulation of EPSI, it was proposed to include missing data as yet another piece of unknown that must be found to be consistent with this model, forcing the algorithm to alternate between estimating the surface-free (primary) Green’s function, the missing data, and the source. While the initial results of this explicit reconstruction technique in EPSI has been successful, it fails to take advantage of the tightly-coupled relationship between the missing data and the surface-free Green’s function, by ignoring their partial derivatives with respect to each other in the inversion process. Our main contribution in this work is to come up with a formulation where the effects of missing data—i.e., a mask acting on the data matrix zeroing entries where data is missing—are explicitly accounted for in the forward model of EPSI through scattering terms at the surface boundary, modeled by repeated autoconvolution with the observed part of the wavefield.

Our work is organized as follows. First, we briefly summarize the EPSI formulation as an alternating optimization problem inverting for the source and surface-free Green’s by promoting sparsity via the ℓ_1 -norm on the latter. We demonstrate the effect of data gaps in the near-offset on EPSI predictions and contrast it with its impact on SRME. Next, we discuss the method proposed by van Groenestijn and Verschuur (2009), which extends EPSI

to additionally invert for missing parts of the upgoing wavefield. We then contrast it with our proposed method, where multiple contributions from the missing data are directly accounted for using scattering terms of the estimated model of the Green’s function on just the observed part of the wavefield. These terms account for the impact of missing data on the prediction of surface-related multiples. The main difference of our method is that it allows formation of exact gradients for the full Green’s function using, using correlating-type interferometry explain data that we have, and Born scattering at the surface to emulate data that we don’t have. We conclude by demonstrating the efficacy of this method on both synthetic and field data sets.

THEORY

Similar to SRME, the multiple model of EPSI relies on wavefield convolution to add one roundtrip through the Earth’s subsurface to all wave paths in the observed data. It assumes the existence of a wavefield kernel $g(\mathbf{x}, \omega; \mathbf{x}_a)$ for all frequencies ω that, under an surface integral, produces all the multiples p_m generated by that surface from a noiseless, densely-sampled pressure wavefield $p(\mathbf{x}, \omega; \mathbf{x}_{\text{src}})$. The physical significance of g will be discussed below, for now we simply require it to satisfy

$$p_m(\mathbf{x}, \omega; \mathbf{x}_{\text{src}}) = \int_S g(\mathbf{x}, \omega; \mathbf{x}') r(\mathbf{x}, \mathbf{x}') p(\mathbf{x}', \omega; \mathbf{x}_{\text{src}}) d\mathbf{x}'. \quad (1)$$

Throughout this paper S is treated as both the multiple-generating free surface and the observation surface on which \mathbf{x}' lies. This assumption can be relaxed by careful treatment of the wavefield directionality of p and g based on where we position S in relation to both surfaces (see Frijlink et al., 2011, for more details). The operator r models the reflection of the free surface, and is typically assumed to be isotropic and well approximated by a Dirac delta function $\delta(\mathbf{x} - \mathbf{x}')$ scaled by a reflectivity coefficient of -1 , although this can again possibly be relaxed and estimated (AlMatar, 2010).

The surface integration in expression 1 with g effectively sends all wave paths in p *exactly* one more time through the Earth’s subsurface using Huygen’s principle. This transforms all detected primary events to the first-order surface multiples, as well as all later n -th ordered multiples to the $n+1$ order (Anstey and Newman, 1966; Riley and Claerbout, 1976). In order for equation 1 to only produce multiple events, both p and g should be free of direct arrivals, as they would map to primary events.

The kernel g that satisfies this property can be physically interpreted as the normal derivative (at surface S) of a particular wave equation’s associated Green’s function. The propagation medium of this wave equation has no multiple-generating discontinuity at S , but is otherwise identical to the Earth’s subsurface which produced p . Thus, g can be seen as the primary wavefield p_o component of p , while also separated from the source wavelet q with the relation $p_o(\mathbf{x}, \omega; \mathbf{x}_{\text{src}}) = g(\mathbf{x}, \omega; \mathbf{x}_{\text{src}}) q(\omega)$. To keep this paper manageable in scope, we assume that all shots have roughly the same source function. We also refer to g as the Green’s function instead of its normal derivative, and simply assume that all surface integrals involving it includes a well-known obliquity factor.

In SRME, g is typically approximated with the observed wavefield p . This leads to an incorrect model of the surface multiples in terms of amplitude and spectral content, and relies on subsequent adaptive subtraction (Verschuur, 1992) in order to match it correctly

with the recorded multiple events. On the other hand, EPSI tries to directly solve for a discretized wavefield kernel g that can simultaneously explain both the multiples through expression 1 as well as the primaries using a jointly-estimated wavelet. As we will show, this fundamental difference allows us to address deficiencies in the acquisition of p by augmenting the functionals involved in EPSI’s variational problem.

The EPSI forward model

With the assumptions given above, we have an expression that maps g to both the primary and the surface-related multiple events in the total wavefield p . We represent this model as a time-domain operator $M(g, q; p)$ that is a function of g and the source wavelet q , parameterized by the data:

$$\begin{aligned} M(g, q; p) &:= \mathcal{F}_\omega^{-1} [g(\mathbf{x}, \omega; \mathbf{x}_{\text{src}})q(\omega) + \int_S g(\mathbf{x}, \omega; \mathbf{x}')r(\mathbf{x}, \mathbf{x}')p(\mathbf{x}', \omega; \mathbf{x}_{\text{src}}) d\mathbf{x}'] \quad (2) \\ &= p_o(\mathbf{x}, t; \mathbf{x}_{\text{src}}) + p_m(\mathbf{x}, t; \mathbf{x}_{\text{src}}) \\ &= p(\mathbf{x}, t; \mathbf{x}_{\text{src}}), \end{aligned}$$

where \mathcal{F}_ω^{-1} represents the continuous inverse Fourier transform.

In practice, EPSI follows the discretize-then-optimize paradigm, which attempts to solve directly for discretized representations \mathbf{g} and \mathbf{q} (discretized entities will be denoted by bold font) using numeric algorithms. We are thus only concerned with the discretized version of expression 2. Assuming a regularized sampling of source and receiver coordinates, the surface integral in expression 1 can also be written as simple matrix multiplications between monochromatic slices of data volumes, following the detail-hiding notation of Berkhout and Pao (1982). We indicate a mono-frequency matrix view of a wavefield by writing it in upper-case letters, and omitting writing the implied frequency dependence for brevity. For example, a mono-frequency data matrix for the wavefield \mathbf{p} (of size $n_t \times n_{\text{rcv}} \times n_{\text{src}}$) is written simply as \mathbf{P} (of size $n_{\text{rcv}} \times n_{\text{src}}$). We adhere to the convention of row (column) indices of a data matrix corresponding to the discretized receiver (source) positions. Using this notation, the discretized forward model is thus

$$\begin{aligned} M(\mathbf{g}, \mathbf{q}; \mathbf{p}) &= \mathbf{F}_\omega^{-1} [M_\omega(\mathbf{G}, \mathbf{Q}; \mathbf{P})], \quad (3) \\ M_\omega(\mathbf{G}, \mathbf{Q}; \mathbf{P}) &:= \mathbf{GQ} + \mathbf{GRP}. \end{aligned}$$

The operator \mathbf{F}_ω^{-1} is the (padded) discrete inverse Fourier transform that prevents convolutional wraparound effects in the time domain. The term \mathbf{GQ} produces the primary events, while the \mathbf{GRP} term produces surface multiples. Note that the discretized notation allows for a more arbitrary source function compared to the continuous expression, but we reconcile the two by parameterizing $\mathbf{Q} = q(\omega)\mathbf{I}$. Similarly, the reflectivity operator is made isotropic with reflection coefficient -1 , as above, by assuming $\mathbf{R} = -\mathbf{I}$. For clarity, we show in Figure 1 a common shot gather example of all the wavefields involved in the preceding expressions.

Simultaneously solving for \mathbf{g} as well as the source wavelet \mathbf{q} will in most cases, in correspondence to nondeterministic blind-deconvolution problems, admit non-unique solutions (Biggs, 1998; Bronstein et al., 2005; Ahmed et al., 2012). Thus, obtaining a useful solution requires additional regularization on the unknowns. Based on the argument that a discretized representation of the Green’s function \mathbf{g} for a wave-equation (in 3-D) resemble a wavefield

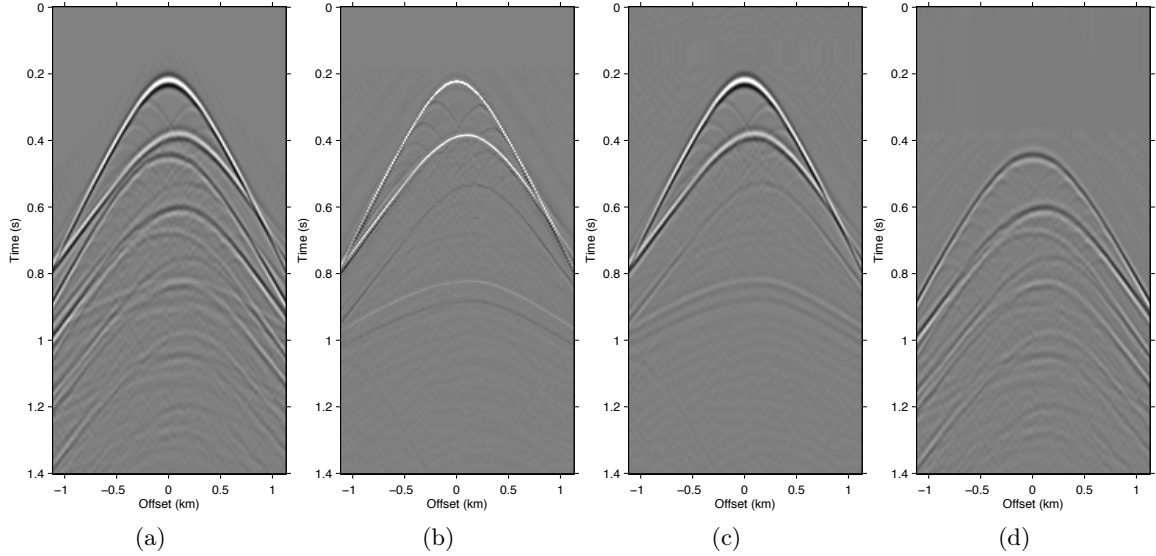


Figure 1: Shot gathers of the various wavefields involved in this paper. (a) the complete pressure data \mathbf{p} , (b) the true primary Green's function \mathbf{g} . Producing this is the primary objective of the EPSI problem. (c) the EPSI primary model produced by the \mathbf{GQ} term in expression 3. (d) the EPSI multiple model produced by the \mathbf{GRP} term in expression 3.

with impulsive wavefronts, the Robust EPSI algorithm (REPSI, Lin and Herrmann, 2013) proposes to find the *sparsest* possible \mathbf{g} (in the time domain) through ℓ_1 -norm minimization. Specifically, it solves the following optimization problem:

$$\begin{aligned} \min_{\mathbf{g}, \mathbf{q}} \|\mathbf{g}\|_1 \quad \text{subject to} \quad & f(\mathbf{g}, \mathbf{q}; \mathbf{p}) \leq \sigma, \\ & f(\mathbf{g}, \mathbf{q}; \mathbf{p}) := \|\mathbf{p} - M(\mathbf{g}, \mathbf{q}; \mathbf{p})\|_2, \end{aligned} \quad (4)$$

where $f(\mathbf{g}, \mathbf{q}; \mathbf{p})$ is the misfit functional, which we to seek minimize below a tolerance σ , and the modeling operator $M(\mathbf{g}, \mathbf{q}; \mathbf{p})$ is as defined in expression 3. As described in Lin and Herrmann (2013), a good solution to this problem can be found by employing an alternating optimization method that, after a proper initialization for \mathbf{g} , switches between updates on \mathbf{q} and \mathbf{g} while gradually relaxing an ℓ_1 -norm constraint on \mathbf{g} in a deterministic way.

