Nonstationary $\mathbf{L}_1$ adaptive subtraction with application to internal multiple attenuation

Scott Keating, Jian Sun, Pan Pan and Kris Innanen

ABSTRACT

A method for adaptive subtraction in conjunction with internal multiple prediction by inverse scattering series is investigated. An L_1, nonstationary adaptive subtraction is found to minimize mistaken matching to primaries while still allowing for a large degree of multiple removal. This adaptive subtraction is tested on both synthetic and physical modeling data. The method proves to be effective, and avoids unwanted matching to primaries.

INTRODUCTION

Multiple reflections are an inevitability in seismic data, and in many cases they are of significant amplitude compared to the primaries in the data. While certain seismic techniques have been devised to gain meaningful information from multiple reflections, the vast majority of seismic methods are designed to work on primary only data, and for these methods multiples are undesirable. Consequently, multiple removal is an important process that is critical to the processing and interpretation of many data sets. Most classical multiple removal methods assume some difference between primaries and multiples and exploit this difference to separate and remove the multiples. For example, predictive deconvolution relies on the assumption that multiples, and not primaries, display periodicity (Peacock and Treitel (1969)), while tau-p filtering methods depend on moveout differences between primaries and multiples at any given arrival time (Foster and Mosher (1992)). While these methods are computationally cheap and often very effective, they struggle to remove multiples which badly violate their assumptions. Alternate methods of multiple removal involve wavefield prediction and subtraction. These methods include surface related multiple elimination (SRME) and inverse scattering series prediction (Weglein et al. (1997)). These methods are usually more robust, as they often make fewer assumptions than the discrimination based removal methods. While these methods can be very robust, approximations such as series truncation or the use of lower dimensional algorithms, together with the complicating factors in real data such as noise, residual ghosts and incompletely deconvolved source wavelets mean that the predicted multiple wavefields will not match the measured multiples exactly. This introduces the need for adaptive subtraction, which performs some kind of matching between the predicted and measured fields prior to subtraction.

For internal multiple removal, which is a particular problem in land seismic data, the inverse scattering series is an attractive method of prediction. Inverse sacttering multiple removal methods are being developed in a variety of domains, such as the wavenumber-pseudodepth domain (Pan (2015)), the tau-p domain (Sun and Innanen (2014)), and the space-time domain (Innanen (2014)). In this report, we aim to focus on adaptive subtraction techniques which work well for these internal multiple predictions.

ENERGY MINIMIZING ADAPTIVE SUBTRACTION

Before the multiples generated by wavefield prediction methods can be subtracted from the data, it is often necessary to correct the prediction such that it more accurately represents the multiples in the data. This correction is often applied by convolving the prediction with a filter, as this can allow for amplitude and phase differences between the prediction and measured data to be corrected (Abma et al. (2005), Wang (2003), Verschuur et al. (2005)). The result of adaptive subtraction, **a**, in one dimension is then obtained by

$$\mathbf{a} = \mathbf{d} - \mathbf{m} * \mathbf{f} = \mathbf{d} - \mathbf{M}\mathbf{f} \quad , \tag{1}$$

where **d** is the vector of measured data, **m** is the vector of multiple predictions, **M** is the convolution matrix of the predicted multiples, and **f** is the filter used.

The important problem in adaptive subtraction then becomes how to design this filter such that the ideal matching between the predicted and measured multiples is achieved. Perhaps the simplest measure of this matching is the energy removed by the adaptive subtraction (Verschuur et al. (2005)). If the multiples and primaries do not overlap at any point in the data, then the result of subtraction where the multiples are completely removed has less energy than any subtraction where the multiples are not completely removed (Verschuur et al. (2005)). Provided that the multiples and primaries do not overlap, the filter which minimizes the energy of our result should then be ideal. Minimizing the energy of our result is achieved by minimizing the L_2 norm of a, so to determine our filter, we minimize

$$||\mathbf{a}||_2 = ||\mathbf{d} - \mathbf{M}\mathbf{f}||_2 = \sum_{n=1}^{N} \mathrm{a}_n^2 \quad , \tag{2}$$

where a_n is the nth element of **A**, and N is the total number of samples in **d**. We can do this by solving the least squares equation

$$\mathbf{f} = (\mathbf{M}^T\mathbf{M})^{-1}\mathbf{M}^T\mathbf{d} \quad . \tag{3}$$

Once this filter has been determined, the result of the adaptive subtraction can be obtained using Equation 1. Figures 1 and 2 demonstrate the matching effects of an energy minimizing filter in a simple two interface model.

When solving for **f**, we first have to choose how long the filter will be. This parameter directly affects the capacity of the adaptive subtraction to match the predicted multiple to the measured data. If, for example, we restrict ourselves to a one point filter, the filter will only have the capacity to perform amplitude scaling on our prediction. On the other hand, if we choose a sufficiently long filter, our filter will be able to match any signal exactly. We must then compromise between choosing a filter that is too short, and incapable of matching well enough to remove multiples, and choosing a filter too long, and able to do unwanted matching to primaries.

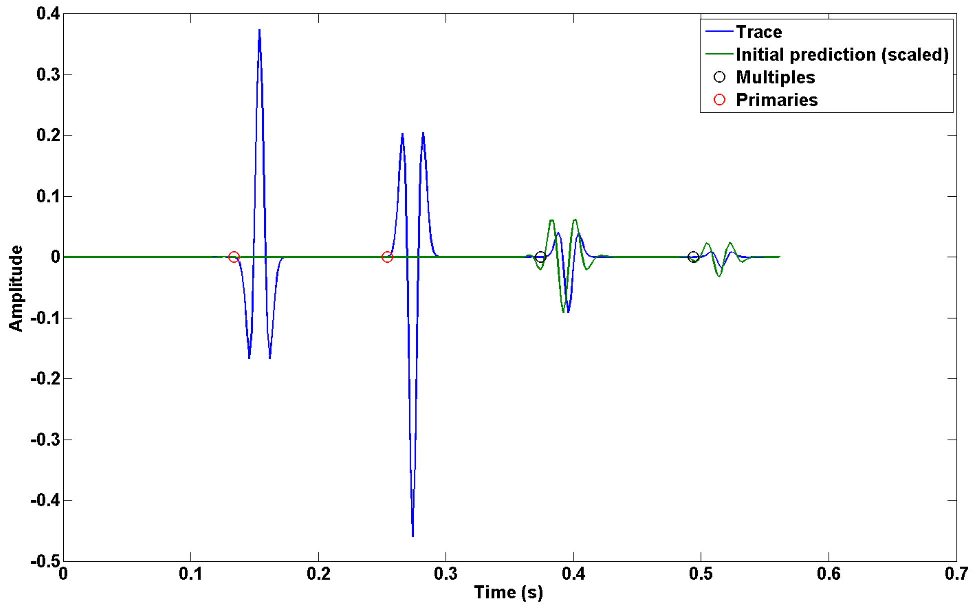

FIG. 1. Trace from a simulated shot record with a two interface model in blue, and multiple prediction after scaling in green.

