

MULTIDIMENSIONAL SEISMIC NOISE ATTENUATION

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF GEOPHYSICS
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Antoine Guitton

May 2005

© Copyright by Antoine Guitton
All Rights Reserved

printed as Stanford Exploration Project No. 121
by permission of the author

Copying for all internal purposes of the sponsors
of the Stanford Exploration Project is permitted

I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Jon F. Claerbout
(Principal Adviser)

I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Biondo L. Biondi

I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Gary Mavko

Approved for the University Committee on Graduate Studies.

Abstract

This dissertation addresses the problem of seismic noise attenuation by using inversion. Two types of noise are considered. First, spiky noise, which is attenuated with a robust error measure called the Huber norm. This norm treats small residual with the ℓ^2 norm and large residuals with the ℓ^1 norm. It is continuous everywhere and can be minimized with a standard gradient-based solver. In this thesis, the well-known quasi-Newton method L-BFGS is used. The Huber norm and the L-BFGS solver are flexible enough to be used in many different situations, such as velocity analysis and regridding of noise bathymetry data. Another powerful application of the Huber norm is adaptive subtraction of surface-related multiples. In this case, weak multiples in the vicinity of strong primaries are better removed with the Huber norm than with ℓ^2 .

Second, coherent noise, which is attenuated by making the residual components of any least-squares fitting of contaminated data independent and identically distributed (IID). To achieve this goal, a weighting and a modeling approach are introduced. The weighting approach aims at approximating the inverse data covariance operator (or matrix) with multidimensional prediction-error filters (PEFs). The modeling approach introduces a coherent noise modeling operator inside the inversion. The modeling technique converges usually better than the weighting approach and yields smaller residuals.

One advantage of the weighting approach, however, is that it can be used to separate non-stationary signal and non-stationary noise with PEFs. In this case, the separation is called pattern-based because it involves the multivariate noise and signal spectra that PEFs approximate. For removal of surface-related multiples, this technique proves being more robust to

modeling inadequacies than adaptive subtraction with the ℓ^2 norm, as exemplified on marine 2-D and 3-D field data examples.

Preface

The electronic version of this thesis¹ makes the included programs and applications available to the reader. The markings [ER], [CR], and [NR] are promises by myself about the reproducibility of each figure result. Reproducibility is a way of organizing computational research that allows both the author and the reader of a publication to verify the reported results. Reproducibility facilitates the transfer of knowledge within SEP and between SEP and its sponsors.

ER denotes Easily Reproducible and are the results of processing described in the paper. I claim that you can reproduce such a figure from the programs, parameters, and makefiles included in the electronic document. The programs are present either locally in the different directories (Chapters 4 and 5) or in the Alib directory which contains all the programs for the pattern-based approach of Chapters 6 and 7, the adaptive subtraction of Chapter 3 and the non-linear solver of Chapter 2. The data must either be included in the electronic distribution, be easily available to all researchers (e.g., SEG-EAGE data sets), or be available in the SEP data library². I assume you have a UNIX workstation with Fortran, Fortran90, C, X-Windows system and the software downloadable from our website (SEP makerules, SEPlib, and the SEP latex package), or other free software such as SU. Before the publication of the electronic document, someone other than me tested my claim by destroying and rebuilding all ER figures. Some ER figures may not be reproducible by outsiders because they depend on data sets that are too large to distribute.

¹<http://sepwww.stanford.edu/public/docs/sep121>

²http://sepwww.stanford.edu/public/docs/sepdata/lib/toc_html/

CR denotes Conditional Reproducibility. I certify that the commands are in place to reproduce the figure if certain resources are available. SEP staff have only attempted to make sure that the makefile rules exist and the source codes referenced are provided. The primary reasons for the CR designation is that the processing requires 2 hours or more. In Chapter 7, proprietary data from Compagnie Générale de Géophysique cannot be used without written permission. All the Figures displaying this dataset are CR. Migration codes used in Chapters 6 and 7 are available in the Plib directory (courtesy of Paul Sava).

M denotes a figure that may be viewed as a movie in the web version of the report. A movie may be either ER or CR.

NR denotes Non-Reproducible figures. SEP discourages authors from flagging their figures as NR except for figures that are used solely for motivation, comparison, or illustration of the theory, such as: artist drawings, scannings, or figures taken from SEP reports not by the authors or from non-SEP publications.

Our testing is currently limited to LINUX 2.4 (using either the Intel or Portland Group Fortran90 compiler), but the code should be portable to other architectures. Reader's suggestions are welcome. For more information on reproducing SEP's electronic documents, please visit <http://sepwww.stanford.edu/research/redoc/>.

Acknowledgments

My interest for exploration geophysics has grown over the years for a simple reason: guidance from mentors who were passionate about their work, and most importantly, who believed in my abilities. First, Laurence Nicolétis at the Institut Français du Pétrole, who got me exposed to the oil industry and its challenges. Laurence was and still is a source of inspiration for her enthusiasm and energy. Second, Guillaume Cambois at Compagnie Générale de Géophysique (CGG), who taught me many valuable lessons about seismic processing and seismic data. Guillaume also encouraged my decision to start a Ph.D. at Stanford and helped me getting a sponsorship from CGG. This funding provided me with financial resources for the first three years. Finally, Jon Claerbout who kept challenging me with new research ideas. After long hours spent at analyzing processing results on a computer screen, scrutinizing every single detail, Jon passed on me his taste for field data examples. The challenges can be quite intimidating, but the gratification might me invaluable.

SEP offers an outstanding research environment by allowing students investigating pretty much anything at will. Biondo Biondi has always been very supportive of my research and kept giving me valuable advices and suggestions. Biondo offered me the opportunity to go to Delft during the summer of 2001, which turned out to be a very stimulating and enriching working experience. This trip was made possible thanks to the financial support of the SMAART joint venture. Diane Lau has spared me many headaches by solving many administrative problems. She made my life so much easier on many different occasions; I am grateful to her for this.

During my stay at SEP, I had the chance to work with many different people. Most importantly, many of them became my friends. Among them, James Rickett, Louis and Sylvie

Vaillant, Morgan Brown, Paul and Diana Sava, and Jesse Lomask deserve a special mention. James taught me all the (dirty) tricks to survive SEP. I always considered James as my mentor at SEP. Louis started three months after me at SEP and left after the master. I shared with Morgan a passion for outdoor activities that lead us from the top of North Dome to the bottom of the Norwegian fjords. Paul and Diana had the “pleasure” to have me for dinner almost everyday when Diana and I were working at Chevron. During that summer, I taught Diana how to drive “a la French”. I am happy to report that she is still accident free! Jesse has been one of my closest friend inside and outside SEP.

Many of my colleagues at Stanford made my life as a student an incredibly enjoyable experience: Nizar Chemingui, Sergey Fomel, Bob and Marie Clapp, Seth Haines, Daniel Rosales, Gabriel Alvarez, Bill Curry, Brad Artman, Nick Vlad, Alejandro Valenciano, Guojian Shan, Jeff Schragge, Kevin Wolf and Laura Chiaramonte.

Finally, I would like to thank my parents, my sister Marie, my brothers Williams and Yohanne, and also Thierry, Mohamed, and Annabel for their love and support during all these years. They helped me to keep everything in perspective, provided all the motivation I needed to accomplish my work, and much more.

Contents

Abstract	v
Preface	vii
Acknowledgments	ix
1 Introduction	1
2 Robust inversion using the Huber norm	9
3 Adaptive subtraction of multiples with the ℓ^1 norm	25
4 Noise attenuation by filtering and modeling	39
5 Interpolation of bathymetry data	61
6 Multiple attenuation: Theory and Practice	81
7 Multiple attenuation: A 3-D field data example	123
8 Conclusions	157

A	Algorithm for minimizing the Huber function	161
B	Least-squares solution of the modeling approach	167
C	Estimation of nonstationary PEFs	175
D	A surface-related multiple prediction theory	183
	Bibliography	187

List of Figures

1.1	Land seismic record with few bad geophones.	2
1.2	Marine seismic record with two water bottom reflections WB1 and WB2. WB1 is generated by the airgun towed by a seismic vessel. WB2 is created by a noisy source within the water column at offset 5000 m. The origin of this source is unknown.	2
1.3	Shot record from a land acquisition in Saudi Arabia. The strong ground roll, although containing valid near-surface information, is often attenuated. The signal on the top needs to be isolated for further processing.	3
1.4	Shot record from a marine acquisition in the Gulf of Mexico. The first reflections (WB) on the top, also known as primaries, are immediately followed by multiple reflections (WBM1, WBM2 and WBM3). An important step in seismic processing is to attenuate these multiples before migration to isolate the primaries.	4
2.1	Error measure with $\alpha = 3$ proposed by Huber (Huber, 1973) (dash line) and the ℓ^2 norm (solid line).	11
2.2	Synthetic data used for the inversion. (a) The true velocity model represented in slowness. (b) The noise-free input data used for the inversion. (c) The same data but with noise added (four spikes).	16

2.3	Result of the inversion with the ℓ^2 norm for the noise-free data. (a) Inverted slowness field. (b) Remodeled data. (c) Difference between the input (Figure 2.2b) and remodeled data.	17
2.4	Result of the inversion with the ℓ^2 norm for the data with noise. (a) Inverted slowness field. (b) Remodeled data. (c) Difference between the input (Figure 2.2c) and remodeled data. The four spikes create artifacts in both inverted model and reconstructed data space.	18
2.5	Result of the inversion with the ℓ^2 norm with damping for the data with noise. (a) Inverted slowness field. (b) Remodeled data. (c) Difference between the input (Figure 2.2c) and remodeled data. The three panels are cleaner than in Figure 2.4 but some artifacts remain, however.	19
2.6	Result of the robust inversion with the Huber norm for the data with noise. (a) Inverted slowness field. (b) Remodeled data. (c) Difference between the input (Figure 2.2c) and remodeled data. The Huber norm behaves like a pure ℓ^1 norm since all artifacts have disappeared.	20
2.7	The field data used for the inversion. Notice the amplitude anomalies at near offset and the time shift near offset 2 km.	20
2.8	The result of the inversion with the ℓ^2 norm for the field data. (a) Inverted slowness field. (b) Remodeled data. (c) Difference between the input (Figure 2.7) and remodeled data. The horizontal stripes in the velocity panel are created by the amplitude anomalies at short offsets.	21
2.9	The result of the inversion with the ℓ^2 norm and regularization for the field data. (a) Inverted slowness field. (b) Remodeled data. (c) Difference between the input (Figure 2.7) and remodeled data. The model is much cleaner than in Figure 2.8a, but the horizontal events remain.	22
2.10	The result of the robust inversion with the Huber norm for the field data. (a) Inverted slowness field. (b) Remodeled data. (c) Difference between the input (Figure 2.7) and remodeled data.	22

3.1	(a) The data with one primary on the left, and three multiples on the right. (b) The multiple model that we want to adaptively subtract from (a).	28
3.2	(a) The signal estimated with the ℓ^2 norm. (b) The noise estimated with the ℓ^2 norm.	28
3.3	Shaping filter estimated for the 1-D problem with the ℓ^2 norm. This filter is not a single spike at $lag = 0$	29
3.4	Shaping filter estimated for the 1-D problem with the ℓ^1 norm. This filter is a single spike at $lag = 0$	29
3.5	(a) The signal estimated with the ℓ^1 norm. (b) The noise estimated with the ℓ^1 norm.	31
3.6	(a) A synthetic shot gather infested with internal multiples. (b) The internal multiples model obtained using the Common Focus Point technology (Berkhout and Verschuur, 1999). This model matches the internal multiples in (a).	32
3.7	(a) The estimated primaries with the ℓ^2 norm. (b) The estimated internal multiples with the ℓ^2 norm. Ideally, (b) should look like Figure 3.6b, but it does not.	33
3.8	(a) The estimated primaries with the ℓ^1 norm. (b) The estimated internal multiples with the ℓ^1 norm. Beside some edge-effects, (b) resembles closely Figure 3.6b. The adaptive subtraction worked very well.	34
3.9	(a) The difference between the exact multiples (Figure 3.6b) and the subtracted multiples with the ℓ^2 norm. (b) The difference between the exact multiples (Figure 3.6b) and the subtracted multiples with the ℓ^1 norm.	35
3.10	(a) Stack infested with multiples. (b) The multiple model computed with the Delft modeling approach. The subtraction is done poststack.	36
3.11	(a) The estimated primaries with ℓ^1 norm adaptive subtraction. (b) The estimated primaries with ℓ^2 norm subtraction. The primary at 1.6s is very attenuated with the ℓ^2 norm. The ℓ^1 technique preserves its amplitude better.	36

3.12	(a) The estimated multiples with the ℓ^1 norm subtraction. (b) The estimated multiples with the ℓ^2 norm subtraction. The ℓ^2 norm tends to over-fit some multiples that creates some leaking of primaries in the estimated noise.	37
4.1	Some basic properties of the prediction-error filters (PEF). (a) A PEF is estimated from a dataset. (b) The estimated PEF is convolved with the input data leading to a white residual. (c) A spike is deconvolved with the estimated PEF leading to a pattern close to the input data in (a).	48
4.2	A synthetic data example with (a) the data and its (f, k) spectrum, (b) the true signal and its (f, k) spectrum, and (c) the noise. The spectra of signal and the noise overlap, which makes simple filtering (i.e., without inversion) likely to fail.	51
4.3	(a) Estimated model space after inversion of the noisy data in Figure 4.2a. (b) Remodeled data from the inverted model. (c) Residual \mathbf{r}_d after inversion. The residual is not IID and a lot of noise contaminates the model space.	52
4.4	(a) (f, k) spectrum of the PEF. It is minimum at the noise location but varies a lot everywhere else. (b) (f, k) spectrum of the projection filter. It is minimum at the noise location and flat otherwise. The projection filter is a better filter than the prediction filter.	53
4.5	Inversion result with the filtering technique. (a) Estimated model. (b) Remodeled data. (c) Unweighted residual, which shows the coherent noise only. (d) Residual. The estimated model is artifacts free and the residual is IID. The movie shows the four panel as a function of the iterations.	54
4.6	Inversion result with the modeling technique. (a) Estimated model \mathbf{m}_s . (b) Remodeled data $\mathbf{H}\mathbf{m}_s$. (c) Remodeled noise $\mathbf{A}^{-1}\mathbf{m}_n$. (d) Residual. The residual is IID and the model space is clean. Because the signal and noise modeling operators tend to overlap, a little bit of noise leaks in the estimated signal. This effect is very small, however.	55

4.7	A CMP gather from a land survey. The ground-roll is very strong throughout the section.	56
4.8	Result of least-squares inversion of the data in Figure 4.7. (a) Estimated model space. The noise level is too high to see any useful information. (b) The remodeled data. The low noise level in this panel indicates that most of the energy in (a) resides in the null space of the operator. (c) Data residual. The residual is not IID.	57
4.9	Inversion result with the filtering technique. (a) Estimated model. (b) Remodeled data. (c) Unweighted residual, showing the ground-roll only. (d) Residual. The residual is not exactly IID. Some coherent energy remains. . .	58
4.10	Inversion result with the modeling technique. (a) Estimated model \mathbf{m}_s . (b) Remodeled data $\mathbf{H}\mathbf{m}_s$. (c) Remodeled noise $\mathbf{A}^{-1}\mathbf{m}_n$. (d) Residual. The residual is IID.	59
4.11	A comparison of convergence between the filtering and modeling approach for the land data example.	60
5.1	Depth of the Sea of Galilee along the vessel tracks. On one traverse across the lake, the depth record is “u” shaped. A few “v” shaped tracks result from vessel turn-arounds. All the points used for building the maps are shown here. Each point corresponds to one depth measurement inside the lake. Note that the input data are displayed as “seismic traces” for display purpose only. The long signal is broken into 23 strips of 5718 depth measurements. The horizontal axis corresponds to a measurement number. The depth appears as one long track although the surveying was done in several episodes that do not always continue the same track.	63
5.2	Simple binning of the raw data in Figure 5.1. The ship tracks and empty bins are visible and need to be accounted for in the inversion process. The north points to the top.	65

5.3	(a) $\mathbf{p} = \mathbf{Hh}$ estimated with equation (5.4) in a least-squares sense after 50 iterations, which simulates a least-squares solution with damping. (b) $\mathbf{p} = \mathbf{Hh}$ estimated with equation (5.6) in a ℓ^1 sense. The spikes have been correctly attenuated. Some interesting features are shown by the arrows: AS points to few ancient shores, O points to some outliers, T points to a few tracks and R points to a ridge.	69
5.4	(a) View of the bottom of the lake ($\mathbf{h} = \mathbf{H}^{-1}\mathbf{p}$) with the ℓ^2 norm after 50 iterations, which simulates a least-squares solution with damping. (b) View of the bottom of the lake with the ℓ^1 norm. Note that with the ℓ^1 norm, the spikes have been attenuated. Each line represents one east-west track every 500 meters.	70
5.5	(a) Estimated \mathbf{p} without attenuation of the tracks, i.e., equation (5.5). (b) Estimated \mathbf{p} with the derivative along the tracks, i.e., equation (5.7). (c) Estimated \mathbf{p} without tracks, i.e., equation (5.9). (d) Recorder drift in model space $\mathbf{B}'\mathbf{Lq}$.	74
5.6	Close-ups of the western shore of the Sea of Galilee. (a) Estimated \mathbf{p} without attenuation of the tracks, i.e., equation (5.5). (b) Estimated \mathbf{p} with the derivative along the tracks, i.e., equation (5.7). (c) Estimated \mathbf{p} without tracks, i.e., equation (5.9). (d) Recorder drift in model space $\mathbf{B}'\mathbf{Lq}$	75
5.7	(a) Input data acquired between A1 and B1 in Figure 5.2. The ship is approximately moving bottom to top going east from A1 to B1. (b) $\mathbf{BH}^{-1}\mathbf{p}$ estimated after inversion, i.e., the estimated noise-free data. (c) Estimated drift after inversion. (d) Data residual after inversion. The horizontal axis represents the measurement number.	77
5.8	(a) Input data acquired between A2 and B2 in Figure 5.2. The ship is approximately moving right to left going south from A2 to B2. (b) $\mathbf{BH}^{-1}\mathbf{p}$ estimated after inversion, i.e., the estimated noise-free data. (c) Estimated drift after inversion. (d) Data residual after inversion. The horizontal axis represents the measurement number.	78

6.1	prediction-error filters in (a) 2-D and (b) 3-D. The first coefficient in black is always one.	86
6.2	Stratigraphic interval velocity model of the Sigsbee2B dataset.	94
6.3	Two constant offset sections (h=1125 ft) of the Sigsbee2B dataset with (a) and without (b) free surface condition. The multiples are very strong below 5 s. The weak horizontal striping in (a) comes from a source effect only present with the free surface condition modeling. Arrow WB shows the water-bottom reflection, WBM1 the first order surface-related multiple for the water-bottom, WBM2 the second order surface-related multiple for the water-bottom, and P a strong primary.	95
6.4	Migrated images at zero-offset for the data with (a) and without (b) free surface condition. Comparing with Figure 6.3, the multiples appear much weaker below the salt after migration. However, some reflectors near 22 kft are hidden in (a). Arrow WBM1 shows the first order water-bottom multiple after migration. Arrow N shows some noise associated with the migration of multiples beneath the salt body.	96
6.5	Two constant offset panels at h=1125 ft. for (a) the estimated primaries and (b) the difference with the true primaries. The true primaries and multiples are used to estimate the PEFs. Arrow P shows a primary that could be mistaken for a multiple.	97
6.6	(a) Migration result after multiple attenuation when the true primaries and multiples are used to estimate the PEFs. (b) Difference between (a) and Figure 6.4b. The estimated primaries are almost exact.	98
6.7	Two constant offset sections (h=1125 ft) after multiple attenuation with the Spitz approximation using (a) 2-D and (b) 3-D filters.	99
6.8	Two migration results of the estimated primaries with (a) 2-D and (b) 3-D filters. The circles show areas where multiples are better attenuated with 3-D filters than with 2-D filters.	100

6.9	Close-up of Figure 6.8 showing two migrated images when (b) 2-D and (c) 3-D filters are used. The true primaries are shown in (a). Arrow '1' points to primaries that are attenuated with the pattern-based approach. The circles show an area where the 3-D filters are the most effective at removing the multiples.	101
6.10	Illustration of a pegleg multiple with a bounce on a scatter point recorded at zero-offset. These events are difficult to model because of the lack of far and/or crossline offsets.	102
6.11	Constant offset section (h=500 m) of the data with multiples. DTs point to the tails of diffracted multiples. These pegleg multiples bounce once on the seafloor and once on the scatter. Arrow WB shows the water bottom, WBM1 the first order surface-related multiple, and WBM2 the second order surface-related multiple.	104
6.12	Constant offset section (h=550 m) of the estimated multiples. The multiples are accurately modeled except for the diffracted multiples, shown as DT, for which the limited range of offsets and number of shots hamper any attempt at modeling the diffraction tails. Arrow WBM1 shows the first order surface-related multiple, and WBM2 the second order surface-related multiple.	109
6.13	Constant offset section (h=550 m) of the estimated primaries with adaptive subtraction. The arrows point to locations where multiples are still present. Arrow WBM1 shows the remaining energy for the first order surface-related multiple.	110
6.14	constant offset section (h=550 m) of the estimated primaries with pattern recognition. Multiples are better attenuated than in Figure 6.13. Arrow WBM1 shows the remaining energy for the first order surface-related multiple.	111

6.15	Constant offset sections (h=700 m) for (a) the input data, (b) the multiple model, (c) the estimated primaries with adaptive subtraction, and (d) estimated primaries with the pattern-based approach. Arrows point to locations where the pattern-based approach attenuates multiples significantly better than the adaptive subtraction.	112
6.16	Constant offset sections (h=4550 m) for (a) the input data, (b) the multiple model, (c) the estimated primaries with adaptive subtraction, and (d) estimated primaries with the pattern-based approach. Arrow A points to aliasing effects due to the offset sampling of the shot gathers. The pattern-based approach attenuates the multiples better than the adaptive subtraction in '1' and '2'. . .	113
6.17	Constant offset sections (h=3300 m) for (a) the input data, (b) the multiple model, (c) the estimated primaries with adaptive subtraction, and (d) estimated primaries with the pattern-based approach. '1' points to a primary that the pattern-recognition preserves very well. '2' points to an event that is attenuated with the pattern-based approach but not with the adaptive subtraction in (c). Though not shown here, a close inspection of the corresponding shot gathers suggests that '2' is actually a multiple.	114
6.18	Constant offset sections (h=5050 m) for (a) the input data, (b) the multiple model, (c) the estimated primaries with adaptive subtraction, and (d) estimated primaries with the pattern-based approach. '1' and '5' show events better preserved with the pattern-based method. '2' and '1' are better recovered with the adaptive subtraction. '4' seems to point to a primary that the adaptive subtraction is able to save. Because the area is contaminated with strong multiples, it is difficult to know without a stratigraphic model if '3' is a primary or not. . .	115
6.19	Stacked section of a 2-D line from the Gulf of Mexico. Surface-related multiples appear below 3 s. Arrow WB indicates the water bottom, TS the top of salt, BS the base of salt, WBM1 the first order water-bottom multiple, TSWBM the top of salt/water bottom pegleg multiple, TSM the top of salt/top of salt multiple, and WBM2 the second order water-bottom multiple.	116

6.20	Shot gathers outside the salt boundaries at 4500 m for (a) the input data, (b) the estimated primaries, (c) the multiple model with SRMP, and (d) the estimated multiples.	117
6.21	Shot gathers inside the salt boundaries at 12000 m for (a) the input data, (b) the estimated primaries, (c) the multiple model with SRMP, and (d) the estimated multiples.	118
6.22	A comparison of stacked sections for (a) the input data, (b) the estimated primaries, (c) the multiple model with SRMP, and (d) the estimated multiples. Some primaries (shown as P) in (b) are recovered while some diffracted multiples (shown as DM), not properly modeled by SRMP, are remaining.	119
6.23	comparison of migrated sections in the angle domain for (a) the input data with multiples, (b) the estimated primaries from the pattern-based approach, and (c) the estimated primaries with adaptive subtraction. Arrows M point to multiples.	120
6.24	Angle domain common image gathers of the Gulf of Mexico dataset for (a) the input data, (b) the estimated primaries with the pattern-based approach and, (c) the estimated primaries with adaptive subtraction. This gather position is X=4000 m, outside the salt boundaries. Primaries are flat (shown as P) and multiples are curved (shown as M).	121
7.1	Nearoffset section of the binned data for the CGG dataset. The cube is 70 km long by 2 km wide. Note the crossline dips in the right panel and the complex overburden where salt bodies are present close to the surface. The top part is a time slice at 2.4 s. Arrow WB shows the sea bottom, TS a top of salt reflection, WBM1 a first order water-bottom multiple, and WBM2 a second order water-bottom multiple.	126

7.2	Acquisition geometry for one sail line. S points to the locations of one source going from 710 km to 642 km. 1,2,3 and 4 points to the receiver lines for the four different streamers. The receiver locations (black lines) are shown for nine shots for the entire sail line. The data corresponding to streamer 1 are processed in this Chapter.	129
7.3	Near offset section (h=250 m) from streamer 1 in Figure 7.2. TS points to top of salt reflections. B points to sedimentary basins. WBM1 points to a first-order surface related multiple and WBM2 points to a second-order surface related multiple. The S/N ratio is extremely low below the salt.	131
7.4	Subset of the offset and CMP distributions. The bin size is 12.5 m along the CMP axis and 25 m along the offset axis. The CMP gathers are interpolated along the offset axis to recover the missing shots and the near offset information.	132
7.5	Illustration of the radon transform for one CMP gather. (a) Input CMP gather showing strong primaries around 2 s, first-order multiples near 4 s, and second-order multiples near 6 s. (b) Masking operator \mathbf{M} . The black stripes are located at the known data locations ($M_{i,i} = 1$) while the rest of the mask is set to zero. (c) Radon domain with Cauchy regularization, i.e., equation (7.4). (d) Radon domain without Cauchy regularization, i.e., equation (7.3).	133
7.6	(a) Input gather with missing traces. (b) Reconstructed gather with the Cauchy regularization. (c) Reconstructed gather without regularization.	134
7.7	F-K spectra of (a) the input CMP gather in Figure 7.6a with missing traces, (b) the reconstructed CMP gather with the Cauchy regularization in Figure 7.6b, (c) the reconstructed CMP gather without regularization in Figure 7.6c. The Cauchy regularization helps to preserve the steep dips better. The steep dips correspond to the aliased energy near 20 Hz.	135
7.8	Interpolation results in the shot domain with (a) and without (b) Cauchy regularization. Shots one and four are known while shots two and three are new. The near offset traces have been interpolated for all the shots. No AGC has been applied to these gathers.	136

7.9	Close-ups of common-offset gathers in the vicinity of first-order multiples at 250 m offset for (a) the input data, (b) the predicted multiples with the sparseness constraint, and (c) the predicted multiples without regularization. The horizontal axis represents the midpoint position. Arrow A points to artifacts in the multiple prediction that are caused by interpolation errors.	138
7.10	Close-ups of common-offset gathers in the vicinity of first-order multiples at 4.425 km offset for (a) the input data, (b) the predicted multiples with the sparseness constraint, and (c) the predicted multiples without regularization. The horizontal axis represents the midpoint position. 1 points to the water-bottom multiple that is better modeled when the Cauchy regularization is used for interpolation. Arrow A points to artifacts similar to those in Figure 7.9c. The two circles point to aliasing artifacts created during the multiple prediction.	139
7.11	Shot gather at 661.65 km for (a) the input data, (b) the predicted multiples with the sparseness constraint, and (c) the predicted multiples without regularization. D points to diffracted multiples that are not properly modeled in either (b) or (c). AA points to aliasing artifacts that create the fictitious events in the circles of Figure 7.10.	140
7.12	A constant offset section at 500 m offset for (a) the input data, (b) the multiple model, (c) the estimated primaries with adaptive subtraction, and (d) the estimated primaries with the pattern-based approach. The horizontal axis represents the midpoint position. 1 and 2 points to area where the multiples are not accurately predicted and where the pattern-based approach performs better than adaptive subtraction.	144

7.13	A shot gather ($X=664$ km) for (a) the input data, (b) the multiple model, (c) the estimated primaries with adaptive subtraction, and (d) the estimated primaries with the pattern-based approach. 1 shows multiples that are not properly predicted and not completely removed. 2 shows an event that is predicted at the wrong location but still attenuated in (d). 3 shows an area where the pattern-based approach attenuates more multiples than adaptive subtraction.	145
7.14	A shot gather ($X=693$ km) for (a) the input data, (b) the multiple model, (c) the estimated primaries with adaptive subtraction, and (d) the estimated primaries with the pattern-based approach. 1 points to a primary that is better preserved in (d).	146
7.15	Migrated images for (a) the data, (b) the estimated primaries with adaptive subtraction, (c) the difference between (a) and (b), i.e., the estimated multiples, and (d) the multiple model with SRMP. Multiples in the circle are not well attenuated.	148
7.16	Migrated images for (a) the data, (b) the estimated primaries with the pattern-based approach, (c) the difference between (a) and (b), i.e., the estimated multiples, and (d) the multiple model with SRMP. Multiples in the circle are better attenuated in (b) than in Figure 7.15b. In addition, the primaries look more continuous everywhere.	149
7.17	Migrated images for (a) the data, (b) the estimated primaries with adaptive subtraction, (c) the difference between (a) and (b), i.e., the estimated multiples, and (d) the multiple model with SRMP. The base of salt is indicated by an arrow in (a). Some subsalt multiples are still present. Arrow WB shows the water bottom, TS the top of salt, and BS the base of salt.	150
7.18	Migrated images for (a) the data, (b) the estimated primaries with the pattern-based approach, (c) the difference between (a) and (b), i.e., the estimated multiples, and (d) the multiple model with SRMP. The base of salt is better preserved than in Figure 7.17b and the subsalt multiples are well attenuated. Arrow WB shows the water bottom, TS the top of salt, and BS the base of salt.	151

7.19	ADCIG at $X=658.975$ for (a) the data, (b) the estimated primaries with adaptive subtraction, and (c) the estimated primaries with the pattern-based approach. The arrow at 4.6 s. shows the base of salt reflection. The second arrow shows multiples at high angles with strong curvatures. The pattern-based approach yields a cleaner gather.	152
7.20	ADCIG at $X=674.325$ for (a) the data, (b) the estimated primaries with adaptive subtraction, and (c) the estimated primaries with the pattern-based approach. The two arrows show primaries that are more continuous after pattern-based subtraction of the multiples.	153
B.1	A geometric interpretation of the noise filter when \mathbf{n} and \mathbf{s} are not orthogonal.	173
B.2	A geometric interpretation of the noise filter when \mathbf{n} and \mathbf{s} are orthogonal. . .	174
D.1	The wavefield is emitted at s and recorded at g . The multiple bounces somewhere at g'	184

Chapter 1

Introduction

As for any measurement of physical properties, seismic data are contaminated with a large variety of noise sources. This noise comes in different flavors and complexities that require careful and specific treatment at different stages of the processing workflow. This thesis addresses the problem of noise attenuation of seismic data using inversion, thus preserving as well as possible the integrity of the signal.

WHAT IS SEISMIC NOISE ?

For seismic data, noise sources can be separated into two categories. The first category of noise comes from experimental errors. These errors comprise any unexpected perturbation of the recording environment during data acquisition. A geophone can have malfunctions (i.e., Figure 1.1) or the recording systems can have glitches creating erratic noise in the seismic record. Wind motion or cable vibrations can generate random noise. Outside factors, such as mammal activity and/or drilling rigs for marine acquisition might also contaminate seismic records. These noise sources create more coherent energy on the data and can be misinterpreted as true signal. As an illustration Figure 1.2 displays a shot record from the Gulf of Mexico with reflections generated from an unknown submarine location. The second category

Figure 1.1: Land seismic record with few bad geophones. [intro-wz.11](#) [ER]

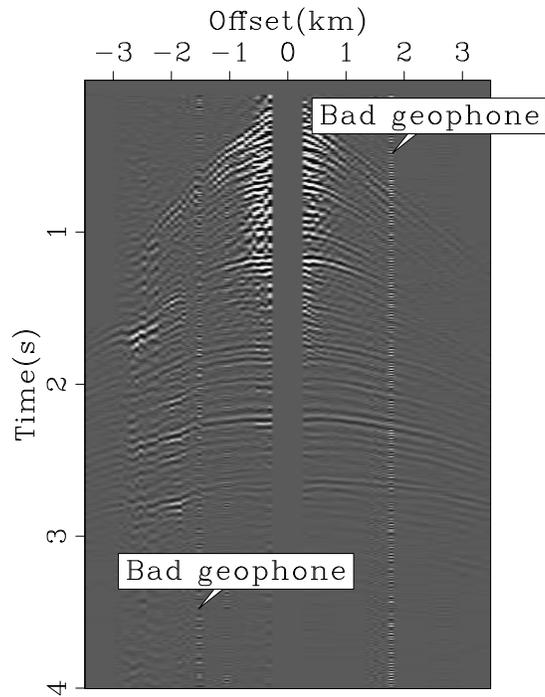
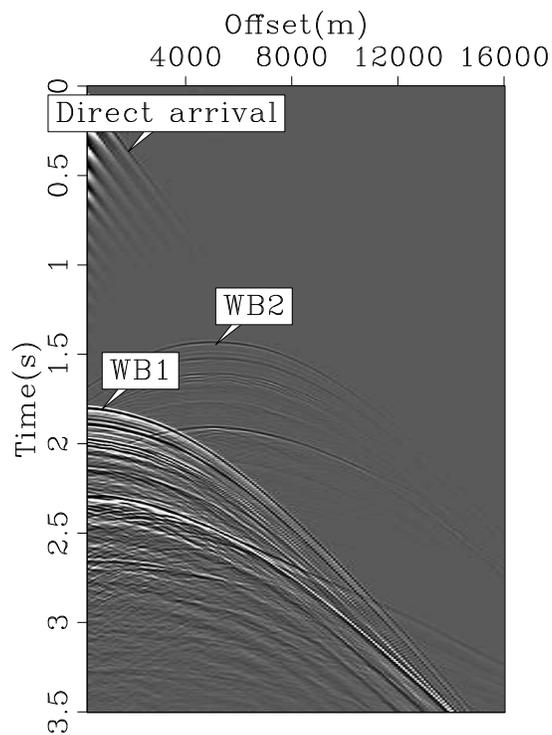


Figure 1.2: Marine seismic record with two water bottom reflections WB1 and WB2. WB1 is generated by the airgun towed by a seismic vessel. WB2 is created by a noisy source within the water column at offset 5000 m. The origin of this source is unknown. [intro-gulf1](#) [ER]



of noise comes from modeling uncertainties (Tarantola, 1987). In seismic, modeling uncertainties occur when the physical description and parameterization of the earth is incomplete. This incomplete description is motivated by the inherent complexity of wavefield propagation in the subsurface. Therefore, the complex seismic signal is often separated into different propagation modes that are then easier to understand and use. The downside of this approach is that most of the recorded signal is consequently obliterated. For instance with land data, the ground roll (Figure 1.3), which might contain useful near surface information, is attenu-

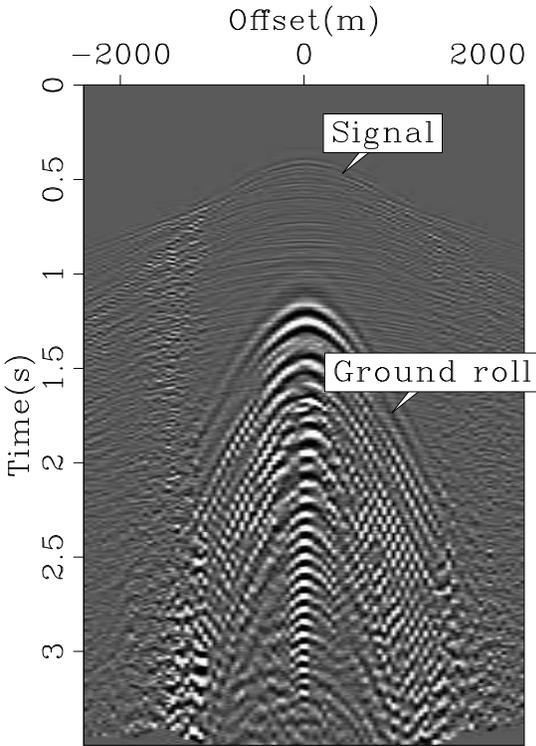
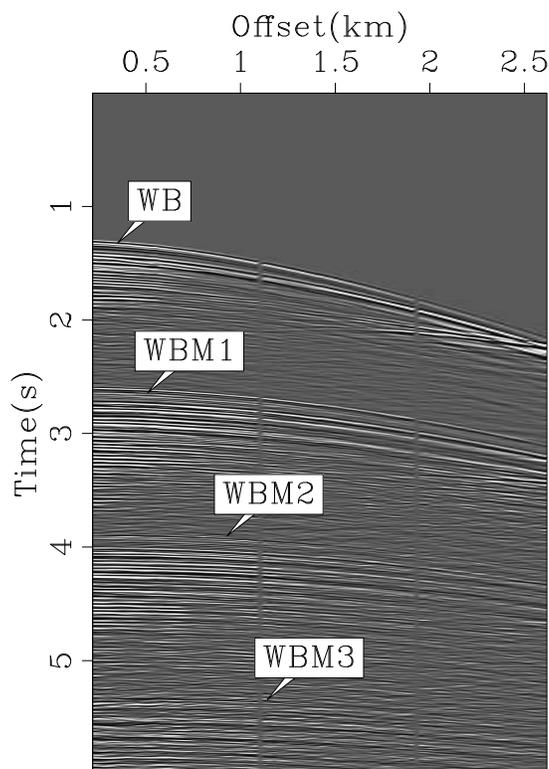


Figure 1.3: Shot record from a land acquisition in Saudi Arabia. The strong ground roll, although containing valid near-surface information, is often attenuated. The signal on the top needs to be isolated for further processing. [intro-dune](#) [ER]

ated to unravel seismic reflections. Multiple reflections are a main source of coherent energy in the seismic record that also need to be separated from the primaries (i.e, reflections with one bounce in the subsurface) before migration (Figure 1.4). Whereas experimental errors rarely bring any useful information, recent advances in inversion and migration are slowly pushing the limits of modeling uncertainties by incorporating and saving as much signal as possible within the processing workflow. Among these techniques, full waveform inversion (Tarantola, 1986; Sirgue and Pratt, 2004) bears promises by using all the subsurface reflections to derive an accurate earth model. Reverse time migration with the two-way wave equation

Figure 1.4: Shot record from a marine acquisition in the Gulf of Mexico. The first reflections (WB) on the top, also known as primaries, are immediately followed by multiple reflections (WBM1, WBM2 and WBM3). An important step in seismic processing is to attenuate these multiples before migration to isolate the primaries. [intro-wz.30](#) [ER]



(Baysal et al., 1983; Youn and Zhou, 2001) can theoretically migrate both multiples and primaries at the same time. One-way wave equation techniques can also migrate multiples but none-the-less require separation of primaries and multiples (Berkhout and Verschuur, 1994; Guitton, 2002b). Recently, Brown (2004) demonstrated how inversion can both image and separate multiples. Passive-seismic imaging (Rickett and Claerbout, 1999) can yield accurate depth maps by crosscorrelating noisy traces without any knowledge of the source location or spectrum.

Although very promising, most of these new techniques are still at an experimental level. Therefore, strategies for seismic noise attenuation are needed to preserve the signal of interest. In this thesis, spurious and coherent noise are attenuated using inverse theory. In particular, by analyzing the statistics of the noise, e.g., probability density functions (PDF) and patterns, we can decide what type of norm and what type of covariance operator are needed for the inversion. Covariance operators act as weighting functions of the data misfit (residual). They insure convergence of any inversion scheme toward independent and identically distributed

(IID) residual components, a central assumption in least-squares inversion. In this thesis, covariance operators are approximated with multidimensional non-stationary prediction-error filters (PEFs).

MULTIDIMENSIONAL SEISMIC NOISE ATTENUATION

Among the different sources of noise in seismic data, noise bursts from experimental errors and coherent noise from modeling uncertainties are very common. By identifying the statistical nature of both noises, inversion theory provides the tools to properly recover the signal of interest.

The Huber norm

Chapter 2 introduces the Huber norm (Huber, 1973). The Huber norm is an hybrid ℓ^1/ℓ^2 error measure that is robust to outliers. Because it is differentiable everywhere, the Huber norm can be minimized with a gradient-based algorithm. In Chapter 2, I propose minimizing the Huber norm with a quasi-Newton method called L-BFGS. This technique can solve any non-linear problem where local-minima are sought. Its limited-memory requirements make it also very attractive for solving large-scale problems. On a velocity estimation problem, this Chapter illustrates that the Huber norm with the L-BFGS solver is robust to outliers, thus providing an alternative to the ℓ^1 norm.

Adaptive subtraction of multiples

Chapter 3 illustrates on a multiple attenuation problem how the Huber norm can help to better separate primaries from multiples. Adaptive subtraction of multiples is a two steps process where multiples are first predicted, leading to an accurate model of the noise, and then adaptively subtracted from the data by estimating matching filters. When weak multiples are present in the surroundings of strong primaries, multiples can be matched to primaries when the ℓ^2 norm is utilized for the filter estimation step. In this Chapter, I prove that the Huber

norm provides a robust measure for computing filter coefficients that is less sensitive to the relative strengths of both multiples and primaries, thus preserving the signal better than the ℓ^2 norm.

Noise attenuation by filtering and modeling

Chapter 4 tackles the noise problem generated by modeling uncertainties. Using the mathematical framework of least-squares inversion, it is theoretically possible to eliminate the noise present in the data by estimating the data covariance operator (Tarantola, 1987). The data covariance operator can be seen as a filtering operator that operates in the data space. Building on Nemeth (1996), it is also possible to separate noise and signal by incorporating a noise modeling operator within the inversion. In both techniques, the residual components become independent and identically distributed (IID). In Chapter 4, I review both approaches and show that multidimensional prediction-error filters (PEFs) can approximate covariance and modeling operators. In addition, I unravel the mathematical links between the two approaches and establish that the noise modeling technique can be algebraically reduced to a filtering method with projection operators.

Interpolation of bathymetry data

Chapter 5 introduces bathymetry data from the Sea of Galilee. The goal is to produce a depth map of the sea bottom. Because this dataset is contaminated with spurious noise and systematic errors (from different weather conditions), both the Huber norm (Chapter 2) and the modeling approach (Chapter 4) are used to produce an artifact-free image of the Sea of Galilee. This dataset illustrates that a careful tuning of any inversion scheme, with appropriate operators, can ultimately produce a result that is robust to the noise level present in the data.

Multiple attenuation: Theory and Practice

Chapter 6 expands the filtering technique of Chapter 4 to the problem of multiple attenuation with 2-D data. The basic idea is to separate multiples from primaries according to their pattern,

or multivariate spectra. To do so, I approximate covariance operators for both the data and the model with multidimensional prediction-error filters. These PEFs can be seen as proxies for the primaries and multiples. This method is a two steps technique. First, the PEFs for the multiples and the primaries are estimated. The noise PEFs (multiples) are estimated from a multiple model, generally derived by convolving shot gathers (Verschuur et al., 1992). To estimate the signal PEFs, I introduce the Spitz approximation (Spitz, 1999). Then, the multiples and primaries are separated in a least-squares sense with the PEFs for covariance operators. One important result of this Chapter is that 3-D filters lead to a better multiple removal than 2-D filters. Comparing with the adaptive subtraction technique of Chapter 3, I also show that the pattern-based technique is less sensitive to modeling errors present in the multiple model.

Multiple attenuation: A 3-D field data example

Chapter 7 introduces a Gulf of Mexico 3-D dataset donated by the Compagnie Générale de Géophysique (CGG), where multiples are attenuated with the pattern-based technique of Chapter 6. With 3-D data, obtaining an accurate model of the multiples becomes the biggest challenge because the acquisition geometry is not regular. Therefore, the method of Verschuur et al. (1992) cannot be directly applied on the data without interpolation. Worst, the data need to be interpolated and extrapolated on a regular and fine mesh in both inline and crossline directions. With 3-D data, the cost of such processes can be prohibitive and alternatives are needed. In this Chapter, I propose doing the interpolation in the inline direction only and model the multiples in a 2-D sense, one streamer at a time. The shots are interpolated with a radon-based technique on common-mid point (CMP) gathers. I show that the resulting noise model from the 2-D prediction, although not perfect, still encapsulates the main features of the true multiples present in the data. Because the pattern-based method is robust to modeling uncertainties, the primaries are well recovered regardless of the approximations made during the model building.

Chapter 2

Robust inversion using the Huber norm

SUMMARY

The “Huber function” (or “Huber norm”) is one of several robust error measures which interpolates between smooth (ℓ^2) treatment of small residuals and robust (ℓ^1) treatment of large residuals. Since the Huber function is differentiable, it may be minimized reliably with a standard gradient-based optimizer. I propose to minimize the Huber function with a quasi-Newton method that has the potential of being faster and more robust than conjugate-gradient when solving non-linear problems. Tests with a linear inverse problem for velocity analysis with both synthetic and field data suggest that the Huber function gives far more robust model estimates than does least-squares with or without damping.

INTRODUCTION

Robust error measures such as the ℓ^1 norm have found a number of applications in geophysics. As measures of data misfit, they show considerably less sensitivity to large measurement errors than least-squares (ℓ^2) measures. Since geophysical inverse problems are generally ill-posed, relatively noise insensitive misfit measures can yield far more stable estimates of Earth parameters than does ℓ^2 norm (Claerbout and Muir, 1973; Taylor et al., 1979; Chapman and

Barrodale, 1983; Scales and Gersztenkorn, 1987; Scales et al., 1988). This insensitivity to large noise has a statistical interpretation: robust measures are related to long-tailed density functions in the same way that ℓ^2 is related to the short-tailed gaussian density function (Tarranto, 1987).

A simple choice of robust measure is the ℓ^1 norm: denoting the residual (misfit) components by $r_i, i = 1, \dots, N$ (N being the number of data points), the ℓ^1 norm misfit function of the residual vector is $\sum_{i=1}^N |r_i|$. This function is not smooth: it is singular where any residual component vanishes. As a result, numerical minimization is difficult. Various approaches based for example on a linear programming viewpoint (Barrodale and Roberts, 1980) or iterative smoothing (Scales et al., 1988), have been used with success but require considerable tuning. Moreover, the singularity implies that small residuals are “taken as seriously” as large residuals, which may not be appropriate in all circumstances.

These drawbacks of the ℓ^1 norm have led to various proposals which combine robust treatment of large residuals with gaussian treatment of small residuals. These proposals are known as “hybrid ℓ^1/ℓ^2 ” methods. For example, Bube and Langan (1997) apply an iteratively reweighted least-squares (IRLS) method to minimize a hybrid objective function on a tomography problem. More recently, Zhang et al. (2000) use an IRLS procedure to locate bed boundaries from electromagnetic data.

In this chapter, a hybrid ℓ^1/ℓ^2 error measure (or norm) proposed by Huber (1973) is presented:

$$M_\alpha(r) = \begin{cases} \frac{r^2}{2\alpha} & 0 \leq |r| \leq \alpha, \\ |r| - \frac{\alpha}{2} & \alpha < |r|, \end{cases} \quad (2.1)$$

where α is the threshold between the ℓ^1 and ℓ^2 norm. The function $\sum_{i=1}^N M_\alpha(r_i)$ is called the “Huber misfit function”, or Huber function for short. Figure 2.1 shows the Huber norm as a function of the residual. It is smooth near zero residual, weights small residuals by mean square and treats large residuals with ℓ^1 . Because it is differentiable everywhere, it is reasonable to suppose that the Huber function is easier to minimize than ℓ^1 while still robust against large residuals. Definition of the misfit *via* the Huber function results in a nonlinear

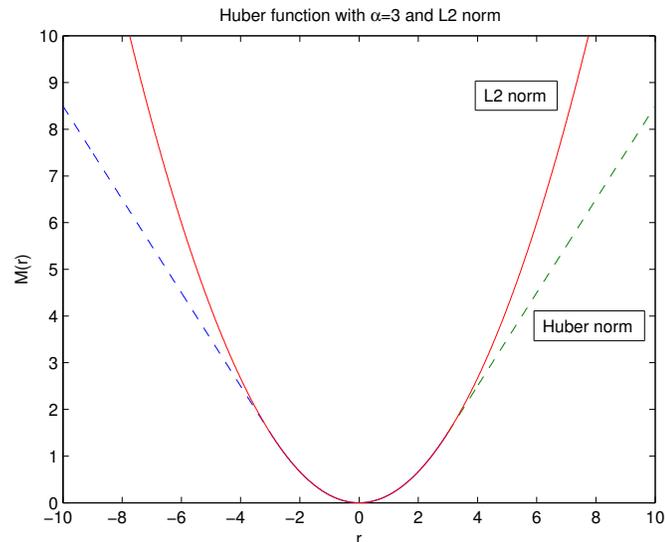


Figure 2.1: Error measure with $\alpha = 3$ proposed by Huber (Huber, 1973) (dash line) and the ℓ^2 norm (solid line). `huber-huber` [NR]

optimization problem because any residual component r_i close to the threshold α can oscillate between the ℓ^1 and ℓ^2 norm. In the first section following this introduction, I propose solving the optimization problem with a quasi-Newton method called limited-memory BFGS (Broyden, 1969; Fletcher, 1970; Goldfarb, 1970; Shanno, 1970; Nocedal, 1980). In the second section, this method is tested to estimate the root mean square (rms) velocity (or slowness) of noisy common midpoint (CMP) and shot gathers for synthetic and field seismic data. There the Huber norm is compared with the ℓ^2 norm with and without regularization.

IMPLEMENTATION OF A NONLINEAR ALGORITHM

In this section, I propose minimizing any inverse problem involving the Huber norm with a quasi-Newton method. First consider the linear system $\mathbf{A}\mathbf{m} = \mathbf{d}$ where \mathbf{m} is a vector of model parameters to be estimated, \mathbf{d} a vector of observed data and \mathbf{A} a matrix or a seismic operator.

The unknown model vector \mathbf{m} is estimated such that

$$f(\mathbf{m}) = \|\mathbf{A}\mathbf{m} - \mathbf{d}\|_{Huber} = \|\mathbf{r}\|_{Huber} = \sum_{i=1}^N M_{\alpha}(r_i) \quad (2.2)$$

is minimum. Because of the definition in equation (2.1), although \mathbf{A} is linear, the minimization of $f(\mathbf{m})$ in equation (2.2) is not a linear problem: a particular point in the residual can oscillate between the ℓ^1 and ℓ^2 norm for different iterations. This difficulty can be overcome by designing a nonlinear solver that will converge to a minimum of $f(\mathbf{m})$. Some specially adapted Huber minimizers have been suggested (Ekblom and Madsen, 1989; Li and Swetist, 1998). One of the questions for this study was whether a standard non-linear method, as opposed to a special solver, would perform satisfactorily in Huber estimation. I show that it does. In addition, these specially adapted minimizers work only when \mathbf{A} is a matrix and not an operator.

In this Chapter, a quasi-Newton method is chosen. The quasi-Newton method is an iterative process where the solution to the problem is updated as follows:

$$\mathbf{m}_{k+1} = \mathbf{m}_k - \lambda_k \mathbf{H}_k^{-1} \nabla f(\mathbf{m}_k), \quad (2.3)$$

where \mathbf{m}_{k+1} is the updated solution at iteration $k + 1$, λ_k the step length computed by a line search that ensures a sufficient decrease of $f(\mathbf{m})$ and \mathbf{H}_k an approximation of the Hessian (or second derivative.)

Quasi-Newton methods intend to approximate the Hessian without explicitly computing it. The Hessian is then re-estimated and improved at each iteration of the descent method (Gill et al., 1986). Quasi-newton methods can be equivalent to conjugate-gradient on a quadratic function and they are often more robust than conjugate-gradient on general nonlinear problems. They also tend to require less iterations (Gill et al., 1986). However, one major drawback of the quasi-Newton methods is that they require the storage of the approximate Hessian in memory, which can be troublesome for large scale problems.

In Appendix A, a computationally effective method that does not require the storage of the Hessian is presented. This method is called limited-memory BFGS (L-BFGS) (Nocedal,

1980). With this technique, only a limited number of solution step and gradient step vectors are stored as the iterations go on. For each new iteration, these vectors are combined to form the approximate Hessian (see Appendix A for details). In the case where all the solution step and gradient step vectors are kept in memory at every iteration, the L-BFGS method becomes equivalent to the BFGS update (Nocedal, 1980). With the L-BFGS method, the storage needed is reduced compare to BFGS and makes it affordable to use for image estimation.

One computational burden is the line search algorithm that ensures sufficient decrease of the misfit function. This comes from the need to evaluate the misfit function many times before finding a good step length λ_k . The value $\lambda_k = 1$ is tested first (Liu and Nocedal, 1989), thus saving substantial computational time.

Another source of improvement is by scaling the Hessian at each iteration (Liu and Nocedal, 1989). This scaling greatly improves the performances of the quasi-Newton method. Liu and Nocedal (1989) show on numerical examples for large-scale problems that the L-BFGS method with the scaling of the Hessian and an appropriate line search algorithm (see Appendix A for details) is usually faster than conjugate-gradient using the Polak-Ribière formula (Kelley, 1999).

One problem with using quasi-Newton method, however, is that the Huber function is not twice continuously differentiable. This assumption is at the heart of the convergence properties of the L-BFGS method. Nonetheless, the L-BFGS update requires the computation of the gradient only (see Appendix A). Furthermore, given that the approximate Hessian is not an exact representation of the real one, the violation to this initial condition are expected to be mild. Examples in the next section show that the method is never-the-less giving satisfying results and is robust to outliers, as expected.

In this section, I have proposed an efficient algorithm that will minimize any misfit function using the Huber norm. This algorithm is a limited-memory BFGS technique that saves computational time and memory requirement by (1) limiting the number of vectors kept in memory for the update of the Hessian \mathbf{H}_k , (2) testing a default value for the step length λ_k , and (3) scaling the Hessian at each iteration. In the next section, this algorithm is tested on a

geophysical problem and compare the Huber norm with the ℓ^2 norm with and without regularization.

APPLICATION OF THE HUBER NORM: VELOCITY ESTIMATION WITH NOISY DATA

In this section, the proposed algorithm is used on a geophysical inverse problem which is velocity estimation with noisy data. Some possible applications of a robust solver are, for example, tomography (Bube and Langan, 1997) and deconvolution of noisy data (Chapman and Barrodale, 1983). My goal is to demonstrate that the Huber function with the L-BFGS method gives a robust estimate of the model parameters when outliers are present in the data. The velocity estimation problem with the Huber norm has potential applications when multiples need to be separated from the signal in the velocity domain (Lumley et al., 1995; Kostov and Nichols, 1995). More conventional multiple attenuation techniques using the parabolic radon transform (Kabir and Marfurt, 1999; Herrmann et al., 2000) can also benefit from using the Huber norm.

The “velocity domain” representation of seismic data using the hyperbolic radon transform (HRT) is an alternative to the standard common midpoint (CMP) gather. Transformation of CMP data into the velocity domain (producing a velocity model or panel of the data) exhibits clearly the moveout inherent in the data and therefore, forms a convenient basis for velocity analysis as a linear inverse problem.

Thorson and Claerbout (1985) were the first to define the forward and adjoint operators of the HRT, formulating it as an inverse problem in which the velocity domain is the unknown space. In their approach the forward operator \mathbf{L} maps the model space (velocity domain) into the data space (CMP gathers). This transformation is a superposition of hyperbolas in the data space. The adjoint operator \mathbf{L}' , the HRT, maps the data space into the model space. This transformation is a summation over hyperbolic trajectories in the data space (related to the velocity stack as defined by Taner and Koehler (1969)). With $d(t, x)$ being a CMP gather and

$m(\tau, s)$ the corresponding velocity model, the forward operation is

$$d(t, x) = \sum_{s=s_{min}}^{s_{max}} w_o m(\tau = \sqrt{t^2 - s^2 x^2}, s), \quad (2.4)$$

and the adjoint transformation

$$m(\tau, s) = \sum_{x=x_{min}}^{x_{max}} w_o d(t = \sqrt{\tau^2 + s^2 x^2}, x), \quad (2.5)$$

where x is the offset (x_{min} and x_{max} being the offset range), s the slowness (s_{min} and s_{max} being the range of slownesses investigated), τ the two-way zero offset travel time, and w_o a weighting function that compensates to some extent for geometrical spreading and other effects (Claerbout and Black, 1997).

Having defined the forward operator \mathbf{L} and its adjoint \mathbf{L}' , the inverse problem can be posed. Inverse theory helps us to find a velocity panel which synthesizes a given CMP gather *via* the operator \mathbf{L} . In equations, given data \mathbf{d} (CMP gather), we want to solve for the model \mathbf{m} (velocity panel)

$$\mathbf{Lm} = \mathbf{d}, \quad (2.6)$$

which leads in a least-squares sense to the linear system (“normal equations”)

$$\mathbf{L}'\mathbf{Lm} = \mathbf{L}'\mathbf{d}. \quad (2.7)$$

This system is easy to solve if $\mathbf{L}'\mathbf{L} \approx \mathbf{I}$, i.e., if \mathbf{L} is close to unitary. Unfortunately, \mathbf{L} is far from an unitary operator (Sacchi and Ulrych, 1995; Kabir and Marfurt, 1999). In addition, the number of equations and unknowns may be large, making an iterative data-fitting approach reasonable.

Consequently, with E being a misfit measurement function, our goal is to iteratively calculate the model \mathbf{m} that minimizes the misfit function

$$f(\mathbf{m}) = E(\mathbf{Lm} - \mathbf{d}). \quad (2.8)$$

One possibility for E is the ℓ^2 norm (least-squares inversion). The misfit function is then usually minimized with conjugate-gradient. Another possible approach is of course by taking the Huber norm along with the L-BFGS method introduced in the preceding section. The two norms are compared in the next section for velocity estimation problems. The results show that the Huber norm gives the expected ℓ^1 behavior when outliers (non-gaussian noise) are present in the data.

Synthetic data results

Figure 2.2 displays the synthetic model. In Figure 2.2a I show the ideal velocity space with five events at different slownesses. In Figure 2.2b the five corresponding hyperbolas in the CMP domain are shown. Finally in Figure 2.2c, four spikes are added to the CMP gather in Figure 2.2b to make gaussian statistics unsuitable. The energy of the four spikes is five times the energy of the five hyperbolas. Our goal is to find the velocity field \mathbf{m} that will best fit a CMP gather \mathbf{d} via the HRT.

Figure 2.3 shows the result of the inversion when the ℓ^2 norm is used for the noise-free data. On the left the velocity space is displayed with its five events. The focusing is not perfect and some artifacts appear (Sacchi and Ulrych, 1995). The middle panel shows the remodeled data after inversion and the right panel displays the residual (difference between the input and remodeled data). I conclude that the inversion reached a minimum since no coherent energy is left in the residual: the data fitting is very good. Because the input data have gaussian statistics, the performance of the least-squares inversion was expected. I now use the same inversion scheme with the ℓ^2 norm but with the “contaminated” CMP gather (Figure 2.2c). Notice that no regularization was applied with the least-squares method. The final result is shown in Figure 2.4. In this case, as expected, the ℓ^2 inversion creates a number of artifacts both in the model and data space. In the model space (Figure 2.4a), the spikes are

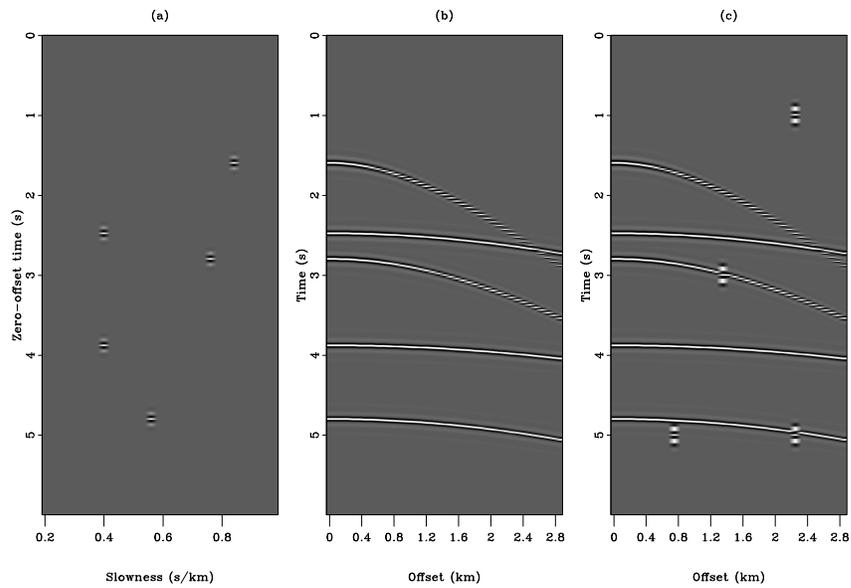


Figure 2.2: Synthetic data used for the inversion. (a) The true velocity model represented in slowness. (b) The noise-free input data used for the inversion. (c) The same data but with noise added (four spikes). `huber-model` [ER]

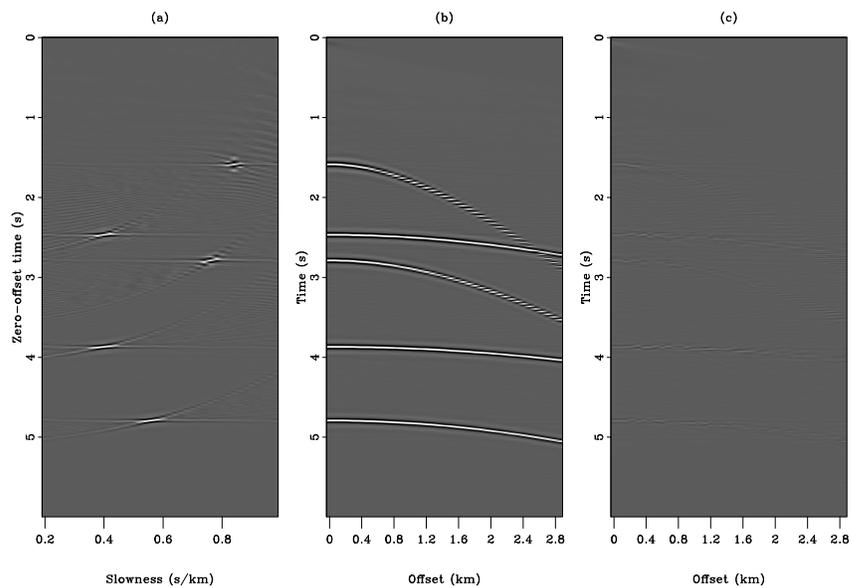


Figure 2.3: Result of the inversion with the ℓ^2 norm for the noise-free data. (a) Inverted slowness field. (b) Remodeled data. (c) Difference between the input (Figure 2.2b) and remodeled data. `huber-res-nospike-HUBER` [ER]

mapped into curves whereas the hyperbola are mapped into nearly focused points. In the data space, after remodeling (Figure 2.4b), spikes are not remodeled correctly and are smearing on the neighboring traces. In Figure 2.5 I use least-squares with a simple damping in the regularization. The model and data space are cleaner but the difference between the input and the remodeled data or residual is still large (comparing Figure 2.5c to Figure 2.3c.) In addition, artifacts in the inverted slowness field and the reconstructed data can be seen.

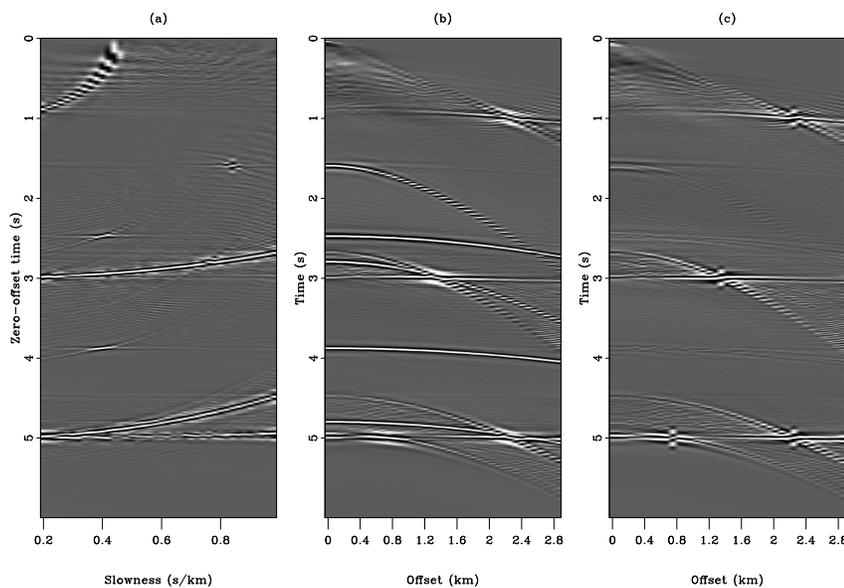


Figure 2.4: Result of the inversion with the ℓ^2 norm for the data with noise. (a) Inverted slowness field. (b) Remodeled data. (c) Difference between the input (Figure 2.2c) and remodeled data. The four spikes create artifacts in both inverted model and reconstructed data space.

