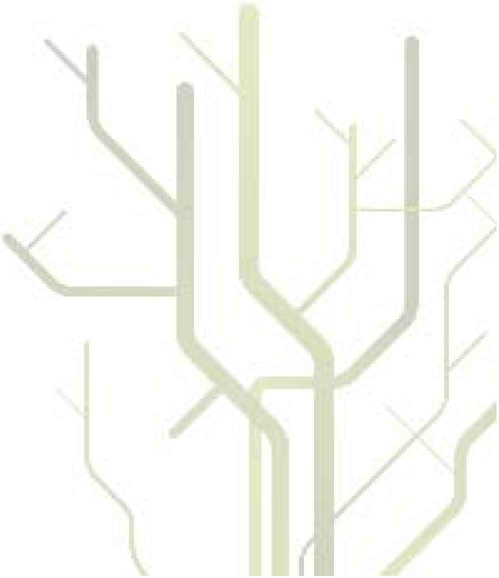
UNIVERSITY OF TROMSØ UIT

FACULTY OF SCIENCE
DEPARTMENT OF PHYSICS AND TECHNOLOGY

Signal Processing Techniques for the Enhancement of Marine Seismic Data

Charlotte Sanchis

A dissertation for the degree of
Philosophiae Doctor
November 12th, 2010



Abstract

The goal of marine seismic exploration is to obtain an image of the subsurface. Thus, an air gun is used as an impulsive source to generate pressure waves that travel down into the subsurface, where they are reflected. Some of the reflected energy is recorded by a long seismic streamer towed behind a vessel, together with unwanted noise and among others, weather and flow related noise. More precisely, streamers are subjected to forces from the vessel, from steering birds and from the motion of the sea. Minimizing the effects of unwanted noise is one of the main challenges faced by seismic data processors. In addition, because geophysicists try to handle more and more complex subsurface structures, it has become necessary to come up with more elaborate signal enhancement techniques.

The derivation of signal enhancement algorithms usually requires statistical assumptions about the noise to be attenuated. For this reason, we start with a statistical analysis of seismic background noise realized with recent marine streamers. State of the art techniques such as the Thomson's multitaper estimator for spectral analysis and Parzen amplitude estimator for density estimation are used. Thereafter, we propose different signal enhancement methods. We first focus on single-azimuth stacking methodologies and propose novel stacking schemes, using either the concept of weighted sum or a novel approach by means of Kalman filter. We then address the problem of random noise attenuation and present an innovative application of sparse code shrinkage and independent component analysis. Finally, we consider the attenuation of flow-generated nonstationary coherent noise and seismic interferences noise. We suggest a multiple-input adaptive noise canceller that utilizes a normalized least mean squares algorithm with a variable normalized step size that is derived as a function of instantaneous frequency.

All the novel methods presented in this thesis have been tested on synthetic and real marine data, and compared against standard existing signal enhancement methods. Thus, we show that the enhanced stacking method yields superior results compared to conventional stacking, and in particular, it exhibits cleaner and better defined reflected events as well as a larger number of reflections in deep waters. A comparison with more advanced stacking methods is also discussed. Furthermore, we demonstrate that sparse code shrinkage is an efficient random noise attenuation method that targets the noise components to shrink them while preserving the signals. By its capacity to enhance signals in both shallow and deep waters, the normal inverse Gaussian density appears to be the best choice of density, but the sparse density, the mixture of generalized Gaussian density and non-parametric Gaussian kernel estimate have proven to be acceptable choices as well. In addition, the comparison with local singular value decomposition and *f*-*x* deconvolution shows that sparse code shrinkage is a valuable method for removing background noise and estimating signals in both shallow and deep waters. Finally, the multiple-input adaptive noise canceller using a normalized least mean squares algorithm with a variable normalized step size attenuates the coherent noise successfully, when used either by itself or in combination with a time-frequency median filter, depending on the noise repartition and its spectral characterization. The comparison with a time-frequency median filter and a high-pass Butterworth filter shows that it yields significantly improved results. Furthermore, we show that the multiple-input adaptive noise canceller can find some applications for seismic interferences attenuation, after manipulation of the input data set. Its efficiency remains however limited at this stage of development.

Preface

This thesis contains seven chapters. Chapter 1 gives an history and the general concepts of seismic acquisition, a brief review of a few existing and comparative noise attenuation methods, as well as a description of my own contribution. Chapters 2, 3, 5 and 6 contain original work that has been published or submitted in various journals and conferences, while chapter 4 presents a brief review of principal component analysis and a related noise attenuation method, the singular value decomposition. Finally, chapter 7 is a summary of the thesis with some suggestions to future work.

Chapter 1: Introduction

This chapter presents a history of seismic acquisition, as well as the general concepts and a brief review of a few existing and comparative noise attenuation methods. In addition, a description of the work presented in this thesis is provided.

Chapter 2: Seismic noise analysis

In this chapter, we define several statistical tools and use them for a statistical analysis of pure seismic noise. Furthermore, we propose an automatic clustering of the noise level using an existing method, but that provides an original quality estimation tool for geophysicists. Part of this work has been presented in the conference paper

- C. Sanchis and A. Hanssen, "Statistical analysis of noise in towed streamer arrays", *Proceedings* 16^{th} *European Signal Processing Conference*, Lausanne, Switzerland, August 25-29, 2008.

Chapter 3: A novel enhanced local correlation stacking method

In this chapter, we present novel stacking schemes and compare them to the conventional technique. This chapter contains the journal paper

- C. Sanchis and A. Hanssen, "Enhanced local correlation stacking method", submitted to *Geophysics* in February 2010.

Chapter 4: Principal components analysis, brief review

This chapter is a brief review of principal components analysis and a related concept, the singular value decomposition. Because principal components analysis is used in chapter 5, we think it is necessary to remind the reader about this concept. In addition, the related concept of singular value decomposition, used in this thesis as comparative method of noise attenuation, is described as well.

Chapter 5: Sparse code shrinkage for signal enhancement of seismic data

This chapter presents an innovative application of sparse code shrinkage and independent component analysis for signal to noise ratio enhancement of seismic signals, as well as an investigation of different parametric and a non-parametric density model. This chapter contains the conference and journal papers

- C. Sanchis and A. Hanssen, "Attenuation of seismic noise by sparse code shrinkage", *Proceedings* 72^{nd} *European Association of Geoscientists & Engineers Conference*, Barcelona, Spain, June 14-17, 2010.
- C. Sanchis and A. Hanssen, "Sparse code shrinkage for signal enhancement of seismic data", submitted to *Geophysics* in May 2010.

Chapter 6: Multiple-input adaptive noise canceller to attenuate non-stationary coherent noise

This chapter presents an application of the multiple-input adaptive noise canceller to the attenuation of flow-generated non-stationary coherent noise and seismic interferences noise. This filter uses a normalized least mean squares algorithm with a variable normalized step size that has been derived to filter the seismic data efficiently. The work presented in this chapter will shortly be submitted.

Chapter 7: Conclusions and suggestions for future work

This chapter contains a short summary of this monograph, along with some suggestions to future work.

Acknowledgement

The work presented in this thesis has been carried out under the supervision of Professor Alfred Hanssen at the Department of Physics and Technology, University of Tromsø, Norway. I would like to thank Professor Hanssen for his involvement, his great enthusiasm and his large effort toward the doctoral students. I highly appreciated his guidance throughout my doctoral studies. He was also a great support for me and my family during our pleasant stay in Tromsø.

I also wish to thank Peder Berentzen, Thomas Elboth, Dr. Xiao-Ping Li, Pierre-Yves Raya and Ottar Sandvin who regularly helped me during my thesis, and more generally all my colleagues at Fugro Geoteam, Fugro Seismic Imaging and Fugro Multi Client Services for providing seismic data, helping me with Fugro internal software, or simply for giving me useful advices and feedbacks about my work. I do not forget Didrik Lilja, Terje Steinsbø and Jakub Warszawski for an instructive offshore trip. I am also grateful to people from the university of Tromsø for interesting scientific discussions and for their pleasant company during my stay in Tromsø. Finally, I wish to thank Aapo Hyvärinen, Robert Jenssen and Tor Arne Øigård for very useful scientific discussions on specific problems.

I would like to thank the Research Council of Norway for support under the PETROMAKS project no. 175921/S30, as well as my employer Fugro Geoteam for giving me the opportunity of pursuing doctoral studies and granting me during that period.

Finally, I would like to express my love and my gratitude to my husband, Arnaud, who encouraged me to pursue doctoral studies and supported me throughout this thesis.

List of symbols

a	Local parameter
$A(t)$	Instantaneous amplitude of the analytic signal
$\mathbf{A}$	ICA mixing matrix
α	Local parameter
b	Local parameter
$b\{x\}$	Bias of x
$\mathbf{B}$	Matrix used in multitaper spectral analysis
β	Local parameter
$\beta(n)$	Variable normalized step size
c_k	Shape parameter of the MGG density
d	Local sparse density parameter
$d_{j,k}$	Euclidean distance between the j^{th} and k^{th} vectors
$\overline{d_{j,k}}$	Average Euclidean distance between the j^{th} and k^{th} vectors
$\mathbf{D}$	Distance matrix of clustering
δ	Local parameter
$e(n)$	Error value
e_j	Norm of the j^{th} trace segment
$\overline{e_j}$	Smoothed norm of the j^{th} trace segment
η	Local parameter
ϵ	Local parameter
$E\{x\}$	Expected value of x
$\mathcal{E}_K$	Energy of the kernel
$\mathbf{E}_k$	k^{th} eigenimage
f	Frequency
$f(s)$	Probability density function
f_B	Half-bandwidth resolution of the multitaper estimator
$f_d(t)$	Fold of the data at time t
F_S	Sampling frequency
$g(x)$	Shrinkage function
G_i	i^{th} subset of the normalization scheme
$\gamma_{i,j}(t)$	local correlation coefficient of the i^{th} trace and j^{th} CMP gather
γ_k	Scale parameter of the MGG density
Γ_3	Skewness
Γ_4	Kurtosis
Γ	Gamma function
h	Smoothing parameter of the Parzen estimator
$\mathcal{H}$	Hilbert transform
$H(z)$	System function of a digital filter
$i(s)$	Negative log-density of s

$i'(s)$	Score function
I	Interval of traveltime occurences where $\Gamma(t)$ is beyond the threshold
$\mathbf{I}_N$	$N \times N$ identity matrix
j	the square root of -1
k	Local parameter
k_i	Kalman gain of the i^{th} trace
$K(\xi)$	Smoothing kernel
$K_h(\xi)$	Smoothing kernel with parameter h
K_0	Modified Bessel function of the second kind with index 0
K_1	Modified Bessel function of the second kind with index 1
$\kappa^{(n)}$	Cumulant of order n
$\ln$	Natural logarithm
λ_i	i^{th} eigenvalue
$\boldsymbol{\Lambda}$	Matrix of eigenvalues λ_i
L_w	Window length
L	Filter length
m	Local parameter
$m_s^{(n)}$	Moment of order n about the origo of s
μ	Local parameter
M	Number of traces
M_C	Number of traces contained in a CMP gather
N	Number of samples
N_B	Block length in samples
N_G	Number of subset for the normalization scheme
Nf_0	Time-bandwidth parameter of the multitaper estimator
ν	Noise component
$\boldsymbol{\nu}(n)$	Noise vector made of the noise components ν
$\mathcal{N}(m, \sigma^2)$	Normal distribution of mean m and variance σ^2
ω	Angular frequency
ω_i	S/N based weight of the i^{th} trace
p	Number of singular values
$p(s)$	Local variable of the NIG function
$\mathbf{p}_i$	Principal component
$\mathbf{P}$	PCA transformation matrix
P_i	Kalman filter estimation covariance matrix of the i^{th} trace
Φ_1	First threshold value of instantaneous frequency
Φ_2	Second threshold value of instantaneous frequency
$\phi_k(t)$	Instantaneous frequency
$\psi(t)$	Mexican hat wavelet
$q(s)$	Local variable of the NIG function
$Q_{v,i}$	Kalman filter observation noise covariance matrix of the i^{th} trace
$\overline{Q}(f)$	Total spectral window

r	Rank of a matrix
r_i	Cross-correlation at zero lag of the i^{th} trace
$\mathbb{R}$	Set of real numbers
$\rho_{s,i}$	Signal power of the i^{th} trace
ρ	Local parameter
σ^2	Variance
σ_K^2	Local parameter of the Gaussian kernel estimator
ς_i	Singular value
$\mathbf{\Sigma}$	Diagonal matrix of singular values
$\mathbf{\Sigma}_G$	Sum of local correlation coefficients
s	Signal component
$\hat{s}$	Estimate of s
$\mathbf{s}$	Signal or source vector
sgn	Sign function
$S(f)$	True power spectrum
$\widehat{S}_{MT}(f)$	Multitaper power spectrum estimate
$\mathbf{S}$	Sparse components matrix
$\mathbf{T}$	Linkage information matrix of clustering
t	Time variable
T_G	Ensemble containing traveltimes of local maxima
T_S	Sampling period
τ_i	Traveltime of a local maximum
θ_k	Local parameter of the MGG density
$\theta(t)$	Instantaneous phase of analytic signal
$\mathbf{U}$	Matrix of eigenvectors of $\mathbf{X}\mathbf{X}^H$
$\mathbf{v}$	Taper vector
$\mathbf{V}$	Matrix of eigenvectors of $\mathbf{X}^H\mathbf{X}$
$V(f)$	Discrete Fourier transform of the taper $v[n]$
var $\{x\}$	Variance of x
$w_{i,j}$	Weight of the i^{th} trace and j^{th} CMP gather based on local-correlation
$\mathbf{w}_i(n)$	Tap-weight vector
$\mathbf{W}$	ICA unmixing matrix
$x(t)$	Seismic trace component
$\tilde{x}(t)$	Hilbert transform of $x(t)$
$\bar{x}(t)$	Stacked component
$\mathbf{x}$	Seismic trace column vector
$\bar{\mathbf{x}}$	Stacked trace vector made of stacked components $\bar{x}(t)$
$X(f)$	Fourier transform of $x(t)$
$\mathbf{X}$	Data matrix
$\hat{\mathbf{X}}$	Estimate of data matrix
$\mathbf{X}_{nf}$	"Noise-free" realization of data matrix
$\mathbf{X}_{LP}$	Low-pass subspace

$\mathbf{X}_{BP}$	Band-pass subspace
$\mathbf{X}_{HP}$	High-pass subspace
$\mathbf{X}_0$	Noise-free data matrix $\mathbf{X}$
$\bar{\mathbf{y}}$	Stacked trace vector made of stacked components $\bar{y}(t)$
ξ	Local parameter
$\mathbb{Z}$	Set of integer numbers
$\mathbf{x}^H$, $\mathbf{X}^H$	Hermitian transpose of a vector $\mathbf{x}$ or matrix $\mathbf{X}$
$\mathbf{x}^T$, $\mathbf{X}^T$	Transpose of a vector $\mathbf{x}$ or matrix $\mathbf{X}$
x^*	Complex conjugate of x
$\|\|\mathbf{X}\|\|_F$	Frobenius norm of a matrix $\mathbf{X}$
$*$	Convolution operator
∇	Gradient operator

List of abbreviations

AGC	Automatic gain control
CDP	Common depth point
CMP	Common midpoint
CPU	Central processing unit of time
DPSS	Discrete prolate spheroidal sequence
EM	Expectation - maximization method
FIR	Finite impulse response
F-K	Frequency-wavenumber domain
ICA	Independent component analysis
IMSE	Integrated mean square error
KL	Karhunen-Loève
LMO	Leave-me-out method
LMS	Least mean squares
MAP	Maximum a posteriori
MGG	Mixture of generalized Gaussian
ML	Maximum likelihood
MSE	Mean square error
MT	Multitaper
NIG	Normal inverse Gaussian
NLMS	Normalized least mean squares
NMO	Normal moveout
PCA	Principal component analysis
pdf	Probability density function
RMS	Root mean square
SRME	Surface-related multiple elimination
SVD	Singular value decomposition
S/N	Signal to noise ratio
UPGMA	Unweighted pair group method with arithmetic mean
WOSA	Weighted overlapped segment averaging

Contents

Abstract i

Preface iii

Acknowledgements v

List of symbols vii

List of abbreviations xi

Table of Contents xiii

1 Introduction **1**
- 1.1 Seismic acquisition: history and motivations . . . 1
- 1.2 Definitions and general concepts used in geophysics . . . 3
- 1.3 Discussion about existing noise-attenuation methods . . . 4
- 1.4 Present work and contributions . . . 5

2 Seismic noise analysis **7**
- 2.1 Background information and test data description . . . 7
- 2.2 Multitaper Spectral Analysis . . . 8
  - 2.2.1 Brief Outline of Multitaper Method . . . 8
  - 2.2.2 Application: Spectral analysis of real seismic noise recordings . . . 11
- 2.3 Clustering of power spectrum . . . 12
  - 2.3.1 Average linkage cluster methodology . . . 12
  - 2.3.2 Application: Cluster of real pure-noise recordings . . . 15
- 2.4 Parzen Amplitude Estimates . . . 19
  - 2.4.1 Moments of the Amplitude Probability . . . 20
  - 2.4.2 Application: Amplitude Probability Density Estimates . . . 21
- 2.5 Summary . . . 23

3 A novel enhanced stacking method **25**
- 3.1 Introduction . . . 25
- 3.2 Reference traces . . . 27

3.2.1 Conventional stacking 27
3.2.2 Signal-to-noise ratio estimation 27
3.2.3 Stacking using Kalman filter 29
3.3 Enhanced local correlation stacking method 30
3.3.1 Unscaled local correlation 30
3.3.2 Weight determination 30
3.3.3 Summary 33
3.4 Applications 34
3.4.1 Application on synthetic data 34
3.4.2 Application on real data 39
3.5 Summary 48

4 Principal components analysis, brief review 53
4.1 Introduction 53
4.2 Definition and properties 54
4.3 Singular value decomposition 56
4.3.1 Theorem 56
4.3.2 Properties 56
4.4 Application to image processing and seismic data 58
4.4.1 Theoretical aspects 58
4.4.2 Application to synthetic and real data 58
4.5 Summary 60

5 Sparse code shrinkage for signal enhancement of seismic data 63
5.1 Introduction 63
5.2 Maximum a posteriori estimation of sparse coded signal 65
5.2.1 ICA representation 65
5.2.2 Maximum likelihood estimator 66
5.3 Shrinkage functions 67
5.3.1 Sparse densities 68
5.3.2 The normal inverse Gaussian density (NIG) 69
5.3.3 Mixture of Generalized Gaussian densities (MGG) 72
5.3.4 Gaussian kernel density estimator 72
5.4 Sparse code shrinkage algorithm 73
5.4.1 Noise-free realization of the data 73
5.4.2 Estimation of noise variance 74
5.4.3 Algorithm 74
5.5 Applications 76
5.5.1 Test on synthetic data 76
5.5.2 Test on real data 79
5.6 Discussion 81
5.6.1 Parameter settings 81
5.6.2 Possible improvements of sparse code shrinkage 81

5.7 Summary 82

6 Multiple-input adaptive noise canceller to attenuate non-stationary coherent noise 93
6.1 Introduction 93
6.2 Theory 94
6.2.1 Multiple-input adaptive noise canceller 94
6.2.2 The normalized least mean squares algorithm 95
6.2.3 Variable normalized step size parameter 97
6.2.4 Other comparative methods 99
6.3 Applications 99
6.3.1 Data conditioning 99
6.3.2 Application to swell noise attenuation: first example 100
6.3.3 Application to swell noise attenuation: second example 101
6.3.4 Application to seismic interferences attenuation 107
6.4 Discussion 110
6.5 Summary 114

7 Conclusions and suggestions for future work 115

Appendix 119

Bibliography 121

Chapter 1

Introduction

1.1 Seismic acquisition: history and motivations

The goal of marine seismic exploration is to obtain an image of the subsurface. Seismic acquisition is the artificial generation and recording of seismic data. For that purpose, a seismic source, typically a compressed air gun for marine surveys and a dynamite explosion or a vibrator unit for land surveys, is used as an impulsive source to generate pressure waves that travel down into the Earth passing through Earth layers with different seismic responses and filtering effects. Some energy returns to the surface due to refraction or reflections, to be recorded as seismic data. The reflected seismic energy is recorded by receivers, which record the seismic signals as electronic waveforms. Hydrophones are one type of receiver, used to measure changes in pressure as sound travels through water. They are usually towed a few meters underwater inside streamers that are typically 3 to 10 km long. Hydrophones can also be suspended vertically or laid on the sea floor. Different types and configurations of receivers are used depending on local environmental conditions and the underground geological features that are to be imaged. The layout of the receivers is also designed to minimize the effect of noise that can otherwise mask the seismic signal. So long as appropriate receiver configurations are used, special computer processing techniques can be applied to the recorded data to remove unwanted noise and enhance the seismic signal. In this thesis, all the real seismic data are acquired by underwater, horizontal streamers towed behind a seismic vessel, as shown in Figure 1.1. A more detailed description of a seismic streamer used for data acquisition is given in section 2.1.

Reflections and refractions of seismic waves at geologic interfaces within the Earth were first observed on recordings of earthquake-generated seismic waves. The basic model of the Earth's deep interior is based on observations of earthquake-generated seismic waves transmitted through the Earth's interior (e.g., the Mohorovicic discontinuity). The use of human-generated seismic waves to map in detail the geology of the upper few kilometers of the Earth's crust followed shortly after and has developed mainly due to commercial enterprise, particularly the petroleum industry. The Canadian inventor Reginald Fessenden was the first to conceive of using reflected seismic waves to infer geology. He filed patents on the method in 1917 while working on methods of detecting submarines during World War I. Due to the war, he was unable to follow up on the idea. However, John Clarence Karcher discovered seismic reflections independently while working for the United States Bureau of Standards (now the National Institute of Standards and Technology) on methods of sound ranging to detect artillery. In discussion with colleagues, the idea developed that these reflections could aid in exploration for petroleum. With several others, many affiliated with the University of Oklahoma, Karcher helped to form the Geological Engineering Company, incorporated in Oklahoma in April, 1920. The first field tests were conducted near Oklahoma City, Oklahoma in 1921. The company soon folded due to a drop in the price of oil. In 1925, oil prices had rebounded, and Karcher helped to form Geophysical Research Corporation (GRC) as part of the oil company Amerada. In 1930, Karcher left GRC and helped to found Geophysical Service Incorporated (GSI). GSI was one of the most successful seismic contracting companies for over 50 years and was the parent of an even more successful company, Texas Instruments. Early GSI employeee Henry Salvatori left that company in 1933 to found another major seismic contractor, Western Geophysical. As of 2005, after several mergers and acquisitions, the heritages of GSI and Western Geophysical still exist, along with several pioneering European companies. Many other companies using reflection seismology in hydrocarbon exploration, hydrology, engineering studies, and other applications have been formed since the method was first invented, and Fugro is one of them. Reflection seismology has also found applications in non-commercial research by academic and government scientists around the world.

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. An hydrophone can have malfunctions 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, nearby seismic vessels and 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.

The second category of noise comes from modeling uncertainties (Tarantola, 1987). In seismic, modeling uncertainties occur when the physical description and the 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, 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. Whereas experimental errors rarely bring any useful information, recent advances are slowly pushing the limits of modeling uncertainties by incorporating and saving as much signal as possible within the processing workflow.

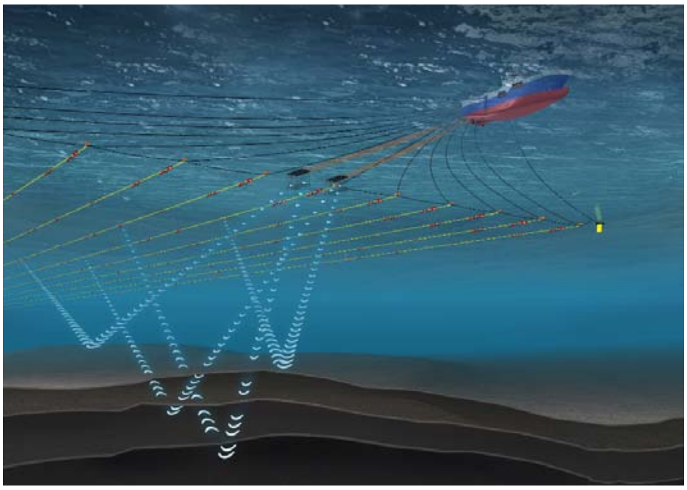


Figure 1.1: Acquisition process of seismic data.

1.2 Definitions and general concepts used in geophysics

This section gives the definitions of general concepts used in seismic processing.

- Common mid point (CMP) gather: Recorded data are sorted into gather where all traces have a common midpoint, i.e., the position on the surface mid-way between the source and the receiver. If the reflector is sub-horizontal and the medium is isotropic, then the CMP will be situated directly perpendicular above the common depth point (CDP) on the reflector. An example is shown in Figure 1.2 (b).

- Flow-generated noise: Noise resulting from the interaction between the streamer and the surrounding flow, i.e., Swell noise, turbulence-induced noise.

- Migration: Process by which geophysical events (reflected events) are geometrically re-located in either space or time to the location where the event occurred in the subsurface rather than the location that it was recorded at the surface, thereby creating an accurate image of the subsurface. This process is necessary to overcome the limitations of geophysical methods imposed by areas of complex geology, such as: faults, salt bodies, folding, etc. The end result is that the migrated image typically resolves these areas of complex geology much better than non-migrated images.

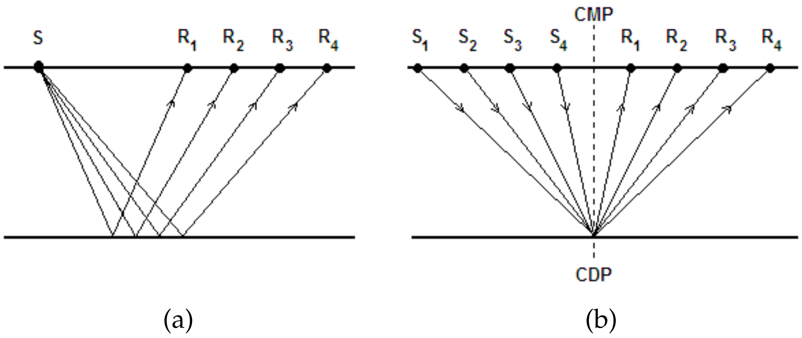
- Multiple reflection: Wave reflections with at least one downward reflection. Multiples travel at lower average velocities than primaries, and therefore, multiples with greater NMO are misaligned contrary to primaries.

- Normal moveout (NMO): The increase in reflection time due to an increase in distance from source to receiver for a horizontal reflecting interface in a homogeneous medium of constant velocity. Thus, if T_0 and T_x denote the two-way reflection time for a zero offset trace and a trace of offset distance x respectively, and if c denotes the compressional wave velocity of the medium, then the NMO time increment $\Delta T_{NMO}(x)$ is

$$\Delta T_{NMO}(x) = T_x - T_0 = T0(\sqrt{1 + \frac{x^2}{(cT_0)^2}} - 1) \tag{1.1}$$

The normal moveout correction consists of the subtraction of the time increment $\Delta T_{NMO}(x)$ from each record time T_x.

- Primary reflection: Wave reflection with only one upward reflection in the subsurface.
- Seismic interferences: Waves generated by other seismic vessel operating in the same area.
- Shot gather: Recorded data are sorted in such a way that all traces have the same source position. An example is shown in Figure 1.2 (a).
- Stacking: In a conventional sense, stacking consists of the simple sum of CDP traces to produce a single composited trace. This method increases reflection signal-to-noise ratios by a factor approaching $\sqrt{M}$, where M is the "fold" or number of CDP traces summed.



(a) (b)

Figure 1.2: (a) An example of a shot gather. (b) An example of a CMP gather: In case of a horizontal reflector, the common mid point on the surface coincides with the common depth point on the reflector in the subsurface.

1.3 Discussion about existing noise-attenuation methods

Throughout this thesis, we propose novel and enhanced noise attenuation methods, and compare them with existing ones. Thus, we dedicate Chapter 4 to a brief review of principal components analysis and singular value decomposition. Principal components analysis is used in sparse code shrinkage for dimensions reduction and for this reason, we want to make sure that the concept is known from the readers. A particular case, the singular value decomposition, is a coherency-based technique of signal enhancement that is well known from the seismic community, and briefly described as well in Chapter 4.

In addition, we will refer to the time-frequency median filter described by Elboth et al. (2010). This filter uses a sliding window that scans the input gather both in space and time. Within this window, a spectral estimate of all traces is achieved and the amplitude estimates at each frequency are compared with the amplitude of a presumed good trace within the chosen window. This technique is mostly used for random noise attenuation, and it is used as comparative method in Chapter 6.

Predicted deconvolution, also known as *f*-*x* deconvolution or *f*-*x* prediction, is one of the most common techniques in seismic to attenuate noise. It finds some applications when random noise is present or when coherent noise can be viewed in a random domain. It is based on a quite simple idea: linear or quasi-linear events in the time-space domain manifest themselves as a superposition of harmonics in frequency-space (*f*-*x*) domain, and can then be modeled optimally by an autoregressive moving average model. Because instabilities may occur, a long autoregressive process is usually used. Thereafter, the signal at each frequency is modeled by an autoregressive model and resolved by a least squares prediction filter. Thus, it is predicted by the nearby traces and noise is the unpredicted factor. Finally, the filtered signal is transformed back to the time-space domain. A more exhaustive description and history is given by Ulrych and Sacchi (2005). However, in this thesis, we use the *f*-*x* deconvolution code contained in the seismic data processing package SeismicLab and developed by Sacchi and Signal Analysis and Imaging Group (2008) as comparative method in Chapter 5.

Finally, it is a common practice in seismic processing to remove any low-frequency components, generally considered as noise, by means of a high-pass filter with cutoff frequency between $[2, 5]$ Hz. We will use such tool, more specifically a second order high-pass Butterworth filter, in Chapter 6 as comparative method for low-frequency noise attenuation. The Butterworth filter is a well known type of filter designed to have a frequency response which is as flat as possible in the passband and rolls off towards zero in the stopband (Bianchi and Sorrentino, 2007). The characteristics of the used filter, i.e., the cutoff frequency, the system function and the magnitude response are given in Chapter 6.

1.4 Present work and contributions

This thesis concentrates on statistical modeling and analysis of background noise, and on development of signal enhancement algorithms that are able to tackle random or coherent noise. Thus, we start in Chapter 2 with the statistical analysis of seismic noise. Noise characterization has been performed in the past, with focus either on low-frequencies, on hydrophone streamer noise or on ocean noise. More recent marine streamers are now used in seismic acquisition, and it is therefore pertinent to extend the previous results with noise characterization completed from modern equipments. Thus, the analysis reveals that the seismic noise is almost symmetric and slightly platykurtic, and that it yields a composite power-law. These results will be used at a latter stage in this thesis. Furthermore, the inhomogeneity of seismic noise along a streamer is emphasized, and in that purpose, an automatic clustering of the noise level is proposed as an original quality estimation tool for geophysicists. Chapter 3 focuses on single-azimuth stacking methodologies, as popular signal to noise ratio enhancement techniques among the seismic industry. We propose novel stacking schemes, using either a weighted sum or a Kalman filter, and compare them to the conventional stacking technique. We dedicate then the Chapter 5 to an innovative application of sparse code shrinkage and independent component analysis for signal to noise ratio enhancement of seismic signals. Here, we propose a method to obtain the "noise-free" realization of the data that allows the derivation of fitted shrinkages, and we investigate different parametric and a non-parametric density models in order to determine the model that suits the best the seismic data. Eventually, Chapter 6 presents an application of the multiple-input adaptive noise canceller to the attenuation of flow-generated non-stationary coherent noise and seismic interferences noise. This filter uses a normalized least mean squares algorithm with a variable normalized step size. The determination of the variable step size parameter is not an easy task and classical methods fail because they cannot adapt to the characteristics of the seismic noise. Here, we propose to derive this parameter as a function of instantaneous frequency for each seismic trace and we show that it makes the multiple-input adaptive noise canceller be a credible alternative for the attenuation of flow-generated non-stationary coherent noise and possibly seismic interferences noise.

This thesis provides the necessary mathematical background and the theory behind each proposed method, as well as their strengths and weaknesses. It also contains many examples of applications, based on both synthetic and real marine data sets.

Finally, the experimental part of this thesis consists in a week offshore, spent on seismic vessel to manipulate standard seismic acquisition equipments and collect my own data. This experience has been really instructive and fruitful, and provided me a better understanding of the possible noise sources.

Chapter 2

Seismic noise analysis

2.1 Background information and test data description

Statistical characterization of seismic noise has been performed in the past, with focus either on low frequencies or on hydrophone streamer noise (Schoenberger and Mifsud, 1974). Statistical properties of ocean noise have been studied as well, (Brockett et al., 1987) and (Pflug et al., 1997). More recent marine streamers are now used in seismic acquisition. It is therefore pertinent to extend the previous results with noise characterization completed from modern equipments. Results could be used for future work on noise reduction. In this chapter, we consider typical pure-noise recordings and we perform a statistical analysis of these data set. Thus, we start with a spectral analysis using Thomson's multitaper estimator that yields the derivation of a composite power-law. We then use a clustering method that reveals the inhomogeneity of the noise along a streamer. Finally, we estimate the distribution of noise amplitudes using a Parzen kernel estimator, and use the third and fourth order central moments, respectively the skewness and the kurtosis, to characterize the amplitude distribution.

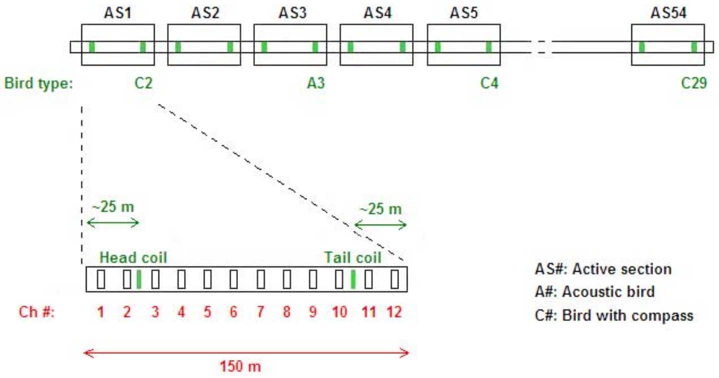


Figure 2.1: Streamer layout.

Data consists in passive hydrophones recordings, i.e., in the absence of impulsive sources. It has been acquired during a Fugro campaign in the Mediterranean sea. The vessel's speed was 5 knots and weather conditions were relatively good with 1 to 1.5 meters waves. Nominal streamer depth is 8 meters. The noise data set was obtained using a Sentinel® solid streamer which consists in a large number of hydrophones, grouped in sub-sections of 12.5 meters long. Within each of 12.5 meters subsections, inputs from hydrophones are summed and a trace representing recorded signal is produced and referred to as a channel. As illustrated in Figure 2.1, the seismic streamer is divided into 54 so-called active sections. Each active section is 150 meters long and contains 12 groups of hydrophones, providing 12 channels with noise records. Altogether, the streamer has 648 channels. There are two possible locations where extra units with depth keeping and ranging functionalities can be attached, at ~ 25 meters from the head and ~ 25 meter from the tail of each active section. Whether they cause extra noise, has to be determined.

Noise pressure values are recorded every 2 milliseconds and for a duration of 6 seconds, simultaneously by the 648 channels. They are gathered as one noise recording. Such a recording process is repeated 10 times at approximatively 6 seconds interval, so that a total of 10 noise recordings are available for the analysis.

2.2 Multitaper Spectral Analysis

It is well known that the standard periodogram estimator of the power spectral density is statistically inconsistent (Percival and Walden, 1993). The shortcomings of the periodogram may be mitigated by tapering in time to reduce spectral leakage, and by smoothing in frequency or time to reduce the estimator variance. The state of the art tapering technique is David J. Thomson's multitaper estimator. This estimator reduces spectral leakage subject to certain optimality criteria, while simultaneously reducing variance through an averaging procedure over an orthonormal set of tapers. A time-frequency resolution bandwidth parameter must be chosen by the user. In practice, the analysis of seismic records is not very sensitive to the actual choice of the bandwidth parameter, as long as an adequate compromise between smoothing and frequency resolution is identified. We regard the multitaper estimator to be the natural choice for our kind of data, and several examples of multitaper estimates will be shown in this chapter.

2.2.1 Brief Outline of Multitaper Method

As the multitaper method may not be well known among seismic data analysts, we will now briefly outline the estimator before actually applying it to streamer data.

The multitaper (MT) technique proposed by Thomson (1982), follows ideas from Slepian (1978). This method combines the use of optimal data tapers, with averaging over a set of power spectral estimates.

Discrete Prolate Spheroidal Sequences

Thomson (1982) proposed to apply some stringent optimality criteria when selecting data tapers. He suggested to consider tapers that maximizes the "spectral concentration", or the energy contained in the mainlobe relative to the total energy of the taper. One therefore seeks the taper $v[n]$ with a discrete Fourier transform $V(f)$, that maximizes the window energy ratio

$$\lambda = \frac{\int\limits_{-f_B}^{f_B} |V(f)|^2 \, df}{\int\limits_{-1/2}^{1/2} |V(f)|^2 \, df} \tag{2.1}$$

where f_B is the wanted resolution half-bandwidth (a design parameter) of the taper. An ideal taper would therefore have $\lambda \simeq 1$ and f_B as small as possible (but note that $f_B > 1/N$).

Expressing $V(f)$ by its discrete Fourier transform, $V(f) = \sum_{n=0}^{N-1} v[n] \exp(-j2\pi f n)$ and maximizing the above functional with respect to $v[n]$, Slepian (1978) showed that the optimal taper $\mathbf{v} = [v[0], v[1], \ldots, v[N-1]]^T$ obeys the eigenvalue equation

$$\mathbf{B}\mathbf{v} = \lambda \mathbf{v} \tag{2.2}$$

where the matrix $\mathbf{B}$ has elements $[\mathbf{B}]_{nm} = \sin[2\pi f_B(n-m)]/[\pi(n-m)]$, for $n, m = 0, 1, \ldots, N-1$. Note that equation 2.2 is an N-dimensional eigenvector/eigenvalue problem, thus giving N eigenvector/eigenvalue pairs, $(\mathbf{v}_k, \lambda_k)$, where $k = 0, 1, \ldots, N-1$. The interpretation is thus that we obtain a *sequence* of orthogonal tapers (eigenvectors), $\mathbf{v}_k$, each with a corresponding spectral concentration measure λ_k. The first taper $\mathbf{v}_0$ has a spectral concentration λ_0. Then, $\mathbf{v}_1$ maximizes the ratio in equation 2.1 subject to being orthogonal to $\mathbf{v}_0$, and with $\lambda_1 < \lambda_0$. Continuing, we can thus form up to N orthogonal tapers $\mathbf{v}_0, \mathbf{v}_1, \ldots, \mathbf{v}_{N-1}$, with $0 < \lambda_{N-1} < \lambda_{N-2} < \cdots < \lambda_0 < 1$. Only tapers with $\lambda_k \simeq 1$ can be applied, since $\lambda_k \ll 1$ implies a large undesirable leakage.

It is usually safe to apply tapers up to order $k = 2Nf_B - 1$ (Percival and Walden, 1993). It is customary to standardize the tapers such that they are orthonormal, $\mathbf{v}_k^T \mathbf{v}_{k'} = \delta_{k,k'}$, where $\delta_{k,k'}$ is the Kronecker delta. The solutions $\mathbf{v}_k$ are referred to as "Discrete Prolate Spheroidal Sequences" (DPSS) (Slepian, 1978). These optimal tapers are not expressible in closed form. The eigenvalue equation 2.2 must therefore be regarded as the defining equation for these tapers. Recent versions of the Matlab® Signal Processing Toolbox include efficient solvers for the DPSS problem.

Basic Multitaper Spectral Estimators

The simplest definition of an MT estimate is simply the arithmetic average of K tapered "eigenspectra"

$$\widehat{S}_{MT}(f) = \frac{1}{K}\sum_{k=0}^{K-1}\widehat{S}_{MT}^{(k)}(f) \tag{2.3}$$

where the "eigenspectrum" of order k is defined by

$$\widehat{S}_{MT}^{(k)}(f) = \left|\sum_{n=0}^{N-1} v_k[n]x[n]\exp(-j2\pi fn)\right|^2 \tag{2.4}$$

where $v_k[n]$ denotes the elements of DPSS-taper of order k. Also data adaptive averaging schemes exist, see Thomson (1982) and Percival and Walden (1993). The adaptive averaging is necessary in several applications, and will be employed also in this chapter.

The averaging of tapered spectral estimates, in equation 2.3, leads to a decrease of the variance relative to any individual spectral estimates. Asymptotically, one can show that Thomson (1982)

$$\text{var}\{\widehat{S}_{MT}(f)\} \simeq (1/K)S^2(f), \tag{2.5}$$

where $S(f)$ is the true power spectrum.

The expectation of the MT estimator is

$$E\{\widehat{S}_{MT}(f)\} = \int_{-1/2}^{1/2}\overline{Q}(f-f')S(f')df', \tag{2.6}$$

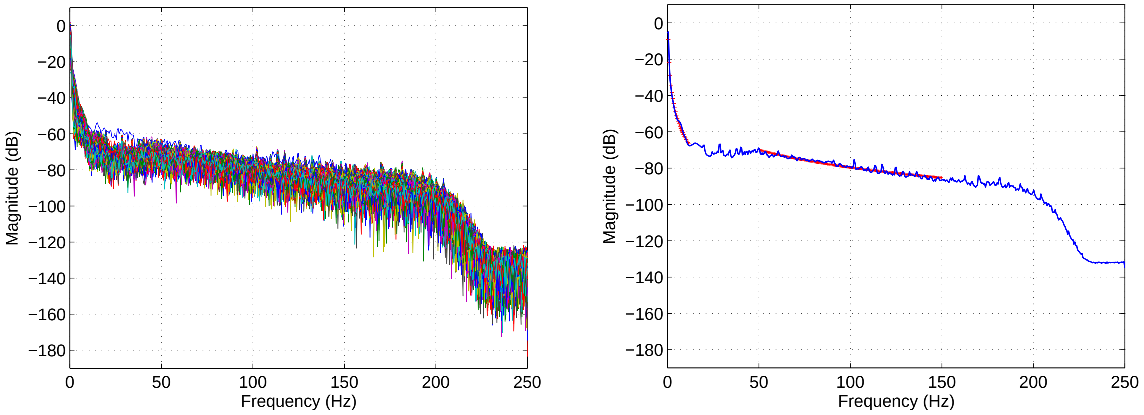
where the so-called *total spectral window* is given by

$$\overline{Q}(f) = \frac{1}{K}\sum_{k=0}^{K-1}Q_k(f), \tag{2.7}$$

and the spectral window of order k is given by

$$Q_k(f) = \left|\sum_{n=0}^{N-1} v_k[n]\exp(-j2\pi fn)\right|^2. \tag{2.8}$$

Note that Bronez (1992) compared the leakage, variance and frequency resolution for the DPSS MT method with that of a standard weighted overlapped segment averaging (WOSA). He found that the MT method always performed better than the WOSA for each of the measures, when the other two measures were required to be equal for both estimation methods.