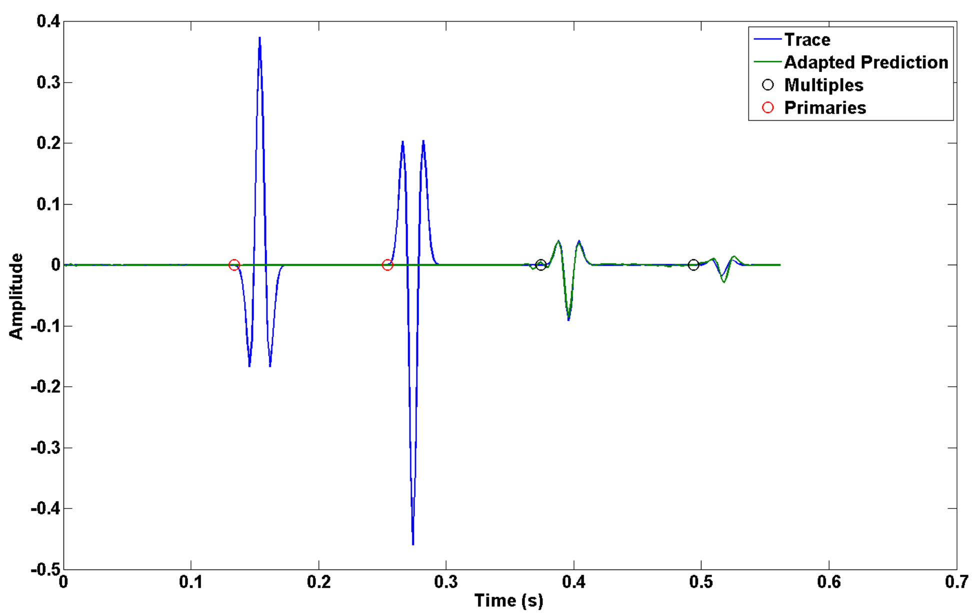

FIG. 2. Trace from a simulated shot record with a two interface model in blue, and multiple prediction after convolution with energy minimizing filter in green. In this case, the energy minimizing filter has successfully matched our prediction to the observed multiples.

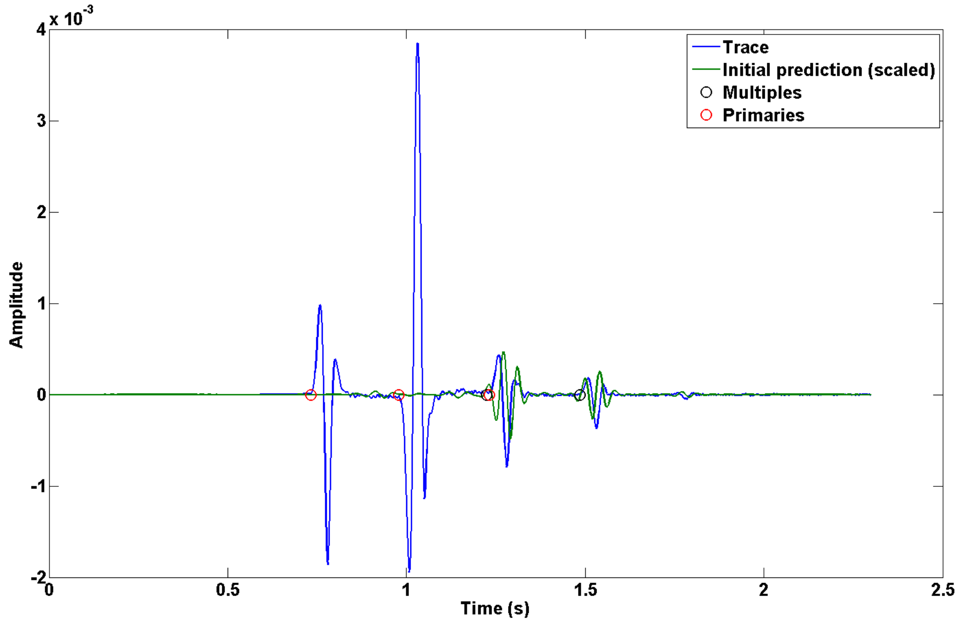

FIG. 3. Trace from a simulated shot record with a three interface model in blue, and multiple prediction after amplitude scaling in green. The third primary arrives at the same time as a multiple, causing significant overlap.

One significant problem with the energy minimizing adaptive subtraction is that we cannot in general safely make the assumption that multiples and primaries do not overlap. This means that when we minimize the energy of our result **a**, we are minimizing the energy of primaries and multiples together. Of course, this is potentially problematic if the energy minimizing subtraction causes primary energy to be removed, which in general will be the case in the event that multiples and primaries overlap. This problem is significantly compounded by the fact that primaries are typically greater in amplitude than multiples, and thus much greater in energy, which is proportional to the square of amplitude. It is not only possible then that primaries can have energy removed in an L_2 subtraction, they will actually be prioritized over removal of multiples. Not only can this cause the removal of primary energy, it can also result in the removal of multiple energy to be done poorly. In Figures 3 and 4, an energy minimizing adaptive subtraction is applied to a three interface model, where the arrival times of one of the primaries and a multiple roughly coincide. Here, the energy minimizing filter adversely affects the prediction; the adapted prediction does not match the observed multiples, and furthermore it will remove primary energy when subtracted.

While the non-orthogonality of primaries and multiples seems a major obstacle to energy minimizing adaptive subtraction, there are factors which help to lessen its importance. Overlap between primaries and predicted multiples is random; the primaries at a given time can alter the prediction of multiples at later times, but have no effect whatsoever on the multiples predicted at the same time. On the other hand, the similarities between the multiple prediction and the actual multiples should be highly systematic. The filter required to match the prediction to the multiples should then be very similar between different multiple events, whereas a filter which matches a multiple prediction to a primary in one region of overlap will not in general share any similarity with one in another region of overlap.

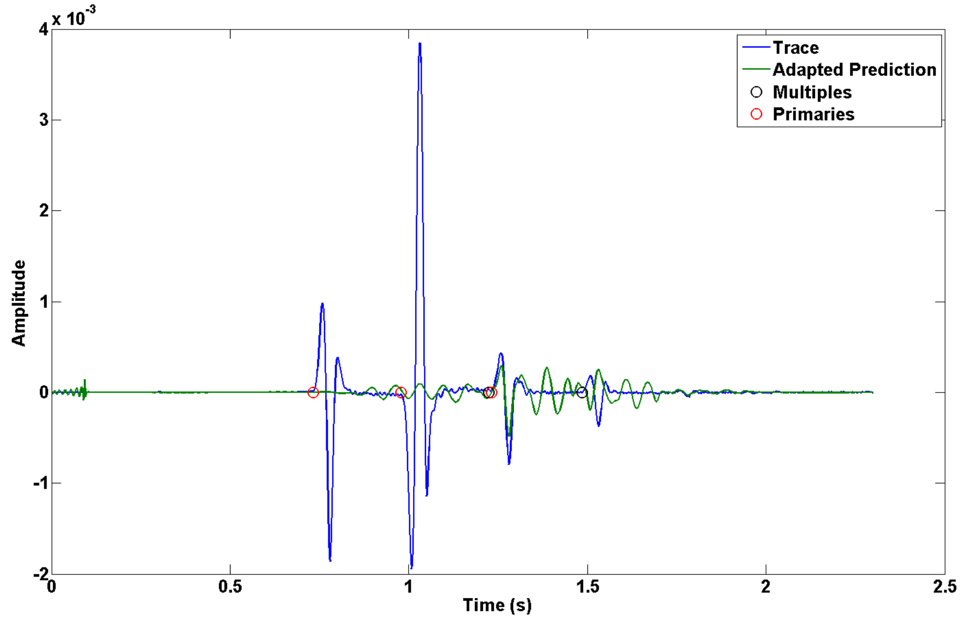

FIG. 4. Trace from a simulated shot record with a three interface model in blue, and multiple prediction after convolution with energy minimizing filter in green. Subtraction of this adapted prediction will not remove multiples, and will remove primary energy.