`huber-res-spike-L2-HUBER` [ER]

Figure 2.6 displays the result of the inversion with the Huber norm. The outcome of the inversion is insensitive to the spiky events, like a pure ℓ^1 norm misfit function. The residual (Figure 2.6c) exhibits the four spikes very clearly. This result demonstrates that the proposed algorithm, although not specifically designed to minimize the Huber function, converges to a satisfying solution. The next section shows inversion results with noisy field data.

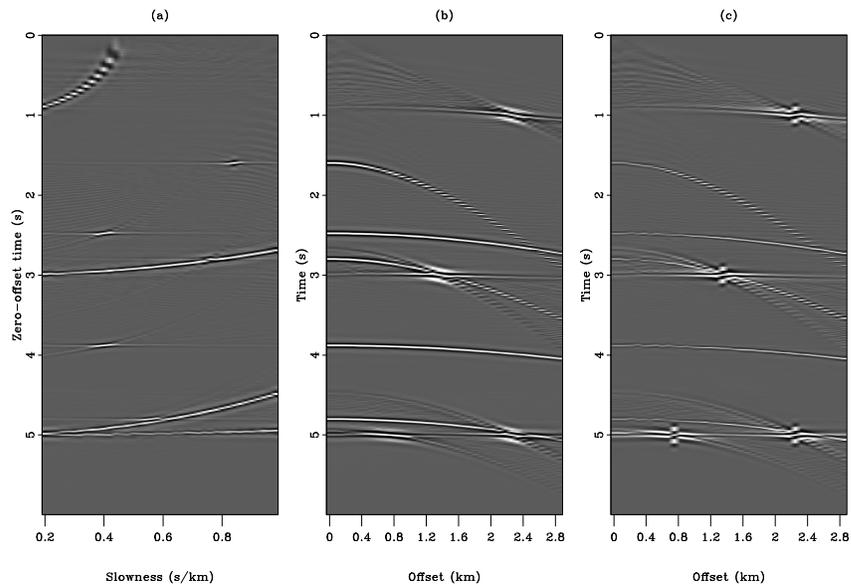


Figure 2.5: Result of the inversion with the ℓ^2 norm with damping for the data with noise. (a) Inverted slowness field. (b) Remodeled data. (c) Difference between the input (Figure 2.2c) and remodeled data. The three panels are cleaner than in Figure 2.4 but some artifacts remain, however. `huber-res-spike-L2-reg-HUBER` [ER]

Field data results

The proposed algorithm is now tested on a field data example. For this purpose, a shot gather from a land-data survey in the Middle East is selected. The trajectories of the events in Figure 2.7 look “hyperbolic” enough to be inverted with our method. Note that in theory, the data should be sorted into CMP gathers before doing the inversion. This dataset is particularly interesting because it has amplitude anomalies at short offset and a low velocity coherent noise that is probably due to guided energy in the near-surface. I could attenuate the amplitude anomalies by applying an Automatic Gain Control (AGC) on the data before inversion. A better approach is by weighting the residual with a function mimicking AGC.

This gather is first inverted with the ℓ^2 norm without regularization (Figure 2.8). The left panel displays the velocity domain obtained after the least-squares inversion. The main velocity event is masked with horizontal stripes coming from the short-offset amplitude anomalies. The reconstructed data (middle panel) show spurious noise at large offset and other inversion

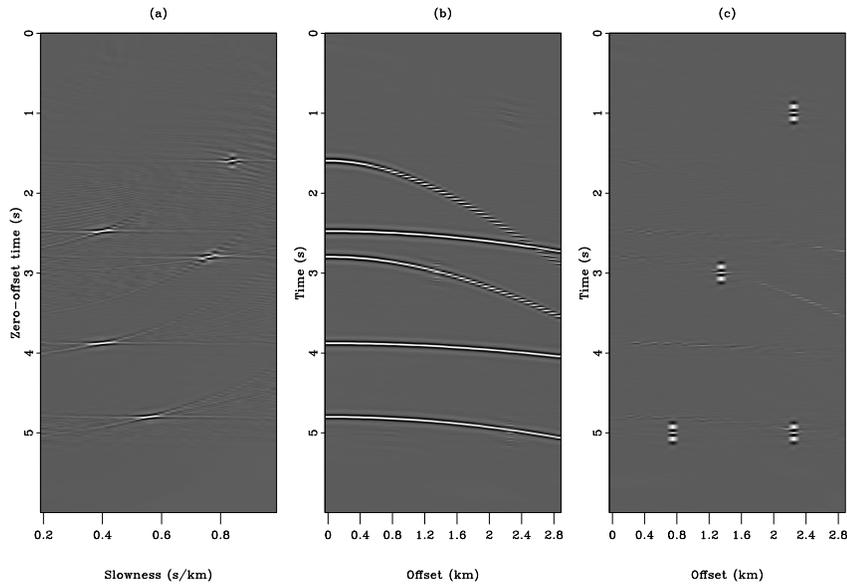
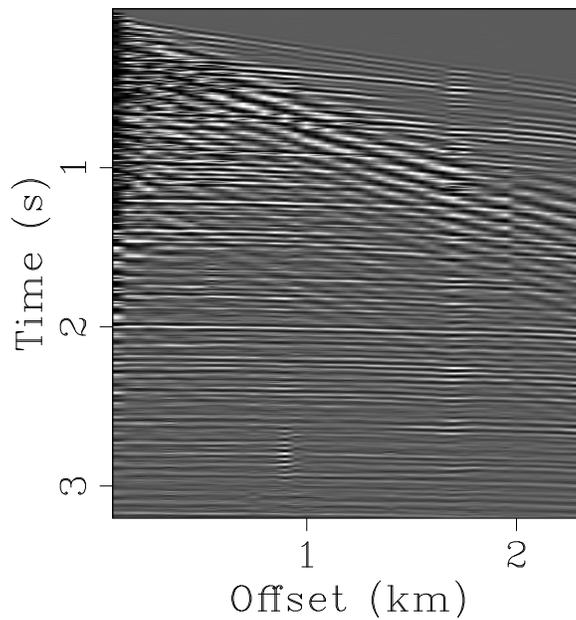


Figure 2.6: Result of the robust inversion with the Huber norm for the data with noise. (a) Inverted slowness field. (b) Remodeled data. (c) Difference between the input (Figure 2.2c) and remodeled data. The Huber norm behaves like a pure ℓ^1 norm since all artifacts have disappeared. `huber-res-spike-HUBER` [ER]

Figure 2.7: The field data used for the inversion. Notice the amplitude anomalies at near offset and the time shift near offset 2 km. `huber-model2` [ER]



artifacts. I now show in Figure 2.9 the result of the damped least-squares. The inverted slowness field is much cleaner, but the horizontal stripes remain. We also have the same velocity from the top to the bottom of Figure 2.9a. This shows that the data are contaminated with multiples generated in the near surface. Figure 2.10 displays the inversion result with the Huber norm and demonstrates the robustness of this method. A very well focused velocity corridor is obtained as opposed to the ℓ^2 result in Figure 2.8a. In addition, the horizontal stripes have disappeared.

The synthetic and field data examples demonstrate that the Huber norm can be an efficient alternative to the ℓ^2 norm when outliers (or non-gaussian noise) are present in the data.

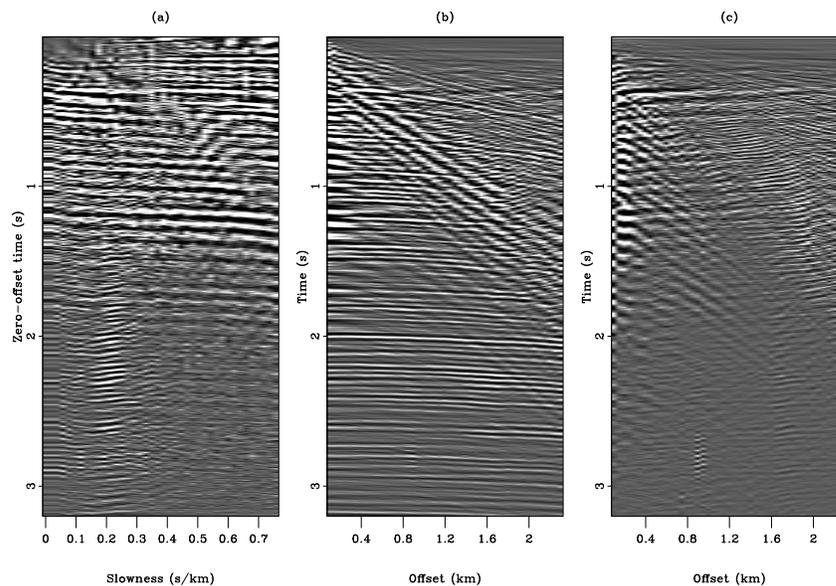


Figure 2.8: The result of the inversion with the ℓ^2 norm for the field data. (a) Inverted slowness field. (b) Remodeled data. (c) Difference between the input (Figure 2.7) and remodeled data. The horizontal stripes in the velocity panel are created by the amplitude anomalies at short offsets. `huber-res-wz08-L2-HUBER` [ER]

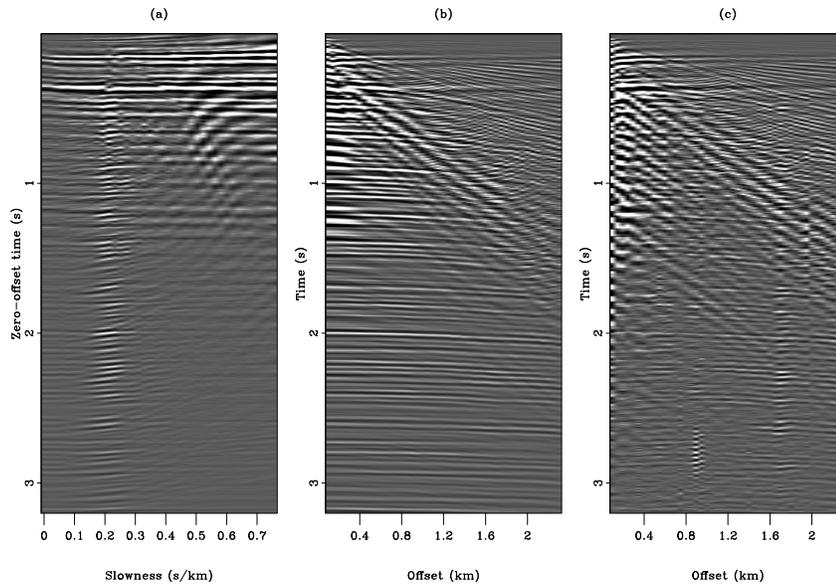


Figure 2.9: The result of the inversion with the ℓ^2 norm and regularization for the field data. (a) Inverted slowness field. (b) Remodeled data. (c) Difference between the input (Figure 2.7) and remodeled data. The model is much cleaner than in Figure 2.8a, but the horizontal events remain. `huber-res-wz08-L2-reg-HUBER` [ER]

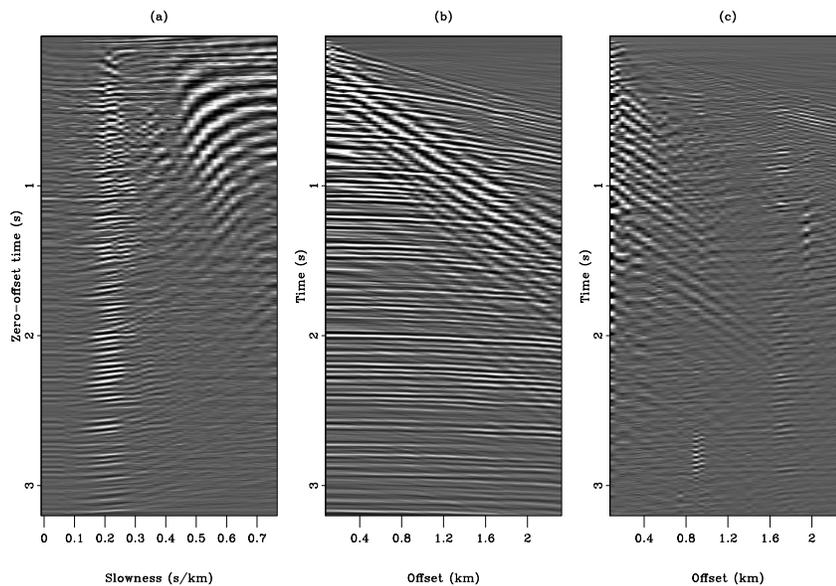


Figure 2.10: The result of the robust inversion with the Huber norm for the field data. (a) Inverted slowness field. (b) Remodeled data. (c) Difference between the input (Figure 2.7) and remodeled data. `huber-res-wz08-0.082-HUBER` [ER]

CONCLUSION

Since geophysical inverse problems are often ill-posed due to the presence of inconsistent data, high amplitude anomalies and outliers, relative insensitivity to noise is a desirable characteristic of an inversion method. The Huber function is a compromise misfit measure between the ℓ^1 and ℓ^2 norm. It not only improves robustness in the presence of noise and outliers with a ℓ^1 measure, but also keeps smoothness for small residuals with a ℓ^2 measure.

In this Chapter I have proposed minimizing the Huber function with a quasi-Newton method called limited-memory BFGS. This method has the potential of being faster and more robust than conjugate-gradient for solving non-linear problems. Tests with noisy synthetic and field data examples demonstrate that our method is robust to outliers present in the data space, as expected.

The possible applications of the Huber norm are endless in geophysics. I illustrate in Chapter 5 how the Huber norm helps removing spikes in bathymetry data. In addition, I show in Chapter 3 how the Huber norm helps separating primaries and multiple reflections better.

ACKNOWLEDGMENTS

This work was started while William Symes from Rice University was on sabbatical at Stanford.

Chapter 3

Adaptive subtraction of multiples with the ℓ^1 norm

SUMMARY

A strategy for multiple removal consists in estimating a model of the multiples and then adaptively subtracting this model to the data by estimating shaping filters. A possible and efficient way of computing these filters is by minimizing the difference or misfit between the input data and the filtered multiples in a least-squares sense. Therefore, the signal is assumed to have minimum energy and to be orthogonal to the noise. Some problems arise when these conditions are not met. For instance, for strong primaries with weak multiples, the multiple model might be matched to the signal (primaries) and not to the noise (multiples). Consequently, when the signal does not exhibit minimum energy, I propose using the ℓ^1 norm as opposed to the ℓ^2 norm for the filter estimation step. This choice comes from the well-known fact that the ℓ^1 norm is robust to “large” amplitude differences when measuring data misfit. The ℓ^1 norm is approximated with the Huber norm (Chapter 2) minimized with a quasi-Newton method. This technique is an excellent approximation to the ℓ^1 norm. I illustrate this method with synthetic and field data where internal multiples are attenuated. I show that the ℓ^1 norm leads to a much improved attenuation of the multiples when the minimum energy assumption is violated. In particular, the multiple model is fitted to the multiples in the data only, while preserving the

primaries.

INTRODUCTION

A classical approach for attenuating multiples consists of building a multiple model, e.g., as described by Verschuur et al., (1992) or Berkhout and Verschuur (1997), and adaptively subtracting this model from the multiple infested-data by estimating shaping filters (Dragoet, 1995; Liu et al., 2000; Rickett et al., 2001). The estimation of the shaping filters is usually done in a least-squares sense making these filters relatively easy to compute. Implicitly, by using the ℓ^2 norm, the resulting signal, after the filter estimation step, is assumed to be orthogonal to the noise and has minimum energy. These assumptions might not hold and other methods, such as pattern-based approaches (Spitz, 1999; Guitton et al., 2001) have been proposed to circumvent these limitations. For instance, when a strong primary is surrounded by weaker multiples, the multiple model will match the noise (multiples) as well as the signal (primaries) such that the difference between the data and the filtered multiple model is minimum in a least-squares sense. Consequently, some primary energy might leak in the estimated multiples and *vice versa*. Therefore, a new criterion or norm for the filter estimation step is needed.

In this paper, I propose estimating the shaping filters with the ℓ^1 norm instead of the ℓ^2 norm, thus relaxing the need for the signal to have minimum energy. This choice is driven by the simple fact that the ℓ^1 norm is robust to “outliers” (Claerbout and Muir, 1973) and large amplitude anomalies. Because the ℓ^1 norm is singular where any residual component vanishes, the Huber norm with L-BFGS solver of Chapter 2 is chosen instead. This method gives an excellent approximation of the ℓ^1 norm.

In this first section following the introduction, some limitations of the least-squares criterion on a simple 1-D problem are illustrated. I then introduce the proposed approach based on the ℓ^1 norm to improve the multiple attenuation results. In a second synthetic example, internal multiples are attenuated with the ℓ^2 and ℓ^1 norms. Finally, I utilize shaping filters on a multiple contaminated gather from a seismic survey showing that the ℓ^1 norm leads to a substantial attenuation of the multiples.

PRINCIPLES OF ℓ^1 NORM AND ℓ^2 NORM SUBTRACTION

In this section, I demonstrate with a 1-D example that the attenuation of multiples with least-squares adaptive filtering is not effective when strong primaries are located in the neighborhood. This simple example helps to better understand the behavior of the adaptive scheme in more complicated cases.

Shaping filters and the ℓ^2 norm

This section illustrates some limitations of the ℓ^2 norm for the estimation of shaping filters. In figure 3.1, a simple 1-D problem is shown. On the top we have four events corresponding to one primary (on the left) and three multiples (on the right). Note that the primary has higher amplitude than the multiples. On the bottom we show a multiple model that exactly corresponds to the real multiples. Our goal is to estimate one shaping filter \mathbf{f} that minimizes the objective function

$$e_2(\mathbf{f}) = \|\mathbf{M}\mathbf{f} - \mathbf{d}\|^2, \quad (3.1)$$

where \mathbf{M} is the matrix representing the convolution with the time series for the multiple model (Figure 3.1b) and \mathbf{d} the time series for the data (Figure 3.1a). Now, if the filter \mathbf{f} is estimated with enough degrees of freedom (enough coefficients) to minimize equation (3.1), we obtain for the estimated primaries, i.e., $\mathbf{d} - \mathbf{M}\mathbf{f}$, Figure 3.2a, and for the estimated multiples, i.e., $\mathbf{M}\mathbf{f}$, Figure 3.2b. The estimated primaries do not resemble the primary in Figure 3.1a. I show the corresponding shaping filter in Figure 3.3. This filter is not an unit spike at $lag = 0$ as expected. The problem stems from the least-squares criterion which yields an estimated signal that has, by definition, minimum energy. In this 1-D case, the total energy in the estimated signal (Figure 3.2a) is $e_2 = 2.4$, which is less than the total energy of the primary alone ($e_2 = 4$). This is a fundamental problem if we use the ℓ^2 norm to estimate the shaping filter. In the next section, I show that the ℓ^1 norm should be used if amplitude differences exist between primaries and multiples.

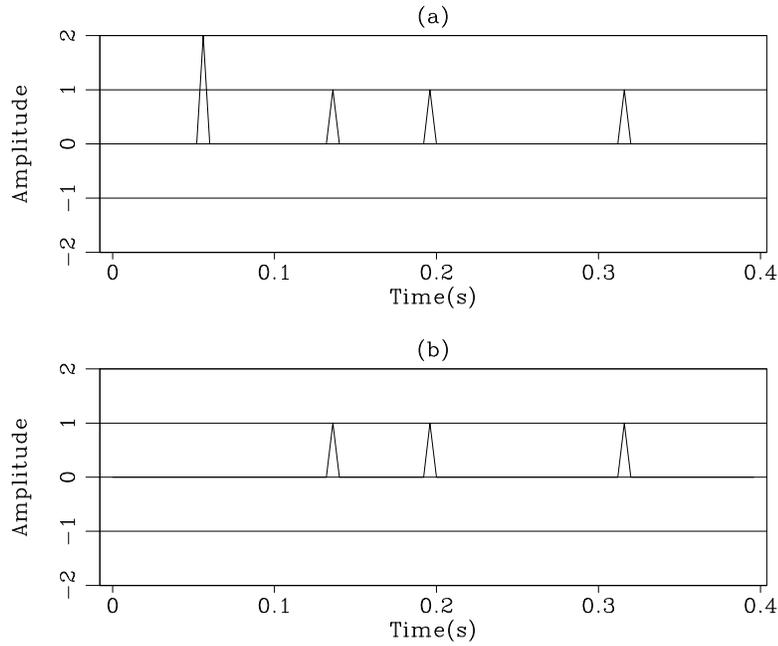


Figure 3.1: (a) The data with one primary on the left, and three multiples on the right. (b) The multiple model that we want to adaptively subtract from (a). `adaptive-datmul` [ER]

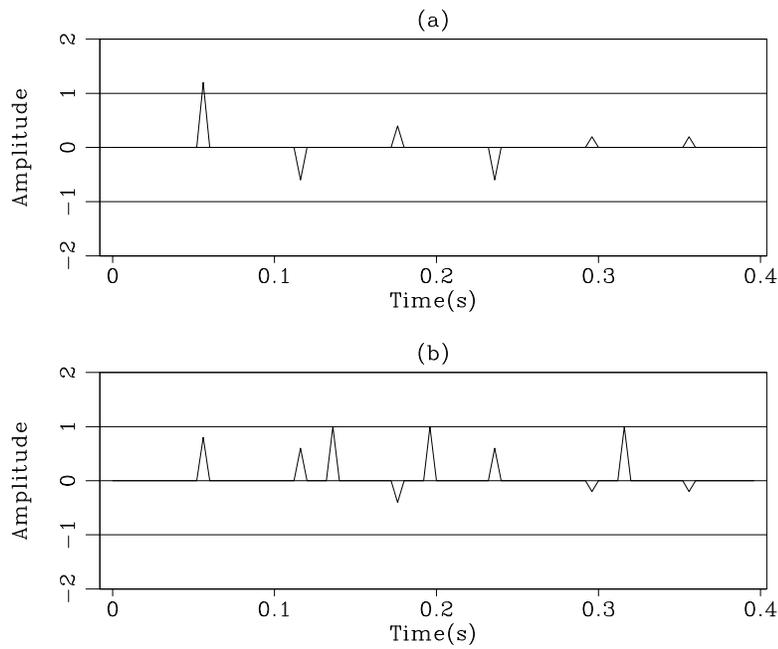


Figure 3.2: (a) The signal estimated with the ℓ^2 norm. (b) The noise estimated with the ℓ^2 norm. `adaptive-1DI2` [ER]

Shaping filters and the ℓ^1 norm

The strong primary in Figure 3.1 can be seen as an outlier that attracts much of the solver's efforts during the filter estimation. Consequently, some of the signal leaks into the noise.

Figure 3.3: Shaping filter estimated for the 1-D problem with the ℓ^2 norm. This filter is not a single spike at $lag = 0$. `adaptive-filterl2` [ER]

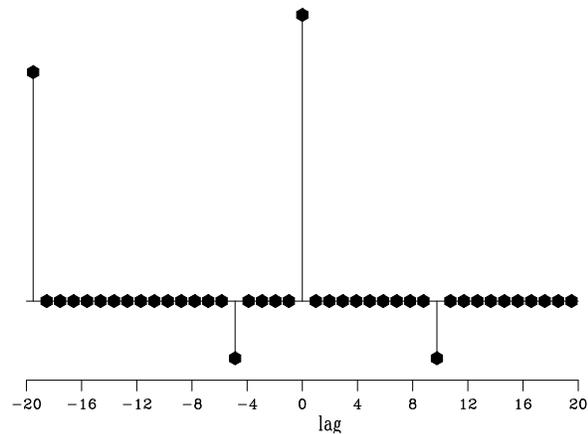
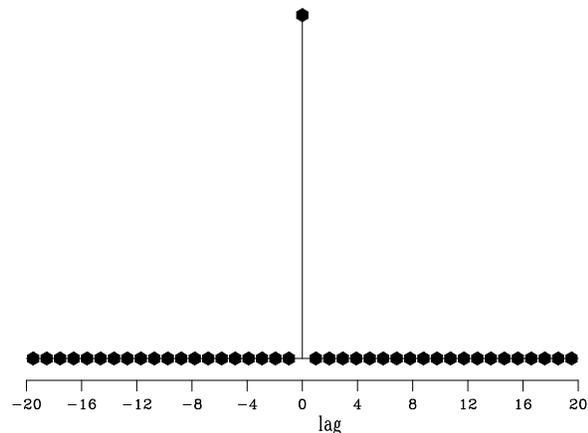


Figure 3.4: Shaping filter estimated for the 1-D problem with the ℓ^1 norm. This filter is a single spike at $lag = 0$. `adaptive-filterl1` [ER]



Because the ℓ^1 norm is robust to outliers, I propose estimating the filter coefficients with it. This insensitivity to large “noise” has a statistical interpretation: robust measures are related to long-tailed density functions in the same way that ℓ^2 is related to the short-tailed gaussian density function (Tarantola, 1987). In this section, I show that the ℓ^1 norm solves the problem highlighted in the preceding section.

Now our goal is to estimate one shaping filter \mathbf{f} that minimizes the objective function

$$e_1(\mathbf{f}) = \|\mathbf{M}\mathbf{f} - \mathbf{d}\|_{\ell^1}. \quad (3.2)$$

The function in equation (3.2) is singular where any residual component vanishes implying that the derivative of $e_1(\mathbf{f})$ is not continuous everywhere. Unfortunately, most of our optimization techniques, e.g., conjugate-gradient or Newton methods, assume that the first derivative of the objective function is continuous in order to find its minimum. Therefore, specialized techniques have been developed to either minimize or approximate the ℓ^1 norm. For instance, various approaches based on a linear programming viewpoint have been used with success, e.g., Barrodale and Roberts (1980). Another idea is to minimize a hybrid ℓ^1/ℓ^2 norm with an iteratively reweighted least-squares (IRLS) method (Gersztenkorn et al., 1986; Scales and Gersztenkorn, 1988; Bube and Langan, 1997).

Alternatively, the Huber norm is utilized. This technique yields a good approximation of the ℓ^1 norm. Therefore, the objective function we minimize becomes

$$e_1(\mathbf{f}) = \|\mathbf{M}\mathbf{f} - \mathbf{d}\|_{Huber}, \quad (3.3)$$

The minimization of the Huber norm is performed with a quasi-Newton method (Appendix A). The only parameter that needs to be set is α in equation (2.1). Consistent with the strategy of Chapter 2, I set

$$\alpha = \frac{\max|\mathbf{d}|}{100}. \quad (3.4)$$

From now on, I call ℓ^1 norm the Huber norm. In text section, I show that the ℓ^1 norm leads to the desired result for the filter estimation problem.

Results on a simple 1-D example

For the simple 1-D example under consideration, the filter coefficients are now estimated with the ℓ^1 norm approximated with the Huber function. In Figure 3.5, I display the result of the

adaptive subtraction when the ℓ^1 norm is utilized to estimate a single shaping filter (equation (3.3) with a small ϵ). The estimated signal in Figure 3.5a resembles the true signal very well,

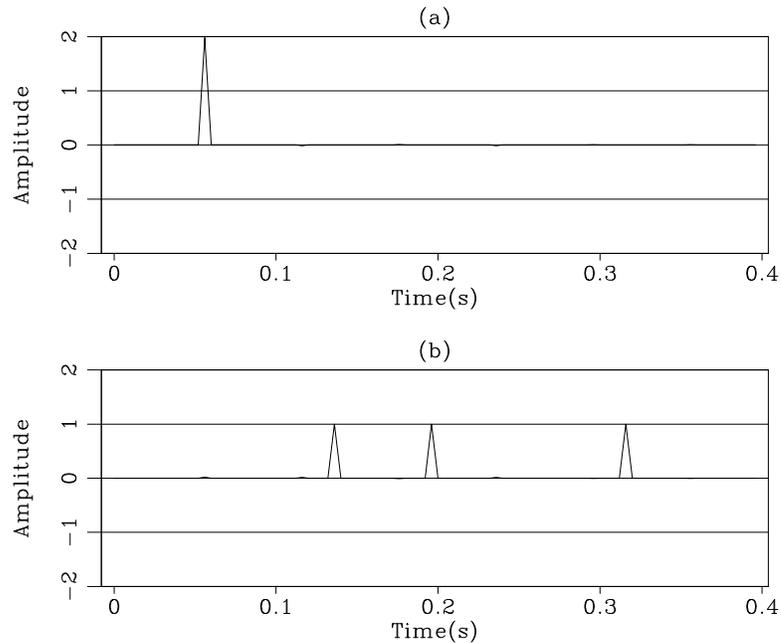


Figure 3.5: (a) The signal estimated with the ℓ^1 norm. (b) The noise estimated with the ℓ^1 norm. adaptive-1D11 [ER]

and so does the estimated noise. It is easy to check that the energy (in a ℓ^1 sense) in Figure 3.5a ($e_1 = 2$) is less than the energy (in a ℓ^1 sense) in Figure 3.2a ($e_1 = 3.2$). Figure 3.4 shows the shaping filter associated with the ℓ^1 norm. This filter is a spike at $lag = 0$. This simple 1-D example demonstrates that the ℓ^1 norm should be utilized whenever significant amplitude differences exist between multiples and primaries. In the next section, a synthetic example where internal multiples are attenuated illustrates the properties of the ℓ^1 subtraction.

2-D DATA EXAMPLE: ATTENUATION OF INTERNAL MULTIPLES

In this section I illustrate the efficiency of the ℓ^1 norm when internal multiples are attenuated in 2-D.

The synthetic data

Figure 3.6a shows a synthetic shot gather for a 1-D medium. This gather is corrupted with internal multiples only. Figure 3.6b displays the internal multiple model. For the sake of

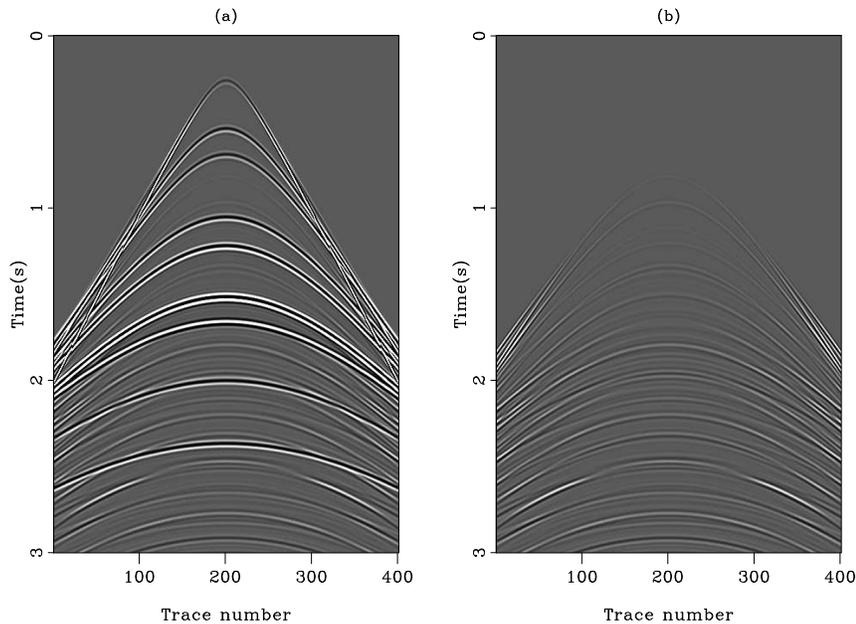


Figure 3.6: (a) A synthetic shot gather infested with internal multiples. (b) The internal multiples model obtained using the Common Focus Point technology (Berkhout and Verschuur, 1999). This model matches the internal multiples in (a). adaptive-inter [ER]

focusing only on the multiple subtraction, and not on the prediction, this internal multiple model is exact and could be directly subtracted from the data in Figure 3.6a. Note that the amplitude of the internal multiples is significantly lower than the amplitude of the primaries, making the ℓ^2 norm unsuitable for estimating the shaping filters.

Adaptive filtering with non-stationary helical filters

To handle the inherent non-stationarity of seismic data, I estimate a bank of non-stationary filters using helical boundary conditions (Mersereau and Dudgeon, 1974; Claerbout, 1998). This approach has been successfully utilized by Rickett et al. (2001) to attenuate surface-related multiples. As described before [equation (3.2)], I use the Huber norm to approximate

the ℓ^1 norm and a standard conjugate-gradient solver with the ℓ^2 norm. The filter coefficients vary smoothly across the output space by introducing a regularization term inside equation (3.3) (Crawley, 2000; Rickett et al., 2001). The misfit function to minimize becomes

$$e_1(\mathbf{f}) = |\mathbf{M}\mathbf{f} - \mathbf{d}|_{Huber} + \epsilon^2 \|\mathbf{R}\mathbf{f}\|^2, \quad (3.5)$$

where \mathbf{f} is the unknown vector of filter coefficients for the non-stationary matching filters and

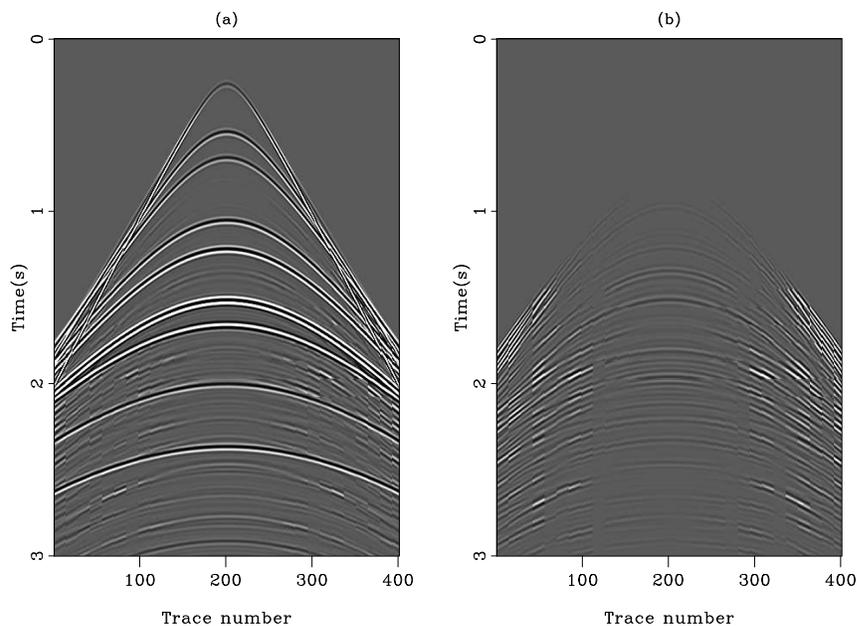


Figure 3.7: (a) The estimated primaries with the ℓ^2 norm. (b) The estimated internal multiples with the ℓ^2 norm. Ideally, (b) should look like Figure 3.6b, but it does not. adaptive-interl2 [ER]

\mathbf{R} is a regularization operator. The Helix derivative (Claerbout, 1998) is chosen for \mathbf{R} . In the following results, the non-stationary filters are 1-D. The same number of coefficients per filter are estimated with both ℓ^2 and ℓ^1 norms.

Adaptive subtraction results

Figure 3.7a shows the estimated primaries when the ℓ^2 norm is used to compute the shaping filters. Figure 3.7b displays the estimated internal multiples. As expected, because of the local

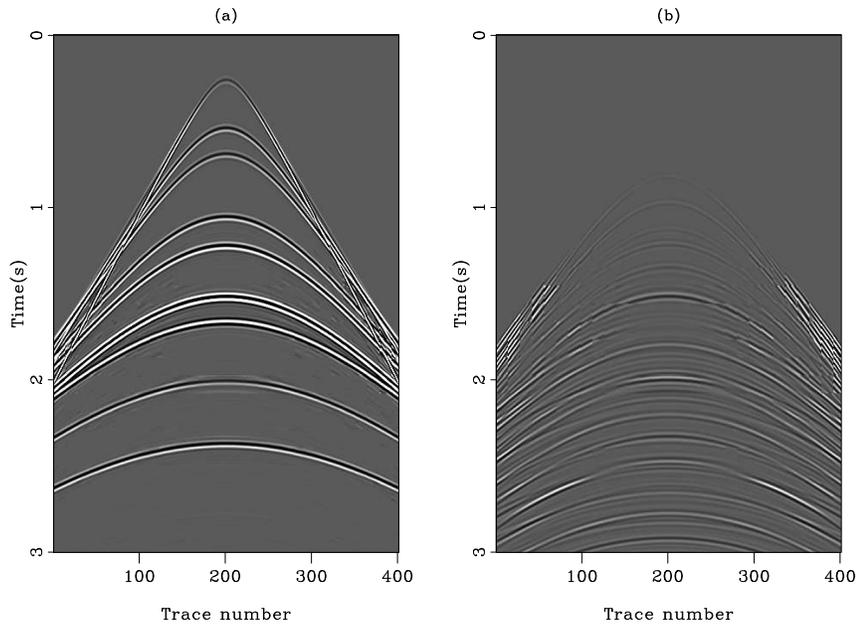


Figure 3.8: (a) The estimated primaries with the ℓ^1 norm. (b) The estimated internal multiples with the ℓ^1 norm. Beside some edge-effects, (b) resembles closely Figure 3.6b. The adaptive subtraction worked very well. `adaptive-interl1` [ER]

amplitude differences between the signal (primaries) and the noise (multiples), the adaptive subtraction fails and we retrieve the behavior explained in the preceding section with the 1-D example. Now, in Figure 3.8, we see the beneficial effects of the ℓ^1 norm. Figure 3.8a shows the estimated primaries and Figure 3.8b the estimated multiples. The noise subtracted matches very well the internal multiple model in Figure 3.6b, as anticipated. Note that with this dataset, α had to be changed to $\alpha = \max|\mathbf{d}|/1000$. As a final comparison, Figure 3.9 displays the difference between the internal-multiple model (Figure 3.6b) and the subtracted multiples with the two norms. The ℓ^1 norm (Figure 3.9b) matches the multiple model much better than the ℓ^2 norm (Figure 3.9a).

POSTSTACK LAND DATA MULTIPLE REMOVAL EXAMPLE

In this section I attenuate in the poststack domain surface-related multiples with shaping filters that I estimate with the ℓ^2 and ℓ^1 norm. These filters are non-stationary. Figure 3.10a shows the

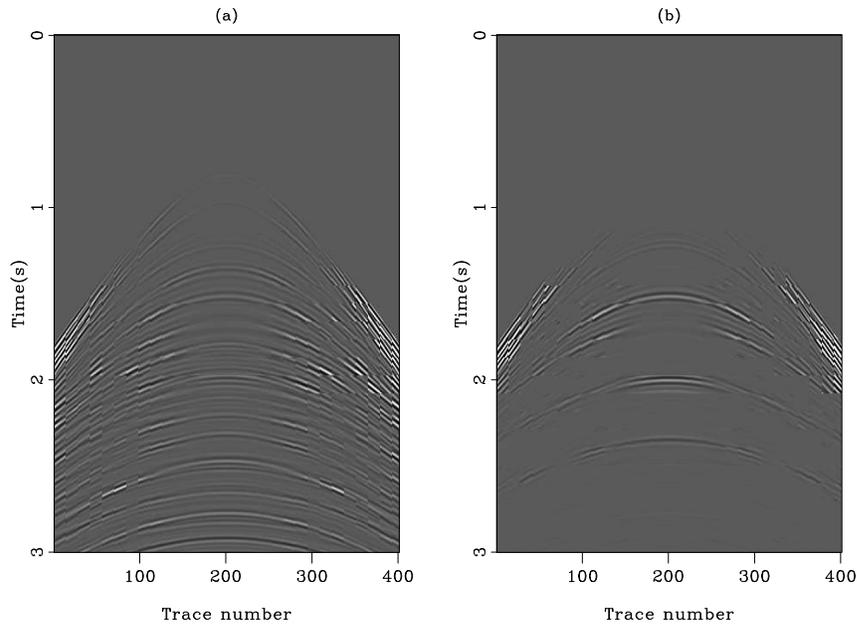


Figure 3.9: (a) The difference between the exact multiples (Figure 3.6b) and the subtracted multiples with the ℓ^2 norm. (b) The difference between the exact multiples (Figure 3.6b) and the subtracted multiples with the ℓ^1 norm. adaptive-comp-diffL1L2 [ER]

multiple-infested data. Figure 3.10b displays the multiple model computed with a data-driven modeling approach (Kelamis and Verschuur, 2000). Note that for this gather, the amplitude differences between the primaries and the multiples are not very strong. My goal is to illustrate the use of the ℓ^1 norm in a more general case when surface-related multiples are present in the data. I specifically focus on the event at 1.6s in Figure 3.10a. This event is a primary that needs to be preserved by the subtraction procedure.

The amplitude of the primary at 1.6s is well preserved with the ℓ^1 norm in Figure 3.11a. Again, $\alpha = \max|\mathbf{d}|/1000$ gave a satisfying result. However, the amplitude of this primary is attenuated with the ℓ^2 norm as displayed in Figure 3.11b. Figure 3.12 shows a comparison between the subtracted multiples with the ℓ^1 (Figure 3.12a) and the ℓ^2 norm (Figure 3.12b). I conclude that the ℓ^2 norm tends to subtract too much energy. This last example proves that the estimation of shaping filters can always be done with the ℓ^1 norm. An advantage of the inversion scheme with the Huber norm is that only one parameter (i.e., α) controls the $\ell^1 - \ell^2$ behavior. Thus it is easy to switch from one norm to another. Of course, it remains difficult

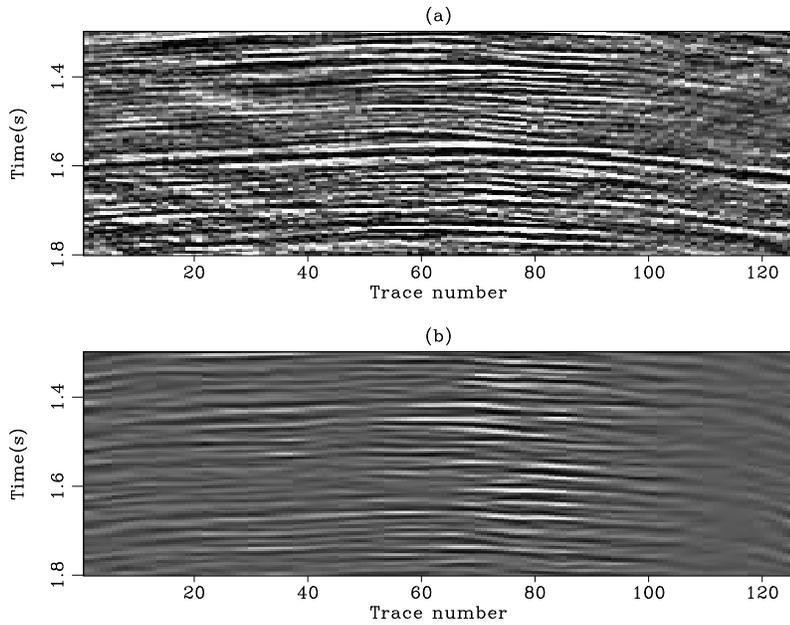


Figure 3.10: (a) Stack infested with multiples. (b) The multiple model computed with the Delft modeling approach. The subtraction is done poststack. `adaptive-win3` [ER]

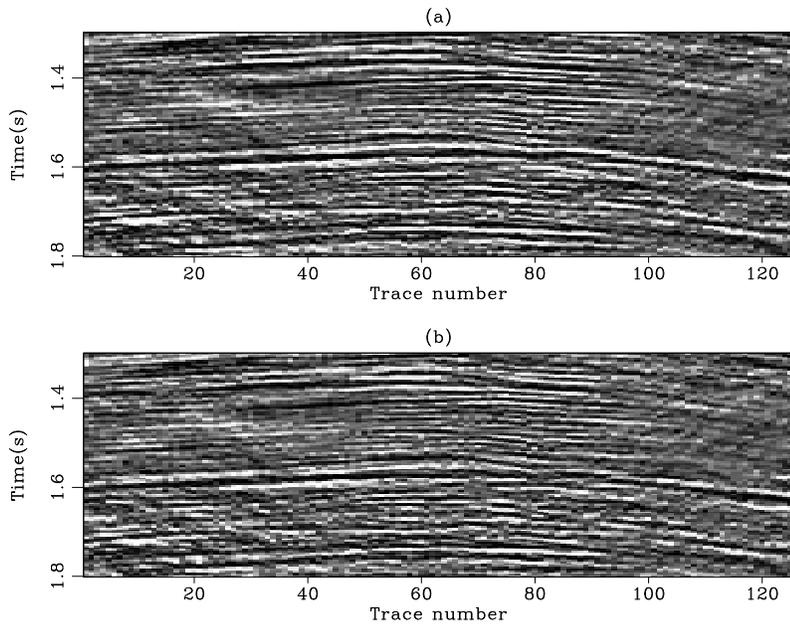


Figure 3.11: (a) The estimated primaries with ℓ^1 norm adaptive subtraction. (b) The estimated primaries with ℓ^2 norm subtraction. The primary at 1.6s is very attenuated with the ℓ^2 norm. The ℓ^1 technique preserves its amplitude better. `adaptive-win` [ER,M]

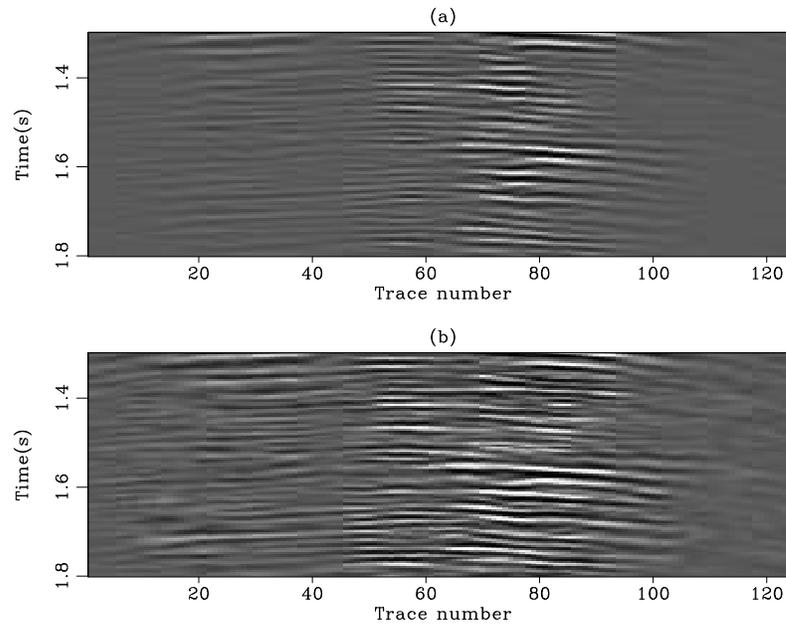


Figure 3.12: (a) The estimated multiples with the ℓ^1 norm subtraction. (b) The estimated multiples with the ℓ^2 norm subtraction. The ℓ^2 norm tends to over-fit some multiples that creates some leaking of primaries in the estimated noise. `adaptive-win2` [ER,M]

to assert if the subtracted multiples with the ℓ^1 norm are more similar to the actual multiples than the subtracted multiples with the ℓ^2 norm. This judgment is only based on qualitative considerations for few known primary reflectors that need to be preserved.

CONCLUSIONS

When a model of the multiples is adaptively subtracted from the data in a least-squares sense, we implicitly assume that the signal (primaries) has minimum energy and is orthogonal to the noise (multiples). This paper demonstrates that the minimum energy assumption might not hold and that another norm, the ℓ^1 norm, should be used instead. Based on Chapter 2, I propose using the Huber norm for the filter estimation problem. I demonstrate with 1-D and 2-D data examples that the Huber norm with a small threshold α gives a much improved multiple attenuation results when the signal has not minimum energy: the multiples are well separated and the primaries are preserved. This Chapter exemplifies that the Huber norm can be used for

many geophysical applications whenever robustness is sought. Chapter 5 will further illustrate this point with an interpolation problem of noisy data.

ACKNOWLEDGMENTS

I would like to thank Saudi Aramco for providing the land data. This project was started with Eric Verschuur during my stay at Delft University in the summer of 2001. For this, I would like to thank SMAART JV for their financial support and the members of the DELPHI consortium for their hospitality and help.

Chapter 4

Noise attenuation by filtering and modeling

SUMMARY

An important assumption in least-squares inversion is that the residual has independent and identically distributed components (IID). If the data are contaminated with noise that the operator does not model properly, the residual will not be IID, thus affecting the estimated model. Theoretically, the data covariance matrix mitigates the effects of modeling uncertainties by filtering the bad data points. In practice, the data covariance operator is often approximated with diagonal operators with similar variance for each point, which can be a valid assumption in the presence of white noise only. For more complicated noise sources with multidimensional spectral components, more elaborated covariance matrices are needed to obtain IID residuals. In this Chapter, two methods that lead to IID residuals are introduced. One method replaces the data covariance matrix with prediction-error filters that absorb the noise spectrum in the data residual. The other method improves the operator by incorporating a noise-modeling component in it. This method can be reduced to a weighting of the data residual with projection filters. On a synthetic and field data examples, while both methods achieve similar results in terms of estimated models, the modeling approach produces better residual panels (i.e., smallest error and smallest coherent energy left).

INTRODUCTION

In seismology, we can express many of the processing steps as linear operators. These operators perform a mapping of one domain, usually a model of the earth parameterized in terms of velocity, reflectivity, into another domain, usually seismic data sorted into CMP or shot gathers. This mapping is called modeling because it models the seismic data. Usually we desire the opposite of modeling, i.e, given the seismic data, we want to retrieve earth parameters that will help us to understand the physical properties of the subsurface. In many cases the adjoint of a modeling operator is used to estimate the model. For some operators like the Fourier transform the adjoint is the exact inverse, for others, the vast majority, the adjoint is not the true inverse but rather an approximation of the inverse.

Nowadays amplitude-preserving processing is an absolutely necessary task for migration, amplitude versus angle (AVA) analysis or 4D interpretation; undoing the modeling part with approximate inverses is then dangerous. Inversion theory provides us with methods to compute a “good” inverse that will honor the seismic data. Pioneering work by Tarantola (1987) shown the usefulness of inversion for earthquake location and tomography. Ever since, inversion has been at the heart of many seismic processing breakthroughs like least-squares migration (Nemeth, 1996), high-resolution radon transforms (Thorson and Claerbout, 1985; Sacchi and Ulrych, 1995) or projection filtering (Soubaras, 1994; Abma and Claerbout, 1995). A very popular method of inversion is the least-squares approach, which can be related to a Bayesian estimation of the model parameters.

It is well understood that inversion in a least-squares sense is very sensitive to the noise level present in the data. By noise I mean abnormally large or small residual components, or outliers which are better described by long-tailed probability density functions (PDF) as opposed to short-tailed gaussian PDF (Chapter 2), and coherent noise that the seismic operator is unable to model. The noise will spoil any analysis based on the result of the inversion and affect the amplitude recovery of the input data. From a more statistical point of view if the data residual, which measures the quality of the data fitting, is corrupted by outliers or coherent noise in the data, it will not have independent and identically distributed (IID) components. In other words, the residual will not have a white spectrum.

In this Chapter, two methods that yield IID data residuals are presented. One method approximates the inverse data covariance operator with prediction-error filters (PEFs), which amounts to a filtering (or weighting) of the data residual. The other method introduces a noise modeling operator that separates signal and coherent noise. I prove that this method amounts to a filtering of the data residual with *projection* filters.

A SHORT REVIEW OF INVERSE PROBLEMS

In this section I review some basic notions on inversion. The least-squares criterion comes directly from the hypothesis that the PDF of each observable data and each model parameter is gaussian. These assumptions lead to the general discrete inverse problem (Tarantola, 1987). Finding \mathbf{m} is then equivalent to minimizing the quadratic function (or cost/objective function)

$$f(\mathbf{m}) = (\mathbf{Lm} - \mathbf{d})' \mathbf{C}_d^{-1} (\mathbf{Lm} - \mathbf{d}) + (\mathbf{m} - \mathbf{m}_{\text{prior}})' \mathbf{C}_m^{-1} (\mathbf{m} - \mathbf{m}_{\text{prior}}), \quad (4.1)$$

where ($'$) represents the (Hermitian) transpose, \mathbf{m} is a mapping of the data (unknown of the inverse problem), \mathbf{L} a seismic (linear) operator, \mathbf{d} the seismic data, \mathbf{C}_d and \mathbf{C}_m the data and model covariance operators and $\mathbf{m}_{\text{prior}}$ a model given a priori.

The covariance matrix \mathbf{C}_d combines experimental errors and modeling uncertainties. Modeling uncertainties describe the difference between what the operator can predict and what is contained in the data. Thus the covariance matrix \mathbf{C}_d is often called the noise covariance matrix (Sacchi and Ulrych, 1995). It is often assumed that, (1) the variances of the model and of the noise are uniform, (2) the covariance matrices are diagonal, i.e., the model and data components are uncorrelated, and (3) no prior model is known in advance. Given these approximations the objective function becomes

$$f(\mathbf{m}) = (\mathbf{Lm} - \mathbf{d})' (\mathbf{Lm} - \mathbf{d}) + \epsilon^2 \mathbf{m}' \mathbf{m}, \quad (4.2)$$

where $\epsilon = \sigma_d^2 / \sigma_m^2$ is a function of the of the noise and model variances σ_d and σ_m . The model perturbation $\mathbf{m}' \mathbf{m}$ reduces to a damping of the cost function. In practice, this damping is used to compensate for numerical instabilities when the parameters (\mathbf{m}) are poorly constrained.

The prior assumptions leading to equation (4.2) are often too strong when dealing with seismic data because the variance of the noise/model may be not uniform and the components of the noise/model not independent. For simplicity I rewrite the objective function in equation (4.2) in terms of “fitting goals” for \mathbf{m} as follows:

$$\begin{aligned}\mathbf{0} &\approx \mathbf{r}_d = \mathbf{Lm} - \mathbf{d}, \\ \mathbf{0} &\approx \epsilon \mathbf{r}_m = \epsilon \mathbf{m},\end{aligned}\tag{4.3}$$

where \mathbf{r}_d is the vector of data residuals and \mathbf{r}_m is the vector of model residuals. With this notation, it is straightforward to rewrite equation (4.2) as

$$f(\mathbf{m}) = \|\mathbf{r}_d\|^2 + \epsilon^2 \|\mathbf{r}_m\|^2.\tag{4.4}$$

The first equality in equation (4.3) stresses the need for \mathbf{Lm} to fit the input data \mathbf{d} . The second equality is often called the regularization or “model styling” term. This term can be useful for imposing a-priori knowledge on the model parameters (Fomel and Claerbout, 2003; Clapp et al., 2004). When the assumptions leading to equation (4.3) are respected, the estimated model $\hat{\mathbf{m}}$ that minimizes equation (4.4) is the maximum likelihood model (Tarantola, 1987).

As stressed before, an important assumption made in equation (4.4) is that the data and model errors \mathbf{r}_d and \mathbf{r}_m are IID. In the situation where coherent noise contaminates the data, these assumptions are violated and the covariance operators cannot be approximated with diagonal operators anymore.

In this Chapter, omitting the model residual vector \mathbf{r}_m in the analysis, I show that a filtering (or weighting) operator \mathbf{W} can be introduced in equation (4.3) such that $\mathbf{W}'\mathbf{W} \approx \mathbf{C}_d^{-1}$. This operator can take the form of a prediction-error filter or a projection filter.

PROPOSED SOLUTIONS TO ATTENUATE COHERENT NOISE

This section proposes two methods for approximating the noise covariance operator \mathbf{C}_d^{-1} when coherent noise is present in the data. But first, let us define \mathbf{d} as the sum of signal \mathbf{s} and coherent

noise \mathbf{n} as follows:

$$\mathbf{d} = \mathbf{s} + \mathbf{n}, \quad (4.5)$$

where \mathbf{n} is made of the inconsistent part (or modeling uncertainties part) of the data \mathbf{d} for any given operator \mathbf{L} . The goal of this Chapter is to estimate earth parameters that are not biased by the noise present in the data.

A filtering method

The filtering method is motivated by the need for estimating the noise covariance operator when the data are noisy. The noise covariance operator has deep theoretical roots for least-squares estimation but its estimation remains somewhat obscure or difficult to achieve. This is why in practice, this matrix is approximated with diagonal operators. In this section I give a method that computes non-diagonal covariance matrices using prediction-error filters.

When coherent noise is present in the data, residual variables are no longer IID. The coherent noise will add “color” to the spectrum of the residual. The key idea is to recognize that the goal of the inverse covariance matrix $\mathbf{C}_{\mathbf{d}}^{-1}$ is to absorb this spectrum. Now, as Claerbout and Fomel (2002) assert:

Clearly, the noise spectrum is the same as the data covariance only if we accept the theoretician’s definition that $E(\mathbf{d})=\mathbf{Fm}$. There is no ambiguity and no argument if we drop the word “variance” and use the word “spectrum”.

This statement is the basis of the filtering method. It says that the data residuals (squared) should be weighted by their inverse multivariate spectrum for optimal convergence. Because a prediction-error filter (PEF) whitens data from which it was estimated (Burg, 1975), it approximates the inverse power spectrum. Thus a PEF (squared) with the inverse spectrum of the coherent noise accomplishes the role of the inverse covariance matrix $\mathbf{C}_{\mathbf{d}}^{-1}$ in equation (4.1).

The fitting goals in equation (4.3) become, omitting the regularization term,

$$\mathbf{0} \approx \mathbf{r}_d = \mathbf{A}(\mathbf{L}\mathbf{m} - \mathbf{d}), \quad (4.6)$$

where \mathbf{A} is a PEF that whitens the coherent noise. Therefore, the noise covariance matrix is approximated as follows:

$$\mathbf{A}'\mathbf{A} \approx \mathbf{C}_d^{-1}. \quad (4.7)$$

This method is solely based on the whitening properties of the PEFs; other types of filters can be used as well. The next section introduces a noise modeling approach. I show that this method reduces to a weighting of the residual with a projection filter.

A modeling approach

Instead of removing the noise by filtering (i.e., equation (4.6)), the forward operator \mathbf{L} can be improved to model both noise and signal components. This technique treats the coherent noise as components of the data. Therefore, a coherent noise modeling operator \mathbf{L}_n and signal modeling operator \mathbf{L}_s are introduced to give (Nemeth, 1996) $\mathbf{L} = (\mathbf{L}_s \ \mathbf{L}_n)$ in equation (4.1). The model space \mathbf{m} then becomes the vector $\mathbf{m}' = (\mathbf{m}_s \ \mathbf{m}_n)$, where \mathbf{m}_n is the model space for the noise and \mathbf{m}_s is the model space for the signal. Introducing an adequate noise modeling operator \mathbf{L}_n , the data residual \mathbf{r}_d will become IID and \mathbf{C}_d^{-1} can be approximated more safely with a diagonal operator with constant variance. In terms of fitting goal and omitting the regularization term, we have now, assuming $\mathbf{C}_d = \sigma_d^2 \mathbf{I}$ where \mathbf{I} is the identity matrix:

$$\mathbf{0} \approx \mathbf{r}_d = \mathbf{L}_s \mathbf{m}_s + \mathbf{L}_n \mathbf{m}_n - \mathbf{d}. \quad (4.8)$$

The cost function becomes

$$f(\mathbf{m}_s, \mathbf{m}_n) = \|\mathbf{L}_s \mathbf{m}_s + \mathbf{L}_n \mathbf{m}_n - \mathbf{d}\|^2, \quad (4.9)$$

and the estimated inverse for \mathbf{m} (see Appendix B for details)

$$\begin{pmatrix} \hat{\mathbf{m}}_s \\ \hat{\mathbf{m}}_n \end{pmatrix} = \begin{pmatrix} (\mathbf{L}'_s \overline{\mathbf{R}}_n \mathbf{L}_s)^{-1} \mathbf{L}'_s \overline{\mathbf{R}}_n \\ (\mathbf{L}'_n \overline{\mathbf{R}}_s \mathbf{L}_n)^{-1} \mathbf{L}'_n \overline{\mathbf{R}}_s \end{pmatrix} \mathbf{d}, \quad (4.10)$$

with $\overline{\mathbf{R}}_s = \mathbf{I} - \mathbf{L}_s (\mathbf{L}'_s \mathbf{L}_s)^{-1} \mathbf{L}'_s$ and $\overline{\mathbf{R}}_n = \mathbf{I} - \mathbf{L}_n (\mathbf{L}'_n \mathbf{L}_n)^{-1} \mathbf{L}'_n$. Both rows in equations (4.10) are the solutions of a weighted least-squares problem (Menke, 1989) for the following fitting goals:

$$\begin{aligned} \mathbf{0} &\approx \mathbf{r}_{ds} = \overline{\mathbf{R}}_n (\mathbf{L}_s \mathbf{m}_s - \mathbf{d}), \text{ and} \\ \mathbf{0} &\approx \mathbf{r}_{dn} = \overline{\mathbf{R}}_s (\mathbf{L}_n \mathbf{m}_n - \mathbf{d}), \end{aligned} \quad (4.11)$$

where \mathbf{r}_{ds} and \mathbf{r}_{dn} are the residuals for the noise and signal components. Equation (4.11) is true because $\overline{\mathbf{R}}_s$ and $\overline{\mathbf{R}}_n$ are (1) projection operators, and (2) signal and noise filtering operators, respectively (see Appendix B for details). It is important to realize that in practice, the projection operators in equation (4.11) are not directly estimated and \mathbf{m}_s and \mathbf{m}_n are rather computed iteratively from the fitting goal in equation (4.8).

What is interesting in equation (4.11), however, is that the first fitting goal is very similar to equation (4.6), the difference stemming from the choice of weighting (or filtering) operator. The modeling approach can be then interpreted as a weighting of the data residual with a projection filter that annihilates coherent noise. Therefore, the noise covariance matrix is approximated as follows:

$$\overline{\mathbf{R}}_n \overline{\mathbf{R}}_n = \overline{\mathbf{R}}_n \approx \mathbf{C}_d^{-1}. \quad (4.12)$$

To summarize, the filtering approach attempts approximating the noise covariance operator with prediction-error filters, thus using the property that \mathbf{C}_d is the power spectrum of the noise. In contrary, the modeling approach tackles the noise problem at its source by trying to model both the noise and signal simultaneously. Nonetheless, this technique can still be seen as a way of approximating \mathbf{C}_d with projection filters (i.e., equation (4.12)).

Numerous authors (Soubaras, 1994, 1995; Abma, 1995b; Abma and Claerbout, 1995; Ozdemir et al., 1999) have proved that projection filters were more desirable for signal/noise

separation than simple prediction-error filters. The main reason is that the spectrum of projection filters is in the range of zero to one. Therefore, the modeling approach should be used as much as possible for coherent seismic noise attenuation. As an illustration, Chapter 5 demonstrates on an interpolation problem of noisy data the benefits of the modeling approach compared to the filtering one.

In the next section, practical considerations are addressed for both noise filtering and noise modeling approaches. In particular, strategies for choosing the filters and operators are detailed. In addition, a pattern-based approach for signal/noise separation is briefly introduced.

COHERENT NOISE FILTERING AND MODELING IN PRACTICE

The core of the filtering method is the PEF \mathbf{A} that destroys the coherent noise present in the residual. In order to accommodate the possible complexity in space and time of the undesirable events, these filters can have any dimensions thanks to the helical boundary conditions (Mersereau and Dudgeon, 1974; Claerbout, 1998). In addition, these filters can be made non-stationary in any directions (Margrave, 1998; Crawley, 2000; Claerbout and Fomel, 2002). Chapters 6 and 7 will illustrate these possibilities.

If a noise model is known in advance, a PEF can be estimated from it and used directly for the inversion of \mathbf{m} in equation (4.6). If a noise model is not known in advance, I propose the following algorithm based on the iterative inversion of \mathbf{m} :

1. Set $\mathbf{A} = \mathbf{I}$ and $\mathbf{m} = \mathbf{0}$.
2. Iterate until the data residual \mathbf{r}_d contains mostly coherent noise.
3. Estimate a new PEF \mathbf{A} from the residual.
4. Set $\mathbf{m} = \mathbf{0}$ and iterate.
5. Inspect the residual. If white then stop else reestimate the PEF \mathbf{A} from the unweighted residual $\mathbf{Lm} - \mathbf{d}$ and go to (4).