We will return to discuss the algorithmic aspect of solving problem (4) in more detail after describing below the effects of missing data on $M(\mathbf{g}, \mathbf{q}; \mathbf{p})$.

Effects of incomplete data coverage

Up to this point we have assumed that the observed wavefield \mathbf{p} has complete coverage over all source-receiver coordinates. We now consider the case where \mathbf{p} is spatially well-sampled enough for the multiple model in Equation 3 to have no aliasing issues, but otherwise has significant gaps in its sampling coverage, such as a near-offset gap. To describe the effects of coverage issues in field acquisition geometries, we adhere to the notation of van Groenestijn and Verschuur (2009) that segregates all entries of \mathbf{p} into actual sampled data \mathbf{p}' and missing

data \mathbf{p}'' , both of equal dimensions to \mathbf{p} , such that $\mathbf{p} = \mathbf{p}' + \mathbf{p}''$. Remember that \mathbf{p} is the hypothetical fully sampled (in terms of both source and receiver locations) data at a fixed, regular sampling grid. We can model the available data \mathbf{p}' using a masking operator \mathbf{K} that eliminates (by setting to zero) traces belonging to unrecorded source-receiver coordinates. Similarly, we can map to the missing data \mathbf{p}'' by the compliment of this mask $\mathbf{K}_c := \mathbf{I} - \mathbf{K}$ (a stencil operator). This gives us

$$\mathbf{p} = \mathbf{p}' + \mathbf{p}'' = \mathbf{K}\mathbf{p} + \mathbf{K}_c\mathbf{p}.$$

The separation can also be done on the monochromatic data-matrices through a binary masking matrix \mathbf{A} , with the same dimensions as the data-matrices \mathbf{G} and \mathbf{P} . The elements of \mathbf{A} is defined 1 at sampled positions and 0 otherwise. In this case applying the mask is equal to a Hadamard product $\mathbf{P}' = \mathbf{K}\mathbf{P} := \mathbf{A} \circ \mathbf{P}$. Likewise, the unacquired part of the data-matrix comes from applying the complement stencil $\mathbf{P}'' = \mathbf{K}_c\mathbf{P}$.

When the difference between \mathbf{p} and \mathbf{p}' is large, the incomplete coverage can become a significant source of error in the calculation of the surface multiples, especially at the near-offset locations (Verschuur, 2006). The most well-known example of this is missing near-offsets, as illustrated in Figure 2. The effects of these gaps on the multiple model are felt throughout the wavefield volume due to the surface integral that underlies the prediction process, which results in a non-stationary spatial convolution.

The approximation $\mathbf{G} \approx \mathbf{P}$ that SRME doubles the impact of missing data during acquisition. Comparing Figure 1d with Figures 2b and 2c, it is clear that using \mathbf{P}' in place of \mathbf{G} greatly exasperates kinematic errors in the multiple prediction, resulting in completely unrecognizable wavefronts even when we have data available up to 100 m. Due to this effect, SRME usually requires some type of trace interpolation as a preprocessing step, since adaptive subtraction alone is not able to account for this degree of error.

However, on the flip side, the non-local effect of the prediction error ultimately enables us to probe the (actions of the) free surface inside the data gap based solely on measurements outside of it. Since EPSI inherently relies on wavefield inversion, there is enough degrees of freedom in the process to take advantage of this. van Groenestijn and Verschuur (2009) proposes to explicitly reconstruct \mathbf{p}'' from intermediate estimates of \mathbf{g} , such that the accuracy of M improves as the algorithm converges. Adapting this approach to the REPSI formulation means extending the definition of M_ω with the unobserved traces as another variable in the forward modeling operator

$$M_\omega(\mathbf{G}, \mathbf{Q}, \mathbf{P}''; \mathbf{P}') := \mathbf{G}\mathbf{Q} + \mathbf{G}\mathbf{R}(\mathbf{P}' + \mathbf{P}''), \quad (5)$$

which leads to a more complicated version of problem 4, with \mathbf{p}'' as an additional unknown:

$$\begin{aligned} \min_{\mathbf{g}, \mathbf{q}, \mathbf{p}''} \|\mathbf{g}\|_1 \quad \text{subject to} \quad & f(\mathbf{g}, \mathbf{q}, \mathbf{p}''; \mathbf{p}') \leq \sigma, \\ f(\mathbf{g}, \mathbf{q}, \mathbf{p}''; \mathbf{p}') := & \|\mathbf{p}' + \mathbf{p}'' - M(\mathbf{g}, \mathbf{q}, \mathbf{p}''; \mathbf{p}')\|_2. \end{aligned} \quad (6)$$

A major issue with this method of independently updating \mathbf{g} and \mathbf{p}'' in a cyclically alternating fashion is that we end up needing to include the intermediate and incorrect estimates of \mathbf{p}'' into the observation term of the misfit functional. Otherwise, the parts of \mathbf{G} inside the unobserved locations will tend to converge very slowly, if at all, since its gradients

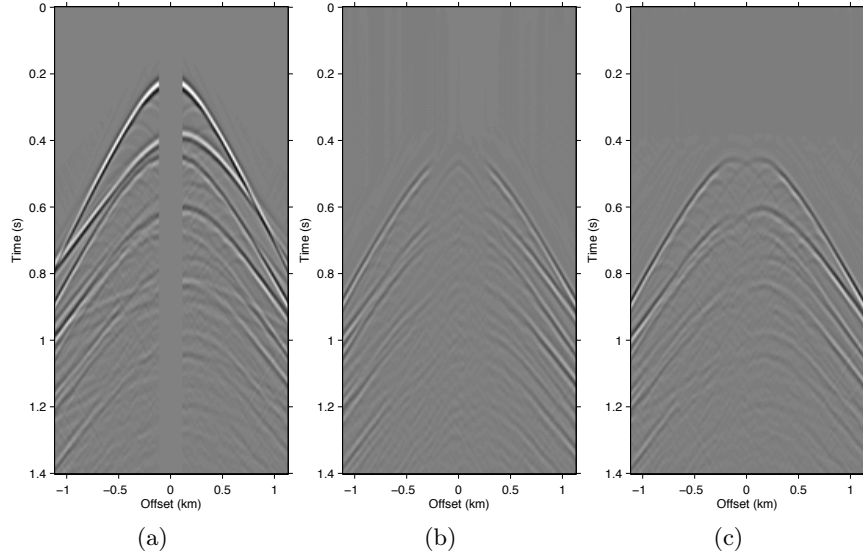


Figure 2: The effects of a 100 m near-offset gap in the observed data \mathbf{p}' shown in (a) on both SRME and EPSI multiple predictions. (b) a shot gather of the SRME multiple prediction produced by auto-convolution of the observed data $\mathbf{P}'\mathbf{P}'$. (c) a shot gather of the EPSI multiple prediction produced by convolution with the correct, fully sampled, primary Green's function $\mathbf{G}\mathbf{P}'$. Despite knowing the true primary Green's function, the EPSI multiple prediction still contains errors caused by the missing near-offsets, although it is much more accurate compared the SRME prediction which is twice affected by the near-offset gap.

would not be driven by the residuals that belong to the primaries, which only appears in \mathbf{p}'' inside the gap.

This approach also ignores the fact that \mathbf{g} and \mathbf{p}'' are tightly coupled in an obvious way. The strongest parts of these two wavefields (the primaries) overlap and relate to one another through convolution with the wavelet. The full relation is implicitly defined by

$$\mathbf{P}'' = \mathbf{K}_c \mathbf{P} = \mathbf{K}_c [\mathbf{G}\mathbf{Q} + \mathbf{G}\mathbf{R}\mathbf{P}' + \mathbf{G}\mathbf{R}\mathbf{P}''] \quad (7)$$

where it is evident that \mathbf{P}'' can be almost completely characterized by \mathbf{G} . Therefore, we would expect that $\partial\mathbf{p}''/\partial\mathbf{g}$ and $\partial\mathbf{g}/\partial\mathbf{p}''$ should not be ignored when independently updating \mathbf{p}'' and \mathbf{g} . While $\partial\mathbf{p}''/\partial\mathbf{g}$ is straightforward to compute, the expression for $\partial\mathbf{g}/\partial\mathbf{p}''$ is, conversely, quite tricky and necessarily involves \mathbf{Q}^{-1} (deconvolution by the unknown wavelet function \mathbf{q}). Motivated by this observation, in this paper we seek to remove \mathbf{p}'' as an inversion variable all-together, and instead account for its multiple contribution with terms that only involve \mathbf{g} , \mathbf{q} , and the known data \mathbf{p}' .

Deterministic correction of surface multiple prediction by scattering terms

We start by attempting to eliminate \mathbf{p}'' from the EPSI modeling function (expression 5) by recursively substituting it with the right-hand side of expression 7. This results in a new modeling function that does not depend on \mathbf{p}'' , but at the cost of having infinitely many

terms in a series expansion

$$\begin{aligned} \widetilde{M}_\omega(\mathbf{G}, \mathbf{Q}; \mathbf{P}') &= \mathbf{K}[(\mathbf{GQ} + \mathbf{GRP}') \\ &\quad + \mathbf{GRK}_c(\mathbf{GQ} + \mathbf{GRP}')] \end{aligned} \quad (8)$$

$$\begin{aligned} &\quad + \mathbf{GRK}_c(\mathbf{GRK}_c(\mathbf{GQ} + \mathbf{GRP}')) \\ &\quad + \mathcal{O}(\mathbf{G}^4)] \end{aligned} \quad (9)$$

$$:= \mathbf{K} \sum_{n=0}^{\infty} (\mathbf{GRK}_c)^n (\mathbf{GQ} + \mathbf{GRP}'). \quad (10)$$

This expression is no longer linear in \mathbf{G} . The $n=1$ term (Equation 8) is a second-order term since it is effectively a function of \mathbf{G}^2 , and similarly the $n=2$ term (Equation 9) is a third-order term.