This means that a filter which matches the prediction to one multiple should work well at all multiples, whereas a filter which matches the prediction to a primary will not match the prediction to other primaries, and may actually increase mismatch in these areas. So, if there are many multiples and many regions of primary-multiple overlap, the problems associated with energy minimizing filters should be reduced.

In Figure 6, the effect of an energy minimizing filter is shown on a many multiple trace, with several times where primary-multiple overlap occurs. As compared with the initial prediction in Figure 5, the energy minimizing filter does a fairly good job of matching the prediction to the multiples. It is clear, however, that the matching could be better. We may expect that the quality of the matching could be improved by increasing the filter length. The result of a longer filter is shown in Figure 7. If anything, this filter does a worse job of matching the multiples than the short filter. Also evident is that there is slightly more unwanted matching to the third primary. Increasing the filter length increases the matching capability of the filter, but the energy minimizing solution here was to match the primary more closely, not the multiples. This limits the level of multiple subtraction we can hope to achieve in this case. Clearly, interfering primaries harm the energy minimizing adaptive subtraction even in this case.

ALTERNATIVES TO THE ENERGY MINIMIZING FILTER

While the energy minimizing adaptive subtraction works to some extent, it clearly puts significant restrictions on how well we are able to match multiples before matching to primaries becomes problematic. These problems arise due to the tendency of the energy minimizing subtraction to match predicted multiples to measured primaries. This suggests that an ideal filter would be designed based on the measured multiples alone, or some pri-

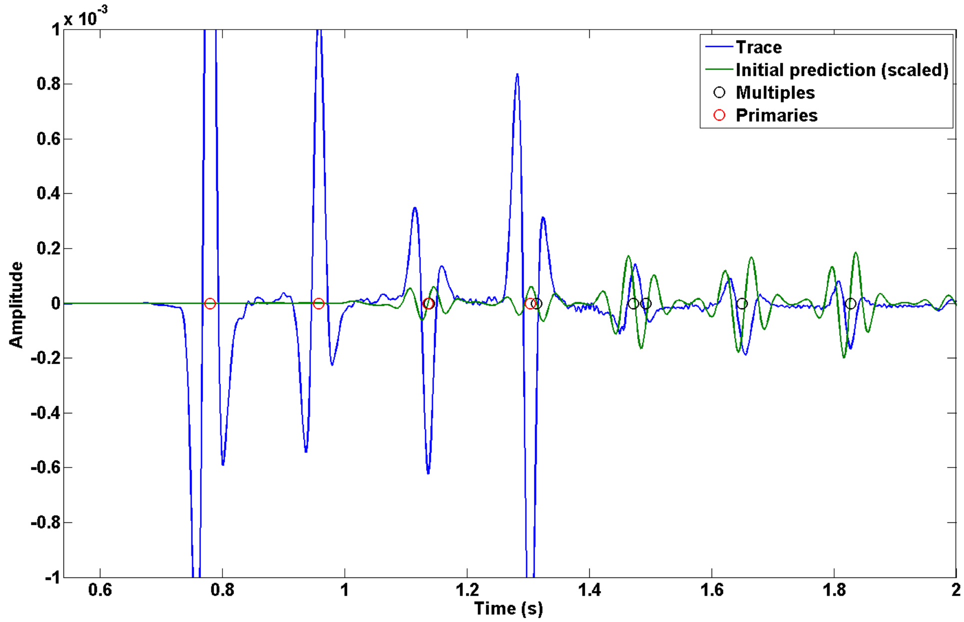

FIG. 5. Trace from a simulated shot record in blue, and multiple prediction after scaling in green. Significant overlap between primary and multiple occurs at two primary arrival times.

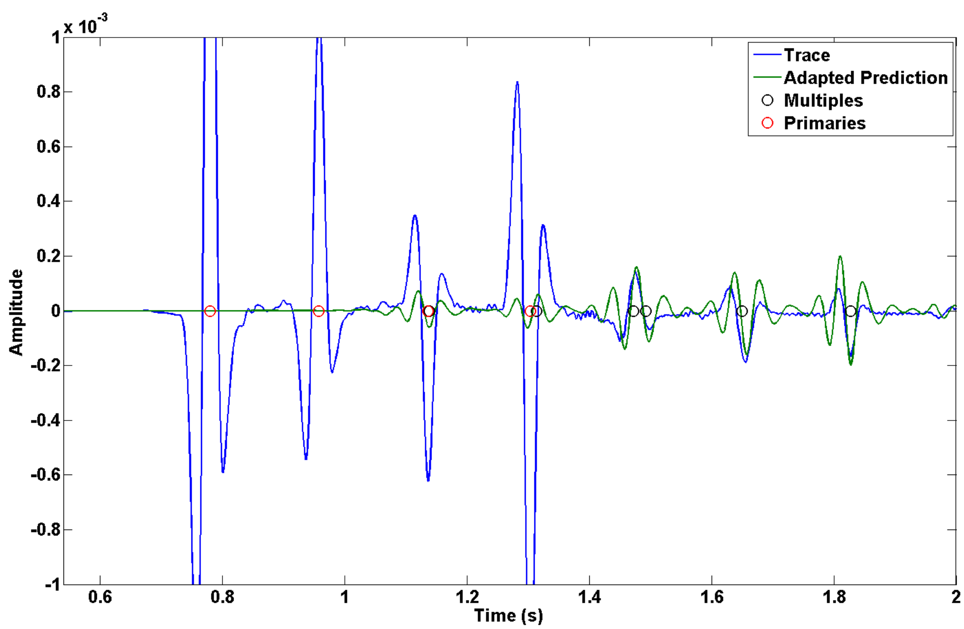

FIG. 6. Trace from a simulated shot record in blue, and multiple prediction after convolution with short energy minimizing filter in green. Arrival times are approximately corrected, but significant errors remain.