(a)

(b)

Figure 2.2: Power spectrum estimates: (a) Overlay of the 648 power spectrum estimates and (b) averaged power spectrum estimate with linear fit of the $(0, 12]$ and $[50, 150]$ Hz intervals (red lines).

2.2.2 Application: Spectral analysis of real seismic noise recordings

The multitaper estimator is used to produce the spectrum estimate of the pure-noise recordings described in section 2.1, i.e., 6 seconds signals recorded simultaneously by 648 channels. Parameters for computation of the estimator are chosen to be: fast Fourier transform of length $N = 1024$ samples, time-bandwidth product $Nf_0 = 4$ and Thomson adaptive weighting as averaging scheme (sampling frequency F_s is 500 Hz). The power spectrum is estimated for each channel, which means that a total of 648 spectrum estimates are produced. Figure 2.2 (a) shows the overlay of the 648 estimated power spectra plotted on the $[0, F_s/2]$ interval. It appears that the difference in noise level is significant and can reach 50 dB re(W/Hz) at 1m. However, all power spectrum curves have the same basic shape. In order to obtain a better overview of the particular shape of this curve, the averaged estimate is computed and shown in Figure 2.2 (b). We are now able to identify three distinct regions in the averaged noise spectrum:

- Low frequency region: $[0, 12)$ Hz, with a magnitude attenuation of 125 dB, which contains most of the noise.

- Intermediate frequency region: $[12, 190)$ Hz, which is pretty flat and has a magnitude attenuation of 49 dB.

- High frequency region: $[190, 250]$ Hz, where we can observe the effects of the a hardware lowpass filter, with cut-off frequency at 200 Hz and attenuation of 370 dB/oct.

Where $[a, b) = \{x \in \mathbf{R} | a \leq x < b\}$.

Because seismic data of interest are normally within the $[0, 150]$ Hz interval, we choose to determine power-laws on the $(0, 12]$ Hz and the $[50, 150]$ Hz intervals. Power-laws are of the type $S(f) \sim f^n$. Generalizing slightly, we aim to perform a piecewise fit of our spectrum to the spectral shape $S(f) = 10^{n_2} f^{n_1}$, which is equivalent to

$$\log_{10}(S(f)) = n_1 \log_{10}(f) + n_2$$

where n_1 and n_2 are coefficients to be determined. Here, n_1 is the spectral decay coefficient of interest, while n_2 is a scaling parameter of little interest in the present chapter. The parameter n_1 can be estimated by plotting the region of interest on log-log scale, and using basic linear fitting tools to measure the slope. With such a procedure, we get the following estimates, drawn in Figure 2.2 (b) by the red curves:

- For the first region, $(0, 12]$ Hz:
$$n_1 = -4.15 \tag{2.9}$$
- For the second region, restricted to $[50, 150]$ Hz:
$$n_1 = -3.24 \tag{2.10}$$

Such spectral analysis has been repeated with other noise recordings, from the same seismic campaign and from other seismic campaigns. Similar characterization of the seismic noise has been repeatedly observed.

2.3 Clustering of power spectrum

The spectrum analysis performed in the previous section yielded a relatively wide envelope of the power spectrum estimates. They indeed differ with their location along the streamer. In order to visualize how the noise level is distributed along a streamer, we propose to cluster the power spectrum estimates using the average linkage cluster method. In this section, we first review this method and then apply it to the power spectrum estimates of section 2.2.2. Note that a more exhaustive review of cluster analysis is given by Romesburg (2004).

2.3.1 Average linkage cluster methodology

The first step consists of finding the similarity or dissimilarity between every pair of observations contained in the data set. To do so, we use the Euclidean distance as a measure of similarity between observations. Let us consider a *m*-by-*n* data matrix **Z**:

$$\mathbf{Z} = \begin{pmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \\ \vdots \\ \mathbf{z}_m \end{pmatrix} \tag{2.11}$$

where $\mathbf{z}_j = [x_{j,1} x_{j,2} \dots x_{j,n}]$ represents in the present case the j^{th} power spectrum estimate of dimension *n*. The Euclidean distance between the j^{th} and k^{th} power spectrum estimates is defined as

$$d_{j,k} = \sqrt{(\mathbf{z}_j - \mathbf{z}_k)(\mathbf{z}_j - \mathbf{z}_k)^T}. \tag{2.12}$$

The Euclidian distance between every pair of observations is calculated, and the result of this computation stored in a matrix commonly known as the distance or dissimilarity matrix where the element (j,k) corresponds to the distance between j^{th} and k^{th} observation in the original data set.

The second step consists of grouping the observations into a binary, hierarchical cluster tree. We consider here an agglomerative algorithm that takes the distance information generated at the first step and link pairs of observations that have the closest proximity into binary clusters, i.e, clusters made up of two observations. Then, these newly formed clusters are linked to each other and to other observations to create bigger clusters, using average distances $\overline{d_{j,k}}$ to determine the cluster order, until all the observations in the original data set are linked together in a hierarchical tree. The average distance $\overline{d_{j,k}}$ is defined as $\frac{1}{n_j n_k} \sum_{n=1}^{n_j} \sum_{q=1}^{n_k} d_{jp,kq}$, where n_j and n_k denote the number of observations in cluster j and k respectively, and jp and kq denote respectively the n^{th} observation in cluster j and the q^{th} observation in cluster k. This method is also referred to as average linkage clustering or UPGMA. The hierarchical cluster tree is stored into a linkage information matrix, where each row identifies a link between observations or formed clusters. More precisely, the first two columns denote the observations that have been linked, while the third column contains the average distance between these observations $\overline{d_{j,k}}$. Index values from 1 to *m* denote the *m* observations of the original data set. When two observations are grouped into a new cluster, it is assigned an index value starting from the value *m*+1, and increasing each time a new cluster is formed.

Finally, at the third step we determine where to cut the hierarchical tree into clusters. Many different ways to divide the data, natural or arbitrary, are possible. Here, we choose to create a constant number of clusters. The power spectrum estimates are time-dependent and their values might change with time. However, what we wish to emphasize is the relative difference between the power spectrum estimates, and not their values in itself. If we use an arbitrary threshold to cut the hierarchical tree, we obtain a set of clusters that would not always be able to emphasize the relative difference between the power spectrum estimates because of their time-dependence. Choosing to create a constant number of clusters will hopefully emphasize this desired feature.

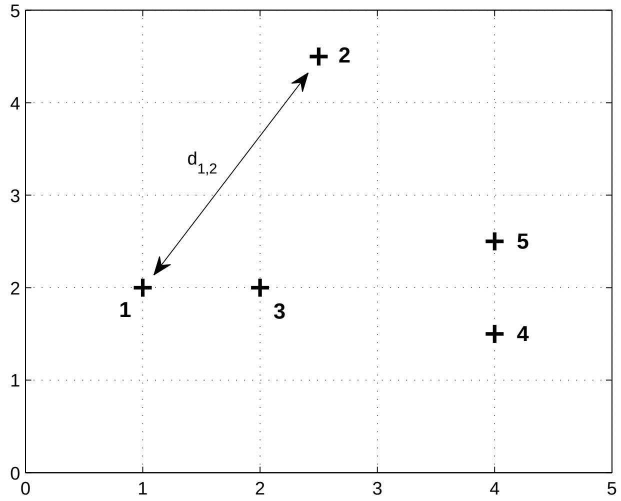


Figure 2.3: Example with 5 observations.

Example Let us consider a data set made of 5 observations of 2 dimensions each, shown in Figure 2.3 and stored in the 5-by-2 matrix **Z**

$$\mathbf{Z} = \begin{pmatrix} 1 & 2 \\ 2.5 & 4.5 \\ 2 & 2 \\ 4 & 1.5 \\ 4 & 2.5 \end{pmatrix}. \tag{2.13}$$

The distance matrix is then given by:

$$\mathbf{D} = \begin{pmatrix} 0 & 2.92 & 1.00 & 3.04 & 3.04 \\ 2.92 & 0 & 2.55 & 3.35 & 2.50 \\ 1.00 & 2.55 & 0 & 2.06 & 2.06 \\ 3.04 & 3.35 & 2.06 & 0 & 1.00 \\ 3.04 & 2.50 & 2.06 & 1.00 & 0 \end{pmatrix}. \tag{2.14}$$

The hierarchical cluster tree is given into a linkage information matrix **T**. Here, the first row shows that observations 4 and 5, which have the closest proximity, i.e., $\overline{d_{4,5}} = d_{4,5} = 1$, are grouped first, followed by observations 1 and 3 with $\overline{d_{1,3}} = d_{1,3} = 1$. The third row indicates that the next cluster is formed by observations 6 and 7, where observation 6 and 7 are the newly formed binary cluster created by the grouping of observations 4 and 5, and observations 1 and 3 respectively. The average distance between these two newly formed clusters is $\overline{d_{6,7}} = 2.55$. Finally, the last row shows that observations 8, the newly formed cluster made up of objects 6 and 7, and 2 are grouped together with an average distance $\overline{d_{2,8}} = 2.83$. A graphical illustration of the way the observations are grouped into a hierarchy of clusters is shown in Figure 2.4.

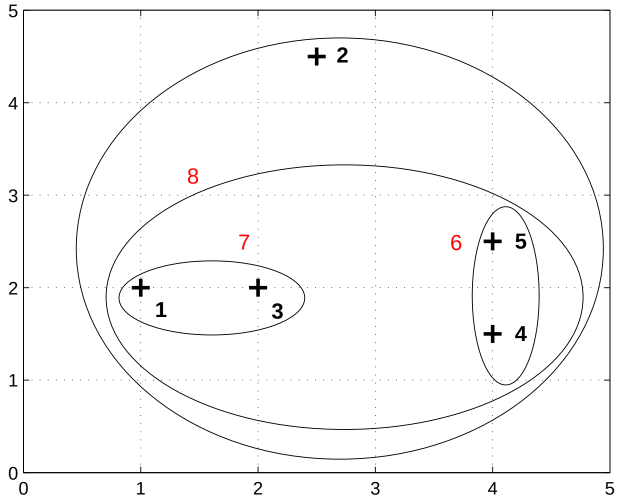


Figure 2.4: Hierarchy of clusters.

$$\mathbf{T} = \begin{pmatrix} 4 & 5 & 1.00 \\ 1 & 3 & 1.00 \\ 6 & 7 & 2.55 \\ 2 & 8 & 2.83 \end{pmatrix}. \tag{2.15}$$

In this example, we wish to create 2 clusters. For that purpose, we find the smallest distance at which the cut through the hierarchical tree leaves 2 clusters, namely 2.55. Thus, the first cluster contains the observations 1, 3, 4, and 5, and the second one contains the observation 2.

2.3.2 Application: Cluster of real pure-noise recordings

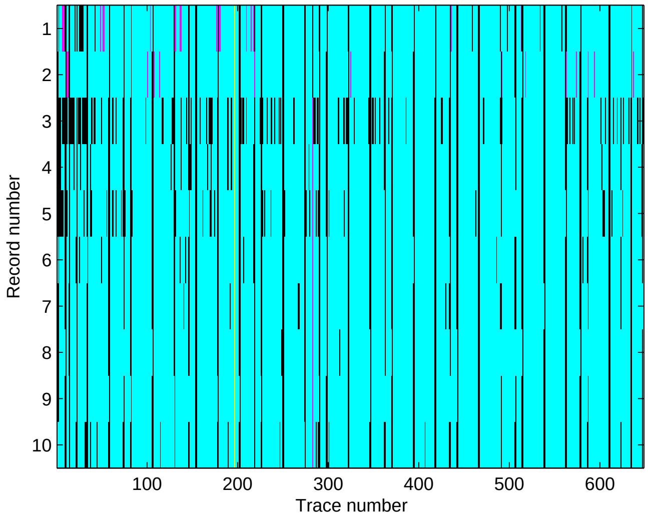
Let us consider two sets of pure noise recordings using two different types of streamer: a solid cable (Sentinel® solid) for the first data set and a fluid-filled cable (Sentinel® ALS fluid) for the second data set. The first data set is the one described in section 2.1.

Let us first apply the average linkage cluster method to the first data set. For each of the 10 records, we cluster a set of 648 power spectrum estimates within the frequency interval $[0, 50]$ Hz. The number of clusters to be created is set to 4. We present the results into a cluster matrix, shown in Figure 2.5 (a), that shows the repartition of the different clusters with respect to the trace and record numbers. Each color refers to one of the four clusters. Figure 2.5 (b) shows the power spectrum estimates displayed in their cluster color for the record number 7. The yellow cluster identifies a single trace with a much higher noise level, through all records. We can suspect that this channel is definitively not working as it should. The pink cluster clearly emphasizes that the trace 283 has a noise level higher than most of the other traces through all records, meaning that this trace does not work properly neither. This cluster also includes a couple of other traces recorded by the first three records. The blue cluster, that includes the majority of the estimates, identifies the estimates with the lowest noise level within $[0, 50]$ Hz. These estimates are typical estimates of background noise. Finally, the black cluster is made of traces whose power spectrum estimates are within the range of well-behaving channels, but in the upper part, i.e., with a noise level higher than typical background noise, especially within $[5, 25]$ Hz. The matrix of clusters in Figure 2.5 (a) exhibits some vertical, periodic black lines that correspond to the same (black) cluster. These lines are often two traces wide and are repeated every 24 traces, starting from traces 10-11 and up to traces 634-635. These locations match exactly with the location of the streamer birds, hanged between two consecutive channels, e.g., between channels 10 and 11. In a similar manner, we observe vertical lines every 72 channels, starting from traces 2-3 and up to traces 578-579, in addition to a line at trace 15 for most of the records. Once again, these locations match exactly with the location where acoustic units are hanged to the streamer.

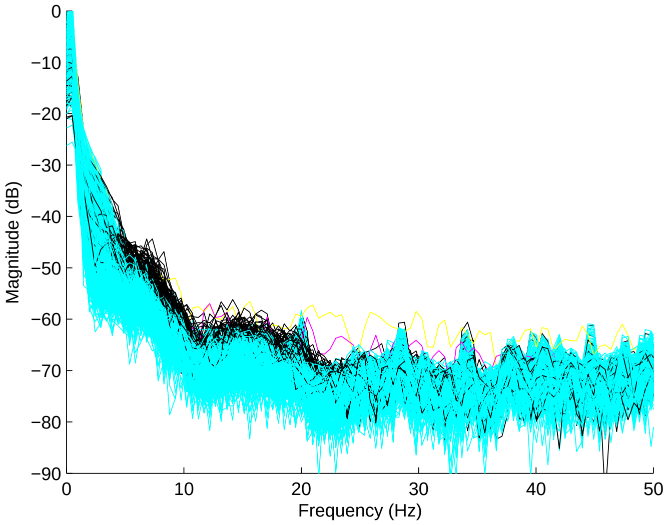
Therefore, the cluster matrix is a powerful and convenient tool to emphasize the noise repartition along the streamer. In our present example, the cluster matrix readily shows that the channels 197 and 283 are not working properly and that units hanged to the streamer generate additional noise that is recorded by the nearby channels in the frequency range $[5, 25]$ Hz. We have repeated the clustering application, setting the number of clusters to be created to 5 or 6. As a consequence, the blue cluster representing the background noise, is divided into subclusters without revealing any specific feature.

Let us then apply the average linkage cluster method to a second data set recorded by hydrophones laid along a fluid-filled cable. This set of data consists of 10 pure-noise recordings recorded consecutively under fairly acceptable weather conditions with 2 meters waves. Each recording contains data recorded simultaneously by 6 parallel streamers of 720 channels each. This means that we need to cluster 720 power spectrum estimates, restricted to the $[0, 50]$ Hz frequency range, for each streamer and each recording, which represents a total of 60 pure-noise records. The number of clusters to be created is still set to 4.

The cluster matrix is shown in Figure 2.6 (a) where data records number in y-ordinate are sorted by streamer, i.e., records 1 to 10 are recorded by the first streamer, records 11 to 20 are recorded by the second streamer, and so on. Figure 2.6 (b) shows the power spectrum estimates displayed in their cluster color for the record number 1. Here, the cluster matrix clearly reveals features that are similar to the one observed in the first example, with a blue cluster containing the background noise and periodic vertical black lines that match exactly with the location of the birds and the acoustic units along the streamer: Birds are hanged every 24 channels along the streamer, from traces 36-37 to traces 708-709, while the acoustic units are hanged every 240 channels along the streamer, at channel locations 239-240, 479-480 and 719-720. In addition, two other identified clusters represented in yellow and red, show that the noise level recorded by the



(a) Cluster matrix.

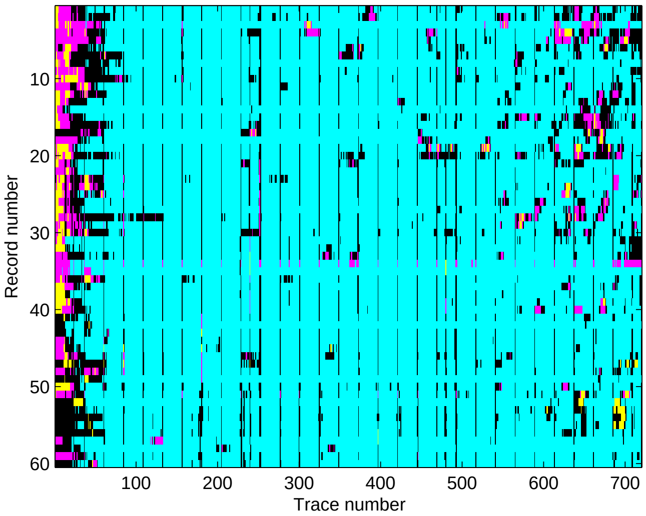


(b) Clustered power spectrum estimates.

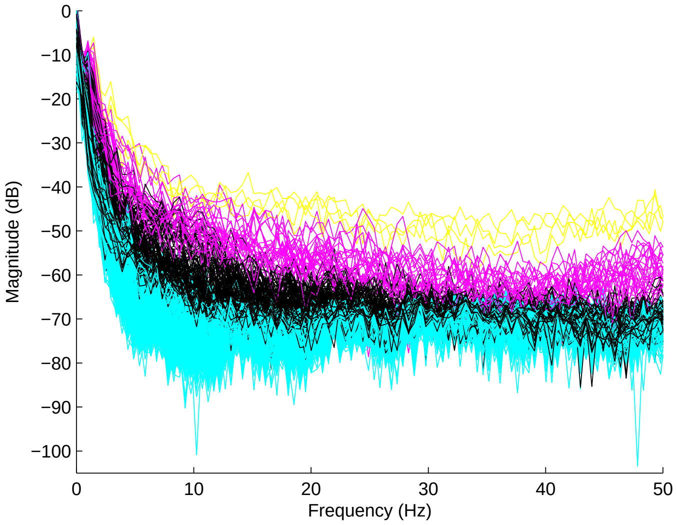
Figure 2.5: Clustering of 10 pure noise records.

first 25 channels is much higher than for any other channel, and that the noise level is progressively decreasing to reach its background level from channel 85. A higher noise level may also be recorded by the last 100 channels, especially for streamers 1 to 3.

Thus, through this second example, the cluster matrix appears to be a very useful tool to emphasize the noise repartition along the streamer and the detection of defective channels. These two examples emphasize that units hanged to the streamer generate additional noise recorded by the nearby channels. Furthermore, the clustering pattern is usually consistent from streamer to streamer and from record to record, meaning that it is valid at least during a short time window of few minutes and for all streamers. Finally, we observe a major difference between these two examples. When using fluid-filled cables, the hydrophones located at the head and at the tail of the streamers are in a much noisier environment than the ones located in the middle part. This is not the case when using solid cables.



(a) Cluster matrix.



(b) Clustered power spectrum estimates.

Figure 2.6: Clustering of 60 pure noise records.

2.4 Parzen Amplitude Estimates

The statistical distribution of the noise amplitude fluctuations, i.e., the probability density function (pdf) of the measured signal amplitudes, is an important tool for characterization of the data.

The histogram is the simplest and most widely used estimator of the pdf. Despite its widespread use, the histogram has several fundamental drawbacks. Basically, the histogram amounts to counting the number of amplitude levels that falls within a specified bin $x_0 + mh \leq x < x_0 + (m+1)h$, where x_0 is a chosen origin, h is a chosen bin width, and m is an integer. The histogram estimator is therefore a discontinuous and quantized piecewise constant function. Moreover, the histogram is sensitive to the choice of origin. The histogram estimator is particularly problematic when analyzing short data segments.

To overcome the many non-desirable properties of the histogram, we have chosen to estimate the probability density function by means of the so-called Parzen kernel estimator. The kernel estimator of density has good statistical properties, and it is well suited for the analysis of short data segments. Basically, the kernel estimator is constructed by placing a smooth and symmetric normalized function (a "kernel") with its origin at each data point. By summing this collection of normalized functions, we obtain a smooth and statistically consistent estimate of the probability density.

In general, the kernel estimator of the pdf at amplitude x, given N data samples $x_0, x_1, \ldots, x_{N-1}$, can be written as (Silverman, 1986; Wand and Jones, 1995)

$$\widehat{f}(x) = \frac{1}{N} \sum_{n=0}^{N-1} K_h\left(x - x_n\right). \tag{2.16}$$

Here, $K_h(\xi) \equiv (1/h)K(\xi/h)$, where $K(\xi)$ is the so-called smoothing kernel, and h is a scaling parameter that controls the degree of smoothing.

For equation 2.16 to be a valid estimator, one must require that $K(\xi) \geq 0$, $\forall \xi$, and $\int_{-\infty}^{\infty} K(\xi)$ $d\xi = 1$. In addition, it is reasonable to restrict $K(\xi)$ to the class of symmetric kernels, $K(-\xi) = K(\xi)$. In practice, the resulting estimator is not very sensitive to the detailed shape of the kernel $K(\xi)$. The estimate however depends strongly on the value of the parameter h.

A good standard choice of the smoothing kernel is the Gaussian $K(\xi) = (1/\sqrt{2\pi})$ $\exp(-\xi^2/2)$. The choice of the smoothing parameter h is non-trivial, and several techniques exist for estimating a value of h that obeys some optimality criterion. In our study, we have chosen the value of h by minimizing the integrated mean-squared error with referral to a standard family of distributions, i.e., the Gaussian distribution.

It is straightforward to show that the expected value of the basic kernel estimator in equation 2.16 can be written as (Silverman, 1986; Wand and Jones, 1995)

$$E\left\{\widehat{f}(x)\right\} = K_h(x) * f(x), \tag{2.17}$$

where $*$ denotes the convolution operator, and $f(x)$ is the true amplitude probability density function (pdf). Thus, we may interpret the pdf estimator in equation 2.16 as a smoothed version of the true (but unknown) underlying pdf. As a consequence, the kernel estimate of the pdf does not exhibit the unphysical discontinuities present in the naive histogram based pdf estimator.

The bias of the estimator can readily be approximated by

$$b\left\{\widehat{f}(x)\right\} \equiv E\left\{\widehat{f}(x)\right\} - f(x) \approx \frac{\sigma_K^2}{2} h^2 f''(x), \tag{2.18}$$

where $\sigma_K^2 \equiv \int_{-\infty}^{\infty} \xi^2 K(\xi) d\xi$. Assuming that the data are statistically independent, a useful approximation for the estimator variance is given by Silverman (1986),

$$\text{var}\left\{\widehat{f}(x)\right\} \approx \frac{\mathcal{E}_K}{Nh} f(x), \tag{2.19}$$

where $\mathcal{E}_K \equiv \int_{-\infty}^{\infty} K^2(\xi) d\xi$ is the energy of the kernel.

A standard criterion for choosing h, is to seek the value that minimizes the integrated mean-square-error (IMSE) defined as

$$\text{IMSE}\left\{\widehat{f}(x)\right\} = \int_{-\infty}^{\infty} \left[\text{var}\left\{\widehat{f}(x)\right\} + b^2\left\{\widehat{f}(x)\right\}\right] dx. \tag{2.20}$$

Inserting equations 2.18 and 2.19 into equation 2.20, and assuming that the true density $f(x)$ is Gaussian with variance σ^2, one readily obtains that the IMSE-optimal value of h is given by

$$h^* = \left[\frac{8\pi^{1/2}\mathcal{E}_K}{3\sigma_K^4 N}\right]^{1/5} \sigma. \tag{2.21}$$

If the smoothing kernel $K(\xi)$ is Gaussian, it is easy to show that we obtain $h^* = (4/3)^{1/5} N^{-1/5} \sigma$.

2.4.1 Moments of the Amplitude Probability

Since the full pdf contains statistical information at a detailed level, it is often convenient to consider higher order moments of the data. The coefficient of skewness (or simply the skewness) is the dimensionless third order central moment defined by $\Gamma_3 = \mu_3/\mu_2^{3/2}$. The kurtosis (or coefficient of excess) is the dimensionless fourth order central moment defined by $\Gamma_4 = \mu_4/\mu_2^2 - 3$. Here, the central moment of order n for a stochastic variable X is defined by $\mu_n = E\left\{[X - E\{X\}]^n\right\}$. Symmetric densities have $\Gamma_3 = 0$, while densities skew to the right (left) have $\Gamma_3 > 0$ ($\Gamma_3 < 0$). The kurtosis is a flatness measure relative to the Gaussian. If $\Gamma_4 < 0$, the density is termed platykurtic, and the density is flatter than the Gaussian close to the maximum.

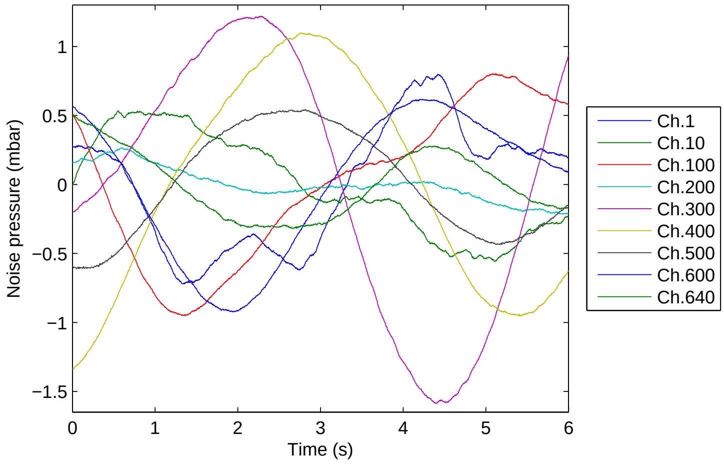


Figure 2.7: A few noise records with respect to time.

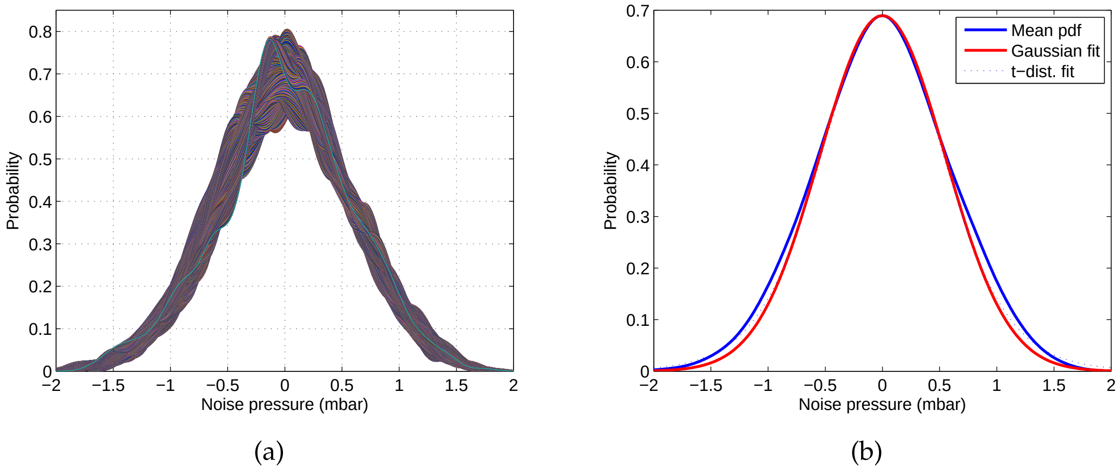
2.4.2 Application: Amplitude Probability Density Estimates

Focus on a single data record

Consider the data set described in Section 2.1. Figure 2.7 shows a few time series from channels located far apart along the cable, more precisely, channels 1, 10, 100, 200, 300, 400, 500, 600 and 640, which span over 8.1 kilometers in this case. The x-axis represents the time in seconds, while the y-axis represents the noise pressure in millibar. We observe that the very first ones (channels 1 and 10) and the last one (channel 640) are corrupted with low amplitude noise. However, all the time series present the same slow up and down fluctuations, whose amplitude can reach significant values.

To go further into the characterization of the noise amplitude, let us now use the Parzen kernel estimator to estimate its statistical distribution. The smoothing kernel is chosen to be the Gaussian function $K(\xi) = (1/\sqrt{2\pi})\exp(-\xi^2/2)$ and the smoothing parameter is chosen as $h = 0.102$, as described in Section 2.4.

We can now estimate the distribution of the noise with respect to time. Since noise measurements are available for 3000 time samples, then 3000 probability density functions are estimated based on data recorded by 648 channels. The overlay of the 3000 probability density estimates is shown in Figure 2.8 (a). If we look at distribution estimates one after the other, we observe that they change very progressively over time. So, even if differences among noise distributions are not negligible, they remain reasonable, and considering the averaged probability density function in order to characterize seismic noise distribution is acceptable. Figure 2.8 (b) shows the averaged probability density function together with fitted Gaussian and t-distributions. The averaged pdf is slightly skewed to the right and none of the two standard distributions can fit with this



(a)

(b)

Figure 2.8: Probability density functions: (a) Overlay of the 3000 pdf estimates, (b) Averaged pdf estimate (solid thick) shown together with Gaussian fit (solid, red) and t-distribution fit (dashed line).

skewness.

Extension to 10 data records

The previous estimation of the noise amplitude distribution is repeated with data recorded under the same conditions few seconds or minutes later. The smoothing parameters remain unchanged. Averaged pdfs resulting from estimations for each data record are plotted in Figure 2.9. All pdf curves have a similar shape, but they differ in the maximum amplitude and in their flatness.

Skewness and kurtosis coefficients, third and fourth order central moments respectively, describe the distortion of pdfs relatively to the normal distribution. They are calculated as described in Section 2.4.1 for each noise record and the results are shown in Table 2.1. Distribution resulting from the 1^{st} record presents the biggest right-skewed value, as observed in Figure 2.8 (b). Distributions resulting from the 2^{nd}, 5^{th} and 10^{th} records present negative kurtosis numbers whereas the ones resulting from the 6^{th} and 8^{th} records present small, positive kurtosis numbers. In Figure 2.9, they are respectively the flattest and more peaked curves of the graph. All skewness numbers fall into the $[-0.05, 0.14]$ interval, which indicates that distributions are almost symmetric. Kurtosis numbers fall into the $[-0.44, 0.15]$ interval, which means that some distributions are slightly flatter than the normal distribution. The coefficient values confirm the impression given by the figures, and therefore, they allow us to conclude that distributions of these seismic noise records are almost symmetric and slightly platykurtic.

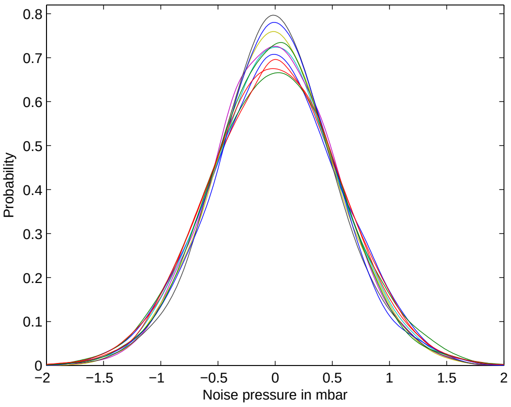


Figure 2.9: Overlay plot of the mean pdf of 10 noise recordings.

X	Rec 1	Rec 2	Rec 3	Rec 4	Rec 5	Rec 6	Rec 7	Rec 8	Rec 9	Rec 10
Γ_3	0.14	-0.04	0.13	-0.05	0.11	-0.05	0.00	-0.02	0.08	0.11
Γ_4	-0.02	-0.12	-0.03	0.03	-0.20	0.11	-0.01	0.15	-0.10	-0.44

Table 2.1: Estimated skewness and kurtosis coefficients for 10 noise records.

2.5 Summary

We first analyzed some pure noise recordings using solid streamers. The multitaper power spectral estimate of the noise processes yields a composite power law spectrum. The Parzen kernel estimation and higher-order moment analysis of the noise amplitudes, shows that the noise is slightly platykurtic and almost symmetric.

We then applied the average linkage cluster method to pure noise recordings using either solid or fluid-filled streamers. These examples have revealed that units hanged to the streamer generate additional noise recorded by the nearby channels. They also have shown that when using fluid-filled cables the hydrophones located at the head and at the tail of the streamers are in a much noisier environment than the ones located in the middle part, while such tendency is not observed when using solid cables. Furthermore, it appeared that the cluster matrix was a very useful tool to emphasize the noise repartition along the streamer and the detection of defective channels.

Finally, because the noise characteristics are changing slowly and they are rather consistent between streamers, then they can be considered as homogeneous and stationary during at least a short time window of a few minutes.

Chapter 3

A novel enhanced stacking method

3.1 Introduction

Enhancement of signals embedded in background noise is an important task in seismic data processing. By improving the quality of the seismic images, the results of subsequent processing or interpretation are much easier. The quality of the seismic image can be improved by means of different methods, and stacking is one of them.

Conventional stacking is performed by averaging a set of normal moveout (NMO) corrected or migrated traces from different shot records with a common reflection point, usually the common mid-point (CMP), that contributes equally to the stacked trace. Thus, let us denote by $\mathbf{X} = [\mathbf{x}_1, ..., \mathbf{x}_M]$ an input gather that consists in M traces of N samples $\mathbf{x}_k$, $k = 1, .., M$. If $f_d(t)$ is the fold of the input gather at a given traveltime t ($f_d(t) \leq M$), then the stacked trace $\bar{\mathbf{x}} = [\bar{x}(1), ..., \bar{x}(N)]^T$ resulting from that input gather $\mathbf{X}$ is obtained from

$$\bar{x}(t) = \frac{1}{f_d(t)} \sum_{i=1}^{f_d(t)} x_i(t), \qquad t = 1, ..., N. \tag{3.1}$$

In that way, the noise is attenuated and the quality of the stacked trace has been improved. This method is however optimal in the sense that the signal-to-noise ratio (S/N) is maximized only if all traces have identical amplitudes, identical S/N and are coincident (or correlated), and if the noise is stationary and statistically independent of the noise of any other trace and of the signal (Mayne, 1962; Neelamani et al., 2006; Robinson, 1970).

A NMO-corrected or migrated seismic record usually contains signal components that are not properly aligned and noise components that are correlated or not necessarily normally distributed. In order to improve the stacking technique, lots of solutions based on weighted stacking have been proposed where the weights are derived according to different optimality criteria. If we denote by $w_i(t)$ the normalized weight associated to the i^{th} trace at a given traveltime t, then the weighted stacked trace $\bar{\mathbf{x}}$ is given by

$$\bar{x}(t) = \sum_{i=1}^{f_d(t)} w_i(t) x_i(t), \qquad t = 1, ..., N. \tag{3.2}$$

One of the most direct approaches has been presented by Robinson (1970). He suggested a statistical approach where optimal weights are derived from estimated signal to noise ratio for each trace. The estimation technique has to be robust for the stacked data to be significantly improved, which is not the case with the technique proposed by Robinson. According to him, the method does not produce any significant improvement on real data.

In order to better reflect any local dependency, some time-dependent weights derivation methods have been proposed. Pruett (1982) used weights that are inversely function of the distance between each trace amplitude and a reference trace, and also function of the amplitudes variance at a given traveltime. Watt and Bednar (1983) and Haldorsen and Farmer (1989) proposed a trimmed mean stack where extreme trace amplitudes are not included in the averaging process. Grion and Mazzotti (1998) presented two weighting schemes based either on the local S/N estimation using singular value decomposition eigenvalues, or the local coherency. Trickett (2007) used a maximum-likelihood estimation stacking where the stacked value is function of the estimated probability distribution of the noise. When the noise is erratic, this method yields a much cleaner stack, but when the noise is Gaussian, this method is identical to the conventional stack. More recently, Rashed (2008) has developed smart stacking that computes the alpha-trimmed mean of amplitudes and evaluates weights as a function of the inverse distance between amplitude and alpha-trimmed mean, for each traveltime sample of the CMP. This method does not perform very well as soon as seismic reflections are not perfectly aligned. Liu et al. (2009) proposed that weights represent a measure of the local correlation between each trace of the CMP and a conventional stacking reference trace. The idea is very appealing because it overcomes many problems such as signal misalignment. The quality of the final stacked data is however dependent on the quality of the reference trace. In addition, this method has the major drawback of not preserving physically meaningful amplitudes because of the weight normalization. Finally, Bayless and Brigham (1970) and Crump (1974) introduced the use of discrete Kalman filters for seismic data processing with application to the deconvolution problem. However, no application to stacking have been published so far.

In this chapter, we propose a new method that improves Liu's method by introducing two other possible reference traces and by developing a consistent way of normalizing the weights. We derive traveltime-dependent weights that are measures of the local correlation between traces of the CMP and a preprocessed reference trace. Weights are then normalized to yield meaningful amplitudes and eventually weighted traces are stacked. Three possible reference traces are investigated and tested: (1) stacked trace resulting from conventional stacking, (2) stacked trace using S/N-estimation, and (3) stacked trace output by Kalman filters. We begin, in Section 3.2, with the presentation and derivation of three possible so-called reference trace methods. We show that they are self-content and they can be used as independent stacking methods. They can, in addition, be used for the calculation of local-correlation weights and the derivation of en enhanced weighted stacking method, as in Section 3.3. Finally, the different presented stacking methods are tested on both synthetic and real marine seismic data in Section 4.4. We show that this new stacking method is very robust, it significantly improves the final stacked section and it can reveal signal components that would not be discernible with conventional stacking.

3.2 Reference traces

In this section, we present three different stacking methods, that are candidates for the reference stack required for the derivation of the enhanced local correlation stacking method. Two of them are well-known methods, conventional and S/N-estimation based stacking, while the third one, that uses a Kalman filter, is fairly new.

3.2.1 Conventional stacking

In conventional stacking, all the traces contribute equally to the stacked trace. If we denote by $f_d(t)$ the fold of the stack at a given traveltime t ($f_d(t) \leq M$), then the stacked trace $\bar{\mathbf{y}} = [\bar{y}(1), ..., \bar{y}(t), ..., \bar{y}(N)]^T$ is obtained from

$$\bar{y}(t) = \frac{1}{f_d(t)} \sum_{i=1}^{f_d(t)} x_i(t), \qquad t = 1, ..., N. \tag{3.3}$$

3.2.2 Signal-to-noise ratio estimation

The signal-to-noise ratio of the traces contained in a CMP gather often differs from each other since it usually decreases with increased offset distance. Therefore, we can improve the quality of the stacked trace by using some S/N information and more precisely, by using a weighted average where each weight is directly proportional to the signal-to-noise ratio of each individual trace. For that purpose, we need to estimate the values of the noise variance and the signal norm for each trace. Rietsch (1980) estimates S/N by comparing pairs of traces. It is a good technique but it requires lot of computer operations. The leave-me-out (LMO) method (Neelamani et al., 2006) is only reliable when signals are aligned with constant amplitude. Here, S/N estimation is only part of a preprocessing step to derive a reference trace and therefore, the accuracy of the S/N estimates is not as crucial as in the aforementioned methods. For this reason, we choose to derive rather good estimates of the noise variance and the signal norm, but which are reasonable from a computational point of view.

We achieve the noise variance estimation by providing a reference trace $\bar{\mathbf{x}}$ processed, e.g., according to conventional stacking. For each trace $\mathbf{x}_i$ of the CMP gather, we calculate a scaling factor defined as $\eta_i = \frac{\max(|\mathbf{x}_i|)}{\max(|\bar{\mathbf{x}}|)}$ that specifically scales the reference trace for the input trace. By subtracting it from the input trace, we get a rough estimate of the noise vector. Then, we select a traveltime interval where the data fold is maximum (i.e., nonzero data are recorded for all traces), and compute the variance of the noise series over the selected traveltime interval as

$$\sigma_i^2 = E[|\mathbf{x}_i - \eta_i \bar{\mathbf{x}}|^2] - (E[\mathbf{x}_i - \eta_i \bar{\mathbf{x}}])^2 \qquad \forall i = 1, \ldots, M. \tag{3.4}$$

This way of estimating the noise variance is quite robust if the traveltime interval where data are recorded for all nonzero traces is long enough, which is usually the case. The leave-me-out method (Neelamani et al., 2006) has been tested as well with a simulation on synthetic data, but because of some signals were not perfectly aligned across the traces, it yielded poor estimates.

To estimate the signal power we calculate the autocorrelation of each trace at zero lag and use the noise variance estimates. We adopt the following trace model

$$\mathbf{x}_i = a_i \mathbf{s} + \boldsymbol{\nu}_i \qquad \forall i = 1, \ldots, M, \tag{3.5}$$

where a_i is the signal amplitude of the i^{th} trace, $\mathbf{s}$ is the signal vector and $\boldsymbol{\nu}_i$ is the corrupting noise vector of the i^{th} trace. Assuming that the noise is zero-mean and uncorrelated with the signal component, then the autocorrelation of the i^{th} trace at zero lag is given by

$$r_i = E\{|\mathbf{x}_i|^2\} = a_i^2 |\mathbf{s}|^2 + \sigma_i^2 \tag{3.6}$$

and therefore, an estimate of the signal power is obtained by

$$\rho_{s,i}^2 = r_i - \sigma_i^2. \tag{3.7}$$

S/N-estimation weights are then defined as

$$\omega_i = \frac{\rho_{s,i}^2}{\sigma_i^2} \tag{3.8}$$

and the S/N-estimation based stacked trace $\bar{\mathbf{y}}$ is obtained from

$$\bar{y}(t) = \frac{1}{\sum_{i=1}^{f_d(t)} \omega_i} \sum_{i=1}^{f_d(t)} \omega_i x_i(t), \qquad t = 1, ..., N. \tag{3.9}$$

Note that ω_i must not be mixed up with w_i defined in equation 3.2.