Since multiples that we observe in nature will decay in amplitude as a function of its order, we can assume from physical arguments that the spectral norm (i.e., largest singular value) of \mathbf{G} is $\|\mathbf{G}\| < 1$, and therefore the infinite-term series expansion 10 converges to

$$\mathbf{KP} = \mathbf{K}(\mathbf{I} - \mathbf{GRK}_c)^{-1}(\mathbf{GQ} + \mathbf{GRP}'). \quad (11)$$

This expression states that we can recover the original EPSI forward model $\mathbf{P} = \mathbf{GQ} + \mathbf{GRP}' + \mathbf{GRP}''$ if it is somehow possible to invert the data mask \mathbf{K} by some hypothetical exact data interpolation, which validates our recursive substitution derivation above.

Although Equation 11 is not practically useful, it admits an useful interpretation of our approach: that of a Born series solution to a Lippmann-Schwinger relation between the total wavefield \mathbf{P} and an ‘‘incident wavefield’’ composed of the erroneous EPSI prediction $(\mathbf{GQ} + \mathbf{GRP}')$ from incomplete data. The perturbation in this case is the missing contribution coming from within the acquisition gap. This view makes it clear that each term in expression 10 is a term in the Born series expansion due to the free-surface reflector within the confines of the acquisition gap (as bound by the stencil \mathbf{K}_c). One direct consequence of this interpretation is that we can expect all orders of surface multiples in every term of expression 10, because the incident wavefield is not only composed of the primaries \mathbf{GQ} but also parts of all subsequent orders of surface multiples, as produced by the term \mathbf{GRP}' .

This scattering-based interpretation is perhaps most commonly seen in literature on the inverse scattering series (ISS) approach to multiple removal (Weglein et al., 1997, 2003). However, unlike typical applications of ISS methodology where it is used on subsurface reflectors for internal multiple removal, in this paper we know exactly that the scattering boundary is at the free surface, and have only reflection effect from this scatterer. Perhaps more significantly, we use the derived multiples as part of the modeling function in the context of a full data-fitting inversion problem for the primary estimation, instead of an adaptive subtraction scheme.

As with all series expansion expressions, it is important to investigate how the approximation error behaves as a function of the number of terms used. Since computing these terms involve partial wavefield convolutions that have significant computation costs (although not as much as the full wavefield convolution, as we will discuss later in the paper), it is practically desirable to limit the number of terms as much as possible. We now claim what is perhaps the main insight of this paper: that a very small number of these terms, even just two or three, can be enough to account for most of the missing multiple contributions.

Effects of term truncation on accuracy

As discussed in the previous section, the whole series up to the n -th term of expression 10 accounts for all of the missing n -th order multiples coming from the part of the free surface confined within the acquisition gap. It also includes some part of all multiples of $n+1$ -th order and above, although for near-offset gaps these contributions may be small. Figure 3 shows this with shot-gather representations of the different terms in expression 10 for the synthetic dataset shown in Figure 2a. We can see that the majority of these terms is composed of the n -th order surface multiple by comparing the wavefront shapes and traveltimes of panels 3b and 3c to the true primaries shown in Figure 1c. The arrows in panel 3b shows two apexes of the missing contribution of the second-order surface multiple, which we can see faintly in the $n=1$ term (Equation 8) and very prominently the $n=2$ term (Equation 9). The sum of these two terms, shown in panel 3d, perfectly accounts for the total missing contribution to the wavefront without relying on any data inside the gap.

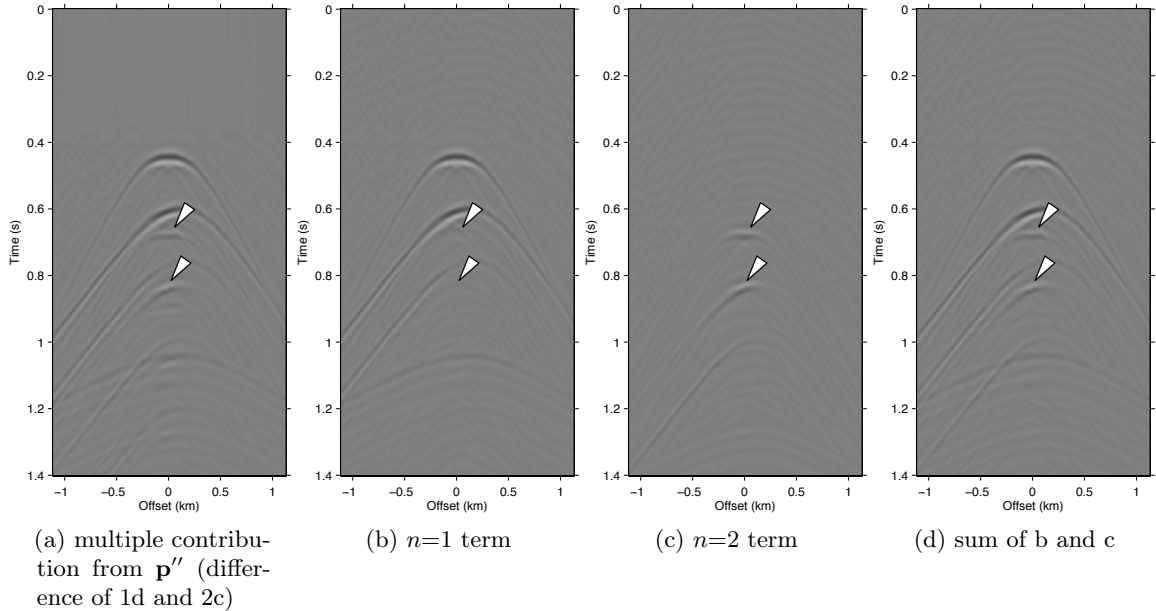


Figure 3: Shot-gathers of various multiple contribution terms in the auto-convolution based forward modeling operator $\widetilde{M}(\mathbf{G}, \mathbf{Q}; \mathbf{P}')$ (shown in expression 10) when applied to the synthetic dataset with 100 m missing near-offsets in \mathbf{p}'' shown in Figure 2. (a) shows the total error in multiple prediction due to missing near-offsets. Panels (b) and (c) are respectively the first two terms of $\widetilde{M}(\mathbf{G}, \mathbf{Q}; \mathbf{P}')$ involving auto-convolutions with \mathbf{g} (expressions 8 and 9). (d) shows the sum of panels (b) and (c). Comparing panels (a) and (d), it is evident that just the first two of the higher order terms of Equation 10 is enough to model most of the significant EPSI multiple prediction errors due to the missing data.

Figure 3 also visually demonstrates that the sum of just the first three terms of the series expansion 10 is enough to model most of the significant multiple contributions from the missing data \mathbf{p}'' . Panel 3d, which shows the sum of the $n=1$ and $n=2$ terms, explains all the missing contributions from \mathbf{p}'' (panel 3a) up to the second-order surface multiple, and also contains some kinematic imprints of all higher order ones. Together, these two terms were

able to bring the normalized prediction error (defined by $\|\mathbf{p}_m - \widetilde{M}(\mathbf{g}, \mathbf{q}; \mathbf{p}')\|/\|\mathbf{p}_m\|$, where \mathbf{p}_m is the true surface multiple) down from $\approx 22\%$ down to $\approx 3\%$. Figure 4 summarizes the effectiveness of these first two terms as a function of the nearest available offset.

From the general trend of the curves in Figure 4 we can see that, for a range of typical near-offset gap sizes, just including the $n=1$ term can reduce the normalized multiple modeling error of \widetilde{M} by more than one half (corresponding to a 6 dB improvement), and a further halving of the error is achieved by including up to the $n=2$ term. The model used here the same ones as used by van Groenestijn and Verschuur (2009), with a gently-varying water column depth around 200 m throughout the model, and a water-bottom reflectivity coefficient of ≈ 0.41 (seafloor p-wave velocity 2.0 km/s, density 1.8 g/cm³).

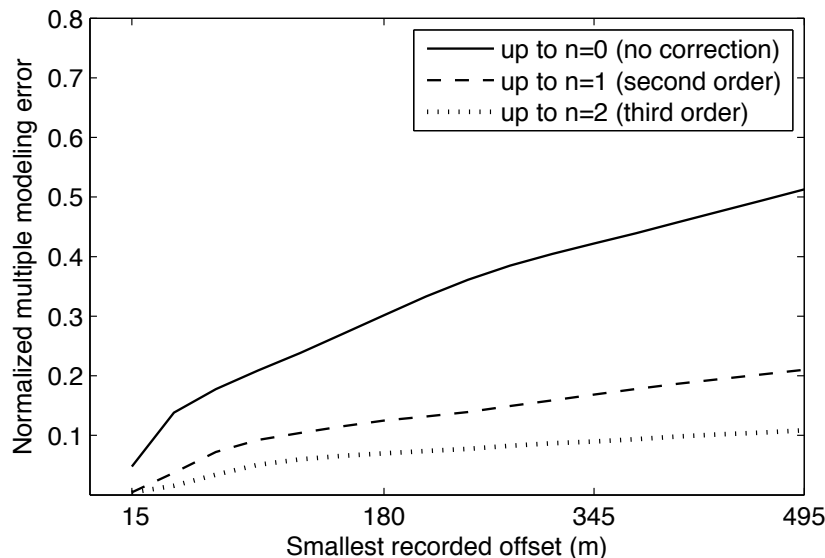


Figure 4: Multiple modeling error for the nonlinear series expansion modeling operator (expression 10) as a function of term truncation and the offset gap in the data. The different lines plot this normalized modeling error from truncating all terms from expression 10 with n higher than that indicated in the legend.

ALGORITHMIC CONSIDERATIONS

With the new augmented forward-modeling operator $\widetilde{M}(\mathbf{G}, \mathbf{Q}; \mathbf{P}')$ as defined in Equation 10 using data-matrix notation, we solve a new optimization problem by using it in place of the uncorrected modeling operator in the original REPSI problem (c.f. Equation 6):

$$\min_{\mathbf{g}, \mathbf{q}} \|\mathbf{g}\|_1 \quad \text{subject to} \quad f(\mathbf{g}, \mathbf{q}; \mathbf{p}') \leq \sigma, \quad (12)$$

$$f(\mathbf{g}, \mathbf{q}; \mathbf{p}') := \|\mathbf{p}' - \widetilde{M}(\mathbf{g}, \mathbf{q}; \mathbf{p}')\|_2.$$

The main challenge that we will address in this section is that, compared to the original REPSI problem with fully-sampled data, this new one loses the linear relationship of the modeling function with \mathbf{g} . In Lin and Herrmann (2013), we relied on this bilinear property of $M(\mathbf{g}, \mathbf{q}; \mathbf{p}')$ in terms of \mathbf{g} and \mathbf{q} to solve problem (4) using an alternating coordinate-descent

method. Below we first briefly recount how the original REPSI problem is solved, followed by the proposal of two possible ways to modify our solution strategy to account for the additional nonlinearity due to the existence of higher-order terms in $\underline{M}(\mathbf{G}, \mathbf{Q}; \mathbf{P}')$.