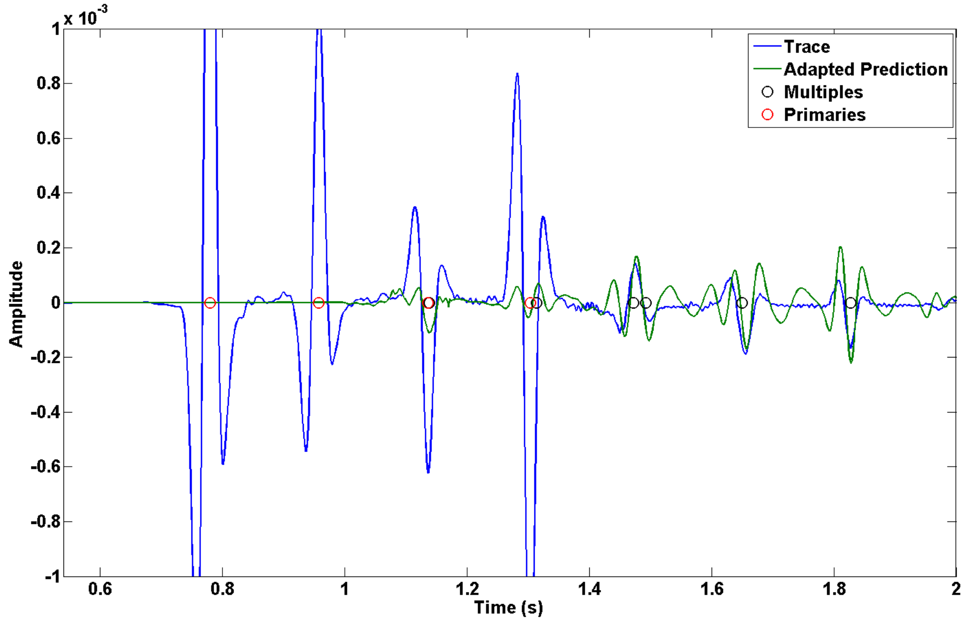

FIG. 7. Trace from a simulated shot record in blue, and multiple prediction after convolution with long energy minimizing filter in green. Matching to multiples is not noticeably improved over the short filter case, while slightly more primary matching is evident.

mary free portion of the data. Such a filter requires, however, that we are able to somehow discriminate between primaries and multiples. As discussed in the introduction, wavefield prediction and adaptive subtraction are typically used in situations where it is very difficult to distinguish between multiples and primaries based on some simple difference. Consequently, a more realistic criterion for a good adaptive subtraction might be that primaries are considered, but at an equal or reduced weighting as compared to multiples, and factors such as the random overlap between primaries and the prediction are relied on to further reduce the impact of primaries on the adaptive subtraction. A good candidate for this criterion is an L_1 minimizing adaptive subtraction.

$\mathbf{L_1}$ NORM

As mentioned previously, the energy minimizing adaptive subtraction is that which minimizes the L_2 norm of the result. As this minimizes the square of the amplitude of the subtracted result **a**, this causes the higher amplitude primaries to be evaluated at much greater weighting than the multiples, and prioritized in the minimization. Guitton and Verschuur (2004) suggest that we instead minimize the L_1 norm, as it is then the absolute value of the result which is minimized, as described by

$$||\mathbf{a}||_1 = ||\mathbf{d} - \mathbf{Mf}||_1 = \sum_{n=1}^{N} |\mathrm{a}_n| \quad . \tag{4}$$

By minimizing the L_1 norm, the weighting of high amplitude primaries is dramatically reduced as compared to the L_2 norm, which should help to mitigate the problems we observe in the energy minimizing adaptive subtraction. To find the filter which minimizes the L_1

norm, we solve the nonlinear normal equations (Bube and Langan (1997)) given by

$$\mathbf{M}^T\mathbf{W}\mathbf{M}\mathbf{f} = \mathbf{M}^T\mathbf{W}\mathbf{d} \quad , \tag{5}$$

where $\mathbf{W}$ is the diagonal matrix whose elements W_{ii} are related to the residual at time i by

$$W_{ii} = |r_i|^{-1} \quad , \tag{6}$$

and where the residual $\mathbf{r}$ is given by

$$\mathbf{r} = \mathbf{M}\mathbf{f} - \mathbf{d} \quad . \tag{7}$$

So, the expression for the L_1 minimizing filter is

$$\mathbf{f} = (\mathbf{M}^T\mathbf{W}\mathbf{M})^{-1}\mathbf{M}^T\mathbf{W}\mathbf{d} \quad . \tag{8}$$

Unfortunately, the elements of W are singular wherever the residual is zero. This makes the minimization of the L_1 norm problematic to compute, and impractical for our purposes. It is then desirable for the filter to minimize some other function that maintains the reduced weighting of high amplitude signals like the L_1 norm, but is better behaved as the residual approaches zero.

$\mathbf{L_1/L_2}$ HYBRID NORM

While there are several approaches to creating better behaved functions that emulate the L_1 norm, the approach pursued here is the L_1/L_2 hybrid norm of Bube and Langan (1997). The idea is to create a function which behaves like the L_1 norm for large residuals, where the L_1 behaviour is useful, and smoothly transitions to L_2 behaviour for small residuals, where the L_1 behaviour is problematic. Specifically, the L_1/L_2 norm minimizes

$$J = \sum_{n=1}^{N} j_n = \sum_{n=1}^{N} \sqrt{1 + \left(\frac{r_n}{\sigma}\right)^2} - 1 \quad , \tag{9}$$

where σ is a parameter chosen to control the transition from L_1 to L_2 behaviour. It can be easily seen that in the limit of very large or very small residuals, this expression becomes

$$j_n = \begin{cases} \frac{1}{2}\left(\frac{r_n}{\sigma}\right)^2 & , \text{ for } r_n \ll \sigma \\ \left|\frac{r_n}{\sigma}\right| & , \text{ for } r_n \gg \sigma \quad . \end{cases} \tag{10}$$

So, for very large residuals, we minimize the same L_1 expression as in Equation 4, and for very small residuals, we minimize the L_2 expression in Equation 2. As in the L_1 case, the filter which minimizes this expression can be found by solving

$$\mathbf{f} = (\mathbf{M}^T\mathbf{W}\mathbf{M})^{-1}\mathbf{M}^T\mathbf{W}\mathbf{d} \quad , \tag{11}$$

but with the weighting matrix W altered, so that

$$W_{ii} = \left(\frac{1}{1 + \left(\frac{r_i}{\sigma}\right)^2}\right)^{\frac{1}{2}} \quad . \tag{12}$$

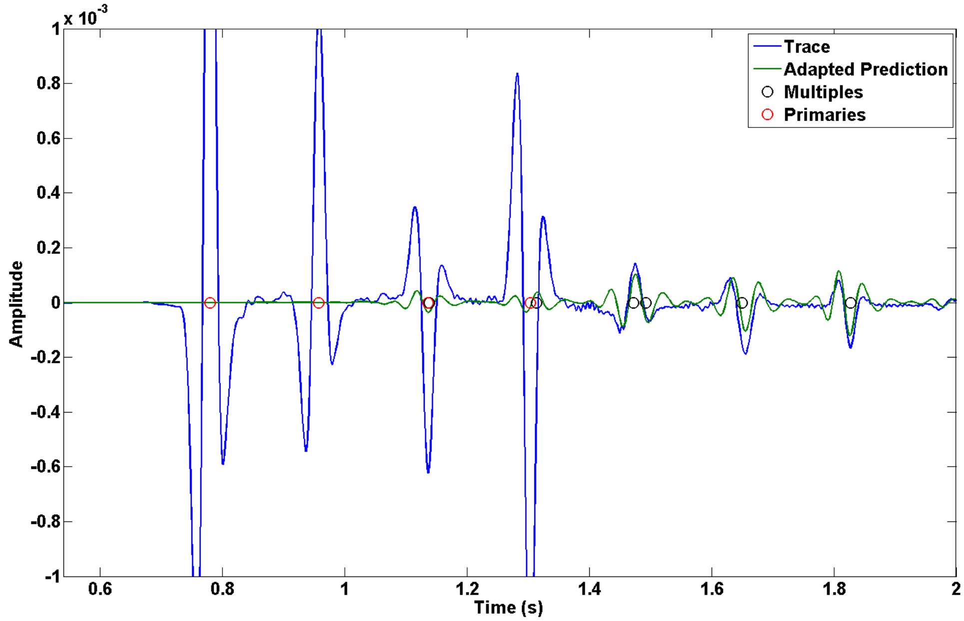

FIG. 8. Trace from a simulated shot record in blue, and multiple prediction after convolution with a short, hybrid norm minimizing filter in green. Notable improvements in matching to observed multiples can be seen as compared to the energy minimizing norm in Figure 6.