3.2.3 Stacking using Kalman filter

The Kalman filter is a very powerful recursive estimation algorithm adapted to nonstationary processes. The idea is to consider values of each trace at a given traveltime t as a dataset, and use the Kalman filter to extract the signal value $s(t)$ hidden in that dataset at that given traveltime. However, for the Kalman filter to be implemented, we need to determine the noise variance and the amplitude decay of each trace.

Let us consider the trace model given in equation 3.5 and let us assume that signal and noise are uncorrelated and that noise sequences are mutually uncorrelated. In order to determine the amplitude factor a_i, we calculate the autocorrelation of each trace at zero lag. According to the previous assumptions, it yields

$$r_i = E\{\mathbf{x}_i \mathbf{x}_i\} = a_i^2 r_s + \sigma_i^2 \qquad \forall i = 1, \dots, M, \tag{3.10}$$

where r_s is the autocorrelation of the signal **s** and σ_i^2 is the noise variance. Let us use the noise variance estimates $\widehat{\sigma_i^2}$ described in the Section 3.2.2. These estimated values are reasonably accurate but they are usually scaled by an unknown multiplicative factor. Therefore, we need to rescale the noise variances so that $\sigma_i^2 \equiv \frac{\widehat{\sigma_i^2}}{\max(\widehat{\sigma_i^2}/r_i)}$. Thus, if we arbitrarily set $a_1 = 1$, then $r_s = r_1 - \sigma_1^2$ in equation 3.10 and the amplitude factors a_i, $i = 1, ..., M$ are defined by

$$a_i = \sqrt{\frac{r_i - \sigma_i^2}{r_1 - \sigma_1^2}} \qquad \forall i = 1, \dots, M. \tag{3.11}$$

In order to determine the seismic signal $s(t)$ at a given traveltime, we assume it to be constant through all traces. In that way, the i^{th} trace consists in a seismic signal scaled in amplitude, $a_i s_i(t)$, and corrupted by noise $\nu_i \sim \mathcal{N}(0, \sigma_i^2)$. I.e., we use the following model at any traveltime t:

$$\text{State equation:} \quad s_i(t) = s_{i-1}(t) \tag{3.12}$$

$$\text{Observation equation:} \quad x_i(t) = a_i s_i(t) + \nu_i(t) \tag{3.13}$$

$$\text{Observation noise covariance matrix:} \quad Q_{\nu,i} = \sigma_i^2. \tag{3.14}$$

The equations of the Kalman filter are as follows (Hayes, 1996):

Initialization:

$$s_1(t) = x_1(t) \tag{3.15}$$

$$P_1(t) = [x_1(t) - y(t)]^2, \tag{3.16}$$

Iterative computation: For $i = 2, \ldots, M$, compute the following iterations

$$\hat{s}_{i|i-1}(t) = s_{i-1}(t) \tag{3.17}$$

$$P_{i|i-1}(t) = P_{i-1}(t) \tag{3.18}$$

$$k_i = P_{i|i-1}(t)a_i[a_i^2 P_{i|i-1}(t) + Q_{\nu,i}]^{-1} \tag{3.19}$$

$$\hat{s}_i(t) = \hat{s}_{i|i-1}(t) + k_i[x_i(t) - a_i\hat{s}_{i|i-1}(t)] \tag{3.20}$$

$$P_i(t) = [1 - k_i a_i]P_{i|i-1}(t). \tag{3.21}$$

Note that the Kalman filter needs quite a few iterations to converge, so in practice, if the data fold at a given travel time is too small (typically less than 6), then we use the value of the reference trace $\bar{\mathbf{y}}$ at that traveltime in the output trace.

3.3 Enhanced local correlation stacking method

3.3.1 Unscaled local correlation

Let us denote by $\mathbf{X}$ a NMO-corrected CMP gather that consists in M traces of N samples and by $\mathbf{x}_i$ the i^{th} trace of that gather. In addition, denote by $\bar{\mathbf{y}}$ a reference trace that has to be correlated with $\mathbf{x}_i$. This reference trace can be any of the ones presented in the previous section. Given two vectors $\mathbf{x}_i$ and $\bar{\mathbf{y}}$, a possible measurement of local similarity between two signals is the sliding-window unscaled local correlation coefficient, defined as

$$\gamma_i(t) = \sum_{k=t-L_w/2}^{t+L_w/2} x_i(k)\bar{y}(k), \tag{3.22}$$

where L_w is the window length.
In the particular cases where $t < \frac{L_w}{2}$ and $t > N + \frac{L_w}{2}$ where N is the length of shot record, a few signals components $x_i(k)$ and $\bar{y}(k)$ remain undefined. We overcome this problem by replicating the defined signals contained in the window.

If a trace is locally very similar to the reference trace, the local correlation coefficient is high and should therefore be given a larger weight. On the contrary, a trace that locally differs significantly from the reference trace does not have a positive contribution to the stacked trace and must therefore be discriminated. In that case, the local correlation coefficient is small and so should the weight. Therefore, coherent signals yield larger correlation value compared to incoherent noise. In addition, we use an unscaled coefficient so that it exhibits both the correlation and the signals amplitudes. Given that amplitudes of seismic reflectors are usually larger than noise, then the choice of unscaled coefficient allows detection of signals that are not perfectly aligned.

3.3.2 Weight determination

Once a reference trace $\bar{\mathbf{y}}$ has been estimated and the local correlation coefficients $\gamma_i(t)$ are calculated, we are ready for the determination of the weights. As expected, we notice that coherent signals yield large correlation values compared to incoherent noise. It is actually these relative differences among correlation values that we want to preserve when deriving weights. Thus, we have set up a simple and robust algorithm that detects traveltime occurrences $T_G = \{\tau_1, \ldots, \tau_{N_G}\}$ where the sum of local correlation coefficients $\sum_{i=1}^{M} \gamma_i(t)$ has local maxima, so that our weight normalization scheme eventually satisfies

$$\begin{cases} \sum_{i=1}^{f_d(t)} w_i(t) = 1 & \text{if } t \in T_G, \\ \sum_{i=1}^{f_d(t)} w_i(t) < 1 & \text{otherwise.} \end{cases} \tag{3.23}$$

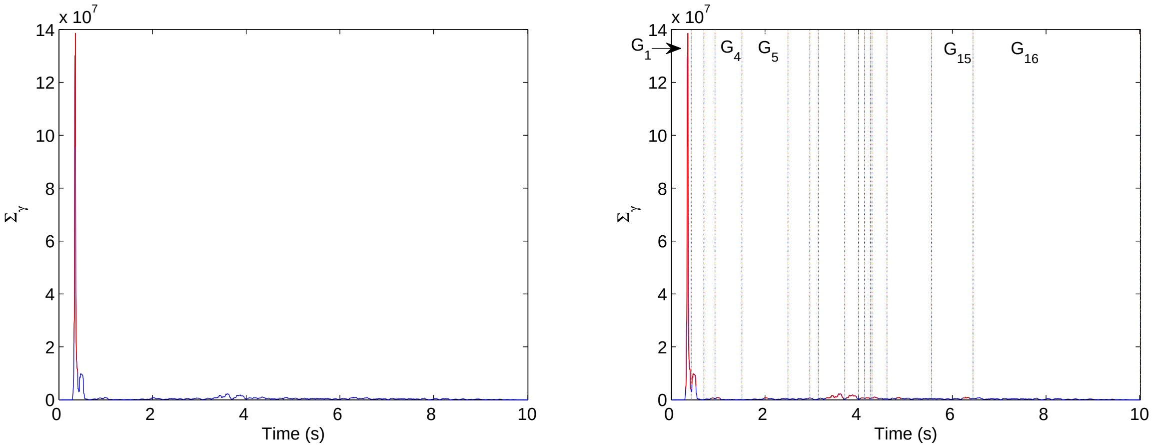
Detection of large amount of coherent signals

In case of coherent signals, correlation values $\gamma_i(t)$ will be large for several traces at a given traveltime t. Therefore, in order to detect the traveltime occurrences of coherent signals, we first calculate the sum of correlation values over channels, and then use an iterative algorithm that separates coherent from incoherent values:

- Calculate $\mathbf{\Sigma}_\gamma(t) = \sum_{i=1}^{M} \gamma_i(t)$. Note that $\mathbf{\Sigma}_\gamma(t)$ is also the Radon transform (Helgason, 1999) when projected to the traveltime axis. Coherent signals yield local maxima in the $\mathbf{\Sigma}_\gamma(t)$ series.
- Perform the iterative computational procedure that detects local maxima:
  1. Initialization: Set $\epsilon_0 = 0$ and find $\tau_1 = \arg\max_t \mathbf{\Sigma}_\gamma(t)$ and the interval $I = \{\tau_1 - N_1, \ldots, t0, \ldots, \tau_1 + N_2\}$ containing τ_1 and where (N_1, N_2) are defined so that: $\mathbf{\Sigma}_\gamma(\tau_1 - N_1) > \mathbf{\Sigma}_\gamma(t)$ and $\mathbf{\Sigma}_\gamma(\tau_1 + N_2) > \mathbf{\Sigma}_\gamma(t)$, $\forall t \notin I$. I.e., once the time occurrence τ_1 of the maximum is identified, adjacent time values are added to form an ensemble until the next local maximum value is found.
  2. Calculate the mean and standard deviation of $\mathbf{\Sigma}_\gamma$ for samples that do not belong to I, i.e., $m = E\{\mathbf{\Sigma}_\gamma(t)\}$ and $\sigma = \sqrt{E\{(\mathbf{\Sigma}_\gamma(t) - m)^2\}}, \forall t \notin I$, respectively.
  3. Calculate the threshold value $\epsilon = m + \delta\sigma$, where typical value of δ is about 3 – 3.5.
  4. Update values of interval I so that $I = \{t, \mathbf{\Sigma}_\gamma(t) > \epsilon\}$.
  5. If $|\epsilon - \epsilon_0| < \alpha\epsilon_0$, stop iterations. Else, $\epsilon_0 = \epsilon$ and go back to step 2. Here, a typical value of α is about 0.01 – 0.1;
- Now, I contains traveltime values of N_G local maxima of $\mathbf{\Sigma}_\gamma(t)$, denoted by $\tau_1, \ldots,$ τ_{N_G}, and their neighborhood. Use I to identify $\tau_1, \ldots, \tau_{N_G}$, and then grid the whole traveltime span so that each subset G_i, $i = 1, \ldots, N_G$ contains a single local maximum at time τ_i.

In the case where the threshold value ϵ is small, it may happen that several local maxima of very different amplitudes end up in the same subset, typically in the first or second subset. If so, one can choose to systematically re-apply this procedure to these subsets, constrained such that several local maxima should be detected in each of them.

An illustration of this iterative procedure is shown in Figure 3.1. This example is applied to real marine data and uses the parameter value $\alpha = 0.05$ and $\delta = 2.5$. In both panels of the figure we show $\mathbf{\Sigma}_\gamma$ as a function of traveltime where the red-colored parts of the curves correspond to the traveltime values that belongs to the ensemble I: (a) At the initialization step, (b) At the end of the iterative process, with subsets $G_1, \ldots, G_{16}$ delimited by vertical dotted lines.



(a) At the initialization step.

(b) At the end of the iterative process.

Figure 3.1: Determination of the ensemble I (made of traveltimes that belong to the red part) and traveltime subsets G_i.

Weight calculation

We apply the following normalization equation to the local correlation coefficients.
For all $i = 1, ..., M$, and for each subset $G_k,\ k = 1, \ldots, N_G$,

$$w_i(t) = \frac{\gamma_i(t)}{\max_{t \epsilon G_k} \mathbf{\Sigma}_\gamma(t)}, \qquad \forall t \in G_k. \tag{3.24}$$

As a result, the amplitude of coherent signals is preserved while the amplitude of incoherent signals is attenuated.

Norm smoothing/equalization

Because the detection process of coherent signal may not lead to a homogeneous stacked image, it may be necessary to smooth locally the norm of the stacked traces. We use a nonlinear smoothing filter that bridges small gaps in norm content and reduces sharp transitions in signal amplitude among stacked traces (Gonzalez and Woods, 2008). The smoothing filter is described as follows:

- Span the post-stack seismic section with a sliding time window. Our window contains 40 time samples for each of the CMP stacked traces and the overlap between windows is set to 33 %.
- Calculate the norm of each trace segment:

$$e_j = ||\bar{\mathbf{x}}_j^S|| \qquad \text{for } j = 1, \ldots, M_C, \tag{3.25}$$

  where M_C indicates the number of CMP stacked traces and superscript S stands for the current window.
- Smooth the norm curve: For each of the norm values, get a median value by applying a one-dimension median filter of order L, calculate the standard deviation $\sigma_{j,S}$ of a dataset containing L norm values and centered around that norm value, remove the norm value if it does not fall into the interval median value $\pm\frac{\sigma_{j,S}}{3}$, and proceed to a linear interpolation of the norm vector to replace the removed values. Thus, we obtain a smoothed norm value denoted by $\overline{e_j}$.
- Rescale the norm of each individual trace segment:

$$\overline{\bar{\mathbf{x}}_j^S} = \frac{\overline{e_j}}{e_j}\bar{\mathbf{x}}_j^S. \tag{3.26}$$

An example of the repartition of the norm content along the CMP stacked traces before (blue curve) and after (red curve) smoothing is shown in Figure 3.2. Here, the parameter L is set to 65 samples. We readily observe that sharp transitions in signal amplitude have disappeared after smoothing. Application to the output of our enhanced local correlation stacking method using the S/N-estimation as reference trace is shown in Figure 3.3. The white stripes caused by smaller norm in Figure (a) have been filled in to become almost invisible in (b).

3.3.3 Summary

The presented stacking method is summed up and generalized to the j^{th} CMP gather as follows:

- Estimate a reference trace $\bar{\mathbf{y}}_j$ using either conventional stacking, S/N estimation based stacking or Kalman based stacking.
- Calculate the local correlation coefficients $\gamma_{i,j}(t),\ i = 1, ...M.$
- Run the iterative coherent signal detection procedure.

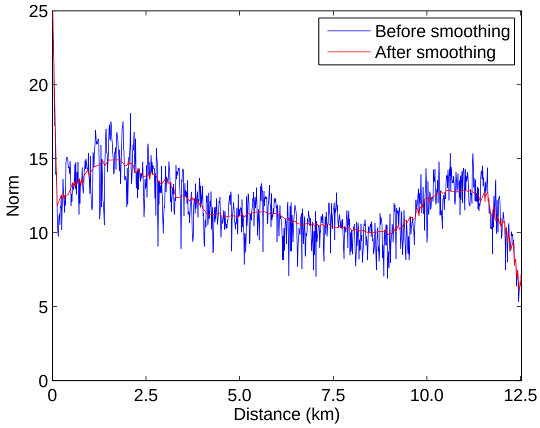


Figure 3.2: Example of smoothing: Before smoothing (blue curve) and after smoothing (red curve).

- Calculate the weights $w_{i,j}(t)$.
- Calculate the final stacked trace $\bar{\mathbf{x}}_j$ from the j^{th} CMP gather, according to $\bar{\mathbf{x}}_j(t) = \sum_{i=1}^{f_d(t)} w_{i,j}(t)x_{i,j}(t)$.
- Apply local smoothing if necessary.

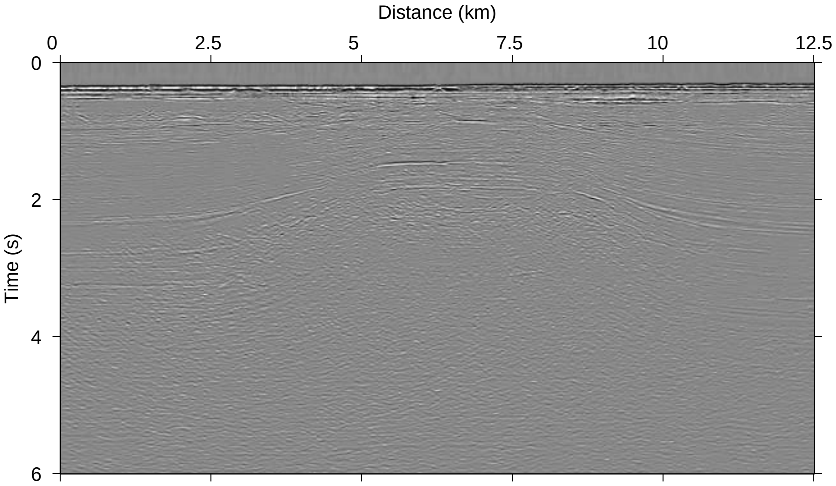
3.4 Applications

In this section, we test and compare the different presented stacking methods on both synthetic and real marine seismic data. The different methodologies presented in the previous sections are summed up in Table 3.1. The purpose is to explicitly define the labeling of the different methodologies for future reference.

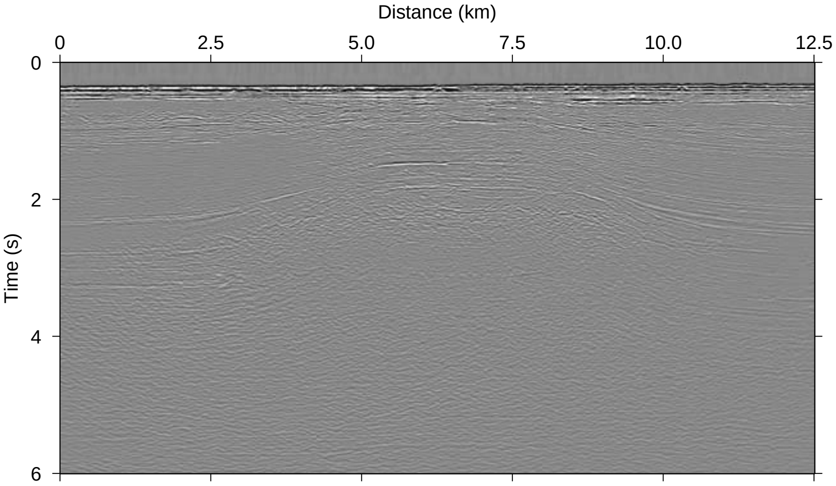
3.4.1 Application on synthetic data

Let us consider two synthetic data sets shown in Figure 3.4 (a) and Figure 3.5 (a), respectively, that consist of $M = 20$ traces and $N = 885$ samples, with four NMO-corrected reflections that results from the convolution of a Ricker wavelet with a reflection sequence of decreasing amplitude and corrupted by different types of noise. The Ricker wavelet is also known in the literature as the Mexican hat wavelet and it is defined by $\psi(t) = \frac{1}{\sqrt{2\pi}\beta^3}(1 - \frac{t^2}{\beta^2})e^{-\frac{t^2}{2\beta^2}}$ with parameter $\beta^2 = 0.14$ in this example (Chui, 1992).

In Figure 3.4, the corrupting noise is zero-mean white Gaussian noise, where the noise variance is 0.04 for the first trace and 0.01 for the other ones. The maximum and



(a) Before norm equalization.



(b) After norm equalization.

Figure 3.3: Application of smoothing on the output of enhanced stacking using S/N-estimation reference trace.

minimum spike amplitudes is 3.41 and 1.42 for the first trace and 0.73 and 0.31 for the 20^{th} trace. In Figure 3.5, the additive noise is correlated along traces and contaminated by randomly distributed noisy spikes. The correlated noise is generated by filtering white Gaussian noise with an autoregressive process of order 1 having a system function $H(z) = \frac{1}{1-0.7z^{-1}}$. The randomly distributed noisy spikes are obtained from the convolution of the Ricker wavelet described earlier with a normally distributed white noise

Methods	Description	Fig. 1 & 2	Fig. 6	Fig. 7	Fig. 8
Conventional stacking	$\bar{y}(t) = \frac{1}{f_d(t)} \sum_{i=1}^{f_d(t)} x_i(t)$ Sum of traces, where all traces are treated equally.	B	A	a	
S/N-estimation stacking	$\bar{y}(t) = \sum_{i=1}^{f_d(t)} \frac{\omega_i}{\sum_{i=1}^{f_d(t)} \omega_i} x_i(t)$ Weighted sum of traces, where each trace is attributed a weight ω_i proportional to its S/N ratio.	C	B	b	
Kalman stacking	$\bar{y}(t) = \lim_{i \to f_d(t)} \hat{s}_i(t)$ Successive trace values are used to convergence to the stacked value.	D	C	c	
Enhanced stacking	$\bar{x}(t) = \sum_{i=1}^{f_d(t)} w_i(t) x_i(t)$ with $w_i(t) = \frac{\sum_{k=t-L_w/2}^{t+L_w/2} x_i(k)\bar{y}(k)}{\max_{t \epsilon G_k} \mathbf{\Sigma}_G(t)}, \forall t \in G_k.$ Weighted sum of traces, where each weight results from the local correlation between each trace and a reference trace $\bar{y}(t)$.	E, F	D, E		a,b,c

Table 3.1: Summary of the different methodologies.

sequence $\sim \mathcal{N}(0,1)$, whose amplitude is normalized and where any amplitude less than 0.49 is set to zero. Because of its complexity, this noise is much more challenging to handle. In addition, the third reflection on the first three traces has a small amplitude (i.e., the same as trace 12), and the fourth reflection on the first three traces has a much larger amplitude than all the other traces. The maximum and minimum spike amplitudes is 3.41 and 0.66 for the first trace and 0.73 and 0.12 for the 20^{th} trace.

For both Figure 3.4 (a) and Figure 3.5 (a), the second reflection has been NMO undercorrected on purpose so that it is not perfectly aligned. The delay between the first and the last trace is 53 samples. The third reflection is recorded by traces 1 to 12 only. The length of the sliding-window parameter L_w is set to 20 samples, i.e., 20 ms, and the weight normalization parameters α and δ are set to 0.01 and 3.5 respectively. Figure 3.4 (b) and Figure 3.5 (b) show the first input trace without additive noise (A), the reference traces resulting from S/N-estimation (C) and Kalman filter (D), the stacked traces resulting from conventional stacking (B), our enhanced stacking method using output S/N-estimation reference (E) and our enhanced stacking method using output Kalman reference (F), and for comparison, manually scaled Liu stacking method without regularization (Liu et al., 2009) with threshold value 0.7 (G), LMO-based weighted stacking (Neelamani et al., 2006) (H) and smart stacking (Rashed, 2008) with parameters $\alpha = 0.2$ and $s = 1$ (I), for the first and second synthetic data sets respectively. Note that the manual scaling of Liu's stacked trace is realized by matching the maximum amplitude of the stacked trace with the maximum amplitude of the first synthetic input trace. In Table 3.2, we give the signal-to-noise ratio in dB for the first and second synthetic data sets according to the following definition:

$$\mathrm{S/N} = 10 \log \frac{\sum_{t=1}^{N} x_0^2(t)}{\sum_{t=1}^{N} [x_0(t) - \bar{x}(t)]^2} \tag{3.27}$$

where $\mathbf{x}_0$ is the first input trace without additive noise, $\bar{\mathbf{x}}$ is the stacked trace and N the number of samples contained in $\mathbf{x}_0$ and $\bar{\mathbf{x}}$. Because the signal amplitudes decay with offset and because signals are not necessarily recorded by all traces, we prefer to compare the stacked traces with the first trace without additive noise.

Our first comment is how poor the conventional, LMO-based and smart stacking perform as soon as the seismic events are not perfectly aligned, as in the case of the second reflection. The resulting signal is barely visible while the amplitude is almost conserved with other methods. If a signal is recorded by a couple of traces only, as in the case of the third reflection, then conventional, LMO-based and smart stacking average out signals and output a reflection of smaller amplitude. Finally, the first and fourth reflections show that even if all reflections are perfectly aligned, the stacked amplitudes of these three methods are underestimated. Both reference traces (C and D) are noisier than the other traces and especially the one resulting from conventional stacking. On the other hand, they have larger reflection amplitudes than conventional stacking which explains their high S/N, 9.35 dB and 9.62 dB respectively, compared to 4.31 dB for conventional stacking in the case of additive Gaussian noise. Hence, S/N-estimation and in particular Kalman reference traces are the best candidates for the enhanced stacking method.

The comparison of the enhanced stacking traces (E and F) with their respective reference trace (C and D) reveals one important feature: Although keeping in mind that the noise level is higher, we notice that whatever the type of additive noise, the amplitude of the reference traces for the first, third and fourth reflections are larger or equal, i.e., for perfectly aligned signals that are recorded by all channels or by a significant number of channels (third reflection). However, as soon as the reflection recorded by the traces is not perfectly aligned, as in the case of the second reflection, the amplitude for the reference traces is clearly attenuated, while it is preserved, although slightly distorted, for the enhanced stacking traces. Note that the second reflection of enhanced stacking traces can be larger than the noise-free trace (A), this is caused by negative weights. We also notice that the reflection amplitudes obtained with the Kalman filter (D) are larger than the ones obtained with S/N-estimation (C). With the S/N-estimation method, most of the weights are given to the first few traces because each S/N estimate is calculated for an entire trace. Since the first few traces are usually the ones with the highest amplitudes, they yield the largest S/N estimates. Therefore, it is mostly the amplitudes of these first few traces that are considered while averaging. On the contrary, the Kalman filter takes into account all the traces scaled by an amplitude factor, meaning that a much larger number of traces are effectually involved in the stacking process. Thus, for synthetic data corrupted by Gaussian noise, the S/N is greater for the reference traces (9.35 dB and 9.62 dB) than for the traces resulting from enhanced stacking (8.29 dB and

8.31 dB), but they however remain much larger than for conventional stacking (4.31 dB), LMO-based stacking (3.31 dB) and smart stacking (3.35 dB). Note the particular case of manually-scaled Liu stacking method without regularization. The S/N reaches 9.47 dB only after we have performed the arbitrary manual scaling. Therefore, this S/N value does not have any meaning here. In addition, enhanced stacking is the only method that conserves the amplitude of reflections that are not perfectly aligned from trace to trace. Finally, Liu's method stacks reasonably well those reflections that are aligned, but not the ones that are slightly delayed relative to each other. In addition, the stacked trace has to be scaled manually to recover real amplitudes, which is a major drawback.

Increasing the complexity of additive noise degrades the quality of all the estimated stacked traces. The most common changes, visible in Figure 3.5, consist of signal distortion, enhancement of noisy spikes, and an increased noise level. This degradation is also remarkable by looking at the S/N of the stacked trace, shown in Table 3.2, where all methods except LMO-based stacking are more or less affected. There are several reasons for this degradation. First, we notice in Figure 3.5 (b) that all the stacked traces contain a larger amount of noise, except S/N-estimation and Kalman enhanced stacked traces (E and F) and Liu's stacked trace (G) that manage indeed to keep a very low noise level. The signal amplitudes of conventional stacking (B), LMO-based stacking (H) and smart stacking (I) methods are about the same as for additive Gaussian noise and are not really affected by the noise complexity. On the opposite side, S/N-estimation and Kalman reference traces (C and D), and Kalman enhanced stacked trace (F) exhibit smaller amplitudes, especially at the second reflection, although keeping larger amplitudes than the other methods. The amplitudes of Kalman reference trace and enhanced stacked trace (D and F) have been significantly attenuated in comparison with additive Gaussian noise, and they are obviously much more affected than S/N-estimation by complex noise. A reason is that the Kalman filter is derived under the assumption of Gaussian observation noise, and therefore, the use of correlated noise deteriorates the signal estimate. In the third and fourth reflections, the traces containing the largest signals differ. However, the reflection amplitudes of the Kalman reference trace are still slightly larger than for S/N-estimation. Liu's stacked trace, manually scaled, is affected by signal distortion when the complexity of the noise increases. In addition, it exhibits a signal amplitude that is abnormally large at the second reflection. We notice, however, that a signal distortion is also clearly visible on the fourth reflection affecting all the tested methods. Finally, only the S/N-estimation enhanced stacked trace (E) looks unaffected by the change of additive noise. Therefore, enhanced stacking is a robust method that ensures the enhancement of seismic signals whatever the type of additive noise.

We obtain the best combination by using the S/N-estimation reference trace. When we combine the enhanced stacking with the Kalman reference trace, we obtain very good signal estimates as long as the additive noise is Gaussian, but we observe some amplitude attenuation when the additive noise is more complex. S/N-estimation and Kalman reference traces have also proven to be effective, but are much more sensitive to the complexity of additive noise than enhanced stacking. They might then be subject to signal distortion, increase of noise level and signal amplitude attenuation. However, they yield an acceptable estimate of the stacked trace. Liu's stacking method without regularization does not handle the reflections misalignment as well as enhanced stacking, enhances some noisy spikes, and is also affected when the complexity of the noise increases. Finally, conventional stacking, LMO-based stacking and smart stacking cannot handle situations where the seismic reflections are not properly aligned, their stacked reflections have lower amplitudes than the other methods. They are nevertheless not so much affected by the complexity of the additive noise.

	Conventional		S/N-estimation		Kalman		Scaled	LMO	Smart
	ref.	stacked	ref.	stacked	ref.	stacked	Liu	stacking	stacking
Synthetic data set 1	4.31	7.91	9.35	8.29	9.62	8.31	9.47	3.31	3.35
Synthetic data set 2	4.08	7.76	6.20	7.93	5.97	6.71	7.44	3.57	3.04

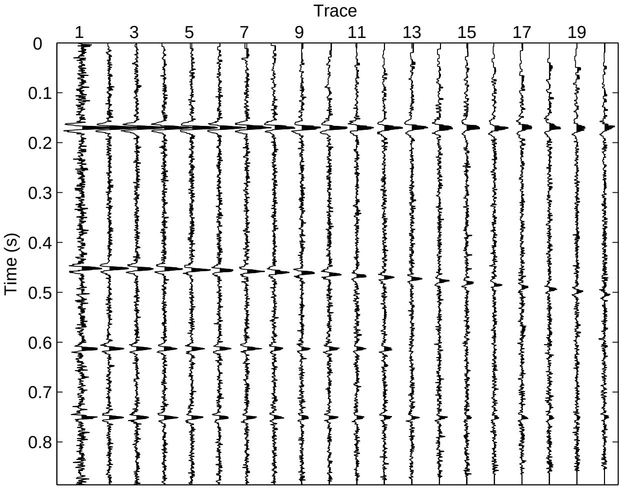
Table 3.2: S/N (dB) of the stacked traces.

3.4.2 Application on real data

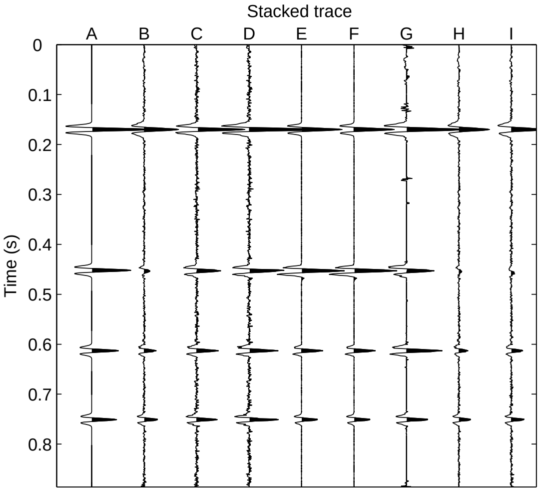
Stacking applied to a single CMP gather

Let us first consider the case of a single CMP gather that contains up to 51 traces simultaneously, shown in Figure 3.6. Time-frequency denoising (Elboth et al., 2010) and normal move-out correction have been applied to this data set. Since data gathers still contain lots of multiples, the shallow waters part of the records have been removed by muting. The sliding-window length L_w is set to 20 samples, i.e. 80 ms, and the weight normalization parameters are chosen to be $\alpha = 0.01$ and $\delta = 3.5$, respectively.

Figure 3.7 shows a zoom of the CMP gather in the traveltime range 3.52 – 4.80 seconds (a) and a zoom in the same time range (b) of the reference traces resulting from S/N-estimation (blue curve) and Kalman filter (red curve), and the stacked trace obtained from conventional stacking (green curve). Clearly, the conventional stacking trace hardly exhibits any structure and the signal amplitude is rather low. On the contrary, Kalman and S/N-estimation reference traces yield greater signal amplitudes, but they differ from each other. The Kalman reference trace is able to emphasize any coherent signal recorded by any traces that are not necessarily the ones with small offsets (i.e., it does not emphasize any specific part of the signal). Let us for instance look at traveltime 3.59 s, where a relatively big almost coherent upward spike (shaded in black) is recorded by the 4^{th} to 7^{th} traces. For the first three traces, signal values are much smaller at that traveltime sample. However, in Figure 3.7 (b) we see that a significant spike is found by the Kalman filter, and a much smaller spike is identified by the S/N-estimation reference trace. Similar observations can be made between 4.28 s and 4.32 s, and around 4.72 s. This can be explained by the fact that the Kalman filter takes into account all the traces scaled by an amplitude factor to converge to the stacked trace,



(a)

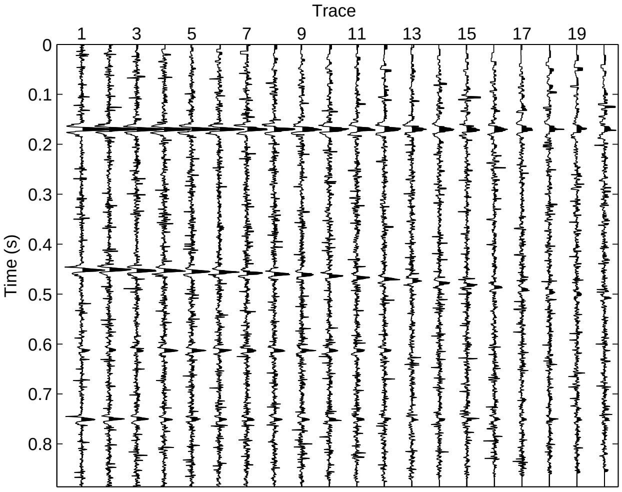


(b)

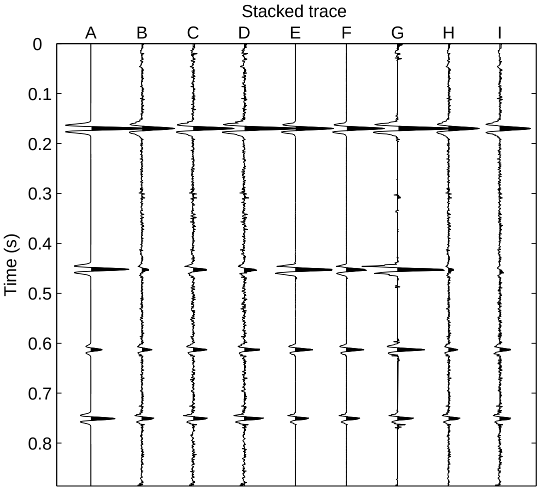
Figure 3.4: (a) Synthetic NMO corrected data set corrupted by white Gaussian noise and (b) Stacked traces: A) First input trace without additive noise, B) Conventional stacking, C) S/N-estimation reference trace, D) Kalman reference trace, E) Enhanced stacking using S/N-estimation reference, F) Enhanced stacking using Kalman reference, G) Rescaled Liu stacking, H) LMO-based stacking and I) Smart stacking.

meaning that any signal recorded for a larger offset is considered as well. On the contrary, the S/N-estimation method puts most of its weights in the first few traces, and does not give much credit to signals recorded for larger offsets.

Let us now have a look at shallow water, in the traveltime range 1.04 – 1.72 sec-



(a)



(b)

Figure 3.5: (a) Synthetic NMO corrected data set corrupted by correlated spiky noise and (b) Stacked traces: A) First input trace without additive noise, B) Conventional stacking, C) S/N-estimation reference trace, D) Kalman reference trace, E) Enhanced stacking using S/N-estimation reference, F) Enhanced stacking using Kalman reference, G) Rescaled Liu stacking, H) LMO-based stacking and I) Smart stacking.

onds, as illustrated in Figure 3.8. The S/N-estimation reference method puts most of its weights on the first two traces (for instance, weights of the first two traces at traveltime 1.04 s and 1.72 s represent respectively 51 % and 38 % of all weights) and therefore, coherent signals recorded by these first two traces are emphasized. In the top figure, we

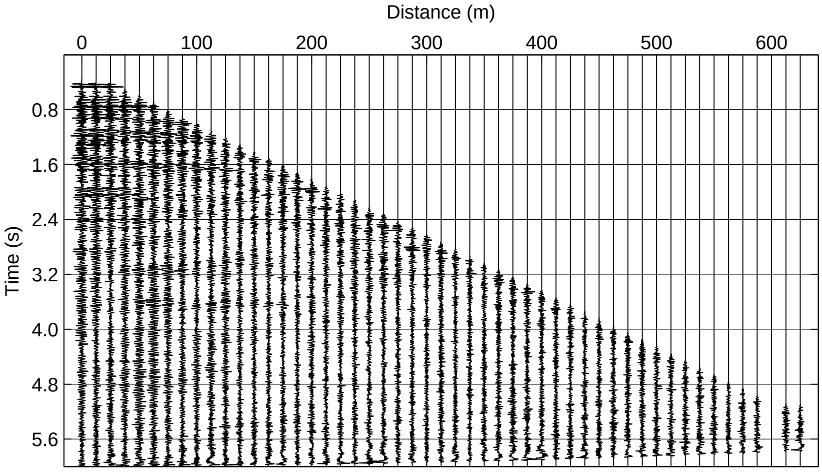
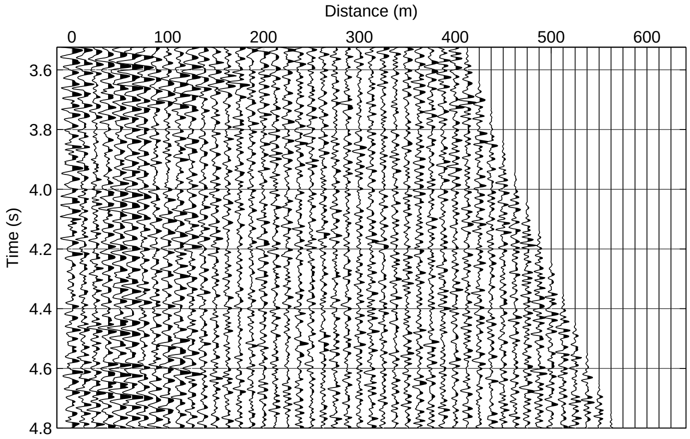


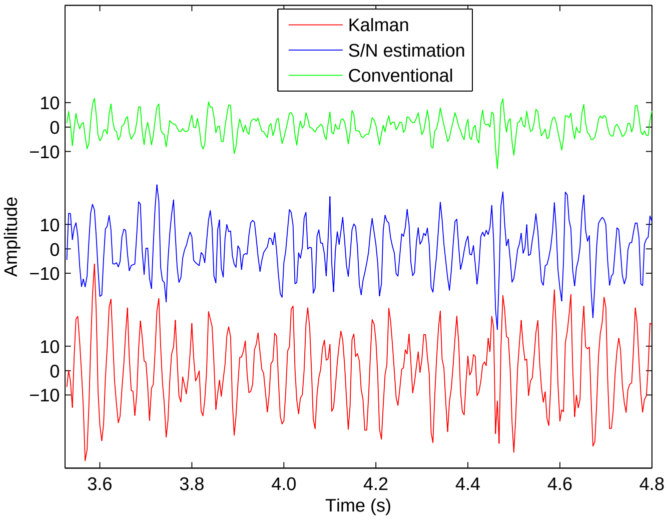
Figure 3.6: 51 fold marine seismic CMP gather.

can observe a few signal events that have a significant amplitude at the first two or three traces, such as at traveltime 1.31 s or 1.37 s. We also notice that most of the signal events are not aligned past the first three traces. If we look at the stacked traces at these traveltime occurrences, we notice S/N-estimation reference trace exhibits a larger amplitude than the Kalman reference trace which suffers from the misalignment across the traces.

Let us now compare our enhanced stacking method (using S/N-estimation and Kalman reference traces) with conventional stacking, Liu's method without regularization (Liu et al., 2009) using a threshold value of 0.7, and smart stacking (Rashed, 2008) with parameters $\alpha = 0.2$ and $s = 1$. In Figure 3.9, we show the reference traces resulting from S/N-estimation (B) and Kalman filter (C), the stacked traces resulting from conventional stacking (A), enhanced stacking method using output S/N-estimation reference (D) and enhanced stacking method using output Kalman reference (E), manually-scaled Liu stacking method without regularization (F) and smart stacking (G). Because of sharp thresholding, the stacked trace obtained with Liu's method removes too much information and looks relatively chaotic (F). In addition, it does not locate coherent signals as accurately as the other methods do. Smart stacking (G) seems to average out most of the coherent signals after traveltime 1.2 s and conventional stacking (A) after traveltime 2 s. Stacked traces resulting from enhanced stacking (D and E) contain less noise than in their respective reference traces (B and C) while exhibiting as much coherent signals. Thus, enhanced stacking is successful in enhancing coherent signals and attenuating random noise. Most of the coherent signals are visible in both (D) and (E), but they differ in a few cases such as at traveltime 1.2 s, 2.8 s and 3.6 s. Actually, enhanced stacking reflects the differences that were already visible on the reference traces (B and C).



(a) Input CMP gather.

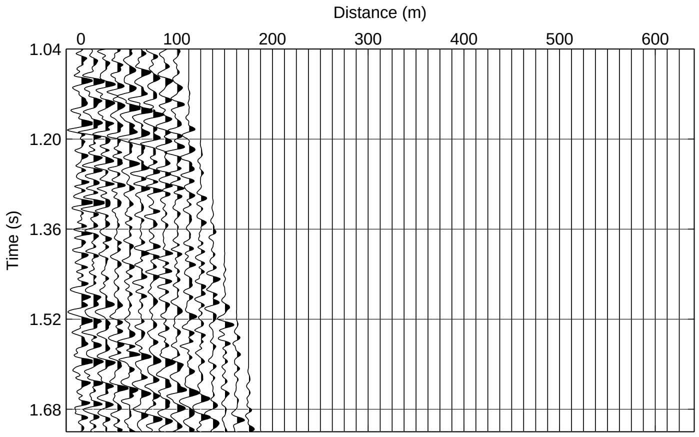


(b) Comparison of Kalman based and S/N estimation based reference traces with conventional stacking.