The existing approach to solving the original REPSI problem (4) is an alternating update of \mathbf{g} and \mathbf{q} . Each new iteration of \mathbf{g} is formed by a ℓ_1 -norm constrained problem (called a Lasso problem, Tibshirani, 1996) to give a sparse wavefront, which has a broadband, impulsive appearance in time. We then use it to estimate the wavelet \mathbf{q} by matching the impulsive wavefronts against the data. The number of sparse wavefronts allowed into \mathbf{g} is kept small at the early iterations to capture only the most significant primary events, quickly giving us the best estimate for the wavelet \mathbf{q} . Later iterations will then gradually increase the number of events allowed to better fit the remaining data residual.

This relaxation of sparsity constraint happens in a theoretically rigorous way for REPSI. The sparsity of \mathbf{g} is directly influenced by the ℓ_1 -norm constraint value of the Lasso problems. The initial ℓ_1 -norm constraints start at a small value that is designed to be an underestimate of the true minimum. It is then gradually increased throughout the iterations in a way such that it does not exceed the ℓ_1 -norm of the true solution, yet approaches it quickly. This scheme is enabled by the convex properties of the Pareto trade-off curve for Lasso problems (van den Berg and Friedlander, 2008; Hennenfent et al., 2008; Daubechies et al., 2008).

We now roughly outline the main algorithm as described in Lin and Herrmann (2013), using the notation of this paper and omitting some initialization details. We first set our solution candidates $\tilde{\mathbf{p}}_0$ and $\tilde{\mathbf{q}}_0$ to zero vectors, and the inversion residue (prediction error) \mathbf{e} to \mathbf{p} . Then we iterate over the following steps with iteration counter k .

1. Select a new ℓ_1 norm constraint $\tau_k > \tau_{k-1}$, which is evaluated through a closed-form expression from the current residue \mathbf{e}_k , the target misfit σ , and the ℓ_1 -norm of the current solution \mathbf{g}_k .
2. Obtain \mathbf{g}_{k+1} by solving through a spectral projected gradient method (van den Berg and Friedlander, 2008; Birgin et al., 2000), using the previous solution \mathbf{g}_k as the initial guess

$$\min_{\mathbf{g}} \|\mathbf{p} - M(\mathbf{g}, \mathbf{q}_k; \mathbf{p})\|_2 \text{ subject to } \|\mathbf{g}\|_1 < \tau_k. \quad (13)$$

3. Do an exact line-search scaling on \mathbf{g}_{k+1} using the multiple term in $M(\mathbf{g}, \mathbf{q}_k; \mathbf{p})$ on \mathbf{p} to minimize the effect of the ℓ_1 constrain on the amplitude of \mathbf{g}_{k+1} .
4. Obtain \mathbf{q}_{k+1} by solving a matched-filtering problem

$$\min_{\mathbf{q}} \|\mathbf{p} - M(\mathbf{g}_{k+1}, \mathbf{q}; \mathbf{p})\|_2$$

5. Form a new residue $\mathbf{e}_{k+1} = \mathbf{p} - M(\mathbf{g}_{k+1}, \mathbf{q}_{k+1}; \mathbf{p})$ and check for convergence conditions.
6. (If explicitly reconstructing data as in Problem 6) Solve for the missing part of the data \mathbf{p}''_{k+1} , using the previous solution \mathbf{p}''_k as the initial guess

$$\min_{\mathbf{p}''} \|\mathbf{p}' + \mathbf{p}'' - M(\mathbf{g}, \mathbf{q}_k; \mathbf{p}' + \mathbf{p}'')\|_2$$

and form new data by $\mathbf{p} = \mathbf{p}' + \mathbf{p}''_{k+1}$.

For steps 1 and 2 we depend on the notion of a smooth, monotonic Pareto curve (sometimes known as the L-curve) for problem 4 to give us sensible ℓ_1 -norm constraint parameters τ_k . The smooth and monotonic property is generally no longer valid when the forward system is nonlinear in \mathbf{g} , so below we will discuss two possible approaches to work around this issue, and propose how it can be applied to the above strategy to derive suitable methods for solving 12.

Modified Gauss-Newton approach

Several existing works on regularized inversion of auto-convolution functions rely on either the Gauss-Newton method or more generally the Levenberg-Marquardt method (Fleischer et al., 1999; Fleischer and Hofmann, 1996). In the same vein, we introduce an approach inspired by Li et al. (2012), which gives heuristically sparse solutions to the new nonlinear REPSI problem (12) using a modified Gauss-Newton method.

The crux of this method is to always ensure that the Green's function updates $\Delta\mathbf{g}$ are as sparse as possible for our a given amount of decrease in the objective. This is achieved most effectively by taking as updates the solution to a Lasso problem formed around the Jacobian \mathbf{J}_k of \widetilde{M} evaluated at $(\mathbf{g}_k, \mathbf{q}_k)$:

$$\Delta\mathbf{g}_{k+1} = \underset{\mathbf{g}}{\operatorname{argmin}} \|\mathbf{e}_k - \mathbf{J}_k\mathbf{g}\|_2 \quad \text{s.t.} \quad \|\mathbf{g}\|_1 \leq \tau_k, \quad (14)$$

where \mathbf{e}_k is the current nonlinear residue $\mathbf{p} - \widetilde{M}(\mathbf{g}_k, \mathbf{q}_k)$. The ℓ_1 -norm constraint τ_k in this case is again obtained deterministically by the optimization parameters of problem 14, using the same expressions used in step 1. However, since we are explicitly solving for updates $\Delta\mathbf{g}$ instead of \mathbf{g} itself, the previous solution \mathbf{g}_k (and hence its ℓ_1 -norm) for this linearized problem is always zero. Each update formed this way will thus tend to have a much smaller ℓ_1 -norm constraint τ_k compared to the final ℓ_1 -norm of the true solution, but the sum of all the updates may have a final norm that exceeds it. The assumption here is that our series of sparse updates will actually sum into a sparse final solution, which will depend on the Jacobians evolving slow enough over the iterations, so that all updates $\Delta\mathbf{g}$ kinematically all consists of very similar wavefront sets. In practice, we find that this assumption is rarely a problem if the acquisition gaps are not too large, and that we indeed usually end up with sparse solutions for \mathbf{g} .

The action of the Jacobian \mathbf{J}_k is straightforward to obtain, but actually computing it will pose a challenge. When we explicitly write it out using the data-matrix notation, it becomes evident that the computational effort required escalates very quickly with the order of the series expansion:

$$\begin{aligned} \mathbf{J}_{k\omega}\mathbf{G} = & \mathbf{K}[(\mathbf{G}\mathbf{Q}_k + \mathbf{G}\mathbf{R}\mathbf{P}') \\ & + \mathbf{G}_k\mathbf{R}\mathbf{K}_c(\mathbf{G}\mathbf{Q}_k + \mathbf{G}\mathbf{R}\mathbf{P}') \\ & + \mathbf{G}\mathbf{R}\mathbf{K}_c(\mathbf{G}_k\mathbf{Q}_k + \mathbf{G}_k\mathbf{R}\mathbf{P}') \\ & + \mathbf{G}_k\mathbf{R}\mathbf{K}_c(\mathbf{G}_k\mathbf{R}\mathbf{K}_c(\mathbf{G}\mathbf{Q}_k + \mathbf{G}\mathbf{R}\mathbf{P}')) \\ & + \mathbf{G}\mathbf{R}\mathbf{K}_c(\mathbf{G}_k\mathbf{R}\mathbf{K}_c(\mathbf{G}_k\mathbf{R}\mathbf{K}_c(\mathbf{G}_k\mathbf{Q}_k + \mathbf{G}_k\mathbf{R}\mathbf{P}')) \\ & + \mathbf{G}_k\mathbf{R}\mathbf{K}_c(\mathbf{G}\mathbf{R}\mathbf{K}_c(\mathbf{G}_k\mathbf{Q}_k + \mathbf{G}_k\mathbf{R}\mathbf{P}')) \\ & + \dots]. \end{aligned} \quad (15)$$

The above expression is only written up to the $n=2$ nonlinear term (expression 9) of \widetilde{M} . Comparing all the terms here to the original nonlinear expression 10, we can see that due to partial derivatives, each of the $\mathcal{O}(\mathbf{G}^n)$ terms expand into n linearized terms in the Jacobian. In aggregate, this effectively introduces a quadratic complexity in the number of nonlinear terms when computing the action. Although some reuse is possible in the calculation of these terms, this approach may nonetheless present a prohibitive amount of computation overhead compared to the original problem. In the next subsection we will discuss an alternative scheme that forgoes using the Jacobian and will instead linearize the problem in some other way.

To summarize the modified Gauss-Newton approach to solving problem 12, we essentially modify steps 2 and 3 of the original algorithm above. In step 2 of each iteration we explicitly compute the updates $\mathbf{g}_{k+1} = \mathbf{g}_k + \Delta\mathbf{g}_{k+1}$ with $\Delta\mathbf{g}_{k+1}$ given by solving problem 14. In step 3 we then perform the exact line-search on the update $\Delta\mathbf{g}_{k+1}$ itself against the current residual, much like in the original algorithm, this mitigates the amplitude loss from the ℓ_1 -norm regularization while keeping its sparsifying effect.

Relinearization by substitution

As shown in expression 15, having to compute the action of the Jacobian of the operator \widetilde{M} may be unacceptably costly, considering that each wavefield multiplication in these expressions is computationally equivalent to one full SRME multiple prediction step. Since our non-linearity only comes from essentially having quadratic and higher powers of \mathbf{G} , a possible alternative to avoid dealing with the Jacobian is to simply substitute the previous iteration \mathbf{g}_k as a constant in the $(\mathbf{GRK}_c)^n$ part of Equation 10. This returns us to the paradigm of inverting a linear operator in \mathbf{G} at step 2:

$$\begin{aligned} \widetilde{M}_{k\omega}(\mathbf{G}) = & \mathbf{K}[(\mathbf{GQ}_k + \mathbf{GRP}') \\ & + \mathbf{G}_k\mathbf{RK}_c(\mathbf{GQ}_k + \mathbf{GRP}') \\ & + \mathbf{G}_k\mathbf{RK}_c(\mathbf{G}_k\mathbf{RK}_c(\mathbf{GQ}_k + \mathbf{GRP}')) \\ & + \dots]. \end{aligned} \quad (16)$$

Similar to Gauss-Newton, the approach also relinearizes the problem at step 2, except that we do not have to find each update explicitly. Instead, we simply replicate step 2 of the original REPSI algorithm (solving the Lasso problem in Equation 13 starting from the previous solution \mathbf{g}_k) using the approximate but linear modeling operator of expression 16. In effect, we now solve the following in step 2:

$$\min_{\mathbf{g}} \|\mathbf{p} - \widetilde{M}_k(\mathbf{g})\|_2 \text{ subject to } \|\mathbf{g}\|_1 < \tau_k. \quad (17)$$

Theoretically speaking, we have to give up some nice properties of Gauss-Newton with this method, such as possible quadratic convergence and guarantees of reaching a stationary point. But in practice we find this method, which we call “relinearization” for the purpose of this paper, to converge similarly to the Gauss-Newton approach while also having much smaller per-iteration compute costs. In fact, the computational overhead here is marginal compared to the original algorithm, as we explain in the Discussions section. However, to achieve this result we find that it is necessary to manually precondition the first few gradients, where we have zero contribution coming from the missing primary part of \mathbf{p}'' .