Equation 11 is nonlinear, as the weighting matrix W is a function of the filter used, $\mathbf{f}$, due to its dependence on the residual (Equation 7). Consequently, we are not able to solve Equation 11 directly. Instead, we have to solve Equation 11 iteratively. To begin, we solve Equation 11 for $\mathbf{f}$, using an identity matrix for $\mathbf{W}$. Using the filter obtained in this way, we can determine the residuals through Equation 7. This in turn allows us to determine $\mathbf{W}$ through Equation 12.This updated $\mathbf{W}$ allows us to calculate a more accurate filter, by again using equation 11. As we iterate this procedure, we obtain more and more accurate values for $\mathbf{f}$, until eventually we converge to the filter which minimizes the hybrid norm.

Figure 8 demonstrates the effect of using a hybrid norm minimizing filter. The initial prediction used was that from Figure 5. Clearly, the hybrid norm minimizing filter offers substantial improvements over the energy minimizing filter in Figure 6. Additionally, the reduced weighting of the primaries mean that we can now safely increase the length of the filter used to improve the matching to multiples without matching primaries. An example of a longer filter using the hybrid norm is shown in Figure 9.

EXTENSION TO TWO DIMENSIONS

Equation 1 is applicable on a trace-by-trace basis, but it is often desirable to design and apply the correction to groups of traces. This can offer several benefits, and makes the prediction more robust and reliable. Perhaps the greatest benefit of designing the filter considering all dimensions of the data is the reduced matching of the multiple prediction to primary data. This effect is due to the fact that while multiples and primaries may overlap in one particular trace, they will not in general overlay one another in exactly the same way at all source and receiver offsets. A filter which matches the predicted multiple to a primary on one trace then will not, in general, match the prediction to the primary

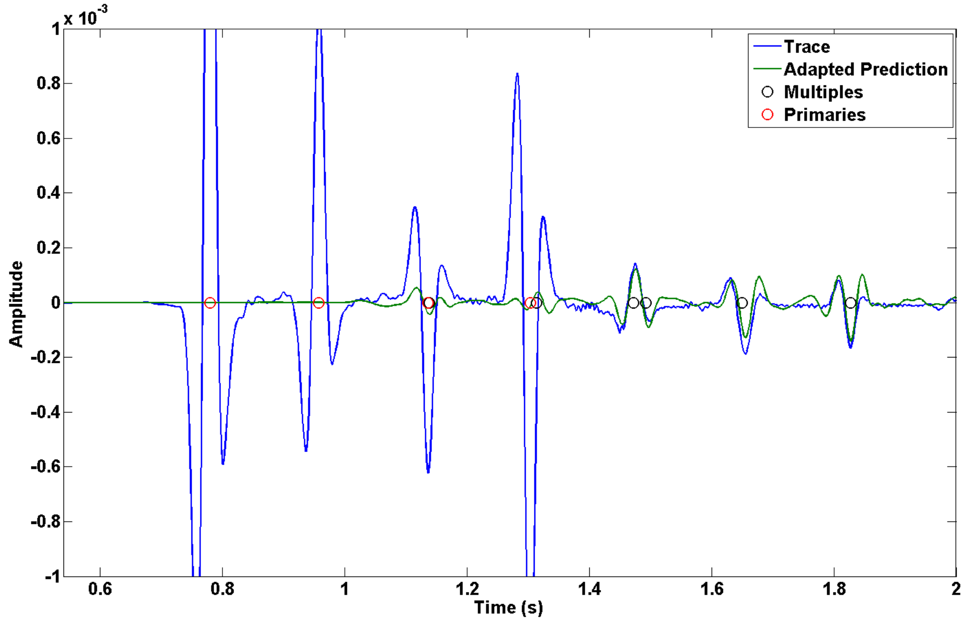

FIG. 9. Trace from a simulated shot record in blue, and multiple prediction after convolution with a long, hybrid norm minimizing filter in green. The longer filter allows for an improved level of matching to the observed multiples over the short filter case (Figure 8) without noticeably increasing the amount of primary energy removed.

on another, provided that the multiple and primary have some difference in their spatial behaviour, even if there is still some overlap between them.

In order to solve for the filter in two dimensions (e.g. offset and time), we must extend Equations 1 and 11 from one dimension. In two dimensions, the adaptive subtraction result becomes

$$\mathbf{A}_{xt} = \mathbf{D}_{xt} - \mathbf{M}_{xtp}\mathbf{f}_p \quad , \tag{13}$$

where $\mathbf{A}_{xt}$ is the result of the adaptive subtraction, $\mathbf{D}_{xt}$ is the measured data, $\mathbf{M}_{xtp}$ is a tensor such that for fixed x, $\mathbf{M}_{xtp}$ is the convolution matrix for the prediction at offset x, $\mathbf{f}_p$ is a filter of length p, and summation occurs over the repeated indices in each term. The energy minimizing expression for the filter is then

$$\mathbf{f}_p = (\mathbf{M}_{abp}\mathbf{M}_{abq})^{-1}\mathbf{M}_{xtq}\mathbf{D}_{xt} \quad . \tag{14}$$

Similarly, the expression for the L_1, or hybrid norm minimizing filter is given by

$$\mathbf{f}_p = (\mathbf{M}_{abp}\mathbf{W}_{abAB}\mathbf{M}_{ABq})^{-1}\mathbf{M}_{xtq}\mathbf{W}_{xtXT}\mathbf{D}_{XT} \quad , \tag{15}$$

where $\mathbf{W}_{xtXT}$ is nonzero only where $x = X$ and $t = T$, and the elements $\mathbf{W}_{xtxt}$ are the same as elements W_{ii} in equation 6 or 12 as appropriate, with residuals r_i replaced with the residuals in two dimensions r_{xt}.

NONSTATIONARITY

When wavefield prediction methods make assumptions about the subsurface, the validity of these assumptions may vary with depth, offset, or both. Furthermore, seismic data can often be inherently nonstationary. These factors mean that a single filter may not be adequate to match the predicted multiples to the multiples actually measured in a seismic experiment. It can be desirable to instead have a filter which varies with the offset and time of the recordings. This can be achieved by calculating a different filter for every point in the measured data. The filter at each point is created by choosing some window about the point, and solving for the filter which minimizes the hybrid norm on that window. The window used here was Gaussian in shape to minimize edge effects, and give greater weighting to data near the point for which the filter was calculated. The rate at which this filter is allowed to vary is controlled by the size of the window; if the window covers the entire data set the filter will be stationary, while if it covers only a few points it will vary quickly.