This nonlinear process tries to refine the PEF estimation by bootstrapping in order to obtain the best possible coherent noise filter \mathbf{A} (Clapp, 2003). It is also based on the assumption that the coherent noise will stay in the residual after a certain number of iterations. Therefore, the residual needs to be closely watched at each iteration, which requires a certain level of human interpretation.

The modeling approach

Before giving strategies on how to choose the modeling operators \mathbf{L}_s and \mathbf{L}_n , an important parameter needs to be introduced. Depending on the signal and noise that need to be attenuated, the two model spaces \mathbf{m}_s and \mathbf{m}_n can have very different units. This difference can basically bias the inversion toward the noise or the signal if nothing is done to balance the two models. Therefore, a physically meaningful fitting goal for the modeling approach is

$$\mathbf{0} \approx \mathbf{r}_d = \mathbf{L}_s \mathbf{m}_s + \gamma \mathbf{L}_n \mathbf{m}_n - \mathbf{d}, \quad (4.13)$$

where γ is a constant that compensates for differences in the units of \mathbf{m}_s and \mathbf{m}_n . This parameter can be seen as a way of putting the common components of the data in whichever domain we choose, noise or signal, in the presence of crosstalks. There is no acceptably simple way to select γ but by trial and error, e.g., in Chapter 5. One possible strategy, however, consists in computing the ratio of the norm (in a ℓ^2 sense) of the two gradients for \mathbf{m}_s and \mathbf{m}_n as follows:

$$\gamma = \frac{\|\mathbf{L}'_s \mathbf{d}\|}{\|\mathbf{L}'_n \mathbf{d}\|}. \quad (4.14)$$

This ratio can be evaluated before starting the inversion (Guitton, 2000). Now for the choice of modeling operators \mathbf{L}_s and \mathbf{L}_n , it is important to have them modeling different parts of the data space. If they overlap, a regularization operator is needed to improve the signal/noise separation (Nemeth, 1996). There are a large number of operators that can model the signal and noise depending on the data. For instance, Nemeth et al. (2000) shows that migration can be used.

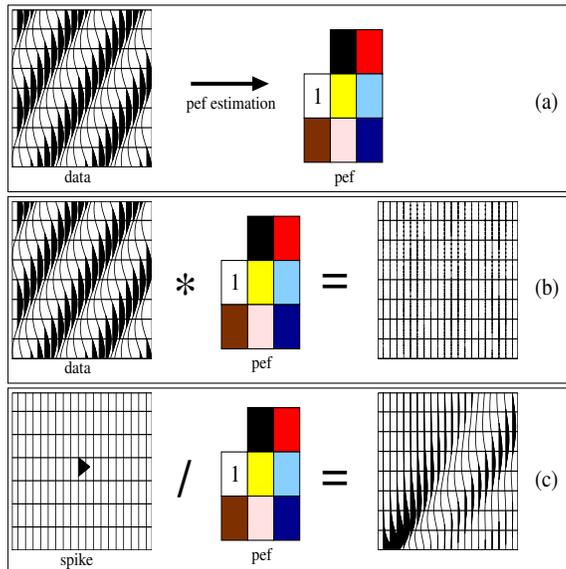
PEFs can play this role as well because they have the inverse spectrum of the data from

which they were estimated. As an illustration, Figure 4.1a displays a simple monochromatic plane wave with a given dip. If a multidimensional PEF is estimated from this dataset, then the PEF will kill this plane wave (Figure 4.1b). Now, looking at the inverse impulse response of this filter (i.e., deconvolution) in Figure 4.1c, notice that an event with similar frequency content and dip than in Figure 4.1a is recovered. Thus, PEFs can serve as either filters when convolved with the residual (filtering approach) or modeling operators when used with deconvolution.

PEFs are minimum-phase filters (Claerbout, 1976; Robinson and Treitel, 2000), which makes their inverse (with deconvolution) stable. In the stationary case where one filter is needed, inverse PEFs can play the role of modeling operators, and strategies similar to the one used for the filtering approach can be employed to estimate them. In the non-stationary case, however, the non-stationary deconvolution of minimum-phase filters is not guaranteed to be stable (Rickett, 2001). This defect prevents us from reliably using inverse non-stationary PEFs as modeling operators.

Figure 4.1: Some basic properties of the prediction-error filters (PEF). (a) A PEF is estimated from a dataset. (b) The estimated PEF is convolved with the input data leading to a white residual. (c) A spike is deconvolved with the estimated PEF leading to a pattern close to the input data in (a).

`iid-fbpef` [NR]



From filtering to covariance-pattern based approach

The filtering approach for coherent noise removal can lead to a covariance-pattern based approach. This technique is often called pattern-based approach for short. The basic idea is to separate the noise and signal present in the data according to their pattern, or multivariate spectra. To identify and approximate these patterns, multidimensional PEFs can be used.

I now show briefly how this technique works. Building on equation (4.6) we introduce a regularization operator to obtain

$$\begin{aligned}\mathbf{0} &\approx \mathbf{r}_d = \mathbf{A}(\mathbf{L}\mathbf{m} - \mathbf{d}), \\ \mathbf{0} &\approx \epsilon\mathbf{r}_m = \epsilon\mathbf{R}\mathbf{m},\end{aligned}\tag{4.15}$$

and minimize

$$f(\mathbf{m}) = \|\mathbf{r}_d\|^2 + \epsilon^2\|\mathbf{r}_m\|^2.\tag{4.16}$$

In this case, ϵ is a trade-off parameter between data fitting and model smoothing. In the pattern-based approach, the unknown parameter is the signal $\mathbf{m} = \mathbf{s}$ and the modeling operator \mathbf{L} becomes the identity. This choice of \mathbf{L} comes from the fact that both the signal and the noise reside simultaneously in the data space. In the filtering approach, \mathbf{A} is a PEF that annihilates the noise components. Let us call it $\mathbf{A} = \mathbf{N}$. For the regularization operator, we need to find an operator for the signal such that $\mathbf{R}\mathbf{s} \approx \mathbf{0}$. By definition, a PEF \mathbf{S} that annihilates the signal components is a good candidate for \mathbf{R} . From these choices, we end up with

$$\begin{aligned}\mathbf{0} &\approx \mathbf{r}_d = \mathbf{N}(\mathbf{s} - \mathbf{d}), \\ \mathbf{0} &\approx \epsilon\mathbf{r}_s = \epsilon\mathbf{S}\mathbf{s}.\end{aligned}\tag{4.17}$$

Solving equation (4.17) in a least-squares sense gives for the estimated signal (Abma, 1995a):

$$\hat{\mathbf{s}} = (\mathbf{N}'\mathbf{N} + \epsilon^2\mathbf{S}'\mathbf{S})^{-1}\mathbf{N}'\mathbf{N}\mathbf{d}.\tag{4.18}$$

Note that $\mathbf{N}'\mathbf{N}$ and $\mathbf{S}'\mathbf{S}$ are the inverse multidimensional spectra for the noise and signal, respectively, which explains why this technique of signal/noise separation is called pattern-based. As

stated by Claerbout and Fomel (2000), PEFs are important where they are small. Analyzing the noise filter in equation (4.18)

$$\mathbf{F} = (\mathbf{N}'\mathbf{N} + \epsilon^2\mathbf{S}'\mathbf{S})^{-1}\mathbf{N}'\mathbf{N}, \quad (4.19)$$

we see that

$$\begin{aligned} \mathbf{F} &= \mathbf{0} && \text{if } \mathbf{N}'\mathbf{N} = \mathbf{0} \text{ and} \\ \mathbf{F} &= \mathbf{1} && \text{if } \mathbf{N}'\mathbf{N} \gg \epsilon^2\mathbf{S}'\mathbf{S}. \end{aligned} \quad (4.20)$$

The first condition is met when the noise is strong (PEF \mathbf{N} small), thus attenuating the noise, and the second when the signal is strong (PEF \mathbf{S} small), thus preserving it. The filter \mathbf{F} is also a projection filter (Soubaras, 1994; Sacchi and Kuehl, 2001).

In the pattern-based approach, \mathbf{N} and \mathbf{S} need to be estimated before the actual separation. This is the main challenge of this technique, i.e, how to choose good noise and signal models. Chapter 6 and 7 illustrate the pattern-based technique for surface-related multiple attenuation. In this case, strategies exist to estimate the noise and signal PEFs. In the next section, I illustrate the filtering and modeling techniques on synthetic and field data examples.

SYNTHETIC AND FIELD DATA EXAMPLES

Figure 4.2a displays a synthetic CMP gather composed of hyperbolas (the signal) and a monochromatic plane-wave (the noise). The true signal and noise are shown in Figures 4.2b and 4.2c with their (f, k) spectra. It is interesting to notice that the noise and signal overlap in the (f, k) domain. This can be seen in the CMP gather where the noise has (1) similar dips to some hyperbola tails, and (2) similar frequency content. This overlap makes signal preservation particularly difficult to achieve with simple filtering techniques such as bandpass. From this dataset, the goal is to estimate in a least-squares sense a velocity panel with the hyperbolic radon transform of Chapter 2 ($\mathbf{L} = \mathbf{H}$). This velocity panel can be later used for velocity analysis or multiple attenuation (Thorson and Claerbout, 1985; Foster and Mosher, 1992; Alvarez and Larner, 2004). A conjugate-gradient (CG) solver is used for the iterations.

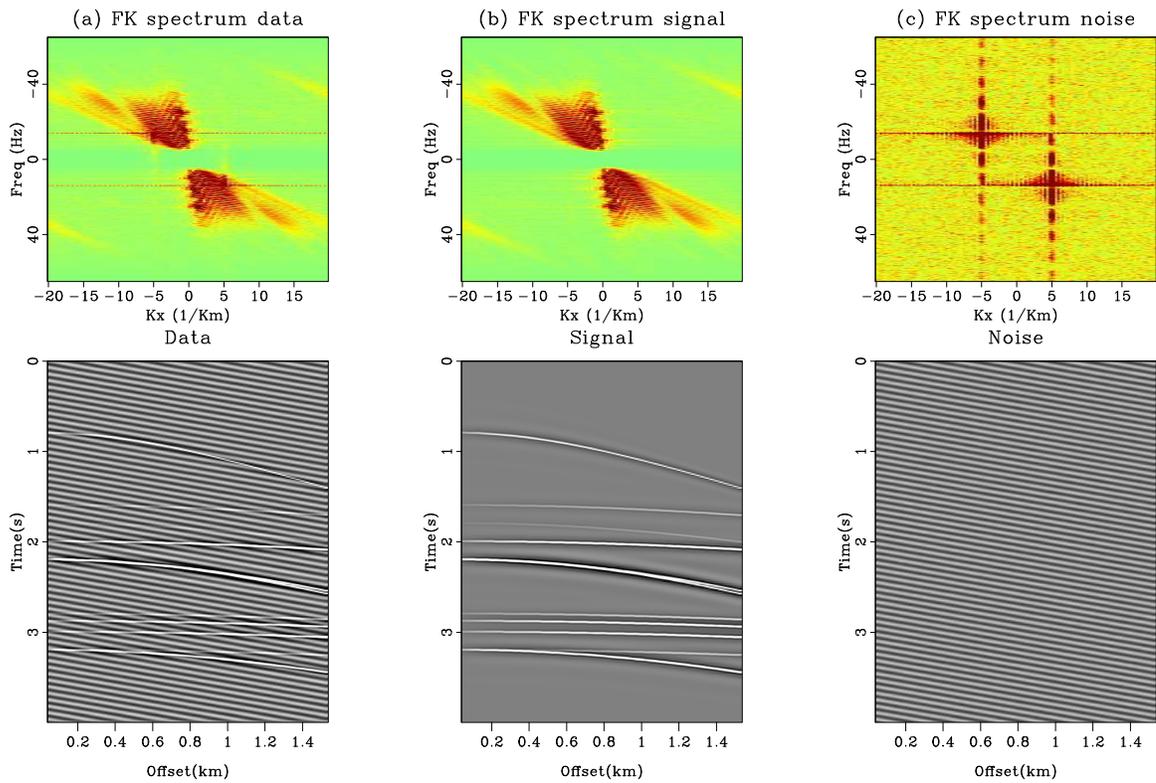


Figure 4.2: A synthetic data example with (a) the data and its (f, k) spectrum, (b) the true signal and its (f, k) spectrum, and (c) the noise. The spectra of signal and the noise overlap, which makes simple filtering (i.e., without inversion) likely to fail. `iid-datasynt` [ER]

Figure 4.3a illustrates the effects of the coherent noise on the model space after inversion of the data in Figure 4.2a. Some energy corresponding to both the signal and the noise is present in Figure 4.3a, but overall, it is quite difficult to recognize events corresponding to the signal. The remodeled data in Figure 4.3b (i.e., $\mathbf{H}\hat{\mathbf{m}}$) indicate that the inversion is fitting both the noise and signal, which creates artifacts in Figure 4.3a. In addition, note that the residual (i.e., $\mathbf{d} - \mathbf{H}\hat{\mathbf{m}}$) in Figure 4.3c is not IID: a lot of coherent energy remains. Therefore, the filtering and modeling techniques are needed to obtain (1) IID residuals, (2) a better velocity panel, and (3) noise-free remodeled data. In this example, the noise is assumed to be known. For the

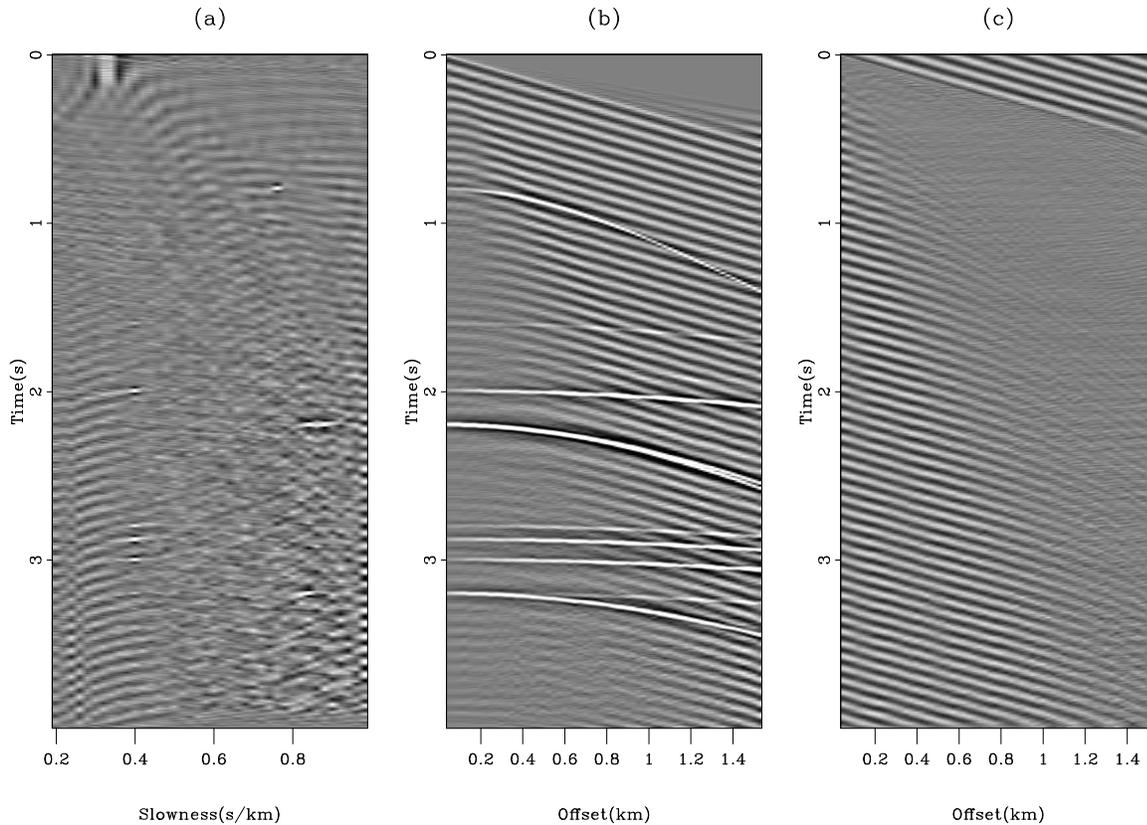


Figure 4.3: (a) Estimated model space after inversion of the noisy data in Figure 4.2a. (b) Remodeled data from the inverted model. (c) Residual \mathbf{r}_d after inversion. The residual is not IID and a lot of noise contaminates the model space. `iid-synth-nopef` [ER]

filtering approach, a 5×3 PEF is used for \mathbf{A} in equation (4.6). For the modeling approach, $\mathbf{L}_s = \mathbf{H}$ and $\mathbf{L}_n = \mathbf{A}^{-1}$, \mathbf{A} being the same PEF that for the filtering approach.

First, as an illustration on how a simple prediction-filter differs from a projection filter, Figure 4.4a displays the spectrum of the PEF \mathbf{A} only and Figure 4.4b the spectrum of the projection filter $\mathbf{I} - \mathbf{L}_n(\mathbf{L}'_n\mathbf{L}_n)^{-1}\mathbf{L}'_n$ with $\mathbf{L}_n = \mathbf{A}^{-1}$. The PEF is estimated from the known noise, and we notice that its spectrum in Figure 4.4a is the smallest at the noise location. The amplitude varies a lot with local minima and a maximum amplitude far from one, however. The spectrum of the projection filter in Figure 4.4b is better behaved: it equals zero at the noise location and is flat almost everywhere. The maximum energy is not exactly one because a damping term was added for the division. Figure 4.5 displays the result of the inversion

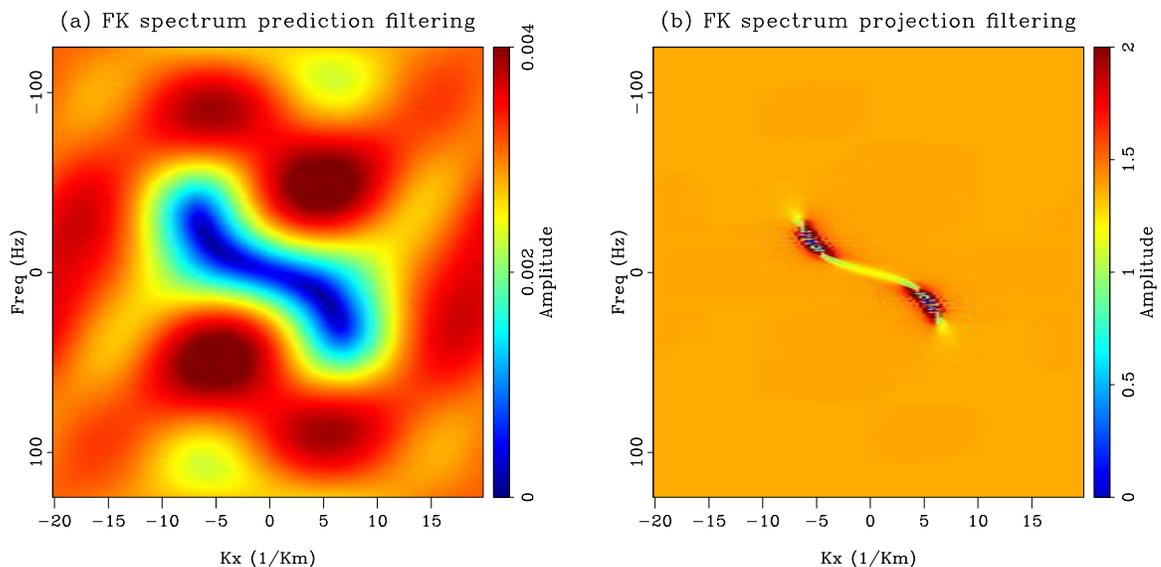


Figure 4.4: (a) (f, k) spectrum of the PEF. It is minimum at the noise location but varies a lot everywhere else. (b) (f, k) spectrum of the projection filter. It is minimum at the noise location and flat otherwise. The projection filter is a better filter than the prediction filter.

iid-pefs [ER]

with the filtering approach. The estimated model in Figure 4.5a is now noise-free and the hyperbolas are well focused in the velocity space. The residual in Figure 4.5d has no coherent energy left and is IID. The remodeled data in Figure 4.5b are also basically noise free, and so is the unweighted residual $(\mathbf{d} - \mathbf{H}\hat{\mathbf{m}})$ in Figure 4.5c. Therefore, the filtering technique worked very well and delivered the expected results. Figure 4.6 shows the result of the modeling approach. Remember that the inverse of the 5×3 PEF estimated from the true noise is used for the noise operator \mathbf{L}_n and that the hyperbolic radon transform is used to model the signal.

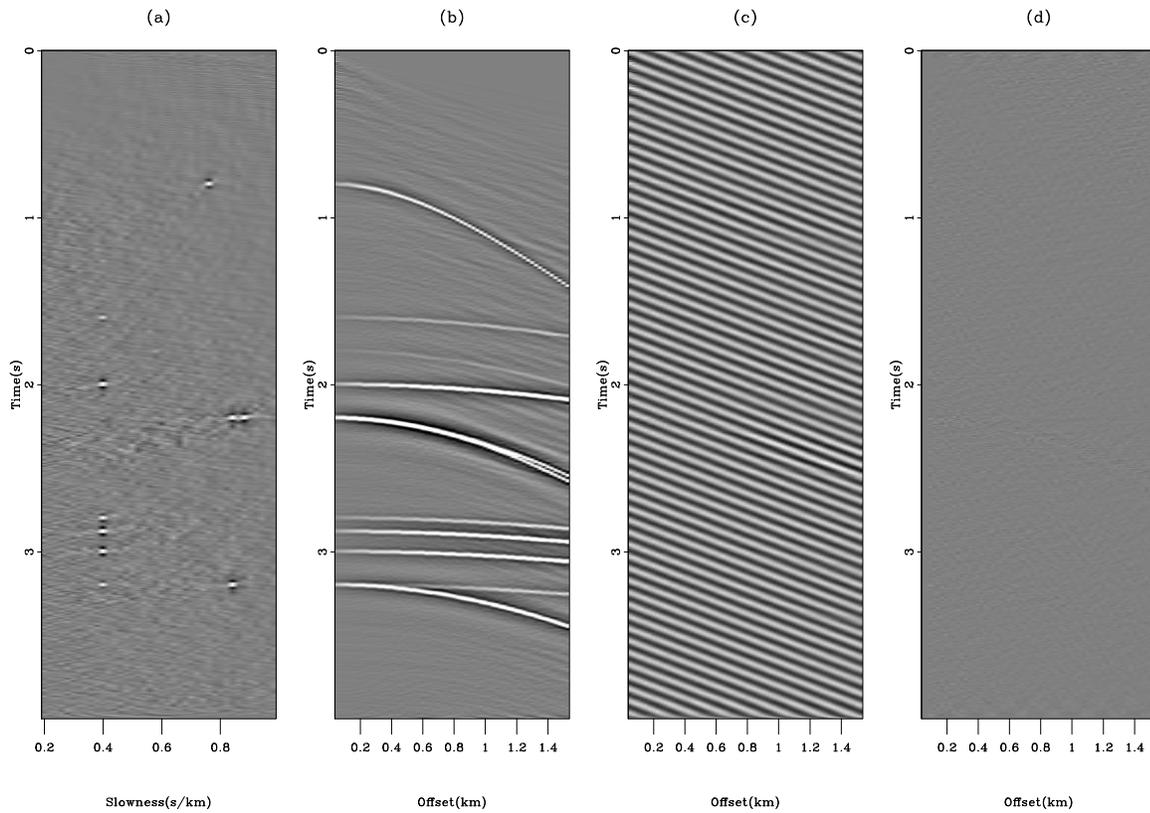


Figure 4.5: Inversion result with the filtering technique. (a) Estimated model. (b) Remodeled data. (c) Unweighted residual, which shows the coherent noise only. (d) Residual. The estimated model is artifacts free and the residual is IID. The movie shows the four panel as a function of the iterations. `iid-c-synth` [ER,M]

The balancing parameter λ in equation (4.13) is chosen by trial and error.

The estimated model after inversion is shown in Figure 4.6a. Similar to what we observed with the filtering approach, the model is noise free and easy to interpret. The remodeled data in Figure 4.6b show a little bit of coherent noise. The problem stems from the fact that the two modeling operators overlap: the inverse PEF can model some tails of hyperbolas and the radon transform can fit some of the coherent noise (because the velocity of the noise is within the range of the slowness scan). Although not crucial here, a regularization term would improve this result (Nemeth, 1996). The estimated noise of Figure 4.6c proves that the separation worked very well. A little bit of one hyperbola has been absorbed, however. The residual in Figure 4.6d is IID which confirms that the goals of this technique were reached.

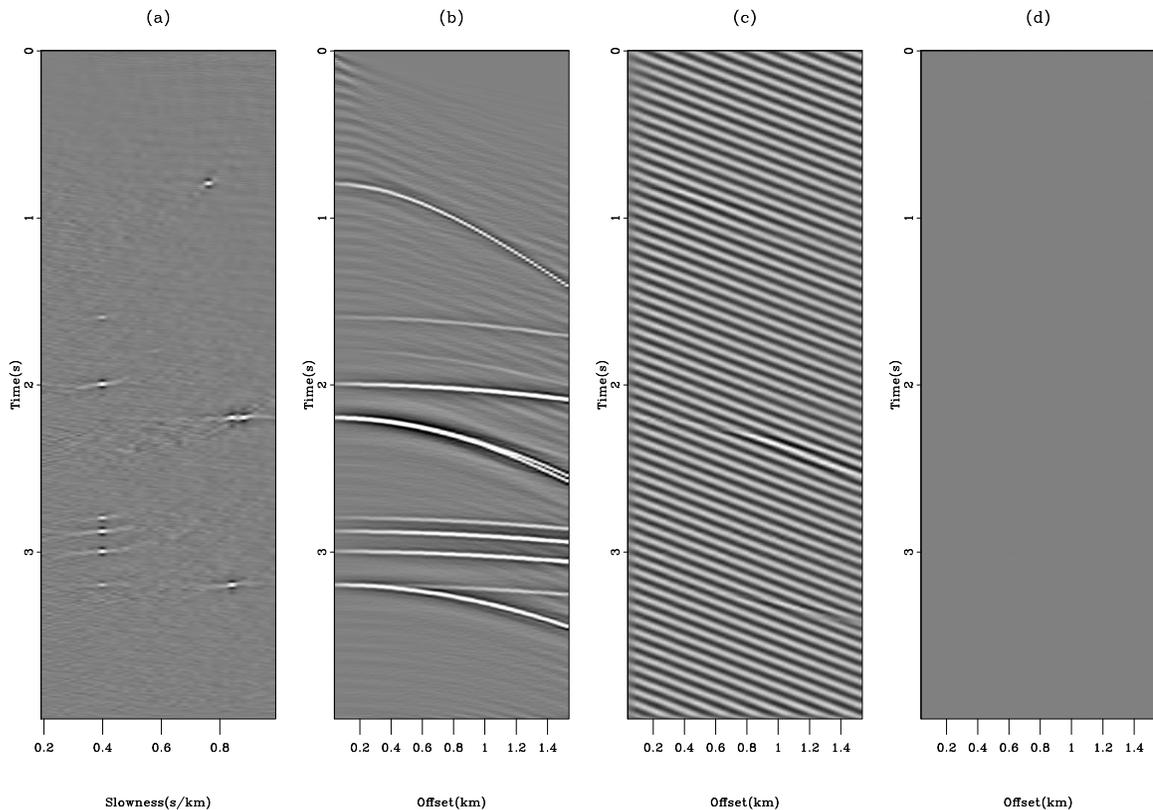


Figure 4.6: Inversion result with the modeling technique. (a) Estimated model \mathbf{m}_s . (b) Remodeled data \mathbf{Hm}_s . (c) Remodeled noise $\mathbf{A}^{-1}\mathbf{m}_n$. (d) Residual. The residual is IID and the model space is clean. Because the signal and noise modeling operators tend to overlap, a little bit of noise leaks in the estimated signal. This effect is very small, however. `IID-compsy` [ER,M]

Land data example

Figure 4.7 displays a CMP gather from a land acquisition. This gather is contaminated by a strong ground-roll with weak signal. The goal is to test the two proposed techniques with the same operators as in the preceding example, i.e., hyperbolic radon transform to obtain velocity information and a 5×3 PEF for the filtering and modeling of the noise. By doing inversion

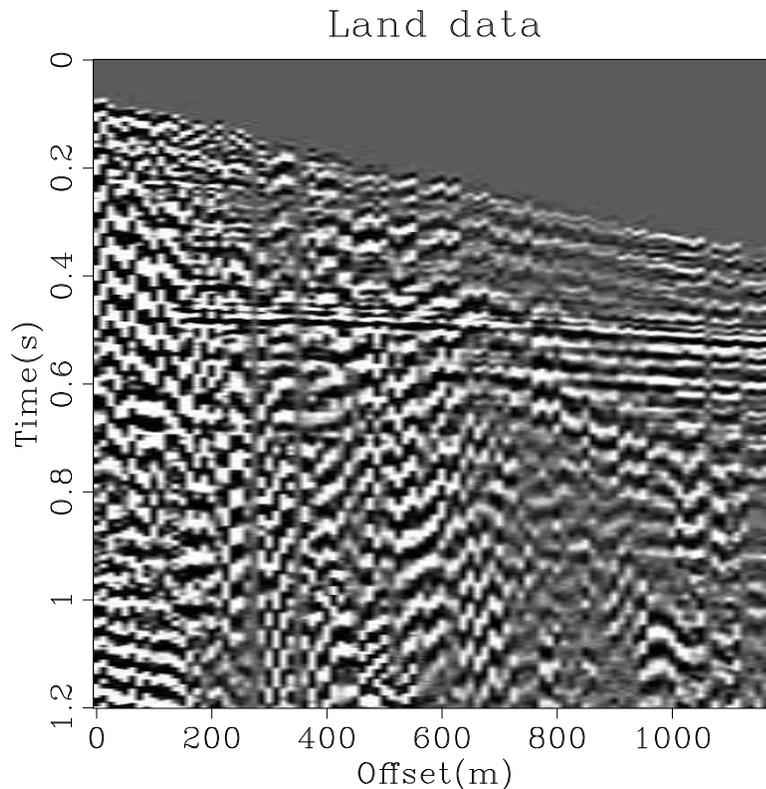


Figure 4.7: A CMP gather from a land survey. The ground-roll is very strong throughout the section. `iid-amoco` [ER]

without taking care of the coherent noise, the velocity panel in Figure 4.8a is obtained. The model space is contaminated with strong noise coming from the residual in Figure 4.8c. When noise is present in the data, least-squares with a simple damping can help to mitigate the noise effects. This was not done in this case, however. The reconstructed data in Figure 4.8b is relatively noise free, which indicates that most of the energy in the model space (Figure 4.8a) is in the null space of the operator. For the filtering and modeling techniques to work,

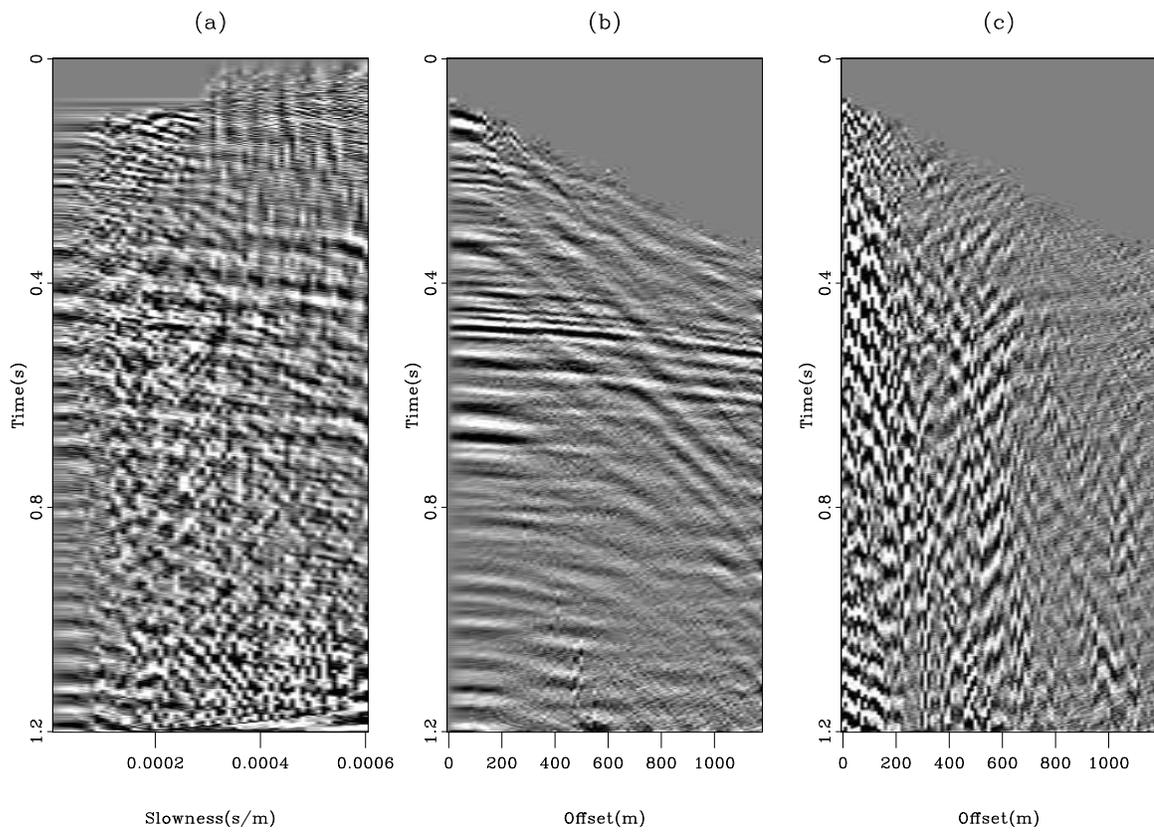


Figure 4.8: Result of least-squares inversion of the data in Figure 4.7. (a) Estimated model space. The noise level is too high to see any useful information. (b) The remodeled data. The low noise level in this panel indicates that most of the energy in (a) resides in the null space of the operator. (c) Data residual. The residual is not IID. `iid-amoco-nopef` [ER]

a noise model is needed. As suggested in a preceding section, the noise model is estimated from the residual of least-squares inversion assuming that $\mathbf{A} = \mathbf{I}$ (i.e., Figure 4.8c). From this noise model, a 5×3 PEF is estimated. Figure 4.9a displays the result of the inversion after 40 iterations of CG. The few hyperbolas present in the data are clearly mapped in the model space without any artifact. The remodeled data in Figure 4.9b show almost no ground-roll remaining. The unweighted residual in Figure 4.9c ($\mathbf{Hm} - \mathbf{d}$) contains only coherent noise and the residual in Figure 4.9d is almost IID: a little bit of ground-roll is still present. The

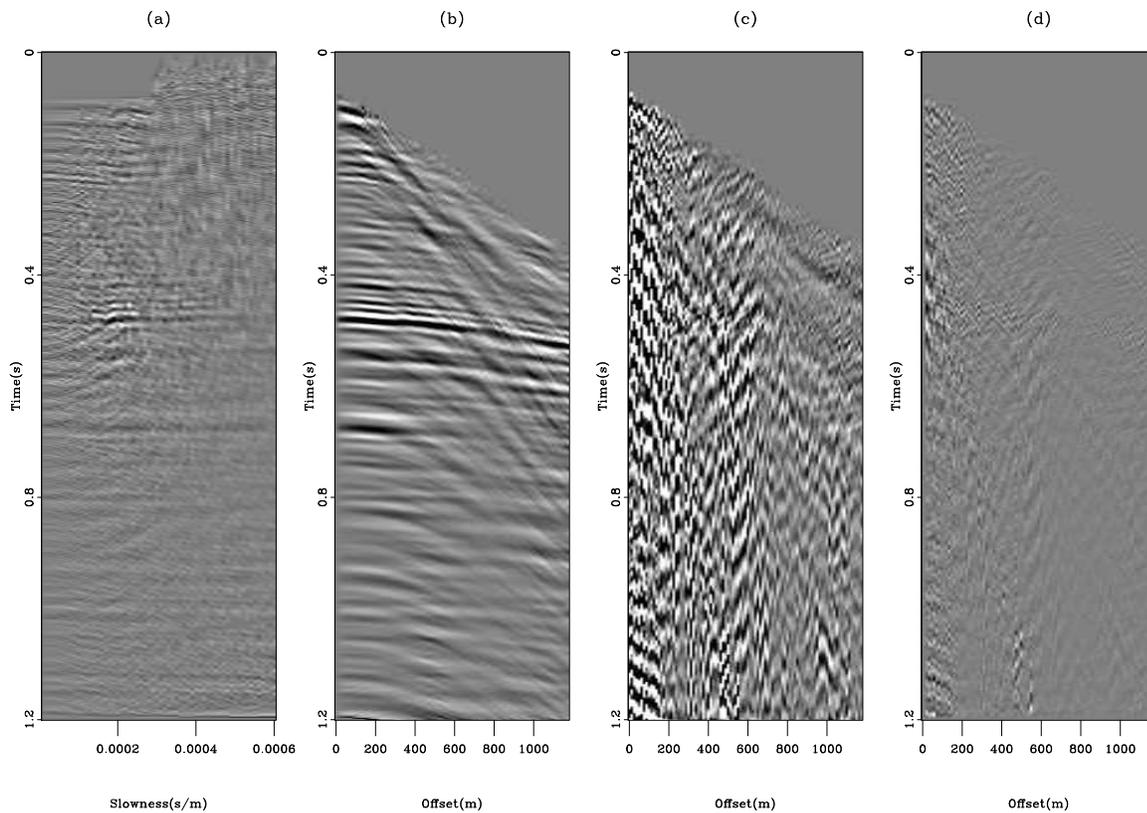


Figure 4.9: Inversion result with the filtering technique. (a) Estimated model. (b) Remodeled data. (c) Unweighted residual, showing the ground-roll only. (d) Residual. The residual is not exactly IID. Some coherent energy remains. `iid-c-amoco` [ER,M]

modeling technique leads to similar results to the filtering technique, as shown in Figure 4.10. The main difference comes from the residual in Figure 4.10d. With the modeling method, the residual is smaller than with the filtering technique and is IID. As a final comparison between the two techniques, Figure 4.11 shows how the objective function (normalized) decreases in

both cases. The modeling technique has the best convergence properties.

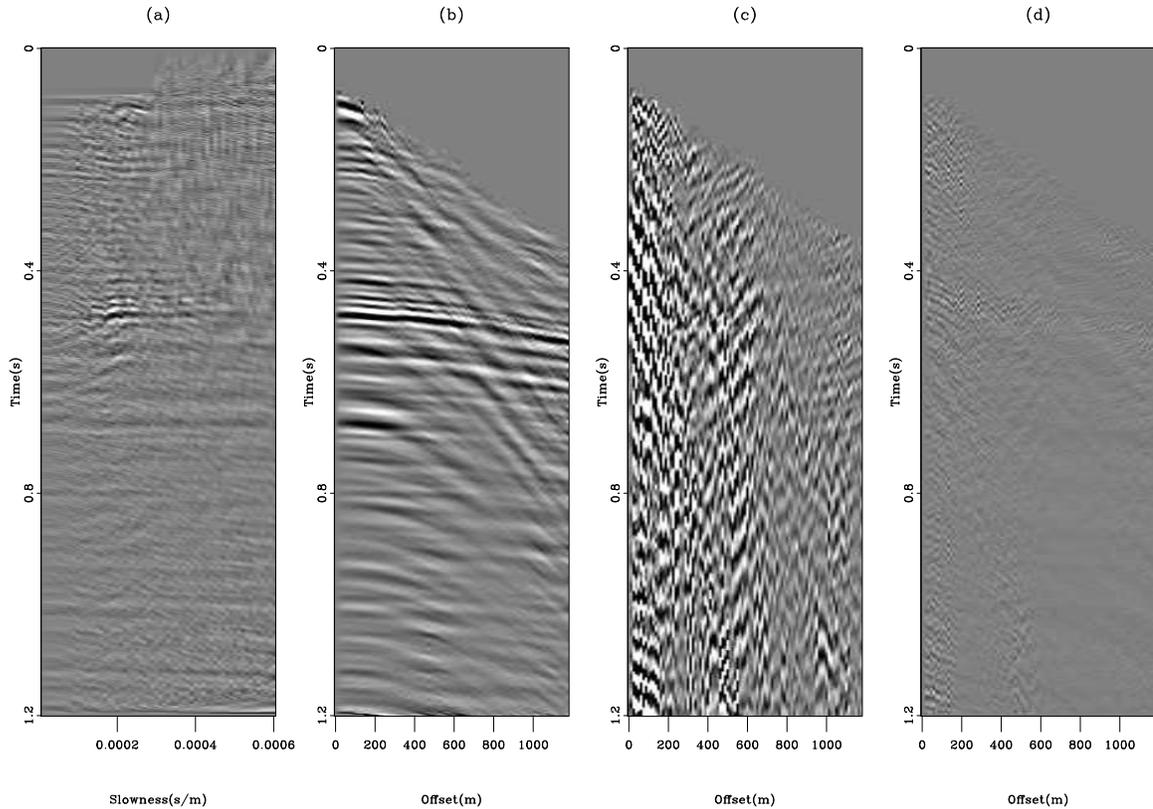


Figure 4.10: Inversion result with the modeling technique. (a) Estimated model \mathbf{m}_s . (b) Remodeled data $\mathbf{H}\mathbf{m}_s$. (c) Remodeled noise $\mathbf{A}^{-1}\mathbf{m}_n$. (d) Residual. The residual is IID. iid-compsa [ER,M]

CONCLUSION

Two methods that attenuate coherent noise present in the data are derived from the need to have IID residual components. In the presence of noise, having IID residual guarantees the inversion to converge toward a model that will be consistent with the signal only. One method weights the residual with a PEF estimated from a noise model. The PEF (squared) plays the role of the data covariance operator. The second method introduces a noise modeling part such that the data covariance operator is not needed anymore. This method can be interpreted as a

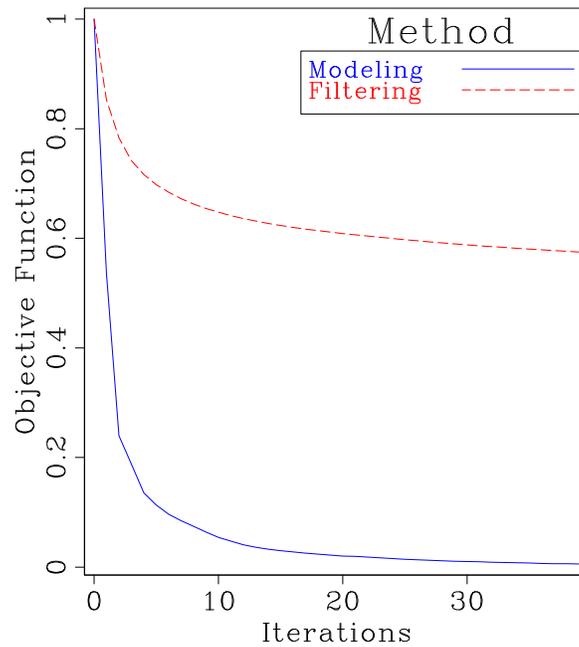


Figure 4.11: A comparison of convergence between the filtering and modeling approach for the land data example. `iid-amococonv` [ER]

weighting of the data residual with projection operators. On a velocity analysis problem, both techniques perform similarly. In terms of convergence properties, the modeling approach gives better results with less energy left in the residual. Added to the fact that projection operators have excellent filtering abilities, the modeling approach should be the method of choice when possible. The next Chapter illustrates on an interpolation problem of noisy data failures of the filtering method and successes of the modeling one.

ACKNOWLEDGMENTS

I would like to thank BP for the land data example.

Chapter 5

Interpolation of bathymetry data

SUMMARY

A bathymetry survey from the Sea of Galilee is processed. This dataset is contaminated with non-gaussian noise in the form of spikes inside the lake. There is drift on the depth measurements leading to vessel tracks in the preliminary depth images. The drift comes from different seasonal and human conditions during the data acquisition, e.g., wind and water levels. I derive an inversion scheme that produces a much reduced noise map of the Sea of Galilee. This inversion scheme includes preconditioning and the Huber norm to eliminate the non-gaussian noise (Chapter 2). The ship tracks are removed by adding a modeling operator inside the inversion (Chapter 4) that accounts for the drift in the data. My images of the Sea of Galilee show ancient shorelines and inside the lake, rifting features.

INTRODUCTION

In this problem, we are given depth sounding data from the Sea of Galilee. The Sea of Galilee is unique because it is a *fresh-water lake below sea-level*. It is connected to a transform fault system (Levant fault zone) along the eastern margin of the Mediterranean Sea. The ultimate goal is to produce not only a good map of the depth to bottom, but images useful for the

purpose of identifying archaeological, geological, or geophysical details of the sea bottom. In particular, we hope to be able to identify some ancient shorelines around the lake and meaningful geological features inside the lake. The ancient shorelines could unravel early settlements of archaeological interest or old fishing ports.

The raw data (Figure 5.1), irregularly distributed across the surface, are 132,044 triples, (x_i, y_i, z_i) , where x_i ranges over about 12 km and where y_i ranges over about 20 km. The data need to be interpolated to a regular grid using inversion to facilitate the processing, such as noise removal, and also to create a map that can be easily analyzed for identifying artifacts and geology.

The pertinence of this dataset to our daily geophysical problems is three fold. First, we often have to do interpolation of seismic maps (Britze, 1998), potential field data (Guspi and Introcaso, 2000) or other measurements to compensate for the sparseness and irregularities of acquisition geometries. Second, as seen in the raw data in Figure 5.1, some noise bursts related to spurious electronic signals (glitches) and/or positioning errors need to be accounted for in the inversion scheme. This problem is common, for example, in tomography (Bube and Langan, 1997), deconvolution of noisy data (Chapman and Barrodale, 1983) and velocity analysis (Chapter 2) where outliers can degrade the final model if we assume a gaussian distribution of the noise. Third, the final image of the Sea of Galilee will display the vessel tracks because the measurements on the lake were made on different days, with different weather and human conditions. We can directly link this problem to the goal of removing the acquisition footprint with 3-D seismic data (Duijndam et al., 2000; Schuster and Liu, 2001; Chemingui and Biondi, 2002). Therefore, the interpolation of the data from the Sea of Galilee becomes a spiky noise and a coherent noise attenuation problem. There is a long list of students at the Stanford Exploration Project who attempted to produce a satisfying map of the sea bottom. Fomel and Claerbout (1995) introduced the ℓ^1 norm via Iteratively Reweighted Least Squares (IRLS) to eliminate the noise bursts present in the data. Recently, Brown (2001) attempted to remove acquisition tracks by estimating the systematic error between tracks at crossing points. Karpushin and Brown (2001) used a bank of prediction-error filters (PEFs) to whiten the residual along tracks. However, in most of these results, there is a loss of resolution that hampers our goal of seeing small features in the final image.

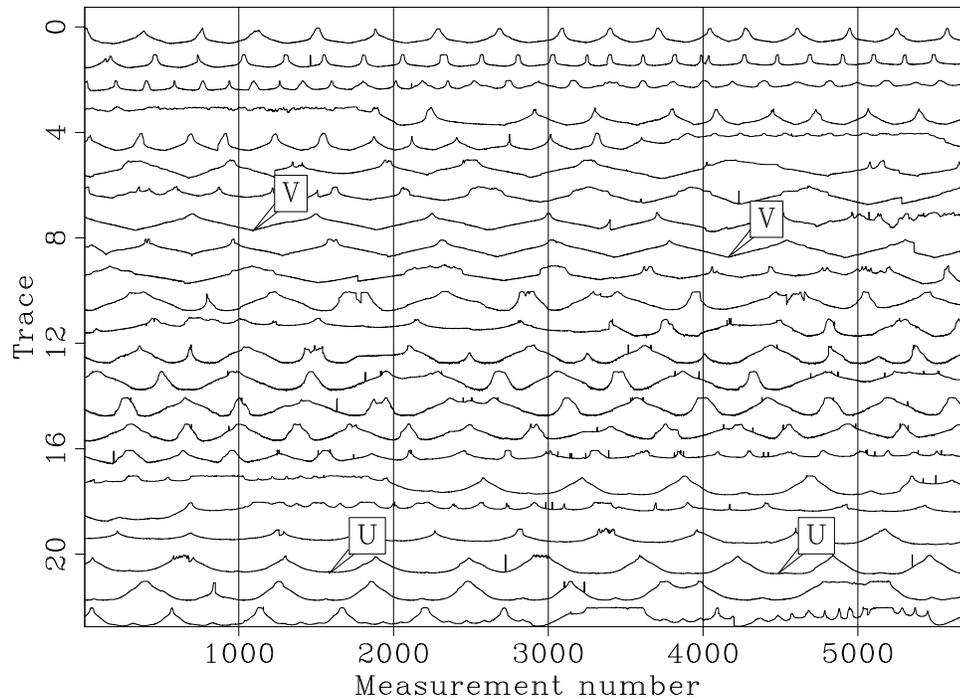


Figure 5.1: Depth of the Sea of Galilee along the vessel tracks. On one traverse across the lake, the depth record is “u” shaped. A few “v” shaped tracks result from vessel turn-arounds. All the points used for building the maps are shown here. Each point corresponds to one depth measurement inside the lake. Note that the input data are displayed as “seismic traces” for display purpose only. The long signal is broken into 23 strips of 5718 depth measurements. The horizontal axis corresponds to a measurement number. The depth appears as one long track although the surveying was done in several episodes that do not always continue the same track. `galilee-fig1` [ER]

Techniques developed in the preceding Chapters, i.e., the Huber norm from Chapter 2 and the modeling approach of Chapter 4, are both used to tackle the noise problems encountered with the mapping of this dataset. This data example also illustrates that the modeling of the tracks gives better results than filtering.

ATTENUATION OF THE NOISE BURSTS AND GLITCHES

Now, a formulation of the regridding problem is developed. Let \mathbf{h} be an abstract vector containing as components the water depth over a 2-D spatial mesh and \mathbf{d} be an abstract vector whose successive components represent depth along the vessel tracks. One way to grid irregular data is to minimize the length of the residual vector $\mathbf{r}_d(\mathbf{h})$

$$\mathbf{0} \approx \mathbf{r}_d = \mathbf{B}\mathbf{h} - \mathbf{d} \quad (5.1)$$

where \mathbf{B} is a 2-D linear interpolation operator and \mathbf{r}_d is the data residual. Note that sinc or spline interpolants could be used as well, but a simple linear interpolation operator is chosen to focus the analysis on the noise attenuation problem only. This fitting goal simply requires that the gridded data \mathbf{h} take on appropriate values where the data \mathbf{d} was collected. The bin size is 60 by 50 meters so that the number of data-points per bin is roughly constant and the aspect ratio of the lake is preserved in the number of samples in the east-west and north-south directions. A simple binning (without interpolation or inversion) of the raw data (Figure 5.1) is shown in Figure 5.2. Note that the north points to the top of the map in Figure 5.2. A coarser mesh would avoid the empty bins but lose resolution. As we refine the mesh for more detail, the number of empty bins grows as does the care needed in devising a technique for filling them. The black lines in Figure 5.2 are the ship tracks. Notice that some data points are outside the contour of the water. These must represent navigation errors. Unless data is collected everywhere on a very fine mesh, and depending on how the grid is parameterized, the regridding may leave holes on the mesh. We can eliminate the holes by adding some

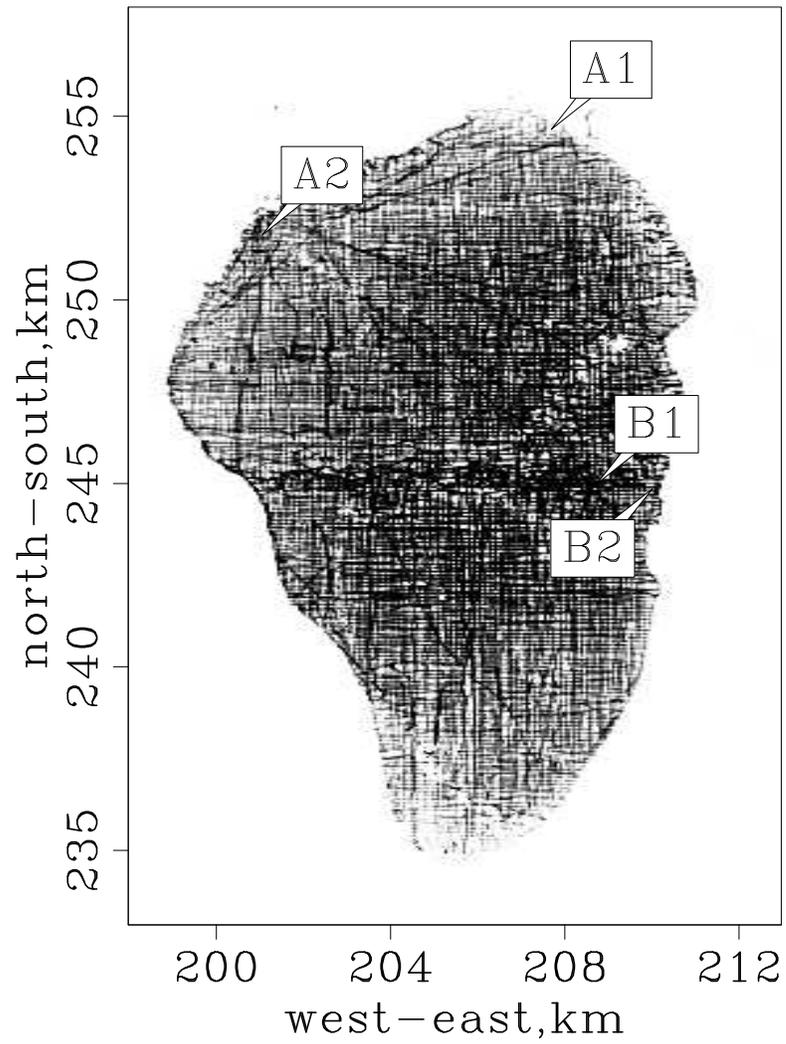


Figure 5.2: Simple binning of the raw data in Figure 5.1. The ship tracks and empty bins are visible and need to be accounted for in the inversion process. The north points to the top.

`galilee-fig1b` [ER]

regularization, like

$$\begin{aligned}\mathbf{0} &\approx \mathbf{r}_d = \mathbf{B}\mathbf{h} - \mathbf{d} \\ \mathbf{0} &\approx \epsilon\mathbf{r}_h = \epsilon\nabla\mathbf{h}\end{aligned}\tag{5.2}$$

where $\nabla = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}\right)$ and \mathbf{r}_h is the model space residual. We then minimize the misfit function

$$f(\mathbf{h}) = \|\mathbf{r}_d\|^2 + \epsilon^2\|\mathbf{r}_h\|^2\tag{5.3}$$

to estimate the interpolated map of the lake. The second term in equation (5.3) becomes Laplace's equation. In theory (Tarantola, 1987), the regularization operator (squared) should be the inverse model covariance operator given an a-priori model \mathbf{h}_0 . Since we do not have any a-priori model, I simply chose the gradient operator ∇ as a way of saying that the bottom of the lake is smooth. However, as pointed out by Harlan (1995), the regularization and the data fitting goal in equation (5.2) contradict each other. One equation tends to add details in the final map whereas the second one (the regularization) tends to smooth it. We can more easily balance these two goals by preconditioning the problem (Fomel, 2001).

Preconditioning for accelerated convergence

A generally available preconditioning method is to change variables so that the regularization operator becomes an identity matrix (Claerbout and Fomel, 2002). The gradient ∇ in equation (5.2) has no inverse, but its spectrum $-\nabla'\nabla$, which appears in equation (5.3), can be factored ($-\nabla'\nabla = \mathbf{H}'\mathbf{H}$) into triangular parts \mathbf{H} and \mathbf{H}' where \mathbf{H} is known as the Helix derivative and ($'$) denotes the adjoint. This \mathbf{H} is invertible by deconvolution (Claerbout, 1998). The fitting goals in equation (5.2) can be then rewritten

$$\begin{aligned}\mathbf{0} &\approx \mathbf{r}_d = \mathbf{B}\mathbf{H}^{-1}\mathbf{p} - \mathbf{d} \\ \mathbf{0} &\approx \epsilon\mathbf{r}_p = \epsilon\mathbf{p},\end{aligned}\tag{5.4}$$

with $\mathbf{p} = \mathbf{H}\mathbf{h} \approx \nabla\mathbf{h}$ and \mathbf{r}_p is the residual for the new variable \mathbf{p} . I then minimize the misfit function

$$g(\mathbf{p}) = \|\mathbf{r}_d\|^2 + \epsilon^2 \|\mathbf{r}_p\|^2 \quad (5.5)$$

and finally compute $\mathbf{h} = \mathbf{H}^{-1}\mathbf{p}$ to estimate the interpolated map of the lake. Experience shows that iterative solution for \mathbf{p} converges much more rapidly than iterative solution for \mathbf{h} thus showing that \mathbf{H} is a good choice for preconditioning. There is no simple way of knowing beforehand what is the best value of ϵ . Practitioners like to see solutions for various values of ϵ which can be computationally expensive. Practical exploratory data analysis is more pragmatic: without a simple clear theoretical basis, analysts generally begin from $\mathbf{p} = \mathbf{0}$ and then abandon the fitting goal $\mathbf{0} \approx \epsilon\mathbf{r}_p = \epsilon\mathbf{p}$ (Crawley, 2000; Rickett et al., 2001; Trad et al., 2003). Implicitly, they take $\epsilon = 0$. Then they examine the solution as a function of iteration, imagining that the solution at larger iterations corresponds to smaller ϵ , and that the solution at smaller iterations corresponds to larger ϵ . In all my computations, I follow this approach and omit the regularization in the estimation of the depth maps.

ℓ^1 norm

One main problem with the Galilee data is the presence of outliers in the middle of the lake. These spikes could be attenuated by editing or applying running median filters. However, the former involves human interpretation and the later might compromise small details by flattening and distorting the signal (Claerbout and Fomel, 2002). Therefore, inversion appears to be the best compromise by eliminating the spikes while honoring the data in an automated way. I introduce in equation (5.5) the Huber norm defined in Chapter 2 and minimize

$$g(\mathbf{p}) = |\mathbf{r}_d|_{Huber} + \epsilon^2 \|\mathbf{r}_p\|^2 \quad (5.6)$$

For the Huber threshold [equation (2.1)], I selected $\alpha = 10$ cm, which corresponds to the measurement error of the sounder. Figure 5.3a shows \mathbf{p} estimated with the ℓ^2 norm [equation (5.5)] after 50 iterations, which simulates a least-squares solution with damping. Note that the scale bar is not displayed whenever \mathbf{p} is plotted because its values are of little interest for

this analysis. Although \mathbf{p} appears to be a variable of mathematical interest only, in fact, the solution \mathbf{h} is so smooth that we have difficulty viewing it. We could view the two components of $\nabla\mathbf{h}$ but it happens that \mathbf{p} is a roughened version of \mathbf{h} . In addition, it is more convenient to view \mathbf{p} than the two images $\partial\mathbf{h}/\partial x$ and $\partial\mathbf{h}/\partial y$ because it is only a single component vector. We can see considerable spurious noise in the map of Figure 5.3a. In addition, we can see the vessel tracks in the north part of the map.

Figure 5.3b displays \mathbf{p} estimated with the ℓ^1 norm [equation (5.6)]. Most of the glitches are attenuated showing vessel tracks only. Some ancient shorelines in the west part and south part of the Sea of Galilee are now easy to identify (shown as “AS” in Figure 5.3b). In addition, we also start to see a “valley” in the middle of the lake (shown as “R” in Figure 5.3b). This feature is also present in Figure 5.3a where no attempts were made to remove the spikes. Therefore, this can be either a geological feature that represents the on-going rifting in this area or a track. The next section will prove that this valley is not a processing artifact or some noise not accounted for in our inversion scheme. The data outside the sea have been also partially removed. The tracks (shown as “T” in Figure 5.3) are still clearly visible after the attenuation of the outliers because they do not fit the model of the noise we are trying to remove.

Figures 5.4a,b show the bottom of the Sea of Galilee ($\mathbf{h} = \mathbf{H}^{-1}\mathbf{p}$) after inversion. Each line represents one east-west line of the interpolated data every 500 meters. The ℓ^1 result is a great improvement over the ℓ^2 maps. The glitches inside and outside the sea have disappeared. It is also pleasing to see that the ℓ^1 norm gives us positive depths everywhere. Although not everywhere visible in Figure 5.4, it is interesting to notice that we produce topography outside the lake. Indeed, the effect of regularization is to produce synthetic topography which is a natural continuation of the lake floor surface.

I have shown that the combined utilization of preconditioning and the Huber norm removes the spikes in the depth map of the Sea of Galilee. In the next section, I propose removing the ship tracks by introducing an operator in equation (5.6) that will model the coherent noise created by different weather and human conditions during the acquisition of the data.