Improving convergence by gradient trace weighting

If we explicitly write out a (negative) gradient step of the Lasso problem in step 2 (where the residue is at \mathbf{e}_k) using the approximate operator in expression 16 up to the $n=1$ term, we obtain, assuming $\mathbf{R} = -\mathbf{I}$ and using H for Hermitian adjoint,

$$(\mathbf{K}\mathbf{E}_k)\mathbf{Q}_k^H - (\mathbf{K}\mathbf{E}_k)\mathbf{P}'^H - \mathbf{K}_c\mathbf{G}_k^H(\mathbf{K}\mathbf{E}_k)\mathbf{Q}_k^H + \mathbf{K}_c\mathbf{G}_k^H(\mathbf{K}\mathbf{E}_k)\mathbf{P}'^H.$$

Recall that \mathbf{E}_k is the data-matrix representation of the current residue error \mathbf{e}_k . From this we can see that \mathbf{g} is not being updated equally inside the acquisition mask compared to the outside of it. In fact, the first term is actually zero inside the acquisition mask (i.e., $\mathbf{K}_c\mathbf{K} = 0$ so $\mathbf{K}_c(\mathbf{K}\mathbf{E}_k)\mathbf{Q}_k^H = \mathbf{0}$), due to the residue being zero inside the mask, and the fact that the source wavelet applies a purely time domain convolution which does not spatially “spread” the wavefield. Since this term contains all the contribution from primary wavefields, which tends to be the strongest part of the wavefield, there is immediately a large loss of amplitude for the updates inside the mask. The job of filling this void then naturally falls to the subsequent terms, which all involve some sort of wavefield convolution and can potentially generate events inside the mask. However, in the first few gradients where \mathbf{G}_k is small in amplitude (due to the ℓ_1 -norm constraints), the third term and above are also unhelpfully close to zero. This creates a convergence “lag” where traces of \mathbf{g} outside of the mask must already be well reconstructed in order for the inside of the mask to begin updating, which is problematic because the outside traces are susceptible to overfit on multiples that should ultimately be explained by contributions coming from the inside of the mask.

We can mitigate the inherent imbalance on the early gradients through a simple weighting to all traces inside the acquisition gap. This can be seen as applying a diagonal right preconditioner to the linearized modeling operator. For the Gauss-Newton approach, this kind of scaling is automatically handled by the implicit Hessian, but for the relinearization method we must form and apply this preconditioner manually. An averaged, trace-independent version can be obtained by finding an exact line-search for a scaling α of $\mathbf{K}_c\mathbf{g}_{k+1}$ which maximally reduces the residual left over from the contributions of $\mathbf{K}\mathbf{g}_{k+1}$ under the action of the next forward operator \widetilde{M}_{k+1} (not that it is linearized around the newly obtained \mathbf{g}_{k+1}). In other words, we seek the scalar α , which solves the following problem

$$\min_{\alpha} \|(\mathbf{p}' - \mathbf{K}\widetilde{M}_{k+1}(\mathbf{K}\mathbf{g}_{k+1})) - \alpha\mathbf{K}\widetilde{M}_{k+1}(\mathbf{K}_c\mathbf{g}_{k+1})\|_2^2.$$

The computation of this scaling can be effectively free in the process of computing the residual of iteration $k+1$, by producing the new predicted wavefield due to the two parts of \mathbf{g}_{k+1} separately:

$$\begin{aligned} \mathbf{e}_{\text{outside}} &= \mathbf{p}' - \mathbf{K}\widetilde{M}_{k+1}(\mathbf{K}\mathbf{g}_{k+1}) \\ \tilde{\mathbf{p}}_{\text{inside}} &= \mathbf{K}\widetilde{M}_{k+1}(\mathbf{K}_c\mathbf{g}_{k+1}) \\ \alpha_{k+1} &= \tilde{\mathbf{p}}_{\text{inside}}^T \mathbf{e}_{\text{outside}} / \|\tilde{\mathbf{p}}_{\text{inside}}\|_2^2. \end{aligned} \tag{18}$$

Note that all the computational cost for this scaling is in the last line, which only involves the vector inner products between $\tilde{\mathbf{p}}_{\text{inside}}$ and $\mathbf{e}_{\text{outside}}$, as well as ℓ_2 norms. All the terms needed to form these two wavefields are already computed once we have the final residue for this iteration. As an aside, this scaling should be calculated after the line-search over all of \mathbf{g} described in step 3, to minimize the impact of the accuracy of the current wavelet estimate

(simply ignoring the primary $\mathbf{G}\mathbf{Q}$ term when forming $\tilde{\mathbf{p}}_{\text{inside}}$ and $\mathbf{e}_{\text{outside}}$ will have a similar effect).

After obtaining α , we use it as a right-preconditioner $\alpha\mathbf{K}_{\mathbf{c}}$ in step 2 of the next iteration, where we now solve

$$\min_{\tilde{\mathbf{g}}} \|\mathbf{p} - \tilde{M}_k(\alpha\mathbf{K}_{\mathbf{c}}\tilde{\mathbf{g}})\|_2 \text{ subject to } \|\tilde{\mathbf{g}}\|_1 < \tau_k. \quad (19)$$

This will effectively scale all gradient traces inside the stencil by α . After the inversion is complete, we obtain the new estimate of the Green’s function by $\mathbf{g}_{k+1} = \alpha\mathbf{K}_{\mathbf{c}}\tilde{\mathbf{g}}$, which removes the imprint of the right preconditioner. As the iteration count increases and \mathbf{g} becomes more accurate, we should see α trend towards 1.

Figure 5 demonstrates the effectiveness of this method. Panel 5a shows the first gradient update for \mathbf{g} (assuming we use the true wavelet \mathbf{q}) we would have obtained from a full-sampled data (shown in Figure 1a). Panel 5b shows the same gradient step that we would obtain using data with missing near-offsets (Figure 2a). We can see that this update contains mostly correct kinematic information about the wavefield, but is much weaker in amplitude for traces inside the near-offset gap. After applying the preconditioner $\alpha\mathbf{K}_{\mathbf{c}}$, with α computed using the inaccurate gradient in Panel 5b as the approximate solution, we see in Panel 5c that this amplitude imbalance is effectively mitigated.

To summarize, in the “relinearization” approach we essentially replace the steps 2 and 3 of the original algorithm with the following modified steps:

2. Solve the modified Lasso problem (Equation 19) instead of the original one (Equation 13). This replaces the wavefield modeling function with the approximately linearized version of \tilde{M} , defined in expression 16, and adds a right preconditioner $\alpha_k\mathbf{K}_{\mathbf{c}}$. Once solution is obtained, apply $\alpha_k\mathbf{K}_{\mathbf{c}}$ to invert the preconditioner and get \mathbf{g}_{k+1} .
3. Compute the scaling for all of \mathbf{g}_{k+1} as in the original. After applying the scaling, compute the next preconditioner scaling α_{k+1} using expression 18.

NUMERICAL EXAMPLES

In this section we demonstrate the efficacy of our proposed methods, focusing on the missing near-offset problem. We compare the results given by both the Gauss-Newton and the relinearization approach, using up to the $n = 2$ term (expression 9) in the augmented nonlinear modeling operator $\tilde{M}(\mathbf{g}, \mathbf{q}; \mathbf{p}')$. We also compare our results to the explicit data reconstruction approach (problem 6) based on the method suggested by van Groenestijn and Verschuur (2009), as well as a parabolic Radon-domain interpolation done as a preprocessing step. We first show results using the synthetic data shown in 1a, then move on to a field data example from the North Sea. Although we only show 2D acquisition examples, note that our method makes no assumptions on the coordinate system used for the data matrices, so it is trivially extendable to a 3D survey geometry.

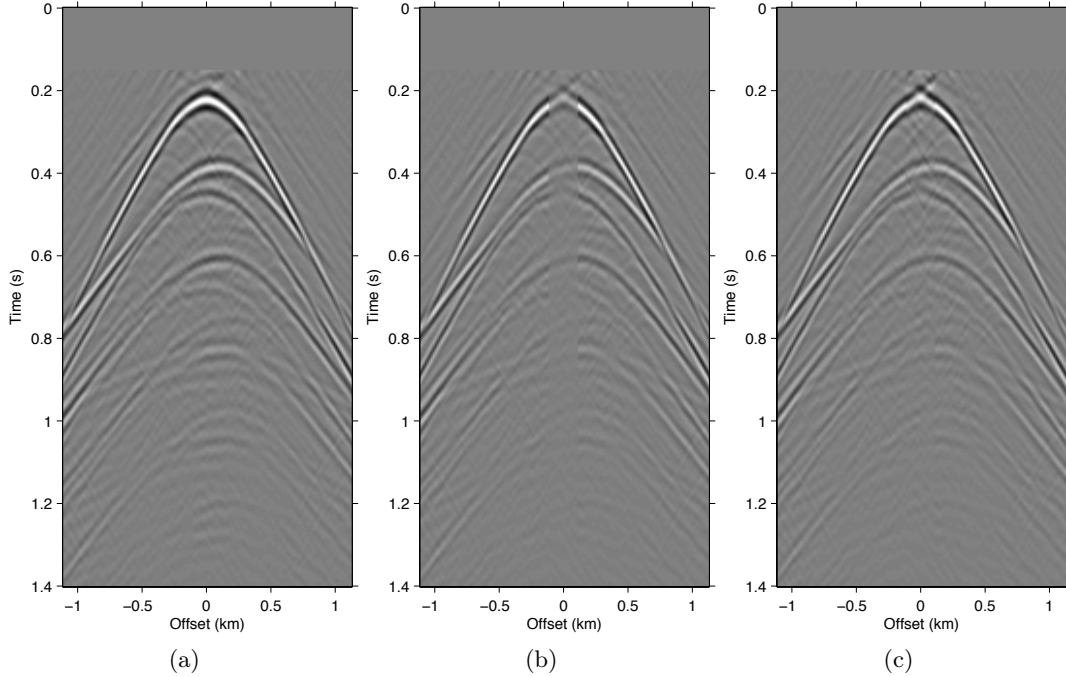


Figure 5: Early gradient updates for \mathbf{g} need to be preconditioned, otherwise converge will be slow inside the acquisition mask due to amplitude imbalances. **(a)** the exact first gradient for \mathbf{g} from fully sampled data, given exact \mathbf{q} . **(b)** the first gradient from data with missing near-offsets (shown in Figure 2a) using the relinearized forward model \tilde{M}_ω . Inside the near-offset gap, the events are correct but greatly diminished in amplitude. **(c)** the gradient shown in (b) after applying the preconditioning scaling described. Our average trace-independent scaling is effective in balancing the amplitude of the updates inside the near-offset gap to be more comparable to the outside.

Synthetic data example

In this section we focus on how effectively the proposed methods in this paper handle different sizes of near-offset gaps, and how it fares compared to explicitly inverting for the missing data \mathbf{p}'' . We use the same 5 km fixed-spread synthetic data used in all the figures so far in the paper, a shot gather of which is shown in Figure 1a. For reference, the corresponding true primary model is shown in Figure 1c. This model appears in almost all existing EPSI literature today and serves as an effective benchmark.