Like the filter length, the window size for the nonstationary filter controls how well the prediction is matched to the data. If the window chosen is too large, the filter may not be able to adequately match the nonstationary changes in the data or prediction. If instead the window is too small, the filter will be able to closely match any prediction to the signal, and will consequently remove primary energy.

An example is shown in Figures 10-14. In Figure 11, a shot record is shown, generated from a model with a series of layers dipping at 0.5° (Figure 10). The multiple prediction used for this example was created using a 1.5D algorithm, that is, the internal multiple prediction assumes two dimensional propagation through flat layers. Our slightly dipping case is a reasonably mild violation of the 1.5 D assumption, and we might expect for the 1.5 D prediction algorithm to produce a fairly accurate result. Indeed, the prediction generated in Figure 12 is quite similar to the multiples observed in the data. If we apply a stationary adaptive subtraction, however, the result obtained is worse than we might expect (Figure 13). Multiple attenuation is limited in some areas in this case, and is significantly variable with offset. This is because, while the prediction is very similar to the observed multiples throughout, the filter required to match the one to the other is not constant, especially with varying offset. Adaptive subtraction is repeated with a nonstationary filter in Figure 14. It is easy to perceive that this allows for substantial improvement in the subtraction.

PHYSICAL MODELING EXAMPLE

To demonstrate the effect of the adaptive subtraction described thus far, some an application to physical modeling data is shown in this section. The physical modeling was done on a scale model made of layers of water, polyvinyl chloride, Plexiglas and aluminum (Pan (2015)). For source and receiver, piezoelectric transducers were used. To create the physical modeling data, a receiver is simulated at each of 120 locations, with a single, constant location source used for each. This simulates a two dimensional shot gather. This physical modeling shot gather is shown in Figure 15. The corresponding prediction obtained by an inverse scattering series method is shown in Figure 16. The result of applying adaptive subtraction given the data and prediction is shown in Figure 17. While most of the multiples in the data are near or below the noise level of the data even before the subtraction, the prominent multiple just after 2 s is significantly attenuated by the adaptive subtraction. This occurs despite the strong overlap of one of the predicted internal multiples and the high amplitude free surface multiple just before 2.5 s, which has the potential to significantly impair an energy minimizing adaptive subtraction, just as interference with a primary would.

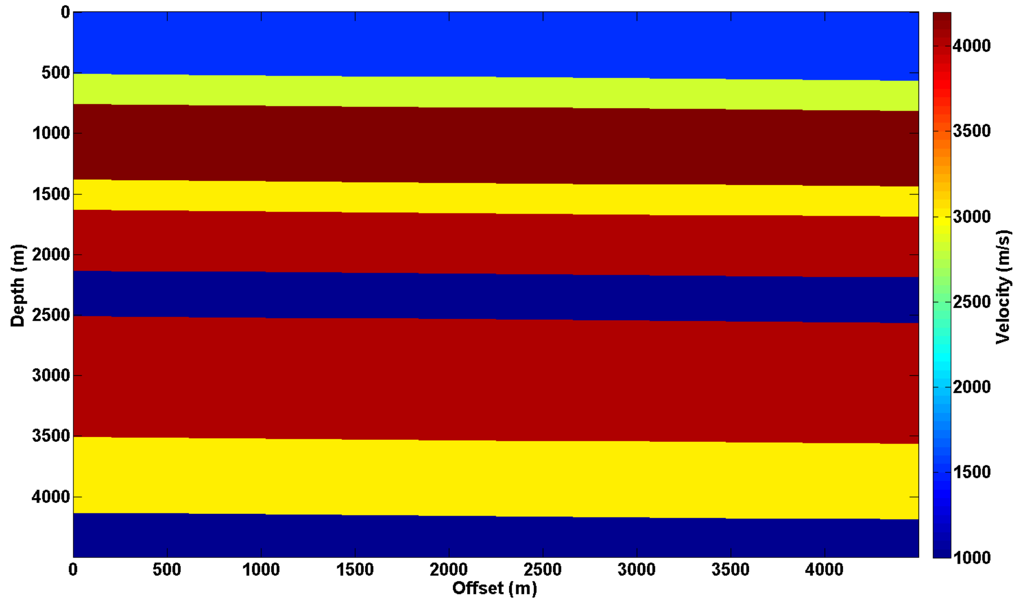

FIG. 10. Velocity model with slightly dipping layers.

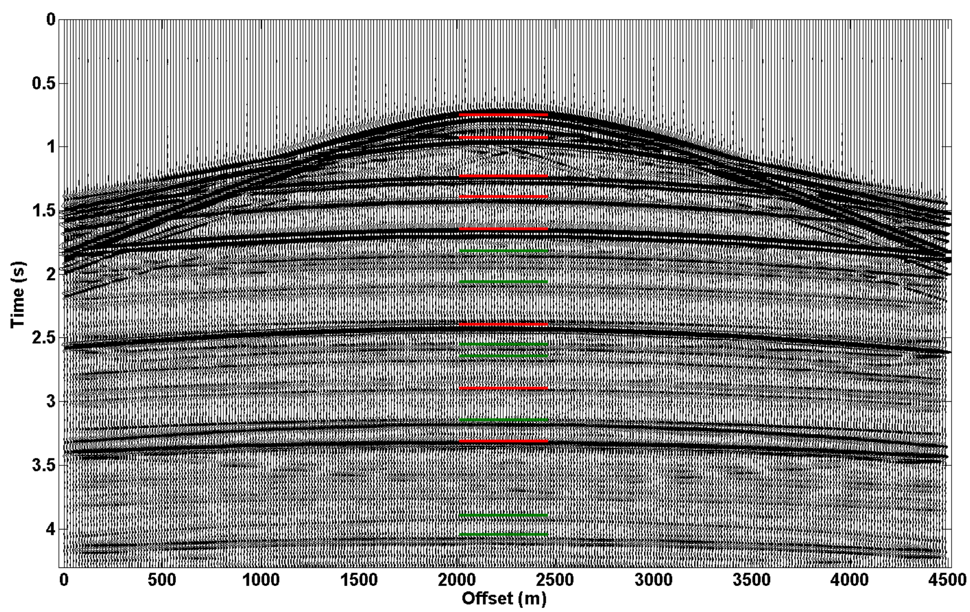

FIG. 11. Shot gather with small dip reflectors. The zero offset primary arrival times are marked in red, some notable multiples in green.