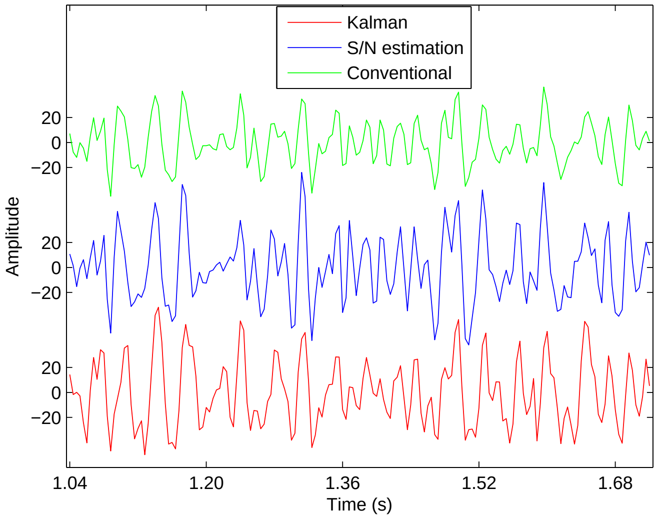
Figure 3.7: Zoom in the traveltime range 3.53 – 4.80 s.

Stacking applied to a 2D line

We apply enhanced stacking with its three possible reference traces (conventional stacking, S/N-estimation reference and Kalman filter) to a set of 1001 prestack migrated gathers of marine seismic sub-salt data (fold is 162), after time-frequency denoising (Elboth et al., 2010) and normal move-out correction. The migration process has been achieved by a 2D Kirchhoff prestack time algorithm (Bancroft et al., 1998; Schneider,



(a) Input CMP gather.



(b) Comparison of Kalman based and S/N estimation based reference traces with conventional stacking.

Figure 3.8: Zoom in the traveltime range 1.04 – 1.72 s.

1978). The length of the sliding-window parameter L_w is set to 20 samples, i.e., 80 ms, and the weight normalization parameters are chosen to be $\alpha = 0.05$ and $\delta = 2.5$. Figure 3.10 shows the stack sections made of reference traces: conventional stacking (a), S/N-estimation method (b) and Kalman filter (c) that uses the conventional reference trace to estimate the noise variance that is a necessary model parameter. Compared to conventional stacking section (a), the section (b) and in a major importance section (c), enhance signal reflections, especially for distances below 2 km and beyond 10 km,

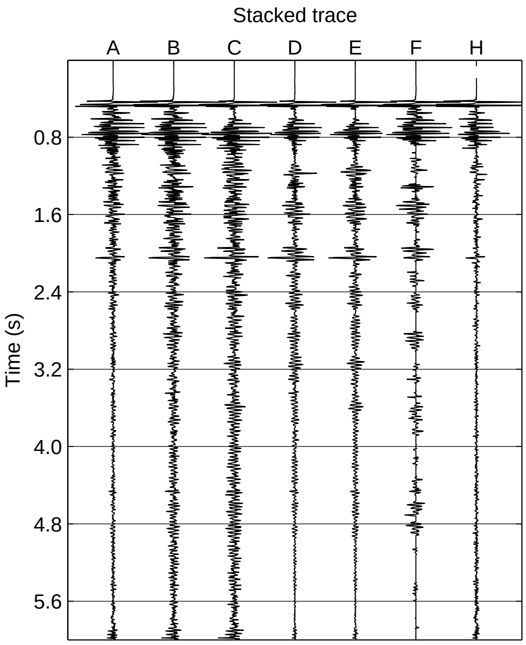


Figure 3.9: Stacked trace: A) Conventional stacking, B) S/N-estimation reference trace, C) Kalman reference trace, D) Enhanced stacking using S/N-estimation reference, E) Enhanced stacking using Kalman reference, F) Rescaled Liu stacking and G) Smart stacking.

and for shallow water. They also enhance deeper events and emphasize the edges of the sub-salt structure. The S/N-estimation section (b) reveals some continuous events in deep water, beyond 4.5 s. The Kalman filter stacked section (c), however, clearly achieves more contrast than the other two, but also emphasizes the artefact caused by the migration process in the sub-salt structure. In order to avoid these shortcomings, more advanced migration processes would be more helpful.

These three reference methods are used in the enhanced local correlation stacking and the outputs are shown in Figure 3.11. Note that we use the norm equalization to remove some artefacts caused by the weight normalization process. Here, we show that enhanced stacking tries to emphasize events that have been detected by the reference method. Thus, when using conventional stacking and S/N estimation as reference methods, enhanced stacking emphasizes the strongest reflections, especially in shallow water, along the sub-salt structure edges, and for deeper reflections at near and far offsets. The artefact caused by the migration process under the sub-salt structure is nevertheless emphasized as well (Figures a and b). However, we observe the opposite effect when using the Kalman filter as reference method (c). Enhanced stacking attenuates the lowest amplitudes that consist mostly of noise and undesirable artefact, but including however small amplitude events in the time interval 1 – 3 s. As a consequence, we obtain a better image where the reflected events of large amplitudes appear clearer, but where a few events of smaller amplitudes are attenuated. The comparison of the three enhanced stacked sections shows that it is the one using S/N estimation as reference method (b) that clearest exhibits the reflected events, especially in deep waters and in the distance interval 4 – 10 km. The edges of the sub-salt structure are clearer and even a few continuous events are visible beyond 4.5 s under the sub-salt structure. Finally, by comparing the enhanced stacked section using S/N estimation as reference method with the Kalman reference section, we observe that they both present advantages and drawbacks. Clearly, the Kalman reference method reveals a much larger number of reflections, but it emphasizes noise and migration artefacts as well. On the other hand, the enhanced stacked section using S/N estimation as reference method does not emphasize as many reflections, but it shows a few reflected events under the sub-salt structure, and the noise level contained in the seismic image has been attenuated. We think that the combination of the presented methods with other pre-stack data enhancement filters and other stacking approaches may attenuate some of these shortcomings. A discussion of alternate methodologies is however outside the scope of the present discussion.

In order to quantify the improvement, we select four regions of the stacked section image and calculate the standard deviation of the signal amplitude histogram of these regions. The first region contains shallow water reflections from 0 to 1 s in time and from 4 to 9 km in distance, the second region contains deeper water reflections from 2 to 4 s in time and from 0 to 2.5 km in distance, the third region contains deep water reflections below the salt structure from 5 to 6 s in time and from 5 to 8 km in distance, and the fourth region contains the right hand side edge of the salt structure from 1.5 to 3 s in time and from 8 to 10 km in distance.

Let z be a random variable denoting the signal amplitude and let $f(z_i), i = 0, \ldots, L_I - 1$ be the corresponding normalized histogram, where L_I is the number of distinct signal amplitude levels. Then, the standard deviation of the signal amplitude histogram is given by:

$$\sigma = \sqrt{\sum_{i=0}^{L_I-1} (z_i - m)^2 f(z_i)} \tag{3.28}$$

where m is the average signal amplitude $m = \sum_{i=0}^{L_I-1} z_i f(z_i)$ (Gonzalez and Woods, 2008). In Table 3.3, we give the relative standard deviation, defined as the ratio of standard deviation calculated for any method divided by the standard deviation of conventional stacking. As observed, the enhanced stacking methods emphasize the largest reflections the most (region 1). The enhanced stacking method using S/N estimation as reference method achieves a contrast as good as the Kalman reference method under the salt structure. Else, these numbers confirm that conventional stacking yields the smallest variability in intensity, i.e., the smallest contrast, while Kalman reference method achieves the largest variability.

	Reference section			Enhanced stacking section		
	Conventional	S/N-estimation	Kalman	Conventional	S/N-estimation	Kalman
Region 1	1.00	0.95	2.72	3.58	3.39	3.59
Region 2	1.00	1.84	4.18	1.95	2.44	2.08
Region 3	1.00	1.19	4.89	3.61	4.80	4.42
Region 4	1.00	2.21	4.70	1.87	2.80	1.91

Table 3.3: Relative standard deviation of three selected regions contained in the stacked sections.

Computational cost

Let us now consider the computational cost for each method. Table 3.4 gives the relative CPU time for each method. Here, the CPU time is calculated per trace to be stacked, each trace containing 2500 time samples. As expected, conventional stacking that only consists of adding traces together is much faster to compute than all the other methods. But, what is interesting here is the comparison of S/N estimation and Kalman methods. The processing of a Kalman reference trace takes a factor of 36.5 longer than a S/N estimation reference trace and a factor of 1.7 longer time than the enhanced stacked trace using S/N estimation as reference trace. A factor of 1.7 becomes significant when processing huge volume of data. Therefore, another advantage of the enhanced stacked section using S/N estimation as reference method is that it is less computationally demanding than the Kalman reference trace.

	Reference section			Enhanced stacking section		
	Conventional	S/N-estimation	Kalman	Conventional	S/N-estimation	Kalman
CPU time per trace (ms)	0.14	0.74	27.0	15.1	15.9	41.9

Table 3.4: CPU time per trace to be stacked in milliseconds.

Homogeneity of the seismic image

The stacking methods presented here do not manipulate the trace amplitudes in itself that are left untouched. It is actually the averaging process that yields enhanced, meaningful amplitudes. Therefore, whatever the method we use, the signal amplitudes are consistent from CMP to CMP as long as the method parameters are close to optimal. Our experience showed us that these parameters may only be non-optimal for a few stacked traces where the signal amplitudes are either under-estimated or over-estimated. To overcome this problem, we apply the norm smoothing systematically. Thus, the lateral homogeneity of the three enhanced stacked sections in Figure 3.11 is satisfactory. In addition, we underlined that the presented methods are capable of revealing seismic events, both in shallow and deep water, and that this is achieved in a natural manner without any manipulation of amplitudes. Using automatic gain control (AGC) for instance, would not lead to consistent amplitudes, it would not necessarily be able to reveal as many events, and finally, it would even emphasize the migration artefact in the sub-salt structure, which is definitively not desirable.

Choice of parameters

Finally, because only three parameters (L_w, α and δ) with a limited number of meaningful values are involved in the enhanced stacking method, the effort to choose these parameters is relatively small. The parameter that depend strongest on the data is δ. If we consider the nicest and the most problematic stacked traces, then a few trial-and-errors on these two stacked traces are usually enough to find the set of optimal parameters.

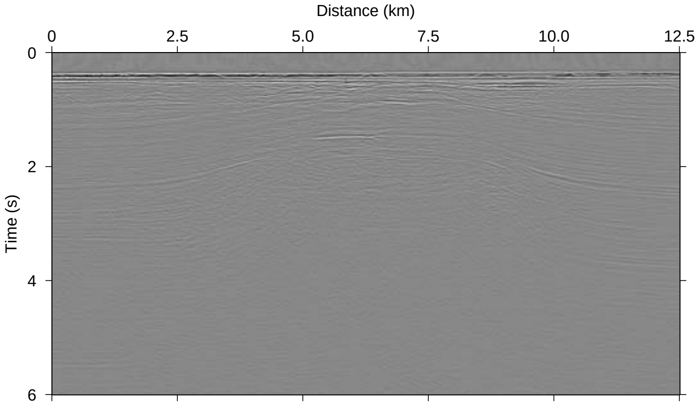
3.5 Summary

We have proposed and demonstrated a novel stacking method that utilizes local correlation and a new weight normalization scheme. The basis of the method is to apply local correlation between each individual trace and a chosen reference trace, as a measure of weight that is then fed into a weighted stacking. Three different reference traces have been proposed, and the corresponding outputs of the stacking method were compared on synthetic and real data.

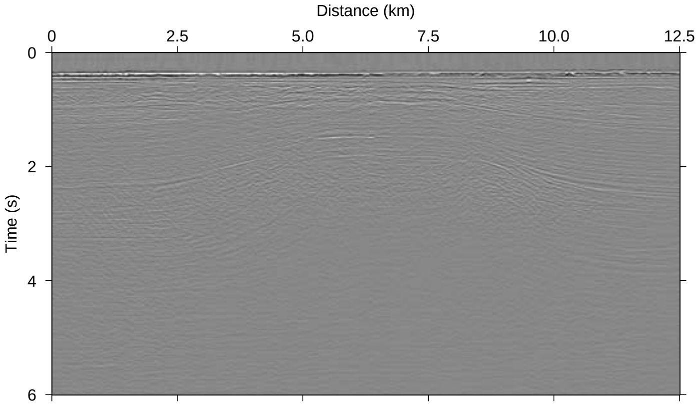
As shown, the use of S/N-estimation reference trace and even more significantly, the use of Kalman reference trace yielded consistently superior results compared to equal weight stacking. In particular, the superior reference traces exhibited cleaner and better defined reflected events, as well as the ability to see a larger number of reflections in deep waters. We documented that the use of an S/N-estimation reference trace has the ability to enhance coherent signals recorded by short offset traces, while a Kalman reference trace has the ability to emphasize any coherent signal recorded by any traces. For this reason, misalignment does not degrade S/N-estimation reference trace as much as Kalman reference trace.

Furthermore, the enhanced stacking method applied together with norm smoothing yielded superior results compared to equal weight stacking, norm smoothing ensuring a satisfactory homogeneous seismic image. We showed that the best combination is obtained when using S/N estimation as reference method. It exhibits more reflected events, and in our present example, it exhibits a few continuous reflected events under the sub-salt structure. However, we showed that enhanced stacking emphasizes the strongest reflections while discriminating weaker ones, which deteriorates the longitudinal homogeneity of the seismic image. Thus, the use of Kalman filter reference method should be preferred to enhanced stacking using S/N estimation as reference method, since it achieves a better overall seismic image contrast and reveals many more, cleaner and better defined reflected events. However, this result is obtained at the cost of a higher noise level and a larger processing time. Therefore, choosing between one of these two methods is a trade-off between more revealed information and cleaner seismic image. Finally, we think that the presented methods may be combined with other pre-stack data enhancement filters and other stacking approaches, and therefore, contribute to the development of seismic imaging improvement techniques.

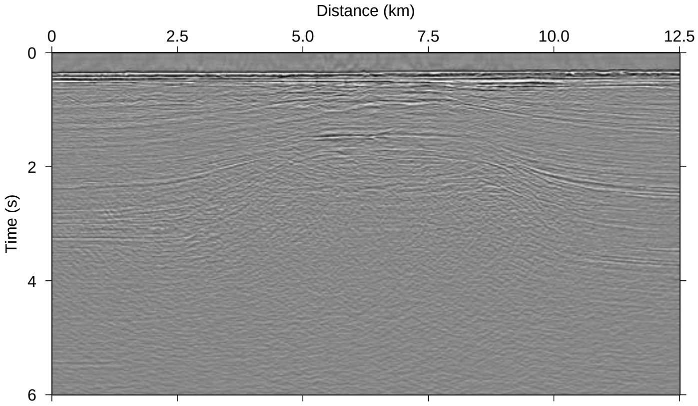
We suggest that future versions of the proposed stacking method should explore the use of mutual information as a generalized statistical measure of signal correlation. Another possible improvement consists of deriving S/N estimates locally instead of globally to improve the S/N-estimation method. Equivalently, deriving trace amplitude locally instead of globally should improve the Kalman filter method.



(a) Conventional stacking.

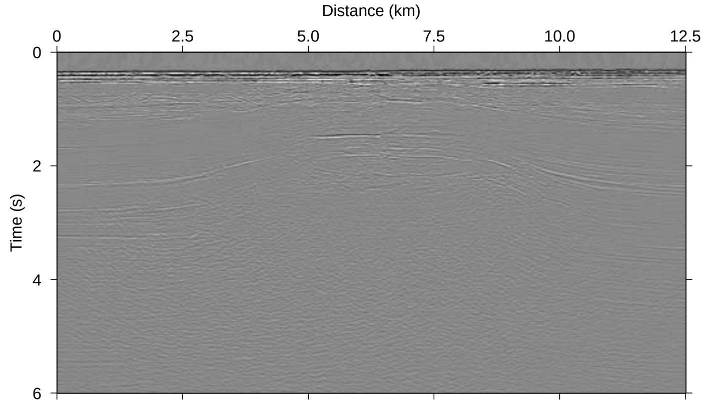


(b) S/N-estimation method.

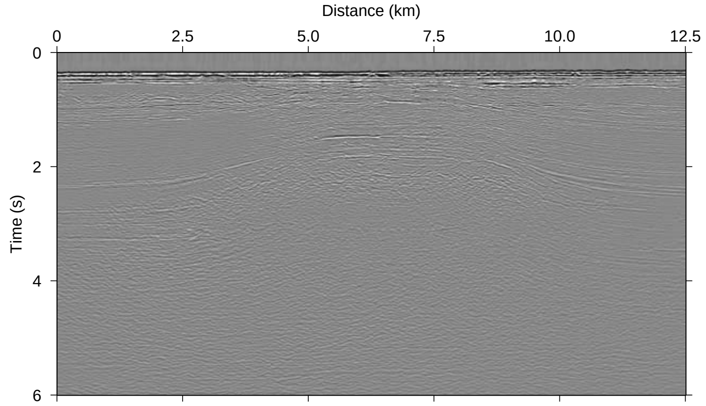


(c) Kalman method.

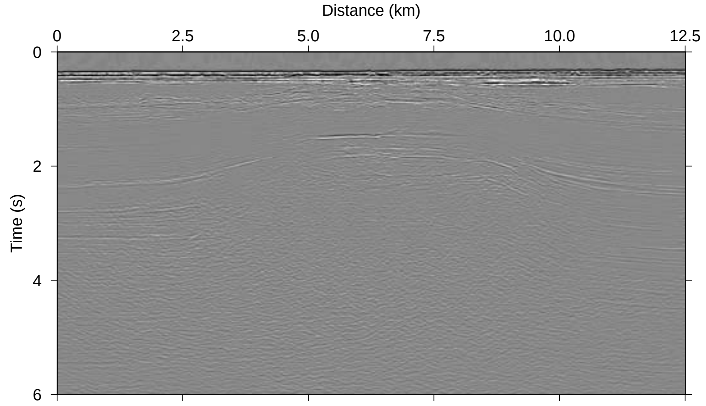
Figure 3.10: Stack sections made of reference traces.



(a) Conventional stacking.



(b) S/N-estimation method.



(c) Kalman method.

Figure 3.11: Enhanced local-correlation stacking sections using the following reference traces.

Chapter 4

Principal components analysis, brief review

4.1 Introduction

This chapter describes a well known eigenvector-based multivariate analysis technique, the decomposition into principal components, with a particular focus on the singular value decomposition (SVD). Although we will be mainly concerned by its application to seismic processing, this technique has a wide range of applicability such as dimension reduction, patterns recognition and image processing.

The earliest descriptions of the technique which is now known as principal components analysis (PCA) appear to have been given by Pearson (1901) and Hotelling (1933). The two papers adopted different approaches. Pearson (1901) was concerned with finding lines and planes which best fitted a set of points in N-dimensional space, and the geometric optimization problems which he considered lead also to principal components. On the other hand, Hotelling (1933) focused on the standard algebraic derivation, where his "components" are chosen to maximize their successive contributions to the total of the variances of the original variables. Some other important references within the subject are given by Anderson (1963), that discussed the asymptotic sampling distributions of the coefficients and variances of the sample principal components. Rao (1964) gave a large number of new ideas concerning the uses, interpretations and extensions of PCA, while Gower (1966) discussed some links between PCA and various other statistical techniques, and also provided a number of geometric insights. Finally, Jeffers (1967) emphasized the practical side of PCA by discussing two case studies.

In the perspective of enhancing the signal contained in seismic images, decomposition into principal components is a well known technique that tries to maximize the information content of recorded data. First publications that have investigated the applications of particular linear transformation, such as the Karhunen-Loève (KL) transformation, were published in the seventies. In particular, a paper from Ready and Wintz (1973) dealing with information extraction and S/N improvement in multispectral imagery. The first publication that introduced and applied these transformations to seismic data was the paper by Hemon and Mace (1978). Freire and Ulrych (1988) applied the KL transformation in a different manner which is in fact the singular value decomposition to the processing of vertical seismic profiling data. SVD has been since applied to various other problems such as residual statics corrections (Ulrych et al., 1988) and seismic signal enhancement (Bekara and Van der Baan, 2007; Vrabie et al., 2004). A review of the applications of principal components analysis to seismic processing is given in Ulrych and Sacchi (2005).

4.2 Definition and properties

The central idea of principal component analysis (PCA) is to reduce the dimensionality of a data set which consists of a large number of interrelated variables, while retaining as much as possible of the variation present in the data set. Reducing the complexity of a data set reveals the sometimes hidden, simplified structure that often underlies it, and is therefore, a powerful tool for analyzing data. The principal components are given by an orthogonal linear transformation of a set of variables, such that in the new coordinate system, the greatest variance by any projection of the data comes to lie on the first coordinate (called the first principal component), the second greatest variance on the second coordinate, and so on (Jolliffe, 2002).

Suppose that we have M independent observations with zero empirical mean of a N-element random vector $\mathbf{x}$, denoted by $\mathbf{x}_1, \ldots, \mathbf{x}_M$, stored columnwise in a $N \times M$ data matrix $\mathbf{X}$. By empirical mean, we meant that the mean of each element of $\mathbf{x}$ has been subtracted to be zero, i.e., the (i, j)th element of $\mathbf{X}$ is adjusted such that $x_{i,j} \leftarrow x_{i,j} - \frac{1}{M}\sum_{j=1}^{M} x_{i,j}$. Let us consider the sample covariance matrix of $\mathbf{X}$ defined as $\frac{1}{M-1}\mathbf{X}\mathbf{X}^H$. Note that the eigenvectors of $\frac{1}{M-1}\mathbf{X}\mathbf{X}^H$ and $\mathbf{X}\mathbf{X}^H$ are identical, and the eigenvalues of $\frac{1}{M-1}\mathbf{X}\mathbf{X}^H$ are simply $\frac{1}{M-1}$ times the eigenvalues of $\mathbf{X}\mathbf{X}^H$. Because of these relationships, it is convenient to work in terms of eigenvectors and eigenvalues of $\mathbf{X}\mathbf{X}^H$, instead of $\frac{1}{M-1}\mathbf{X}\mathbf{X}^H$. The eigenvectors and eigenvalues of $\mathbf{X}\mathbf{X}^H$ are obtained from $(\mathbf{X}\mathbf{X}^H)\mathbf{P} = \mathbf{P}\mathbf{\Lambda}$, where $\mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_N]$ contains the N eigenvectors of $\mathbf{X}\mathbf{X}^H$, and $\mathbf{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_N)$ is a diagonal matrix containing the N eigenvalues of $\mathbf{X}\mathbf{X}^H$. Thus, the PCA transformation of $\mathbf{X}$ is given by

$$\mathbf{Y} = \mathbf{P}^H\mathbf{X}. \tag{4.1}$$

Note that $\mathbf{P}$ is an orthonormal matrix, containing the principal components of $\mathbf{X}$. Thus, the first principal component corresponds to the eigenvector with the largest eigenvalue, the second principal component corresponds to the eigenvector with the second largest eigenvalue, and so on.

In addition, the following interpretation follows:

- The PCA transformation is equivalent to the Karhunen-Loève transformation since the sample covariance matrix of $\mathbf{Y}$ is diagonal and made of the eigenvalues of the sample covariance matrix of $\mathbf{X}$:

$$\begin{aligned}\frac{1}{M-1}\mathbf{Y}\mathbf{Y}^H &= \frac{1}{M-1}\mathbf{P}^H\mathbf{X}\mathbf{X}^H\mathbf{P} && (4.2)\\ &= \frac{1}{M-1}\mathbf{P}^H\mathbf{P}\mathbf{\Lambda} && (4.3)\\ &= \frac{1}{M-1}\mathbf{\Lambda}. && (4.4)\end{aligned}$$

- PCA projects the data along the directions where the data varies the most.

- These directions are determined by the eigenvectors of $\mathbf{X}\mathbf{X}^H$ corresponding to the largest eigenvalues.

- The magnitude of the eigenvalues corresponds to the variance of the data along the eigenvector directions.

To illustrate the determination of the principal components, let us consider the simple case where $N = 2$. The advantage of $N = 2$ is, of course, that the data can be plotted in two dimensions. Figure 4.1 shows a set of 10 two-dimensional observations plotted by a blue cross, and its two principal component vectors scaled by the square root of their respective eigenvalues. It is clear that there is a greater variation in the direction of $\mathbf{p}_1$ than in either of the original variables, but very little variation in the direction of $\mathbf{p}_2$.

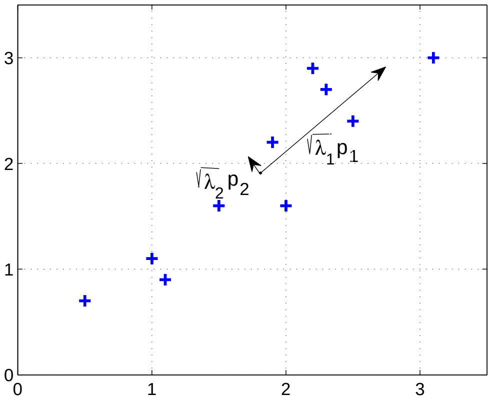


Figure 4.1: Example of principle components when $N = 2$.

Application to seismic data: When applied to a NMO-corrected seismic data set, the search of principal components with largest variance and presumably highest S/N, is expected to yield the direction of laterally coherent signals. Therefore, any spread deviating from coherent signals direction must be noise. Thus, reducing the dimensions of this data set by choosing the principal components with largest variance should lead to a seismic data set attenuated in noise, and therefore with a higher S/N.

4.3 Singular value decomposition

This section describes the singular value decomposition (SVD) which is relevant to PCA in several aspects. The importance of the SVD for PCA is twofold. First, it provides a computationally efficient method of finding the principal components, since it gives us the eigenvectors and the square roots of the eigenvalues of $\mathbf{X}\mathbf{X}^H$, and hence the principal components and their variances. Secondly, the SVD provides additional insight into what a PCA actually does, and gives useful means, both graphical and algebraic, of representing the results of a PCA. More details about SVD are given in Golub and van Loan (1996) and Jolliffe (2002).

4.3.1 Theorem

Any real $N \times M$ matrix **X** with rank r made of M observations of N elements can be written as

$$\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H \tag{4.5}$$

where

U is an $N \times N$ unitary matrix, i.e., $\mathbf{U}^H\mathbf{U} = \mathbf{I}_N$,

V is an $M \times M$ unitary matrix, i.e., $\mathbf{V}^H\mathbf{V} = \mathbf{I}_M$,

$\mathbf{\Sigma}$ is an $N \times M$ diagonal matrix with diagonal elements $\langle\mathbf{\Sigma}\rangle_{ii} = \varsigma_i > 0, i = 1, 2, \ldots, r$. The numbers ς_i are known as the *singular values* of **X** and are usually arranged in decreasing order as $\varsigma_1 \geqslant \varsigma_2 \geqslant \ldots \geqslant \varsigma_r > 0$.

The proof of this theorem is given by Manolakis et al. (2005).

4.3.2 Properties

1. The diagonal matrix $\mathbf{\Sigma}$ is made of a square block of r singular values ($r \leq M$) and blocks of zeros. Therefore, in the product $\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$, the last $N - r$ columns of **U** and $M - r$ columns of **V** are superfluous because they interact only with blocks of zeros in $\mathbf{\Sigma}$. This leads to the following *thin SVD* representation of **X**

   $$\mathbf{X} = \mathbf{U}_r\mathbf{\Sigma}_r\mathbf{V}_r^H, \tag{4.6}$$

   where

$\mathbf{U}_r = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_r]$, $\mathbf{u}_k \in \mathbb{R}^{N\times 1}$, consists in the first r columns of $\mathbf{U}$ and contains the *left singular vectors* of $\mathbf{X}$,

$\mathbf{V}_r = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r]$, $\mathbf{v}_k \in \mathbb{R}^{M\times 1}$, consists in the first r columns of $\mathbf{V}$ and contains the *right singular vectors* of $\mathbf{X}$,

$\mathbf{\Sigma}_r = \text{diag}(\varsigma_1, \varsigma_2, \dots, \varsigma_r)$ is a diagonal matrix containing the r singular values.

2. The SVD of the data matrix $\mathbf{X}$ can be expressed as the sum of cross products weighted by the singular values

$$\mathbf{X} = \sum_{k=1}^{r} \varsigma_k \mathbf{u}_k \mathbf{v}_k^H. \tag{4.7}$$

3. The cross product

$$\mathbf{E}_k = \mathbf{u}_k \mathbf{v}_k^H, \qquad k = 1, \dots, r, \tag{4.8}$$

is the k^{th} eigenimage of the matrix $\mathbf{X}$. The r eigenimages form an orthonormal basis for the representation of the matrix $\mathbf{X}$.

4. The product $\mathbf{X}\mathbf{X}^H$ may be considered as a weighted estimate of the correlation matrix at lag zero of the matrix $\mathbf{X}$. Therefore, using the eigenimages associated to the highest singular values is equivalent to look for correlated structure into the matrix $\mathbf{X}$.

5. Correspondance between SVD and PCA:

The vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_M$ are eigenvectors of $\mathbf{X}^H\mathbf{X}$.

The vectors $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_N$ are eigenvectors of $\mathbf{X}\mathbf{X}^H$ and equal to $\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_N$.

The squares of the singular values $\varsigma_1^2, \varsigma_2^2, \dots, \varsigma_r^2$ of $\mathbf{X}$ are the first r nonzero eigenvalues of $\mathbf{X}^H\mathbf{X}$ and $\mathbf{X}\mathbf{X}^H$, that are

$$\lambda_i(\mathbf{X}^H\mathbf{X}) = \lambda_i(\mathbf{X}\mathbf{X}^H) = \varsigma_i^2(\mathbf{X}). \tag{4.9}$$

6. In Freire and Ulrych (1988), the authors decompose the data matrix $\mathbf{X}$ in three orthogonal subspaces

$$\mathbf{X} = \mathbf{X}_{LP} + \mathbf{X}_{BP} + \mathbf{X}_{HP}. \tag{4.10}$$

The low-pass subspace $\mathbf{X}_{LP}$ associated to the highest p singular values, contains source waves with a higher degree of sensor to sensor correlation.

The band-pass subspace $\mathbf{X}_{BP}$ associated to the next q singular values, is composed of events with a little degree of sensor to sensor correlation.

The high-pass subspace $\mathbf{X}_{HP}$ associated to the last $r - p - q$ singular values, is the noise subspace.

The purpose of this kind of decomposition is to isolate only the highest sensor to sensor correlated waves in the low-pass subspace $\mathbf{X}_{LP}$. This is possible if there is a large gap between the first p singular values and the others. Mari et al. (1999) shows that $p = 1$ for a perfect wave alignment and $p = 2$ for a non-perfect wave alignment.

4.4 Application to image processing and seismic data

4.4.1 Theoretical aspects

Let us now consider the application of SVD to image processing. In this instance, the situation encountered is essentially different. From a stochastic point of view, we now have one realization, **X**, of a two dimensional random process. PCA and SVD of **X** will be the same only if we assume the separability of the covariance matrix into a product of the covariance matrices of the rows and columns of the image, and if these covariance estimates are computed from the single realization which is our image (Gerbrands, 1981). It is clear that in this case, the use of stochastic terminology like PCA is not appropriate. On the other hand, eigenimage decomposition, which depends on a deterministic decomposition using the SVD is perfectly descriptive.

Seismic data that consists of M traces of N samples each can be represented by a $N \times M$ matrix **X**. In this case, the columns of $\mathbf{U}_r$ form an orthonormal basis to represent space and $\mathbf{u}_k$ is called "propagation vector". Similarly, the columns of $\mathbf{V}_r$ form an orthonormal basis to represent time and $\mathbf{v}_k$ is called "normalized wavelet" (Vrabie et al., 2004). When considering NMO corrected data, laterally coherent events create a linear dependence among the traces. Thus, the data shows a pretty high degree of trace-to-trace correlation and therefore, only a few eigenimages associated to the highest singular values are necessary to get a good estimate $\hat{\mathbf{X}}$. Typically, only the highest p singular values are necessary to estimate **X**, and therefore, $\hat{\mathbf{X}} = \mathbf{X}_{LP}$.

Note that Vrabie et al. (2004) proposed a modified SVD that combines SVD with independent component analysis (ICA) to avoid some artefacts caused by the orthogonality criterion imposed to the propagation vectors. In the specific case of seismic data, the best results in terms of signal enhancement are obtained when SVD is applied to NMO corrected data. The number of singular values necessary to describe the subspace containing aligned events is usually chosen to be equal to one and in that case, SVD-ICA is strictly equivalent to SVD.

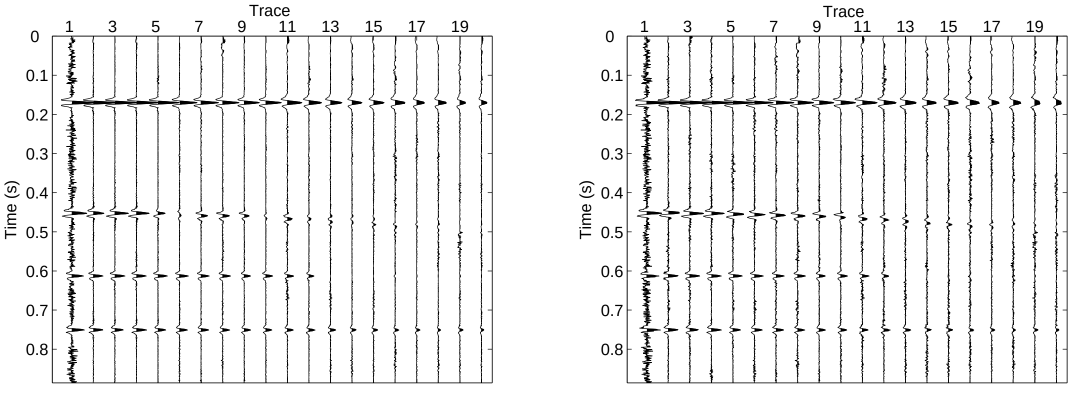
4.4.2 Application to synthetic and real data

Let us apply the SVD locally to a synthetic data set similar to the first set described in Chapter 3, Section 3.4.1, which consists of four NMO-corrected reflections recorded by $M = 20$ traces of $N = 885$ samples each. In its local form, SVD is applied to a small data window, extracted from the input data set by a sliding window that scans the input gather in space and time. If the sliding windows overlap, then we take the result to be the mean of all estimated traces. Local SVD, contrary to global SVD (i.e., applied to the entire input gather at once), can cope with short and quickly varying events, and therefore, yields a better signal estimation (Bekara and Van der Baan, 2007). In the following application, the sliding window consists of 10 traces and 50 samples per trace, and has 50% overlap both along time and traces between adjacent windows. We fix the number p of singular values to be used for the reconstruction of the estimated data set to 1 or 2. In addition, we use the mean-square error (MSE), defined as

$$\text{MSE} = \frac{1}{MN}||\hat{\mathbf{X}} - \mathbf{X}_0||_F, \tag{4.11}$$

to quantify the quality of the estimated data set. The MSE of input gather is 11.4×10^{-3}.

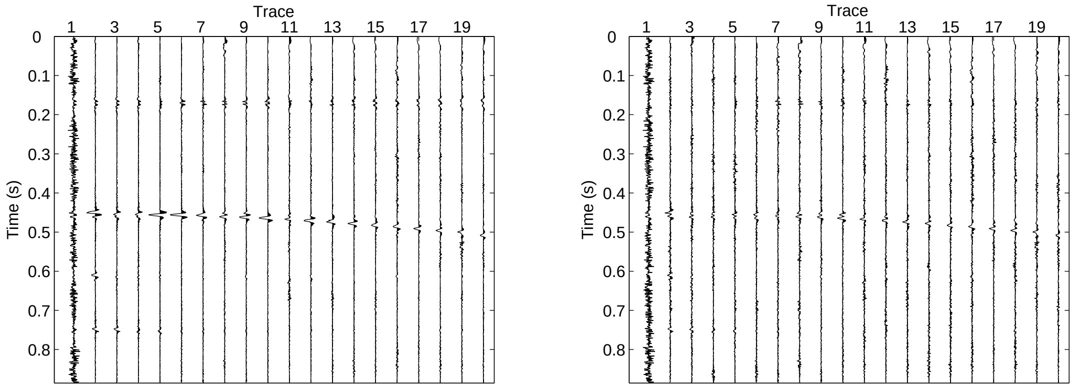
Figure 4.2 shows the output gather obtained after application of SVD locally, when using $p = 1$ singular value (a) and $p = 2$ singular values (b). Figure 4.3 shows the estimation error, i.e., the difference between output $\hat{\mathbf{X}}$ and noise-free $\mathbf{X}_0$ gathers after application of SVD locally for $p = 1$ singular value (a) and $p = 2$ singular values (b). When only one singular value is used, local SVD partly fails to retrieve the second reflection that is under NMO corrected. It also fails to retrieve completely the other reflections, as observed in Figure 4.3 (a), but attenuates random noise pretty well as long as its variance is about the same. Using instead the first two singular values to reconstruct the estimated data set, yields a much better reflections retrieval but at the cost of higher random noise, as in Figure 4.2 (b) and 4.3 (b). The MSE of output gathers is 4.50×10^{-3} for both $p = 1$ and $p = 2$, and are therefore significantly reduced. The choice of p is however a trade-off between reflections retrieval and random noise attenuation.



(a) $p = 1$ singular value.

(b) $p = 2$ singular values.

Figure 4.2: Signal enhancement of a stacked section using local SVD: Output gather for (a) $p = 1$ and (b) $p = 2$.



(a) $p = 1$ singular value.

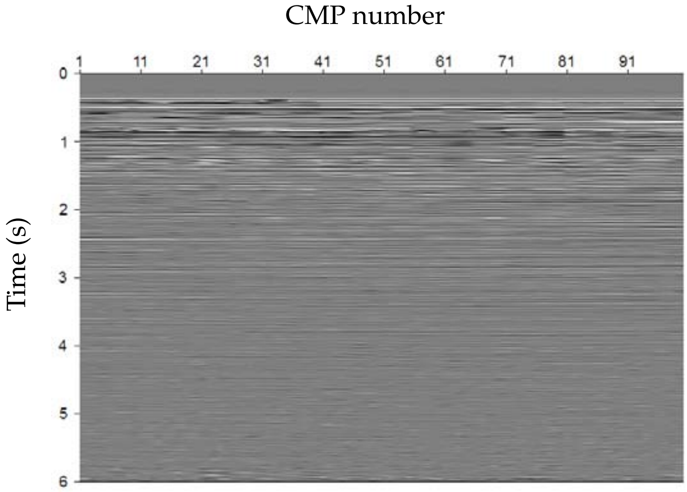
(b) $p = 2$ singular values.

Figure 4.3: Signal enhancement of a stacked section using local SVD: Difference between output and noise-free gathers for (a) $p = 1$ and (b) $p = 2$.

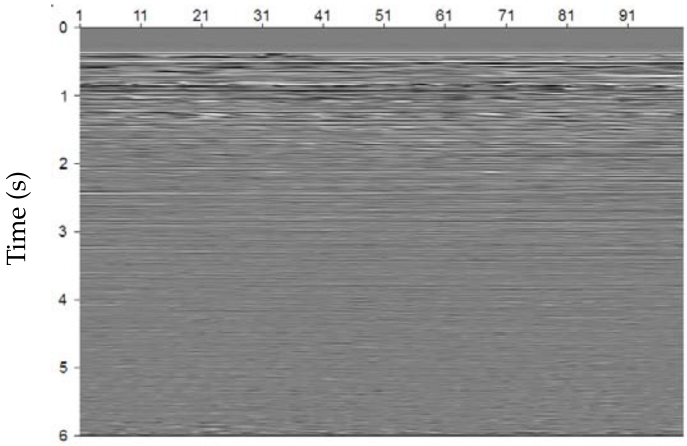
An application of local SVD on real data is shown in Chapter 5, Section 5.5.2. In this case, four singular values are used to reconstruct a stacked section. The result is compared with other noise attenuation methods that are *f-x* deconvolution and sparse code shrinkage. Let us consider the same data set and compare the outputs for $p = 2$ and $p = 4$ singular values, shown respectively in Figure 4.4 (a) and (b). Their residual plots, i.e., the difference between output and input plots, are shown in Figure 4.5. These two examples clearly illustrate the trade-off between reflections retrieval and random noise attenuation: when $p = 2$, the denoising ability is obvious but lots of seismic reflections are attenuated, especially in shallow waters. Therefore, one would prefer to increase the number of singular values to limit the signal leakage even if the noise attenuation ability decreases. Thus, in this example, using $p = 4$ singular values is necessary to ensure a good signal recovery.

4.5 Summary

In this chapter, we presented a multivariate analysis technique as a method to extract coherent events from a seismic data set. Thus, we begun with the description of principal component analysis and explained in what way it was related to singular value decomposition. SVD is found to be more computationally efficient to find the principal components, and provides also some additional insights, especially graphical and algebraic. These methods find the directions of greater variation in the data, and by restricting the dimensions of the data to those directions, we should be able to extract seismic coherent events. However, we showed through different examples on both synthetic and real data, that more than two singular values are usually necessary to prevent lots of seismic reflections from being attenuated, allowing more noise to be included in



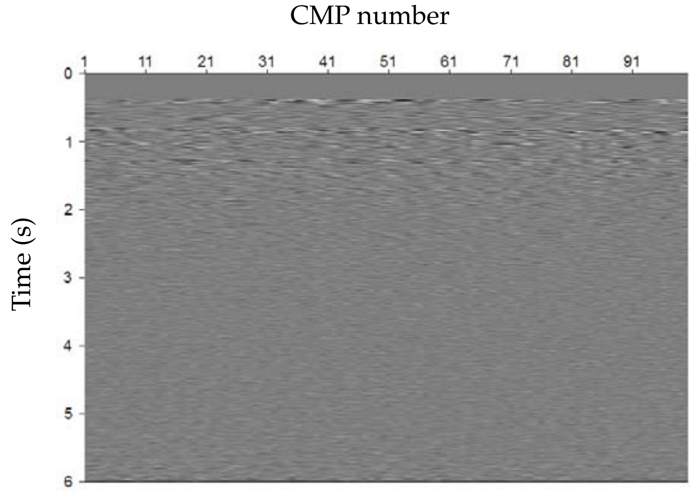
(a) $p = 2$ singular values.



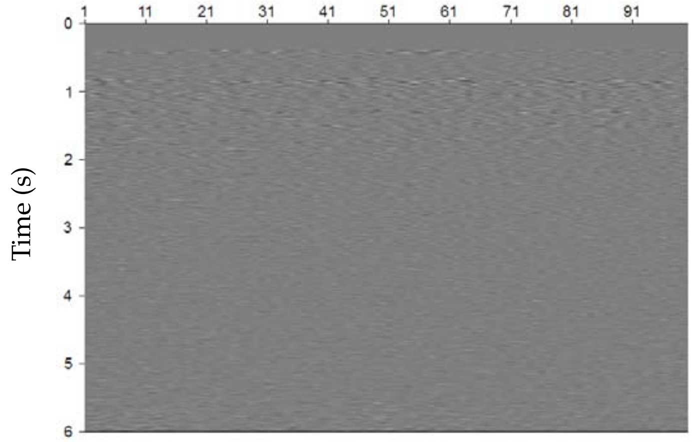
(b) $p = 4$ singular values.