Figure 6 shows a comparison of methods for a 45 m near-offset gap (in both positive and negative offsets). This is an extremely mild offset gap and typically poses no major problems for SRME workflows. In this case, the data reconstruction (Panel 6b), the Gauss-Newton (Panels 6c and 6d), and the relinearization strategies (Panel 6e) all performed very similarly and gave good results without needing to interpolate the near-offsets in preprocessing. All the results plotted here is the direct primary estimation (the final \mathbf{g} from the inversion convolved with the final \mathbf{q} from the same inversion), as the effects of reconstruction inside the offset gap is clearer this way. In practical applications, usually a multiple model would be formed with

the solution \mathbf{g} instead, so it can be subtracted from the data under perhaps more scrutiny.

Comparing the Gauss-Newton results allows us to examine the effects of term truncation in the nonlinear modeling operator \widetilde{M} . Panel 6c shows the Gauss-Newton solution using up to the $n=1$ term (second-order in \mathbf{g}) in \widetilde{M} . As discussed in the theoretical section, this term alone cannot fully account for the second-order surface multiples, and thus some part of the first order multiple is put into \mathbf{g} by the algorithm to make up some of the difference. This is especially apparent inside the offset gap, where \mathbf{g} is less constrained by the observed data.

Panel 6d shows the solution we can obtain by just using one more term in \widetilde{M} , up to the $n=2$ term which is third-order in \mathbf{g} . We see that the imprint from the first-order multiple is much less severe in this result. This result, notably, is already very close to the solution from fully sampled data as shown in Figure 1c, even though only two additional terms from the series expansion is used. The results of the relinearization method (also using up to the $n=2$ term) shown in Panel 6e is almost identical to the Gauss-Newton solution, despite being much faster to compute. In fact it is arguably cleaner, perhaps due to less overfit of the remaining multiple prediction errors from the lack of higher order terms.

Figure 7 shows the opposite end of the near-offset gap severity scale with a 225 m offset gap. As we can see in panel 7a, the majority of the near-offset wavefront curvature information is missing, which poses a serious challenge for standard workflow tools to accurately reconstruct near-offset traces. In this case the explicit data reconstruction (Figure 7b) fails to remove some of the later surface multiples, such as those after 0.6 s. The augmented nonlinear modeling operator approach introduced in this paper all arguably fared better in comparison. The Gauss-Newton approach up to the second-order term (Figure 7c) is better at recovering the wavefronts inside offset gap, even building up some of the diffraction events, but also similarly fails to remove some of the later surface multiples. The multiple removal is improved greatly by including up to the third-order term (Figure 7d), although some still remain, again possibly due to overfit. The relinearization method remains the overall winner in this case (Panel 7e), producing the cleanest primary estimation that is nearly identical to the reference primary outside of the near-offset gap, while also computing it much faster than the Gauss-Newton method.

Field data example

We now demonstrate the effectiveness of our proposed methods using a North Sea shallow-water marine 2D seismic line with a 100 m near-offset gap. Pre-processing has been applied to convert the collected dual-sensor data to an upgoing pressure wavefield using the method described in Cambois et al. (2009). Data regularization and interpolation were carried out as part of the preprocessing, so in this example we will compare our method with the existing parabolic Radon interpolated near-offsets. A 3D-to-2D correction factor \sqrt{t} has also been applied after data regularization.

Figure 8 show the moveout-corrected stacks of the field data experiments. Panel 8a shows the data itself. All near-offsets are excluded from the stack to keep comparisons consistent. Panel 8b shows the results obtain from using the original REPSI on the regularized data where the near-offset is pre-interpolation using parabolic Radon. Panel 8c show results obtained by inverting the third order (keeping up to $n=2$ terms) nonlinear operator $\widetilde{M}(\mathbf{g}, \mathbf{q}; \mathbf{p}')$ using the relinearization method, while panel 8d shows the same using the Gauss-Newton method.

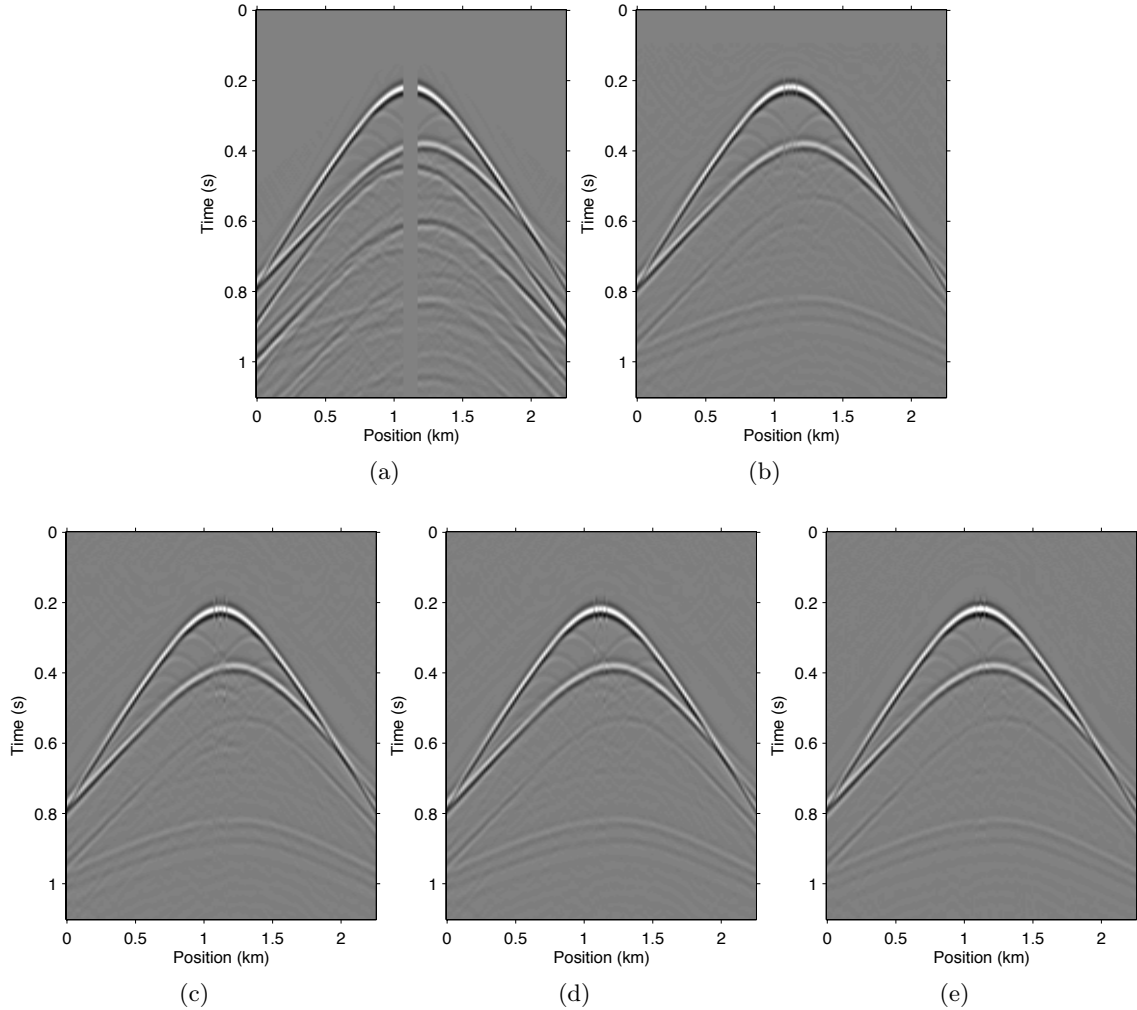


Figure 6: Direct primary estimation results from a 5 km fixed-spread synthetic data with a nearest recorded offset of 45 m, for which a shot-gather at 2.5 km is shown in **(a)** up to 1 km offset. **(b)** the estimated primary using explicit data reconstruction (solving problem 6). **(c)** estimated primary using the Gauss-Newton method described in this paper, using up to the $n=1$ term (second-order in \mathbf{g}) in \widetilde{M} . **(d)** the results obtained with the same method as (c), but including up to the $n=2$ term (third-order in \mathbf{g}). **(e)** estimated primary using the relinearization strategy, including up to the $n=2$ term. All the methods produce similar quality results for this small offset gap, although we clearly see the benefit of using the $n=2$ term in the nonlinear forward model when comparing (c) and (d). The relinearization strategy produced slightly better results over the Gauss-Newton methods, despite also being computationally faster.

Our methods are successful at improving on the results obtained by using unmodified EPSI on pre-interpolation data. In shallow water data most pre-interpolation methods are known to be inaccurate, often under-estimating near-offset amplitudes and incorrectly assessing wavefront curvature at the apex. Comparing panel 8b with panels 8c and 8d shows that our methods exhibit a much improved ability to remove large water-bottom