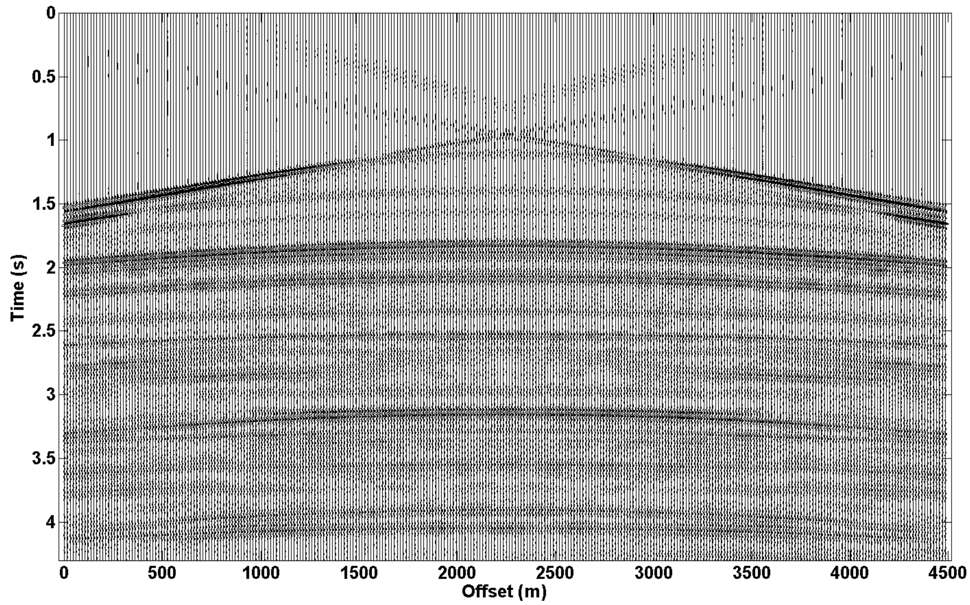

FIG. 12. 1.5D internal multiple prediction. Multiples predicted are very similar to those observed.

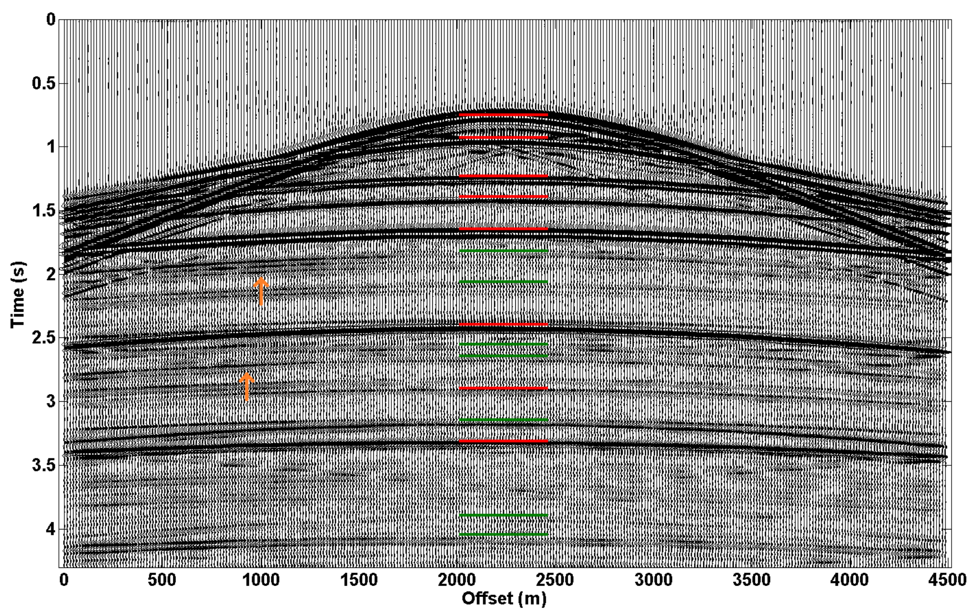

FIG. 13. Shot gather with small dip reflectors after stationary adaptive subtraction. The zero offset arrival times are marked in red, some notable multiples in green. The quality of the subtraction varies with offset. Arrows highlight major differences between the stationary and nonstationary cases.

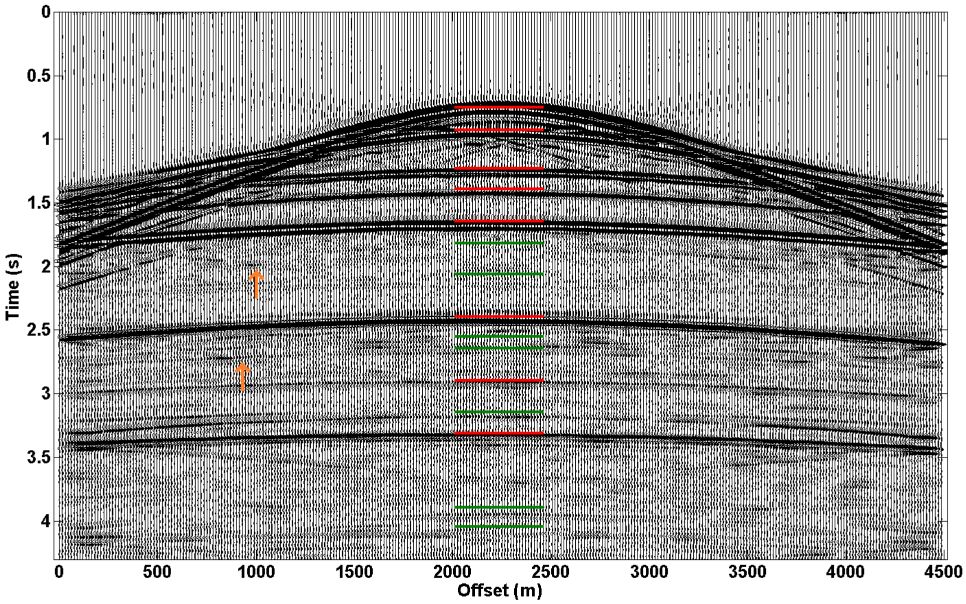

FIG. 14. Shot gather with small dip reflectors after nonstationary adaptive subtraction. The zero offset arrival times are marked in red, some notable multiples in green. Significant improvements can be seen in comparison to the stationary case. Arrows highlight major differences between the stationary and nonstationary cases.

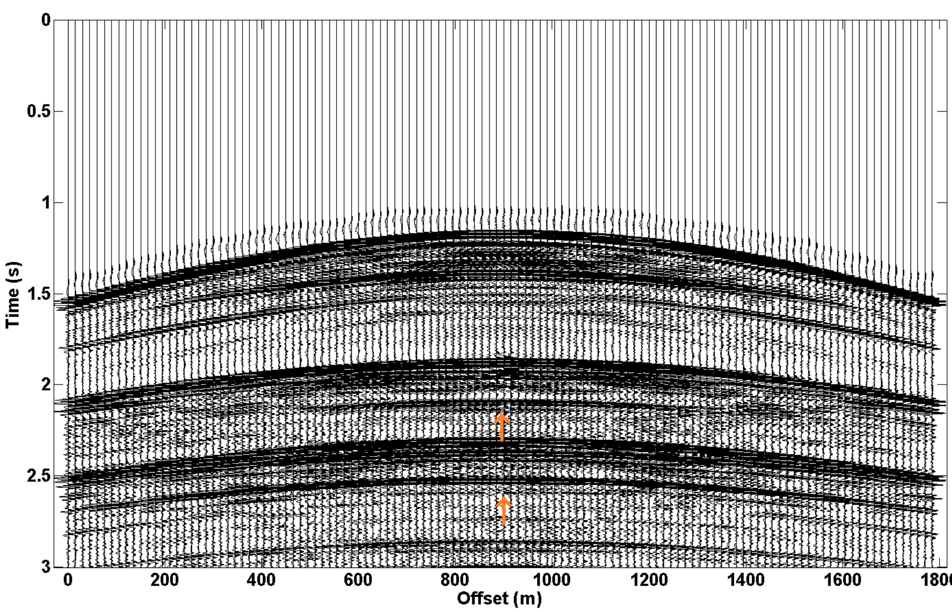

FIG. 15. Physical modeling shot gather. Arrows highlight visible multiples.