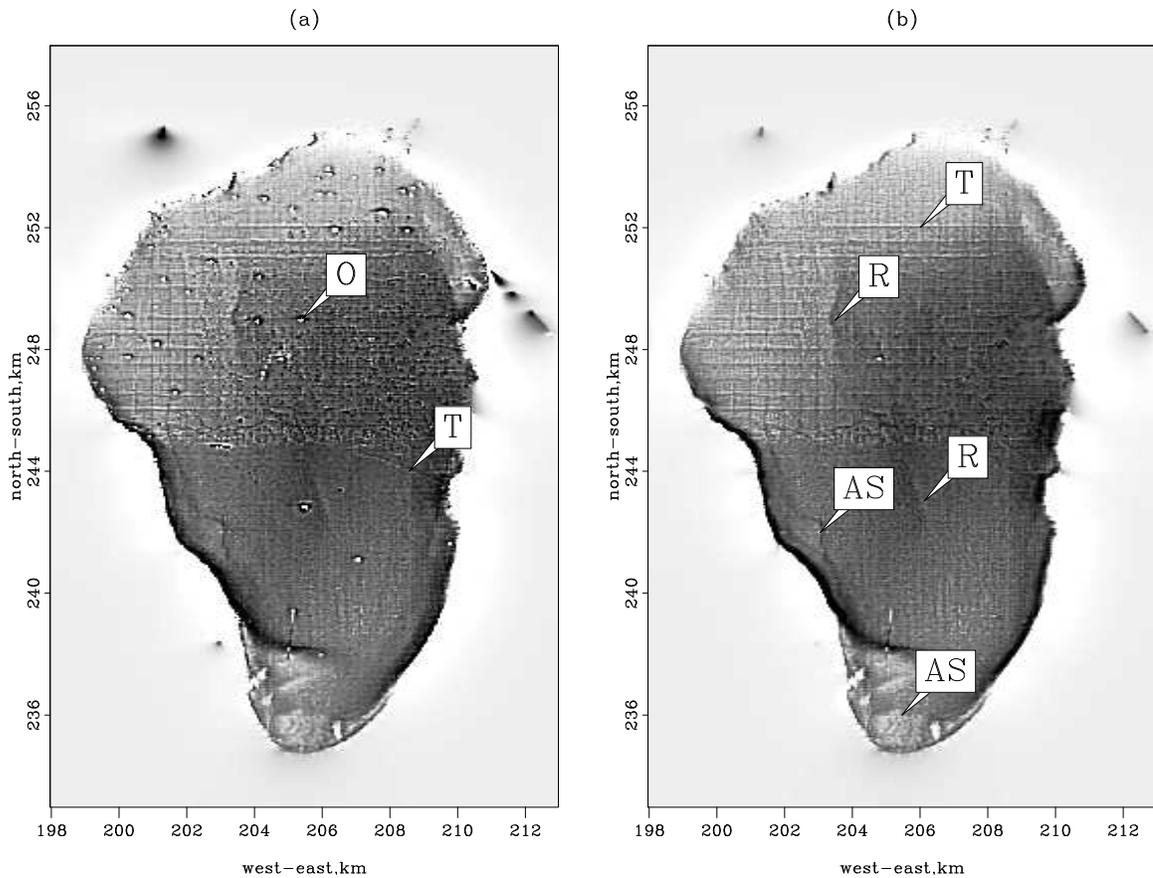


Figure 5.3: (a) $\mathbf{p} = \mathbf{H}\mathbf{h}$ estimated with equation (5.4) in a least-squares sense after 50 iterations, which simulates a least-squares solution with damping. (b) $\mathbf{p} = \mathbf{H}\mathbf{h}$ estimated with equation (5.6) in a ℓ^1 sense. The spikes have been correctly attenuated. Some interesting features are shown by the arrows: AS points to few ancient shores, O points to some outliers, T points to a few tracks and R points to a ridge. galilee-fig2 [ER,M]

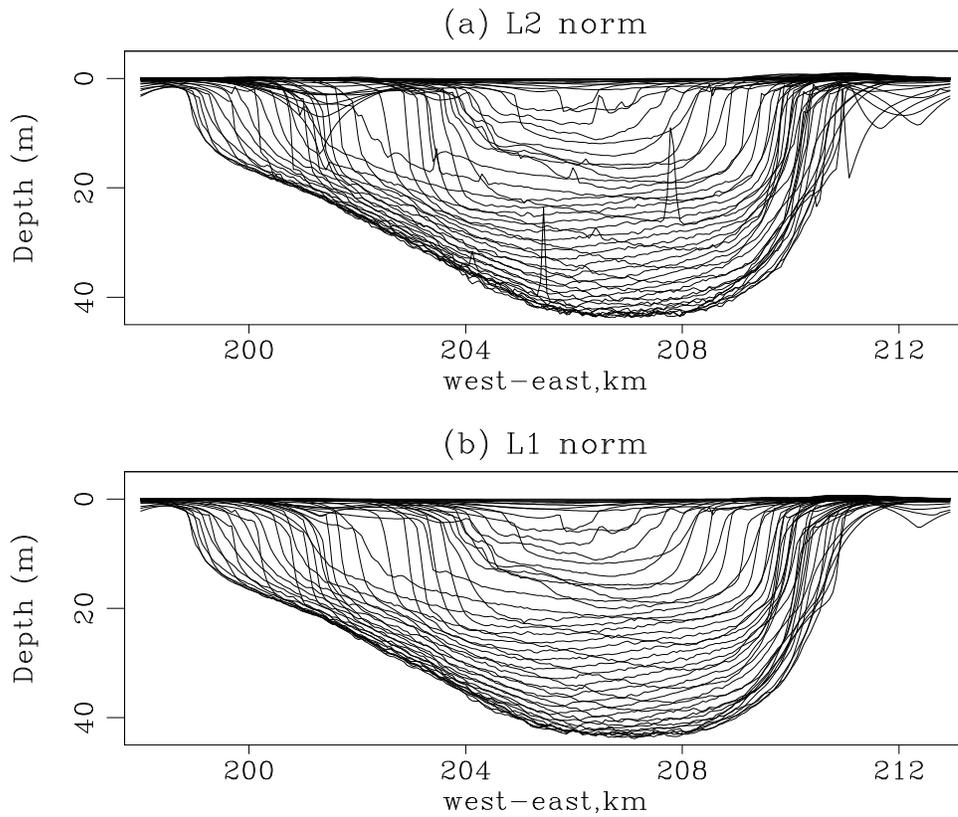


Figure 5.4: (a) View of the bottom of the lake ($\mathbf{h} = \mathbf{H}^{-1}\mathbf{p}$) with the ℓ^2 norm after 50 iterations, which simulates a least-squares solution with damping. (b) View of the bottom of the lake with the ℓ^1 norm. Note that with the ℓ^1 norm, the spikes have been attenuated. Each line represents one east-west track every 500 meters. galilee-fig3 [ER,M]

ATTENUATION OF THE SHIP TRACKS

We are now half way to a noise-free image. As seen in Figure 5.3, the vessel tracks overwhelm possible small details at the bottom of the Sea of Galilee. By small details, we mean objects of 100 meters and more, according to our bin size. In this section, I propose a strategy based on the assumption that the inconsistency between tracks comes mainly from different human and seasonal conditions during the acquisition. Unfortunately, no record of the weather and the time of the year the data were acquired is available. I presume that the depth differences between different acquisition tracks must be small and relatively smooth in time. That is why I propose introducing a leaky-integration operator (Claerbout, 1992) to model these “secular” data variations within our inversion scheme.

Abandoned strategy for attenuating tracks

An earlier strategy to remove the ship tracks is to filter the residual during the inversion as follows (Fomel and Claerbout, 1995):

$$\begin{aligned} \mathbf{0} &\approx \mathbf{r}_d = \frac{d}{ds}(\mathbf{B}\mathbf{H}^{-1}\mathbf{p} - \mathbf{d}) \\ \mathbf{0} &\approx \epsilon \mathbf{r}_p = \epsilon \mathbf{p}, \end{aligned} \tag{5.7}$$

where $\frac{d}{ds}$ is the derivative along the track and minimizing

$$g(\mathbf{p}) = |\mathbf{r}_d|_{\ell^1} + \epsilon^2 \|\mathbf{r}_p\|^2. \tag{5.8}$$

The purpose of the derivative is to remove the drift, which is assumed to have a zero frequency component. The derivative poses two types of problem, however. First, it attenuates the bathymetry as well, which has frequency components very close to zero. Second, it creates more bad data points for the high frequency noise. Both effects are illustrated in Figure 5.5b where almost all the details inside the lake have disappeared after the minimization of equation (5.8). In addition, the map is more noisy due to the aggravating effect of the derivative on bad data points.

Recently, Brown (2001) proposed estimating systematic errors between tracks by analyzing measurements at points where the acquisition swaths cross. This approach has the advantage of preserving the resolution of the depth map compared to the derivative along the tracks. Brown (2001) uses this idea as a preprocessing step, however. Based on Brown's idea (2001) and following Chapter 4, I propose introducing an operator that will adaptively model and subtract the systematic shift within the inversion scheme.

In the next section, I show that by incorporating a modeling operator for the drift in the data, the ship tracks can be effectively removed without any loss of resolution in the estimated depth map.

A new fitting goal

Now, I develop the new idea of removing the tracks by adaptively subtracting them within our inversion scheme. Building on Chapter 4, I introduce a modeling operator for the ship tracks inside our fitting goal in equation (5.5) as follows:

$$\begin{aligned} \mathbf{0} &\approx \mathbf{r}_d = \mathbf{B}\mathbf{H}^{-1}\mathbf{p} + \mathbf{L}\mathbf{q} - \mathbf{d} \\ \mathbf{0} &\approx \epsilon_1 \mathbf{r}_p = \epsilon_1 \mathbf{p} \\ \mathbf{0} &\approx \epsilon_2 \mathbf{r}_q = \epsilon_2 \mathbf{q} \end{aligned} \tag{5.9}$$

where \mathbf{L} is a drift modeling operator (leaky integration), \mathbf{q} is a new variable of the inversion. The following misfit function can then be minimized

$$g_2(\mathbf{p}, \mathbf{q}) = |\mathbf{r}_d|_{Huber} + \epsilon_1^2 \|\mathbf{r}_p\|^2 + \epsilon_2^2 \|\mathbf{r}_q\|^2 \tag{5.10}$$

where $\mathbf{h} = \mathbf{H}^{-1}\mathbf{p}$ estimates the interpolated map of the lake. Again, I set $\epsilon_1 = \epsilon_2 = 0$ and do not iterate to completion. The explicit elimination of the two regularization parameters ϵ_1 and ϵ_2 needs to be accounted for. In practice, in the presence of crosstalks, they allow us to put the common components of the data in whichever space we choose. To accommodate my choice of $\epsilon_1 = \epsilon_2 = 0$, a constant $\gamma = \epsilon_1/\epsilon_2$ is added to the modeling operator for the drift (similar to equation (4.13) in Chapter 4). Then, for the operator \mathbf{L} , I choose a leaky integration operator

such that $\mathbf{y} = \gamma \mathbf{L} \mathbf{q}$ is the portion of data value \mathbf{d} that results from drift. Consistent with the way I use a rough variable \mathbf{p} to represent the smooth water depth \mathbf{h} , I now represent (for the purpose of speeding iteration) \mathbf{y} by a rougher function \mathbf{q} . The operator \mathbf{L} has the following recursive form

$$y_s = \rho y_{s-1} + q_s \quad s \text{ increasing along the data track.} \quad (5.11)$$

The parameter ρ controls the decay of the integration. For $\rho = 1$, leaky integration represents causal integration. The operator \mathbf{L} is then appropriate to model the secular variations implied by the different season and human conditions during the data acquisition. We simply have to choose a value of ρ that best represents the variations between the different tracks. This task is rather difficult to achieve: if ρ is too small, we might not be able to remove the drift and if ρ is too big, we might remove the drift and the bathymetry. Therefore, ρ was carefully selected by starting from $\rho = 0.999$, interpolating with this value, looking at the final result, and decreasing ρ by 0.001 if necessary. I repeated this process until all the tracks were attenuated. At the end of this exhaustive search, the value $\rho = 0.99$ removes the tracks while preserving the bathymetry. I keep this value of $\rho = 0.99$ for all remaining results involving track attenuation. I show that the operator \mathbf{L} removes most of the vessel tracks present in Figure 5.3.

The choice of γ in equation (5.9) is also critical. I tried different values by starting from a very small number and increasing it slowly. I then chose the smallest value that removed enough tracks in the final image ($\gamma = 0.08$). Nemeth et al. (2000) demonstrates that the noise (the tracks) and signal (the depth) can be separated in equation (5.9) if the two operators \mathbf{L} and $\mathbf{B}\mathbf{H}^{-1}$ do not model similar components of the data space. Figures 5.5a and 5.5c display a comparison of the estimated \mathbf{p} with or without the attenuation of the vessel tracks. It is delightful that Figure 5.5c is essentially track-free without any loss of details compared to Figure 5.5a. The difference plot in Figure 5.5d between the two results corroborates this and does not show any geological feature. A close-up of Figure 5.5 is displayed in Figure 5.6. The differences between the proposed techniques are clearly visible.

Comparing Figure 5.5c and Figure 5.5b, we see that the drift-modeling strategy [equation

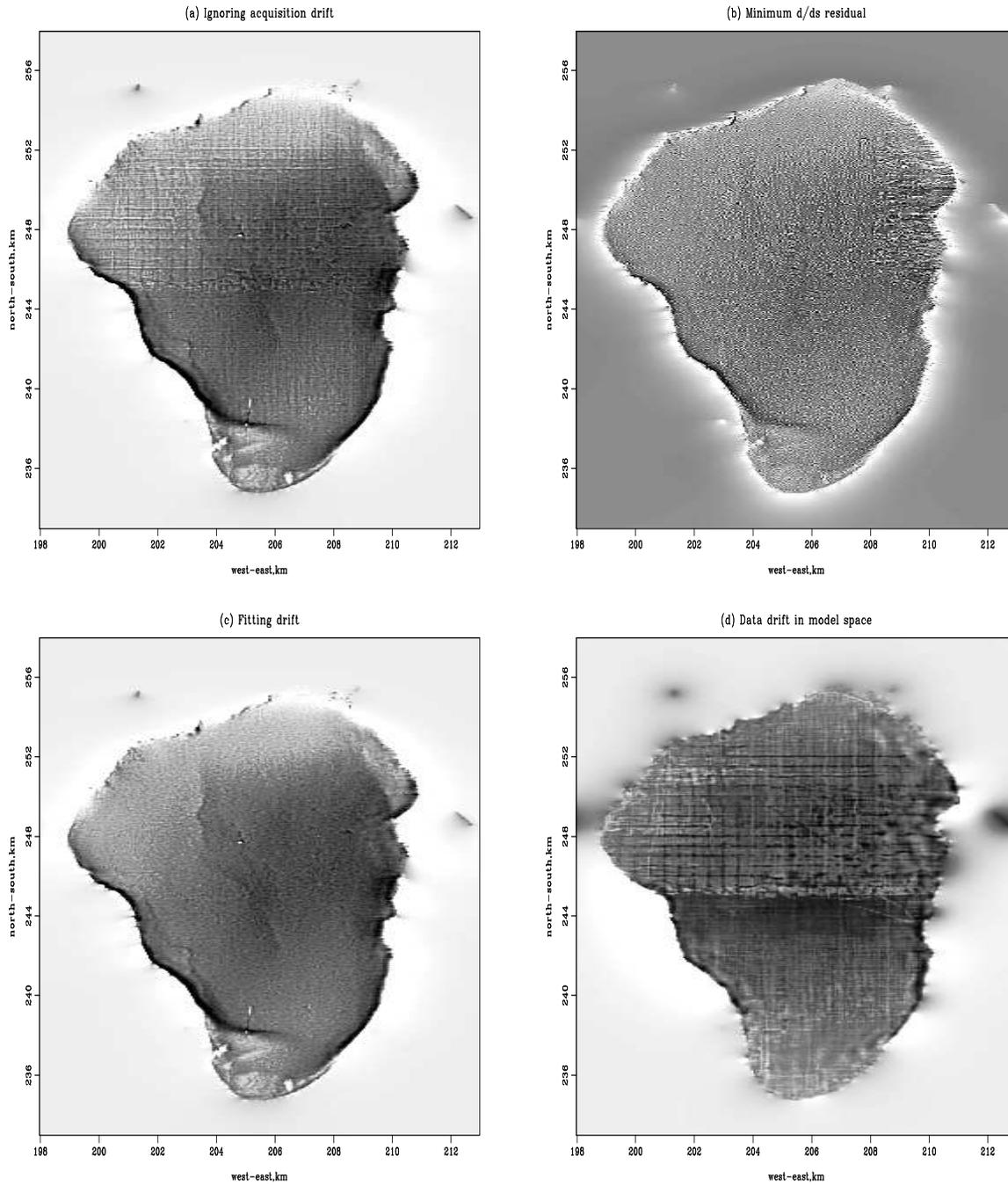


Figure 5.5: (a) Estimated \mathbf{p} without attenuation of the tracks, i.e., equation (5.5). (b) Estimated \mathbf{p} with the derivative along the tracks, i.e., equation (5.7). (c) Estimated \mathbf{p} without tracks, i.e., equation (5.9). (d) Recorder drift in model space $\mathbf{B}^T \mathbf{L} \mathbf{q}$. galilee-fig4 [ER,M]

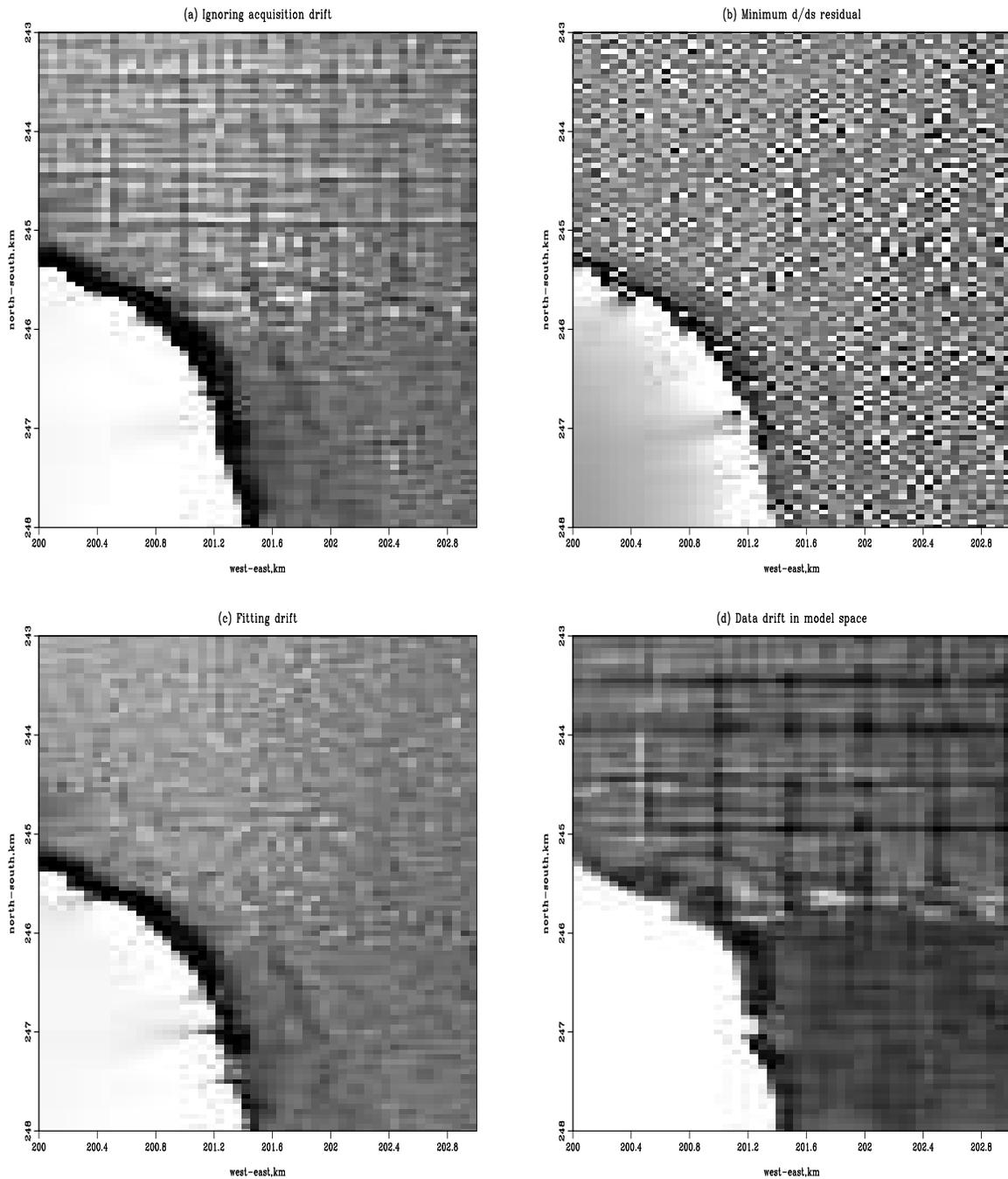


Figure 5.6: Close-ups of the western shore of the Sea of Galilee. (a) Estimated \mathbf{p} without attenuation of the tracks, i.e., equation (5.5). (b) Estimated \mathbf{p} with the derivative along the tracks, i.e., equation (5.7). (c) Estimated \mathbf{p} without tracks, i.e., equation (5.9). (d) Recorder drift in model space $\mathbf{B}'\mathbf{L}\mathbf{q}$. galilee-fig4b [ER,M]

5.9] works much better than the noise-filtering strategy [equation 5.7]. One possible explanation for the difference between the two results is that the modeling approach is more adaptive than the filtering of the residual. Indeed, by introducing the modeling operator, we basically look for the best \mathbf{q} that models the drift of the data on each track at each point. The price to pay is an increase of the number of unknowns in equation (5.9). The reward is a surgically removed acquisition footprint. Notice that we can identify the ancient shorelines in the west and east parts of the lake very well.

To better understand what is done, Figures 5.7 and 5.8 show some segments of the input data (\mathbf{d}), the estimated noise-free data ($\mathbf{BH}^{-1}\mathbf{p}$), the estimated secular variations ($\gamma\mathbf{Lq}$) and the residual ($\mathbf{BH}^{-1}\mathbf{p} + \gamma\mathbf{Lq} - \mathbf{d}$) after inversion. The estimated noise-free data in Figures 5.7b and 5.8b show no remaining spikes. The effect of the track attenuation is more difficult to see because the amplitude of the drift is much smaller than the amplitude of the measurements. Notice in Figure 5.7c that the estimated drift seems to have reasonable amplitudes: the average drift is around 15 cm for an accuracy of about 10 cm for the measurements. We also observe that the estimated drift is relatively constant throughout Figure 5.7c. Now, looking at the estimated drift for another portion of the data (Figure 5.8c), notice that the drift has more variance than in Figure 5.7c and oscillates between 0 to 2 m, which is too much. In addition, the estimated drift seems to follow the bathymetry of the lake in Figure 5.8a. Decreasing γ would attenuate the drift component with the effect of increasing the tracks in the final image, however.

Looking closely at the residual (Figure 5.8d), the drift is large where the data are noisy (Figure 5.8a). It is possible that the day of acquisition was very windy, which is not a rare weather condition for the Sea of Galilee (Volohonsky et al., 1983). Thus, the wind forces the water to pile-up on one side of the lake which can explain the lower water level on the other side. A rapid calculation shows that the seiche period for the Sea of Galilee is roughly 40 mn (assuming a lake length of 20 km and an average depth of 30 m), which is well within a day of data acquisition. In addition, the strong wind in the middle of the lake induces noisy measurements because of the waves and of the erratic movement of the ship. It is also possible that the depth sounder was not working properly that day and had problems to correctly measure the deepest part of the lake. These causes could probably explain the shape

and amplitude of the estimated drift in Figure 5.8c, but we can't be absolutely sure. It is very unfortunate that no daily logs of the survey were kept in order to better interpret these results, especially for such a noisy dataset.

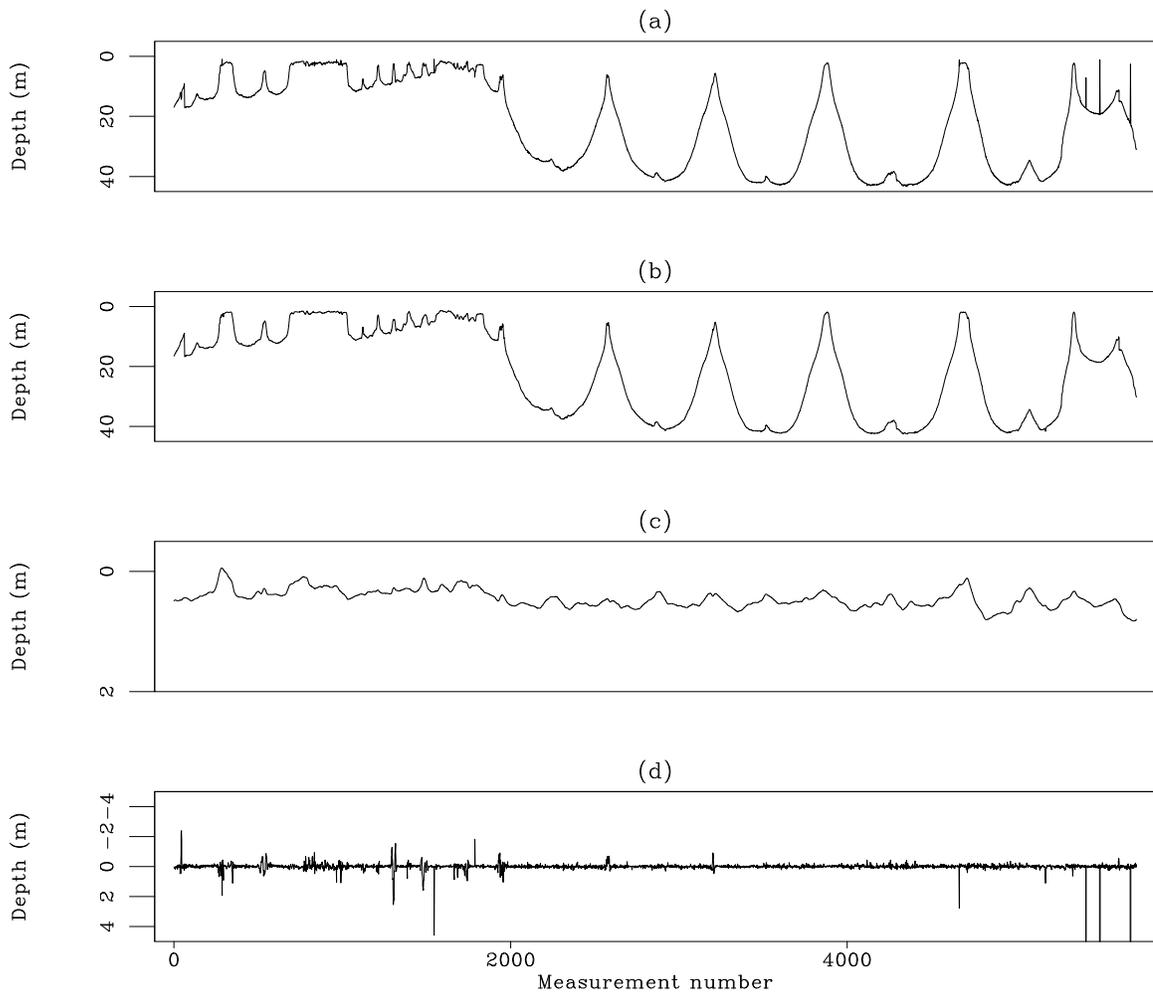


Figure 5.7: (a) Input data acquired between A1 and B1 in Figure 5.2. The ship is approximately moving bottom to top going east from A1 to B1. (b) $\mathbf{BH}^{-1}\mathbf{p}$ estimated after inversion, i.e., the estimated noise-free data. (c) Estimated drift after inversion. (d) Data residual after inversion. The horizontal axis represents the measurement number. galilee-fig5 [ER,M]

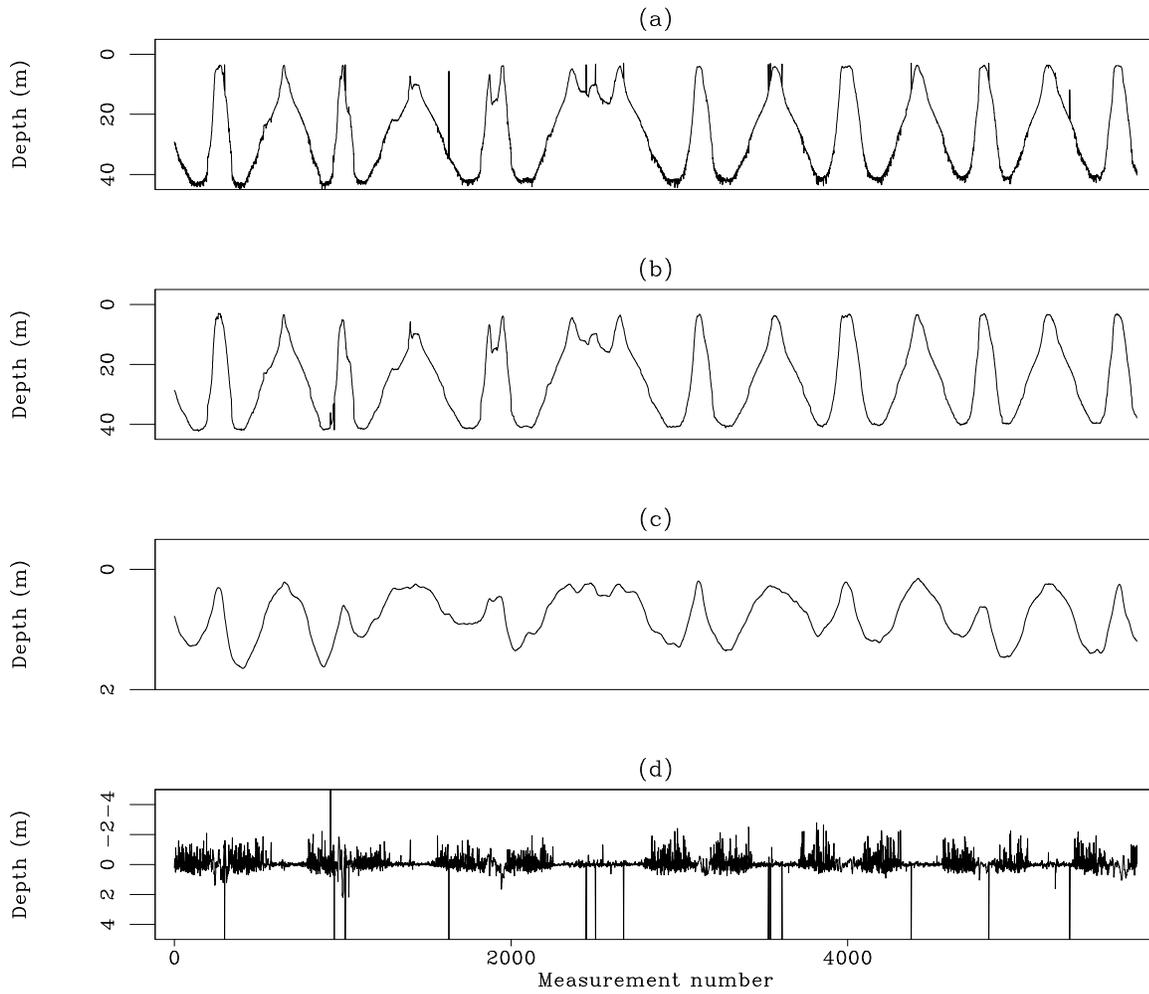


Figure 5.8: (a) Input data acquired between A2 and B2 in Figure 5.2. The ship is approximately moving right to left going south from A2 to B2. (b) $\mathbf{BH}^{-1}\mathbf{p}$ estimated after inversion, i.e., the estimated noise-free data. (c) Estimated drift after inversion. (d) Data residual after inversion. The horizontal axis represents the measurement number. galilee-fig6 [ER]

CONCLUSION

The interpolation of the Sea of Galilee dataset on a regular grid is addressed as a noise attenuation problem where the Huber norm (Chapter 2) and the noise modeling approach (Chapter 4) simultaneously eliminate the damaging effects of bad data points and tracks. This data set exemplifies the need for identifying the correct statistics in order to choose the right norm but also the need for modeling the noise as often as possible. In the next Chapter, the pattern-based approach of Chapter 4 is used to remove surface related multiples.

ACKNOWLEDGMENTS

I thank Zvi Ben-Avraham at Tel-Aviv University for providing the dataset.

Chapter 6

Multiple attenuation: Theory and Practice

SUMMARY

Primaries (signal) and multiples (noise) exhibit often different kinematics and amplitudes (i.e., patterns) in time and space. Multidimensional prediction-error filters (PEFs) approximate these patterns to separate noise and signal in a least-square sense. These filters are time-space variant to handle the non-stationarity of multi-offset seismic data. PEFs for the primaries and multiples are estimated from pattern models. In an ideal case where accurate pattern models of both noise and signal exist, the pattern-based method recovers the primaries while preserving their amplitudes. In the more general case, the pattern model of the multiples is obtained by using the data as prediction operators. The pattern model of the primaries is obtained by convolving the noise PEFs with the input data. In this situation, 3-D PEFs are preferred to separate (in prestack data) the multiples properly and to preserve the primaries. Comparisons of the proposed method with adaptive subtraction with a ℓ^2 norm demonstrate that for a given multiple model, the pattern-based approach attenuates the multiples and recovers the primaries generally better. In addition, tests on a 2-D line from the Gulf of Mexico demonstrate that the proposed technique copes fairly well with modeling inadequacies present in the multiple prediction.

INTRODUCTION

In the presence of complex geology where multipathing, illumination gaps, and coherent noise are prevalent, the most advanced techniques need to be used for both preprocessing and imaging. For multiple attenuation, Weglein (1999) shows that current technology may be divided into filtering methods, which exploit the periodicity and the separability (move-out discrepancies) of the multiples, and wavefield methods, which first predict, then subtract the multiples (Verschuur et al., 1992; Dragoset and MacKay, 1993; Weglein et al., 1997).

Traditionally, filtering techniques are chosen because of their robustness and low cost. However, filtering techniques have some limitations when tackling multiples in complex media. For example predictive deconvolution in the ray parameter domain fails when the water bottom is not flat (Treitel et al., 1982). Furthermore, numerous authors (Matson et al., 1999; Bishop et al., 2001; Paffenholz et al., 2002) have presented cases where wavefield approaches such as surface-related multiple elimination (SRME) attenuate multiples much better than filtering techniques such as radon-based methods.

Wavefield techniques start usually with a prediction step where surface-related multiples are modeled from the data with (Wiggins, 1988; Lokshtanov, 1999) or without any subsurface information. Then the multiples are subtracted from the data. Both aspects of the multiple attenuation procedure are important but most of today's efforts are concentrated on the prediction, not on the subtraction.

With SRME (Verschuur et al., 1992), two significant assumptions are usually made for the subtraction. First, it is assumed that the signal has minimum energy, leading to the adaptive subtraction of the multiples with a ℓ^2 norm. This assumption might not hold where primaries and multiples interfere (Spitz, 1999), however. Chapter 3 shows that when primaries are much stronger than multiples, the ℓ^1 norm should be used instead. Second, it is assumed that the multiples are accurately modeled. This point relies on the acquisition or the interpolation/extrapolation of the data to provide the necessary traces for the prediction step. In practice, however, the data are never acquired densely enough and the interpolation schemes are never perfect, especially with sparse acquisition geometries. For instance, 3-D seismic

marine data, although well sampled in the inline direction, lack of crossline offsets. Consequently, the prediction is not as accurate as required. Therefore, other subtraction techniques are desirable when adaptive subtraction fails to recover the primaries and when the multiple model is not precise enough. As stated by Berkhout (2004, personal communication), the subtraction step is the weakest component of SRME, and more work needs to be done in this direction.

A new class of multiple attenuation techniques has recently emerged to circumvent some of the limitations of adaptive subtraction and of the modeling. These techniques are called pattern-based because they discriminate primaries from multiples according to their multidimensional spectra (Manin and Spitz, 1995; Guitton and Cambois, 1998; Brown and Clapp, 2000; Fomel, 2002). This Chapter presents one implementation of a pattern-based approach based on prediction-error filters (PEFs). By construction, PEFs approximate the inverse spectrum of the data from which they are estimated (Burg, 1975). Therefore, PEFs can serve as proxies for the patterns of both primaries and multiples and used for signal-noise separation (Guitton, 2002a).

This Chapter begins with a description of non-stationary PEFs estimation. The helical boundary conditions (Mersereau and Dudgeon, 1974; Claerbout, 1998) are used to estimate 2-D and 3-D filters. Both noise (multiples) and signal (primaries) PEFs are needed for the attenuation. Because the signal is usually unknown, a method that needs only the noise and the input data to derive a pattern model for the signal is presented. Having estimated the noise and signal PEFs, the multiple subtraction method (pattern-based) that separates primaries from multiples in a least-squares sense is then described.

To illustrate the efficiency of the pattern-based approach, surface-related multiples are attenuated for the Sigsbee2B synthetic dataset. The results of multiple attenuation are analyzed after migration to assess the effects of the proposed technique on the primaries. This example illustrates that the pattern-based approach can lead to a very good elimination of the multiples if an accurate pattern model for both primaries and multiples is available. In addition, this dataset shows that 3-D PEFs preserve the primaries better than 2-D filters.

Then, adaptive and pattern-based subtractions are compared on a synthetic data example

provided by BP. This example proves that when multiples and primaries are spatially uncorrelated (i.e., different patterns), PEFs attenuate multiples better than adaptive subtraction. Finally, I show multiple attenuation results on a Gulf of Mexico 2-D line before and after migration. This last example illustrates that the pattern-based method is robust to model inadequacies.

FILTER ESTIMATION

The key assumption of the proposed multiple attenuation technique is that primaries and multiples have different multidimensional spectra that PEFs can approximate (Claerbout, 1992; Spitz, 1999). These filters can be interpreted as multidimensional dip filters. The estimation of these filters is based on the spatial predictability of seismic events. Therefore, PEFs for primaries and multiples are needed prior to the subtraction step. This section describes how these filters are estimated.

The PEFs used here are time-space domain non-stationary filters to cope with the variability of seismic data with time, offset and shot position. Implementing non-stationary filters is not an easy task. A possible solution is to break-up the dataset into patches and estimate a filter for each patch. However, reassembling these patches create edge effects in the overlapping zones. Alternatively, non-stationary convolution or combination can be used to estimate one filter per data point (Margrave, 1998) or more realistically, per micro-patch (Crawley, 2000). A micro-patch is made of a succession of data points that share the same filter. With this technique, the filters can vary smoothly across the dataset while leaving almost no edge effect.

A detailed description of the non-stationary filters is given in Appendix C. Here I present the most important steps of the filter estimation procedure only. Calling \mathbf{Y} the non-stationary combination matrix (Margrave, 1998) with the data vector \mathbf{y} from which we want to estimate the filters and \mathbf{a} the unknown PEFs coefficients, one way to estimate the PEFs is to minimize the length of residual vectors \mathbf{r}_y (Claerbout and Fomel, 2002):

$$\mathbf{0} \approx \mathbf{r}_y = \mathbf{Y}\mathbf{a}. \quad (6.1)$$

This last expression is also called fitting goal. By definition of the PEFs, the first coefficient is always one (Figure 6.1). To take this requirement into account, equation (6.1) becomes

$$\mathbf{0} \approx \mathbf{r}_y = \mathbf{Y}\mathbf{K}\mathbf{a} + \mathbf{y}, \quad (6.2)$$

where \mathbf{K} is a masking operator that forces the first coefficient of the PEFs to be one.

If one filter is used per data point, the matrix of unknown coefficients \mathbf{a} can be enormous, making the problem very underdetermined. This difficulty can be overcome in two ways. Firstly, the filter is kept constant inside a micro-patch. Secondly, a smoothing operator \mathbf{R} is introduced to penalize strong variations between neighboring filters. Both strategies are considered here. Introducing the Laplacian operator \mathbf{R} , equation (6.2) is augmented as follows:

$$\begin{aligned} \mathbf{0} &\approx \mathbf{r}_y = \mathbf{Y}\mathbf{K}\mathbf{a} + \mathbf{y}, \\ \mathbf{0} &\approx \epsilon \mathbf{r}_a = \epsilon \mathbf{R}\mathbf{a}, \end{aligned} \quad (6.3)$$

where ϵ is a trade-off parameter between coefficient estimation and filter smoothing. The Laplacian operator smoothes along two (i.e., time/offset) or three axes (i.e., time/offset/shot) depending on the PEF's dimensions. A least-squares estimate of the PEF's coefficients leads to the following objective function:

$$f(\mathbf{a}) = \|\mathbf{r}_y\|^2 + \epsilon^2 \|\mathbf{r}_a\|^2, \quad (6.4)$$

where \mathbf{a} is estimated iteratively with the conjugate-gradient method and ϵ is chosen by trial and error. Note that the Huber norm of Chapter 2 is not used here.

The amplitudes of seismic data vary across offset, shot and time. Large amplitude variations can be troublesome with least-squares inversion because they tend to bias the final result (Claerbout, 1992) by ignoring or over-fitting some areas of the data. Therefore, it is important to make sure that these amplitude variations do not affect our processing. One solution is to apply a weight to the data prior to the inversion like Automatic Gain Control (AGC) or a geometrical spreading correction. Alternatively, a weight \mathbf{W} can be incorporated inside the fitting

goals in equation (6.3):

$$\begin{aligned} \mathbf{0} &\approx \mathbf{r}_y = \mathbf{W}(\mathbf{Y}\mathbf{K}\mathbf{a} + \mathbf{y}), \\ \mathbf{0} &\approx \epsilon \mathbf{r}_a = \epsilon \mathbf{R}\mathbf{a}. \end{aligned} \quad (6.5)$$

This weight can be interpreted as a change of norm consistent with the data, similar to ℓ^1 norm. In this Chapter, \mathbf{W} is obtained by computing for each shot gather

$$W_{i,i} = \frac{y_{AGC_i}}{y_i} \quad (6.6)$$

where \mathbf{y}_{AGC} is the data vector with AGC applied. If division by zero occurs, the weight is set to zero. This weight can also incorporate a mute zone where no data are present. Thanks to the helical boundary conditions (Mersereau and Dudgeon, 1974; Claerbout, 1998), the PEFs can have any dimension. In this Chapter, I use 2-D (time/offset) and 3-D (time/offset/shot position) filters (Figure 6.1) and demonstrate that 3-D filters lead to the best noise attenuation results.

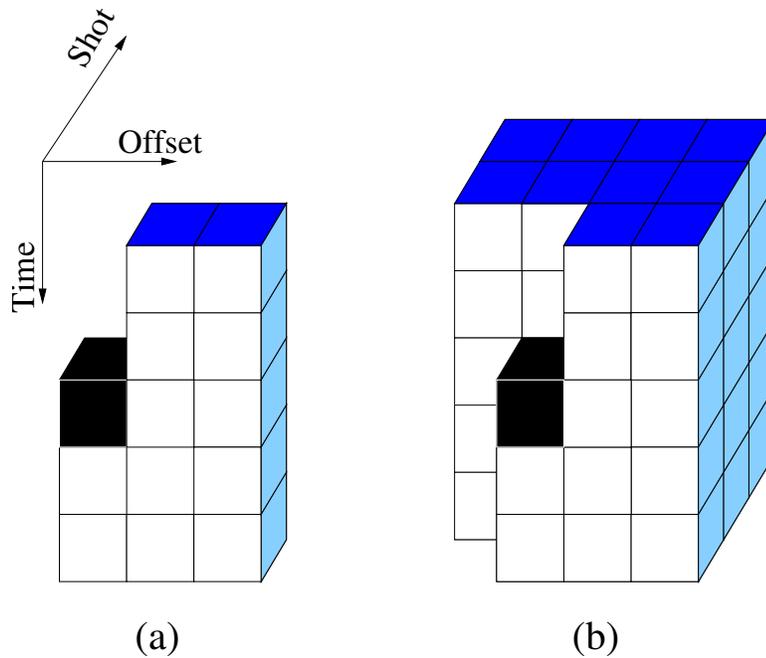


Figure 6.1: prediction-error filters in (a) 2-D and (b) 3-D. The first coefficient in black is always one. `mul2d-pef` [NR]

When 2-D filters are used, the multiple attenuation is performed on one shot gather at a time. When 3-D filters are used, the multiple attenuation is performed on one macro-gather at a time. A macro-gather is a cube made of adjacent shots with all the offsets and time samples. There is an overlapping zone of five shot gathers between neighboring macro-patches. When the multiple attenuation is done, the macro-gathers are reassembled to form the final result. The next section describes how to choose the model \mathbf{Y} in equation (6.5) when PEFs for multiples and primaries are estimated.

SIGNAL AND NOISE PEFS ESTIMATION

For multiple attenuation, non-stationary PEFs \mathbf{N} and \mathbf{S} for the multiples and the primaries, respectively, are needed. Therefore, models for the noise and the signal are necessary. For surface-related multiples, the multiple model may be provided by Surface-Related Multiple Prediction (SRMP) (Verschuur et al., 1992; van Dedem, 2002), which yields kinematically correct prestack model of the multiples, especially with 2-D data (see Appendix D for details). In 3-D, with marine acquisition geometry, the crossline offsets need to be heavily interpolated first. The interpolation results have then a direct impact on the prediction, as illustrated in Chapter 7.

Amplitude-wise, an accurate surface-related multiple model can be derived if (1) the source wavelet is known, (2) the surface source and receiver coverage is large and dense enough, and (3) all the terms of the series that model different orders of multiples are incorporated (Verschuur et al., 1992). Often in practice, a single convolution of the input data (i.e., one term of the series) is usually performed, giving a multiple model with erroneous relative amplitudes for high-order multiples (see Appendix D, equation (D.12)). Dragoset and Jericevic (1998) detail the possible flaws introduced in the prediction due to acquisition parameters.

Because PEFs estimate patterns, incorrect relative amplitudes and kinematic errors can affect multiple suppression. However, as we shall see later, 3-D filters seem to cope better with noise modeling inadequacies.

Signal PEFs are more difficult to estimate since the primaries are usually unknown. As a possible solution to this problem, Spitz (1999) estimates a signal PEF \mathbf{S} by deconvolving a data PEF \mathbf{D} , estimated from the data, by a noise PEF \mathbf{N} . With this process, Spitz assumes that

$$\mathbf{D} = \mathbf{S}\mathbf{N}. \quad (6.7)$$

I call equation (6.7) the Spitz approximation. Note that \mathbf{D} , \mathbf{N} , and \mathbf{S} are matrices for the combinations with the non-stationary PEFs (Margrave, 1998). These matrices are very sparse and are never formed in practice (Claerbout and Fomel, 2002). Equation (6.7) can be retrieved by considering a simple 1-D example using the Z-transform notations (Claerbout, 1976) for a data PEF $D_{1D}(Z)$, a signal PEF $S_{1D}(Z)$, and a noise PEF $N_{1D}(Z)$. Extension to more dimensions is straightforward using the helical boundary conditions (Claerbout and Fomel, 2002). Because PEFs have the inverse spectrum of the data from which they have been estimated (Burg, 1975), we have (omitting Z for clarity purposes)

$$\frac{1}{D_{1D}^* D_{1D}} = \frac{1}{S_{1D}^* S_{1D}} + \frac{1}{N_{1D}^* N_{1D}}, \quad (6.8)$$

where (*) is the complex conjugate. Equation (6.8) simply states that the spectrum of the data is equal to the spectrum of the noise plus the spectrum of the signal. Equation (6.8) can be written as follows:

$$D_{1D}^* D_{1D} = \frac{S_{1D}^* N_{1D}^* S_{1D} N_{1D}}{N_{1D}^* N_{1D} + S_{1D}^* S_{1D}}. \quad (6.9)$$

Now, because PEFs are important where they are small (i.e., where they attenuate seismic events), the denominator can be neglected:

$$D_{1D}^* D_{1D} \approx S_{1D}^* N_{1D}^* S_{1D} N_{1D}, \quad (6.10)$$

which leads to the Spitz approximation in equation (6.7). The PEFs \mathbf{D} and \mathbf{N} can be easily estimated because the data vector and a noise model are often available. However, estimating the signal PEFs requires a potentially unstable non-stationary deconvolution $\mathbf{S} = \mathbf{D}\mathbf{N}^{-1}$ (Rickett, 2001) in equation (6.7). To avoid the deconvolution step, the noise PEFs are convolved

with the data:

$$\mathbf{u} = \mathbf{N}\mathbf{d}, \quad (6.11)$$

where \mathbf{u} is the result of the convolution. Estimating the PEFs \mathbf{U} for \mathbf{u} gives by definition of the PEFs (Claerbout and Fomel, 2002):

$$\mathbf{0} \approx \mathbf{U}\mathbf{u}. \quad (6.12)$$

Then, from the Spitz approximation in equation (6.7), we have

$$\mathbf{0} \approx \mathbf{U}\mathbf{u} = \mathbf{U}\mathbf{N}\mathbf{d} = \mathbf{D}\mathbf{d} = \mathbf{S}\mathbf{N}\mathbf{d}, \quad (6.13)$$

and $\mathbf{U} = \mathbf{S}$. Therefore, by convolving the data with the noise PEFs, signal PEFs consistent with the Spitz approximation can be computed. Again, an important assumption is that signal and noise are uncorrelated. The PEFs for the primaries (\mathbf{S}) and the multiples (\mathbf{N}) are estimated directly from the data and the model of the multiples. These filters approximate the multidimensional spectra of the noise and signal.

MULTIPLE ATTENUATION

When the noise and signal PEFs have been determined, the task of multiple attenuation follows. First consider that the seismic data \mathbf{d} are the sum of signal (primaries) and noise (multiples):

$$\mathbf{d} = \mathbf{s} + \mathbf{n}, \quad (6.14)$$

where \mathbf{s} is the signal we want to preserve and \mathbf{n} the noise we wish to attenuate.

By definition of the signal and noise PEFs, the following relationships hold:

$$\begin{aligned} \mathbf{N}\mathbf{n} &\approx \mathbf{0}, \\ \mathbf{S}\mathbf{s} &\approx \mathbf{0}. \end{aligned} \quad (6.15)$$

Equations (6.14) and (6.15) can be combined to solve a constrained problem to separate signal from noise as follows:

$$\begin{aligned} \mathbf{0} &\approx \mathbf{r}_n = \mathbf{Nn}, \\ \mathbf{0} &\approx \epsilon \mathbf{r}_s = \epsilon \mathbf{Ss}, \\ \text{subject to } &\leftrightarrow \mathbf{d} = \mathbf{s} + \mathbf{n}. \end{aligned} \quad (6.16)$$

The scalar ϵ is related to the S/N ratio. In practice, however, ϵ is estimated by trial and error. Replacing \mathbf{n} by $\mathbf{s} - \mathbf{d}$ in equation (6.16) yields

$$\begin{aligned} \mathbf{0} &\approx \mathbf{r}_n = \mathbf{N}\mathbf{s} - \mathbf{N}\mathbf{d}, \\ \mathbf{0} &\approx \epsilon \mathbf{r}_s = \epsilon \mathbf{S}\mathbf{s}. \end{aligned} \quad (6.17)$$

Sometimes it is useful to add a masking operator for the noise and signal residuals \mathbf{r}_n and \mathbf{r}_s when performing the noise attenuation. For example, in areas of the data where no multiples are present, the signal should be preserved. For instance, a mute zone can be taken into account very easily. Calling \mathbf{M} this masking operator, the fitting goals in equation (6.17) are weighted as follows:

$$\begin{aligned} \mathbf{0} &\approx \mathbf{r}_n = \mathbf{M}(\mathbf{N}\mathbf{s} - \mathbf{N}\mathbf{d}), \\ \mathbf{0} &\approx \epsilon \mathbf{r}_s = \epsilon \mathbf{M}\mathbf{S}\mathbf{s}. \end{aligned} \quad (6.18)$$

Solving for \mathbf{s} in a least-squares sense leads to the objective function

$$f(\mathbf{s}) = \|\mathbf{r}_n\|^2 + \epsilon^2 \|\mathbf{r}_s\|^2. \quad (6.19)$$

It is interesting to look at the least-squares inverse for \mathbf{s} :

$$\hat{\mathbf{s}} = (\mathbf{N}'\mathbf{M}\mathbf{N} + \epsilon^2 \mathbf{S}'\mathbf{M}\mathbf{S})^{-1} \mathbf{N}'\mathbf{M}\mathbf{N}\mathbf{d}, \quad (6.20)$$

where ($'$) stands for the adjoint. Because \mathbf{M} is a diagonal matrix of zeros and ones, $\mathbf{M}'\mathbf{M} = \mathbf{M}$. In practice, all computations are done in the time domain. In the Fourier domain, equation (6.20) demonstrates that the least-squares estimate of \mathbf{s} is obtained by combining the spectra of both noise and signal. Abma (1995a) shows that equation (6.20) is similar to Wiener filtering

for random noise attenuation. Soubaras (1994) uses a very similar approach for random noise attenuation and more recently for coherent noise attenuation (Soubaras, 2001). Because the size of the data space can be quite large, s is estimated iteratively with the conjugate-gradient method.

Therefore, multiple attenuation with prediction-error filters (in 2-D or 3-D) is two-step process where (1) noise and signal PEFs are estimated and (2) signal and noise are separated according to their multidimensional spectra. The next section illustrates this technique with the Sigsbee2B dataset.

MULTIPLE ATTENUATION WITH THE SIGSBEE2B DATASET

The Sigsbee2B dataset was designed to generate strong surface-related multiples, the surface being the water surface. Figure 6.2 shows the true stratigraphic interval velocity model for this dataset. The data were created with a 2-D acoustic finite difference modeler with constant density. Two datasets were generated: one with a free surface in Figure 6.3a (FS) and one without a free surface in Figure 6.3b (NFS). Both receiver and source ghosts are included in the modeling with or without free surface (<http://www.delphi.tudelft.nl/SMAART/S2Breadme.htm>). We can then directly subtract the two datasets to obtain a very accurate prestack model of the surface-related multiples, without the need for SRMP.

In complex geology, multiple attenuation results should be assessed after migration; then, the effects of the multiple attenuation technique on the amplitudes of the primaries in angle domain common-image gathers or on migrated images at zero offset (Figure 6.4) can be inspected. For the Sigsbee2B dataset, a split-step double square-root (DSR) migration code with three reference velocities is used (Stoffa et al., 1990; Popovici, 1996). It is interesting to see that in Figure 6.4a the multiples are very weak after migration below the salt compared to the constant offset sections in Figure 6.3a. In particular, the water bottom multiple seems to disappear. This is because the multiples are extremely distorted by the migration process in the vicinity of the complex salt structure. Compared with the migration of the primaries only in Figure 6.4b, the multiples in Figure 6.4a are masking a lot of primaries in the deepest part of the model and need to be removed.

Two important tests are carried out in this section. Firstly, because the true primaries and multiples are known, the noise and signal PEFs can be ideally estimated without SRMP or the Spitz approximation and used for the separation. Secondly, in the more realistic case where only the model of the multiples is known, noise attenuation results are shown with 2-D or 3-D filters. The next section demonstrates that when an accurate model of the noise and signal is available, the signal can be recovered with minimum distortion.

Estimating biases

A bias is a processing footprint left by the multiple attenuation technique, e.g., edge effects from the non-stationary PEFs. In an ideal but unrealistic case, a model for both the primaries and the multiples might be available. In this case, a bias is also any difference between the true primaries and the estimated primaries after attenuation of the multiples. In this section I demonstrate that the bias is minimum with the pattern-based approach.

For the pattern model of the primaries, the answer, i.e., the data modeled without the free surface condition, is used. For the multiples, the difference between the FS (Figure 6.3a) and NFS (Figure 6.3b) datasets is used. Because the noise and signal PEFs are estimated from accurate pattern models, only 2-D filters are estimated. 3-D filters can help if the primaries and multiples are correlated in time and offset but uncorrelated across shot position. With 2-D filters, the attenuation is performed one shot gather at a time. Figure 6.5a displays the estimated primaries and Figure 6.5b the difference with the true primaries (Figure 6.3b). The bias introduced by the attenuation method is very small. 3-D filters would have given better results where the difference between Figure 6.5a and 6.3b is the strongest (e.g., near 20 km).

Looking now at the same estimated primaries after migration in Figure 6.6a, we see again that the attenuation gives a very good result with little bias. Some energy is visible in the difference plot in Figure 6.6b where no multiples are actually present, however. These artifacts have two origins. First, below 4,000 m, some primaries are affected by the multiple attenuation process, especially at far offset where primaries and multiples overlap. Second, above 4,000 m, the amplitude of the reflections for the sea floor and the top of salt are slightly different between the FS and NFS datasets. These differences are migrating at the reflector positions in

Figure 6.6b but with a very small energy, however.

From these results it appears that the quality of the multiple attenuation depends essentially on the filters. If accurate models for both the primaries and multiples are available, the primaries are recoverable with their true amplitude. Therefore, we should always try to find the best models for the signal and the noise. In practice, a very accurate model of the multiples can often be estimated with the auto-convolutional process of the Delft approach (Verschuur et al., 1992). For the primaries, the next section shows that the Spitz approximation gives a very good model if 3-D filters are used for multiple removal.

Testing the Spitz approximation

Now I assume that only a model of the multiples is known. The Spitz approximation in equation (6.7) shows how the PEFs for the signal can be estimated. The primaries are recovered with 2-D and 3-D filters. Figure 6.7 displays two constant offset sections after multiple attenuation with 2-D and 3-D PEFs. 3-D PEFs give by far the best results and attenuate multiples very well.

After migration, we see again in Figure 6.8 that the 3-D PEFs attenuate the multiples more effectively. The circles in Figure 6.8 surround areas where the 3-D filters are the most effective. A close-up in Figure 6.9 demonstrates in more detail (e.g., within the circles) how the two results with 2-D or 3-D filters differ below the salt. Events are more continuous and preserved better with 3-D filters. Comparing with the true reflectors in Figure 6.9a, important primaries (shown at '1' in Figure 6.9a) are attenuated with both 2-D and 3-D filters.

These important observations could not have been made before migration in the prestack domain because the primaries are much weaker than the surface-related multiples below the salt. This illustrates that for complex geology, the quality of a multiple removal technique should be assessed in the image space as often as possible. The fact that some primaries are attenuated in Figure 6.9 should motivate us in devising improved strategies for building more accurate noise and signal models.

The fact that 3-D PEFs attenuate the multiples better than 2-D PEFs is not surprising.

With higher dimensions, primaries and multiples are less likely to be correlated. Therefore, the noise and signal PEFs are less prone to annihilate similar data components. This is particularly important with the Spitz approximation which implicitly assumes that primaries and multiples are uncorrelated.

The next section compares the pattern-based approach with adaptive subtraction on a synthetic dataset provided by BP.