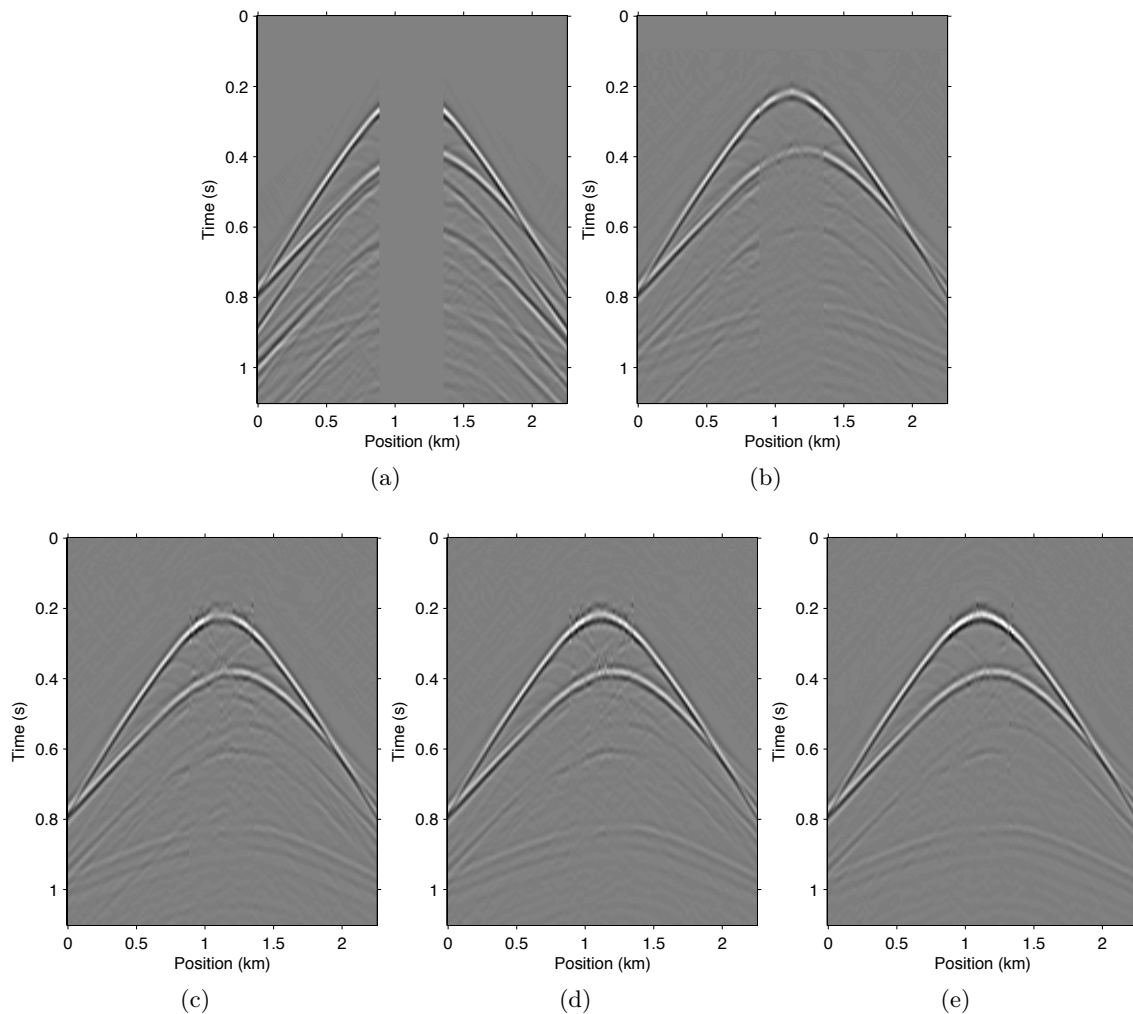


Figure 7: Direct primary estimation results from a 5 km fixed-spread synthetic data with a nearest recorded offset of 225 m, for which a shot-gather is shown in **(a)**. The panels in this figure show the same strategies as the ones shown in Figure 6. **(b)** the estimated primary using explicit data reconstruction (solving problem 6). **(c)** estimated primary using the Gauss-Newton method described in this paper, using up to the $n=1$ term (second-order in \mathbf{g}) in \tilde{M} . **(d)** the results obtained with the same method as (c), but including up to the $n=2$ term (third-order in \mathbf{g}). **(e)** estimated primary using the relinearization strategy, including up to the $n=2$ term. The explicit data reconstruction result in (b) fails to reject some of the later surface multiples, while the methods introduced in the paper manages to do a better job. As in Figure 6, we readily see the benefit of including up to the $n=2$ term, and also that the relinearization strategy manages to produce the best cleanest result despite being faster to compute than Gauss-Newton.

multiples that is evident at 0.25 s, 0.55s, and 0.8 s of the stack. This is especially evident by looking at the multiples removed from the pre-interpolation result (panel 8e) and comparing it to the multiples removed using the relinearization result (panel 8f). By discarding the pre-interpolated near-offset and using the augmented modeling operator \widetilde{M} to account for its effects, we are able to get a much better defined multiple model with more accurate amplitude. Pre-interpolation also led to some ringing kinematic artifacts in the multiple model near the large diffraction events at 3 km and 1.0 s, which is not present in our methods.

Panel 8g shows a difference plot between inverting the second order (keeping up to $n=2$ terms) and the third order forward operator using relinearization. We see that the impact of using the additional term is already very small, and mainly consists of the imprint of second-order surface multiples. The comparison between relinearization and Gauss-Newton is more interesting (panel 8h). It is not very clear from the difference plot which method produced a better multiple model, although when carefully examining the primary stacks we see that the amplitude of the first water-bottom multiple (at 0.25 s) is slightly over-estimated for Gauss-Newton. This is possibly due to the over-fitting issues discussed earlier in the synthetic example section. Nevertheless, the relinearization method once again provided a very good result at a fraction of the compute time required for Gauss-Newton.

DISCUSSION

The synthetic results shown in the previous section demonstrated that our proposed scattering-based correction to the multiple modeling with incomplete data can outperform the existing explicit data reconstruction approach used in literature, especially when large gaps are encountered in the near-offset. As far as we are aware, our proposed method is also the first practical formulation where a scattering series is rigorously used in a variational problem for primary estimation, with all its numerical approximations explicitly stated and explored. The real data example also demonstrates a significant improvement over using parabolic Radon interpolation for the near-offset. These improvements are tangible regardless of the algorithmic approach used to invert the augmented and non-linear modeling operator.

Perhaps more interestingly, we find that the Gauss-Newton method did not converge appreciably faster than the conceptually simpler and computationally cheaper relinearization method, despite its more solid theoretical convergence guarantees. This might be because the number of total gradient updates typically involved in REPSI is already fairly small (typically around 50), since the projected gradient method that we employ to solve the Lasso problems forms an implicit Hessian of the problem by using spectral step lengths (Birgin et al., 2000). The Gauss-Newton Hessian might not have been able to improve much more on this scaling information. On the other hand, we find to our surprise that the relinearization strategy often produced better results. One possible explanation is that in some of the later iterations, the Gauss-Newton method is more prone to over-fitting \mathbf{g} , particularly inside the trace mask where it is less sensitive to data misfit, to account for some of the inherent modeling error due to term truncation.

Although we have only shown near-offset acquisition gap effects in this paper, the derivations in this paper is valid for missing data located anywhere on the acquisition geometry. We impose no special structure on our diagonal masking operator \mathbf{K} , so in theory the missing traces can be anywhere. Near-offset traces are well-understood to be responsible

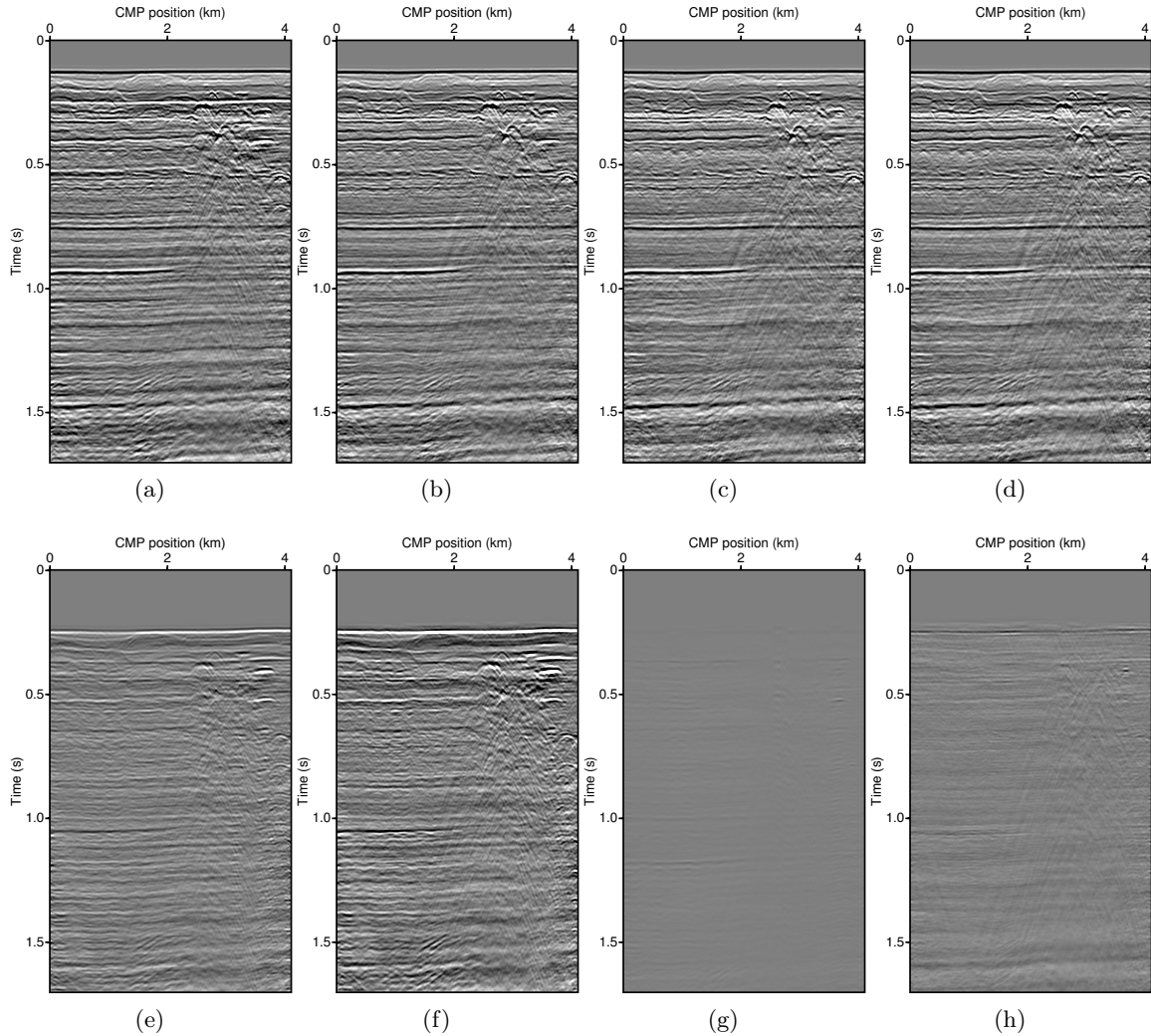


Figure 8: Moveout-corrected stacks (near-offsets excluded) of the North Sea field data with 100 m near-offset gap after conservative multiple removal, with the multiple model generated from the Green's function obtained using the methods introduced in this paper. **(a)** the field data. **(b)** the results using parabolic Radon near-offset interpolation and the unmodified REPSI algorithm. **(c)** the results obtained by discarding the near-offset traces and using the relinearization method with up to the $n=2$ term in \widetilde{M} . **(d)** the same as (c) but using the Gauss-Newton method. **(e)** the final multiple model from the Radon interpolated data result shown in (b). **(f)** the final multiple model from the relinearization method result shown in (c). **(g)** difference plot between solving the relinearization problem using up to the $n=1$ term versus up to the $n=2$ term. **(h)** difference plot between the solutions obtain from the relinearization method and the Gauss-Newton method.

for the majority of the constructively interfering part of the multiple contribution gather, so we conveniently use it as a worst-case scenario where the presence of \mathbf{K} can introduce the largest errors.

In theory, our scattering-based method to account for missing data works only when we can trace ray paths that reaches the inside of the masked location from the observed data positions, using the estimated Green’s function. This would imply that missing traces near the edge of the acquisition grid would be less effectively mitigated, which we do observe in practice. However, we find that the resulting edge effects tend to be overshadowed in magnitude by what would typically be produced in imaging algorithms.

Another important exception to the generality of our method is that the methods in this paper cannot deal with regular undersampling, even if it strictly counts as missing data in our framework. This is mainly because the mechanisms employed here also depends on wavefield interferometry, which is fundamentally erroneous when strong, coherent spatial aliasing is present in the data.

Finally, we note that one additional advantage of our approach over explicit data reconstruction is the possibility of exploiting any existing regularization schemes for \mathbf{g} , such as the curvelet-based estimation shown in Lin and Herrmann (2013), although a thorough investigation is outside the scope of this paper. This has the potential to improve the accuracy of the recovery inside the gap using constraints such as wavefront continuity and lateral smoothness. While it is also possible to regularize explicitly reconstructed data, it is much easier in our approach to take advantages of regularizations that are already in place.