Figure 4.4: Signal enhancement of a stacked section using local SVD: Results for (a) $p = 2$ and (b) $p = 4$.

the reconstructed image. Thus, we demonstrated the noise attenuation ability of this method as well as its limitations.



(a) Residual plot for $p = 2$.



(b) Residual plot for $p = 4$.

Figure 4.5: Signal enhancement of a stacked section using local SVD: Residual plots for (a) $p = 2$ and (b) $p = 4$.

Chapter 5

Sparse code shrinkage for signal enhancement of seismic data

5.1 Introduction

Denoising seismic signals buried in noise is one of the main issues in seismic. Some of the most popular methods used for that purpose are *f*-*x* predictive deconvolution (Canales, 1984), median filtering (Bednar, 1983), and decomposition into principal components whose one main application is the local singular value decomposition (Andrews and Patterson, 1976; Ulrych et al., 1988). In this chapter, we suggest a new approach by using sparse code shrinkage and we propose a method to obtain the necessary "noise-free" realization of the data. Sparse code shrinkage consists in applying a soft thresholding or shrinkage to the sparse representation of the data. The noise-free realization, that is in fact noise-attenuated and that aims at having the same statistical properties as the seismic signal, allows us to derive shrinkage functions based on a density model and the statistical properties of the noise-free sparse code. Small amplitudes are thought to originate from zero-valued components influenced by noise and are therefore suppressed, while large amplitudes are preserved (Hyvärinen et al., 1999).

Sparse coding is a method that is commonly used for noise reduction in image processing (Hyvärinen et al., 1999; Jenssen et al., 2001) and that has recently found some applications in seismic for random noise attenuation and multiple removal (Kaplan et al., 2009). Kaplan et al. (2009), however, did not provide any noise-free realization of the data and they used instead a unique shrinkage function with arbitrary parameters to threshold the sparse representation of the data. Sparse code shrinkage is closely related to the method of wavelet shrinkage, but has the important benefit over wavelet methods that both the features and the shrinkage parameters are estimated directly from the data. The wavelet representation, on the other hand, relies heavily on certain mathematical properties that may be only weakly relate to the properties of the data (Donoho et al., 1995).

Sparse coding (Barlow, 1994; Olshausen and Field, 1996) is a method that finds a representation of data in a way that in this representation, each of the components is only rarely significantly active at the same time. Equivalently, this means that each component has a sparse distribution. A zero-mean random variable is called sparse when its distribution has a peak at zero and heavy tails. For all practical purposes, sparsity is equivalent to super-Gaussianity (Hyvärinen and Oja, 1997) or leptokurtosis (positive kurtosis) (Kendall and Stuart, 1958). Sparse coding is an adaptive method, meaning that the transform matrix is estimated for a given class of data so that the transformed components are as sparse as possible. Such a representation is closely related to independent component analysis (ICA). It has been proven that the estimation of an ICA data model can be reduced to the search for uncorrelated directions in which the components are as non-Gaussian as possible (Comon, 1994). If the independent components are sparse, this amounts to the search for uncorrelated directions that have as sparse distributions as possible. Thus, the estimation of ICA model for sparse data is roughly equivalent to sparse coding if the components are constrained to be uncorrelated (Hyvärinen et al., 1999). Therefore, we will in this chapter use the constrained ICA model as a representation for sparsely coded data. ICA has already found some successful applications in geophysics. Vrabie et al. (2004) combined ICA with SVD in order to relax the orthogonality constraint of the propagation vectors, and thus he obtained a better separation of the source waves in the signal subspace. Lu (2006) presented an adaptive multiple subtraction algorithm based on ICA, where the problem of adaptive multiple subtraction was expressed as a blind source separation problem with two mixtures (the seismic data and the predicted multiples) of two or more sources (primaries and multiples). Finally, Van der Baan (2006) used ICA to achieve PP/PS wavefield separation by exploiting statistical differences between P and S waves.

A seismic data set consists of a collection of seismic traces that have a very high degree of correlation when seismic events are aligned. Its transformation into an ICA representation yields some components that partially contain some seismic signals and others that consist mostly of noise. Therefore, we can expect super-Gaussian densities, i.e., densities with positive normalized kurtosis, for the components that partially contain some seismic signals, while the others should follow a Gaussian distribution (Sanchis and Hanssen, 2008). This difference in probability density function is a very important feature since it also constitutes the basics for a signal detector. In addition, knowledge of the a posteriori probability density makes it possible to design a shrinkage function and therefore, a maximum a posteriori (MAP) estimator for the seismic signal. However, the user has to provide a density model that suits the ICA-transformed data and that will be fed into the maximum a posteriori estimator. Thus, we begin in Section 5.2 with a brief review of the theory of the MAP estimator for any model of ICA-transformed data. In Section 5.3, we will present two alternatives to provide a density model. The first solution consists of fitting the ICA-transformed data with three possible parametric models: Laplace and mildly sparse densities (Hyvärinen et al., 1999), normal inverse Gaussian density (NIG) (Jenssen et al., 2001; Øigård et al., 2005) and mixture of generalized Gaussian densities (MGG) (Mohamed and Jaïdane-Saïdane, 2009). In the second solution, we consider a non-parametric density model that is estimated directly from the data by mean of a Gaussian kernel estimator. Finally, in Section 5.5, we test the performance of sparse code shrinkage on both synthetic and real marine seismic data, together with two others denoising methods that are *f-x* deconvolution and local SVD.

5.2 Maximum a posteriori estimation of sparse coded signal

5.2.1 ICA representation

The independent component analysis is a statistical and computational technique used to reveal hidden factors that underlie sets of random variables, measurements, or signals. ICA is based on the assumption that data signals are linear mixtures of some unknown variables, while the mixing system is also unknown. The unknown variables are assumed non-Gaussian and mutually independent, and they are called the independent components of the observed data. These independent components, also called sources, can be found by ICA. In other words, ICA finds a transformation of signal mixtures that produces independent signals sources, on the assumption that each of the independent signal sources is generated by a different physical process. Let us consider *n* independent random variables denoted by $s_1, \ldots, s_n$ and *m* observations or linear mixtures $x_1, \ldots, x_m$ of these *n* independent variables so that:

$$x_j = a_{j1}s_1 + a_{j2}s_2 + \ldots + a_{jn}s_n, \text{ for } j = 1, \ldots, m. \tag{5.1}$$

We denote the set of observations by $\mathbf{x} = (x_1, \ldots, x_m)^T$, the set of sources by $\mathbf{s} = (s_1, \ldots, s_n)^T$ and the matrix with elements a_{ij} by **A**. Then, the transformation from **s** to **x** defines the *mixing* matrix **A** such that:

$$\mathbf{x} = \mathbf{A}\mathbf{s}. \tag{5.2}$$

The problem consists in recovering the source vector **s** from the observations vector **x**, by searching for an (*n* x *m*) *unmixing* matrix **W** that makes the components of **s** as sparse as possible:

$$\mathbf{s} = \mathbf{W}\mathbf{x}. \tag{5.3}$$

When estimating the unmixing matrix **W**, we use the FastICA algorithm developed by Hyvärinen and Oja (1997) with the hyperbolic tangent nonlinearity and the symmetric decorrelation approach. The ICA representation is then obtained by multiplying the unmixing matrix **W** with the input signals.

Sparse coding is closely related to independent component analysis, as the transform makes the components of **s** as non-Gaussian and jointly statistically independent as possible. Thus, the estimation of the ICA data model can be reduced to the search for uncorrelated directions in which the components are as non-Gaussian as possible (Comon, 1994). If the independent components are sparse, this amounts to the search for uncorrelated directions that have as sparse distributions as possible. We consider here that supergaussianity is sufficient to provide sparse coding when data are assumed to be generated by a noisy ICA model. This is not exactly true, as supergaussianity is necessary but not sufficient to enforce sparsity, but it has been admitted in a variety of simulations and applications (Kreutz-Delgado et al., 1999). Thus, if the components are constrained to be uncorrelated, the estimation of ICA model for sparse data is roughly equivalent to sparse coding (Hyvärinen et al., 1999).

When observing noisy data, sparse coding separates the statistics of the noise free signal from the noise in the transform domain. In the case of a signal $\mathbf{x}$ corrupted by white Gaussian noise $\mathbf{n}$ of zero mean and variance σ^2, then the ICA transformation yields

$$\mathbf{y} = \mathbf{W}(\mathbf{x} + \mathbf{n}) = \mathbf{s} + \mathbf{v} \tag{5.4}$$

where $\mathbf{s}$ is maximally sparse and $\mathbf{v}$ is the ICA transform of the noise. To ensure the noise components to be Gaussian and uncorrelated, $\mathbf{W}$ needs to be orthogonal. This can be achieved by applying $\mathbf{W} \leftarrow \mathbf{W}(\mathbf{W}^T\mathbf{W})^{-1/2}$ (Hyvärinen et al., 1999). However, one could also consider the case where $\mathbf{W}$ has fewer rows than columns, FastICA for example supports such estimation. Then, one estimates a representation, consisting of the rows of $\mathbf{W}$, which is such that the rows of $\mathbf{W}$ are orthogonal to each other. The orthogonalization step is then achieved by $\mathbf{W} \leftarrow (\mathbf{W}\mathbf{W}^T)^{-1/2}\mathbf{W}$.

5.2.2 Maximum likelihood estimator

In this subsection, we consider only scalar random variables, but the results can be generalized for n-dimensional random vectors (Hyvärinen et al., 1999). The multivariate case seems however to be of little importance in practice because sparse components are considered separately. Let us consider the scalar model

$$x = s + \nu \tag{5.5}$$

where s and ν are two independent random variables that denote respectively the non-Gaussian sparse component and the Gaussian noise of zero mean and variance σ^2. The probability density function of ν is denoted by $f_\nu(x)$.The maximum a posteriori likelihood estimator $\hat{s}$ of s is the value of s that maximizes $f(s|x)$, the conditional density of the observation s given x, also called the posterior probability of s. Thus, the MAP estimate of s is given by

$$\frac{\partial}{\partial s} \ln f(\hat{s}|x) = 0. \tag{5.6}$$

According to the Bayes rule for densities, the conditional density of the observation s given x is expressed as

$$f(s|x) = \frac{f(x|s)f(s)}{f(x)}. \tag{5.7}$$

Maximizing the posterior probability of s is equivalent to maximizing the product $f(x|s)f(s)$ since $f(x)$ is independent of the parameter s. Furthermore (proof given in

Appendix),

$$f(x|s) = f_\nu(x - s). \tag{5.8}$$

Therefore, the MAP estimate $\hat{s}$ is obtained by inserting the product $f(x|s)f(s)$ and the form of conditional density of x given s, equation 5.8, into the maximum likelihood equation 5.6:

$$\frac{\partial}{\partial s}[\ln f_\nu(x - \hat{s}) + \ln f(\hat{s})] = 0. \tag{5.9}$$

Let us denote the negative log-density of s by $i(s) = -\ln f(s)$ and the score function of s by $i'(s) = \frac{\partial}{\partial s} i(s)$ (i is assumed to be strictly convex and differentiable). Then, equation 5.9 becomes

$$\frac{\hat{s} - x}{\sigma^2} + i'(\hat{s}) = 0. \tag{5.10}$$

The MAP estimate $\hat{s}$ is a solution of equation 5.10, i.e., a function of the observed signal x exists so that $\hat{s} = g(x)$. This function g is the shrinkage function that discriminates the noise and enhances the sparse component into the noisy observation x. For certain densities, equation 5.10 cannot be solved in closed form, but it is always possible to use the first-order approximation $i'(\hat{s}) \simeq i'(x)$ in equation 5.10, to obtain the MAP estimator

$$\hat{s} = x - \sigma^2 i'(x). \tag{5.11}$$

However, i' is often discontinuous or it can even be singular at 0 which implies that the first-order approximation is quite inaccurate near 0. To overcome this problem of "over-shrinkage", as described by Efron and Morris (1975), we use the following modification (Hyvärinen et al., 1999):

$$\hat{s} = \text{sgn}(x) \max(0, |x| - \sigma^2 |i'(x)|). \tag{5.12}$$

Thus, equation 5.12 is an approximation of the maximum likelihood (ML) estimator of a non-Gaussian random variable corrupted by Gaussian noise. Hyvärinen et al. (1999) showed that the ML estimator has very good denoising capability and it can lead to significant improvement for non-Gaussian (e.g., sparse) variables. Thus, the more non-Gaussian the variable s is, the better Gaussian noise is reduced. However, if s is Gaussian, then allowing nonlinearity in the estimation does not improve the performance and the ML estimator is then equal to the best linear estimator in the mean-square sense.

5.3 Shrinkage functions

To use the estimator defined by equation 5.11 in practice, the densities of the $\mathbf{s}_i$ need to be modeled with a parametrization that is rich enough. In this section, we propose different parametrizations that we think could describe the densities encountered in seismic. We first present sparse densities that are able to model different degrees of super-Gaussianity and different scales, i.e., variances. Sparse densities are commonly used with sparse code shrinkage. We continue with the normal inverse Gaussian density. This density has a flexible parametrization that makes it possible to model a wider variety of distributions than sparse densities. Indeed, the normal inverse Gaussian density models a large variety of unimodal shapes with various decay rates of the tail and skewness properties. Finally, we think that multi-modal densities should be investigated as well. The mixture of Generalized Gaussian densities can fit super-Gaussian densities and is therefore an appropriate density model candidate for seismic components. In addition, we test a non-parametric model that does not require us to make any assumption about the parametric family and therefore, that allows various shapes of densities. Thus, we use a Gaussian kernel estimator to estimate the density $f(s)$.

5.3.1 Sparse densities

Two density models are suggested to suit super-Gaussian densities with the Laplace distribution as the limit between both models. The first density model is suitable for super-Gaussian densities that are not sparser than the Laplace distribution. Its probability density function is given by

$$f(s) = C\exp(-as^2/2 - b|s|), \tag{5.13}$$

where *a*,*b* > 0 are parameters to be estimated and *C* is an irrelevant scaling factor. To estimate the parameters *a* and *b*, Hyvärinen et al. (1999) assumes that the score function, i.e., the derivative of the negative log-density of s, $i'(s) = as + \mathrm{sgn}(s)b$ is a linear combination of two functions where one of them is linear. Thus, after some tedious algebraic manipulations, Hyvärinen et al. (1999) obtains

$$\begin{aligned} b &= \frac{2f_s(0)E\{s^2\} - E\{|s|\}}{E\{s^2\} - E\{|s|\}^2}, && (5.14) \\ a &= \frac{1}{E\{s^2\}}[1 - bE\{|s|\}], && (5.15) \end{aligned}$$

where $E\{\}$ denotes the expected value but can be replaced here by sample average, whereas $f_s(0)$ is the value of the density function of s at zero,and can be estimated, for example, by using a Gaussian kernel estimator (Silverman, 1986) at 0. Let σ^2 be the noise variance, then the corresponding shrinkage function *g* for this model is given by

$$g(x) = \frac{1}{1 + a\sigma^2}\mathrm{sgn}(x)\max(0, |x| - b\sigma^2). \tag{5.16}$$

The second density model is suitable for super-Gaussian densities that are sparser than the Laplace distribution. Its probability density function is given by

$$f(s) = \frac{1}{2d}\frac{(\alpha+2)[\alpha(\alpha+1)/2]^{\alpha/2+1}}{[\sqrt{\alpha(\alpha+1/2)} + |s/d|]^{\alpha+3}} \tag{5.17}$$

where *d* is a scale parameter and α a sparsity parameter. Using the same assumption as in the first density model, Hyvärinen et al. (1999) obtains

$$\begin{aligned} d &= \sqrt{E\{s^2\}}, && (5.18)\\ \alpha &= \frac{2-k+\sqrt{k(k+4)}}{2k-1} \text{ with } k = d^2 f_s(0)^2. && (5.19) \end{aligned}$$

Let again σ^2 be the noise variance, if we set $a = \sqrt{\alpha(\alpha+1)/2}$ and $Q = (|x| + ad)^2 - 4\sigma^2(\alpha+3)$, then we shrink using:

$$g(x) = \begin{cases} \operatorname{sgn}(x)\max\left(0, \frac{|x|-ad}{2} + \frac{1}{2}\sqrt{Q}\right) & \text{if } Q \geq 0,\\ 0 & \text{if } Q < 0. \end{cases} \quad (5.20)$$

The choice of first or second models can be made by calculating the moments of the distributions, i.e., if

$$\sqrt{E\{s^2\}} f_s(0) < \frac{1}{\sqrt{2}}, \quad (5.21)$$

then the first model should be used. Otherwise, use the second model (Hyvärinen et al., 1999).

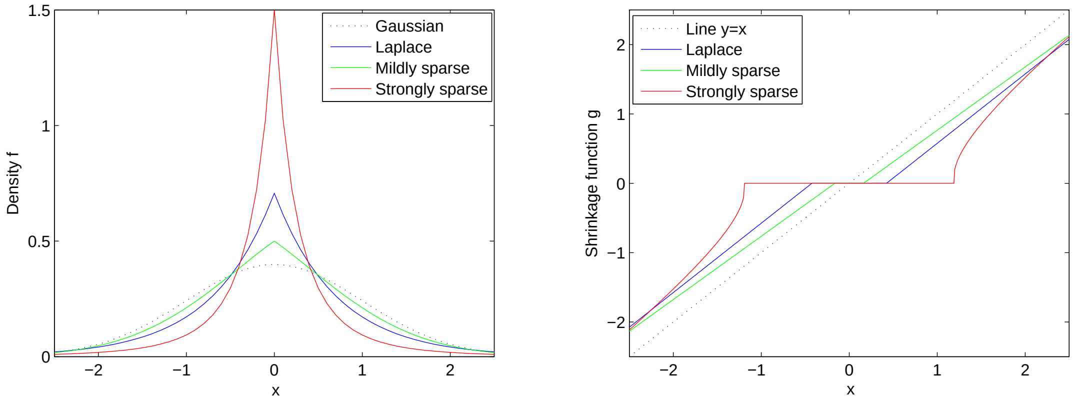
In order to give some insight as to what the specific density functions imply, we show in Figure 5.1 (a) three sparse density functions: mildly sparse, Laplace and strongly sparse, together with the Gaussian distribution for comparison. All the densities were normalized to unit variance. We show in Figure 5.1 (b) the shrinkage functions associated with the three sparse distributions, together with the line $y = x$ for comparison. In this plot, the noise variance σ^2 was fixed to 0.3. Small amplitudes are considered to originate from zero-valued components influenced by noise and are therefore suppressed, while large amplitudes are preserved. In addition, the shrinkage functions reduce the absolute value of its argument by a certain amount which depends on the noise variance σ^2. Finally, the sharper the distribution is, the more likely small values are to be pure noise.

5.3.2 The normal inverse Gaussian density (NIG)

Another possible model for sparsely coded data is the NIG density. The NIG is suitable for modeling heavy-tailed random process (Hanssen and Øigård, 2001a). In addition, it is characterized by a set of four parameters that makes it to be very flexible. NIG has recently found many applications in electrical engineering and in particular, it was used as wavelet coefficients of synthetic aperture radar images (Solbø and Eltoft, 2004).
The NIG density is a variance-mean mixture of a Gaussian density with an inverse Gaussian. Its probability density function is given by

$$f(s) = \frac{\alpha\delta}{\pi} \frac{\exp\left[p(s)\right]}{q(s)} K_1[\alpha q(s)] \quad (5.22)$$

where



(a) (b)

Figure 5.1: Example of super-Gaussian densities (a) and their associated shrinkage function (b). For comparison, the Gaussian distribution is drawn in dot line in (a), and the line $y = x$ is drawn in dot line in (b). The sharper the distribution is, the more likely are small values to be pure noise and set to zero.

$K_1(s)$ is the modified Bessel function of the second kind with index 1,

$p(s) = \delta\sqrt{\alpha^2 - \beta^2} + \beta(s - \mu)$,

$q(s) = \sqrt{(s - \mu)^2 + \delta^2}$, $0 \leq |\beta| < \alpha$, $\delta > 0$ and $-\infty < \mu < \infty$.

The NIG density is completely defined by a set of four parameters $\alpha, \beta, \delta, \mu$ and is very suitable for unimodal, positive kurtosis data. Here, α controls the steepness of density, β controls the skewness, δ is a scale parameter, while μ is the mean value of samples s. There are two ways of deriving the parameter values: direct calculation and expectation-maximization (EM) algorithm. A direct calculation is presented by Hanssen and Øigård (2001b). They derive a cumulant based estimator of the NIG parameters, where a cumulant of order n for the random variable s is defined as

$$\kappa^{(n)} = (-j)^n \frac{d^n \Psi_s(\omega)}{d\omega^n}|\omega = 0 \quad \text{where} \quad \Psi_s(\omega) = \ln E(e^{j\omega s}). \tag{5.23}$$

The cumulants are easily computed thanks to their relationship with moments. If we define the moment of order n about the origo of the random variable s by $m_s^{(n)} = E\{s^n\}$, then the cumulants of order 1 to 4 are obtained from:

$$\begin{aligned}
\kappa^{(1)} &= m_s^{(1)} && (5.24)\\
\kappa^{(2)} &= m_s^{(2)} - (m_s^{(1)})^2 && (5.25)\\
\kappa^{(3)} &= m_s^{(3)} - 3m_s^{(1)}m_s^{(2)} + 2(m_s^{(1)})^3 && (5.26)\\
\kappa^{(4)} &= m_s^{(4)} - 3(m_s^{(2)})^2 - 4m_s^{(1)}m_s^{(3)} + 12(m_s^{(1)})^2 m_s^{(2)} - 6(m_s^{(1)})^4. && (5.27)
\end{aligned}$$

Thus, first, we form the estimates $\kappa^{(1)}$, $\kappa^{(2)}$, $\kappa^{(3)}$ and $\kappa^{(4)}$ of the four lowest order cumulants from data samples. Thereafter, the third and fourth order standardized moments, namely the skewness and the kurtosis, are estimated by

$$\hat{\gamma}_3 = \frac{\kappa^{(3)}}{\left[\kappa^{(2)}\right]^{3/2}} \qquad \text{and} \qquad \hat{\gamma}_4 = \frac{\kappa^{(4)}}{\left[\kappa^{(2)}\right]^{2}}. \tag{5.28}$$

In addition, we form two auxiliary variables as

$$\xi = 3\left(\hat{\gamma}_4 - \frac{4}{3}\hat{\gamma}_3^2\right)^{-1} \qquad \text{and} \qquad \rho = \frac{\hat{\gamma}_3}{3}\sqrt{\xi}. \tag{5.29}$$

Finally, Hanssen and Øigård (2001b) show that the four parameters can be estimated by

$$\hat{\delta} = \sqrt{\kappa^{(2)}\xi(1-\rho^2)} \tag{5.30}$$

$$\hat{\alpha} = \frac{\xi}{\hat{\delta}\sqrt{(1-\rho^2)}} \tag{5.31}$$

$$\hat{\beta} = \hat{\alpha}\rho \tag{5.32}$$

$$\hat{\mu} = \kappa^{(1)} - \rho\sqrt{\kappa^{(2)}\xi}. \tag{5.33}$$

This estimation technique yields statistically consistent estimators. However, it is important that data sets are large enough in order to get a fairly accurate estimation of the parameters. Alternatively, Øigård et al. (2005) proposes a simple EM algorithm for the estimation of the parameters of the multivariate NIG distribution. The heavy-tailed multivariate normal inverse Gaussian (MNIG) distribution is a variance-mean mixture of a multivariate Gaussian with a univariate inverse Gaussian distribution. The conventional way of estimating the MNIG parameters has been the maximum likelihood estimation method, but because of the complexity of the likelihood function, parameter estimation by direct maximization is exceedingly difficult. To overcome this problem, Øigård et al. (2005) proposed a fast and accurate multivariate expectation-maximization (EM) algorithm for maximum likelihood estimation of the scalar, vector, and matrix parameters of the MNIG distribution. Thus, restricting this EM algorithm to the univariate case provides us the NIG distribution parameters. We will prioritize the EM algorithm in our implementation of the sparse code shrinkage code. However, if it does not converge within a reasonable number of iterations, then we use the cumulant based parameter estimation technique described above.

The score function of the NIG density is given by

$$i'(s) = \frac{\alpha(s-\mu)}{q(s)}\left(\frac{K_0[\alpha q(s)]}{K_1[\alpha q(s)]} + \frac{2}{\alpha q(s)}\right) - \beta \tag{5.34}$$

where K_0 is the modified Bessel function of the second kind with index 0. The maximum a posteriori probability estimate is obtained by inserting the score function in equation 5.34 into the MAP estimate in equation 5.12.

5.3.3 Mixture of Generalized Gaussian densities (MGG)

When examining the distribution of some of the sparse components, we notice that some of them look bivariate because it contains both noise (that is usually normally distributed) and signals (that has a different mean and statistical properties). For these reasons, we believe that a multi-modal density model should be tested. The density function of a mixture of Q generalized Gaussian distributions is given by:

$$f_\theta(s) = \sum_{k=1}^{Q} \alpha_k f_k(s|\theta_k) \quad \text{where} \quad f_k(s|\theta_k) = \frac{\gamma_k c_k}{\Gamma(1/c_k)}\, e^{-\gamma_k^{c_k} |s-\mu_k|^{c_k}}. \tag{5.35}$$

Here,

$\gamma_k = \frac{1}{\sigma_k} \frac{\Gamma(3/c_k)^{1/2}}{\Gamma(1/c_k)}$ is a scale parameter,

Γ represents the Gamma function,

$\theta_k = (\alpha_k, \mu_k, \sigma_k, c_k), \quad k = 1, \ldots, Q,$

α_k is the k^{th} mixture weight that satisfies $\alpha_k \geq 0$ and $\sum_{k=1}^{Q} \alpha_k = 1$,

μ_k is the k^{th} mean,

σ_k is the k^{th} standard deviation,

c_k is the k^{th} shape parameter that can characterize a large class of distribution including super-Gaussian distribution.

We use an expectation-maximization (EM) algorithm that has been reformulated in the case of MGG model (Mohamed and Jaïdane-Saïdane, 2009) for the estimation of coefficients and the distribution's parameters. The score function of the MGG density is given by

$$i'(s) = -\frac{\sum_{k=1}^{Q} \alpha_k f_k'(s|\theta_k)}{\sum_{k=1}^{Q} \alpha_k f_k(s|\theta_k)} \qquad \text{with} \qquad f_k'(s|\theta_k) = -\gamma_k^{c_k} c_k\, |s-\mu_k|^{c_k-1} f_k(s|\theta_k) \tag{5.36}$$

and the maximum a posteriori probability estimate is obtained by inserting the score function in equation 5.36 into the MAP estimate in equation 5.12.

5.3.4 Gaussian kernel density estimator

We motivate the investigation of a non-parametric density by the desire to test the case where no prior assumption about the parametric family are made. The final results will then be compared with the parametric models. The histogram is the simplest and most widely used estimator of densities. Despite its widespread use, the histogram has several fundamental drawbacks. In particular, it yields a discontinuous function and it is sensitive to the choice of origin (Silverman, 1986). To overcome these many non-desirable properties, we have chosen to estimate the density by means of the so-called Parzen kernel estimator. The kernel estimator of density has good statistical properties, and it is well suited for the analysis of short data segments. Basically, the kernel estimator is constructed by placing a smooth and symmetric normalized function (a "kernel") with its origin at each data point. By summing this collection of normalized functions, we obtain a smooth and statistically consistent estimate of the density. A more exhaustive description of the kernel estimator and its statistical properties is given in Section 2.4.

The kernel estimator of the pdf at amplitude s, given N data samples $s_0, s_1, \ldots, s_{N-1}$, can be written as (Silverman, 1986; Wand and Jones, 1995)

$$\widehat{f}(s) = \frac{1}{N} \sum_{n=0}^{N-1} K_h \left(s - s_n\right). \tag{5.37}$$

Here, $K_h(\xi) \equiv (1/h)K(\xi/h)$, where $K(\xi)$ is the so-called smoothing kernel and h is a scaling parameter that controls the degree of smoothing. For this application, we choose the Gaussian smoothing kernel $K(\xi) = (1/\sqrt{2\pi}) \exp(-\xi^2/2)$ and the value of h is obtained by minimizing the integrated mean-squared error with referral to the Gaussian distribution.

The score function of the density estimate when using a Gaussian kernel is given by

$$i'(s) = -\frac{\widehat{f}'(s)}{\hat{f}(s)} \qquad \text{with} \qquad \widehat{f}'(s) = -\frac{1}{N} \sum_{n=0}^{N-1} \frac{(s - s_n)}{h^2} K_h \left(s - s_n\right) \tag{5.38}$$

and the maximum a posteriori probability estimate is obtained by inserting the score function in equation 5.38 into the MAP estimate in equation 5.12.

5.4 Sparse code shrinkage algorithm

5.4.1 Noise-free realization of the data

We need a realization of noise-free data to get an accurate estimation of the density model parameters. In seismic, such a realization is impossible to obtain. The underwater environment is naturally noisy and in addition, measurements always generate extra noise. To overcome this problem, we seek a noise-reduced realization of data that has the same statistical properties as the seismic signals. We can naturally assume that the better this estimation is, the more accurate the parameters estimates are. However, accuracy of model parameters is not a critical criterion in sparse code shrinkage. We use a novel stacking method with signal-to-noise ratio based reference trace (C. Sanchis and A. Hanssen, 2010, Enhanced stacking method based on local correlation: Geophysics, Submitted), applied locally. This method consists of a time-dependent weighted average that utilizes local correlation between each individual trace and a chosen reference trace as a measure of weight. Here, the reference trace results from weighted stacking, where the weights are proportional to the estimated S/N of each individual trace. This method, which enhances the coherent signals and attenuates the random ones, preserves the statistical properties of the seismic signals. In addition, it presents several other advantages: It is easy to implement, it is not computationally demanding, and it has proven to estimate events that are not completely flat quite well. The entire seismic section is scanned by a sliding window that typically consists of 3 normal move-out (NMO) corrected traces and 50 time samples, with possible overlap between windows. The stacking method is applied to the traces contained in each window to produce one averaged trace. If the sliding windows overlap, then we take the result to be the mean of all averaged traces. By decreasing the window size and overlap percentage, we reduce the averaged volume of data to be averaged and therefore, increase the variance. On the other hand, we increase the resolution in the sense that it better reflects the local situations. Thus, there is a trade-off between resolution and variance. Our experience showed us that the resolution achieved by averaging 3 trace segments was sufficient and the noise variance acceptable in order to obtain a good noise-free realization. In addition, there is another reason to choose a small number of windowed traces, that is the necessity to obtain a noise-free realization that has about the same dimensions as the section to be denoised. Because of the averaging, we have available fewer averaged traces than the input seismic section. It is however reasonable to assume that the missing traces have the same statistical properties as the last averaged one. In addition, note that the same noise-free realization can be used for denoising of any CMP gather as long as the geology remains unchanged.

5.4.2 Estimation of noise variance

An estimation of noise variance is required for the estimation of the density parameters. We suggest two ways of estimating it. First, we consider the difference between input data and "noise-free" data to be a rough estimate of the noise, we calculate an unbiased estimation of the standard deviation and we correct this number by a factor $\frac{2}{3}$ known from experience for "noise-free" generated by our local stacking method. This correction is required since it still contains noise. Another alternative consists in taking the median absolute deviation of the y_j corresponding to the sparsest noise free component s_j, and dividing it by 0.6745 (Donoho et al., 1995).

5.4.3 Algorithm

The estimation of the ICA transform matrix **W** requires a large number N of realizations of "noise-free" random data vectors. To aid in this estimation, we define a small window that slides over the whole "noise-free" gather with possible overlap. A row-by-row scanning of each window transforms a two-dimensional signal into a one-dimensional $(m \times 1)$ column vector $\mathbf{x}_i$. Thus, we reorganize the "noise-free" gather into a set of N

vectors in $\mathbf{X}_{nf} = [\mathbf{x}_1, ..., \mathbf{x}_N]$. These vectors are used to determine the ICA transform matrix **W** by means of any ICA algorithm that is then orthogonalized. The $(m \times 1)$ sparse vectors in $\mathbf{S} = [\mathbf{s}_1, ..., \mathbf{s}_N]$ result from the ICA transformation of vectors in **X** such that $\mathbf{S} = \mathbf{W}\mathbf{X}_{nf}$. The first element of each vector $\mathbf{s}_i, i = 1, ..., N$ is a realization of the first sparse component, the second element of each vector is a realization of the second sparse component, and so on. Each sparse component is then fed into a density parameter estimator which defines a shrinkage function that is fitted to the noise-free sparse component. Once all the shrinkage functions have been determined, we transform each noisy data vector **x+n** into its sparse representation **y**. Denoising of each component of **y** is then achieved by applying the shrinkage function that is associated to the same noise-free sparse component. The sparse code shrinkage algorithm is summarized as follows:

1. Generate a "noise-free" realization of the data by applying the enhanced stacking method locally to NMO corrected data, and reorganize the estimated "noise-free" gather into a set of vectors $\mathbf{X}_{nf} = [\mathbf{x}_1, ..., \mathbf{x}_N]$.

2. Apply any ICA algorithm to $\mathbf{X}_{nf}$ to determine the transformation matrix **W** so that if $\mathbf{S} = \mathbf{W}\mathbf{X}_{nf}$, then **S** defines a set of independent components.

3. Orthogonalize the ICA transformation matrix **W** to ensure that noise components are Gaussian and uncorrelated.

4. For each sparse component $\mathbf{s}_j, j = 1, ..., m$ from $\mathbf{S} = \mathbf{W}\mathbf{X}_{nf}$, choose a density model and estimate its parameters and the associated shrinkage function g_j.

5. Reorganize the noisy seismic section into a set of noisy vectors and apply the sparse transformation to every noisy vector **x+n** :
$$\mathbf{y} = \mathbf{W}(\mathbf{x}+\mathbf{n}). \tag{5.39}$$

6. For each component **y**, apply the shrinkage function g_j to y_j to obtain $\hat{s}_j = g_j(y_j), j = 1, ..., m$.

7. Invert the sparse transformation to obtain a denoised estimate of **x**:
$$\hat{\mathbf{x}} = \mathbf{W}^T\hat{\mathbf{s}}. \tag{5.40}$$

8. Reshape the set of denoised vectors $[\hat{\mathbf{x}}_1, ..., \hat{\mathbf{x}}_N]$ back to two-dimensional seismic section.

Note that the entire input data set is normalized to zero-mean and unit variance so that the method is not affected by the global statistics of the data. Therefore, we provide the normalized noise variance as input parameter. In addition, each noisy data vector **x+n** is again normalized to zero-mean and unit variance in order to compensate for possible non-stationarity of the data.

5.5 Applications

In this section, we test sparse code shrinkage with the different density models presented earlier, together with another possibility, that is the choice of the mildly sparse density model with fixed parameters $a = b = 0.5$. This application of sparse code shrinkage was used by Kaplan et al. (2009) while no "noise-free" realization was available. In our real data examples, we will compare sparse code shrinkage to two other common denoising methods that are *f-x* deconvolution and local SVD. The Matlab code for *f-x* deconvolution comes from the seismic data processing package SeismicLab (Sacchi and Signal Analysis and Imaging Group, 2008). In order to obtain the "noise-free" realization, we use the enhanced stacking method applied locally with a sliding window that contains 3 traces and 50 time samples, with 50 % overlap along time between windows.

5.5.1 Test on synthetic data

We first examine the performance of sparse code shrinkage for different density models with a synthetic example shown in Figure 5.2 (a). It is composed of 20 traces and 885 samples, with four NMO-corrected reflections that result from the convolution of a Ricker wavelet with a reflection sequence of decreasing amplitude and corrupted by different types of noise. The Ricker wavelet is also known in the literature as the Mexican hat wavelet and it is defined by $\psi(t) = \frac{1}{\sqrt{2\pi\beta^3}}(1 - \frac{t^2}{\beta^2})e^{-\frac{t^2}{2\beta^2}}$ with parameter $\beta^2 = 0.14$ in this chapter (Chui, 1992). The true signal without additive noise is shown in Figure 5.2 (b). The corrupting noise is zero-mean white Gaussian noise, where the noise variance is 0.04 for the first trace and 0.01 for the other ones. The maximum and minimum spike amplitudes are 3.41 and 1.42 for the first trace respectively, and 0.73 and 0.31 for trace 20 respectively. In addition, the second reflection has been NMO under-corrected on purpose so that it is slightly hyperbolic, with a delay between the first and the last trace equal to 43 ms. The goal here is to investigate the consequences of this inaccuracy on the final estimate. Finally, the third reflection is recorded by traces 1 to 12 only.

We extract data windows that contain 3 traces and 64 time samples (here 64 ms), and has 80 % overlap both along time and traces, to reorganize the seismic section into 192×1 vectors. The reason why we choose so few traces is to better estimate short and quickly varying events, and in particular, hyperbolic events. Several authors have applied ICA to image data, and found that one component represents the local mean. It was noted in Hyvärinen et al. (1999) that this component actually does not belong to a sparse density, and that it has a large variance associated with it. Therefore, we ignore this component by subtracting the local mean from each 192×1 data vector used in the experiment and drop one dimension by principal component analysis (PCA), before proceeding. Thus, the dataset consists of vectors of dimension 191×1. In this manner, we had 1980 realizations of "noise-free" data $\mathbf{X}_{nf}$ to be used in estimating the transform matrix $\mathbf{W}$ and the densities of the transformed components, and in calculating shrinkage functions. The estimated normalized noise variance is set to $\widehat{\sigma_N^2} = 0.7$. To quantify the estimation error, we calculate the mean square error, defined as $E[(\hat{\mathbf{X}} - \mathbf{X}_0)^2]$, where $\hat{\mathbf{X}}$ and $\mathbf{X}_0$ denote the estimated data set and the true noise-free data set respectively, given in Table 5.1.

The "noise-free" realization $\mathbf{X}_{nf}$ used to estimate the transform matrix $\mathbf{W}$ and the fitted shrinkage functions are shown in Figure 5.3 (a), and the estimation error, i.e., the difference between estimated and noiseless section, is shown in Figure 5.3 (b). As expected, local stacking allows the detection of events that are not perfectly flat, and yields a realization that is noise attenuated and whose signals are distorted and attenuated. This realization achieves a rather poor signal reconstruction, but because of its denoising ability, it is a rather good candidate for the "noise-free" realization.

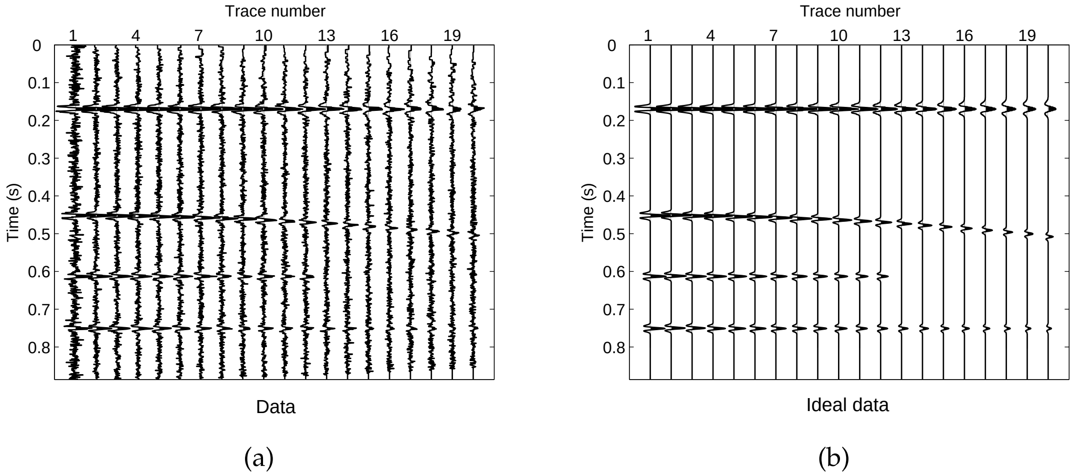
Application of sparse code shrinkage to the synthetic section and using different density models yields the following results: the estimated sections and the estimation error when using parametric models of densities are shown in Figure 5.3 (c) and (d) for sparse density, in Figure 5.3 (e) and (f) for NIG density, and in Figure 5.4 (a) and (b) for MGG density. When assessing the reconstructed section, we observe that these parametric models perform similarly. The estimation of events is reasonably good, even if we observe little signal leakage for the first event and for the slightly hyperbolic event. These two events actually are the ones that are the most distorted in the "noise-free" realization. Corrupting noise is significantly attenuated, even for the first trace. Thus, the MSE obtained with these three densities are almost equal (between 39×10^{-4} and 42×10^{-4}), and significantly reduced compared to the input section (120×10^{-4}). Figure 5.4 (c) and (d) shows the estimated section and estimation error for the non-parametric Gaussian kernel density estimate. There are very few differences with parametric densities: the first and second reflections are slightly more accurately estimated than with the parametric methods, but the corrupting noise is less attenuated. Thus, the MSE is a little larger than for the parametric models: 51×10^{-4} against 39×10^{-4}. Finally, whatever the choice of density, it is important to notice that the quality of the reconstructed signals achieved by sparse code shrinkage is much more accurate than in the "noise-free" realization.

Let us now imagine that we do not have access to any "noise-free" realization. We have therefore no other choice that using an arbitrary density model with fixed parameters, e.g., the mildly sparse density with non-fitted parameters $a = b = 0.5$. The estimated section and the estimation error are shown in Figure 5.4 (e) and (f). The events are recovered although distorted and attenuated, but the additive noise is hardly attenuated. Because the shrinkage functions are not specifically fitted to the data, we need to choose a unique set of parameters (a, b) that allows the event reconstruction. Thus, small parameter values preserve the events, but limit the shrinkage ability of the noisy sparse components. The MSE is twice that of other density models.