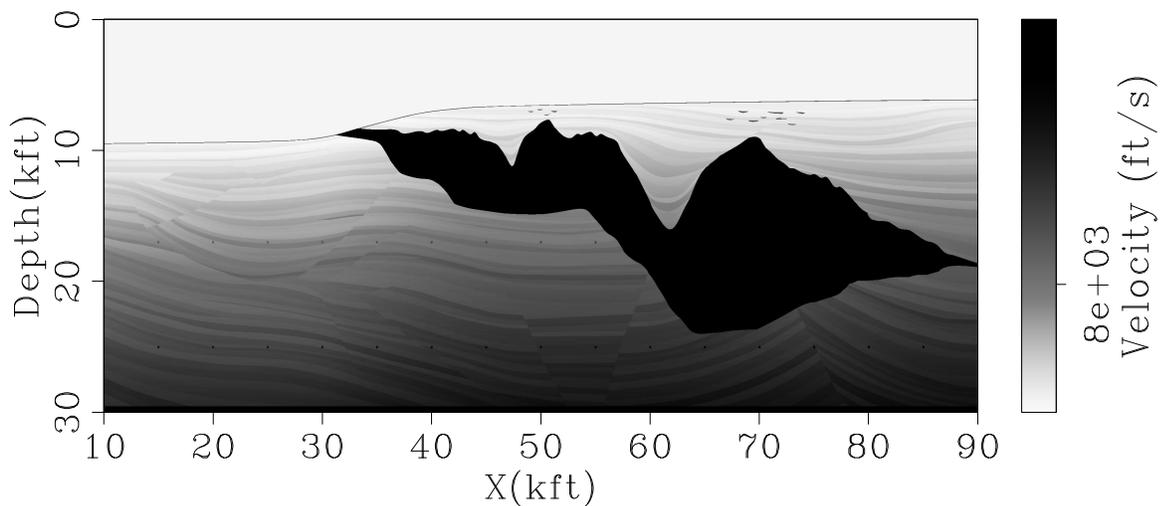


Figure 6.2: Stratigraphic interval velocity model of the Sigsbee2B dataset. `mul2d-stratigraphy` [ER]

ADAPTIVE VS. PATTERN BASED SUBTRACTION

The goal of this section is to compare the pattern-based method with the more conventional adaptive subtraction on a synthetic dataset provided by BP. These two methods have very different properties. In one hand, adaptive subtraction assumes implicitly that the signal has minimum energy. In addition, the separation is very fast because the matching filters and the primaries are usually estimated simultaneously with Wiener-Levinson type methods. On the other hand, the proposed approach assumes that the signal and noise have different patterns. Unfortunately, high dimension filters for primaries and multiples need to be estimated first

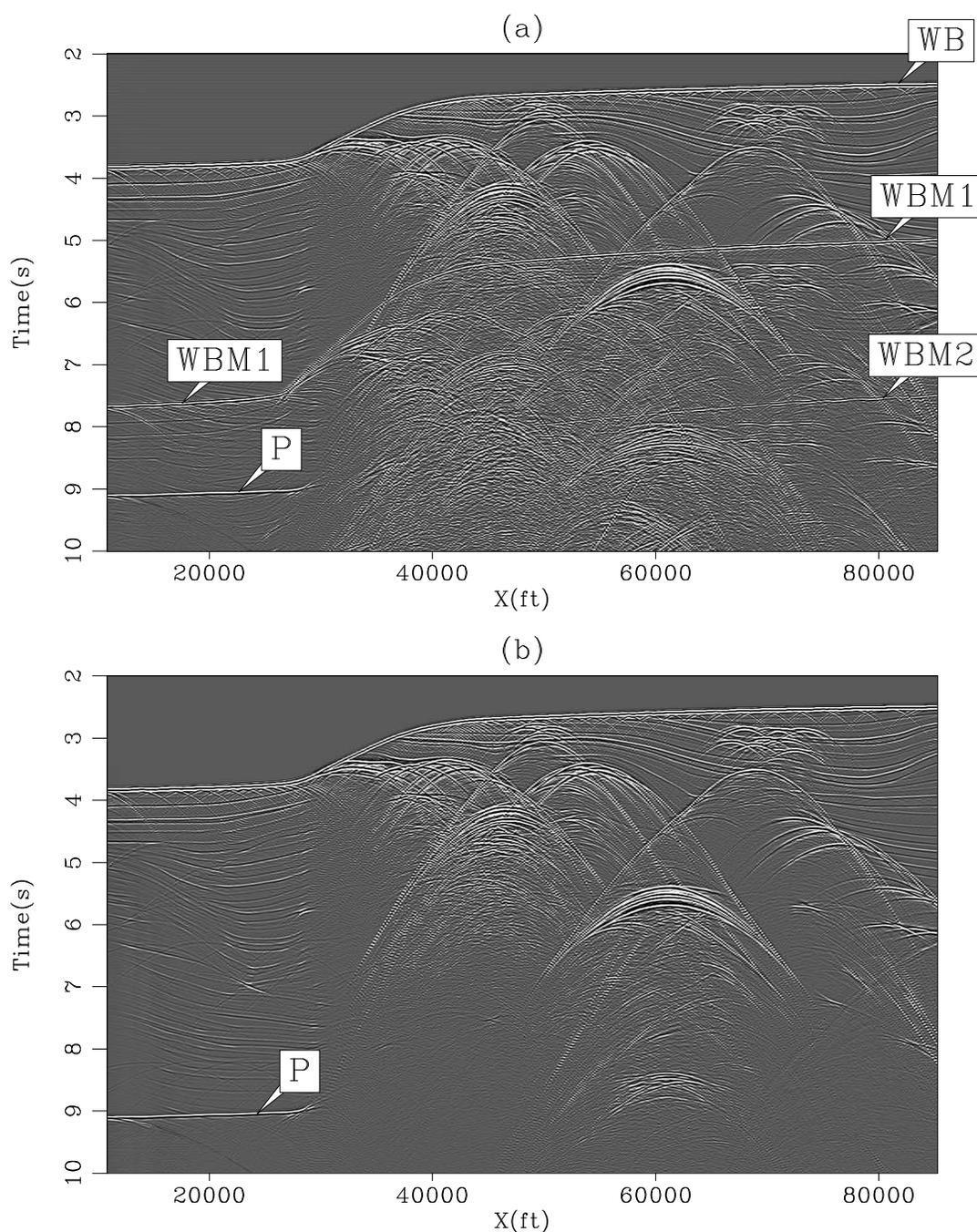


Figure 6.3: Two constant offset sections ($h=1125$ ft) of the Sigsbee2B dataset with (a) and without (b) free surface condition. The multiples are very strong below 5 s. The weak horizontal striping in (a) comes from a source effect only present with the free surface condition modeling. Arrow WB shows the water-bottom reflection, WBM1 the first order surface-related multiple for the water-bottom, WBM2 the second order surface-related multiple for the water-bottom, and P a strong primary. `mul2d-datasignal` [ER,M]

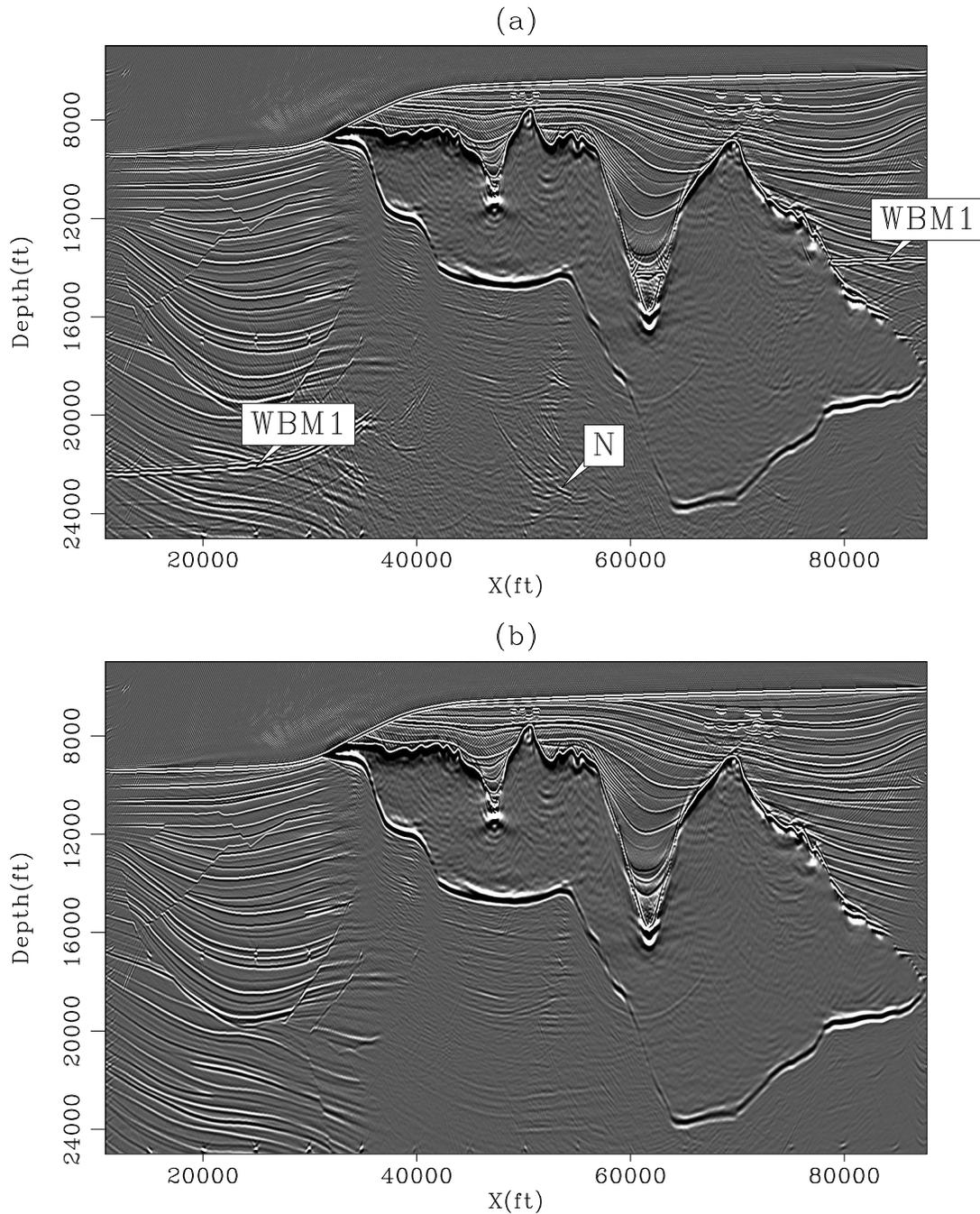


Figure 6.4: Migrated images at zero-offset for the data with (a) and without (b) free surface condition. Comparing with Figure 6.3, the multiples appear much weaker below the salt after migration. However, some reflectors near 22 kft are hidden in (a). Arrow WBM1 shows the first order water-bottom multiple after migration. Arrow N shows some noise associated with the migration of multiples beneath the salt body. `mul2d-datasignal-mig` [CR,M]

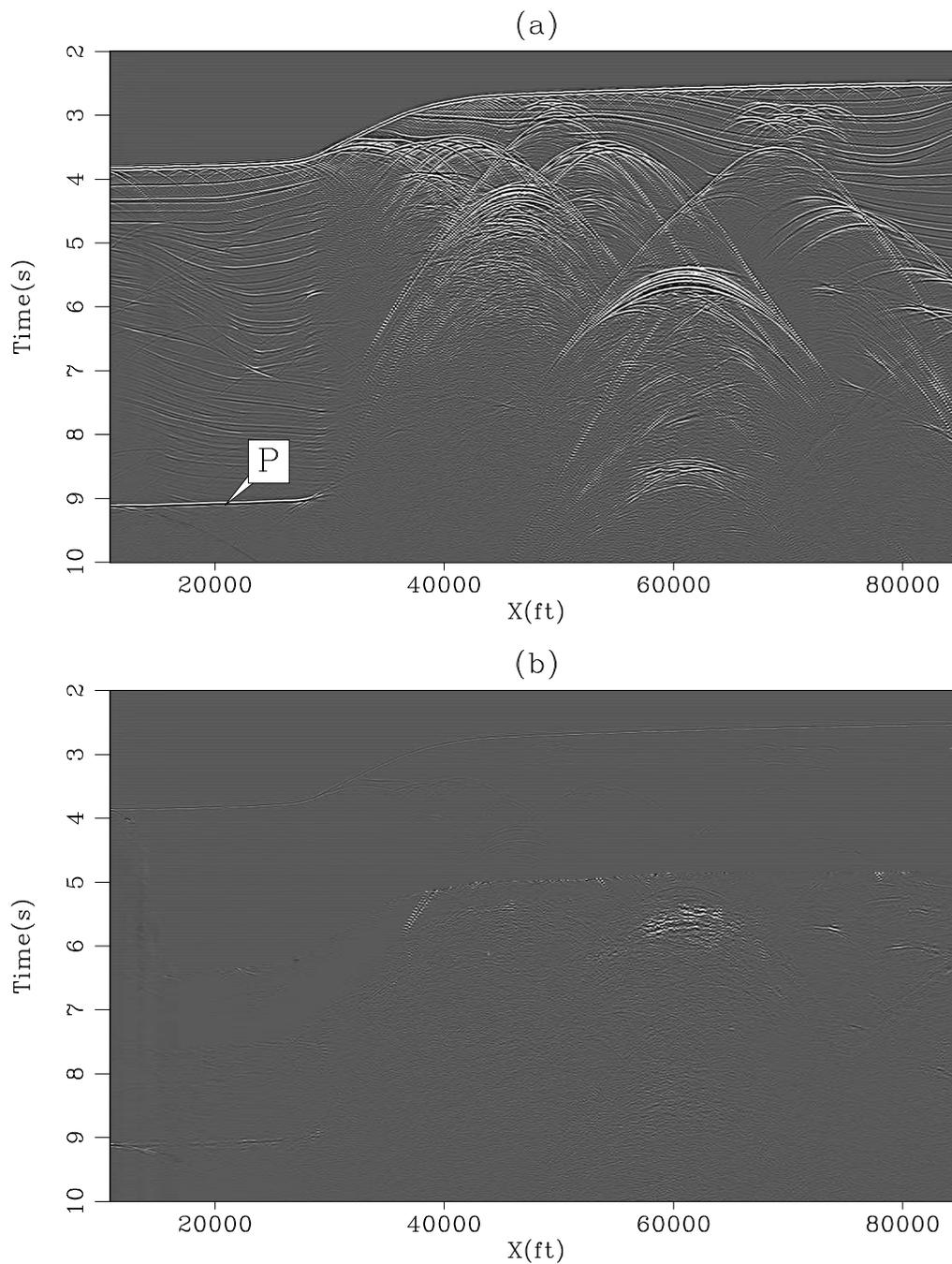


Figure 6.5: Two constant offset panels at $h=1125$ ft. for (a) the estimated primaries and (b) the difference with the true primaries. The true primaries and multiples are used to estimate the PEFs. Arrow P shows a primary that could be mistaken for a multiple. mul2d-signal-true [CR,M]

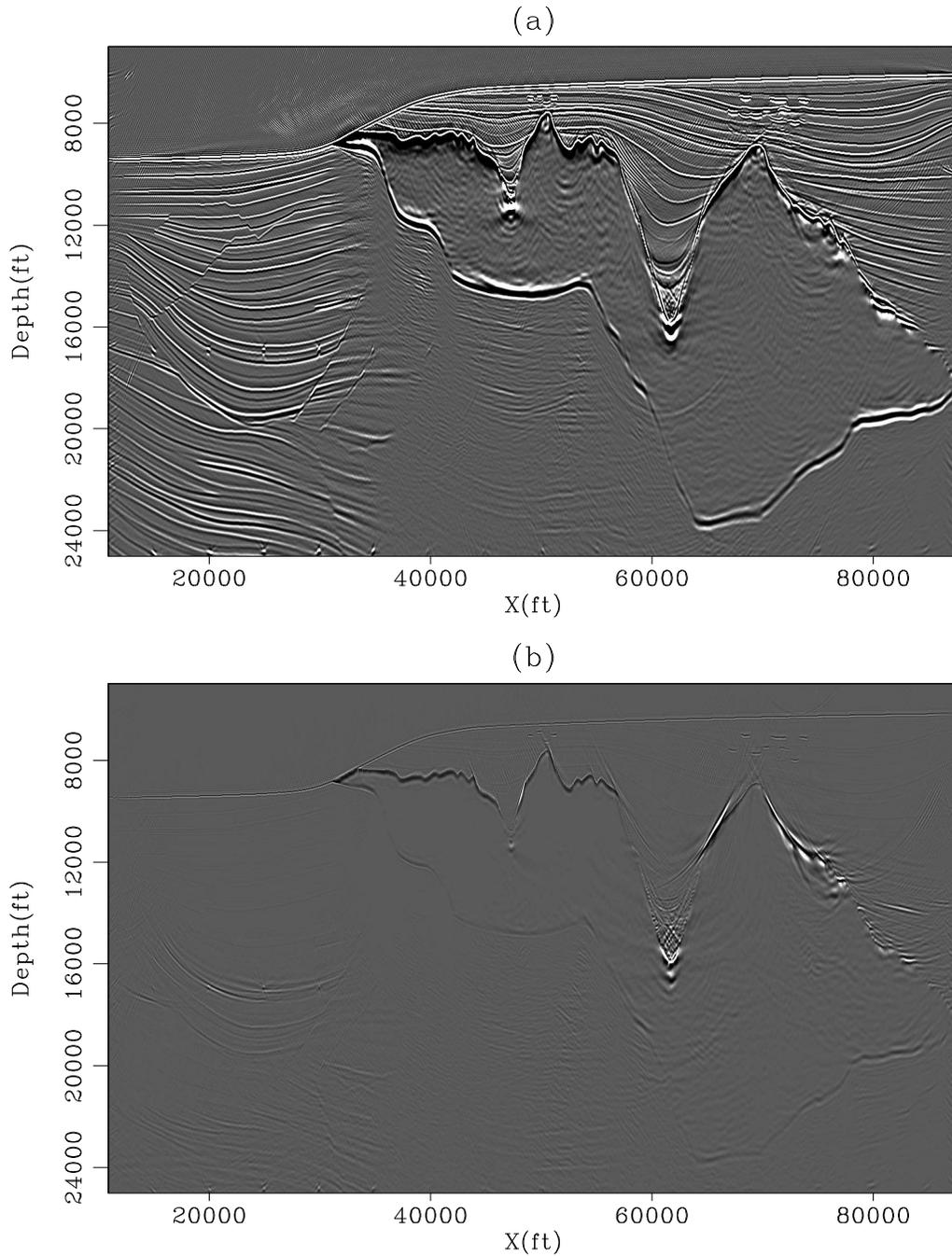


Figure 6.6: (a) Migration result after multiple attenuation when the true primaries and multiples are used to estimate the PEFs. (b) Difference between (a) and Figure 6.4b. The estimated primaries are almost exact. `mul2d-signal-true-mig` [CR]

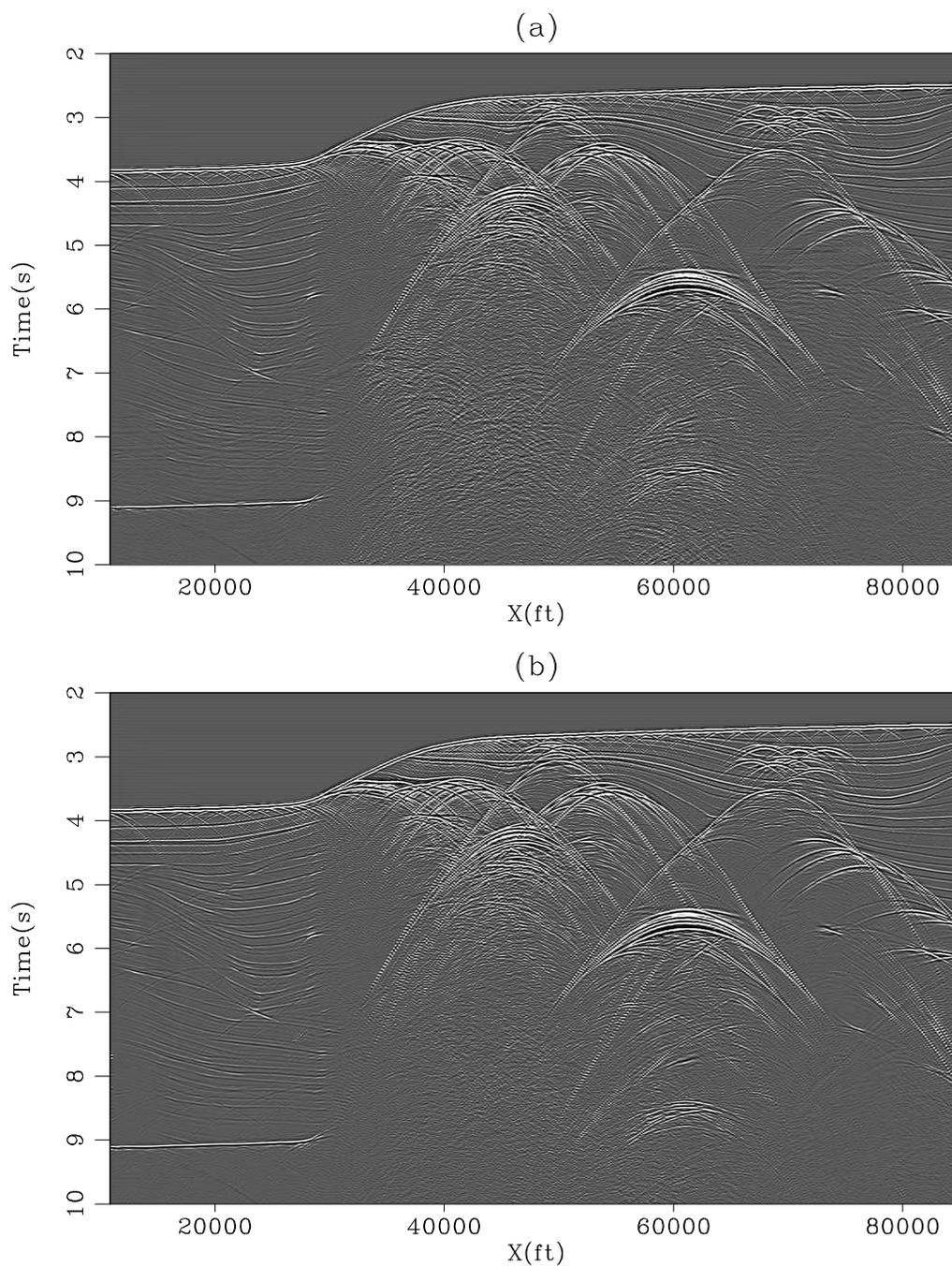


Figure 6.7: Two constant offset sections ($h=1125$ ft) after multiple attenuation with the Spitz approximation using (a) 2-D and (b) 3-D filters. `mul2d-signal-2D-3D-PEF` [CR]

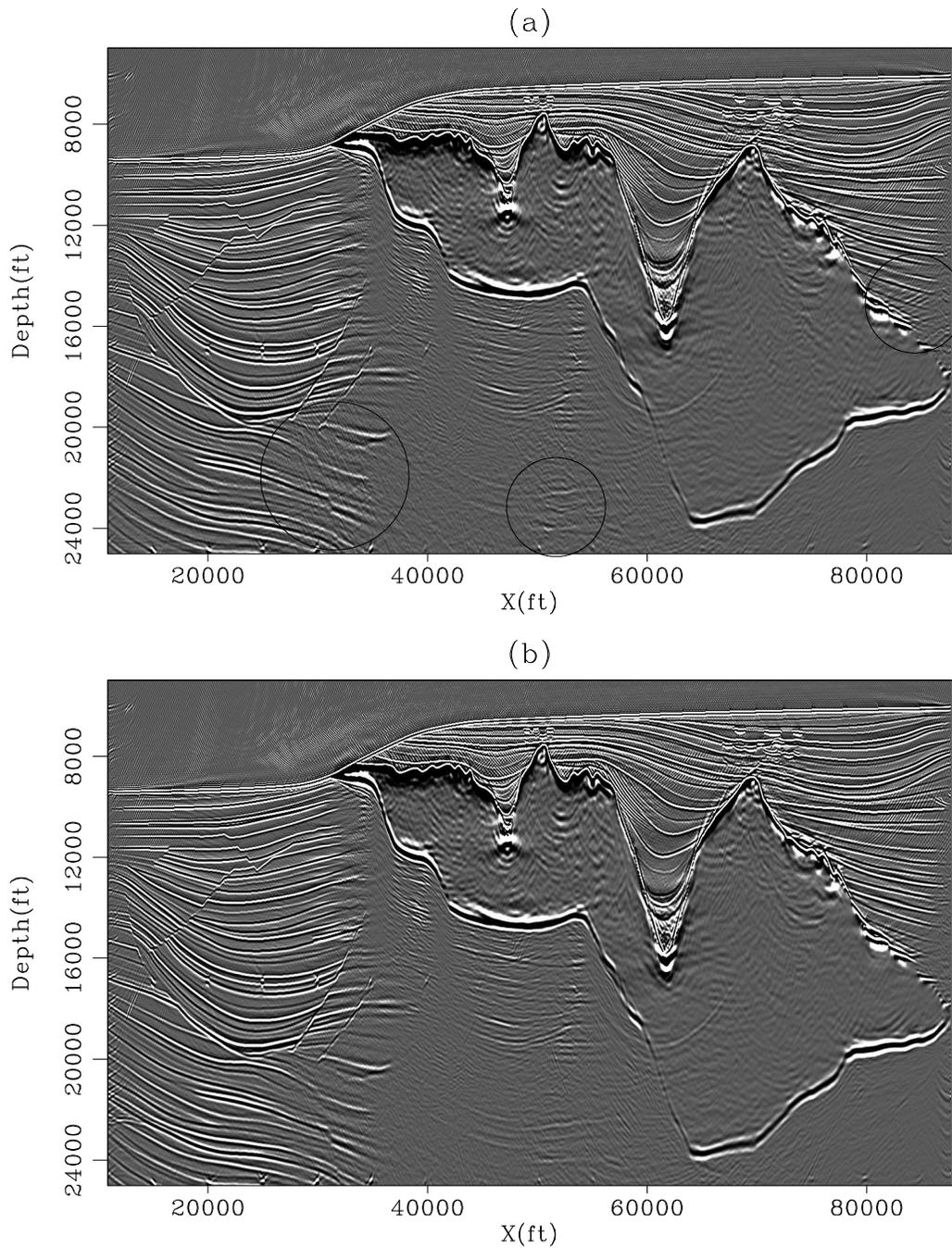


Figure 6.8: Two migration results of the estimated primaries with (a) 2-D and (b) 3-D filters. The circles show areas where multiples are better attenuated with 3-D filters than with 2-D filters. `mul2d-signal-2D-3D-PEF-mig` [CR,M]

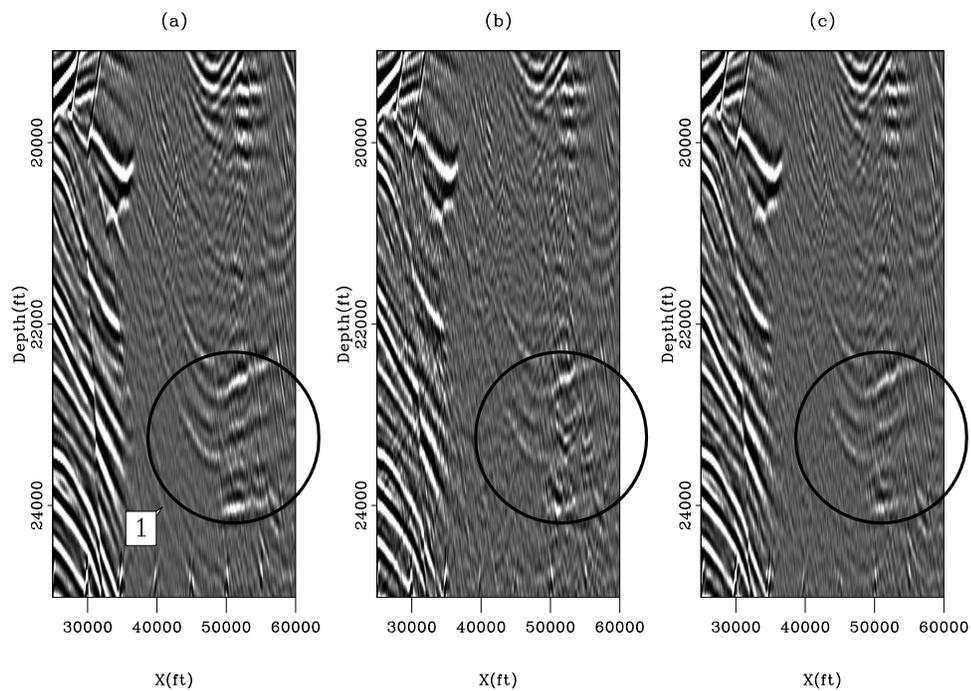


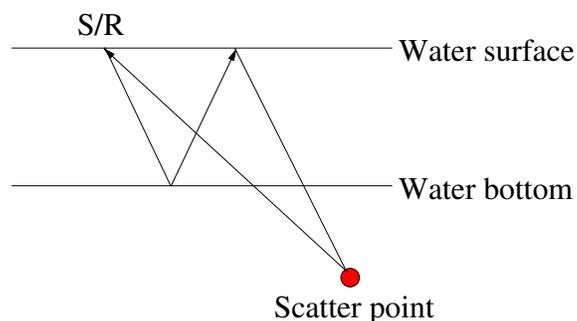
Figure 6.9: Close-up of Figure 6.8 showing two migrated images when (b) 2-D and (c) 3-D filters are used. The true primaries are shown in (a). Arrow '1' points to primaries that are attenuated with the pattern-based approach. The circles show an area where the 3-D filters are the most effective at removing the multiples. mul2d-signal-2D-3D-PEF-small-mig [CR,M]

before starting the separation, which adds to the total cost of the method. Therefore, the pattern-based approach can be much slower than adaptive subtraction.

Computing considerations aside, the two methods can perform very differently according to the geological setting. Here, both are tested in a salt environment with a synthetic dataset. This dataset was primarily designed to conduct blind tests for velocity estimation methods. Consequently, no structural information is known. The adaptive subtraction technique used in this section is based on the estimation of 2-D, time/space varying matching filters (Rickett et al., 2001). The filters are computed for one shot gather at a time. With the pattern-based approach, 3-D filters are used for the multiple attenuation. Ideally, 3-D filters should be also used with adaptive subtraction. However, matching filters are generally not estimated that way.

The multiple model is computed with SRMP. Therefore, only multiples bouncing at least once at the water surface will be modeled and subtracted. The synthetic model has an offset spacing of 12.5 m and a shot separation of 50 m. To make the multiple prediction work, the offset axis is sampled down to 50 m. Figure 6.11 shows one constant offset section from -15,000 m to +15,000 m with primaries and multiples. This section of the dataset is particularly interesting because of the diffractions visible throughout. A possible interpretation of these diffractions is the presence of salt bodies with a rugose top (similar to what we see with the Sigsbee2B dataset). The multiple model is shown in Figure 6.12 for the same offset. DT points to diffraction tails where the model is not properly rendering the multiples in the data. Figure 6.10 illustrates on a zero-offset example why diffractions are difficult to predict. The main reasons are that far offset data are not recorded and/or the scatter points are out of plane. Besides these few imperfections, the model looks very faithful to the actual multiples.

Figure 6.10: Illustration of a pegleg multiple with a bounce on a scatter point recorded at zero-offset. These events are difficult to model because of the lack of far and/or crossline offsets. `mul2d-scatter` [NR]



The result of adaptive subtraction is shown for one offset section in Figure 6.13 and the result of pattern-based subtraction is shown in Figure 6.14. The adaptive subtraction is doing a decent job everywhere. However, some multiples are still visible. For example, '1' in Figure 6.13 points to a location where multiples overlap with primaries and are not attenuated. In contrast, the pattern-based subtraction (i.e., Figure 6.14) seems to do a better job attenuating these events. The same is true for arrows '2' and '5'. The diffracted multiples (arrows '3' and '4') are also attenuated more effectively with the pattern-based approach.

Because no velocity analysis was conducted with this dataset, no stacks are presented. Alternatively, close-ups of constant offset sections are shown to illustrate strengths and weaknesses of the two different approaches. Figure 6.15 shows a comparison between the input data, the multiple model, the estimated primaries with the adaptive subtraction and the estimated primaries with the pattern-based technique. The offset is 700 m. As shown by the arrows, the pattern-based method performs generally better. The same conclusions hold in Figure 6.16. Note in Figure 6.16b aliasing artifacts due to the coarse sampling of the offset axis for the multiple prediction (van Dedem, 2002).

Sometimes, it can be rather difficult to see if multiples are removed or not by simply looking at 2-D planes. Figure 6.17c shows one event at '2' that seems to be a primary. However, by looking at the shot gathers (not shown here), it appears that this event is a multiple that the pattern-based approach was able to attenuate.

One shortcoming of the pattern-recognition technique is that it relies on the Spitz approximation to provide a signal model if nothing else is available. By construction, the signal and noise filters will span different components of the data space. Therefore, the estimated primaries and multiples are uncorrelated. This simple fact suggests that with the Spitz approximation, higher dimension filters are preferred because primaries and multiples have fewer chances to look similar.

Figure 6.18 shows an example where primaries are damaged by the pattern-based method. For instance in Figure 6.18a, we see at '2' a primary that is attenuated by the PEFs (Figure 6.18d) but well preserved by the adaptive subtraction (Figure 6.18c). Here the primaries and multiples (Figure 6.18b) exhibit similar patterns and the signal may have minimum energy.

Using the Spitz approximation, event '2' is identified as noise and removed as such. For event '3', it is quite difficult to say if multiples are removed in Figure 6.18d or if primaries are preserved in Figure 6.18c. Looking at the corresponding shot gathers did not help to make a decision because the multiples are very strong. Event '4' is preserved with the adaptive subtraction and '1' and '5' are well recovered with the pattern-based approach.

This synthetic example indicates that the pattern-based approach tend to attenuate the multiples more accurately than adaptive subtraction when the multiples are not correlated with the primaries. This illustrates that higher dimensions filters should be preferred to discriminate the noise and signal more effectively. The next section shows how the pattern-based approach performs on a field dataset from the Gulf of Mexico.

GULF OF MEXICO EXAMPLE

The pattern-based approach is tested on a 2-D line from the Mississippi Canyon. This dataset has been extensively used in the past to benchmark multiple attenuation techniques (special edition of *The Leading Edge*, January 1999). A stacked section of this dataset is shown in Figure 6.19. Strong surface-related multiples are visible below 3 s. The tabular salt near the water bottom generates peg-legs in the data. The shot and receiver spacing is 26.6 m. and the first hydrophone is 100 m. away from the source. Ideally, before using SRMP, the missing short offset traces should be interpolated for improved separation (Kabir and Verschuur, 1995). No interpolation was performed here, thus creating wrong amplitudes in the multiple model, especially at short offset where amplitudes are zero. Note that SRMP creates a model for all multiples with at least one bounce at the sea surface. For this dataset both adaptive subtraction and pattern-based techniques are compared after migration.

The multiple attenuation starts with the estimation of 3-D PEFs from SRMP and the Spitz approximation. Then multiples and primaries are separated according to their multidimensional spectra. Figure 6.20 displays the multiple attenuation result for one shot gather outside the salt boundaries ($X=4500$ m). The multiple model in Figure 6.20c demonstrates that the short offset traces are not accurately modeled. However, the estimated primaries (Figure

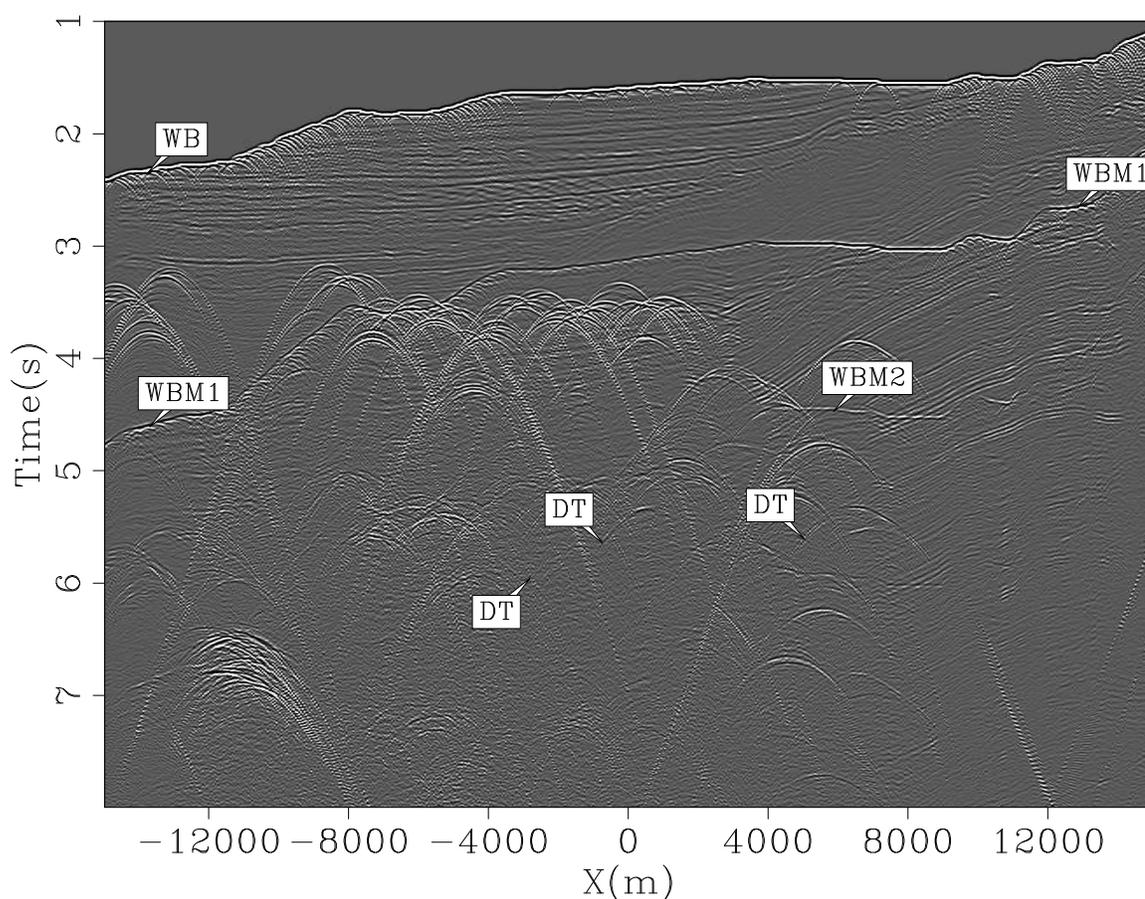


Figure 6.11: Constant offset section ($h=500$ m) of the data with multiples. DTs point to the tails of diffracted multiples. These pegleg multiples bounce once on the seafloor and once on the scatter. Arrow WB shows the water bottom, WBM1 the first order surface-related multiple, and WBM2 the second order surface-related multiple. `mul2d-data` [ER]

6.20b) and multiples (Figure 6.20d) show that the subtraction is working at short and far offsets. This robustness to modeling errors is increased by the smoothing of the filter coefficients during the PEFs estimation.

A second shot gather is shown in Figure 6.21 from inside the salt boundaries ($X=12000$ m). The estimated primaries (Figure 6.21b) and multiples (Figure 6.21d) are again separated well. However, looking closely at the primaries in Figure 6.21b, some diffracted multiples seem to have leaked in. These diffractions (Figure 6.10) can be either out of plane events or in

plane events that SRMP cannot model due to limitations imposed by the acquisition geometry (Kabir, 2003).

Stacked sections of the input data, estimated multiples and primaries are displayed in Figure 6.22. These stacks are located below the salt, as shown in Figure 6.22a. The estimated primaries in Figure 6.22b are still contaminated with diffracted multiples (shown as M). The multiples are attenuated very well otherwise. Comparing the stack of the multiple model (Figure 6.22c) with the stack of the extracted multiples (Figure 6.22d) indicates that no prominent primaries have been attenuated.

Comparison with adaptive subtraction in the image space

In addition to the pattern-based approach, the multiples are attenuated with adaptive subtraction. Instead of comparing both results in the data space, the two techniques are now analyzed in the image space after migration. A split-step DSR migration is used with three reference velocities. The results of the migrations are then studied in the angle domain on common-image gathers (ADCIG). These angle gathers are generated with the method of Sava and Fomel (2003) after migration (Stolt and Weglein, 1985; Weglein and Stolt, 1999). Angle gathers can be useful for AVO analysis (Prucha et al., 1999; Kuhl and Sacchi, 2003) and can form a convenient basis for multiple attenuation as well (Sava and Guitton, 2005; Alvarez et al., 2004).

Figure 6.23 shows the migration results in the angle domain for the input data (Figure 6.23a), the estimated signal with the pattern-based approach (Figure 6.23b) and the estimated signal with adaptive subtraction (Figure 6.23c). By looking at the constant angle section ($X=11000$ m), it appears that more multiples have been removed with the pattern-based approach, especially below the salt where the multiples are the strongest. The common angle gathers corroborate this: a lot more energy from the multiples is present with the adaptive subtraction. Note that some multiple energy remains in both Figure 6.23b and Figure 6.23c where the salt body is present. These multiples have strong curvature and low-frequency content. These events are probably internal multiples bouncing between the sea bottom and the top of salt. Another pass of multiple attenuation in the image space (Sava and Guitton, 2005) could

eliminate these reflections.

Finally, Figure 6.24 shows a comparison of ADCIGs for the input data with multiples (Figure 6.24a), the estimated primaries with the pattern-based approach (Figure 6.24b) and the estimated primaries with adaptive subtraction (Figure 6.24c) outside the salt boundaries ($X=4000$ m). It illustrates once again that the pattern-based approach outperform the adaptive subtraction method with a cleaner panel.

The Gulf of Mexico example demonstrates that the pattern-based approach is an effective tool for multiple attenuation in complex geology. Although the multiple model obtained with SRMP presented some obvious flaws (short offset amplitudes), the proposed approach is able to attenuate the multiples while preserving the primaries. In addition, comparisons in the image space after migration on ADCIGs show that the pattern-based approach gives cleaner panels than adaptive subtraction.

CONCLUSION

Multiple attenuation can be cast as a problem where events are separated according to their patterns (e.g., multidimensional spectra). A pattern is made up of both the kinematic and amplitude information that PEFs can approximate. The pattern-based technique is a two-step procedure. Firstly, non-stationary time/space domain PEFs are estimated for both primaries and multiples. These filters are estimated from a model of the multiples usually computed with SRMP and a model of the primaries computed with the Spitz approximation. The Spitz approximation assumes that the noise and signal are uncorrelated. Secondly, multiples are separated from the primaries in a least-squares sense according to their multidimensional spectra.

As illustrated with the Sigsbee2B dataset, this approach has the potential to separate primaries and multiples very well as long as an accurate model exists for the PEFs estimation. When no model of the primaries exist, the Spitz approximation, which convolves the data with the noise PEFs, leads to a very good attenuation of the multiples if high dimensions filters are used (i.e., 3-D vs. 2-D). Indeed, primaries and multiples are less likely to look similar.

Comparing adaptive and pattern-based subtraction indicates that the latter removes the

multiples almost always better, except in areas where the primaries and multiples are correlated. An important property of the pattern-based approach is that it seems to cope well with modeling inadequacies. For instance, with the Gulf of Mexico dataset, although not properly reconstructed, short-offset multiples were nonetheless removed.

Multiple attenuation can be viewed as a two steps process where multiples are first predicted and then subtracted. Both steps are equally important but most of the efforts are usually concentrated on the prediction step and not the subtraction. Since in practice it remains impossible to get a perfect multiple model due to the limitations of the acquisition geometry and interpolation/extrapolation techniques, new subtraction methods are needed. The pattern-based method presented in this Chapter is a successful tool for removing coherent energy in seismic data. This technique offers a viable alternative to adaptive subtraction by being less sensitive to errors in the multiple model. In addition, compared to adaptive subtraction, the pattern-based technique alleviates the strong assumption that primaries have minimum energy.

Chapter 7 illustrates the pattern-based method with a 3-D field data example where the acquisition geometry limits our ability to predict surface-related multiples accurately.

ACKNOWLEDGMENTS

I thank SMAART-JV for providing the Sigsbee2B dataset, BP for the second synthetic dataset and WesternGeco for the Gulf of Mexico dataset. I thank also Paul Sava for providing the migration code.

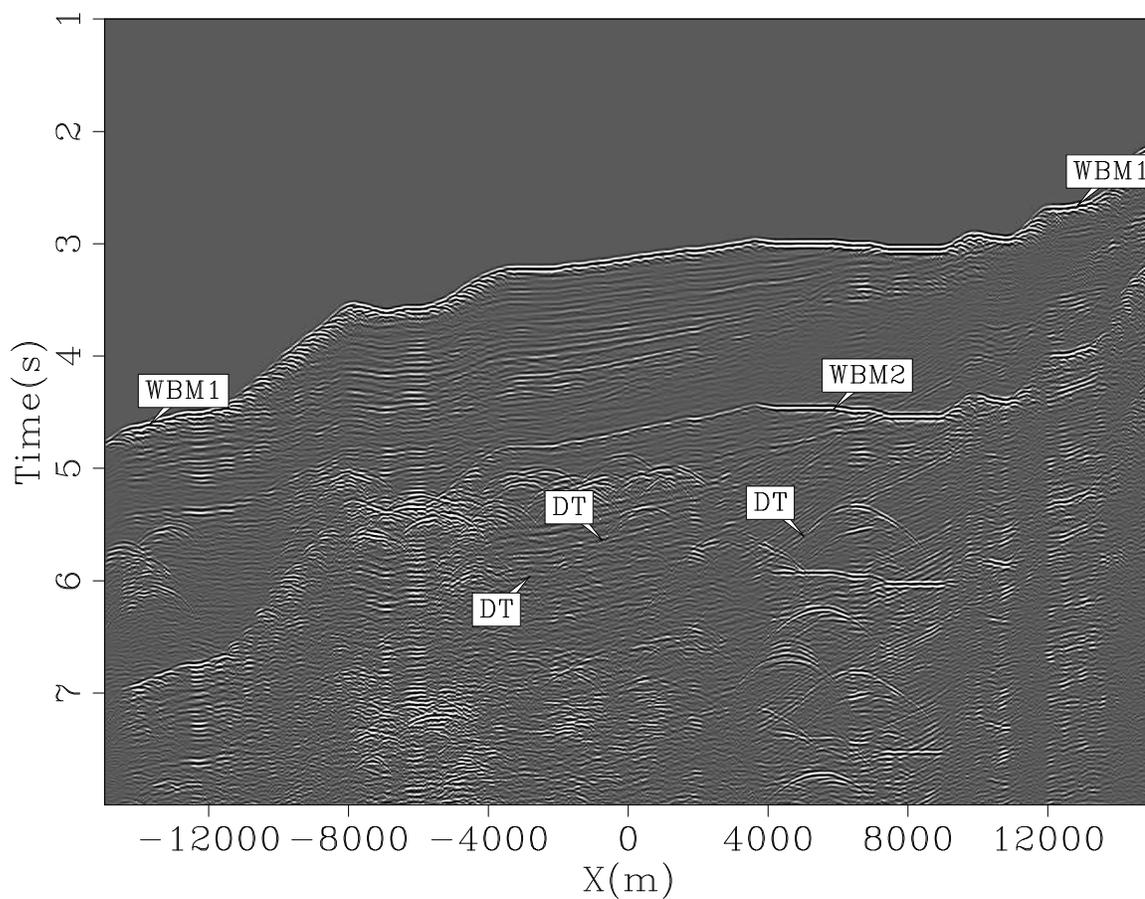


Figure 6.12: Constant offset section ($h=550$ m) of the estimated multiples. The multiples are accurately modeled except for the diffracted multiples, shown as DT, for which the limited range of offsets and number of shots hamper any attempt at modeling the diffraction tails. Arrow WBM1 shows the first order surface-related multiple, and WBM2 the second order surface-related multiple. `mul2d-mult` [CR]

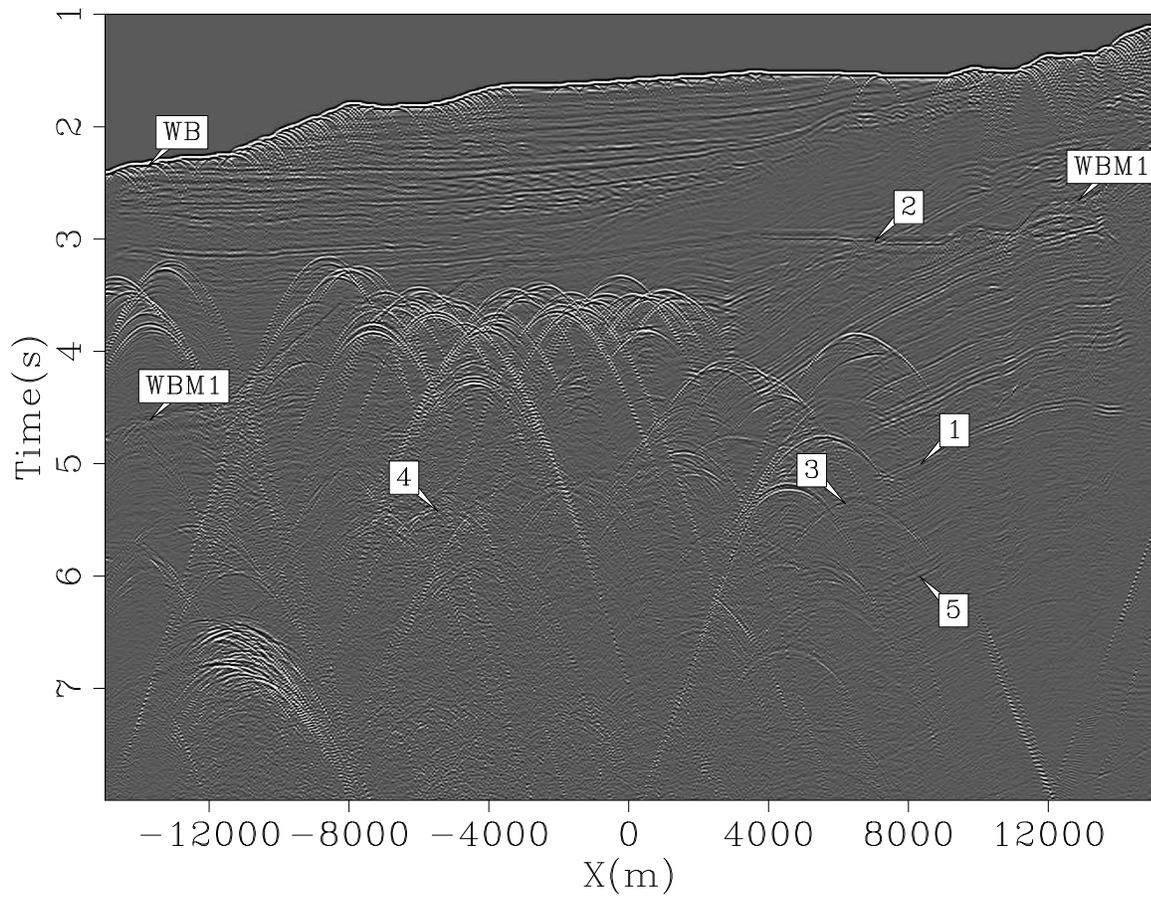


Figure 6.13: Constant offset section ($h=550$ m) of the estimated primaries with adaptive subtraction. The arrows point to locations where multiples are still present. Arrow WBM1 shows the remaining energy for the first order surface-related multiple. `mul2d-dsign` [CR]

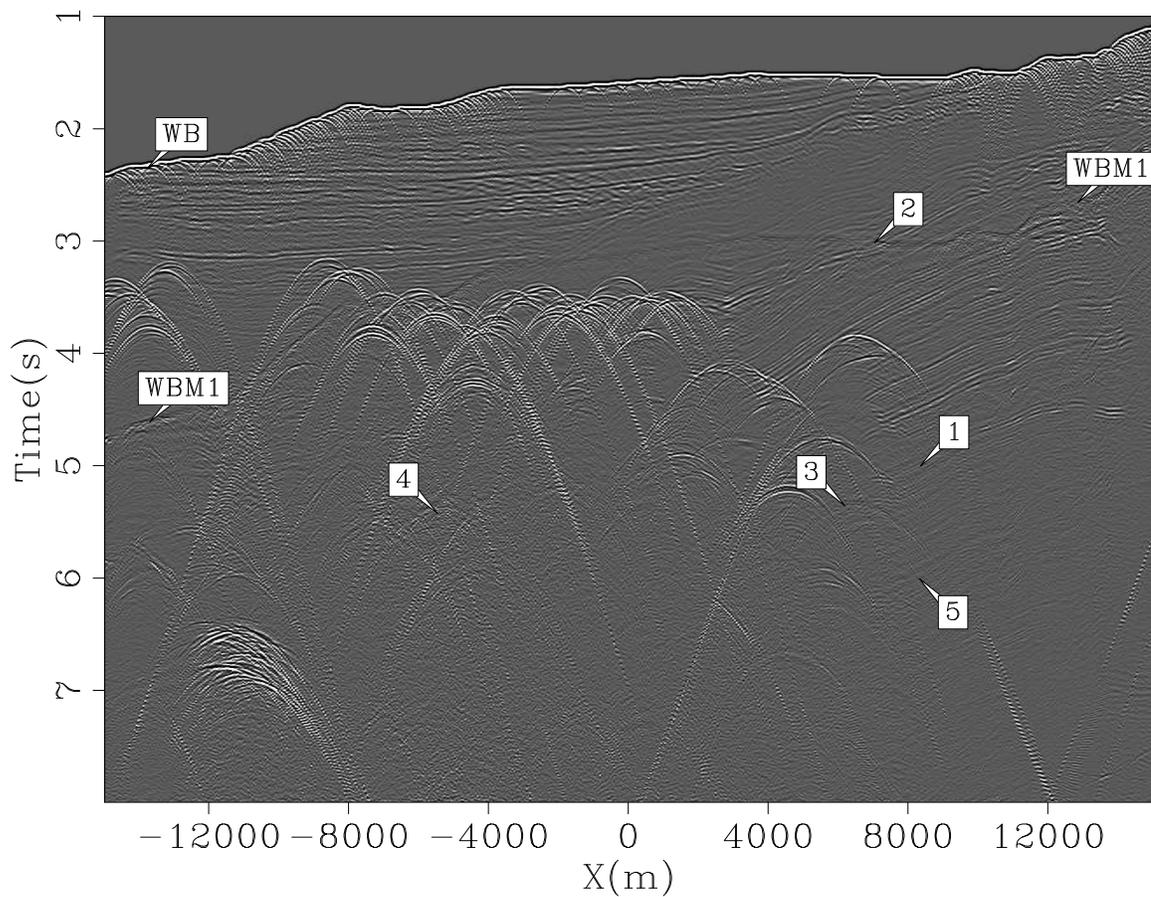


Figure 6.14: constant offset section ($h=550$ m) of the estimated primaries with pattern recognition. Multiples are better attenuated than in Figure 6.13. Arrow WBM1 shows the remaining energy for the first order surface-related multiple. mul2d-sign [CR]

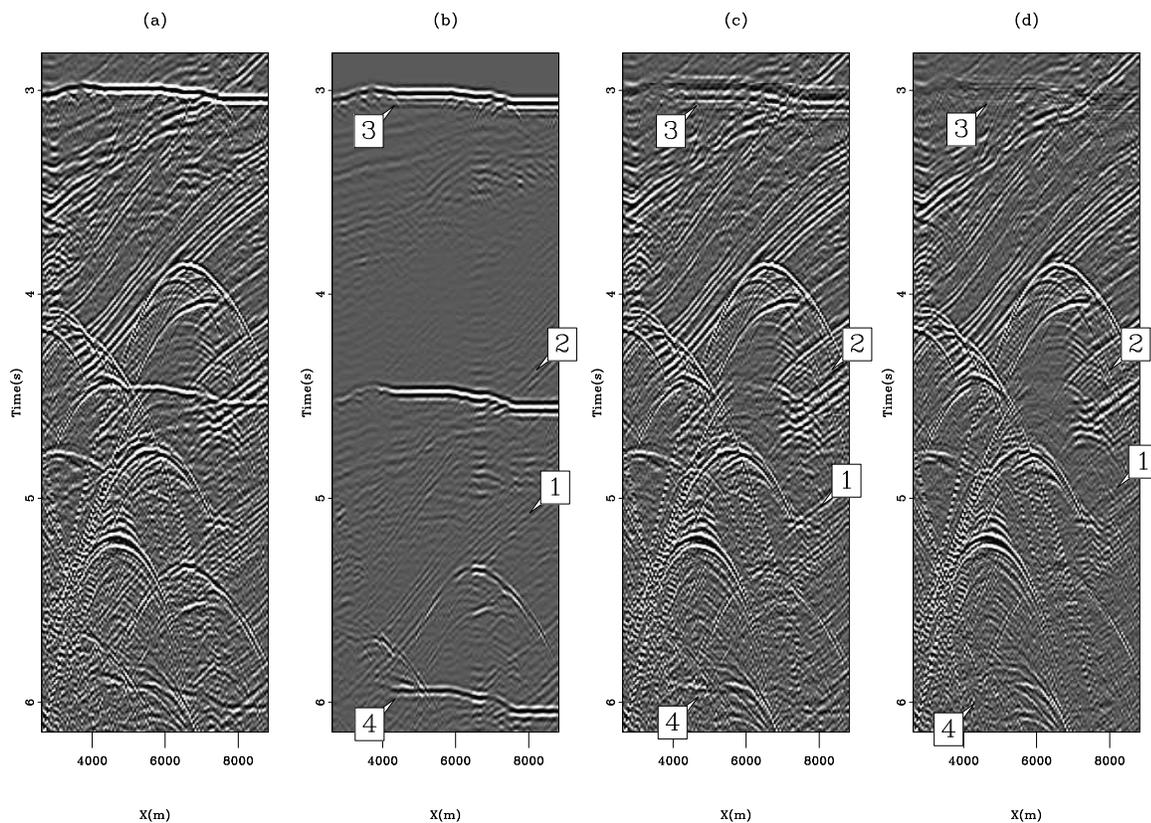


Figure 6.15: Constant offset sections ($h=700$ m) for (a) the input data, (b) the multiple model, (c) the estimated primaries with adaptive subtraction, and (d) estimated primaries with the pattern-based approach. Arrows point to locations where the pattern-based approach attenuates multiples significantly better than the adaptive subtraction. `mul2d-compwin1` [CR,M]

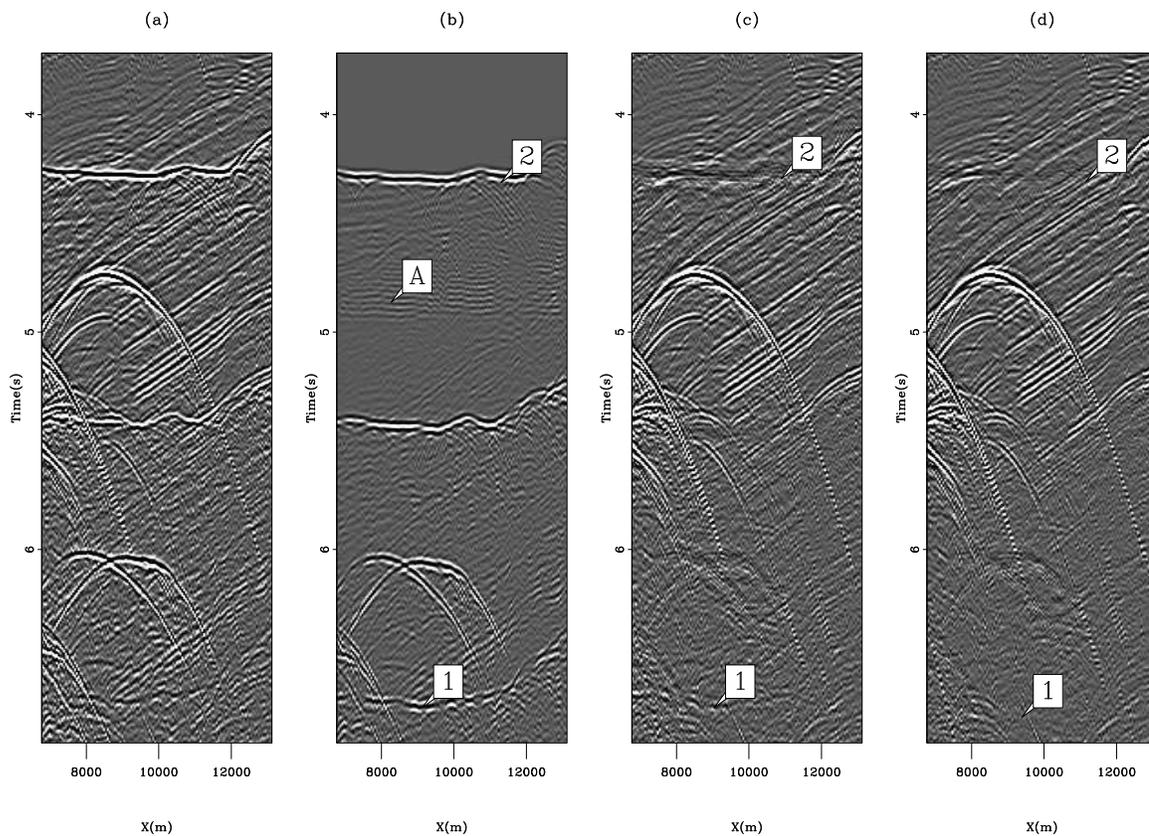


Figure 6.16: Constant offset sections ($h=4550$ m) for (a) the input data, (b) the multiple model, (c) the estimated primaries with adaptive subtraction, and (d) estimated primaries with the pattern-based approach. Arrow A points to aliasing effects due to the offset sampling of the shot gathers. The pattern-based approach attenuates the multiples better than the adaptive subtraction in '1' and '2'. `mul2d-compwin6` [CR,M]

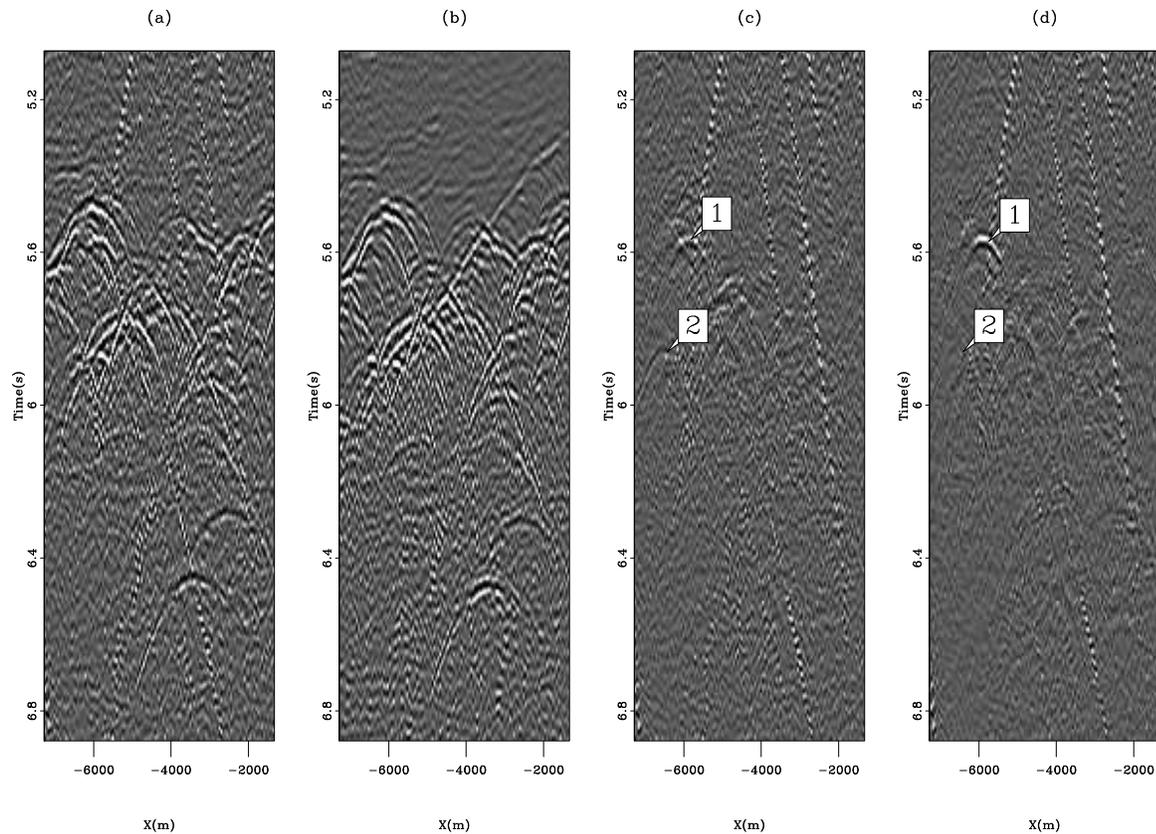


Figure 6.17: Constant offset sections ($h=3300$ m) for (a) the input data, (b) the multiple model, (c) the estimated primaries with adaptive subtraction, and (d) estimated primaries with the pattern-based approach. '1' points to a primary that the pattern-recognition preserves very well. '2' points to an event that is attenuated with the pattern-based approach but not with the adaptive subtraction in (c). Though not shown here, a close inspection of the corresponding shot gathers suggests that '2' is actually a multiple. `mul2d-compwin8` [CR,M]

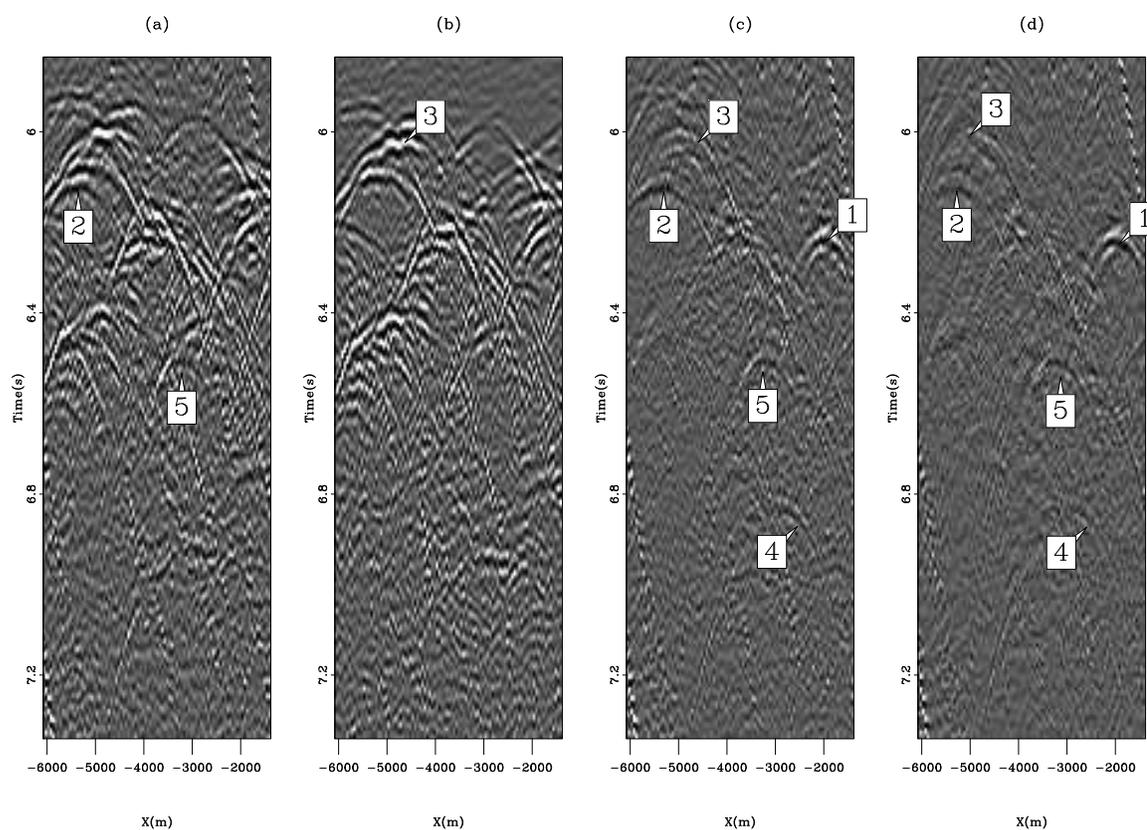


Figure 6.18: Constant offset sections ($h=5050$ m) for (a) the input data, (b) the multiple model, (c) the estimated primaries with adaptive subtraction, and (d) estimated primaries with the pattern-based approach. '1' and '5' show events better preserved with the pattern-based method. '2' and '1' are better recovered with the adaptive subtraction. '4' seems to point to a primary that the adaptive subtraction is able to save. Because the area is contaminated with strong multiples, it is difficult to know without a stratigraphic model if '3' is a primary or not.