Computation costs

In each of the scattering terms \mathbf{GRK}_c in expression 10, the wavefield convolution is done after stencilling out the traces that lie inside unobserved locations. Therefore the computation cost of applying each of the higher-order terms will depend on the ratio of unobserved to observed locations, and can be much smaller than the convolution cost for the whole wavefield if we are only interested in near-offsets. Since we only have to apply a single \mathbf{GRK}_c to the current term to obtain the subsequent subsequent term, the marginal cost of using higher-order terms is small, and the aggregate cost of computing expression 10 will only be a fractional overhead to the uncorrected modeling operator.

The story is different when applying the adjoint of the relinearized operator (expression 16) to form the gradient for \mathbf{g} in the relinearization step. In this case each of the scattering terms add a $\mathbf{K}_c \mathbf{R}^H \mathbf{G}^H$ to the previous-order term. Although we still involve the stencil \mathbf{K}_c before computing the next term, there is a special case when computing the $n=1$ term, where we do not have a stencil between \mathbf{G}^H and the residual wavefield, and instead need to compute one full wavefield convolution over the whole grid. The cost of the (linearized) adjoint is thus two times the cost compared to the uncorrected modeling operator, plus the same fractional overhead for higher order terms as in the forward.

However, note that the method in this paper avoids the cost of updating and store the unknown data \mathbf{p}'' as described in step 6 of the original algorithm. The gradients for \mathbf{p}'' also require a full wavefield convolution, so the overall cost of updating the unknown wavefield is actually absorbed into the cost of updating \mathbf{g} . In fact, if we only use the $n=1$ term, then the overall computational cost of the relinearization strategy is identical to the original

REPSI algorithm with explicit data reconstruction (assuming the same number of gradient updates is applied on both \mathbf{p}'' and \mathbf{g}), while being slightly more memory efficient due to not having to store and compare \mathbf{p}'' . As discussed in the algorithms section, the cost of using the Gauss-Newton approach is significantly higher due to the partial derivatives in the Jacobian. Depending on the number of terms used, we find that the Gauss-Newton approach adds a two-to-three time computational overhead compared to the relinearization approach.

CONCLUSIONS

We have presented a modification of the REPSI problem that accounts for large gaps in the acquisition grid (as a function of both source and receiver coordinates), such as the near-offsets, without having to explicitly interpolate or invert for the missing near-offset data. The main idea of this method is to modify the forward modeling operator to fully explain the multiples in the observed data as long as the correct surface-free Green’s function is obtained, even though the data operator itself is not completely sampled. This is achieved by augmenting the forward model for the observed wavefield with a truncated scattering series that approximately mimic the action of the free surface reflector within the acquisition gap. Inverting this modified operator can thus recover the primary Green’s function without involving the unobserved data in any way.

Part of our main contribution is demonstrating that just two terms in the scattering series is enough to reach a useable accuracy for the REPSI inversion problem. We demonstrated on both synthetic and real data that this level of approximation already leads to significant improvements over existing methods. The scattering terms involve the same wavefield convolution kernel as the original REPSI problem, so not much effort is required to implement its action. Furthermore, because the effective aperture of the multiple contribution for these scattering terms is limited to the inside of the acquisition gap, each additional term imposes only a small computational overhead compared to the original problem.

Unlike the original REPSI problem, the augmented modeling operator is no longer linear with respect to the primary Green’s function; each term in the scattering series involve higher powers of the primary Green’s function. We presented two possible modifications to the REPSI algorithm for dealing with this nonlinearity: a Gauss-Newton type approach (which involves a costly computation of the action of the Jacobian), and a straightforward relinearization approach that simply fixes the primary Green’s function used in higher-order scattering terms at its previous iteration value. A simple scaling preconditioner allows the relinearization approach to perform similarly to the Gauss-Newton approach while keeping the same computational complexity as the original REPSI algorithm with explicit missing-data inversion, making it a suitable candidate for large industry-scale multiple-removal problems.

ACKNOWLEDGEMENTS

We would like to acknowledge Gert-Jan van Groenestijn and Eric Verschuur for discussion with led to our involvement in the EPSI problem. We also thank PGS for permission to use the field dataset. Our Lasso solver is derived from SPGL_1 , which is the work of Michael Friedlander and Ewout van den Berg. This work was financially supported in part by the CRD Grant DNOISE II (CDRP J 375142-08). This research was carried out as part of

the SINBAD II project with support from the following organizations: BG Group, BGP, Chevron, ConocoPhillips, CGG, DownUnder GeoSolutions, Hess, ION, Petrobras, PGS, Statoil, Total SA, Sub Salt Solutions, WesternGeco, and Woodside.

REFERENCES

- Ahmed, A., B. Recht, and J. Romberg, 2012, Blind Deconvolution using Convex Programming: 40.
- AlMatar, M. H., 2010, Estimation of Surface-free Data by Curvelet-domain Matched Filtering and Sparse Inversion: Master's thesis, University of British Columbia.
- Amundsen, L., 2001, Elimination of free-surface related multiples without need of the source wavelet: *Geophysics*, **66**, 327–341.
- Anstey, N. A., and P. Newman, 1966, Part I: The sectional auto-correlogram and Part II: The sectional retro-correlogram: *Geophysical Prospecting*, **14**, 389–426.
- Berkhout, A. J., and Y. H. Pao, 1982, Seismic Migration - Imaging of Acoustic Energy by Wave Field Extrapolation: *Journal of Applied Mechanics*, **49**, 682.
- Berkhout, A. J., and D. J. Verschuur, 1997, Estimation of multiple scattering by iterative inversion, Part I: Theoretical considerations: *Geophysics*, **62**, 1586–1595.
- Biggs, D. S. C., 1998, Accelerated iterative blind deconvolution: PhD thesis, University of Auckland, Auckland.
- Birgin, E. G., J. M. Martinez, and M. Raydan, 2000, Nonmonotone spectral projected gradient methods on convex sets: *SIAM Journal on Optimization*, **10**, 1196–1211.
- Bronstein, A., M. Bronstein, and M. Zibulevsky, 2005, Relative optimization for blind deconvolution: *IEEE Transactions on Signal Processing*, **53**, 2018–2026.
- Cambois, G., D. Carlson, C. Jones, M. Lesnes, W. Söllner, and H. Tabti, 2009, Dual-sensor streamer data: Calibration, acquisition QC and attenuation of seismic interferences and other noises: *SEG Technical Program Expanded Abstracts*, 142–146.
- Daubechies, I., M. Fornasier, and I. Loris, 2008, Accelerated projected gradient method for linear inverse problems with sparsity constraints: *Journal of Fourier Analysis and Applications*, **14**, 764–792.
- Fleischer, G., R. Gorenflo, and B. Hofmann, 1999, On the Autoconvolution Equation and Total Variation Constraints: *ZAMM*, **79**, 149–159.
- Fleischer, G., and B. Hofmann, 1996, On inversion rates for the autoconvolution equation: *Inverse Problems*, **12**, 419–435.
- Fokkema, J. T., and P. M. van den Berg, 1993, *Seismic applications of acoustic reciprocity*: Elsevier Science.
- Frijlink, M. O., R. G. van Borselen, and W. Söllner, 2011, The free surface assumption for marine data-driven demultiple methods: *Geophysical Prospecting*, **59**, 269–278.
- Hennenfent, G., L. Felon, and F. J. Herrmann, 2010, Nonequispaced curvelet transform for seismic data reconstruction: A sparsity-promoting approach: *Geophysics*, **75**, WB203–WB210.
- Hennenfent, G., E. van den Berg, M. P. Friedlander, and F. J. Herrmann, 2008, New insights into one-norm solvers from the Pareto curve: *Geophysics*, **73**, A23.
- Herrmann, F. J., and G. Hennenfent, 2008, Non-parametric seismic data recovery with curvelet frames: *Geophysical Journal International*, **173**, 233–248.
- Kabir, M. N., and D. Verschuur, 1995, Restoration of missing offsets by parabolic Radon transform1: *Geophysical Prospecting*, **43**, 347–368.

- Kreimer, N., A. Stanton, and M. D. Sacchi, 2013, Tensor completion based on nuclear norm minimization for 5D seismic data reconstruction: *Geophysics*, **78**, V273–V284.
- Kumar, R., C. D. Silva, O. Akalin, A. Y. Aravkin, H. Mansour, B. Recht, and F. J. Herrmann, 2015, Efficient matrix completion for seismic data reconstruction: *Geophysics*.
- Li, X., A. Y. Aravkin, T. van Leeuwen, and F. J. Herrmann, 2012, Fast randomized full-waveform inversion with compressive sensing: *Geophysics*, **77**, A13–A17.
- Lin, T. T. Y., and F. J. Herrmann, 2013, Robust estimation of primaries by sparse inversion via one-norm minimization: *Geophysics*, **78**, R133–R150.
- Ma, J., 2013, Three-dimensional irregular seismic data reconstruction via low-rank matrix completion: *Geophysics*, **78**, V181–V192.
- Majdański, M., C. Kostov, E. Kragh, I. Moore, M. Thompson, and J. Mispel, 2011, Attenuation of free-surface multiples by up/down deconvolution for marine towed-streamer data: *Geophysics*, **76**, V129–V138.
- Riley, D. C., and J. F. Claerbout, 1976, 2-D multiple reflections: *Geophysics*, **41**, 592–620.
- Tibshirani, R., 1996, Regression shrinkage and selection via the lasso: *Journal of the Royal Statistical Society Series B Methodological*, **58**, 267–288.
- van den Berg, E., and M. P. Friedlander, 2008, Probing the Pareto frontier for basis pursuit solutions: *SIAM Journal on Scientific Computing*, **31**, 890–912.
- van Groenestijn, G. J. A., and D. J. Verschuur, 2009, Estimating primaries by sparse inversion and application to near-offset data reconstruction: *Geophysics*, **74**, A23–A28.
- Verschuur, D. J., 1992, Adaptive surface-related multiple elimination: *Geophysics*, **57**, 1166.
- , 2006, Seismic multiple removal techniques: Past, present and future: *EAGE Publications*.
- Verschuur, D. J., and A. J. Berkhout, 1997, Estimation of multiple scattering by iterative inversion, Part II: Practical aspects and examples: *Geophysics*, **62**, 1596–1611.
- Weglein, A. B., F. V. Araújo, P. M. Carvalho, R. H. Stolt, K. H. Matson, R. T. Coates, D. Corrigan, D. J. Foster, S. A. Shaw, and H. Zhang, 2003, Inverse scattering series and seismic exploration.
- Weglein, A. B., F. A. Gasparotto, P. M. Carvalho, and R. H. Stolt, 1997, An inverse-scattering series method for attenuating multiples in seismic reflection data: *Geophysics*, **62**, 1975–1989.
- Ziolkowski, A. M., D. B. Taylor, and R. G. K. Johnston, 1999, Marine seismic wavefield measurement to remove sea-surface multiples: *Geophysical Prospecting*, **47**, 841–870.