Let us now find out in what extend the accuracy of the "noise-free" realization influences the quality of the final estimate. We still use the enhanced stacking method applied locally to obtain another "noise-free" realization, but with the less favorable window parameters: 2 traces of 50 time samples (here 50 ms), with 20 % overlap along time between windows, and we add to the resulting section some zero-mean white

Gaussian noise of variance 2.5×10^{-3}. This new "noise-free" realization is shown in Figure 5.5 (a). Compared to the first "noise-free" realization shown in Figure 5.3 (a), the estimated signals are more attenuated and distorted, especially for the first and second reflections, and the noise level is larger. The MSE of this section is increased by 41 % to reach 99×10^{-4}. We apply sparse code shrinkage with the sparse density to this second "noise-free" realization and show the estimated data set in Figure 5.5 (b). The noise attenuation is as good as in Figure 5.3 (a), but the reconstructed signals for the first and second reflections is slightly deteriorated, exactly as in the "noise-free" realization. Thus, the MSE is hardly larger and equal to 45×10^{-4} against 39×10^{-4} for the first "noise-free" realization. Finally, we obtain similar results when using the other density models. Thus, the accuracy of the "noise-free" realization allows some flexibility. But the most important requirement to yield reconstructed signals of good quality, is that events are contained in the realization, and possibly as little distorted as possible.

We show through this synthetic example that the noise attenuation ability of sparse code shrinkage is promising as long as the shrinkage functions are specifically fitted to the data. In addition, the choice of parametric or non-parametric models is not decisive since they yield very small differences. The parametric models yield slightly better noise attenuation, while the non-parametric model yields slightly more accurate signal estimates. Finally, the quality of events contained in the "noise-free" realization influences the quality of reconstructed signals in the final estimate, but sparse code shrinkage achieves a much more accurate signal reconstruction than the "noise-free" realization. Finally, sparse code shrinkage remains relatively insensitive to a larger amount of noise in the "noise-free" realization, and handles relatively well slightly hyperbolic events.



(a) (b)

Figure 5.2: Signal enhancement of synthetic section: (a) noisy data and (b) true noise free data.

	MSE (x10^{-4})
Synthetic input data	119
"Noise-free" realization	70
Sparse	39
NIG	42
MGG	42
Gaussian kernel estimate	51
Mildly sparse with $a = b = 0.5$	84

Table 5.1: Mean square error of inputs and estimated outputs using different density models. Sparse, NIG and MGG densities achieve the lowest MSE.

5.5.2 Test on real data

We consider here a NMO-corrected CMP gather of marine seismic data from the North sea contaminated by random noise and multiples, shown in Figure 5.6 (a). The area that is the most contaminated by high amplitude multiples has been muted away, but some multiples can still be observed at small offsets. We apply on this section sparse code shrinkage with the presented density models, as well as two other noise attenuation methods: local singular value decomposition (SVD) described by Bekara and Van der Baan (2007) and *f-x* deconvolution described by Canales (1984). The NMO correction is necessary to obtain the "noise-free" realization shown in Figure 5.6 (b), and for application of local SVD. For sparse code shrinkage and local SVD, the sliding-window contains 10 traces and 32 time samples, with a percentage overlap of 50 % overlap both in time and offset. The normalized noise variance used in sparse code shrinkage is estimated to 0.6. Like for synthetic examples, the local mean is subtracted from each data vector and its dimension is reduced by PCA. *f-x* deconvolution is applied to the frequency range $[1, 120]$ Hz and uses an autoregressive filter of order 20 as well as a pre-whitening factor $\mu = 0.01$. For local SVD, dip steering is applied prior to SVD and eventually, the first four eigenimages are kept for the reconstruction of the window. Figure 5.6 (b) and (c) shows the "noise-free" realization and its residual plot respectively. Coherent signals at small and far offsets and some multiples (within $[0, 2.5]$ s) are partly estimated, but lots of signals have been attenuated within the $[0, 2.5]$ s interval and at far offsets. Thus, the signal reconstruction is relatively limited. Figure 5.7 shows the results after signal enhancement using sparse code shrinkage, and Figure 5.9 (a) and (c), the results after signal enhancement using *f-x* deconvolution and local SVD respectively. The residual plots for sparse code shrinkage are shown in Figure 5.8, and for *f-x* deconvolution and local SVD in Figure 5.9 (b) and (d) respectively. For sparse code shrinkage, we obtain results similar to the synthetic data example, with however small differences between the parametric models. The sparse density (Figure 5.7 (a)) and Gaussian kernel density (Figure 5.7 (d)) perform about the same for both the signal estimation and noise attenuation. The MGG density (Figure 5.7 (c)) and even more, the NIG density (Figure 5.7 (b)) are a little more efficient for noise attenuation, but they attenuate some very small parts of multiples (Figure 5.8 (b) and (c), around 1 s).

Obviously, MGG and even more, NIG densities set the sparse components of small amplitude to zero more often than the sparse density does. This means that fitting the seismic sparse components to MGG and even more, to NIG density models, yields distributions that are more super-Gaussian than the ones fitted to the sparse distribution. Thus, as long as the slight attenuation of multiples stays within an acceptable limit, we suggest to choose the density model that yields the best signal enhancement, i.e., the NIG density, even though MGG and sparse densities would be an acceptable choice as well. Again, we notice that sparse code shrinkage applied with shrinkage functions that are not specifically fitted to the data (Figure 5.7 (e) and Figure 5.8 (e)) cannot target the sparse noise components only, and therefore, affects both noise and coherent events in a same manner. Finally, *f-x* deconvolution is very efficient in noise suppression and in estimating the flat events at short offsets (Figure 5.9 (a)), but it also attenuates a non-negligible part of the multiples (Figure 5.9 (b), within the $[0, 2]$ s interval). Local SVD performs about the same as sparse code shrinkage with sparse density in term of noise suppression, but it attenuates a few events in the shallow waters, short offset area (Figure 5.9 (d)). The noise suppression along the muted area is rather poor, but this can be partly avoided by applying SVD before muting the data.

In Figure 5.10 (a), we consider finally a stacked section. Here, sparse code shrinkage is tested with only two density models: the sparse density that is the most commonly used one and the NIG density that has given among the best results so far. Sparse code shrinkage is compared with local SVD and *f-x* deconvolution. Sparse code shrinkage extracts data windows of 6 traces and 90 time samples with an overlap of 60 %, and the normalized noise variance is 0.1. SVD is applied locally using a sliding window that consists of 10 traces and 32 time samples with an overlap of 50 %. Within each window, dip steering is applied prior to SVD and the first four eigenimages are kept for the reconstruction of the window. *f-x* deconvolution uses an autoregressive filter of order 20, a pre-whitening factor $\mu = 0.01$, and it is applied to the frequency range $[1, 120]$ Hz. Figure 5.10 (b) to (e) shows the results after signal enhancement using sparse code shrinkage with sparse and NIG densities (b and c), *f-x* deconvolution (d) and local SVD (e). The residual plots for these different techniques are shown in Figure 5.11.

We obtain results that are similar to the previous examples. The NIG density yields slightly harsher shrinkage than the sparse density. It attenuates the strongest reflections a little more (Figure 5.11 (b) around 1 s), but also attenuates more noise. Thus, thin structures are enhanced around 1.5 s for small CMPs, between 2.5 and 3 s, and around 5 s, and the reflections in shallow waters are estimated with a better resolution. Both NIG and sparse densities however tend to erroneously estimate some noiselike signal in the muted area from 0.2 s to 0.5 s. The comparison with *f-x* deconvolution and local SVD shows that all these methods have good background noise attenuation ability. *f-x* deconvolution exhibits more continuous reflections, e.g. between 1 and 2 s, because of better interpolation capabilities, and it enhances signals in shallow waters (time < 2 s) with the same resolution as sparse code shrinkage with NIG density. However, it affects more the strongest reflections that are partly attenuated (Figure 5.11 (c)), it fails to reveal thin structures between 2 to 5 s, and it erroneously estimates some noiselike signal in a muted area (time < 0.5 s) in a more important manner than sparse code shrinkage (Figure 5.11 (c)). Local SVD also slightly attenuates the strongest reflections as well, but in a minor importance. Increasing the number of eigenimages to reconstruct this section would deteriorate its noise attenuation ability. Some strong structures are blurred in shallow waters, as well as thinner ones between 2.5 and 3 s, and around 5 s. However, the thin structures between 3.5 and 4.5 s are enhanced. Therefore, in this example, sparse code shrinkage using NIG density offers the best compromise between noise attenuation and signal enhancement.

5.6 Discussion

5.6.1 Parameter settings

The transformation of seismic data into sparse code requires them to be reorganized into a set of realizations, where each realization consists of a reorganized data set extracted from a sliding window. Thus, we discuss in this paragraph how the choice of the sliding window parameters affects the performance of sparse code shrinkage. The best resolution is achieved with a small number of traces, but this is usually obtained at the cost of a less effective noise reduction and longer processing time because of the larger number of realizations N to be treated. The window length, i.e., number of samples, is less crucial. We obtain a better resolution and events continuity with a large number, i.e., when the length of each realization is long enough. However, a too large number is not necessary since it is reduced by PCA. Finally, the overlap percentage influences the number of realizations N, as well as the averaging of the overlapping areas. Again, we obtain a better resolution with a larger number of realizations N, but at the cost of longer processing time. Thus, a median overlap percentage, around 50 - 60 % is a good compromise.

The noise variance parameter should be estimated from data, but the user is however free to adjust this parameter. Increasing the noise variance influences both the range of values set to zero and the slope of the shrinkage function. Thus, the larger the noise variance parameter is, the larger is the range of values that are considered to be noise and therefore set to zero, and the more severe the components attenuation is.

5.6.2 Possible improvements of sparse code shrinkage

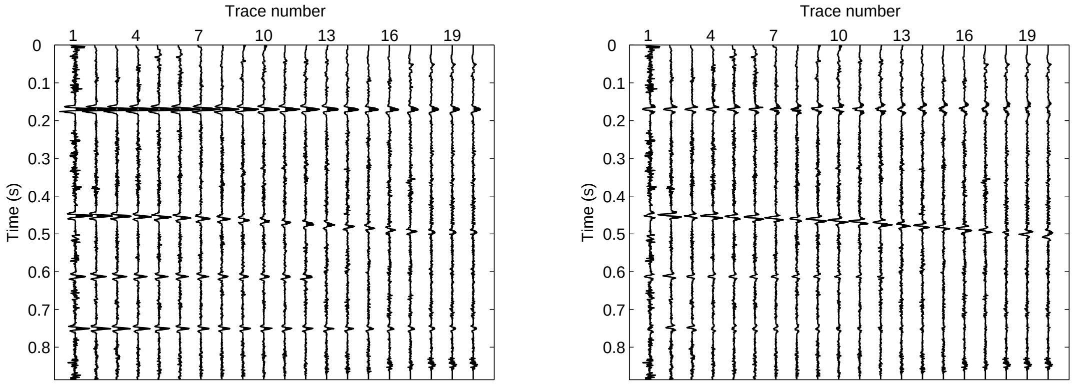
We suggested one method to obtain a "noise-free" realization of the seismic section, but others may be suitable as well. Thus, the signal reconstruction achieved by sparse code shrinkage may be improved by improving the quality of signals contained in that realization. In addition, some other super-Gaussian density models can be investigated. In particular, the generalized Laplacian density and the mixture of two normal inverse Gaussian might be suitable to describe the seismic components.

5.7 Summary

This chapter contains a review of sparse coding as a method to remove random noise in seismic data. Sparse coding has been derived for three possible parametric models of density: sparse, normal inverse Gaussian and mixture of generalized Gaussian densities, and for a novel non-parametric Gaussian kernel density estimate. Sparse coding requires a "noise-free" realization to provide data-driven shrinkage functions. We showed indeed that non-fitted shrinkages could not simultaneously reconstruct signals and attenuate the noise efficiently. By applying an enhanced stacking method locally, we proposed a realization that is good enough for the derivation of fitted shrinkages.

Sparse code shrinkage has proven to be an efficient method that targets the noise components to shrink them while preserving the signals. By its capacity to enhance signals in both shallow and deep waters, the NIG density appears to be the best choice of density. However, sparse, MGG and non-parametric estimate densities have proven to be acceptable choices as well. The flexible parametrization of the NIG density allows it to model a larger number of unimodal densities, and therefore, suits seismic data very well. In addition, the level of attenuation of the sparse coded data can be easily adjusted by the user through the noise variance parameter.

In the presented examples, sparse code shrinkage, local SVD and *f-x* deconvolution have shown to be effective methods for removing background noise. However, *f-x* deconvolution appeared to be the one that affects certain high amplitude signals the most (although within an acceptable limit), but achieves however a good resolution in shallow waters. Local SVD slightly blurred the signals in shallow waters, but was better than *f-x* deconvolution at enhancing signals in deeper waters. Finally, sparse code shrinkage estimated signals in both shallow and deep waters pretty accurately. Like local SVD, sparse code shrinkage requires NMO-corrected data but it can however handle some slightly hyperbolic events. In addition, *f-x* deconvolution and of minor importance, sparse code shrinkage, map energy to muted areas where none existed before. Finally, sparse code shrinkage requires more computational load than local SVD and *f-x* deconvolution, mostly because of the determination of the ICA transformation matrix. Therefore, based on the analysis and examples presented in this chapter, we recommend sparse code shrinkage as a valuable alternative method for the denoising of seismic data.

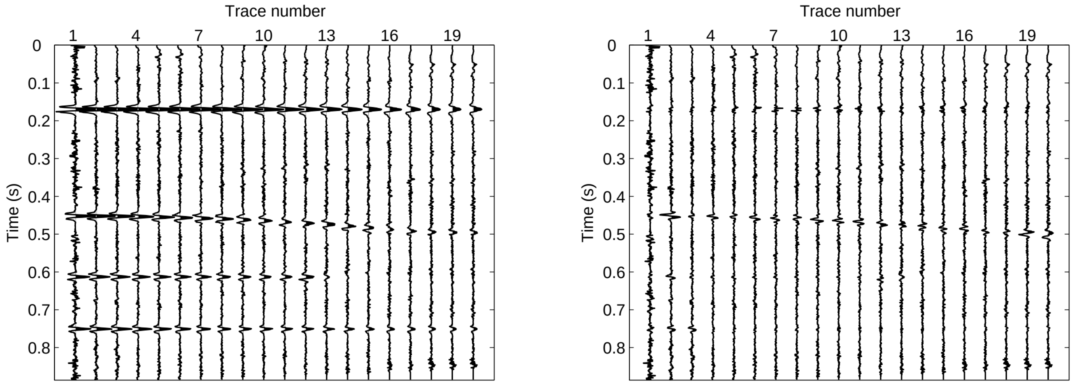


Noise−free realization: result

(a)

Noise−free realization: estimation error

(b)

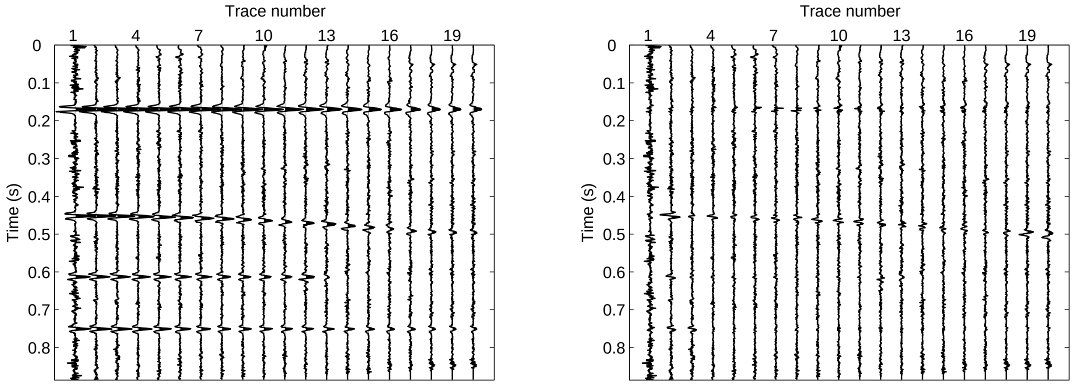


Sparse density: result

(c)

Sparse density: estimation error

(d)



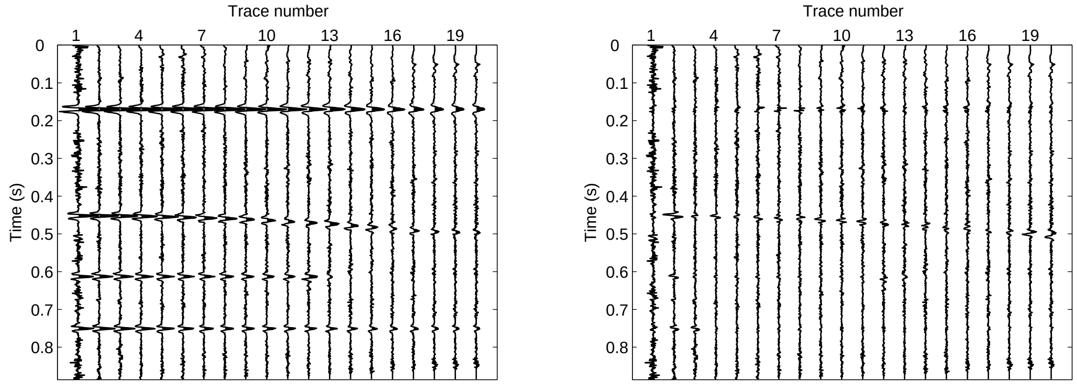
NIG density: result

(e)

NIG density: estimation error

(f)

Figure 5.3: Estimation of "noise-free" realization for sparse code shrinkage: (a) results and (b) estimation error. Sparse code shrinkage applied to a synthetic section and using different density models: sparse density: (c) section estimate and (d) estimation error, and NIG density: (e) section estimate and (f) estimation error.

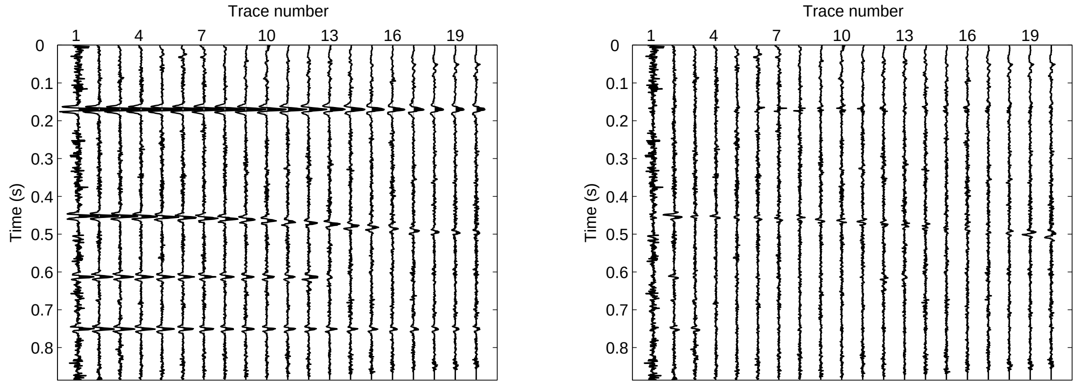


MGG density: result

(a)

MGG density: estimation error

(b)

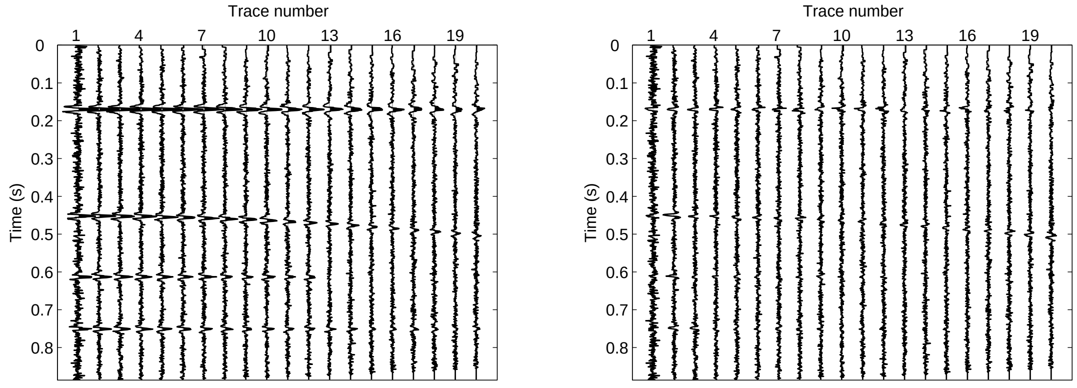


Gaussian kernel density: result

(c)

Gaussian kernel density: estimation error

(d)



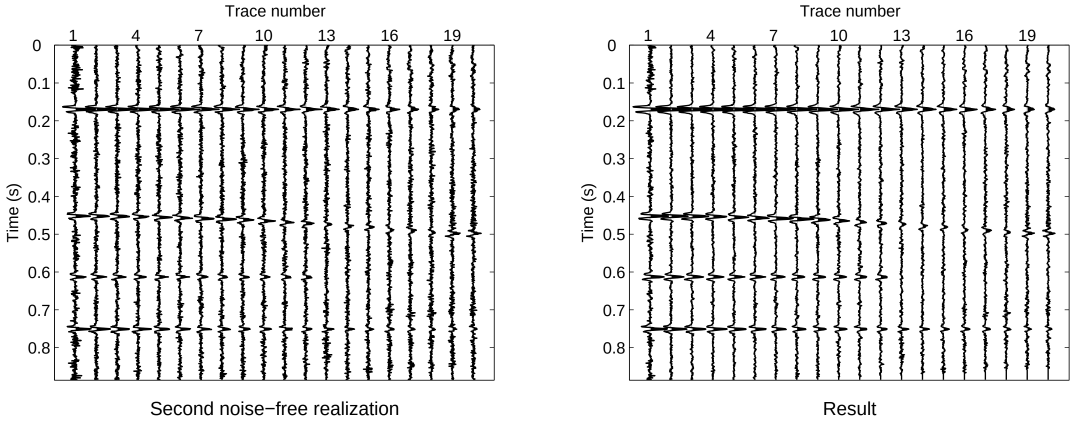
Mildly sparse density with a=b=0.5: result

(e)

Mildly sparse density with a=b=0.5: estimation error

(f)

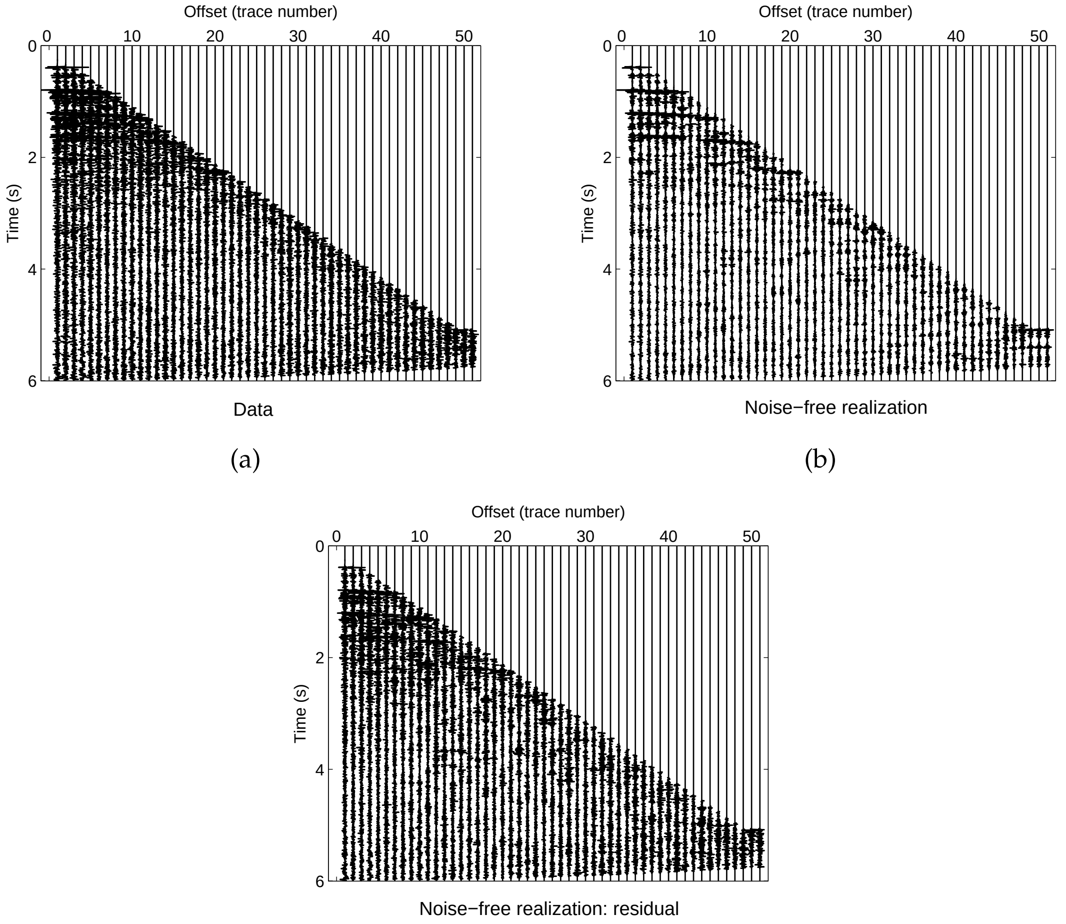
Figure 5.4: Sparse code shrinkage applied to a synthetic section and using different density models. MGG density: (a) section estimate and (b) estimation error, non-parametric Gaussian kernel estimated density: (c) section estimate and (d) estimation error, and mildly sparse density with fixed parameters $a = b = 0.5$: (e) section estimate and (f) estimation error.



(a)

(b)

Figure 5.5: Influence of the quality of the "noise-free" realization: (a) second "noise-free" realization and (b) section estimate after sparse code shrinkage with sparse density.

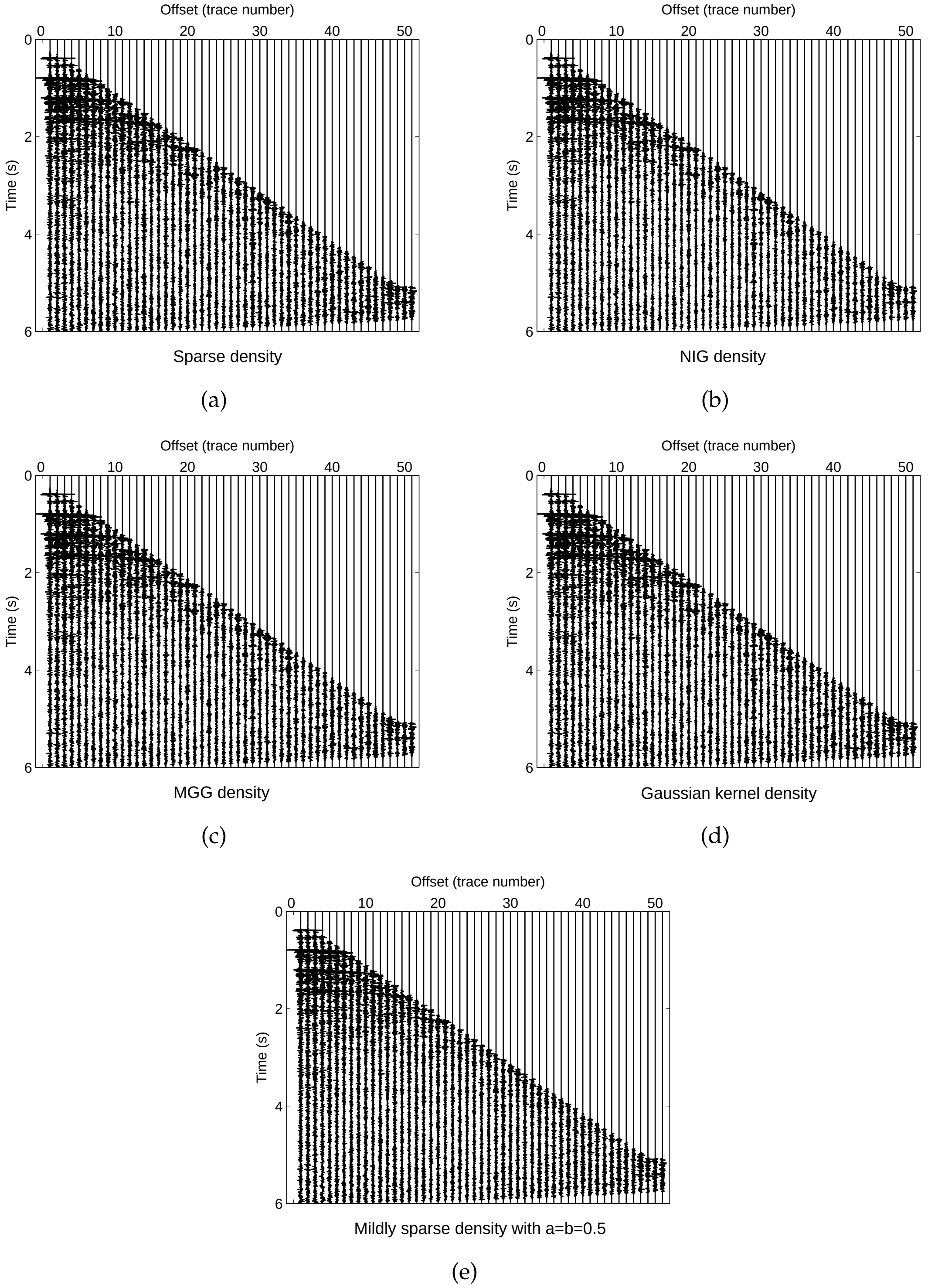


(a)

(b)

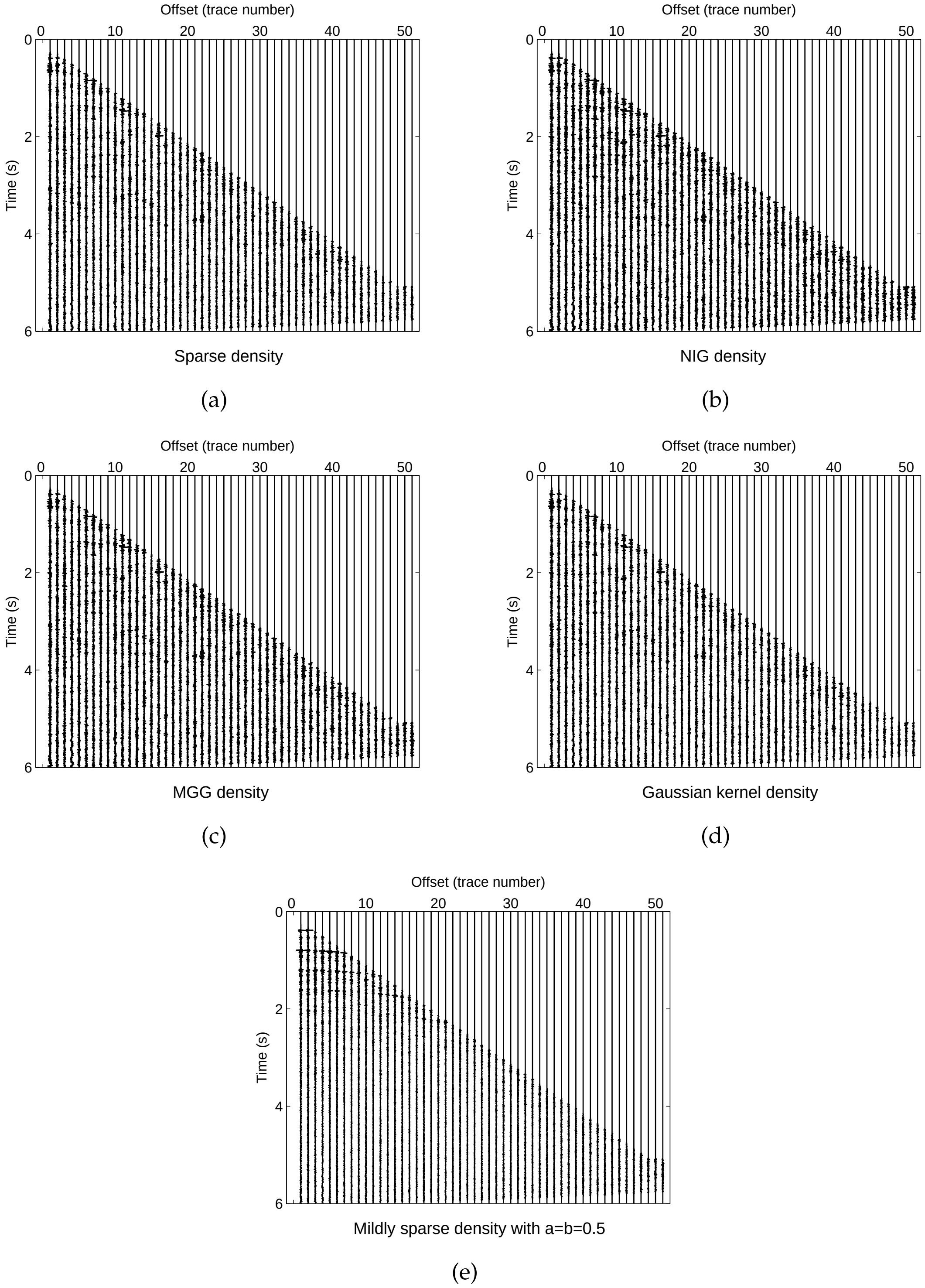
(c)

Figure 5.6: Signal enhancement of unstacked data. (a) NMO-corrected CMP gather, and the "noise-free" realization used in sparse code shrinkage: (b) estimated section and (c) residual plot. In the "noise-free" realization, coherent signals and multiples are only partly estimated.



(a) (b) (c) (d) (e)

Figure 5.7: Signal enhancement of unstacked data using sparse code shrinkage with different density models: (a) sparse density, (b) NIG density, (c) MGG density, (d) Gaussian kernel estimator, and (e) mildly sparse density with fixed parameters $a = b = 0.5$. Use of fitted shrinkages is necessary for sparse code shrinkage to successfully enhance signals.



(a) (b) (c) (d) (e)

Figure 5.8: Residual plots for the sparse code shrinkage using the different density models: (a) sparse density, (b) NIG density, (c) MGG density, (d) Gaussian kernel estimator, and (e) mildly sparse density with fixed parameters $a = b = 0.5$.

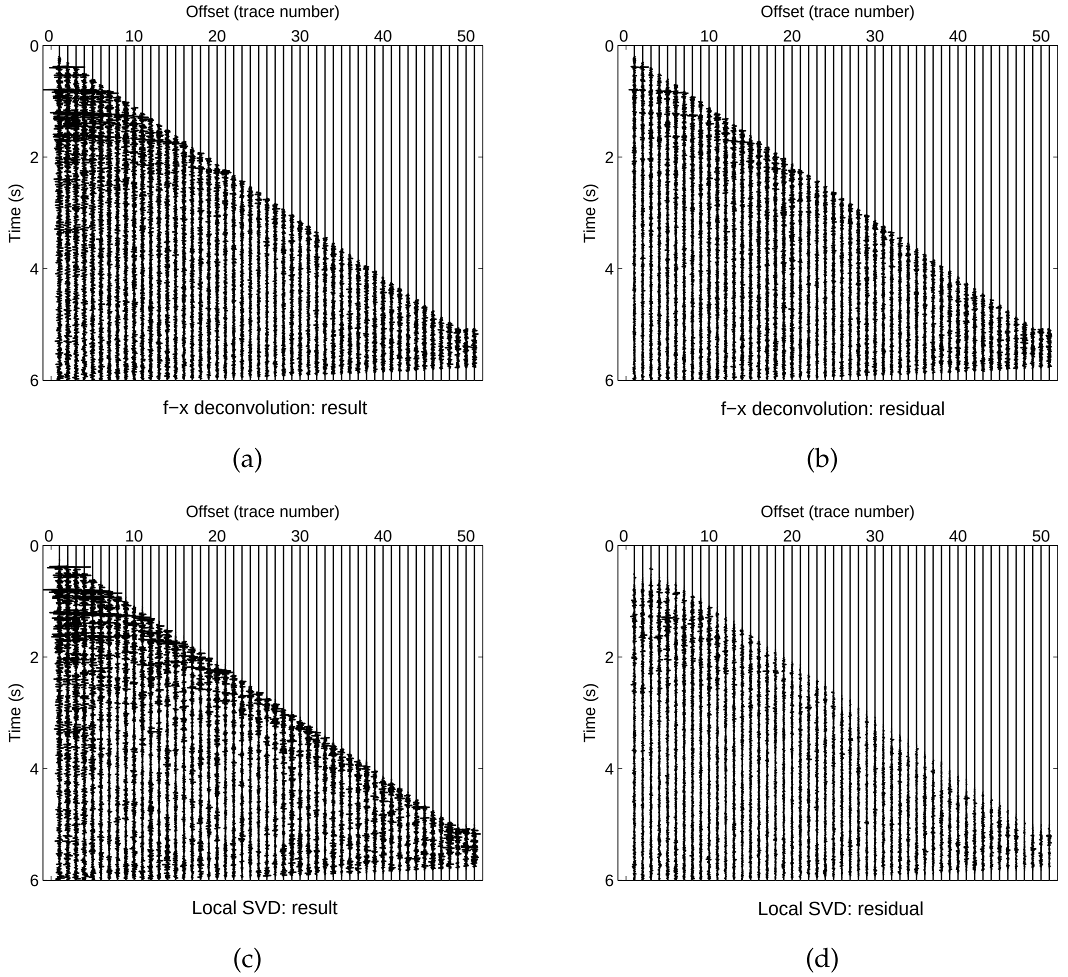
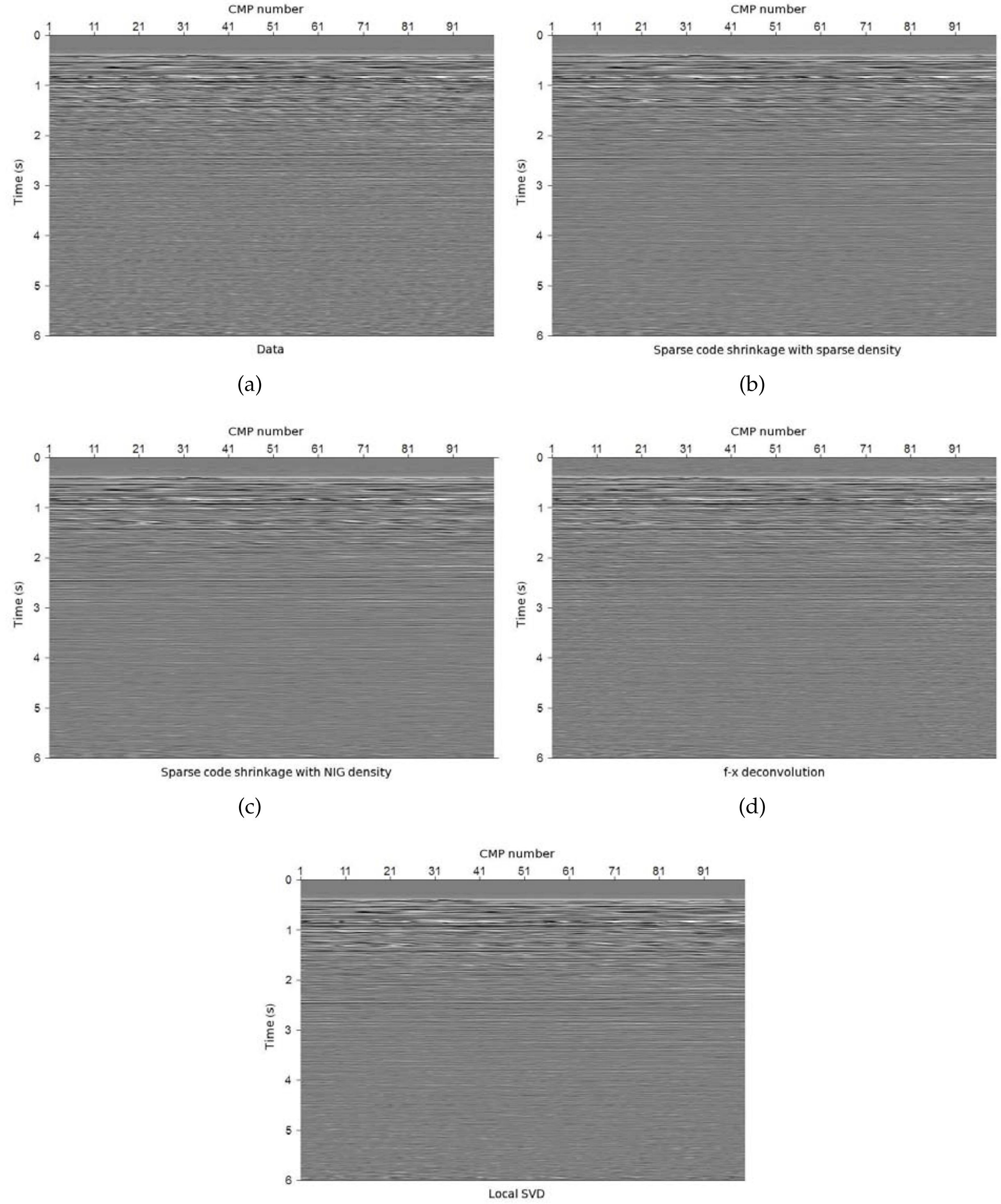
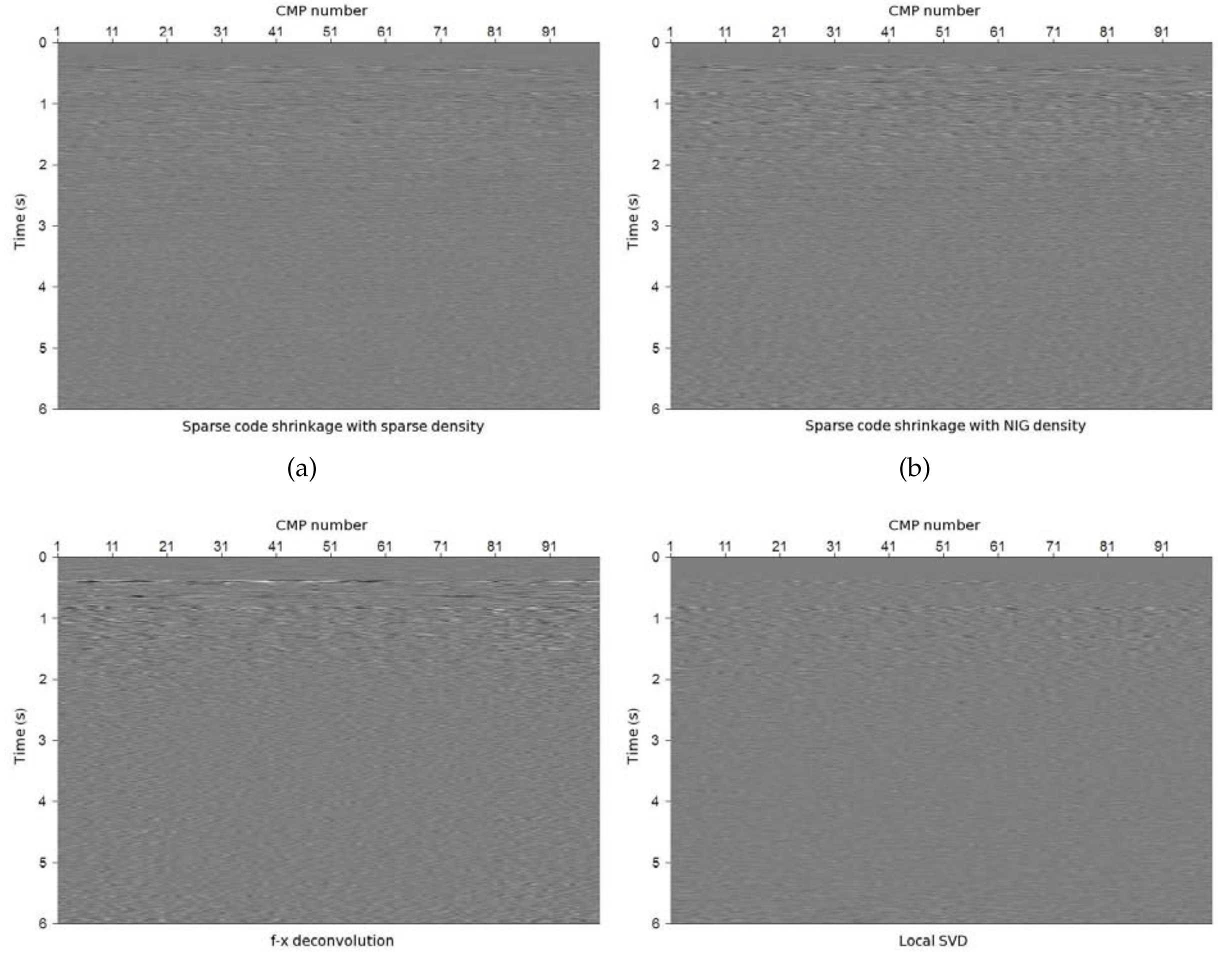


Figure 5.9: Other signal enhancement techniques, results and residual plots. *f-x* deconvolution : (a) estimated section and (b) residual plot. Local SVD: (c) estimated section and (d) residual plot.