`mul2d-compwin12` [CR,M]

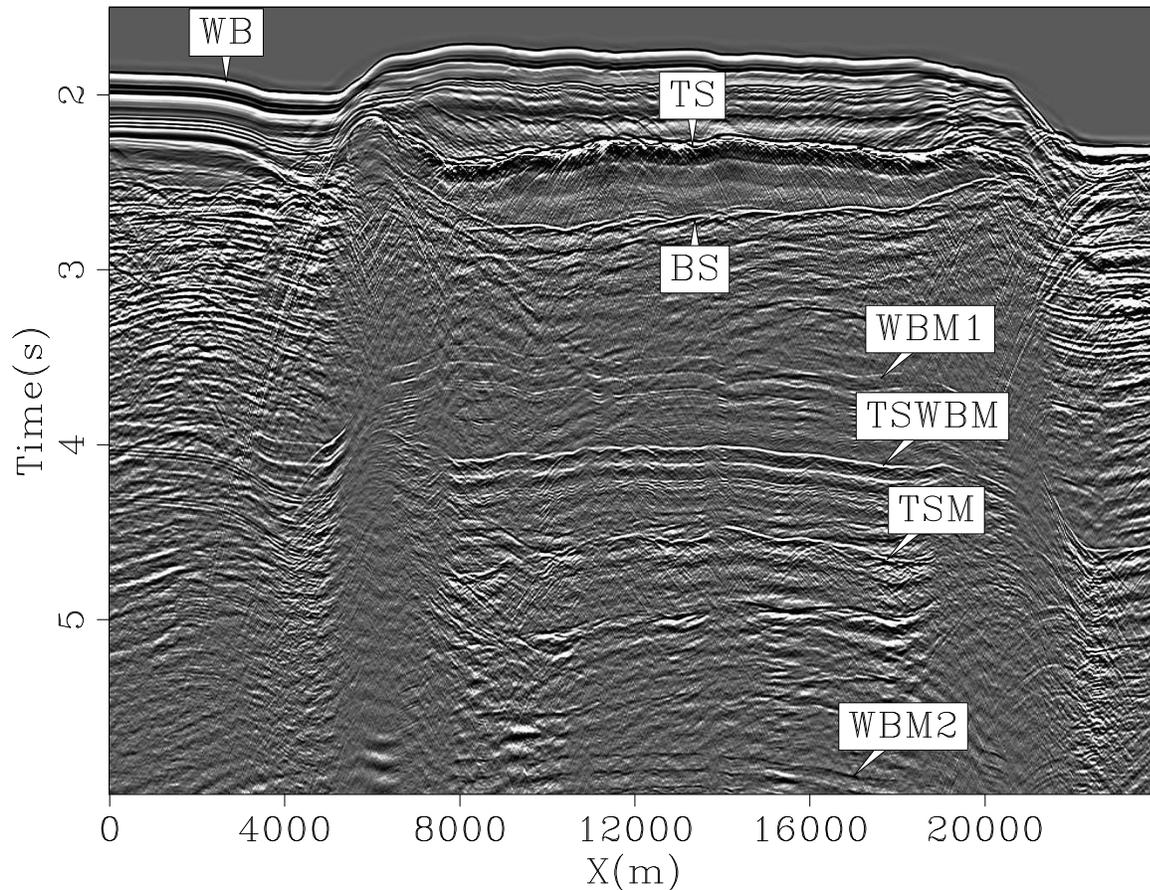


Figure 6.19: Stacked section of a 2-D line from the Gulf of Mexico. Surface-related multiples appear below 3 s. Arrow WB indicates the water bottom, TS the top of salt, BS the base of salt, WBM1 the first order water-bottom multiple, TSWBM the top of salt/water bottom pegleg multiple, TSM the top of salt/top of salt multiple, and WBM2 the second order water-bottom multiple. `mul2d-stack.dat` [CR]

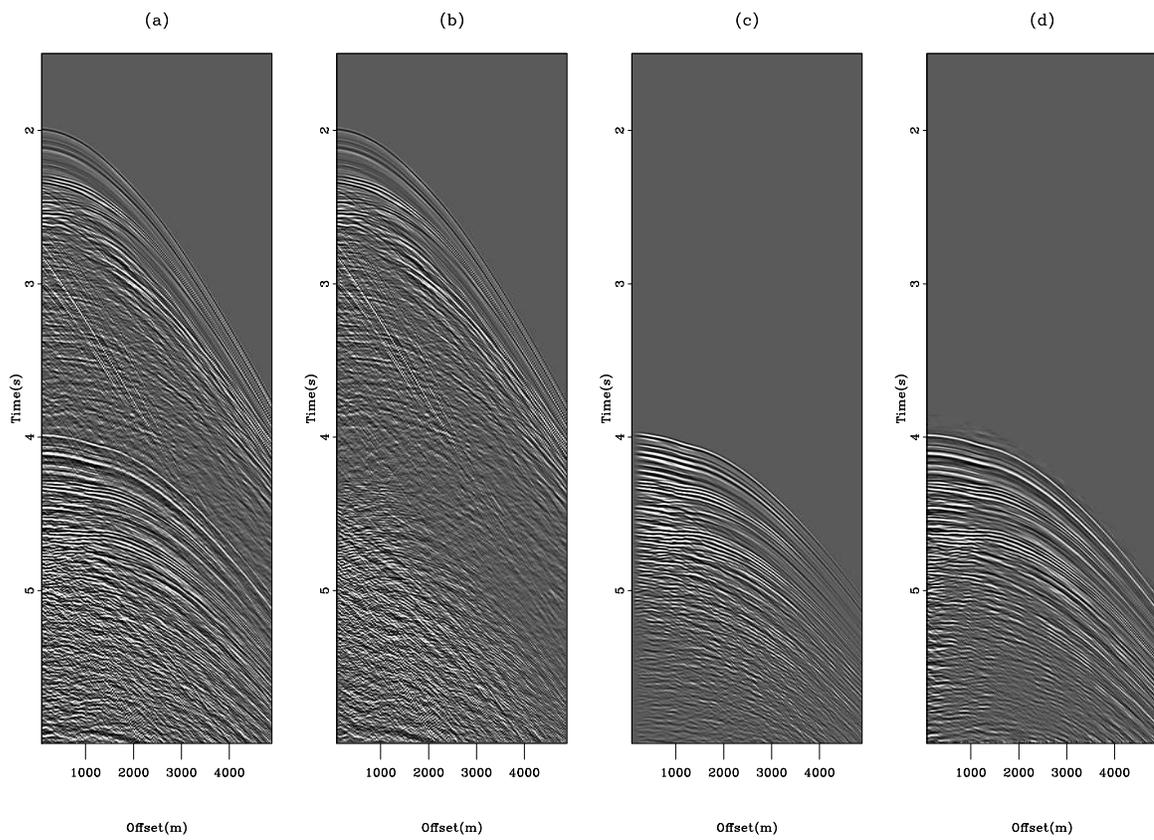


Figure 6.20: Shot gathers outside the salt boundaries at 4500 m for (a) the input data, (b) the estimated primaries, (c) the multiple model with SRMP, and (d) the estimated multiples. `mul2d-comp.shot.4500` [ER]

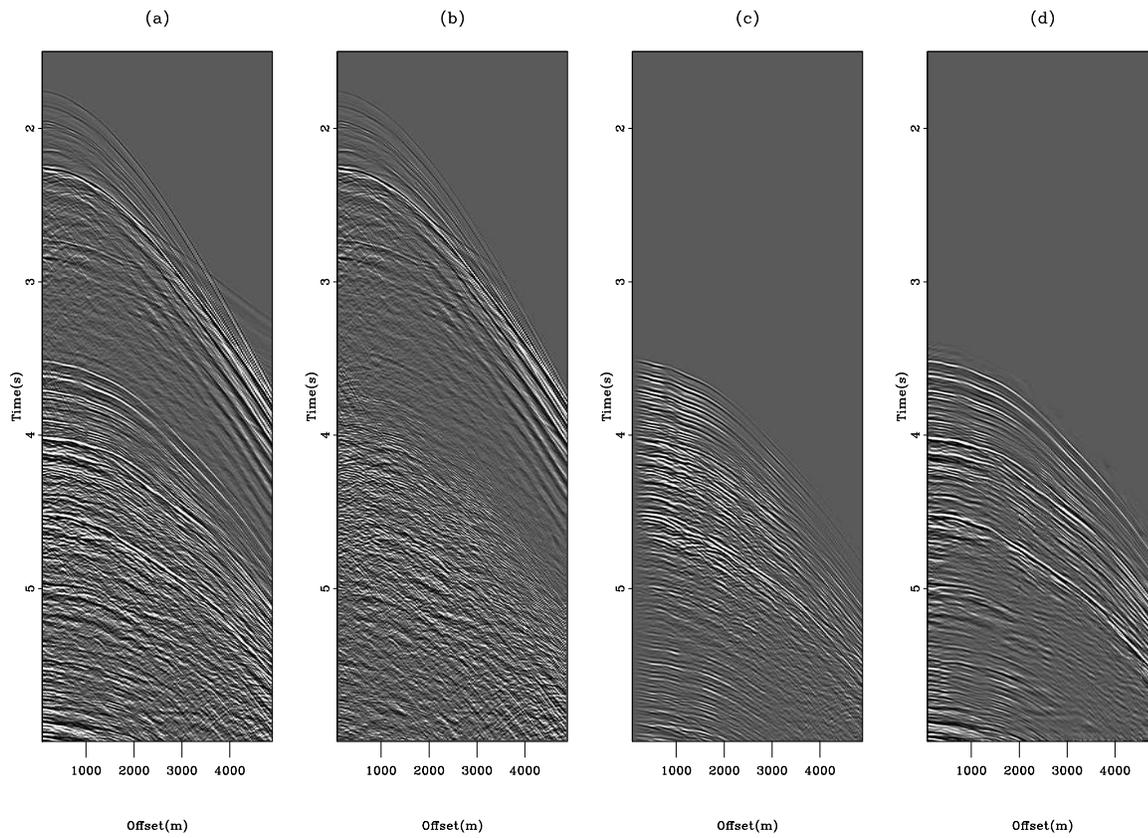


Figure 6.21: Shot gathers inside the salt boundaries at 12000 m for (a) the input data, (b) the estimated primaries, (c) the multiple model with SRMP, and (d) the estimated multiples.

`mul2d-comp.shot.12000` [ER]

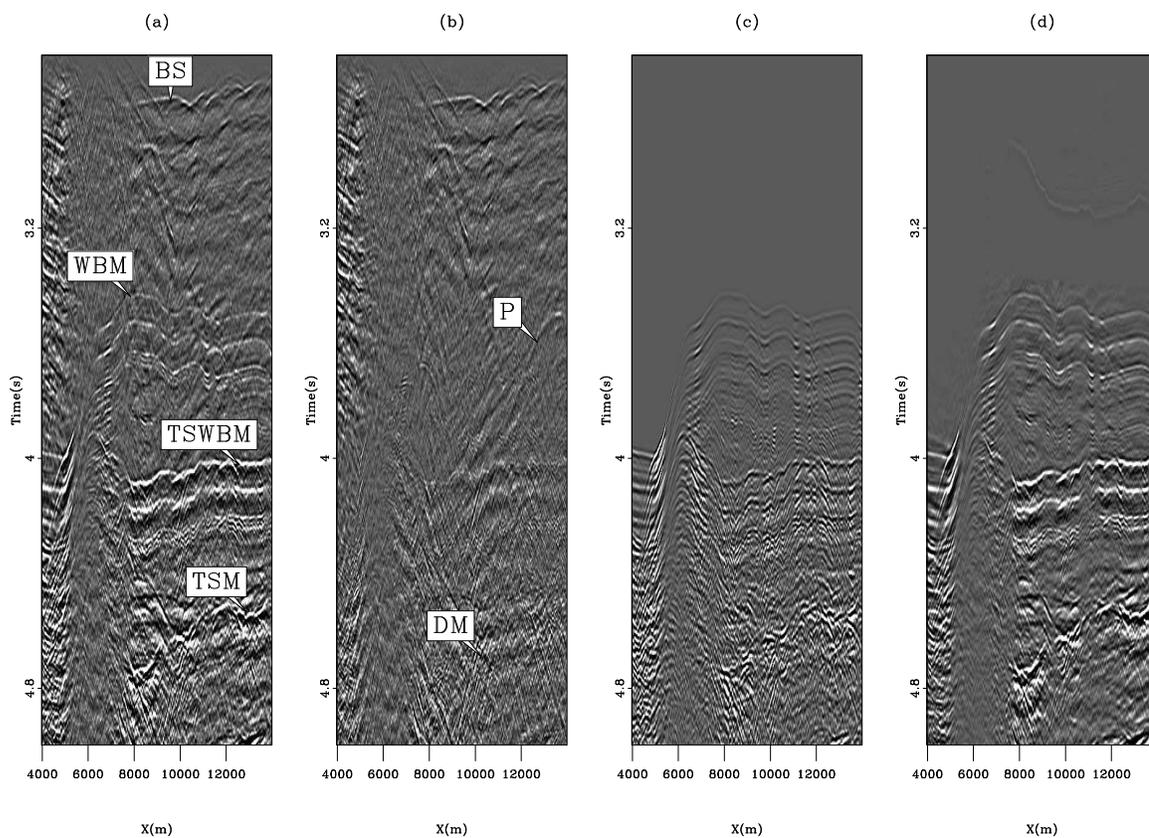


Figure 6.22: A comparison of stacked sections for (a) the input data, (b) the estimated primaries, (c) the multiple model with SRMP, and (d) the estimated multiples. Some primaries (shown as P) in (b) are recovered while some diffracted multiples (shown as DM), not properly modeled by SRMP, are remaining. `mul2d-comp.stack` [CR]

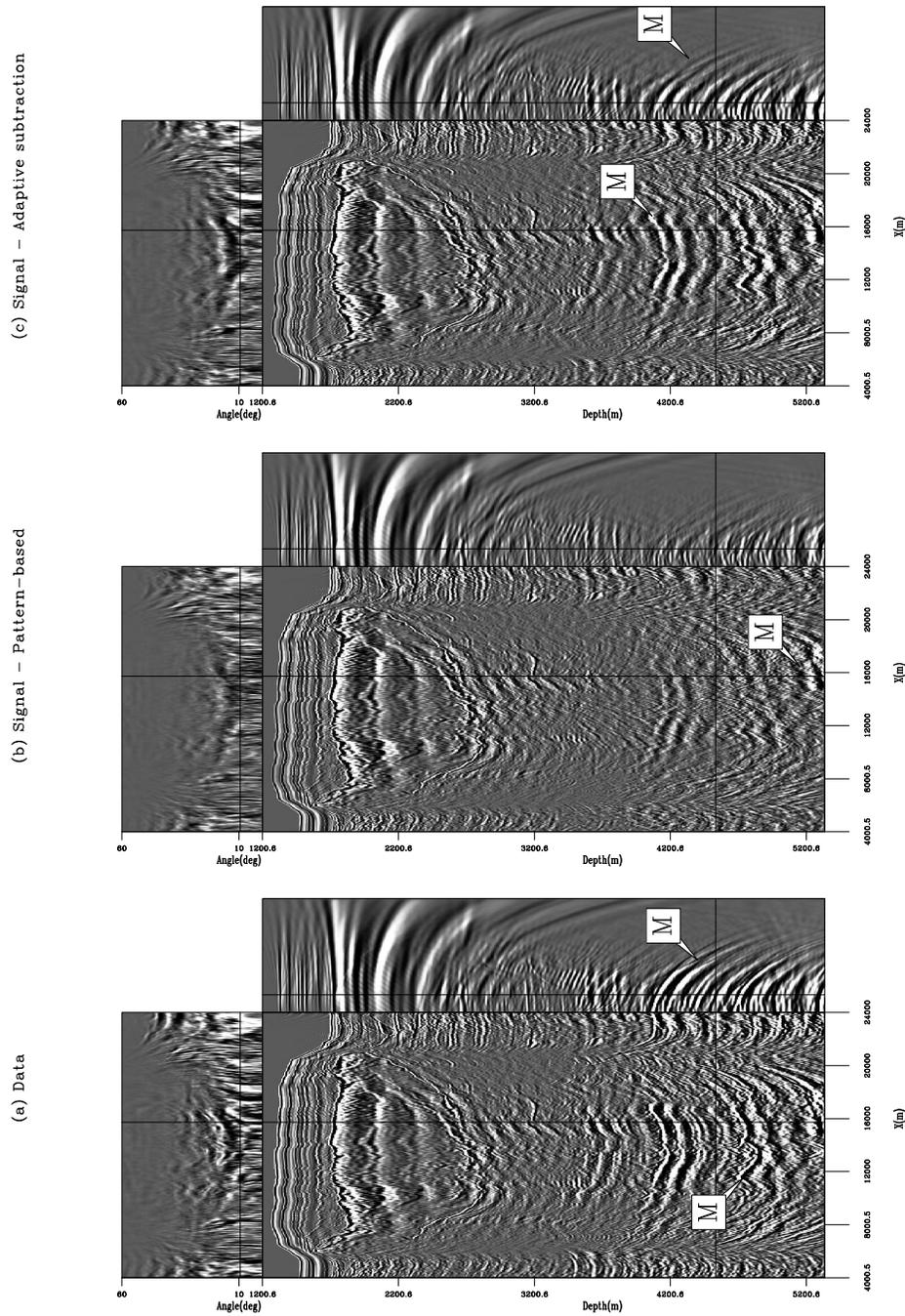


Figure 6.23: comparison of migrated sections in the angle domain for (a) the input data with multiples, (b) the estimated primaries from the pattern-based approach, and (c) the estimated primaries with adaptive subtraction. Arrows M point to multiples. `mul2d-cube.mig` [CR,M]

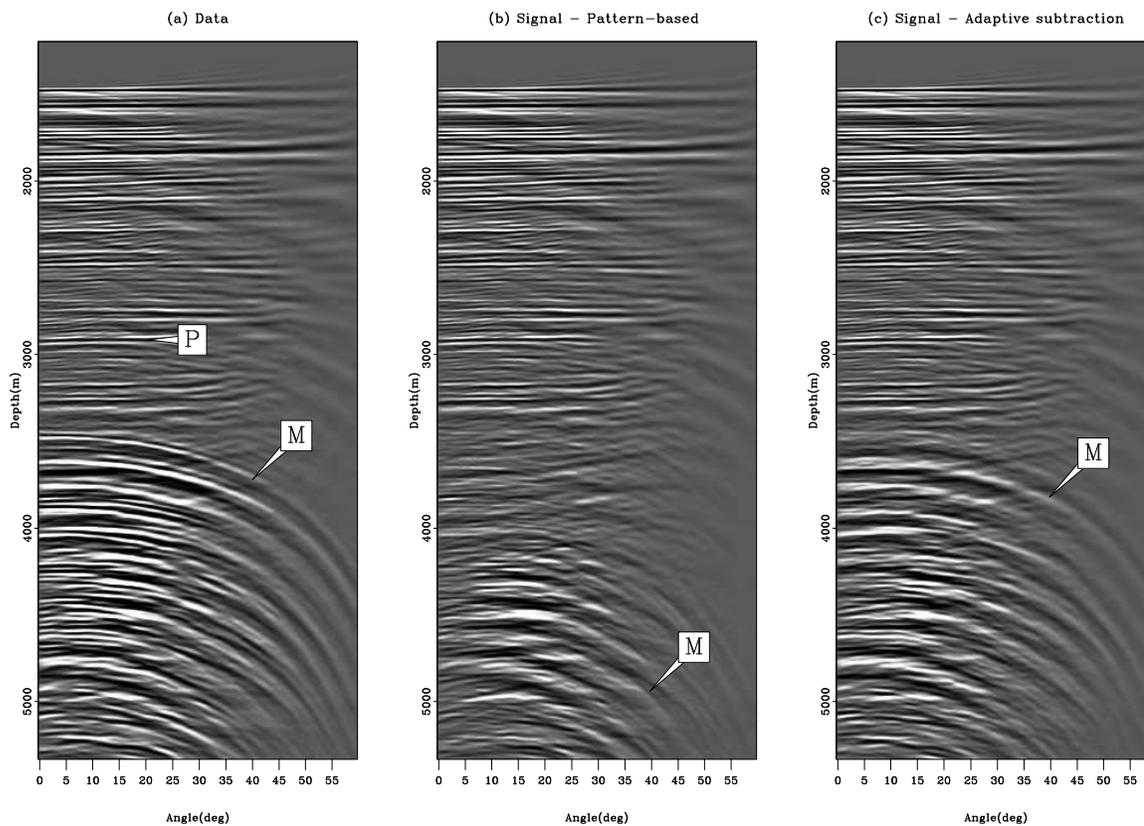


Figure 6.24: Angle domain common image gathers of the Gulf of Mexico dataset for (a) the input data, (b) the estimated primaries with the pattern-based approach and, (c) the estimated primaries with adaptive subtraction. This gather position is $X=4000$ m, outside the salt boundaries. Primaries are flat (shown as P) and multiples are curved (shown as M). `mul2d-sedi.mig` [CR,M]

Chapter 7

Multiple attenuation: A 3-D field data example

SUMMARY

Surface-related multiples are attenuated for one sail line and one streamer of a 3-D dataset donated by the Compagnie Générale de Géophysique (CGG). The survey was carried out in the Gulf of Mexico in the Green Canyon area where salt intrusions close to the water-bottom are present. Because of the complexity of the subsurface, a wavefield method incorporating the full 3-D volume of the data for multiple removal is necessary. This method comprises a modeling of the multiples where the data are used as a prediction operator and a subtraction step where the model of the multiples is usually adaptively removed from the data with matching filters. The accuracy of the multiple model depends on the source/receiver coverage at the surface. When this coverage is not dense enough, the multiple model contains errors that make the subtraction more difficult to succeed. In these circumstances, one can either (1) improve the modeling step by interpolating the missing traces, (2) improve the subtraction step by designing methods that are less sensitive to modeling errors, or (3) both. For this dataset, the second option is investigated by predicting the multiples in a 2-D sense (as opposed to 3-D) and performing the subtraction with a pattern-based approach. Because some traces and shots are missing for the 2-D prediction, the data are interpolated in the inline direction with an

hyperbolic radon transform with and without sparseness constraints. The interpolation with sparseness constraint yields the best multiple model. For the subtraction, the pattern-based method proves to be more effective than adaptive subtraction with matching filters at removing surface-related multiples when the multiple model is not accurate.

INTRODUCTION

The surface-related multiple elimination (SRME) technique (Verschuur et al., 1992) is a wave-field method that (1) uses the data as prediction operators to generate a model of the multiples (the noise), and (2) adaptively subtract the model from the data in order to estimate the primaries (the signal). Both steps are equally important for successfully attenuate the multiples. The prediction of the multiples requires a dense sampling of both shots and receivers across the whole survey. With modern acquisition geometries, especially in 3-D, this condition is rarely met in practice. In marine seismic, for example, a boat usually tows four to twelve streamers with two to six shots. While the shots and receivers are usually well sampled in the inline direction, e.g., every twenty five meters for the receivers and every fifty meters for the shots, they are not in the crossline direction where the spacing between adjacent sail lines might go up to 150 meters, or more. This creates a lack of crossline offsets that can damage the prediction. In addition, cable feathering, cross-line aperture, and other factors (Dragoset and Jericevic, 1998) make the acquisition geometry irregular. Therefore, 3-D geometries cause serious problems to the applicability of SRME.

To make SRME work with 3-D data, both the prediction and the subtraction steps should be looked at. The prediction can be improved by devising strategies for regularizing the data by interpolation or extrapolation. The subtraction can be improved by designing methods that are less sensitive to modeling errors. However, most of today's research is focused on the prediction. For instance, Biersteker (2001) interpolates the missing data before the multiple prediction to insure dense sampling of both shots and receivers. This technique is computationally expensive but leads to very good prediction results because the multiples can be estimated in a 3-D sense. Hill et al. (2002) predict the multiples with a beam method from a 3-D model of the subsurface. This method is much faster than the full interpolation because it

relies on a depth model to kinematically predict the multiples. However, being model based, this technique can only attenuate a limited number of multiples from interpreted 3-D surfaces. Recently, van Dedem (2002) showed how a 3-D Fresnel zone reconstruction method could postpone the interpolation step within the multiple prediction itself. Once the multiple model is estimated, these authors rely on adaptive subtraction with matching filters to subtract the estimated model. However, this technique is known to suffer from leakage problems (Spitz, 1999; Berkhout and Verschuur, 2004) and to be quite sensitive to modeling errors (i.e., Chapter 6).

Although being an important part of the process, very little is done to improve the subtraction step. For 3-D data, Dragoset et al. (2000) propose predicting the multiples streamer by streamer in 2D and subtracting them with flexible adaptive subtraction schemes with time-varying filters. The main problem with these methods is that they can damage primaries. The obvious advantage is that dealing with the prediction in a 2-D sense makes the computation of the multiples affordable for large surveys, relying on the subtraction step to deal with slight variations between the modeled and recorded multiples. The major drawback is that out of plane multiples, e.g., diffracted multiples, are not properly modeled, if at all. Herrmann and Verschuur (2004) separate multiples from primaries in the Curvelet domain by thresholding the data where the noise is present.

In 2003, CGG donated 3-D marine data from a non-exclusive survey in the Gulf of Mexico from the Green Canyon area (Figure 7.1). The acquisition consists of four streamers placed every 150 m with a flip-flop source interval of 37.5 m, the same source being repeated every 75 m. The receiver spacing is 25 m and the maximum offset equals 8.1 km. In addition, the presence of salt bodies with complex structure generate multiple events bouncing outside the 2-D plane between a given source and receiver. Thus, the acquisition geometry and the geology make the multiple removal with SRME very challenging. The goal of this Chapter is to illustrate on one seismic line for one streamer a multiple attenuation strategy based on a 2-D prediction of the multiples (Dragoset et al., 2000) and a subtraction with the pattern-based approach of Chapter 6. This subtraction technique is robust to modeling errors and does not suffer from leakage problems. Application of this method to this dataset proves that the pattern-based method gives far better multiple attenuation results than adaptive subtraction

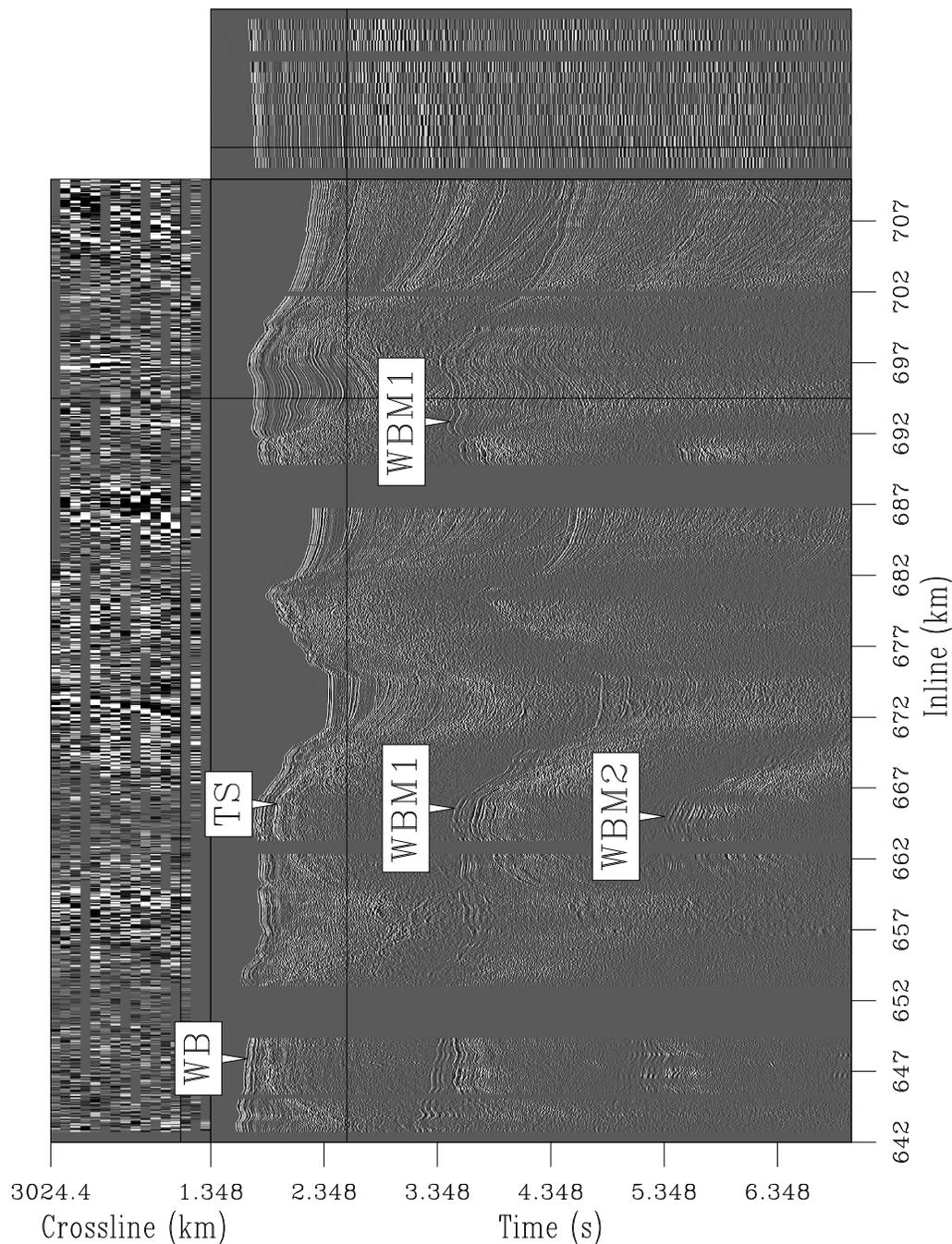


Figure 7.1: Nearoffset section of the binned data for the CGG dataset. The cube is 70 km long by 2 km wide. Note the crossline dips in the right panel and the complex overburden where salt bodies are present close to the surface. The top part is a time slice at 2.4 s. Arrow WB shows the sea bottom, TS a top of salt reflection, WBM1 a first order water-bottom multiple, and WBM2 a second order water-bottom multiple. `mul3d-truenearoffset` [NR]

with matching filters.

This Chapter starts by a description of an interpolation technique for shots and near-offset traces used before 2-D surface-related multiple prediction (SRMP) (Verschuur et al., 1992). The interpolation is done in the CMP domain with a hyperbolic radon transform with and without sparseness constraints. Inversion with the sparseness constraint yields better interpolation results. Then, the multiple model obtained from the 2-D prediction is presented. Compared with the actual multiples present in the data, kinematic and amplitude errors are clearly visible in the multiple model. Finally, the multiples are attenuated with the pattern-based approach of Chapter 6 and with adaptive subtraction. The data are then prestack migrated with and without multiples for the two subtraction techniques. Comparisons of the migration results illustrate that the pattern-based technique gives the best multiple attenuation results while preserving the primaries.

INTERPOLATION

Before creating a multiple model with 2-D SRMP, shots and near-offset traces need to be reconstructed. This section presents an interpolation technique based on the hyperbolic radon transform (HRT). The HRT is performed with and without sparseness constraints to assess the effect of such a choice on the multiple prediction.

Theory

First, the shot gathers are sorted into CMP gathers. Therefore, shot interpolation becomes a trace interpolation problem in the CMP domain. The traces to be interpolated are replaced with zeroed traces. Let's call \mathbf{d} a given CMP gather to be interpolated. The goal is to find a model space \mathbf{m} that minimizes the difference \mathbf{r}_d between the known data and the modeled data with an HRT operator \mathbf{L} , i.e.,

$$\mathbf{0} \approx \mathbf{r}_d = \mathbf{Lm} - \mathbf{d}. \quad (7.1)$$

Because missing traces are present in the data, a masking operator \mathbf{M} is introduced such that the model will only fit the known data as follows:

$$\mathbf{0} \approx \mathbf{r}_d = \mathbf{M}(\mathbf{L}\mathbf{m} - \mathbf{d}). \quad (7.2)$$

The masking operator equals one where the data are known, zero otherwise. Solving for \mathbf{m} in a least-squares sense leads to the following objective function

$$f(\mathbf{m}) = \|\mathbf{M}(\mathbf{L}\mathbf{m} - \mathbf{d})\|^2. \quad (7.3)$$

In most cases, the data to be interpolated are aliased, creating strong artifacts in \mathbf{m} . An efficient way to mitigate these artifacts is by introducing a regularization operator in equation (7.3) that will enforce sparseness in the model space. This regularization isolates the strongest events in the model space while ignoring the weakest. To achieve this, a Cauchy regularization (Sacchi and Ulrych, 1995; Trad et al., 2003) is introduced in equation (7.3) as follows:

$$f(\mathbf{m}) = \|\mathbf{M}(\mathbf{L}\mathbf{m} - \mathbf{d})\|^2 + \epsilon^2 \sum_{i=1}^n \ln \left(1 + \frac{m_i^2}{\hat{m}^2} \right), \quad (7.4)$$

where n is the size of the model space, ϵ is a trade-off parameter between data fitting and model sparseness and \hat{m} is a constant that controls how much damping is applied to the elements of the model space. In practice, both ϵ and \hat{m} are estimated by trial and error.

By definition, introducing the Cauchy norm in equation (7.4) makes the problem of finding \mathbf{m} non-linear. To take the nonlinearity of the objective function into account, $f(\mathbf{m})$ is minimized with the quasi-Newton method introduced in Chapter 2. In practice, this choice leads to satisfying results after a few number of iterations (≈ 25).

Once a model is estimated, the interpolated CMP gathers \mathbf{d}_{int} are obtained by forward modeling the data from the estimated model space $\tilde{\mathbf{m}}$ and replacing the modeled traces by the known traces as follows:

$$\mathbf{d}_{\text{int}} = \mathbf{d} + (\mathbf{I} - \mathbf{M})\mathbf{L}\tilde{\mathbf{m}}, \quad (7.5)$$

where \mathbf{I} is the identity matrix. This approach preserves the known data exactly. Once the CMP gathers are interpolated, the next step consists in resorting the data into shot gathers.

In the next section, the interpolation technique is applied to the data. This example illustrates that the shots can be effectively interpolated with the radon-based technique.

Shot interpolation of the CGG dataset

Figure 7.2 shows the acquisition geometry for one particular sail line. The first streamer (shown as 1) is extracted and used for this field data example. The ship is moving from right to left. Surface currents generate strong cable feathering around 642 km in the inline direction. The receiver positions are displayed for nine shots only every eight kilometers in Figure 7.2.

A near offset section for streamer 1 in Figure 7.2 is shown in Figure 7.3. Salt intrusions make the S/N ratio very low below three seconds. Two orders of surface-related multiples (WBM1 and WBM2) are present in the data. The trace spacing is 75 m. The goal is to interpolate the shots every 25 m and to recover the near offset information before 2-D SRMP.

The shot gathers are first sorted into CMP gathers. The binning parameters are illustrated in Figure 7.4. The bin size is 25×12.5 m. The goal of the interpolation is to fill-up the empty bins shown in Figure 7.4. From the CMP gathers, a masking operator \mathbf{M} is built. This masking operator is set to zero where traces are missing and to one where traces are present.

From the CMP gathers and the masking operator, a model $\tilde{\mathbf{m}}$ is estimated by minimizing equations (7.4) and (7.3). Figures 7.5c and 7.5d show $\tilde{\mathbf{m}}$ when the sparseness constraint is or is not applied, respectively. Most of the aliasing artifacts caused by the missing traces in Figure 7.5 are well attenuated when the Cauchy regularization is used. Note that the remaining artifacts in Figure 7.5c could be attenuated by increasing \hat{m} in equation (7.4) with the possible effect of damaging some useful signal, e.g., below 5 s in Figure 7.5a.

Once $\tilde{\mathbf{m}}$ is estimated, the interpolated CMP gathers can be created with equation (7.5). Figure 7.6 illustrates the interpolation result for one CMP gather. The sparseness constraint (Figure 7.6b) gives a cleaner result and preserve the steep dips better than the radon transform without regularization (Figure 7.6c). Note that the reconstructed traces are less noisy than the

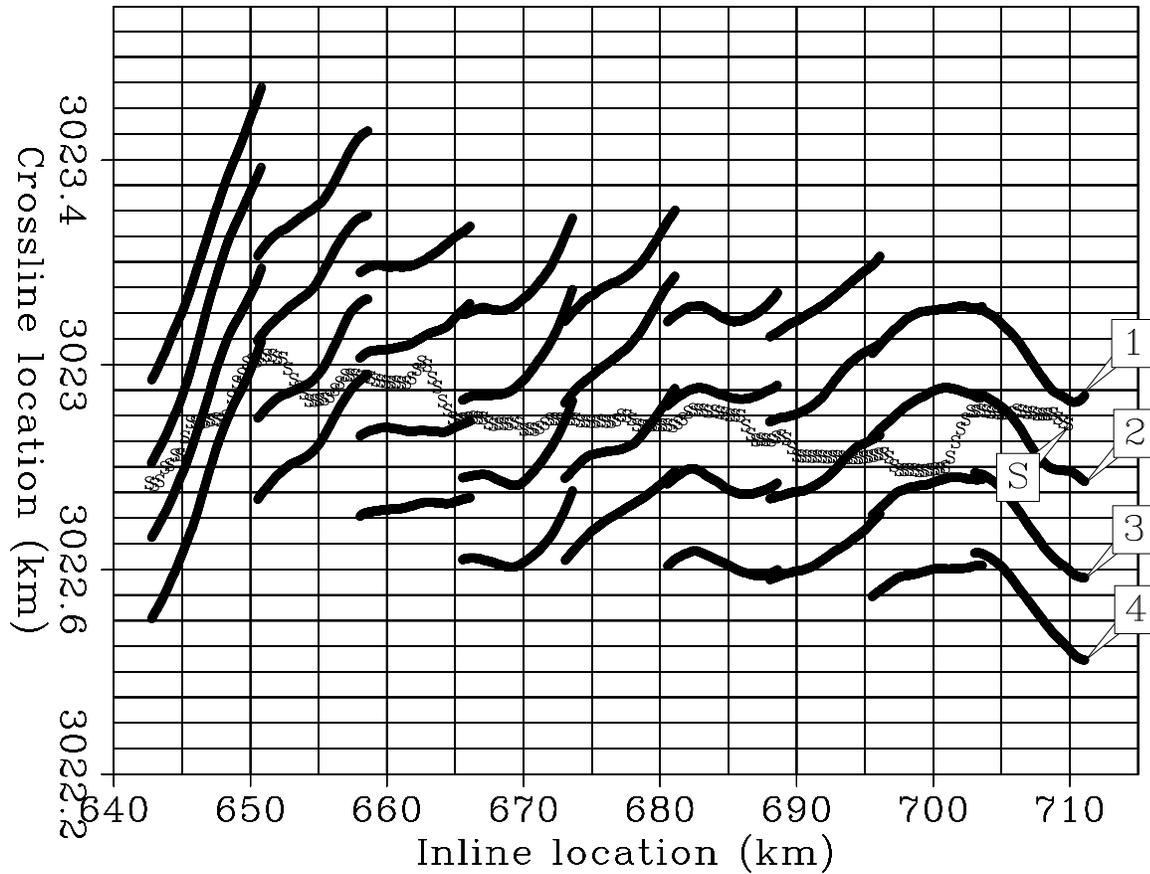


Figure 7.2: Acquisition geometry for one sail line. S points to the locations of one source going from 710 km to 642 km. 1,2,3 and 4 points to the receiver lines for the four different streamers. The receiver locations (black lines) are shown for nine shots for the entire sail line. The data corresponding to streamer 1 are processed in this Chapter. `mul3d-shotline-10` [NR]

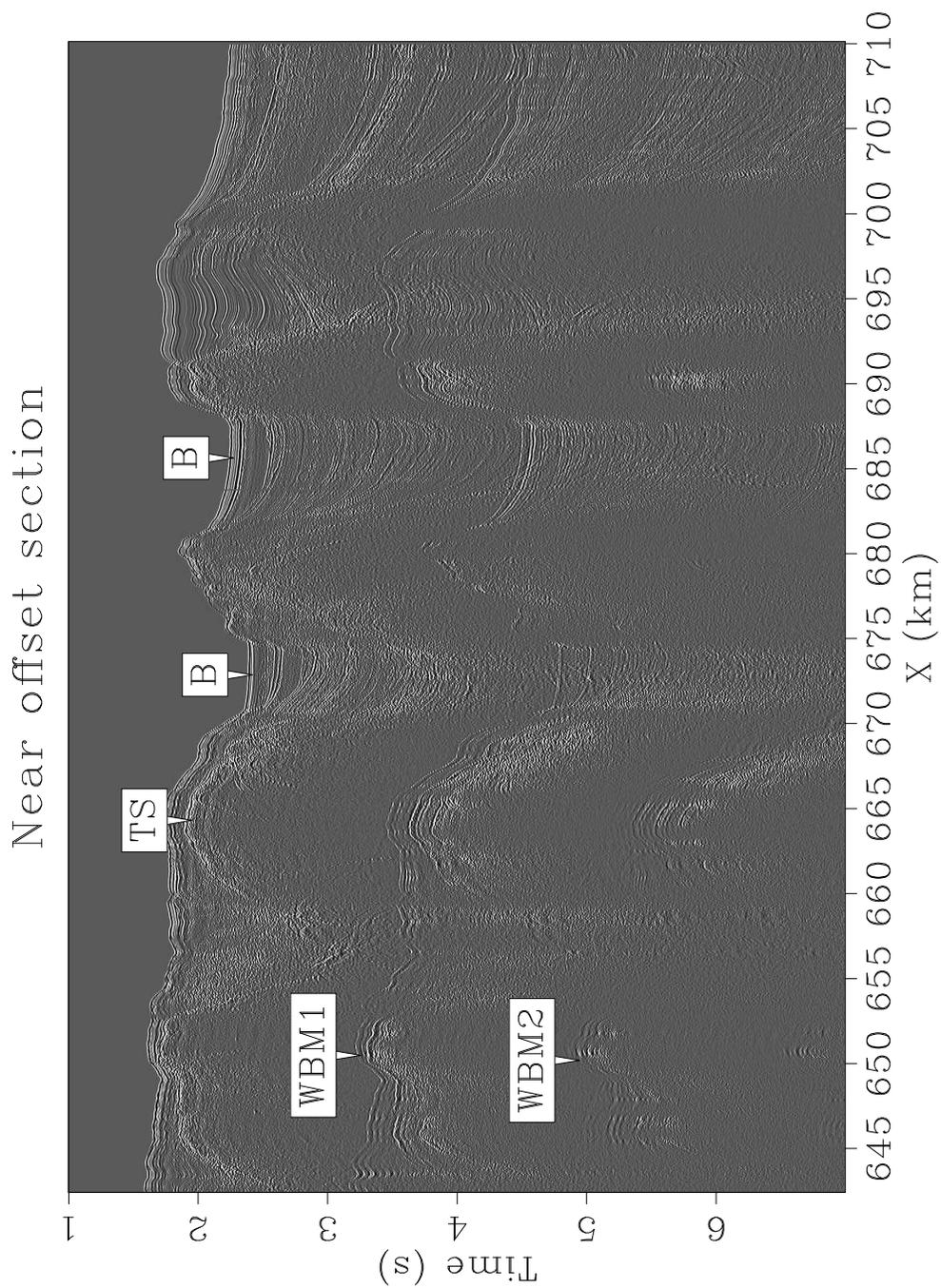


Figure 7.3: Near offset section ($h=250$ m) from streamer 1 in Figure 7.2. TS points to top of salt reflections. B points to sedimentary basins. WBM1 points to a first-order surface related multiple and WBM2 points to a second-order surface related multiple. The S/N ratio is extremely low below the salt. `mul3d-nearoffset` [CR]

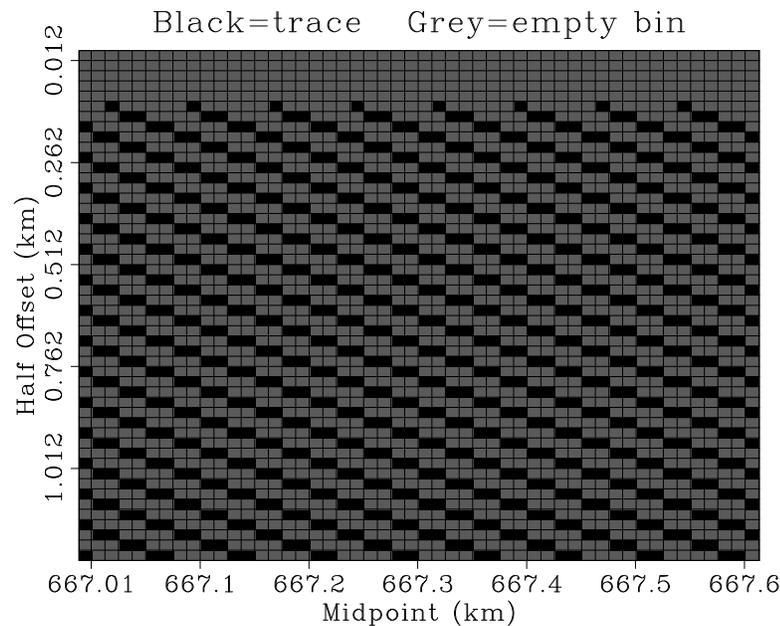


Figure 7.4: Subset of the offset and CMP distributions. The bin size is 12.5 m along the CMP axis and 25 m along the offset axis. The CMP gathers are interpolated along the offset axis to recover the missing shots and the near offset information. `mul3d-cmp-geo` [CR]

known data and that adding some white noise might be needed (Gulunay, 2003). Because the shots are interpolated for multiple prediction only, no processing is applied to correct for this defect.

To better understand the effect of the Cauchy regularization on the steep dips, Figures 7.7a, 7.7b, and 7.7c show the F-K spectra of the CMP gathers in Figure 7.6 for the input data, the reconstructed data with sparseness constraint and the reconstructed data without regularization, respectively. Most of the steep dips are attenuated when no regularization is applied during the inversion. This effect is clearly shown in the 15-25 Hz band where the sparse interpolation shows some aliased energy for events going slower than 1700 m/s (Figure 7.7b). The same events are attenuated when no regularization is applied in Figure 7.7c. Therefore, inversion with Cauchy regularization is preferred for data interpolation in the CMP domain. Note that the aliased energy could be attenuated in the CMP domain by sampling the offset axis on a

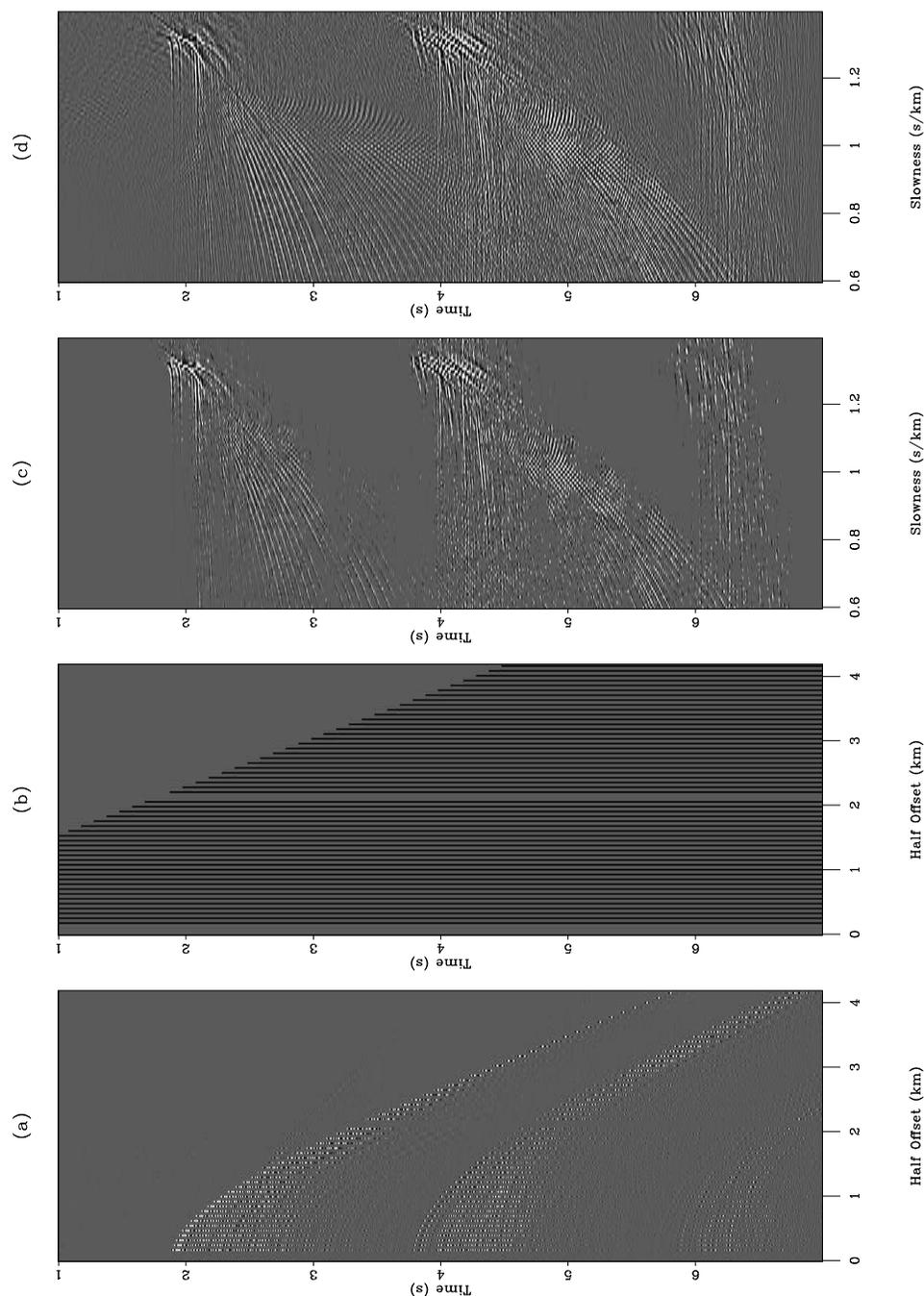


Figure 7.5: Illustration of the radon transform for one CMP gather. (a) Input CMP gather showing strong primaries around 2 s, first-order multiples near 4 s, and second-order multiples near 6 s. (b) Masking operator \mathbf{M} . The black stripes are located at the known data locations ($M_{i,i} = 1$) while the rest of the mask is set to zero. (c) Radon domain with Cauchy regularization, i.e., equation (7.4). (d) Radon domain without Cauchy regularization, i.e., equation (7.3). `mul3d-hrt` [CR]

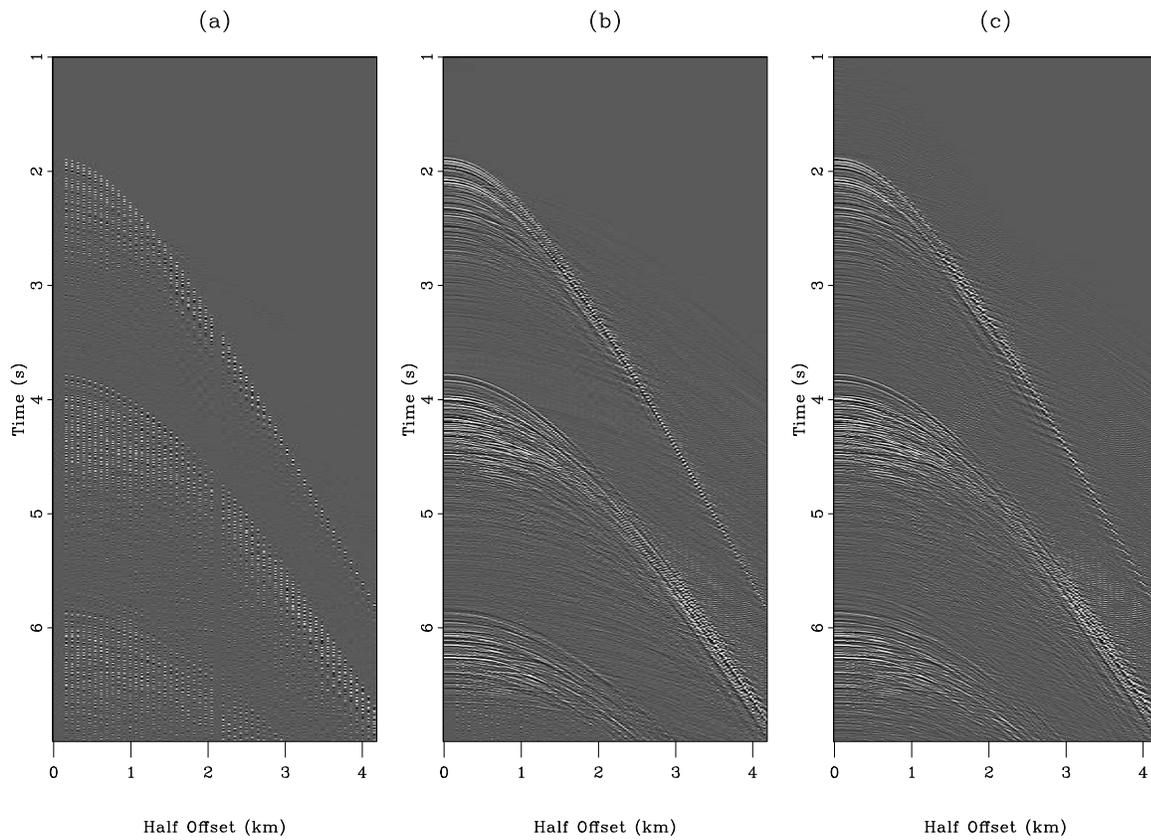


Figure 7.6: (a) Input gather with missing traces. (b) Reconstructed gather with the Cauchy regularization. (c) Reconstructed gather without regularization. `mul3d-cmp-recons` [CR,M]

thinner grid. However, because SRMP is performed in the shot domain, it is difficult to anticipate the effects of this aliasing on the multiple prediction result. In addition, by resampling the CMP offset axis, more CMP gathers would be needed to maintain a uniform grid in the shot domain for the multiple prediction. This extra-requirement would increase the total cost and would add more strain on the interpolation technique.

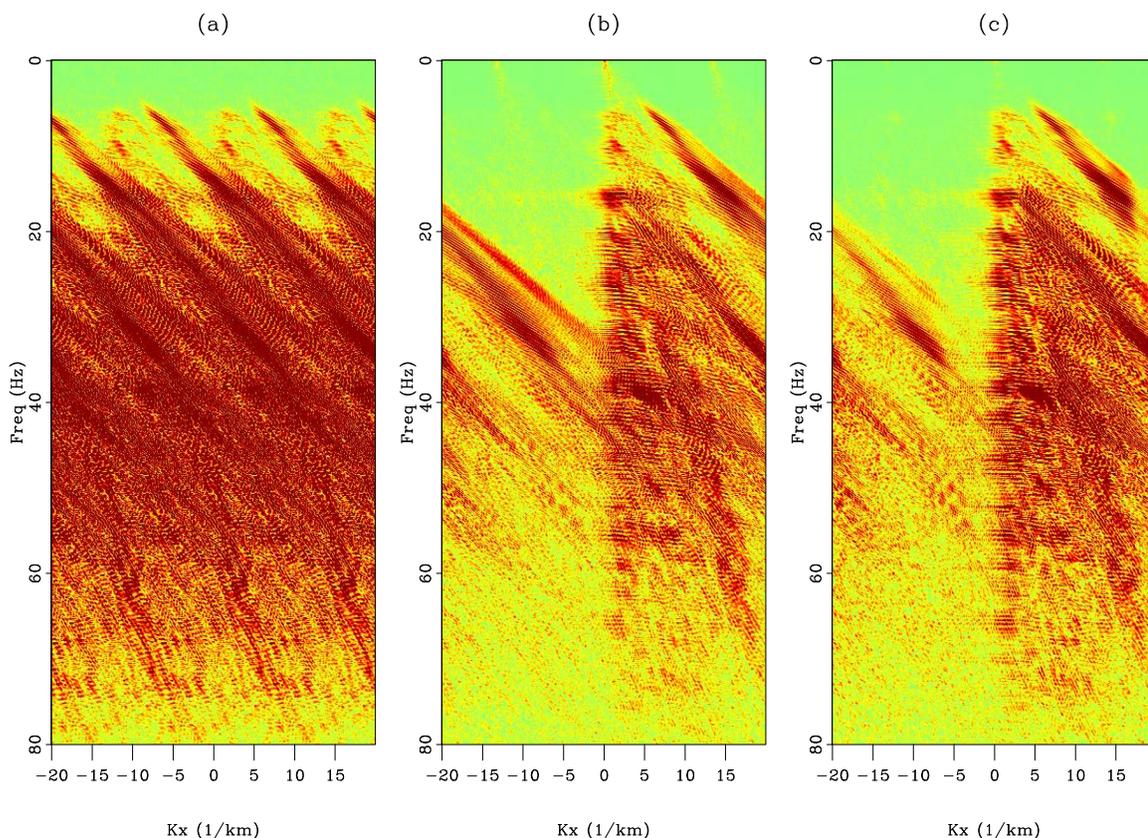


Figure 7.7: F-K spectra of (a) the input CMP gather in Figure 7.6a with missing traces, (b) the reconstructed CMP gather with the Cauchy regularization in Figure 7.6b, (c) the reconstructed CMP gather without regularization in Figure 7.6c. The Cauchy regularization helps to preserve the steep dips better. The steep dips correspond to the aliased energy near 20 Hz. `mul3d-cmp-fk` [CR,M]

Finally, the interpolated CMP gathers are resorted into shot gathers. Figure 7.8 displays four shot gathers after interpolation. Shots one and four are original shots from the survey while shots two and three are new. Note that the near offset traces have been interpolated for all the shots (known and interpolated). Similar to what we observed in the CMP domain, the

Cauchy regularization helped to preserve the steep dips better in Figure 7.8a. In addition, the noise level is quite high for the interpolation result without regularization in Figure 7.8b. Below 5 s, the interpolation without sparseness constraint gives better results. However, for the multiple prediction, these events are of little interest. Overall, despite a fairly coarse acquisition geometry, the radon-based interpolation yields accurate results for a successful multiple prediction.

MULTIPLE PREDICTION RESULTS

Having interpolated the shots to a 25 m grid makes the multiple prediction more accurate. Figure 7.9 shows a comparison for one offset of the input data and the predicted multiples with and without sparse interpolation. The predicted multiples in Figures 7.9b and 7.9c look almost identical to the true multiples in Figure 7.9a. Some artifacts, shown as A in Figure 7.9c, are nevertheless present in the multiple model obtained from the interpolated data without regularization. These artifacts come from the convolution of the noise visible in Figure 7.8b before the water-bottom reflection with coherent energy. Some of them could be attenuated by applying a mute on the interpolated shot gathers.

The multiple model looks relatively accurate for the whole section. However, some important discrepancies exist between the true and modeled multiples in some places. For instance, Figure 7.10 displays the multiple prediction results in an area where the model does not match the observed multiples very well. In Figure 7.10b, the water-bottom multiple (shown as 1) is clearly modeled better with the interpolated data with sparseness constraint than in Figure 7.10c without regularization. The two circles in Figure 7.10b and 7.10c highlight aliasing artifacts that are also visible in Figures 7.11b and 7.11c. These artifacts would disappear by having a smaller sampling of both shot and offset axes. Therefore, these aliasing artifacts are stronger when no Cauchy regularization is applied for the interpolation.

Looking now at shot gathers in Figure 7.11, it appears that the modeling of multiples worked well for the main reflection events. Energetic diffracted multiples are almost totally absent in the predictions of Figures 7.11b and 7.11c, however. Diffracted multiples are generally more difficult to model because they require a very dense surface coverage of sources

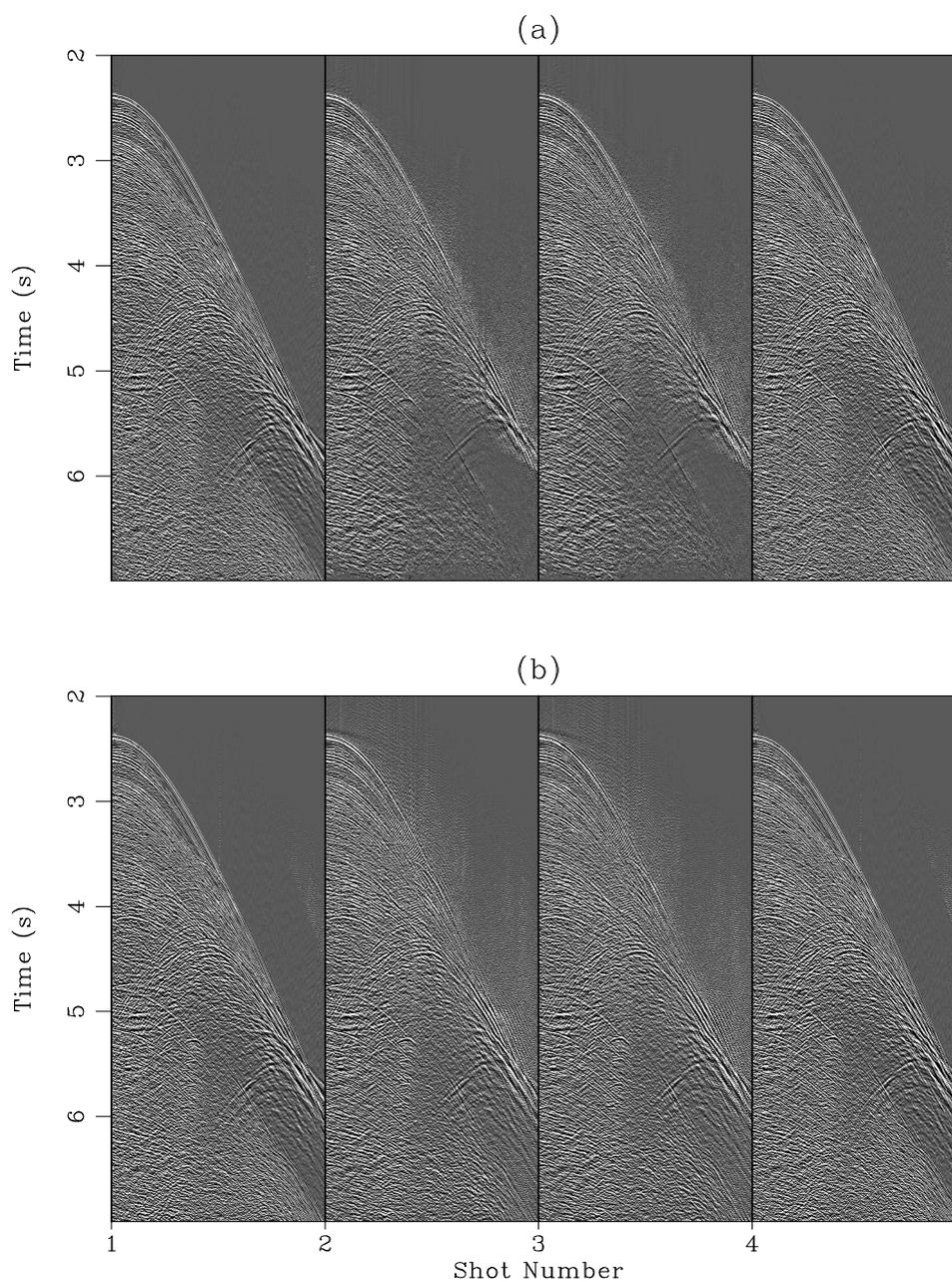


Figure 7.8: Interpolation results in the shot domain with (a) and without (b) Cauchy regularization. Shots one and four are known while shots two and three are new. The near offset traces have been interpolated for all the shots. No AGC has been applied to these gathers. `mul3d-shot-recons` [CR,M]

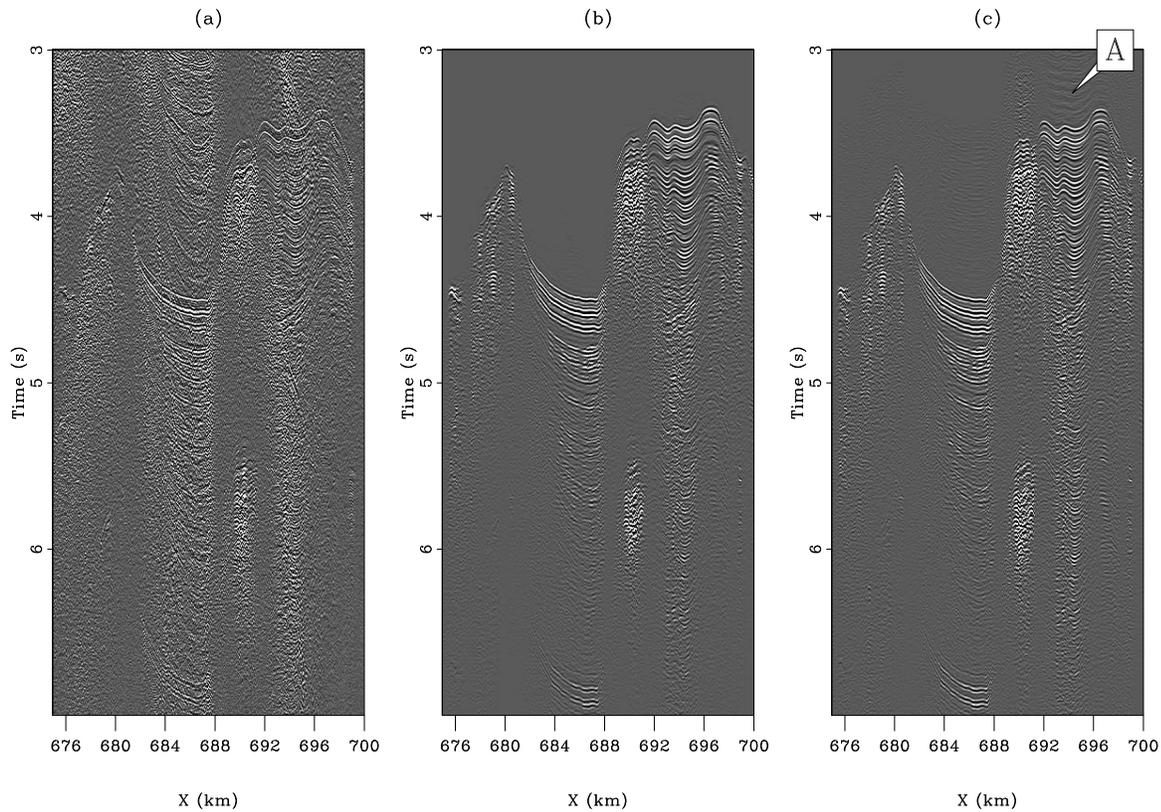


Figure 7.9: Close-ups of common-offset gathers in the vicinity of first-order multiples at 250 m offset for (a) the input data, (b) the predicted multiples with the sparseness constraint, and (c) the predicted multiples without regularization. The horizontal axis represents the midpoint position. Arrow A points to artifacts in the multiple prediction that are caused by interpolation errors.