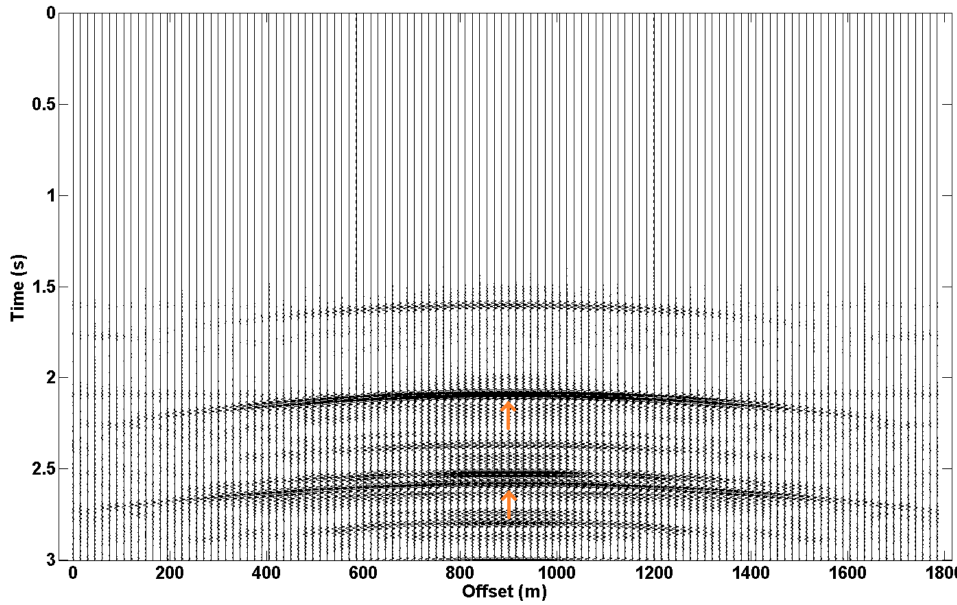

FIG. 16. Physical modeling multiple prediction. Arrows highlight multiples visible in the measured data.

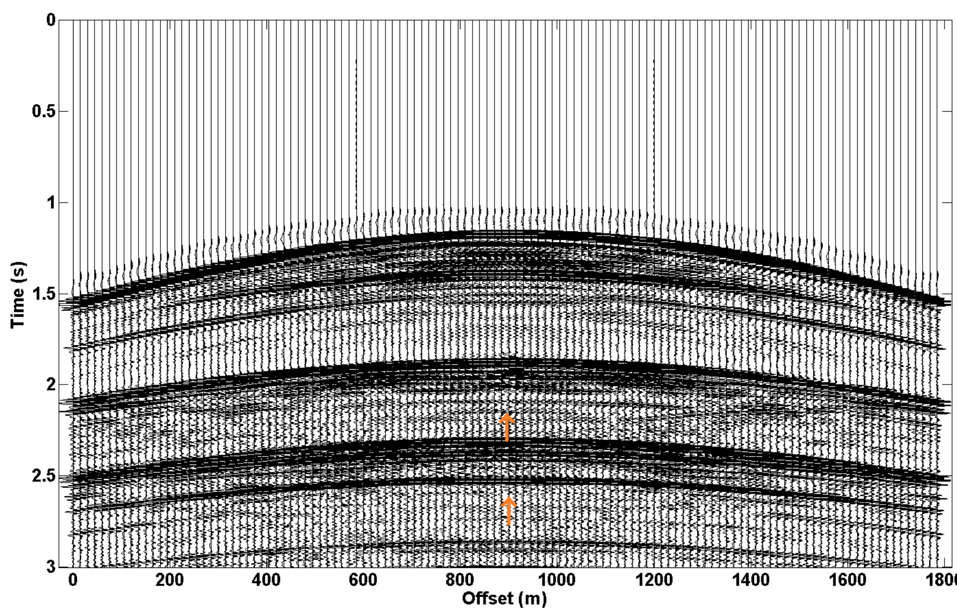

FIG. 17. Physical modeling shot gather after adaptive subtraction of the predicted multiples. Arrows highlight multiples visible in the measured data.

CONCLUSION

Inverse scattering multiple prediction is a powerful method for predicting multiples, but the predictions it generates require some correction before they can be successfully removed from the data. One means of applying such a correction is by convolution with a filter. While in general matching schemes will choose filters that increase matching to both primaries and multiples, we can improve matching to multiples without unwanted matching to primaries by applying a nonstationary, two dimensional hybrid norm minimizing filter in our adaptive subtraction. This was shown to be an effective adaptive subtraction technique, both when applied to synthetic data and when applied to physical modeling data.

FUTURE WORK

Future work for this project will include more specifically designing adaptive subtractions for each of the different prediction domain implementations of the inverse scattering prediction algorithm. Adaptive subtraction on fully two or even three dimensional predictions is also a point of interest. Lastly, prediction and subtraction of multiples in real land seismic data is another priority.

ACKNOWLEDGEMENTS

The authors thank the sponsors of CREWES for continued support. This work was funded by CREWES industrial sponsors and NSERC (Natural Science and Engineering Research Council of Canada) through the grant CRDPJ 461179-13. Scott Keating was also supported by the Queen Elizabeth II scholarship.

REFERENCES

Abma, R., Kabir, N., Matson, K., Shaw, S., and McLain, B., 2005, Comparison of adaptive subtraction methods for multiple attenuation: The Leading Edge.

Bube, K., and Langan, R., 1997, Hybrid l1/l2 minimization with applications to tomography: Geophysics, **62**.

Foster, D., and Mosher, C., 1992, Suppression of multiple reflections using the radon transform: Geophysics, **57**.

Guitton, A., and Verschuur, D., 2004, Adaptive subtraction of multiples using the l1-norm: Geophysical Prospecting, **52**.

Innanen, K., 2014, Internal multiple prediction in the plane wave domain: CREWES annual report, **26**.

Pan, P., 2015, 1.5d internal multiple prediction: an application on synthetic data, physical modeling data and land data sythetics: M.Sc. thesis, Univ. of Calgary.

Peacock, K., and Treitel, S., 1969, Predictive deconvolution: Theory and practice: Geophysics, **34**.

Sun, J., and Innanen, K., 2014, Internal multiple prediction in the plane wave domain: CREWES annual report, **26**.

Verschuur, D., Berkhout, A., and Wapenaar, C., 2005, Adaptive surface-related multiple elimination: Geophysics, **68**.

Wang, Y., 2003, Multichannel subtraction using an expanded multichannel matching filter: Geophysics, **57**.

Weglein, A., Gasparratto, F., Carvalho, P., and Stolt, R., 1997, An inverse-scattering series method for attenuating multiples in seismic reflection data: Geophysics, **62**.