(a) (b) (c) (d) (e)

Figure 5.10: Signal enhancement of stacked data. (a) A stacked section. Results using different techniques: (b) sparse code shrinkage using sparse density, (c) sparse code shrinkage using NIG density, (d) *f*-*x* deconvolution, and (e) local SVD.



(a) (b) (c) (d)

Figure 5.11: Residual plots for the different techniques. (a) Sparse code shrinkage with sparse density, (b) sparse code shrinkage with NIG density, (c) *f-x* deconvolution, and (d) local SVD.

Chapter 6

Multiple-input adaptive noise canceller to attenuate non-stationary coherent noise

6.1 Introduction

Attenuation of coherent noise is a major concern in seismic. Typically, weather generated noise and more specifically swell noise, usually characterized by low frequency and high amplitude, is considered as coherent in seismic (D'Agosto et al., 2003). Noise generated by surface waves is generally attenuated by high-pass and band-pass filters, while swell noise can be attenuated by time-frequency median filtering (Bekara et al., 2008; Elboth et al., 2010). Time-frequency median filtering compares the signal amplitudes in the frequency domain and assumes any amplitude beyond a given threshold to be noisy. This method is however very dependent on the noise repartition in the input gather and on the chosen threshold, especially if it is chosen arbitrarily. Finally, another conventional way of attenuating the coherent high-amplitude noises is to apply a F-K filter to the data within a given range. The drawback with this algorithm is that this filter is deterministic, which introduces some limitations when signal and noise have overlaps in the F-K domain. The multichannel approach has been described by Özbek (2000). He proposed a multichannel time-adaptive interference canceller where the normal equation is resolved by means of a principal components decomposition. Wang et al. (2008) implemented a multichannel Wiener filter in the frequency domain. By using a non-adaptive resolution method, this algorithm cannot easily adapt to the changes in statistics and the noise attenuation may be limited.

In this chapter, we consider a multiple-input adaptive noise canceller as a solution to remove non-stationary coherent noise. This filter uses multiple noise sequences to provide an estimate of the noise contained in the primary channel, and then subtracts it to obtain an estimate of the primary channel whose coherent noise component has been attenuated. There is a variety of different algorithms for the design and implementation of adaptive filters. In this chapter, we focus on the normalized least mean squares (NLMS) algorithm used with a variable normalized step size. The NLMS has proven indeed to be a simple, robust and effective algorithm that has found a variety of applications in signal processing (Hayes, 1996; Haykin, 2001; Widrow and Stearns, 1985). We will show that the variable normalized step size is necessary for the filter to respond quickly to changes in signal statistics.

First, in Section 6.2 we present the structure of the multiple-input adaptive noise canceller, we develop the theory about the normalized least mean squares algorithm and we derive a method to obtain a variable step size parameter. In Sections 6.3.2 and 6.3.3, we consider an application to swell noise attenuation on two real marine seismic data sets, and compare against two other noise attenuation methods: time-frequency median filtering and a second order high-pass Butterworth filter. We show that the multiple-input adaptive noise canceller achieves a very good attenuation of coherent noise while preserving the seismic reflections in the case of localized, high amplitudes swell noise that differs significantly from seismic reflections. This method yields better results than time-frequency median filtering and second order high-pass Butterworth filter when dealing with such noise. However, the multiple-input adaptive noise canceller can only partly attenuate the coherent noise when this one is somewhat correlated with seismic reflections. Such noise is however much more difficult to handle, and the other two comparative methods also show some limitations in their attempt of attenuating it. However, we demonstrate that combining time-frequency median filtering with the multiple-input adaptive noise canceller yields a very good attenuation of coherent noise. Finally, in Section 6.3.4, we consider an application to seismic interference attenuation, for which the efficiency is admittedly limited.

6.2 Theory

6.2.1 Multiple-input adaptive noise canceller

Let us consider a seismic trace or primary channel whose value at time sample *n* is denoted by $x_k(n)$. The trace signal consists of the sum of a seismic signal $s_k(n)$ corrupted by noise $\nu_k(n)$, such that

$$x_k(n) = s_k(n) + \nu_k(n). \tag{6.1}$$

The multiple-input adaptive noise canceller uses a set of M noise sequences $\boldsymbol{\nu}_1(n), \ldots,$ $\boldsymbol{\nu}_M(n)$, where $\boldsymbol{\nu}_i(n) = [\nu_i(n)\, \nu_i(n-1)\, \ldots\, \nu_i(n-L+1)]^T$, to predict the noise contained in the primary channel at time sample *n*, and then subtract it from the primary. The block diagram is shown in Figure 6.1. Thus, if the input noise sequences are correlated with the corrupting noise $\nu_k(n)$ but uncorrelated with the seismic signal $s_k(n)$, then the multiple-input adaptive noise canceller provides an estimate of the noise $\hat{\nu}_k(n) = \sum_{i=1}^{M} \hat{\nu}_i(n)$, that is subtracted from the primary channel $x_k(n)$ to form an estimate of the seismic signal $s_k(n)$, given by the error value $e(n)$:

$$e(n) = s_k(n) + \nu_k(n) - \sum_{i=1}^{M} \hat{\nu}_i(n) = \hat{s}_k(n). \tag{6.2}$$

The estimate of the noise component, $\hat{\nu}_k(n)$, is obtained by finding a set of M optimal time-dependent tap-weight vectors $\mathbf{w}_1(n), \ldots, \mathbf{w}_M(n)$. As optimality criterion, we mean here the minimization of the mean-square error $E\{|e(n)|^2\}$ at any time. Under the previous assumptions, minimizing the mean-square error $E\{|e(n)|^2\}$ is equivalent to minimizing $E\{|\nu_k(n) - \sum_{i=1}^{M} \hat{\nu}_i(n)|^2\}$, and therefore the error signal $e(n)$ forms the mean-square estimate of $s_k(n)$. In order to determine the set of tap-weight vectors that minimizes the mean-square error, we use the normalized least mean squares algorithm. The choice of an adaptive filter is justified by the nonstationary character of the seismic data where a non-adaptive filter fails to adapt to the changes of the signal statistics. Finally, as input noise sequences we use either pure noise recordings recorded shortly before data has been acquired if they are available, or we extract noise sequences from data records prior the first seismic reflection arrival.

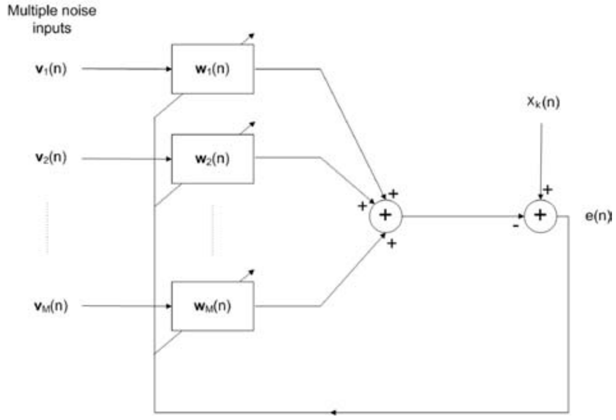


Figure 6.1: Block diagram of the multiple inputs adaptive noise canceller.

6.2.2 The normalized least mean squares algorithm

The least mean squares algorithm is a simple, robust and efficient algorithm that has found a variety of applications in signal processing. In the least mean squares algorithm, the minimization of the mean-square error is actually replaced by the minimization of the instantaneous squared error $|e(n)|^2$, which does not require any ensemble average to be known.

Let us consider an adaptive finite impulse response (FIR) filter of order L that produces an output

$$\hat{\nu}_i(n) = \mathbf{w}_i^H(n)\boldsymbol{\nu}_i(n), \tag{6.3}$$

where $\mathbf{w}_i(n) = [w_{i,0}(n)\, w_{i,1}(n)\, \dots\, w_{i,L-1}(n)]^T$ is the i^{th} tap-weight vector of the adaptive filter and $\boldsymbol{\nu}_i(n) = [\nu_i(n)\, \nu_i(n-1)\, \dots\, \nu_i(n-L+1)]^T$ is the i^{th} tap-input vector made of the most recent L samples of input noise sequence $\nu_i(n)$. If we consider M input noise sequences and if we denote the k^{th} primary channel by $x_k(n)$, then the error between the primary channel and the multiple noise inputs is given by

$$e(n) = x_k(n) - \sum_{i=1}^{M} \boldsymbol{\nu}_i^T(n)\mathbf{w}_i^*(n). \tag{6.4}$$

The LMS algorithm updates the i^{th} tap-weight vector by the gradient of the instantaneous squared error (Widrow and Stearns, 1985), such as

$$\mathbf{w}_i(n+1) = \mathbf{w}_i(n) - \mu\nabla|e(n)|^2 = \mathbf{w}_i(n) - \mu\, e(n)\nabla e^*(n). \tag{6.5}$$

Using the error signal given in equation 6.4, the gradient of the conjugate error is given by

$$\nabla e^*(n) = -[\nu_i^*(n)\, \nu_i^*(n-1)\, \dots\, \nu_i^*(n-L+1)]^T, \tag{6.6}$$

and therefore, the update equation of the i^{th} tap-weight vector can be written as

$$\mathbf{w}_i(n+1) = \mathbf{w}_i(n) + \mu\, e(n)\boldsymbol{\nu}_i^*(n). \tag{6.7}$$

Because the selection of the step size μ can be tricky, we prefer the use of the normalized LMS instead by setting

$$\mu = \frac{\beta}{\epsilon + ||\boldsymbol{\nu}_i^*(n)||^2} \tag{6.8}$$

where $0 < \beta < 2$ is the normalized step-size and ϵ is a small positive number called regularization parameter. The normalization by $||\boldsymbol{\nu}_i^*(n)||$ makes us avoid the problem of noise amplification, while ϵ represents the regularization parameter added to overcome situations where $||\boldsymbol{\nu}_i^*(n)||$ is too small.

Therefore, we can summarize the normalized LMS algorithm for a FIR adaptive filter of order L using M input noise sequences and applied to the k^{th} seismic trace, as follows:

Initialization: $\mathbf{w}_i(0) = \mathbf{0}$ for $i = 1, \dots, M$. (6.9)

Computation: For $n = 0, 1, 2, \dots$ (6.10)

$$e(n+1) = x_k(n+1) - \sum_{i=1}^{M} \boldsymbol{\nu}_i^T(n+1)\mathbf{w}_i^*(n) \tag{6.11}$$

$$\mathbf{w}_i(n+1) = \mathbf{w}_i(n) + \frac{\beta}{\epsilon + ||\boldsymbol{\nu}_i^*(n+1)||^2}\, e(n+1)\boldsymbol{\nu}_i^*(n+1) \tag{6.12}$$

$$\hat{s}_k(n+1) = x_k(n+1) - \sum_{i=1}^{M} \boldsymbol{\nu}_i^T(n+1)\mathbf{w}_i^*(n+1). \tag{6.13}$$

Equation 6.12 is the tap-weight update equation while equation 6.13 gives the a posteriori error signal that represents the estimate of the seismic signal.

6.2.3 Variable normalized step size parameter

The normalized step-size parameter β of the NLMS algorithm that governs the convergence speed and the steady-state excess of mean-square error must be properly selected. It is shown that the rate of convergence is proportional to the step-size, while the steady-state excess of mean-square error is inversely proportional to the step-size. Thus, a fixed β value means a tradeoff between rate of convergence, steady-state excess of mean-square error and the ability of the filter to track signals as their statistics change (Haykin, 2001; Zhao et al., 2009). Preliminary experiments on seismic data showed us that if β is large, the correction of $\mathbf{w}_i$ is large and as a consequence, the low-frequency noise is significantly attenuated as well as certain seismic reflections. On the opposite, if β is small, the correction of $\mathbf{w}_i$ is small and then, the estimated signal follows rather accurately the seismic reflections but fails to attenuate the low-frequency noise significantly.

Since we seek a filter that conserves the seismic reflections while attenuating coherent low-frequency noise, the use of a variable normalized step size is necessary. We first tested a well-known variable step size algorithm described by Kwong and Johnston (1992) where the step size adjustment is controlled by the square of the prediction error. It appeared to be unsuccessful when applied to seismic data because using the error as a metric yields the same shortcomings as for constant step size. There is, however, one metric that differentiates seismic reflections from the coherent low-frequency noise, which is the instantaneous frequency. Seismic reflections have indeed instantaneous frequencies that are larger than coherent low-frequency noise. Thus, we look for an instantaneous frequency threshold value that differentiates seismic reflections from coherent low-frequency noise. For instantaneous frequencies smaller than the threshold value, low-frequency noise is detected and a larger normalized step size should be used. For instantaneous frequencies larger than the threshold value, seismic reflections are detected and a smaller normalized step size should be used.

For this application, we calculate the instantaneous frequency $\phi(t)$ by means of the Hilbert transform. The Hilbert transform $\tilde{x}(t)$ of a real signal $x(t)$ is the conjugate of this signal, i.e., a version of this signal with a 90° phase shift. They create a so-called analytic signal $z(t)$ that can be written with an amplitude and a phase, and whose phase derivative can be identified as the instantaneous frequency. Thus, we define the analytic signal $z(t)$ of $x(t)$ as

$$z(t) = x(t) + j\tilde{x}(t) \tag{6.14}$$

where j is the square root of -1. The imaginary part $\tilde{x}(t)$ contains the Hilbert transform, defined as

$$\tilde{x}(t) = \mathcal{H}[x(t)] = \frac{1}{\pi} p.v. \int_{-\infty}^{\infty} \frac{x(\tau)}{t-\tau} \mathrm{d}\tau \tag{6.15}$$

where p.v. denotes the Cauchy principal value, provided this integral exists as a principal value. Note that the Hilbert transform also satisfies $\mathcal{H}[\mathcal{H}[x(t)]] = -x(t)$. In addition, $x(t)$ and $\tilde{x}(t)$ have the same amplitude and frequency content, and $\tilde{x}(t)$ includes phase information that depends on the phase of $x(t)$ (Boashash, 1992; Claerbout, 1976; Hahn, 1996). The Hilbert transform is useful in calculating instantaneous attributes of a time series and in particular, the instantaneous frequency (Boashash, 1992; Cohen, 1995; Luo et al., 2003). Thus, given the real signal $x(t) = A(t)\cos\theta(t)$, we may express the analytic signal in the compact form $z(t) = A(t)e^{j\theta(t)}$, where

$$A(t) = \sqrt{x^2(t) + \tilde{x}^2(t)} \tag{6.16}$$

is referred to as the instantaneous amplitude and

$$\theta(t) = \arctan\frac{\tilde{x}(t)}{x(t)} \tag{6.17}$$

is referred to as the instantaneous phase of the analytic signal. Therefore, the instantaneous frequency, that is the time rate of change of the instantaneous phase angle, is given by

$$\phi(t) = \frac{\theta(t+T_s) - \theta(t)}{2\pi T_s} \tag{6.18}$$

where T_s denotes the sampling period. Using a discrete-time notation $t = nT_s$, $n \in \mathbb{Z}$, the instantaneous frequency is given by

$$\phi(n) = \frac{\theta(n+1) - \theta(n)}{2\pi T_s}. \tag{6.19}$$

In our method, we first set up to two threshold values for instantaneous frequency, denoted by Φ_1 and Φ_2. In order to choose these two values, we calculate the instantaneous frequency for all the noise input sequences to obtain the range of possible values, and then choose the thresholds so that they include a high percentage of the frequency values. Thereafter, we need to calculate the instantaneous frequencies $\phi_k(n)$ for each seismic trace $x_k(n)$. Because they usually are very spiky, it is necessary to smooth the frequency time series before applying thresholds. To do so, we use the robust locally weighted regression proposed by Cleveland (1979). This method is an extension of smoothing techniques based on local fitting of polynomials where a robust fitting procedure is used that guards against deviant points distorting the smoothed points. Note that this smoothing function is built in Matlab® (2010). Thus, we are now able to define the normalized step-size parameter value β as follows

$$\beta(n) = \begin{cases} \beta_0 & \text{if } \phi_k(n) \geq \Phi_1 \\ \beta_1 & \text{if } \Phi_2 \leq \phi_k(n) < \Phi_1 \\ \beta_2 & \text{if } \phi_k(n) < \Phi_2. \end{cases} \tag{6.20}$$

Note that the normalized step-size parameter has become time dependent and it is therefore denoted by $\beta(n)$ here. The equation 6.12 must be subsequently updated and becomes

$$\mathbf{w}_i(n+1) = \mathbf{w}_i(n) + \frac{\beta(n)}{\epsilon + ||\boldsymbol{\nu}_i^*(n+1)||^2}\, e(n+1)\boldsymbol{\nu}_i^*(n+1). \tag{6.21}$$

6.2.4 Other comparative methods

In order to evaluate the performances of the variable step size multiple-input adaptive noise canceller, we will test our data set against two other low-frequency noise attenuation methods, that are a local time-frequency median filter and a second order high-pass Butterworth filter. In this chapter, we use a time-frequency median filter as described by Elboth et al. (2010). This filter uses a sliding window that scans the input gather both in space and time. Within this window, a spectral estimate of all traces is achieved and the amplitude estimates at each frequency are compared with the amplitude of a presumed good trace within the chosen window. Here, time-frequency median filtering is applied within the frequency range 0 - 15 Hz where every amplitude estimate greater than 3.2 times a reference amplitude corresponding here to the lower quartile of the amplitude spectrum is attenuated.

The Butterworth filter is a well known type of filter designed to have a frequency response which is as flat as possible in the passband and rolls off towards zero in the stopband (Bianchi and Sorrentino, 2007). We use in our applications a second order high-pass Butterworth filter with cutoff frequency at 2 Hz. It is described by its system function $H(z) = \frac{0.97-1.93z^{-1}+0.97z^{-2}}{1-1.93z^{-1}+0.93z^{-2}}$ and its magnitude response is shown in Figure 6.2.

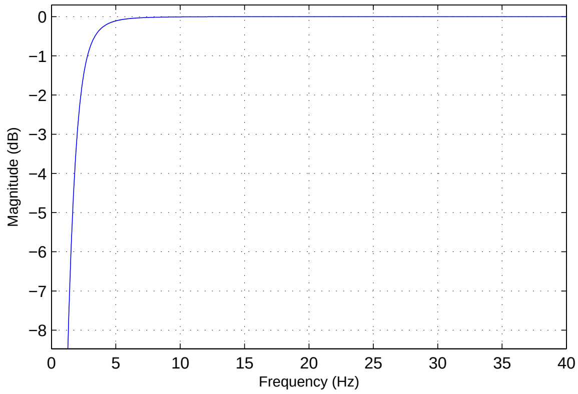


Figure 6.2: Magnitude response of a second order high-pass Butterworth filter with cutoff frequency at 2 Hz (Zoom in the first 40 Hz).

6.3 Applications

6.3.1 Data conditioning

In order to avoid convergence problems with the NLMS algorithm, we wish to have input noise sequences that have the same number of samples as the primary channel. If it is not the case, we replicate the noise sequences as many times as necessary and smooth up to five data samples just before and after the replication time by mean of cubic spline interpolation to avoid sharp transitions. Finally, we shift each of these new input noise sequences so that the replication time differs from sequence to sequence.

In addition, in spite of the use of a time-adaptive algorithm, it appears that it cannot always follow the rapidly changing statistics of input traces. In that case, a solution to boost our method consists of dividing each input trace into small overlapping segments of N_B samples each and reinitializing the algorithm at the beginning of each segment. The overlapping samples are then discarded to avoid the non-steady state of the first few iterations.

6.3.2 Application to swell noise attenuation: first example

We test the multiple-input adaptive noise canceller on two sets of real marine seismic data. The first one, shown in Figure 6.3 (a), is a marine shot gather that consists of 200 traces and that is greatly corrupted by swell noise in a very localized manner. We extract a data window prior the first seismic reflection arrival in the offset interval 2 – 2.36 km and in the time interval 0 – 1.12 s, shown in Figure 6.3 (b). It is then replicated to form 30 input noise sequences of length equal to the traces length.

We first consider a variable step size NLMS algorithm of order $L = 50$. The block length N_B is set to 0.6 s with 10 % overlap between blocks and ϵ is set to 0.0001. In this example, we use two threshold values Φ_1 and Φ_2 by counting up respectively 99 % and 97 % of all instantaneous frequency values. The normalized step size values are set to $\beta_0 = 0.00005$, $\beta_1 = 0.005$ and $\beta_2 = 0.015$. In order to optimize the noise attenuation, we apply the variable step size NLMS algorithm three times successively with the same parameters. The output gather after the first application is shown in Figure 6.4 (a) and after the third application in Figure 6.4 (b). The residual plot, i.e., the difference between the input and output gathers after three applications of the filter is shown in Figure 6.4 (c). We observe that most of the swell noise has been removed while the seismic reflections are unaffected. However, three successive applications of the filter to this data set are necessary to attenuate the largest amplitudes of swell noise.

We now compare the performances of the variable step size multiple-input adaptive noise canceller against the local time-frequency median filter. The time-frequency median filter is applied to the entire data set a first time, and then a second time to data recorded after 1.8 s to avoid the attenuation of seismic reflections. The output gather is shown in Figure 6.5 (a) while the residual plot is shown in Figure 6.5 (b). A few seismic reflections have been attenuated before 2.5 s, but a significant part of swell noise, including the largest swell amplitudes, have been removed. However, some swell noise is still visible in the output gather after 5 s, and in particular for offset distances between 0.3 and 1.0 km. This is caused by the trace-to-trace comparison used by this technique. Applying the time-frequency median filter one more time may remove some more data, which is not desired. In addition, the amount of swell noise left is much more important than for the multiple-input adaptive noise canceller. Let us now apply the second-order high-pass Butterworth filter to remove the low-frequency components of this data set.

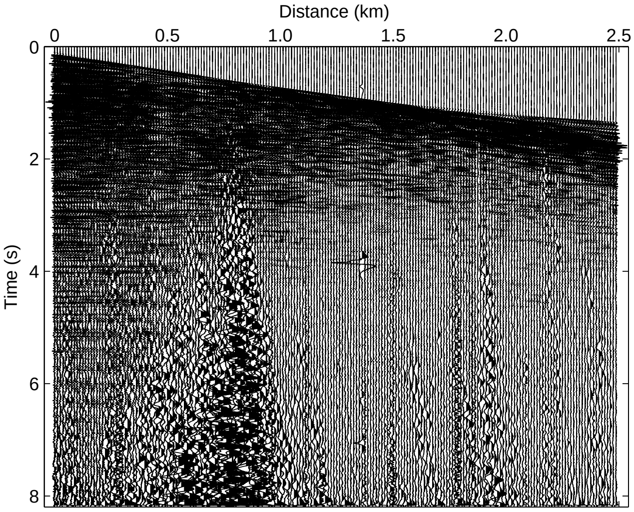
Again, the choice of a larger cutoff frequency or a larger filter order, as well as successive applications of this high-pass filter, would remove too much seismic signal. Figure 6.6 (a and b) shows the output gather and the residual plot, respectively. The attenuation of swell noise, especially the highest amplitudes, is very limited and in addition, a few seismic reflections are attenuated as well before 2 s. Such filter is definitively not a valuable alternative to handle this type of noise.

We now combine the time-frequency median filter with the multiple-input adaptive noise canceller. We think the first method may attenuate the largest swell amplitudes, while the second one should attenuate most of the coherent noise left. Keeping the same parameters for both methods, we first apply the time-frequency median filter to the entire input gather followed by the multiple-input adaptive noise canceller. The output gather and the residual plot are shown respectively in Figure 6.7 (a) and (b). The noise attenuation is not as efficient as in Figure 6.4 (b) where only the multiple-input noise canceller is used, with more leftovers in deep waters for all offsets. In addition, more seismic reflections are attenuated as well. Consequently, there is no benefits to combine the time-frequency median filter with the multiple-input noise canceller for the attenuation of this type of noise, and the multiple-input noise canceller used by itself provides the best solution here.

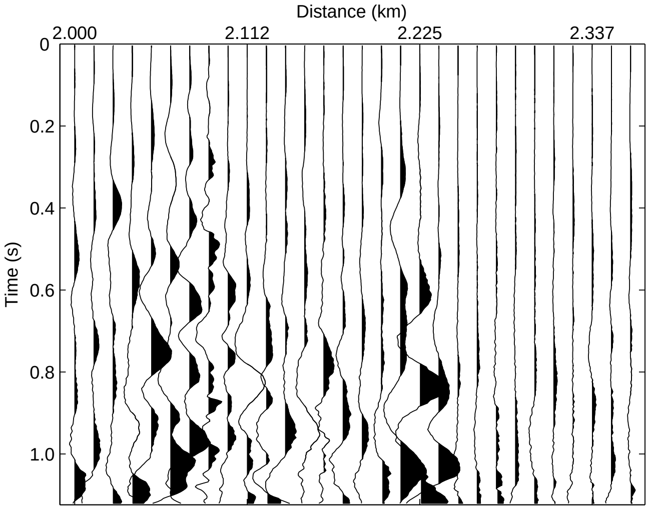
6.3.3 Application to swell noise attenuation: second example

Let us consider another marine shot gather shown in Figure 6.8 (a) made of 96 traces and corrupted by seismic interferences visible after 1 s, and by swell noise, visible around offsets 0.25 km, 0.5 km and beyond 1 km. Contrary to the previous data set, the frequency content of the swell noise does not differ so dramatically from the frequency content of seismic signals. In addition, it is recorded along the entire traces and characterized by smaller amplitudes. We extract a data window prior the first seismic reflection arrival in the offset interval 0.937 – 1.112 km and in the time interval 0 – 0.6 s, and replicate it to form 15 input noise sequences of length equal to the traces length shown in Figure 6.8 (b).

We consider a variable step size NLMS algorithm of order $L = 50$, where data are treated as a single block, i.e., N_B is equal to the trace length. In this example, we use a unique instantaneous frequency threshold value, $\Phi_1 = \Phi_2 = 70$ Hz. The normalized step size values are $\beta_0 = 0.000005$ and $\beta_1 = \beta_2 = 0.001$, while $\epsilon = 0.0001$. In order to optimize the noise attenuation, we apply the variable step size NLMS algorithm three times successively with the same parameters. The output gather after the first application is shown in Figure 6.9 (a) and after the third application in Figure 6.9 (b). The residual plot, i.e., the difference between the input and output gathers after three applications of the filter is shown in Figure 6.9 (c). After three applications, lots of swell noise has been removed, more specifically around 0.25 km, 0.4 – 0.5 km, 0.7 km and beyond 0.9 km, while the seismic reflections have been preserved. We observed however that the output gather still contains low-frequency noise, especially along a few chaotic traces at



(a)

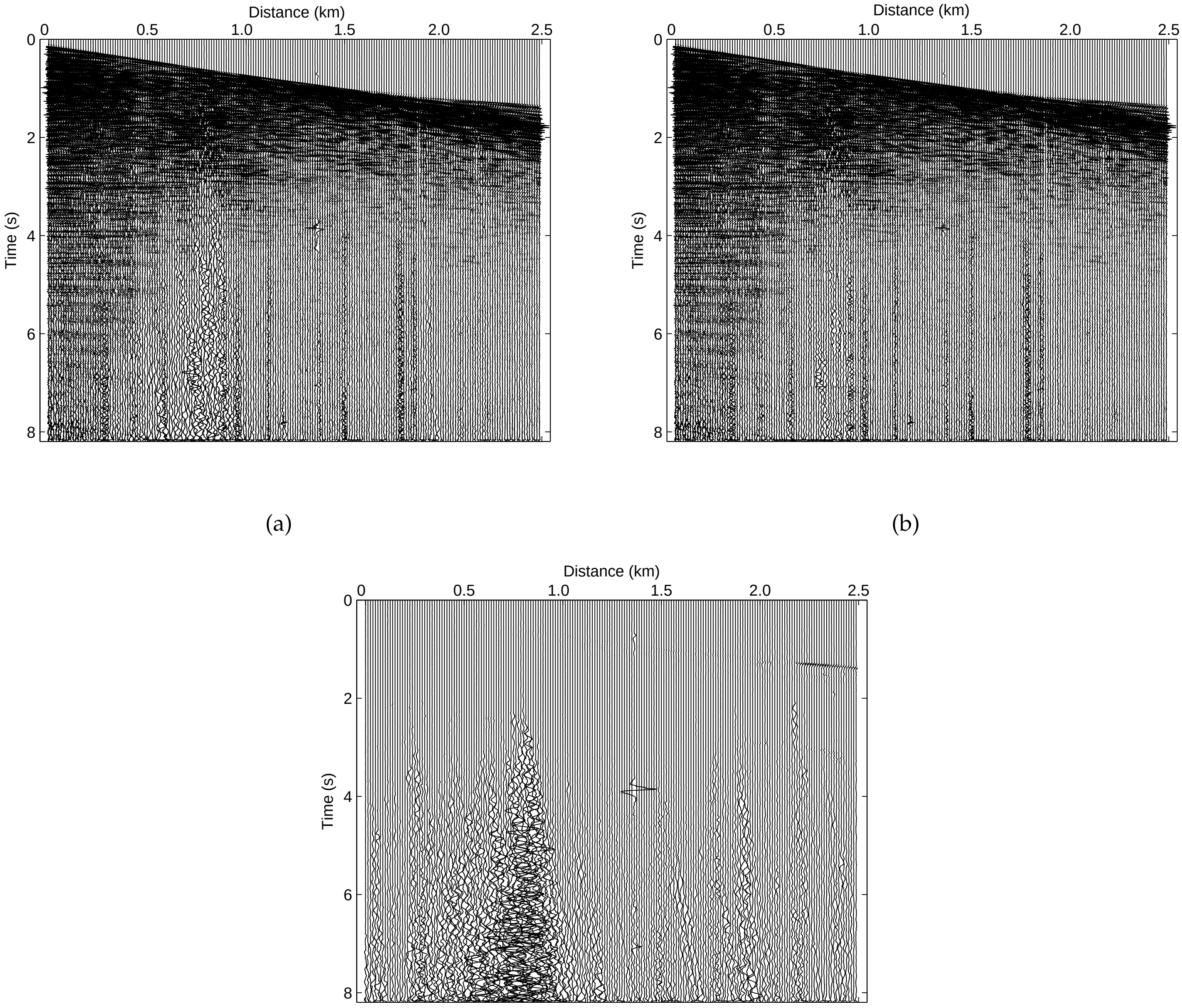


(b)

Figure 6.3: First example: (a) input marine shot gather and (b) 30 extracted input noise sequences.

offsets 0.28 km, 0.4 km, 1.1 km and 1.2 km. Finally, like in the first example, three successive applications of this filter are necessary to attenuate significantly the low-frequency noise. This filter however is not able to attenuate the seismic interferences.

Clearly, distinct swell noise that is localized within an input gather and that is characterized by a high amplitudes, low-frequency content as in the first example, is much more successfully handled by the multiple-input adaptive noise canceller. First, be-



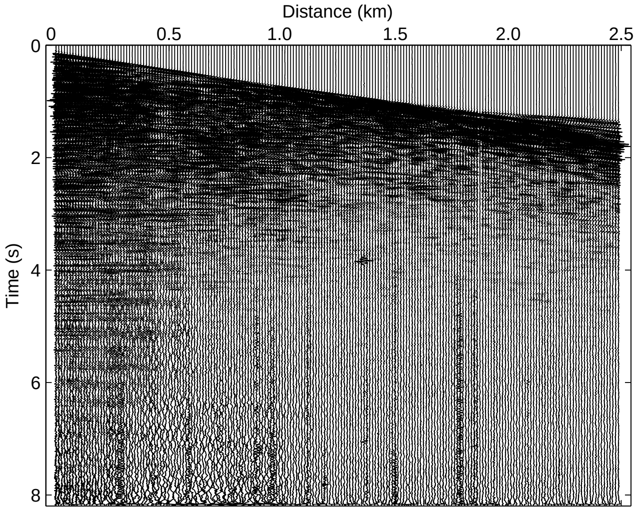
(a)

(b)

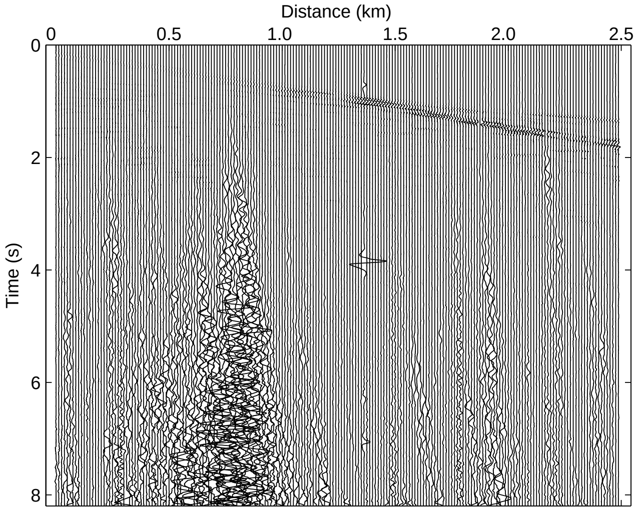
(c)

Figure 6.4: Multiple-input noise canceller: output shot gather after (a) one application of the filter, and (b) three applications of the filter. (c) Residual plot after three applications of the filter.

cause swell noise and seismic reflections have a rather different range of instantaneous frequencies, it is easier for the user to set optimal threshold values. Then, because the noise is clearly uncorrelated with the seismic reflections, its attenuation by the multiple-



(a)

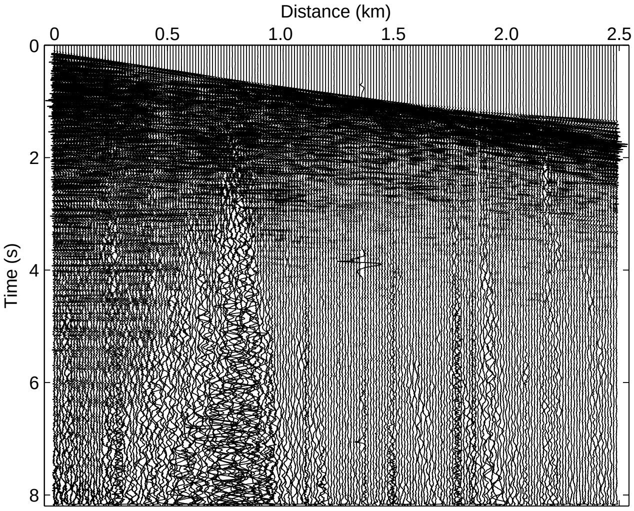


(b)

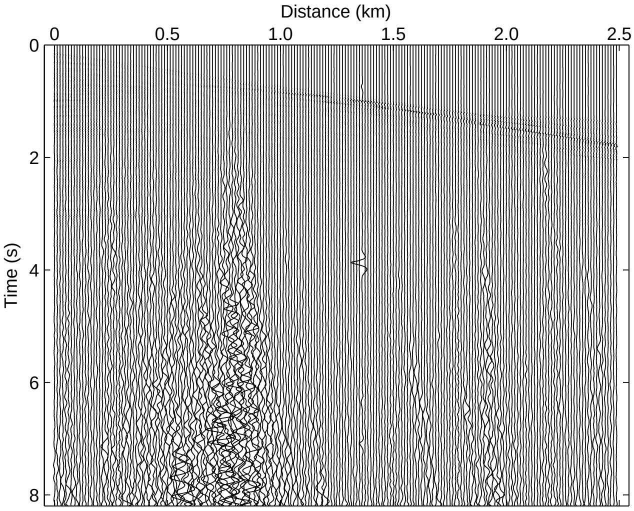
Figure 6.5: Time-frequency median filtering: (a) output shot gather and (b) residual plot.

input adaptive noise canceller is very efficient. On the contrary, when the low-frequency noise is somewhat correlated with seismic reflections, as in the second example, the multiple-input adaptive noise canceller has a limited efficiency.

For comparison, we apply the time-frequency median filter within the frequency range 0 - 12 Hz to the data set shown in Figure 6.8 (a). Successive applications of this filter do not bring further improvement. Figure 6.10 (a and b) shows the output gather and the residual plot, respectively. The highest amplitudes of swell noise are much better



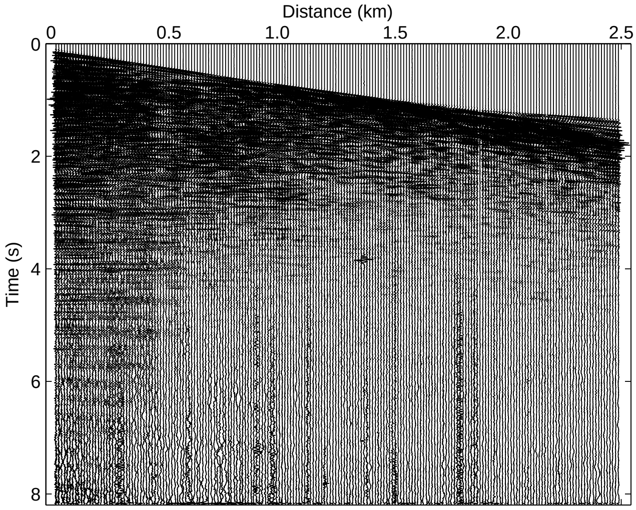
(a)



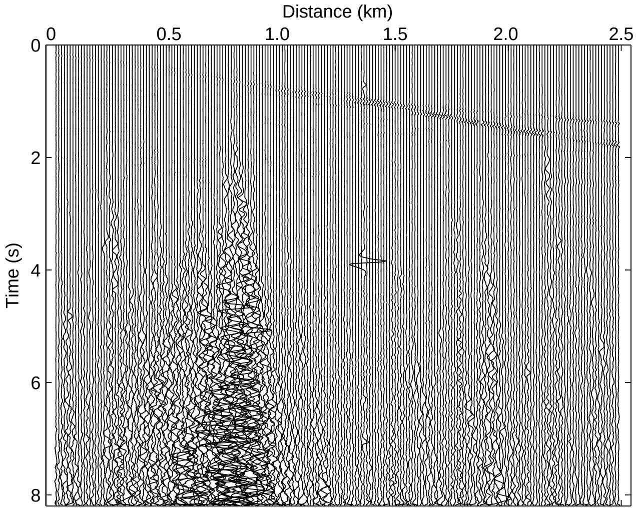
(b)

Figure 6.6: High-pass Butterworth filter: (a) output shot gather and (b) residual plot.

attenuated at for instance distances 0.26 km, 0.40 km and 1.15 km, but the smallest ones are almost left unchanged. This is explained by the fact that the time-frequency median filter compares each frequency amplitude with the ones from the neighbor traces that are also contaminated by noise of similar amplitude spectrum. In such a case, small frequency amplitudes are not considered as abnormal and therefore, they are not attenuated. Since the multiple-input noise canceller better attenuates small swell noise amplitudes, then both methods are potentially complementary. Thus, we investigate



(a)



(b)

Figure 6.7: Time-frequency median filtering + multiple-input noise canceller: (a) output shot gather and (b) residual plot.

now the combination of the time-frequency median filter with the multiple-input adaptive noise canceller. We apply the multiple-input adaptive noise canceller to the output of the time-frequency median filter with the same parameters as in Figure 6.9. The resulting output gather and its residual plot are shown in Figure 6.11. This combination yields significant improvement since small swell noise amplitudes are much better attenuated at distances 0.25 – 0.5 km and beyond 1 km. Therefore, the time-frequency median filter combined with the multiple-input noise canceller yields a real improvement on this data example.

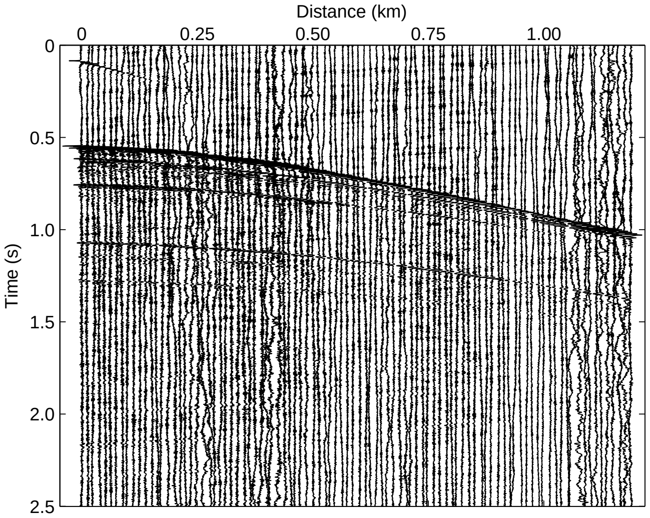
Finally, we apply the second order high-pass Butterworth filter to the second data set to remove the low-frequency content. Again, the choice of a larger cutoff frequency or a larger filter order would remove too much seismic signals, as well as successive applications of this high-pass filter that attenuate certain seismic reflections. Figure 6.12 (a and b) shows the output gather and its residual plot, respectively. The high-pass filter presents the same limitations as the multiple-input adaptive noise canceller concerning the attenuation of high amplitude swell noise. This last method, nevertheless, better attenuates specific traces such as at offsets 0.25 km and 1.2 km, but also attenuates the primary reflection as well. Thus, even if the high-pass Butterworth filter presents the advantage of having a very small computational complexity, it yields some artefact that can be partly resolved by the multiple-input adaptive noise canceller and even better, by the combination of the time-frequency median filter with the multiple-input noise canceller, at the cost however of a larger computational complexity. Finally, we notice that none of the tested methods are able to attenuate the seismic interferences contained in the input data set.