`mul3d-window1-recons` [CR,M]

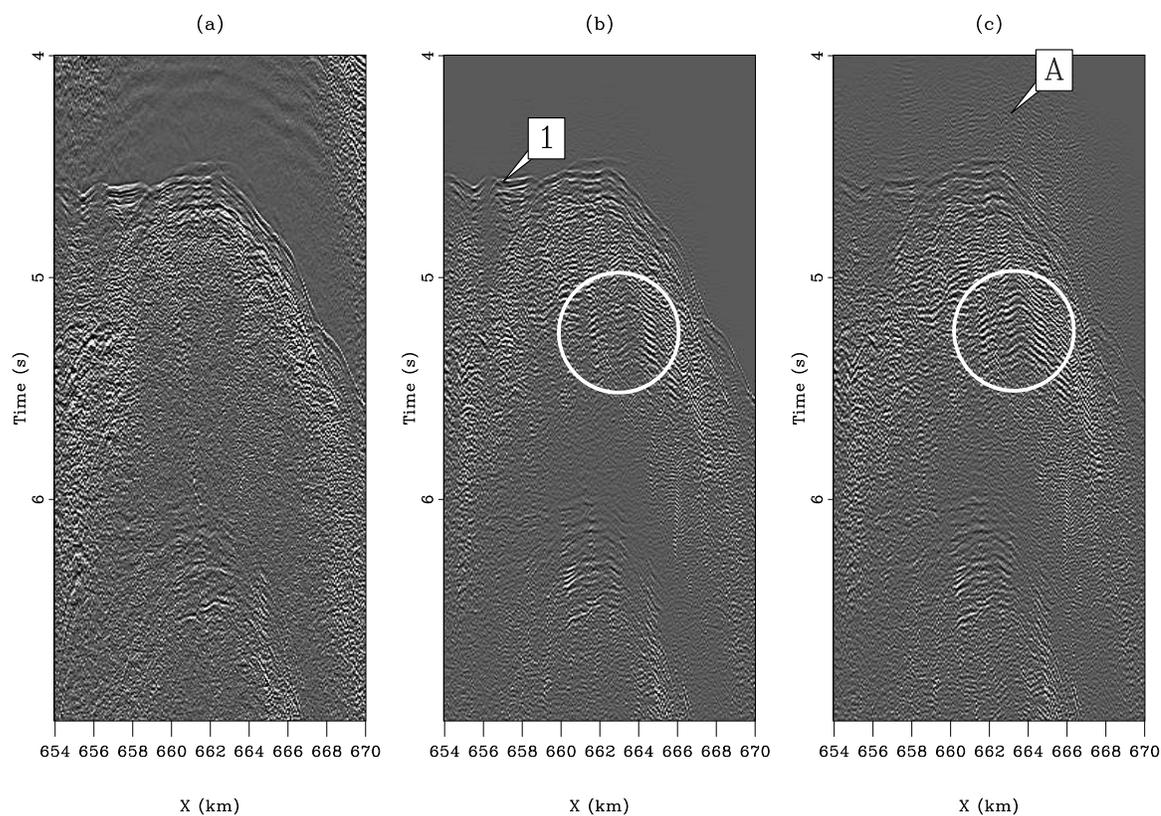


Figure 7.10: Close-ups of common-offset gathers in the vicinity of first-order multiples at 4.425 km offset for (a) the input data, (b) the predicted multiples with the sparseness constraint, and (c) the predicted multiples without regularization. The horizontal axis represents the midpoint position. 1 points to the water-bottom multiple that is better modeled when the Cauchy regularization is used for interpolation. Arrow A points to artifacts similar to those in Figure 7.9c. The two circles point to aliasing artifacts created during the multiple prediction.

`mul3d-window2-recons` [CR,M]

and receivers in order to recover all the dips (Figure 6.10). This effect is amplified in 3-D. Approximating the sail line as a continuous 2-D survey and ignoring 3-D effects lead to large errors in the model of diffracted multiples.

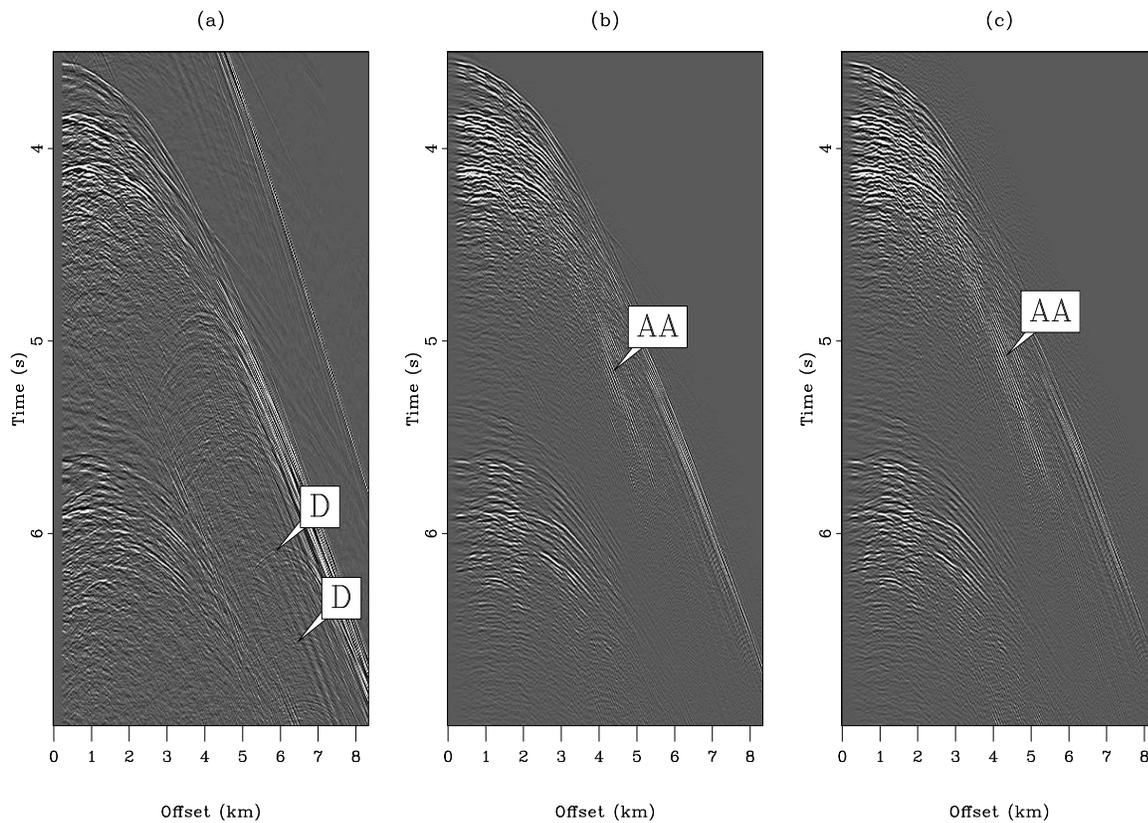


Figure 7.11: Shot gather at 661.65 km for (a) the input data, (b) the predicted multiples with the sparseness constraint, and (c) the predicted multiples without regularization. D points to diffracted multiples that are not properly modeled in either (b) or (c). AA points to aliasing artifacts that create the fictitious events in the circles of Figure 7.10. `mul3d-window3-recons` [CR,M]

The estimation of a multiple model with a 2-D prediction scheme for 3-D field data, although being accurate in most places, suffers from kinematic and amplitude errors. For some events, e.g., diffracted multiples, the modeling fails completely. In addition, the shot interpolation technique needs to be chosen carefully to minimize its impact on the final multiple model. Here, the radon-based approach with a sparseness constraint yields the best multiple model. Given this imperfect model, it is now the goal of the subtraction to come up with the

best estimated primaries. In the next section, multiple attenuation results are presented with a pattern-based and adaptive subtraction technique.

MULTIPLE SUBTRACTION

In this section, the multiple model computed in the preceding section is subtracted from the data with two techniques. The model is obtained after shot interpolation with the sparseness constraint. The first technique is a pattern-based method introduced in Chapter 6 that separates primaries from multiples according to their multivariate spectra. These spectra are approximated with prediction-error filters. The second technique adaptively subtracts the multiple model from the data by estimating non-stationary matching filters (see Chapter 3). The two methods are now briefly described.

Pattern-based approach

The pattern-based approach separates noise and signal according to their multivariate spectra. In this Chapter, the patterns are approximated with 3-D prediction-error filters (PEFs). Calling \mathbf{N} the PEFs for the multiples estimated from the multiple model with SRMP, \mathbf{S} the PEFs for the signal estimated from the Spitz approximation (Spitz, 1999) and \mathbf{s} the primaries, the objective function to minimize becomes

$$g(\mathbf{s}) = \|\mathbf{MN}(\mathbf{s} - \mathbf{d})\|^2 + \epsilon^2 \|\mathbf{MSs}\|^2, \quad (7.6)$$

where \mathbf{M} is a masking operator that preserves the signal where no multiples are present. The signal is estimated iteratively with a conjugate gradient method, and not with the Huber norm of Chapter 2. In few words, the Spitz approximation consists in (1) convolving the noise PEFs \mathbf{N} with the data \mathbf{d} , and (2) estimating the signal PEFs \mathbf{S} from the convolution result. As demonstrated by Abma (1995a), the least-squares inverse of \mathbf{s} in equation (7.6) is an optimal Wiener filter. Similar to Chapter 6, the dataset is divided into macro-patches of 50 consecutive shots with an overlap of 5 shots before the separation. Then the macro-patches are reassembled to form the final result.

Adaptive subtraction

The goal of adaptive subtraction is to estimate the non-stationary filters \mathbf{f} that minimize the objective function

$$g(\mathbf{f}) = \|\mathbf{P}\mathbf{f} - \mathbf{d}\|^2, \quad (7.7)$$

where \mathbf{P} represents the non-stationary convolution with the multiple model obtained with SRMP (i.e., Chapter 3) and \mathbf{d} are the input data. These filters are estimated in a least-squares sense for one shot gather at a time. Note that in practice, a regularization term is usually added in equation (7.7) to enforce smoothness between filters. This strategy is similar to the one used in Chapter 3. The residual vector $\mathbf{P}\mathbf{f} - \mathbf{d}$ contains the estimated primaries.

Comparing the two methods

Both techniques have their advantages and disadvantages. With adaptive subtraction, the primaries are assumed to have minimum energy because the filters are estimated in a least-squares sense [equation (7.7)]. In addition, adaptive subtraction is sensitive to modeling errors (Chapter 6). The strengths of adaptive subtraction are the computing cost and its ease of use, where the filter and patch sizes are only needed. On the contrary, the pattern-based approach is more costly (three to four times) because three inversions need to be run: one for estimating the noise filters \mathbf{N} , one for estimating the signal filters \mathbf{S} , and one for estimating the primaries [equation (7.6)]. The choice of parameters can be also cumbersome because filter and patch sizes are needed for the two sets of filters (noise and signal). In addition, a signal model usually estimated with the Spitz approximation is needed. The strengths of the pattern-based approach are its robustness to modeling uncertainties and the fact that it does not assume that the signal has minimum energy. These two qualities are particularly important wherever the multiple model is not accurate enough, as it is the case for the dataset presented in this Chapter.

Field data results

For this dataset, the choice of parameters such as filter sizes, patch sizes, number of iterations and Lagrange multipliers (e.g., ϵ in equation (7.6)) was determined by trial and error on small portions of the data for both methods. Experience shows that the Lagrange multipliers are the only parameters to be changed from one dataset to another.

Prestack results are presented in Figures 7.12, 7.13, and 7.14 where panel (a) displays the input data, (b) the multiple model, (c) the estimated primaries with adaptive subtraction, and (d) the estimated primaries with the pattern-based approach. Figure 7.12 displays a constant offset section at 500 m offset. The arrows in 1 and 2 point to locations where the multiple model is not accurate. At these locations, the pattern-based approach attenuates the multiples better than adaptive subtraction. Figure 7.13 displays multiple removal results for one shot gather ($X=664$ km). Arrow 1 shows an area where the multiples are poorly predicted and poorly attenuated by both methods. At this location, the pattern-based techniques attenuates more multiples, however. Arrow 2 points to a diffracted event that is present in the multiple model but slightly shifted in time and offset. The adaptive subtraction is unable to remove this event while the pattern-based technique attenuates it. Arrow 3 shows another location where the multiples are better attenuated with the pattern-based approach. Note that some multiples, especially diffracted multiples, are still present in the estimated primaries because these events are totally absent in the model. Finally, Figure 7.14 displays a shot gather where the multiple attenuation worked well for both techniques. This gather ($X=693$ km) comes from an area where the acquisition geometry and the geology are the most appropriate for the 2-D prediction, which explains why both techniques work very well. Arrow 1 points to one event that is still better preserved with the pattern-based method.

Now, the data and the estimated primaries for both subtraction techniques are migrated with a split-step double square-root (DSR) migration algorithm (see Chapter 6). Figures 7.15 and 7.16 show a comparison of the migrated data (shown in (a)), the estimated primaries (shown in (b)), the estimated multiples (shown in (c)), and the multiple model from SRMP (shown in (d)) for the adaptive and pattern-based subtraction techniques, respectively. These results are close-ups of a sedimentary basin. The pattern-based approach (Figure 7.16b) yields

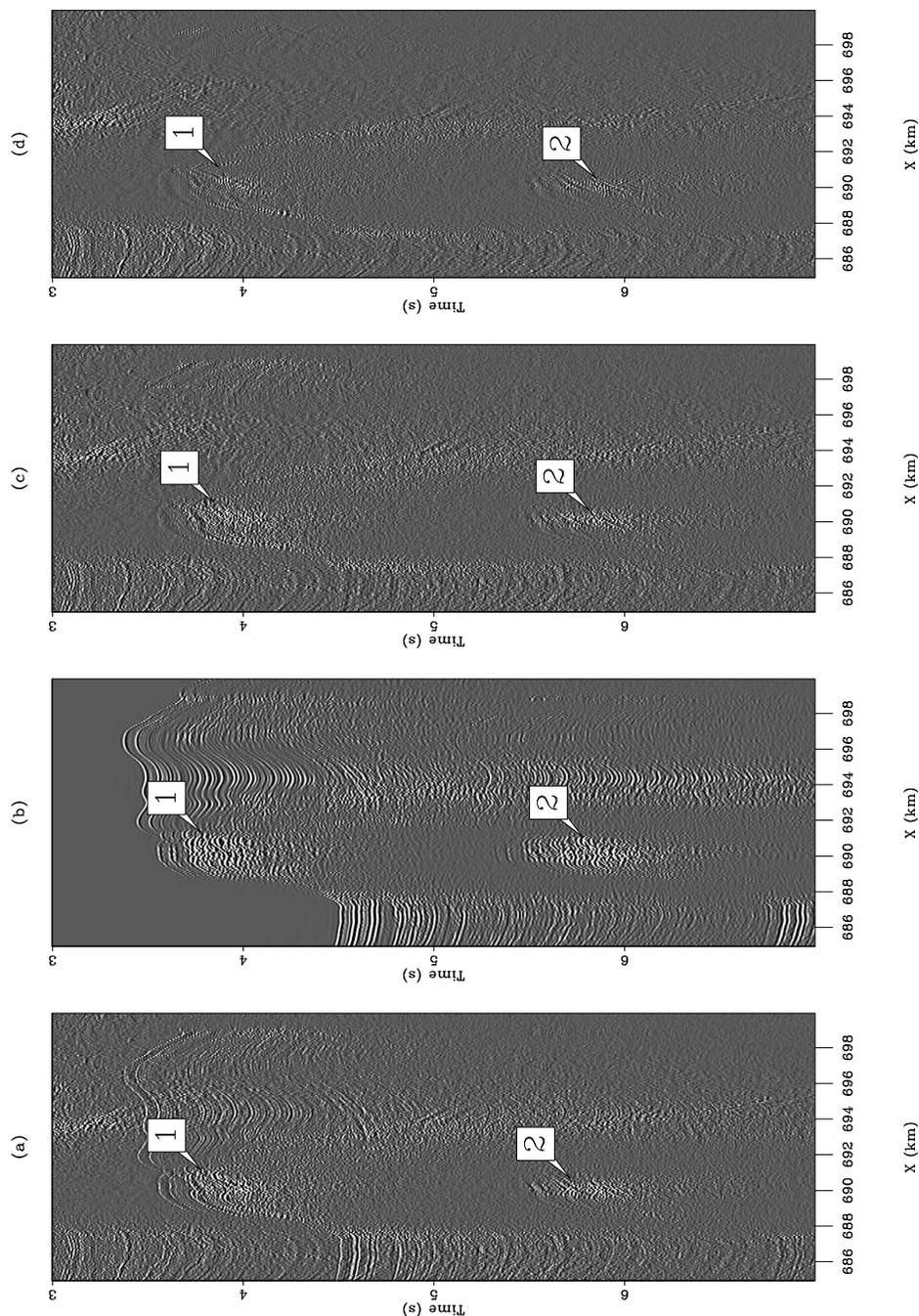


Figure 7.12: A constant offset section at 500 m offset for (a) the input data, (b) the multiple model, (c) the estimated primaries with adaptive subtraction, and (d) the estimated primaries with the pattern-based approach. The horizontal axis represents the midpoint position. 1 and 2 points to area where the multiples are not accurately predicted and where the pattern-based approach performs better than adaptive subtraction. `mul3d-window1-signal` [CR,M]

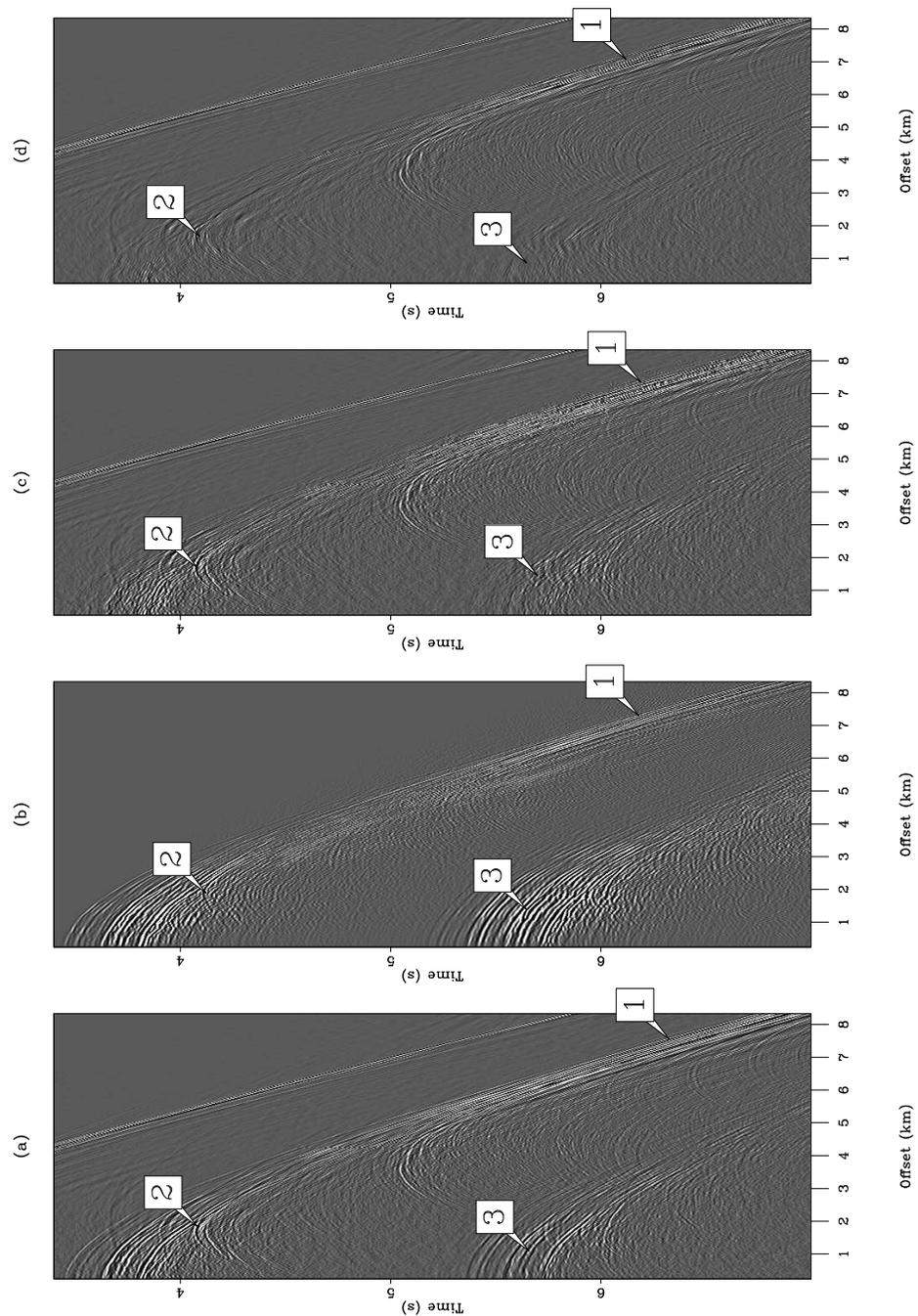


Figure 7.13: A shot gather ($X=664$ km) for (a) the input data, (b) the multiple model, (c) the estimated primaries with adaptive subtraction, and (d) the estimated primaries with the pattern-based approach. 1 shows multiples that are not properly predicted and not completely removed. 2 shows an event that is predicted at the wrong location but still attenuated in (d). 3 shows an area where the pattern-based approach attenuates more multiples than adaptive subtraction. `mul3d-window3-signal` [CR,M]

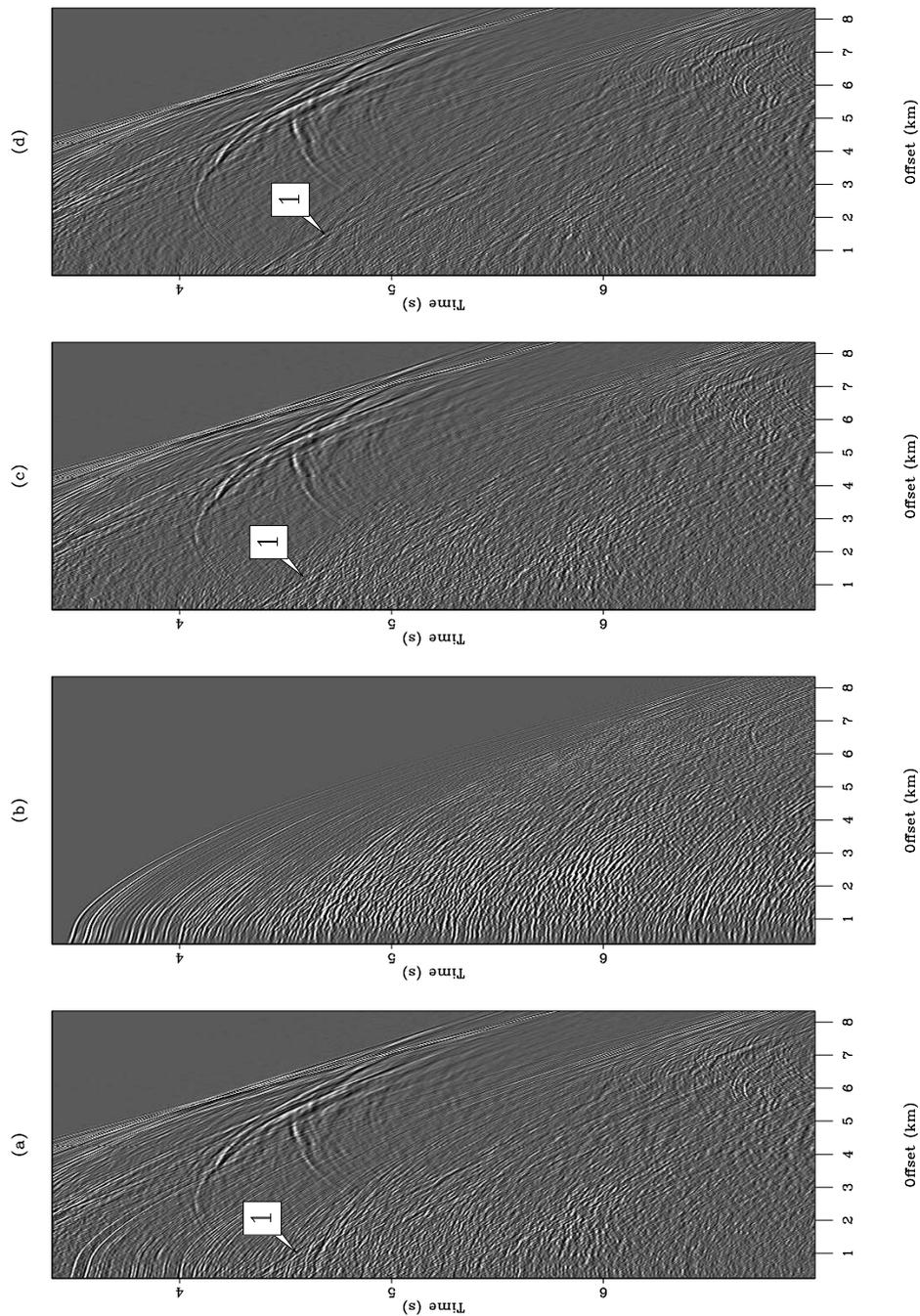


Figure 7.14: A shot gather ($X=693$ km) for (a) the input data, (b) the multiple model, (c) the estimated primaries with adaptive subtraction, and (d) the estimated primaries with the pattern-based approach. 1 points to a primary that is better preserved in (d). `mul3d-window4-signal` [CR,M]

cleaner multiple-free panels. In addition, the multiples inside the circle are better attenuated with the pattern-based approach than with the adaptive-subtraction technique (Figure 7.15b). One reason for these differences comes from the modeling inaccuracies of 2-D SRMP, as exemplified in Figures 7.15c and 7.16c. There, it appears that modeled multiples do not match the true multiples in the migrated data exactly (e.g., Figure 7.15a). The pattern-based approach is able to cope better with the differences between true and modeled multiples.

Now, Figures 7.17 and 7.18 display multiple attenuation results in the presence of a salt body. Again, the base of salt reflection is cleaner and more continuous with the pattern-based approach (Figure 7.18b) than with adaptive subtraction (Figure 7.17b). Figures 7.17d and 7.18d show the migrated multiple model. The model exhibits important kinematic differences with the true multiples. For instance, the water-bottom multiple at 4.2 km is slightly shifted horizontally in the model. The pattern-based approach is able to remove this multiple albeit modeling errors.

Finally, angle-domain common image gathers (ADCIG) are created with the method of Sava and Fomel (2003) after migration (Stolt and Weglein, 1985; Weglein and Stolt, 1999). Figures 7.19a, 7.19b, and 7.19c show three ADCIG for the data, the estimated primaries with adaptive subtraction, and the estimated primaries with the pattern-based approach, respectively. This ADCIG is located within the salt boundaries. The primary at 4.6 s. is the base of salt reflection. Events with large curvatures in Figure 7.19b are multiples. Again, the pattern-based approach produces a better panel. The method of Sava and Guitton (2005) could be used to remove the remaining multiples with a Radon-based technique. As a final example, Figure 7.20 show a comparison of ADCIGs outside the salt body, within the sedimentary basin of Figures 7.16 and 7.15. The primaries are more continuous and better preserved with the pattern-based approach (Figure 7.20b) than with adaptive subtraction (Figure 7.20c).

DISCUSSION

With 3-D data, attenuating surface-related multiples with wavefield techniques such as SRME is particularly challenging. The difficulty stems from the lack of source/receiver coverage, the lack of crossline offsets, and the coarse source/receiver spacing at the surface for the

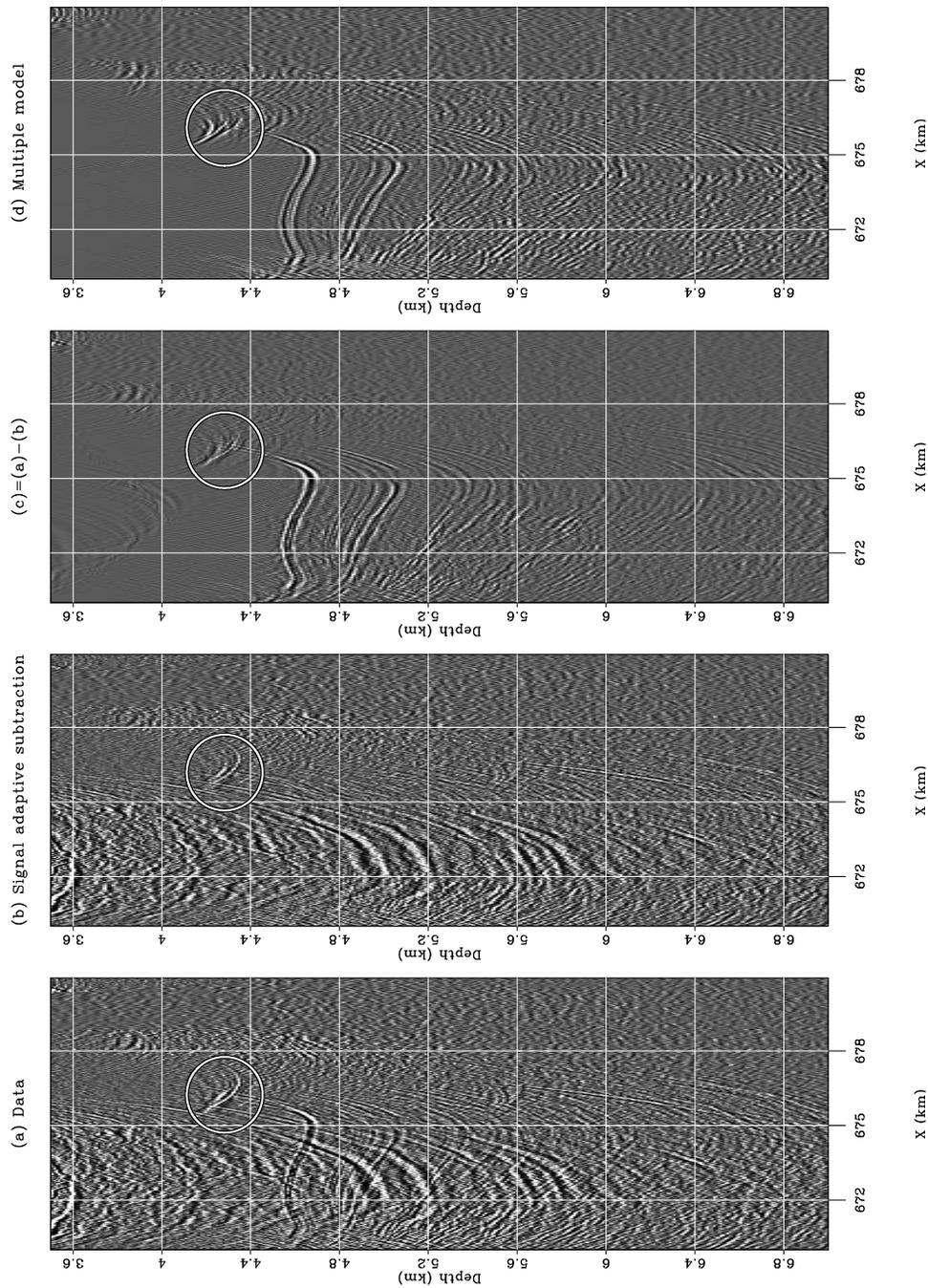


Figure 7.15: Migrated images for (a) the data, (b) the estimated primaries with adaptive subtraction, (c) the difference between (a) and (b), i.e., the estimated multiples, and (d) the multiple model with SRMP. Multiples in the circle are not well attenuated. `mul3d-mig_adaptive` [CR,M]

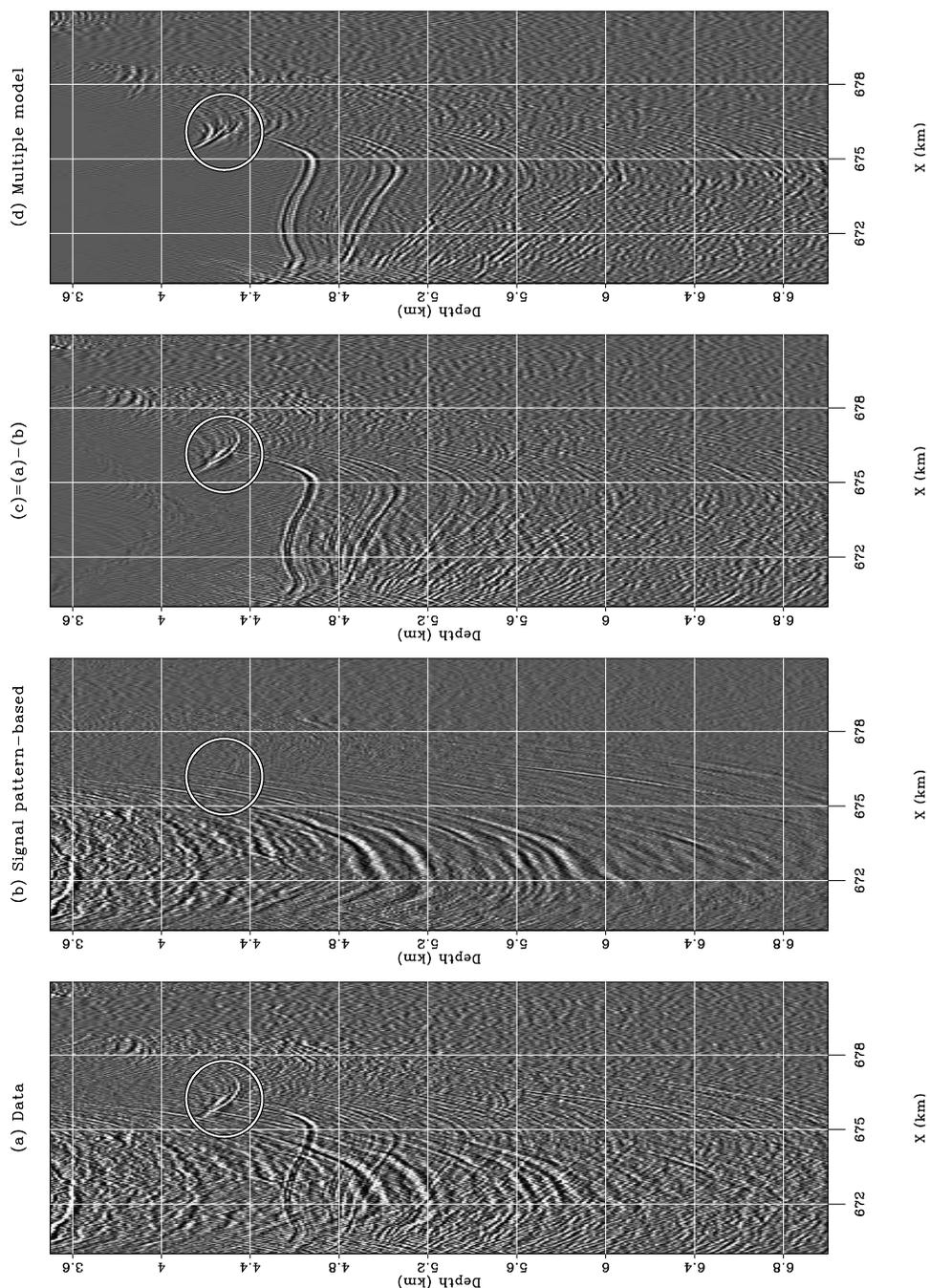


Figure 7.16: Migrated images for (a) the data, (b) the estimated primaries with the pattern-based approach, (c) the difference between (a) and (b), i.e., the estimated multiples, and (d) the multiple model with SRMP. Multiples in the circle are better attenuated in (b) than in Figure 7.15b. In addition, the primaries look more continuous everywhere. `mul3d-mig_pattern` [CR,M]

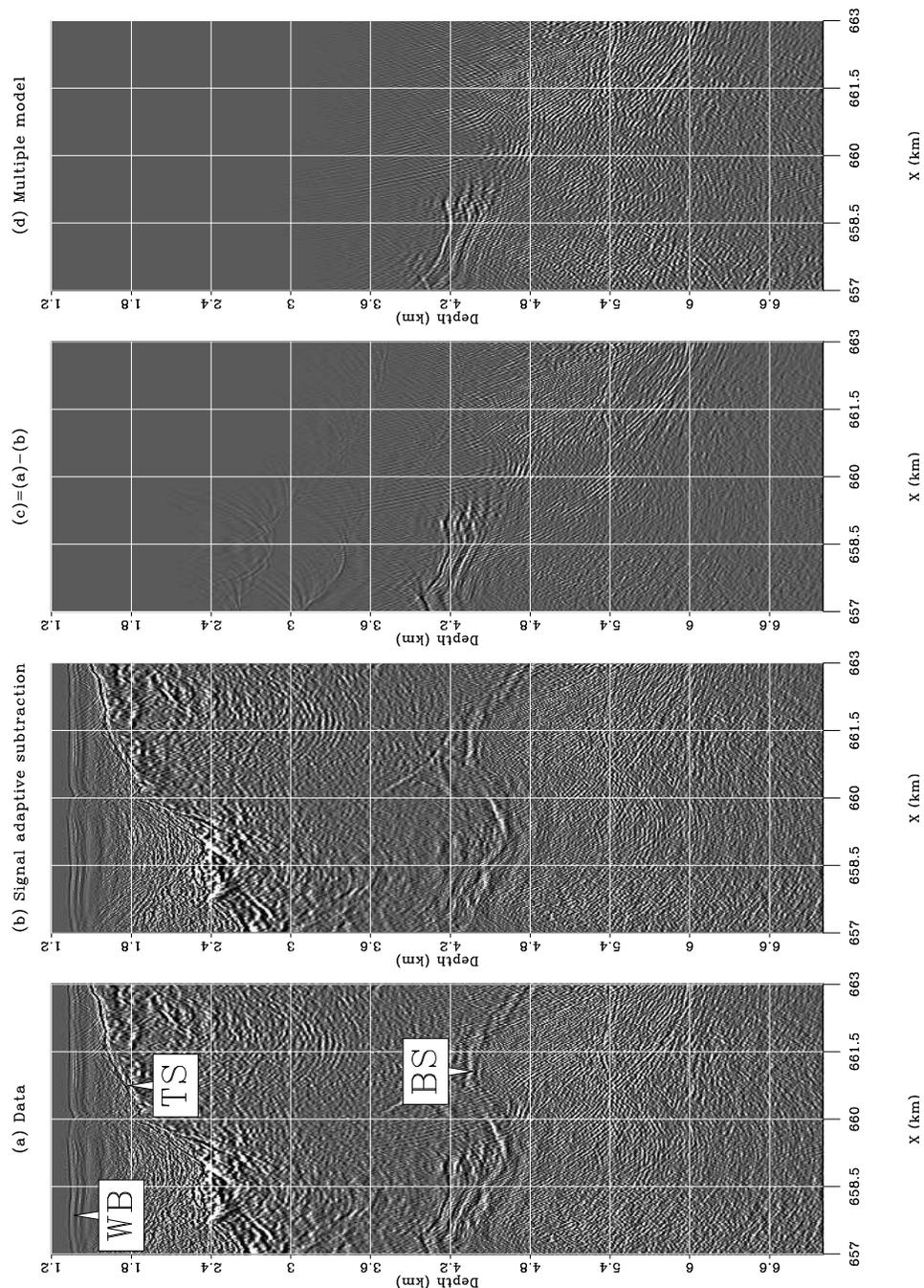


Figure 7.17: Migrated images for (a) the data, (b) the estimated primaries with adaptive subtraction, (c) the difference between (a) and (b), i.e., the estimated multiples, and (d) the multiple model with SRMP. The base of salt is indicated by an arrow in (a). Some subsalt multiples are still present. Arrow WB shows the water bottom, TS the top of salt, and BS the base of salt. `mul3d-mig_adaptive1` [CR,M]

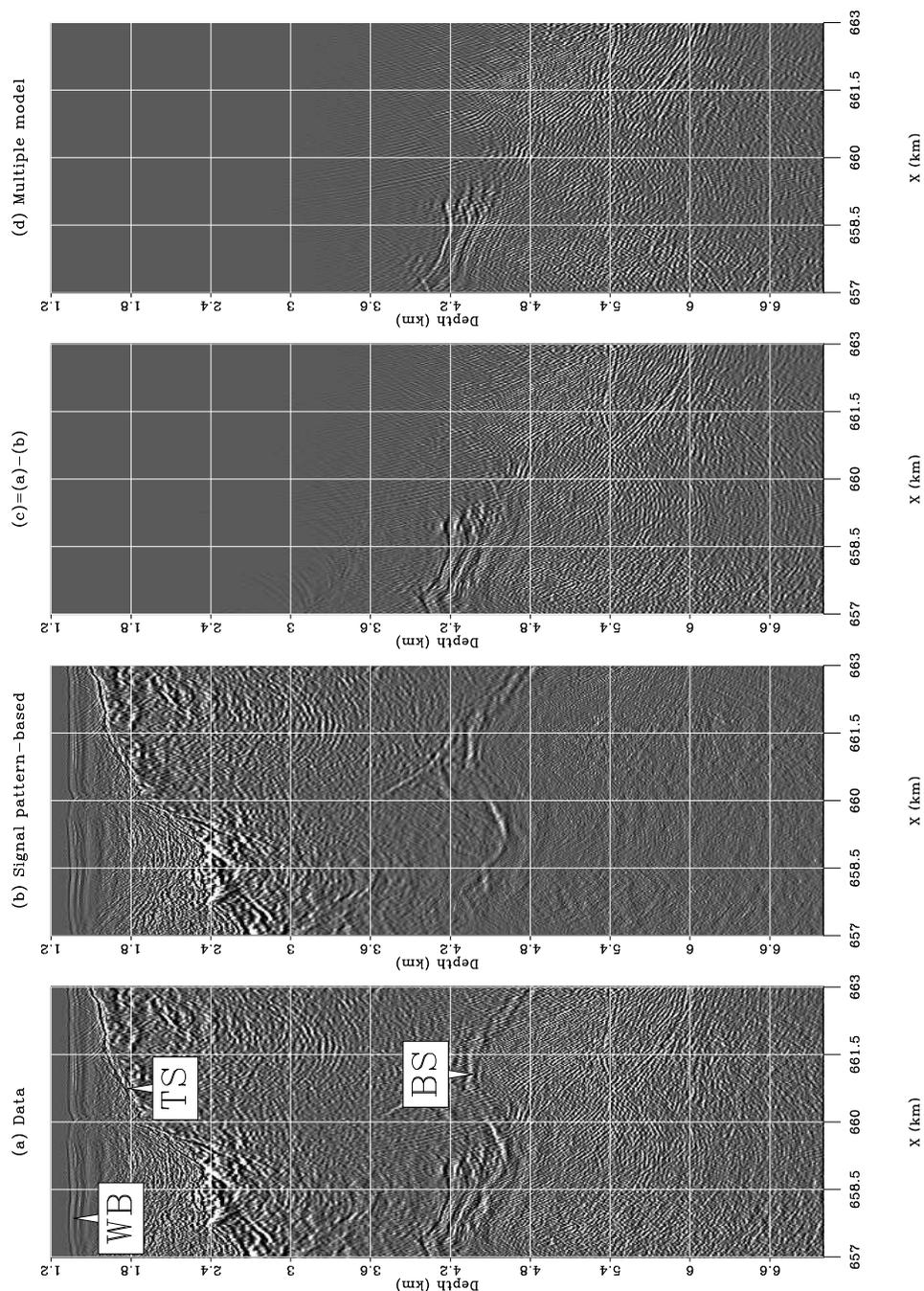


Figure 7.18: Migrated images for (a) the data, (b) the estimated primaries with the pattern-based approach, (c) the difference between (a) and (b), i.e., the estimated multiples, and (d) the multiple model with SRMP. The base of salt is better preserved than in Figure 7.17b and the subsalt multiples are well attenuated. Arrow WB shows the water bottom, TS the top of salt, and BS the base of salt. `mul3d-mig_pattern1` [CR,M]

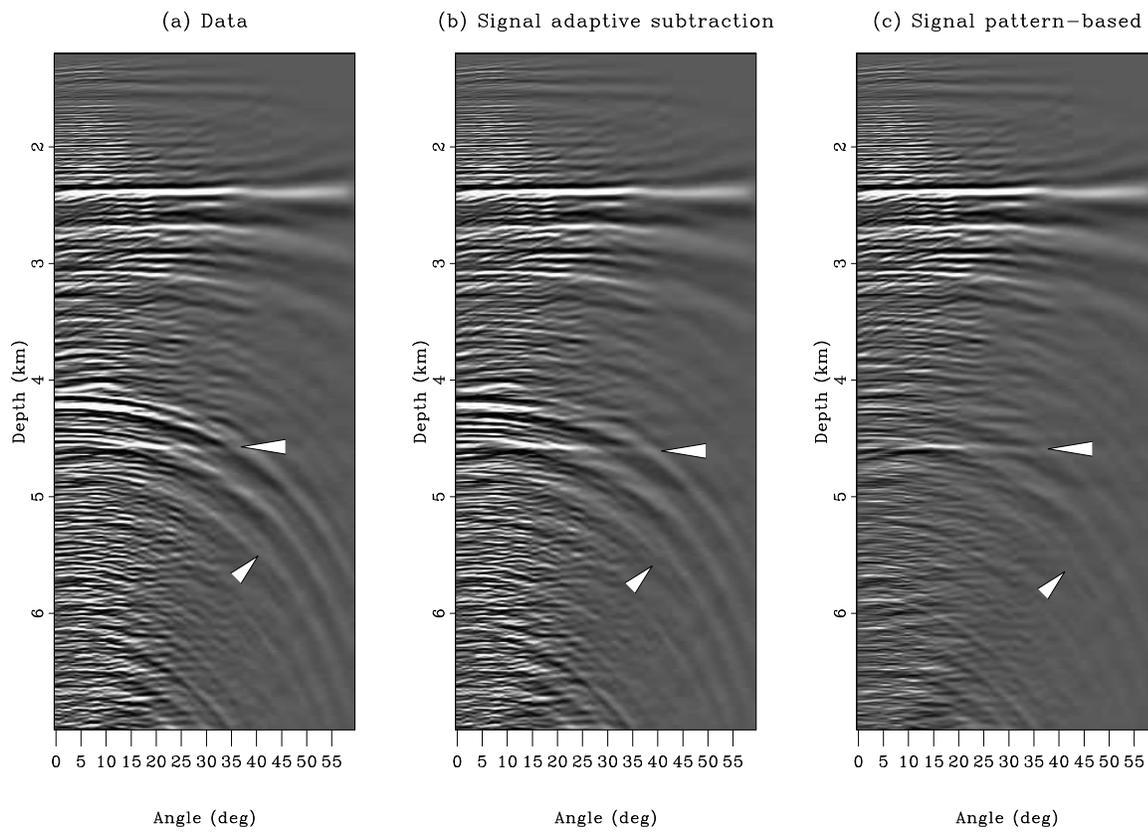


Figure 7.19: ADCIG at $X=658.975$ for (a) the data, (b) the estimated primaries with adaptive subtraction, and (c) the estimated primaries with the pattern-based approach. The arrow at 4.6 s. shows the base of salt reflection. The second arrow shows multiples at high angles with strong curvatures. The pattern-based approach yields a cleaner gather. `mul3d-ang1` [CR,M]

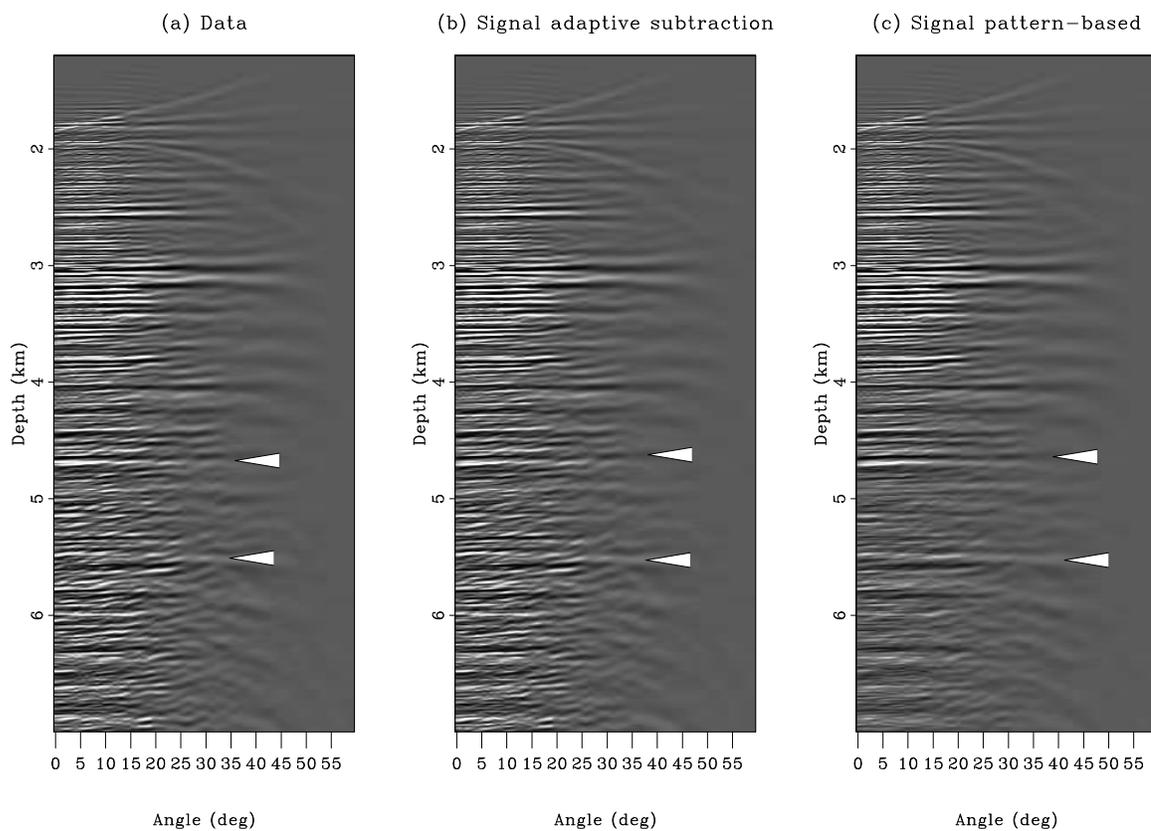


Figure 7.20: ADCIG at $X=674.325$ for (a) the data, (b) the estimated primaries with adaptive subtraction, and (c) the estimated primaries with the pattern-based approach. The two arrows show primaries that are more continuous after pattern-based subtraction of the multiples.

`mul3d-ang2` [CR,M]

prediction of multiples. Making SRME work with 3-D data can be achieved by interpolating/extrapolating the missing data and designing subtraction techniques that are less sensitive to modeling errors due to the acquisition geometry. However, most of today's efforts are focused on the prediction step relying on the adaptive subtraction to optimally estimate the primaries. This Chapter presented a multiple removal strategy on one streamer from one sail line of a 3-D survey in the Gulf of Mexico that focuses on the subtraction step only.

Due to the acquisition geometry, the shots and the near offset traces are first interpolated in 2-D with a radon-based technique with and without sparseness constraints. Interpolation is a vital element of the whole multiple attenuation procedure because it affects directly the quality of the multiple prediction. For this reason, interpolation with the sparseness constraint gives the best results by recovering most of the steep dips present in the data and by giving the fewest reconstruction artifacts.

Then, from the interpolated shots, a multiple model is estimated. Because some dips were not well reconstructed during the interpolation process, some aliasing artifacts are present in the multiple model. Again, interpolation with the sparseness constraint yields the best multiple model. However, important discrepancies exist between the actual and modeled multiples. For instance, some diffracted multiples are either poorly or not recovered by the 2-D prediction. This result should be expected wherever the geology is complex and the acquisition geometry is irregular. This observation makes a strong case for 3-D prediction.

Finally, the estimated multiple model from the 2-D prediction is subtracted from the data using adaptive subtraction and the pattern-based approach. Because this model has some flaws, the adaptive subtraction does not remove the multiples as well as the pattern-based method. In particular, the migrated images of the estimated primaries exemplify the weaknesses of adaptive subtraction where the prediction is the least accurate, e.g., below the salt.

Therefore with 3-D data, because it remains very difficult to obtain an exact multiple model from the recorded data, both the prediction and the subtraction should be looked at and improved to obtain better attenuation results. On one hand, deriving interpolation strategies in order to recover the missing traces is vital to obtain accurate multiple models. The interpolation schemes must be carefully tuned to leave as few artifacts as possible. The radon-based

reconstruction with sparseness constraints proved to be quite effective. On the other hand, deriving intelligent and flexible subtraction techniques is also fundamental. Here, the pattern-based approach offers an alternative to the more conventional adaptive subtraction technique by being robust to modeling inadequacies.

ACKNOWLEDGMENTS

I would like to thank CGG Houston for providing the 3-D dataset and the velocity model, and Paul Sava for providing the migration tools.

Chapter 8

Conclusions

Seismic noise attenuation is an important part of the processing workflow before a migrated image can be obtained and interpreted. This thesis tackles the noise problem with inversion to preserve any signal present in the data. In particular, spiky and coherent noise are attenuated by looking at the statistics (i.e., PDF and pattern) of the undesirable events.

For spiky events, Chapter 2 introduces the Huber norm. The Huber norm behaves like the ℓ^1 norm for large residuals and like the ℓ^2 norm for small residuals. Because it is continuous and differentiable everywhere, the Huber norm is minimized with a quasi-Newton method called L-BFGS. This technique maintains most of the convergence properties of the standard BFGS method for non-linear problems while keeping the memory requirements very low. This last point is particularly important for seismic processing where the volume of data can be quite large.

In this thesis, the combination Huber norm/L-BFGS solver proves being very versatile. For instance, in Chapter 2, velocity scans are derived from data with bad traces and noise bursts. In Chapter 5, small geological features at the bottom of the Sea of Galilee are unraveled from extremely noisy data. In Chapter 3, multiples and primaries are better separated by estimating matching filters with the Huber norm. This list of applications is not exhaustive. It is my belief that many more geophysical problems could benefit from using the Huber norm with the L-BFGS solver.

For coherent noise attenuation, I introduce in Chapter 4 two strategies based on the observation that coherent noise generally results from an approximate modeling of the seismic data. These techniques achieve an important goal of least-squares inversion: to have IID residuals. One strategy approximates the inverse data covariance operator with multidimensional prediction-error filters (PEFs) by weighting the data residual. The other strategy incorporates a coherent noise modeling term inside the inversion. Compared to the weighting approach, the modeling technique yields the smallest residual with the best convergence properties. One advantage of the weighting approach with PEFs, however, is its ability to separate non-stationary noise and signal according to their pattern. This pattern-based approach is very effective at separating primaries and multiples as illustrated in Chapters 6 and 7. In particular, I demonstrate that the pattern-based approach with 3-D PEFs is more robust to modeling uncertainty than adaptive subtraction, an industry standard for this task.

Therefore, the main contributions of this thesis are four fold:

1. It presents a particular implementation of a robust solver for the inversion of data contaminated with spiky noise. This implementation is based on the Huber norm and the L-BFGS method (Chapters 2 and 5).
2. It develops processing techniques for coherent noise attenuation based on the need to have IID residuals (Chapters 4 and 5).
3. It establishes multidimensional prediction-error filters with helical boundary conditions as an extremely flexible tool for noise filtering and/or modeling (Chapters 4, 6, and 7).
4. It highlights the limitations of the standard adaptive subtraction technique for the multiple attenuation problem by introducing two viable alternatives: (1) matching filters estimation with the Huber norm (Chapter 3), and (2) multiple subtraction with a pattern-based approach (Chapters 6 and 7).

One important problem omitted in this thesis that directly links to my work, however, is how to build stable inverse non-stationary PEFs. In practice, patching solves it but in a very crude manner. More elegant solutions to that problem should be developed. Valcarce et al. (2000)

give conditions for building stable inverse filters, but their approach seems to be difficult to apply in practice. Beylkin (1995) proposes transforming infinite impulse response filters (IIR) into finite response filters (FIR). This approach could impact our way of doing non-stationary deconvolution. If stable non-stationary PEFs can be computed, then the modeling approach, which has very good signal/noise separation properties, could be also used for non-stationary coherent noise removal. There, inverse non-stationary PEFs would play the role of modeling operators, as illustrated in Chapter 4 for the stationary case.

Appendix A

Algorithm for minimizing the Huber function

I present a method for solving nonlinear problems that I later use to minimize the Huber function.

A QUASI-NEWTON METHOD FOR SOLVING NONLINEAR PROBLEMS

The method I present in this paper is suitable for smooth functions where local minima exist. It is not a method for global optimization where the global minimum is sought. We define \mathbf{m}^* a local minimizer for $f(\mathbf{m})$ and we assume that $f(\mathbf{m})$ and \mathbf{m}^* satisfy the “standard requirements”:

1. f is twice differentiable,
2. $\nabla f(\mathbf{m}^*) = 0$,
3. $\nabla^2 f(\mathbf{m}^*)$ is positive definite , i.e., $\mathbf{m}'\nabla^2 f(\mathbf{m}^*)\mathbf{m} > 0$ for all $\mathbf{m} \in \Re^N$ ($'$ denotes the adjoint).

where N is the dimension of the model vector \mathbf{m} and \Re^N the real space for the model vector \mathbf{m} . Any vector \mathbf{m}^* that satisfies the standard requirements is a local minimizer of $f(\mathbf{m})$.

Newton's method is an iterative process where the solution to the problem is updated as follows:

$$\mathbf{m}_{k+1} = \mathbf{m}_k - \lambda_k \mathbf{H}_k^{-1} \nabla f(\mathbf{m}_k), \quad (\text{A.1})$$

where \mathbf{m}_{k+1} is the updated solution at iteration $k + 1$, λ_k the step length computed by a line-search that ensures a sufficient decrease of $f(\mathbf{m})$ and $\mathbf{H}_k = \nabla^2 f(\mathbf{m}_k)$ the Hessian (or second derivative). In many circumstances the inverse of the Hessian can't be computed directly. It happens for example when the matrix \mathbf{H} is too big or when operators are used rather than matrices. Fortunately we might be able to compute an approximation of the Hessian of $f(\mathbf{m})$. This strategy gives birth to quasi-Newton methods where the way in which the Hessian is computed determines the method (Kelley, 1999).

A possible update of the Hessian is given by the BFGS technique (Broyden, 1969; Fletcher, 1970; Goldfarb, 1970; Shanno, 1970). The BFGS update is given by

$$\mathbf{H}_{k+1} = \mathbf{H}_k + \frac{\mathbf{y}\mathbf{y}'}{\mathbf{y}'\mathbf{s}} - \frac{(\mathbf{H}_k\mathbf{s})(\mathbf{H}_k\mathbf{s})'}{\mathbf{s}'\mathbf{H}_k\mathbf{s}}, \quad (\text{A.2})$$

where $\mathbf{s} = \mathbf{m}_{k+1} - \mathbf{m}_k$ and $\mathbf{y} = \nabla f(\mathbf{m}_{k+1}) - \nabla f(\mathbf{m}_k)$. In practice, however, we rather write the previous equation in terms of the inverse matrices. We have then

$$\mathbf{H}_{k+1}^{-1} = \left(\mathbf{I} - \frac{\mathbf{s}\mathbf{y}'}{\mathbf{y}'\mathbf{s}} \right) \mathbf{H}_k^{-1} \left(\mathbf{I} - \frac{\mathbf{y}\mathbf{s}'}{\mathbf{s}'\mathbf{H}_k\mathbf{s}} \right) + \frac{\mathbf{s}\mathbf{s}'}{\mathbf{y}'\mathbf{s}}. \quad (\text{A.3})$$

In addition, we use the history of the iterations to compute the new Hessian rather than a full storage of the matrix \mathbf{H}_k^{-1} . This requires that a gradient step vector \mathbf{y} and a solution step vector \mathbf{s} are kept in memory after each iteration. Consequently this method might not be affordable with large data and model space. In the next section a modified version of the BFGS method that limits the storage needed to compute the update of the Hessian is proposed.

THE LIMITED MEMORY BFGS METHOD

Nocedal (1980) derives a technique that partially solves the storage problem caused by the BFGS update. Instead of keeping all the \mathbf{s} and \mathbf{y} from the past iterations, we update the Hessian using the information from the l previous iterations, where l is given by the end-user. This implies that when the number of iterations is smaller than l , we have the usual BFGS update, and when it is larger than l , we have a limited memory BFGS (L-BFGS) update.

I give the updating formulas of the Hessian as presented by Nocedal (1980). First, we define

$$\rho_i = 1/\mathbf{y}'_i \mathbf{s}_i, \mathbf{v}_i = (I - \rho_i \mathbf{y}_i \mathbf{s}'_i) \text{ and } \mathbf{H}^{-1} = \mathbf{B}.$$

As described above, when k , the iteration number, obeys $k + 1 \leq l$, where l is the storage limit, we have the BFGS update

$$\begin{aligned} \mathbf{B}_{k+1} &= \mathbf{v}'_k \mathbf{v}'_{k-1} \cdots \mathbf{v}'_0 \mathbf{B}_0 \mathbf{v}_0 \cdots \mathbf{v}_{k-1} \mathbf{v}_k \\ &\quad + \mathbf{v}'_k \cdots \mathbf{v}'_1 \rho_0 \mathbf{s}_0 \mathbf{s}'_0 \mathbf{v}_1 \cdots \mathbf{v}_k \\ &\quad \vdots \\ &\quad + \mathbf{v}'_k \rho_{k-1} \mathbf{s}_{k-1} \mathbf{s}'_{k-1} \mathbf{v}_k \\ &\quad + \rho_k \mathbf{s}_k \mathbf{s}'_k. \end{aligned} \tag{A.4}$$

For $k + 1 > l$ we have the limited memory update

$$\begin{aligned} \mathbf{B}_{k+1} &= \mathbf{v}'_k \mathbf{v}'_{k-1} \cdots \mathbf{v}'_{k-l+1} \mathbf{B}_0 \mathbf{v}_{k-l+1} \cdots \mathbf{v}_{k-1} \mathbf{v}_k \\ &\quad + \mathbf{v}'_k \cdots \mathbf{v}'_{k-l+2} \rho_{k-l+1} \mathbf{s}_{k-l+1} \mathbf{s}'_{k-l+1} \mathbf{v}_{k-l+2} \cdots \mathbf{v}_k \\ &\quad \vdots \\ &\quad + \mathbf{v}'_k \rho_{k-1} \mathbf{s}_{k-1} \mathbf{s}'_{k-1} \mathbf{v}_k \\ &\quad + \rho_k \mathbf{s}_k \mathbf{s}'_k. \end{aligned} \tag{A.5}$$

These equations show how the update of the Hessian is calculated.

Usually the L-BFGS method is implemented with a line search for the step length λ_k to ensure a sufficient decrease of the misfit function. Convergence properties of the L-BFGS method are guaranteed if λ_k in equation (A.2) satisfies the *Wolfe conditions* (Kelley, 1999):

$$f(\mathbf{x}_k + \lambda_k \mathbf{d}_k) \leq f(\mathbf{x}_k) + \mu \lambda_k \nabla f(\mathbf{x}_k)' \mathbf{d}_k, \quad (\text{A.6})$$

$$|\nabla f(\mathbf{x}_k + \lambda_k \mathbf{d}_k)' \mathbf{d}_k| \geq \nu |\nabla f(\mathbf{x}_k)' \mathbf{d}_k|. \quad (\text{A.7})$$

ν and μ are constants to be chosen a priori and $\mathbf{d}_k = -\mathbf{B}_k \nabla f(\mathbf{m}_k)$. For ν and μ we set $\nu = 0.9$ and $\mu = 10^{-4}$ as proposed by Liu and Nocedal (1989). Equation (A.6) is a sufficient decrease condition that all line search algorithms must satisfy. Equation (A.7) is a curvature condition. The line search algorithm has to be carefully designed since it absorbs most of the computing time. I programmed a line search based on the More and Thuente (1994) method. Because the line search is time consuming, the step length $\lambda_k = 1$ is always tested first. This procedure saves a lot of computing time and is also recommended by Liu and Nocedal (1989). I now give the algorithm used to minimize any objective function involving nonlinear problems.

AN EFFICIENT ALGORITHM FOR SOLVING NONLINEAR PROBLEMS

The solver works as follows:

1. Choose \mathbf{m}_0 , l , \mathbf{B}_0 . Set $k = 0$.

2. Compute

$$\mathbf{d}_k = -\mathbf{B}_k \nabla f(\mathbf{m}_k), \quad (\text{A.8})$$

$$\mathbf{m}_{k+1} = \mathbf{m}_k + \lambda_k \mathbf{d}_k, \quad (\text{A.9})$$

where λ_k meets the Wolfe conditions.

3. Let $\hat{l} = \min\{k, l - 1\}$. Update \mathbf{B}_0 $\hat{l} + 1$ times using the pairs $\{\mathbf{y}_i, \mathbf{s}_i\}_{j=k-\hat{l}}^k$, i.e., let

$$\begin{aligned} \mathbf{B}_{k+1} &= \mathbf{v}'_k \mathbf{v}'_{k-1} \cdots \mathbf{v}'_{k-\hat{l}} \mathbf{B}_0 \mathbf{v}_{k-\hat{l}} \cdots \mathbf{v}_{k-1} \mathbf{v}_k \\ &\quad + \mathbf{v}'_k \cdots \mathbf{v}'_{k-\hat{l}+1} \rho_{k-\hat{l}} \mathbf{s}_{k-\hat{l}} \mathbf{s}'_{k-\hat{l}} \mathbf{v}_{k-\hat{l}+1} \cdots \mathbf{v}_k \\ &\quad \vdots \end{aligned} \tag{A.10}$$

$$\begin{aligned} &\quad + \mathbf{v}'_k \rho_{k-1} \mathbf{s}_{k-1} \mathbf{s}'_{k-1} \mathbf{v}_k \\ &\quad + \rho_k \mathbf{s}_k \mathbf{s}'_k. \end{aligned} \tag{A.11}$$

4. Set $k = k + 1$ and go to 2 if the residual power is not small enough.

The update \mathbf{B}_{k+1} is not formed explicitly; instead we compute $\mathbf{d}_k = -\mathbf{B}_k \nabla f(\mathbf{x}_k)$ with an iterative formula (Nocedal, 1980). Liu and Nocedal (1989) propose scaling the initial symmetric positive definite \mathbf{B}_0 at each iteration as follows:

$$\mathbf{B}_k^0 = \frac{\mathbf{y}'_k \mathbf{s}_k}{\|\mathbf{y}_k\|_2^2} \mathbf{B}_0. \tag{A.12}$$

This scaling greatly improves the performances of the method. Liu and Nocedal (1989) show that the storage limit for large-scale problems has little effects. A common choice for l is $l = 5$. In practice, the initial guess \mathbf{B}_0 for the Hessian is the identity matrix \mathbf{I} ; then it might be scaled as proposed in equation (A.12). The nonlinear solver as detailed in the previous algorithm converges to a local minimizer \mathbf{m}^* of $f(\mathbf{m})$.

MINIMIZING THE HUBER FUNCTION

So far I have introduced a general method for solving nonlinear problems. In this section I show how this algorithm can be used when the Huber function is utilized for measuring the data misfit. In fact, we only need to derive the gradient of the objective function in equation (2.2). The gradient can be written in the following compact form (Li and Swetist, 1998):

$$\nabla f(\mathbf{m}) = \mathbf{L}'(\mathbf{L}\mathbf{m} - \mathbf{d})_{-\alpha}^\alpha, \tag{A.13}$$

where $\mathbf{z}_{-\alpha}^\alpha$ is a vector with i th component

$$z_i \leftarrow \max\{-1, \min\{1, z_i/\alpha\}\}. \quad (\text{A.14})$$

A last difficulty arises in the choice of the threshold α in equation (2.1). This value remains constant during the iterations. α is also the only parameter to choose a priori for different problems. I have not derived any analytical expression for α , but based on previous works with IRLS methods (Darche, 1989), it seems that

$$\alpha = \frac{\max|\mathbf{d}|}{100} \quad (\text{A.15})$$

is a good practical choice. Another possible solution is to set α at the 98-th percentile of the data (Claerbout, 2000, Personal communication).

Appendix B

Least-squares solution of the modeling approach

Solving for the model \mathbf{m} in the modeling approach requires the inversion of a 2×2 block matrix. I derive the basic steps of this procedure in the following section. Then I derive the least-squares inverse of \mathbf{m} in equation (4.8). Finally, I give a geometric interpretation of the least-squares inverse.

INVERSION OF A 2×2 BLOCK MATRIX

Let us define the 2×2 block matrix \mathbf{M} as follows:

$$\mathbf{M} = \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix}, \quad (\text{B.1})$$

where \mathbf{A} , \mathbf{B} , \mathbf{C} , and \mathbf{D} are matrices. First, we consider the matrix equation

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix} \begin{pmatrix} \mathbf{E} \\ \mathbf{F} \end{pmatrix} = \begin{pmatrix} \mathbf{G} \\ \mathbf{H} \end{pmatrix}. \quad (\text{B.2})$$

If we multiply the top row by $-\mathbf{CA}^{-1}$ and add it to the bottom, we have

$$(\mathbf{D} - \mathbf{CA}^{-1}\mathbf{B})\mathbf{F} = \mathbf{H} - \mathbf{CA}^{-1}\mathbf{G}. \quad (\text{B.3})$$

Then we can easily find \mathbf{F} and \mathbf{E} . The quantity $(\mathbf{D} - \mathbf{CA}^{-1}\mathbf{B})$ is called the *Schur complement* of \mathbf{A} and, denoted as \mathbf{S}_A , appears often in linear algebra (Demmel, 1997). The derivation of \mathbf{F} and \mathbf{E} can be written in a matrix form

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{CA}^{-1} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_A \end{pmatrix} \begin{pmatrix} \mathbf{I} & \mathbf{A}^{-1}\mathbf{B} \\ \mathbf{0} & \mathbf{I} \end{pmatrix}, \quad (\text{B.4})$$

which resembles an *LDU* decomposition of \mathbf{M} . Alternatively, we have the *UDL* decomposition

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix} = \begin{pmatrix} \mathbf{I} & \mathbf{BD}^{-1} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{S}_D & \mathbf{0} \\ \mathbf{0} & \mathbf{D} \end{pmatrix} \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{D}^{-1}\mathbf{C} & \mathbf{I} \end{pmatrix}, \quad (\text{B.5})$$

where $\mathbf{S}_D = \mathbf{A} - \mathbf{BD}^{-1}\mathbf{C}$ is the Schur complement of \mathbf{D} . The inversion formulas are then easy to derive as follows:

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{I} & -\mathbf{A}^{-1}\mathbf{B} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{A}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_A^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{CA}^{-1} & \mathbf{I} \end{pmatrix} \quad (\text{B.6})$$

and

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{D}^{-1}\mathbf{C} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{S}_D^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{I} & -\mathbf{BD}^{-1} \\ \mathbf{0} & \mathbf{I} \end{pmatrix}. \quad (\text{B.7})$$

The decomposition of the matrix \mathbf{M} offers opportunities for fast inversion algorithms. The final expressions for \mathbf{M} are

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix}^{-1} = \begin{pmatrix} [\mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{BS}_A^{-1}\mathbf{CA}^{-1}] & -\mathbf{A}^{-1}\mathbf{BS}_A^{-1} \\ -\mathbf{S}_A^{-1}\mathbf{CA}^{-1} & \mathbf{S}_A^{-1} \end{pmatrix} \quad (\text{B.8})$$

and

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{S}_D^{-1} & -\mathbf{S}_D^{-1}\mathbf{B}\mathbf{D}^{-1} \\ -\mathbf{D}^{-1}\mathbf{C}\mathbf{S}_D^{-1} & [\mathbf{D}^{-1} + \mathbf{D}^{-1}\mathbf{C}\mathbf{S}_D^{-1}\mathbf{B}\mathbf{D}^{-1}] \end{pmatrix}. \quad (\text{B.9})$$

Equations (B.8) and (B.9) yield the matrix inversion lemma

$$(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} = \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{B}(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{C}\mathbf{A}^{-1}. \quad (\text{B.10})$$

INVERSION OF THE HESSIAN

Using the results above, the least-squares estimate of \mathbf{m} in equation (4.8) is derived. Assuming that $\mathbf{C}_d = \sigma_d^2 \mathbf{I}$, the fitting goal is

$$\mathbf{0} \approx \mathbf{r}_d = \mathbf{L}\mathbf{m} - \mathbf{d}, \quad (\text{B.11})$$

with $\mathbf{L} = (\mathbf{L}_s \ \mathbf{L}_n)$ and $\mathbf{m}' = (\mathbf{m}_s \ \mathbf{m}_n)$. The normal equations are given by

$$\begin{pmatrix} \mathbf{L}'_s\mathbf{L}_s & \mathbf{L}'_s\mathbf{L}_n \\ \mathbf{L}'_n\mathbf{L}_s & \mathbf{L}'_n\mathbf{L}_n \end{pmatrix} \begin{pmatrix} \mathbf{m}_s \\ \mathbf{m}_n \end{pmatrix} = \begin{pmatrix} \mathbf{L}'_s\mathbf{d} \\ \mathbf{L}'_n\mathbf{d} \end{pmatrix}, \quad (\text{B.12})$$

where \mathbf{m}_s and \mathbf{m}_n are the unknowns. The least-square estimate $\hat{\mathbf{m}}_s$ of \mathbf{m}_s can be derived from the bottom row of equation (B.8). The least-square estimate $\hat{\mathbf{m}}_n$ of \mathbf{m}_n can be derived from the top row of equation (B.9). We have, then,

$$\hat{\mathbf{m}}_s = (\mathbf{L}'_s\mathbf{L}_s - \mathbf{L}'_s\mathbf{L}_n(\mathbf{L}'_n\mathbf{L}_n)^{-1}\mathbf{L}'_n\mathbf{L}_s)^{-1}\mathbf{L}'_s\mathbf{d} - (\mathbf{L}'_s\mathbf{L}_s - \mathbf{L}'_s\mathbf{L}_n(\mathbf{L}'_n\mathbf{L}_n)^{-1}\mathbf{L}'_n\mathbf{L}_s)^{-1}\mathbf{L}'_s\mathbf{L}_n(\mathbf{L}'_n\mathbf{L}_n)^{-1}\mathbf{L}'_n\mathbf{d}, \quad (\text{B.13})$$

$$\hat{\mathbf{m}}_n = (\mathbf{L}'_n\mathbf{L}_n - \mathbf{L}'_n\mathbf{L}_s(\mathbf{L}'_s\mathbf{L}_s)^{-1}\mathbf{L}'_s\mathbf{L}_n)^{-1}\mathbf{L}'_n\mathbf{d} - (\mathbf{L}'_n\mathbf{L}_n - \mathbf{L}'_n\mathbf{L}_s(\mathbf{L}'_s\mathbf{L}_s)^{-1}\mathbf{L}'_s\mathbf{L}_n)^{-1}\mathbf{L}'_n\mathbf{L}_s(\mathbf{L}'_s\mathbf{L}_s)^{-1}\mathbf{L}'_s\mathbf{d}, \quad (\text{B.14})$$

which can be simplified as follows:

$$\hat{\mathbf{m}}_s = (\mathbf{L}'_s(\mathbf{I} - \mathbf{L}_n(\mathbf{L}'_n\mathbf{L}_n)^{-1}\mathbf{L}'_n)\mathbf{L}_s)^{-1}\mathbf{L}'_s(\mathbf{I} - \mathbf{L}_n(\mathbf{L}'_n\mathbf{L}_n)^{-1}\mathbf{L}'_n)\mathbf{d}, \quad (\text{B.15})$$

$$\hat{\mathbf{m}}_n = (\mathbf{L}'_n(\mathbf{I} - \mathbf{L}_s(\mathbf{L}'_s\mathbf{L}_s)^{-1}\mathbf{L}'_s)\mathbf{L}_n)^{-1}\mathbf{L}'_n(\mathbf{I} - \mathbf{L}_s(\mathbf{L}'_s\mathbf{L}_s)^{-1}\mathbf{L}'_s)\mathbf{d}. \quad (\text{B.16})$$

$\mathbf{L}_n(\mathbf{L}'_n\mathbf{L}_n)^{-1}\mathbf{L}'_n$ is the coherent noise resolution matrix, whereas $\mathbf{L}_s(\mathbf{L}'_s\mathbf{L}_s)^{-1}\mathbf{L}'_s$ is the signal resolution matrix (Tarantola, 1987). Denoting $\overline{\mathbf{R}}_s = \mathbf{I} - \mathbf{L}_s(\mathbf{L}'_s\mathbf{L}_s)^{-1}\mathbf{L}'_s$ and $\overline{\mathbf{R}}_n = \mathbf{I} - \mathbf{L}_n(\mathbf{L}'_n\mathbf{L}_n)^{-1}\mathbf{L}'_n$ yields the following simplified expression for $\hat{\mathbf{m}}_s$ and $\hat{\mathbf{m}}_n$:

$$\begin{pmatrix} \hat{\mathbf{m}}_s \\ \hat{\mathbf{m}}_n \end{pmatrix} = \begin{pmatrix} (\mathbf{L}'_s\overline{\mathbf{R}}_n\mathbf{L}_s)^{-1}\mathbf{L}'_s\overline{\mathbf{R}}_n \\ (\mathbf{L}'_n\overline{\mathbf{R}}_s\mathbf{L}_n)^{-1}\mathbf{L}'_n\overline{\mathbf{R}}_s \end{pmatrix} \mathbf{d}. \quad (\text{B.17})$$

By property of the resolution operators, $\overline{\mathbf{R}}_n$ and $\overline{\mathbf{R}}_s$ perform noise and signal filtering, i.e.,

$$\begin{aligned} \overline{\mathbf{R}}_n\mathbf{d} &= \overline{\mathbf{R}}_n\mathbf{s}, \\ \overline{\mathbf{R}}_s\mathbf{d} &= \overline{\mathbf{R}}_s\mathbf{n}, \end{aligned} \quad (\text{B.18})$$

if the noise and signal are well predicted by the noise and signal modeling operators. Nemeth (1996) demonstrates that the inverse of the Hessian in equation (B.12) is well conditioned if the noise and signal operators are orthogonal, meaning that they predict distinct parts of the model space without overlapping. If overlapping occurs, a model regularization term can improve the signal/noise separation.

GEOMETRIC INTERPRETATION OF THE NOISE AND SIGNAL FILTERS

In this section, a geometric interpretation of the noise and data filters which appear in equation (B.17) is given. The properties of the resolution operators are well known (Tarantola, 1987). The goal here is to extend these properties to the particular case of the subtraction scheme proposed in equation (B.17). But first, it is useful to give some definitions.

Definitions

I give a set of definitions that will help us to better understand the properties of the noise and signal filters in equation (B.17).

Definition 1: an operator \mathbf{P} is a projector if

$$\mathbf{P}\mathbf{P} = \mathbf{P}. \quad (\text{B.19})$$

Definition 2: two operators \mathbf{P} and \mathbf{Q} are complementary operators if

$$\mathbf{P} + \mathbf{Q} = \mathbf{I}. \quad (\text{B.20})$$

Definition 3: two operators \mathbf{P} and \mathbf{Q} are mutually orthogonal if

$$\mathbf{P}\mathbf{Q} = \mathbf{Q}\mathbf{P} = \mathbf{0}. \quad (\text{B.21})$$

Definition 4: two vectors \mathbf{u} and \mathbf{v} are orthogonal if

$$\mathbf{u}'\mathbf{v} = \mathbf{v}'\mathbf{u} = 0. \quad (\text{B.22})$$

General properties of the noise and signal filters

Following the preceding definitions, we can define the noise and signal filters more precisely. But first, recall that

$$\begin{pmatrix} \hat{\mathbf{m}}_n \\ \hat{\mathbf{m}}_s \end{pmatrix} = \begin{pmatrix} (\mathbf{L}'_n \overline{\mathbf{R}}_s \mathbf{L}_n)^{-1} \mathbf{L}'_n \overline{\mathbf{R}}_s \\ (\mathbf{L}'_s \overline{\mathbf{R}}_n \mathbf{L}_s)^{-1} \mathbf{L}'_s \overline{\mathbf{R}}_n \end{pmatrix} \mathbf{d},$$

with

$$\begin{aligned} \overline{\mathbf{R}}_s &= \mathbf{I} - \mathbf{L}_s (\mathbf{L}'_s \mathbf{L}_s)^{-1} \mathbf{L}'_s, \\ \overline{\mathbf{R}}_n &= \mathbf{I} - \mathbf{L}_n (\mathbf{L}'_n \mathbf{L}_n)^{-1} \mathbf{L}'_n. \end{aligned} \quad (\text{B.23})$$

$\overline{\mathbf{R}}_s$ and $\overline{\mathbf{R}}_n$ are signal and noise filtering operators respectively. If we define

$$\begin{aligned}\overline{\mathbf{R}}_s &= \mathbf{I} - \mathbf{R}_s, \\ \overline{\mathbf{R}}_n &= \mathbf{I} - \mathbf{R}_n,\end{aligned}\tag{B.24}$$

with $\mathbf{R}_s = \mathbf{L}_s(\mathbf{L}'_s\mathbf{L}_s)^{-1}\mathbf{L}'_s$ and $\mathbf{R}_n = \mathbf{L}_n(\mathbf{L}'_n\mathbf{L}_n)^{-1}\mathbf{L}'_n$ the signal and noise resolution operators, we deduce that \mathbf{R}_s and $\overline{\mathbf{R}}_s$, \mathbf{R}_n and $\overline{\mathbf{R}}_n$ are complementary operators (definition 2).

It can be shown that $\overline{\mathbf{R}}_s$, $\overline{\mathbf{R}}_n$, \mathbf{R}_s and \mathbf{R}_n are projectors. Indeed, for \mathbf{R}_s and $\overline{\mathbf{R}}_s$, we have

$$\begin{aligned}\mathbf{R}_s\mathbf{R}_s &= \mathbf{L}_s(\mathbf{L}'_s\mathbf{L}_s)^{-1}\mathbf{L}'_s\mathbf{L}_s(\mathbf{L}'_s\mathbf{L}_s)^{-1}\mathbf{L}'_s, \\ &= \mathbf{L}_s(\mathbf{L}'_s\mathbf{L}_s)^{-1}\mathbf{L}'_s, \\ \mathbf{R}_s\mathbf{R}_s &= \mathbf{R}_s,\end{aligned}\tag{B.25}$$

and

$$\begin{aligned}\overline{\mathbf{R}}_s\overline{\mathbf{R}}_s &= (\mathbf{I} - \mathbf{R}_s)(\mathbf{I} - \mathbf{R}_s), \\ &= \mathbf{I} - 2\mathbf{R}_s + \mathbf{R}_s, \\ \overline{\mathbf{R}}_s\overline{\mathbf{R}}_s &= \overline{\mathbf{R}}_s.\end{aligned}\tag{B.26}$$

Thus, \mathbf{R}_s and $\overline{\mathbf{R}}_s$ are projectors as defined in definition 1. The same proofs work for \mathbf{R}_n and $\overline{\mathbf{R}}_n$.

We can prove also that $\overline{\mathbf{R}}_s$ and \mathbf{R}_s , $\overline{\mathbf{R}}_n$ and \mathbf{R}_n are mutually orthogonal. For \mathbf{R}_s and $\overline{\mathbf{R}}_s$, we have

$$\begin{aligned}\overline{\mathbf{R}}_s\mathbf{R}_s &= (\mathbf{I} - \mathbf{R}_s)\mathbf{R}_s, \\ &= (\mathbf{R}_s - \mathbf{R}_s), \\ \overline{\mathbf{R}}_s\mathbf{R}_s &= \mathbf{0}.\end{aligned}\tag{B.27}$$

Hence, $\overline{\mathbf{R}}_s$ and \mathbf{R}_s , $\overline{\mathbf{R}}_n$ and \mathbf{R}_n are complementary, mutually orthogonal projectors.

Geometric interpretation

The operators \mathbf{R}_n and \mathbf{R}_s are the noise and signal resolution operators. They describe how well the predictions match the noise and signal (Menke, 1989). In the following equations, we consider that

$$\mathbf{R}_n \mathbf{n} = \mathbf{n} \quad (\text{B.28})$$

and

$$\mathbf{R}_s \mathbf{s} = \mathbf{s}, \quad (\text{B.29})$$

meaning that each component of the data has been predicted. These equalities will help us to build a comprehensive geometric interpretation for the different operators. Based on equations (B.28) and (B.29), we have for the data vector \mathbf{d} the following equalities:

$$\begin{aligned} \mathbf{R}_s \mathbf{d} &= \mathbf{R}_s \mathbf{n} + \mathbf{s} \\ \mathbf{R}_n \mathbf{d} &= \mathbf{R}_n \mathbf{s} + \mathbf{n}, \end{aligned} \quad (\text{B.30})$$

and

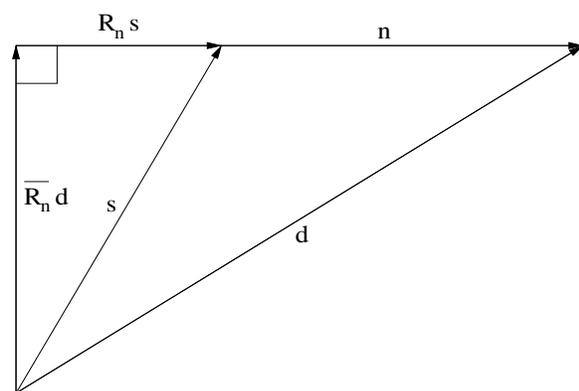
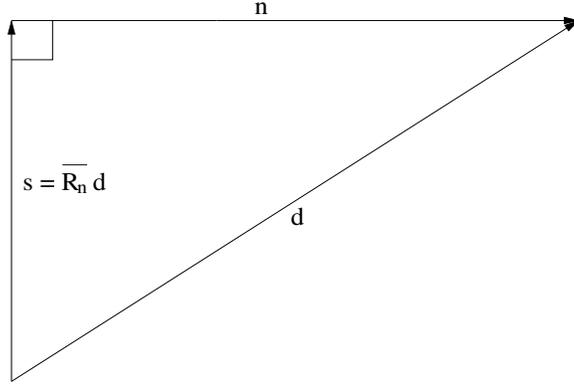


Figure B.1: A geometric interpretation of the noise filter when \mathbf{n} and \mathbf{s} are not orthogonal. append-geom11
[NR]

$$\begin{aligned} \overline{\mathbf{R}_s \mathbf{d}} &= \overline{\mathbf{R}_s \mathbf{n}} \\ \overline{\mathbf{R}_n \mathbf{d}} &= \overline{\mathbf{R}_n \mathbf{s}}. \end{aligned} \quad (\text{B.31})$$

Figure B.2: A geometric interpretation of the noise filter when \mathbf{n} and \mathbf{s} are orthogonal. append-geom21
[NR]



In the following equations, we prove that $\|\overline{\mathbf{R}}_n \mathbf{s}\|^2 + \|\mathbf{R}_n \mathbf{s} + \mathbf{n}\|^2 = \|\mathbf{d}\|^2$:

$$\begin{aligned}
 \|\overline{\mathbf{R}}_n \mathbf{s}\|^2 + \|\mathbf{R}_n \mathbf{s} + \mathbf{n}\|^2 &= \mathbf{s}' \overline{\mathbf{R}}_n' \overline{\mathbf{R}}_n \mathbf{s} + \mathbf{s}' \mathbf{R}_n' \mathbf{R}_n \mathbf{s} + \mathbf{s}' \mathbf{R}_n' \mathbf{n} + \mathbf{n}' \mathbf{R}_n \mathbf{s} + \\
 &\quad \mathbf{n}' \mathbf{n} \\
 &= \mathbf{d}' \overline{\mathbf{R}}_n \mathbf{d} + \mathbf{s}' \mathbf{R}_n \mathbf{s} + \mathbf{n}' \mathbf{R}_n \mathbf{s} + \mathbf{s}' \mathbf{R}_n \mathbf{n} + \\
 &\quad \mathbf{n}' \mathbf{R}_n \mathbf{n} \\
 &= \mathbf{d}' \overline{\mathbf{R}}_n \mathbf{d} + \mathbf{d}' \mathbf{R}_n \mathbf{d} \\
 &= \mathbf{d}' \mathbf{d} \\
 \|\overline{\mathbf{R}}_n \mathbf{s}\|^2 + \|\mathbf{R}_n \mathbf{s} + \mathbf{n}\|^2 &= \|\mathbf{d}\|^2.
 \end{aligned} \tag{B.32}$$

Similarly, we have $\|\overline{\mathbf{R}}_s \mathbf{n}\|^2 + \|\mathbf{R}_s \mathbf{n} + \mathbf{s}\|^2 = \|\mathbf{d}\|^2$. If we use equations (B.30) and (B.31), the last two equalities can be written as follows:

$$\begin{aligned}
 \|\overline{\mathbf{R}}_n \mathbf{d}\|^2 + \|\mathbf{R}_n \mathbf{d}\|^2 &= \|\mathbf{d}\|^2, \\
 \|\overline{\mathbf{R}}_s \mathbf{d}\|^2 + \|\mathbf{R}_s \mathbf{d}\|^2 &= \|\mathbf{d}\|^2.
 \end{aligned} \tag{B.33}$$

Hence, $\overline{\mathbf{R}}_n \mathbf{d}$, $\mathbf{R}_n \mathbf{d}$ and \mathbf{d} form a right triangle with hypotenuse \mathbf{d} and legs $\overline{\mathbf{R}}_n \mathbf{d}$ and $\mathbf{R}_n \mathbf{d}$, as depicted in Figure B.1; similarly, $\overline{\mathbf{R}}_s \mathbf{d}$, $\mathbf{R}_s \mathbf{d}$ and \mathbf{d} form a right triangle with hypotenuse \mathbf{d} and legs $\overline{\mathbf{R}}_s \mathbf{d}$ and $\mathbf{R}_s \mathbf{d}$. If \mathbf{n} and \mathbf{s} are orthogonal, \mathbf{s} is in the null space of \mathbf{R}_n and $\overline{\mathbf{R}}_n \mathbf{d} = \overline{\mathbf{R}}_n \mathbf{s} = \mathbf{s}$ (Figure B.2). Similarly, \mathbf{n} is in the null space of \mathbf{R}_s and $\overline{\mathbf{R}}_s \mathbf{d} = \overline{\mathbf{R}}_s \mathbf{n} = \mathbf{n}$.

Appendix C

Estimation of nonstationary PEFs

As shown by Margrave (1998), the generalization of stationary convolution to the non-stationary case can be done in two different ways called non-stationary convolution and non-stationary combination. As identified by Rickett (2001), our programs at the Stanford Exploration Project implement non-stationary combination. This section intends to explain in a matrix form the filter estimation step for non-stationary PEFs for both non-stationary convolution and combination.

DEFINITIONS

I call \mathbf{A} the convolution or combination operator with a bank of non-stationary filters. For the non-stationary convolution, the filters are in the column of \mathbf{A}_{conv} (one filter corresponds to one point in the input space) whereas for the non-stationary combination, the filters lie in the rows of \mathbf{A}_{comb} (one filter corresponds to one point in the output space). For the convolution matrix, I define $a_{i,j}$ as the i^{th} coefficient of the filter for the j^{th} data point in the input space. For the combination matrix, I define $a_{i,j}$ as the j^{th} coefficient of the filter for the i^{th} data point in the

output space. Therefore, for the non-stationary convolution we have

$$\mathbf{A}_{\text{conv}} = \begin{pmatrix} 1 & 0 & 0 & \vdots \\ a_{1,0} & 1 & 0 & \vdots \\ a_{2,0} & a_{1,1} & 1 & \vdots \\ a_{3,0} & a_{2,1} & a_{1,2} & \vdots \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix} \quad (\text{C.1})$$

and for the non-stationary combination we have

$$\mathbf{A}_{\text{comb}} = \begin{pmatrix} 1 & 0 & 0 & \vdots \\ a_{1,1} & 1 & 0 & \vdots \\ a_{2,2} & a_{2,1} & 1 & \vdots \\ a_{3,3} & a_{3,2} & a_{3,1} & \vdots \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix} \quad (\text{C.2})$$

The size of both matrices is $(n \times m)$ where n is the size of an output vector (\mathbf{x}) and m the size of an input vector (\mathbf{y}) if

$$\mathbf{A}\mathbf{y} = \mathbf{x} \quad (\text{C.3})$$

The helical boundary conditions allow to generalize this 1-D convolution to higher dimensions. We can rewrite equation (C.3) for the convolution as follows:

$$y_k = x_k + \sum_{i=1}^{\min(nf-1, k-1)} a_{i, (k-i)} x_{k-i} \quad (\text{C.4})$$

where nf is the number of filter coefficients, and for the combination

$$y_k = x_k + \sum_{i=1}^{\min(nf-1, k-1)} a_{i, k} x_{k-i}. \quad (\text{C.5})$$

In the next section, I show how the non-stationary PEFs are estimated.

FILTER ESTIMATION

When PEFs are estimated, the matrix \mathbf{A} is unknown. If \mathbf{y} is the data vector from which we want to estimate the filters, we minimize the vector \mathbf{r}_y as follows:

$$\mathbf{0} \approx \mathbf{r}_y = \mathbf{A}\mathbf{y} \quad (\text{C.6})$$

which can be rewritten

$$\mathbf{0} \approx \mathbf{r}_y = \mathbf{Y}\mathbf{a}, \quad (\text{C.7})$$

where \mathbf{Y} is the matrix representation of the non-stationary convolution or combination with the input vector \mathbf{y} . The transition between equations (C.6) and (C.7) is not simple. In particular, the shape of the matrix \mathbf{Y} is quite different if we are doing non-stationary convolution or combination. For the convolution, we have

$$\mathbf{Y}_{\text{conv}} = \left(\mathbf{Y}_{\text{conv}}^0 \mid \mathbf{Y}_{\text{conv}}^1 \mid \mathbf{Y}_{\text{conv}}^2 \mid \dots \right), \text{ where} \quad (\text{C.8})$$

$$\mathbf{Y}_{\text{conv}}^0 = \begin{pmatrix} y_0 & 0 & 0 & \vdots \\ 0 & y_0 & 0 & \vdots \\ 0 & 0 & y_0 & \vdots \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix}, \mathbf{Y}_{\text{conv}}^1 = \begin{pmatrix} 0 & 0 & 0 & \vdots \\ y_1 & 0 & 0 & \vdots \\ 0 & y_1 & 0 & \vdots \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix}, \mathbf{Y}_{\text{conv}}^2 = \begin{pmatrix} 0 & 0 & 0 & \vdots \\ 0 & 0 & 0 & \vdots \\ y_2 & 0 & 0 & \vdots \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix} \text{ etc...}$$

We see that for the convolution case, the $\mathbf{Y}_{\text{conv}}^k$ matrices are diagonal operators, translating the need for one filter to be applied to one input point. The size of the matrix \mathbf{Y}_{conv} is $(n \times (m \times nf))$ where nf is the number of filter coefficients. Now, for the combination, we have

$$\mathbf{Y}_{\text{comb}} = \left(\mathbf{Y}_{\text{comb}}^0 \mid \mathbf{Y}_{\text{comb}}^1 \mid \mathbf{Y}_{\text{comb}}^2 \mid \dots \right), \text{ where} \quad (\text{C.9})$$

$$\mathbf{Y}_{\text{comb}}^0 = \begin{pmatrix} y_0 & 0 & 0 & \vdots \\ 0 & 0 & 0 & \vdots \\ 0 & 0 & 0 & \vdots \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix}, \mathbf{Y}_{\text{comb}}^1 = \begin{pmatrix} 0 & 0 & 0 & \vdots \\ y_1 & y_0 & 0 & \vdots \\ 0 & 0 & 0 & \vdots \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix}, \mathbf{Y}_{\text{comb}}^2 = \begin{pmatrix} 0 & 0 & 0 & \vdots \\ 0 & 0 & 0 & \vdots \\ y_2 & y_1 & y_0 & \vdots \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix} \text{ etc...}$$

We see that for the combination case, the $\mathbf{Y}_{\text{combo}}^k$ matrices are row operators, translating the need for one filter to be constant for one output point. The size of \mathbf{Y}_{comb} is equal to the size of \mathbf{Y}_{conv} . For the vector \mathbf{a} in equation (C.7) we have

$$\mathbf{a} = \begin{pmatrix} \mathbf{a}_0 \\ \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \end{pmatrix} \text{ with } \mathbf{a}_k = \begin{pmatrix} 1 \\ a_{1,k} \\ a_{2,k} \\ \vdots \\ a_{nf-1,k} \end{pmatrix} \quad (\text{C.10})$$

where nf is the number of coefficients per filter. This definition of \mathbf{a} is independent of \mathbf{Y} . We might want to have one filter common to different input or output points instead of one filter per point. In that case, the matrix \mathbf{Y} is obtained by adding successive \mathbf{Y}^k matrices depending on how many points have a similar filter. Note that in the stationary case, for both the convolution and the combination case we have $\mathbf{a}_0 = \mathbf{a}_1 = \mathbf{a}_2 = \dots = \mathbf{a}$ and

$$\mathbf{Y}\mathbf{a} = (\mathbf{Y}_0 + \mathbf{Y}_1 + \mathbf{Y}_2 + \dots)\mathbf{a}. \quad (\text{C.11})$$

Therefore, for the matrix \mathbf{Y} , we have to add all the \mathbf{Y}^k matrices together. If we take advantage of the special structure of \mathbf{Y}_k for the convolution and the combination, we obtain for the

stationary case

$$\mathbf{A}\mathbf{y} = \mathbf{Y}\mathbf{a} = \begin{pmatrix} y_0 & 0 & 0 & 0 & \vdots \\ y_1 & y_0 & 0 & 0 & \vdots \\ y_2 & y_1 & y_0 & 0 & \vdots \\ y_3 & y_2 & y_1 & y_0 & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix} \begin{pmatrix} 1 \\ a_1 \\ a_2 \\ a_3 \\ \vdots \end{pmatrix}, \quad (\text{C.12})$$

which is the matrix formulation of the stationary convolution. With the definitions given in equations (C.8), (C.9) and (C.10), the fitting goal in equation (C.7) can be rewritten

$$\mathbf{0} \approx \mathbf{r}_y = \mathbf{Y}_{\text{conv}}^0 \mathbf{a}_0 + \mathbf{Y}_{\text{conv}}^1 \mathbf{a}_1 + \mathbf{Y}_{\text{conv}}^2 \mathbf{a}_2 + \dots \quad (\text{C.13})$$

or

$$\mathbf{0} \approx \mathbf{r}_y = \mathbf{Y}_{\text{comb}}^0 \mathbf{a}_0 + \mathbf{Y}_{\text{comb}}^1 \mathbf{a}_1 + \mathbf{Y}_{\text{comb}}^2 \mathbf{a}_2 + \dots \quad (\text{C.14})$$

Each vector \mathbf{a}_k has one constrained coefficient. We can then rewrite equations (C.13) and (C.14) as follows:

$$\mathbf{0} \approx \mathbf{r}_y = \mathbf{y} + \mathbf{Y}_{\text{conv}}^0 \mathbf{M}\mathbf{a}_0 + \mathbf{Y}_{\text{conv}}^1 \mathbf{M}\mathbf{a}_1 + \mathbf{Y}_{\text{conv}}^2 \mathbf{M}\mathbf{a}_2 + \dots \quad (\text{C.15})$$

and

$$\mathbf{0} \approx \mathbf{r}_y = \mathbf{y} + \mathbf{Y}_{\text{comb}}^0 \mathbf{M}\mathbf{a}_0 + \mathbf{Y}_{\text{comb}}^1 \mathbf{M}\mathbf{a}_1 + \mathbf{Y}_{\text{comb}}^2 \mathbf{M}\mathbf{a}_2 + \dots \quad (\text{C.16})$$

with

$$\mathbf{M} = \begin{pmatrix} 0 & 0 & 0 & \vdots \\ 0 & 1 & 0 & \vdots \\ 0 & 0 & 1 & \vdots \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix} \quad (\text{C.17})$$

The definition of \mathbf{M} assumes that the first coefficient of each filter is known. Note that \mathbf{M} is equal for both convolution and combination methods. Having defined the matrix \mathbf{M} , we can now rewrite equation (C.7) as follows:

$$\mathbf{0} \approx \mathbf{r}_y = \mathbf{YKa} + \mathbf{y} \quad (\text{C.18})$$

where the square matrix \mathbf{K} is

$$\mathbf{K} = \left(\begin{array}{c|c|c|c} \mathbf{M} & \mathbf{0} & \mathbf{0} & \vdots \\ \hline \mathbf{0} & \mathbf{M} & \mathbf{0} & \vdots \\ \hline \mathbf{0} & \mathbf{0} & \mathbf{M} & \vdots \\ \hline \vdots & \vdots & \vdots & \vdots \end{array} \right). \quad (\text{C.19})$$

The next step consists of estimating the filter coefficients in a least-squares sense.

REGULARIZATION OF THE FILTER COEFFICIENTS

The number of coefficients to estimate is usually much greater than the number of data points. This makes the problem very under-determined. A solution is to introduce more equations in equation (C.18) as follows:

$$\begin{aligned} \mathbf{0} &\approx \mathbf{r}_y = \mathbf{YKa} + \mathbf{y} \\ \mathbf{0} &\approx \epsilon \mathbf{r}_a = \epsilon \mathbf{Ra}, \end{aligned} \quad (\text{C.20})$$

where ϵ is a constant to be chosen, usually by trial and error. The second term in equation (C.20) improves the conditioning of our problem and is called regularization. In the filter estimation problem, it is reasonable that \mathbf{R} penalizes strong variations between filter coefficients. Hence, \mathbf{R} is usually a gradient or a Laplacian. Crawley (2000) proposes smoothing the filter coefficients along radial directions. This proposal is valid for shot or common mid point gathers only where constant dips are roughly aligned along radial segments. For instance, if \mathbf{R} is a

gradient operator we have

$$\mathbf{R} = \left(\begin{array}{c|c|c|c} \mathbf{I} & -\mathbf{I} & \mathbf{0} & \vdots \\ \hline \mathbf{0} & \mathbf{I} & -\mathbf{I} & \vdots \\ \hline \mathbf{0} & \mathbf{0} & \mathbf{I} & \vdots \\ \hline \vdots & \vdots & \vdots & \vdots \end{array} \right) \quad (\text{C.21})$$

with \mathbf{I} the identity matrix. Equation (C.20) can be solved for \mathbf{a} in a least-squares sense. We then want to minimize the objective function

$$f(\mathbf{a}) = \|\mathbf{r}_y\|^2 + \epsilon^2 \|\mathbf{r}_a\|^2 \quad (\text{C.22})$$

which gives for \mathbf{a} the least-squares estimate

$$\hat{\mathbf{a}} = -(\mathbf{K}'\mathbf{Y}'\mathbf{Y}\mathbf{K} + \epsilon^2\mathbf{R}'\mathbf{R})^{-1}\mathbf{K}'\mathbf{Y}'\mathbf{y}. \quad (\text{C.23})$$

Because of the number of unknowns and of the sparseness of the problem, we use a conjugate-gradient method to estimate our PEFs.

Appendix D

A surface-related multiple prediction theory

This section details how multiples are generated by convolution of shot gathers. For a 1-D earth, the convolution can be done directly in the $f - k$ domain. For 2-D and 3-D earth, the convolution becomes nonstationary. In addition, one convolution of the shot gathers tends to overpredict high-order multiples.

If $p(g|s)$ represents a single frequency component of the primary reflected wavefield recorded at g after an impulsive shot at s , then the *first-order* surface-related multiple, $m(g|s)$, can be computed with a Kirchhoff-style integral over the reflection surface:

$$m(g|s) = \int_{\text{surface}} p(g|g') p(g'|s) dg'. \quad (\text{D.1})$$

Equation (D.1) is expensive to evaluate, especially for large 3-D data sets, but nevertheless widely-used for multiples modeling.

ONE-DIMENSIONAL EARTH AND IMPULSIVE SOURCE

Let's define u_0 as the primary wavefield and u_1 the surface-related, first-order multiple wavefield recorded at the surface. If the earth varies only as a function of depth, then u will not depend on both s and g , but only on the offset, $h = g - s$. In this one-dimensional case,

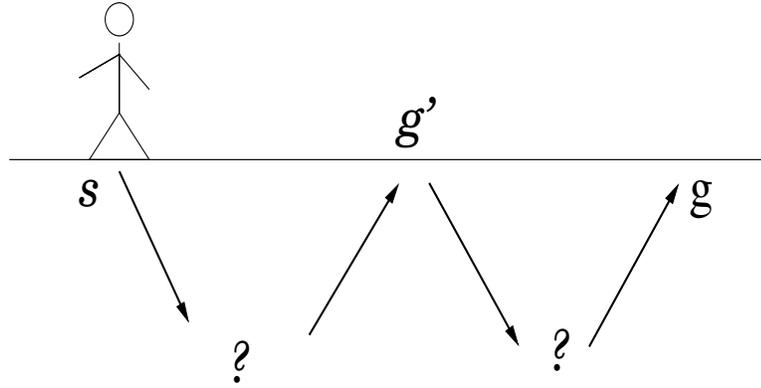


Figure D.1: The wavefield is emitted at s and recorded at g . The multiple bounces somewhere at g' . append-hombre [NR]

equation (D.1) becomes

$$u_1(g-s) = \int u_0(g-g') u_0(g'-s) dg' \quad (\text{D.2})$$

$$u_1(h) = \int u_0(h-h') u_0(h') dh', \quad (\text{D.3})$$

where $h' = g' - s$. Equation (D.3) clearly represents a convolution, so can be computed by multiplication in the Fourier domain such that

$$U_1(k_h) = U_0(k_h)^2, \quad (\text{D.4})$$

where $U_i(k_h)$ is the Fourier transform of $u_i(h)$ defined by

$$U_i(k_h) = \int u_i(h) e^{-2\pi i k_h h} dh. \quad (\text{D.5})$$

Equation (D.4) can be extended to deal with a smoothly varying earth by considering common shot-gathers (or common midpoint gathers) independently, and assuming the earth is locally one-dimensional in the vicinity of the shot, e.g., Rickett and Guitton (2000):

$$U_1(k_h, s) = U_0(k_h, s)^2. \quad (\text{D.6})$$

In practice, however, the primaries are not known and u_0 is replaced by the data with primaries and multiples. A similar approach has been used by Kelamis and Verschuur (2000) for attenuating surface-related multiples on land data for relatively flat geology.

2- AND 3-D EARTH

In the general case, modeling multiples becomes more expensive. Equation (D.3) is not valid anymore (except for smoothly varying media), and the convolution becomes nonstationary (shot gathers are different from one location to another). Hence, the wavefield is not only a function of offset, h , but also depends on another spatial coordinate such as shot location s . In 3-D, the integral in equation D.1 spans the entire acquisition plane (van Dedem, 2002), which makes the prediction very expensive. Introducing the nonstationary convolution, equation (D.1) can be written as

$$u_1(h, s) = \int u_0(h - h', s + h') u_0(h', s) dh'. \quad (\text{D.7})$$

Now, following Dragoset and Jericevic (1998) for 2-D prediction, we introduce some amplitude corrections in the previous equation:

$$\begin{aligned} u_0(h - h', s + h') &= F_{t \rightarrow \omega}[\sqrt{t}u_0(h - h', s + h', t)], \\ u_0(h', s) &= (1 - i)\sqrt{\frac{\omega}{4\pi}} F_{t \rightarrow \omega}[\sqrt{t}u_{0g}(h', s, t)]. \end{aligned} \quad (\text{D.8})$$

Replacing u_0 by the data with primaries and multiples, equation (D.7) with the amplitude correction is used throughout this thesis to model surface-related multiples in 2-D.

LIMITATIONS OF THE MULTIPLE PREDICTION USING REAL DATA

In the real life, the source is not impulsive. In addition, multiples are computed directly from the data and not from the primary wavefield. In that, a model of the multiples is obtained by convolving the recorded data with primaries and multiples as opposed to the primaries only. Hence, the relative amplitude of first order multiples with respect to higher order multiples

is not preserved. To illustrate this last point, consider the surface-related multiple modeling equation (Verschuur et al., 1992)

$$u_r = u_0 - W^{-1}u_m, \quad (\text{D.9})$$

where u_r is the recorded wavefield at the surface with primaries and surface-related multiples, W the source wavelet, and u_m the surface-related multiple wavefield given by

$$\begin{aligned} u_m &= u_0 \otimes u_0 - W^{-1}u_0 \otimes u_0 \otimes u_0 \\ &\quad + W^{-2}u_0 \otimes u_0 \otimes u_0 \otimes u_0 \dots \end{aligned} \quad (\text{D.10})$$

$$u_m = u_1 + u_2 + u_3 + \dots \quad (\text{D.11})$$

where \otimes represents the nonstationary convolution and u_i the i -th order multiples. If we use equation (D.7), replacing u_0 by u_r , we obtain for the approximated multiple field \tilde{u}_m

$$\tilde{u}_m = u_1 + 2u_2 + 3u_3 + \dots \quad (\text{D.12})$$

Comparing equation (D.11) and equation (D.12), we notice that higher order multiples in equation (D.12) are multiplied by a coefficient that is difficult to correct for. Therefore higher order multiples have the correct kinematics, but the wrong amplitudes (Wang and Levin, 1994; Hugonnet, 2002). Hence, the modeling scheme that consists in convolving the shot gathers only once explicitly overpredict high-order multiples (amplitude wise) but models them with the correct pattern. As shown by Berkhout and Verschuur (1997) and Verschuur and Berkhout (1997), this single convolution can be interpreted as a first iteration of the recursive formulation of SRME.

Bibliography

- Abma, R., 1995a, Least-squares separation of signal and noise with multidimensional filters: Ph.D. thesis, Stanford University. URL <http://sepwww.stanford.edu/public/docs/sep88/index.html>.
- Abma, R., and Claerbout, J., 1995, Lateral prediction for noise attenuation by t-x and F-X techniques: *Geophysics*, **60**, 1887–1896.
- Abma, R. L., 1995b, Enhanced random noise removal by inversion *in* 65th Ann. Internat. Mtg. Soc. of Expl. Geophys., 1340–1341.
- Alvarez, G., Biondi, B., and Guitton, A., 2004, Attenuation of diffracted multiples with an apex-shifted tangent-squared radon transform in image space: *SEP*– **115**, 139–152.
- Alvarez, G., and Larner, K., 2004, Relative performance of moveout-based multiple-suppression methods for amplitude variation with offset (avo) analysis and common mid-point (cmp) stacking: *Geophysics*, **69**, 275–285.
- Barrodale, I., and Roberts, F. D. K., 1980, Algorithm 552 : Solution of the constrained l1 linear approximation problem: *ACM Transactions on Mathematical Software*, **6**, 231–235.
- Baysal, E., Kosloff, D. D., and Sherwood, J. W. C., 1983, Reverse time migration: *Geophysics*, **48**, 1514–1524.
- Berkhout, A., and Verschuur, D., 1999, Removal of internal multiples *in* 69th Ann. Internat. Mtg. Soc. of Expl. Geophys., 1334–1337.

- Berkhout, A. J., and Verschuur, D. J., 1994, Multiple technology: Part 2, migration of multiple reflections *in* 64th Ann. Internat. Mtg. Soc. of Expl. Geophys., 1497–1500.
- , 1997, Estimation of multiple scattering by iterative inversion, part I: Theoretical considerations: *Geophysics*, **62**, 1586–1595.
- , 2004, Imaging multiple reflections: The concept: 74th Annual International Meeting, SEG, Expanded Abstracts, 1273–1276.
- Beylkin, G., 1995, On factored FIR approximation of IIR filters: *Applied and Computational Harmonic Analysis*, **2**, 293–398.
- Biersteker, J., 2001, MAGIC: Shell's surface multiple attenuation technique *in* 71st Ann. Internat. Mtg. Soc. of Expl. Geophys., 1301–1304.
- Bishop, K., Keliher, P., Paffenholz, J., Stoughton, D., Michell, S., Ergas, R., and Hadidi, M., 2001, Investigation of vendor demultiple technology for complex subsalt geology *in* 71st Ann. Internat. Mtg. Soc. of Expl. Geophys., 1273–1276.
- Britze, P., 1998, Interpreter's Corner - Seismic mapping using Trendform gridding: *The Leading Edge*, **17**, 606–608.
- Brown, M., 2001, Estimation of systematic errors in tracked datasets using least squares crossing point analysis: *SEP*— **110**, 123–132.
- , 2004, Least-squares joint imaging of multiples and primaries: Ph.D. thesis, Stanford University.
- Brown, M., and Clapp, R., 2000, T-x domain, pattern-based ground-roll removal *in* 70th Ann. Internat. Mtg. Soc. of Expl. Geophys., 2103–2106.
- Broyden, C. G., 1969, A new double-rank minimization algorithm: *AMS Notices*, **16**, 670.
- Bube, K. P., and Langan, R. T., 1997, Hybrid λ_1/λ_2 minimization with applications to tomography: *Geophysics*, **62**, 1183–1195.

- Burg, J. P., 1975, Maximum entropy spectral analysis: Ph.D. thesis, Stanford University. URL <http://sepwww.stanford.edu/oldreports/sep06/index.html>.
- Chapman, N. R., and Barrodale, I., 1983, Deconvolution of marine seismic data using the l_1 norm: *Geophys. J. Roy. Astr. Soc.*, **72**, 93–100.
- Chemingui, N., and Biondi, B., 2002, Seismic data reconstruction by inversion to common offset: *Geophysics*, **67**, 1575–1585.
- Claerbout, J., 1976, *Fundamentals of Geophysical Data Processing*: Blackwell Scientific Publications.
- , 1992, *Earth Sounding Analysis, Processing versus Inversion*: Blackwell Scientific Publications.
- , 1998, Multidimensional recursive filters via a helix: *Geophysics*, **63**, 1532–1541.
- Claerbout, J., and Black, J. L., 1997, *Basic Earth Imaging*: Class notes, <http://sepwww.stanford.edu/sep/prof/index.html>.
- Claerbout, J., and Fomel, S., 2000, Spitz makes a better assumption for the signal PEF: *SEP*—**103**, 211–219.
- , 2002, *Image Estimation by Example: Geophysical soundings image construction*: Class notes, <http://sepwww.stanford.edu/sep/prof/index.html>.
- Claerbout, J. F., and Muir, F., 1973, Robust modeling with erratic data: *Geophysics*, **38**, 826–844.
- Clapp, R., Biondi, B., and Claerbout, J. F., 2004, Incorporating geologic information into reflection tomography: *Geophysics*, **69**, 533–546.
- Clapp, R. G., 2003, Multiple realizations and data variance: Successes and failures: *SEP*—**113**, 231–246.
- Crawley, S., 2000, *Seismic trace interpolation with nonstationary prediction-error filters*: Ph.D. thesis, Stanford University. URL http://sepwww.stanford.edu/public/docs/sep104/paper_html/index.html.

- Darche, G., 1989, Iterative L1 deconvolution: SEP Annual Report, **61**, 281–301.
- van Dedem, E. J., 2002, 3D surface-related multiple prediction: Ph.D. thesis, Delft University of Technology.
- Demmel, J. W., 1997, Applied Numerical Linear Algebra: SIAM.
- Dragoset, B., 1995, Geophysical applications of adaptive-noise cancellation *in* 65th Ann. Internat. Mtg. Soc. of Expl. Geophys., 1389–1392.
- Dragoset, B., and MacKay, S., 1993, Surface multiple attenuation and subsalt imaging *in* 63rd Ann. Internat. Mtg. Soc. of Expl. Geophys., 1099–1102.
- Dragoset, B., Pattberg, D., Bell, J., and Vauthrin, R., 2000, Surface multiple attenuation: A 3-D example *in* 70th Ann. Internat. Mtg. Soc. of Expl. Geophys., 1957–1960.
- Dragoset, W. H., and Jericevic, Z., 1998, Some remarks on surface multiple attenuation: Geophysics, **63**, 772–789.
- Duijndam, A. J. W., Volker, A. W. F., and Zwartjes, P. M., 2000, Acquisition footprint reduction - least squares migration versus reconstruction *in* 62nd Mtg. Eur. Assn. Geosci. Eng., Session:P0128.
- Eklblom, H., and Madsen, K., 1989, Algorithms for non-linear huber estimation: BIT, **29**, 60–76.
- Fletcher, R., 1970, A new approach to variable metric methods: Comput. J., **13**, 317–322.
- Fomel, S., 2001, Three-dimensional seismic data regularization: Ph.D. thesis, Stanford University. URL http://sepwww.stanford.edu/public/docs/sep107/paper_html/index.html.
- , 2002, Applications of plane-wave destruction filters: Geophysics, **67**, 1946–1960.
- Fomel, S., and Claerbout, J., 1995, Searching the Sea of Galilee: The splendors and miseries of iteratively reweighted least squares: SEP— **84**, 259–270.

- , 2003, Multidimensional recursive filter preconditioning in geophysical estimation problems: *Geophysics*, **68**, 577–588.
- Foster, D. J., and Mosher, C. C., 1992, Suppression of multiple reflections using the Radon transform: *Geophysics*, **57**, 386–395.
- Gersztenkorn, A., Bednar, J. B., and Lines, L. R., 1986, Robust iterative inversion for the one-dimensional acoustic wave equation: *Geophysics*, **51**, 357–368.
- Gill, P. H., Murray, W., and Wright, M. H., 1986, *Practical Optimization*: Academic Press.
- Goldfarb, D., 1970, A family of variable metric methods derived by variational means: *Math. Comp.*, **24**, 23–26.
- Guittou, A., 2000, Coherent noise attenuation using Inverse Problems and Prediction Error Filters: *SEP*— **105**, 27–48.
- , 2002a, Coherent noise attenuation using inverse problems and prediction-error filters: *First Break*, **20**, 161–167.
- , 2002b, Shot-profile migration of multiple reflections *in* 72nd Ann. Internat. Mtg. Soc. of Expl. Geophys., 1296–1299.
- Guittou, A., Brown, M., Rickett, J., and Clapp, R., 2001, Multiple attenuation using a t - x pattern-based subtraction method *in* 71st Ann. Internat. Mtg. Soc. of Expl. Geophys., 1305–1308.
- Guittou, A., and Cambois, G., 1998, Prestack elimination of complex multiples: A Gulf of Mexico subsalt example *in* 68th Ann. Internat. Mtg. Soc. of Expl. Geophys., 1329–1332.
- Gulunay, N., 2003, Seismic trace interpolation in the Fourier transform domain: *Geophysics*, **68**, 355–369.
- Guspi, F., and Introcaso, B., 2000, A sparse spectrum technique for gridding and separating potential field anomalies: *Geophysics*, **65**, 1154–1161.

- Harlan, W. S., 1995, Regularization by Model Parameterization: Available on the world wide web, <http://www.billharlan.com/pub/papers/regularization.pdf>.
- Herrmann, F. J., and Verschuur, E., 2004, Curvelet-domain multiple elimination with sparseness constraints: 74st Annual International Meeting, SEG, Expanded Abstracts, 1333–1336.
- Herrmann, P., Mojesky, T., Magesan, M., and Hugonnet, P., 2000, De-aliased, high-resolution Radon transforms *in* 70th Ann. Internat. Mtg. Soc. of Expl. Geophys., 1953–1956.
- Hill, N., Langan, R., Nemeth, T., Zhao, M., and Bube, K., 2002, Beam methods for predictive suppression of seismic multiples in deep water *in* 72nd Ann. Internat. Mtg. Soc. of Expl. Geophys., 2118–2121.
- Huber, P. J., 1973, Robust regression: Asymptotics, conjectures, and Monte Carlo: *Ann. Statist.*, **1**, 799–821.
- Hugonnet, P., 2002, Partial surface related multiple elimination *in* 72nd Ann. Internat. Mtg. Soc. of Expl. Geophys., 2102–2105.
- Kabir, M., 2003, Weighted subtraction for diffracted multiple attenuation *in* 73rd Ann. Internat. Mtg. Soc. of Expl. Geophys., 1941–1944.
- Kabir, M. M. N., and Verschuur, D. J., 1995, Restoration of missing offsets by parabolic Radon transform: *Geophys. Prosp.*, **43**, 347–368.
- Kabir, N. M. M., and Marfurt, K. J., 1999, Toward true amplitude multiple removal: *The Leading Edge*, **18**, 66–73.
- Karpushin, A., and Brown, M., 2001, Whitening track residuals with PEFs in IRLS approach to the Sea of Galilee: *SEP*– **110**, 133–140.
- Kelamis, P. G., and Verschuur, D. J., 2000, Surface-related multiple elimination on land seismic data—Strategies via case studies: *Geophysics*, **65**, 719–734.
- Kelley, C. T., 1999, Iterative methods for optimization: *SIAM in applied mathematics*.

- Kostov, C., and Nichols, D., 1995, Moveout-discriminating adaptive subtraction of multiples *in* 65th Ann. Internat. Mtg. Soc. of Expl. Geophys., 1464–1467.
- Kuhl, H., and Sacchi, M. D., 2003, Least-squares wave-equation migration for AVP/AVA inversion: *Geophysics*, **68**, 262–273.
- Li, W., and Swetist, J. J., 1998, The linear l1 estimator and the Huber M-estimator: *SIAM J. Optim.*, **8**, 457–475.
- Liu, D. C., and Nocedal, J., 1989, On the limited memory BFGS method for large scale optimization: *Mathematical Programming*, **45**, 503–528.
- Liu, F., Sen, M. K., and Stoffa, P. L., 2000, Dip selective 2-D multiple attenuation in the plane-wave domain: *Geophysics*, **65**, 264–274.
- Lokshtanov, D., 1999, Multiple suppression by data-consistent deconvolution: *The Leading Edge*, **18**, 115–119.
- Lumley, D. E., Nichols, D., and Rekdal, T., 1995, Amplitude-preserved multiple suppression *in* 65th Ann. Internat. Mtg. Soc. of Expl. Geophys., 1460–1463.
- Manin, M., and Spitz, S., 1995, 3D attenuation of targeted multiples with a pattern recognition technique *in* 57th Mtg. Eur. Assn. of Expl. Geophys., Session:B046.
- Margrave, G. F., 1998, Theory of nonstationary linear filtering in the Fourier domain with application to time-variant filtering: *Geophysics*, **63**, 244–259.
- Matson, K. H., Paschal, D., and Weglein, A. B., 1999, A comparison of three multiple-attenuation methods applied to a hard water-bottom data set: *The Leading Edge*, **18**, 120–126.
- Menke, W., 1989, *Geophysical Data Analysis: Discrete Inverse Theory*: Academic Press.
- Mersereau, R. M., and Dudgeon, D. E., 1974, The Representation of Two-Dimensional Sequences as One-Dimensional Sequences: *IEEE Trans. Acoust., Speech, Signal Processing*, **22**, 320–325.

- More, J. J., and Thuente, J., 1994, Line search algorithms with guaranteed sufficient decrease: *ACM Transactions on Mathematical Software*, **20**, 286–307.
- Nemeth, T., 1996, Imaging and filtering by Least-Squares migration: Ph.D. thesis, The university of Utah.
- Nemeth, T., Sun, H., and Schuster, G. T., 2000, Separation of signal and coherent noise by migration filtering: *Geophysics*, **65**, 574–583.
- Nocedal, J., 1980, Updating quasi-Newton matrices with limited storage: *Mathematics of Computation*, **95**, 339–353.
- Ozdemir, A., Ozbek, A., Ferber, R., and Zerouk, K., 1999, f-xy projection filtering using helical transformation *in* 69th Ann. Internat. Mtg. Soc. of Expl. Geophys., 1231–1234.
- Paffenholz, J., McLain, B., Zasko, J., and Keliher, P., 2002, Subsalt multiple attenuation and imaging: Observations from the Sigsbee2B synthetic dataset *in* 72nd Ann. Internat. Mtg. Soc. of Expl. Geophys., 2122–2125.
- Popovici, A. M., 1996, Prestack migration by split-step DSR: *Geophysics*, **61**, 1412–1416.
- Prucha, M., Biondi, B., and Symes, W., 1999, Angle-domain common image gathers by wave-equation migration *in* 69th Ann. Internat. Mtg. Soc. of Expl. Geophys., 824–827.
- Rickett, J., 2001, Spectral factorization of wavefields and wave operators: Ph.D. thesis, Stanford University. URL http://sepwww.stanford.edu/public/docs/sep109/paper_html/index.html.
- Rickett, J., and Claerbout, J., 1999, Acoustic daylight imaging via spectral factorization: Helioseismology and reservoir monitoring: *The Leading Edge*, **18**, 957–960.
- Rickett, J., and Guitton, A., 2000, Multi-dimensional Fourier transforms in the helical coordinate system: *SEP*– **105**, 167–176.
- Rickett, J., Guitton, A., and Gratwick, D., 2001, Adaptive Multiple Subtraction with Non-Stationary Helical Shaping Filters *in* 63rd Mtg. Eur. Assn. Geosci. Eng., Session: P167.

- Robinson, E. A., and Treitel, S., 2000, Geophysical Signal Analysis: Society of Exploration Geophysicists.
- Sacchi, M., and Kuehl, H., 2001, ARMA Formulation of FX Prediction Error Filters and Projection Filters: *Journal of Seismic Exploration*, **9**, 185–197.
- Sacchi, M. D., and Ulrych, T. J., 1995, High-resolution velocity gathers and offset space reconstruction: *Geophysics*, **60**, 1169–1177.
- Sava, P., and Guitton, A., 2005, Multiple attenuation in the image space: *Geophysics*, **70**, v10–v20.
- Sava, P. C., and Fomel, S., 2003, Angle-domain common-image gathers by wavefield continuation methods: *Geophysics*, **68**, 1065–1074.
- Scales, J. A., and Gersztenkorn, A., 1987, Robust methods in inverse theory, *in* Scales, J. A., Ed., Geophysical imaging, symposium of geophysical society of Tulsa: Soc. of Expl. Geophys. 25–50.
- , 1988, Robust method in inverse theory: *Inverse Problems*, **4**, 1071–1091.
- Scales, J. A., Gersztenkorn, A., Treitel, S., and Lines, L. R., 1988, Robust optimization methods in geophysical inverse theory *in* 58th Ann. Internat. Mtg. Soc. of Expl. Geophys., Session:S7.1.
- Schuster, G. T., and Liu, Z., 2001, Seismic array theorem and rapid calculation of acquisition footprint noise: *Geophysics*, **66**, 1843–1849.
- Shanno, D. F., 1970, Conditioning of quasi-Newton methods for function minimization: *Math. Comp.*, **24**, 647–657.
- Sirgue, L., and Pratt, R. G., 2004, Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies: *Geophysics*, **69**, 231–248.
- Soubaras, R., 1994, Signal-preserving random noise attenuation by the F-X projection *in* 64th Ann. Internat. Mtg. Soc. of Expl. Geophys., 1576–1579.

- , 1995, Prestack random and impulsive noise attenuation by F-X projection filtering *in* 65th Ann. Internat. Mtg. Soc. of Expl. Geophys., 711–714.
- , 2001, Dispersive noise attenuation for converted wave data *in* 71st Ann. Internat. Mtg. Soc. of Expl. Geophys., 802–805.
- Spitz, S., 1999, Pattern recognition, spatial predictability, and subtraction of multiple events: *The Leading Edge*, **18**, 55–58.
- Stoffa, P. L., Fokkema, J. T., de Luna Freire, R. M., and Kessinger, W. P., 1990, Split-step Fourier migration: *Geophysics*, **55**, 410–421.
- Stolt, R. H., and Weglein, A. B., 1985, Migration and inversion of seismic data: *Geophysics*, **50**, 2458–2472.
- Taner, M. T., and Koehler, F., 1969, Velocity spectra - digital computer derivation and applications of velocity functions: *Geophysics*, **34**, 859–881.
- Tarantola, A., 1986, A strategy for nonlinear elastic inversion of seismic reflection data: *Geophysics*, **51**, 1893–1903.
- , 1987, *Inverse Problem Theory*: Elsevier.
- Taylor, H. L., Banks, S. C., and McCoy, J. F., 1979, Deconvolution with the L1 norm: *Geophysics*, **44**, 39–52.
- Thorson, J. R., and Claerbout, J. F., 1985, Velocity stack and slant stochastic inversion: *Geophysics*, **50**, 2727–2741.
- Trad, D., Ulrych, T., and Sacchi, M., 2003, Latest views of the sparse Radon transform: *Geophysics*, **68**, 386–399.
- Treitel, S., Gutowski, P. R., and Wagner, D. E., 1982, Plane-wave decomposition of seismograms: *Geophysics*, **47**, 1375–1401.
- Valcarce, R. L., Dasgupta, S., Tempo, R., and Fu, M., 2000, Exponential Asymptotic Stability of Time-Varying Inverse Prediction Error Filters: *IEEE Transactions on Signal Processing*, **48**, 1928–1936.

- Verschuur, D. J., and Berkhout, A. J., 1997, Estimation of multiple scattering by iterative inversion, part II: Practical aspects and examples: *Geophysics*, **62**, 1596–1611.
- Verschuur, D. J., Berkhout, A. J., and Wapenaar, C. P. A., 1992, Adaptive surface-related multiple elimination: *Geophysics*, **57**, 1166–1177.
- Volohonsky, H., Kaplunovsky, A., and Serruya, S., 1983, Storms on Lake Kinneret: observations and mathematical model: *Ecological Modelling*, **18**, 141–153.
- Wang, Y., and Levin, S. A., 1994, An investigation into eliminating surface multiples: *SEP*—**80**, 589–602.
- Weglein, A. B., 1999, Multiple attenuation: an overview of recent advances and the road ahead (1999): *The Leading Edge*, **18**, 40–44.
- Weglein, A. B., Gasparotto, F. A. F., Carvalho, P. M., and Stolt, R. H., 1997, An inverse-scattering series method for attenuating multiples in seismic reflection data: *Geophysics*, **62**, 1975–1989.
- Weglein, A. B., and Stolt, R. H., 1999, Migration-inversion revisited (1999): *The Leading Edge*, **18**, 950–952.
- Wiggins, J. W., 1988, Attenuation of complex water-bottom multiples by wave equation-based prediction and subtraction: *Geophysics*, **53**, 1527–1539.
- Youn, O., and Zhou, H.-W., 2001, Depth imaging with multiples: *Geophysics*, **66**, 246–255.
- Zhang, Z., Chunduru, R. K., and Jervis, M. A., 2000, Determining bed boundaries from inversion of EM logging data using general measures of model structure and data misfit: *Geophysics*, **65**, 76–82.