6.3.4 Application to seismic interferences attenuation

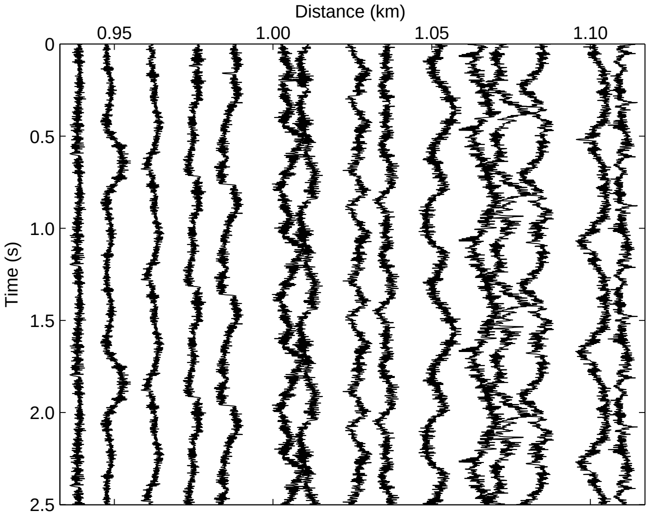
The multiple-input noise canceller, in its present form, can also be used for the attenuation of seismic interferences when applied along the dip of the interferences. It basically consists of "rotating" the input gather so that the interferences are aligned vertically, as a normal trace usually is. We then pick-up a few trace segments containing interferences that will be considered as noise sequences, run the multiple-input noise canceller for all the rotated traces, and finally rotate them back to their original configuration.

Let us now consider the real data set used in the second example shown in Figure 6.8 (a), and apply the multiple-input noise canceller to attenuate the rather strong seismic interferences recorded after 1.5 s, or the smaller ones recorded before 0.5 s. The dip of the interferences is about 2 ms upward per trace. To form the input noise sequences, we extract 9 rotated traces whose original time location is around 2 s at offset 0 km. Frequency thresholds are set to $\Phi_1 = 125$ Hz and $\Phi_2 = 75$ Hz, and the normalized step size values are $\beta_0 = 0.00005$, $\beta_1 = 0.01$ and $\beta_2 = 0.02$. The other parameters are left unchanged. We apply this filter to the output of time-frequency median filter combined with the multiple-input noise canceller in Figure 6.11 in order to evaluate the best possible result that can be obtained.

Figure 6.13 (a) shows the output gather and (b) the residual plot, i.e., the difference between the input gather shown in Figure 6.8 (a) and the output gather. Only some of the strongest interferences located after 1 s have been been attenuated, while a tiny part of the primary reflection has been attenuated as well. By increasing the value of β_2, we experienced that the attenuation of seismic interferences is more significant but at the cost of a more important attenuation of the primary reflection. Note that no other



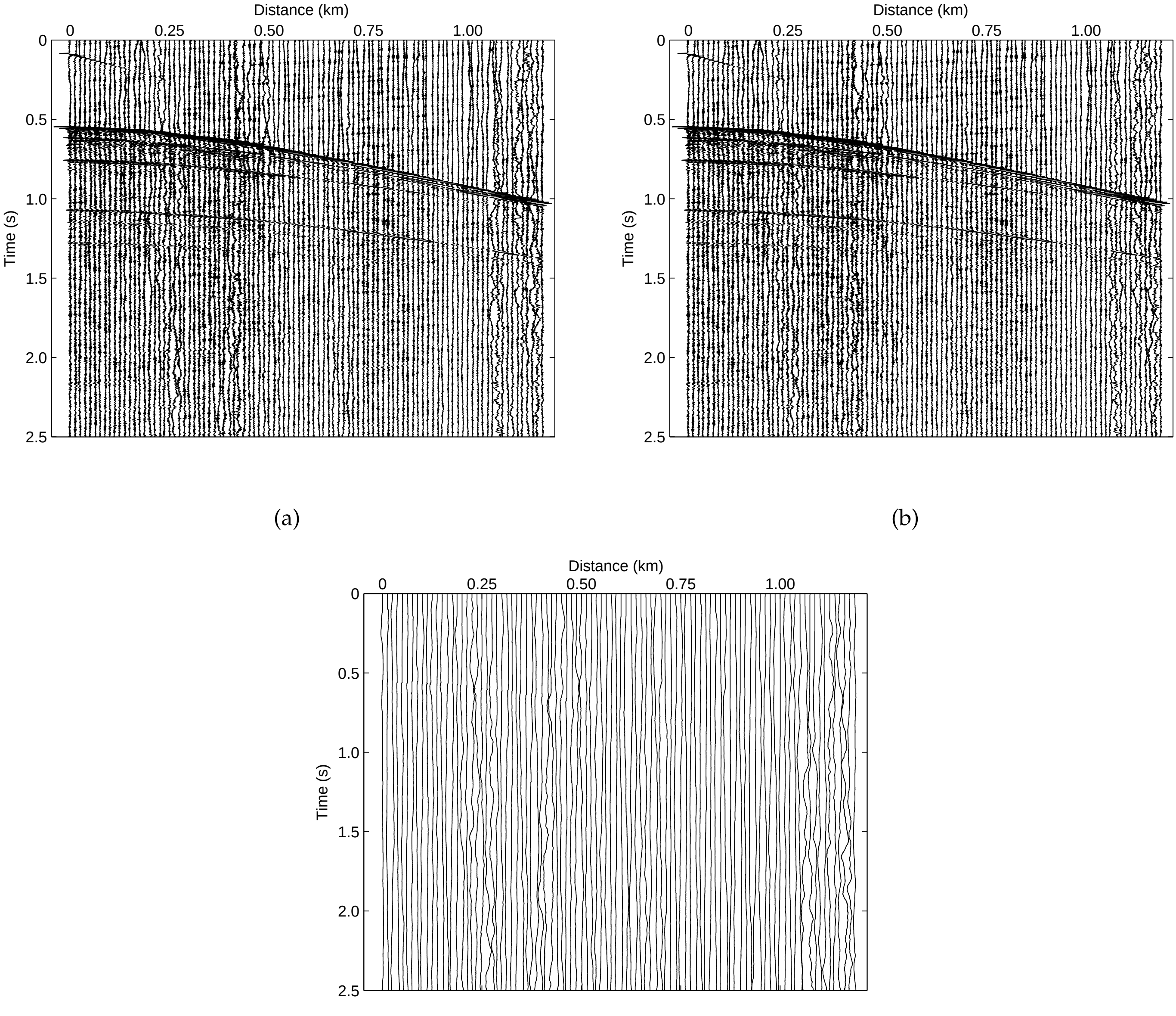
(a)



(b)

Figure 6.8: Second example: (a) input marine shot gather and (b) 15 extracted input noise sequences.

seismic reflections are attenuated. We think however that a possible way to overcome this problem consists of applying the multiple-input noise canceller to a limited section of the data that does not include the primary reflection. The attenuation of seismic interferences is more tricky than the attenuation of swell noise, as the frequency spectra of the seismic interferences and the seismic reflections are relatively similar. Thus, using the instantaneous frequency as a metric that can differentiate the seismic interferences



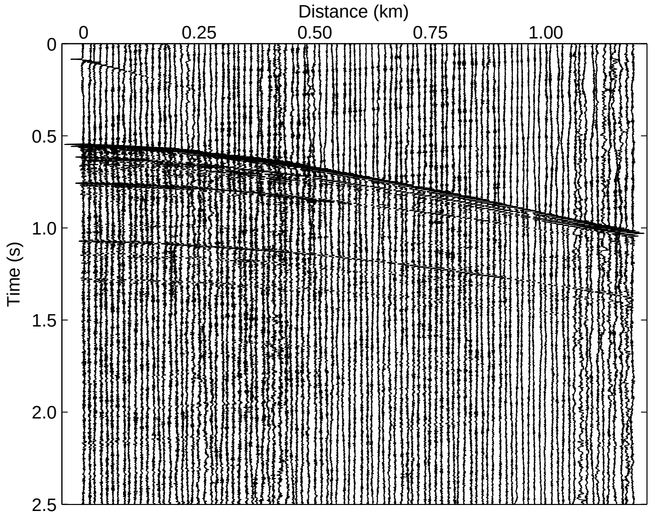
(a)

(b)

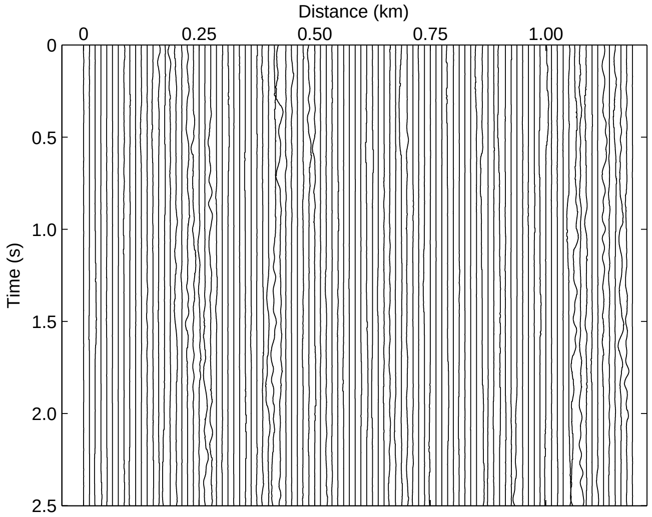
(c)

Figure 6.9: Multiple-input noise canceller: output shot gather after (a) one application of the filter and (b) three applications of the filter. (c) Residual plot after three applications of the filter.

from standard seismic reflections is no longer optimal, and therefore the efficiency is limited.



(a)

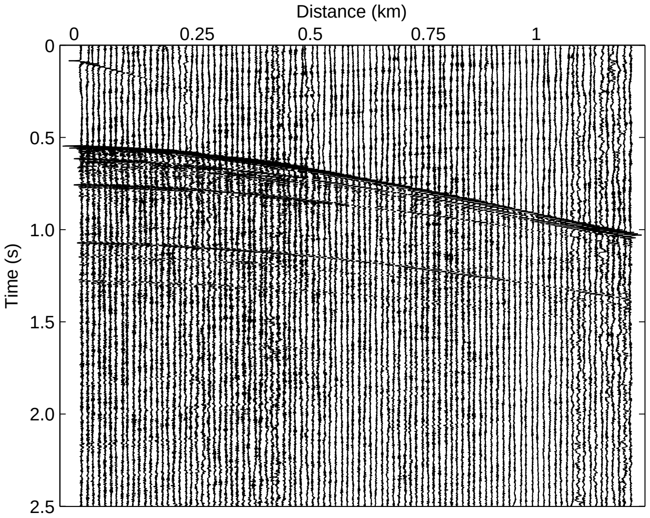


(b)

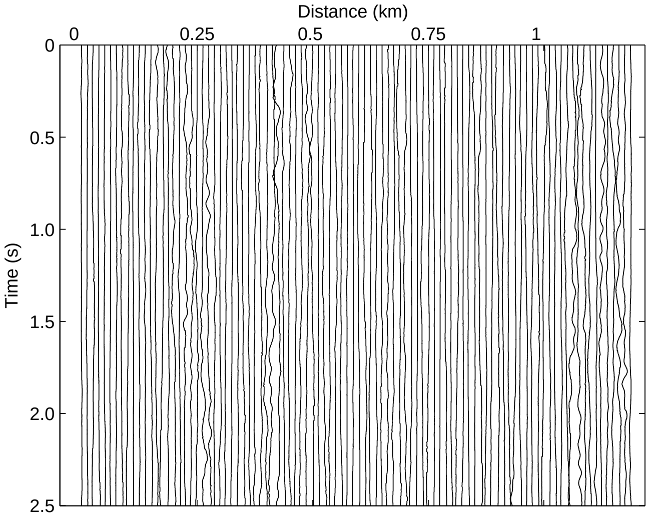
Figure 6.10: Time-frequency median filtering: (a) output shot gather and (b) residual plot.

6.4 Discussion

The user effort to set up the parameters is dependent on the data set to be treated. The key parameters are the instantaneous frequency threshold values, Φ_1 and Φ_2, and the normalized step size values β_0, β_1 and β_2. Here, β_0 is usually low to keep high frequency components unchanged, while β_1 and β_2 are much larger and control the rate



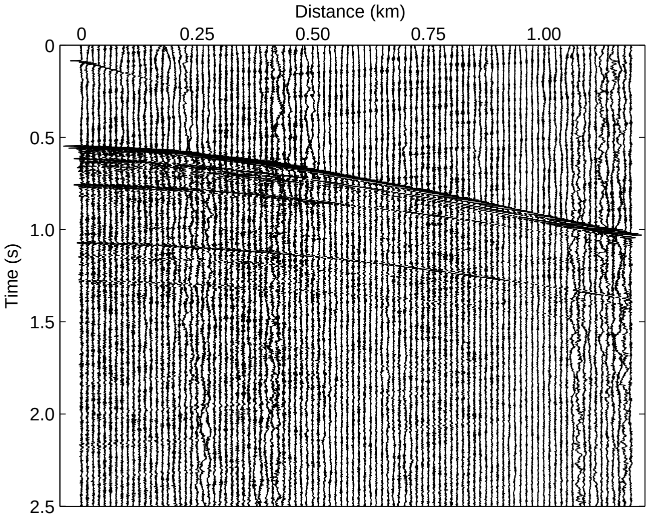
(a)



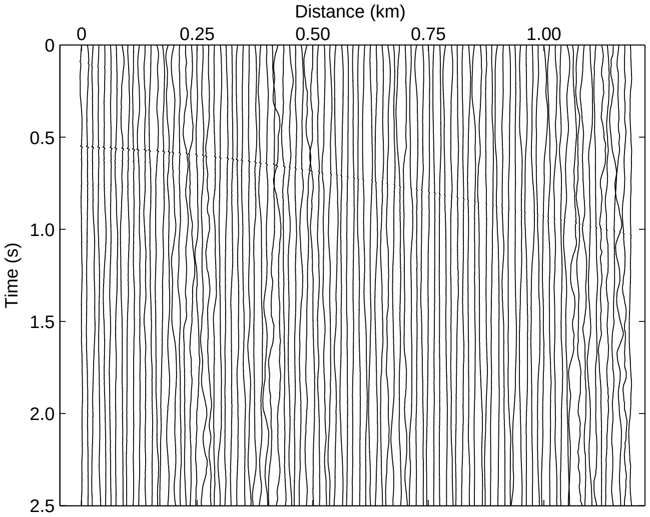
(b)

Figure 6.11: Time-frequency median filtering + multiple-input noise canceller: (a) output shot gather and (b) residual plot.

of convergence, i.e., the level of signal attenuation. In the case where the swell noise is localized within the input gather and characterized by high amplitudes low-frequency signals that are uncorrelated with the seismic reflections, as in the first data example, the instantaneous frequencies of the noise and the seismic reflections differ significantly. Therefore, the threshold values Φ_1 and Φ_2 are easy to find, typically corresponding to a high percentage of all instantaneous frequency values. In the case where the swell



(a)

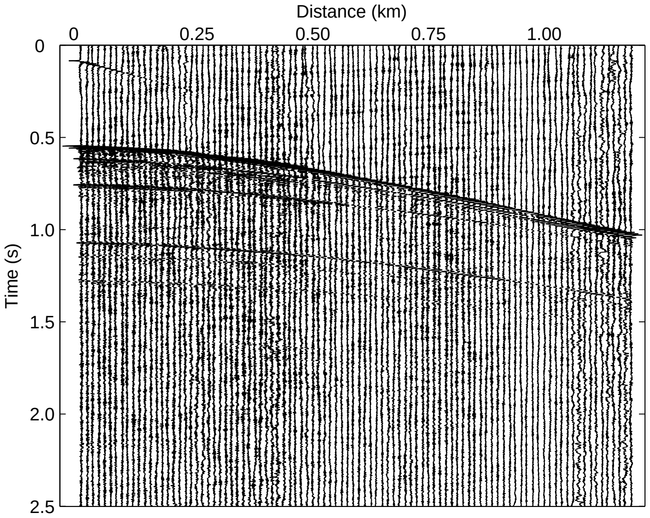


(b)

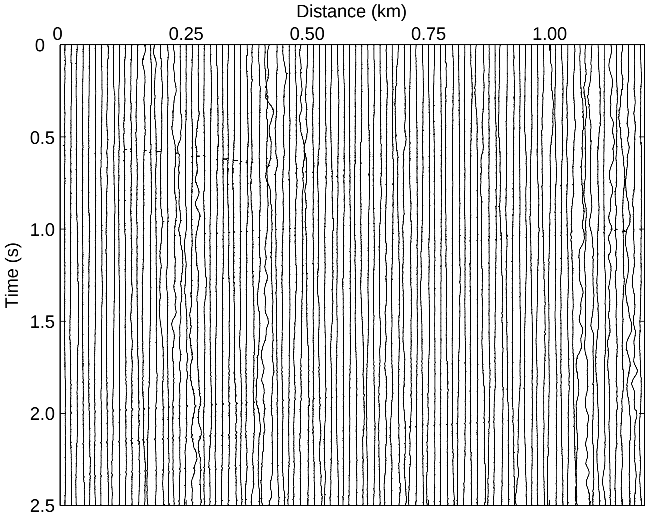
Figure 6.12: High-pass Butterworth filter: (a) output shot gather and (b) residual plot.

noise is somewhat correlated with seismic reflections, as in the second data example, the determination of the threshold value parameters is more tricky and requires a few trial-and-errors to find optimal values.

The computational cost of the multiple-input adaptive noise canceller is satisfactory. This method is definitively more complex than a standard second order high-pass filter, but it requires less operations than the time-frequency median filter, and it is therefore, less computationally demanding. The main reason is that the multiple-input adaptive



(a)



(b)

Figure 6.13: Multiple-input noise canceller application for seismic interferences attenuation: (a) output shot gather and (b) residual plot.

noise canceller treats all the traces independently and does not perform any trace-to-trace comparison as the time-frequency median filter does.

6.5 Summary

We have proposed a method to remove coherent noise that utilizes the filter structure of the multiple-input adaptive noise canceller, and an adaptive normalized least mean squares algorithm with variable step size to resolve the tap-weights coefficients. This method has been compared to the time-frequency median filter and a second-order high-pass Butterworth filter.

We showed that the use a variable step size is a necessary condition for this method to adapt to the changing statistics of the data and therefore, successfully attenuate coherent noise. By choosing the appropriate set of normalized step size values, the multiple-inputs adaptive noise canceller achieves a very good attenuation of coherent noise while preserving the seismic reflections in the case where the swell noise is localized within the input gather and characterized by high amplitudes low-frequency signals that are uncorrelated with the seismic reflections. In such a case, we showed that this method was more efficient than the time-frequency median filter by its ability to attenuate coherent noise even when the neighbor traces are corrupted by similar noise. However, we found that when the low-frequency noise is somewhat correlated with the seismic reflections, the multiple-input adaptive noise canceller affects mostly the small noise amplitudes whereas the time-frequency median filtering was better at attenuating the highest noise amplitudes. We showed then that the combination of the time-frequency median filter with the multiple-input adaptive noise canceller yielded significant improvements. Thus, the multiple-input adaptive noise canceller is a powerful and efficient filter to attenuate swell noise, that can be used either by itself or in combination with the time-frequency median filter, depending of the noise configuration. In addition, we demonstrated in both cases that the high-pass Butterworth filter had limited effects on coherent noise attenuation. Finally, we showed that the multiple-input adaptive noise canceller could find some applications for seismic interferences attenuation, after manipulation of the input data set. However, because the frequency spectra of the seismic interferences and the seismic reflections are relatively similar, the efficiency of this filter is limited.

Finally, this method is relatively easy to implement, it has an acceptable computational cost, and the user effort to set up the parametrization is reasonable. However, one requirement for this method is that the reference noise is available and well correlated with the noise contained in the primary channel. As future work, we may consider its application in the frequency domain.

Chapter 7

Conclusions and suggestions for future work

First, the statistical modeling and analysis of seismic background noise have been investigated. They have revealed that the seismic noise is almost symmetric and slightly platykurtic, and that it yields a composite power-law spectrum. Furthermore, we have proposed a clustering procedure of the noise level as an original quality estimation tool for geophysicists. Thus, the cluster matrix emphasizes defective channels as well as the inhomogeneity of seismic noise along a streamer, and in particular a higher noise level at the head of the streamer and nearby the units hanged to the streamer.

In Chapter 3, we focused on single-azimuth stacking methodologies. We have proposed a novel enhanced stacking method that utilizes local correlation between each individual trace and a chosen reference trace, as a measure of weight, and a new weight normalization scheme. Three different reference traces have been proposed based on (i) conventional stacking, (ii) weighed-stacking using S/N-estimation weights, and (iii) Kalman filter. The use of a Kalman filter is an innovative approach for this application as well as the weight normalization scheme in the enhanced local correlation stacking method. We showed that the use of S/N-estimation reference trace and even more significantly, the use of Kalman reference trace yield consistently superior results compared to conventional stacking. In particular, the superior reference traces exhibit cleaner and better defined reflected events, as well as the ability to see a larger number of reflections in deep waters. We documented that the use of an S/N-estimation reference trace has the ability to enhance coherent signals recorded by short offset traces, while a Kalman reference trace has the ability to emphasize any coherent signal recorded by any traces. For this reason, misalignment does not degrade S/N-estimation reference trace as much as Kalman reference trace. Furthermore, we showed that the enhanced stacking method yields superior results compared to conventional stacking. The best combination is obtained when using S/N estimation as reference method. It exhibits more reflected events, and in our application on sub-salt structure, it reveals a few continuous reflected events under the sub-salt structure. However, we showed that enhanced stacking emphasizes the strongest reflections while discriminating weaker ones, which deteriorates the longitudinal homogeneity of the seismic image. Thus, the use of Kalman filter reference method should be preferred to enhanced stacking using S/N estimation as reference method, since it achieves a better overall seismic image contrast and reveals many more, cleaner and better defined reflected events. However, this result is obtained at the cost of a higher noise level and a larger processing time. Therefore, choosing between one of these two methods is a trade-off between more revealed information and cleaner seismic image. As future work, we suggest to explore the use of mutual information as a generalized statistical measure of signal correlation. Another possible improvement of the presented methods consists of deriving S/N estimates locally instead of globally to improve the S/N-estimation method. Equivalently, deriving the trace amplitudes locally instead of globally should improve the Kalman filter method.

In Chapter 5, we presented an innovative application of sparse code shrinkage and independent component analysis for signal to noise ratio enhancement of seismic signals. More specifically, we proposed to use the enhanced stacking method presented in Chapter 3 applied locally to obtain the "noise-free" realization of the data that allows the derivation of data-driven shrinkages. We showed that using data-fitted shrinkages is a necessary condition for sparse code shrinkage to be a valuable alternative as signal enhancement method. In addition, we investigated several density models, either non-parametric or parametric. Thus, we have derived the shrinkage theory for the non-parametric Gaussian kernel density estimate and for the parametric models using the sparse, the normal inverse Gaussian and the mixture of generalized Gaussian densities. We showed that sparse code shrinkage is an efficient method that targets the noise components to shrink them while preserving the signals. By its ability to enhance signals in both shallow and deep waters, the normal inverse Gaussian density appears to be the best choice of density. However, the sparse density, the mixture of generalized Gaussian density and non-parametric Gaussian kernel estimate have proven to be acceptable choices as well. The flexible parametrization of the normal inverse Gaussian density allows it to model a larger number of unimodal densities, and therefore, it suits seismic data very well. In addition, the level of attenuation of the sparse coded data can be easily adjusted by the user through the noise variance parameter. Finally, the comparison of sparse code shrinkage with local SVD and *f-x* deconvolution has shown that sparse code shrinkage was an efficient method for removing background noise and estimating signals in both shallow and deep waters. However, this is achieved at the cost of a higher computational load, mostly because of the determination of the ICA transformation matrix. As future work, we suggest the investigation of more elaborate "noise-free" realizations, as the signal reconstruction achieved by sparse code shrinkage may be enhanced by improving the quality of signals contained in this realization. In addition, some other super-Gaussian density models can be investigated, and in particular, the generalized Laplacian density and the mixture of two normal inverse Gaussian might good candidates.

Finally, we presented in Chapter 6 an application of the multiple-input adaptive noise canceller to the attenuation of flow-generated nonstationary coherent noise and seismic interferences. This filter uses a normalized least mean squares algorithm with a variable normalized step size. We showed that the use a variable step size is a necessary condition for this method to adapt to the changing statistics of the data, and therefore successfully attenuates coherent noise. The determination of the variable step size parameter is not however an easy task and classical methods fail because they cannot adapt to the characteristics of the seismic noise. In this chapter, we proposed to derive this parameter as a function of instantaneous frequency for each seismic trace, as it is a metric that differentiates relatively well the seismic reflections from the coherent low-frequency noise. The comparison with a time-frequency median filter and a second-order high-pass Butterworth filter revealed that the multiple-input adaptive noise canceller achieves a very good attenuation of coherent noise while preserving the seismic reflections in the case of localized, high amplitudes swell noise that is uncorrelated with the seismic reflections. This method is indeed more efficient than the time-frequency median filter by its ability to attenuate coherent noise even when the neighbor traces are corrupted by similar noise. However, we found that when the low-frequency noise is somewhat correlated with the seismic reflections, the multiple-input adaptive noise canceller affects mostly the small noise amplitudes whereas the time-frequency median filtering is better at attenuating the highest noise amplitudes. We showed then that the combination of the time-frequency median filter with the multiple-input adaptive noise canceller yields significant improvements. Thus, the multiple-input adaptive noise canceller is a powerful and efficient filter to attenuate swell noise, that can be used either by itself or in combination with the time-frequency median filter, depending of the noise configuration. In addition, we demonstrated in both cases that the high-pass Butterworth filter has limited effects on coherent noise attenuation. Finally, we showed that the multiple-input adaptive noise canceller could find some applications for seismic interferences attenuation, after manipulation of the input data set. However, because the frequency spectra of the seismic interferences and the seismic reflections are relatively similar, the efficiency of the filter remains limited. As future work, we may consider the derivation of such filter in the frequency domain.

Appendix

Proof of equation 5.8: $f(x|s) = f_\nu(x - s)$

Consider the model

$$x = s + \nu$$

where s and ν are two independent random variables. The change of variables (s, x) to (u, v) is defined by the linear transform G such that

$$\begin{cases} u = x - s \\ v = s, \end{cases}$$

which is equivalent to

$$\begin{cases} x = u + v \\ s = v. \end{cases}$$

The probability density of the new variables, $f_{u,v}$, is given by

$$f_{u,v}(u, v) = f_{s,x}(G^{-1}(u, v))|J^{-1}|$$

where J^{-1} is the inverse of the Jacobian matrix given by

$$\begin{pmatrix} \frac{\partial s}{\partial u} & \frac{\partial s}{\partial v} \\ \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}.$$

Thus, $|J^{-1}| = 1$ and $f_{u,v}(u, v) = f_{s,x}(G^{-1}(u, v)) = f_{s,x}(v, u + v)$.
Using again the variables (s, x) instead of (u, v) in the latest equation, we obtain

$$f_{s,\nu}(s, \nu) = f_{s,x}(s, x).$$

Moreover, s and ν are independent, which yields

$$f_{s,\nu}(s, \nu) = f_s(s)f_\nu(\nu) = f_{s,x}(s, x).$$

Bayes rule for densities gives

$$f_{s,x}(x|s) = \frac{f_{s,x}(s, x)}{f_s(s)} = \frac{f_s(s)f_\nu(\nu)}{f_s(s)} = f_\nu(\nu) = f_\nu(x - s).$$

Finally, by dropping the subscripts s and x, we obtain

$$f(x|s) = f_\nu(x - s).$$

Bibliography

T. W. Anderson. *Asymptotic theory for principal component analysis*. *Annals of Mathematical Statistics*, **34**: 122–148, 1963.

H. Andrews and C. Patterson. *Singular value decomposition and digital image processing*. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, **24**: 26–53, 1976.

J. C. Bancroft, H. D. Geiger and G. F. Margrave. *The equivalent offset method of prestack time migration*. *Geophysics*, **63**(6): 2042–2053, 1998.

H. B. Barlow. *What is the computational goal of the neocortex?*. In *C. Koch, and J. L. Davis, eds., Large-Scale Neuronal Theories of the Brain*, pp. 1–22. MIT Press, Cambridge, MA, 1994.

J. W. Bayless and E. O. Brigham. *Application of the* Kalman *filter to continuous signal restoration*. *Geophysics*, **35**(1): 2–23, 1970.

J. B. Bednar. *Applications of median filtering to deconvolution, pulse estimation, and statistical editing of seismic data*. *Geophysics*, **48**(12): 1598–1610, 1983.

M. Bekara and M. Van der Baan. *Local singular value decomposition for signal enhancement of seismic data*. *Geophysics*, **72**: V59–V65, 2007.

M. Bekara, A. Ferreira and M. Van der Baan. *A statistical technique for high amplitude noise detection: Application to swell noise attenuation*. *SEG Technical Program Expanded Abstracts*, **27**(1): 2601–2605, 2008.

G. Bianchi and R. Sorrentino. *Electronic Filter Simulation and Design*. McGraw-Hill Professional, 2007.

B. Boashash. *Estimating and interpreting the instantaneous frequency of a signal - part 1: Fundamentals*. *Proceedings of the IEEE*, **80**(4): 520–538, 1992.

P. Brockett, M. Hinich and G. Wilson. *Nonlinear and non-*Gaussian *ocean noise*. *Journal of the Acoustical Society of America*, **82**: 1386–1394, 1987.

T. Bronez. *On the performance advantage of multitaper spectral analysis*. *IEEE Trans. Signal Proc.*, **40**(12): 2941–2946, 1992.

L. L. Canales. *Random noise reduction*. *SEG Technical Program Expanded Abstracts*, **3**(1): 525–527, 1984.

C. Chui. *An Introduction to Wavelets*. Academic Press, 1992.

J. Claerbout. *Fundamentals of Geophysical Data Processing*. McGraw-Hill, 1976.

W. S. Cleveland. *Robust locally weighted regression and smoothing scatterplots*. *Journal of the American Statistical Association*, **74**: 829–836, 1979.

L. Cohen. *Time-frequency analysis*. Prentice Hall PTR, Upper Saddle River, N.J, 1995.

P. Comon. *Independent component analysis-a new concept?*. *Signal Processing*, **36**: 287–314, 1994.

N. D. Crump. *A Kalman filter approach to the deconvonlution of seismic signals*. *Geophysics*, **39**(1): 1–13, 1974.

C. D'Agosto, K. Marfurt and J. Steven. *Modeling and removal of ground roll from horizontal component of C-waves*. *SEG Technical Program Expanded Abstracts*, **22**(1): 1981–1983, 2003.

D. L. Donoho, I. M. Johnstone, G. Kerkyacharian and D. Picard. *Wavelet shrinkage: Asymptopia?*. *Journal of the Royal Statistical Society, Series B*, **57**: 301–337, 1995.

B. Efron and C. Morris. *Data analysis using Stein's estimator and its generalizations*. *J. American Statistical Association*, **70**: 311–319, 1975.

T. Elboth, I. V. Presterud and D. Hermansen. *Time-frequency seismic data de-noising*. *Geophysical prospecting*, 2010.

S. L. M. Freire and T. J. Ulrych. *Application of singular value decomposition to vertical seismic profiling*. *Geophysics*, **53**(6): 778–785, 1988.

J. Gerbrands. *On the relationship between* SVD, KLT, *and* PCA. *Pattern Recognition*, **14**(6), 1981.

G. Golub and C. van Loan. *Matrix computations, Third Edition*. The Johns Hopkins University Press, London, 1996.

R. C. Gonzalez and R. E. Woods. *Digital image processing, Third Edition*. Prentice Hall, 2008.

J. C. Gower. *Some distance properties of latent root and vector methods used in multivariate analysis*. *Biometrika*, **53**: 325–338, 1966.

S. Grion and A. Mazzotti. *Stacking weights determination by means of* SVD *and cross-correlation*. *SEG Technical Program Expanded Abstracts*, **17**(1): 1135–1138, 1998.

S. L. Hahn. *Hilbert transforms in signal processing*. Artech House, Inc., Boston, 1996.

J. B. U. Haldorsen and P. A. Farmer. *Suppression of high-energy noise using an alternative stacking procedure*. *Geophysics*, **54**(2): 181–190, 1989.

A. Hanssen and T. A. Øigård. *The normal inverse* Gaussian *distribution: a versatile model for heavy-tailed stochastic processes*. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 6, pp. 3985–3988. Salt Lake City, UT, May 2001a.

A. Hanssen and T. A. Øigård. *The normal inverse* Gaussian *distribution, as a flexible model for heavy tailed processes*. In *Proceedings IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing*. Baltimore, Maryland, USA, June 2001b.

M. H. Hayes. *Statistical digital signal processing and modeling*. Wiley, 1996.

S. Haykin. *Adaptive Filter Theory, 4th Edition*. Prentice Hall, September 2001.

S. Helgason. *The Radon Transform (Progress in Mathematics), Second Edition*. Birkhäuser, 1999.

C. Hemon and D. Mace. *Essai d'une application de la transformation de* Karhunen-Loève *au traitement sismique*. *Geophysical Prospecting*, **26**: 600–626, 1978.

H. Hotelling. *Analysis of a complex of statistical variables into principal components*. *Journal of Educational Psychology*, **24**: 417–441, 498–520, 1933.

A. Hyvärinen and E. Oja. *A fast fixed-point algorithm for independent component analysis*. *Neural Computation*, **9**: 1483–1492, 1997.

A. Hyvärinen, P. Hoyer and E. Oja. *Sparse code shrinkage: Denoising of nongaussian data by maximum likehood estimation*. *Neural Computation*, **11**: 1739–1768, 1999.

J. N. R. Jeffers. *Two case studies in the application of principal component analysis*. *Applied Statistics*, **16**: 225–236, 1967.

R. Jenssen, T. A. Øigård, T. Eltoft and A. Hanssen. *Sparse code shrinkage based on the normal inverse* Gaussian *density model*. In *Proceedings of International Workshop on Independent Component Analysis and Blind Signal Separation*, pp. 212–217. San Diego, 2001.

I. Jolliffe. *Principal* Component *A*nalysis*, Second Edition*. Springer, 2002.

S. T. Kaplan, M. D. Sacchi and T. J. Ulrych. *Sparse coding for data-driven coherent and incoherent noise attenuation*. *SEG Technical Program Expanded Abstracts*, **28**(1): 3327–3331, 2009.

M. Kendall and A. Stuart. *The advanced theory of statistics*. Charles Griffin & Company, 1958.

K. Kreutz-Delgado, B. Rao, K. Engan, T. W. Lee and T. Sejnowski. *Convex/schur-convex (*CSC*) log-priors and sparse coding*. In *6th Joint Symposium on Neural Computation*, pp. 65–71. 1999.

R. Kwong and E. Johnston. *A variable step size* LMS *algorithm*. *IEEE Transactions on Signal Processing*, **40**: 1633–1642, 1992.

G. Liu, S. Fomel, L. Jin and X. Chen. *Stacking seismic data using local correlation*. *Geophysics*, **74**: V43–V48, 2009.

W. Lu. *Adaptive multiple subtraction using independent component analysis*. *Geophysics*, **71**(5): S179–S184, 2006.

Y. Luo, S. Al-Dossary, M. Marhoon and M. Alfaraj. *Generalized* H*ilbert transform and its applications in geophysics*. *The Leading Edge*, **22**(3): 198–202, 2003.

D. Manolakis, V. Ingle and S. Kogon. *Statistical and Adaptive Signal Processing*. Artech House, 2005.

J. Mari, F. Glangeaud and F. Coppens. *Signal processing for geologists and geophysicists*. Technip, Paris, 1999.

Matlab®. The MathWorks, Inc., Boston, MA, R2010a edition, 2010.

W. H. Mayne. *Common reflection point horizontal data stacking techniques*. *Geophysics*, **27**(6): 927–938, 1962.

O. M. M. Mohamed and M. Jaïdane-Saïdane. *On the parameters estimation of the generalized* Gaussian *mixture model*. In *Proceedings of the* 17^{th} *European Signal Processing Conference (EUSIPCO 2009)*. Glasgow, Scotland, August 2009.

R. Neelamani, T. A. Dickens and M. Deffenbaugh. *Stack-and-denoise: A new method to stack seismic datasets*. *SEG Technical Program Expanded Abstracts*, **25**(1): 2827–2831, 2006.

T. A. Øigård, A. Hanssen, R. Hansen and F. Godtliebsen. EM*-estimation and modeling of heavy-tailed processes with the multivariate normal inverse* Gaussian *distribution*. *Signal Processing*, **85**: 1655–1673, 2005.

B. A. Olshausen and D. J. Field. *Natural image statistics and efficient coding*. *Network*, **7**(2): 333–339, 1996.

A. Özbek. *Multichannel adaptive interference cancelling*. *SEG Technical Program Expanded Abstracts*, **19**(1): 2088–2091, 2000.

K. Pearson. *On lines and planes of closest fit to systems of points is space*. *Philosophical Magazine*, **2**: 559–572, 1901.

D. Percival and A. T. Walden. *Spectral Analysis for Physical Applications: Multivariate and Conventional Univariate Techniques*. Cambridge: Cambridge Univ. Press, 1993.

L. A. Pflug, P. Jackson, G. E. Ioup and J. W. Ioup. *Variability in higher order statistics of measured shallow-water shipping noise*. In *Proceedings of the 1997 IEEE Signal Processing Workshop on Higher-Order Statistics (SPW-HOS '97)*, p. 400. IEEE Computer Society, Washington, DC, USA, 1997.

R. C. Pruett. *Long period multiple reflection suppression and enhanced velocity discrimination using a weighted stack*. *SEG Technical Program Expanded Abstracts*, **1**(1): 10–12, 1982.

C. Rao. *The use and interpretation of principal component analysis in applied research*. *Sankhya A*, **26**: 329 –358, 1964.

M. A. Rashed. *Smart stacking: A new CMP stacking technique for seismic data*. *The Leading Edge*, **27**(4): 462–467, 2008.

P. Ready and P. Wintz. *Information extraction,* SNR *improvement, and data compression in multispectral imagery*. *IEEE Transactions on Communications*, **21**(10): 1123 – 1131, October 1973.

E. Rietsch. *Estimation of the signal-to-noise ratio of seismic data with an application to stacking*. *Geophysical Prospecting*, **28**: 531–550, 1980.

J. C. Robinson. *Statistically optimal stacking of seismic data*. *Geophysics*, **35**(3): 436–446, 1970.

H. C. Romesburg. *Cluster Analysis for Researchers*. Lulu press, 2004.

M. D. Sacchi and Signal Analysis and Imaging Group. *Seismic*Lab. http://www-geo.phys.ualberta.ca/saig/SeismicLab, accessed 9 July 2010, 2008.

C. Sanchis and A. Hanssen. *Statistical analysis of noise in towed streamer arrays*. In *Proceedings of the 2008 European Signal processing conference (EUSIPCO 2008)*, pp. 200–205. Lausanne, Switzerland, August 2008.

W. A. Schneider. *Integral formulation for migration in two and three dimensions*. *Geophysics*, **43**(1): 49–76, 1978.

M. Schoenberger and J. F. Mifsud. *Hydrophone streamer noise*. *Geophysics*, **39**(6): 781–793, 1974.

B. W. Silverman. *Density Estimation for Statistics and Data Analysis*. Chapman & Hall, 1986.

D. Slepian. *Prolate spheroidal wave functions,* Fourier *analysis, and uncertainty - part* V*: The discrete case*. *Bell System Technical Journal*, **57**: 1371–1430, 1978.

S. Solbø and T. Eltoft. *Homomorphic wavelet-based statistical de-speckling of* SAR *images. IEEE Transactions on Geoscience & Remote Sensing*, **42**: 711–721, April 2004.

A. Tarantola. *Inverse problem theory*. Elsevier Science Publ. Co., Amsterdam, 1987.

D. J. Thomson. *Spectrum estimation and harmonic analysis*. In *Proc. IEEE*, volume 70 of *Institute of Electrical and Electronics Engineers, Inc. Conference*, pp. 1055–1096. 1982.

S. Trickett. *Maximum-likelihood-estimation stacking. SEG Technical Program Expanded Abstracts*, **26**(1): 2640–2643, 2007.

T. J. Ulrych and M. Sacchi. *Volume 36: Information-based* I*nversion and* P*rocessing with* A*pplications (*H*andbook of* G*eophysical* E*xploration:* S*eismic* E*xploration)*. Elsevier, 2005.

T. J. Ulrych, S. Freire and P. Siston. *Eigenimage processing of seismic sections. SEG Technical Program Expanded Abstracts*, **7**(1): 1261–1265, 1988.

M. Van der Baan. PP/PS W*avefield separation by independent component analysis. Geophysical Journal International*, **166**(1): 339–348, 2006.

V. D. Vrabie, J. I. Mars and J.-L. Lacoume. *Modified singular value decomposition by means of independent component analysis. Signal Processing*, **84**: 645–652, 2004.

M. Wand and M. Jones. *Kernel Smoothing*. Chapman & Hall, 1995.

J. Wang, F. Tilmann, R. S. White, H. Soosalu and P. Bordoni. *Application of multichannel* W*iener filters to the suppression of ambient seismic noise in passive seismic arrays. The Leading Edge*, **27**(2): 232–238, 2008.

T. Watt and J. B. Bednar. *Role of the alpha-trimmed mean in combining and analyzing seismic common-depth-point gathers. SEG Technical Program Expanded Abstracts*, **2**(1): 276–277, 1983.

B. Widrow and S. D. Stearns. *Adaptive signal processing*. Prentice-Hall, Inc., Engelwood Cliffs, NJ, USA, 1985.

S. Zhao, Z. Man, S. Khoo and H. R. Wu. *Variable step-size* LMS *algorithm with a quotient form. Signal Processing*, **89**(1): 67–76, 2009.